View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by DSpace@MIT

PROCEEDINGS OF SPIE

SPIEDigitalLibrary.org/conference-proceedings-of-spie

Photon-efficient super-resolution laser radar

Dongeek Shin, Jeffrey H. Shapiro, Vivek K. Goyal

Dongeek Shin, Jeffrey H. Shapiro, Vivek K. Goyal, "Photon-efficient super-resolution laser radar," Proc. SPIE 10394, Wavelets and Sparsity XVII, 1039409 (24 August 2017); doi: 10.1117/12.2273208

SPIE.

Event: SPIE Optical Engineering + Applications, 2017, San Diego, California, United States

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 3/16/2018 Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

Photon-efficient super-resolution laser radar

Dongeek Shin[a], Jeffrey H. Shapiro[a], and Vivek K Goyal[b]

[a]Massachusetts Institute of Technology
[b]Boston University

ABSTRACT

The resolution achieved in photon-efficient active optical range imaging systems can be low due to non-idealities such as propagation through a diffuse scattering medium. We propose a constrained optimization-based framework to address extremes in scarcity of photons and blurring by a forward imaging kernel. We provide two algorithms for the resulting inverse problem: a greedy algorithm, inspired by sparse pursuit algorithms; and a convex optimization heuristic that incorporates image total variation regularization. We demonstrate that our framework outperforms existing deconvolution imaging techniques in terms of peak signal-to-noise ratio. Since our proposed method is able to super-resolve depth features using small numbers of photon counts, it can be useful for observing fine-scale phenomena in remote sensing through a scattering medium and through-the-skin biomedical imaging applications.

Keywords: deconvolution, depth imaging, inverse problem, laser radar, photon counting, super-resolution

1. INTRODUCTION

Laser radar (LADAR, also known as LIDAR) imaging systems recover the depth map of a scene, typically by using a raster-scanning pulsed light source and a single time-resolved photodetector. Unlike other depth imaging solutions, such as stereo vision,[1] structured light sensors,[2] and amplitude-modulated time-of-flight cameras,[3] laser radar can deliver high accuracy and maximum range at the same time due to its ability to directly measure object distance from the travel time of a short, powerful pulse.[4] Thus, it finds its use in many engineering applications including robot navigation,[5] bathymetry,[6] 3D city reconstruction,[7] and object shape scanning.[8]

Traditionally, by operating the light source at high flux, the optical pulse signal that is backreflected from a scene pixel and incident at the photodetector is recorded using analog-to-digital conversion; from the scaling and the mean time-of-flight of the observed pulse signal, one can infer reflectivity and depth at that particular pixel. When the light incident at the detector is very weak, however, conventional photodetectors fail to pick up signal due to the low detector sensitivity. In such a condition, highly sensitive photon-counting detectors, such as avalanche photodiodes (APDs)[9] and photomultipliers (PMTs),[10] are used to resolve individual photon detections. The backreflected pulse signal is then observed as a photon-count histogram.

There exist well-known methods that rely on pixelwise photon counting statistics and perform maximum-likelihood (ML) estimation (or some variant of it) to recover depth.[11,12] However, conventional pixelwise depth estimation performs well only under high photon count assumptions, and using low-light photon-count data results in noisy depth maps. Several methods have been developed to mitigate the noise in the low-flux photon-counting process by exploiting the spatial correlations in natural images, including methods that use denoising,[13] Markov random field models,[14] and convex optimization.[15–17] One core assumption behind these existing low-light depth imaging frameworks is that the spatial spot size of the illumination is small enough that we can assume that each pixel measurement has addressed a different scene patch. However, in practice, this independent-pixel assumption can break down due to non-idealities in the illumination source and imaging conditions. For example, when imaging through a scattering medium, such as a diffuser, the illumination spot on the hidden target behind the scatterer is constrained to be large and often non-uniform.[18,19] In such cases, the reconstructed depth map overestimates the sizes of object features; this is an artifact known as "blooming" in the LIDAR literature.[20] A similar problem occurs in the dual imaging framework known as flash LIDAR, where instead of using a scanning source and a single detector, one uses a diffuse source and an array of detectors.[21,22] Here, detector array lens aberrations lead to a non-trivial point spread function (PSF),[23] which acts as a convolution operator and spatially blurs the set of measurements.

Wavelets and Sparsity XVII, edited by Yue M. Lu, Dimitri Van De Ville, Manos Papadakis, Proc. of SPIE
Vol. 10394, 1039409 · © 2017 SPIE · CCC code: 0277-786X/17/$18 · doi: 10.1117/12.2273208

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 3/16/2018 Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

The problem of recovering a high-resolution depth map from low-resolution measurements is known as the depth super-resolution problem. There are several approaches for depth super-resolution, each exploiting different physical and image constraints. One of them is the upsampling approach, which converts the depth map at low-resolution to high-resolution by exploiting, for example, the information in the high-resolution reflectivity image or in multi-view image acquisition.[24,25] Another, which is more relevant to us, is image deconvolution-based super-resolution. This approach exploits the fact that many image acquisition constraints that lead to low-resolution data can be modeled using a convolution operation. Several deconvolution-based depth super-resolution techniques have been developed for commodity depth cameras.[26,27] For laser radar, a framework that uses spatiotemporal deconvolution of the photon histogram data for accurate depth imaging has been demonstrated.[28] The limitation of this framework is that the accuracy of high-resolution depth reconstruction degrades in the low-photon count regime.

