Photon-Efficient Computational 3-D and Reflectivity Imaging With Single-Photon Detectors

Citation: Shin, Dongeek, et al. "Photon-Efficient Computational 3-D and Reflectivity Imaging With Single-Photon Detectors." IEEE Transactions on Computational Imaging 1, 2 (June 2015): 112–125 © 2015 Institute of Electrical and Electronics Engineers (IEEE)

As Published: http://dx.doi.org/10.1109/TCI.2015.2453093

Publisher: Institute of Electrical and Electronics Engineers (IEEE)

Persistent URL: http://hdl.handle.net/1721.1/111032

Version: Original manuscript: author's manuscript prior to formal peer review

Photon-Efficient Computational 3D and Reflectivity Imaging with Single-Photon Detectors

Dongeek Shin, Ahmed Kirmani, Vivek K Goyal, and Jeffrey H. Shapiro

***Abstract*—Capturing depth and reflectivity images at low light levels from active illumination of a scene has wide-ranging applications. Conventionally, even with single-photon detectors, hundreds of photon detections are needed at each pixel to mitigate Poisson noise. We develop a robust method for estimating depth and reflectivity using on the order of 1 detected photon per pixel averaged over the scene. Our computational imager combines physically accurate single-photon counting statistics with exploitation of the spatial correlations present in real-world reflectivity and 3D structure. Experiments conducted in the presence of strong background light demonstrate that our computational imager is able to accurately recover scene depth and reflectivity, while traditional maximum-likelihood based imaging methods lead to estimates that are highly noisy. Our framework increases photon efficiency 100-fold over traditional processing and also improves, somewhat, upon first-photon imaging under a total acquisition time constraint in raster-scanned operation. Thus our new imager will be useful for rapid, low-power, and noise-tolerant active optical imaging, and its fixed dwell time will facilitate parallelization through use of a detector array.**

***Index Terms*—3D imaging, computational imaging, convex optimization, first-photon imaging, LIDAR, low light-level imaging, Poisson noise, single-photon detection, time-of-flight imaging.**

I. Introduction

A light detection and ranging (LIDAR) system [1] builds a histogram of photon counts over time. The time delay and amplitude of the photon-count histogram, relative to the transmitted pulse's temporal profile, contain object depth and reflectivity information, respectively, about the illuminated scene. LIDAR signal-acquisition time must be long enough to collect the 10^2 to 10^3 photons per pixel (ppp) needed to generate the finely-binned histogram required for accurate scene 3D and reflectivity images.

In this paper, we expound upon an active optical imaging framework that recovers accurate reflectivity and 3D images simultaneously using on the order of 1 detected ppp averaged over the scene. This framework was introduced in [2] and builds upon an approach initiated in [3]. Like the first-photon imaging (FPI) method of [3], our computational imager avoids the formation of histograms and instead uses probabilistic modeling at the level of individual detected photons.

This material is based upon work supported in part by the US National Science Foundation under Grant No. 1161413, a Samsung Scholarship, and a Microsoft PhD Fellowship.

This work will be presented in part at the IEEE International Conference on Image Processing, Paris, France, October 2014.

D. Shin, A. Kirmani, and J. H. Shapiro are with the Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139 USA.

V. K. Goyal is with the Department of Electrical and Computer Engineering, Boston University, Boston, MA 02215 USA.

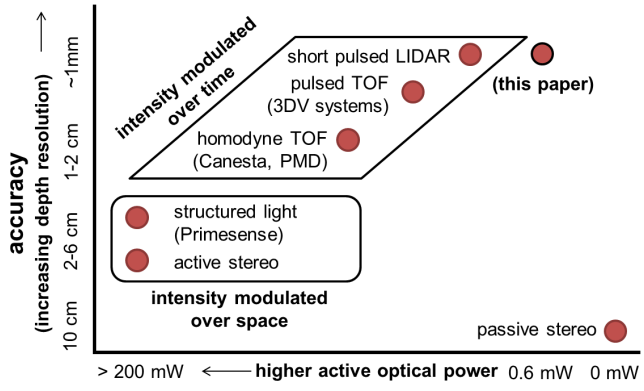
Both methods combine physically accurate probabilistic modeling of single-photon detection with exploitation of the spatial correlations present in real-world scenes to achieve accurate 3D and reflectivity imaging when very little backreflected light reaches the detector, as will be the case with low optical-power active imagers [4]. The method introduced here uses deterministic dwell times, which is both more convenient for raster scanning and amenable to parallelization through the use of a detector array. This ease of applicability comes with somewhat improved performance over FPI when compared at equal total acquisition times in raster-scanned operation.

A. Prior Work

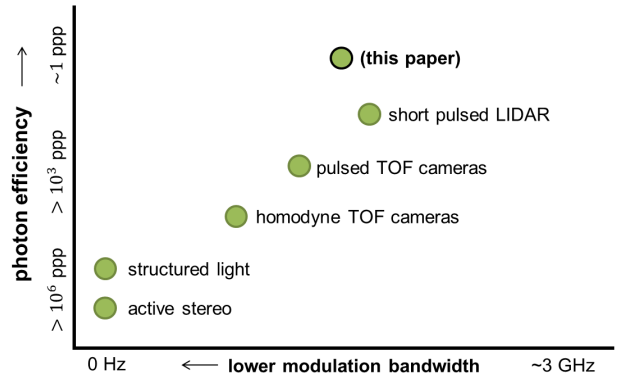
1) Active imaging methods: Active 3D imaging systems differ in how they modulate their transmitted power, leading to a variety of trade-offs in accuracy, modulation frequency, optical power, and photon efficiency; see Figure 1 for a qualitative summary. Temporal modulation enables distance measurement by the time-of-flight (TOF) principle. Examples of TOF acquisition systems, ordered by increasing modulation bandwidth (decreasing pulse duration), include homodyne TOF cameras [5], pulsed TOF cameras [6], and picosecond laser radar systems [7]. Spatial modulation techniques include structured light [8] and active stereo imaging [9]. These spatial-modulation techniques have low photon efficiencies because they use an always-on optical source, whereas pulsed-TOF systems have higher photon efficiencies because they use sources that are on only for short intervals. Additionally, the systems using temporal modulation have better accuracy than those using spatial modulation. The advantage of spatial modulation tends to be cheaper sensing hardware, since high-speed sampling is not required.

The most photon-efficient TOF imagers—those requiring the fewest photons for accurate imaging—use single-photon avalanche diode (SPAD) detectors [10]. Earlier efforts in SPAD-based 3D imaging from on the order of 1 detected ppp are reported in [11]–[13]. The framework presented here improves upon these works in part due to the use of estimated reflectivity. This translates to SPAD-based imagers with lower optical power and lower system bandwidth without sacrificing image quality. There also has been significant recent interest in compressive methods for 3D imaging, with [14]–[16] and without [17] single-photon detection. While compressive methods may reduce some measures of acquisition cost, they do not generally improve photon efficiency.

2) Optoelectronic techniques for low light levels: In low light-level scenarios, a variety of optoelectronic techniques



(a) Accuracy vs. power trade-offs.



(b) Photon efficiency vs. modulation bandwidth trade-offs.

Fig. 1. Qualitative comparison of state-of-the-art active optical 3D sensing technologies. Photon efficiency is defined as photons per pixel (ppp) necessary for centimeter-accurate depth imaging.

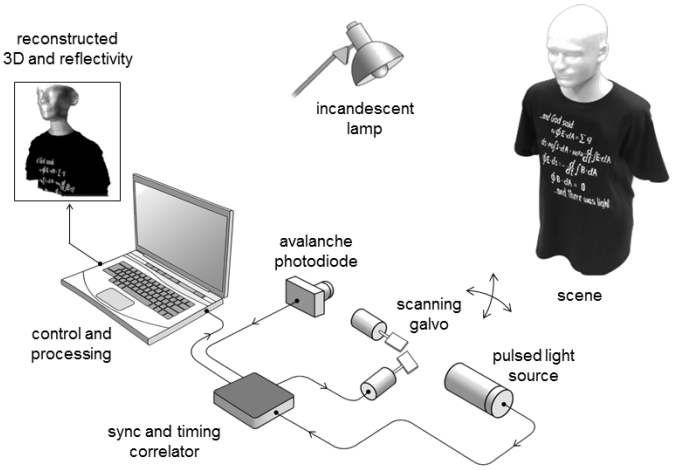


Fig. 2. Experimental imaging setup used with random dwell time in [3] and with constant dwell time here. A pulsed light source illuminates the scene in a raster scan pattern. The backscattered light is collected by a time-resolved single-photon detector. Each spatial location is illuminated with exactly N light pulses (fixed dwell time). An incandescent lamp injects background light which corrupts the information-bearing signal. The photon detection times and the total photon count are recorded at every image pixel. This dataset is used to estimate the 3D structure and reflectivity. The setup is analogous to having a floodlight illumination source and an array of single-photon counting detectors operating at a fixed dwell time.

are employed for robust imaging. Active imagers use lasers with narrow spectral bandwidths and spectral filters to suppress background light and minimize the Poisson noise it creates. However, optical filtering alone cannot completely eliminate background light, and it also causes signal attenuation. Range-gated imaging [18] is another common technique, but this method requires a priori knowledge of object location. Furthermore, a SPAD may be replaced with a superconducting nanowire single-photon detector (SNSPD) [19], which is much faster, has lower timing jitter, and has lower dark-count rate than a SPAD. However, SNSPDs have much smaller active areas and hence have narrower fields of view than SPAD-based systems with the same optics.

