Computational Imaging with Small Numbers of Photons

by

Dongeek Shin

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Electrical Engineering and Computer Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2016

© Massachusetts Institute of Technology 2016. All rights reserved.

Author . . . .
Department of Electrical Engineering and Computer Science
January 29, 2016

Certified by . . . .
Jeffrey H. Shapiro
Julius A. Stratton Professor, Massachusetts Institute of Technology
Thesis Supervisor

Certified by . . . .
Vivek K Goyal
Associate Professor, Boston University
Thesis Supervisor

Accepted by . . . .
Leslie A. Kolodziejski
Chair, Department Committee on Graduate Theses

Computational Imaging with Small Numbers of Photons

by

Dongeek Shin

Submitted to the Department of Electrical Engineering and Computer Science
on January 29, 2016, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Electrical Engineering and Computer Science

Abstract

The ability of an active imaging system to accurately reconstruct scene properties in low light-level conditions has wide-ranging applications, spanning biological imaging of delicate samples to long-range remote sensing. Conventionally, even with time-resolved detectors that are sensitive to individual photons, obtaining accurate images requires hundreds of photon detections at each pixel to mitigate the shot noise inherent in photon-counting optical sensors.

In this thesis, we develop computational imaging frameworks that allow accurate reconstruction of scene properties using small numbers of photons. These frameworks first model the statistics of individual photon detections, which are observations of an inhomogeneous Poisson process, and express a priori scene constraints for the specific imaging problem. Each yields an inverse problem that can be accurately solved using novel variations on sparse signal pursuit methods and regularized convex optimization techniques. We demonstrate our frameworks' photon efficiencies in six imaging scenarios that have been well-studied in the classical settings with large numbers of photon detections: single-depth imaging, multi-depth imaging, array-based time-resolved imaging, super-resolution imaging, single-pixel imaging, and fluorescence imaging. Using simulations and experimental datasets, we show that our frameworks outperform conventional imagers that use more naive observation models based on high light-level assumptions. For example, when imaging depth, reflectivity, or fluorescence lifetime, our implementation gives accurate reconstruction results even when the average number of detected signal photons at a pixel is less than 1, in the presence of extraneous background light.

Thesis Supervisor: Jeffrey H. Shapiro
Title: Julius A. Stratton Professor, Massachusetts Institute of Technology

Thesis Supervisor: Vivek K Goyal
Title: Associate Professor, Boston University

Acknowledgments

I would like to first thank my advisors Jeffrey Shapiro and Vivek Goyal for giving me opportunites to grow as an independent researcher during my doctoral program years. They both were extremely supportive of me exploring dark places, handing me the light only when it seemed necessary, and gave me time to independently come up with interesting research ideas. From many discussions I had with Jeff, I learned the process of shaping rough research ideas to a well-defined interesting problem that can be worked on. Vivek gave me the insight on how to classify good research problems and really going for the ones with high positive impact in the community. Jeff and Vivek both were always open to hearing and talking about any type of excitement, doubt, or concern I bring, related to academics or not, and I thank them again for that support. I thank my committee members Franco Wong and Bill Freeman for appreciating the thesis material and for their valuable feedback that improved the thesis as a whole.

I thank the past and present members of Signal Transformation and Information Representation Group and the Optical and Quantum Communications Group at the Research Laboratory of Electronics for creating a friendly environment and for always allowing me to discuss very rough research ideas with them. Especially, I would like to thank Feihu Xu and Dheera Venkatraman for leading the experimental side of the low-light imaging projects, and Ahmed Kirmani and Andrea Colaço for their work on the fundamental aspects of low-light imaging that I built my thesis work upon.

I also thank Rudi Lussana, Federica Villa, and Franco Zappa at the Polytechnic University of Milan for providing the MIT team with the single-photon camera equipment and Rudi for being at MIT to help us with the initial setup. I would like to thank the Bawendi Group as well for providing us with the samples that were used to generate preliminary fluorescence imaging results.

I thank Greg Wornell, for giving me the opportunity to be a teaching assistant for his inference course and letting me learn how teaching can inspire one in many other ways that working on a research problem cannot. I also would like to thank the

members of his Signals, Inference, and Algorithms Group for many valuable feedback.

Finally, I would like to dedicate this thesis to my family for their never-ending love and support.

This research was supported in part by NSF grant No. 1422034, 1161413, and a Samsung Scholarship.

Contents

1 Introduction **27**

1.1 A Unifying Viewpoint 29

2 Single-Reflector Depth Imaging **33**

2.1 Overview of Problem 33

2.2 Single-Photon Imaging Setup 35

2.3 Forward Imaging Model 36

2.4 Solving the Inverse Problem 39

2.5 Results 42

2.6 Summary and Discussion 47

3 Multi-Depth Imaging **49**

3.1 Overview of Problem 49

3.2 Single-Photon Imaging Setup 52

3.3 Forward Imaging Model 52

3.4 Solving the Inverse Problem 56

3.5 Results 58

3.6 Summary and Discussion 69

4 Array-Based Time-Resolved Imaging **73**

4.1 Overview of Problem 73

4.2 Single-Photon Imaging Setup 74

4.3 Forward Imaging Model 76

4.4 Solving the Inverse Problem 78
4.5 Results 83
4.6 Summary and Discussion 93

5 Super-Resolution Imaging 97
5.1 Overview of Problem 97
5.2 Single-Photon Imaging Setup 100
5.3 Forward Imaging Model 100
5.4 Solving the Inverse Problem 103
5.5 Results 111
5.6 Summary and Discussion 122

6 Single-Pixel Imaging 123
6.1 Overview of Problem 123
6.2 Single-Photon Imaging Setup 125
6.3 Forward Imaging Model 125
6.4 Solving the Inverse Problem 126
6.5 Results 132
6.6 Summary and Discussion 135

7 Fluorescence Imaging 141
7.1 Overview of Problem 141
7.2 Single-Photon Imaging Setup 143
7.3 Forward Imaging Model 145
7.4 Solving the Inverse Problem 147
7.5 Results 151
7.6 Summary and Discussion 152

8 Conclusions and Final Remarks 155

A Table of Notations 159

B Dynamic Range of Single-Photon Detectors 161

C Convexity of Negative Log Poisson Likelihood 169

D Approximation of Poisson Likelihood for Single-Photon Depth Imaging 171

E Operation of SPAD Detector Array 173

F SNR of the Histogram Sum Variable 181

G Asymptotic Efficiency of Least-Squares Single-Pixel Imaging 185

List of Figures

1-1 Single-photon active imaging setup with common components. The figure specifically depicts a setup with scanning source and a single-pixel detector. 30

2-1 An illustration of the data acquisition procedure for one illumination pulse. A pulsed optical source illuminates a scene pixel with photon-flux waveform $s(t)$. The flux waveform $r(t)$ that is incident on the detector consists of the pixel return $as(t-2d/c)$—where a is the pixel reflectivity, d is the pixel depth, and c is light speed—plus the background-light flux b. The rate function $\lambda(t)$ driving the photodetection process equals the sum of the pixel return and background flux, scaled by the detector efficiency η, plus the detector's dark-count rate b_d. The record of detection times from the pixel return (or background light plus dark counts) is shown as blue (or red) spikes, generated by the Poisson process driven by $\lambda(t)$. 36

2-2 Illustration of $\mathcal{S}_2$. Our non-convex union-of-subspaces constraint set describes model-based signal sparsity that is specific to the LIDAR imaging setup. 39

2-3 Simulation of pixelwise depth imaging results using single photon observations for two SBR levels. (a) Depth RMSE of log-matched filter and proposed method for SBR = 200. (b) Depth RMSE of log-matched filter and proposed method for SBR = 100. 43

2-4 Experimental setup with a raster-scanning source and a single SPAD detector. The red arrows show the path of the optical signal from laser source, and the green arrows show the path of the electrical signal that indicates whether a photon has been detected or not. 44

2-5 Experimental pixelwise depth imaging results using single photon observations. The number of photon detections at every pixel was set to be 15. The figure shows the (a) photograph of imaged face, (b) ground-truth depth, (c) depth from log-matched filtering, which is approximately ML and (d) depth using our method. The absolute depth-error maps for ML and our framework are shown in (e) and (f), respectively. 45

2-6 Depth recovery performance of our algorithm at a face pixel (SBR = 6.7 and pixel coordinates $(81, 272)$) and a depth boundary pixel (SBR = 1.5 and pixel coordinates $(237, 278)$) for varying number of photon detections. 46

3-1 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. 50

3-2 (Top) Full-waveform single-photon imaging setup for depth estimation of multiple objects. In this example, a pulsed optical source illuminates a scene pixel 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_s-pulse rate function $N_s\lambda(t)$ 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. 53

3-3 Illustration of a shrinkage-thresholding operation used as a step in ISTA (left) and the shrinkage-rectification operation used as a step in our SPISTA (right) that includes the nonnegativity constraint. Here the operations map scalar v to scalar z (variables only used for illustration purposes), with regularization parameter τ. 57

3-4 Illustration of steps of Algorithm 3 using experimental photon-count data for the single-pixel multi-depth example of partially-occluding reflectors in Fig. 3-9. (a) The raw photon count vector $\mathbf{y}$ from a pixel that contains two reflectors at around time bins 2600 and 3500. Other than the photon detections describing the two targets of interest, we observe extraneous photon detections from background and dark counts. (b) The output solution of SPISTA in Algorithm 2. Note that the extraneous background and detector dark counts are suppressed. (c) The final solution of Algorithm 3 that groups depths of SPISTA output. 60

3-5 Simulated performance of MoG-based method and proposed framework of Algorithm 3 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 corresponds to an unnormalized mean-squared error of $cT_p/2 = 4.5$ cm. The plots also include error bars indicating the ± 1 standard errors. 61

3-6 Simulated results of mean estimates of the number of reflectors produced by Algorithm 2 at a pixel, when the RMS pulsewidth is set to give $cT_p = 2$ cm. Here we show plots when the reflectivity ratio between the first and second target is (a) 1 (blue line), (b) 1/2 (sky blue line), (c) 1/4 (yellow line), and (d) 1/8 (red line). 63

3-7 For different values of signal amplitude a (reflectivity multiplied by the peak value of pulse waveform) and regularization parameter τ, the plots show success-1 in black and success-2 in red. Note that, by definition, success-1 upper-bounds success-2. 64

3-8 (Left) Photograph of the mannequin placed behind a partially-scattering object from the single-photon imager's point of view. (Right) Experimental results for estimating the mannequin's depth map through the partially-reflective material using MoG-based and our estimators, given that our imaging setup is at $z = 0$. The EM algorithm for MoG fitting used $K = 2$. Our multi-depth results were generated using the parameters $\tau = 0.1$, $\delta = 10^{-4}$, $\epsilon = 0.1$, and $\hat{\mathbf{x}}^{(0)} = \mathbf{S}^T\mathbf{y}$. 67

3-9 Single-photon imaging setup for estimating multi-depth from partial occlusions at depth boundary pixels. Sample data from 38 photon detections is shown below for the pixel $(94, 230)$ where partial occlusions occur. We show experimental results of multi-depth recovery for this scene using the MoG-based and our methods in Figure 3-10. 70

3-10 Experimental results of depth reconstruction for a 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. The EM algorithm for MoG fitting used $K = 2$. Our multi-depth results were generated using the parameters $\tau = 0.2$, $\delta = 10^{-4}$, $\epsilon = 0.1$, and $\hat{\mathbf{x}}^{(0)} = \mathbf{S}^T\mathbf{y}$. 71

4-1 Single-photon array imaging framework. (a) SPAD-array imaging setup. A repetitively-pulsed laser flood-illuminates the scene of interest. Laser light reflected from the scene plus background light is detected by a SPAD camera. Photon detections at each pixel are time tagged relative to the most recently transmitted pulse and recorded. The raw photon-detection data were processed on a standard laptop computer to recover the scene's 3D structure and reflectivity. (b) Example of 3D structure and reflectivity reconstruction of mannequin and flower scene using the baseline single-photon imager from [1]. (c) Example of 3D structure and reflectivity reconstruction of mannequin and flower scene from our processing. Large portions of the mannequin's shirt and facial features that were not visible in the baseline image are revealed using our method. Both results in (b) and (c) were generated using an average of ∼1 detected signal photon per pixel. 75

4-2 Photon-count histograms after censoring extraneous detections for the state-of-the-art pseudo-array method [2] (red) and our method (blue). The raw-data photon histogram is also given for reference (dashed black). The blue block indicates the ground truth depth values of objects in the scene scaled by $c/2$. By exploiting the scene's longitudinal sparsity, our method rejects more extraneous detections than does the pseudo-array method, which relies on transverse correlations. The greater the number of extraneous detections that survive censoring, the greater the amount of regularization that will occur in depth estimation, which will lead, in turn, to oversmoothing the depth image. 82

4-3 Stages of 3D structure and reflectivity reconstruction algorithm. (a) Raw time-tagged photon detection data are captured using the SPAD camera setup. Averaged over the scene, the number of detected signal photons per pixel was $\sim$1, as was the average number of background-light detections plus dark counts. (b) Step 1: raw time-tagged photon detections are used to accurately estimate the scene's reflectivity by solving a regularized optimization problem. (c) Step 2: to estimate 3D structure, extraneous (background-light plus dark-count) photon detections are first censored, based on the longitudinal sparsity constraint of natural scenes, by solving a sparse deconvolution problem. (d) Step 3: the uncensored (presumed to be signal) photon detections are used for 3D structure reconstruction, by solving a regularized optimization problem. 84

4-4 3D structure and reflectivity reconstructions of the mannequin and flower scene. (a)–(d) Results of imaging 3D structure and reflectivity using the filtered histogram method, the state-of-the-art pseudo-array imaging method, our proposed framework, and the ground-truth proxy obtained from detecting 550 signal photons per pixel. For visualization, the reflectivity estimates are overlaid on the reconstructed depth maps for each method. The frontal views, shown here, provide the best visualizations of the reflectivity estimates. (e)–(h) Results of imaging 3D structure and reflectivity from (a)–(d) rotated to reveal the side view, which makes the reconstructed depth clearly visible. The filtered histogram image is too noisy to show any useful depth features. The pseudo-array imaging method successfully recovers gross depth features, but, in comparison with the ground truth estimate in (h), it overestimates the dimensions of the mannequin's face by several cm and oversmooths the facial features. Our SPAD-array-specific method in (g), however, gives high-resolution depth and reflectivity reconstruction at low flux. (i)–(k) The depth error maps obtained by taking the absolute difference between estimated depth and ground truth depth show that our method successfully recovers the scene structure with sub-pulse-width resolution of less than $c\Delta/2 \approx 6$ cm, while existing methods fail to do so. 85

4-5 Effect of varying the regularization parameters in our 3D structure and reflectivity reconstruction algorithm for the mannequin and flower scene. The optimal parameter set was $\{\tau_A, \tau_Z\} = \{2.6, 4.3\}$. 88

4-6 Imaging results for the watering can and basketball scene. Notice the stripes of the basketball being visible when using our method. 89

4-7 Calibration results for the SPAD camera and scene parameters. (a) A 32×32 image of photon counts when the SPAD camera observes a weakly-reflecting planar wall with the laser off (left) was used to generate a $32{\times}32$ binary mask indicating hot-pixel locations with white markers (middle). The hot pixel mask was obtained by thresholding the photon-count image with an appropriate threshold chosen from the photon-count histogram (right). (b) Laser pulse shape. Extraneous photon detections were suppressed by time-gating near the roundtrip delay to a $1 \times 1\,\mathrm{m}$ calibration target. (c) The non-constant 384×384 background-light plus dark-count rate matrix $\mathbf{B}$ for the mannequin and flower scene. 91

4-8 Relationship between RMS pulse duration and depth-recovery accuracy. (a) Plot of depth-recovery accuracy using our algorithm (Step 2 and Step 3) versus RMS pulse duration T_p obtained by simulating a low-flux SPAD imaging environment with an average of 10 detected signal photons per pixel and 390 ps time bins. Depth recovery is deemed a success at a pixel if estimated depth is within 3 cm of ground truth, and the depth recovery accuracy of a method is computed by the mean success rate over all pixels. (b) Ground-truth depth map used in the simulations. (c) Estimated depth map for $T_p = 0.3\,\mathrm{ns}$. (d) Estimated depth map for $T_p = 1.1\,\mathrm{ns}$. (e) Estimated depth map for $T_p = 2.4\,\mathrm{ns}$. When T_p is too short, there is a systematic bias in the estimated depths, although random errors are minimal. When T_p is too long, the estimated depths are very noisy. In all our SPAD-array experiments we used $T_p \approx 1\,\mathrm{ns}$, which is in the sweet spot between durations that are too short or too long. We emphasize, however, that our algorithm is not tuned to a particular pulse-width and can be performed using any T_p value. 94

4-9 Photon efficiency of proposed framework. (a) Plots of RMS depth error (log scale) versus average number of detected signal photons per pixel for imaging the mannequin's face using our proposed framework and the baseline pixelwise processor. Our method consistently realizes sub-pulse-width performance throughout the low-flux region shown in the plot, whereas the baseline approach's accuracy is more than an order of magnitude worse, owing to its inability to cope with extraneous detections. (b) Summary of photon efficiency versus acquisition speed (not the computational speed) for existing 3D structure and reflectivity imagers, where fps denotes frames per second and ppp denotes photons per pixel. 95

5-1 (a) Scanning LIDAR setup. A repetitively pulsed source illuminates the scene in a scanning manner, where at each scan point multiple reflectors at different depths may be illuminated. The detector records the photon-count histogram of the backreflected response. (b) From one scanning pixel to another, there can be overlap in illumination due to physical constraints. (c) The interpixel overlaps can come from either illumination non-idealities, such as its finite beam-width (top), or scene constraints such as the presence of strongly scattering media that generate random speckle illumination patterns (bottom). We define the two-dimensional intensity pattern cast on the scene as the transverse imaging kernel in our problem, and denote it using the two-dimensional function h centered at $(0,0)$. 101

5-2 Illustration of the acquisition model when imaging a one-dimensional scene. The PSF from scene illumination is modeled by $\mathbf{H}$ and the temporal pulse waveform and detector response is modeled by $\mathbf{S}$. Both $\mathbf{H}$ and $\mathbf{S}$ are Gaussian convolution matrices in this example. $\mathbf{B}$ models the extraneous background and detector dark count response. The set of photon-count histograms observed at each scan pixel after N_s illumination trials is represented by $\mathbf{Y}$. 103

5-3 Illustration of $\mathcal{S}_1(N, m)$ for $m = 3$. Since every row of $\mathbf{X} \in \mathcal{S}_1(N, m)$ belongs to a set of 1-sparse signals, our constraint set $\mathcal{S}_1(N, m)$ is a product space of N of 1-sparse signal sets. 104

5-4 Reflectivity reconstruction results (with mean absolute errors (MAE)) for a 19×19 MIT logo scene when SBR $= +\infty$. We compare the pixelwise ML, deconvolved ML, and the proposed frameworks (Approach 1 and 2) for different PSFs and values of signal photons per pixel (sppp). (b–e) Using Gaussian PSF and sppp $= 10^6$. (f–i) Using Gaussian PSF and sppp $= 10^4$. (j–m) Using Bernoulli PSF and sppp $= 10^3$. (n–q) Using Bernoulli PSF and sppp $= 10^2$. 114

5-5 Depth reconstruction results (with mean absolute errors (MAE)) for a 19×19 MIT logo scene when SBR $= +\infty$. We compare the pixelwise ML, deconvolved ML, for different PSFs and values of signal photons per pixel (sppp). (b–e) Using Gaussian PSF and sppp $= 10^6$. (f–i) Using Gaussian PSF and sppp $= 10^4$. (j–m) Using Bernoulli PSF and sppp $= 10^3$. (n–q) Using Bernoulli PSF and sppp $= 10^2$. 115

5-6 Reflectivity reconstruction results (with mean absolute errors (MAE)) for a 19×19 MIT logo scene when SBR $= 1$. We compare the pixelwise ML, deconvolved ML, and the proposed frameworks (Approach 1 and 2) for different PSFs and values of signal photons per pixel (sppp). (b–e) Using Gaussian PSF and sppp $= 10^6$. (f–i) Using Gaussian PSF and sppp $= 10^4$. (j–m) Using Bernoulli PSF and sppp $= 10^3$. (n–q) Using Bernoulli PSF and sppp $= 10^2$. 116

5-7 Depth reconstruction results (with mean absolute errors (MAE)) for a 19×19 MIT logo scene when SBR $= 1$. We compare the pixelwise ML, deconvolved ML, and the proposed frameworks (Approach 1 and 2) for different PSFs and values of signal photons per pixel (sppp). (b–e) Using Gaussian PSF and sppp $= 10^6$. (f–i) Using Gaussian PSF and sppp $= 10^4$. (j–m) Using Bernoulli PSF and sppp $= 10^3$. (n–q) Using Bernoulli PSF and sppp $= 10^2$. 117

5-8 Reflectivity and depth reconstruction results for a simulated ultra low-light LIDAR setup with MIT logo scene for SBR $= +\infty$, with high native resolution of 270×275 (thus, $N = 74250$). Here, the mean number of photons per pixel was 9.8. We compare the pixelwise ML, ML after Richardson-Lucy (RL) deconvolution, and results of our Approach 2 for both reflectivity and depth recovery. Here $d_{\max}$ is 18 cm. 120

5-9 Reflectivity and depth reconstruction results for a simulated ultra low-light LIDAR setup with MIT logo scene for SBR $= 1$, with high native resolution of 270×275 (thus, $N = 74250$). Here, the mean number of photons per pixel was 9.8. We compare the pixelwise ML, ML after Richardson-Lucy (RL) deconvolution, and results of our Approach 2 for both reflectivity and depth recovery. Here $d_{\max}$ is 18 cm. 121

6-1 Active single-pixel imaging setup. For the kth measurement, the constant flux sent from source ($s(t) = 1/T_r,\ t \in [0, T_r)$) and reflected from $\mathbf{x}$ is spatially-modulated using the pattern $\mathbf{a}_k$, such that the flux incident on the single-photon detector is $\mathbf{a}_k^T\mathbf{x}$. The observation $\mathbf{y}_k \sim \text{Poisson}(\mathbf{a}_k^T\mathbf{x})$ is made by the photodetector, and the process is repeated for M measurements. Although this figure depicts the spatial light modulator as a reflector, such as a programmable mirror device, it may work in transmission mode by using a random diffraction grating instead. 125

6-2 Plots of EDFs $f_{\lambda(\mathbf{W})}(x)$ for different q values according to the Marchenko–Pastur law. 130

6-3 Plots of logarithmic mse (dashed black with 'x' markers), mse-rmt (red with 'o' markers), and mse-baseline (blue with diamond markers) for different values of the Bernoulli probability $p \in \{0.2, 0.5, 0.8\}$ and signal dimension $N \in \{20, 100\}$. Each log-MSE plot is shown over various values of q, the ratio between the number of signal dimensions (N) to the number of observations (M). 133

6-4 Rice single-pixel camera setup. This photograph is taken from the Rice single-pixel camera project website: `http://dsp.rice.edu/cscamera`. 134

6-5 Least-squares single-pixel imaging estimates of letter 'R' for increasing values of M using the Rice single-pixel dataset. Here, $N = 16 \times 16 = 256$ and the ground truth image was generated using the least squares solution with $M = 4290$, since it is an asymptotically efficient estimator. 136

6-6 Comparison of MSE (plot in log scale) estimates for least-squares imaging method on experimental Rice single-pixel camera data of letter 'R'. Here $N = 16 \times 16 = 256$. 137

6-7 Plot of nrmse-rmt over M and P for $N = 50$ and $p = 0.5$. The red line gives the contour for nrmse-rmt $= 0.4$. 138

7-1 Single-photon fluorescence imaging setup as described in Section 7-1 (Top) and an illustration of the photodetection process (Bottom). In this illustration, we have one photon detection (marked in red) resulting from the (i, j)-th pixel's second pulse excitation of the fluorophore sample, given that $N_s = 3$. Note that we define the start of the fluorescence process as time $t = 2d_{i,j}/c$, which is defined by the calibrated sample-to-imager distance. This distance offset in defining the decay signal does not affect the lifetime measurements, as exponential processes are memoryless. 144

7-2 Example large photon count histogram of $\lambda'(t)$ from a pixel of a quantum dot sample with lifetime of $\sim$7 ns, used to visualize the properties of fluorescence signals. The total number of detections contains information about fluorescence intensity and the rate of decay of the histogram contains information about fluorescence lifetime. The residual photon counts uniformly distributed over $[0, 64)$ ns are the extraneous background and dark counts. 146

7-3 Experimental recovery results of (top) fluorescence intensity normalized to have maximum value of 1 and (bottom) lifetime for two different quantum dot scenes with different mean numbers of photons-per-pixel (ppp) using the pixelwise maximum-likelihood estimator and the proposed framework. 152

B-1 Plots of probability function in Eq. (B.6) for different values of a. For increasing a, the probability function deviates away from the normalized rate function $\mathbf{x}$ shown as the dashed black line. 164

B-2 Plots of $\mathbf{x}$ (dashed blue line, labeled as "truth"), $\mathbf{y}/n_s$ (solid black line, labeled as "dead"), and $\hat{\mathbf{x}}_{\text{fix}}$ (solid red line, labeled as "estimate"), which was obtained using Algorithm 7. 166

B-3 Plots of log-RMSE of $\mathbf{y}$ and $\hat{\mathbf{x}}_{\text{fix}}$ for increasing number of photon detections at three different laser illumination powers. 167

E-1 SPAD-camera image acquisition terminology for RS-mode operation. 174

E-2 Experimental SPAD-array imaging setup. 178

E-3 Scanning scheme to obtain increased image area and image resolution using the 32×32-pixel SPAD array. (a) Array translation by increments of a full array size (4.8 mm) along both axes to image multiple tiles of the entire array. (b) Zoom-in view of individual SPAD pixels showing the active area with no sub-pixel scanning, (c) 2×2 sub-pixel scanning by translating along each axis by $75\,\mu$m, hence multiplying resolution by 2 in each dimension. 179

F-1 Depth variation vs. SNR. The solid blue line shows Eq. (F.7) and the dashed red line indicates $\sqrt{N_s}$, which is an upper bound to SNR, for various values of depth-var. 183

G-1 Plots of crb and mse generated using Monte Carlo simulations of Eq. (G.14) and Eq. (G.16). Observe that as M increases, the crb becomes a tighter lower bound for mse. 188

List of Tables

3.1 Processing times for MoG and proposed methods for experimental imaging through a partially-reflecting object. 68

3.2 Processing times for MoG and proposed methods for experimental imaging of a partially-occluding object. 69

Chapter 1

Introduction

Modern imaging systems enable us to better understand our physical environment by providing capabilities that exceed those of the human eye. Active imaging systems in particular use their own light sources and detectors to recover useful scene information, such as 3D structure, object reflectivity, fluorescence, etc. In order to suppress noise inherent in the optical sensing process, they usually require huge amounts of light to form their images. For example, a commercially available flash camera typically collects more than 10^9 photons (10^3 photons per pixel in a 1 megapixel image) to provide the user with a single photograph [3]. However, in sensing a macroscopic scene at a long standoff distance, as well as in imaging of microscopic biological samples, physical limitations on the amount of optical flux available and sensor integration time preclude the collection of such a large number of photons. In those cases, when the light incident on the detector is very low, we must resort to using sensitive single-photon imaging systems that are able to resolve individual photon detections. A key challenge in such scenarios is to make use of a small number of photon detections to accurately recover the desired scene information. Driven by such constraints, it is then natural to ask the following question: how many photons do we really need to form an image?

In this thesis, we address that question by investigating an imaging framework in which small numbers of photon detections are used as raw observations. Mainly, we show how computation and signal processing play an integral role in the photon-efficient image formation process for the following imaging scenarios, which are well-known in the classical high light-level regime.

- **Single-reflector depth imaging:** Pulsed illumination onto an opaque, reflective object leads to backscattered light, which is incident on the imager's detector. A time-resolved sensor is able to record the backreflected optical flux and use its time-of-flight information to infer the object depth on a pixel-by-pixel basis. This is the fundamental problem of light detection and ranging (LIDAR) [4].
- **Multi-depth imaging:** Pulsed illumination of a complicated scene produces reflections from many surfaces in a pixel, generating a complicated temporal profile for the light arriving at the detector. Using a time-resolved sensor, we can observe the combined response of the multi-reflections. Pixelwise reconstruction of multiple object depths by analyzing the combined response is the multi-depth imaging problem [5].
- **Array-based time-resolved imaging:** A detector array used for active scene 3D and reflectivity imaging typically suffers from low temporal resolution, because each detector element in the array can only be engineered to have a moderate sampling rate, as compared to those for single-detector-based scanning imagers. By mitigating this array-specific artifact, we can accurately reconstruct the scene's 3D and reflectivity characteristics from array measurements. This is the problem of array-based time-resolved imaging [6].
- **Single-pixel imaging:** When photodetectors are expensive, such as those operating at infrared or ultraviolet wavelengths or those capable of detecting individual photons, a single-pixel imaging system that reconstructs spatially-resolved reflectivity images using one photodetector, one source, and a spatial light modulator is a desirable imager [7].
- **Super-resolution depth imaging:** Spatial resolution of an active 3D imager is limited by its transverse pixelation degree. Is it possible to improve the spatial

resolution of the reconstructed depth map beyond the native pixel resolution? This is the problem of super-resolution depth imaging [8].

- **Fluorescence imaging:** Fluorescent markers are used to label and track the locations of molecular processes or samples. The aim in fluorescence imaging is to recover spatially-resolved images of the sample fluorescence intensity and lifetime that give information about its molecular properties [9].

1.1 A Unifying Viewpoint

The single-photon active imaging setup has three common components for all of the previously described imaging scenarios. (See Figure 1-1 for an illustration of the setup.)

1. **Source:** The single-photon imaging setup includes a light source that illuminates the scene with a temporally varying waveform, such as a pulse signal. The source can either be a laser that illuminates one pixel at a time as part of a raster-scanning process of the scene, or a floodlight source that sends out light to all pixels simultaneously.
2. **Single-photon detector:** Depending on the imaging application, we can either use a single-pixel photon counting detector or an array of single-photon detectors to record the detection times of individual photons arriving from the scene illuminated by a source.
3. **Optics:** Optical components such as lenses and spatial light modulators are often required to focus and control the light transport, in addition to the source and the single-photon detector.

We can then describe the common single-photon imaging setup employing those three components as follows. We use $s(t)$ to denote the photon flux of a finite-duration waveform emitted into the scene at time $t = 0$, and T_p to denote its root mean square (RMS) duration. We take N_s to be the total number of pulses employed per pixel and

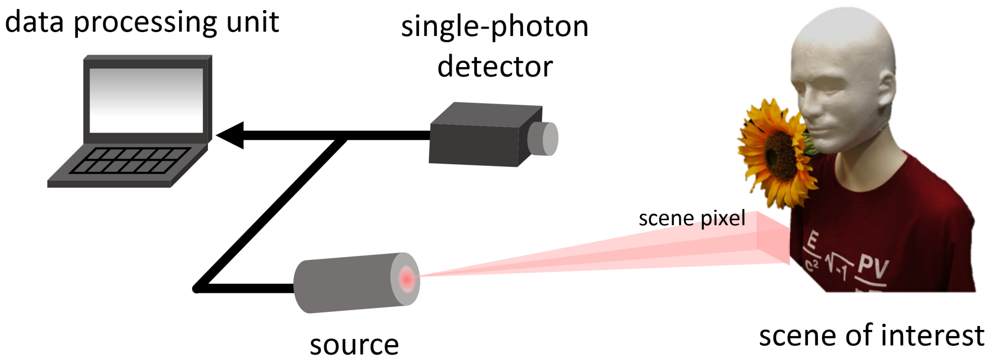


Figure 1-1: Single-photon active imaging setup with common components. The figure specifically depicts a setup with scanning source and a single-pixel detector.

T_r be the repetition period of the pulse illumination. (See Appendix A for a complete table of notations used in this thesis.)

To build up our model for the photodetection statistics, it is convenient to restrict our attention to a single scene pixel for now. Let $I(t)$ be the scene impulse response function that we are interested in knowing at the particular pixel. For example, $I(t)$ is an impulse function when imaging a single reflector, a multi-peaked function when imaging an object through diffuse media, and a decaying exponential function with unknown lifetime when imaging target fluorescence. When $s(t)$ illuminates the pixel in question, the photon flux $r(t)$ incident on the single-photon detector, is

$$r(t) = (s * I)(t) + b(t), \qquad t \in [0, T_r), \tag{1.1}$$

where $(s*I)(t)$ is the convolution between the pulse waveform and the scene response, and $b(t)$ represents the unwanted background response, such as sunlight or light from a competing active imager. The rate function $\lambda(t)$ generating the photon detections at the single-photon detector is then

$$\lambda(t) = \eta(r * I_d)(t) + b_d, \qquad t \in [0, T_r), \tag{1.2}$$

where $\eta \in (0, 1]$ is the detector's quantum efficiency, b_d is the detector's dark count rate, and $I_d(t)$ is the normalized detector response function ($\int_0^{T_r} I_d(t)\, dt = 1$). Defining $\tilde{s}(t) = (s * I_d)(t)$ and $\tilde{b}(t) = (b * I_d)(t)$, we can rewrite Eq. (1.2) as $\lambda(t) = \eta(\tilde{s} * I)(t) + \eta\tilde{b}(t) + b_d$.

A time-correlated single-photon detector is capable of determining the time-of-detection of a photon within an accuracy of Δ seconds. We assume $\Delta \ll T_r$ and that T_r is divisible by Δ. Then, we see that $m = T_r/\Delta$ is the number of time bins that contain our observed photon detection times. In other words, m is the size of the photon-count vector $\mathbf{y}$, where $\mathbf{y}_k \in \{0, 1, 2, \ldots\}$ for $k = 1, 2, \ldots, m$. Using the probabilistic theory of low-flux photon-counting, the number of photon detections in the kth time bin after N_s pulse illuminations of a motionless scene is an integer random variable that is Poisson distributed:

$$\mathbf{y}_k \sim \text{Poisson}\left(N_s \int_{(k-1)\Delta}^{k\Delta} \lambda(t)\, dt\right), \qquad k = 1, 2, \ldots, m, \tag{1.3}$$

where we have assumed that T_r is long enough to preclude pulse aliasing and that $b(t)$ is constant over the total acquisition time $N_s T_r$. Also, due to our low-flux assumption, the dead time effect of single-photon detectors is negligible (See Appendix B). The signal-to-noise (SNR) ratio, defined as the ratio of mean to the standard deviation, of the photon-count at kth bin is then

$$\mathsf{snr}(\mathbf{y}_k) = \sqrt{N_s \int_{(k-1)\Delta}^{k\Delta} \lambda(t)\, dt}, \qquad k = 1, 2, \ldots, m. \tag{1.4}$$

We see that, as the number of illuminations increases, the signal-to-noise ratio also increases at a rate of $\sqrt{N_s}$. However, in order to operate in the regime of high photon efficiency, we are interested in accurately recovering information about the scene response $I(t)$ from single-photon observations $\mathbf{y}$ obtained using a small N_s.

In order to form an image using a small number of photon detections, we propose a unifying single-photon imaging framework, which models the distribution of single-photon observations combined with physical constraints on scene parameters. Below we outline the elements of our framework, viz., the five-step procedure we will employ in all that follows.

- **Step I:** Derive the forward single-photon measurement model using the photodetection process from Eq. (1.3), for a specific single-photon imaging problem.
- **Step II:** Identify constraints on the scene parameters that we aim to recover for the specific single-photon imaging problem.
- **Step III:** Combine the photodetection model and physical constraints derived in Steps I and II, respectively, to formulate an optimization program that solves the inverse problem of recovering the parameters of scene response $I(t)$ (for a single pixel or multiple pixels, depending on the problem).
- **Step IV:** Regularize and relax the optimization problem that is formulated in Step III for computational efficiency, if necessary, while least perturbing the optimal solution.
- **Step V:** Design an algorithm that efficiently solves the final optimization problem in Step IV.

We emphasize that our low light level imaging framework does not require any new optical technology, although it will certainly benefit from continued advances in single-photon detectors. Instead, it relies on the high computational power that can be made available in modern imaging systems. In the chapters that follow, we will apply our five-step framework for the single-photon imaging regime to the six imaging problems that were described earlier, and develop photon-efficient approaches for each of them.

Chapter 2

Single-Reflector Depth Imaging

2.1 Overview of Problem

A conventional LIDAR system, which uses a pulsed light source and a single-photon detector, forms a depth image pixelwise using the histograms of photon detection times. The acquisition times for such systems are made long enough to detect hundreds of photons per pixel for the finely binned histograms these systems require to do accurate depth estimation.

Prior art: The conventional LIDAR technique of estimating depth using histograms of photon detections is accurate when the number of photon detections is high, since the photon histogram can be considered as an observation of the backreflected waveform. In the low photon-count regime, the depth solution is noisy due to shot noise. It has been shown that image denoising methods, such as wavelet thresholding, can improve the performance of scene depth recovery [10]. In other work, an imaging model that incorporates occlusion constraints was proposed to recover an accurate depth map [11]. However, these imaging algorithms implicitly assume that the observations are either noiseless or Gaussian distributed. Thus, at low photon-counts, where photon-detection statistics are highly non-Gaussian, their performance degrades significantly [2].

First-photon imaging (FPI) [12] is a framework that allows high-accuracy depth imaging using only the first detected photon at every pixel. It demonstrated that centimeter-accurate depth recovery is possible by combining the non-Gaussian statistics of first-photon detection with spatial correlations of natural scenes. The FPI framework uses an imaging setup that includes a raster-scanning light source and a lensless single-photon detector. More recently, a photon-efficient imaging framework that uses a detector array, in which every pixel has the same acquisition time, has also been proposed [13, 14]. It too relies on exploiting the spatial correlations of natural scenes.

We observe two common limitations that exist in the prior active imaging frameworks for depth reconstruction.

- **Over-smoothing:** Many of the frameworks assume spatial smoothness of the scene to mitigate the effect of shot noise. In some imaging applications, however, it is important to capture fine spatial features that only occupy a few image pixels. Using methods that assume spatial correlations may lead to erroneously over-smoothed images that wash out the scene's fine-scale features. In such scenarios, a robust *pixelwise* imager is preferable.

- **Calibration:** Many imaging methods assume a calibration step to measure the amount of background flux existing in the environment. This calibration mitigates bias in the depth estimate caused by background-photon or dark-count detections, which have high temporal variance. In practical imaging scenarios, however, the background response could vary in time, and continuous calibration may not be practical. Furthermore, many methods assume background flux does not vary spatially. Thus, a *calibrationless* imager that performs simultaneous estimation of scene parameters and spatially-varying background flux from raw photon detections is useful.

Summary of our approach: In this chapter, we propose a novel framework for depth acquisition that is applied pixelwise and without background calibration [15]. At each pixel, our imager estimates the background response along with scene depth from photon detections. Our framework uses a Poisson observation model for the photon detections plus a union-of-subspaces constraint on the scene's discrete-time flux at any single pixel, similar to the occlusion-based imaging framework in [11]. However, our union-of-subspaces constraint is defined for both the signal and background waveform parameters that generate photon detections; whereas the framework in [11] assumes a system observing a noiseless signal waveform in high light-level operation, without corruption by photon shot noise.

Using the derived imaging model, we propose a greedy signal pursuit algorithm that accurately solves for the scene parameters at each pixel. We evaluate the photon efficiency of this framework using experimental single-photon data. In the presence of strong background light, we show that our pixelwise imager gives an absolute depth error that is 6.1 times lower than that of the conventional pixelwise log-matched filter.

2.2 Single-Photon Imaging Setup

Figure 2-1 illustrates our imaging setup, for one illumination pulse, when the scene is illuminated in raster-scanning manner and a single-element photon detector is employed. (Alternatively, to reduce the time needed to acquire a depth map, our framework can be applied without modification when the scene is flood illuminated and a detector array is used.) A focused optical source, such as a laser, illuminates a pixel in the scene with the pulse waveform $s(t)$ that starts at time 0 and has root-mean-square pulsewidth T_p. This illumination is repeated every T_r seconds for a sequence of N_s pulses. The single-photon detector, in conjunction with a time correlator, is used to time stamp individual photon detections, relative to the time at which the immediately preceding pulse was transmitted. These detection times are observations of a time-inhomogeneous Poisson process whose rate function combines contributions from pixel return, background light, and dark counts. They are used to estimate scene depth for the illuminated pixel. This pixelwise acquisition process is repeated for $N_x \times N_y$ image pixels by raster scanning the light source in the transverse directions.

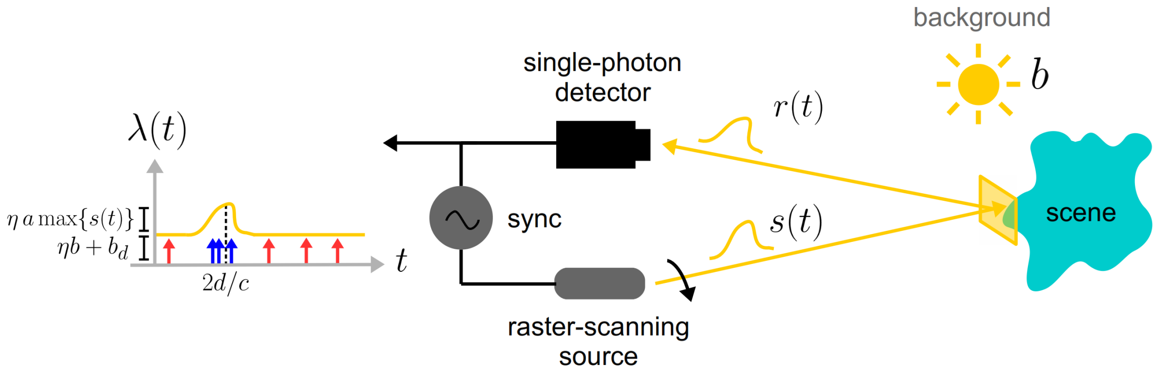


Figure 2-1: An illustration of the data acquisition procedure for one illumination pulse. A pulsed optical source illuminates a scene pixel with photon-flux waveform $s(t)$. The flux waveform $r(t)$ that is incident on the detector consists of the pixel return $as(t-2d/c)$—where a is the pixel reflectivity, d is the pixel depth, and c is light speed—plus the background-light flux b. The rate function $\lambda(t)$ driving the photodetection process equals the sum of the pixel return and background flux, scaled by the detector efficiency η, plus the detector's dark-count rate b_d. The record of detection times from the pixel return (or background light plus dark counts) is shown as blue (or red) spikes, generated by the Poisson process driven by $\lambda(t)$.

2.3 Forward Imaging Model

In this section, we introduce a model relating photon detections and scene parameters. For simplicity of exposition and notation, we focus on one pixel; this model is repeated for each pixel of a raster-scanning or array-detection setup.

Let a, d, and b be unknown scalar values that represent reflectivity, depth, and background flux at the given pixel. The reflectivity value includes the effects of radial fall-off, view angle, and material properties. Then, after illuminating the scene pixel with a single pulse $s(t)$, the backreflected waveform that is incident at the single-photon detector is

$$r(t) = as(t-2d/c) + b, \qquad t \in [0, T_r). \tag{2.1}$$

Comparing to our general light transport equation in Eq. (1.1), Eq. (2.1) specifically assumes that the scene response is generated by a single reflector ($I(t)$ is scaled and shifted Dirac delta function) and that the background light is constant.

Photodetection statistics: Using Eq. (2.1), we observe that the rate function that generates the photon detections is

$$\lambda(t) = \eta\left(as(t - 2d/c) + b\right) + b_d, \qquad t \in [0, T_r), \tag{2.2}$$

where $\eta \in (0, 1]$ is the quantum efficiency of the detector and $b_d \geq 0$ is the dark-count rate of the single-photon detector. Here, we have assumed an ideal detector response: $I_d(t) = \delta(t)$.

Recall that Δ is the time bin duration of the single-photon detector. Then, we define $m = T_r/\Delta$ to be the total number of time bins that capture photon detections. Let $\mathbf{y}$ be the vector of size $m \times 1$ that contains the photon counts at each time bin after we illuminate the pixel N_s times with pulse waveform $s(t)$. Then, from low-flux photon-counting theory in Eq. (1.3), we have that

$$\mathbf{y}_k \sim \text{Poisson}\left(N_s \int_{(k-1)\Delta}^{k\Delta} \left[\eta(as(t - 2d/c) + b) + b_d\right] dt\right), \tag{2.3}$$

for $k = 1, \ldots, m$. Note that we have assumed a stationary reflector and that our total pixelwise acquisition time $N_s T_r$ is short enough that b is constant during that period. We have also assumed that the low-flux condition $\sum_{k=1}^{m} \mathbf{y}_k \ll N_s$ holds, so that the effect of the single-photon detector's reset (dead) time can be neglected. We wish to reach an approximation in which the Poisson parameter of $\mathbf{y}_k$ is given by the product of a known matrix and an unknown (and constrained) vector.

Choose $n \in \mathbb{Z}^+$ such that $\Delta' = T_r/n$ is deemed adequate resolution for the estimated time of flight. (Our interest is in high-resolution imaging, where $n \geq m$, and hence $\Delta' \leq \Delta$.) Since $2d/c \in [0, T_r)$, $\mathbf{v} \in \mathbb{R}^{n \times 1}$ defined by

$$\mathbf{v}_j = \begin{cases} N_s \eta a, & \text{if } 2d/c \in [(j-1)\Delta', j\Delta'); \\ 0, & \text{otherwise,} \end{cases} \qquad j = 1, 2, \ldots, n, \tag{2.4}$$

has exactly one nonzero entry. Using this vector,

$$N_s \eta a s(t - 2d/c) \approx \sum_{j=1}^{n} \mathbf{v}_j s\big(t - \big(j - \tfrac{1}{2}\big)\,\Delta'\big) \tag{2.5}$$

is a good approximation when Δ' is small enough, because the sum has one nonzero term and $2d/c$ has been quantized to an interval of length Δ'. Substituting (2.5) into the Poisson parameter expression in (2.3) gives

$$\sum_{j=1}^{n} \left(\int_{(k-1)\Delta}^{k\Delta} s\big(t - \big(j - \tfrac{1}{2}\big)\,\Delta'\big)\,dt \right) \mathbf{v}_j + N_s \Delta(\eta b + b_d).$$

Then, defining $\mathbf{1}_{m\times 1}$ to be an $m \times 1$ vector of 1's, we can rewrite (2.3) as

$$\mathbf{y}_k \sim \text{Poisson}\big(\,(\mathbf{S}\mathbf{v} + B\mathbf{1}_{m\times 1})_k\,\big), \tag{2.6}$$

for $k = 1, \ldots, m$, where

$$\mathbf{S}_{i,j} = \int_{(i-1)\Delta}^{i\Delta} s\big(t - \big(j - \tfrac{1}{2}\big)\,\Delta'\big)\,dt, \tag{2.7}$$

$$B = N_s \Delta(\eta b + b_d), \tag{2.8}$$

for $i = 1 \ldots, m$ and $j = 1, \ldots, n$. Finally, defining $\mathbf{A} = [\mathbf{S}, \mathbf{1}_{m\times 1}]$ and $\mathbf{x} = [\mathbf{v}^T, B]^T$, we can further rewrite (2.6) as

$$\mathbf{y}_k \sim \text{Poisson}\big(\,(\mathbf{A}\mathbf{x})_k\,\big). \tag{2.9}$$

So far, we have simplified the pixelwise single-photon observation model, such that the photon-count vector $\mathbf{y} \in \mathbb{N}^{m\times 1}$ is a linear measurement of scene response vector $\mathbf{x} \in \mathbb{R}_+^{(n+1)\times 1}$ corrupted by signal-dependent Poisson noise.

Scene parameter constraints: We defined our $(n+1) \times 1$ signal $\mathbf{x}$ to be a concatenation of $\mathbf{v}$, which is the scene response vector of size n, and B, which is the scalar representing background flux plus dark counts. Since $\mathbf{v}$ has exactly one

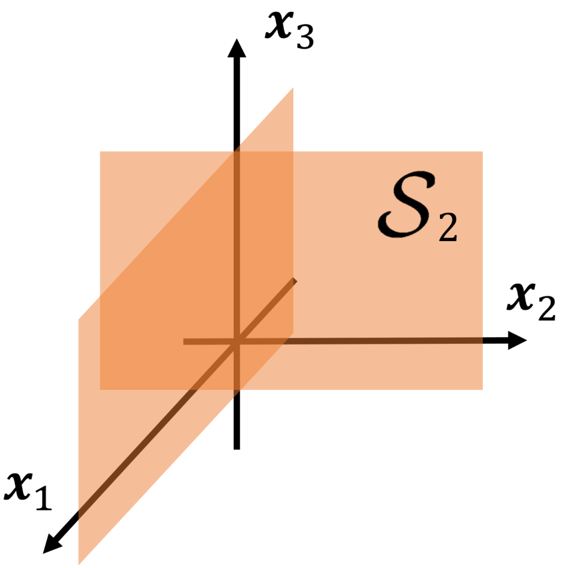


Figure 2-2: Illustration of $\mathcal{S}_2$. Our non-convex union-of-subspaces constraint set describes model-based signal sparsity that is specific to the LIDAR imaging setup.

nonzero entry, $\mathbf{x}$ lies in the union of n subspaces defined as

$$\mathcal{S}_n = \bigcup_{k=1}^{n} \left\{ \mathbf{x} \in \mathbb{R}^{n+1} : \mathbf{x}_{\{1,2,\ldots,n\}\setminus\{k\}} = 0 \right\}, \tag{2.10}$$

where each subspace is of dimension 2. Figure 2-2 illustrates $\mathcal{S}_n$ for $n = 2$.

2.4 Solving the Inverse Problem

Using accurate photodetection statistics and scene constraints, we have interpreted the problem of robust single-photon depth imaging as a noisy linear inverse problem, where the signal of interest $\mathbf{x}$ lies in the union-of-subspaces $\mathcal{S}_n$. Using (2.9), the observed photon count histogram $\mathbf{y}$ has the probability mass function

$$f_Y(\mathbf{y}; \mathbf{A}, \mathbf{x}) = \prod_{k=1}^{m} \frac{e^{-(\mathbf{A}\mathbf{x})_k} (\mathbf{A}\mathbf{x})_k^{\mathbf{y}_k}}{\mathbf{y}_k!}. \tag{2.11}$$

Thus, neglecting terms in the negative log-likelihood function that are dependent on $\mathbf{y}$ but not on $\mathbf{x}$, we can define the objective function

$$\mathcal{L}(\mathbf{x};\mathbf{A},\mathbf{y}) = \sum_{k=1}^{m} \left[(\mathbf{A}\mathbf{x})_k - \mathbf{y}_k \log{(\mathbf{A}\mathbf{x})_k} \right]. \tag{2.12}$$

This objective can be proved to be convex in $\mathbf{x}$.

We solve for $\mathbf{x}$ by minimizing $\mathcal{L}(\mathbf{x};\mathbf{A},\mathbf{y})$ with the constraint that $\mathbf{x}$ lies in the union-of-subspaces $\mathcal{S}_n$. Also, because photon flux is a non-negative quantity, the minimization results in a more accurate estimate when we include a non-negative signal constraint. In summary, the optimization problem that we want to solve can be written as

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & \mathcal{L}(\mathbf{x};\mathbf{A},\mathbf{y}) \\ \text{s.t.} \quad & \mathbf{x} \in \mathcal{S}_n, \\ & \mathbf{x}_i \geq 0, \quad i = 1, \ldots, (n+1). \end{aligned} \tag{2.13}$$

To solve our optimization problem in which the union-of-subspaces selects a sparse support, we propose an algorithm that is inspired by an existing fast iterative algorithm for sparse signal pursuit. Compressive sampling matching pursuit (CoSaMP) [16] is a greedy algorithm that finds a K-sparse approximate solution to an underdetermined linear inverse problem. CoSaMP iterates until it finds a solution that agrees with the observed data (according to some convergence metric), while the solution is a linear combination of K columns of the forward matrix $\mathbf{A}$. Adapting CoSaMP to our imaging framework is interesting for two particular reasons. First, unlike algorithms that only add to the solution support, never culling, CoSaMP has solution stability and accuracy properties that compete with globally-optimal ℓ_1-based convex optimization methods for sparse approximation [17]. Second, CoSaMP has been shown to be adaptable to applications in which the signal being estimated has a structured support [18], as is true for the union-of-subspaces model. Thus, we are motivated to modify the CoSaMP algorithm to our specific use case, where we are interested in recovering a sparse solution in the union-of-subspaces $\mathcal{S}_N$, instead of a K-sparse one, using photon-noise corrupted data.