In this paper, we propose a depth super-resolution framework for laser radar that incorporates detector, scene, and imaging models. Our contributions can be summarized with the following points.

- **Modeling:** Our model combines the statistics of photon-counting noise, the spatiotemporal kernel, and single-reflector-per-pixel constraints of reflective depth imaging to formulate the super-resolution imaging problem as a single optimization problem.
- **Algorithms:** To solve the optimization problem for scene reflectivity and depth reconstruction, we propose two algorithms:
  1. ML solver: We develop a heuristic approximation to the ML estimator, inspired by the compressive sampling matching pursuit (CoSaMP) algorithm for ℓ_0-norm-constrained optimization,[29] that uses moderate photon counts to locate the sparse object locations.
  2. Regularized ML solver: The ML solution performs poorly at low photon counts and has high computational complexity. By exploiting small total variation (sparse gradients) of natural reflectivity and depth scenes, we improve the accuracy of the reconstruction. With certain reparametrization of variables, we show that the TV-regularized optimization problem for reflectivity and depth reconstruction is convex and can be solved using simple first-order gradient descent methods.
- **Simulations:** Using a simulated forward imaging kernel, we show that our proposed algorithms, especially Algorithm 2, outperform conventional depth super-resolution methods such as (1) pixelwise ML estimation, (2) pixelwise ML followed by TV-regularized Gaussian deconvolution,[30] and (3) spatiotemporal Richardson-Lucy data deconvolution[31] followed by pixelwise ML, in the style of McMahon et al.[28]

2. PHOTON-COUNTING DEPTH IMAGING SETUP AND MODEL

Let $s(t)$ be a non-negative function that starts at time 0 and has root-mean-square pulsewidth T_p. An optical source, such as a laser, illuminates the scene with power waveform $s(t)$ at scanning angle (θ_x, θ_y). Let T_r be the total acquisition period for one scanning angle. Then the time-correlated photon-counting detector is able to record the photon-count histogram of the spatially-distorted backreflected waveform with $n_t = T_r/\Delta$ bins, where Δ is the bin width. The observed histogram consists of multiple modes from a depth-rich scene, because the spatial kernel mixes responses from multiple reflectors that are at different depths. For scenes with partially-reflecting scatterers, we time gate and reject the first-bounce response coming from the scatterer itself so that the observed data always describes only the scene behind the scatterer. This acquisition process is repeated for $n_x \times n_y$ scanning angles, and the objective is to form an $n_x \times n_y$ spatially-resolved intensity and depth map pair.

For the scanning angle (θ_x, θ_y), the detector rate function $v(t; \theta_x, \theta_y)$, which generates the photon counts, is derived using the linearity of light transport:

$$v(t;\theta_x,\theta_y) = \int_{(\theta'_x,\theta'_y)\in \mathrm{FOV}} h_{xy}(\theta_x-\theta'_x,\theta_y-\theta'_y)\, a(\theta'_x,\theta'_y)\, h_t(t-2d(\theta'_x,\theta'_y)/c)\, d\theta'_x d\theta'_y + b, \tag{1}$$

for $t \in [0, T_r)$. Here FOV denotes the rectangular field-of-view of the scanner, c is the speed of light, $(a(\theta'_x, \theta'_y), d(\theta'_x, \theta'_y))$ is the (intensity, depth) pair at scanning direction (θ'_x, θ'_y), and b is the extraneous background and

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 3/16/2018 Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

dark count rate. We have h_{xy} and h_t denoting the spatial and temporal kernels that originate from scatterers and pulsewidth plus detector jitter, respectively. Intuitively, (1) shows that the scene impulse response, which contains intensity and depth information, is convolved in space by h_{xy} and in time by h_t, leading to our low-resolution observations. Here, we have assumed that $a(\theta'_x, \theta'_y)$ multiplicatively combines object reflectivity, effect of radial fall-off of optical power, and surface angles.