3) Image denoising: For depth imaging using SPAD data, it is typical to first find a maximum likelihood (ML) estimate of scene depth using a time-inhomogeneous Poisson process model for photon detection times and then apply a denoising method. The ML estimate is obtained independently at each pixel, and the denoising is able to exploit the scene's spatial correlations. This two-step approach commonly assumes a Gaussian noise model, which is befitting because of the optimal behavior of ML with large numbers of data samples. However, at low light levels, performing denoising well is more challenging due to the signal-dependent nature of Poisson noise. In Section VI, we compare our technique with the state-of-the-art denoising methods that use sparsity-promoting regularization. Our superior performance is due in part to our novel method for classifying detection events as being due to signal (backscattered light) or noise (background light and dark counts).

4) First-photon imaging: First-photon imaging [3] is a method that allows accurate 3D and reflectivity reconstruction using only the first detected photon at every pixel in a raster-scanned scene. FPI combines accurate first-photon detection statistics with the spatial correlations existing in natural scenes to achieve robust low light-level imaging. The raster-scanning process of FPI, however, makes the dwell time at each pixel a random variable. Thus, FPI does not extend naturally to operation using SPAD arrays—since simultaneous measurement implies equal dwell times—thus precluding the dramatic speedup in image acquisition that such arrays enable.

In this paper, we develop models and methods analogous to FPI that apply when there is a fixed dwell time at each pixel. In the experimental configuration depicted in Figure 2, we demonstrate that the performance of the new method is similar to or slightly better than FPI when compared for equal total acquisition time in raster-scanned operation. Furthermore, with an M-fold increase in laser power and an M-element SPAD array, our fixed dwell-time framework can provide this same robust imaging M-times faster than a single-detector raster-scanned system.

B. Main contributions

1) Modeling: We introduce a physically accurate model for the signal produced by a SPAD under low light-level conditions that incorporates an arbitrary illumination pulse

shape, background (ambient) light contribution, dark counts, and the inhomogeneous Poisson process characteristics (shot noise from the quantum nature of light) given a fixed acquisition time.

2) Algorithmic: We provide a method for computational reconstruction of depth and reflectivity from noisy photon-detection data. Our technique combines a shot-noise model for single photon-counting statistics with the high degree of spatial correlation present in real-world scenes.

3) Experimental: We experimentally demonstrate that our proposed 3D imager's photon efficiency is more than 100 times higher than that of traditional ML estimation. We also show that our 3D imager achieves sub-pulse-width depth resolution under short acquisition times, in which 54% of the pixels have missing data (no photon detections), and at high background levels, when any given photon detection has approximately probability 0.5 of originating from ambient light.

C. Outline

The remainder of the paper is organized as follows. Section II introduces the LIDAR-like imaging configuration that we consider. The key probabilistic models for the measured data are derived in Section III. These models are related to conventional image formation in Section IV, and they are the basis for the novel image formation method in Section V. Section VI presents experimental results for the novel method, and Section VII provides additional discussion and conclusions. An appendix presents performance bounds based on our modeling.

The methods detailed in this paper were initially presented in abbreviated form in [2]. The present manuscript provides additional context, details on derivations, performance bounds, and many more experimental results.

II. Imaging Setup

Figure 3 depicts the signal-acquisition model underlying our imager. We aim to form reflectivity image $\boldsymbol{\alpha} \in \mathbb{R}_+^{n\times n}$ and depth image $\mathbf{z} \in \mathbb{R}_+^{n\times n}$ of the scene. We index the scene pixels as (i,j), where $i,j = 1, 2, \ldots, n$. The distance to pixel (i,j) is $z_{i,j} \geq 0$ and its reflectivity, $\alpha_{i,j} \geq 0$, includes the effect of radial fall-off, view angle, and material properties.

A. Illumination

We use a periodically pulsed laser to illuminate the scene in a raster-scanned manner. The repetition period is T_r and the waveform of a single pulse is denoted by $s(t)$. Physically, $s(t)$ is the photon-flux waveform of the pulse emitted at $t = 0$ measured in counts/sec (cps). To avoid distance aliasing, we assume $T_r > 2z_{\max}/c$, where $z_{\max}$ is the maximum scene depth and c is the speed of light. With conventional processing, the root mean-square (RMS) pulse width T_p governs the achievable depth resolution in the absence of background light [20]. As is typically done in depth imaging, we assume that $T_p \ll 2z_{\max}/c$.

B. Detection

A SPAD detector provides time-resolved single-photon detections [10], called *clicks*. Its quantum efficiency η is the fraction of photons passing through the pre-detection optical filter that are detected. Each detected photon is time stamped within a time bin of duration Δ, measuring a few picoseconds, that is much shorter than T_p.

C. Data Acquisition

Each pixel (i,j) is illuminated with N laser pulses. The total acquisition time (dwell time) is thus $T_a = NT_r$. We record the total number of photon detections $k_{i,j}$, along with their detection times $\{t_{i,j}^{(\ell)}\}_{\ell=1}^{k_{i,j}}$, where the latter are measured relative to the immediately-preceding transmitted pulse. We also shine background light, with photon flux b_λ at the operating optical wavelength λ, onto the detector.

III. Probabilistic Measurement Model

Illuminating pixel (i,j) with the pulse $s(t)$ results in backreflected light with photon flux $r_{i,j}(t) = \alpha_{i,j}s(t - 2z_{i,j}/c) + b_\lambda$ at the detector. The measurement of photon flux is through photon detections, and carefully modeling the relationships between the measured quantities and the reflectivity and depth variables is central to our imaging method.

A. Poisson Statistics

The photon detections produced by the SPAD in response to the backreflected light from transmission of $s(t)$ constitute an inhomogeneous Poisson process with time-varying rate function $\eta r_{i,j}(t)$. To these photon detections we must add the detector dark counts, which come from an independent homogeneous Poisson process with rate d. Lumping the dark counts together with the background-generated counts yields the observation process at the SPAD's output, viz., as shown in Figure 3, an inhomogeneous Poisson process with rate function

$$\begin{aligned} \lambda_{i,j}(t) &= \eta r_{i,j}(t) + d \\ &= \eta\alpha_{i,j}s(t - 2z_{i,j}/c) + (\eta b_\lambda + d), \end{aligned} \tag{1}$$

when only a single pulse is transmitted. Figure 3 shows the rate function $\lambda_{i,j}(t)$ for the pulse-stream transmission.

Define $S = \int s(t)\,dt$ and $B = (\eta b_\lambda + d)T_r$ as the total signal and background count per pulse-repetition period, where we have used—and will use in all that follows—background counts to include dark counts as well as counts arising from ambient light. We assume that B is known, because it is straightforward to measure it before we begin data acquisition. The derivations to follow assume $\eta\alpha_{i,j}S + B \ll 1$, meaning that the photon-flux per pixel per pulse-repetition period is much less than 1, as would be the case in low light-level imaging where an imager's photon efficiency is paramount.

B. Distributions of Numbers of Detected Photons

A SPAD detector is not *number-resolving*, meaning that it reports at most one click from detection of a signal pulse. Using Poisson process properties [21], we have that the