Algorithm 1 Single-photon depth imaging using a union-of-subspaces model

```
Input: $\mathbf{y}$, $\mathbf{A}$, $\delta$
Output: $\hat{\mathbf{x}}_{\text{UOS}}$

Initialize $\hat{\mathbf{x}}^{(0)} \leftarrow \vec{0}$, $\mathbf{u} \leftarrow \mathbf{y}$, $k \leftarrow 0$
repeat
    $k \leftarrow k + 1$
    $\mathbf{x}' \leftarrow \mathbf{A}^T \mathbf{u}$                                          ▷ Compute adjoint
    $\Omega \leftarrow \text{supp}((\mathbf{x}'_{1:n})_{[1]}) \cup \text{supp}(\hat{\mathbf{x}}^{(k-1)}_{1:n}) \cup \{n+1\}$      ▷ Merge support
    $\mathbf{w}|_{\Omega} \leftarrow \mathbf{A}^{\dagger}_{\Omega}\mathbf{y}, \quad \mathbf{w}|_{\Omega^c} \leftarrow \mathbf{0}$
    $\hat{\mathbf{x}}^{(k)} \leftarrow \mathcal{T}_0\left([(\mathbf{w}_{1:n})^T_{[1]},\ \mathbf{w}_{n+1}]^T\right)$        ▷ Threshold and update solution
    $\mathbf{u} \leftarrow \mathbf{y} - \mathbf{A}\hat{\mathbf{x}}^{(k)}$
until $\|\hat{\mathbf{x}}^{(k-1)} - \hat{\mathbf{x}}^{(k)}\|_2^2 < \delta$
$\hat{\mathbf{x}}_{\text{UOS}} \leftarrow \hat{\mathbf{x}}^{(k)}$                                  ▷ Output converged solution
```

Our proposed greedy algorithm, inspired by CoSaMP, is given in Algorithm 1. We define $\mathcal{T}_0(\mathbf{x})$ to be the thresholding operator that sets all negative entries of $\mathbf{x}$ to zero, $\text{supp}(\mathbf{x})$ to be the set of indices of $\mathbf{x}$'s nonzero elements, and $\mathbf{x}_{[k]}$ to be the vector that approximates $\mathbf{x}$ with its k largest terms. Also, we take $\mathbf{A}_S$ to be a matrix generated using the columns of $\mathbf{A}$ described by the index set S. Finally, we use $\mathbf{A}^T$ and $\mathbf{A}^{\dagger}$ to denote the transpose and pseudo-inverse of matrix $\mathbf{A}$, respectively.

In Algorithm 1, for computational efficiency we have approximated $\mathcal{L}(\mathbf{x}; \mathbf{A}, \mathbf{y})$ with the ℓ_2-loss $\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2$, which is the first-order Taylor expansion (up to a constant) of $\mathcal{L}(\mathbf{x}; \mathbf{A}, \mathbf{y})$ with respect to $\mathbf{x}$ (see Appendix D for derivation). Because CoSaMP also assumes an ℓ_2-loss function, the only change from CoSaMP to our algorithm is then the update stage; instead of picking out the best k terms, we pick out the two terms from the intermediate solution based on the union-of-subspaces and non-negativity constraints. We iterate the algorithm until the solution meets the convergence criterion: $\|\hat{\mathbf{x}}^{(k-1)} - \hat{\mathbf{x}}^{(k)}\|_2^2 < \delta$, where δ is a small number.

Many sparse pursuit algorithms, such as CoSaMP, are successful given the assumption that $\mathbf{A}$ is incoherent. In our setup, however, the forward matrix $\mathbf{A}$ is highly coherent due to Δ' being small and the pulse waveform $s(t)$ being smooth. Nevertheless, because the linear system's degree of underdetermination is extremely mild ($\mathbf{A} \in \mathbb{R}^{n\times(n+1)}$) and the sparsity level is fixed to a small number ($\dim(\mathcal{S}_n) = 2$) relative to the signal dimension (typically exceeding 100), our single-photon imaging algorithm recovers the scene parameters of interest in a robust manner.

2.5 Results

Simulations: We simulated the single-photon imaging setup using Eq. (2.9) to study the performance of the proposed depth imaging method. It would be valuable to compare our imaging method with the ML estimator for scene parameters $\{a, d, b\}$. Unfortunately, due to nonzero background flux, ML estimation requires minimizing a non-convex cost function, leading to a solution without convergence and accuracy guarantees. Thus, zero background is assumed conventionally, even for data that is contaminated by background, such that the ML depth estimate reduces to the simple log-matched filter [19]:

$$\hat{d}_{\mathrm{ML}} = \frac{c\Delta'}{2}\left(\underset{i\in\{1,\ldots,n\}}{\arg\max} \log \mathbf{S}_i^T \mathbf{y}\right). \tag{2.14}$$

Note that this estimator is equivalent to a one-step greedy algorithm (where a union-of-subspaces constraint is irrelevant) of minimizing $\mathcal{L}(\mathbf{x}; \mathbf{A}, \mathbf{y})$ for a 1-sparse solution. We use (2.14) as the baseline depth estimator that is compared with our proposed estimator using the union-of-subspaces model.

Figure 2-3 shows the root mean square error (RMSE) of depth reconstruction at two different signal-to-background (SBR) levels. The SBR is defined as the ratio of the probability of a detection coming from signal to the probability of a detection coming from background light plus dark counts:

$$\mathrm{SBR} = \frac{\sum_{k=1}^{m} (\mathbf{S}\mathbf{v})_k}{\sum_{k=1}^{m} (B\mathbf{1}_{m\times 1})_k}. \tag{2.15}$$

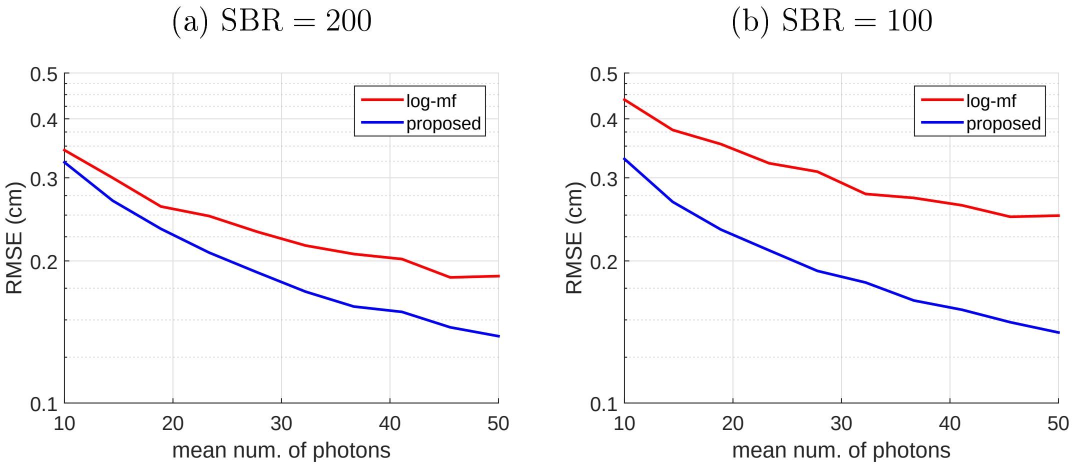


Figure 2-3: Simulation of pixelwise depth imaging results using single photon observations for two SBR levels. (a) Depth RMSE of log-matched filter and proposed method for SBR = 200. (b) Depth RMSE of log-matched filter and proposed method for SBR = 100.

Figure 2-3 was generated using $m = n = 801$ and using the RMS duration T_p of a Gaussian pulse such that $cT_p/2 = 1$ cm. For each number of photon detections, 4000 Monte Carlo trials of Eq. (2.6) were run. We observe that our method improves over conventional log-matched filtering for both SBR levels, with greater improvements seen in the lower SBR case.

Experiments: To experimentally validate our imaging framework, we used a dataset collected for the First-Photon Imaging project [12]; this dataset and others are available from [20]. Figure 2-4 shows a photograph of the single-photon imaging setup. The single-photon imaging setup used a pulsed laser diode with pulsewidth $T_p = 270$ ps (such that $cT_p/2 = 4$ cm) and repetition period $T_r = 100$ ns. A two-axis galvo was used to scan 350 × 350 pixels of a mannequin face at a distance of about 4 m. A lensless single-photon avalanche diode (SPAD) detector with quantum efficiency $\eta = 0.35$ was used for detection. The background light level was set using an incandescent lamp. The original mannequin data from [20] had the background count rate approximately equal to the signal count rate. Our experiment uses the cropped data showing only the mannequin's face, where the background count rate was approximately 0.1 of the average signal count rate. Although we used a raster-scanning

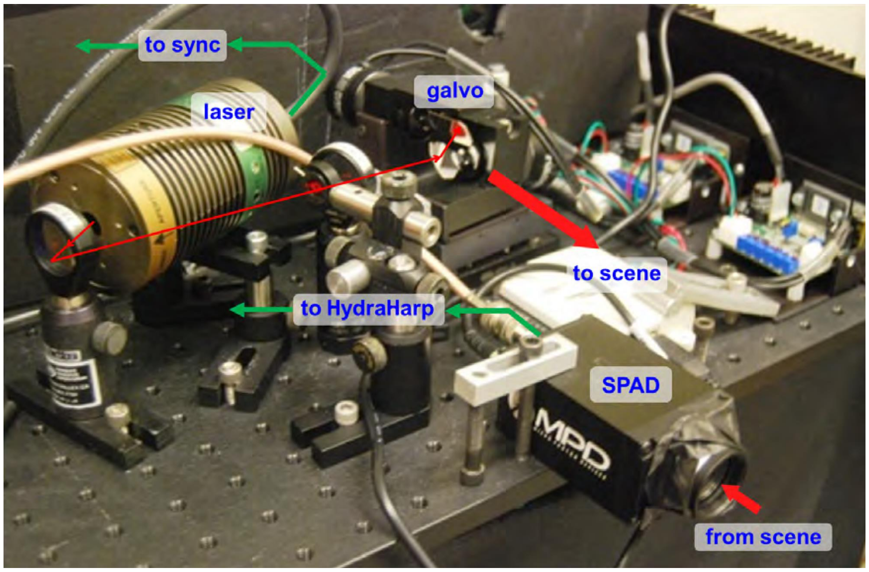


Figure 2-4: Experimental setup with a raster-scanning source and a single SPAD detector. The red arrows show the path of the optical signal from laser source, and the green arrows show the path of the electrical signal that indicates whether a photon has been detected or not.

setup for our single-photon imaging experiments, since our imaging algorithm is applied pixelwise, it can be also used for imaging with a floodlight illumination source and a single-photon detector array. However, the current framework does not explicitly deal with array limitations such as the low resolution of time-correlation inherent in single-photon camera operations; Chapter 4 develops an array imaging framework to address that point.

Figure 2-5 shows the results of recovering depth of the mannequin face using single-photon observations. The kernel matrix $\mathbf{S}$ was obtained by an offline measurement of the pulse shape. Note that this measurement depends only on the source, not on properties of the scene. The ground-truth depth, shown in Figure 2-5(b), was generated separately by using background-calibrated ML estimation from 200 photon detections at each pixel.

In our depth imaging experiment, the number of photon detections at each pixel was set to 15. We observe that, due to extraneous background photon detections and dark counts, the log-matched filter estimate in Figure 2-5(c) (average absolute error =

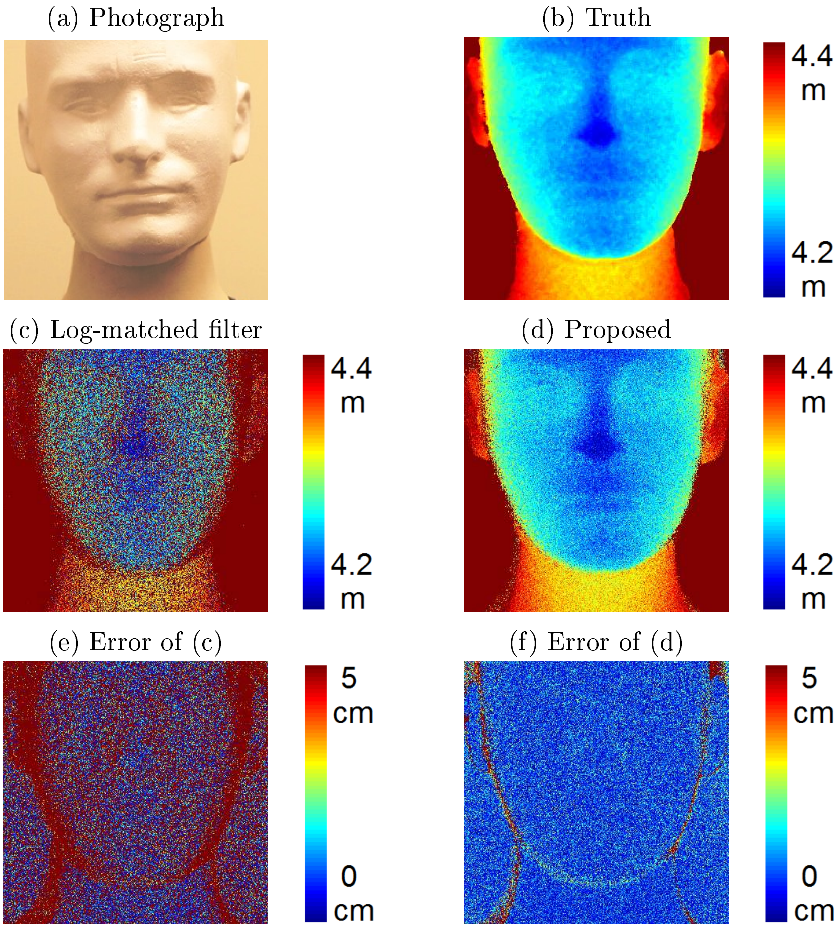


Figure 2-5: Experimental pixelwise depth imaging results using single photon observations. The number of photon detections at every pixel was set to be 15. The figure shows the (a) photograph of imaged face, (b) ground-truth depth, (c) depth from log-matched filtering, which is approximately ML and (d) depth using our method. The absolute depth-error maps for ML and our framework are shown in (e) and (f), respectively.

10.3 cm) is corrupted with high-variance noise and the facial features of the mannequin are heavily obscured. On the other hand, our estimate, shown in Figure 2-5(d), shows high-accuracy depth recovery (average absolute error = 1.7 cm). As shown by the error maps in Figure 2-5(e), (f), both methods fail in depth recovery in the face boundary regions, where very little light is reflected back from the scene to the single-photon detector. This failure is because the SBR is very low in such regions. Also, we observe that our estimated average background level over all pixels (last entry of $\hat{\mathbf{x}}_{\text{UOS}}$ from our algorithm) was $\hat{B} = 1.4 \times 10^{-3}$, which is very close to the calibrated true background level $B = 1.3 \times 10^{-3}$. This background flux information,

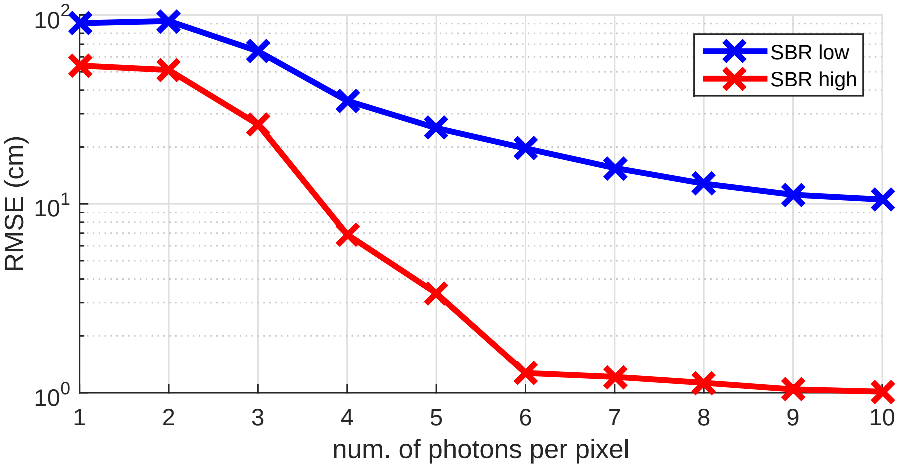


Figure 2-6: Depth recovery performance of our algorithm at a face pixel (SBR $= 6.7$ and pixel coordinates $(81, 272)$) and a depth boundary pixel (SBR $= 1.5$ and pixel coordinates $(237, 278)$) for varying number of photon detections.

which is valuable in evaluating the confidence of image estimates, is not available when using the conventional log-matched filtering.

Figure 2-6 shows, for a pixel in facial region and another at the face boundary, which have SBRs of 6.7 and 1.5 respectively, how our depth reconstruction algorithm performs as the number of photon detections increase. We observe that the algorithm performs better for higher SBR overall, and that the rate of decrease in depth error with increasing number of photon detections is faster for high SBR than for low SBR, especially in the very low-flux regime (2 to 5 detections). Also, we observe a floor in depth error for increasing number of photon detections, at around 1 cm when SBR is 6.7. Knowing that $c\Delta'$ was around 1 cm in our experimental imaging setup, we see, at low SBR and low photon counts, it is hard for the algorithm to achieve depth resolution below $c\Delta'$. In this experiment, we had $m = n = 801$. Also, we set $\delta = 10^{-4}$ and the average number of iterations until convergence was measured to be 2.1 over all pixels. Code and data used to generate results can be downloaded from [21].

Computational complexity: Solving the ML problem by performing likelihood evaluations over the cube $\{a_i, b_j, d_k\}_{i,j,k=1}^{n}$ takes $O(n^3)$ searches. On the other hand, the log-ML solution, which is a convolution operation between the pulse waveform and the photon-count histogram, takes $O(wn)$ flops, where w is the number of depth bins defined by the support of the finite-duration pulse waveform.

The most computationally expensive step in our Algorithm 1 is when the adjoint solution is being obtained, since $\mathbf{A}$ is size $m \times (n+1)$. However, we see that due to the Toeplitz structure of the submatrix $\mathbf{S}$ of $\mathbf{A}$, we can reduce computing a single iteration of our Algorithm to $O(wn)$ flops, assuming $m = n$. When tested empirically, using both experimental and simulated data, we noticed that the number of iterations required for our algorithm to converge was close to 2, which is our fixed LIDAR signal sparsity level. Thus, the computational complexity of our algorithm is around $O(wn)$, which is still linear in the depth bin dimension n that typically is the largest number of all imaging parameters.

2.6 Summary and Discussion

In this chapter, we presented an imaging framework for background-calibrationless, pixelwise depth reconstruction using single-photon observations. Our imaging model combined photon detection statistics with the discrete-time flux constraints expressed using a union-of-subspaces model. Then, using our imaging model, we developed a greedy algorithm that recovers scene depth by solving a constrained optimization problem.

Our pixelwise imaging framework can be used in low light-level imaging applications, in which the scene being imaged has fine features and filtering techniques that exploit patchwise smoothness can potentially wash out those details. For example, it can be useful in applications such as airborne remote sensing [22], where the aim is to recover finely-featured 3D terrain maps.

The proposed single-photon imaging framework can be extended to multiple-depth estimation, where more than one reflector may be present at each pixel. In the case of estimating depths of K reflectors at a pixel, one may use a more general K-sparsity assumption instead of the 1-sparsity assumption when defining the union-of-subspaces constraint. However, K is typically unknown. By attempting to fix the value of K, this naive extension fails to accurately image scenes with number of reflectors varying from one pixel to another. In the next chapter, we see how the generalized multi-depth estimation problem can be addressed by using a convex optimization variant of our framework developed in this chapter, such that the number of reflectors does not have to be specified for multi-depth imaging.

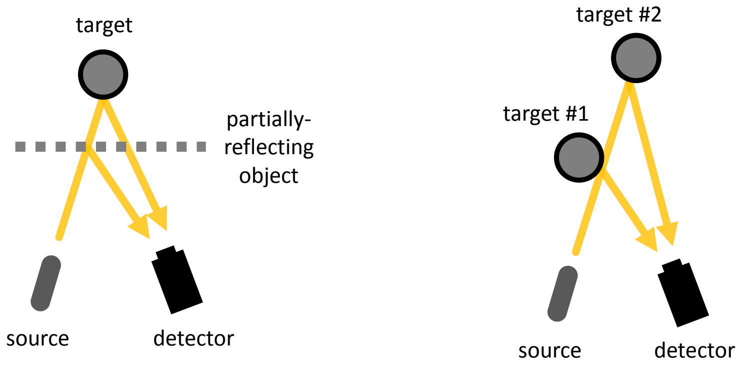
Chapter 3

Multi-Depth Imaging

3.1 Overview of Problem

As noted in the previous chapter, the ability to acquire a scene's 3D structure is important in many applications, spanning from biometrics [23] to terrestrial mapping [24]. Active depth imagers use time-of-flight information of a backreflected light signal to construct a scene's depth map. However, as illustrated in Figure 3-1, 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 recover the multiple depths present in the field-of-view. This is known as the problem of multi-depth reconstruction from full-waveform measurements [5].

Prior art: Conventional time-of-flight imaging sensors, such as the amplitude-modulated continuous-wave (AMCW) modules, aim to reconstruct multi-depth profiles by the methods of transient imaging [25–27]. For low-power and long-range 3D imaging applications, a sensitive single-photon detector is used in the active imaging setup for low-light level operations [2, 12, 13, 28]. Recent advances in single-photon imaging system design allow time-of-flight imaging at long ranges and at low fluxes [29, 30]. Photon-count histogramming methods have been used for pixelwise reconstruction of depths using the time-correlated single-photon counting (TCSPC) imag-



(a) Partially-reflecting object (b) Partially-occluding object

Figure 3-1: 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.

ing setup [14, 31]. For example, the high sensitivity of time-correlated single-photon imaging systems were demonstrated when full photon histogram measurements were used to track pulses of light in flight [32]. For multi-depth imaging of scene reflectors in particular, one may choose to identify the peaks in the photon histogram by brute-force search over depth 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 [33, 34]. Assuming accurate waveform measurement, the compressive depth acquisition camera (CoDAC) [35] framework also exploits parametric deconvolution, 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 photon detections is high enough to form a histogram that resembles the backreflected waveform. The problem of recovering the multi-depth information is generally difficult in low-flux scenarios due to the low signal-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. Multi-depth estimation by learning of the mixture model is achieved either by using the expectation-maximization (EM) algorithm [36] or the Markov chain Monte Carlo (MCMC) method [37, 38]. 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.

Summary of our approach: In this chapter, we develop a multi-depth imaging framework that combines the statistics of single-photon detections with the sparsity of the scene's temporal response at a pixel. Using our model for the measured data, we show that pixelwise multi-depth reconstruction can be interpreted as a sparse deconvolution problem. We propose an adaptation of the iterative soft thresholding algorithm (ISTA) [39, 40] for our single-photon imaging setup to solve the convex optimization problem that arises from discretizing and relaxing the sparse deconvolution problem.

In single-photon imaging experiments, we demonstrate the performance of the proposed framework in the two experimental configurations shown in Figure 3-1: a scene with a partially-reflective object and a scene with a partially-occluding object. In both cases, we show that our framework outperforms the existing MoG-based method at low-light levels in the presence of strong background light. For example, as detailed in Section 3.5, 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 a RMS depth error of 48.7 cm, this is an improvement by a factor of 4.2.

3.2 Single-Photon Imaging Setup

Figure 3-2 depicts the single-photon imaging setup for estimating the 3D structure of scene. A focused optical source illuminates a scene patch with a sequence of pulses. 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 multiple reflectors in 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.

3.3 Forward Imaging Model

We first derive the relationship between the scene's reflector depths and our photon-detection data. To avoid unnecessary notation, we characterize the light transport and detection statistics for a single pixel; the same model applies at each pixel. Recall that a time-correlated single-photon detector records the time-of-detection of a photon within a timing resolution of Δ seconds. Then, $m = T_r/\Delta$ is the number of time bins that may contain photon detections. In other words, m is the dimension of the photon count vector $\mathbf{y}$ with each of its entries distributed as

$$\begin{aligned} \mathbf{y}_k &\sim \text{Poisson}\left(N_s \int_{(k-1)\Delta}^{k\Delta} \lambda(t)\,dt\right) && (3.1)\\ &= \text{Poisson}\left(\underbrace{N_s\eta \int_{(k-1)\Delta}^{k\Delta} (I * s)(t)\,dt}_{\substack{\text{Mean count of}\\ \text{signal photons}}} \;+\; \underbrace{N_s\Delta(\eta b + b_d)}_{\substack{\text{Mean count of}\\ \text{background photons}\\ \text{plus dark counts}}}\right), \text{ for } k = 1, \ldots, m, && (3.2) \end{aligned}$$

where $I(t)$ is the scene impulse response and $s(t)$ is the pulse waveform for $t \in [0, T_r)$. Typically, we have $T_p \ll T_r$, where the pulse duration is much shorter than the inter-pulse time interval. Here, we have taken the detector response to be ideal, i.e., $I_d(t) = \delta(t)$. We would like to reach an approximation in which the Poisson parameter of $\mathbf{y}_k$ is given by the product of a known matrix and an unknown deterministic vector

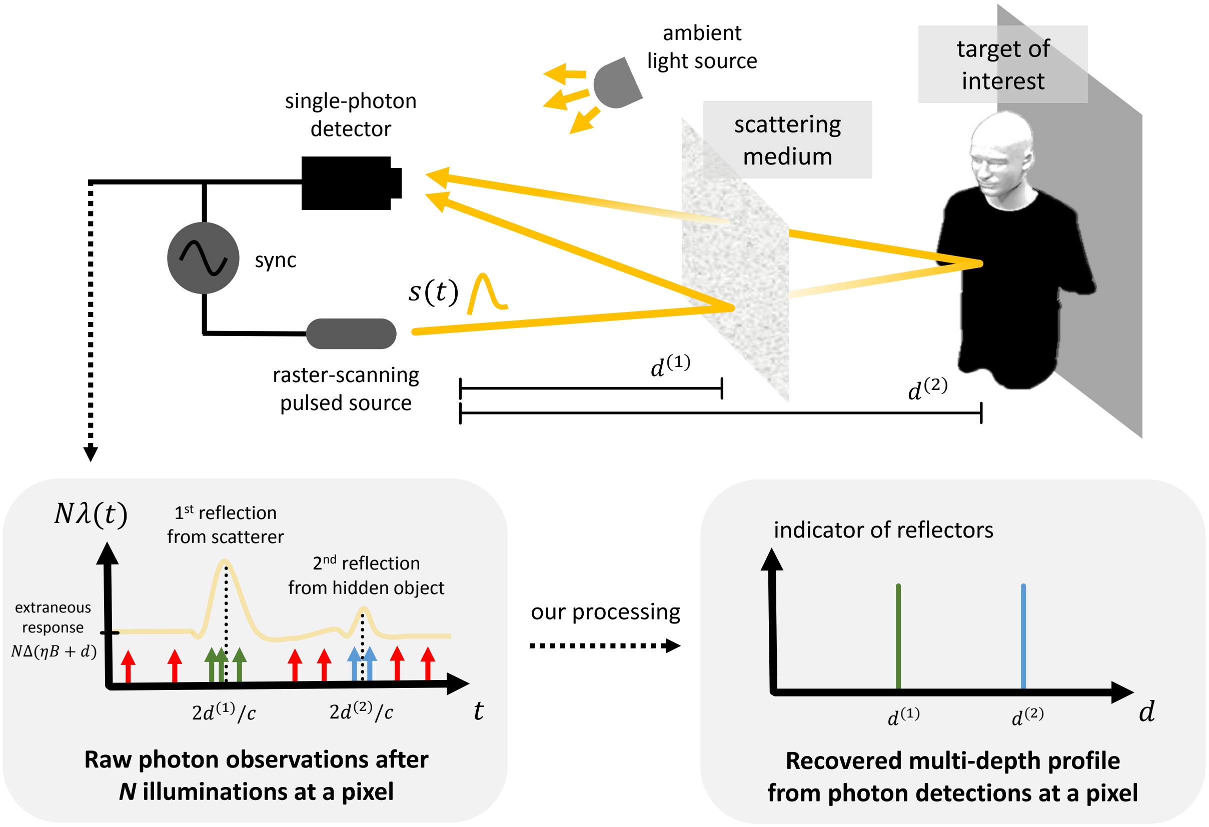

ambient
light source
target of
interest
single-photon
detector
scattering
medium
sync
$s(t)$
raster-scanning
pulsed source
$d^{(1)}$
$d^{(2)}$
$N\lambda(t)$
1st reflection
from scatterer
2nd reflection
from hidden object
extraneous
response
$N\Delta(\eta B + d)$
$2d^{(1)}/c$
$2d^{(2)}/c$
t
Raw photon observations after
N illuminations at a pixel
our processing
indicator of reflectors
$d^{(1)}$
$d^{(2)}$
d
Recovered multi-depth profile
from photon detections at a pixel

Figure 3-2: (Top) Full-waveform single-photon imaging setup for depth estimation of multiple objects. In this example, a pulsed optical source illuminates a scene pixel 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_s-pulse rate function $N_s\lambda(t)$ 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.

representing the scene response at a pixel.

Similar to the discrete modeling procedure in Chapter 2, we will approximate $(I * s)(t)$ with a sampling period of Δ', where $\Delta' = T_r/n$ for some $n \in \mathbb{Z}^+$. We can approximate the first term in the Poisson parameter in Eq. (3.2) as

$$\begin{aligned}
N_s\eta \int_{(k-1)\Delta}^{k\Delta} (I * s)(t)\,dt &= \int_{(k-1)\Delta}^{k\Delta} \int_0^{T_r} N_s\eta I(y)s(t-y)\,dy\,dt && (3.3)\\
&\overset{(a)}{=} \int_{(k-1)\Delta}^{k\Delta} \sum_{j=1}^{n} \int_{(j-1)\Delta'}^{j\Delta'} N_s\eta I(y)s(t-y)\,dy\,dt && (3.4)\\
&= \sum_{j=1}^{n} \int_{(k-1)\Delta}^{k\Delta} \int_{(j-1)\Delta'}^{j\Delta'} I(y)\,N_s\eta s(t-y)\,dy\,dt && (3.5)\\
&\overset{(b)}{\approx} \sum_{j=1}^{n} \int_{(k-1)\Delta}^{k\Delta} \int_{(j-1)\Delta'}^{j\Delta'} \frac{\mathbf{x}_j}{\Delta'}\,\frac{\mathbf{S}_{k,j}}{\Delta}\,dy\,dt && (3.6)\\
&= \sum_{j=1}^{n} \mathbf{S}_{k,j}\,\mathbf{x}_j, && (3.7)
\end{aligned}$$

where (a) follows from partitioning $[0, T_r)$ into n subintervals and (b) from replacing $I(y)$ and $N_s\eta s(t-y)$ by constant approximations on $(y,t) \in [(j-1)\Delta', j\Delta') \times [(k-1)\Delta, k\Delta)$; specifically, we define

$$\begin{aligned}
\mathbf{x}_j &= \int_{(j-1)\Delta'}^{j\Delta'} I(y)\,dy, \quad \text{for} \ \ j = 1, \ldots n && (3.8)\\
\mathbf{S}_{k,j} &= \frac{1}{\Delta'} \int_{(k-1)\Delta}^{k\Delta} \int_{(j-1)\Delta'}^{j\Delta'} N_s\eta s(t-y)\,dy\,dt, \quad \text{for} \ \ k = 1, \ldots m, \ j = 1, \ldots, n. && (3.9)
\end{aligned}$$

Note that the quality of the approximation (b) will depend on the size of Δ'. Finally, using the derived approximations, the observation model of Eq. (3.2) can be rewritten in a concise matrix-vector form as the following:

$$\mathbf{y}_k \sim \text{Poisson}\left((\mathbf{S}\mathbf{x} + B\mathbf{1}_{m\times 1})_k \right), \qquad \text{for } k = 1, 2, \ldots, m, \tag{3.10}$$

where $\mathbf{x}$ is an $n \times 1$ vector, $\mathbf{S}$ is an $m \times n$ matrix, and $\mathbf{1}_{m\times 1}$ is an $m \times 1$ vector of ones, and

$$B = N_s \Delta(\eta b + b_d). \tag{3.11}$$

Observation likelihood expressions: Our goal of multi-depth reconstruction is to accurately estimate $\mathbf{x}$ from $\mathbf{y}$. Using Eq. (3.10), the photon count histogram vector $\mathbf{y}$ has the probability mass function

$$p_Y(\mathbf{y}; \mathbf{x}, \mathbf{S}, B) = \prod_{k=1}^{m} \frac{\exp\left\{-(\mathbf{S}\mathbf{x} + B\mathbf{1}_{m\times 1})_k\right\} (\mathbf{S}\mathbf{x} + B\mathbf{1}_{m\times 1})_k^{\mathbf{y}_k}}{\mathbf{y}_k\,!}. \tag{3.12}$$

We can thus write the negative log-likelihood of $\mathbf{x}$ as

$$\mathcal{L}(\mathbf{x}; \mathbf{y}, \mathbf{S}, B) = -\log p_Y(\mathbf{y}; \mathbf{x}, \mathbf{S}, B) \cong \sum_{k=1}^{m} \left[(\mathbf{S}\mathbf{x})_k - \mathbf{y}_k \log\left(\mathbf{S}\mathbf{x} + B\mathbf{1}_{m\times 1}\right)_k\right], \tag{3.13}$$

where $\cong$ 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 $\mathbf{x}$ (see Appendix C).

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 [37]:

$$I(t) = \sum_{i=1}^{K} a^{(i)} \delta\big(t - 2d^{(i)}/c\big)\,, \qquad t \in [0, T_r), \tag{3.14}$$

where $a^{(i)}$ and $d^{(i)}$ are respectively the reflectivity and depth values of the ith 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

$$d_s = \min_{i=1,\ldots,K-1} \left| d^{(i)} - d^{(i+1)} \right|. \tag{3.15}$$

Assuming the K-reflector model and $c\Delta'/2 < d_s$, we see that exactly K elements of $\mathbf{x}$ are nonzero and those entries have values $\{a^{(i)}\}_{i=1}^{K}$.

3.4 Solving the Inverse Problem

The multipath profile can be estimated by minimizing the negative log-likelihood function, while including a signal support constraint that the number of nonzero elements in $\mathbf{x}$ must be equal to K. However, the ℓ_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 [41]. In order to design a computationally feasible algorithm, we apply the convex relaxation whereby the ℓ_1-norm serves as a proxy for the ℓ_0-norm [42]. We see later in this section that this relaxation is also useful given the fact that K varies from one pixel to another for a scene (See Section 3.5). Our proposed imaging framework also constrains the reflectivity estimates to be nonnegative. Thus, we obtain the multi-depth profile estimate $\hat{\mathbf{x}}_{\mathrm{OPT}}$ by solving the following ℓ_1-penalized and constrained likelihood optimization problem:

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & \mathcal{L}(\mathbf{x}; \mathbf{y}, \mathbf{S}, B) + \tau \|\mathbf{x}\|_1 \\ \text{subject to} \quad & \mathbf{x}_k \geq 0, \quad k = 1, 2, \ldots, n, \end{aligned} \tag{3.16}$$

where $\tau > 0$ is the variational parameter controlling the degree of penalizing the non-sparsity of the estimate. Because $\mathcal{L}(\mathbf{x}; \mathbf{y}, \mathbf{S}, B)$ and the ℓ_1-norm are both convex functions in $\mathbf{x}$ and the nonnegative cone is a convex set, the minimization problem given in (3.16) is a convex optimization problem.

ISTA is a celebrated algorithm for rapidly solving the ℓ_1-penalized constrained likelihood optimization problem when the data is corrupted by additive white Gaussian noise. However, instead of using a Gaussian likelihood, (3.16) is derived based on the model for single-photon observations. Thus, we modify the first step of ISTA that

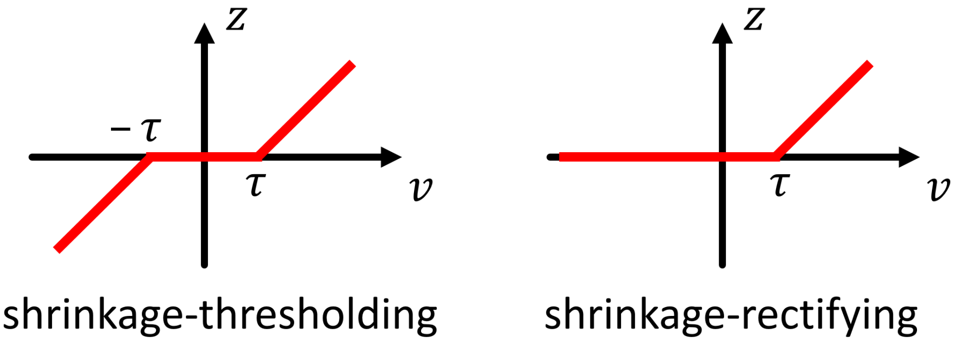


Figure 3-3: Illustration of a shrinkage-thresholding operation used as a step in ISTA (left) and the shrinkage-rectification operation used as a step in our SPISTA (right) that includes the nonnegativity constraint. Here the operations map scalar v to scalar z (variables only used for illustration purposes), with regularization parameter τ.

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. (3.13). We can compute the gradient of our negative log-likelihood function as

$$\begin{aligned}\nabla_{\mathbf{x}}\mathcal{L}(\mathbf{x};\mathbf{y},\mathbf{S},b) &= \sum_{k=1}^{m} \nabla_{\mathbf{x}}\left[(\mathbf{S}\mathbf{x})_k - \mathbf{y}_k \log\left(\mathbf{S}\mathbf{x} + B\mathbf{1}_{m\times 1}\right)_k\right] && (3.17)\\ &= \sum_{k=1}^{m} \left[(\mathbf{S}^T)_k - \frac{\mathbf{y}_k}{(\mathbf{S}\mathbf{x} + B\mathbf{1}_{m\times 1})_k}(\mathbf{S}^T)_k\right] && (3.18)\\ &= \mathbf{S}^T\left[\mathbb{1}_m - \operatorname{div}(\mathbf{y},\, \mathbf{S}\mathbf{x} + B\mathbf{1}_{m\times 1})\right], && (3.19)\end{aligned}$$

where we used $(\mathbf{S}^T)_k$ to denote the kth column of $\mathbf{S}^T$ and $\operatorname{div}(\cdot,\cdot)$ 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. 3-3.

Our modified ISTA algorithm is thus given in Algorithm 2. After solving (3.16) using Algorithm 2, we apply post-processing on $\hat{\mathbf{x}}_{\mathrm{SPISTA}}$ 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 3. Also, Fig. 3-4 shows the intermediate SPISTA output and the final output of Algorithm 3, given single-pixel photon count observations obtained from our experiments of imaging partially-occluding objects shown in Fig. 3-9.

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 (2.13). This is unlike existing histogram-based 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 (2.13). 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.

3.5 Results

Simulations 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 [43] and looking through a transparent object [44] 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 used the pulsewidth-normalized root mean-square

Algorithm 2 Single-photon iterative soft-thresholding algorithm (SPISTA)

```
Input: $\mathbf{y}, \mathbf{S}, B, \tau, \hat{\mathbf{x}}^{(0)}, \delta$
Output: $\hat{\mathbf{x}}_{\text{SPISTA}}$

Initialize $k \leftarrow 0$
repeat
    $k \leftarrow k + 1$
    $\hat{\mathbf{x}}^{(k)} \leftarrow \hat{\mathbf{x}}^{(k-1)} - \mathbf{S}^T \left[ \mathbf{1}_{m \times 1} - \text{div}(\mathbf{y},\, \mathbf{S}\hat{\mathbf{x}}^{(k-1)} + B\mathbf{1}_{m \times 1}) \right]$    ▷ Gradient descent
    $\hat{\mathbf{x}}^{(k)} \leftarrow \mathcal{T}_0(\hat{\mathbf{x}}^{(k)} - \tau)$    ▷ Shrinkage-rectification
until $\|\hat{\mathbf{x}}^{(k-1)} - \hat{\mathbf{x}}^{(k)}\|_2^2 < \delta$
$\hat{\mathbf{x}}_{\text{SPISTA}} \leftarrow \hat{\mathbf{x}}^{(k)}$    ▷ Output converged solution
```

Algorithm 3 Computational multi-depth single-photon imaging

Input: $\mathbf{y}, \mathbf{S}, B, \tau, \hat{\mathbf{x}}^{(0)}, \delta, \epsilon$

Output: $\hat{\mathbf{x}}_{\text{OPT}}$, the sparse multi-depth vector

1. Convex optimization:
Obtain $\hat{\mathbf{x}}_{\text{SPISTA}}$ by solving (3.16) with SPISTA$(\mathbf{y}, \mathbf{S}, B, \tau, \hat{\mathbf{x}}^{(0)}, \delta)$. Set $\hat{\mathbf{x}}_{\text{OPT}} = \hat{\mathbf{x}}_{\text{SPISTA}}$.

2. Residual filtering:
Identify the index set $L = \{i \in \{1, 2, \ldots, n\} : (\hat{\mathbf{x}}_{\text{OPT}})_i < \epsilon\}$, where ϵ is a small positive number. Set $(\hat{\mathbf{x}}_{\text{OPT}})|_L = 0$ to filter out residuals.

3. Depth grouping:
Identify index set in which the adjacent bins of $\hat{\mathbf{x}}_{\text{OPT}}$ are all non-negative. Replace them with their average depth bin.

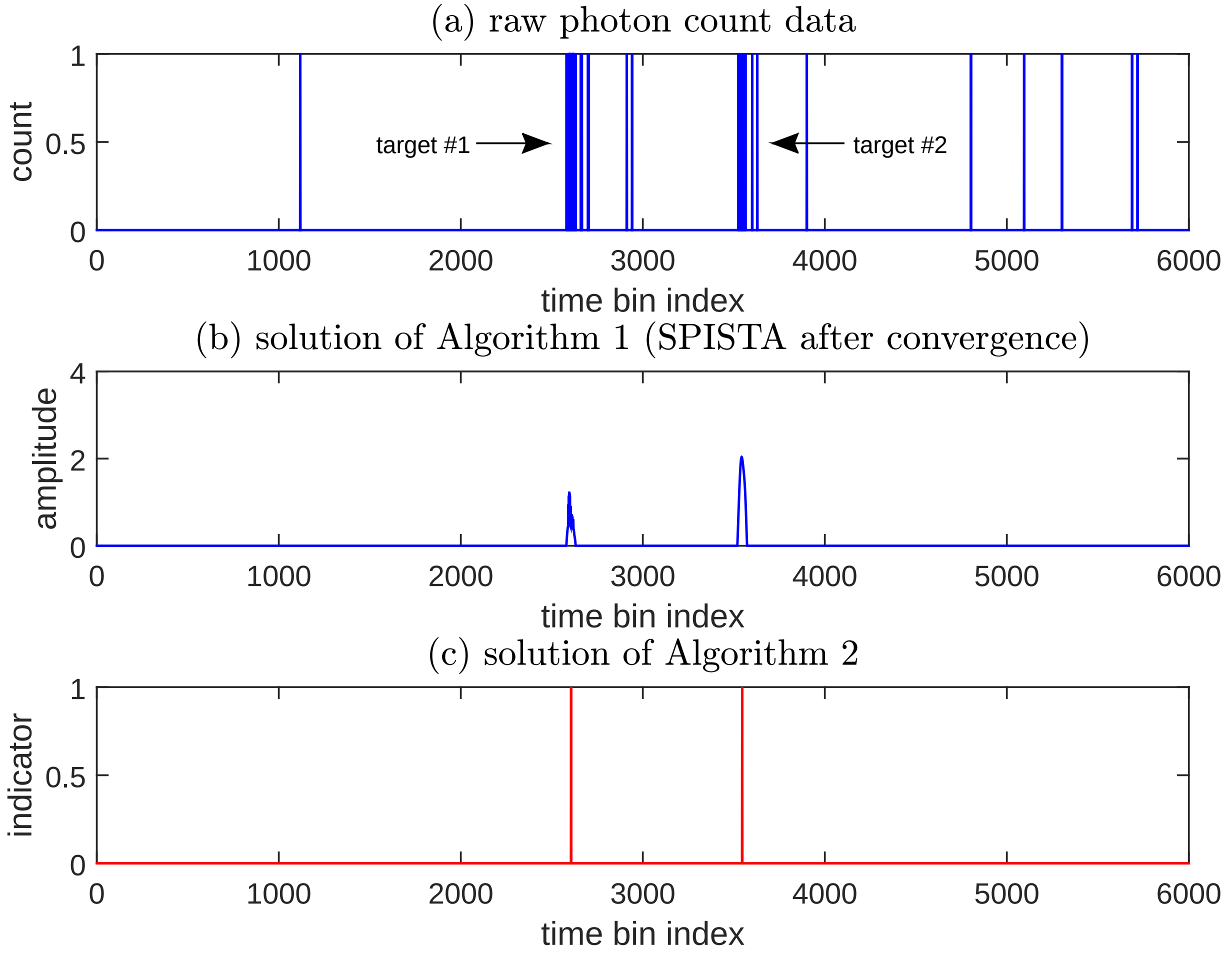


Figure 3-4: Illustration of steps of Algorithm 3 using experimental photon-count data for the single-pixel multi-depth example of partially-occluding reflectors in Fig. 3-9. (a) The raw photon count vector $\mathbf{y}$ from a pixel that contains two reflectors at around time bins 2600 and 3500. Other than the photon detections describing the two targets of interest, we observe extraneous photon detections from background and dark counts. (b) The output solution of SPISTA in Algorithm 2. Note that the extraneous background and detector dark counts are suppressed. (c) The final solution of Algorithm 3 that groups depths of SPISTA output.

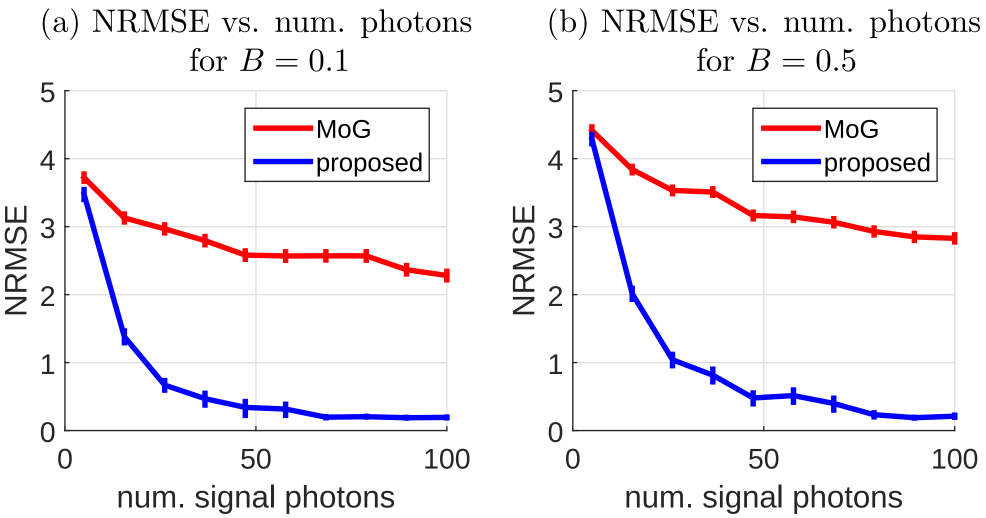


Figure 3-5: Simulated performance of MoG-based method and proposed framework of Algorithm 3 in recovering signals with $K = 2$ for two different background levels. Signal photon detections are detections originating from scene response and do not include the background-light plus dark-count detections. Note that the units of NRMSE are in meters, after being normalized by the pulsewidth; 1 NRMSE corresponds to an unnormalized mean-squared error of $cT_p/2 = 4.5$ cm. The plots also include error bars indicating the ±1 standard errors.

error (NRMSE) to quantify the recovery performance for the two-reflector signal:

$$\text{NRMSE}\left(\{d_1, d_2\}, \{\hat{d}_1, \hat{d}_2\}\right) = \frac{1}{cT_p/2}\sqrt{\mathbb{E}\left[\frac{1}{2}\left(\left(d_1 - \hat{d}_1\right)^2 + \left(d_2 - \hat{d}_2\right)^2\right)\right]}, \quad (3.20)$$

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 3-5 shows Monte Carlo simulated performance results of pixelwise two-reflector 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 observation samples, we used the EM algorithm. Simulation parameters were set as the following: the number of detector time bins $m = 100$, the number of discretized depth bins $n = 100$, RMS pulsewidth $T_p = 0.3$ ns, and bin width $\Delta = 1$ ns. The pulse shape was discrete Gaussian. The number of Monte Carlo trials for the simulation was 2000. For each Monte Carlo trial, two entries out of n were generated in a uniformly random fashion. (A two-path reflector profile was chosen from n-choose-2 combinations at random.) Both selected entries were set to 1, in order to simulate two reflectors with unit reflectivities. For our algorithm, we chose the regularization parameter $\tau = B$, based on a heuristic that higher penalty is required for higher background levels. We chose the convergence parameter $\delta = 0.01$, and the residual filtering parameter $\epsilon = 0.1$. Also, our initialization $\hat{\mathbf{x}}^{(0)}$ was chosen to be $\mathbf{y}$. We see that for both low and high background levels, our proposed framework uniformly outperforms the existing MoG method for various numbers of photons backreflected from the scene. For example, for both $b = 0.1$ and $b = 0.5$, the difference in RMSE between our framework and MoG is around 2 given 10 signal photon detections. This translates to RMS depth error reduction of 9 cm, since 1 NRMSE equals $cT_p/2 = 4.5$ cm. Also, our method successfully achieves sub-pulsewidth depth accuracy (NRMSE less than 1) when the number of signal photons exceeds $\sim$30, while the MoG method fails to do so.

In this simulation, we required an average of 85 SPISTA iterations per pixel. The average per pixel processing time of Algorithm 3 was measured to be $\sim$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 $\sim$0.019 seconds. All processing was done using a laptop with Intel(R) Core(TM) i7-4500u CPU running at 1.80 GHz.

Simulations of 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 3-6 shows simulation results that describe how the number of reflectors estimated by our algorithm (number of nonzero elements in $\hat{\mathbf{x}}_{\mathrm{OPT}}$) 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

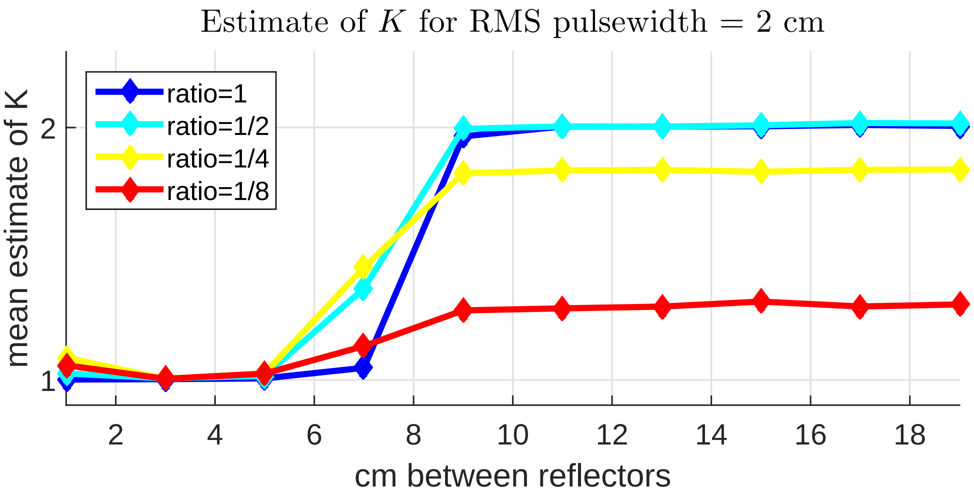


Figure 3-6: Simulated results of mean estimates of the number of reflectors produced by Algorithm 2 at a pixel, when the RMS pulsewidth is set to give $cT_p = 2$ cm. Here we show plots when the reflectivity ratio between the first and second target is (a) 1 (blue line), (b) 1/2 (sky blue line), (c) 1/4 (yellow line), and (d) 1/8 (red line).

at 100 by changing N, the number of illumination trials, in Eq. (3.9). 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, $\delta = 10^{-2}$, $\epsilon = 0.1$, $\tau = 0.5$, and $b = 0$. Algorithm 2 was applied to 2000 independent simulated experiments and the results in Fig. 3-6 show the mean number of estimated reflectors. In Fig. 3-6, for all relative reflectivity settings, we observe that if the distance between two reflectors is too small (around $3cT_p$), then our algorithm falsely recognizes two reflectors as a single reflector. Even when there is a large separation between reflectors (larger than $3cT_p$), if the ratio between target reflectivities is too small as 1/8, then the number of reflectors is most likely to be falsely recognized as 1 by our Algorithm 3.

Effect of ℓ_1 relaxation on estimating number of reflectors: Our framework relies on the relaxation of the ℓ_0-norm constraint into a ℓ_1-norm constraint to formulate a convex optimization problem. By doing this, our algorithm does not explicitly require a parameter specifying $\hat{K}$, which is an estimate of number of scene reflectors, as is the case for EM algorithms for MoG estimation. Instead, our framework employs τ, a continuous regularization parameter, to set the degree of penalizing non-sparse estimates of $\mathbf{x}$.