In practice, the spatiotemporal domain is discrete in both pixels and time bins, so we express the continuous-domain convolution (1) using matrix operations. Let $n_x \times n_y$ be the number of pixels and $\bar{\mathbf{A}}, \bar{\mathbf{D}} \in \mathbb{R}_+^{n_x \times n_y}$ be the high-resolution scene intensity and depth pair that we aim to recover. Then, let $\bar{\mathbf{X}}$ be a $n_x \times n_y \times n_t$ scene impulse response matrix, where at its (i,j)-th scanning pixel is a vector that has one non-zero entry, whose value is $\bar{\mathbf{A}}_{i,j}$ and index is $\bar{\mathbf{T}}_{i,j} = \text{round}(2\bar{\mathbf{D}}_{i,j}/(c\Delta))$. (Here the assumption is that $\bar{\mathbf{D}}_{i,j} \gg c\Delta$ and integer rounding offers a good approximation.) Let $\mathbf{B}$ be a $(b\Delta)$-constant matrix of size $n_x \times n_y \times n_t$. Let $\mathbf{h}_{xy}$ and $\mathbf{h}_t$ be the $k_x \times k_y$ and length-k_t discrete kernel approximation of h_{xy} and h_t, respectively. Finally, let $\mathbf{h}$ be the $k_x \times k_y \times k_t$ spatiotemporal kernel that is the outer product of $\mathbf{h}_{xy}$ and $\mathbf{h}_t$.

By the theory of photodetection,[11] our $n_x \times n_y \times n_t$ photon histogram matrix is distributed as

$$\boldsymbol{Y} \sim \text{Poisson}\left(\mathbf{h} * \bar{\mathbf{X}} + \mathbf{B}\right), \tag{2}$$

where $*$ denotes the convolution operator in the spatiotemporal domain and Poisson$(\cdot)$ is defined entrywise. Our depth super-resolution problem is to estimate $\bar{\mathbf{X}}$, which contains intensity and depth information, from low-resolution photon histogram data $\mathbf{Y}$, the spatiotemporal kernel $\mathbf{h}$, and extraneous response $\mathbf{B}$.

3. MAXIMUM-LIKELIHOOD RECONSTRUCTION

Define $\mathcal{L}_X(\mathbf{X}; \mathbf{Y}, \mathbf{h}, \mathbf{B})$ to be the negative log-likelihood function of $\mathbf{X}$ derived from (2). Also, define $\mathcal{S}_1(n_x, n_y, n_t)$ as the set of $n_x \times n_y \times n_t$ sparse matrices in which each of the size-n_t vector slices in the third dimension contains exactly one non-zero entry. Then, we aim to solve the maximum-likelihood deconvolution problem:

$$\begin{aligned} \underset{\mathbf{X}}{\text{minimize}} \quad & \mathcal{L}_X(\mathbf{X}; \mathbf{Y}, \mathbf{h}, \mathbf{B}) \\ \text{subject to} \quad & \mathbf{X} \in \mathcal{S}_1(n_x, n_y, n_t), \\ & \mathbf{X}_{i,j,k} \geq 0, \ \forall i, j, k, \end{aligned} \tag{3}$$

where $\mathbf{X}_{i,j,k} \geq 0$ models the non-negativity of light intensity. The negative log Poisson likelihood cost is a convex function and the non-negativity is a convex constraint so the global minimizer could be found efficiently, if those two were the only constraints in the optimization program. However, $\mathcal{S}_1(n_x, n_y, n_t)$ is a non-convex set of sparse matrices. Thus, we aim to develop a reconstruction framework inspired by well-known sparse signal pursuit methods that use the non-convex ℓ_0-norm constraint set to promote sparsity of solution.

We first express our observation model in (2) in the familiar matrix-vector form. We can flatten the size-$(n_x \times n_y \times n_t)$ matrices $\mathbf{Y}, \mathbf{X}, \mathbf{B}$ into two-dimensional ones, $f(\mathbf{Y}), f(\mathbf{X}), f(\mathbf{B})$, of size $n_x n_y \times n_t$, by column-first grouping the transverse pixel dimensions together. Then, (2) can be expressed using a triple matrix product: $f(\boldsymbol{Y}) \sim \text{Poisson}\left(\mathbf{H}_{xy} f(\mathbf{X}) \mathbf{H}_t^T + f(\mathbf{B})\right)$, where $\mathbf{H}_{xy}$ is the $n_x n_y \times n_x n_y$ block-convolution matrix generated using $\mathbf{h}_{xy}$ and $\mathbf{H}_t$ is the $n_t \times n_t$ convolution matrix from $\mathbf{h}_t$. Furthermore, using the identity $\text{vec}(\mathbf{H}_t f(\mathbf{X})^T \mathbf{H}_{xy}^T) = (\mathbf{H}_{xy} \otimes \mathbf{H}_t)\text{vec}(f(\mathbf{X})^T)$, where vec$(\cdot)$ denotes the column-first matrix vectorization operator and $\otimes$ denotes the Kronecker product, we rewrite the matrix optimization problem in (3) as a vector optimization problem by defining $\mathbf{y} = \text{vec}(f(\mathbf{Y})^T)$, $\mathbf{x} = \text{vec}(f(\mathbf{X})^T)$, $\mathbf{b} = \text{vec}(f(\mathbf{B})^T)$, $\mathbf{H} = \mathbf{H}_{xy} \otimes \mathbf{H}_t$:

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & \mathcal{L}_x(\mathbf{x}; \mathbf{y}, \mathbf{H}, \mathbf{b}) \\ \text{subject to} \quad & \mathbf{x} \in \mathcal{S}_{n_t,1}(n), \\ & \mathbf{x}_i \geq 0, \ \forall i, \end{aligned} \tag{4}$$

where we used $\mathcal{L}_x(\mathbf{x}; \mathbf{y}, \mathbf{H}, \mathbf{b})$ as the negative log-likelihood derived from $\mathbf{y}$ being the observation of a Poisson channel with rate $\mathbf{H}\mathbf{x} + \mathbf{b}$, defined $n = n_x n_y n_t$, and used $\mathcal{S}_{n_t,1}(n)$ as the set of size-n sparse vectors, where each of its $n_x n_y$ subvector blocks of size n_t has only one non-zero entry.

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 3/16/2018 Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

Algorithm 1 ML reconstruction by greedy pursuit

Input: $\mathbf{y}$, $\mathbf{H}$, $\mathbf{b}$, δ
Output: $\hat{\mathbf{x}}^{\mathrm{ML}}$
Initialize $\mathbf{x}^{(0)} \leftarrow \vec{0}$, $\mathbf{r} \leftarrow \mathbf{y} - \mathbf{b}$, $k \leftarrow 0$;
repeat
  $k \leftarrow k + 1$;
  $\hat{\mathbf{x}} \leftarrow \mathbf{H}^T\mathbf{r}$;
  $\Omega \leftarrow \mathrm{supp}\left(\mathbb{S}(\hat{\mathbf{x}}, n_t)\right) \cup \mathrm{supp}\left(\mathbf{x}^{(k-1)}\right)$;
  $\mathbf{u}|_{\Omega} \leftarrow \mathbb{N}(\mathbf{H}_{\Omega}, \mathbf{y}, \mathbf{b})$; $\mathbf{u}|_{\Omega^c} \leftarrow \mathbf{0}$;
  $\mathbf{x}^{(k)} \leftarrow \mathbf{u}$;
  $\mathbf{r} \leftarrow \mathbf{y} - (\mathbf{H}\mathbf{x}^{(k)} + \mathbf{b})$;
until $\|\mathbf{x}^{(k-1)} - \mathbf{x}^{(k)}\|_2^2 < \delta$
$\hat{\mathbf{x}}^{\mathrm{ML}} \leftarrow \mathbf{x}^{(k)}$;

Our proposed greedy algorithm for (4) is outlined in Algorithm 1. In the algorithm, which is similar to a previous modification of CoSaMP for photon-efficient imaging,[32] $\mathbb{S}(\hat{\mathbf{x}}, n_t)$ is the n-sparse vector generated by picking the best 1-sparse vector for every size-n_t subvector of the intermediate solution $\hat{\mathbf{x}}$, which may be dense; $\mathrm{supp}(\mathbf{x})$ is the set of indices of non-zero entries of vector $\mathbf{x}$; $\mathbf{H}_{\Omega}$ is the submatrix of $\mathbf{H}$ that is obtained by horizontally concatenating its columns indexed by set Ω; and $\mathbb{N}(\mathbf{H}_{\Omega}, \mathbf{y}, \mathbf{b})$ as the solution to the following non-negative least-squares program:

$$\begin{aligned} \underset{\mathbf{v}}{\text{minimize}} \quad & \|\mathbf{y} - (\mathbf{H}_{\Omega}\mathbf{v} + \mathbf{b})\|_2^2 \\ \text{subject to} \quad & \mathbf{v}_i \geq 0, \ \forall i. \end{aligned} \tag{5}$$

Using $\hat{\mathbf{x}}^{\mathrm{ML}}$, the n-sparse output vector of Algorithm 1, we can obtain our scene intensity and depth estimates, $\hat{\mathbf{A}}^{\mathrm{ML}}, \hat{\mathbf{D}}^{\mathrm{ML}}$, for each pixel by reading out the value and index of the non-zero entry in each size-n_t block vector.

4. REGULARIZED MAXIMUM-LIKELIHOOD RECONSTRUCTION

Algorithm 1 solves (4) and estimates a sparse vector of size $n_x n_y n_t$. However, (4) does not include any natural image models: intensity and depth images are typically compressible due to their spatial correlations. Thus, Algorithm 1 has limited photon efficiency, because its deconvolution relies solely on the single-reflector-per-pixel constraint. Moreover, Algorithm 1 carries the limitation that $\mathbf{H}$ is a large $n_x n_y n_t \times n_x n_y n_t$ matrix. Because of the adjoint computation step of $\mathbf{H}^T\mathbf{r}$ at each iteration, the number of flops of a single iteration of Algorithm 1 is at least $(n_x n_y n_t)^2$. Given that practical LIDAR image resolution can go up to 1 megapixels, computational time depending at least quadratically on $n_x n_y$ is too large for applications that require fast processing. Thus, we propose an alternative depth super-resolution imaging framework that addresses the above two points: an algorithm that incorporates spatial priors and that is computationally efficient.