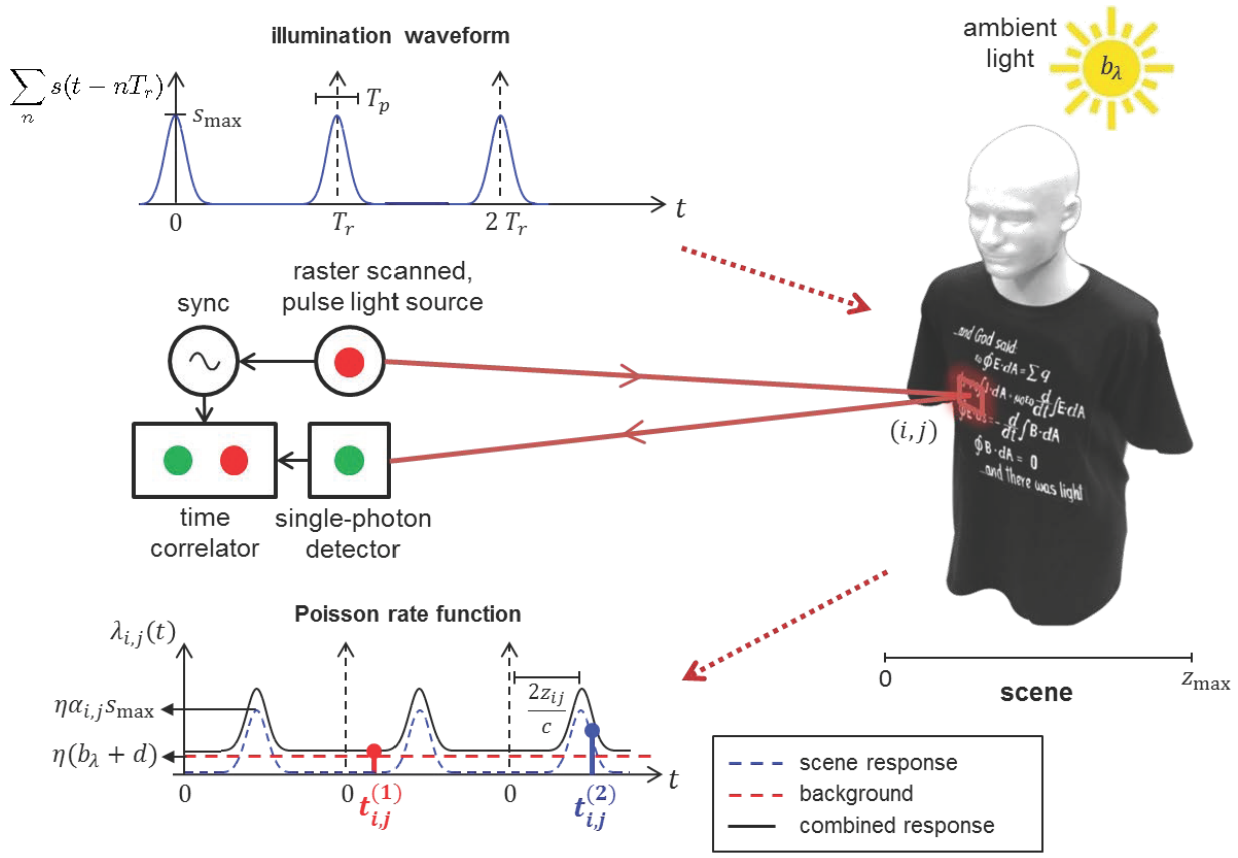


Fig. 3. Summary of observation model. Rate function of inhomogeneous Poisson process combining desired scene response and noise sources is shown. Here, $N = 2$ and $k_{i,j} = 2$. A background count (red) occurred after the second pulse was transmitted, and a signal count (blue) occurred after the third pulse was transmitted.

probability of the SPAD detector *not* recording a detection at pixel (i, j) from one illumination trial is

$$P_0(\alpha_{i,j}) = \exp[-(\eta\alpha_{i,j}S + B)]. \tag{2}$$

Because we illuminate with a total of N pulses, and the low-flux condition ensures that multiple detections per repetition interval can be neglected, the number of detected photons $K_{i,j}$ is binomially distributed with probability mass function

$$\begin{aligned}&\Pr\left[K_{i,j} = k_{i,j}; \alpha_{i,j}\right] \\ &\quad = \binom{N}{k_{i,j}} P_0(\alpha_{i,j})^{N-k_{i,j}} \left[1 - P_0(\alpha_{i,j})\right]^{k_{i,j}},\end{aligned}$$

for $k_{i,j} = 0, 1, \ldots, N$.

In the ultimate low-flux limit in which $\eta\alpha_{i,j}S + B \to 0^+$ with $N \to \infty$ such that $N\{1-\exp[-(\eta\alpha_{i,j}S+B)]\} = C(\alpha_{i,j})$ is held constant, $K_{i,j}$ converges to a Poisson random variable [22] with probability mass function

$$\Pr[K_{i,j} = k_{i,j};\, \alpha_{i,j}] = \frac{C(\alpha_{i,j})^k}{k\,!} \exp[-C(\alpha_{i,j})].$$

C. Distributions of Single-Photon Detection Times

At pixel (i, j), the single-photon detection time $T_{i,j}$ recorded by the SPAD detector is localized to a time bin of duration Δ. Because the SPAD detector only provides timing information for the first (and, in the low-flux regime, *only*) detected photon in a single pulse-repetition interval, the probability of a SPAD click in $[t_{i,j}, t_{i,j} + \Delta)$, given there was a click in that repetition interval, is

$$\begin{aligned}&\Pr[\text{no click in } [0, t_{i,j}),\ \text{click in } [t_{i,j}, t_{i,j}+\Delta) \,|\, \text{click in } [0, T_r)] \\ &\overset{(a)}{=} \frac{\Pr[\text{no click in } [0, t_{i,j})]\,\Pr[\text{click in } [t_{i,j}, t_{i,j}+\Delta)]}{\Pr[\text{click in } [0, T_r)]} \\ &\overset{(b)}{=} \frac{1}{1 - \exp[-(\eta\alpha_{i,j}S + B)]} \times \\ &\quad \left\{\exp\left[-\int_0^{t_{i,j}} \left(\eta\alpha_{i,j}s\left(\tau - \frac{2z_{i,j}}{c}\right) + \frac{B}{T_r}\right) d\tau\right]\right. \\ &\quad \left. - \exp\left[-\int_0^{t_{i,j}+\Delta} \left(\eta\alpha_{i,j}s\left(\tau - \frac{2z_{i,j}}{c}\right) + \frac{B}{T_r}\right) d\tau\right]\right\},\end{aligned}$$

where (a) uses the independent increments property of the Poisson process and (b) uses Equation (2). The probability density function of $T_{i,j} \in [0, T_r)$, the continuous time-of-detection random variable, is then obtained by evaluating the

preceding probability on a per unit time basis as $\Delta \to 0^+$:

$$
\begin{aligned}
& f_{T_{i,j}}(t_{i,j};\, \alpha_{i,j}, z_{i,j}) \\
&= \frac{1}{1-\exp[-(\eta\alpha_{i,j}S+B)]} \times \\
&\quad \lim_{\Delta\to 0^+} \frac{1}{\Delta}\left\{\exp\left[-\int_0^{t_{i,j}}\left(\eta\alpha_{i,j}s\left(\tau-\frac{2z_{i,j}}{c}\right)+\frac{B}{T_r}\right)d\tau\right]\right. \\
&\quad \left. -\exp\left[-\int_0^{t_{i,j}+\Delta}\left(\eta\alpha_{i,j}s\left(\tau-\frac{2z_{i,j}}{c}\right)+\frac{B}{T_r}\right)d\tau\right]\right\} \\
&= \frac{\eta\alpha_{i,j}s(t_{i,j}-2z_{i,j}/c)+B/T_r}{1-\exp[-(\eta\alpha_{i,j}S+B)]} \\
&\quad \times \exp\left[-\int_0^{t_{i,j}}\left(\eta\alpha_{i,j}s(\tau-2z_{i,j}/c)+\frac{B}{T_r}\right)d\tau\right] \\
&\overset{(a)}{=} \frac{\eta\alpha_{i,j}s(t_{i,j}-2z_{i,j}/c)+B/T_r}{\int_0^{T_r}[\eta\alpha_{i,j}s(t_{i,j}-2z_{i,j}/c)+B/T_r]\,dt} \\
&= \frac{\eta\alpha_{i,j}S}{\eta\alpha_{i,j}S+B}\left(\frac{s(t_{i,j}-2z_{i,j}/c)}{S}\right)+\frac{B}{\eta\alpha_{i,j}S+B}\left(\frac{1}{T_r}\right), \qquad (3)
\end{aligned}
$$

where (a) follows from $\eta\alpha_{i,j}S + B \ll 1$.

A detection could be a signal count or a background count The detection statistics result from the merging of the Poisson processes corresponding to these sources. Under our low-flux assumption, the detection time for a signal count from a single pulse-repetition interval is characterized by the normalized time-shifted pulse shape. On the other hand, the detection time for a background count in that interval is uniformly distributed on $[0, T_r)$. Thus the probability density function in Equation (3) is a mixture distribution, with mixture weights

$$
\begin{aligned}
\Pr[\,\text{Detection at } (i,j) \text{ is signal}\,] &= \frac{\eta\alpha_{i,j}S}{\eta\alpha_{i,j}S+B}, \\
\Pr[\,\text{Detection at } (i,j) \text{ is noise}\,] &= \frac{B}{\eta\alpha_{i,j}S+B}.
\end{aligned}
$$

IV. Conventional Image Formation

A. Pixelwise ML Reflectivity Estimation

Given the total observed photon count $k_{i,j}$ at pixel (i,j), the constrained ML (CML) reflectivity estimate is

$$
\begin{aligned}
\hat{\alpha}_{i,j}^{\text{CML}} &= \arg\max_{\alpha_{i,j}\geq 0} \Pr[\,K_{i,j}=k_{i,j};\alpha_{i,j}\,] \\
&= \max\left\{\frac{1}{\eta S}\left[\log\left(\frac{N}{N-k_{i,j}}\right)-B\right],\, 0\right\}.
\end{aligned}
$$

where $\log$ is the natural logarithm. Traditionally, the normalized photon-count value is used as the reflectivity estimate [23],

$$
\tilde{\alpha}_{i,j} = \frac{k_{i,j}}{N\eta S}. \qquad (4)
$$

Note that the normalized count value estimate is equal to the CML estimate under the Poisson approximation to the binomial distribution when $B = 0$.