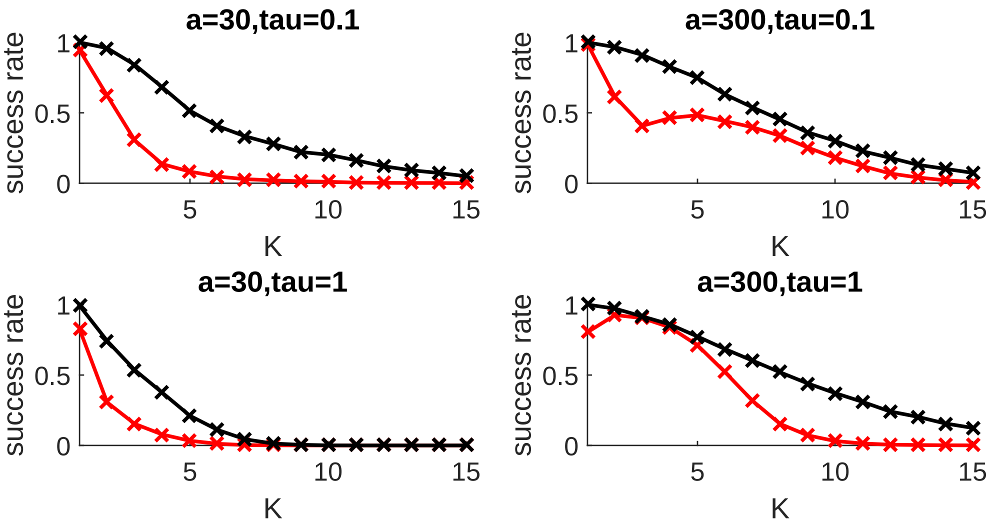


Figure 3-7: For different values of signal amplitude a (reflectivity multiplied by the peak value of pulse waveform) and regularization parameter τ, the plots show success-1 in black and success-2 in red. Note that, by definition, success-1 upper-bounds success-2.

Our use of a regularization parameter naturally leads to the following question. What is the range of number of reflectors that a single τ value can cover in our algorithm? In other words, how robust is our algorithm to using a single τ value for accurate imaging of both a scene with two reflectors and a scene with ten reflectors, for example? To answer this question, we first define two success rates for our multi-depth estimate:

$$\mathsf{success\text{-}1} = \mathbb{E}[\mathbb{1}_{K=\hat{K}}], \tag{3.21}$$

$$\mathsf{success\text{-}2} = \mathbb{E}[\mathbb{1}_{\mathrm{supp}(\mathbf{x})=\mathrm{supp}(\hat{\mathbf{x}}_{\mathrm{OPT}})}]. \tag{3.22}$$

The values of success-1 and success-2 measure the probabilities that our framework will output the correct number of reflectors and the correct depth values, respectively. Note that both success-1 and success-2 have maximum values of 1. Moreover, success-1 must always equal or exceed success-2, since if the depth estimates are correct, then the estimated the number of reflectors must be correct.

Figure 3-7 shows success-1 and success-2 over increasing numbers of simulated reflectors at a pixel (K ranging from 1 to 15) for two different regularization parameters τ and signal amplitude values a for $B = 0$. In this simulation, we chose $m = n = 300$, $\delta = 10^{-5}$, $\epsilon = \max(\hat{\mathbf{x}}_{\text{SPISTA}})/5$. and a Gaussian pulse illumination with RMS duration equal to 1 detector time bin. For each subplot, success-1 and success-2 were computed using Monte Carlo simulations of 3000 trials, where each trial randomly sampled the reflector depths from n-choose-K settings as in the previous sections. We make two observations from our simulation plots.

1. There is a graceful degradation in the success probabilities as the number of reflectors increase for all τ and a values. This is in contrast to the conventional methods like EM, as they will give rise to non-zero success probabilities only when $K = \hat{K}$, and not for a range of K's.
2. The optimal values of τ depend on the amplitude a. For example, if the amplitude is low ($a = 30$), then we see that choosing $\tau = 0.1$ is better than setting $\tau = 1$, as it leads to higher success probabilities for the entire range of simulated reflector numbers. On the other hand, if the amplitude is high ($a = 300$), then $\tau = 1$ seems to be a more optimal choice than $\tau = 0.1$, at least for $K \in \{2 \ldots, 6\}$.

We also demonstrated the performance of the proposed multi-depth imager using the experimental single-photon imaging setup employed in Chapter 2. 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. In our experiments, 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 7.9×10^{-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 2×10^4 was used for detection of photons. A PicoQuant model HydraHarp 400 time-correlator was used to record the detection times of individual photons. We injected extraneous background light using an incandescent lamp. We chose to use the raster-scanning setup simply due to availability of equipment; 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. The experimental data collection was performed by collaborator F. Xu.

For all measurements, we obtained the pulse waveform matrix $\mathbf{S}$ 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 $4 \times 2/c$ 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 $\mathbf{S}$. The value of B was obtained by taking a baseline measurement of the incandescent light with laser light not present in scene. Code and data are available in [45].

Experimental imaging through a partially-reflective object: Figure 3-8 shows experimental results of imaging through a partially-reflective object, which is the multi-depth imaging scenario in Fig. 3-1(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 100×100 in this experiment.

We see that the existing MoG-based method fails to recover useful depth features

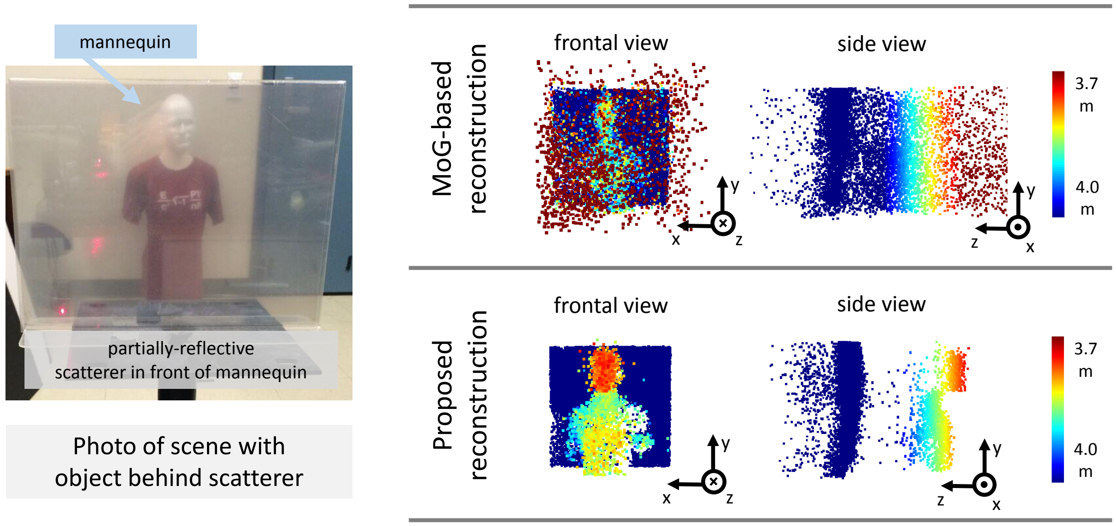


Photo of scene with
object behind scatterer

Figure 3-8: (Left) Photograph of the mannequin placed behind a partially-scattering object from the single-photon imager's point of view. (Right) Experimental results for estimating the mannequin's depth map through the partially-reflective material using MoG-based and our estimators, given that our imaging setup is at $z = 0$. The EM algorithm for MoG fitting used $K = 2$. Our multi-depth results were generated using the parameters $\tau = 0.1$, $\delta = 10^{-4}$, $\epsilon = 0.1$, and $\hat{\mathbf{x}}^{(0)} = \mathbf{S}^T\mathbf{y}$.

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 the ground truth depth map of the mannequin by using the log-matched filtering solution [46] 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. Then, we computed that while the conventional MoG solution gave 48.7 cm of RMS depth error, 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 depth pixels 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 processing times for MoG and the proposed methods are

	Per-pixel processing time	Total processing time
MoG	0.003 seconds	30 seconds
Proposed	0.036 seconds	360 seconds

Table 3.1: Processing times for MoG and proposed methods for experimental imaging through a partially-reflecting object.

given in Table 3.1.

Experimental imaging of 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 D. Venkatraman for the first-photon imaging project [12] with the same raster-scanning setup. Our experimental scene consisted of a sunflower and the background wall as shown in Figure 3-9. This experiment models the multi-depth imaging scenario in Figure 3-1(b). Here we have a higher raster-scanning resolution of 300×300, 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's petals partially occlude the wall behind it. In our data, the probabilities of a photon originating from the scene and from background light plus 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 3-10 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 multipath imager successfully distinguishes the sunflower's petals from its leaves and the wall behind it, even though there exists 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 the 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

	Per-pixel processing time	Total processing time
MoG	0.0057 seconds	8.5 minutes
Proposed	0.0014 seconds	2 minutes

Table 3.2: Processing times for MoG and proposed methods for experimental imaging of a partially-occluding object.

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 processing times for MoG and the proposed methods are given in Table 3.2. Note that the MoG method in both depth accuracy and time complexity performs worse in the partial-occluder imaging experiment than in the previous partial-reflector imaging experiment. The reason is because, in the previous experiment, the number of reflectors was fixed to be 2 (first bounce from plexiglass and second bounce from mannequin scene) for all pixels, but it varies to be either 1 or 2 from pixel to another here. Because the MoG method relies on recovering two modes, the model mismatch is larger in the partial-occluder imaging experiment, and thus leading to poorer performance.

3.6 Summary and Discussion

In this chapter, we presented a robust framework for reconstructing a scene's multi-depth profile using low light-level data from a single-photon detector. Our novel imaging method accurately models the single-photon detection statistics from multiple reflectors in the scene, while exploiting the fact that multipath profiles can be expressed as sparse signals. Using our signal model, the multipath estimation problem is one of sparse deconvolution. We designed an algorithm inspired by ISTA that reaches the globally optimal solution of the sparse deconvolution problem's convex

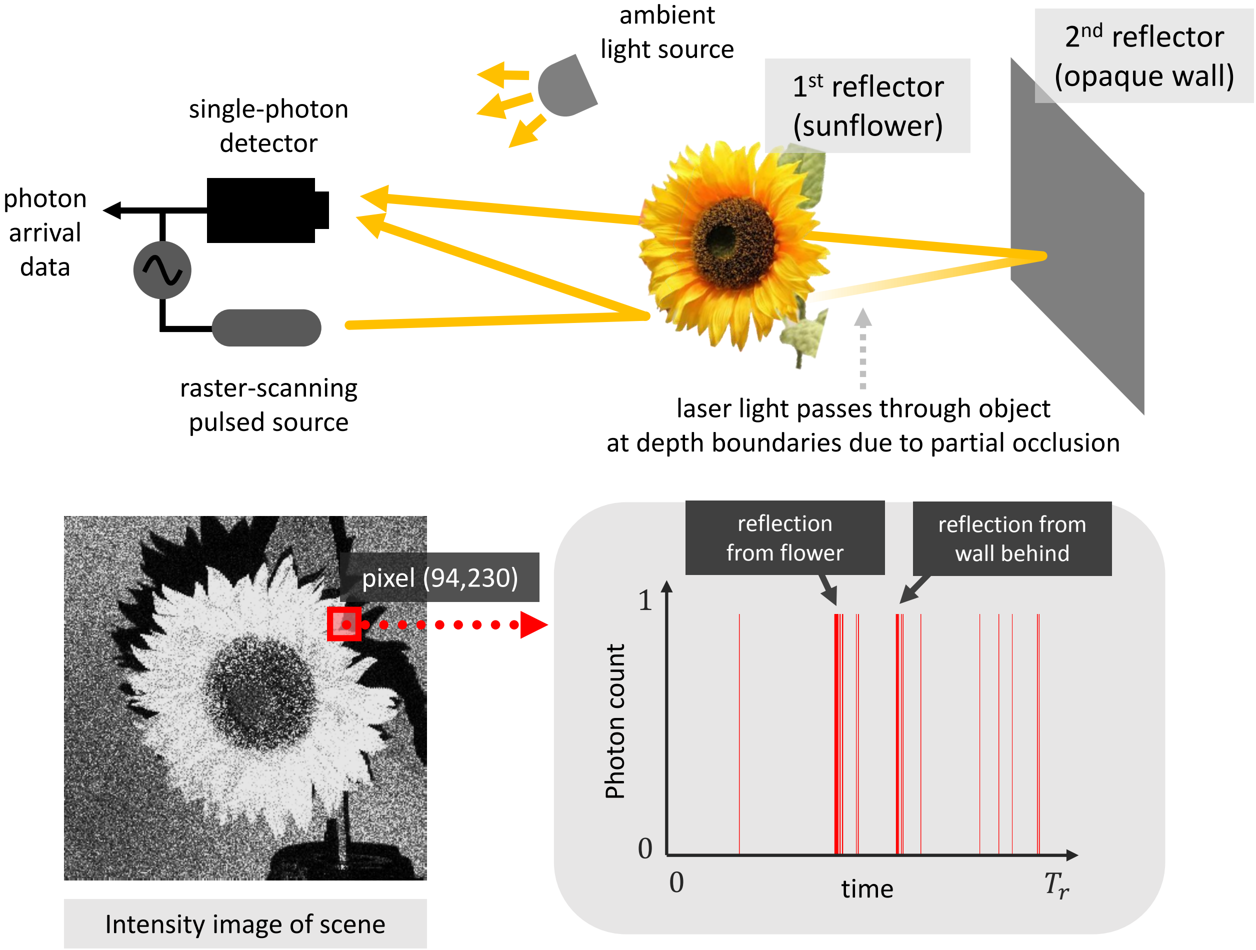

ambient light source
1st reflector (sunflower)
2nd reflector (opaque wall)
single-photon detector
photon arrival data
raster-scanning pulsed source
laser light passes through object at depth boundaries due to partial occlusion
pixel (94,230)
Intensity image of scene
reflection from flower
reflection from wall behind
Photon count
1
0
0
time
T_r

Figure 3-9: Single-photon imaging setup for estimating multi-depth from partial occlusions at depth boundary pixels. Sample data from 38 photon detections is shown below for the pixel $(94, 230)$ where partial occlusions occur. We show experimental results of multi-depth recovery for this scene using the MoG-based and our methods in Figure 3-10.

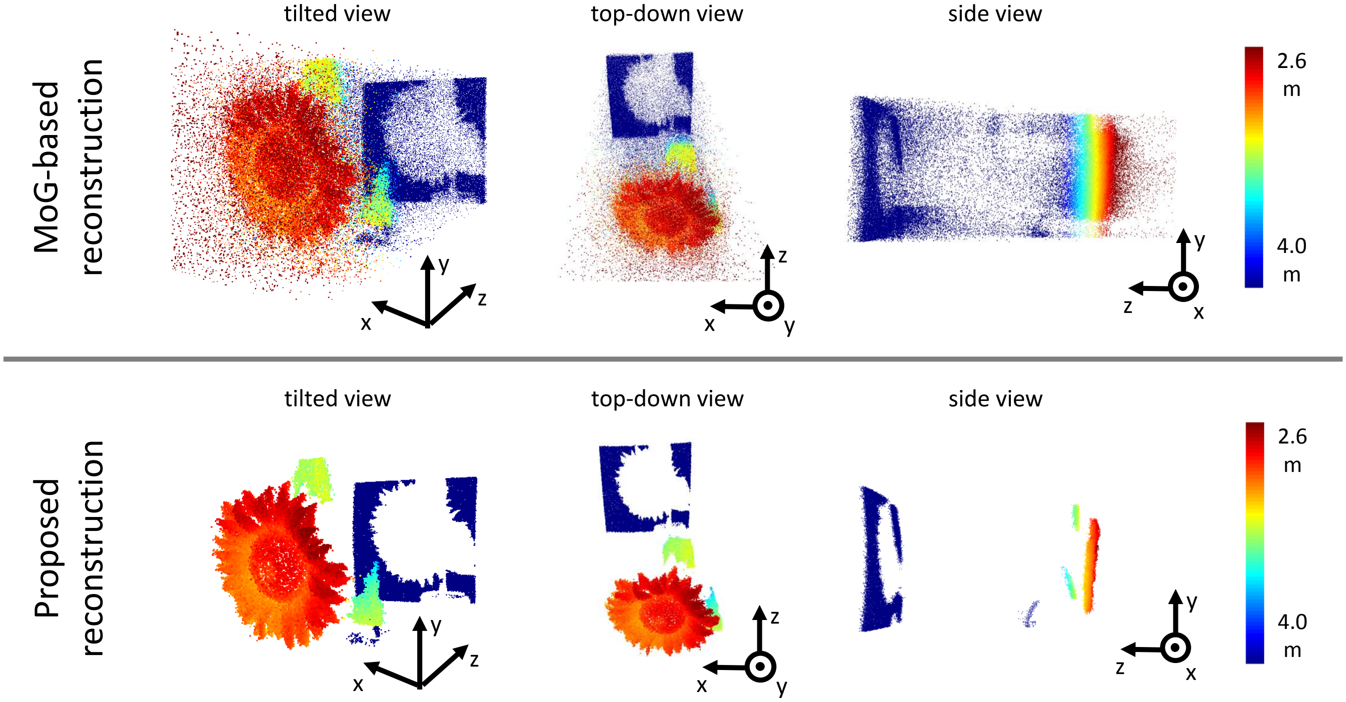

MoG-based reconstruction
tilted view
top-down view
side view
2.6 m
4.0 m
x
y
z
Proposed reconstruction
tilted view
top-down view
side view
2.6 m
4.0 m
x
y
z

Figure 3-10: Experimental results of depth reconstruction for a 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. The EM algorithm for MoG fitting used $K = 2$. Our multi-depth results were generated using the parameters $\tau = 0.2$, $\delta = 10^{-4}$, $\epsilon = 0.1$, and $\hat{\mathbf{x}}^{(0)} = \mathbf{S}^T\mathbf{y}$.

relaxation in a computationally efficient manner, relative to brute force search methods. For example, the average runtime of our algorithm to estimate the multiple depths of a pixel of the mannequin scene in Figure 3-8 was ∼36 ms.

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 [47].

For future work, it is of interest to study how applying post-processing techniques, such as 3D point cloud filtering, can improve 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 and thus enable a more accurate multi-depth reconstruction framework.

Chapter 4

Array-Based Time-Resolved Imaging

4.1 Overview of Problem

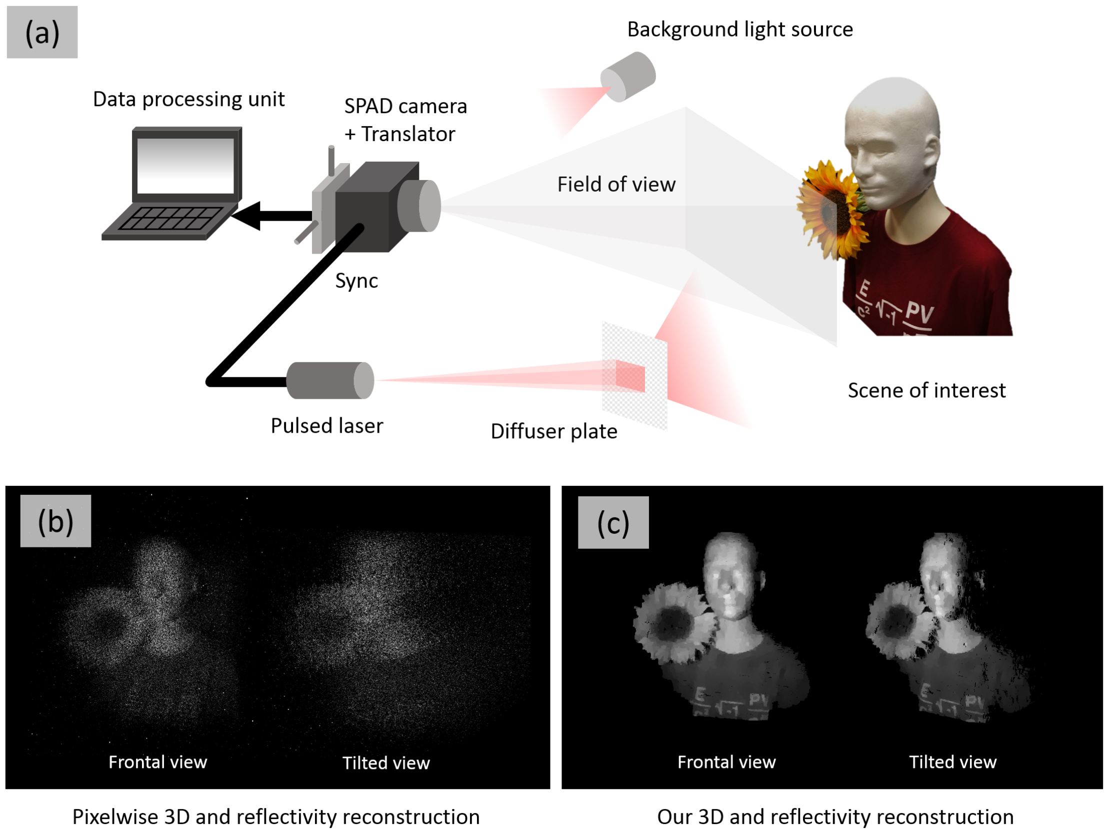
As demonstrated in the previous two chapters, accurately estimating a scene's 3D structure at low light levels can be done using time-of-flight data collected from a raster scanning source and a time-correlated single SPAD detector. However, unlike conventional raster-scanning imagers, which use single-pixel photon counters capable of ∼10-ps time tagging, a typical array setup requires highly parallelized time-to-digital conversions that limit its photon time-tagging accuracy to ∼1 ns [48]. In this chapter, we develop an array-specific algorithm that converts coarsely time-binned photon detections to highly accurate scene depth and reflectivity by exploiting both transverse smoothness and longitudinal sparsity of natural scenes.

Prior art: The state-of-the-art depth imaging technique in terms of achieving high photon-efficiency was established by the authors of first-photon imaging (FPI) [12], who demonstrated accurate 3D and reflectivity recovery from the *first* detected photon at each pixel. Their experimental imaging setup, which was employed for our imaging purposes in the previous two chapters, used a raster scanning source and a time-correlated single SPAD detector. Using this setup, the first-photon imaging framework acquired exactly one photon detection at every pixel, making each pixel's acquisition time a random variable. Consequently, FPI is not applicable to operation using a SPAD-detector array [49–54]—all of whose pixels must have the same acquisition time—thus precluding FPI's reaping the dramatic image-acquisition speedup that such arrays afford [1, 32, 55]. Although there have been extensions of FPI to the fixed acquisition-time operation needed for array detection [2, 14], both their theoretical modeling and experimental validations were still limited to raster scanning with a single SPAD detector. As a result, they ignored the limitations of currently available SPAD-array systems—much poorer time-tagging performance and pixel-to-pixel variations of SPAD properties—implying that these initial fixed acquisition-time (pseudo-array) frameworks will yield sub-optimal depth and reflectivity reconstructions when used with low-light experimental data from an actual SPAD array.

Summary of our approach: Here we propose and demonstrate a photon-efficient 3D structure and reflectivity imaging technique that can deal with the aforementioned constraints that SPAD array imagers impose. We give the first experimental demonstration of accurate time-correlated SPAD-array imaging of natural scenes obtained from ∼1 detected signal photon per pixel on average. Unlike prior work, our framework achieves high photon efficiency by exploiting the scene's structural information in both the transverse and the longitudinal domains to censor extraneous (background-light plus dark-count) detections from the SPAD array detector. Because our new imager achieves highly photon-efficient imaging in a short data-acquisition time, it paves the way for dynamic and noise-tolerant active optical imaging applications such as monitoring an ultrafast biological process that is sensitive to light.

4.2 Single-Photon Imaging Setup

Our experimental setup is illustrated in Figure 4-1. The illumination source was a pulsed laser diode (PicoQuant LDH series with a 640 nm center wavelength) whose original output-pulse duration was increased to a root-mean-square (RMS) value of $T_p \approx 1\,\text{ns}$. The laser diode was pulsed at a $T_r \approx 50\,\text{ns}$ repetition period set by the SPAD array's trigger output. A diffuser plate spatially spread the laser pulses


(a)
Background light source
Data processing unit
SPAD camera
+ Translator
Field of view
Sync
Scene of interest
Pulsed laser
Diffuser plate
(b)
Frontal view
Tilted view
(c)
Frontal view
Tilted view

Pixelwise 3D and reflectivity reconstruction

Our 3D and reflectivity reconstruction

Figure 4-1: Single-photon array imaging framework. (a) SPAD-array imaging setup. A repetitively-pulsed laser flood-illuminates the scene of interest. Laser light reflected from the scene plus background light is detected by a SPAD camera. Photon detections at each pixel are time tagged relative to the most recently transmitted pulse and recorded. The raw photon-detection data were processed on a standard laptop computer to recover the scene's 3D structure and reflectivity. (b) Example of 3D structure and reflectivity reconstruction of mannequin and flower scene using the baseline single-photon imager from [1]. (c) Example of 3D structure and reflectivity reconstruction of mannequin and flower scene from our processing. Large portions of the mannequin's shirt and facial features that were not visible in the baseline image are revealed using our method. Both results in (b) and (c) were generated using an average of $\sim$1 detected signal photon per pixel.

to flood illuminate the scene of interest. An incandescent lamp injected unwanted background light into the camera. The lamp's power was adjusted so that (averaged over the region that was imaged) each detected photon was equally likely to be due to signal (back-reflected laser) light or background light. A standard Canon FL-series photographic lens focused the signal plus background light on the SPAD array. Each photon detection from the array was time tagged relative to the time of the most recently transmitted laser pulse and recorded.

Our SPAD array [52, 54], covering a 4.8×4.8 mm footprint, consists of 32×32 pixels of fully independent Si single-photon avalanche diodes and CMOS-based electronic circuitry that includes a time-to-digital converter for each SPAD detector. The SPAD within each $150 \times 150\,\mu$m pixel has a 30-μm-diameter circular active region, giving the array a 3.14% fill factor. At the 640 nm operating wavelength each array element's photon detection efficiency is $\sim$20% and its dark-count rate is $\sim$100 Hz at room temperature. To extend the region that could be imaged and increase the number of pixels, we used multiple image scans to form a larger-size composite image. In particular, we mounted the SPAD array on a feedback-controlled, two-axis motorized translation stage, to produce images with $N_x \times N_y = 384 \times 384$ pixels.

The SPAD array has a $\Delta = 390$ ps time resolution set by its internal clock rate. We set each acquisition frame length to $65\,\mu$s, with a gate-on time of $16\,\mu$s and a gate-off time of $49\,\mu$s for limiting power dissipation of the chip and for data transfer. At the start of each frame the SPAD array was set to trigger the laser to generate pulses at a $\sim$20 MHz repetition rate. Hence, in the $16\,\mu$s gate-on time of each frame, about 320 pulses illuminated the scene. (See Appendix E for more details on the SPAD detector array operation.)

4.3 Forward Imaging Model

We define $\mathbf{Z}, \mathbf{A} \in \mathbb{R}^{N_x \times N_y}$ to be the scene's 3D structure and reflectivity that we aim to recover, and we let $\mathbf{B} \in \mathbb{R}^{N_x \times N_y}$ be the average rate of background-light plus dark-count detections. Flood illumination of the scene at time $t = 0$ with a photon-flux pulse $s(t)$ then results in the following Poisson-process rate function for (i,j)-th pixel of the composite image:

$$\lambda_{i,j}(t) = \eta_{i,j}\, \mathbf{A}_{i,j}\, s(t - 2\mathbf{Z}_{i,j}/c) + \mathbf{B}_{i,j}, \qquad t \in [0, T_r), \tag{4.1}$$

where $\eta_{i,j} \in (0,1]$ is the (i,j)th detector's photon detection efficiency, and c is the speed of light. Observe that, unlike the raster-scanner's imaging model used in the previous two chapters (Eq. (2.2)), the quantum efficiency and background plus dark-count response are pixel-dependent in our array setup.

Fabrication imperfections of the SPAD array cause some pixels to have inordinately high dark-count rates ($\mathbf{B}_{i,j} \gg \eta_{i,j}\, \mathbf{A}_{i,j} \int_0^{T_r} s(t)\, dt$), making their detection times uninformative in our imaging experiments because they are predominantly from dark counts. Thus we performed camera calibration to determine the set $\mathcal{H}$ of these "hot pixels " (2% of all pixels in our experiment reported below) so that their outputs could be ignored in the processing of the imaging data.

Recall our definition of the total number of time bins as $m = T_r/\Delta$ and let $\mathbf{Y}_{i,j,k}$ be the observed number of photon counts in the kth time bin for pixel (i,j) after N_s pulsed-illumination trials. By the theory of low-flux photon counting from Eq. (1.3), we have that $\mathbf{Y}_{i,j,k}$'s statistical distribution is

$$\mathbf{Y}_{i,j,k} \sim \text{Poisson}\left(N_s \int_{(k-1)\Delta}^{k\Delta} \lambda_{i,j}(t)\ dt \right), \tag{4.2}$$

for $k = 1, 2, \ldots, m$, where we have assumed that the pulse repetition period is long enough to preclude pulse aliasing artifacts. Again, we operate in a low-flux condition such that $\sum_{k=1}^{m} \mathbf{Y}_{i,j,k}$, the total number of detections at pixel (i,j), is much less than N_s, the total number of illumination pulses, to avoid dead time effects of single-photon detector array.

Our imaging problem is then to construct accurate image estimates, $\hat{\mathbf{A}}_{\text{array}}$ and $\hat{\mathbf{Z}}_{\text{array}}$, of the scene's reflectivity $\mathbf{A}$ and 3D structure $\mathbf{Z}$ using the sparse photon detection data $\mathbf{Y} \in \mathbb{R}^{N_x \times N_y \times m}$.

4.4 Solving the Inverse Problem

In the low-flux regime, wherein there are very few detections and many of them are extraneous, an algorithm that relies solely on the aforementioned pixelwise photodetection statistics has very limited robustness. For example, we saw from Figure 2-6 of Chapter 2 that accurate pixelwise depth imaging methods can be robust when we detect more than 10 photons per pixel. Instead of only using the pixelwise photodetection statistics, we aim to achieve higher photon efficiency by exploiting the structural constraints that are present in most natural scenes. Our 3D structure and reflectivity reconstruction algorithm exploits such scene constraints in three steps.

- **Step 1 – Estimating scene reflectivity:** Using the statistics of the raw photon data cube $\mathbf{Y}_{i,j,k}$, we have that $\mathbf{C}_{i,j}$, the total number of photon detections at pixel (i,j), is the sum of statistically independent Poisson random variables and hence it is itself Poisson, i.e.,

$$\begin{aligned}\mathbf{C}_{i,j} \equiv \sum_{k=1}^{m} \mathbf{Y}_{i,j,k} &\sim \text{Poisson}\left(\sum_{k=1}^{m} N_s \int_{(k-1)\Delta}^{k\Delta} \lambda_{i,j}(t)\, dt\right) \qquad (4.3)\\ &= \text{Poisson}\left(N_s \int_{0}^{T_r} \lambda_{i,j}(t)\, dt\right). \qquad (4.4)\end{aligned}$$

Using the expression given in the previous section for $\lambda_{i,j}(t)$ we get

$$N_s \int_0^{T_r} \lambda_{i,j}(t)\, dt = N_s(\eta_{i,j}\mathbf{A}_{i,j}S + m\Delta\mathbf{B}_{i,j}), \qquad (4.5)$$

where $S = \int_0^{T_r} s(t)\, dt$. We can thus write $\mathbf{C}_{i,j}$'s probability mass function in terms of $\mathbf{A}_{i,j}$, as

$$f_C(\mathbf{C}_{i,j}; \mathbf{A}_{i,j}) = \frac{[N_s(\eta_{i,j}\mathbf{A}_{i,j}S + m\Delta\mathbf{B}_{i,j})]^{\mathbf{C}_{i,j}} \exp[-N_s(\eta_{i,j}\mathbf{A}_{i,j}S + m\Delta\mathbf{B}_{i,j})]}{\mathbf{C}_{i,j}!}, \qquad (4.6)$$

for $\mathbf{C}_{i,j} = 0, 1, 2, \ldots$. We let $\mathcal{L}_A(\mathbf{A}_{i,j}; \mathbf{C}_{i,j}) = -\log f_C(\mathbf{C}_{i,j}; \mathbf{A}_{i,j})$ be the negative log-likelihood of $\mathbf{A}_{i,j}$, and, for convenience, we set $S = 1$, because reflectivity images can be scaled arbitrarily.

For robust reflectivity estimation, we first ignore data from the hot-pixel set $\mathcal{H}$ and then take our reflectivity image to be the non-negative $\hat{\mathbf{A}}_{\text{array}}$ that minimizes the negative log-likelihood of the remaining pixels plus a term that penalizes non-smoothness, i.e., we find $\hat{\mathbf{A}}_{\text{array}}$ by solving the following optimization problem:

$$\begin{aligned} \hat{\mathbf{A}}_{\text{array}} &= \arg\min_{\mathbf{A}} \left[\left(\sum_{(i,j)\notin\mathcal{H}} \mathcal{L}_A(\mathbf{C}_{i,j}; \mathbf{A}_{i,j}) \right) + \tau_A \, \|\mathbf{A}\|_{\text{TV}} \right] \qquad (4.7) \\ &\text{subject to} \qquad \mathbf{A}_{i,j} \geq 0, \qquad \text{for all } i, j. \end{aligned}$$

Here, τ_A is the regularization parameter penalizing non-smoothness in the reflectivity image and $\|\cdot\|_{\text{TV}}$ is the total-variation (TV) norm, which is known to be effective in characterizing the spatial correlations in natural images [56, 57]. The objective function in (4.7) is convex, so computing the reflectivity image can be done efficiently using projected-gradient methods [58].

In the experiment described below, we chose the regularization parameter by first generating reflectivity images with $\tau_A = \tau'_A/(1-\tau'_A)$ for $\tau'_A \in \{0.1, 0.2, \ldots, 0.9\}$ and then choosing the τ_A value that minimized ℓ_2 distance from the ground truth reflectivity. In practical imaging scenarios, where the ground truth is not available, the regularization parameter is typically chosen using cross-validation methods [47].

- **Step 2 – Censoring extraneous detections:** Unlike reflectivity estimation, whose optimization problem is convex even in the presence of extraneous detections from background light and dark counts, the optimization problem for depth estimation when there are extraneous detections is not convex, and hence not computationally tractable [2]. Thus, in order to obtain a tractable problem for depth estimation, we first censor extraneous detections.

Let $\mathbf{z}_\Delta = \text{hist}(\mathbf{Z}_\Delta)$ be the one-dimensional histogram of the scene's depth values, $\mathbf{Z}$, in m bins of width $c\Delta/2$ m. A natural assumption for many applications, such as ground-to-air LIDAR, is that the opaque reflective objects in scene are individually clustered in depth, so that $\mathbf{z}_\Delta$ is sparse. Thus, by locating the non-zero elements of $\mathbf{z}_\Delta$ we can be very effective in censoring extraneous detections, because they are uniformly distributed over $[0, T_r)$.

We obtain a proxy estimate for $\mathbf{z}_\Delta$ using our raw photon data cube as follows. We compute $\tilde{\mathbf{z}}_\Delta = \text{hist}(c\mathbf{T}/2)$, where $\mathbf{T}$ is the set of detection times obtained by ignoring hot pixels, summing $\mathbf{Y}_{i,j,k}$ over its transverse-coordinate indices, and then subtracting out the average background-light plus dark-count response, viz.,

$$\tilde{\mathbf{z}}_\Delta = \sum_{(i,j)\notin\mathcal{H}} (\mathbf{Y}_{i,j,k} - \Delta\mathbf{B}_{i,j}), \qquad k = 1, \ldots, m. \tag{4.8}$$

We note that $\tilde{\mathbf{z}}_\Delta$ is a noisy version of $\mathbf{z}_\Delta$ convolved with $s_z \propto s(2z/c)$, where $s(t)$ is the laser's pulse shape and $\int s_z \, dz = 1$. Thus, the optimization problem that solves for the sparse signal $\hat{\mathbf{z}}_\Delta$ from $\tilde{\mathbf{z}}_\Delta$ is

$$\begin{aligned} \hat{\mathbf{z}}_\Delta \;=\; \underset{\mathbf{z}_\Delta}{\arg\min} \qquad & \sum_{k=1}^{K} \|\tilde{\mathbf{z}}_\Delta - (\mathbf{s}_z * \mathbf{z}_\Delta)_k\|_2^2 \\ \text{subject to} \qquad & \|\mathbf{z}_\Delta\|_0 = K, \qquad (\mathbf{z}_\Delta)_k \geq 0, \text{ for all } k, \end{aligned} \tag{4.9}$$

where $\mathbf{s}_z$ is s_z discretized into bins of width $c\Delta/2$, $*$ is the discrete one-dimensional convolution operator, and K is the number of non-zero elements in the reconstructed $\hat{\mathbf{z}}_\Delta$. In our experiments, we chose $K = 2$ because our scene of interest consisted of two objects (i.e., a mannequin and a flower) with a small separation in depth. In practice, one can always use a K value that overestimates the true number of depth clusters, since we are not interested in locating the exact depth cluster centroids; we are only using Step 2 to coarsely identify relevant depth regions in space to filter out extraneous detections.

The optimization program in (4.9) is a discrete sparse-deconvolution problem that we solve approximately using a modified version of orthogonal matching pursuit (OMP) [59]. OMP is a fast and robust sparse signal-pursuit algorithm, which we modified to orthogonally project the intermediate solution onto the non-negativity constraint set at every iteration.

Let $\text{supp}(\hat{\mathbf{z}}_\Delta) = \{k : \hat{\mathbf{z}}_{\Delta_k} \neq 0, k = 1, 2, \ldots, m\}$ be the support of the solution to (4.9). Because signal-photon detections at pixel (i, j) are mostly confined to T_p s around $2Z_{i,j}/c$, we use

$$\mathcal{S}_{T_p} = \{k : |k - k'| < T_p/\Delta, \text{ for any } k' \in \text{supp}(\hat{\mathbf{z}})_\Delta\}. \tag{4.10}$$

to generate the photon-sparse data cube after censoring, denoted $\tilde{\mathbf{Y}} \in \mathbb{R}^{N_x \times N_y \times m}$, via

$$\tilde{\mathbf{Y}}_{i,j,k} = \begin{cases} \mathbf{Y}_{i,j,k}, & \text{if } k \in \mathcal{S}_{T_p} \text{ and } (i,j) \notin \mathcal{H}, \\ 0, & \text{otherwise.} \end{cases} \tag{4.11}$$

Our extraneous-detection censoring mechanism differs significantly from the ones used in first-photon imaging [12] and pseudo-array imaging [13]. Those prior works exploited transverse correlations in a scene to discriminate extraneous detections from signal detections, whereas our method relies on the longitudinal sparsity of natural scenes when they are subjected to the coarse time-binning inherent in a SPAD array. Figure 4-2 shows the increased censoring of extraneous detections—for the mannequin and flower dataset—that results from using our method instead of the pseudo-array method.

- **Step 3 – Estimating scene depth:** Assuming that $\tilde{\mathbf{Y}}_{i,j,k}$ contains only the signal photon detections, the detection time-bin $T_{i,j}$ at pixel $(i, j) \notin \mathcal{H}$ has

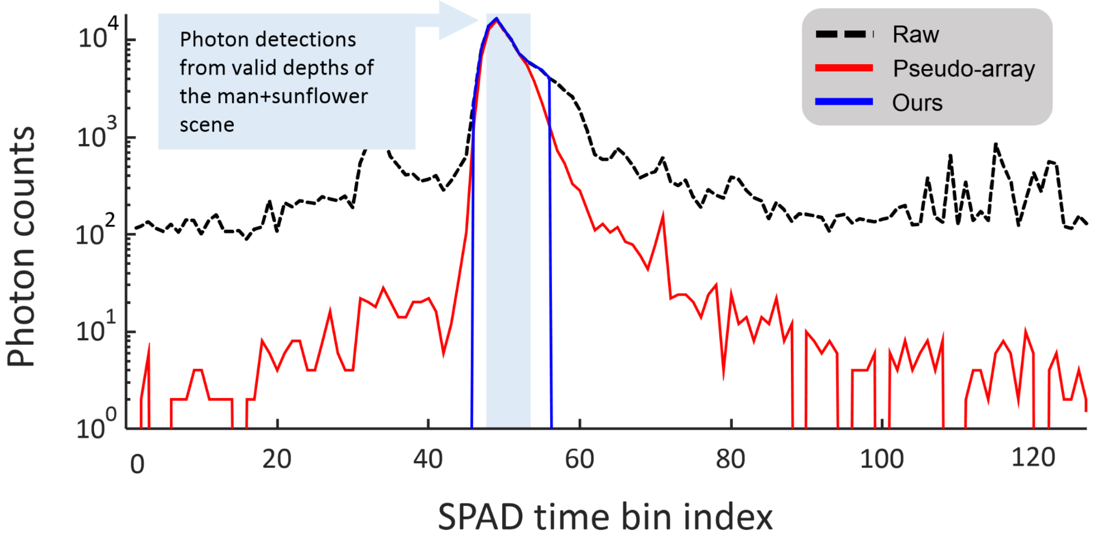


Figure 4-2: Photon-count histograms after censoring extraneous detections for the state-of-the-art pseudo-array method [2] (red) and our method (blue). The raw-data photon histogram is also given for reference (dashed black). The blue block indicates the ground truth depth values of objects in the scene scaled by $c/2$. By exploiting the scene's longitudinal sparsity, our method rejects more extraneous detections than does the pseudo-array method, which relies on transverse correlations. The greater the number of extraneous detections that survive censoring, the greater the amount of regularization that will occur in depth estimation, which will lead, in turn, to oversmoothing the depth image.

probability mass function

$$f_T(T_{i,j}; \mathbf{Z}_{i,j}) = \frac{\displaystyle\int_{(T_{i,j}-1)\Delta}^{T_{i,j}\Delta} s(t - 2\mathbf{Z}_{i,j}/c)\, dt}{\displaystyle\sum_{T'_{i,j}=1}^{m} \int_{(T'_{i,j}-1)\Delta}^{T'_{i,j}\Delta} s(t - 2\mathbf{Z}_{i,j}/c)\, dt}, \qquad T = 1, 2, \ldots, m. \tag{4.12}$$

Approximating the laser's pulse shape $s(t)$ by a Gaussian, and using left-Riemann sums to approximate the integrals, we can reduce the negative log-likelihood function $\mathcal{L}_Z(\mathbf{Z}_{i,j}; T_{i,j}) = -\log f_T(T_{i,j}; \mathbf{Z}_{i,j})$ to the simple ℓ_2-loss function:

$$\mathcal{L}_Z(\mathbf{Z}_{i,j}; T_{i,j}) = \begin{cases} \|T_{i,j}\Delta - 2\mathbf{Z}_{i,j}/c\|_2^2, & (i,j) \notin \mathcal{H}, \\ 0, & \text{otherwise}. \end{cases} \tag{4.13}$$

For robust depth estimation, we take our depth image to be the non-negative $\hat{\mathbf{Z}}$ that minimizes the negative log-likelihood plus a term that penalizes non-smoothness, viz., we find $\hat{\mathbf{Z}}$ by solving the following optimization problem:

$$\begin{aligned} \hat{\mathbf{Z}}_{\text{array}} &= \arg\min_{\mathbf{Z}} \left[\sum_{(i,j)\notin\mathcal{H}} \sum_{T_{i,j}\in\mathbf{U}_{i,j}} \mathcal{L}_Z(\mathbf{Z}_{i,j}, T_{i,j}) \right] + \tau_z \, \|\mathbf{Z}\|_{\text{TV}} \\ &\text{subject to} \qquad \mathbf{Z}_{i,j} \geq 0, \qquad \text{for all } i, j, \end{aligned} \tag{4.14}$$

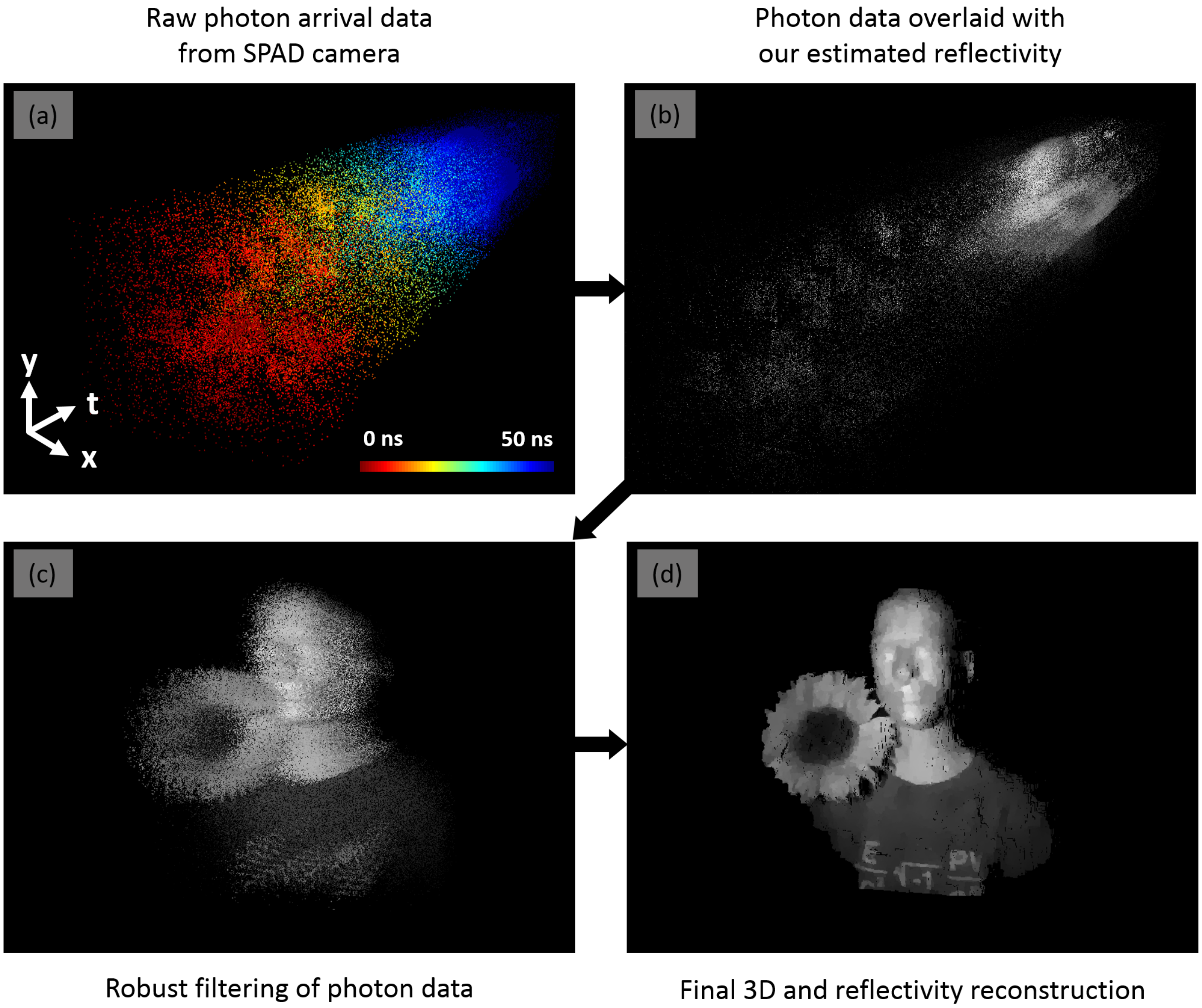
where $\mathbf{U}_{i,j}$ is the set of uncensored detection times that can be obtained using $\tilde{\mathbf{Y}}$ at pixel (i,j) and τ_z is a regularization parameter penalizing non-smoothness in the depth image. The problem in (4.14) is convex, so it can be solved efficiently using projected-gradient methods.

Similar to what was done for reflectivity estimation, in our experiments, we chose τ_z by first generating depth images using $\tau_z = \tau_z'/(1-\tau_z')$ for $\tau_z' \in \{0.1, 0.2, \ldots, 0.9\}$ and then choosing the τ_z value that minimized the ℓ_2 distance from the ground truth depth. As before, cross-validation methods can be employed when ground truth is unavailable. Figure 4-3 illustrates the steps of our algorithm using experimental photon count dataset obtained using the SPAD array imaging setup.

4.5 Results

Experiments: Using the SPAD array imaging setup described in Section 4.2, we performed experiments of using a small number of photon detections to recover scene reflectivity and depth. Experimental data collection was performed by collaborator F. Xu.

Figure 4-4 shows experimental results of 3D structure and reflectivity reconstructions for a scene comprised of a mannequin and sunflower when, averaged over the scene, there was ∼1 signal photon detected per pixel and ∼1 extraneous (background-light plus dark-count) detection per pixel. The image resolution was 384×384 for this



Raw photon arrival data from SPAD camera

Photon data overlaid with our estimated reflectivity

Robust filtering of photon data

Final 3D and reflectivity reconstruction

Figure 4-3: Stages of 3D structure and reflectivity reconstruction algorithm. (a) Raw time-tagged photon detection data are captured using the SPAD camera setup. Averaged over the scene, the number of detected signal photons per pixel was $\sim$1, as was the average number of background-light detections plus dark counts. (b) Step 1: raw time-tagged photon detections are used to accurately estimate the scene's reflectivity by solving a regularized optimization problem. (c) Step 2: to estimate 3D structure, extraneous (background-light plus dark-count) photon detections are first censored, based on the longitudinal sparsity constraint of natural scenes, by solving a sparse deconvolution problem. (d) Step 3: the uncensored (presumed to be signal) photon detections are used for 3D structure reconstruction, by solving a regularized optimization problem.

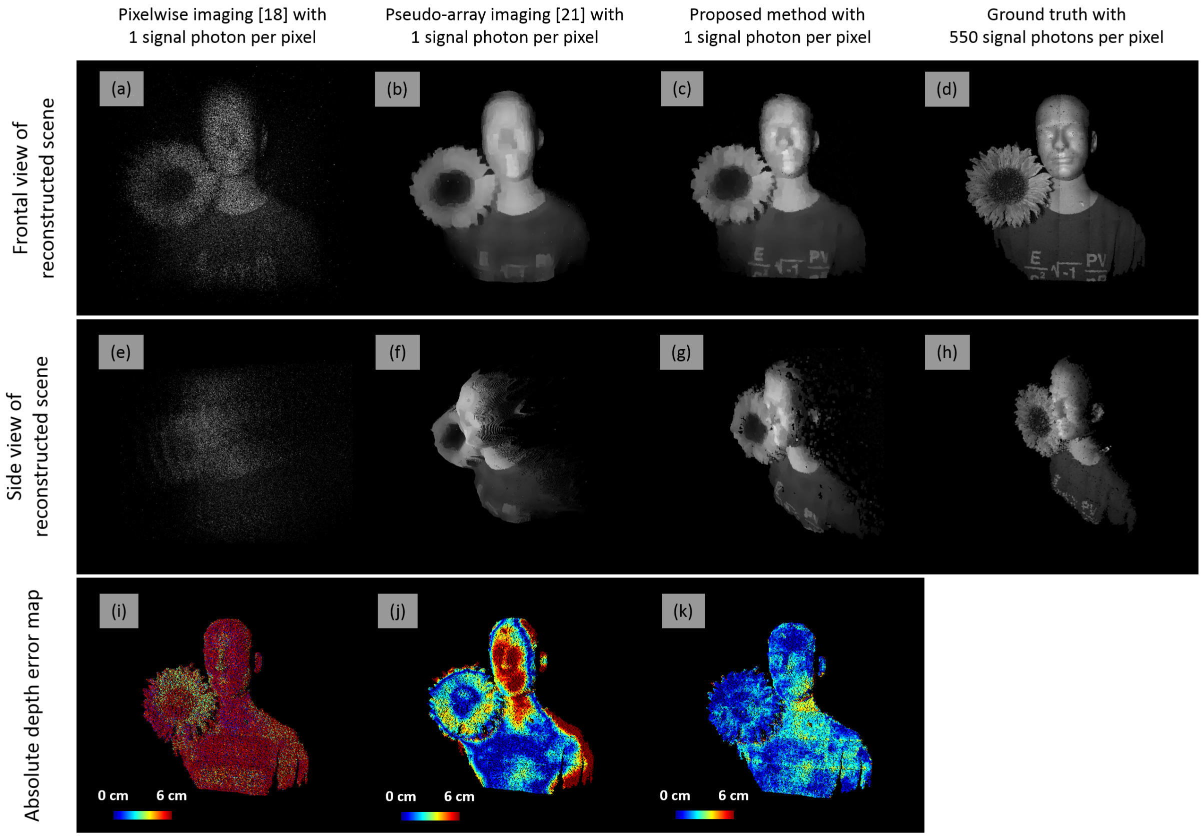

Pixelwise imaging [18] with
1 signal photon per pixel
Pseudo-array imaging [21] with
1 signal photon per pixel
Proposed method with
1 signal photon per pixel
Ground truth with
550 signal photons per pixel
Frontal view of
reconstructed scene
Side view of
reconstructed scene
Absolute depth error map
(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
(i)
(j)
(k)
0 cm
6 cm
0 cm
6 cm
0 cm
6 cm

Figure 4-4: 3D structure and reflectivity reconstructions of the mannequin and flower scene. (a)–(d) Results of imaging 3D structure and reflectivity using the filtered histogram method, the state-of-the-art pseudo-array imaging method, our proposed framework, and the ground-truth proxy obtained from detecting 550 signal photons per pixel. For visualization, the reflectivity estimates are overlaid on the reconstructed depth maps for each method. The frontal views, shown here, provide the best visualizations of the reflectivity estimates. (e)–(h) Results of imaging 3D structure and reflectivity from (a)–(d) rotated to reveal the side view, which makes the reconstructed depth clearly visible. The filtered histogram image is too noisy to show any useful depth features. The pseudo-array imaging method successfully recovers gross depth features, but, in comparison with the ground truth estimate in (h), it overestimates the dimensions of the mannequin's face by several cm and oversmooths the facial features. Our SPAD-array-specific method in (g), however, gives high-resolution depth and reflectivity reconstruction at low flux. (i)–(k) The depth error maps obtained by taking the absolute difference between estimated depth and ground truth depth show that our method successfully recovers the scene structure with sub-pulse-width resolution of less than $c\Delta/2 \approx 6$ cm, while existing methods fail to do so.

experiment. We compare our proposed method with the baseline pixelwise imaging method that uses filtered histograms [1] and the state-of-the-art pseudo-array imaging method [2]. From the visualization of reflectivity overlaid on depth, we observe that the baseline pixelwise imaging method (Figure 4-4(a),(e)) generates noisy depth and reflectivity images without useful scene features, owing to the combination of low-flux operation and high background detections plus detector dark counts. In contrast, the existing pseudo-array method—which exploits transverse spatial correlations but presumes constant $B_{i,j}$—gives a reflectivity image that presents overall object features but suffers oversmoothing from its effort to mitigate hot pixel contributions (Figure 4-4(b)). Furthermore, because the pseudo-array method presumes the 10-ps-class time tagging of a single-element SPAD that is used in raster-scanning setups, its depth image fails to reproduce the 3D structure of the mannequin's face from the ns-class time tagging afforded by our SPAD array. In particular, it overestimates the head's dimensions and oversmooths the facial features (Figure 4-4(f)), whereas our array-specific method accurately captures the scene's 3D structure and reflectivity (Figure 4-4(c),(g)). This accuracy can be seen by comparing our framework's result with the high-flux pixelwise depth and reflectivity images (Figure 4-4(d),(h))—obtained by detecting 550 signal photons per pixel and performing time-gated pixelwise processing—that serve as ground-truth proxies for the scene's actual depth and reflectivity. For the fairest comparisons in Figure 4-4, each algorithm—baseline pixelwise processing, pseudo-array processing, and our new framework—had its parameters tuned to minimize the mean-squared degradation from the ground-truth proxies. The depth error maps in Figure 4-4(i)–(k) quantify the resolution improvements from our imager over the existing ones for this low-flux imaging experiment. Recall that the time bin duration of each pixel of the SPAD camera is $\Delta = 390\,\text{ps}$, corresponding to $c\Delta/2 \approx 6\,\text{cm}$ depth resolution. Our imager successfully recovers depth with sub-bin-duration resolution, while existing methods fail to do so.

