Photon-Efficient Computational Imaging with Single-Photon Avalanche Diode (SPAD) Arrays

by

Sarah Rumbley

S.B., Massachusetts Institute of Technology (2012)

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

Master of Engineering in Electrical Engineering and Computer Science

at the

Massachusetts Institute of Technology

September 2015

© Massachusetts Institute of Technology 2015. All rights reserved.

Author .....
Department of Electrical Engineering and Computer Science
September 8, 2015

Certified by .....
Jeffrey H. Shapiro
Julius A. Stratton Professor of Electrical Engineering
Thesis Supervisor
September 8, 2015

Accepted by .....
Christopher J. Terman
Chairman, Master of Engineering Thesis Committee

Photon-Efficient Computational Imaging with Single-Photon Avalanche Diode (SPAD) Arrays

by

Sarah Rumbley

Submitted to the Department of Electrical Engineering and Computer Science
on September 8, 2015, in partial fulfillment of the
requirements for the degree of
Master of Engineering in Electrical Engineering and Computer Science

Abstract

Single-photon avalanche diodes (SPADs) are highly sensitive photodetectors that enable LIDAR imaging at extremely low photon flux levels. While conventional image formation methods require hundreds or thousands of photon detections per pixel to suppress noise, a recent computational approach achieves comparable results when forming reflectivity and depth images from on the order of 1 photon detection per pixel. This method uses the statistics underlying photon detections, along with the assumption that depth and reflectivity are spatially correlated in natural scenes, to perform noise censoring and regularized maximum-likelihood estimation. We expand on this research by adapting the method for use with SPAD arrays, accounting for the spatial non-uniformity of imaging parameters and the effects of crosstalk. We develop statistical models that incorporate these non-idealities, and present a statistical method for censoring crosstalk detections. We show results that demonstrate the performance of our method on simulated data with a range of imaging parameters.

Thesis Supervisor: Jeffrey H. Shapiro
Title: Julius A. Stratton Professor of Electrical Engineering

Acknowledgments

I owe much gratitude to those who have advised, supported, and encouraged me during my studies and research over the past year.

I would like to thank my advisor, Prof. Jeffrey Shapiro, who has not only overseen my Master's research, but has also served as my academic advisor since my undergraduate years at MIT. At two major turns in my academic life, he has encouraged me to pursue my interests in engineering where I had little previous experience: first, when I switched my undergraduate major from music to electrical engineering and computer science, and second, when I returned to MIT after two years away from engineering. I am incredibly grateful to him for granting me the opportunity to be a part of his research group, and for teaching me to be a better engineer, researcher, and problem solver. For all of his caring support, dedicated mentorship, and practical life advice over the years, he will always have my deepest appreciation.

MIT Lincoln Laboratory's Advanced Concepts Committee provided funding for my research, and I am very grateful to them for having made my Master's studies possible, and for the opportunity to participate in this area of research.

I would also like to thank Dr. Dale Fried for generously and eagerly being available to share his expertise on LIDAR systems and SPAD arrays. His answers to my many questions greatly contributed to my understanding how SPAD arrays work on a physical level, which was essential to developing accurate statistical models. His years of practical experience with LIDAR helped me to ensure that my models and simulations were representative of real systems.

From my research group, I would like to thank Dongeek Shin and Dheera Ventrakaman for patiently familiarizing me with their research as the foundation for mine, and for providing me with their data and code to get started. Their thorough understanding of the project's technical aspects and their enthusiasm about their research were a great help and a motivating example to me. I was also fortunate to have two very pleasant, good-natured officemates: Feihu Xu, who was always willing to offer helpful technical input and ideas, and

Jane Heyes, whose companionship, sense of humor, and natural talent for pep talks made our office a much better place.

Apart from the academic community, I am very grateful to all of my family and friends whose encouragement and prayers were a much-appreciated support from day to day. A special thanks goes to Stephen and Rebecca Bruso—much of my work leading to this thesis took place at their dining table and was fueled by their endless supply of tea, coffee, and moral support.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Traditional light detection and ranging (LIDAR) systems must detect hundreds to thousands of back-reflected photons at each pixel to obtain accurate reflectivity and depth images [1]. This requirement places a lower limit on the size, weight, and power (SWaP) of LIDAR systems. One line of advancement for LIDAR is to develop systems with higher photon efficiency—that is, systems that require less back-reflected laser light to form reliable images. Higher photon efficiency means that imaging systems can use lower laser power, or can capture images more quickly or from longer distances. Such characteristics have pertinent applications in airborne and satellite systems [1, 2], where SWaP is at a premium, or in biological imaging [3], where low flux is needed to protect delicate organisms.

On the hardware side, efforts to increase photon efficiency have led to the use of highly sensitive photodetectors known as Geiger-mode avalanche photodiodes (APDs) or single-photon avalanche diodes (SPADs) [4]. These devices can detect individual photons and record their arrival times with ~ 10 ps precision. In recent years, SPAD technology has progressed from single-pixel detectors to detector arrays, enabling data collection at thousands of pixels in parallel.

Over the past several years, photon-counting LIDAR systems have been primarily developed for depth imaging, and most require at least tens of photon detections per pixel using conventional image formation methods [5]. At extremely low photon flux levels, the

data is corrupted by the Poisson noise inherent in the photon detection process, as well as background noise from ambient light and internal noise in the detector hardware. These high noise levels are the reason that traditional imaging techniques rely on collecting a large number of photon detections. We seek to push the limits of photon efficiency through better computational processing of the noisy data.

The novel computational imaging frameworks developed by Kirmani *et al.* [6] and Shin *et al.* [7] use physically correct statistical models and spatial regularization to construct reflectivity and depth images from data with a mean photon count on the order of one photon detection per pixel (ppp). Their frameworks have been experimentally tested with an imaging setup that uses focused-beam illumination and a single-pixel SPAD detector to raster scan the entire scene, as in Figure 1-1. In this setup, the important system parameters that factor into the statistical models are the laser beam's intensity and the detector's quantum efficiency and dark count rate. We extend the framework from [7] to process data collected from an array-based system that uses floodlight illumination and a SPAD array. In an array-based setup, the parameters used in the statistical model generally vary from pixel to pixel—the laser beam's intensity profile across the array's field of view is non-uniform, and there is some variation in the quantum efficiencies and dark count rates of the individual detector pixels. Array data is also affected by crosstalk, a phenomenon in which an avalanche at one pixel triggers spurious detections at other pixels in the array. These non-idealities affect the photon detection data, and must be accounted for in order to accurately estimate scene reflectivity and depth from photon-limited data. In this thesis, we present our developments in adapting the computational framework from [7] to account for the non-idealities present in an array-based LIDAR system. We also conduct a more detailed analysis of the performance of this framework to better understand how various imaging parameters and scene characteristics affect the results.

The remainder of Chapter 1 will describe prior work in both conventional imaging methods and the photon-efficient computational framework upon which our research builds. In Chapter 2, we will discuss the non-ideal parameters in an array-based LIDAR system, and

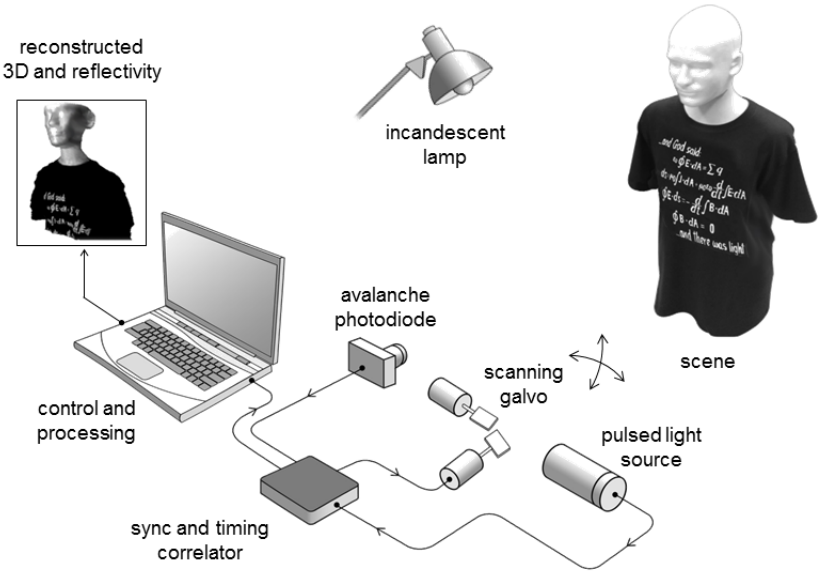


Figure 1-1: Imaging setup used in [6] and [7].

develop our modified image formation algorithm. In Chapter 3, we will describe our methods for simulating photon detection data. In Chapter 4, we will present our results and analyze our algorithm's performance on simulated data with a range of imaging parameters. In Chapter 5, we will conclude by discussing the significance of our research and suggesting future work to build upon our results.

1.1 Prior Work

1.1.1 Operating Principles and Parameters of SPAD Detectors

A SPAD detector operates in Geiger mode, in which an avalanche photodiode is reverse-biased past the breakdown voltage. At such a large reverse bias, the electric field in the active detector area is so strong that an incident photon can trigger a self-sustaining avalanche breakdown current by impact ionization. This avalanche current acts as a very large gain, such that the current induced by the absorption of a single photon is large enough to be detected and timestamped by the accompanying circuitry. The probability that an incident photon triggers a detectable avalanche current is referred to as the detector's *quantum effi-*

ciency, or *photon detection efficiency*. The amount of uncertainty in the registered detection times is termed *timing jitter*, and is usually on the order of 10-100 ps. Internal noise in the detector causes *dark counts*, or false detections, primarily due to avalanches triggered by thermally-generated carriers [8].

Because the avalanche breakdown is self-sustaining, it renders the detector inactive until it is quenched and re-armed after some interval of time (the *reset* or *dead time*). When time-correlated with a pulsed laser, the detector registers either no detections or a single detection (but not multiple detections) for each laser pulse. We refer to the fact that a detection will prevent any later detections in the same pulse frame as a *blocking effect*.

1.1.2 Conventional Imaging Methods

Reflectivity

Reflectivity images are conventionally formed by constructing a histogram of the amount of light received at each pixel over a fixed acquisition time. Reflectivity is generally treated as a relative measurement normalized over the image. This histogram method is computationally straightforward, but requires hundreds or thousands of photon detections to suppress noise and obtain an adequately large dynamic range.

Depth

Several methods exist for forming depth images from LIDAR data, depending on the hardware configuration and the type of data acquired. Of these methods, pulsed time-of-flight (TOF) depth imaging has the highest photon efficiency. In pulsed TOF imaging, the photodetector records the arrival times of detected photons relative to the most recent laser pulse, and the depth is calculated based on the photons' round-trip times to the scene and back. Photon-counting detectors like SPADs are well-suited to pulsed TOF imaging, as they have very fine timing resolution and can be time-correlated with a pulsed laser. Currently, the most photon-efficient systems use photon-counting detectors with pulsed TOF imaging, achieving good results from as few as tens of photon detections per pixel.

Sources of noise in the pulsed TOF method include photons from ambient light, false detections from dark current, and crosstalk in a detector array. Theoretically, if this noise were absent, a very accurate depth image could be reconstructed from just one photon detection at each pixel, with accuracy affected only by the duration of the laser pulse and the timing jitter of the detector. The combined uncertainty of these two factors generally corresponds to $\sim$ cm depth resolution. We will show in Chapter 4 that we can use statistical noise suppression methods to approach this hypothetical noiseless scenario, thereby increasing the accuracy of our depth estimate in photon-starved conditions.

1.1.3 Photon-Efficient Computational Framework

The computational frameworks presented in [6] and [7] were developed with the goal of maximizing the photon efficiency of LIDAR systems. Both have been demonstrated to successfully construct reflectivity and depth images from an average of 1-2 photon detections per pixel with significant levels of background noise, with results comparable to or better than conventional methods that require much higher photon counts. They achieve such high photon efficiency by leveraging two key concepts: 1) accurate probabilistic models of the physics underlying photon detections, and 2) the assumption that reflectivity and depth are each spatially correlated in natural scenes.

Here we will briefly develop the probabilistic models used in [7] for a single-pixel SPAD detector, and show how they are used to estimate scene reflectivity and depth.

Probabilistic model for number of photon detections

When collecting data over a fixed dwell time (as in [7]), reflectivity can be estimated from the number of photons detected at each pixel. Photon counting is an inhomogeneous Poisson process with a rate function that comprises photon detections from back-reflected signal, ambient light, and dark counts. For a laser pulse transmitted at $t = 0$, the rate function for

photon detections at a pixel (i, j) is:

$$\lambda_{i,j}(t) = \eta[\alpha_{i,j} S s(t - 2z_{i,j}/c) + b] + d, \text{ for } 0 \leq t \leq T_r, \tag{1.1}$$

where we have defined variables as follows:

T_r : the pulse repetition period (s)

η : the detector's quantum efficiency, $0 \leq \eta \leq 1$

$\alpha_{i,j}$: the scene reflectivity at pixel (i, j), $0 \leq \alpha_{i,j} \leq 1$

S : the expected number of signal photons from one laser pulse that are back-reflected from a unity-reflectivity pixel and arrive at the detector

$s(t)$: the laser pulse's photon flux (Hz), normalized to satisfy $\int_0^{T_r} s(t)dt = 1$

$z_{i,j}$: the scene depth at pixel (i, j)

b : the expected rate (Hz) at which photons from ambient light at the operating wavelength arrive at the detector

d : the detector's dark count rate (Hz).