B. Pixelwise ML Depth Estimation

Using the photon detection-time dataset $\{t_{i,j}^{(\ell)}\}_{\ell=1}^{k_{i,j}}$, the pixelwise constrained ML depth estimate is

$$
\begin{aligned}
\hat{z}_{i,j}^{\text{CML}} &= \arg\max_{z_{i,j}\in[0,cT_r/2)} \prod_{\ell=1}^{k_{i,j}} f_{T_{i,j}}(t_{i,j}^{(\ell)};\, \alpha_{i,j}, z_{i,j}) \\
&= \arg\max_{z_{i,j}\in[0,cT_r/2)} \sum_{\ell=1}^{k_{i,j}} \log\left[\eta\alpha_{i,j}s\left(t_{i,j}^{(\ell)}-\frac{2z_{i,j}}{c}\right)+\frac{B}{T_r}\right],
\end{aligned}
$$

assuming that $k_{i,j} \geq 1$. If $B > 0$, then the ML depth estimate is obtained by solving a non-convex optimization problem. Moreover, ML estimation when $B > 0$ requires the knowledge of the true reflectivity $\alpha_{i,j}$, which is not typically available. Thus, the log-matched filter [21] is instead traditionally used for estimating depth from $k_{i,j}$ photon detections:

$$
\tilde{z}_{i,j} = \arg\max_{z_{i,j}\in[0,cT_r/2)} \sum_{\ell=1}^{k_{i,j}} \log\left[s\left(t_{i,j}^{(\ell)}-2z_{i,j}/c\right)\right]. \qquad (5)
$$

The log-matched filter solution is equal to the CML estimate when $B = 0$.

V. Novel Image Formation

In the limit of large sample size or high signal-to-noise ratio (SNR), the ML estimate converges to the true parameter value [24]. However, when the data is limited or SNR is low—such as in our problem—pixelwise ML solutions yield inaccurate estimates. We compare our 3D imaging method with the baseline normalized-count reflectivity estimate $\tilde{\alpha}_{i,j}$ and the log-matched filter depth estimate $\tilde{z}_{i,j}$, which are ML estimates asymptotically. Along with using the single-photon detection statistics, we exploit the spatial correlations present in real-world scenes by regularizing the ML estimators. Our approach provides significant improvements over pixelwise ML estimators as well as traditional denoising techniques that may exploit scene sparsity but assume additive Gaussian noise. Our computational image formation proceeds in three steps.

Step 1: Reflectivity estimation. The negative log-likelihood of scene reflectivity $\alpha_{i,j}$ given count data $k_{i,j}$ is

$$
\begin{aligned}
&\mathcal{L}_\alpha(\alpha_{i,j}; k_{i,j}) = \\
&(N-k_{i,j})\eta S\alpha_{i,j} - k_{i,j}\log\{1-\exp[-(\eta\alpha_{i,j}S+B)]\}, \qquad (6)
\end{aligned}
$$

after constants independent of $\alpha_{i,j}$ are dropped. Since $\mathcal{L}_\alpha(\alpha_{i,j}; k_{i,j})$ is a strictly convex function in $\alpha_{i,j}$, it is amenable to global minimization using convex optimization, with or without the inclusion of sparsity-based regularization [25]. The penalized ML (PML) estimate for scene reflectivity is obtained from noisy data $\{k_{i,j}\}_{i,j}$ by solving the following convex program:

$$
\hat{\boldsymbol{\alpha}}^{\text{PML}} = \arg\min_{\alpha:\alpha_{i,j}\geq 0} \sum_{i=1}^{n}\sum_{j=1}^{n} \mathcal{L}_\alpha(\alpha_{i,j}; k_{i,j}) + \beta_\alpha\, \text{pen}_\alpha(\boldsymbol{\alpha}),
$$

where $\text{pen}_\alpha(\cdot)$ is a convex function that penalizes the non-smoothness of the reflectivity estimate, and β_α controls the degree of penalization.

Step 2: Rejection of background detections. Direct application of a similar regularized-ML approach to depth estimation using time-of-detection data is infeasible. This is because the background contribution to the likelihood function creates a nonconvex cost function with locally-optimal solutions that are far from the global optimum. Hence, before estimating depth, a second processing step attempts to identify and censor the detections that are due to background.

Background counts do not contain any scene-depth information. Their detection times are mutually independent over spatial locations with variance $T_r^2/12$. In contrast, since light pulses have duration $T_p \ll T_r$ and depths $z_{i,j}$ are correlated over spatial locations, the detection times of signal counts have conditional variance, given data from neighboring positions, that is much lower than $T_r^2/12$. Based on this key observation, our method to censor a noisy detection at (i,j) is as follows: leftmargin=*

1) Compute the rank-ordered mean (ROM) $t_{i,j}^{\text{ROM}}$ for each pixel, which is the median value of the detection times at the 8 neighboring pixels of (i,j) [26]. If $t_{i,j}^{\text{ROM}}$ cannot be computed due to missing data, then set $t_{i,j}^{\text{ROM}} = \infty$.
2) Estimate the set of uncensored detections, $U_{i,j}$, i.e., those presumed to be signal detections, as follows:

$$\left\{ \ell : |t_{i,j}^{(\ell)} - t_{i,j}^{\text{ROM}}| < 2T_p \left(\frac{B}{\eta \hat{\alpha}_{i,j}^{\text{PML}} S + B} \right), 1 \le \ell \le k_{i,j} \right\}.$$

If $k_{i,j} = 0$, then set $U_{i,j} = \varnothing$. It is demonstrated in [26] that the method of rank-ordered means is effective in detecting pixels that are corrupted by high-variance uniform noise. Because background detections are uniformly distributed, we use the ROM method to reject such detections and only keep signal detections for further processing.

Step 3: Depth Estimation. With background detections rejected, the negative log-likelihood function of depth $z_{i,j}$, given uncensored data $\{t_{i,j}^{(\ell)}\}_{\ell \in U_{i,j}}$, is

$$\mathcal{L}_z\Big(z_{i,j};\, \{t_{i,j}^{(\ell)}\}_{\ell \in U_{i,j}}\Big) = - \sum_{\ell \in U_{i,j}} \log\Big[s\Big(t_{i,j}^{(\ell)} - 2z_{i,j}/c\Big)\Big].$$

If $|U_{i,j}| = 0$, then set $\mathcal{L}_z(z_{i,j}; \{t_{i,j}^{(\ell)}\}_{\ell \in U_{i,j}}) = 0$, so that it has no contribution to the scene's negative log-likelihood cost function.

Our framework allows the use of arbitrary pulse shapes, but many practical pulse shapes are well approximated as $s(t) \propto \exp[-v(t)]$, where $v(t)$ is a convex function in t. Then, $\mathcal{L}_z(z_{i,j}; \{t_{i,j}^{(\ell)}\}_{\ell \in U_{i,j}}) = \sum_{\ell \in U_{i,j}} v(t_{i,j}^{(\ell)} - 2z_{i,j}/c)$ is a convex function in $z_{i,j}$. Our penalized ML estimate for the scene depth image is thus obtained using uncensored data and solving the following convex optimization problem:

$$\hat{\mathbf{z}}^{\text{PML}} = \underset{z: z_{i,j} \in [0, cT_r/2)}{\arg\min} \sum_{i=1}^{n} \sum_{j=1}^{n} \mathcal{L}_z\Big(z_{i,j}; \{t_{i,j}^{(\ell)}\}_{\ell \in U_{i,j}}\Big) + \beta_z \, \text{pen}_z(\mathbf{z}),$$

where $\text{pen}_z(\cdot)$ is a convex function that penalizes non-smoothness of the depth estimate, and $\beta_z > 0$ controls the degree of penalization.

VI. Experimental Results

To test the performance of our proposed 3D structure and reflectivity imaging method, we used the dataset collected by D. Venkatraman for [3], which is available from [27]. The experimental setup used to collect data is shown in Figure 2. A pulsed laser diode with pulse width $T_p = 270\,\text{ps}$ and repetition period $T_r = 100\,\text{ns}$ was used as the illumination source. A two-axis galvo was used to raster scan 1000×1000 pixels. A lensless SPAD detector with quantum efficiency $\eta = 0.35$ was used for detection. The background light level was set such that B equaled the scene-averaged value of $\eta \alpha_{i,j} S$. Further details of the experimental setup are given in [3]. Because raster scanning with a fixed dwell time is equivalent to using a floodlight illumination source and a detector array, our experimental results are indicative of what can be accomplished in real-time imaging scenarios using SPAD arrays.

A. Reflectivity Resolution Test

Reflectivity estimation was tested using the linear grayscale reflectivity chart shown in Figure 4(a). Figure 4(e) shows that our method resolves 16 gray levels, performance similar to that of the ground-truth image from Figure 4(b), which required about 1000 photon detections per pixel.