In practical imaging scenarios, the camera will collect extremely low levels of back-reflected laser light at pixels where reflectors have very low reflectivity or are very far away. In our imaging experiment, by covering the wall behind our mannequin and flower with a black cloth, we simulated such an environment in our experiments. Conventional depth estimates, even with a high number of detections per pixel, will be very noisy at such pixels because most of those pixels' detections will be extraneous. Because our Step 2 is highly effective in pixelwise censoring extraneous detections, the regularization and inpainting that we employ in our Step 3 yields depth images with sharp edges in such difficult imaging scenarios.

In Figure 4-5, we show the effect of varying the reflectivity and depth regularization parameters τ_A, τ_Z for the mannequin and sunflower dataset. We see that for very small regularization parameters ($\tau_A = 0.4, \tau_Z = 0.4$), the reconstruction result is noisy due to undersmoothing. For very high regularization parameters ($\tau_A = 9.0, \tau_Z = 9.0$), we find that the reflectivity estimate is almost featureless and the depth estimate has oversmoothing distortions that are visible at the sunflower's edges.

In Figure 4-6, we compare our 3D structure and reflectivity reconstruction method with baseline pixelwise imaging [60] and the pseudo-array method [2, 13] for a scene that consisted of a watering can and a basketball. The mean photon count averaged over all 384×384 pixels was approximately 1.0, and the mean signal photon count per pixel was approximately 0.5. Thus, the signal-to-background ratio was approximately 1, as in the mannequin and sunflower experiment. As the figure shows, our method successfully recovers the undistorted 3D shapes of the two objects and their reflectivity profiles, while the pixelwise method's image is exceedingly noisy and the pseudo-array's is oversmoothed from its attempt to mitigate extraneous-detection noise.

Implementation details of the reconstruction algorithm: Prior to initiating our three-step imaging algorithm, we first performed calibration measurements to: identify $\mathcal{H}$, the SPAD array's set of hot pixels; obtain the average background-light plus dark-count rates for the remaining pixels; and determine the laser pulse's RMS time duration. It turned out that: about 2% of our camera's 1024 pixels were placed in $\mathcal{H}$; the background-light plus dark-count rates were indeed spatially varying across the remaining pixels; and the laser pulse's time duration was $T_p \approx 1$ ns and its normalized shape was reasonably approximated as a Gaussian.

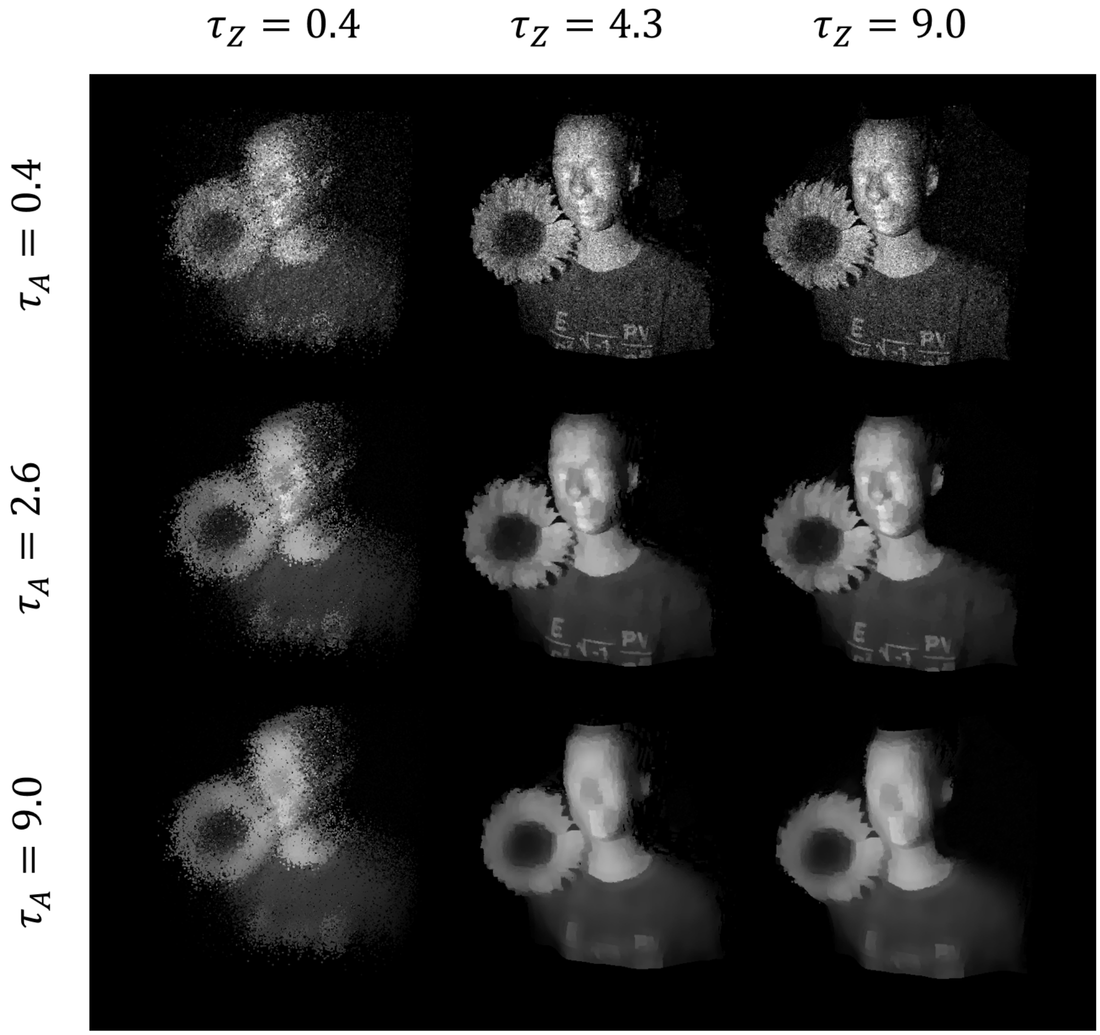

$\tau_Z = 0.4$
$\tau_Z = 4.3$
$\tau_Z = 9.0$
$\tau_A = 0.4$
$\tau_A = 2.6$
$\tau_A = 9.0$

Figure 4-5: Effect of varying the regularization parameters in our 3D structure and reflectivity reconstruction algorithm for the mannequin and flower scene. The optimal parameter set was $\{\tau_A, \tau_Z\} = \{2.6, 4.3\}$.

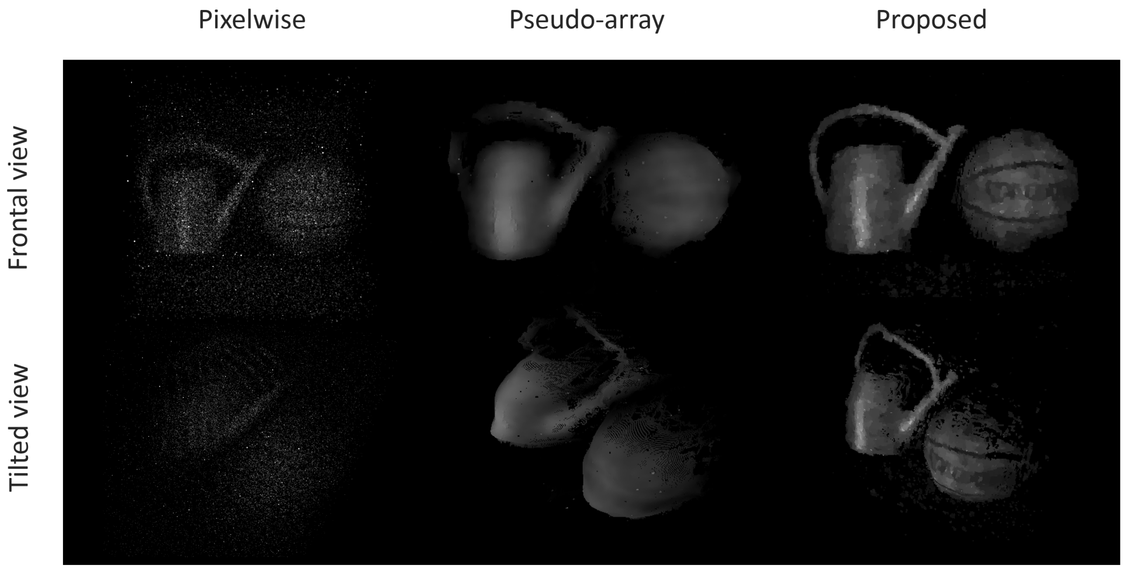

Pixelwise
Pseudo-array
Proposed
Frontal view
Tilted view

Figure 4-6: Imaging results for the watering can and basketball scene. Notice the stripes of the basketball being visible when using our method.

Hot pixels in the 32×32 pixel SPAD array are pixels whose dark-count rates are so high that their outputs are non-informative in estimating scene reflectivity and depth. To identify such pixels, we collected data with the laser turned off and the SPAD array observing a low-reflectivity target (dark wall). Under these conditions, hot pixels are clearly visible in the resulting photon-count image, see Figure 4-7(a), and we get the SPAD array's hot-pixel set by a simple thresholding procedure on that image. With a threshold of 150 counts/s, 2% of the camera's 1024 pixels are labeled as hot pixels. The hot-pixel set $\mathcal{H}$ for the 384×384 pixel composite image is then obtained by appropriate translations of the array's hot pixels.

To accurately measure the laser's pulse shape, we replaced our SPAD array—which has 390 ps time binning—with a single-pixel Micro Photon Devices SPAD detector and a HydraHarp time-correlator that provides 8 ps time binning. We focused the laser on a $1 \times 1\,\mathrm{m}^2$ white planar calibration target placed $z \approx 1\,\mathrm{m}$ from the imaging setup and obtained the photon-count histogram shown in Figure 4-7(b) when 3344 signal photons were detected. Extraneous photon detections were suppressed in this data collection by time-gating around $2z/c$ seconds. The laser's root mean-square (RMS) pulse duration was computed from this histogram to be $T_p \approx 1\,\mathrm{ns}$. For our computational reconstruction algorithm, we apply the Gaussian approximation to the calibrated pulse waveform.

To calibrate the background-light plus dark-count rate matrix $\mathbf{B}$, we turned off the laser and obtained the $\mathbf{B}$ image, shown in Figure 4-7(c), for the mannequin and flower scene. Spatial variations in $\mathbf{B}$ and the presence of hot pixels are clearly visible (note the periodic xy spacing of hot pixels, due to the translational scanning procedure).

We then proceeded to Step 1 of the reconstruction algorithm: we estimated reflectivity $\hat{\mathbf{A}}_{\text{array}}$ by combining the Poisson statistics of photon counts with a TV-norm smoothness constraint on the estimated reflectivity—while censoring the set of hot pixels—to write the optimization as a TV-regularized, Poisson image inpainting problem. This optimization problem is convex in the reflectivity image variable $\mathbf{A}$, which allowed us to solve it with simple projected-gradient methods [58] in a computationally-efficient manner. For a 384×384 image, the processing time of

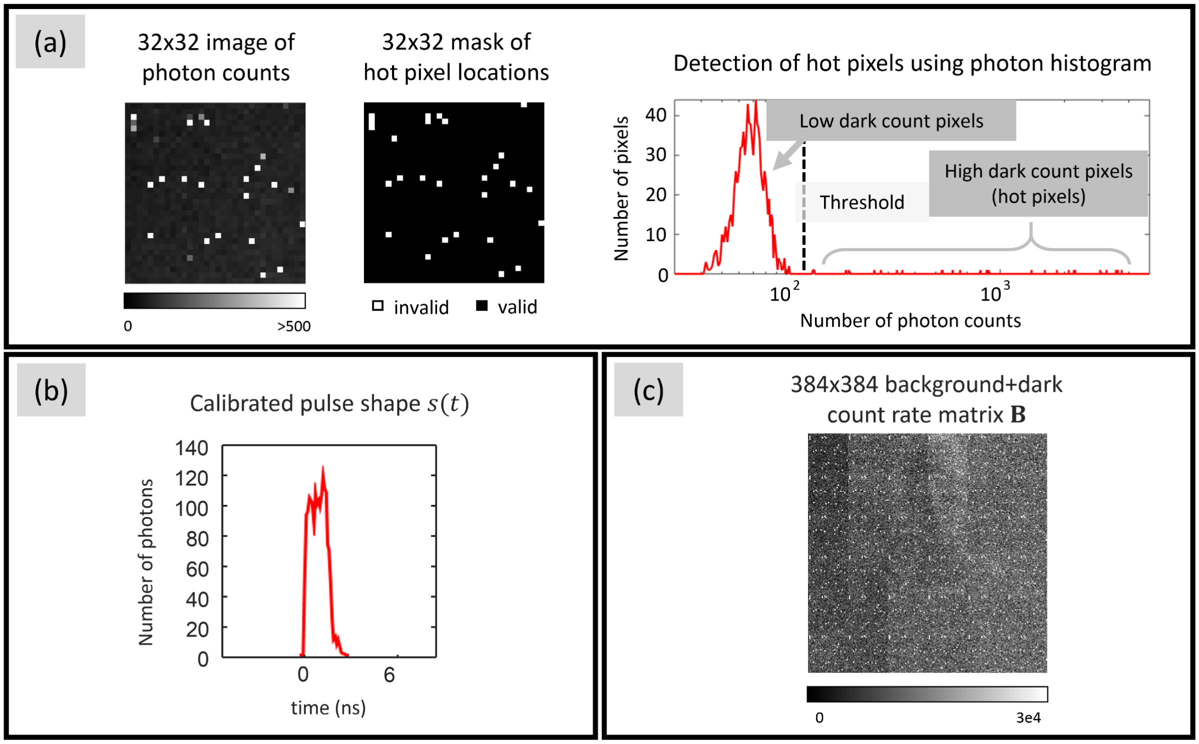


Figure 4-7: Calibration results for the SPAD camera and scene parameters. (a) A 32×32 image of photon counts when the SPAD camera observes a weakly-reflecting planar wall with the laser off (left) was used to generate a 32×32 binary mask indicating hot-pixel locations with white markers (middle). The hot pixel mask was obtained by thresholding the photon-count image with an appropriate threshold chosen from the photon-count histogram (right). (b) Laser pulse shape. Extraneous photon detections were suppressed by time-gating near the roundtrip delay to a 1×1 m calibration target. (c) The non-constant 384×384 background-light plus dark-count rate matrix $\mathbf{B}$ for the mannequin and flower scene.

Step 1 was about 6 seconds on a standard laptop computer.

For Step 2 of the reconstruction algorithm, we filtered the photon-detection dataset to impose the longitudinal constraint that the scene has a sparse set of reflectors. This is because the scaled detection-time histogram $\text{hist}(c\mathbf{T}/2)$ that has been corrected for the average background-light plus dark-count detections per bin is a proxy solution for $\text{hist}(\mathbf{Z}_\Delta)$, where $\text{hist}(\mathbf{Z}_\Delta)$ is a size-m histogram that bins the scene's 3D structure at the camera's $c\Delta/2$ native range resolution. Thus, we used orthogonal matching pursuit (OMP) [59] on $\text{hist}(c\mathbf{T}/2)$, the coarsely-binned histogram of photon detections, to find the non-zero spikes representing the object cluster depths. We then discarded photon detections that implied depth values more than $cT_p/2$ away from the estimated depth values, because they were presumably extraneous detections. For a 384×384 image, the processing time of Step 2 was about 17 seconds on a standard laptop computer.

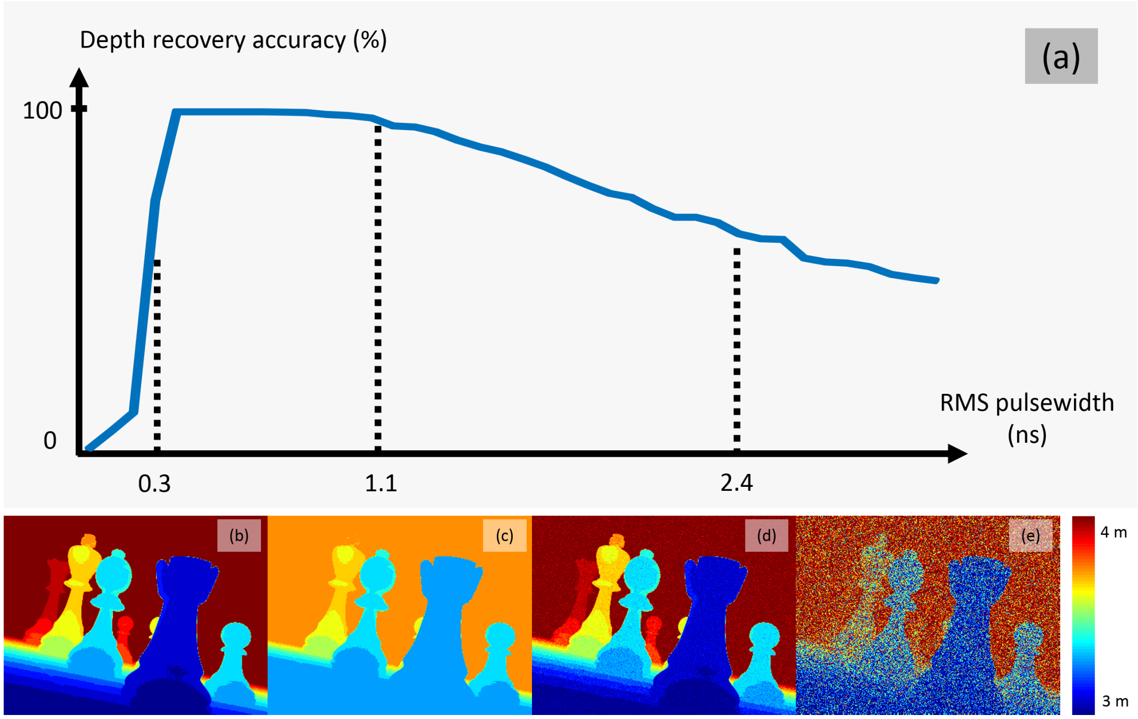
Having censored detections from all hot pixels and, through the longitudinal constraint, censored almost all extraneous detections on the remaining pixels, we treated all the uncensored photon detections as being from backreflected laser light, i.e., that they were all signal-photon detections. For Step 3 of our reconstruction algorithm, we estimated the scene's 3D structure using these uncensored photon detections. Because we operated in the low-flux regime, many of the pixels had no photon detections and thus are non-informative for 3D-structure estimation. A robust 3D estimation algorithm must inpaint these missing pixels using information derived from nearby pixels' photon detection times. Approximating the laser's pulse waveform $s(t)$ by a Gaussian with RMS duration T_p, we solved a TV-regularized, Gaussian image inpainting problem to obtain our depth estimate $\hat{\mathbf{Z}}_{\text{array}}$. This is a convex optimization problem in the depth image variable $\mathbf{Z}$, and projected-gradient methods were used to generate $\hat{\mathbf{Z}}_{\text{array}}$ in a computationally-efficient manner. For a 384×384 image, the processing time of Step 3 was about 20 seconds on a standard laptop computer.

Choice of laser-pulse RMS time duration: For a transform-limited laser pulse, like the Gaussian $s(t)$ that our imaging framework presumes, the RMS time duration T_p is a direct measure of system bandwidth. As such, it has an impact on the depth-imaging accuracy in low-flux operation. This impact is borne out by the simulation results in Figure 4-8, where we see that the pulse waveform with the shortest RMS duration does *not* provide the best depth recovery. In order to improve the reconstruction accuracy even with short RMS duation, frameworks for dithered quantization may be employed [61, 62]. Thus, in our experiments, we broadened the laser's output pulse to $T_p \approx 1\,\text{ns}$. This pulse duration allowed us to resolve depth features well below the $c\Delta/2 \approx 6\,\text{cm}$ value set by the SPAD array's 390-ps-duration time bins.

4.6 Summary and Discussion

We have proposed and demonstrated a SPAD-array-based imaging framework that generates highly-accurate images of a scene's 3D structure and reflectivity from ~1 detected signal photon per pixel, despite the presence of extraneous detections at roughly the same rate from background light and dark counts. By explicitly modeling the limited single-photon time-tagging resolution of SPAD-array imagers, our framework dramatically improves reconstruction accuracy in this low-flux regime as compared to what is achieved with existing methods. The photon efficiency of our proposed framework is quantified in Figure 4-9 where we have plotted the sub-pulse-width RMS depth error it affords in imaging the mannequin's face versus the average number of detected signal photons per pixel. For this task our algorithm realizes cm-class depth resolution down to less then 1 detected signal photon per pixel, while the baseline pixelwise imager's depth resolution is more than an order of magnitude worse because of its inability to cope with extraneous detections.

Because our framework employs a SPAD camera for highly photon-efficient imaging, it opens up new ways to image 3D structure and reflectivity on very short time scales while requiring very few photon detections. Hence it could find widespread use in applications that require accurate imaging using extremely small amounts of light, such as remote terrestrial mapping [24], seismic imaging [63], fluorescence pro-



Ground truth depth map

Estimated depth when RMS duration is 0.3 ns

Estimated depth when RMS duration is 1.1 ns

Estimated depth when RMS duration is 2.4 ns

Figure 4-8: Relationship between RMS pulse duration and depth-recovery accuracy. (a) Plot of depth-recovery accuracy using our algorithm (Step 2 and Step 3) versus RMS pulse duration T_p obtained by simulating a low-flux SPAD imaging environment with an average of 10 detected signal photons per pixel and 390 ps time bins. Depth recovery is deemed a success at a pixel if estimated depth is within 3 cm of ground truth, and the depth recovery accuracy of a method is computed by the mean success rate over all pixels. (b) Ground-truth depth map used in the simulations. (c) Estimated depth map for $T_p = 0.3\,\text{ns}$. (d) Estimated depth map for $T_p = 1.1\,\text{ns}$. (e) Estimated depth map for $T_p = 2.4\,\text{ns}$. When T_p is too short, there is a systematic bias in the estimated depths, although random errors are minimal. When T_p is too long, the estimated depths are very noisy. In all our SPAD-array experiments we used $T_p \approx 1\,\text{ns}$, which is in the sweet spot between durations that are too short or too long. We emphasize, however, that our algorithm is not tuned to a particular pulse-width and can be performed using any T_p value.

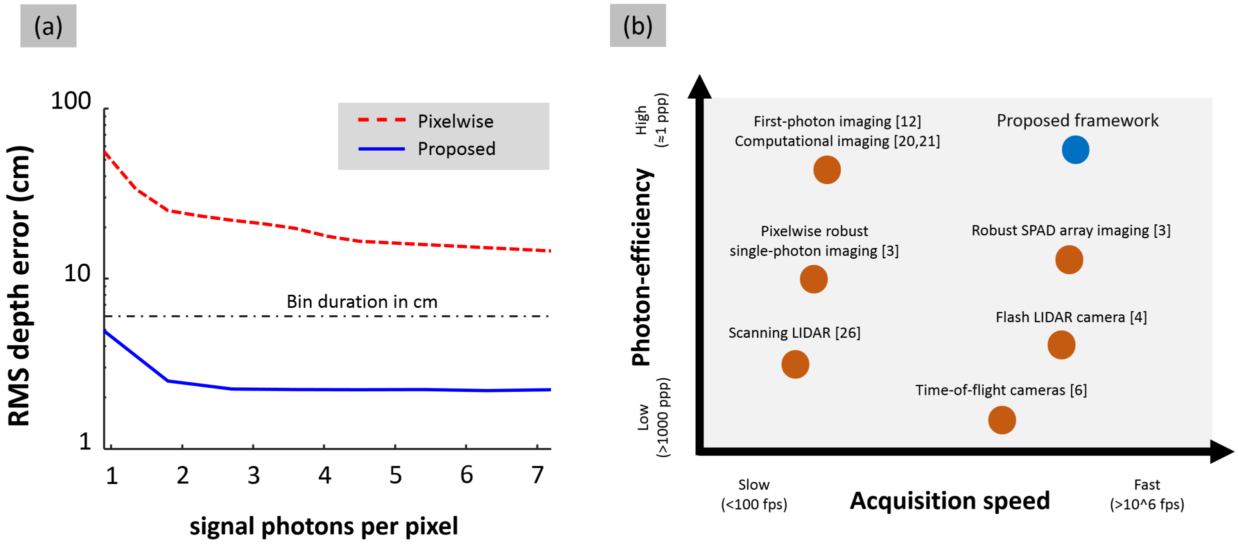


Figure 4-9: Photon efficiency of proposed framework. (a) Plots of RMS depth error (log scale) versus average number of detected signal photons per pixel for imaging the mannequin's face using our proposed framework and the baseline pixelwise processor. Our method consistently realizes sub-pulse-width performance throughout the low-flux region shown in the plot, whereas the baseline approach's accuracy is more than an order of magnitude worse, owing to its inability to cope with extraneous detections. (b) Summary of photon efficiency versus acquisition speed (not the computational speed) for existing 3D structure and reflectivity imagers, where fps denotes frames per second and ppp denotes photons per pixel.

filing [64], and astronomy [65]. We emphasize, in this regard, that our framework affords automatic rejection of ambient-light and dark-count noise effects *without* requiring sophisticated time-gating hardware. It follows that our imager could also enable noise-tolerant 3D vision for self-navigating advanced robotic systems, such as unmanned aerial vehicles (UAV) and exploration rovers [66].

Chapter 5

Super-Resolution Imaging

5.1 Overview of Problem

In the previous chapters, we have developed photon-efficient depth and reflectivity imaging frameworks, using an imaging setup that performs pixel-by-pixel signal acquisition using a photon-counting detector. The core assumption behind these frameworks was that each pixel defined a different scene patch, i.e., in object space pixels were non-overlapping. In practice, this assumption can often break down due to non-idealities in the illumination source and imaging conditions. For example, when using a scanning setup to image very small features of a biological sample, the illumination spot size for a scan point may be appreciably larger then the interpixel separation. Another example is when we are constrained to image the scene through a strongly scattering medium, whose effect is that a random illumination pattern with large spatial spread will be cast on the scene instead of a low-spatial-spread illumination. As a result, measurements from multiple scan points will be correlated [67]. (In Chapter 3, when we were imaging the scene behind a scattering medium, our scanning resolution was low enough that we did not have to deal with the problem of interpixel correlations.) Although the scanning resolution may be high, the effective spatial resolution of the measurements can be low in such imaging scenarios. In this chapter, we address the problem of recovering a high-resolution image from a low-resolution measurements. As usual, we will be interested in recovering depth and reflectivity information from low light-level data.

Prior art: There are several approaches for image super-resolution, each modeling different physical constraints. The first is the upsampling approach, which aims to increase the resolution of the low-resolution image by exploiting a scene's spatial prior, using techniques such as cubic splines [68], example-based learning [69], patch recurrences [70], and depth-intensity correlations [71]. It has also been shown that multiple misaligned low-resolution images at several viewpoints can be used to reconstruct a high resolution image in the context of both digital photography and LIDAR depth imaging [72, 73].

Another approach is based on the method of image deconvolution, which exploits the fact that the image acquisition artifacts that convert measurement data from high-resolution to low-resolution can be modeled using a two-dimensional convolution operation. This image convolution kernel is called the point spread function (PSF). In conventional digital photography, this deconvolution problem for super-resolution is solved to restore focus in blurry photographs [74]. For depth imaging using amplitude modulated continuous wave (AMCW) illuminations, the camera lens-blur PSF is typically modeled as a two-dimensional Gaussian function, and robust deconvolution techniques that use the image smoothness prior to well-condition the deconvolution problem have been developed [75, 76]. Theoretical recovery guarantees of super-resolved signals by the means of sparse deconvolution methods have also been published, both in noiseless and bounded-noise acquisition models [77, 78].

Summary of our approach: In this chapter, we propose a super-resolution imaging framework that uses a photon-count histogram dataset for the purpose of low-light depth and reflectivity reconstruction. When using the scanning LIDAR setup, we can model the illumination patterns as a PSF, and so our super-resolution approach for scanning LIDAR belongs in the set of deconvolution-based super-resolution methods. However, our framework is different from existing deconvolution-based depth super-resolution methods, because it models the effect of photon noise specific to low-light pulsed LIDAR systems that employ photon-counting detectors. Furthermore, we not only consider a PSF that is a decaying function, such as a two-dimensional Gaussian used in classical super-resolution literature, but we also treat randomly-speckled PSFs that are encountered when imaging a scene through strongly scattering media. We propose two approaches for low-light super-resolution imaging.

- **Approach 1:** Our first approach to super-resolution imaging is to exploit the LIDAR constraint that each pixel contains a single reflector at an unknown depth. We develop a greedy algorithm, inspired by the compressive sensing algorithm CoSaMP [16], to accurately recover the super-resolved depth image from low-resolution photon-count data. Using a simulated single-photon imaging setup, we demonstrate that our framework outperforms the conventional maximum-likelihood (ML) and deconvolution imaging techniques for moderate light-levels with $\sim 10^4$ photons per pixel.

- **Approach 2:** As we derive later in this chapter, Approach 1 may not be practical because of its high computational complexity of $O(N^2 m^2)$, where N is the number of pixels and m is the number of depth bins. Moreover, its photon efficiency is limited. In particular, the imaging method starts to fail in simulations with $< 10^4$ photons per pixel and a Gaussian PSF. Our second approach uses a transformed photon-count dataset and computes the depth and reflectivity images in $O(N)$ time. Unlike Approach 1, Approach 2 also gains photon efficiency by exploiting the spatial correlations existing in natural scenes. Using simulated photon-count datasets, we verify that Approach 2 outperforms conventional photon-noise-tolerant deconvolution methods, such as the Richardson-Lucy algorithm [79, 80], and is robust to low light-level operation at < 10 photons per pixel.

5.2 Single-Photon Imaging Setup

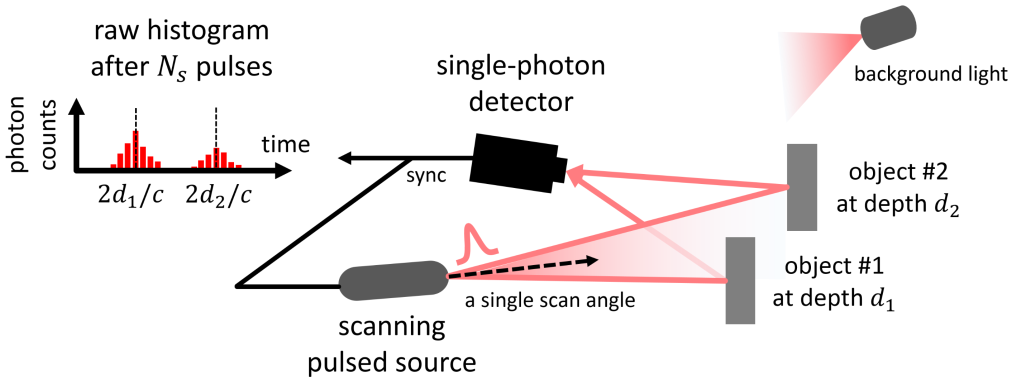
Figure 5-1(a) illustrates our super-resolution imaging setup that uses a scanner and a single-pixel photon-counting detector. (It is possible to formulate the super-resolution problem using the dual setup of having an array detector with a floodlight illumination source. In this array imaging setup, the PSF will be usually defined as a Gaussian function modeling lens aberrations.) An optical source, such as a laser, illuminates the scene with the pulse waveform $s(t)$ aimed at transverse angles of (θ_x, θ_y). Recall that T_r is the acquisition period for one pulse illumination at one scanning angle, and T_p is the RMS duration of $s(t)$. The time-correlated photon-counting detector is able to record the photon-count histogram of the backreflected waveform in $[0, T_r)$ with time-binning accuracy of Δ seconds. The observed photon count histogram has multiple peaks, because the PSF of illumination mixes responses from multiple reflectors with different depths (Figure 5-1(b),(c)). This acquisition process is repeated for N_s pulses per scanning angle and for multiple scanning angles to form a spatially-resolved depth map.

5.3 Forward Imaging Model

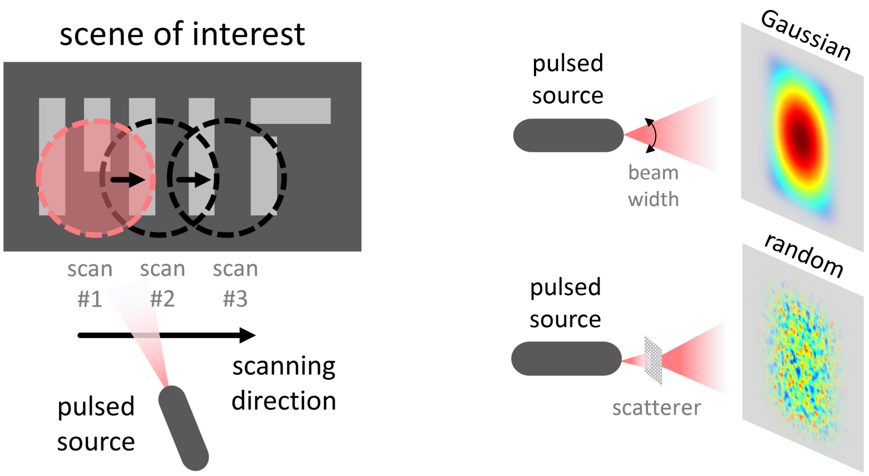
Consider the return from transmission of a single pulse $s(t)$, starting at time 0, that was transmitted at transverse scanning angles (θ_x, θ_y). The optical flux incident at the detector, $r(t; \theta_x, \theta_y)$, is then given by

$$\begin{aligned} r(t; \theta_x, \theta_y) = \int_{(\theta'_x, \theta'_y) \in \text{FOV}} & h(\theta_x - \theta'_x, \theta_y - \theta'_y) \\ & \times a(\theta'_x, \theta'_y) \times s(t - 2d(\theta'_x, \theta'_y)/c)\, d\theta'_x d\theta'_y, \end{aligned} \tag{5.1}$$

for (θ_x, θ_y) in the imaging system's field-of-view (FOV) and $t \in [0, T_r)$. Here we take FOV to be the rectangular angular region, e.g., $\text{FOV} = \{(\theta_x, \theta_y) : \theta_x \in [-80^\circ, 80^\circ], \theta_y \in [-60^\circ, 60^\circ]\}$, we use c to denote the speed of light, and $h(\theta'_x, \theta'_y)$, $a(\theta'_x, \theta'_y)$, $d(\theta'_x, \theta'_y)$ to be the illumination PSF value, reflectivity, depth of scene reflector at scanning di-


raw histogram
after N_s pulses
photon counts
time
$2d_1/c$
$2d_2/c$
single-photon detector
sync
background light
object #2
at depth d_2
object #1
at depth d_1
a single scan angle
scanning
pulsed source

(a) Super-resolution single-photon imaging setup


scene of interest
scan #1
scan #2
scan #3
scanning
direction
pulsed
source
pulsed
source
beam
width
Gaussian
pulsed
source
scatterer
random

(b) Scanning procedure (c) Illumination patterns

Figure 5-1: (a) Scanning LIDAR setup. A repetitively pulsed source illuminates the scene in a scanning manner, where at each scan point multiple reflectors at different depths may be illuminated. The detector records the photon-count histogram of the backreflected response. (b) From one scanning pixel to another, there can be overlap in illumination due to physical constraints. (c) The interpixel overlaps can come from either illumination non-idealities, such as its finite beam-width (top), or scene constraints such as the presence of strongly scattering media that generate random speckle illumination patterns (bottom). We define the two-dimensional intensity pattern cast on the scene as the transverse imaging kernel in our problem, and denote it using the two-dimensional function h centered at $(0, 0)$.

rection (θ'_x, θ'_y), respectively. In our model, we have assumed that $a(\theta'_x, \theta'_y)$ combines object reflectivity properties, effect of radial fall-off, and effect of reflector surface angles.

To express the convolution equation in Eq. (5.1) in discrete matrix-vector form, we let $m = T_r/\Delta$ be the number of detector bins and N be the number of scanned pixels. Then, let $\mathbf{X}$ be an $N \times m$ sparse matrix, whose its ith row contains one non-zero entry for $i \in \{1, \ldots, N\}$ (single opaque reflector constraint), representing the ith pixel's true depth and reflectivity by its index and value. (For now, we focus on recovering a one-dimensional depth image for illustration purposes.) Letting $\mathbf{h}$ be the $N \times 1$ vector obtained by using a discrete approximation of PSF h, we define $\mathbf{H}$ to be the $N \times N$ convolution matrix generated by $\mathbf{h}$. Let $\mathbf{S}$ be the $m \times m$ convolution matrix describing the discretized temporal response coming from the non-zero duration of $s(t)$. Both $\mathbf{H}$ and $\mathbf{S}$ can be calibrated prior to the imaging experiment. Finally, we use $\mathbf{B}$ to denote the combined constant background and dark counts after N_s acquisition trials. We assume $\mathbf{B}$ is also known through offline calibration by measuring the incoming signal while the laser is turned off. By the theory of photodetection, our set of photon-count histogram measurements $\mathbf{Y} \in \mathbb{N}^{N \times m}$ can now be written as

$$\mathbf{Y} \sim \text{Poisson}\left(N_s \mathbf{H}\mathbf{X}\mathbf{S}^T + \mathbf{B}\right), \tag{5.2}$$

where Poisson$(\cdot)$ is defined entrywise. Here, for simplicity of modeling, we have assumed unit detector quantum efficiency and a Dirac delta function for the detector response. (If the quantum efficiency is non-unity, the final estimate can be scaled appropriately and if the detector response is non-trivial, then it can be simply included in $\mathbf{S}$.) Figure 5-2 illustrates Eq. (5.2) using graphical representations of matrices $\mathbf{Y}$, $\mathbf{H}$, $\mathbf{X}$, $\mathbf{S}$, and $\mathbf{B}$ when imaging a one-dimensional scene. From Eq. (5.2), we can write the probability mass function of the random matrix $\mathbf{Y}$ as

$$f_Y(\mathbf{Y}; \mathbf{H}, \mathbf{X}, \mathbf{S}, \mathbf{B}) = \prod_{i=1}^{N} \prod_{j=1}^{m} \frac{(N_s \mathbf{H}\mathbf{X}\mathbf{S}^T + \mathbf{B})_{i,j}^{\mathbf{Y}_{i,j}}}{\mathbf{Y}_{i,j}!} \exp\{-(N_s \mathbf{H}\mathbf{X}\mathbf{S}^T + \mathbf{B})_{i,j}\}. \tag{5.3}$$

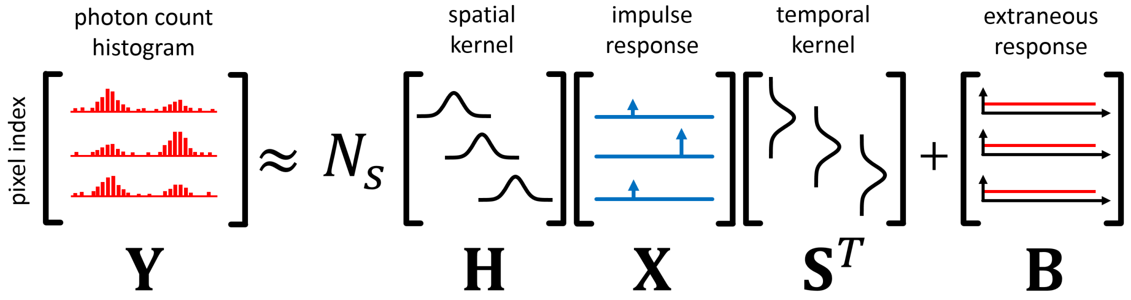


Figure 5-2: Illustration of the acquisition model when imaging a one-dimensional scene. The PSF from scene illumination is modeled by $\mathbf{H}$ and the temporal pulse waveform and detector response is modeled by $\mathbf{S}$. Both $\mathbf{H}$ and $\mathbf{S}$ are Gaussian convolution matrices in this example. $\mathbf{B}$ models the extraneous background and detector dark count response. The set of photon-count histograms observed at each scan pixel after N_s illumination trials is represented by $\mathbf{Y}$.

Our depth super-resolution problem is to recover the scene response of $\mathbf{X}$ from noisy low-spatial-resolution histogram data $\mathbf{Y}$, given the spatial and temporal kernel matrices $\mathbf{H}$ and $\mathbf{S}$, respectively, and the constant matrix $\mathbf{B}$. The value and index of the non-zero entry at the ith row of the estimate of $\mathbf{X}$ will indicate the reflectivity and depth estimates of the ith pixel of images with improved resolution.

5.4 Solving the Inverse Problem

The first approach: Define $\mathcal{S}_1(N, m)$ as the set of $N \times m$ matrices with the sparsity pattern of every row of the matrix containing exactly one non-zero entry. Figure 5-3 illustrates the constraint set $\mathcal{S}_1(N, m)$ for $m = 3$ as a product space of 1-sparse signal sets. We aim to solve the constrained maximum-likelihood (CML) deconvolution problem:

$$\begin{aligned} \underset{\mathbf{X}}{\text{minimize}} \quad & \mathcal{L}(\mathbf{X}; \mathbf{Y}, \mathbf{H}, \mathbf{S}, \mathbf{B}) \\ \text{subject to} \quad & \mathbf{X} \in \mathcal{S}_1(N, m), \quad \mathbf{X}_{i,j} \geq 0, \ \forall i, j, \end{aligned} \tag{5.4}$$

where $\mathcal{L}(\mathbf{X}; \mathbf{Y}, \mathbf{H}, \mathbf{S}, \mathbf{B})$ is the negative log-likelihood function of $\mathbf{X}$ obtained from Eq. (5.3) and $\mathbf{X}_{i,j} \geq 0$ models the non-negativity of light intensity. We have the

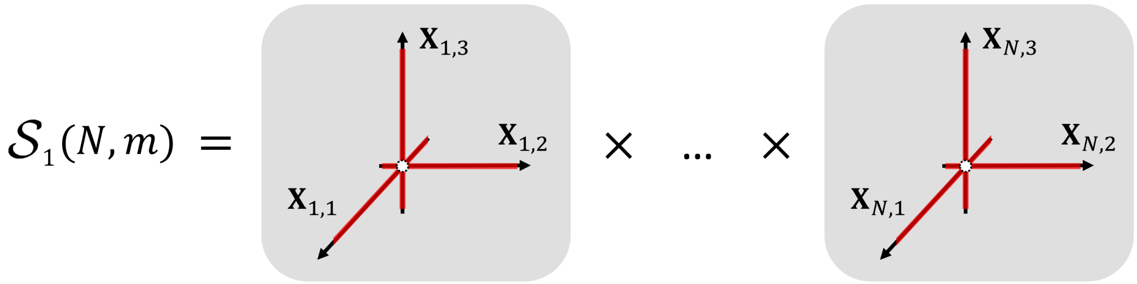


Figure 5-3: Illustration of $\mathcal{S}_1(N,m)$ for $m = 3$. Since every row of $\mathbf{X} \in \mathcal{S}_1(N,m)$ belongs to a set of 1-sparse signals, our constraint set $\mathcal{S}_1(N,m)$ is a product space of N of 1-sparse signal sets.

kronecker product identity that $\text{vec}(\mathbf{S}\mathbf{X}^T\mathbf{H}^T) = (\mathbf{H}\otimes\mathbf{S})\text{vec}(\mathbf{X}^T)$, where $\text{vec}(\cdot)$ denotes matrix vectorization (column-first) and $\otimes$ denotes the kronecker product. Using the kronecker product identity, we write the matrix optimization problem in (5.4) as a vector optimization problem by defining our variables as $\mathbf{x}_v = \text{vec}(\mathbf{X}^T)$, $\mathbf{y}_v = \text{vec}(\mathbf{Y}^T)$, $\mathbf{b}_v = \text{vec}(\mathbf{B})$, and $\mathbf{K} = N_s(\mathbf{H}\otimes\mathbf{S})$:

$$\begin{aligned} \underset{\mathbf{x}_v}{\text{minimize}} \quad & \|\mathbf{y}_v - (\mathbf{K}\mathbf{x}_v + \mathbf{b}_v)\|_2^2 \\ \text{subject to} \quad & \mathbf{x}_v \in \mathcal{S}_{1,m}(Nm), \\ & (\mathbf{x}_v)_i \geq 0, \quad \forall i \in \{1, 2, \dots, Nm\}, \end{aligned} \tag{5.5}$$

where, for computational efficiency, we have replaced the negative log Poisson likelihood function by its squared ℓ_2-norm approximation for computational efficiency (see Appendix D), and used $\mathcal{S}_{1,m}(Nm)$ as the set of size-Nm N-sparse vectors with each of its N sub-block-vectors of size m having only one non-zero entry.

Our problem in (5.5) is a problem of non-negative least squares (NNLS) with a structured signal sparsity constraint $\mathbf{x}_v \in \mathcal{S}_{1,m}(Nm)$. Similar to our algorithm development in Chapter 2 for pixelwise depth imaging, we draw inspiration from the well-known algorithms for sparse signal pursuit, such as CoSaMP [16] and union-of-subspaces-based methods [11, 15], to design a greedy method that solves (5.5). Our proposed approach is given in Algorithm 4.

In Algorithm 4, we define $\hat{\mathbf{x}}_{[1,m]}$ as the N-sparse vector generated by picking the

Algorithm 4 Proposed algorithm for super-resolution LIDAR (Approach 1)

```
Input: $\mathbf{y}_v$, $\mathbf{K}$, $\mathbf{b}_v$, $\delta$
Output: $\hat{\mathbf{x}}_{\mathrm{SR}}$
Initialize $\mathbf{x}^{(0)} \leftarrow \vec{0}$, $\mathbf{y}'_v \leftarrow \mathbf{y}_v - \mathbf{b}_v$, $\mathbf{u} \leftarrow \mathbf{y}'_v$, $k \leftarrow 0$
repeat
    $k \leftarrow k + 1$
    $\hat{\mathbf{x}} \leftarrow \mathbf{K}^T\mathbf{u}$                                          ▷ Compute adjoint solution
    $\Omega \leftarrow \mathrm{supp}\left(\hat{\mathbf{x}}_{[1,m]}\right) \cup \mathrm{supp}\left(\mathbf{x}^{(k-1)}\right)$      ▷ Merge support
    $\mathbf{v}|_\Omega \leftarrow \mathbb{N}(\mathbf{K}_\Omega, \mathbf{y}'_v)$                     ▷ Solve NNLS for deconvolution
    $\mathbf{v}|_{\Omega^c} \leftarrow \mathbf{0}$
    $\mathbf{x}^{(k)} \leftarrow \mathbf{v}$                                             ▷ Update solution
    $\mathbf{u} \leftarrow \mathbf{y}'_v - \mathbf{K}\mathbf{x}^{(k)}$
until $\|\mathbf{x}^{(k-1)} - \mathbf{x}^{(k)}\|_2^2 < \delta$
$\hat{\mathbf{x}}_{\mathrm{SR}} \leftarrow \mathbf{x}^{(k)}$
```

best 1-sparse vector for every size-m subvector of the intermediate solution $\hat{\mathbf{x}}$, which may be dense. We let $\mathbf{K}_\Omega$ be the submatrix of $\mathbf{K}$ that is obtained by horizontally concatenating the columns of $\mathbf{K}$ indexed by set Ω, and we use supp($\mathbf{x}$) to be the indices of vector $\mathbf{x}$'s non-zero entries. We also define $\mathbb{N}(\mathbf{K}_\Omega, \mathbf{y}'_v)$, used in Algorithm 4 solving (5.5), to be the solution that minimizes the well-known non-negative least squares program:

$$\begin{aligned} \underset{\mathbf{v}}{\text{minimize}} \quad & \left\|\mathbf{y}'_v - \mathbf{K}_\Omega \mathbf{v}\right\|_2^2 \\ \text{subject to} \quad & \mathbf{v}_i \geq 0, \quad \forall i \in \{1, 2, \ldots, |\Omega|\}, \end{aligned} \tag{5.6}$$

where $\mathbf{y}'_v$ is the background corrected $\mathbf{y}_v$.

Because the non-negative least squares program has a convex cost and convex constraint set, its optimal solution can be found quickly using projected-gradient methods [81]. Our Algorithm 4 terminates based on the condition that the solution gradient has squared l_2-norm less than $\delta > 0$.

Our super-resolution depth imaging algorithm is similar to existing greedy sparse signal pursuit algorithms as it alternates between gradient descent on the cost function and pseudo-projection of the intermediate solution onto the sparsity constraint set. The core differences between existing sparse signal pursuit algorithms and the proposed algorithm are

1. We compute the structured k-sparse vector specific to our single-reflector-per-pixel constraint in LIDAR, instead of the vanilla k-sparsity constraint.
2. Instead of solving for the least squares estimate, we compute the non-negative least squares estimate using $\mathbb{N}(\mathbf{K}_\Omega, \mathbf{y}'_v)$ to model the non-negativity of light intensity.

Using $\hat{\mathbf{x}}_{\text{SR}}$, which is the N-sparse output vector of Algorithm 4, we obtain our final scene depth reconstruction $\hat{\mathbf{d}}_{\text{SR}}$ by fetching the index of non-zero entry for each of $\hat{\mathbf{x}}_{\text{SR}}$'s m-dimensional sub-block-vector:

$$(\hat{\mathbf{d}}_{\text{SR}})_i = \text{supp}\left((\hat{\mathbf{x}}_{\text{SR}})_{(m(i-1)+1):mi}\right), \quad \forall i = 1, \ldots, N, \tag{5.7}$$

where $(\hat{\mathbf{x}}_{\text{SR}})_{(m(i-1)+1):mi}$ denotes the sub-vector of $\hat{\mathbf{x}}_{\text{SR}}$ created by using its indices $(m(i-1)+1), (m(i-1)+2), \ldots, mi$.