We can integrate the rate function to obtain the expected number of photon detections in one pulse frame:

$$\int_0^{T_r} \lambda_{i,j}(t)dt = \eta \alpha_{i,j} S + B, \tag{1.2}$$

where we have defined B as $(\eta b + d)T_r$, or the expected number of detections due to ambient light and dark counts, which we will collectively refer to as background detections. From the Poisson distribution, we can use this rate function to write the probability of no detections at pixel (i, j) in one pulse frame:

$$P_0(i, j) = e^{-(\eta \alpha_{i,j} S + B)}. \tag{1.3}$$

In each pulse frame, the SPAD will report either no detections or one detection. Thus, under the low-flux assumption, at every pixel (i, j) we can model each pulse frame as a Bernoulli trial, where the probability of success is $1 - P_0(i, j)$. Over N_P pulses at a pixel (i, j), the number of photon detections $K_{i,j}$ is then a binomial random variable:

$$\Pr[K_{i,j} = k; \alpha_{i,j}] = \binom{N_P}{k} P_0(\alpha_{i,j})^{N_P - k} \left[1 - P_0(\alpha_{i,j})\right]^k. \tag{1.4}$$

Probabilistic model for photon detection times

At each pixel (i, j), detection times are treated as independent and identically-distributed continuous random variables $T_{i,j} \in [0, T_r]$. Detection times for photons from back-reflected laser light (i.e., signal photons) are distributed according to the photon flux of the laser pulse, $s(t)$:

$$f_{T_{i,j}}(t; z_{i,j} | \text{detection is signal}) = s(t - 2z_{i,j}/c), \text{ for } t \in [0, T_r] \tag{1.5}$$

Detection times for background detections are uniformly distributed over time from 0 to T_r, where these times are relative to the laser pulse:

$$f_{T_{i,j}}(t | \text{detection is noise}) = \frac{1}{T_r}, \text{ for } t \in [0, T_r]. \tag{1.6}$$

Combining these models for signal and background detections gives a mixture distribution, where the low-flux approximation allows us to ignore blocking effects:

$$\begin{aligned} f_{T_{i,j}}(t; \alpha_{i,j}, z_{i,j}) &= \Pr[\text{detection is signal}; \alpha_{i,j}] \cdot f_{T_{i,j}}(t; z_{i,j} | \text{detection is signal}) \\ &\quad + \Pr[\text{detection is noise}; \alpha_{i,j}] \cdot f_{T_{i,j}}(t | \text{detection is noise}) \\ &= \frac{\eta \alpha_{i,j} S}{\eta \alpha_{i,j} S + B} s(t - 2z_{i,j}/c) + \frac{B}{\eta \alpha_{i,j} S + B} \left(\frac{1}{T_r}\right), \text{ for } t \in [0, T_r]. \end{aligned} \tag{1.7}$$

Image formation

Our detection dataset consists of a set of $k_{i,j}$ detection times at each pixel (i, j): $\{t_{i,j}^{(\ell)}\}_{\ell=1}^{k_{i,j}}$. Using the probabilistic models for the random variables $\{K_{i,j}\}$ and the $\{T_{i,j}^{(\ell)}\}$, the image

formation algorithm from [7] performs maximum-likelihood estimation with spatial regularization to form reflectivity and depth images. The optimization problem consists of a data likelihood term and a regularization term:

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \left[L(\boldsymbol{\theta}; \mathbf{y}) + \tau_{\boldsymbol{\theta}} \text{ pen}(\boldsymbol{\theta}) \right], \tag{1.8}$$

where $L(\boldsymbol{\theta}; \mathbf{y})$ is the negative log-likelihood of the estimated parameter $\boldsymbol{\theta}$ (either reflectivity or depth) given the observed data $\mathbf{y}$. These variables $\boldsymbol{\theta}$ and $\mathbf{y}$ are 2-D matrices over all the pixels in the image, and we can also write the likelihood term in a pixelwise representation for an $N_1 \times N_2$-pixel image:

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \left[\sum_{i=1}^{N_1} \sum_{j=1}^{N_2} L(\theta_{i,j}; y_{i,j}) + \tau_{\boldsymbol{\theta}} \text{ pen}(\boldsymbol{\theta}) \right]. \tag{1.9}$$

The regularization function $\text{pen}(\boldsymbol{\theta})$ penalizes non-smoothness across pixels in $\boldsymbol{\theta}$, and the constant $\tau_{\boldsymbol{\theta}}$ controls the relative strength of this smoothing term. In [6] and [7], using the total variation (TV) norm as a penalty function was found to give the best results in experiments.

The image formation algorithm is comprised of three steps:

1. **Reflectivity estimation.** Reflectivity is estimated based on the pixelwise photon counts $k_{i,j}$:

$$\hat{\boldsymbol{\alpha}} = \arg\min_{\boldsymbol{\alpha}} \left[\sum_{i=1}^{N_1} \sum_{j=1}^{N_2} L(\alpha_{i,j}; k_{i,j}) + \tau_{\boldsymbol{\alpha}} \text{ pen}(\boldsymbol{\alpha}) \right]. \tag{1.10}$$

   From Eq. (1.4), we have the following negative log-likelihood:

$$\begin{aligned} L(\alpha_{i,j}; k_{i,j}) &= -\log \binom{N_P}{k_{i,j}} - (N_P - k_{i,j}) \log P_0(\alpha_{i,j}) - k_{i,j} \log[1 - P_0(\alpha_{i,j})] \\ &= (N_P - k_{i,j})(\eta \alpha_{i,j} S + B) - k_{i,j} \log(1 - e^{-(\eta \alpha_{i,j} S + B)}) \\ &\approx (N_P - k_{i,j})(\eta \alpha_{i,j} S + B) - k_{i,j} \log(\eta \alpha_{i,j} S + B), \end{aligned} \tag{1.11}$$

where log is the natural logarithm, we have dropped the binomial coefficient term that is independent of $\alpha_{i,j}$, and we have simplified the final term using a first-order Taylor-series approximation for low-flux conditions.

This negative log-likelihood is a convex function of $\alpha_{i,j}$, which we can prove by showing that its second derivative with respect to $\alpha_{i,j}$ is always positive (in Appendix A, we include the analogous proof for our modified statistical model). When the penalty function is also convex, the convexity of Eq. (1.10) makes it amenable to efficient implementations of regularized convex optimization. In particular, [7] uses an implementation called Sparse Poisson Intensity Reconstruction Algorithm (SPIRAL) [9].

2. **Noise censoring.** We can estimate depth from the photon detection times, which are distributed according to Eq. (1.7). Unfortunately, the negative log-likelihood of Eq. (1.7) is non-convex because of the background noise term, it cannot be used in convex optimization. Therefore, we perform a hypothesis test to censor background detections before estimating depth. Using our assumption that depth is spatially correlated, we compute the rank-ordered mean (ROM) [10] of detection times in the local neighborhood of each pixel (i, j) to obtain a value $t_{i,j}^{ROM}$ that we expect to be close to the target signal value $2z_{i,j}/c$. We can reasonably infer that detection times $t_{i,j}^{(l)}$ that are far away from $t_{i,j}^{ROM}$ are most likely due to background noise. These detections are excluded from the set of observed data used for depth estimation. We use the following threshold comparison to obtain a set of detections that are presumably due to signal:

$$U_{i,j} = \left\{ \ell : |t_{i,j}^{(\ell)} - t_{i,j}^{ROM}| < 2T_p \left(\frac{B}{\eta \hat{\alpha}_{i,j} S + B} \right), 1 \leq \ell \leq k_{i,j} \right\}, \tag{1.12}$$

where T_p is the full-width to half-maximum (FWHM) laser pulse duration.

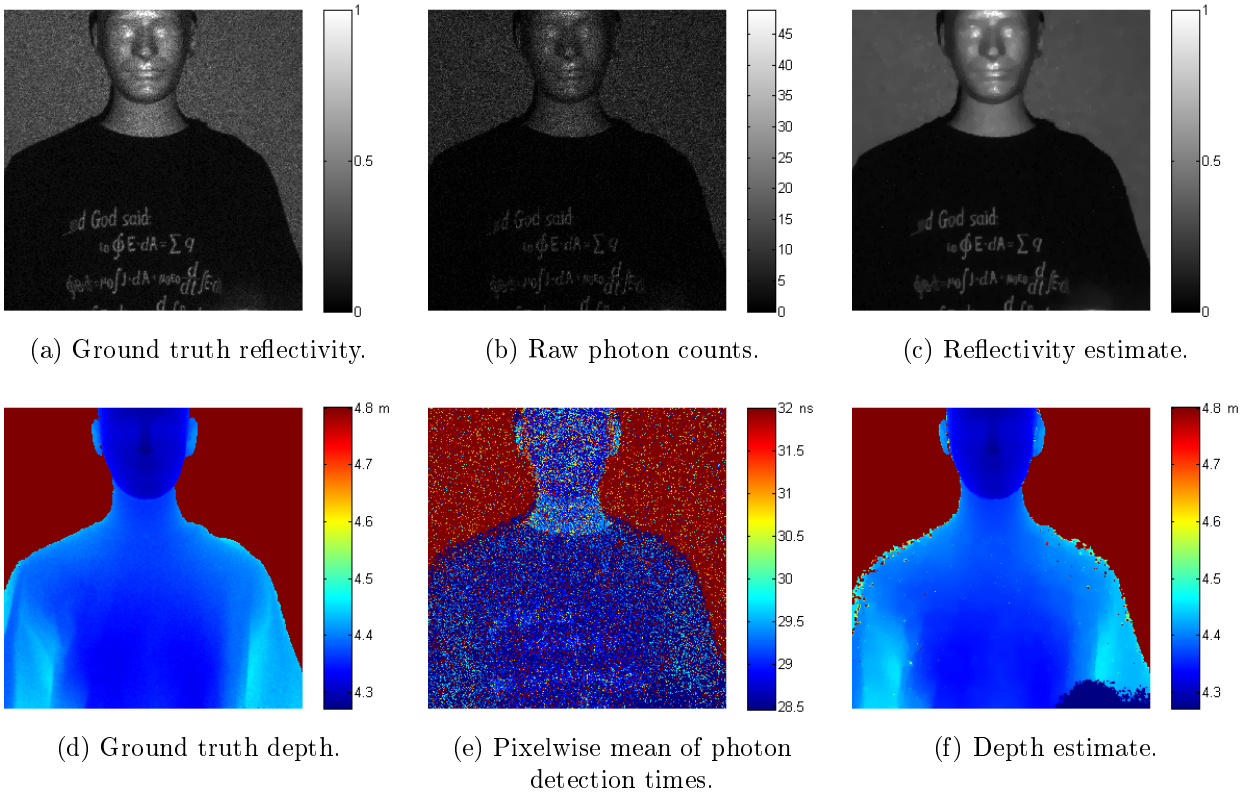
3. **Depth estimation.** After noise censoring, we can assume that the remaining photon detections are due to back-reflected signal, and their arrival times can be used to

estimate depth. Thus, the optimization problem for depth is:

$$\hat{\mathbf{z}} = \arg\min_{\mathbf{z}} \left[\sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \left(- \sum_{\ell \in U_{i,j}} \log s(t_{i,j}^{(\ell)} - \frac{2z_{i,j}}{c}) \right) + \tau_{\mathbf{z}} \, \text{pen}(\mathbf{z}) \right]. \tag{1.13}$$

Assuming that the negative log of the laser pulse shape $s(t)$ is convex, the depth optimization problem is convex and can be solved efficiently using SPIRAL. Laser pulse shapes can commonly be modeled by a gamma distribution, which is known to be logarithmically concave, so its negative log is therefore convex.

Figure 1-2 shows an example of raw data and the reflectivity and depth results achieved in [7], with an average of 4 photon detections per pixel.



(a) Ground truth reflectivity. (b) Raw photon counts. (c) Reflectivity estimate.

(d) Ground truth depth. (e) Pixelwise mean of photon detection times. (f) Depth estimate.

Figure 1-2: Results of reflectivity and depth estimation using photon-efficient computational framework [7].

With this imaging framework as a foundation, our first task, which we will undertake in Chapter 2, is to characterize the non-idealities that we must incorporate into the imaging framework to more accurately represent the physical statistics of an array-based LIDAR system.