We quantified the performance of a reflectivity estimator $\hat{\boldsymbol{\alpha}}$ of a true scene reflectivity $\boldsymbol{\alpha}$ using peak signal-to-noise ratio (PSNR):

$$\text{PSNR}(\boldsymbol{\alpha}, \hat{\boldsymbol{\alpha}}) = 10 \log_{10} \left(\frac{\max_{i,j} \alpha_{i,j}^2}{\sum_{i,j} (\alpha_{i,j} - \hat{\alpha}_{i,j})^2 / n^2} \right)$$

Figure 4 show that our method's PSNR exceeds that of pixelwise ML (Equation (4)) by 16 dB, and it exceeds that of the bilateral-filtered [28] pixelwise ML estimate by 3 dB.

B. Depth Resolution Test

Depth resolution was evaluated with a test target comprising $5\,\text{cm} \times 5\,\text{cm}$ squares of varying thickness mounted on a flat board, as shown by the red-labeled squares in Figure 4(a). The smallest resolvable height (thickness) above the reference level is an indicator of achievable depth resolution. Figure 4(e) shows that our method achieves 4 mm depth resolution, which is comparable to that of the ground truth image (Figure 4(b)), which required 100 detections per pixel, and far superior to the very noisy pixelwise ML image (Equation (5)), and its median-filtered [29] version, which appear in Figures 4(c) and (d), respectively.

We quantified the performance of a depth estimator $\hat{\mathbf{z}}$ of a true scene depth $\mathbf{z}$ using root mean-square error (RMSE):

$$\text{RMSE}(\mathbf{z}, \hat{\mathbf{z}}) = \sqrt{\frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} (z_{i,j} - \hat{z}_{i,j})^2}.$$

At the background level in our experiment, the pixelwise ML estimates have an RMSE of at least 3 m. Because many pixels are missing photon detection-time observations, in order to denoise the pixelwise ML estimate, we first perform bicubic

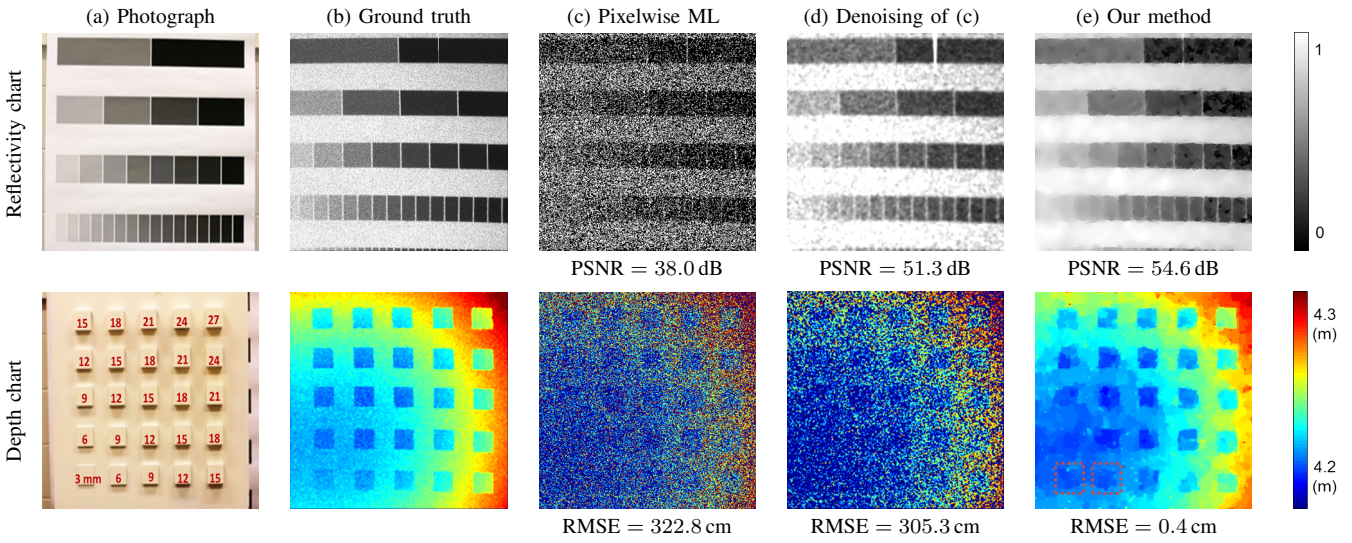


Fig. 4. Resolution test experiments. Reflectivity chart imaging (top) was done using $T_a = 300\,\mu$s and had a mean count per pixel of 0.48. They were scaled to fill the reflectivity interval $[0, 1]$. Depth chart imaging (bottom) was done using $T_a = 6.2\,\mu$s and had a mean count per pixel of 1.1 with 33% of the pixels having missing data, i.e., no detections.

interpolation and then apply median filtering, which is effective in eliminating high-variance noise. The depth resolution of our method (4 mm) corresponds to 760-fold depth error reduction, compared to the denoised estimate.

C. Natural Scenes

Reflectivity and depth images of two natural scenes—a life-size mannequin, and a basketball next to a can—are shown in Figure 5. Ground-truth images, obtained using ML estimation from 200 detections at each pixel, appear in Figure 5(a). The mannequin dataset for pixelwise ML imaging and for our method was generated using acquisition time $T_a = 100\,\mu$s. This dataset had 1.21 detections per pixel averaged over the entire scene with 54% of the pixels having no detections. The basketball-plus-can dataset for pixelwise ML imaging and for our method also had $T_a = 100\,\mu$s, but its mean number of detections per pixel was 2.1, and 32% of its pixels had no detections. All reflectivity images were scaled to fill the interval $[0, 1]$.

Figure 4(b) shows that the pixelwise ML approach gives reflectivity and 3D estimates with low PSNR and high RMSE due to background-count shot noise at low light-levels. Pixels with missing data were imputed with the average of their neighboring 8 pixelwise ML values. Denoising the ML reflectivity estimate using bilateral filtering [28] and the ML depth estimate using median filtering [29] improves the image qualities (Figure 4(c)). However, denoising the 3D structure of the mannequin shirt fails, because this region has very low reflectivity so that many of its pixels have missing data. On the other hand, our framework, which combines accurate photon-detection statistics with spatial prior information, constructs reflectivity and 3D images with 30.6 dB PSNR and 0.8 cm RMSE, respectively (Figure 4(d)). We used the total variation semi-norm [30] as the penalty function in our method, and the penalty parameters were chosen to maximize PSNR for reflectivity imaging and minimize RMSE for 3D imaging.

Figure 6 shows how much photon efficiency we gain over traditional LIDAR systems that use the histogramming approach. The histogramming approach is a pixelwise depth-estimation method that simply searches for the location of the peak in the photon-count histogram of the backreflected pulse. Whereas the log-matched filter is asymptotically ML as $B \to 0^+$, histogramming-based depth estimation method is asymptotically ML as $N \to \infty$. Thus, when T_a is long enough, as is the case in traditional LIDAR, it is effective to use the histogramming-based depth estimation method. Based on PSNR and RMSE values, we see that our framework can allow more than $30\times$ speed-up in acquisition, while constructing the same high-quality 3D and reflectivity images that a traditional LIDAR system would have formed using long acquisition times.

D. Repeatability Test

For each scene, we processed 100 independent dataset and computed the sample RMSE images that approximate $\sqrt{\mathbb{E}[(\alpha_{i,j} - \hat{\alpha}_{i,j}^{\mathrm{PML}})^2]}$ and $\sqrt{\mathbb{E}[(z_{i,j} - \hat{z}_{i,j}^{\mathrm{PML}})^2]}$. The pixelwise RMSE images, provided in Figure 5(e), corroborate the consistent accuracy and high resolution of our computational reflectivity and 3D imager.

E. Effect of System Parameters

Figure 7 shows how the performance of traditional ML and our image-formation methods are affected by changing the acquisition time T_a and the signal-to-background ratio (SBR), defined to be

$$\mathrm{SBR} = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\eta \alpha_{i,j} S}{B}.$$

In our experiment, SBR was modified by changing T_r such that $B = (\eta b_\lambda + d) T_r$ is varied at constant S. Figure 8 provides