We first define a new variable $\mathbf{C} \in \mathbb{N}^{n_x \times n_y}$ to be the total photon count matrix: $\mathbf{C}_{i,j} = \sum_{k=1}^{n_t} \mathbf{Y}_{i,j,k}$. Because the sum of independent Poisson random variables is a Poisson random variable with summed rates, we have

$$\boldsymbol{C} \sim \mathrm{Poisson}(\mathbf{h}_{xy} * \bar{\mathbf{A}} + \mathbf{B}'), \tag{6}$$

where $\mathbf{B}'_{i,j} = b\Delta n_t$. We also define the variable $\mathbf{P} \in \mathbb{R}^{n_x \times n_y}$, that for each pixel is a dot product between the histogram data with a linear function:

$$\mathbf{P}_{i,j} = \sum_{k=1}^{n_t} (k\mathbf{Y}_{i,j,k}). \tag{7}$$

One can intuitively see that this is a way of linearly encoding depth values in a single value. Previously, this histogram sum trick has been exploited in the context of compressed depth imaging.[33]

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 3/16/2018 Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

Algorithm 2 Regularized ML reconstruction by convex optimization

Input: $\mathbf{Y}$, $\mathbf{h}_{xy}$, b
Output: $\hat{\mathbf{A}}^{\mathrm{RML}}$, $\hat{\mathbf{T}}^{\mathrm{RML}}$
$\mathbf{C}_{i,j} \to \sum_{k=1}^{n_t} \mathbf{Y}_{i,j,k}$, $\forall i, j$;
$\mathbf{P}_{i,j} \to \sum_{k=1}^{n_t} (k\mathbf{Y}_{i,j,k})$, $\forall i, j$;
$\hat{\mathbf{A}}^{\mathrm{RML}} \leftarrow$ solution to (10) using inputs $(\mathbf{C}, \mathbf{h}_{xy}, \mathbf{B}')$
$\hat{\mathbf{T}}^{\mathrm{RML}} \leftarrow$ solution to (11) using inputs $(\mathbf{P}, \mathbf{h}_{xy}, \hat{\mathbf{A}}^{\mathrm{RML}}, \mathbf{B}'')$

Recall that we defined $\bar{\mathbf{T}}_{i,j} \in \{1, \ldots, n_t\}^{n_x \times n_y}$ to represent the non-zero index of $\bar{\mathbf{X}}_{i,j}$ at each pixel; from $\bar{\mathbf{T}}$, one can obtain $\bar{\mathbf{D}}$ by scaling each entry by $(c\Delta)/2$. Define $\circ$ to be the entrywise product, also known as the Hadamard product, between matrices. We have

$$\mathbb{E}[\mathbf{P}_{i,j}] = (\mathbf{h}_{xy} * (\bar{\mathbf{A}} \circ \bar{\mathbf{T}}))_{i,j} + \mathbf{B}''_{i,j}, \tag{8}$$

where $\mathbf{B}''$ is a constant matrix of size $n_x \times n_y$ with all entries equal to $b\Delta n_t(n_t+1)/2$. We use the assumption that $\mathbf{P}_{i,j}$ is Poisson distributed as a heuristic and write the approximate observation model

$$\mathbf{P} \sim \mathrm{Poisson}(\mathbf{h}_{xy} * (\bar{\mathbf{A}} \circ \bar{\mathbf{T}}) + \mathbf{B}''). \tag{9}$$

Let $\mathcal{L}_A(\mathbf{A}; \mathbf{C}, \mathbf{h}_{xy}, \mathbf{B}')$ be the negative log-likelihood function derived using (6). Then, we obtain the regularized maximum-likelihood intensity solution $\hat{\mathbf{A}}^{\mathrm{RML}}$ by solving the following optimization problem:

$$\begin{aligned} \underset{\mathbf{A}}{\text{minimize}} \quad & \mathcal{L}_A(\mathbf{A}; \mathbf{C}, \mathbf{h}_{xy}, \mathbf{B}') + \beta_A \, \mathrm{pen}_A(\mathbf{A}) \\ \text{subject to} \quad & \mathbf{A}_{i,j} \geq 0, \ \forall i, j, \end{aligned} \tag{10}$$

where $\mathrm{pen}_A(\mathbf{A})$ is a regularization function enforcing smoothness in the intensity solution and β_A is a positive scalar parameter controlling the degree of regularization. One choice of $\mathrm{pen}_A(\mathbf{A})$ is the TV norm. The choice of the regularizer is application-dependent; for biological imaging of sparsely located cells, one may use the ℓ_1-norm instead. It is straightforward to show that $\mathcal{L}_A(\mathbf{A}; \mathbf{C}, \mathbf{h}_{xy}, \mathbf{B}')$ and $\|\mathbf{A}\|_{\mathrm{TV}}$ are convex functions in $\mathbf{A}$, and thus (10) is a convex optimization problem that can be solved efficiently.[34]