Algorithm 4 solves for the super-resolved depth map using physically accurate models using a non-parametric approach of estimating size-Nm matrix $\mathbf{X}$ that contains information scene depth. This carries the limitation that $\mathbf{K}$ is a large $Nm \times Nm$ matrix. Because of the adjoint computation step, which consists of multiplying $\mathbf{K}^T$ with a vector, the computational time of a single iteration of Algorithm 4 is then $O(N^2m^2)$. Given that LIDAR image sizes can go up to 1 megapixels [82], computational time depending quadratically on N is too large for Algorithm 4 to work for applications requiring fast data processing. In addition, Algorithm 4 has limited photon efficiency, because its deconvolution relies solely on the single-reflector constraint. In particular, it ignores the spatial prior information of depth maps being spatially smooth, which is a safe assumption in high-resolution imaging applications. Thus, we now propose an alternative super-resolution imaging framework that performs depth and reflectivity reconstruction in linear time with respect to N while incorporating spatial priors by using a parametric approach of solving for the size-N depth and reflectivity maps directly, instead of using a non-parametric one.

The second approach: For the sake of illustration, let $\mathbf{y}'$ define a new vector that represents a size-m photon-count histogram measuring the combined responses from reflectors at depths $\{d^{(1)}, \ldots, d^{(\ell)}\}$ with reflectivities $\{a^{(1)}, \ldots, a^{(\ell)}\}$, that is obtained by illuminating the scene with N_s pulses, so that

$$\mathbf{y}' \sim \text{Poisson}\left(N_s \int_{(k-1)\Delta}^{k\Delta} \sum_{i=1}^{\ell} a^{(i)} s(t - 2d^{(i)}/c) dt \right), \quad \text{for } k = 1, 2, \ldots, m. \tag{5.8}$$

Here we assumed zero extraneous background and dark counts and unit quantum efficiency. (In our algorithm, we will background-correct the raw histogram vector in the pre-processing step).

Even when assuming that $N_s \to +\infty$, so that SNR grows without bound, and that the object reflectivities are known exactly, we see that the set of depth parameters of interest $\{d^{(1)}, \ldots, d^{(\ell)}\}$ is hard to estimate as it is non-linearly related to our observations. Consider the following linear sum of the raw photon count histogram $\mathbf{y}'$:

$$\mathbf{p} = \sum_{k=1}^{m} k \, \mathbf{y}'_k. \tag{5.9}$$

In the context of compressive depth imaging, and using assumptions of ideal noiseless sensing conditions, it has been shown that the scalar variable $\mathbf{p}$ is linear in the depth values $\{d^{(1)}, \ldots, d^{(\ell)}\}$ [83]. This can be seen by first recognizing that

$$\mathbf{p} \overset{(a)}{\approx} \sum_{k=1}^{m} \left(k N_s \int_{(k-1)\Delta}^{k\Delta} \sum_{i=1}^{\ell} a^{(i)} s(t - 2d^{(i)}/c) dt \right) \tag{5.10}$$

$$\overset{(b)}{\approx} \frac{N_s}{\Delta} \int_{0}^{T_r} t \sum_{i=1}^{\ell} a^{(i)} s(t - 2d^{(i)}/c) dt, \tag{5.11}$$

where (a) comes from a strong assumption that the acquisition time is large (large

N_s) and (b) assumes fine sampling (small Δ). Then, we have that

$$\begin{aligned} &\frac{N_s}{\Delta}\int_0^{T_r} t\sum_{i=1}^{\ell} a^{(i)} s(t-2d^{(i)}/c)dt && (5.12)\\ &= \frac{N_s}{\Delta}\sum_{i=1}^{\ell} a^{(i)}\int_0^{T_r} ts(t-2d^{(i)}/c)dt = \frac{2N_s}{c\Delta}\left[\sum_{i=1}^{\ell} a^{(i)}d^{(i)}\right], && (5.13)\end{aligned}$$

assuming that $\int ts(t)dt = 0$. This data transformation technique was demonstrated to be powerful in the compressive depth imaging framework of [83], where several $\mathbf{p}$ values are measured independently using a spatial light modulator that defines a forward imaging matrix with a restricted isometry property. (Essentially, Eq. (5.13), which is linear in depth, is observed for many combinations of pixels for depth reconstruction using linear inverse algorithms such as LASSO.)

Our imaging scenario's imaging kernel h mixes backscattered responses from several pixels, and so we can attempt to solve for the high-resolution depth map by defining an $N_x \times N_y$ matrix $\mathbf{P}$ that can be computed from low-light single-photon datasets. In our low-light imaging setup, let $\mathbf{Y}$ be the $N_x \times N_y \times m$ photon count histogram array obtained by performing the raster scanning acquisition over an $N_x \times N_y$ uniform grid. We have that the mean of $\mathbf{P}_{i,j}$ is

$$\begin{aligned} \mathbb{E}[\mathbf{P}_{i,j}] &= \sum_{k=1}^{m} k\mathbb{E}\left[\mathbf{Y}_{i,j,k}\right] && (5.14)\\ &= \sum_{k=1}^{m} kN_s \int_{(k-1)\Delta}^{k\Delta} \sum_{(i',j')\in\mathcal{H}_{i,j}} \mathbf{H}_{i-i',j-j'}\mathbf{A}_{i',j'} s(t-2\mathbf{D}_{i',j'}/c)dt, && (5.15)\end{aligned}$$

where $\mathcal{H}_{i,j}$ is the set of pixel indices defined by the support of the imaging kernel centered at (i,j), $\mathbf{H}$ is the $w \times w$ matrix representing the discretized version of the two-dimensional kernel h, and $\mathbf{A}$ and $\mathbf{D}$ are the $N_x \times N_y$ reflectivity and depth maps of interest. We can further simplify the expression in Eq. (5.15) down to

$$\mathbb{E}[\mathbf{P}_{i,j}] \approx \frac{2N_s}{c\Delta}\sum_{(i',j')\in\mathcal{H}_{i,j}} \mathbf{H}_{i-i',j-j'}\mathbf{A}_{i',j'}\mathbf{D}_{i',j'} = \frac{2N_s}{c\Delta}(\mathbf{H} * (\mathbf{A}\circ\mathbf{D}))_{i,j}, \tag{5.16}$$

where the approximation only requires the fine sampling assumption ($\Delta \to 0^+$), $*$ denotes two-dimensional discrete convolution, and $\circ$ denotes entrywise Hadamard product. Similarly, the variance of $\mathbf{P}_{i,j}$ can be derived as follows:

$$\mathrm{Var}(\mathbf{P}_{i,j}) = \sum_{k=1}^{m} k^2 \mathrm{Var}\left(\mathbf{Y}_{i,j,k}\right). \tag{5.17}$$

$$= \sum_{k=1}^{m} k^2 N_s \int_{(k-1)\Delta}^{k\Delta} \sum_{(i',j')\in\mathcal{A}_{i,j}} \mathbf{H}_{i-i',j-j'}\mathbf{A}_{i',j'} s(t - 2\mathbf{D}_{i',j'}/c)dt, \tag{5.18}$$

$$\approx \frac{N_s}{\Delta^2}(T_p^2(\mathbf{H} * \mathbf{A})_{i,j} + (4/c^2)(\mathbf{H} * (\mathbf{A} \circ \mathbf{D} \circ \mathbf{D}))_{i,j}), \tag{5.19}$$

where Eq. (5.17) is by the independence of photon counts among different time bins, Eq. (5.18) is by the Poisson distribution of photon counts for each time bin, and the approximation in Eq. (5.19) employs the fine sampling assumption. From our mean and variance derivations of our new data variable $\mathbf{P}$, we observe that to solve for $\mathbf{D}$, we require knowledge of $\mathbf{A}$. Also, the variance of the data variable $\mathbf{P}_{i,j}$ is typically larger than its mean, due to its quadratic dependence on $\mathbf{D}_{i,j}$ (see Appendix F for a full analysis of $\mathbf{P}$'s statistics). Finally, by summation of Poisson variables, noting that the total observed photon count variable $\mathbf{C}_{i,j} = \sum_{k=1}^{m} \mathbf{Y}_{i,j,k}$ obeys

$$\mathbf{C}_{i,j} \sim \mathrm{Poisson}((\mathbf{H} * \mathbf{A})_{i,j}), \tag{5.20}$$

we can develop the steps of our photon-efficient and time-efficient super-resolution framework (Approach 2) in the following way.

- **Step 1 – Compute data variables:**

  Compute $\mathbf{C}_{i,j} = \sum_{k=1}^{m} \mathbf{Y}'_{i,j,k}$ and $\mathbf{P}_{i,j} = \sum_{k=1}^{m}(k\mathbf{Y}'_{i,j,k})$ for all i, j, where $\mathbf{Y}' = \max(\mathbf{Y} - \mathbf{B}, 0.0)$ is the background corrected photon dataset. By employing a pre-processing step for background subtraction, instead of direct noise photon filtering mechanisms such as in the previous chapter, we are trading off decrease in photon efficiency for an increase in computational efficiency.

- **Step 2 – Reconstruct reflectivity:**

Compute the super-resolved reflectivity estimate $\hat{\mathbf{A}}_{\text{SR}}$ by finding the maximum-likelihood solution with total variation (TV) regularization:

$$\begin{aligned} \underset{\mathbf{A}}{\text{minimize}} \quad & \mathcal{L}_{\mathbf{A}}(\mathbf{A}; \mathbf{H}, \mathbf{C}) + \tau_{\mathbf{A}} \|\mathbf{A}\|_{\text{TV}} \\ \text{subject to} \quad & \mathbf{A}_{i,j} \geq 0, \quad \forall i, j, \end{aligned} \tag{5.21}$$

where $\mathcal{L}_{\mathbf{A}}(\mathbf{A}; \mathbf{H}, \mathbf{C})$ is the negative log Poisson likelihood function of reflectivity map $\mathbf{A}$, given that the rate parameter of Poisson variable $\mathbf{C}_{i,j}$ is $(\mathbf{H} * \mathbf{A})_{i,j}$, for all i, j, and $\|\mathbf{A}\|_{\text{TV}} = \sum_{i,j} \sum_{(i',j') \in \text{neighbors}(i,j)} |\mathbf{A}_{i,j} - \mathbf{A}_{i',j'}|$ to be the TV norm. Also, $\tau_{\mathbf{A}} > 0$ is used as a regularization parameter that controls the degree of penalizing non-smooth reflectivity map estimates with high TV norms. Since the negative log Poisson likelihood and the TV norm are both convex functions in $\mathbf{A}$, (5.21) is a convex optimization problem whose optimum can be solved efficiently using a combination of gradient descent and fast iterative shrinkage-thresholding algorithm (FISTA) [40]. A general algorithm that solves this convex optimization problem is summarized as pseudocode in Algorithm 5, which is implemented in the software for sparse Poisson intensity reconstruction algorithms (SPIRAL) [58].

- **Step 3 – Reconstruct depth:**

Compute the super-resolved depth estimate $\hat{\mathbf{D}}_{\text{SR}}$ by solving for the approximate maximum-likelihood solution with total variation (TV) regularization:

$$\begin{aligned} \underset{\mathbf{D}}{\text{minimize}} \quad & \mathcal{L}_{\mathbf{D}}(\mathbf{D}; \mathbf{H}, \hat{\mathbf{A}}_{\text{SR}}, \mathbf{P}) + \tau_{\mathbf{D}} \|\mathbf{D}\|_{\text{TV}} \\ \text{subject to} \quad & \mathbf{D}_{i,j} \geq 0, \quad \forall i, j, \end{aligned} \tag{5.22}$$

where $\mathcal{L}_{\mathbf{D}}(\mathbf{D}; \mathbf{H}, \hat{\mathbf{A}}_{\text{SR}}, \mathbf{P})$ is the negative log Poisson likelihood function of the super-resolved depth map $\mathbf{D}$, given that the rate parameter of Poisson variable $\mathbf{P}_{i,j}$ is $(\mathbf{H} * (\hat{\mathbf{A}}_{\text{SR}} \circ \mathbf{D}))_{i,j}$. Observe that we use the Poisson approximation to the distribution of $\mathbf{P}_{i,j}$, knowing that it is a signal-dependent variable with its mean being larger than its standard deviation, and that the super-resolved reflectivity estimate $\hat{\mathbf{A}}_{\text{SR}}$ obtained from Step 2 is used as an input to Step 4. The regularization parameter $\tau_{\mathbf{D}} > 0$ controls the degree of penalizing non-smooth depth map estimates with high TV norms. Since the negative log Poisson likelihood and the TV norm are both convex functions in $\mathbf{D}$, (5.22) is a convex optimization problem. A general algorithm that solves for this convex optimization problem is also described by the pseudocode in Algorithm 5.

All steps of the above framework are summarized in Algorithm 6 as pseudocode.

5.5 Results

To validate our super-resolution imaging framework, we simulated a scanning LIDAR setup at low light-levels. We first compared the depth recovery accuracy of the proposed framework using our Approaches 1 and 2 with conventional methods of pixelwise ML reflectivity and depth estimation and deconvolution-based ML.

Using Eq. (5.2), the baseline ML estimators $\hat{\mathbf{a}}_{\text{ML}}$ and $\hat{\mathbf{d}}_{\text{ML}}$ given that the observations are corrupted by photon noise and assuming an identity matrix PSF ($\mathbf{H} = \mathbf{I}_{N \times N}$) can be written as

$$(\hat{\mathbf{a}}_{\text{ML}})_i = \sum_{k=1}^{m} \mathbf{Y}'_{i,k} \tag{5.23}$$

$$(\hat{\mathbf{d}}_{\text{ML}})_i = \underset{l \in \{1,\ldots,m\}}{\arg\max}\ \mathbf{Y}'_{i,:} \log(\mathbf{S}_{:,l} + \epsilon) \tag{5.24}$$

for $i = 1, \ldots, N$, where $\mathbf{Y}'_{i,:}$ denotes the i-th row of $\mathbf{Y}'$, $\mathbf{S}_{:,l}$ denotes the l-th column of $\mathbf{S}$, and a small number ϵ is added to avoid numerical instability (we use $\epsilon = 2.2 \times 10^{-16}$ using the built-in variable `eps` in MATLAB). The depth estimate is simply the log-matched filter [19]. Another estimator we use for comparison is the deconvolution-

Algorithm 5 Poisson deconvolution by TV-regularized optimization (Deconv-TV)

```
Input: $\mathbf{J}$, $\mathbf{K}$, $\mathbf{L}$, $\delta$, $\tau$, $\hat{\mathbf{X}}^{(0)}$
Output: $\hat{\mathbf{X}}_{\text{dec}}$
Initialize $\mathcal{L}(\mathbf{X}; \mathbf{J}, \mathbf{K}, \mathbf{L})$ to be the negative log Poisson likelihood with data $\mathbf{J}$ and rate parameter $\mathbf{K} * (\mathbf{L} \circ \mathbf{X})$
$k \leftarrow 0$
repeat
    $k \leftarrow k + 1$
    $\alpha^{(k-1)} \leftarrow \text{BB}(\alpha^{(k-1)}, \mathbf{J}, \hat{\mathbf{X}}^{(k-1)}, \delta)$    ▷ Optimal step size by Barzilai-Borwein [84]
    $\hat{\mathbf{X}} \leftarrow \hat{\mathbf{X}}^{(k-1)} - \alpha^{(k-1)} \nabla_{\mathbf{X}} \mathcal{L}(\mathbf{X}; \mathbf{J}, \mathbf{K}, \mathbf{L})|_{\mathbf{X}=\hat{\mathbf{X}}^{(k-1)}}$    ▷ Gradient descent
    $\hat{\mathbf{X}}^{(k)} \leftarrow \text{FISTA}(\hat{\mathbf{X}}, \tau)$    ▷ TV denoising by FISTA [40]
until $\|\mathbf{X}^{(k-1)} - \mathbf{X}^{(k)}\|_2^2 < \delta$
$\hat{\mathbf{X}}_{\text{dec}} \leftarrow \hat{\mathbf{X}}^{(k)}$
```

Algorithm 6 Proposed algorithm for super-resolution LIDAR (Approach 2)

```
Input: $\mathbf{Y}'$, $\mathbf{H}$, $\delta$, $\tau_{\mathbf{A}}$, $\tau_{\mathbf{D}}$
Output: $\hat{\mathbf{A}}_{\text{SR}}$, $\hat{\mathbf{D}}_{\text{SR}}$
Initialize $\mathbf{C}$ and $\mathbf{P}$ as $N_x \times N_y$ empty matrices
for $i = 1, 2, \ldots, N_x$ do
    for $j = 1, 2, \ldots, N_y$ do
        $\mathbf{C}_{i,j} \leftarrow \sum_{k=1}^{m} \mathbf{Y}'_{i,j,k}$
        $\mathbf{P}_{i,j} \leftarrow \sum_{k=1}^{m} (k\mathbf{Y}'_{i,j,k})$    ▷ Compute data variables
    end for
end for
Compute pixelwise ML solutions $\hat{\mathbf{A}}_{\text{ML}}$, $\hat{\mathbf{D}}_{\text{ML}}$ to initialize convex programs
$\hat{\mathbf{A}}_{\text{SR}} \leftarrow \text{Deconv-TV}(\mathbf{C}, \mathbf{H}, \mathbf{I}, \delta_{\mathbf{A}}, \tau_{\mathbf{A}}, \hat{\mathbf{A}}_{\text{ML}})$    ▷ Compute super-resolved reflectivity
$\hat{\mathbf{D}}_{\text{SR}} \leftarrow \text{Deconv-TV}(\mathbf{P}, \mathbf{H}, \hat{\mathbf{A}}_{\text{SR}}, \delta_{\mathbf{D}}, \tau_{\mathbf{D}}, \hat{\mathbf{D}}_{\text{ML}})$    ▷ Compute super-resolved depth
```

based ML method (also known as pseudo-inverse filtering) where we have

$$(\hat{\mathbf{a}}_{\text{ML-DC}})_i = \sum_{k=1}^{m} \left(\mathbf{H}^{\dagger}\mathbf{Y}'\right)_{i,k} \tag{5.25}$$

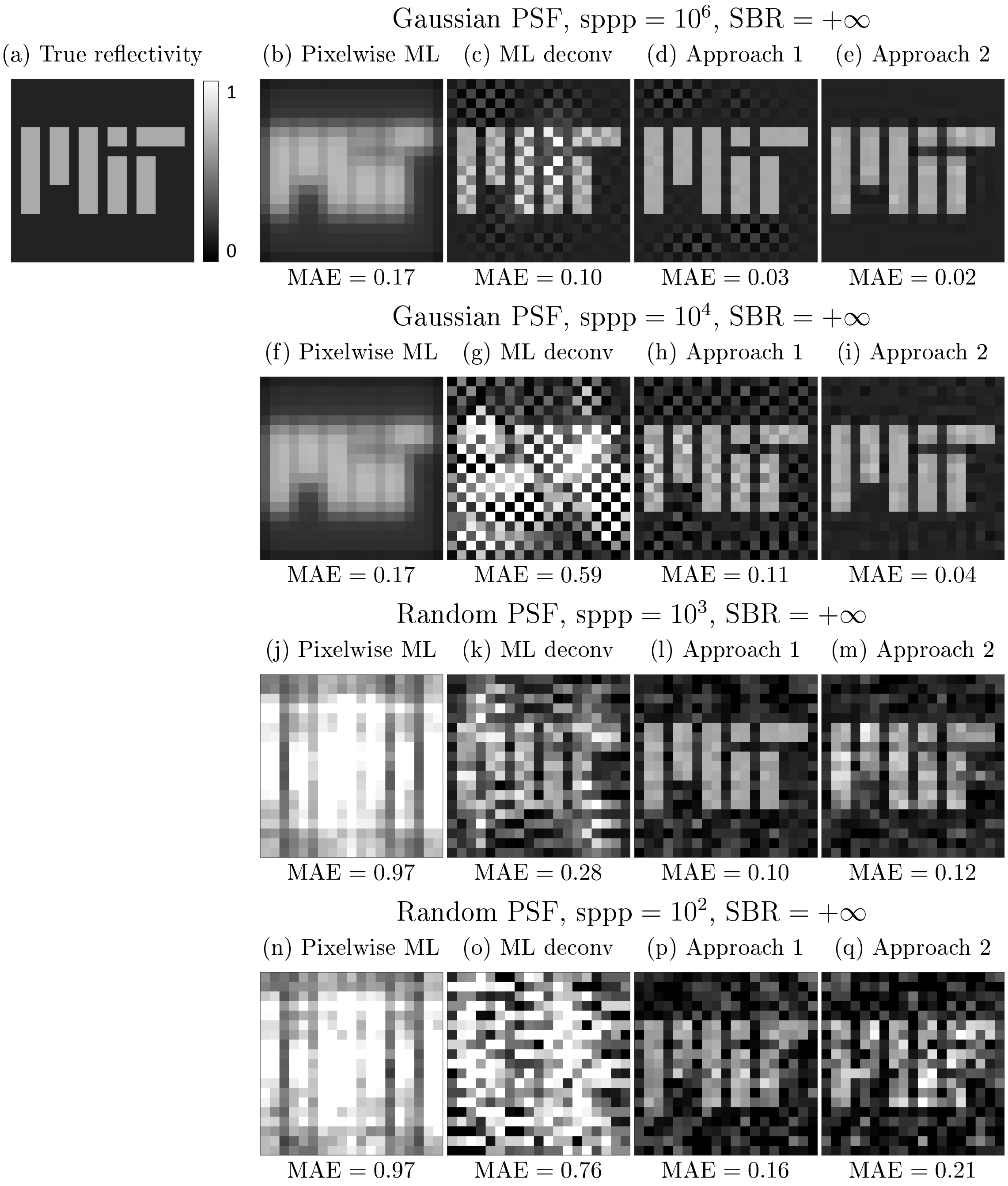
$$(\hat{\mathbf{d}}_{\text{ML-DC}})_i = \underset{l \in \{1,\ldots,m\}}{\arg\max} \ \left(\mathbf{H}^{\dagger}\mathbf{Y}'\right)_{i,:} \log(\mathbf{S}_{:,l} + \epsilon), \tag{5.26}$$

for $i = 1, \ldots, N$, where $\mathbf{H}^{\dagger}$ is the minimum-norm pseudoinverse of the PSF $\mathbf{H}$, which is no longer taken to be the identity matrix. We observe that $\hat{\mathbf{d}}_{\text{ML-DC}}$ is obtained from a two-step approach of first deconvolving the noisy data using the PSF and then estimating depth.

In Figures 5-4 and 5-5, we show the results of reconstructing the reflectivity and depth of the MIT logo using the scanning LIDAR setup with a photon-counting detector. The spatial resolution of the MIT logo image was 19×19 and thus we had $N = 361$. The ground truth reflectivity and depth maps of the MIT logo are shown in Figure 5-4(a) and Figure 5-5(a), respectively. For this simulation, the number of detector time bins was $m = 10$. Note that even with our low-resolution native image, the size of matrix $\mathbf{K}$ in Algorithm 4 of Approach 1 is $Nm \times Nm = 3610 \times 3610$, which is quite large.

Here the illumination PSF was set to be either a 2D symmetric Gaussian or Bernoulli. The root mean square extent of the Gaussian PSF was approximately 1 scanning pixel (the dimensions of Gaussian PSF was 10×10), and the Bernoulli PSF was generated randomly using 0.1 success probability with its dimensions being 19×19. Also, we assumed that the pulsewidth T_p is much smaller than the detector time-bin Δ and set $\mathbf{S}$ to be the identity matrix. For Approach 1, we set $\delta = 0.01$ for all experiments. For Approach 2, we set $\delta = 10^{-4}$, $\tau_{\mathbf{A}} = 1$ and $\tau_{\mathbf{D}}$ was chosen to minimize the absolute error between $\mathbf{D}$ and $\hat{\mathbf{D}}_{\text{SR}}$. In practice, since the true $\mathbf{D}$ is unavailable for the tuning of $\tau_{\mathbf{D}}$, cross-validation methods can be used to train for the regularization parameters prior to the imaging experiment [85]. The code used to generate our results is available at [86].

For the imaging scenarios with the Gaussian PSF in the absence of background



Gaussian PSF, sppp = 10^6, SBR = $+\infty$

(a) True reflectivity (b) Pixelwise ML (c) ML deconv (d) Approach 1 (e) Approach 2

MAE = 0.17 MAE = 0.10 MAE = 0.03 MAE = 0.02

Gaussian PSF, sppp = 10^4, SBR = $+\infty$

(f) Pixelwise ML (g) ML deconv (h) Approach 1 (i) Approach 2

MAE = 0.17 MAE = 0.59 MAE = 0.11 MAE = 0.04

Random PSF, sppp = 10^3, SBR = $+\infty$

(j) Pixelwise ML (k) ML deconv (l) Approach 1 (m) Approach 2

MAE = 0.97 MAE = 0.28 MAE = 0.10 MAE = 0.12

Random PSF, sppp = 10^2, SBR = $+\infty$

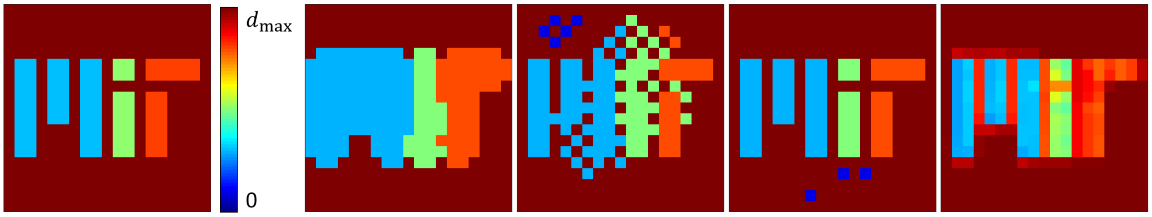
(n) Pixelwise ML (o) ML deconv (p) Approach 1 (q) Approach 2

MAE = 0.97 MAE = 0.76 MAE = 0.16 MAE = 0.21

Figure 5-4: Reflectivity reconstruction results (with mean absolute errors (MAE)) for a 19×19 MIT logo scene when SBR = $+\infty$. We compare the pixelwise ML, deconvolved ML, and the proposed frameworks (Approach 1 and 2) for different PSFs and values of signal photons per pixel (sppp). (b–e) Using Gaussian PSF and sppp = 10^6. (f–i) Using Gaussian PSF and sppp = 10^4. (j–m) Using Bernoulli PSF and sppp = 10^3. (n–q) Using Bernoulli PSF and sppp = 10^2.

Gaussian PSF, sppp = 10^6, SBR = $+\infty$

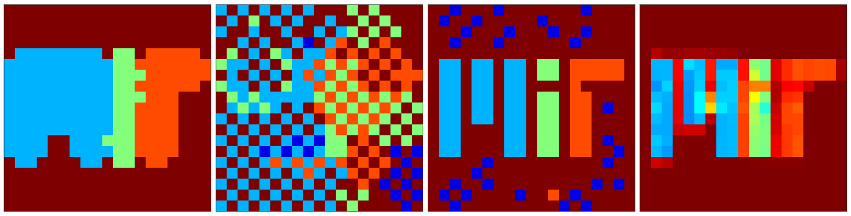
(a) True depth (b) Pixelwise ML (c) ML deconv (d) Approach 1 (e) Approach 2



MAE = 1.24 MAE = 0.70 MAE = 0.07 MAE = 0.26

Gaussian PSF, sppp = 10^4, SBR = $+\infty$

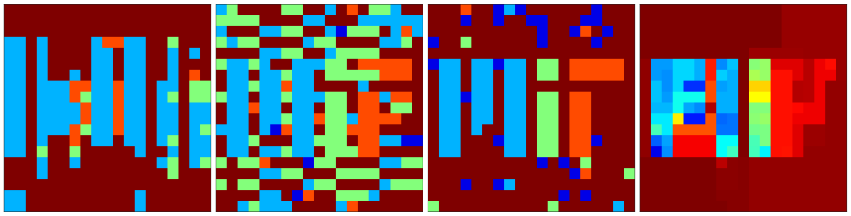
(f) Pixelwise ML (g) ML deconv (h) Approach 1 (i) Approach 2



MAE = 1.22 MAE = 2.67 MAE = 0.75 MAE = 0.23

Random PSF, sppp = 10^3, SBR = $+\infty$

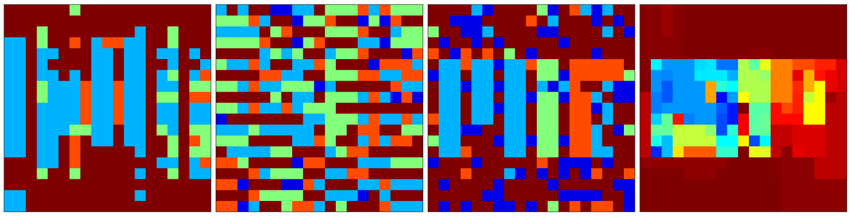
(j) Pixelwise ML (k) ML deconv (l) Approach 1 (m) Approach 2



MAE = 2.51 MAE = 2.11 MAE = 0.84 MAE = 0.49

Random PSF, sppp = 10^2, SBR = $+\infty$

(n) Pixelwise ML (o) ML deconv (p) Approach 1 (q) Approach 2



MAE = 2.68 MAE = 3.01 MAE = 2.22 MAE = 0.89

Figure 5-5: Depth reconstruction results (with mean absolute errors (MAE)) for a 19×19 MIT logo scene when SBR = $+\infty$. We compare the pixelwise ML, deconvolved ML, for different PSFs and values of signal photons per pixel (sppp). (b–e) Using Gaussian PSF and sppp = 10^6. (f–i) Using Gaussian PSF and sppp = 10^4. (j–m) Using Bernoulli PSF and sppp = 10^3. (n–q) Using Bernoulli PSF and sppp = 10^2.

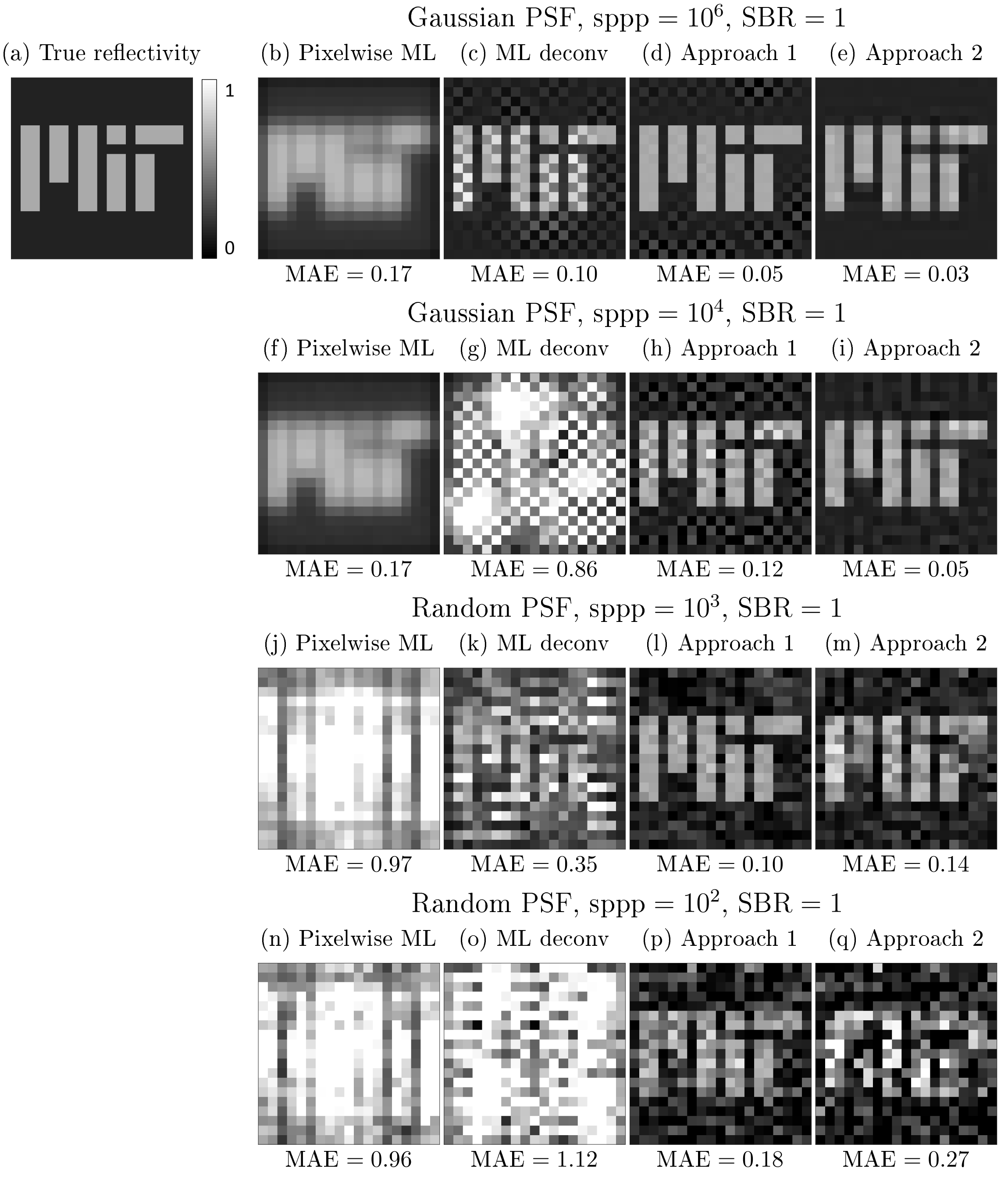
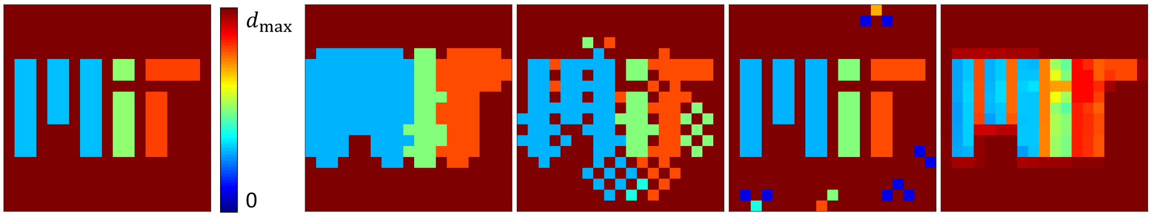


Figure 5-6: Reflectivity reconstruction results (with mean absolute errors (MAE)) for a 19×19 MIT logo scene when SBR = 1. We compare the pixelwise ML, deconvolved ML, and the proposed frameworks (Approach 1 and 2) for different PSFs and values of signal photons per pixel (sppp). (b–e) Using Gaussian PSF and sppp = 10^6. (f–i) Using Gaussian PSF and sppp = 10^4. (j–m) Using Bernoulli PSF and sppp = 10^3. (n–q) Using Bernoulli PSF and sppp = 10^2.

Gaussian PSF, sppp = 10^6, SBR = 1

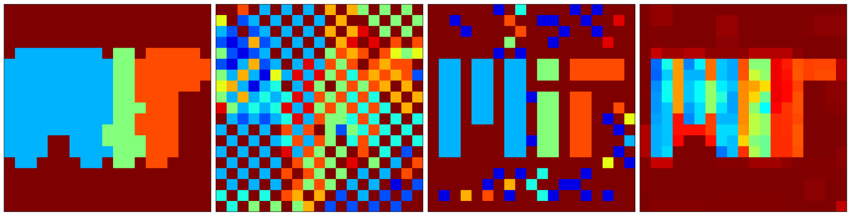
(a) True depth (b) Pixelwise ML (c) ML deconv (d) Approach 1 (e) Approach 2



MAE = 1.23 MAE = 0.67 MAE = 0.27 MAE = 0.30

Gaussian PSF, sppp = 10^4, SBR = 1

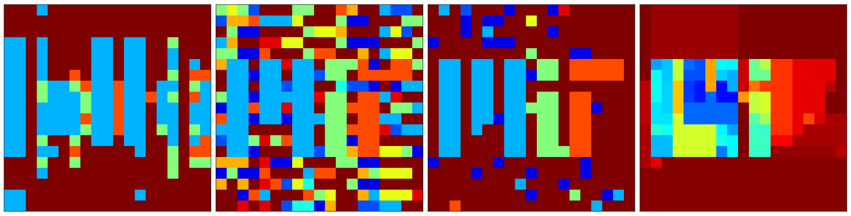
(f) Pixelwise ML (g) ML deconv (h) Approach 1 (i) Approach 2



MAE = 1.22 MAE = 2.84 MAE = 0.91 MAE = 0.42

Random PSF, sppp = 10^3, SBR = 1

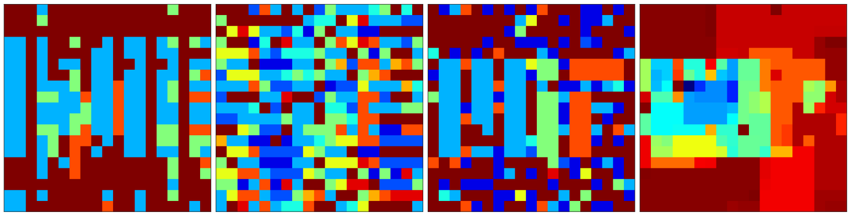
(j) Pixelwise ML (k) ML deconv (l) Approach 1 (m) Approach 2



MAE = 2.43 MAE = 2.51 MAE = 0.84 MAE = 0.68

Random PSF, sppp = 10^2, SBR = 1

(n) Pixelwise ML (o) ML deconv (p) Approach 1 (q) Approach 2



MAE = 2.71 MAE = 3.71 MAE = 2.38 MAE = 1.20

Figure 5-7: Depth reconstruction results (with mean absolute errors (MAE)) for a 19×19 MIT logo scene when SBR = 1. We compare the pixelwise ML, deconvolved ML, and the proposed frameworks (Approach 1 and 2) for different PSFs and values of signal photons per pixel (sppp). (b–e) Using Gaussian PSF and sppp = 10^6. (f–i) Using Gaussian PSF and sppp = 10^4. (j–m) Using Bernoulli PSF and sppp = 10^3. (n–q) Using Bernoulli PSF and sppp = 10^2.

light, we see that conventional pixelwise ML lead to blurry reflectivity and depth estimates of the letters MIT ((b) of Figures 5-4 and 5-5). The two-step approach of deconvolving data and then performing ML estimation partially recovers the high-resolution reflectivity and depth features of the MIT logo when the mean number of photons-per-pixel (ppp) is 10^6 ((c) of Figures 5-4 and 5-5). Compared to the two-step deconvolution approach, our Approach 1 successfully recovers the full reflectivity and depth maps without many corrupted pixels ((d) of Figures 5-4 and 5-5). Approach 2 also gives successful reconstruction with minor errors ((e) of Figures 5-4 and 5-5). The mean absolute error numbers presented for depth reconstruction are in the units of $c\Delta/2$ meters. In other words, if we assume that the time bin length is 1 ns, then the error of Approach 2 in cm is simply $100 \times (0.12 \times c \times 10^{-9})/2 = 1.8$ cm with maximum range of the imager being $d_{\max} = 18$ cm. At lower light-levels (ppp $= 10^4$), the ML+deconvolution approach fails to recover useful features even compared to the ML estimate ((f) of Figures 5-4 and 5-5), owing to a decrease in signal-to-noise ratio ((g) of Figures 5-4 and 5-5). On the other hand, Approach 1 recovers the MIT logo reflectivity and depth while having a small number corrupted pixels ((h) of Figures 5-4 and 5-5) and Approach 2 gives an acceptable reconstruction with smoothing artifacts, especially for the depth map ((i) of Figures 5-4 and 5-5).

For imaging scenarios with the Bernoulli PSF in the absence of background light, we have a similar set of results showing that our framework outperforms conventional LIDAR techniques for high-resolution reflectivity and depth imaging. One observation that we can make is because the matrix condition number of $\mathbf{A}$ from the Gaussian PSF is 3764 and that from the random PSF is 440, our Approach 1 is more robust to photon noise with the random PSF than with the Gaussian PSF. For example, the reflectivity and depth reconstructions of our Approach 1 using Gaussian PSF at ppp $= 10^4$ is similar in quality to that using random PSF at ppp $= 10^3$ ((l) of Figures 5-4 and 5-5). This leads to an interesting optical design question, related to coded aperture photography in classical high light level imaging [87, 88], as to how one should design the illumination PSF for optimal LIDAR performance. At much lower light levels (ppp $= 10^2$), Approach 1 starts to break down with $\sim$1/3 of the reconstructed pixels being noisy ((p) of Figures 5-4 and 5-5). Approach 2 is unable to cope with the data variable $\mathbf{P}$ generated by the random PSF as effectively as Approach 1. ((m),(q) of Figures 5-4 and 5-5). Figures 5-6 and 5-7 perform the same imaging experiment while setting SBR $= 1$. As expected, compared to results from Figures 5-4 and 5-5 we observe that there is a serious deterioration in image reconstruction performance of the estimators at all light levels tested.

Earlier in this chapter, we noted that Approach 1 cannot be used for imaging scenarios in which the number of pixels is large. Even for the previous simulations with $N = 361$, a single run of Algorithm 4 of Approach 1 took around 10 minutes, while a single run of Algorithm 6 of Approach 2 took less than a second. Figure 5-8 and 5-9 shows simulated results on how Approach 2 succeeds in reconstructing high spatial resolution reflectivity and depth maps at low light levels, where the mean number of backscattered photons per pixel was around 9.8, for signal-to-background (SBR) ratios of $+\infty$ and 1. We chose $\mathbf{H}$ to be a Gaussian kernel with a 15-pixel root mean square extent. Here we compare our framework not only with the results of pixelwise ML reflectivity and depth estimate, but also with the Richardson-Lucy deconvolution algorithm [79, 80], which is designed to perform super-resolution on images corrupted with photon noise, applied on pixelwise ML solutions.

In Figure 5-8, the simulations used SBR $= +\infty$ (zero extraneous background and dark counts). Although the results from Richardson-Lucy (Figure 5-8(e),(f)) show corrections of the image blurring effect from the pixelwise ML solutions (Figure 5-8(c),(d)), they still suffer from noise coming from low photon counts. Because our Approach 2 combines approximate likelihood functions derived from photodetection with spatial correlations, it produces more accurate super-resolved image results than the existing methods, which fail to capture all the LIDAR constraints (Figure 5-8(g),(h)). Figure 5-9 shows simulation results when SBR $= 1$ (detecting a photon from background plus dark counts is as likely as detecting a photon returning from scene of interest), and comparing with results in Fig. 5-8 we can observe the similar phenomena in recovery performance for all estimators, but with deterioration from extra photon noise coming from background light.

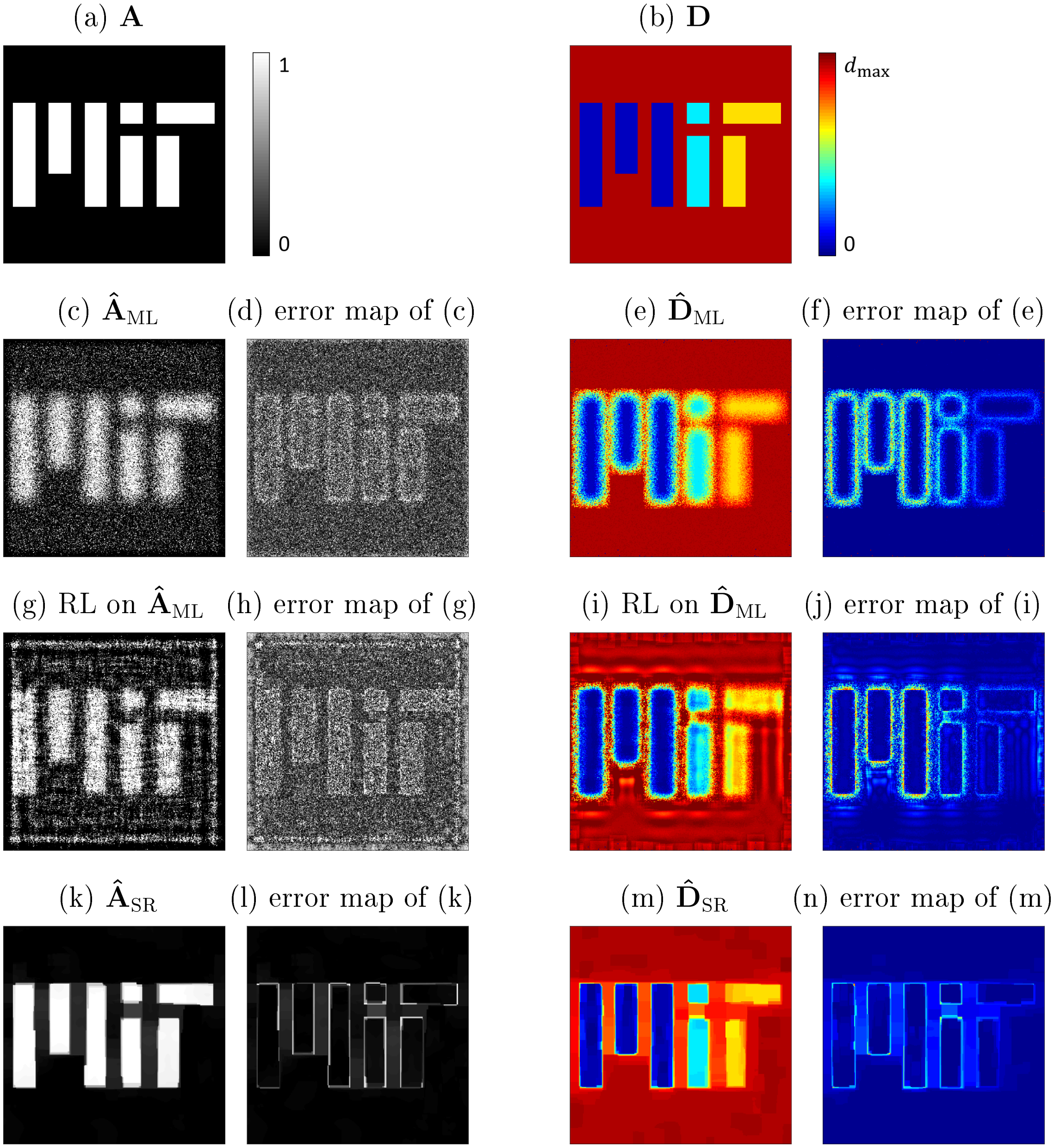


Figure 5-8: Reflectivity and depth reconstruction results for a simulated ultra low-light LIDAR setup with MIT logo scene for SBR $= +\infty$, with high native resolution of 270×275 (thus, $N = 74250$). Here, the mean number of photons per pixel was 9.8. We compare the pixelwise ML, ML after Richardson-Lucy (RL) deconvolution, and results of our Approach 2 for both reflectivity and depth recovery. Here $d_{\max}$ is 18 cm.

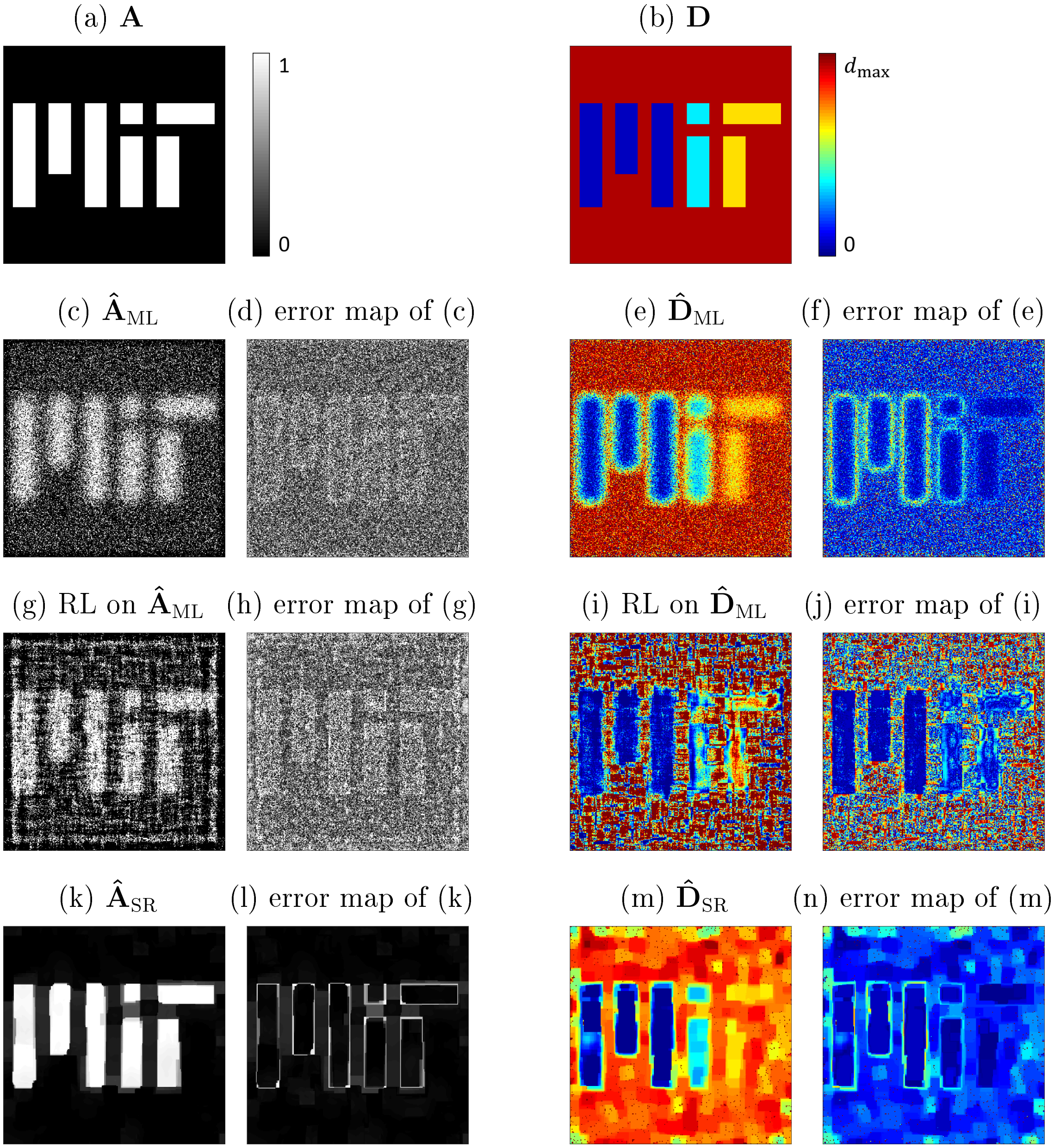


Figure 5-9: Reflectivity and depth reconstruction results for a simulated ultra low-light LIDAR setup with MIT logo scene for SBR = 1, with high native resolution of 270×275 (thus, $N = 74250$). Here, the mean number of photons per pixel was 9.8. We compare the pixelwise ML, ML after Richardson-Lucy (RL) deconvolution, and results of our Approach 2 for both reflectivity and depth recovery. Here $d_{\max}$ is 18 cm.

5.6 Summary and Discussion

In this chapter, we presented a super-resolution LIDAR framework using moderate-flux ($\sim 10^3$ ppp) and low-flux (~ 10 ppp) photon-count measurements. Our framework combined photon detection statistics, illumination PSF constraints, and model-based sparsity constraints of reflective surfaces to formulate an optimization problem for joint deconvolution and depth estimation. We proposed two super-resolution depth imaging approaches. The first approach was a greedy algorithm, inspired by sparse signal pursuit methods, that recovers the super-resolved depth image by solving a model-based sparse optimization problem. The second approach relied on defining new data variables and also incorporating spatial priors, such that the overall data-processing burden is much lower and it can be more photon-efficient than the first approach for high resolution scenes. Using low-light LIDAR simulation results, we demonstrated that both approaches of our framework outperform the approach of combining existing deconvolution methods with pixelwise ML for several low light level conditions, even in the presence of strong background light.

Chapter 6

Single-Pixel Imaging

6.1 Overview of Problem

To form an image consisting of N pixels, a conventional digital camera uses an N-element detector array to take direct light intensity measurements. A single-pixel camera, on the other hand, forms an N-pixel reflectivity image by solving an inverse problem using a single photodetector and M patterns projected from a N-pixel spatial light modulator (SLM) [89], such as a digital micromirror device (DMD) [90]. When a photodetector is expensive, such as one that operates in non-visible wavelengths or is sensitive to individual photons, single-pixel cameras have the potential to be much cheaper than array detection-based cameras [91].

Prior art: When the amount of light reaching the imaging device is low and the detector records the incident light intensity as discrete photon counts, the photodetection process is the dominating noise source [46]. Many works in signal processing, such as the Rice single-pixel camera project, have concentrated on having $M < N$ in the context of compressed sensing [92]. For better reconstruction within the low-flux single-pixel architecture, however, overcomplete measurements ($M > N$) are used [93]. Even for the simplest estimators, analyzing their performance by computing the mean-squared error (MSE) for this imaging architecture is non-trivial. This is due to the difficulties introduced by the Poisson statistics of photon counting and the random, non-Gaussian SLM patterns of the camera. Thus, MSE is conventionally estimated using Monte Carlo simulations [94], which are computationally costly.