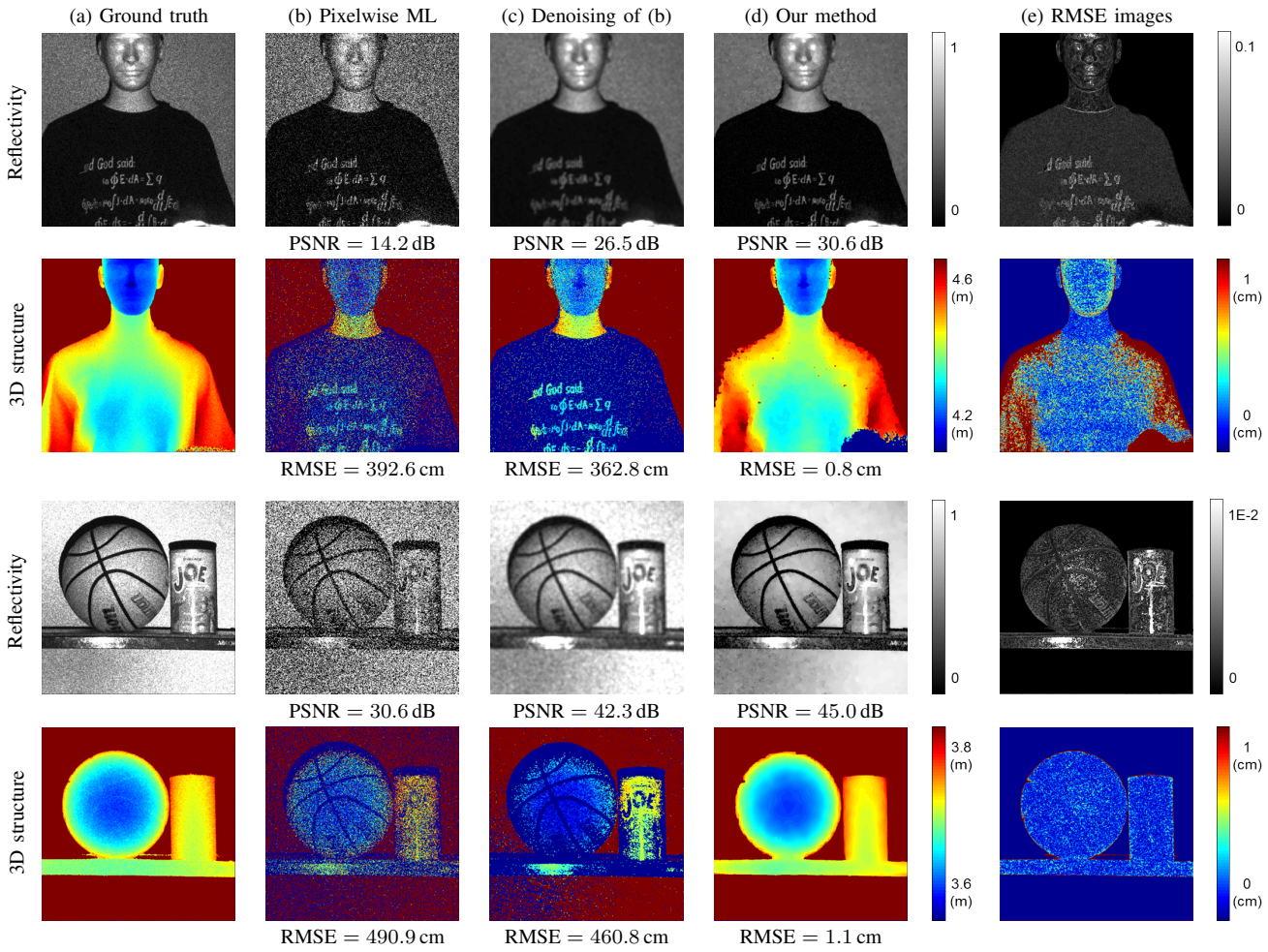


Fig. 5. Experimental results for reflectivity and 3D imaging of natural scenes. We compare the reflectivity and depth images from our proposed method with those from pixelwise ML estimation (see Section III). Pixelwise RMSEs for the reflectivity and 3D images using our method were generated from 100 trials of the experiments. For the mannequin dataset (top), the mean per-pixel count was 1.2 and 55% of the pixels were missing data. For the basketball-plus-can dataset (bottom), the mean per-pixel count was 2.1 and 32% of the pixels were missing data.

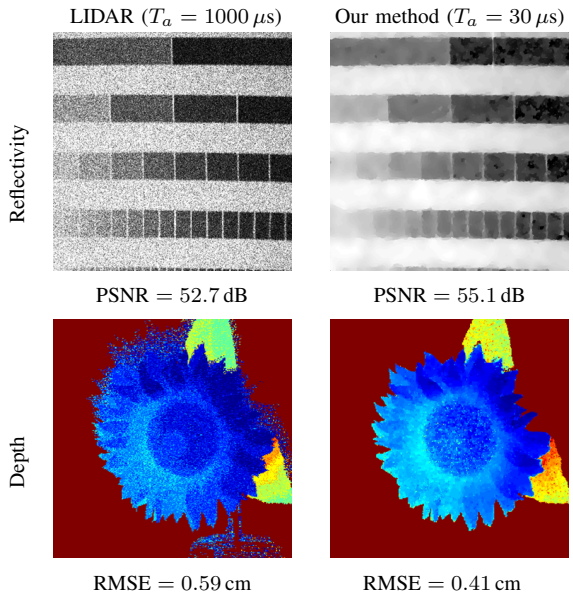
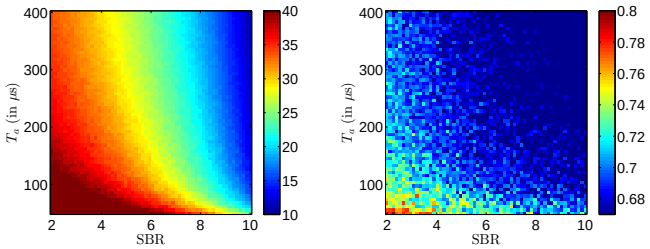


Fig. 6. Comparison between our framework and conventional LIDAR technology.

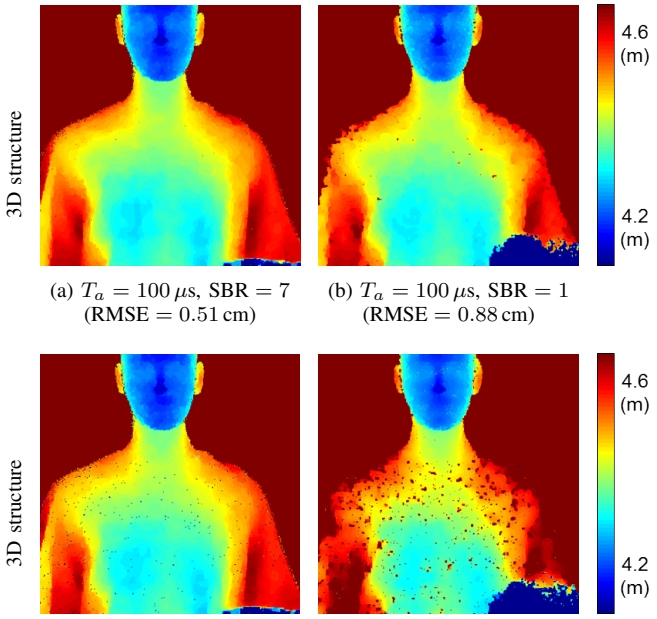


(a) RMSE (in cm) of pixelwise ML

(b) RMSE (in cm) of our method

Fig. 7. RMSE results for 3D imaging. SBR was varied by simulating background levels on the ground-truth mannequin dataset. Note the differences in the colorbar scales.

additional evidence that our method's RMSE decreases monotonically with increasing T_a and SBR, as one would expect. More importantly, it shows that our 3D recovery method is robust under strong background noise and short acquisition times.



(a) $T_a = 100\,\mu$s, SBR $= 7$ (RMSE $= 0.51$ cm)
(b) $T_a = 100\,\mu$s, SBR $= 1$ (RMSE $= 0.88$ cm)
(c) $T_a = 50\,\mu$s, SBR $= 7$ (RMSE $= 5.80$ cm)
(d) $T_a = 50\,\mu$s, SBR $= 1$ (RMSE $= 11.91$ cm)

Fig. 8. Effect of dwell time T_a and signal-to-background ratio (SBR) on our 3D recovery method. For acquisition times of 100 μs and 50 μs, we calculated the mean photon count $k_{i,j}$ over all pixels to be 1.4 and 0.6, respectively.

		FPI	Ours
Mannequin	Mean T_a	244 μs	244 μs
	Mean $k_{i,j}$	1 ppp	2.7 ppp
	Pixels missing data	0%	33%
	PSNR	35 dB	37 dB
	RMSE	0.4 cm	0.3 cm
Sunflower	Mean T_a	15 μs	15 μs
	Mean $k_{i,j}$	1 ppp	8.7 ppp
	Pixels missing data	0%	18%
	PSNR	47 dB	47 dB
	RMSE	0.8 cm	0.5 cm
Basketball and can	Mean T_a	181 μs	181 μs
	Mean $k_{i,j}$	1 ppp	1.7 ppp
	Pixels missing data	0%	24%
	PSNR	44 dB	45 dB
	RMSE	1.1 cm	1.1 cm
Reflectivity chart	Mean T_a	120 μs	120 μs
	Mean $k_{i,j}$	1 ppp	1.7 ppp
	Pixels missing data	0%	27%
	PSNR	54 dB	56 dB
Depth chart	Mean T_a	6.2 μs	6.2 μs
	Mean $k_{i,j}$	1 ppp	1.1 ppp
	Pixels missing data	0%	35%
	RMSE	0.4 cm	0.4 cm

TABLE I
COMPARISON BETWEEN FIRST-PHOTON IMAGING AND OUR FRAMEWORK. NOTE THAT $k_{i,j}$ IS FIXED AND T_a PER PIXEL IS A RANDOM VARIABLE FOR FPI, WHEREAS $k_{i,j}$ IS A RANDOM VARIABLE AND T_a PER PIXEL IS FIXED FOR OUR FRAMEWORK.