To estimate $\bar{\mathbf{T}}$, and thus $\bar{\mathbf{D}}$, from $\mathbf{P}$, we must solve an inverse problem of deconvolution of $\mathbf{h}_{xy}$ and entrywise scaling in $\bar{\mathbf{A}}$. Because the ground truth $\bar{\mathbf{A}}$ is not available, we use $\hat{\mathbf{A}}^{\mathrm{RML}}$ obtained by solving (10) as a plug-in estimator. Let $\mathcal{L}_T(\mathbf{T}; \mathbf{P}, \mathbf{h}_{xy}, \hat{\mathbf{A}}^{\mathrm{RML}}, \mathbf{B}'')$ be the negative log-likelihood function of the scaled depth image derived from (9). We solve for the regularized ML depth solution $\hat{\mathbf{T}}^{\mathrm{RML}}$ using the following optimization program:

$$\begin{aligned} \underset{\mathbf{T}}{\text{minimize}} \quad & \mathcal{L}_T(\mathbf{T}; \mathbf{P}, \mathbf{h}_{xy}, \hat{\mathbf{A}}^{\mathrm{RML}}, \mathbf{B}'') + \beta_T \, \mathrm{pen}_T(\mathbf{T}) \\ \text{subject to} \quad & \mathbf{T}_{i,j} \geq 0, \ \forall i, j, \end{aligned} \tag{11}$$

where, similar to the optimization program for intensity, $\mathrm{pen}_T(\mathbf{T})$ is a regularization function enforcing smoothness in the depth solution and β_t is a positive scalar parameter controlling the degree of regularization. We can show that $\mathcal{L}_T(\mathbf{T}; \mathbf{P}, \mathbf{h}_{xy}, \hat{\mathbf{A}}^{\mathrm{RML}}, \mathbf{B}'')$ is a convex function in $\mathbf{T}$, and thus (11) is a convex optimization problem. Algorithm 2 summarizes the steps of our regularized ML intensity and depth reconstruction framework.

5. SIMULATIONS

To validate our framework, we perform simulations of the photon-counting depth imaging setup. In Fig. 1, we show simulation results of reconstructing the intensity and depth of the MIT logo scene. The resolution of the simulated MIT logo image was 19×19 and thus we had $n = 361$. The ground truth intensity and depth pair is shown in Fig. 1(a,h). In our simulations, the number of detector time bins was $n_t = 10$. The spatial kernel was set to be a 2D symmetric Gaussian with variance 3 pixels, as illustrated in Fig. 1(o). We assumed that the pulsewidth T_p is much smaller than the detector time bin Δ and set $\mathbf{H}_t$ to be identity. The mean number of

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 3/16/2018 Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

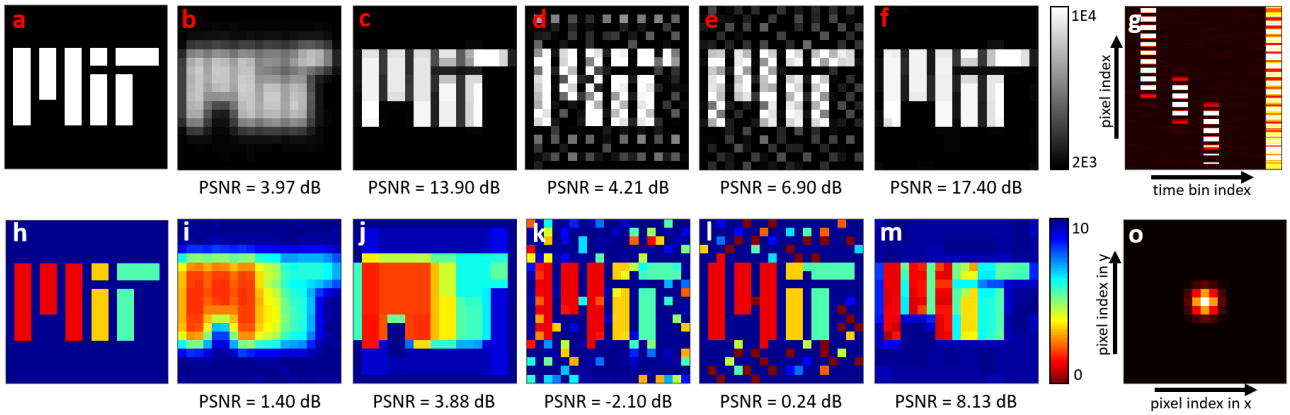

Figure 1: Intensity and depth reconstruction results with PSNR values for a simulated 19×19 MIT logo scene with a Gaussian kernel. (a,h) Ground truth $\bar{\mathbf{A}}, \bar{\mathbf{D}}$. (b,i) Conventional pixelwise ML estimates. (c,j) Gaussian deconvolution as post-processing of pixelwise ML estimates. (d,k) Richardson–Lucy spatiotemporal deconvolution as pre-processing for pixelwise ML estimation. (e,l) $\hat{\mathbf{A}}^{\mathrm{ML}}, \hat{\mathbf{D}}^{\mathrm{ML}}$ from proposed Algorithm 1. (f,m) $\hat{\mathbf{A}}^{\mathrm{RML}}, \hat{\mathbf{D}}^{\mathrm{RML}}$ from proposed Algorithm 2. (g) Visualization of $f(\mathbf{Y})$. (o) The Gaussian kernel $\mathbf{h}_{xy}$.