Chapter 2

Image Formation Algorithm for SPAD Arrays

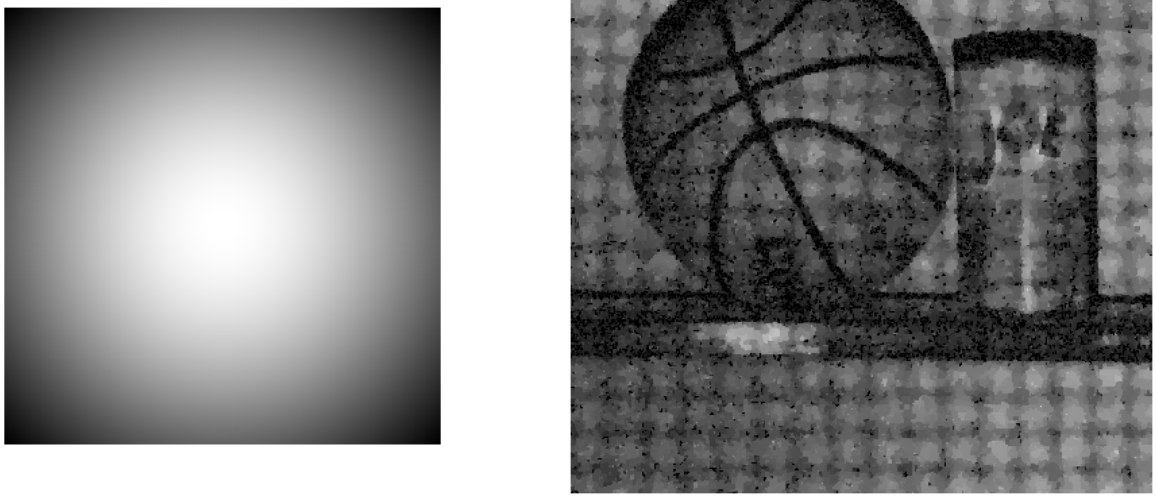
Our modified image formation method is designed for a LIDAR configuration that illuminates a scene with a floodlight laser source and uses a SPAD array to detect photons returned from the illuminated region. We assume that an $M_1 \times M_2$ detector array will scan over a scene without overlap to produce a larger image of size $N_1 \times N_2$. We will distinguish between pixels in the detector array, denoted by $(m, n) \in [1, M_1] \times [1, M_2]$, and pixels in the image, denoted by $(i, j) \in [1, N_1] \times [1, N_2]$.

2.1 Non-Uniform Parameters in Array Setup

2.1.1 Laser Beam Intensity Profile

In an array setup, the laser beam is expanded to illuminate the detector array's entire field of view. Lasers most commonly have a beam profile (spatial cross-section of beam intensity) that is approximately Gaussian, with the illumination brightest at the center of the beam and tapering off radially (Figure 2-1(a)). This 2-D profile can be written as a matrix $\mathbf{S} = \{S_{m,n}\}_{m,n=1}^{m=M_1,n=M_2}$ over all pixels in the array, where, as in Eq. (1.3), $S_{m,n}$ is the

expected number of back-reflected signal photons received at detector pixel (m, n) from a unity-reflectivity object.



(a) Example of Gaussian beam profile.

(b) Reflectivity image with artifacts from beam falloff: 32×32 array scanned over 512×512 image.

Figure 2-1: Laser beam falloff.

The size of the beam is a controllable parameter, and it determines how much $\mathbf{S}$ falls off toward the edges of the array. A larger beam will result in $\mathbf{S}$ being more uniform over the array, but at the cost of wasted photons that fall outside the detector's imaging area. We will define the radius of a laser beam to be the radius at which the intensity is $1-1/e^2 \approx 0.86$ times its peak value. Within this radius, the beam intensity is fairly uniform, and the reflectivity estimation method from [7] performs adequately without accounting for the small amount of non-uniformity. However, if the beam is expanded such that the array's entire imaging area lies within the $1 - 1/e^2$ beam radius, then only $\sim 9\%$ of the laser power falls within the array's field of view. For the same laser power, a narrower beam will result in higher photon counts throughout the array, but the beam profile will be highly non-uniform across the array's imaging area. The number of back-reflected signal photons that are detected at a given pixel is related to beam intensity according to Eq. (1.4). Assuming roughly uniform

scene reflectivity over an illuminated region, we can expect more signal detections at the center of the array and fewer as the radial distance from the center of the beam increases. Without any processing aimed at compensating for beam falloff, the estimated reflectivity image will contain artifacts as shown in Figure 2-1(b). Because we are aiming for photon efficiency, we would like to accommodate some amount of beam falloff with minimal loss of image quality. In determining the beam size and laser power level, we must also be concerned with the decrease in signal-to-background ratio (SBR) as the beam falls off toward the edges and corners of the array's imaging area. Because SBR has a significant impact on our method's performance, an optimal choice of beam size for our method will maximize the beam intensity at the corners of the array's imaging area, where the intensity is lowest and will therefore give the lowest SBR. We calculated this optimal size to be when the standard deviation of the 2-D Gaussian function, σ, is $M/2$, which places $\sim 47\%$ of the laser power within the array's imaging area. A detailed derivation of this optimal beam size is included in Appendix B.

2.1.2 Quantum Efficiency

In a SPAD array, quantum efficiency varies from pixel to pixel by some small amount [11, 12]. Thus, quantum efficiency has a spatial dependency and we can denote its value at each pixel (m, n) as $\eta_{m,n}$. The distribution of an array's pixel-wise quantum efficiencies is approximately Gaussian (Figure 2-2).

2.1.3 Dark Count Rate

Dark count rate can vary throughout a SPAD array, and can also vary significantly from one array to another (Figure 2-3(a)). In general, the distribution of an array's pixel-wise dark count rates is approximately Gaussian, with a small number of outliers termed "hot pixels" that have an unusually high dark count rate (Figure 2-3(b)). At low signal return levels, the data at hot pixels may be overwhelmed by noise and therefore unreliable for image formation.

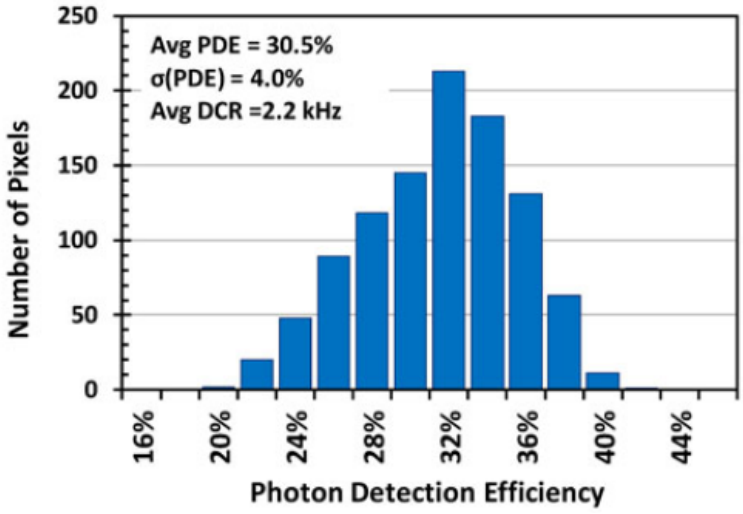


Figure 2-2: Distribution of pixel-wise quantum efficiencies in a SPAD array [12].

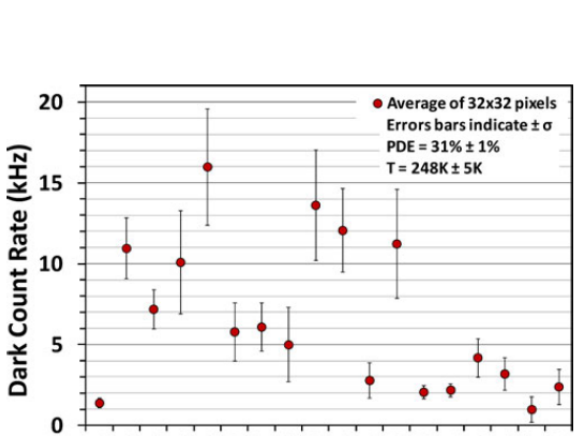
In all that follows, we shall presume, without explicitly noting it, that all hot-pixel data has been excluded from likelihood calculations at the outset.

2.1.4 Dead Pixels

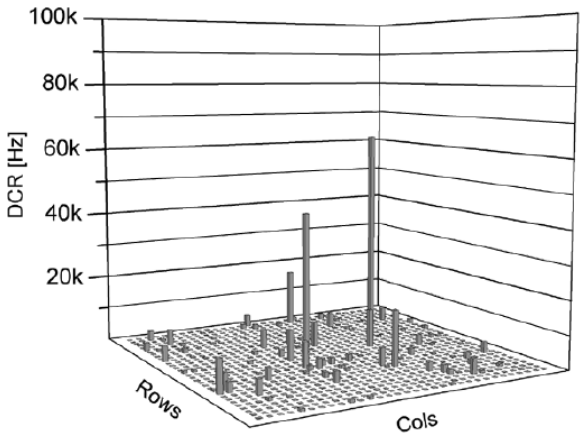
A SPAD array may have a small number of "dead pixels" that never return any detections. These dead pixels can be considered to have a quantum efficiency of 0, and therefore have no contribution to the data likelihood term in our reflectivity estimation. As we will do with hot pixels, we will exclude all dead pixels from our likelihood calculations, without explicit notation to that effect.

2.2 Crosstalk

Optical crosstalk is a phenomenon that generates signal-dependent noise and behaves as a sort of spatio-temporal point spread function. When an incident photon triggers an avalanche current at a pixel in the detector, high-energy "hot carriers" in the avalanche generate secondary photons that are internally reflected within the array and can, in turn, cause avalanches at other pixels. We can model this phenomenon as an independent Pois-



(a) Plot of mean and standard deviation of dark count rate in 18 different 32 × 32 SPAD arrays [12].



(b) Plot of dark count rate at each pixel in a SPAD array [13].

Figure 2-3: Non-uniformity of pixel-wise dark count rates in a SPAD array.

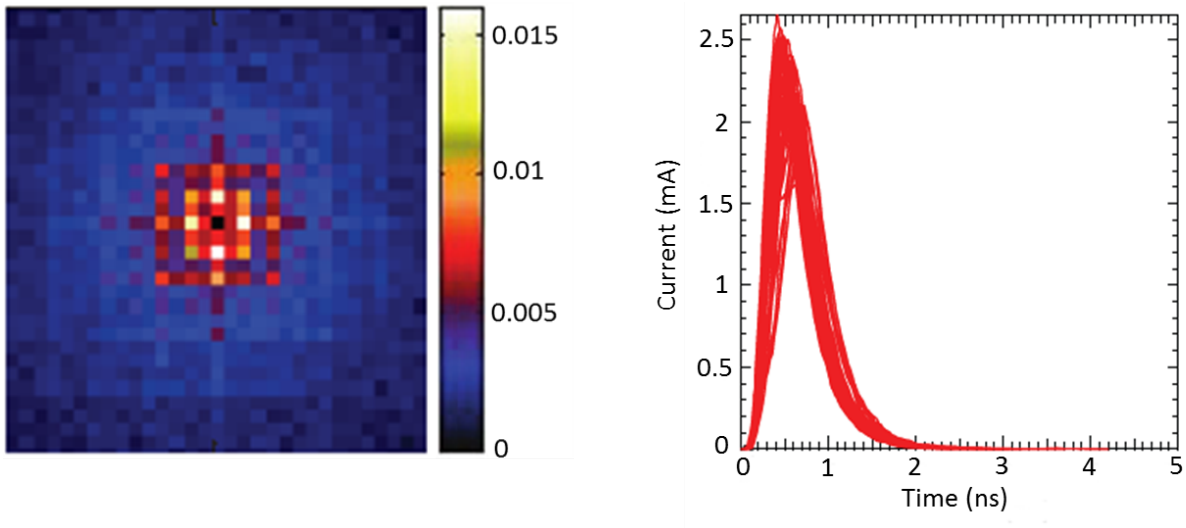
son process that is initiated when an avalanche releases secondary photons into the array. Given an initial detection at pixel (m, n) at time t, the rate at which secondary photons arrive at another pixel (m', n') at a time $t' > t$ is a function of both the spatial offset $(\Delta m, \Delta n) = (m' - m, n' - n)$ and the temporal delay $\Delta t = t' - t$. We can write the rate function as $\lambda_{sp}(\Delta m, \Delta n, \Delta t)$, and we will see shortly that the spatial and temporal components of this rate function are separable under a reasonable approximation.

The spatial component of λ_{sp} depends on the geometry and fabrication details of the device. In general, the spatial component looks like a 2-D Gaussian, but there are certain hardware techniques designed to reduce crosstalk that result in different spatial patterns, as in Figure 2-4(a). According to [14], it is reasonable to assume that this distribution over spatial offsets is the same for every pixel in the array. Thus, we can write $\beta(\Delta m, \Delta n)$ as the expected number of secondary photons generated by an avalanche at a pixel (m, n) that travel to another detector pixel $(m + \Delta m, n + \Delta n)$. Models for $\beta(\Delta m, \Delta n)$ are described in [14] and [15].

The temporal component of λ_{sp} is determined by the delay between the arrival of a photon that triggers an initial avalanche and the arrival of a resulting secondary photon at a different pixel. Secondary photons are generated at a rate proportional to the temporal

response of the avalanche current, shown in Figure 2-4(b). According to this rate function, a secondary photon is generated at some time after the arrival of the initial photon. The secondary photon then travels through the device to reach another pixel. The total time delay is dominated by the delay in generating the secondary photon, compared to which the photon's travel time is negligible ($\sim$ fs) because the device is very small. Thus, the temporal component of λ_{sp} is approximately proportional to the rate of secondary photon generation, which we will write as $\gamma(\Delta t)$, normalized to integrate to 1. With this approximation, the temporal delay is independent of the spatial offset, and we can therefore write the rate of secondary photon arrivals as:

$$\lambda_{sp}(\Delta m, \Delta n, \Delta t) = \beta(\Delta m, \Delta n)\gamma(\Delta t). \tag{2.1}$$



(a) Spatial distribution of crosstalk [14]. (Colorbar represents probability of crosstalk given primary detection at center pixel.)