F. Comparison with First-Photon Imaging

First-photon imaging [3] requires a single detection at each pixel, hence its dwell time on each pixel is a random variable. Our method requires a fixed dwell time on each pixel, hence the number of detections on each pixel is a random variable. So, to compare the performance of first-photon imaging with that of our method, we set the average per pixel dwell time of the former equal to the fixed per pixel dwell time of the latter. That comparison, shown in Table I, between the PSNRs of their reflectivity images and the RMSEs of their depth images, reveals several interesting characteristics. In particular, when our method's image-acquisition time is matched to that of first-photon imaging, a substantial fraction of its pixels have missing data (no detections). Nevertheless, our method successfully deals with this problem and yields performance similar to, or slightly better than, that of first-photon imaging for the five different scenes we have measured.

G. Limitations

Our method's depth image incurs its highest error near the edges of scene objects. The surface normals at these locations are nearly perpendicular to the line of sight, which dramatically reduces SNR. Consequently, these regions have fewer detections, with more of them being background counts, than do the rest of the pixels. Although our method censors depth anomalies near edges, it estimates the missing depths using spatial correlations, leading to loss of subtle depth details. Also, a detected photon may have originated from a multiple reflection, causing estimation inaccuracy. However, for quasi-Lambertian scenes, diffuse scattering causes multiple reflections to be considerably weaker than the direct reflection. Combined with Poisson statistics, this implies an exponentially diminishing probability of photon detection originating from multiple reflections Finally, our method of estimating reflectivity fails if the background is sufficient to provide a detection in each pulse repetition period with high probability. Hence, in our experiments, we employed a suitably narrowband spectral filter so that $\eta\alpha_{i,j}S \approx B \ll 1$ averaged over the scene.

VII. CONCLUSIONS AND FUTURE WORK

We have extended the FPI framework from [3]—which has a random per-pixel dwell time, because it records exactly one detection for each pixel in the scene—to one that has a fixed dwell time per pixel, but records a random number of detections for each pixel in the scene. Both systems combine physically accurate single-photon detection statistics with exploitation of the spatial correlations found in natural scenes. Our new method's fixed dwell time, however, makes it compatible with detector arrays. Hence it is significant that we demonstrated its ability to produce accurate reflectivity and depth images using on the order of 1 detected photon per pixel averaged over the scene, even with significant background light and a substantial fraction of the pixels having no detections. This highly photon-efficient performance motivates the development of accurate and low-power SPAD array-based 3D and reflectivity imagers. Current commercial CMOS-based depth imagers, for example Kinect and TOF cameras, have significantly impacted research in 3D imaging. These sensors offer high depth resolution, but their use is limited due to poor spatial resolution and high power consumption. Our approach offers a potential route to solving these problems.

More generally, our framework can be used in a variety of low light-level imaging applications using photon-counting detectors, such as spatially-resolved fluorescence lifetime imag-

ing (FLIM) [31] and high-resolution LIDAR [10]. Our method naturally extends to imaging at a variety of wavelengths, making it suitable for practical implementations. Furthermore, future advances in optoelectronic methods can improve the accuracy of our 3D and reflectivity imager. In particular, it can benefit from improved background suppression techniques [4] and range-gating methods [18].

APPENDIX

This appendix provides performance analyses for pixelwise estimation. The Cramér-Rao lower bound (CRLB) sets the limit on the mean-square error (MSE) of an unbiased estimator of a parameter. Let x be a scalar continuous parameter in the probability density function $f_Y(y;x)$ of random variable Y. The CRLB for an unbiased estimator, $\hat{x}$, of the parameter x based on observation of Y is the inverse of the Fisher information $J(x)$ [24]:

$$\begin{aligned}\mathbb{E}[(x-\hat{x})^2] &\geq \text{CRLB}(x) = J^{-1}(x) \\ &= \left\{\mathbb{E}\left[\frac{d^2}{d^2x}\left(-\log f_Y(y;x)\right)\right]\right\}^{-1}. \end{aligned} \tag{7}$$

An unbiased estimator $\hat{x}$ is efficient if $\mathbb{E}[(x-\hat{x})^2] = \text{CRLB}(x)$.

A. Mean-Square Error of Reflectivity Estimation

With some algebra, the CRLB for estimating the reflectivity, $\alpha_{i,j}$, at pixel (i,j) can be shown to be

$$\begin{aligned}\text{CRLB}\left(\alpha_{i,j}\right) &= \left\{\mathbb{E}\left[\frac{d^2}{d^2\alpha_{i,j}}\left(-\log \Pr[K_{i,j}=k;\alpha_{i,j}]\right)\right]\right\}^{-1} \\ &= \left\{\mathbb{E}\left[\frac{k\eta^2S^2\exp\left[\eta\alpha_{i,j}S+B\right]}{\left(\exp\left[\eta\alpha_{i,j}S+B\right]-1\right)^2}\right]\right\}^{-1} \\ &= \frac{\exp\left[\eta\alpha_{i,j}S+B\right]-1}{N\eta^2S^2} \\ &\approx \frac{\eta\alpha_{i,j}S+B}{N\eta^2S^2}, \end{aligned} \tag{8}$$

where the approximation makes use of the low-flux condition. As could easily be expected, increasing the number of pulse repetitions, N, collects more photons and hence decreases the CRLB.

Note, however, that we cannot directly use the CRLB result to lower bound the mean-square error of the unconstrained ML reflectivity estimate $\hat{\alpha}_{i,j}^{\text{ML}}$ given by

$$\hat{\alpha}_{i,j}^{\text{ML}} = \frac{1}{\eta S}\left[\log\left(\frac{N}{N-k_{i,j}}\right) - B\right].$$

This is because the ML estimate is biased, ($\mathbb{E}[\hat{\alpha}_{i,j}^{\text{ML}}] \neq \alpha_{i,j}$):

$$\begin{aligned}\mathbb{E}\left[\hat{\alpha}_{i,j}^{\text{ML}}\right] &= \mathbb{E}\left[\frac{1}{\eta S}\log\left(\frac{N}{N-k_{i,j}}\right) - \frac{B}{\eta S}\right] \\ &= \frac{1}{\eta S}\log N - \frac{1}{\eta S}\mathbb{E}\left[\log\left(N-K_{i,j}\right)\right] - \frac{B}{\eta S} \\ &> \frac{1}{\eta S}\log N - \frac{1}{\eta S}\log\left(N-\mathbb{E}[K_{i,j}]\right) - \frac{B}{\eta S} \\ &= \alpha_{i,j}, \end{aligned}$$

where the strict inequality comes from Jensen's inequality and the fact that the logarithm function is strictly concave.

When $N \to \infty$ and $\eta\alpha_{i,j}S + B \to 0^+$ with $N[1-\exp(\eta\alpha_{i,j}S+B)]$ equal to a constant $C(\alpha_{i,j})$, the ML reflectivity estimate is

$$\hat{\alpha}_{i,j}^{\text{ML}} = \frac{k}{N\eta S} - \frac{B}{\eta S}. \tag{9}$$

In this case, the CRLB equals the MSE of the ML reflectivity estimate,

$$\text{CRLB}(\alpha_{i,j}) = \mathbb{E}\left[\left(\alpha_{i,j}-\hat{\alpha}_{i,j}^{\text{ML}}\right)^2\right] = \frac{1}{N}\left(\frac{\alpha_{i,j}}{\eta S} + \frac{B}{\eta^2S^2}\right),$$

We see that the CRLB expression from Poisson likelihood is equal to the first-order Taylor expansion of the CRLB expression of the exact binomial likelihood given by Equation (8).

Knowing that the ML solution in the limiting Poisson distribution is unbiased and efficient, we conclude that the maximum likelihood reflectivity estimate $\hat{\alpha}_{i,j}^{\text{ML}}$ is efficient asymptotically as $(\eta\alpha_{i,j}S+B) \to 0^+$ and $N \to \infty$.

B. Mean-Square Error of 3D Estimation

We again assume that $\eta\alpha_{i,j}S+B \to 0^+$ and $N \to \infty$ such that $N[1-\exp(\eta\alpha_{i,j}S+B)]$ is a constant $C(\alpha_{i,j})$. The CRLB for estimating the depth $z_{i,j}$ is then found to be

$$\begin{aligned}\text{CRLB}(z_{i,j}) &= \left\{\mathbb{E}\left[\frac{d^2}{d^2z_{i,j}}\left(-\log f_{T_{i,j}}(\{t_{i,j}^{(\ell)}\}_{\ell=1}^{k_{i,j}};z_{i,j})\right)\right]\right\}^{-1} \\ &= \left\{\mathbb{E}\left[-\sum_{\ell=1}^{k_{i,j}}\frac{d^2}{d^2z_{i,j}}\log f_{T_{i,j}}(t_{i,j}^{(\ell)};z_{i,j})\right]\right\}^{-1} \\ &= \frac{1}{C(\alpha_{i,j})}\left(\int_0^{T_r}\frac{\dot{p}(t;z_{i,j})^2}{p(t;z_{i,j})}\,dt\right)^{-1}, \end{aligned} \tag{10}$$

where

$$p(t;z_{i,j}) = \frac{\lambda_{i,j}(t)}{\int_0^{T_r}\lambda_{i,j}(\tau)\,d\tau}$$

with $\lambda_{i,j}(t)$ being the single-pulse rate from Equation (1), and $\dot{p}(t;z_{i,j})$ is its derivative with respect to time.