Recently, random matrix theory (RMT) has been used to obtain accurate closed-form characterization of estimation errors in the context of channel identification [95], wireless communications [96], inverse covariance estimation [97], and many more topics in signal processing. Although it may be convenient to directly use RMT-based error analysis results for the single-pixel imaging framework, there are two novel constraints that prevent us from doing this.

1. In the low-flux single-pixel imaging scenario, measurements are described by signal-dependent Poisson observations. This statistical behavior deviates from the conventional noiseless channels or Gaussian channels with fixed-variance, signal-independent noise.

2. In conventional error analysis frameworks, the random variables in the system typically have zero means, allowing direct application of classical RMT results. However, in our single-pixel imaging scenario, the modulation patterns are required to be non-negative as light intensity is non-negative.

Summary of our approach: In this chapter, we derive a closed-form approximation for the MSE of least-squares single-pixel imaging that accounts for the Poisson observation channel and the random SLM patterns. We borrow techniques from RMT to arrive at simple MSE expressions, avoiding the implicit characterizations obtained through the replica method [98, 99] and issues arising from the non-zero mean of the SLM patterns. Our numerical experiments show that our approximation is near exact, while having lower computational complexity than Monte Carlo methods and higher accuracy than a baseline asymptotic MSE computation. Because its image reconstruction performance is well-understood using our framework, the least-squares single-pixel imager is an attractive one to use in low-flux conditions.

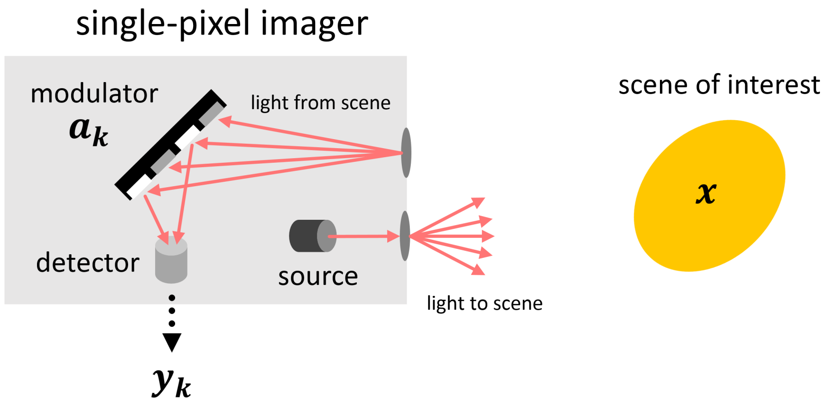


Figure 6-1: Active single-pixel imaging setup. For the kth measurement, the constant flux sent from source ($s(t) = 1/T_r$, $t \in [0, T_r)$) and reflected from $\mathbf{x}$ is spatially-modulated using the pattern $\mathbf{a}_k$, such that the flux incident on the single-photon detector is $\mathbf{a}_k^T\mathbf{x}$. The observation $\mathbf{y}_k \sim \text{Poisson}(\mathbf{a}_k^T\mathbf{x})$ is made by the photodetector, and the process is repeated for M measurements. Although this figure depicts the spatial light modulator as a reflector, such as a programmable mirror device, it may work in transmission mode by using a random diffraction grating instead.

6.2 Single-Photon Imaging Setup

Figure 6-1 illustrates our single-pixel imaging setup. Let $\mathbf{x} \in \mathbb{R}_+^{N\times 1}$ be the scene reflectivity vector that we aim to recover. If the desired image is two-dimensional, one can vectorize it to form $\mathbf{x}$. An SLM with N pixels reflects the light coming from scene at a random subset of the pixels to the detector. Let $\mathbf{a}_k$ denote the $N \times 1$ random vector representing the kth SLM pattern, where each of its entries are i.i.d. on the support $[0, 1]$, as the SLM is a device with no gain, i.e., it cannot amplify the illumination incident upon it.

6.3 Forward Imaging Model

In this chapter, also for practical purposes [91], we assume that each entry of $\mathbf{a}_k$, $k = 1, \ldots, M$, is an i.i.d. Bernoulli variable with success probability of $p \in (0, 1)$. Using low-flux photodetection statistics, we can write our kth photon count observation as

$$\mathbf{y}_k \sim \text{Poisson}\left(\mathbf{a}_k^T\mathbf{x}\right), \qquad k = 1, 2, \ldots, m, \tag{6.1}$$

meaning that it has the probability mass function

$$\Pr(\mathbf{y}_k|\mathbf{a}_k;\mathbf{x}) = \frac{\exp\{-\mathbf{a}_k^T\mathbf{x}\}\left(\mathbf{a}_k^T\mathbf{x}\right)^{\mathbf{y}_k}}{\mathbf{y}_k\,!}, \qquad \text{for } \mathbf{y}_k \in \{0,1,2,\ldots\}, \tag{6.2}$$

and that

$$\Pr((\mathbf{a}_k)_i) = p^{(\mathbf{a}_k)_i}(1-p)^{(1-(\mathbf{a}_k)_i)}, \qquad \text{for } (\mathbf{a}_k)_i \in \{0,1\}, \tag{6.3}$$

where $i = 1, 2, \ldots, N$ given $k \in \{1, \ldots, M\}$. Because we are interested in steady-state illumination for the single-pixel imaging architecture, we assume any background or dark counts are included in $\mathbf{x}$.

We write Eq. (6.1) more compactly as

$$\mathbf{y} \sim \text{Poisson}(\mathbf{A}\mathbf{x})\,, \tag{6.4}$$

where $\mathbf{y}$ is the column vector formed from the M observations of photon counts, $\mathbf{A}$ is the $M \times N$ random matrix that row-concatenates the vectors $\mathbf{a}_1^T, \ldots, \mathbf{a}_M^T$. The inverse problem of single-pixel imaging is to recover the size-N image of $\mathbf{x}$ using the observations $\mathbf{y}$ and SLM pattern matrix $\mathbf{A}$. If $\mathbf{A}$ were replaced by an identity matrix, Eq. (6.4) would be the intensity measurement model for raster-scanned LIDAR with photon-counting detection [100]; methods to exploit spatial correlation for good low-flux performance are presented in [12, 13]. Alternatively, acquisition with square and nonsingular $\mathbf{A}$ not equal to the identity is termed multiplexed imaging [101]. Multiplexed imaging has been the subject of analysis supporting the use of Hadamard matrices [102, 103], and it has been shown that enforcing non-negativity constraints on $\mathbf{x}$ causes a multiplexing advantage to extend to low-flux operation [104].

6.4 Solving the Inverse Problem

Because we are interested in low-flux operations, we will assume $M \geq N$ to allow effective noise mitigation. The linear inverse problem at hand is thus invertible in the classical sense, without employing image priors. Based on Eq. (6.4), we can write the negative log likelihood function of $\mathbf{x}$ given data $\mathbf{y}$ and $\mathbf{A}$ as

$$\mathcal{L}(\mathbf{x}; \mathbf{A}, \mathbf{y}) = \sum_{k=1}^{M} \left[(\mathbf{A}\mathbf{x})_k - \mathbf{y}_k \log(\mathbf{A}\mathbf{x})_k\right], \tag{6.5}$$

where terms independent of $\mathbf{x}$ have been omitted. We observe that $\mathcal{L}(\mathbf{x}; \mathbf{A}, \mathbf{y})$ is a convex function in $\mathbf{x}$ (see Appendix C). Thus, checking the zero-gradient condition suffices to solve for the maximum likelihood solution. We have that the gradient is

$$\nabla_{\mathbf{x}} \mathcal{L}(\mathbf{x}; \mathbf{A}, \mathbf{y}) = \sum_{k=1}^{M} \left(1 - \frac{\mathbf{y}_k}{(\mathbf{A}\mathbf{x})_k}\right) (\mathbf{A}^T)_{:,k}. \tag{6.6}$$

When $\mathbf{A}$ is square ($M = N$) and invertible ($\text{rank}(\mathbf{A}) = N$), the inverse solution $\hat{\mathbf{x}}_{\text{inv}} = \mathbf{A}^{-1}\mathbf{y}$ makes Eq. (6.6) zero as

$$\begin{aligned}
\sum_{k=1}^{M} \left(1 - \frac{\mathbf{y}_k}{(\mathbf{A}\hat{\mathbf{x}}_{\text{inv}})_k}\right) (\mathbf{A}^T)_{:,k} &= \sum_{k=1}^{M} \left(1 - \frac{\mathbf{y}_k}{(\mathbf{A}\mathbf{A}^{-1}\mathbf{y})_k}\right) (\mathbf{A}^T)_{:,k} &\quad (6.7)\\
&= \sum_{k=1}^{M} \left(1 - \frac{\mathbf{y}_k}{\mathbf{y}_k}\right) (\mathbf{A}^T)_{:,k}. &\quad (6.8)\\
&= \sum_{k=1}^{M} (1 - 1)\, (\mathbf{A}^T)_{:,k} = 0. &\quad (6.9)
\end{aligned}$$

However, we are interested in the setup using overcomplete measurements ($M \geq N$), where the inverse solution generalizes to the pseudoinverse least-squares estimator:

$$\hat{\mathbf{x}}_{\text{LS}} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{y}, \tag{6.10}$$

where we have assumed that $\text{rank}(\mathbf{A}) = N$ so that $\mathbf{A}^T\mathbf{A}$ is invertible. We observe that $\hat{\mathbf{x}}_{\text{LS}}$ does not make the gradient vanish, because the projection $\mathbf{P} = \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T$ is not an identity matrix for $M > N$. However, due to the randomness of the imaging matrix $\mathbf{A}$, the projection $\mathbf{P}$ is approximately the identity, and the least-squares method is thus approximately maximum-likelihood. Moreover, it serves as a popular inversion method due to its low computational complexity and its asymptotic efficiency as $M \to \infty$ (see Appendix G). For example, the often-used correlation estimator, also known as the ghost imaging estimator, that employs overcomplete photon-count observations in practical single-pixel imaging systems is a variation of the least-squares estimator [105].

In analyzing the performance of the least-squares imager, we compute its MSE: $\mathsf{mse}(\mathbf{x}, \hat{\mathbf{x}}_{\text{LS}}) = \mathbb{E}[\|\mathbf{x} - \hat{\mathbf{x}}_{\text{LS}}\|_2^2]$. Monte Carlo simulations of the expectation term can be used to obtain an approximation for the MSE. However, the quality of the Monte Carlo approximation is dependent on the number of trials employed, and to compute highly accurate MSE values, long simulation runs are required. Thus, we aim to derive a closed-form approximation to the MSE that allows direct computation.

Approximating the MSE: Our first step is to express the MSE in a form that uses the imaging variables $\mathbf{A}$ (random) and $\mathbf{x}$ (non-random). The Poisson observation vector $\mathbf{y}$ has conditional mean and variance both equal to $\mathbf{A}\mathbf{x}$. Thus, it is possible to write $\mathbf{y} = \mathbf{A}\mathbf{x} + \boldsymbol{\eta}$, where $\boldsymbol{\eta}$ is a zero-mean random vector. Because the measurements $\mathbf{y}_1, \ldots, \mathbf{y}_M$ are independent, the covariance matrix of $\boldsymbol{\eta}$ is diagonal with $\mathbf{A}\mathbf{x}$ as the $M \times 1$ vector defining that diagonal. The error of the least-squares estimator is

$$\mathbf{x} - \hat{\mathbf{x}}_{\text{LS}} = \mathbf{x} - (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T(\mathbf{A}\mathbf{x} + \boldsymbol{\eta}) = -(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\boldsymbol{\eta}, \tag{6.11}$$

so the MSE can be written as

$$\begin{aligned}
\mathsf{mse} &= \mathbb{E}\Big[\left((\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\boldsymbol{\eta}\right)^T (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\boldsymbol{\eta}\Big] && (6.12)\\
&= \mathbb{E}\big[\boldsymbol{\eta}^T\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\boldsymbol{\eta}\big] && (6.13)\\
&\overset{(a)}{=} \mathbb{E}\big[\mathrm{Tr}\left(\boldsymbol{\eta}^T\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-2}\mathbf{A}^T\boldsymbol{\eta}\right)\big] && (6.14)\\
&\overset{(b)}{=} \mathbb{E}\big[\mathrm{Tr}\left(\boldsymbol{\eta}\boldsymbol{\eta}^T\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-2}\mathbf{A}^T\right)\big] && (6.15)\\
&\overset{(c)}{=} \mathrm{Tr}\left(\mathbb{E}\big[\boldsymbol{\eta}\boldsymbol{\eta}^T\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-2}\mathbf{A}^T\big]\right), && (6.16)
\end{aligned}$$

where in (a) the trace of a scalar is the same scalar; in (b) the trace of a product is invariant to cyclic permutation of its factors; and in (c) the trace and expectation operators commute.

Recall that both the SLM pattern $\mathbf{A}$ and the photodetection process are sources of randomness. Denote the MSE conditioned on the choice of $\mathbf{A}$ by cmse. Then

$$\begin{aligned}\mathsf{cmse} &= \operatorname{Tr}\left(\mathbb{E}\big[\boldsymbol{\eta}\boldsymbol{\eta}^T \mid \mathbf{A}\big]\, \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-2}\mathbf{A}^T\right) && (6.17)\\ &= \operatorname{Tr}\left(\operatorname{diag}(\mathbf{A}\mathbf{x})\, \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-2}\mathbf{A}^T\right). && (6.18)\end{aligned}$$

Exact computation of the expectation of cmse over $\mathbf{A}$ is complicated by the lack of independence of $\operatorname{diag}(\mathbf{A}\mathbf{x})$ and $\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-2}\mathbf{A}^T$. We approximate these as independent, giving

$$\mathbb{E}\big[\boldsymbol{\eta}\boldsymbol{\eta}^T\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-2}\mathbf{A}^T\big] \approx \mathbb{E}\big[\boldsymbol{\eta}\boldsymbol{\eta}^T\big]\,\mathbb{E}\big[\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-2}\mathbf{A}^T\big]. \tag{6.19}$$

This approximation can be justified for large N, because $\operatorname{diag}(\mathbf{A}\mathbf{x})$ approaches a scaled identity matrix as N grows without bound. By iterated expectation with conditioning on $\mathbf{A}$,

$$\mathbb{E}\big[\boldsymbol{\eta}\boldsymbol{\eta}^T\big] = \mathbb{E}_{\mathbf{A}}\big[\mathbb{E}\big[\boldsymbol{\eta}\boldsymbol{\eta}^T \mid \mathbf{A}\big]\big] = \mathbb{E}_{\mathbf{A}}[\operatorname{diag}(\mathbf{A}\mathbf{x})] = p\|\mathbf{x}\|_1\,\mathbf{I}_M, \tag{6.20}$$

where $\mathbf{I}_M$ is the $M \times M$ identity matrix and we have used the Bernoulli distribution of $\mathbf{A}$'s entries. We can now write

$$\begin{aligned}\mathsf{mse} &= \operatorname{Tr}\left(\mathbb{E}\big[\boldsymbol{\eta}\boldsymbol{\eta}^T\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-2}\mathbf{A}^T\big]\right) && (6.21)\\ &\overset{(a)}{\approx} \operatorname{Tr}\left(\mathbb{E}\big[\boldsymbol{\eta}\boldsymbol{\eta}^T\big]\,\mathbb{E}\big[\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-2}\mathbf{A}^T\big]\right) && (6.22)\\ &\overset{(b)}{=} p\|\mathbf{x}\|_1 \operatorname{Tr}\left(\mathbb{E}\big[\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-2}\mathbf{A}^T\big]\right), && (6.23)\\ &\overset{(c)}{=} p\|\mathbf{x}\|_1 \operatorname{Tr}\left(\mathbb{E}\big[(\mathbf{A}^T\mathbf{A})^{-1}\big]\right), && (6.24)\end{aligned}$$

where (a) uses the independence approximation (6.19); (b) uses (6.20); and (c) uses both the invariance of trace of a product to cycle permutation of factors and the

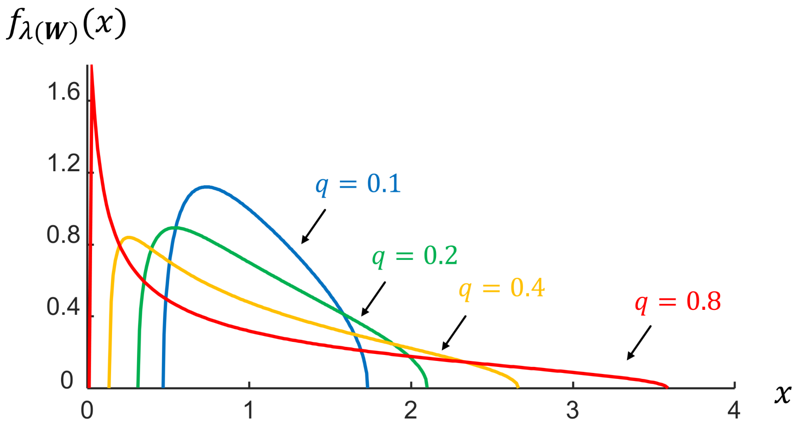


Figure 6-2: Plots of EDFs $f_{\lambda(\mathbf{W})}(x)$ for different q values according to the Marchenko–Pastur law.

commuting of expectation and trace.

So far, we have approximated the MSE of the least-squares estimator as a product of imaging parameters and the trace-expectation of the inverse of a scalar multiple of the *correlation matrix* $\mathbf{C} = \mathbf{A}^T\mathbf{A}/M$. Since the trace of a matrix is the sum of its eigenvalues, we aim to understand the properties of inverse correlation matrix eigenvalues $\lambda(\mathbf{C}^{-1})$ to compute the trace-expectation factor.

RMT characterizes the eigenvalue density function (EDF) of a random matrix with known parameterization. The celebrated Marchenko–Pastur law that derives the EDF of Wishart matrices for $M \geq N$ is given as follows [106]:

(Marchenko–Pastur Law) Let $\mathbf{X}$ be an $M \times N$ random matrix with each entry being an i.i.d. random variable with mean 0 and variance σ^2. Letting $M, N \to \infty$, while keeping $q = N/M \in (0, 1]$ fixed, the EDF $f_{\lambda(\mathbf{W})}$ of the Wishart matrix $\mathbf{W} = \mathbf{X}^T\mathbf{X}/M$ is given by

$$f_{\lambda(\mathbf{W})}(x) = \frac{1}{2\pi\sigma^2}\frac{\sqrt{(q_+ - x)(x - q_-)}}{qx}, \tag{6.25}$$

where $q_\pm = \sigma^2(1 \pm \sqrt{q})^2$ and $x \in [q_-, q_+]$.

Figure 6-2 plots the EDF expression stated by Marchenko–Pastur for different values of $q \in (0, 1]$. If we denote the kth-order moment of the Wishart eigenvalues by

$$\mu_{\lambda(\mathbf{W})}(k; M, N) = \int_{\mathrm{supp}(f)} x^k f_{\lambda(\mathbf{W})}(x)\, dx, \tag{6.26}$$

then the trace-expectation of the kth power of $\mathbf{W}$ can be expressed as the following [96]:

$$\mathrm{Tr}\left(\mathbb{E}\big[\mathbf{W}^k\big]\right) = N\mu_{\lambda(\mathbf{W})}(k; M, N). \tag{6.27}$$

The Marchenko–Pastur law is not directly applicable to simplification of (6.24) because the entries of $\mathbf{A}$ do not have zero means. Thus, we study variations for computing the moments when the correlation matrix is generated using non-negative random vectors.

Let $\mathbf{A}' = \mathbf{A} - \mathbf{D}$, where $\mathbf{D}_{i,j} = p$ for all i, j, so that $\mathbf{A}'$ has zero-mean random entries, and let $\mathbf{C}' = (\mathbf{A}')^T\mathbf{A}'/M$ be our Wishart matrix. Then, because the single nonzero eigenvalue of $\mathbf{D}^T\mathbf{D}$, which is $\lambda(\mathbf{D}^T\mathbf{D}) = MNp^2$, is much larger than $\lambda_{\max}(\mathbf{C}')$, we can approximate the EDF of $\mathbf{C}$ for our *spiked model* [107] as the following:

$$f_{\lambda(\mathbf{C})}(x) \approx \frac{N-1}{N} f_{\lambda(\mathbf{C}')}(x) + \frac{1}{N}\delta(x - MNp^2), \tag{6.28}$$

where $\delta(\cdot)$ is the Dirac delta function. Its kth-order moment can then be written as

$$\mu_{\lambda(\mathbf{C})}(k; M, N) \approx \frac{N-1}{N}\mu_{\lambda(\mathbf{C}')}(k; M, N) + \frac{1}{N}\left(MNp^2\right)^k, \tag{6.29}$$

by the linearity of integrals.

To compute our trace-expectation factor, we are interested in the first negative moment: $k = -1$. Since MNp^2 is significantly larger than q_+, its inverse is very small and negligible. Thus, we can relate the first negative moment of the Marchenko–Pastur distribution to our spiked distribution as

$$\mu_{\lambda(\mathbf{C})}(-1; M, N) \approx \frac{N-1}{N}\mu_{\lambda(\mathbf{C}')}(-1; M, N). \tag{6.30}$$

Combining (6.24), (6.27), and (6.30), we obtain an analytical expression that approximates mse in the constant aspect ratio regime (constant q):

$$\mathsf{mse\text{-}rmt} = \frac{p(N-1)}{M}\|\mathbf{x}\|_1\,\mu_{\lambda(\mathbf{C}')}(-1;M,N). \tag{6.31}$$

Finally, using the Wishart matrix moment computation result that $\mu_{\lambda(\mathbf{C}')}(-1;M,N) = M/(\sigma^2(M-N))$ [96, 108], we have

$$\mathsf{mse\text{-}rmt} = \frac{(N-1)\|\mathbf{x}\|_1}{(1-p)(M-N)}. \tag{6.32}$$

6.5 Results

A well-known baseline approximation method for the MSE is asymptotic in the number of observations while fixing the signal dimension ($q \to 0$ by $M \to \infty$ while N is constant) [109]. Specifically,

$$\begin{aligned}\frac{\mathbf{A}^T\mathbf{A}}{M} &\approx \lim_{M\to\infty}\frac{\sum_{k=1}^{M}\mathbf{a}_k\mathbf{a}_k^T}{M} = \mathbb{E}\left[\mathbf{a}\mathbf{a}^T\right]\\ &= p(1-p)\mathbf{I}_N + p^2\mathbf{1}_{N\times N},\end{aligned} \tag{6.33}$$

where $\mathbf{1}_{N\times N}$ is the $N\times N$ matrix of ones and we have used the mean and variance of a Bernoulli random variable. Since the inverse of the matrix in (6.33) is

$$\frac{1}{p(1-p)}\mathbf{I}_N - \frac{1}{(1-p)(Np+1-p)}\mathbf{1}_{N\times N}, \tag{6.34}$$

(6.24) is approximated with

$$\begin{aligned}\mathsf{mse\text{-}baseline} &= p\|\mathbf{x}\|_1\,\mathrm{Tr}\left(\frac{\left(\mathbb{E}\left[\mathbf{a}\mathbf{a}^T\right]\right)^{-1}}{M}\right)\\ &= \frac{(N-1)p+1-p}{Np+1-p}\,\frac{N\|\mathbf{x}\|_1}{(1-p)M}.\end{aligned} \tag{6.35}$$

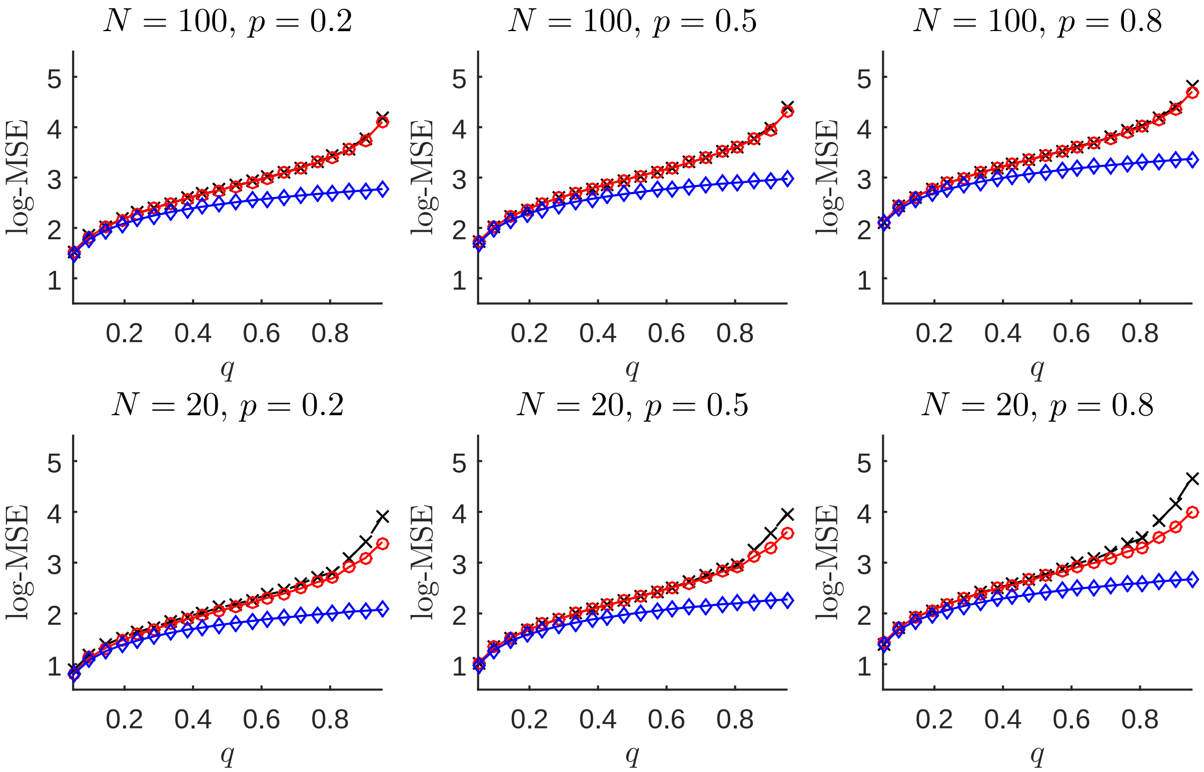


Figure 6-3: Plots of logarithmic mse (dashed black with 'x' markers), mse-rmt (red with 'o' markers), and mse-baseline (blue with diamond markers) for different values of the Bernoulli probability $p \in \{0.2, 0.5, 0.8\}$ and signal dimension $N \in \{20, 100\}$. Each log-MSE plot is shown over various values of q, the ratio between the number of signal dimensions (N) to the number of observations (M).

The first factor in (6.35) is approximately 1 for large N; the second factor is approximately the same as (6.32) when $q \to 0$.

In Figure 6-3, we study how the conventional closed-form approximation mse-baseline ($q \to 0$) and the proposed closed-form approximation mse-rmt (constant q) compare to mse, which is computed using Monte Carlo simulations of 1000 trials for each imaging parameter set. Each entry of the true image $\mathbf{x}$ is drawn i.i.d. from the uniform distribution on $[0, 10]$. The code that was used to generate our results is available from [110].

For a moderate image dimension ($N = 100$), we observe that for different values of the Bernoulli probability p, mse-rmt is an almost exact approximation of mse, while the baseline asymptotic model mse-baseline lower-bounds mse with a non-trivial gap

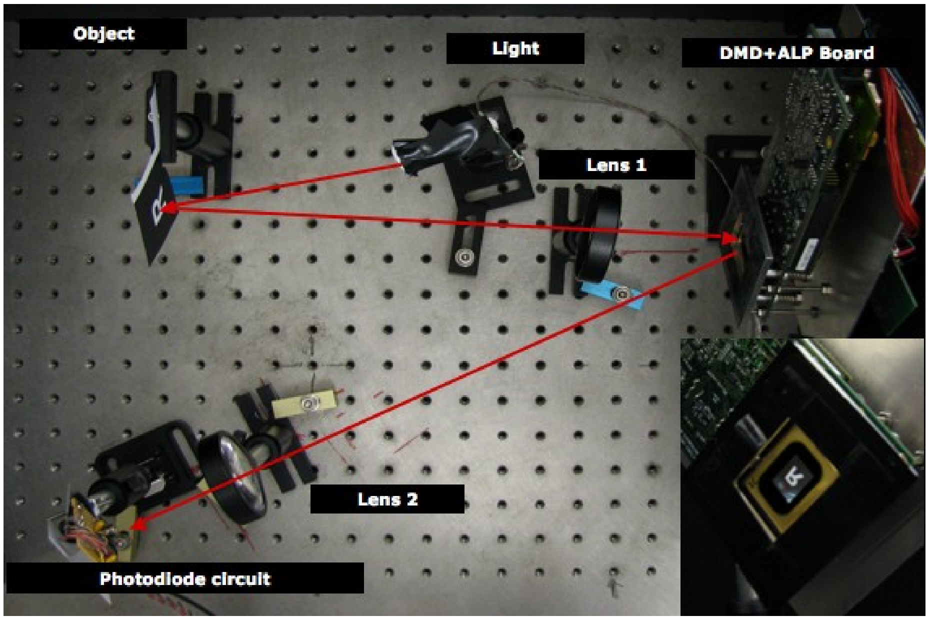


Figure 6-4: Rice single-pixel camera setup. This photograph is taken from the Rice single-pixel camera project website: `http://dsp.rice.edu/cscamera`.

for q values that are not very small ($q > 0.15$).

For the smaller image dimension ($N = 20$) and for high q values, we see that our RMT-based approximation, although it is a better approximation to mse than mse-baseline, deviates from true MSE as well, since our approximations in (6.19) were based on the assumption of the problem size being large.

Results using experimental Rice single-pixel data: We used the public Rice single-pixel camera dataset [111] to experimentally validate our MSE estimation performance. The imaging setup is shown in Figure 6-4. The Rice single-pixel camera uses a classical photodiode instead of a photon-counting detector, and thus the raw measurements $\mathbf{y}'$ are not exactly Poisson distributed:

$$\mathbf{y} \sim \text{Poisson}(\mathbf{A}\mathbf{x}), \tag{6.36}$$

$$\mathbf{y}' = \text{ADC}(\mathbf{y} + \mathbf{e}), \tag{6.37}$$

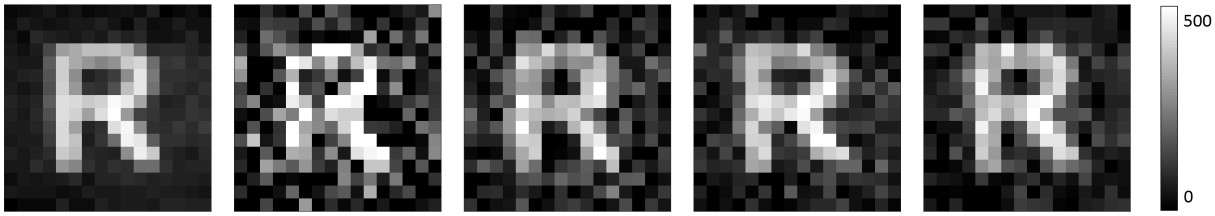
where $\text{ADC}(\cdot)$ describes the analog-to-digital electrical conversion that scales and quantizes the raw photodetection measurements and $\mathbf{e}$ is used to model the additional signal-independent sensor noise vector with zero-mean.

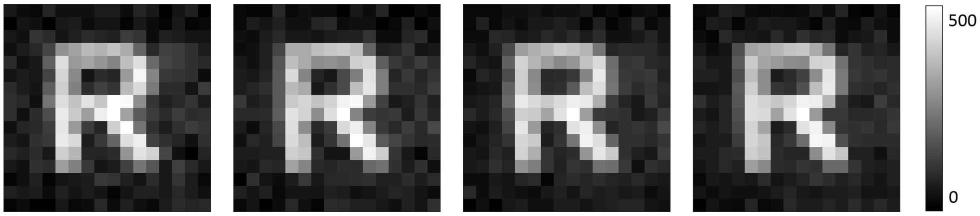
In order to apply our Poisson MSE results derived earlier, we scale the photocurrent observations $\mathbf{y}'$ by the ADC quantization level to obtain $\hat{\mathbf{y}}$, which is an approximation to the photon count data $\mathbf{y}$. Figure 6-5 shows the result of recovering the 16×16 image of letter 'R' by the least-squares solution $\hat{\mathbf{x}}_{\mathrm{LS}}(\hat{\mathbf{y}})$ for different M values. As expected, the quality of imaging improves when larger M values are used. We used the least-squares estimate of $M = 4290$ as ground truth in computing the experimental MSE values. Figure 6-6 compares the experimental MSE (mse) with the estimates of the baseline (mse-baseline) and the proposed (mse-rmt). We observe that even when using $\hat{\mathbf{y}}$, which is an estimate of the photon-count vector from ADC measurements, mse-rmt is a better estimation to mse than mse-baseline. However, due to the model mismatch, we have a bias error in mse-rmt in the experimental result, unlike the simulation results in Figure 6-3.

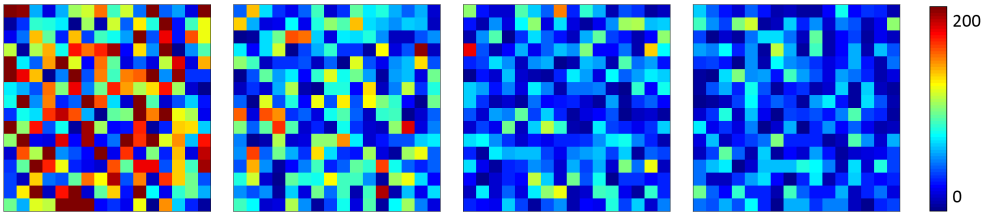
6.6 Summary and Discussion

In this chapter, we derived a remarkably simple closed-form approximation to the MSE of a low-flux least-squares single-pixel imaging system by using moment computation methods from random matrix theory. Unlike conventional Monte Carlo methods (high accuracy, long computational time) and asymptotic approximation techniques (low accuracy, short computational time), our RMT-based MSE approximation achieves both high accuracy and high computational efficiency as shown by numerical experiments.

Analysis for Optical Design: Because our result in Eq. (6.32) gives a good approximation to the MSE in closed-form, it can be used to choose optimal acquisition parameters given imaging constraints. If the single-pixel camera is equipped with its own light source, then we can ask the following optical design question: what is the smallest number of observations that we can make and the lowest illumination power that we can use given that we can tolerate up to a certain error in the image? Defining


(a) Truth (b) $M = 300$ (c) $M = 400$ (d) $M = 500$ (e) $M = 600$


(f) $M = 1200$ (g) $M = 1800$ (h) $M = 2400$ (i) $M = 3000$


(j) error of (b) (k) error of (c) (l) error of (d) (m) error of (e)

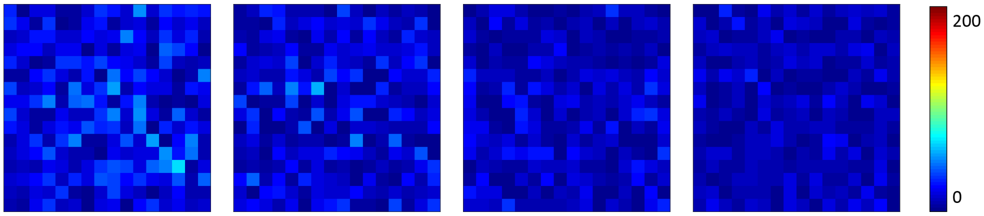

(n) error of (f) (o) error of (g) (p) error of (h) (q) error of (i)

Figure 6-5: Least-squares single-pixel imaging estimates of letter 'R' for increasing values of M using the Rice single-pixel dataset. Here, $N = 16 \times 16 = 256$ and the ground truth image was generated using the least squares solution with $M = 4290$, since it is an asymptotically efficient estimator.

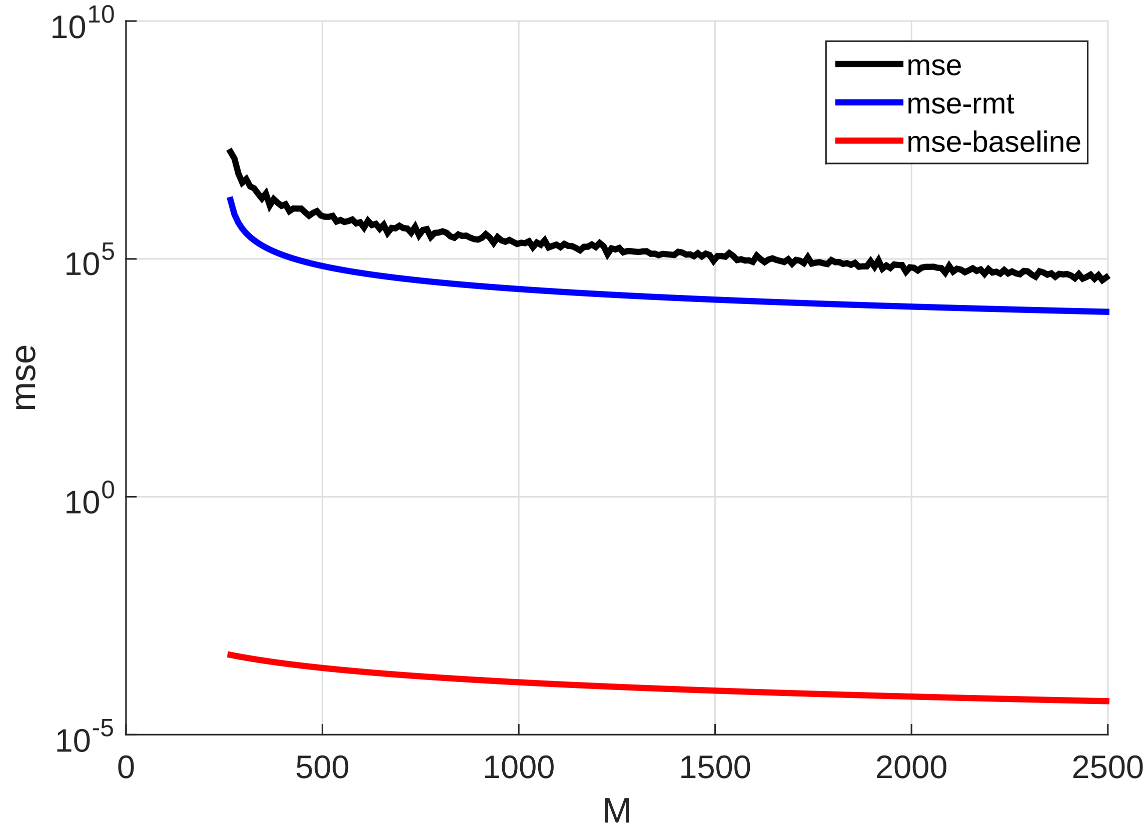


Figure 6-6: Comparison of MSE (plot in log scale) estimates for least-squares imaging method on experimental Rice single-pixel camera data of letter ‘R’. Here $N = 16 \times 16 = 256$.

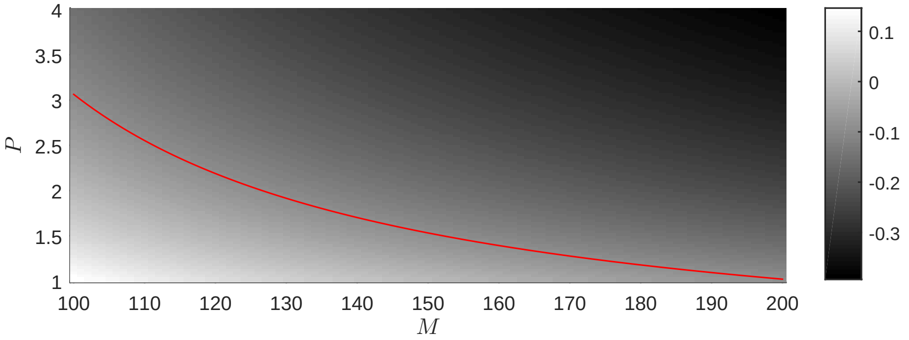


Figure 6-7: Plot of nrmse-rmt over M and P for $N = 50$ and $p = 0.5$. The red line gives the contour for nrmse-rmt $= 0.4$.

normalized root mean square error (NRMSE) as nrmse-rmt $= \sqrt{\text{mse-rmt}}/\|\mathbf{x}\|_1$, our approximation clearly shows the inverse relationship between that normalized error and the number of observations (M) and the amount of optical flux incident at the SLM device ($P = \|\mathbf{x}\|_1$). For example, suppose we have the constraint that we can tolerate up to 1 normalized standard deviation (NRMSE) for image analysis purposes. Then, we can use our NRMSE approximation to determine the optimal acquisition parameters that simultaneously minimizes the number of observations and flux:

$$\left\{(M, P) : \sqrt{\frac{N-1}{(1-p)(M-N)P}} = 1,\ M \in \{N+1, N+2, \ldots\},\ P > 0\right\}, \tag{6.38}$$

for fixed p and N. For example, in Figure 6-7, we illustrate the trade-off between amount of optical flux and number of measurements to achieve nrmse $= 0.4$, given a single-pixel imaging setup with $N = 50$ and $p = 0.5$.

Equation (6.32) also tells us that there is a monotonic decrease in the MSE as p gets smaller. This is intuitive as smaller p implies smaller condition number for $\mathbf{A}$, as its largest singular value is quadratic in p while the other $N-1$ singular values are clustered around 1 from the EDF expression in (6.28). However, Eq. (6.32) assumes that $\hat{\mathbf{x}}_{\text{LS}}$ is well-defined. In other words, what it fails to capture is that small p can lead to rank$(\mathbf{A}) < N$ with high probability. Thus, we cannot conclude that lowering p for MSE improvements is always the solution, because $\mathbf{A}^T\mathbf{A}$ can be non-invertible such that no least-squares solution is defined.

It is of future interest to see how the proposed performance-analysis framework can be extended to other estimators for single-pixel imaging. For example, the Tikhonov estimator regularizes the least-squares estimate by perturbing the correlation matrix with a positive definite matrix $\mathbf{T}$ [112],

$$\hat{\mathbf{x}}_{\text{tik}} = (\mathbf{A}^T\mathbf{A} + \mathbf{T})^{-1}\mathbf{A}^T\mathbf{y}. \tag{6.39}$$

The Tikhonov approach is useful for modeling image priors that enforce smoothness in image estimates. A perturbation analysis, for example, on the correlation matrix may lead to an accurate MSE approximation for the Tikhonov estimator.

Chapter 7

Fluorescence Imaging

7.1 Overview of Problem

The ability to acquire fluorescence lifetime data from a fluorophore is useful for molecular studies involving pH distributions [113] and energy transfer rates [114]. The technique of fluorescence lifetime imaging microscopy (FLIM) aims to recover the lifetimes of fluorophores by exciting them with short optical pulses, measuring the emitted exponentially-decaying light, and using algorithms to infer their lifetimes.

Prior art: There are two main approaches to FLIM: a time-domain technique [115, 116] and a frequency-domain technique [117, 118]. The time-domain technique perturbs the fluorescence sample with an ultrafast pulse, and aims to directly measure the lifetime by using exponential-fitting methods on the fluorescent signal measured by a time-resolved photodiode. The time-tagging accuracy typically is in picoseconds. The frequency domain technique, on the other hand, employs a pulsed source and a time-gated-and-intensified CCD camera setup. It uses homodyne detection techniques (modulating the received signal with a cosine or a rectangular pulse function) such that the sample lifetime can be inferred from the phase shift of the modulated signal. Although the frequency-domain technique has its advantages of being computationally cheap, it is limited to high light-level operation as the demodulation step requires almost noiseless signal for high accuracy reconstruction. In this chapter, we are interested in developing a time-domain FLIM method for extremely low-flux imaging purposes ($\sim$ 1 detected photon per pixel on average), and thus focus on developing a fluorescence imaging framework that achieves high photon-efficiency by employing a single-photon pulsed imaging setup.

A typical low-flux time-domain FLIM method aims to reconstruct the spatially-resolved intensity and lifetime profiles of a fluorophore sample using raster-scanning pulsed illumination source and a single-photon detector [64]. Due to the Poissonian nature of photon counting, the conventional FLIM method requires a long acquisition time to detect many photons (typically up to thousands per pixel) in order to construct a histogram modeling an almost noiseless exponential function for accurate lifetime estimation [46]. Traditionally, pixelwise maximum-likelihood processing and fast decay-rate fitting methods have been employed for lifetime estimation in the time domain [119]. It has been demonstrated that the conventional pixelwise lifetime estimation method leads to accurate results only when the number of photon detections is large enough that amount of Poisson noise in the histogram is negligible. For example, at least 200 detections are required per pixel in order to recover the lifetime with 10% accuracy, when the instrumental response function can be well-approximated with a Dirac delta function. [120].

In order to use time-domain FLIM for accurate imaging with fewer photon detections per pixel, methods of image denoising can be applied to the noisy pixelwise intensity and lifetime image estimates to improve their signal-to-noise ratios [121]. Combining Markov random field based image denoising methods and range-gating techniques, it has been shown that the number of photon counts per pixel can be reduced to $\sim$100 for accurate lifetime reconstruction [122].

Summary of our approach: In this chapter, we propose a method for accurately recovering fluorescence intensity and lifetime images using only a small number of photon detections at each pixel. Unlike existing methods, our proposed framework combines the probabilistic model of single-photon detection from the fluorescence signal with spatial correlations existing in fluorescent scenes. Experimental and simulation results show that our method recovers fluorescence intensity and nanosecond lifetime images accurately using considerably fewer photon detections than what existing pixelwise recovery methods, such as pixelwise maximum-likelihood processing, require.

In particular, we experimentally demonstrate that our proposed framework allows accurate imaging of fluorescent markers using only *one* photon detection on average at each image pixel, due to the following novelties:

1. We construct an accurate pixelwise probabilistic model for the detection times of individual photons from the exponential fluorescence signal.

2. We develop an optimization framework that combines the accurate photodetection model with an image prior that models the spatial correlations of natural fluorescent scenes. Using a transformation of variables, we show that this optimization problem can be convex, which allows it to provide fast and accurate solutions for high-dimensional images.

7.2 Single-Photon Imaging Setup

The low-light fluorescence imaging setup includes a raster-scanning pulsed light source, a single-photon detector, and a time-correlator (see Figure 7-1). The laser source spot-illuminates a pixel of the sample with a total of N_s pulses with repetition period of T_r. For each trial of pulse illumination, the sample is excited so that it goes through a process of fluorescence, and the exponentially-decaying fluorescent signal is emitted towards the single-photon detector. Due to the low-flux condition, the detector with high probability detects zero photons and with low probability detects a single photon per illumination trial. Each photon detection time, relative to the most recent pulse-illumination time, is recorded using a time-correlator. Repeating the process of illumination, detection, and time-correlation over a total of $N_x \times N_y$ image pixels, we have the raw photon arrival dataset that can be used to estimate the intensity and lifetime images of the fluorescent scene.

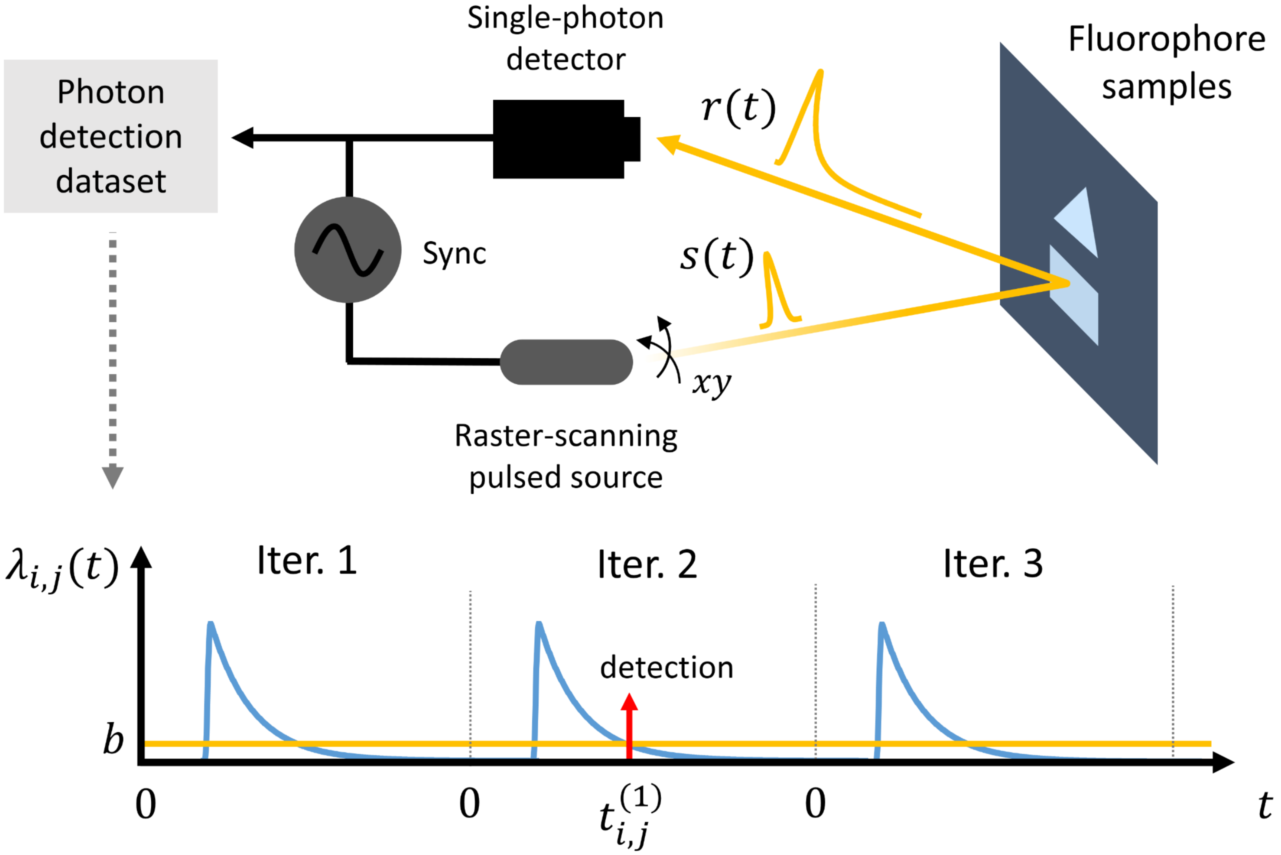

Single-photon
detector
Fluorophore
samples
Photon
detection
dataset
$r(t)$
Sync
$s(t)$
xy
Raster-scanning
pulsed source
$\lambda_{i,j}(t)$
Iter. 1
Iter. 2
Iter. 3
detection
b
0
0
$t_{i,j}^{(1)}$
0
t

Figure 7-1: Single-photon fluorescence imaging setup as described in Section 7-1 (Top) and an illustration of the photodetection process (Bottom). In this illustration, we have one photon detection (marked in red) resulting from the (i, j)-th pixel's second pulse excitation of the fluorophore sample, given that $N_s = 3$. Note that we define the start of the fluorescence process as time $t = 2d_{i,j}/c$, which is defined by the calibrated sample-to-imager distance. This distance offset in defining the decay signal does not affect the lifetime measurements, as exponential processes are memoryless.

7.3 Forward Imaging Model

For the sake of exposition, we focus on a single pixel for developing the statistics of photodetection from fluorescence signals. Assume that a single optical pulse $s(t)$ illuminates the given pixel in the scene at time $t = 0$. Let a and β be the unknown fluorescence amplitude and lifetime values at that pixel. Then, the optical waveform incident at the single-photon detector is

$$r(t) = \left[s(t) * a \exp\left\{ - \left(\frac{t - 2d/c}{\beta} \right) \right\} \mathbb{1}_{t \geq 2d/c} \right] + b, \qquad \text{for } t \in [0, T_r), \tag{7.1}$$

where d is the distance from imager to fluorophore, c is the speed of light, $\mathbb{1}$ is the indicator function, b is the photon count rate of residual background light, and $*$ is the one-dimensional convolution operator over the time variable t. Recall that the rate function determining the photon counts after a single pulse illumination is

$$\lambda(t) = \eta \left[(r * I_d)(t) \right] + b_d, \qquad \text{for } t \in [0, T_r), \tag{7.2}$$

where η is the detector efficiency, $I_d(t)$ is the detector response, and b_d is the dark count rate of the single-photon detector. Assuming the practical imaging scenario in which the RMS duration of $s(t) * I_d(t)$ (typically ~ps) is much smaller than the fluorescent lifetime β (typically ~ns) [123], we use $s(t) * I_d(t) = \delta(t)$ throughout this chapter. Also, let the distance d from the imager to the fluorophore be known through calibration. Because exponential processes are memoryless, we can express Eq. (7.2) using an exponential function with a time support that corrects $[2d/c, T_r)$, such that the start of the fluorescence process is marked by time $t = 0$:

$$\lambda'(t) = \eta \, a \exp\{-t/\beta\} + \tilde{b} \qquad \text{for } t \in [0, T_r'), \tag{7.3}$$

where $T_r' = T_r - 2d/c$ and $\tilde{b} = \eta \, b + b_d$. In our setup, we emphasize that T_r' is always set large enough, relative to β, such that the truncation of the received fluorescence signal is avoided and the signal acquisition trials are statistically independent of each other. Figure 7-2 shows an example of $N_s \lambda'(t)$ obtained by experimental photon-count measurements, with N_s being 100000 (collected by collaborator F. Xu).