(b) Several simulations of avalanche current response [14], which determines temporal distribution of crosstalk.

Figure 2-4: Spatial and temporal crosstalk parameters.

In some arrays and under some operating conditions, crosstalk is negligible [13, 16]. In other cases, however, crosstalk can be a significant source of noise [14], so a photon-efficient system must be able to mitigate the effects of this noise.

The parameters of a SPAD detector depend on many factors related to both the fabrication and operation of the device, especially temperature and bias voltage. In practice, SPAD detectors are characterized by measuring these parameters empirically for a specific device in specific operating conditions. [12], [14], and [15] describe methods for empirically measuring the spatial and temporal distributions of crosstalk within an array. For our purposes, we assume that the beam profile, quantum efficiency, dark count rate, and crosstalk probability distribution are known for a particular LIDAR configuration.

2.3 Image Formation Algorithm Without Crosstalk

We use an image formation algorithm based on the one presented in [7] and outlined in Section 1.1.3. If crosstalk is negligible, we need only to modify the statistical models of photon counts and detection times to accommodate pixel-wise non-uniformities in beam intensity, quantum efficiency, and dark count rate. These modified models can be substituted into the data likelihood calculations and noise censoring threshold without any changes to the image formation algorithm itself.

In the most general case, we can associate every combination of image pixel (i, j) and pulse frame f with its corresponding pixel (m, n) in the detector array. This means that in pulse frame f, the imaging system was positioned such that detector pixel (m, n) received photons from the part of the scene corresponding to pixel (i, j) in the image. The detector pixel (m, n) maps to a beam intensity $S_{m,n}$, quantum efficiency $\eta_{m,n}$, and dark count rate $d_{m,n}$, which are used to compute the data likelihood at each image pixel for reflectivity estimation and noise censoring. This scheme is able to accommodate any scanning pattern.

For notational simplicity, we will proceed with the assumption that the system scans over the scene without overlap and with a uniform dwell time at every image pixel. In this scanning pattern, each image pixel corresponds to only one detector pixel for all pulse frames, and we no longer need the notational distinction between image pixels and detector pixels. Thus, we have $S_{i,j}$, $\eta_{i,j}$, and $d_{i,j}$ for each image pixel (i, j). We can represent our data as a 3-D matrix $\{t_{i,j,f}\}$ of size $N_1 \times N_2 \times N_P$, where the value at (i, j, f) is either a detection

time in the interval $[0, T_r]$, or ∞ to indicate no detection. At each element, the probability that there is no detection follows from Eq. (2.2):

$$P_0(i,j) = e^{-(\eta_{i,j}\alpha_{i,j}S_{i,j}+B_{i,j})}, \tag{2.2}$$

where $B_{i,j} = (\eta_{i,j}b + d_{i,j})T_r$. In our present work, we have assumed, like [6] and [7], that the photon flux from ambient light is spatially uniform. In real imaging applications, this is generally not the case, and recent work by Shin *et al.* addresses non-uniform ambient light flux [17].

From Eq. (2.2), the modified negative log-likelihood for reflectivity estimation is then:

$$L(\alpha_{i,j}; k_{i,j}) = (N_P - k_{i,j})(\eta_{i,j}\alpha_{i,j}S_{i,j} + B_{i,j}) - k_{i,j}\log(\eta_{i,j}\alpha_{i,j}S_{i,j} + B_{i,j}). \tag{2.3}$$

Similarly, the distribution of detection times follows from Eq. (1.7):

$$f_{T_{i,j}}(t; \alpha_{i,j}, z_{i,j}) = \frac{\eta_{i,j}\alpha_{i,j}S_{i,j}}{\eta_{i,j}\alpha_{i,j}S_{i,j} + B_{i,j}} s_{i,j}(t - 2z_{i,j}/c) + \frac{B_{i,j}}{\eta_{i,j}\alpha_{i,j}S_{i,j} + B_{i,j}} \left(\frac{1}{T_r}\right). \tag{2.4}$$

We note that in the no-crosstalk case, these statistics are independent of the pulse frame. Furthermore, we show in Appendix A that the negative log-likelihood equations are still convex, and we can therefore use the same efficient implementation of convex optimization as in [6] and [7].

2.4 Image Formation Algorithm With Crosstalk

If crosstalk is significant, we must take additional steps to account for crosstalk in our statistical models and to censor crosstalk detections. We will first construct probabilistic models that incorporate crosstalk, then propose a censoring method based on the likelihood that a detection is crosstalk.

To avoid notational complexity, we will maintain our assumption of a simple non-overlapping scanning pattern. When considering crosstalk, it is useful to define a term *array-frame* to

denote the $M_1 \times M_2$ set of pixels captured by the array in one pulse frame. Crosstalk effects are confined within a single array-frame—that is, any crosstalk detections triggered by an initial detection will be within the same array-frame as the initial detection. We also note that a crosstalk detection occurs with some time delay *after* the initial detection. Thus, to compute the probability of a crosstalk detection at a pixel (i, j) at some time t, we must add up the contributions of all previous detections in the same array-frame that could cause a crosstalk detection at (i, j) at time t.

2.4.1 Reflectivity Estimation with Crosstalk

Crosstalk is an event-triggered phenomenon. We can say that every avalanche initiates an independent Poisson process when it releases secondary photons into the detector array. At the time of data processing, we have complete knowledge of detection events. Based on this knowledge, we can construct the overall Poisson rate function including the Poisson processes due to crosstalk. The rate function for crosstalk detections at a pixel (i, j) in pulse frame f is the rate at which secondary photons arrive at (i, j), times the probability that an incident photon at (i, j) triggers a detection:

$$\begin{aligned} \lambda^{CT}_{i,j,f}(t) &= \eta_{i,j} \sum_{\substack{(i',j') \in \\ A(i,j,f)}} \lambda_{sp}(i - i', j - j', t - t_{i',j',f}) \\ &= \eta_{i,j} \sum_{\substack{(i',j') \\ \in A(i,j,f)}} \beta(i - i', j - j')\gamma(t - t_{i',j',f}), \end{aligned} \tag{2.5}$$

where $A(i, j, f)$ is the $M_1 \times M_2$ set of image pixel indices in the same array-frame as (i, j, f). This probability contributes to the overall Poisson rate function:

$$\lambda_{i,j,f}(t) = \eta_{i,j} \left(\alpha_{i,j} S_{i,j} s(t - \frac{2z_{i,j}}{c}) + b + \sum_{\substack{(i',j') \in \\ A(i,j,f)}} \beta(i - i', j - j')\gamma(t - t_{i',j',f}) \right) + d_{i,j}. \tag{2.6}$$

Integrated over the pulse frame, we have:

$$\int_0^{T_r} \lambda_{i,j,f}(t)dt = \eta_{i,j}\alpha_{i,j}S_{i,j} + B_{i,j} + \eta_{i,j} \sum_{\substack{(i',j')\in \\ A(i,j,f)}} \beta(i-i', j-j'). \tag{2.7}$$

From this integral, we can write the probability that pixel (i, j) has no detections in pulse frame f:

$$P_0(i,j,f) = \exp\left[-\left(\eta_{i,j}\alpha_{i,j}S_{i,j} + B_{i,j} + \eta_{i,j} \sum_{\substack{(i',j')\in \\ A(i,j,f)}} \beta(i-i', j-j') \right) \right]. \tag{2.8}$$

Let $\mathbb{I}_{i,j,f}$ be an indicator random variable that is 1 if pixel (i, j) has a detection in frame f, and 0 otherwise. Then, the data likelihood for reflectivity is:

$$\begin{aligned} \Pr[\{\mathbb{I}_{i,j,f}\}_{f=1}^{N_P}; \alpha_{i,j}] &= \prod_{f=1}^{N_P} [1 - P_0(i,j,f)]^{\mathbb{I}_{i,j,f}} P_0(i,j,f)^{1-\mathbb{I}_{i,j,f}} \\ &= \prod_{f:\mathbb{I}_{i,j,f}=1} [1 - P_0(i,j,f)] \prod_{f:\mathbb{I}_{i,j,f}=0} P_0(i,j,f). \end{aligned}$$

The negative log-likelihood is:

$$
\begin{aligned}
L[\alpha_{i,j}|\{\mathbb{I}_{i,j,f}\}_{f=1}^{N_P}] &= -\sum_{f:\mathbb{I}_{i,j,f}=1} \log[1-P_0(i,j,f)] - \sum_{f:\mathbb{I}_{i,j,f}=0} \log P_0(i,j,f) \\
&\approx -\sum_{f:\mathbb{I}_{i,j,f}=1} \log\{\eta_{i,j}[\alpha_{i,j}S_{i,j} + \sum_{\substack{(i',j')\in \\ A(i,j,f)}} \beta(i-i',j-j')] + B_{i,j}\} \\
&\quad + \sum_{f:\mathbb{I}_{i,j,f}=0} \{\eta_{i,j}[\alpha_{i,j}S_{i,j} + \sum_{\substack{(i',j')\in \\ A(i,j,f)}} \beta(i-i',j-j')] + B_{i,j}\} \\
&= -\sum_{f:\mathbb{I}_{i,j,f}=1} \log\{\eta_{i,j}[\alpha_{i,j}S_{i,j} + \sum_{\substack{(i',j')\in \\ A(i,j,f)}} \beta(i-i',j-j')] + B_{i,j}\} \\
&\quad + (N_P - k_{i,j})(\eta_{i,j}\alpha_{i,j}S_{i,j} + B_{i,j}) + \eta_{i,j} \sum_{f:\mathbb{I}_{i,j,f}=0} \sum_{\substack{(i',j')\in \\ A(i,j,f)}} \beta(i-i',j-j')
\end{aligned}
\tag{2.9}
$$

where we have again used the low-flux approximation. This data likelihood function is convex with respect to $\alpha_{i,j}$ because the crosstalk term is independent of $\alpha_{i,j}$, and therefore does not affect the non-negativity of the second derivative.

Thus far, we have developed a statistical model that closely represents the physical processes that generate photon detections. Computationally, however, Eq. (2.9) causes the complexity of our reflectivity estimation to scale with the total number of photon detections, rather that the number of image pixels. This means that increasing the number of detections to obtain higher quality images would adversely affect our method's computational efficiency. We found in our experiments that we could achieve comparable image results by considering the expected number of crosstalk detections to be uniform across all pulse frames, still using

our knowledge of the photon count data $k_{i,j}$. In this case, we can define $P_0(i,j)$ as:

$$P_0(i,j) = \exp\left[-\left(\eta_{i,j}\alpha_{i,j}S_{i,j} + B_{i,j} + \frac{\eta_{i,j}}{N_P}\sum_{\Delta m=m-1}^{m-M_1}\sum_{\Delta n=n-1}^{n-M_2}\beta(\Delta m,\Delta n)k_{i-\Delta m,j-\Delta n}\right)\right], \tag{2.10}$$

where (m,n) is the array pixel corresponding to the image pixel (i,j), i.e., $m = ((i-1) \bmod M_1) + 1, n = ((j-1) \mod M_2) + 1$.

2.4.2 Noise Censoring

Like background noise, crosstalk causes the depth likelihood to be non-convex. Thus, before we can estimate depth by convex optimization, we must censor both background and crosstalk detections. We can use the Poisson rate functions of signal, background, and crosstalk detections to compute the likelihood that a detection is due to background or crosstalk:

$$\begin{aligned}\Pr[\text{detection at } (i,j,f) \text{ is noise}] &= \frac{\lambda_{i,j}^{BG} + \lambda_{i,j,f}^{CT}(t_{i,j,f})}{\lambda_{i,j,f}^{SIG}(t_{i,j,f}) + \lambda_{i,j}^{BG} + \lambda_{i,j,f}^{CT}(t_{i,j,f})} \\ &= \frac{B_{i,j}/T_r + \lambda_{i,j,f}^{CT}(t_{i,j,f})}{\eta_{i,j}\alpha_{i,j}S_{i,j}s(t_{i,j,f} - 2z_{i,j}/c) + B_{i,j}/T_r + \lambda_{i,j,f}^{CT}(t_{i,j,f})}.\end{aligned} \tag{2.11}$$

At this step, we already have an estimate for $\alpha_{i,j}$, but we need some preliminary estimate of depth $z_{i,j}$ to suppress crosstalk. In practice, we use $ct_{i,j}^{ROM}/2$ as a depth estimate, just as we did in the no-crosstalk case. We can now rewrite Eq. (2.11) using our initial estimates for reflectivity and depth:

$$\Pr[\text{detection at } (i,j,f) \text{ is noise}] = \frac{B_{i,j}/T_r + \lambda_{i,j,f}^{CT}(t_{i,j,f})}{\eta_{i,j}\hat{\alpha}_{i,j}S_{i,j}s(t_{i,j,f} - t_{i,j}^{ROM}) + B_{i,j}/T_r + \lambda_{i,j,f}^{CT}(t_{i,j,f})}. \tag{2.12}$$

We censor detections for which $\Pr[\text{detection at } (i,j,f) \text{ is noise}] > 0.5$. Because a crosstalk detection occurs with only a short time delay after its originating detection, it is easy to cen-

sor crosstalk that results from background noise, and much more difficult to censor crosstalk that results from signal detections. In low-SBR conditions, we obtained better results by first censoring background detections before attempting to censor crosstalk (see Section 4.4).