We can exactly compute the MSE expression for certain pulse waveforms. For example, if the illumination waveform is a Gaussian pulse $s(t) \propto \exp\left[t^2/2T_p^2\right]$, then using the unconstrained log-matched filter expression, we get

$$\hat{z}_{i,j}^{\text{ML}} = \arg\max_{z_{i,j}}\sum_{\ell=1}^{k_{i,j}}\log\left[s(t_{i,j}^{(\ell)}-2z_{i,j}/c)\right] = \frac{c}{2}\left(\frac{\sum_{\ell=1}^{k_{i,j}}t_{i,j}^{(\ell)}}{k_{i,j}}\right),$$

given $k_{i,j} \geq 1$. If $k_{i,j} = 0$, then a standard pixelwise data imputation is done by making a uniformly random guess over the interval $[0, cT_r/2)$. Assuming $B = 0$, the MSE expression

can be written as

$$\begin{aligned}
\mathbb{E}[(z_{i,j} - \hat{z}_{i,j}^{\mathrm{ML}})^2] &= \mathbb{E}_{K_{i,j}}\{\mathbb{E}[(z_{i,j} - \hat{z}_{i,j}^{\mathrm{ML}})^2 \mid K_{i,j}]\} \\
&= \sum_{k=0}^{\infty} \frac{C^k(\alpha_{i,j}) e^{-C(\alpha_{i,j})}}{k!} \mathbb{E}[(z_{i,j} - \hat{z}_{i,j}^{\mathrm{ML}})^2 \mid K_{i,j} = k] \\
&= e^{-C(\alpha_{i,j})} \left[\left(\frac{cT_r}{2}\right)^2 + \left(z_{i,j} - \frac{cT_r}{4}\right)^2 \right. \\
&\qquad \left. + \sum_{k=1}^{\infty} \frac{C^k(\alpha_{i,j})}{k!} \frac{1}{k} \left(\frac{cT_p}{2}\right)^2 \right] \\
&= e^{-C(\alpha_{i,j})} \left(\underbrace{\left(\frac{cT_r}{2}\right)^2 + \left(z_{i,j} - \frac{cT_r}{4}\right)^2}_{\text{random guess error}} \right. \\
&\qquad \left. + \underbrace{\left(\frac{cT_p}{2}\right)^2 \int_0^{C(\alpha_{i,j})} \frac{\exp[\tau] - 1}{\tau} \, d\tau}_{\text{pulse-width error}} \right). \quad (11)
\end{aligned}$$

As $C(\alpha_{i,j}) \to \infty$, the pulse-width error term in MSE dominates and $\hat{z}_{i,j}^{\mathrm{ML}}$ becomes an efficient estimator.

REFERENCES

[1] B. Schwarz, "LIDAR: Mapping the world in 3d," *Nat. Phot.*, 2010.

[2] D. Shin, A. Kirmani, V. K. Goyal, and J. H. Shapiro, "Computational 3d and reflectivity imaging with high photon efficiency," in *Proc. IEEE Int. Conf. Image Process.*, Paris, France, Oct. 1995, to appear.

[3] A. Kirmani, D. Venkatraman, D. Shin, A. Colaço, F. N. C. Wong, J. H. Shapiro, and V. K. Goyal, "First-photon imaging," *Science*, vol. 343, no. 6166, pp. 58–61, 2014.

[4] A. McCarthy, R. J. Collins, N. J. Krichel, V. Fernández, A. M. Wallace, and G. S. Buller, "Long-range time-of-flight scanning sensor based on high-speed time-correlated single-photon counting," *Appl. Opt.*, vol. 48, no. 32, pp. 6241–6251, 2009.

[5] S. B. Gokturk, H. Yalcin, and C. Bamji, "A time-of-flight depth sensor — system description, issues and solutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2004.

[6] S. Lee, O. Choi, and R. Horaud, *Time-of-Flight Cameras: Principles, Methods and Applications*. Springer, 2013.

[7] A. V. Jelalian, "Laser radar systems," in *EASCON'80; Electronics and Aerospace Systems Conference*, vol. 1, 1980, pp. 546–554.

[8] Z. Zhang, "Microsoft Kinect sensor and its effect," *IEEE Multimedia*, vol. 19, no. 2, pp. 4–10, 2012.

[9] D. A. Forsyth and J. Ponce, *Computer Vision: A Modern Approach*. Prentice Hall, 2002.

[10] B. F. Aull, A. H. Loomis, D. J. Young, R. M. Heinrichs, B. J. Felton, P. J. Daniels, and D. J. Landers, "Geiger-mode avalanche photodiodes for three-dimensional imaging," *Lincoln Lab. J.*, vol. 13, no. 2, pp. 335–349, 2002.

[11] A. Kirmani, D. Venkatraman, A. Colaço, F. N. C. Wong, and V. K. Goyal, "High photon efficiency computational range imaging using spatio-temporal statistical regularization," in *Proc. CLEO*, San Jose, CA, Jun. 2013, paper QF1B.2.

[12] A. Kirmani, A. Colaço, D. Shin, and V. K. Goyal, "Spatio-temporal regularization for range imaging with high photon efficiency," in *SPIE Wavelets and Sparsity XV*, San Diego, CA, Aug. 2013, pp. 88 581F–88 581F.

[13] D. Shin, A. Kirmani, A. Colaço, and V. K. Goyal, "Parametric Poisson process imaging," in *Proc. IEEE Global Conf. Signal Inform. Process.*, Austin, TX, Dec. 2013, pp. 1053–1056.

[14] G. A. Howland, P. B. Dixon, and J. C. Howell, "Photon-counting compressive sensing laser radar for 3d imaging," *Appl. Optics*, vol. 50, no. 31, pp. 5917–5920, Nov. 2011.

[15] A. Colaço, A. Kirmani, G. A. Howland, J. C. Howell, and V. K. Goyal, "Compressive depth map acquisition using a single photon-counting detector: Parametric signal processing meets sparsity," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Providence, RI, Jun. 2012, pp. 96–102.

[16] G. A. Howland, D. J. Lum, M. R. Ware, and J. C. Howell, "Photon counting compressive depth mapping," *Opt. Expr.*, vol. 21, no. 20, pp. 23 822–23 837, Oct. 2013.

[17] A. Kirmani, A. Colaço, F. N. C. Wong, and V. K. Goyal, "Exploiting sparsity in time-of-flight range acquisition using a single time-resolved sensor," *Opt. Expr.*, vol. 19, no. 22, pp. 21 485–21 507, Oct. 2011.

[18] J. Busck and H. Heiselberg, "Gated viewing and high-accuracy three-dimensional laser radar," *Appl. Opt.*, vol. 43, no. 24, pp. 4705–4710, 2004.

[19] G. Goltsman, O. Okunev, G. Chulkova, A. Lipatov, A. Semenov, K. Smirnov, B. Voronov, A. Dzardanov, C. Williams, and R. Sobolewski, "Picosecond superconducting single-photon optical detector," *Appl. Phys. Lett.*, vol. 79, no. 6, pp. 705–707, 2001.

[20] B. I. Erkmen and B. Moision, "Maximum likelihood time-of-arrival estimation of optical pulses via photon-counting photodetectors," in *Proc. IEEE Int. Symp. Inform. Theory*, 2009, pp. 1909–1913.

[21] D. L. Snyder, *Random Point Processes*. Wiley, New York, 1975.

[22] D. P. Bertsekas and J. N. Tsitsiklis, *Introduction to Probability*. Athena Scientific, 2002.

[23] Y. Chen, J. D. Müller, P. T. So, and E. Gratton, "The photon counting histogram in fluorescence fluctuation spectroscopy," *Biophys. J.*, vol. 77, no. 1, pp. 553–567, 1999.

[24] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory*. Prentice Hall PTR, 1998.

[25] Z. T. Harmany, R. F. Marcia, and R. M. Willett, "This is SPIRAL-TAP: Sparse Poisson intensity reconstruction algorithms—theory and practice," *IEEE Trans. Image Process.*, vol. 21, no. 3, pp. 1084–1096, 2012.

[26] E. Abreu, M. Lightstone, S. K. Mitra, and K. Arakawa, "A new efficient approach for the removal of impulse noise from highly corrupted images," *IEEE Trans. Image Process.*, vol. 5, no. 6, pp. 1012–1025, 1996.

[27] "First-photon imaging project," http://www.rle.mit.edu/first-photon-imaging/.

[28] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Proc. 6th Int. Conf. Comput. Vis.*, 1998, pp. 839–846.

[29] R. Jain, R. Kasturi, and B. G. Schunck, *Machine Vision*. McGraw-Hill New York, 1995, vol. 5.

[30] S. Osher, A. Solé, and L. Vese, "Image decomposition and restoration using total variation minimization and the H^{-1} norm," *Multiscale Model. Simul.*, vol. 1, no. 3, pp. 349–370, 2003.

[31] D. O'Connor, *Time-correlated single photon counting*. Academic Press, 1984.