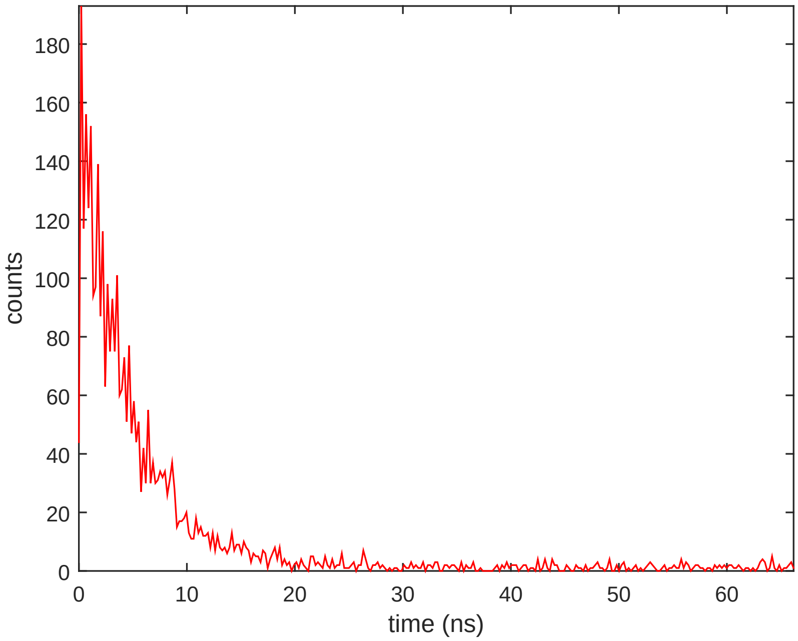


Figure 7-2: Example large photon count histogram of $\lambda'(t)$ from a pixel of a quantum dot sample with lifetime of ∼7 ns, used to visualize the properties of fluorescence signals. The total number of detections contains information about fluorescence intensity and the rate of decay of the histogram contains information about fluorescence lifetime. The residual photon counts uniformly distributed over $[0, 64)$ ns are the extraneous background and dark counts.

We assume the low-flux condition in which the imager is operating at low-light levels such that the mean number of photons of the received fluorescence signal is much less than 1 (see Appendix B). Then, after N_s pulse illuminations, the total number of photon detections C is distributed as

$$C \sim \text{Poisson}\left(N_s \int_0^{T_r'} \lambda'(t)\, dt \right) = \text{Poisson}\left(N_s x + N_s \tilde{b} T_r' \right), \qquad (7.4)$$

where x is the fluorescence intensity that we are interested in recovering. Thus C has the discrete probability mass function

$$f(C; x) = \frac{\exp\{-(N_s x + N_s \tilde{b} T_r')\}(N_s x + N_s \tilde{b} T_r')^C}{C\,!}, \qquad \text{for } C = 0, 1, 2, \ldots \qquad (7.5)$$

Using on the low-flux condition from Appendix B, we can also derive the continuous probability density function of the detection time for an individual photon, as the normalized rate function:

$$p_T(t;\beta) = \frac{\lambda'(t)}{\int_0^{T_r'} \lambda'(t)\,dt} \tag{7.6}$$

$$= \frac{N_s\,\eta\,a\exp\{-t/\beta\} + N_s\,\tilde{b}}{N_s\,\eta\,a\,\int_0^{T_r'}\exp\{-t/\beta\}\,dt + N_s\,\tilde{b}\,T_r'} = \frac{N_s\,\eta\,a\exp\{-t/\beta\} + N_s\,\tilde{b}}{N_s\,\eta\,a\,\beta + N_s\,\tilde{b}\,T_r'} \tag{7.7}$$

for $t \in [0, T_r')$.

7.4 Solving the Inverse Problem

Pixelwise maximum-likelihood estimation: Assuming $\tilde{b} = 0$, using the probability distributions in Eq. (7.5) and Eq. (7.7), we can derive very simple closed-form expressions for the pixelwise maximum-likelihood intensity and lifetime estimates given C photon detections from N_s illuminations:

$$\hat{x}_{\mathrm{ML}} = \underset{x \geq 0}{\arg\max} \quad \log\ f(C;x), \tag{7.8}$$

$$\hat{\beta}_{\mathrm{ML}} = \underset{\beta \geq 0}{\arg\max} \quad \log\ p_T(\{t^{(k)}\}_{k=1}^{C};\beta). \tag{7.9}$$

The lifetime estimate assumes non-zero detections. When using this pixelwise method, we can always label pixels with zero detections as having unknown lifetimes or choose to randomly guess their lifetimes.

For the pixelwise ML intensity estimate, we have

$$\underset{x \geq 0}{\arg\max} \quad \log f(C;x) = \underset{x \geq 0}{\arg\max} \quad \left[C\log(N_s x) - N_s x - \log C!\right]. \tag{7.10}$$

Since $\log f(C;x)$ is a sum of a log, linear, and a constant function in x, $\log f(C;x)$ is convex and the maximizer can be found by simply checking the zero-gradient condition with the non-negativity constraint, which is also a convex set. We find that

$$\frac{d}{dx}\left[C\log(N_s x) - N_s x - \log C!\right] = C/N_s - N_s = 0, \text{ given } x \geq 0 \tag{7.11}$$

$$\Rightarrow \quad \hat{x}_{\text{ML}} = \frac{C}{N_s}, \tag{7.12}$$

which is the scaled photon count. For the pixelwise ML lifetime estimate, we can perform a similar trick by checking the constrained zero-gradient condition and the second-gradient of $\log p_T(\{t^{(k)}\}_{k=1}^{C};\beta)$ around the unique zero-gradient solution is negative, implying that it is also the unique maximizer. With some algebra, we have

$$\hat{\beta}_{\text{ML}} = \frac{1}{C}\sum_{k=1}^{C} t^{(k)}, \tag{7.13}$$

which is simply the mean of photon detection times.

Pixelwise estimators are practical due to their simple closed-form expressions, and their asymptotic efficiency. However, they require a large number of photon detections per pixel to form accurate images. Below we propose a framework that accurately recovers spatially-resolved fluorescence intensity and lifetime images ($\mathbf{X}, \boldsymbol{\beta} \in \mathbb{R}_{+}^{N_x \times N_y}$). using a small number of photon detections, by combining pixelwise single-photon statistics with spatial correlations.

- **Step 1 – Estimation of fluorescent intensity:**

  Let $\mathbf{C}$ be the $N_x \times N_y$ matrix whose (i,j)th entry $\mathbf{C}_{i,j}$ is the observed number of photon detections at pixel (i,j). Then, we solve for the penalized minimum negative log-likelihood fluorescence intensity solution as follows:

$$\hat{\mathbf{X}}_{\text{CML}} = \underset{\mathbf{X}:\mathbf{X}_{i,j}\geq 0}{\arg\min} \left[\sum_{i=1}^{N_x}\sum_{j=1}^{N_y} -\log f(\mathbf{C}_{i,j};\mathbf{X}_{i,j},\tilde{b})\right] + \tau_x \,\text{pen}_x(\mathbf{X}), \tag{7.14}$$

  where $\text{pen}_x(\mathbf{X})$ penalizes the high variations of the intensity solution to enforce spatial smoothness. Here we use $\tau_x > 0$ to control the degree of penalization. Because the negative Poisson log-likelihood is convex in the intensity variable, as shown in the previous section, as long as the model prior $\text{pen}_x(\mathbf{X})$ is a convex function (e.g., the total-variation (TV) norm as used in Chapter 5 or the ℓ_1-norm of the image's wavelet transform, etc.), Eq. (7.14) is a convex optimization problem and thus easy to solve using simple projected gradient descent methods.

- **Step 2 – Detection of fluorophores:**

  Because not all pixels contain fluorophores, before estimating the lifetime image, we employ a thresholding technique to obtain $\hat{\mathbf{M}}$, which is the $N_x \times N_y$ binary image indicating the presence of fluorophore at each pixel, using the accurate intensity image recovered from Step 1.

  $$\hat{\mathbf{M}}_{i,j} = \begin{cases} 1, & \text{if } (\hat{\mathbf{X}}_{\text{CML}})_{i,j} > \epsilon \text{ AND } \mathbf{C}_{i,j} > 0 \\ 0, & \text{otherwise} \end{cases} \qquad \text{for all } i, j, \tag{7.15}$$

  where ϵ is a small number of choice (e.g. $\epsilon = \max_{i,j}(\hat{\mathbf{X}}_{\text{CML}})_{i,j}/1000$). By using data only at pixels with $\hat{\mathbf{M}}_{i,j} = 1$, we are effectively filtering out extraneous background and dark counts that do not contain information about the fluorophore lifetime.

- **Step 3 – Estimation of fluorescent lifetime:**

  A direct application of the penalized minimum log-likelihood estimation as we did in Step 1 for lifetime estimation will lead to the following optimization problem:

  $$\underset{\boldsymbol{\beta}:\boldsymbol{\beta}_{i,j}\geq 0}{\arg\min} \left[\sum_{(i,j):\hat{\mathbf{M}}_{i,j}=1} \sum_{k=1}^{\mathbf{C}_{i,j}} -\log f_T(t_{i,j}^{(k)}; \boldsymbol{\beta}_{i,j}, \tilde{b} = 0) \right] + \tau_\beta \, \text{pen}_\beta(\boldsymbol{\beta}). \tag{7.16}$$

  Here, observe that we are not using the data from all pixels, but only that from ones defined by the set $\{(i,j) \,|\, \hat{\mathbf{M}}_{i,j} = 1\}$.

  Unfortunately, it can be shown that the negative log-likelihood function of the fluorophore's lifetime at a pixel is a not a globally convex function, and thus greedy algorithms such as gradient descent can get stuck in a local minimum point that does not have any solution-accuracy guarantees. However, if the fluorescence lifetime likelihood is instead parametrized by $\boldsymbol{\mu}_{i,j} = 1/\boldsymbol{\beta}_{i,j}$, which is the inverse lifetime, then the second derivative of the negative log-likelihood function at pixel (i,j) is

$$\frac{d^2}{d\boldsymbol{\mu}_{i,j}^2}\left[-\sum_{k=1}^{\mathbf{C}_{i,j}} \log p_T(t_{i,j}^{(k)}; \boldsymbol{\beta}_{i,j}, \tilde{b} = 0)\right] = \mathbf{C}_{i,j}/\boldsymbol{\mu}_{i,j}^2, \tag{7.17}$$

which is always non-negative, since $\boldsymbol{\mu}_{i,j}$ is non-negative. This implies that the negative log-likelihood function is a convex function in $\boldsymbol{\mu}_{i,j}$. Thus, we employ a heuristic of first accurately obtaining the inverse lifetime by solving a convex optimization, and then applying a variable transformation to get our final lifetime image. The inverse lifetime image is obtained by solving the following optimization problem that combines the photon-detection statistics from an exponential distribution parametrized by inverse lifetime, with a spatial correlation prior:

$$\hat{\boldsymbol{\mu}}_{\text{CML}} = \underset{\boldsymbol{\mu}:\boldsymbol{\mu}_{i,j}\geq 0}{\arg\min} \left[\sum_{(i,j):\hat{\mathbf{M}}_{i,j}=1} \sum_{k=1}^{\mathbf{C}_{i,j}} -\log f_T(t_{i,j}^{(k)}; \boldsymbol{\mu}_{i,j})\right] + \tau_\mu \, \text{pen}_\mu(\boldsymbol{\mu}). \tag{7.18}$$

Observe that if the lifetime image is spatially smooth then the inverse lifetime is also spatially smooth, because the inverse function is continuous for positive values. We can thus directly penalize for the non-smooth inverse lifetime images using $\text{pen}_\mu(\boldsymbol{\mu})$. Similar to (7.14), as long as $\text{pen}_\mu(\boldsymbol{\mu})$ is a convex function in $\boldsymbol{\mu}$, we have that (7.18) is a convex optimization problem, which can be solved efficiently. After solving for $\hat{\boldsymbol{\mu}}_{\text{CML}}$, the lifetime image estimate is obtained using a simple variable transformation: $(\hat{\boldsymbol{\beta}}_{\text{CML}})_{i,j} = 1/(\hat{\boldsymbol{\mu}}_{\text{CML}})_{i,j}$ for all i,j. Also, as a post processing step, pixels of $\hat{\boldsymbol{\beta}}_{\text{CML}}$ that have $\hat{\mathbf{M}}_{i,j} = 0$ are labeled as non-fluorescent.

7.5 Results

Experiments: Using the experimental raster-scanning single-photon imaging setup we used in the first Chapter (Figure 2-4), we performed fluorescence imaging experiments with the scene containing fluorescent markers, instead of room-scale objects for depth and reflectivity imaging. On a non-fluorescent black cardboard, we placed small patches of quantum dot fluorescent samples generated to have various lifetimes. Fluorescence sample preparation was done by H. Utzat and Y. Chen from the Bawendi group at MIT, and more details of the chemical properties of the solution can be found in [123]. In terms of imaging setup, we used the exact same raster-scanning pulsed laser system with a single SPAD detector employed in Chapter 2. The scanning resolution of the laser was set to be 100 × 100, $T_r = 100$ ns, and $T_p = 270$ ps. Refer to Chapter 2 for complete details in the single-photon imaging setup. The experimental data collection was performed by F. Xu.

Figure 7-3 compares the performance of pixelwise ML estimators with that of the proposed framework in recoverying fluorescent intensity and lifetime using a small number of photon detections. The first scene uses a sample with lifetime of ~ 15 ns and the second scene uses two samples with lifetimes of 7 and 23 ns. In the lifetime image, only for visualization purposes, we set pixels with no fluorophores as having a non-existent -1 ns lifetime. For both scenes, although the pixelwise method gives the approximate locations of fluorophores in the image domain, it fails to give their accurate intensity and lifetime values that may be useful for analyzing chemical properties. On the other hand, our intensity and lifetime estimates $\mathbf{X}_{\text{CML}}, \boldsymbol{\beta}_{\text{CML}}$ recovers not only the spatial fluorophore features in a robust manner, but also the intensity and lifetime values accurately. We observe that our fluorophore identification step (Step 2) is accurate enough such that our final lifetime images have finite lifetime values only at the pixels of the fluorophores. In our algorithm, we used the intensity and inverse lifetime penalty functions to be the TV-norm with parameters $\tau_A, \tau_\mu = 0.5$, and $\epsilon = 0.03$.

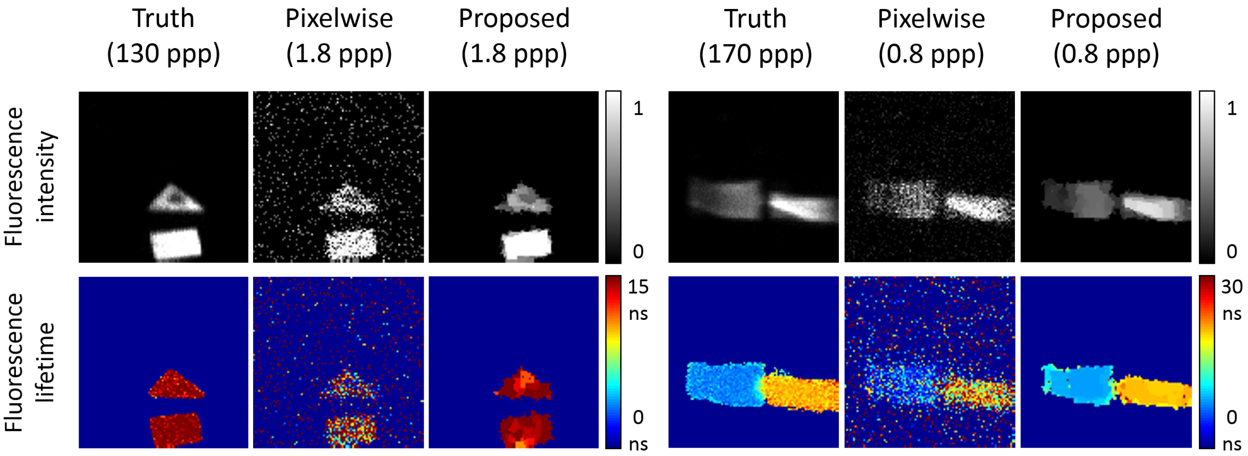


Figure 7-3: Experimental recovery results of (top) fluorescence intensity normalized to have maximum value of 1 and (bottom) lifetime for two different quantum dot scenes with different mean numbers of photons-per-pixel (ppp) using the pixelwise maximum-likelihood estimator and the proposed framework.

7.6 Summary and Discussion

In this chapter, we proposed a method of accurately recovering a scene's fluorescence intensity and lifetime using only a small number of photon detections at each pixel. To solve for intensity and lifetime, we formulated an optimization problem that combined pixelwise photodetection statistics with an image prior that models the spatial correlations of natural fluorescent scenes. Using a transformation-of-variable technique, we showed that this optimization problem can be solved using convex optimization techniques, which allows for fast and accurate computation of high-dimensional images.

Recovering multiple lifetimes: In this chapter, we used the single-lifetime assumption, i.e., that the fluorescence signal is well modeled by a single exponential function at each pixel. However, fluorescent samples such as a cyan fluorescent protein (CFP) are known to have more than one lifetime, and thus emit a light waveform that is a sum of exponential functions at each pixel [9]. Thus, our computational framework must be improved to simultaneously perform non-binary estimation of the number of lifetimes and the recovery of those lifetime values.

Fluorescence imaging through diffusers: Extending the framework to enable accurate imaging in the presence of a diffuser in between the imager and the sample is of high interest, as the diffuser models fluorescence imaging through cell skins. It can be shown that including the diffuser will give an effect of two-way blurring on the observations (the first blurring comes from light going into the diffuser en route to the sample, and the second blurring comes from light coming out of the diffuser to approach the detector). For example, the intensity image observed in the presence of the diffuser, even at very high-flux so that noise is negligible, is given as the spatially-resolved measurements convolved with the two-dimensional kernel of the diffuser that is typically a Gaussian function. The larger the extent of the kernel (width of the Gaussian), the more difficult the inverse problem is to solve. Thus, extending the low-flux super-resolution framework developed in Chapter 5 to the FLIM setup might enable accurate low-light fluorescence imaging through diffusing media.

Chapter 8

Conclusions and Final Remarks

In this thesis, we demonstrated how modeling and computation can play major roles in acquiring high-quality scene information at low light-levels, where instead of measuring continuous light intensities as in the classical high light-level regime, we are constrained by the physical limitation of observing small numbers of photon detections using a single-photon imaging system.

We looked at six different imaging problems in the single-photon regime: single-reflector depth imaging, multi-depth imaging, array imaging, super-resolution imaging, single-pixel photography, and fluorescence imaging. For each low-light imaging problem, we started by deriving an accurate observation model for single-photon detections, characterizing the information the photons carry about the scene of interest. Then, for the scene parameters of interest, we derived their physical constraints, based on what statistics or assumptions we can reasonably make prior to the imaging experiments. Using the imaging model that combines the single-photon statistics and scene parameter constraints, we formulated an inverse problem to recover the scene information in a statistically optimal manner. Finally, we developed fast and accurate algorithms that solve the inverse problem using optimization methods.

In Chapter 2, we showed how discrete-flux modeling of photodetection combined with the union-of-subspaces constaint from single-reflector LIDAR setup lead to a photon-efficient pixelwise depth imaging framework. We then developed a greedy algorithm, inspired by CoSaMP, a ℓ_0-norm-based sparse signal pursuit method, to solve the resulting inverse imaging problem.

Chapter 3 extended the framework from Chapter 2 to imaging multiple reflectors per pixel. We showed that the accurate multidepth imaging framework solves a convex optimization problem that is formulated using photon-count statistics with a sparse-reflector profile constraint. The inverse problem then could be solved using a variation of the iterative shrinkage-thresholding algorithm for LASSO.

Unlike Chapters 2 and 3, which assumed raster-scanning operation with a single detector, Chapter 4 developed a framework that modeled limitations inherent in single-photon cameras with array detectors. There, we showed how combining hardware constraints, photodetection statistics, and image prior models can lead to an accurate array-based depth plus reflectivity imager. Here we were able to solve the inverse problem using a mixture of sparse signal pursuit by OMP and convex optimization methods.

Chapter 5 addressed a particular problem of depth and reflectivity super-resolution. There we showed that a photon-efficient super-resolution technique could be developed by modeling the limitations of optical illumination and sensing and photodetection statistics. The inverse problem could be solved using either a greedy pursuit or convex optimization for deconvolution.

Chapter 6 studied the single-pixel camera architecture, and showed how the accurate photodetection model can be used to reason that the least-squares imager is a useful estimator at low flux as it uses non-compressive measurements. Separately, we derived its closed-form MSE performance at low flux using results from random matrix theory.

Finally, in Chapter 7, we used the photodetection statistics from fluorophore signals and image priors to develop a photon-efficient method of fluorescent imaging. Once again the inverse problem was relaxed so that it could be solved using convex heuristics.

In conclusion, this thesis presented a unifying viewpoint on how accurate modeling and the design of computational reconstruction algorithms can lead to photon-efficient reconstruction of scene properties. Our framework opens up new ways to reconstruct scene properties, such as depth, reflectivity, and fluorescence, in a variety of imaging scenarios while using very few photon detections. Thus, it could find widespread use in engineering applications, spanning from long-range and large-scene remote optical sensing to small and sensitive biological imaging, where the imaging task must be performed fast and accurately using extremely small amounts of light.

Appendix A

Table of Notations

Given below is a list of notations commonly used in this thesis along with their meanings. For illustration purposes, define x to be a real scalar, $\mathbf{x}$ to be a real vector of size $p \times 1$ and $\mathbf{X}$ to be a real matrix of size $p \times q$. Also, we assume that p is an integer divisible by integer p'. Quantities that have physical units have those units indicated in brackets.

I. Notation used for modeling

Notation	**Meaning**
N	number of image pixels
m	number of sampling bins defined by single-photon detector
n	number of discretized depth bins
K	number of reflectors in a scene pixel
Δ	sampling period of time-correlated single-photon detector [s]
N_s	number of independent pulse illuminations
T_r	period of pulsed illumination [s]
T_p	root mean square pulsewidth [s]
$I(t)$	scene impulse response function
$s(t)$	illumination optical flux [photons/s]
$b(t)$	background flux [photons/s]
$r(t)$	optical flux incident at sensor [photons/s]
$\lambda(t)$	rate function of single-photon detector [counts/s]
c	speed of light [m/s]
η	quantum efficiency of single-photon detector
b_d	dark count rate of single-photon detector [counts/s]
$\mathcal{L}(\mathbf{x}; \mathbf{y})$	negative log likelihood function of $\mathbf{x}$ given data $\mathbf{y}$

II. Notation used for inference

Notation	**Meaning**
$\mathbf{x}_i$	value of i-th entry of $\mathbf{x}$
$\mathbf{x}\|_{\mathcal{A}}$	subvector of $\mathbf{x}$ picked out by indices in set $\mathcal{A}$
$\|\mathbf{x}\|_p$	ℓ_p-norm of vector $\mathbf{x}$
$\mathbf{X}_{i,:}$	a $1 \times q$ row vector defined by the i-th row of matrix $\mathbf{X}$
$\mathbf{X}_{:,j}$	a $p \times 1$ column vector defined by the j-th column of matrix $\mathbf{X}$
$\mathbf{X}^T$	transpose of matrix $\mathbf{X}$
$\mathbf{X}^{-1}$	inverse of square, full-rank matrix $\mathbf{X}$
$\mathbf{X}^{\dagger}$	pseudoinverse of full-rank matrix $\mathbf{X}$
$\mathrm{rank}(\mathbf{X})$	rank of matrix $\mathbf{X}$
$\mathbf{X}\|_{\mathcal{A}}$	submatrix of $\mathbf{X}$ defined by columns of $\mathbf{X}$ picked out by indices in set $\mathcal{A}$
$\mathcal{S}_k(p,q)$	set of real $p \times q$ matrices with every row being k-sparse
$\mathbb{1}_{\mathcal{A}}(x)$	indicator function that equals to 1 if $x \in \mathcal{A}$ and 0 otherwise
$\mathbf{I}_p$	identity matrix of size $p \times p$
$\mathbf{0}_{p\times q}$	matrix of size $p \times q$ with all entries being zeros
$\mathbf{1}_{p\times q}$	matrix of size $p \times q$ with all entries being ones
$\mathrm{vec}(\mathbf{X})$	vector of size $pq \times 1$ obtained by orderly stacking the columns of $\mathbf{X}$ vertically
$\mathrm{supp}(\mathbf{x})$	set of indices of $\mathbf{x}$ that contain non-zero values
$\mathbf{x}_{[k]}$	a $p \times 1$ vector with best k-term approximation to $\mathbf{x}$
$\mathbf{x}_{[k,p']}$	a $p \times 1$ vector with every block subvector of size p' approximated with its best k terms
$\mathcal{T}_z(\cdot)$	thresholding operator that sets every entry of input vector less than z to z.
$(f * g)(t)$	one-dimensional convolution of functions $f(t)$ and $g(t)$

Appendix B

Dynamic Range of Single-Photon Detectors

Consider the active imaging setup with an illumination source and an ideal single-pixel photon counter, one that can record multiple photon detections over time from the scene-backreflected light with unit detector efficiency. Then, based on classical photodetection theory, the number of photon detections at the kth time bin of the ideal photon counter is Poisson distributed as [46]

$$\mathbf{y}_k \sim \text{Poisson}\left(\int_{(k-1)\Delta}^{k\Delta} \lambda(t)\, dt\right), \quad \text{for } k = 1, \ldots, m. \tag{B.1}$$

We can always define $\lambda(t) = a\lambda'(t)$ such that $\lambda'(t)$ integrates to 1 and $a = \int_0^{T_r} \lambda(t)dt$ is the effective strength of the rate function. Then we can write the probabilities of having no detection and at least one detection event at time bin k are

$$p_{\text{miss}}(k) = \Pr[\mathbf{y}_k = 0] = \exp\left\{-a \int_{(k-1)\Delta}^{k\Delta} \lambda'(t)\, dt\right\}, \tag{B.2}$$

$$p_{\text{hit}}(k) = \Pr[\mathbf{y}_k \neq 0] = 1 - \exp\left\{-a \int_{(k-1)\Delta}^{k\Delta} \lambda'(t)\, dt\right\}, \tag{B.3}$$

using Eq. (B.1).

In contrast to the ideal photon counter, a practical photon-counting detector, such as the single-photon avalanche diode (SPAD) [124], suffers from what is known as dead (or reset) time, which is the time it takes for the detector to become active and start recording again after a photon detection event. Letting T_d be the dead time of a detector, we assume that $T_d > T_r$.

Based on Eq. (B.2) and (B.3), we can write the probability of the detection at time bin k of such a non-ideal detector suffering from dead time effects as

$$\begin{aligned}
&\Pr[\text{detection at bin } k \,|\, \text{detection in } [0, T_r)] && \text{(B.4)}\\
&= \frac{1}{Z}\left(\prod_{i=1}^{k-1} p_{\text{miss}}(i)\right) p_{\text{hit}}(k) && \text{(B.5)}\\
&= \frac{1}{Z}\exp\left\{-a\int_0^{(k-1)\Delta}\lambda'(t)\,dt\right\}\left(1-\exp\left\{-a\int_{(k-1)\Delta}^{k\Delta}\lambda'(t)\,dt\right\}\right), && \text{(B.6)}
\end{aligned}$$

where Z is the normalization scalar, given by

$$\begin{aligned}
Z &= \sum_{k=1}^{m}\left(\exp\left\{-a\int_0^{(k-1)\Delta}\lambda'(t)\,dt\right\} - \exp\left\{-a\int_0^{k\Delta}\lambda'(t)\,dt\right\}\right) && \text{(B.7)}\\
&= 1-\exp\left\{-a\int_0^{m\Delta}\lambda'(t)\,dt\right\} = 1-\exp\{-a\}. && \text{(B.8)}
\end{aligned}$$

Note that the photodetection events in bins later than k, which is the first bin with a detection event, are considered as “don’t cares” due to the detector dead time. After N_s independent illumination trials, we would have gathered a number of detections. By the dead-time constraint, note that the number of detections must be less than or equal to N_s. In fact, the distribution of the total number of photon detections is

$$\mathbf{c} \sim \text{Binomial}(N_s, 1-\exp\{-a\}), \tag{B.9}$$

as every illumination period defines a Bernoulli trial for detection with success probability $1-\prod_{k=1}^{m} p_{\text{miss}}(k) = 1-\exp\{-a\}$. Here we have assumed that there is sufficient time gap between one illumination trial to its next, such that the dead time effect does not overflow and the two adjacent trials are completely independent.

In our applications, we are interested in the photodetection characteristics of non-ideal detectors suffering from dead time when operating at low-flux condition. We define the vector $\mathbf{x}$ that represents the discretized normalized waveform: $\mathbf{x}_k = \int_{(k-1)\Delta}^{k\Delta} \lambda'(t)\, dt$ for $k = 1, \ldots, m$. Then the behavior of the detection probability under the low-flux condition, $a \to 0^+$ becomes

$$\begin{aligned}
&\Pr[\text{detection at bin } k \,|\, \text{detection in } [0, T_r)] && \text{(B.10)}\\
&\propto \exp\left\{-a\sum_{i=1}^{k-1}\mathbf{x}_i\right\} - \exp\left\{-a\sum_{i=1}^{k}\mathbf{x}_i\right\} && \text{(B.11)}\\
&\overset{a\to 0^+}{\to} \left(1 - a\sum_{i=1}^{k-1}\mathbf{x}_i\right) - \left(1 - a\sum_{i=1}^{k}\mathbf{x}_i\right) && \text{(B.12)}\\
&= a\mathbf{x}_k. && \text{(B.13)}
\end{aligned}$$

Because $\sum_{k=1}^{m}\mathbf{x}_k = 1$ by definition, we conclude that

$$\lim_{a\to 0^+} \Pr[\text{detection at bin } k \,|\, \text{detection in } [0, T_r)] = \mathbf{x}_k. \tag{B.14}$$

The result in Eq. (B.14) states that under the low-flux condition, the distribution from which photon detection times are sampled is proportional to the rate function defined by the detector. Letting $N_s \to +\infty$, we can see that the total number of photon detections is distributed as

$$\mathbf{c} \sim \text{Poisson}(N_s a) = \text{Poisson}\left(N_s a \sum_{k=1}^{n}\mathbf{x}_k\right), \tag{B.15}$$

using the fact that the limiting distribution of the binomial expression in Eq. (B.9) is Poisson. Since $\mathbf{c}_i \sim \text{Poisson}(N_s a\mathbf{x}_i)$ and $\mathbf{c}_j \sim \text{Poisson}(N_s a\mathbf{x}_j)$ are independent for $i \neq j$ for $i, j \in \{1, \ldots, m\}$, we can conclude that

$$\mathbf{y} \sim \text{Poisson}\left(N_s a\mathbf{x}\right), \tag{B.16}$$

where Poisson$(\cdot)$ is defined entrywise. In other words, even when using a non-ideal

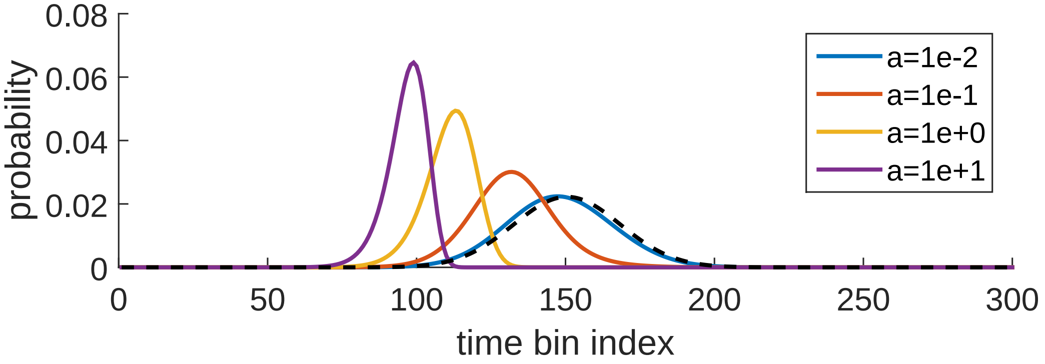


Figure B-1: Plots of probability function in Eq. (B.6) for different values of a. For increasing a, the probability function deviates away from the normalized rate function $\mathbf{x}$ shown as the dashed black line.

detector with dead time effects, as long as we operate in the low-flux regime with sufficiently large N_s, the photodetection statistics is still Poissonian. Also, by gathering a large number of photon detections, the optical waveform $\mathbf{x}$ can be reconstructed from $\mathbf{y}$ with arbitrarily small error.

On the other hand, the above properties no longer hold when we consider imaging scenarios with moderate or high flux conditions. In such cases, because of the dead time effect, the photon count observation $\mathbf{y}$ inaccurately models $\mathbf{x}$, even when $N_s \to +\infty$. Intuitively, we can observe that a high amplitude of an optical signal incident at the detector leads to an early detection event, and thus results in a mismatch between $\mathbf{y}$ and normalized rate function $\mathbf{x}$. Figure B-1 illustrates this phenomenon; as the flux level a increases, the photodetection probability in Eq. (B.6) deviates away from the normalized rate function. Accurate imaging cannot be performed by directly using raw photon count observations, due to the mismatch of the histogram $\mathbf{y}$ and $\lambda'(t)$. By correcting the effect of dead time from $\mathbf{y}$, we hope to increase the amplitude range over which our photon count observations well model the received optical waveform, and thus be used for accurate scene information recovery.

Assuming a large number of illuminations ($N_s \gg 0$) and using Eq. (B.6) and

(B.9), the kth bin value of the count histogram vector can be approximated as

$$\mathbf{y}_k \approx \mathbb{E}[\mathbf{c}]\Pr[\text{detection at bin } k \,|\, \text{detection in } [0, T_r)] \tag{B.17}$$

$$= (N_s(1-\exp\{-a\}))\,\frac{\exp\left\{-a\sum_{i=1}^{k-1}\mathbf{x}_i\right\}(1-\exp\{-a\mathbf{x}_k\})}{1-\exp\{-a\}}. \tag{B.18}$$

$$= N_s \exp\left\{-a\sum_{i=1}^{k-1}\mathbf{x}_i\right\}(1-\exp\{-a\mathbf{x}_k\}), \tag{B.19}$$

for $k = 1, \ldots, m$. Using Eq. (B.19), we propose Algorithm 7 that corrects the dead time effect to estimate an $m \times 1$ waveform $\hat{\mathbf{x}}_{\text{fix}}$ from raw photon-count data $\mathbf{y}$. Our algorithm observes that the summation in the first exponential term of Eq. (B.19) is a cumulative sum, and thus uses the sum of estimates $(\hat{\mathbf{x}}_{\text{fix}})_1, \ldots, (\hat{\mathbf{x}}_{\text{fix}})_{k-1}$ to compute $(\hat{\mathbf{x}}_{\text{fix}})_k$ from $\mathbf{y}_k$ for $1 < k \leq m$, Thus, because of its recursive form, the algorithm has computational complexity which is quadratic in the number of detector time bins. Also, note that the proposed algorithm uses a two-step approach of maximum likelihood estimation of amplitude $\hat{a}$ from Eq. (B.9), and using $\hat{a}$ for iterative estimation of waveform $\hat{\mathbf{x}}_{\text{fix}}$ from Eq. (B.19).

Algorithm 7 Proposed algorithm for dead time effect mitigation

```
Input: $\mathbf{y}$, $N_s$
Output: $\hat{\mathbf{x}}_{\text{fix}}$
$s = 0$, $k = 0$
$\hat{a} = \log(N_s/(N_s - \sum_{i=1}^{m}\mathbf{y}_i))$          ▷ ML amplitude estimate
repeat
    $k \leftarrow k + 1$
    $(\hat{\mathbf{x}}_{\text{fix}})_k \leftarrow -\log(1 - \mathbf{y}_k/(N_s\exp\{-\hat{a}s\}))/\hat{a}$
    $s \leftarrow s + (\hat{\mathbf{x}}_{\text{fix}})_k$
until $k = m$
```

Simulation results: Figure B-2 shows a simulation example of using Algorithm 7 to perform waveform correction for a photon counting detector suffering from dead time. The ground truth waveform distribution $\mathbf{x}$ (blue dashed line, labeled as "truth" in the plot) was taken to be a shifted Gaussian pulse on a pedestal that models the ambient light in the scene. We set $N_s = 10000$ and $a = 2$, such that the mean number

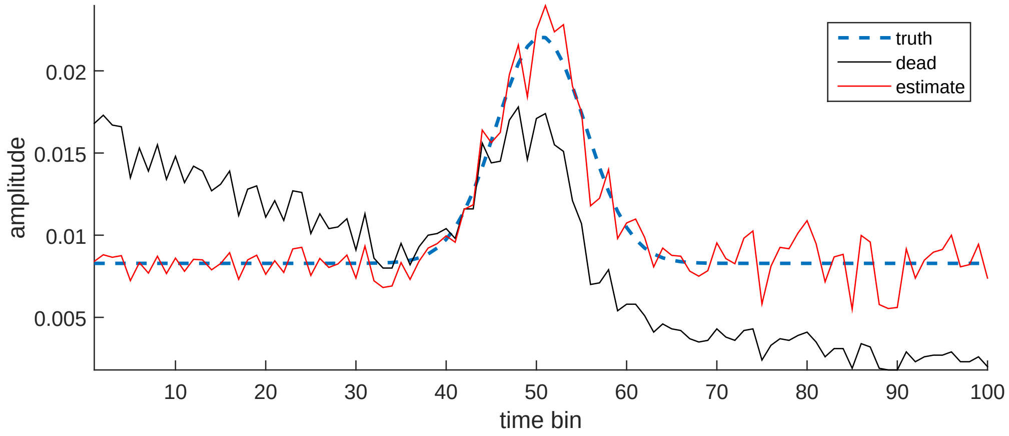


Figure B-2: Plots of $\mathbf{x}$ (dashed blue line, labeled as "truth"), $\mathbf{y}/n_s$ (solid black line, labeled as "dead"), and $\hat{\mathbf{x}}_{\text{fix}}$ (solid red line, labeled as "estimate"), which was obtained using Algorithm 7.

of photon detections was calculated to be around 8000. The scaled raw observation $\mathbf{y}/N_s$ (blue dashed line, labeled as "dead" in the plot), which was corrupted by dead time effects, had root mean square error (RMSE) of 0.005. As expected, the plot shows that the dead time effect translates to a bias in photodetection earlier in the acquisition interval. The corrected waveform $\hat{\mathbf{x}}_{\text{fix}}$ using Algorithm 7 had RMSE of 0.001.

Experimental validation: The dead time phenomenon and the effectiveness of the correction algorithm were also validated by using photon counting experimental data (see Figure B-3). With the single-photon avalanche photodiode detector plus the HydraHarp time-correlator, we were able to record photon arrivals from pulse illuminations on a white board that was 2 meters away. We used three illumination powers: 30 mW (low), 50 mW (moderate), and 70 mW (high). In order to compute reconstruction errors, we calibrated the ground truth rate function $\mathbf{x}$, by gathering high photon counts at very low-flux operations such that the dead time effect is negligible. We found that our algorithm successfully corrects the distortions in observed photon-count histograms to achieve a lower RMSE for the three light levels. At low light levels, because the dead time effect is not that severe, the raw and corrected results are both close to the ground truth, and exhibit low RMSE. On the other hand,

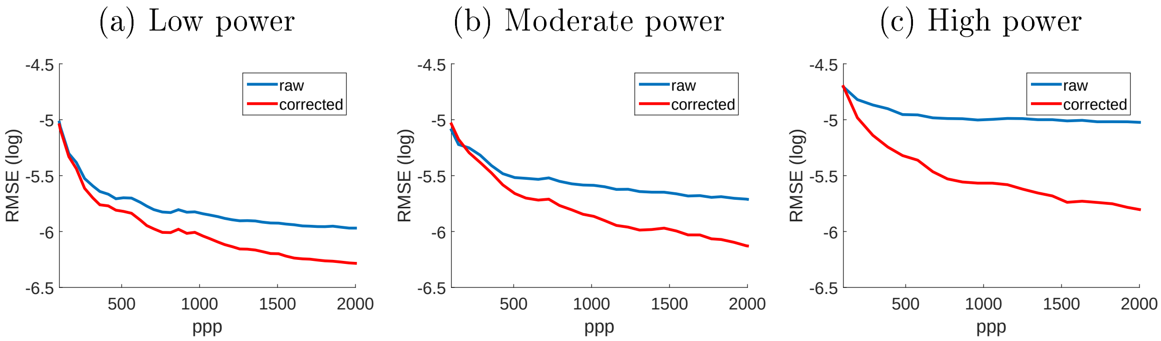


Figure B-3: Plots of log-RMSE of $\mathbf{y}$ and $\hat{\mathbf{x}}_{\text{fix}}$ for increasing number of photon detections at three different laser illumination powers.

at high light levels, the RMSE improvement from our correction method is large.

Appendix C

Convexity of Negative Log Poisson Likelihood

Consider the linear observation model from photodetection:

$$\mathbf{y} \sim \text{Poisson}(\mathbf{A}\mathbf{x} + \mathbf{b}), \tag{C.1}$$

where $\mathbf{y}$ is an $m \times 1$ photon count vector, $\mathbf{A}$ is an $m \times n$ deterministic forward imaging matrix, $\mathbf{x}$ is an $n \times 1$ deterministic vector representing scene information, and $\mathbf{b}$ is an $m \times 1$ deterministic perturbation vector. The negative log-likelihood of $\mathbf{x}$ is then

$$\mathcal{L}(\mathbf{x}; \mathbf{A}, \mathbf{y}) = \sum_{k=1}^{m} \left[(\mathbf{A}\mathbf{x} + \mathbf{b})_k - \mathbf{y}_k \log(\mathbf{A}\mathbf{x} + \mathbf{b})_k + \log(\mathbf{y}_k)! \right]. \tag{C.2}$$

It suffices to show that the Hessian matrix of $\mathcal{L}(\mathbf{x}; \mathbf{A}, \mathbf{y})$ is non-negative to prove that it is convex in $\mathbf{x}$. The gradient of $\mathcal{L}(\mathbf{x}; \mathbf{A}, \mathbf{y})$ is

$$\nabla_{\mathbf{x}} \mathcal{L}(\mathbf{x}; \mathbf{A}, \mathbf{y}) = \mathbf{A}^T \mathbf{1}_{m \times 1} - \sum_{k=1}^{m} \left[\frac{\mathbf{y}_k}{(\mathbf{A}\mathbf{x} + \mathbf{b})_k} (\mathbf{A}^T)_{:,k} \right], \tag{C.3}$$

from which we can compute the Hessian as

$$\nabla^2_{\mathbf{x}} \mathcal{L}(\mathbf{x}; \mathbf{A}, \mathbf{y}) = \mathbf{A}^T \operatorname{diag}(\mathbf{y}) \operatorname{diag}^{-2}(\mathbf{A}\mathbf{x} + \mathbf{b}) \mathbf{A}. \tag{C.4}$$

For any $\mathbf{v} \in \mathbb{R}^{n\times 1}$, we have

$$\begin{aligned}
\mathbf{v}^T \nabla_{\mathbf{x}}^2 \mathcal{L}(\mathbf{x}; \mathbf{A}, \mathbf{y})\mathbf{v} &= (\mathbf{A}\mathbf{v})^T \operatorname{diag}(\mathbf{y}) \operatorname{diag}^{-2}(\mathbf{A}\mathbf{x} + \mathbf{b})(\mathbf{A}\mathbf{v}) \qquad \text{(C.5)} \\
&= \sum_{k=1}^{m} \frac{\mathbf{y}_k}{(\mathbf{A}\mathbf{x} + \mathbf{b})_k^2} (\mathbf{A}\mathbf{v})_k^2. \qquad \text{(C.6)}
\end{aligned}$$

Since $\mathbf{y}$ is a non-negative count vector, the expression in Eq. (C.6) is always non-negative. Thus, the negative log Poisson likelihood is a convex function in $\mathbf{x}$.

Appendix D

Approximation of Poisson Likelihood for Single-Photon Depth Imaging

Consider the general Poisson channel:

$$\mathbf{y} \sim \text{Poisson}(\mathbf{A}\mathbf{x}), \tag{D.1}$$

where $\mathbf{y}$ is an $m \times 1$ observed vector with entrywise Poisson distributed samples, $\mathbf{A}$ is an $m \times n$ non-negative deterministic matrix, and $\mathbf{x}$ is an $n \times 1$ non-negative deterministic vector. Recall that the negative log likelihood function of $\mathbf{x}$ is

$$\mathcal{L}(\mathbf{x}; \mathbf{A}, \mathbf{y}) = \sum_{k=1}^{m} \left[(\mathbf{A}\mathbf{x})_k - \mathbf{y}_k \log(\mathbf{A}\mathbf{x})_k + \log(\mathbf{y}_k!) \right]. \tag{D.2}$$

Then, we can write the first-order Taylor approximation $\mathcal{L}^{(1)}(\mathbf{x}; \mathbf{A}, \mathbf{y})$ of $\mathcal{L}(\mathbf{x}; \mathbf{A}, \mathbf{y})$ about $\mathbf{x} = \mathbf{1}_{n\times 1}$ as follows:

$$\mathcal{L}^{(1)}(\mathbf{x}; \mathbf{A}, \mathbf{y}) \tag{D.3}$$

$$= \sum_{k=1}^{m} \left((\mathbf{A}\mathbf{x})_k - \mathbf{y}_k \left(\log \mathbf{A}_{k,:}\mathbf{1}_{n\times 1} + \frac{1}{\mathbf{A}_{k,:}\mathbf{1}_{n\times 1}} \mathbf{A}_{k,:}(\mathbf{x} - \mathbf{1}_{n\times 1}) \right) + \log(\mathbf{y}_k!) \right) \tag{D.4}$$

$$\cong \underbrace{\sum_{k=1}^{m} (\mathbf{A}\mathbf{x})_k}_{\text{term 1}} - \underbrace{\sum_{k=1}^{m} \frac{\mathbf{y}_k}{\mathbf{A}_{k,:}\mathbf{1}_{n\times 1}} \mathbf{A}_{k,:}\mathbf{x}}_{\text{term 2}} \tag{D.5}$$

where $\cong$ denotes equality up to a constant independent of $\mathbf{x}$. Recall that in Chapter 2, our inverse problem had the specific constraints of pixelwise reconstruction of object depth.

1. All columns of $\mathbf{A}$ have equal ℓ_2-norm and all of its rows has equal sum.
2. $\mathbf{x}$ is 1-sparse.

Then, "term 1" and the denominator in the sum of "term 2" in Eq. (D.5) are constants. It follows that Eq. (D.5) can be rewritten as

$$\begin{aligned}
\mathcal{L}^{(1)}(\mathbf{x};\mathbf{A},\mathbf{y}) &\cong -\sum_{k=1}^{m} \frac{\mathbf{y}_k}{\mathbf{A}_{k,:}\mathbf{1}_{n\times 1}}\mathbf{A}_{k,:}\mathbf{x} && \text{(D.6)}\\
&\propto \sum_{k=1}^{m} -\mathbf{y}_k\mathbf{A}_{k,:}\mathbf{x} && \text{(D.7)}\\
&\propto \sum_{k=1}^{m} -2\mathbf{y}_k\mathbf{A}_{k,:}\mathbf{x} && \text{(D.8)}\\
&\cong \sum_{k=1}^{m} \left[\mathbf{y}_k^T\mathbf{y}_k - 2\mathbf{y}_k\mathbf{A}_{k,:}\mathbf{x} + (\mathbf{A}_{k,:}\mathbf{x})(\mathbf{A}_{k,:}\mathbf{x})\right] && \text{(D.9)}\\
&= \sum_{k=1}^{m} \left(\mathbf{y}_k - \mathbf{A}_{k,:}\mathbf{x}\right)^2 = \|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2. && \text{(D.10)}
\end{aligned}$$

In conclusion, given that Poisson inverse problem includes the the single-reflector constraint for LIDAR, then the maximum likelihood estimator of $\mathbf{x}$ can be approximated using the least-squares solution by using the first-order Taylor expansion of the negative log likelihood.

Appendix E

Operation of SPAD Detector Array

In this appendix, we detail the operational characteristics of the SPAD array used for the imaging experiments in Chapter 4.

SPAD detector array: The 32×32-pixel single-photon avalanche photodiode (SPAD) array was developed by F. Villa *et al.* at the Politecnico di Milano [48, 54, 125]. It has two modes of operation: counting mode and timing mode. In counting mode, a 6-bit counter located at each detector site is incremented each time a photon is detected, allowing for up to 63 photon detections before rolling over to 0. In timing mode, the SPAD camera outputs a digital signal at the start of each acquisition window for synchronization with external devices and for time stamping photon detection events. An internal clock running at 160 MHz controls the timing signals of the SPAD array, and each pixel has a time-to-digital converter (TDC) comprising the same 6-bit counter plus a 4-bit interpolator to provide 10 bits of timing resolution at 390 ps. In our experiment, we operated the SPAD array in the timing mode for simultaneous depth and reflectivity imaging.

Data-acquisition timing terminology and rationale: Figure E-1 illustrates the terminology and time scales for the timing-mode acquisition process. Acquisition begins when the camera issues the trigger signal to the laser diode—for synchronizing its pulse generation schedule—and ends 400 ns later, i.e., the acquisition window set by

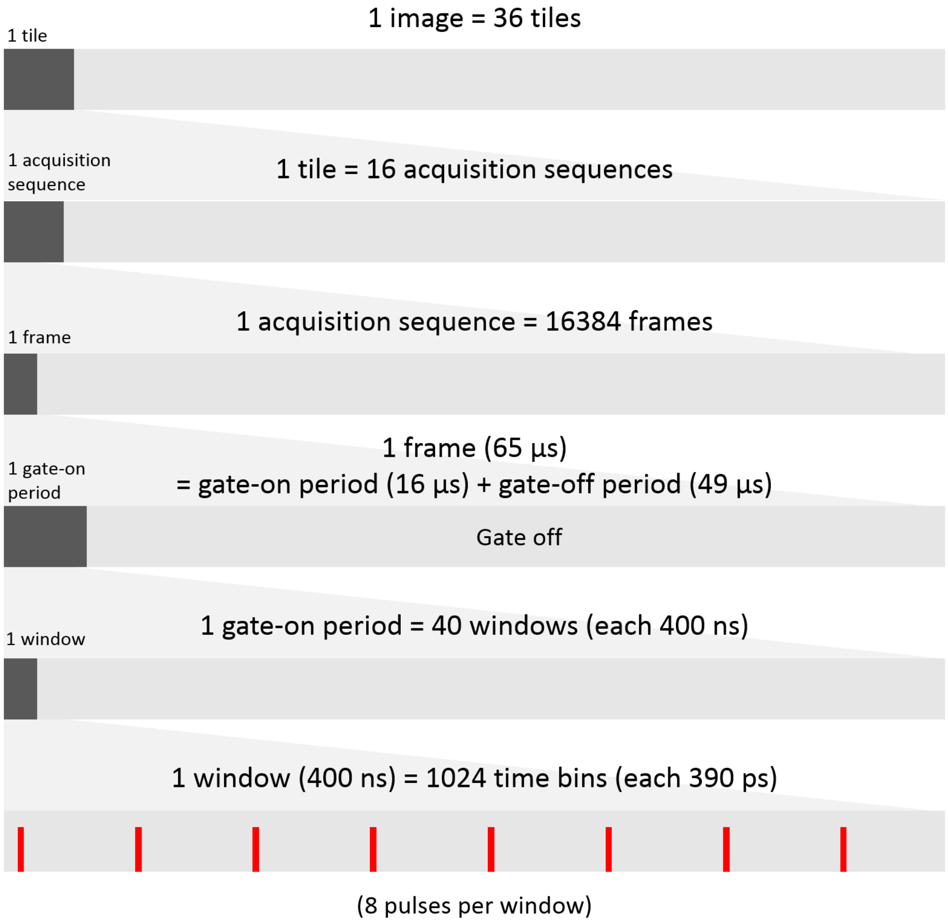

1 image = 36 tiles
1 tile
1 acquisition sequence
1 tile = 16 acquisition sequences
1 frame
1 acquisition sequence = 16384 frames
1 gate-on period
1 frame (65 µs)
= gate-on period (16 µs) + gate-off period (49 µs)
Gate off
1 window
1 gate-on period = 40 windows (each 400 ns)
1 window (400 ns) = 1024 time bins (each 390 ps)
(8 pulses per window)

Figure E-1: SPAD-camera image acquisition terminology for RS-mode operation.

the TDC's 10-bit timing resolution combined with its $\Delta = 390\,\text{ps}$ time-bin duration. The TDC timing results at each pixel are stored locally in a 10-bit memory, and the data from all 1024 memory registers are then transferred to an off-chip random access memory (RAM) from which they are routed to a computer for processing. During the $10\,\mu\text{s}$ needed for this data transfer, no photon detections can be recorded. The resulting 4% duty cycle for a single acquisition-readout cycle severely impacts measurements that require quick acquisition of a large amount of data. In particular, our laser's $T_r = 50\,\text{ns}$ pulse repetition period implies that only 8 laser pulses are sent in an acquisition window, making our experiment's effective pulse rate 800 kHz, instead of 20 MHz, when each acquisition window is immediately followed by data readout.