2.4.3 Depth Estimation

After noise censoring, we proceed to estimate depth with the assumption that the remaining detections are due to back-reflected signal, just as in the no-crosstalk case. Any crosstalk detections that do not get censored are generally indistinguishable from signal detections, and therefore do not severely impact the accuracy of depth estimation.

One parameter we do not account for is timing jitter. There is some delay between the time when a photon arrives at the active area of a detector pixel and the time when a detection is registered. This delay varies with some uncertainty due to randomness in the motion of charge carriers through the device, as illustrated by the variation in the simulated avalanche current responses in Figure 2-4(b). This uncertainty is called the detector's timing jitter, and is often as low as tens of picoseconds. Although the delay itself may be hundreds of picoseconds, the *uncertainty* in the delay is small enough that it does not warrant the computational complexity it would add to our statistical model. We have not explicitly incorporated the overall delay into our models, but it can be corrected by a simple calibration either before or after depth estimation.

Another point of interest is that the readout circuitry quantizes detection times into discrete time bins, which affects depth resolution. While these time bins can be very small ($<$10 ps) for single-pixel detectors, they are generally larger (100 ps - 1 ns) for arrays. For simplicity, we have omitted time bin quantization from our discussion in this chapter, but in practice we can account for it by converting between time units of seconds and time bin number.

Chapter 3

Simulation of Photon Detection Data

In this chapter, we will briefly describe our methods for simulating the photon detection data that we used to test our imaging method.

We begin with a set of imaging parameters that represent the relevant characteristics of a LIDAR system:

- Array size: $M \times M$ pixels
- Laser beam intensity profile: $S_{m,n}$ for all array pixels $(m, n) \in [1, M] \times [1, M]$
- Pixel-wise quantum efficiency: $\eta_{m,n}$ for all array pixels $(m, n) \in [1, M] \times [1, M]$
- Pixel-wise dark count rate: $d_{m,n}$ for all array pixels $(m, n) \in [1, M] \times [1, M]$
- Rate of photon arrivals from ambient light: b, assumed to be spatially uniform
- Number of laser pulses: N_P
- Pulse repetition period: $T_r > 2z_{max}/c$, where z_{max} is the maximum scene depth
- Photon flux of laser pulse: $s(t)$, normalized to satisfy $\int_0^{T_r} s(t)dt = 1$
- Time bin width: T_Δ

We represent a scene with ground truth reflectivity and depth images, which are composed of discrete pixels. In the continuous real world, however, a scene's reflectivity and depth are not necessarily uniform over the imaging area of a single detector pixel, which may be as large as several m^2 for airborne LIDAR systems. To simulate a continuous scene while still using discretized ground truth images, we introduce the concept of a *superpixel* to describe the imaging area of a detector pixel. This superpixel method is important in simulating realistic behavior when the scene reflectivity and/or depth is highly non-uniform within a detector pixel's imaging area, i.e., along edges and when scene features are small compared to the detector pixel's imaging area. We begin with high-resolution ground truth images, in which one $w \times w$ superpixel—for example, a 2×2 square in the ground truth images—maps onto one detector pixel (Figure 3-1). If we want to form $N_1 \times N_2$ reflectivity and depth images with a superpixel size of $w \times w$, then our ground truth images should be $wN_1 \times wN_2$ pixels. For the estimated reflectivity and depth values $\hat{\alpha}_{i,j}$ and $\hat{z}_{i,j}$, we will notate the corresponding ground truth values as $\alpha_{i,j}^{(g)}$ and $z_{i,j}^{(g)}$ for $g = [1, ..., w^2]$.

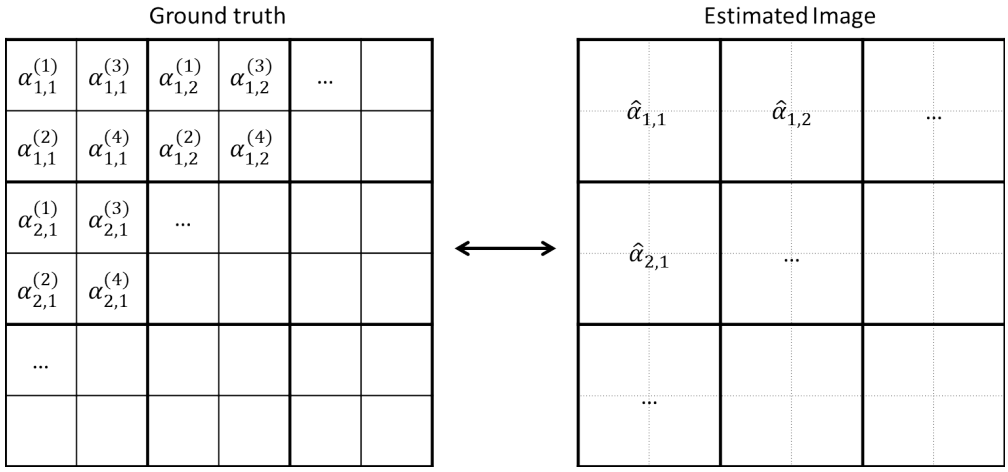


Figure 3-1: Use of superpixels to approximate a spatially continuous parameter.

We can then rewrite Eq. (2.2) as:

$$\begin{aligned} P_0(i,j) &= \left(\prod_{g=1}^{w^2} e^{-(\eta_{m,n}\alpha_{i,j}^{(g)} S_{m,n}^{(g)})} \right) e^{-B_{m,n}}, \\ &= e^{-(\eta_{m,n} \sum_{g=1}^{w^2} (\alpha_{i,j}^{(g)} S_{m,n}^{(g)}) + B_{m,n})} \end{aligned} \tag{3.1}$$

where we understand (m,n) to be the detector pixel corresponding to the image pixel (i,j) as discussed in Section 2.3. We have also used the notation $S_{m,n}^{(g)}$ to apply the superpixel concept to a laser beam intensity profile that is spatially continuous. In our estimation method, however, we can approximate the beam intensity as being uniform within a superpixel, so we will say $S_{m,n}^{(g)} \approx S_{m,n}/w^2$ for $g = [1, ..., w^2]$. We calculate $S_{m,n}$ from a 2-D Gaussian function:

$$S_{m,n} \propto \int_{y=m-1}^{m} \int_{x=n-1}^{n} e^{-\frac{(x-\frac{M}{2})^2 + (y-\frac{M}{2})^2}{2\sigma^2}} dxdy, \text{ for } (m,n) \in [1,M] \times [1,M], \tag{3.2}$$

where x and y are in pixel units. Thus, we can simplify Eq. (3.1):

$$P_0(i,j) \approx e^{-(\eta_{m,n} S_{m,n} \frac{1}{w^2} \sum_{g=1}^{w^2} \alpha_{i,j}^{(g)} + B_{m,n})}. \tag{3.3}$$

With $P_0(i,j)$ in this form, we can see that our estimated reflectivity value at (i,j) will actually be an estimate of the *average* of the scene reflectivity over the corresponding ground truth superpixel, which we will denote as $\bar{\alpha}_{i,j}$.

When the depth is highly non-uniform within a superpixel, we call the resulting image pixel a *multi-range pixel*. The distribution of signal detection times at a multi-range pixel will have multiple peaks corresponding to the different ground truth depth values within the corresponding superpixel. We can rewrite Eq. (1.5) as:

$$f_{T_{i,j}}(t; \alpha_{i,j}^{(g)}, {z_{i,j}^{(g)}}_{g=1}^{w^2} | \text{detection is signal}) = \frac{\sum_{g=1}^{w^2} \alpha_{i,j}^{(g)} s(t - 2z_{i,j}^{(g)}/c)}{\sum_{g=1}^{w^2} \alpha_{i,j}^{(g)}}. \tag{3.4}$$

Although our work does not attempt to address this issue of multi-range pixels, we take note of it as an expected source of error along edges in our depth estimation.

To simulate detections, let us first consider a superpixel within a single pulse frame. Each superpixel (i, j) corresponds to a detector pixel (m, n) and ground truth pixels $g = [1, ..., w^2]$. For each superpixel (i, j), we generate w^2 Bernoulli random variables to represent signal detections, with success probability $1 - e^{-(\eta_{m,n}\alpha_{i,j}^{(g)}S_{m,n}^{(g)})}$ for $g = [1, ..., w^2]$. If a Bernoulli trial returns a success, then we then assign it a detection time from the distribution $s(t - 2z_{i,j}^{(g)}/c)$. We also generate another Bernoulli random variable to represent noise detections with success probability $1 - e^{-B_{m,n}}$, and assign a detection time drawn from a uniform distribution on $[0, T_r]$. If multiple detections occur within the superpixel, then we retain only the detection with the earliest arrival time, since a detector pixel can only fire once in a pulse frame. In this way, we simulate detections over the $M \times M$ array for N_P pulses, and scan over the ground truth scene to obtain detection times $t_{i,j,f}$ for $i = [1, ..., N_1], j = [1, ..., N_2], f = [1, ..., N_P]$. The final step is to quantize the detection times into time bins: $t_{i,j,f} \rightarrow \lceil t_{i,j,f}/T_\Delta \rceil$. We store the quantized detection times in a $N_1 \times N_2 \times N_P$ sparse matrix. We also keep track of which detections are due to signal and which are due to noise, as this information will be useful in assessing the accuracy of our noise censoring method (see Section 4.4).

3.1 Simulation of Crosstalk Detections

After generating a set of primary detections, we may optionally choose to simulate crosstalk. We generate crosstalk based on the parameters $\beta(\Delta m, \Delta n)$ and $\gamma(\Delta t)$ defined in Section 2.2. In our simulations, we experimented with various choices of $\beta(\Delta m, \Delta n)$, including the ones empirically measured for real arrays in [14] and [15]. We used a gamma distribution for the normalized rate of secondary photon generation: $\gamma(\Delta t) = \Gamma(8, 0.1)$, where the parameters of the gamma distribution were selected to match the curves given in Figure 2-4(b), and Δt is in nanoseconds.

Because crosstalk is confined within each array-frame (see Section 2.4), we will consider one array-frame at a time when simulating crosstalk. For each array-frame, our implementation maintains two data structures that hold photon detection data: (1) a priority queue that returns detections in order of detection time, and (2) an $M \times M$ array to store the output

set of detections. We begin by placing all of the primary detections in the array-frame into the priority queue. When we pop a detection off the queue, we first check the output array to see if we have already processed an earlier detection at the same pixel. If there is already an earlier detection stored at that pixel in the output array, then the detection in question is blocked, and we discard it. Otherwise, we store the detection in the output array and proceed to generate crosstalk.

To generate crosstalk from a detection at pixel (m, n) at time t, we will first consider the spatial offsets at which crosstalk can occur. At every other pixel (m', n') in the array, the expected number of crosstalk detections resulting from the primary detection at (m, n) is $\eta_{m',n'}\beta(m' - m, n' - n)$. If this quantity satisfies the low-flux assumption, then we can use it as the probability of success in a Bernoulli trial. If the trial returns a success, then we randomly sample a temporal delay Δt from the distribution $\gamma(\Delta t)$, and we assign the detection time $t' = t + \Delta t$ to the new crosstalk detection at (m', n'). The new detection is then placed in the priority queue. We repeat this process until the priority queue is empty, at which point all detections that are not blocked will be stored in the output array.

In this way, we simulate crosstalk for each array-frame, and combine the output arrays to obtain a full set of detections over all image pixels and pulse frames. Just as we did for signal and noise detections, we keep track of which detections are due to crosstalk so that we can assess the accuracy of our crosstalk censoring method.

Chapter 4

Results and Performance Analysis

4.1 Performance Measurements

We will first discuss how we quantify our algorithm's performance for reflectivity and depth estimation. To quantify the accuracy of a reflectivity estimate, we use mean squared error (MSE) in decibels (dB):

$$\mathrm{MSE}(\hat{\boldsymbol{\alpha}}, \boldsymbol{\alpha}) = 10 \log_{10} \left(\frac{1}{N_1 N_2} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} (\hat{\alpha}_{i,j} - \bar{\alpha}_{i,j})^2 \right), \tag{4.1}$$

where $\hat{\boldsymbol{\alpha}}$ is our estimated reflectivity image, $\boldsymbol{\alpha}$ is the ground truth reflectivity image, and $\bar{\alpha}_{i,j} = \frac{1}{w^2} \sum_{g=1}^{w^2} \alpha_{i,j}^{(g)}$, as in Chapter 3

We quantify the accuracy of a depth estimate by computing the root mean squared error (RMSE) in meters:

$$\mathrm{RMSE}(\hat{\boldsymbol{z}}, \boldsymbol{z}) = \sqrt{\frac{1}{w^2 N_1 N_2} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \sum_{g=1}^{w^2} (\hat{z}_{i,j} - z_{i,j}^{(g)})^2}, \tag{4.2}$$

where $\hat{\boldsymbol{z}}$ is our estimated depth image, and $\boldsymbol{z}$ is the ground truth depth image. Because it does not make sense to average depth values within a superpixel, we have defined the depth

RMSE such that we compare the estimated depth at an image pixel with each of the w^2 ground truth depth values within the corresponding superpixel.