detected photons per pixel was 4.5×10^3, and the signal-to-background ratio (SBR) was 5 (so the mean number of detected signal photons per pixel was 3.7×10^3). This is not a low number of photons per pixel comparable to recent works in which deconvolution is not needed,[14–17] but the results suggest that it is low enough that our explicit Poissonian modeling of observed photon counts improves performance. The flattened photon data cube $f(\mathbf{Y}) \in \mathbb{N}^{n_x n_y \times n_t}$ is illustrated in image form in Fig. 1(g). Note how the three vertical slices representing the letters MIT overlap in the spatial domain due to the non-trivial forward kernel.

Fig. 1(b,i) shows the intensity and depth results from the straightforward pixelwise ML approach. As expected, the image features are blurry and overestimated in size. By performing TV-regularized deconvolution on these, we obtain the results in Fig. 1(c,j). By performing the spatiotemporal data deconvolution as a pre-processing step, we obtain the results in Fig. 1(d,k). Here, due to photon noise, the resulting images are also noisy but some finer object features start to emerge, especially in depth. Fig. 1(e,l) shows the intensity and depth reconstruction results $\hat{\mathbf{A}}^{\mathrm{ML}}, \hat{\mathbf{D}}^{\mathrm{ML}}$ from our Algorithm 1, and we see that there is only a mild improvement compared to Fig. 1(d,k), since Algorithm 1 relies only on the single-reflector-per-pixel constraint and not the image smoothness constraint. Fig. 1(f,m) shows the intensity and depth pair $\hat{\mathbf{A}}^{\mathrm{RML}}, \hat{\mathbf{D}}^{\mathrm{RML}}$ from our Algorithm 2, which regularizes for natural image smoothness. Compared to previous results, here we see the noise has been suppressed while the fine features are preserved for both intensity and depth. Although the depth values at the boundaries of the letter features have been under-estimated due to the 1-pixel separation between each pair of letters, our intensity and depth results in Fig. 1(f,m) has the highest PSNR values among all results.

ACKNOWLEDGMENTS

This work was supported in part by a Samsung Scholarship and by the U.S. National Science Foundation under Grant 1161413 and Grant 1422034.

REFERENCES

[1] I. J. Cox, S. L. Hingorani, S. B. Rao, and B. M. Maggs, "A maximum likelihood stereo algorithm," *Computer Vision and Image Understanding* **63**, pp. 542–567, May 1996.

[2] Z. Zhang, "Microsoft Kinect sensor and its effect," *IEEE MultiMedia* **19**, pp. 4–10, Apr.–June 2012.

[3] S. B. Gokturk, H. Yalcin, and C. Bamji, "A time-of-flight depth sensor – system description, issues and solutions," in *IEEE Conf. Computer Vision and Pattern Recognition Workshop*, June–July 2004.

[4] B. Schwarz, "Lidar: Mapping the world in 3D," *Nat. Photon.* **4**(7), pp. 429–430, 2010.

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 3/16/2018 Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

[5] M. Raibert, K. Blankespoor, G. Nelson, and R. Playter, "BigDog, the rough-terrain quadruped robot," *IFAC Proceedings Volumes* **41**(2), pp. 10822–10825, 2008.

[6] G. C. Guenther, A. G. Cunningham, P. E. LaRocque, and D. J. Reid, "Meeting the accuracy challenge in airborne lidar bathymetry," in *Proc. EARSeL-SIG-Workshop LIDAR*, (Dresden, FRG), June 2000.

[7] R. A. Haugerud, D. J. Harding, S. Y. Johnson, J. L. Harless, C. S. Weaver, and B. L. Sherrod, "High-resolution lidar topography of the Puget Lowland, Washington," *GSA Today* **13**(6), pp. 4–10, 2003.

[8] M. Levoy, K. Pulli, B. Curless, S. Rusinkiewicz, D. Koller, L. Pereira, M. Ginzton, S. Anderson, J. Davis, J. Ginsberg, *et al.*, "The digital Michelangelo project: 3d scanning of large statues," in *Proc. 27th Ann. Conf. Computer Graphics and Interactive Techniques*, pp. 131–144, 2000.

[9] 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.* **13**(2), pp. 335–349, 2002.