The camera has a work-around provision, called readout-skip (RS) mode, that allows acquisition windows to be strung together without gaps by holding off the long readout process. In the experiment we adjoined 40 windows to form a single continuous gate-on period of $16\,\mu$s. The gate-on period is followed by a 49-μs-long gate-off period, for limiting chip power dissipation and for reading out and transferring data from 1024 local memory registers to the computer. As indicated in Figure E-1, a full cycle of a gate-on period and the subsequent gate-off period constitutes a 65-μs-long frame. In our experiment we acquired $1.06\,$s of imaging data—comprised of 16,384 frames—for each camera position in a high-precision scan. These scans, which were performed with a two-axis translation stage, formed a 384×384 pixel composite image, as detailed below.

There is a compromise inherent in operating the SPAD camera in its RS mode: suspending readout until the end of the gate-on period means that at most one photon detection per frame can be recorded at any given pixel, and this limitation could have consequences for depth and reflectivity imaging. Consider a photon detection recorded at a particular pixel in a particular frame. RS-mode operation precludes our knowing in which of that frame's 40 acquisition windows the detection event occurred, even though the 10-bit timing information within that window is intact. Because the laser's pulse-repetition period is chosen to satisfy $cT_r/2 > z_{\max}$, where $z_{\max}$ is the scene's maximum depth, each detection of a back-reflected photon is due to the immediately preceding laser pulse. Hence the quality of timing data collected with RS-mode enabled is the same as would be obtained—over a much longer image-acquisition time—with that mode disabled.

The situation is different for reflectivity measurements, because the number of preceding laser pulses before the detection event within a frame yields useful reflectivity information. We operate in a low-flux scenario, however, so the probability of obtaining a detection event at a particular pixel within a frame of 40 windows is very low. It follows that many frames are needed to acquire the first detection event. RS-mode operation records the frame in which each detection event occurs, but neither the window within that frame in which it occurred, nor the laser pulse within that window which produced that detection. In other words, we retain coarse timing information (number of frames before the detection event happens) but lose the finer timing details (number of laser pulses preceding the detection event within the frame). Given that we use over 16,000 frames per acquisition sequence, RS-mode operation has minimal adverse impact on reflectivity estimation.

Setup and spatial scanning: A diagram of the experimental setup is shown in Figure E-2. The SPAD array occupies a 4.8×4.8 mm footprint that contains 32×32 pixels. We would like to have a larger imaging area, as well as more pixels, in order to more fully explore computational reconstruction methods. Given the limited size of the current SPAD array, we opted to stitch together multiple image acquisitions to form a larger-size composite image. To do so we mounted the SPAD array on a feedback-controlled, two-axis motorized translation stage and translated the array in a square pattern of 6×6 seamless tiles, as shown in Figure E-3(a), giving us a total imaging area of $28.8 \times 28.8\,\text{mm}$ and a resolution of 192×192 pixels. The $30\,\mu\text{m}$ diameter of each pixel's circular active area being less than half the 150-μm pixel pitch provided a way to increase resolution with the same image area by translating the SPAD array along its horizontal (x) and vertical (y) axes to all combinations of 0 and 1/2 pixel-pitch distances before moving on to the next tile, as shown in Figure E-3(c). In other words, we combined 6×6 tiling with factor-of-2 sub-pixel sampling by collecting data at $(x, y) = (L_x + \Delta_x, L_y + \Delta_y)$ for all combinations of $L_x, L_y \in \{0, 4.8, 9.6, 14.4, 19.2, 24\}$ and $\Delta_x, \Delta_y \in \{0, 0.075\}$, where all dimensions are in mm. This gave us a total imaging resolution of $32 \times 6 \times 2 = 384$ pixels along each axis and required a total of $6 \times 6 \times 2 \times 2 = 144$ image acquisition sequences to produce the resulting 384×384 pixel image. Although this approach required scanning, it was still significantly faster than raster scanning with a single-pixel SPAD, which would have required a total of $384^2 = 147,456$ image acquisition sequences for the same resolution.

To obtain sufficient data for comparison with conventional imaging methods, we configured the SPAD camera to output a total of 16,384 frames for each acquisition sequence, which at 65 μs per frame took 1.06 s. MATLAB control programming was used to automate the acquisition procedure and synchronously drive the translation stage to each of the 144 camera positions, outputting a separate file for each position containing photon detection data. To speed up data processing, collaborator D. Venkatraman developed a custom GNU C program to read the SPAD array binary files and output a file containing only a sequence of photon detections tagged by pixel number and time location number. The C program also has both ASCII and binary output modes, which are conveniently and efficiently read in MATLAB.

Mounted in front of the translatable SPAD array was a standard Canon FL-series photographic lens with a focal length of 55 mm and an f/1.2 maximum photographic aperture. We set the lens' aperture to f/2.8, which provided depth of field sufficient to capture details from the object depths in our scene, increased sharpness, and reduced vignetting. The lens, designed for 35 mm film cameras, had an image-circle diameter slightly larger than 35 mm, allowing us to mount the lens at a fixed position and conveniently fit our 24 × 24 mm imaging area entirely within the lens' image circle. Preliminary tests in counting mode showed that at f/2.8 the lens was able to easily and clearly resolve objects as small as one pixel in the 384 × 384 pixel image with almost zero cross-talk.

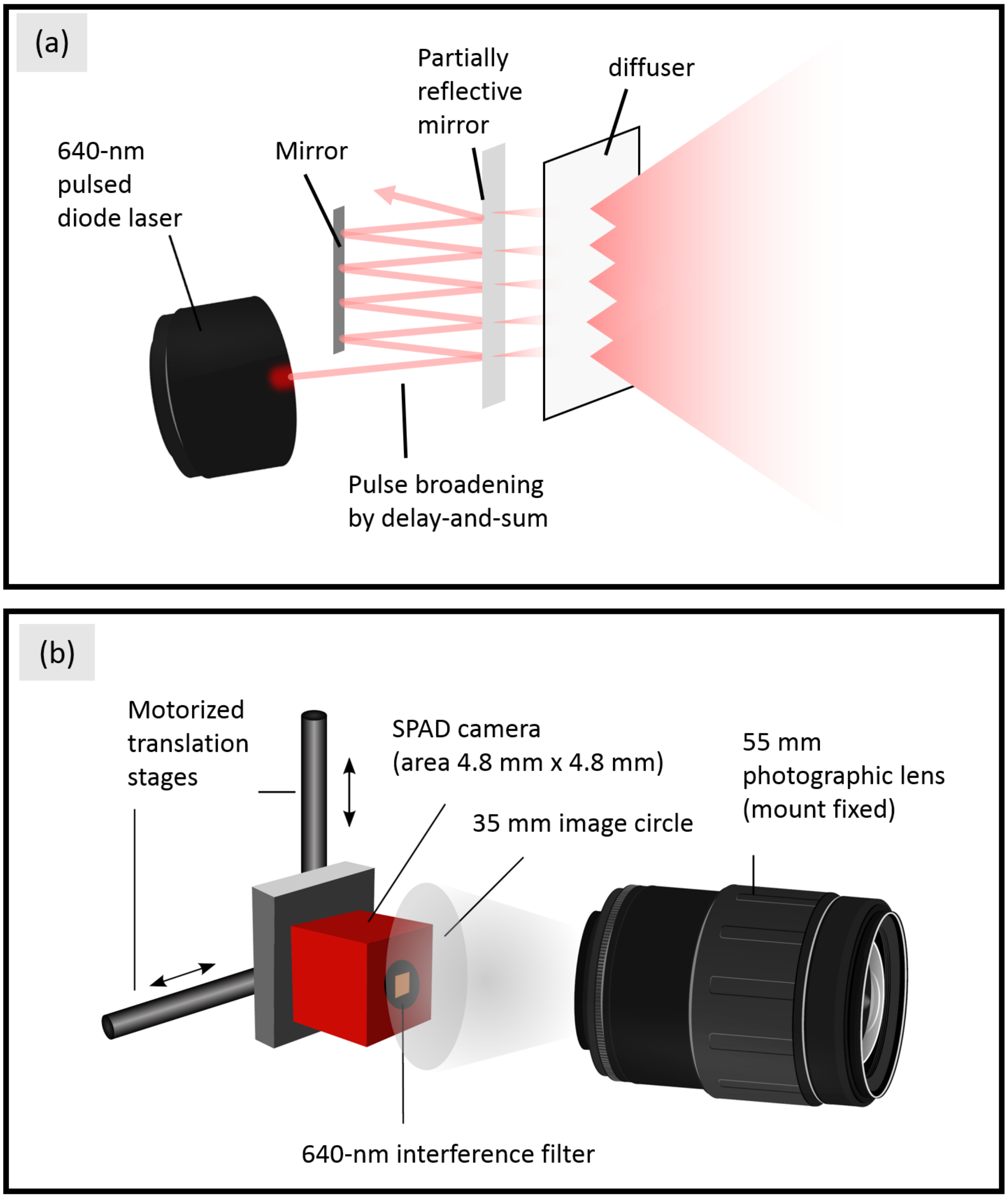

(a)
640-nm pulsed diode laser
Mirror
Partially reflective mirror
diffuser
Pulse broadening by delay-and-sum
(b)
Motorized translation stages
SPAD camera (area 4.8 mm x 4.8 mm)
35 mm image circle
55 mm photographic lens (mount fixed)
640-nm interference filter

Figure E-2: Experimental SPAD-array imaging setup.

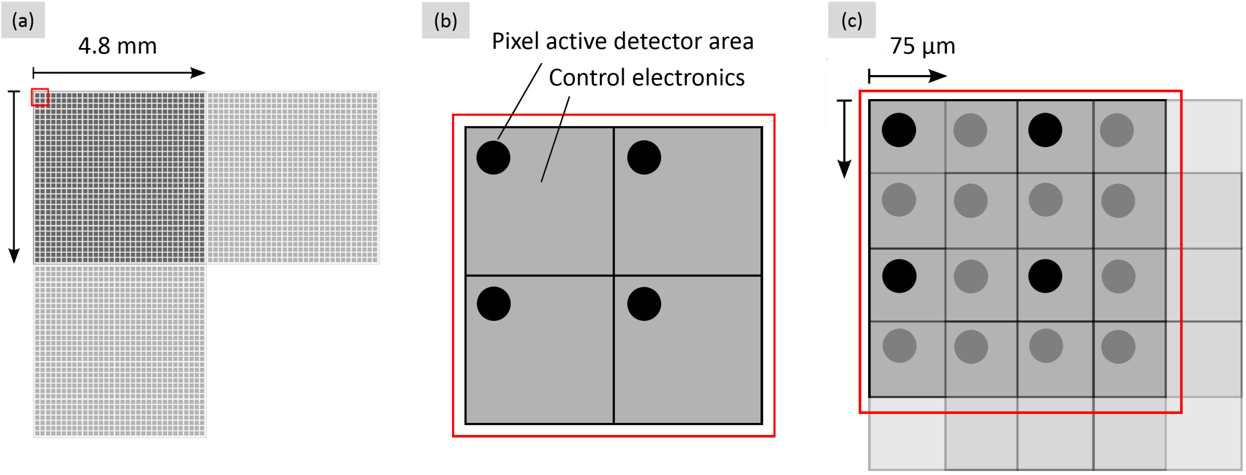

(a)
4.8 mm
(b)
Pixel active detector area
Control electronics
(c)
75 µm

Figure E-3: Scanning scheme to obtain increased image area and image resolution using the 32 × 32-pixel SPAD array. (a) Array translation by increments of a full array size (4.8 mm) along both axes to image multiple tiles of the entire array. (b) Zoom-in view of individual SPAD pixels showing the active area with no sub-pixel scanning, (c) 2×2 sub-pixel scanning by translating along each axis by 75 μm, hence multiplying resolution by 2 in each dimension.

Appendix F

SNR of the Histogram Sum Variable

Recall in Chapter 5, as part of our super-resolution single-photon imaging framework, we defined the variable $\mathbf{P}$, which is a linear sum of the raw photon-count histogram data:

$$\mathbf{P}_{i,j} = \sum_{k=1}^{m} \left(k\, \mathbf{y}_{i,j,k}\right). \tag{F.1}$$

Our interest is in computing the SNR of $\mathbf{P}_{i,j}$:

$$\mathsf{snr}(\mathbf{P}_{i,j}) = \frac{\mathbb{E}[\mathbf{P}_{i,j}]}{\sqrt{\mathrm{Var}(\mathbf{P}_{i,j})}}, \tag{F.2}$$

which tells us how confident we can be with our super-resolution image estimates. In Chapter 5, we saw that

$$\begin{aligned} \mathbb{E}[\mathbf{P}_{i,j}] &= \frac{2N_s}{c\Delta}(\mathbf{H} * (\mathbf{A} \circ \mathbf{D}))_{i,j} && \text{(F.3)} \\ \mathrm{Var}(\mathbf{P}_{i,j}) &= \frac{N_s}{\Delta^2}\left(T_p^2(\mathbf{H} * \mathbf{A})_{i,j} + (4/c^2)(\mathbf{H} * (\mathbf{A} \circ \mathbf{D} \circ \mathbf{D}))_{i,j}\right), && \text{(F.4)} \end{aligned}$$

where $\circ$ denotes the entrywise Hadamard product. We can then have that the SNR for $\mathbf{P}_{i,j}$ is

$$\mathsf{snr}(\mathbf{P}_{i,j}) = \frac{\sqrt{N_s}(\mathbf{H} * (\mathbf{A} \circ \mathbf{D}))_{i,j}}{\sqrt{((cT_p/2)^2(\mathbf{H} * \mathbf{A})_{i,j} + (\mathbf{H} * (\mathbf{A} \circ \mathbf{D} \circ \mathbf{D}))_{i,j})}}. \tag{F.5}$$

One immediate observation we can make from the above expression is that the SNR of $\mathbf{P}_{i,j}$ increases linearly with the square root of the number of illuminations N_s. In other words, by using an arbitrarily long acquisition time ($N_s \to +\infty$), we can always make $\mathbf{P}_{i,j}$ behave almost like a noiseless variable.

However, we are interested in low-flux operations, and thus using small N_s for our acquisition. For the sake of SNR analysis, let us assume a very short pulse ($T_p \to 0^+$), unit reflectivity map, and a blur kernel that is constant over all pixels. Let $\mathcal{H}_{i,j}$ be the index set of pixels that the kernel defines when it is centered at pixel (i, j). Because of the normalization rule of the blur kernel, we have $\mathbf{H}_{i',j'} = 1/|\mathcal{H}_{i',j'}|$ for $(i', j') \in \mathcal{H}_{i,j}$ for any i and j. Then, we can approximate and simplify Eq. (F.5) as

$$\mathsf{snr}(\mathbf{P}_{i,j}) \approx \sqrt{N_s}\frac{\frac{1}{|\mathcal{H}_{i,j}|}\sum_{(i',j')\in\mathcal{H}_{i,j}}\mathbf{D}_{i',j'}}{\sqrt{\frac{1}{|\mathcal{H}_{i,j}|}\sum_{(i',j')\in\mathcal{H}_{i,j}}\mathbf{D}^2_{i',j'}}} \tag{F.6}$$

$$= \sqrt{N_s}\frac{\mathcal{M}_1(\{\mathbf{D}_{i',j'}\}_{(i',j')\in\mathcal{H}_{i,j}})}{\mathcal{M}_2(\{\mathbf{D}_{i',j'}\}_{(i',j')\in\mathcal{H}_{i,j}})}, \tag{F.7}$$

where $\mathcal{M}_p(\{\mathbf{D}_{i',j'}\}_{(i',j')\in\mathcal{H}_{i,j}})$ denotes the generalized mean of power p for the set $\{\mathbf{D}_{i',j'}\}_{(i',j')\in\mathcal{H}_{i,j}}$:

$$\mathcal{M}_p(\{\mathbf{D}_{i',j'}\}_{(i',j')\in\mathcal{H}_{i,j}}) = \left(\frac{1}{|\mathcal{H}_{i,j}|}\sum_{(i',j')\in\mathcal{H}_{i,j}}\mathbf{D}^p_{i',j'}\right)^{1/p}. \tag{F.8}$$

The generalized means of power 1 and 2 obey

$$\mathcal{M}_1(\{\mathbf{D}_{i',j'}\}_{(i',j')\in\mathcal{H}_{i,j}}) \le \mathcal{M}_2(\{\mathbf{D}_{i',j'}\}_{(i',j')\in\mathcal{H}_{i,j}}) \tag{F.9}$$

with equality holding only when all the elements of $\{\mathbf{D}_{i',j'}\}_{(i',j')\in\mathcal{H}_{i,j}}$ are equal. In other words, the SNR value in Eq. (F.7) is maximized, when the depths of pixels defined by the blur kernel h represent a planar surface.

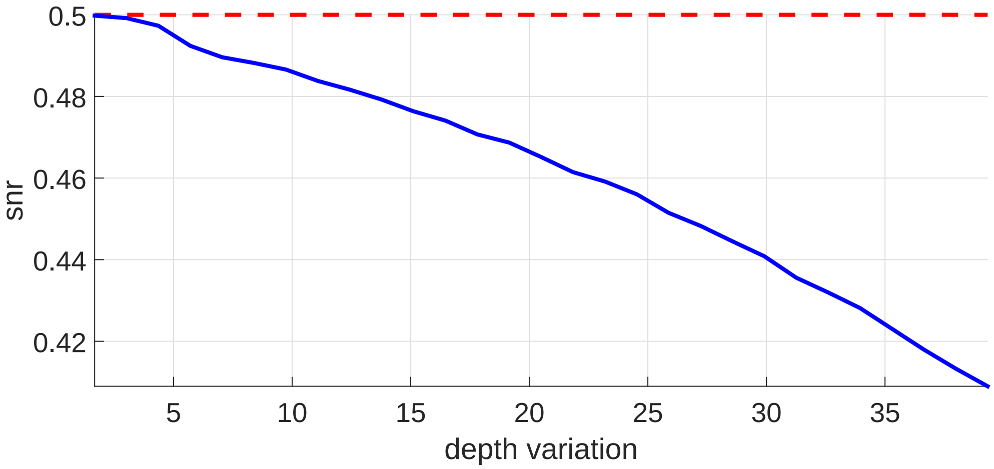


Figure F-1: Depth variation vs. SNR. The solid blue line shows Eq. (F.7) and the dashed red line indicates $\sqrt{N_s}$, which is an upper bound to SNR, for various values of depth-var.

Define the depth variation quantity at pixel (i, j) by

$$\mathsf{depth\text{-}var}(i,j) = \sqrt{\frac{1}{|\mathcal{H}_{i,j}|} \sum_{(i'',j'') \in \mathcal{H}_{i,j}} \left(\mathbf{D}_{i'',j''} - \frac{1}{|\mathcal{H}_{i,j}|} \sum_{(i',j') \in \mathcal{H}_{i,j}} \mathbf{D}_{i',j'} \right)^2}, \tag{F.10}$$

which is simply the standard deviation of depths of reflectors defined by the kernel $\mathbf{H}$ at pixel (i, j). Figure F-1 verifies our intuition that little variation in depths is required to achieve high SNR of the data variable $\mathbf{P}_{i,j}$, by plotting Eq. (F.7) with respect to $\mathsf{depth\text{-}var}(i, j)$. This plot was generated by simulating Eq. (F.7) with reflectors placed in $m = 100$ depth bins in a uniformly random fashion for a 5×1 image. In other words, the dimension of $\mathbf{H}$ was set to be 5 pixels for this simulation.

Appendix G

Asymptotic Efficiency of Least-Squares Single-Pixel Imaging

The Fisher information matrix (FIM) of the probability function $f_{\mathsf{y}}(\mathbf{y};\mathbf{x})$ that is parametrized by a column vector $\mathbf{x}$ is defined as

$$\mathsf{J}(\mathbf{x}) = \mathbb{E}_{\mathbf{y}}\left[\left(\frac{\partial}{\partial \mathbf{x}}\log f(\mathbf{y};\mathbf{x})\right)\left(\frac{\partial}{\partial \mathbf{x}}\log f(\mathbf{y};\mathbf{x})\right)^{T}\right]. \tag{G.1}$$

Intuitively speaking, the FIM measures the average concavity of a likelihood function. The Cramér-Rao bound (CRB) is the trace of the inverse of FIM:

$$\mathsf{crb}(\mathbf{x}) = \mathrm{Tr}\left(\mathsf{J}(\mathbf{x})^{-1}\right). \tag{G.2}$$

The CRB is a particularly useful quantity as it is a lower bound on the variance of unbiased estimators of $\mathbf{x}$ [126]. Then, an estimator $\hat{\mathbf{x}}(\mathbf{y})$ of $\mathbf{x}$ is *efficient* if

I. it is unbiased ($\mathbb{E}[\hat{\mathbf{x}}] = \mathbf{x}$),

II. its mean square error is the Cramér-Rao bound ($\mathsf{mse}(\hat{\mathbf{x}},\mathbf{x}) = \mathsf{crb}(\mathbf{x})$).

In other words, efficient estimators are the optimal unbiased estimators. In regards to our least-squares single-pixel imaging framework in Chapter 6, we can make the following observation.

Observation 1. Consider our single-pixel imaging setup, where $\mathbf{x}$ is an $N \times 1$ non-negative scene reflectivity vector, $\mathbf{A}$ is an $M \times N$ forward imaging matrix with each entry being non-negative and i.i.d., and

$$\mathbf{y} \sim \text{Poisson}(\mathbf{A}\mathbf{x}). \tag{G.3}$$

Then, $\hat{\mathbf{x}}_{\text{LS}} = \mathbf{A}^{\dagger}\mathbf{y}$ is an asymptotically efficient estimator of $\mathbf{x}$, given $M \to +\infty$ for fixed N.

Proof. We first prove its unbiasedness.

$$\begin{aligned}
\mathbb{E}[\hat{\mathbf{x}}_{\text{LS}}] &= \mathbb{E}[\left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T\mathbf{y}] && \text{(G.4)}\\
&\overset{(a)}{=} \sum_{\mathbf{A}'} \mathbb{E}[\left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T\mathbf{y}|\mathbf{A}=\mathbf{A}']\,\Pr[\mathbf{A}=\mathbf{A}'] && \text{(G.5)}\\
&= \sum_{\mathbf{A}'} \left((\mathbf{A}')^T\mathbf{A}'\right)^{-1}(\mathbf{A}')^T\mathbb{E}[\mathbf{y}|\mathbf{A}=\mathbf{A}']\,\Pr[\mathbf{A}=\mathbf{A}'] && \text{(G.6)}\\
&\overset{(b)}{=} \sum_{\mathbf{A}'} \left((\mathbf{A}')^T\mathbf{A}'\right)^{-1}(\mathbf{A}')^T\mathbf{A}'\mathbf{x}\,\Pr[\mathbf{A}=\mathbf{A}'] && \text{(G.7)}\\
&= \sum_{\mathbf{A}'} \mathbf{x}\,\Pr[\mathbf{A}=\mathbf{A}'] && \text{(G.8)}\\
&= \mathbf{x}\sum_{\mathbf{A}'} \Pr[\mathbf{A}=\mathbf{A}'] && \text{(G.9)}\\
&= \mathbf{x}, && \text{(G.10)}
\end{aligned}$$

where (a) uses the law of total expectation and (b) uses the fact that the mean of a Poisson random variable is equal to its rate parameter $\mathbf{A}\mathbf{x}$.

Also, we have that the FIM is

$$\begin{aligned}
\mathsf{J}(\mathbf{x}) &= \mathbb{E}_{\mathbf{y},\mathbf{A}}\left[\left(\frac{\partial}{\partial\mathbf{x}}\log f(\mathbf{y},\mathbf{A};\mathbf{x})\right)\left(\frac{\partial}{\partial\mathbf{x}}\log f(\mathbf{y},\mathbf{A};\mathbf{x})\right)^T\right] && \text{(G.11)}\\
&= \mathbb{E}_{\mathbf{A}}\left[\mathbb{E}_{\mathbf{y}}\left[\left(\frac{\partial}{\partial\mathbf{x}}\log f(\mathbf{y},\mathbf{A};\mathbf{x})\right)\left(\frac{\partial}{\partial\mathbf{x}}\log f(\mathbf{y},\mathbf{A};\mathbf{x})\right)^T\middle|\mathbf{A}=\mathbf{A}'\right]\right] && \text{(G.12)}\\
&= \mathbb{E}_{\mathbf{A}}\left[\mathbf{A}^T\text{diag}(\mathbf{A}\mathbf{x})^{-1}\mathbf{A}\right], && \text{(G.13)}
\end{aligned}$$

and thus

$$\mathsf{crb}(\mathbf{x}) = \mathrm{Tr}\left(\mathbb{E}\left[\mathbf{A}^T \mathrm{diag}(\mathbf{A}\mathbf{x})^{-1}\mathbf{A}\right]^{-1}\right). \tag{G.14}$$

When $M \to +\infty$, the diagonal matrix term effectively becomes an identity matrix scaled with $g(\mathbf{x}) = p\|\mathbf{x}\|_1$. Also, in this limit, the term inside the expectation approaches the correlation matrix, which is deterministic, and by Jensen's equality, the inverse and the expectation operators commute. In other words,

$$\mathsf{crb}(\mathbf{x}) \to g(\mathbf{x})\mathrm{Tr}\left(\mathbb{E}\left[(\mathbf{A}^T\mathbf{A})^{-1}\right]\right). \tag{G.15}$$

We recall from (6.16), that mse is computed as

$$\mathsf{mse} = \mathrm{Tr}\left(\mathbb{E}_{\mathbf{A}}\left[\mathrm{diag}(\mathbf{A}\mathbf{x})\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-2}\mathbf{A}^T\right]\right). \tag{G.16}$$

In the limit of $M \to +\infty$, the diagonal matrix becomes a scaled identity and we can write

$$\begin{aligned} \mathsf{mse} &\to g(\mathbf{x})\mathrm{Tr}\left(\mathbb{E}\left[\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-2}\mathbf{A}^T\right]\right) && \text{(G.17)} \\ &= g(\mathbf{x})\mathrm{Tr}\left(\mathbb{E}\left[(\mathbf{A}^T\mathbf{A})^{-1}\right]\right), && \text{(G.18)} \end{aligned}$$

where the last line used the commutation between trace and expectation and the cyclic property of trace. Comparing Eq. (G.15) and Eq. (G.18), we conclude that our least-squares single-pixel imaging estimator is asymptotically efficient as $M \to +\infty$. □

The asymptotic efficiency is also observable from Figure G-1, which shows the plots of mse and crb for increasing M, given that $\mathbf{x}$ was generated using a uniformly random sample of size 10 with a Bernoulli matrix $\mathbf{A}$.

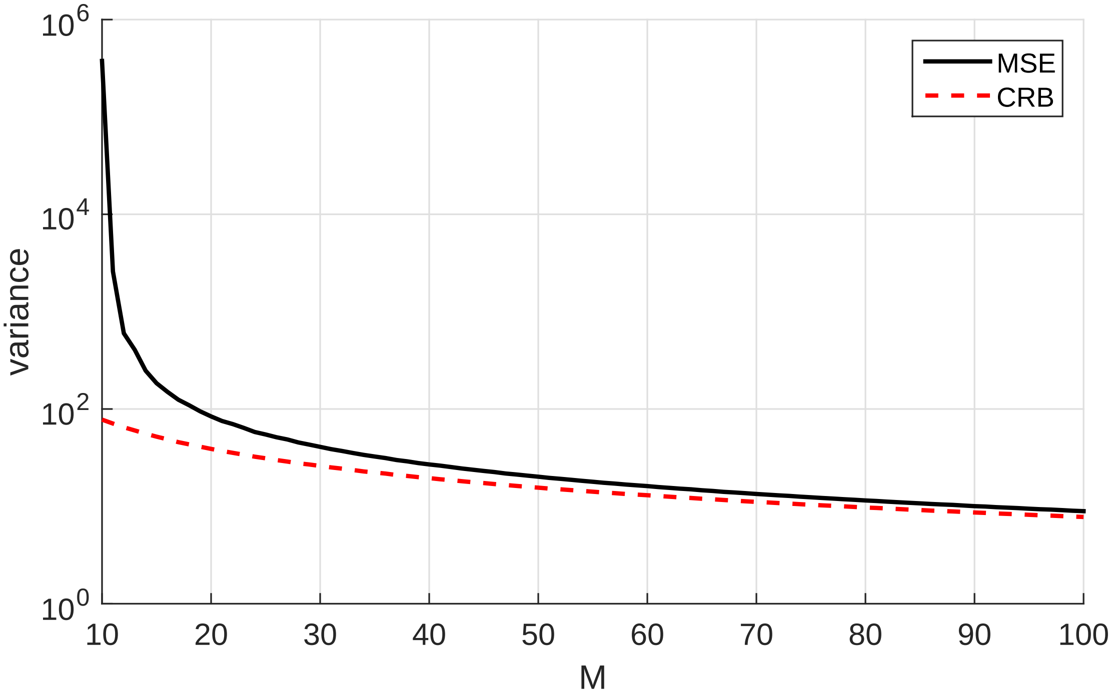


Figure G-1: Plots of crb and mse generated using Monte Carlo simulations of Eq. (G.14) and Eq. (G.16). Observe that as M increases, the crb becomes a tighter lower bound for mse.

Bibliography

[1] G. S. Buller and A. M. Wallace, “Ranging and three-dimensional imaging using time-correlated single-photon counting and point-by-point acquisition,” *IEEE J. Sel. Top. Quantum Electron.*, vol. 13, no. 4, pp. 1006–1015, 2007.

[2] D. Shin, A. Kirmani, V. K. Goyal, and J. H. Shapiro, “Photon-efficient computational 3-D and reflectivity imaging with single-photon detectors,” *IEEE Transactions on Computational Imaging*, vol. 1, no. 2, pp. 112–125, June 2015.

[3] G. C. Holst, *CCD Arrays, Cameras, and Displays*, JCD Publishing, 1998.

[4] R. Stettner, “Compact 3D flash lidar video cameras and applications,” *SPIE Laser Radar Technology and Applications XV*, 2010, pp. 768405–768405.

[5] C. Mallet and F. Bretar, “Full-waveform topographic lidar: state-of-the-art,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 64, no. 1, pp. 1–16, 2009.

[6] M. A. Albota, B. F. Aull, D. G. Fouche, R. M. Heinrichs, D. G. Kocher, R. M. Marino, J. G. Mooney, N. R. Newbury, M. E. O’Brien, B. E. Player, B. C. Willard, and J. J. Zayhowski, “Three-dimensional imaging laser radars with Geiger-mode avalanche photodiode Arrays,” *Lincoln Lab. J.*, vol. 13, no. 2, pp. 351–370, 2002.

[7] J. Romberg, “Imaging via compressive sampling,” *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 14–20, 2008.

[8] P. Milanfar, *Super-Resolution Imaging*, CRC Press, Florida, 2010.

[9] W. Becker, A. Bergmann, M. Hink, K. König, K. Benndorf, and C. Biskup, "Fluorescence lifetime imaging by time-correlated single-photon counting," *Microscopy Research and Technique*, vol. 63, no. 1, pp. 58–66, 2004.

[10] B.-Y. Sun, D.-S. Huang, and H.-T. Fang, "Lidar signal denoising using least-squares support vector machine," *IEEE Signal Process. Lett.*, vol. 12, no. 2, pp. 101–104, 2005.

[11] P. T. Boufounos, "Depth sensing using active coherent illumination," *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2012, pp. 5417–5420.

[12] 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.

[13] D. Shin, A. Kirmani, V. K. Goyal, and J. H. Shapiro, "Computational 3D and reflectivity imaging with high photon efficiency," *Proc. IEEE Int. Conf. Image Process.*, 2014, pp. 46–50.

[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," arXiv e-Print arXiv:1507.02511, 2015.

[15] D. Shin, J. H. Shapiro, and V. K. Goyal, "Single-photon depth imaging using a union-of-subspaces model," *IEEE Signal Process. Lett.*, vol. 22, no. 12, pp. 2254–2258, Dec 2015.

[16] D. Needell and J. A. Tropp, "CoSaMP: iterative signal recovery from incomplete and inaccurate samples," *Applied and Computational Harmonic Analysis*, vol. 26, no. 3, pp. 301–321, 2009.

[17] T. Goldstein and S. Osher, "The split Bregman method for L1-regularized problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 2, pp. 323–343, 2009.

[18] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde, "Model-based compressive sensing," *IEEE Trans. Inform. Theory.*, vol. 56, no. 4, pp. 1982–2001, 2010.

[19] R. E. Blahut, *Principles and Practice of Information Theory*, Addison-Wesley Longman Publishing Co., Inc., 1987.

[20] "GitHub repository for photon-efficient imaging," `https://github.com/photon-efficient-imaging/sample-data`.

[21] "GitHub repository for union-of-subspace imaging," `https://github.com/photon-efficient-imaging/uos-imaging`.

[22] M. Nilsson, "Estimation of tree heights and stand volume using an airborne lidar system," *Remote Sensing of Environment*, vol. 56, no. 1, pp. 1–7, 1996.

[23] K. I. Chang, K. W. Bowyer, and P. J. Flynn, "An evaluation of multimodal 2D+ 3D face biometrics," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 4, pp. 619–624, 2005.

[24] J.-F. Côté, J.-L. 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*, vol. 113, no. 5, pp. 1067–1081, 2009.

[25] F. Heide, M. B. Hullin, J. Gregson, and W. Heidrich, "Low-budget transient imaging using photonic mixer devices," *ACM Transactions on Graphics (TOG)*, vol. 32, no. 4, pp. 45, 2013.

[26] D. Freedman, Y. Smolin, E. Krupka, I. Leichter, and M. Schmidt, "SRA: Fast removal of general multipath for TOF sensors," *European Conference on Computer Vision (ECCV)*, pp. 234–249. Springer, 2014.

[27] H. Qiao, J. Lin, Y. Liu, M. B. Hullin, and Q. Dai, "Resolving transient time profile in ToF imaging via log-sum sparse regularization," *Opt. Lett.*, vol. 40, no. 6, pp. 918–921, 2015.

[28] S. Pellegrini, G. S. Buller, J. M. Smith, A. M. Wallace, and S. Cova, "Laser-based distance measurement using picosecond resolution time-correlated single-photon counting," *Measurement Science and Technology*, vol. 11, no. 6, pp. 712, 2000.

[29] C. Niclass, A. Rochas, P.-A. Besse, and E. Charbon, "Design and characterization of a CMOS 3-D image sensor based on single photon avalanche diodes," *IEEE Journal of Solid-State Circuits*, vol. 40, no. 9, pp. 1847–1854, 2005.

[30] C. Niclass, M. Soga, H. Matsubara, S. Kato, and M. Kagami, "A 100-m range 10-frame/s 340 96-pixel time-of-flight depth sensor in 0.18-CMOS," *IEEE Journal of Solid-State Circuits*, vol. 48, no. 2, pp. 559–572, 2013.

[31] S. Bellisai, D. Bronzi, F. Villa, S. Tisa, A. Tosi, and F. Zappa, "Single-photon pulsed-light indirect time-of-flight 3D ranging," *Opt. Expr.*, vol. 21, no. 4, pp. 5086–5098, 2013.

[32] G. Gariepy, N. Krstajić, R. Henderson, C. Li, R. R. Thomson, G. S. Buller, B. Heshmat, R. Raskar, J. Leach, and D. Faccio, "Single-photon sensitive light-in-flight imaging," *Nat. Commun.*, vol. 6, 2015.

[33] J. Castorena, C. Creusere, and D. Voelz, "Using finite moment rate of innovation for lidar waveform complexity estimation," *Forty Fourth Asilomar Conference on Signals, Systems and Computers*. IEEE, 2010, pp. 608–612.

[34] J. Castorena and C. D. Creusere, "Compressive sampling of lidar: Full-waveforms as signals of finite rate of innovation," *Proceedings of the 20th European Signal Processing Conference*. IEEE, 2012, pp. 984–988.

[35] 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. 21485–21507, 2011.

[36] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 1–38, 1977.

[37] A. M. Wallace, J. Ye, N. J. Krichel, A. McCarthy, R. J. Collins, and G. S. Buller, "Full waveform analysis for long-range 3D imaging laser radar," *EURASIP Journal on Advances in Signal Processing*, vol. 2010, no. 33, 2010.

[38] S. Hernandez-Marin, A. M. Wallace, and G. J. Gibson, "Creating multi-layered 3D images using reversible jump MCMC algorithms," *Advances in Visual Computing*, pp. 405–416. Springer, 2006.

[39] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Communications on Pure and Applied Mathematics*, vol. 57, no. 11, pp. 1413–1457, 2004.

[40] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009.

[41] B. K. Natarajan, "Sparse approximate solutions to linear systems," *SIAM Journal on Computing*, vol. 24, no. 2, pp. 227–234, 1995.

[42] M. Elad, *Sparse and Redundant Representations: from Theory to Applications in Signal and Image Processing*, Springer Science & Business Media, 2010.

[43] S. Fuchs, M. Suppa, and O. Hellwich, "Compensation for multipath in ToF camera measurements supported by photometric calibration and environment integration," *Computer Vision Systems*, pp. 31–41. Springer, 2013.

[44] A. Bhandari, A. Kadambi, R. Whyte, C. Barsi, M. Feigin, A. Dorrington, and R. Raskar, "Resolving multipath interference in time-of-flight imaging via modulation frequency diversity and sparse regularization," *Opt. Lett.*, vol. 39, no. 6, pp. 1705–1708, 2014.

[45] "GitHub repository for multi-depth imaging," `https://github.com/photon-efficient-imaging/full-waveform`.

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

[47] G. H. Golub, M. Heath, and G. Wahba, "Generalized cross-validation as a method for choosing a good ridge parameter," *Technometrics*, vol. 21, no. 2, pp. 215–223, 1979.

[48] F. A. Villa, *Time of Flight Camera for Long Distance 3D Ranging*, Ph.D. thesis, Politecnico di Milano, 2014.

[49] W. Becker, *Advanced Time-Correlated Single Photon Counting Techniques*, Springer, 2005.

[50] J. Richardson, R. Walker, L. Grant, D. Stoppa, F. Borghetti, E. Charbon, M. Gersbach, and R. K. Henderson, "A 32×32 50ps resolution 10 bit time to digital converter array in 130nm CMOS for time correlated imaging," *IEEE Custom Integrated Circ. Conf.*, 2009, pp. 77–80.

[51] J. Richardson, L. Grant, and R. K. Henderson, "Low dark count single-photon avalanche diode structure compatible with standard nanometer scale CMOS technology," *IEEE Photonics Tech. Lett.*, vol. 21, no. 14, pp. 1020–1022, 2009.

[52] F. Villa, R. Lussana, D. Bronzi, S. Tisa, A. Tosi, F. Zappa, D. M. A, D. Contini, D. Durini, S. Weyers, and W. Brockherde, "CMOS imager with 1024 SPADs and TDCs for single-photon timing and 3-D time-of-flight," *IEEE J. Sel. Top. Quantum Electron.*, vol. 20, no. 6, pp. 3804810, Nov. 2014.

[53] D. Bronzi, F. Villa, S. Tisa, A. Tosi, F. Zappa, D. Durini, S. Weyers, and W. Brockherde, "100,000 frames/s 64×32 single-photon detector array for 2-D imaging and 3-D ranging," *IEEE J. Sel. Top. Quantum Electron.*, vol. 20, no. 6, pp. 354–363, 2014.

[54] C. Scarcella, A. Tosi, F. Villa, S. Tisa, and F. Zappa, “Low-noise low-jitter 32-pixels CMOS single-photon avalanche diodes array for single-photon counting from 300 nm to 900 nm,” *Rev. Sci. Instr.*, vol. 84, no. 12, pp. 123112, 2013.

[55] D.-U. Li, J. Arlt, J. Richardson, R. Walker, A. Buts, D. Stoppa, E. Charbon, and R. Henderson, “Real-time fluorescence lifetime imaging system with a 32× 32 0.13 μm CMOS low dark-count single-photon avalanche diode array,” *Opt. Expr.*, vol. 18, no. 10, pp. 10257–10269, 2010.

[56] L. I. Rudin, S. Osher, and E. Fatemi, “Nonlinear total variation based noise removal algorithms,” *Physica D: Nonlinear Phenomena*, vol. 60, no. 1, pp. 259–268, 1992.

[57] J. Shen and T. F. Chan, “Mathematical models for local nontexture inpaintings,” *SIAM J. Appl. Math.*, vol. 62, no. 3, pp. 1019–1043, 2001.

[58] 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.

[59] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, “Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition,” *Conf. Rec. 27th Asilomar Conf. Sig., Sys., & Comput.*, Pacific Grove, CA, Nov. 1993, vol. 1, pp. 40–44.

[60] 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. Optics*, vol. 48, no. 32, pp. 6241–6251, 2009.

[61] S. P. Lipshitz, R. A. Wannamaker, and J. Vanderkooy, “Quantization and dither: A theoretical survey,” *Journal of the Audio Engineering Society*, vol. 40, no. 5, pp. 355–375, 1992.

[62] R. M. Gray and T. G. Stockham, "Dithered quantizers," *IEEE Trans. Inform. Theory.*, vol. 39, no. 3, pp. 805–812, 1993.

[63] 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*, vol. 13, no. 6, pp. 4–10, 2003.

[64] 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.

[65] I. Gatley, D. DePoy, and A. Fowler, "Astronomical imaging with infrared array detectors," *Science*, vol. 242, no. 4883, pp. 1264–1270, 1988.

[66] A. C. Watts, V. G. Ambrosia, and E. A. Hinkley, "Unmanned aircraft systems in remote sensing and scientific research: Classification and considerations of use," *Remote Sensing*, vol. 4, no. 6, pp. 1671–1692, 2012.

[67] J. Bertolotti, E. G. v. Putten, C. Blum, A. Lagendijk, W. L. Vos, and A. P. Mosk, "Non-invasive imaging through opaque scattering layers," *Nature*, vol. 491, no. 7423, pp. 232–234, 2012.

[68] R. G. Keys, "Cubic convolution interpolation for digital image processing," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 29, no. 6, pp. 1153–1160, 1981.

[69] W. T. Freeman, T. R. Jones, and E. C. Pasztor, "Example-based super-resolution," *Computer Graphics and Applications, IEEE*, vol. 22, no. 2, pp. 56–65, 2002.

[70] D. Glasner, S. Bagon, and M. Irani, "Super-resolution from a single image," *Proc. IEEE Conf. Comput. Vis. Pattern Recog.* IEEE, 2009, pp. 349–356.

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

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

[73] S. C. Park, M. K. Park, and M. G. Kang, "Super-resolution image reconstruction: a technical overview," *IEEE Signal Process. Mag.*, vol. 20, no. 3, pp. 21–36, 2003.

[74] T. F. Chan and C.-K. Wong, "Total variation blind deconvolution," *IEEE Trans. Image Process.*, vol. 7, no. 3, pp. 370–375, 1998.

[75] 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," *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2015, pp. 2376–2384.

[76] 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," *IS&T/SPIE Electronic Imaging.* International Society for Optics and Photonics, 2010, number 75330B.

[77] E. J. Candès and C. Fernandez-Granda, "Super-resolution from noisy data," *Journal of Fourier Analysis and Applications*, vol. 19, no. 6, pp. 1229–1254, 2013.

[78] E. J. Candes and C. Fernandez-Granda, "Towards a mathematical theory of super-resolution," *Commun. Pure Appl. Math.*, vol. 67, no. 6, pp. 906–956, 2014.

[79] W. H. Richardson, "Bayesian-based iterative method of image restoration," *J. Opt. Soc. Am.*, vol. 62, no. 1, pp. 55–59, 1972.

[80] L. B. Lucy, "An iterative technique for the rectification of observed distributions," *The Astronomical Journal*, vol. 79, no. 745, 1974.

[81] D. P. Bertsekas, *Nonlinear Programming*, Athena Scientific, Belmont, 1999.

[82] F. Rottensteiner and C. Briese, "A new method for building extraction in urban areas from high-resolution LIDAR data," *International Archives of Photogrammetry Remote Sensing and Spatial Information Sciences*, vol. 34, no. 3/A, pp. 295–301, 2002.

[83] 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. 23822–23837, 2013.

[84] J. Barzilai and J. M. Borwein, "Two-point step size gradient methods," *IMA Journal of Numerical Analysis*, vol. 8, no. 1, pp. 141–148, 1988.

[85] A. Oh and R. Willett, "Regularized non-Gaussian image denoising," *arXiv preprint arXiv:1508.02971*, 2015.

[86] "GitHub repository for super-resolution low-light LIDAR," `https://github.com/photon-efficient-imaging/superresolution`.

[87] C.-K. Liang, T.-H. Lin, B.-Y. Wong, C. Liu, and H. H. Chen, "Programmable aperture photography: multiplexed light field acquisition," *ACM Transactions on Graphics (TOG)*, vol. 27, no. 3, pp. 55, 2008.

[88] A. Levin, R. Fergus, F. Durand, and W. T. Freeman, "Image and depth from a conventional camera with a coded aperture," *ACM Transactions on Graphics (TOG)*. ACM, 2007, vol. 26, p. 70.

[89] R. A. Stoltz, "Image capture with spatial light modulator and single-cell photosensor," May 18 1993, US Patent 5,212,555.

[90] D. Dudley, W. M. Duncan, and J. Slaughter, "Emerging digital micromirror device (DMD) applications," *Micromachining and Microfabrication*. International Society for Optics and Photonics, 2003, pp. 14–25.

[91] M. F. Duarte, M. A. Davenport, D. Takhar, J. N. Laska, T. Sun, K. F. Kelly, and R. G. Baraniuk, "Single-pixel Imaging via Compressive Sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 83–91, 2008.

[92] E. J. Candes, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics*, vol. 59, no. 8, pp. 1207–1223, 2006.

[93] P. A. Morris, R. S. Aspden, J. E. Bell, R. W. Boyd, and M. J. Padgett, "Imaging with a small number of photons," *Nat. Commun.*, vol. 6, 2015, doi: 10.1038/ncomms6913.

[94] C. Z. Mooney, *Monte Carlo Simulation*, vol. 116 of *Quantitative Applications in the Social Sciences*, Sage Publications, Thousand Oaks, CA, 1997.

[95] M. Pajovic and J. C. Preisig, "Performance analysis of the least squares based LTI channel identification algorithm using random matrix methods," *Proc. 49th Ann. Allerton Conf. Commun., Control, Comput.*, 2011, pp. 516–523.

[96] A. M. Tulino and S. Verdú, *Random Matrix Theory and Wireless Communications*, Now Publishers, Hanover, MA, 2004.

[97] D. C. Hoyle, "Accuracy of pseudo-inverse covariance learning—a random matrix theory analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 7, pp. 1470–1481, 2011.

[98] D. Guo and S. Verdú, "Randomly spread CDMA: asymptotics via statistical physics," *IEEE Trans. Inform. Theory.*, vol. 51, no. 6, pp. 1983–2010, June 2005.

[99] S. Rangan, A. Fletcher, and V. K. Goyal, "Asymptotic analysis of MAP estimation via the replica method and applications to compressed sensing," *IEEE Trans. Inform. Theory.*, vol. 58, no. 3, pp. 1902–1923, Mar. 2012.

[100] B. Schwarz, "LIDAR: mapping the world in 3D," *Nat. Photon.*, vol. 4, no. 7, pp. 429–430, 2010.

[101] Y. Y. Schechner, S. K. Nayar, and P. N. Belhumeur, "A theory of multiplexed illumination," *Proc. 9th IEEE Int. Conf. Comput. Vis.*, 2003, vol. 2, pp. 808–815.

[102] Y. Y. Schechner, S. K. Nayar, and P. N. Belhumeur, “Multiplexing for optimal lighting,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 8, pp. 1339–1354, 2007.

[103] N. Ratner, Y. Y. Schechner, and F. Goldberg, “Optimal multiplexed sensing: bounds, conditions and a graph theory link,” *Opt. Expr.*, vol. 15, no. 25, pp. 17072–17092, 2007.

[104] D. Shin, A. Kirmani, and V. K. Goyal, “Low-rate Poisson intensity estimation using multiplexed imaging,” *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Process.*, Vancouver, Canada, May 2013, pp. 1364–1368.

[105] J. H. Shapiro, “Computational ghost imaging,” *Phys. Rev. A*, vol. 78, no. 6, pp. 061802, 2008.

[106] V. A. Marchenko and L. A. Pastur, “Distribution of eigenvalues for some sets of random matrices,” *Matematicheskii Sbornik*, vol. 114, no. 4, pp. 507–536, 1967.

[107] I. M. Johnstone, “On the distribution of the largest eigenvalue in principal components analysis,” *Annals of statistics*, pp. 295–327, 2001.

[108] I. E. Telatar, “Capacity of multi-antenna Gaussian channels,” *European Trans. Telecommun.*, vol. 10, no. 6, pp. 585–595, Dec. 1999.

[109] S. S. Haykin, *Adaptive Filter Theory*, Pearson Higher Education, Upper Saddle River, NJ, fifth edition, 2013.

[110] “GitHub repository for single-pixel imaging analysis,” `https://github.com/photon-efficient-imaging/rmt-for-ghost`.

[111] “Rice single-pixel camera project,” `http://dsp.rice.edu/cscamera`.

[112] A. N. Tikhonov and V. Y. Arsenin, *Solutions of Ill-Posed Problems*, Winston, Washington, 1977.

[113] H.-J. Lin, P. Herman, and J. R. Lakowicz, “Fluorescence lifetime-resolved pH imaging of living cells,” *Cytometry Part A*, vol. 52, no. 2, pp. 77–89, 2003.

[114] I. V. Gopich and A. Szabo, “Theory of the energy transfer efficiency and fluorescence lifetime distribution in single-molecule FRET,” *Proceedings of the National Academy of Sciences*, vol. 109, no. 20, pp. 7747–7752, 2012.

[115] S. Padilla-Parra, N. Audugé, M. Coppey-Moisan, and M. Tramier, “Quantitative FRET analysis by fast acquisition time domain FLIM at high spatial resolution in living cells,” *Biophysical Journal*, vol. 95, no. 6, pp. 2976–2988, 2008.

[116] D. Elson, J. Requejo-Isidro, I. Munro, F. Reavell, J. Siegel, K. Suhling, P. Tadrous, R. Benninger, P. Lanigan, J. McGinty, et al., “Time-domain fluorescence lifetime imaging applied to biological tissue,” *Photochemical & Photobiological Sciences*, vol. 3, no. 8, pp. 795–801, 2004.

[117] G. I. Redford and R. M. Clegg, “Polar plot representation for frequency-domain analysis of fluorescence lifetimes,” *Journal of Fluorescence*, vol. 15, no. 5, pp. 805–815, 2005.

[118] J. R. Lakowicz and I. Gryczynski, “Frequency-domain fluorescence spectroscopy,” *Topics in Fluorescence Spectroscopy*, pp. 293–335. Springer, 1999.

[119] J. Enderlein and R. Erdmann, “Fast fitting of multi-exponential decay curves,” *Optics Communications*, vol. 134, no. 1, pp. 371–378, 1997.

[120] M. Köllner and J. Wolfrum, “How many photons are necessary for fluorescence-lifetime measurements?,” *Chemical Physics Letters*, vol. 200, no. 1, pp. 199–204, 1992.

[121] C.-W. Chang and M.-A. Mycek, “Enhancing precision in time-domain fluorescence lifetime imaging,” *Journal of Biomedical Optics*, vol. 15, no. 5, pp. 056013–056013, 2010.

[122] J. Kim, J. Seok, H. Lee, and M. Lee, “Penalized maximum likelihood estimation of lifetime and amplitude images from multi-exponentially decaying fluorescence signals,” *Opt. Expr.*, vol. 21, no. 17, pp. 20240–20253, 2013.

[123] G. Satat, B. Heshmat, C. Barsi, D. Raviv, O. Chen, M. G. Bawendi, and R. Raskar, "Locating and classifying fluorescent tags behind turbid layers using time-resolved inversion," *Nat. Commun.*, vol. 6, 2015.

[124] L.-Q. Li and L. M. Davis, "Single photon avalanche diode for single molecule detection," *Review of Scientific Instruments*, vol. 64, no. 6, pp. 1524–1529, 1993.

[125] F. Villa, B. Markovic, S. Bellisai, D. Bronzi, A. Tosi, F. Zappa, S. Tisa, D. Durini, S. Weyers, U. Paschen, and W. Brockherde, "SPAD smart pixel for time-of-flight and time-correlated single-photon counting measurements," *IEEE Photonics J.*, vol. 4, no. 3, pp. 795–804, June 2012.

[126] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume 1: Estimation theory*, vol. 1, Prentice Hall, New Jersey, 1993.