Another useful measurement to assess our method's performance for a given ground truth scene is the pixel-wise RMSE over repeated trials of reflectivity and depth estimation with different sets of simulated data. Over R repeated trials, we can compute an RMSE image for reflectivity estimation as:

$$\text{RMSE}_{i,j}(\{\hat{\alpha}_{i,j}^{(\ell)}\}_{\ell=1}^{R}, \bar{\alpha}_{i,j}) = \sqrt{\frac{1}{R}\sum_{\ell=1}^{R}(\hat{\alpha}_{i,j}^{(\ell)} - \bar{\alpha}_{i,j})^2}\ , \text{ for all image pixels } (i,j), \tag{4.3}$$

and similarly for depth estimation. For all three of these quantities—reflectivity MSE, depth RMSE, and pixel-wise RMSE—a lower value means that the estimated image is closer to the ground truth.

4.2 Simulation of Ideal Conditions

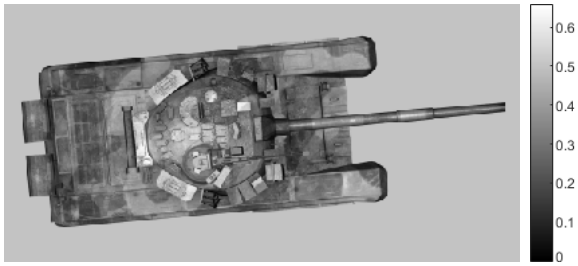
We begin by assessing our algorithm's performance in ideal conditions, by simulating only signal photon detections with uniform array parameters. In real imaging systems, signal-only conditions can be approached when ambient light and dark counts are negligible (very low dark count rates have been achieved in some CMOS arrays [8]). For our purposes, our algorithm's performance with ideal data will serve as a useful benchmark for comparison when we add noise and non-idealities to our simulation. In these simulations and all that follow, unless otherwise specified, the imaging parameters are shown in Table 4.1. The parameters N_P and $S_{i,j}$ were chosen to realize the desired number of photon counts. In later results that include noise, $B_{i,j}$ was chosen to realize the desired SBR. We also included the effect of an optical bandpass filter with a 50% transmission at the operating wavelength, emulating the physical imaging setup in [6] and [7].

In Figure 4-1, we show results from a tank scene with only signal detections. Figure 4-2 shows how performance improves as the mean number of photon detections increases. Figure

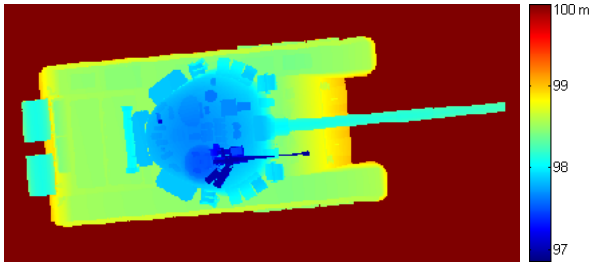
Parameter	Value
$M \times M$	32×32 pixels
$N_1 \times N_2$	256×512 pixels
$w \times w$	4×4 pixels
T_p	40 ps
T_r	1 μs
$\eta_{i,j}$	0.35
T_Δ	100 ps

Table 4.1: Simulation parameters.

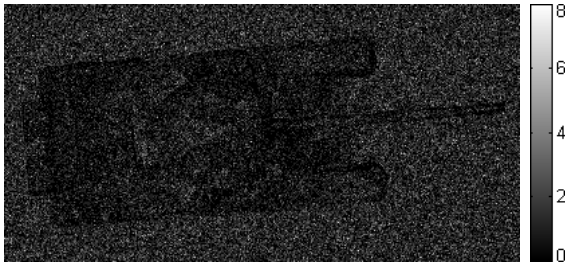
4-3 shows that the error in both reflectivity and depth estimation is primarily due to loss of edge detail and small image features. For reflectivity estimation, the high variance of Poisson noise with low photon counts makes it difficult to simultaneously preserve image details and suppress noise, which we will discuss further in Section 4.3. For depth estimation, the error along depth boundaries can be attributed to the fact that we used 4×4 superpixels in this set of simulations. In other words, 16 pixels in the ground truth images are mapped onto 1 pixel in the estimated image. This means that we have multi-range pixels when the scene depth is highly non-uniform within a 4×4 superpixel. Although our work did not seek to address the issue of multi-range pixels, there is ongoing work in this area using techniques such as range gating and mode-based coincidence processing [18]. When the scene depth is uniform within a superpixel (e.g., $w = 1$), any error in the depth estimate is due to the duration of the laser pulse, the quantization of detection times, the detector's timing jitter, and possible over-smoothing resulting from spatial regularization. Error due to the laser pulse's duration decreases as the number of signal detections increases.



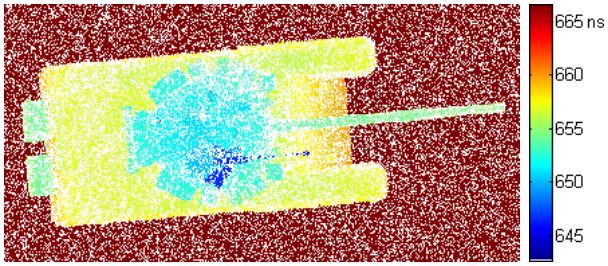
(a) Ground truth reflectivity.



(b) Ground truth depth.

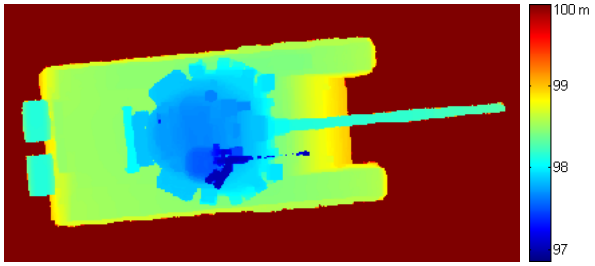


(c) Photon counts with avg 1 ppp.

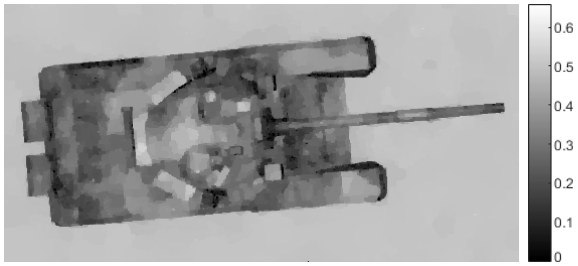


(d) Mean of detection times with avg 1 ppp (white indicates no detections).

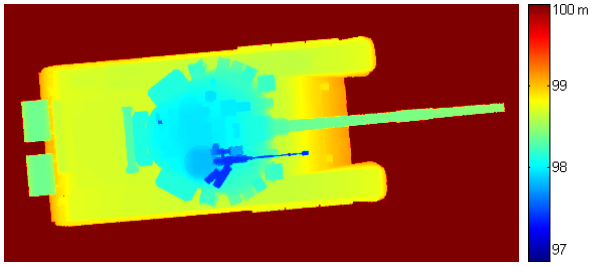
(e) Reflectivity estimate with avg 1 ppp.
MSE = -26.4 dB.



(f) Depth estimate with avg 1 ppp.
RMSE = 12.3 cm.

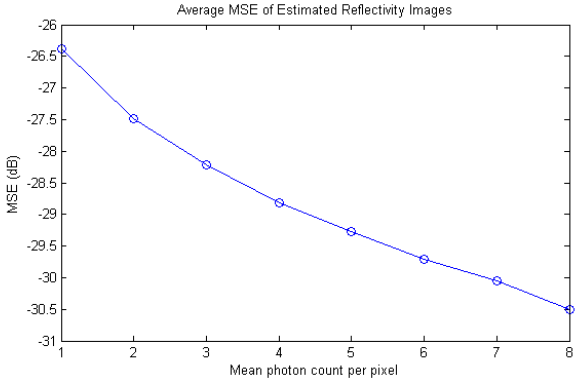


(g) Reflectivity estimate with avg 8 ppp.
MSE = -30.3 dB.

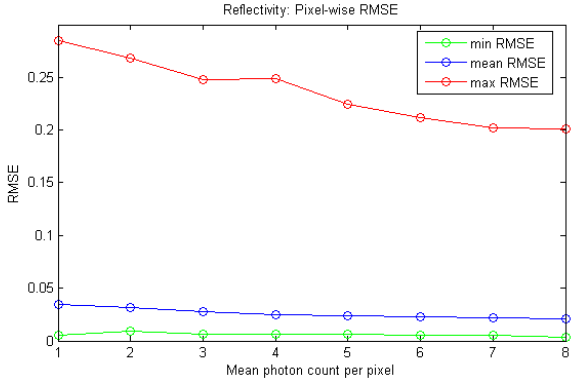


(h) Depth estimate with avg 8 ppp.
RMSE = 7.7 cm.

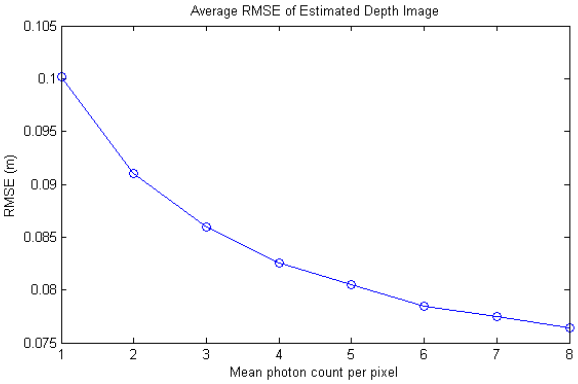
Figure 4-1: Results with only signal detections. Ground truth size: 1024×2048. Superpixel size: 4×4. Image size: 256 × 512.



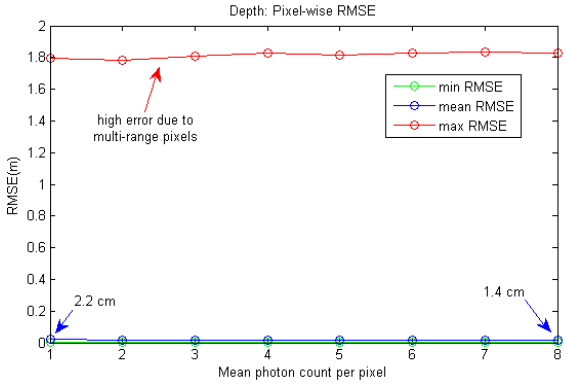
(a) MSE of reflectivity estimate for varying signal levels.



(b) Pixel-wise RMSE values for reflectivity estimates for varying signal levels.



(c) RMSE of depth estimate for varying signal levels.

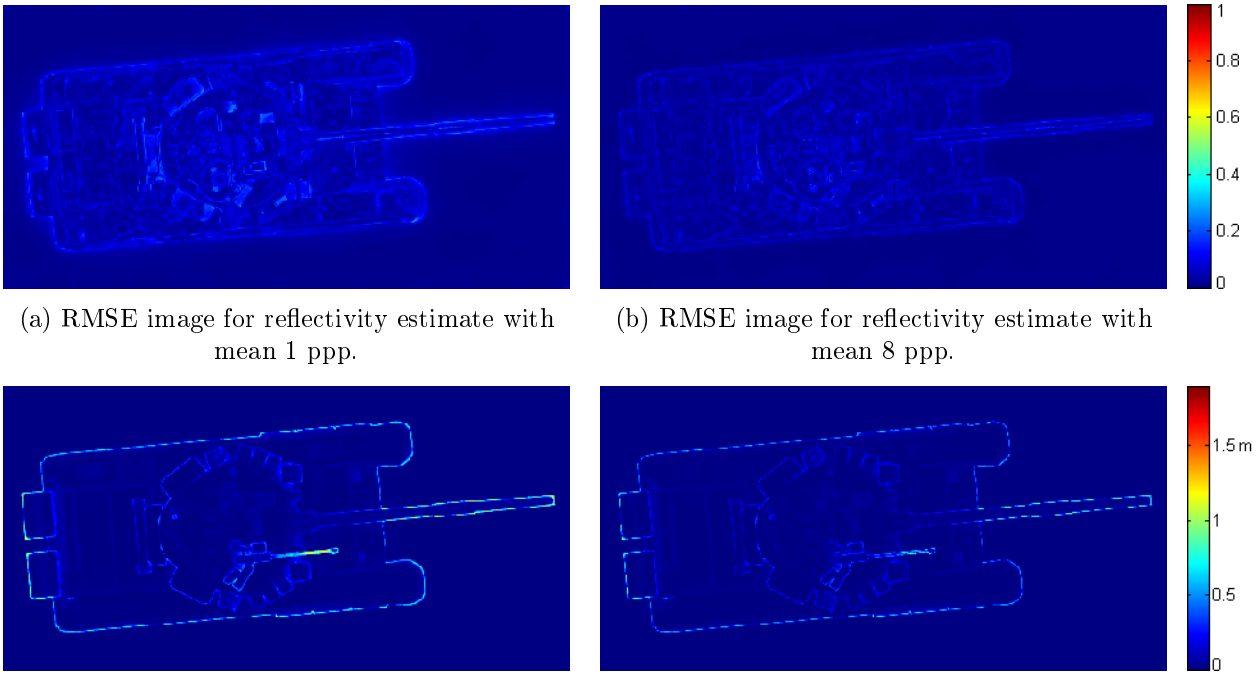


(d) Pixel-wise RMSE values of depth estimates for varying signal levels.