[10] P. Buzhan, B. Dolgoshein, L. Filatov, A. Ilyin, V. Kantzerov, V. Kaplin, A. Karakash, F. Kayumov, S. Klemin, E. Popova, and S. Smirnov, "Silicon photomultiplier and its possible applications," *Nucl. Instr. Meth. Phys. Res. Sect. A* **504**(1), pp. 48–52, 2003.

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

[12] R. Blahut, *Remote Image Formation*, Elsevier, Amsterdam, 2012.

[13] F. Crosby, H. Zhou, and Q. Huynh, "Total variation methods for three dimensional lidar image denoising," *Photogrammetric Engineering & Remote Sensing* **76**(12), pp. 1365–1371, 2010.

[14] Y. Altmann, X. Ren, A. McCarthy, G. S. Buller, and S. McLaughlin, "Lidar waveform-based analysis of depth images constructed using sparse single-photon data," *IEEE Trans. Image Process.* **25**, pp. 1935–1946, May 2015.

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

[16] D. Shin, A. Kirmani, V. K. Goyal, and J. H. Shapiro, "Photon-efficient computational 3D and reflectivity imaging with single-photon detectors," *IEEE Trans. Comput. Imaging* **1**, pp. 112–125, June 2015.

[17] J. Rapp and V. K. Goyal, "A few photons among many: Unmixing signal and noise for photon-efficient active imaging," *IEEE Trans. Comput. Imaging* **3**, 2017. To appear.

[18] J. Bertolotti, E. G. van Putten, C. Blum, A. Lagendijk, W. L. Vos, and A. P. Mosk, "Non-invasive imaging through opaque scattering layers," *Nature* **491**(7423), pp. 232–234, 2012.

[19] O. Katz, E. Small, and Y. Silberberg, "Looking around corners and through thin turbid layers in real time with scattered incoherent light," *Nat. Photon.* **6**(8), pp. 549–553, 2012.

[20] M. J. Olsen, G. V. Roe, C. Glennie, F. Persi, M. Reedy, D. Hurwitz, K. Williams, H. Tuss, A. Squellati, and M. Knodler, *Guidelines for the Use of Mobile LIDAR in Transportation Applications*, vol. 748, Transportation Research Board, 2013.

[21] R. Stettner, "Compact 3D flash lidar video cameras and applications," in *SPIE Laser Radar Technology and Applications XV*, **7684**, 2010.

[22] D. Shin, F. Xu, D. Venkatraman, R. Lussana, F. Villa, F. Zappa, V. K. Goyal, F. N. C. Wong, and J. H. Shapiro, "Photon-efficient imaging with a single-photon camera," *Nat. Commun.* **7**(12046), 24 Jun 2016.

[23] V. N. Mahajan, *Aberration Theory Made Simple*, SPIE Press, Bellingham, WA, second ed., 2011.

[24] Q. Yang, R. Yang, J. Davis, and D. Nistér, "Spatial-depth super resolution for range images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, pp. 1–8, 2007.

[25] S. Schuon, C. Theobalt, J. Davis, and S. Thrun, "LidarBoost: Depth superresolution for ToF 3D shape scanning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, pp. 343–350, 2009.

[26] L. Xiao, F. Heide, M. O'Toole, A. Kolb, M. B. Hullin, K. Kutulakos, and W. Heidrich, "Defocus deblurring and superresolution for time-of-flight depth cameras," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, pp. 2376–2384, 2015.

[27] J. P. Godbaz, M. J. Cree, and A. A. Dorrington, "Blind deconvolution of depth-of-field limited full-field lidar data by determination of focal parameters," in *Proc. SPIE Computational Imaging VII*, **7533**, 2010.

[28] J. R. McMahon, R. K. Martin, and S. C. Cain, "Three-dimensional FLASH laser radar range estimation via blind deconvolution," *J. Appl. Remote Sens.* **4**(1), 2010.

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 3/16/2018 Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

[29] D. Needell and J. A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Appl. Comput. Harmon. Anal.* **26**(3), pp. 301–321, 2009.

[30] J. M. Bioucas-Dias, M. A. Figueiredo, and J. P. Oliveira, "Total variation-based image deconvolution: a majorization-minimization approach," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Process.*, **2**, pp. 861–864, 2006.

[31] W. H. Richardson, "Bayesian-based iterative method of image restoration," *JOSA* **62**(1), pp. 55–59, 1972.

[32] D. Shin, J. H. Shapiro, and V. K. Goyal, "Single-photon depth imaging using a union-of-subspaces model," *IEEE Signal Process. Lett.* **22**, pp. 2254–2258, Dec. 2015.

[33] G. A. Howland, D. J. Lum, M. R. Ware, and J. C. Howell, "Photon counting compressive depth mapping," *Opt. Expr.* **21**(20), pp. 23822–23837, 2013.

[34] 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.* **21**(3), pp. 1084–1096, 2012.

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 3/16/2018 Terms of Use: https://www.spiedigitallibrary.org/terms-of-use