Figure 4-2: Performance with only signal detections, averaged over 50 trials.

4.3 Spatial Resolution Tests

Our method's ability to suppress high noise levels in photon-starved conditions relies on the key assumption that reflectivity and depth are each spatially correlated in natural scenes. This assumption breaks down when the features of interest are small, i.e., only a few pixels in size. Seeking to achieve a spatial resolution of only 1-2 pixels, for example, would equate to abandoning our assumption of spatial correlation in the scene. Our method thus sacrifices some degree of spatial resolution in order to achieve high photon efficiency, and in this section we seek to determine our method's best achievable spatial resolution, in terms of the smallest



(a) RMSE image for reflectivity estimate with mean 1 ppp.

(b) RMSE image for reflectivity estimate with mean 8 ppp.

(c) RMSE image for depth estimate with mean 1 ppp.

(d) RMSE image for depth estimate with mean 8 ppp.

Figure 4-3: Pixel-wise RMSE images for reflectivity and depth estimation over 50 repeated trials.

feature size (in pixel units) that is discernible in the reflectivity and depth images. Naturally, spatial resolution improves as the number of signal detections increases, but since we are interested in exploring the limits of photon efficiency, we conduct our analysis here only for very low photon counts.

To test spatial resolution, we created the test chart shown in Figure 4-4. In these charts, the feature size increases from left to right, and the image "contrast" increases from top to bottom. In this context, the image contrast corresponds to the change in reflectivity and depth across edges, which we will refer to as the dynamic range of the scene's reflectivity and depth, respectively. This chart allows us to determine if and how the spatial resolution of our reflectivity and depth estimation depends on dynamic range. In these test chart simulations, we used $w = 1$ and uniform imaging parameters, with noise added at an SBR of 1.

Our tests show that the factors that determine spatial resolution differ for reflectivity versus depth estimation. For reflectivity, the spatial resolution depends on the dynamic

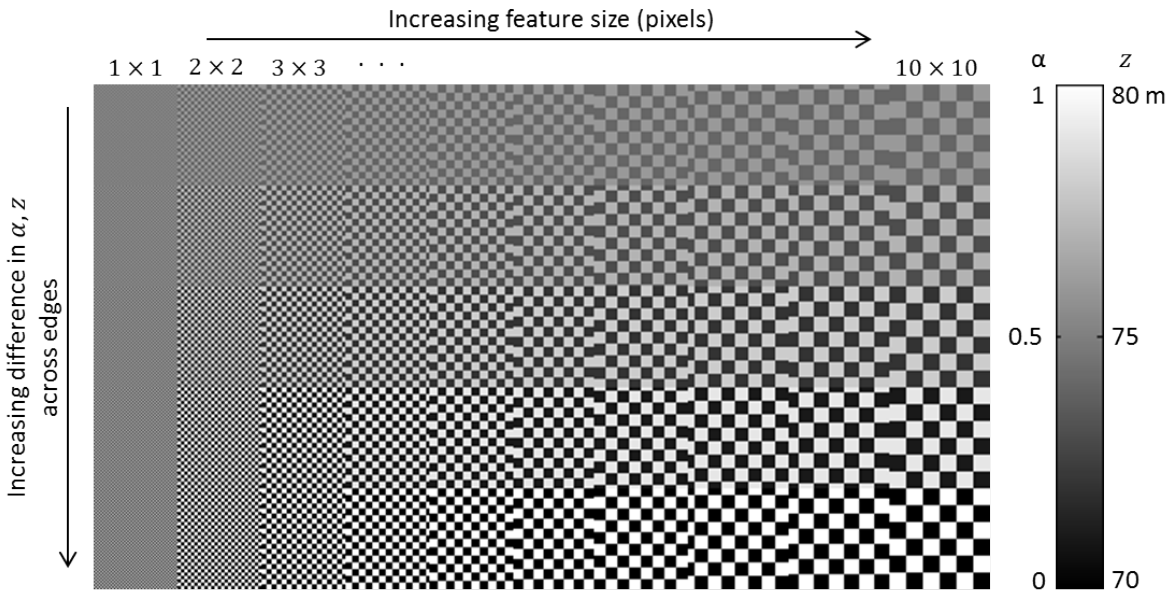
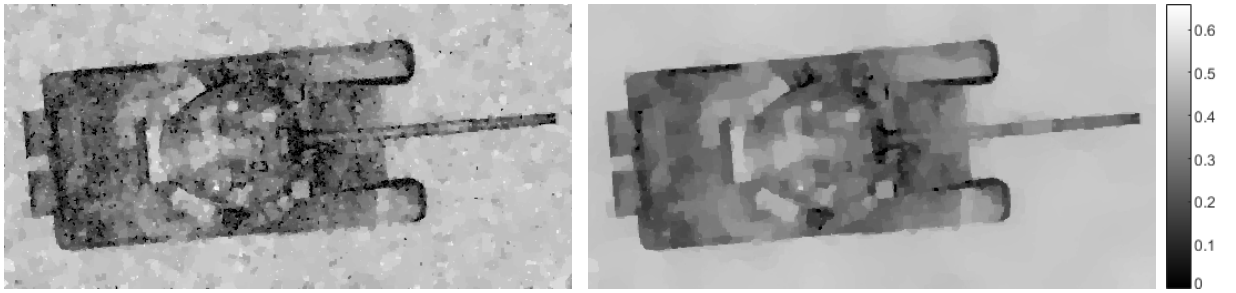


Figure 4-4: Grid chart used to test spatial resolution of both reflectivity and depth estimation. When testing reflectivity, the ground truth depth was held constant at 75 m. When testing depth, the ground truth reflectivity was held constant at 1.

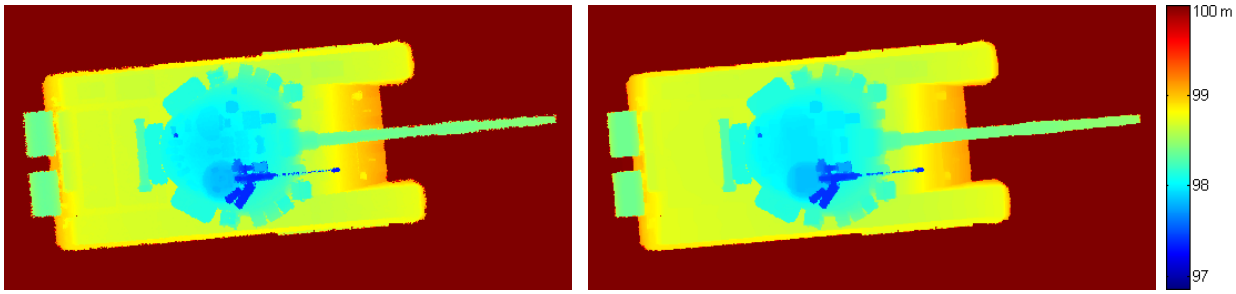
range of the scene's reflectivity, the number of signal detections, the SBR, and the spatial regularization parameter $\tau_{\boldsymbol{\alpha}}$. For depth, the spatial resolution depends primarily on the ROM window size, which plays a large role in determining the set of detections that are used to estimate depth after noise censoring.

As with many denoising algorithms in image processing, we have a tradeoff between the suppression of unwanted noise and the preservation of image details. With photon-limited, noisy data, it is impossible to distinguish small image features from noise unless we have some prior knowledge about what the scene should look like. In fact, the spatial regularization term is a spatial prior from the perspective of Bayesian inference, where the parameters $\tau_{\boldsymbol{\alpha}}$ and $\tau_{\boldsymbol{z}}$ represent our belief about the degree to which reflectivity and depth, respectively, are spatially correlated in the scene. Figures 4-5 demonstrates the effect of these spatial regularization parameters. In Figure 4-6, (a) and (b), we can see more precisely how $\tau_{\boldsymbol{\alpha}}$ affects spatial resolution in reflectivity estimation, where spatial regularization plays a larger role than in depth estimation.



(a) Reflectivity estimate with low $\tau_{\boldsymbol{\alpha}}$.

(b) Reflectivity estimate with high $\tau_{\boldsymbol{\alpha}}$.

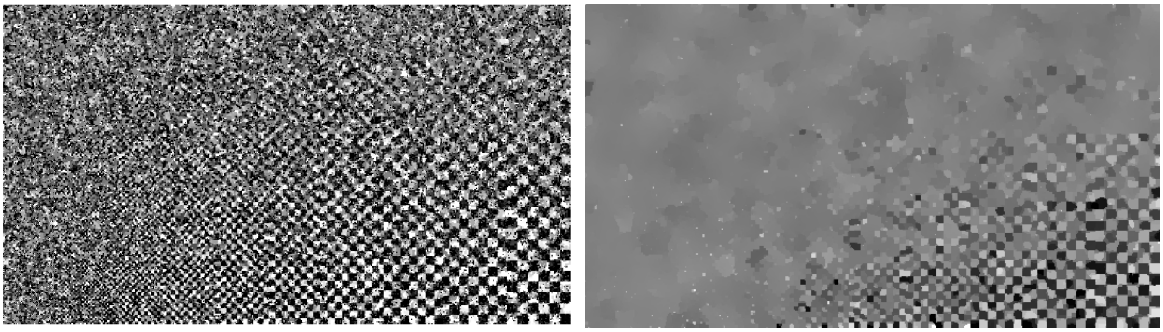


(c) Depth estimate with low $\tau_{\boldsymbol{z}}$.

(d) Depth estimate with high $\tau_{\boldsymbol{z}}$.

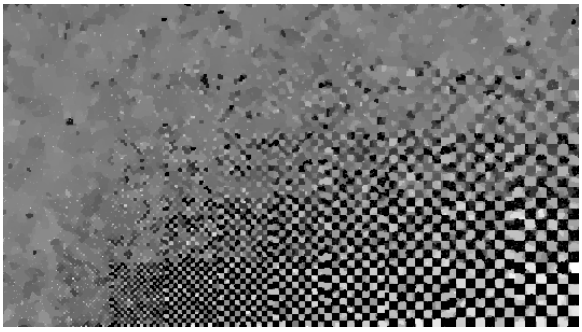
Figure 4-5: Smoothness vs. detail tradeoff: in (a) and (c), smaller image features appear more defined, but the image is noisier. Mean photon count = 3 ppp, signal only.

Besides spatial regularization, our method also suppresses noise by means of statistical noise censoring, where our assumption of spatial correlation is implied in the ROM window size. In their previous work, Kirmani *et al.* [6] and Shin *et al.* [7] performed noise censoring after reflectivity estimation. With an additional step of re-estimating reflectivity after noise censoring, we were able to achieve significantly better spatial resolution, as shown in Figure 4-6(c). For depth estimation, the spatial resolution is inherently limited by the fact that our noise censoring method uses a local neighborhood ROM filter. As we will discuss further in Section 4.4, the noise censoring step is likely to incorrectly remove some signal detections when the depth features are spatially smaller in size than the ROM window. This behavior is seen in Figure 4-7(a), where much of the signal return is removed at grid sizes that are smaller than the 5×5 ROM window. The effects of this behavior are seen in the resulting depth image in Figure 4-7(b), where the depth estimation cannot reliably resolve features smaller than 5×5 pixels. For the data set in these figures (with a mean signal count of 2 ppp and SBR of 1), a 5×5 ROM window gave the best results. Even with the same SBR of 1, a



(a) Low $\tau_{\boldsymbol{\alpha}} = 0.5$.

(b) High $\tau_{\boldsymbol{\alpha}} = 1.5$.

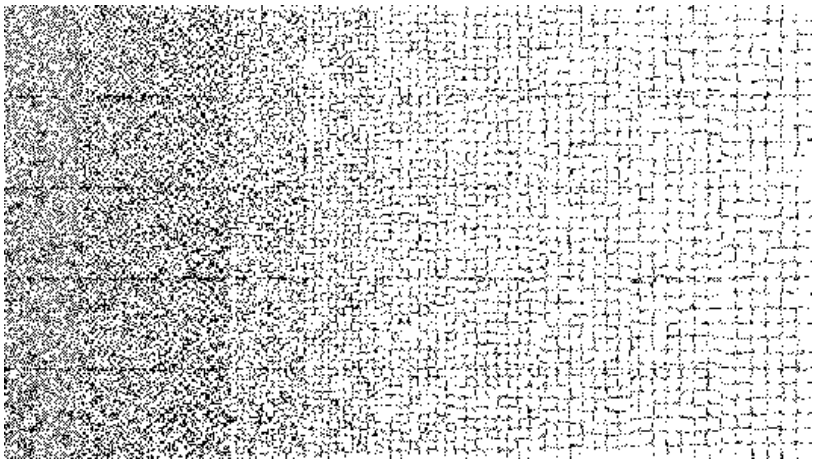


(c) Re-estimated reflectivity after noise censoring, $\tau_{\boldsymbol{\alpha}} = 1.5$.

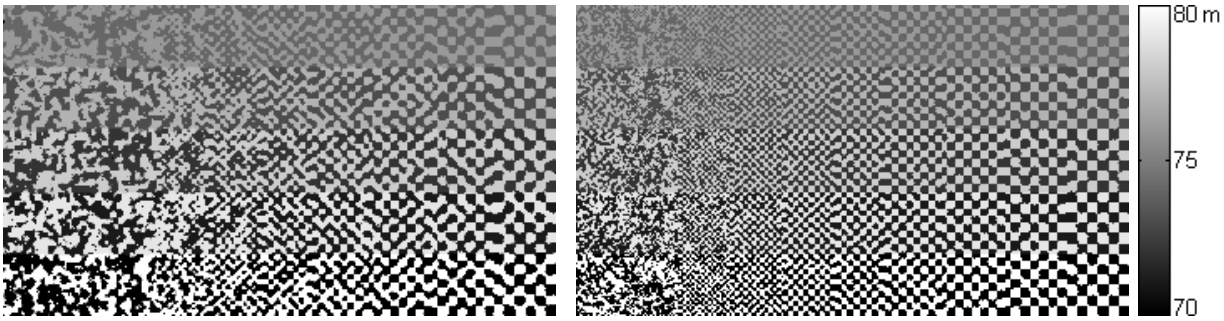
Figure 4-6: Spatial resolution test for reflectivity estimation. Mean signal counts = 2 ppp, SBR = 1.

slightly higher number of signal detections (~ 5 ppp) allows us to use a smaller ROM window (3×3), and thereby improve spatial resolution to $\sim 3 \times 3$ pixels, as shown in Figure 4-7(c).

Interestingly, depth estimation achieves better edge definition at the top of the test chart where the dynamic range is smaller (the change in depth across edges is smaller), whereas it is the opposite for reflectivity, which achieves better edge definition at the bottom of the test chart where the dynamic range is large. This characteristic of depth estimation is a result of the noise censoring method, as it is more likely to incorrectly censor signal detections when there is a very large depth gradient within a ROM window. This behavior becomes more apparent at very low signal flux, as in Figure 4-7(b), where we can see that our depth estimate loses some edge definition at the bottom of the test chart.



(a) Black indicates pixels where signal detections are removed by noise censoring. Mean signal counts = 1 ppp, SBR = 1. ROM window size = 5×5.



(b) Depth estimate. Mean signal counts = 1 ppp, SBR = 1. ROM window size = 5×5.

(c) Depth estimate. Mean signal counts = 5 ppp, SBR = 1. ROM window size = 3×3.

Figure 4-7: Spatial resolution test for depth estimation.

In summary, our spatial resolution tests show that our method is capable of resolving features as small as 3×3 pixels with low signal flux and high noise levels. The tests demonstrate how our method's spatial resolution varies depending on the scene's characteristics and the choice of parameters that represent our assumption of spatial correlation. We also saw that this assumption presents us with a performance tradeoff between noise suppression and spatial resolution. Because of this tradeoff, our discussion of spatial resolution is closely related to the analysis of our noise censoring method in Section 4.4. Gaining a better understanding of the behavior of our noise censoring method will help to explain the spatial resolution characteristics that we have seen in this section.

4.4 Noise Censoring

We will now analyze our method's performance in censoring background detections and crosstalk detections, addressing each of these cases separately. One advantage of using simulated data is that we know whether each detection is due to signal, background, or crosstalk. This knowledge allows us to assess the accuracy of our censoring methods for background and crosstalk detections, respectively. Both background and crosstalk censoring involve two flexible parameters: the ROM window size and the censoring threshold. We will explore how these parameters each affect the behavior of the censoring method and the resulting reflectivity and depth images after censoring.

4.4.1 ROM Filter

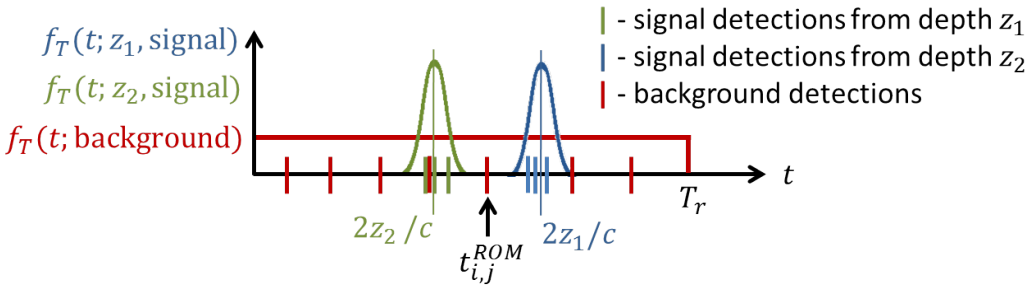
The purpose of the ROM filter is to obtain an initial rough estimate of depth (or precisely, $2z_{i,j}/c$) for use in the hypothesis test for noise censoring. We censor detections whose detection times are beyond some threshold away from the ROM value $t_{i,j}^{ROM}$, so we want $t_{i,j}^{ROM}$ to be as close as possible to the true signal value $2z_{i,j}/c$. There are two main cases in which the ROM value is likely to be a poor estimate of $2z_{i,j}/c$:

1. when the scene depth is highly non-uniform within the ROM window, e.g., near depth boundaries, as in Figure 4-8(a). In this case, the ROM window will contain multiple signal detection times, and $t_{i,j}^{ROM}$ may fall between those times but be close to none of them, especially in the presence of background noise. This behavior is analogous to that of multi-range pixels. Intuitively, a smaller ROM window would help to achieve better performance near depth boundaries, but a window size smaller than 5×5 or 3×3 is infeasible in low-flux and noisy conditions.

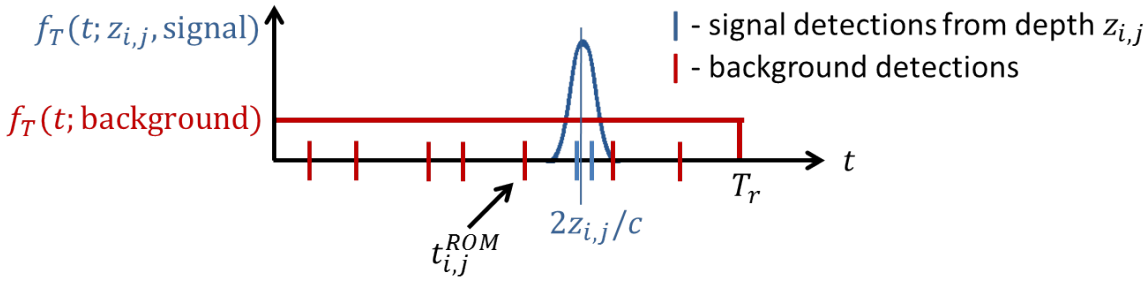
2. when there is more background than signal (low SBR), e.g., in low-reflectivity regions, as in Figure 4-8(b). In this case, $t_{i,j}^{ROM}$ will be biased toward $T_r/2$ because background detection times are uniformly distributed over $[0, T_r]$. If the approximate scene depth is known, accuracy can be improved by using a technique known as *range gating*, in

which we only consider detections within a truncated time range, chosen such that $2z_{i,j}/c$ is near the middle of the time range.

In both of these cases, the ROM estimate may also be improved by using alternative mode-based (rather than median-based) approaches to detect peaks in the distribution of detection times within the ROM window [18].



(a) Non-uniform depth within ROM window.



(b) Low SBR within ROM window.

Figure 4-8: Distribution of detection times for two cases in which the ROM filter may not provide an accurate depth estimate. The marks along the t axis correspond to a sample simulation of signal and background detection times within a ROM window.

In Figure 4-9(a), we show the difference between the ground truth depth, $z_{i,j}$, and the depth corresponding to the ROM value, $ct_{i,j}^{ROM}/2$, verifying that the ROM error is likely to be higher in the two cases of depth boundaries and low SBR.

4.4.2 Censoring Threshold

The choice of the censoring threshold is an adjustable parameter that differs for background censoring versus crosstalk censoring. This threshold determines a tradeoff between censoring

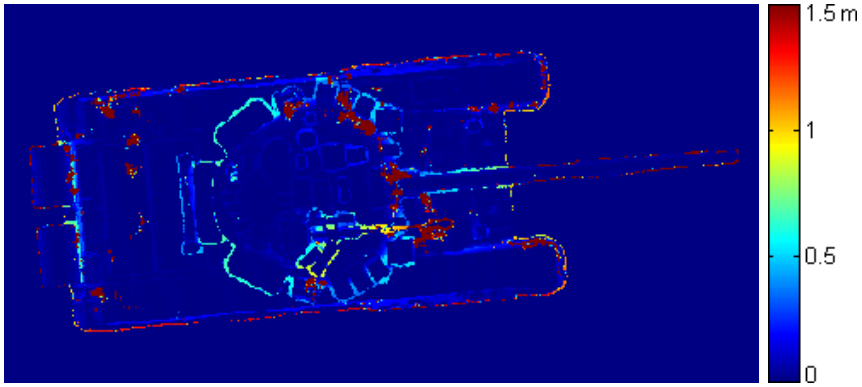
a greater fraction of noise detections and preserving a greater fraction of signal detections. A higher threshold signifies a higher tolerance for error in $t_{i,j}^{ROM}$ as an estimate of $2z_{i,j}/c$, and thus gives a better chance of retaining signal detections near depth boundaries and in low-reflectivity regions. On the other hand, a lower threshold will censor a very high percentage of noise detections, but at the cost of incorrectly removing a greater number of signal detections. This result may be acceptable if the signal return is high enough to begin with, but at low signal levels it can result in small image features being removed entirely. We will also see that in general, a higher threshold gives better results when re-estimating reflectivity after censoring, whereas a lower threshold is more suitable for depth estimation.

Censoring Background Detections

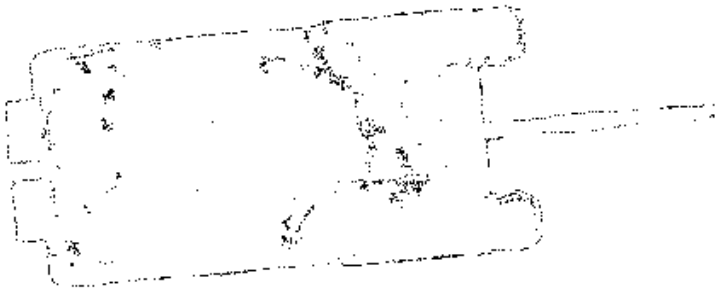
In the no-crosstalk case, Eq. (1.12) gives one possibility of a censoring threshold for background detections. We found that Eq. (1.12) is a very low threshold, and although it censored >99.9% of background detections, it also mistakenly censored $\sim 30\%$ of signal detections. For several different scenes and imaging parameters, we were generally able to find a threshold setting that censored $\sim 99\%$ of noise detections and retained $\sim 99\%$ of signal detections. Figure 4-9(b) shows that the incorrectly censored signal detections are concentrated along edges and in low-reflectivity regions, where $t_{i,j}^{ROM}$ is not close to $2z_{i,j}/c$.

Censoring Crosstalk

With crosstalk, we use a censoring threshold based on the statistical likelihood that a detection is due to either background noise or crosstalk, as in Eq. (2.12). This likelihood is very sensitive to ROM error, especially if the laser's FWHM pulse duration T_p is small compared to the time bin width T_Δ. We can adjust the tolerance of ROM error by substituting a different function for $s(t - t_{i,j}^{ROM})$ in the denominator of Eq. (2.12). In the results shown here, we found that a simple box function with a width of ± 3 time bins gave satisfactory results.



(a) ROM error, calculated as $|z_{i,j} - ct_{i,j}^{ROM}/2|$.



(b) Black indicates pixels where signal detections are removed by background censoring.

Figure 4-9: Background censoring is likely to incorrectly remove some signal detections where the ROM error is high. Mean signal count = 2 ppp, SBR = 1.

In our crosstalk simulations, we used $w = 1$, and the results we present here are for a data set with a mean signal count of 2 ppp, mean background count of 2 ppp, and mean crosstalk count of 0.75 ppp. As shown in Figure 4-10, our crosstalk censoring method incorrectly removed much of the signal return near depth boundaries. The error near edges is more pronounced than in the no-crosstalk case because secondary photons can travel through the array to cause crosstalk detections across depth boundaries. This means that crosstalk may degrade our ROM estimate as far as M pixels away from a depth boundary. We were able to ameliorate this problem to some extent by first censoring background detections, then re-computing $t_{i,j}^{ROM}$ to obtain a significantly improved estimate of $2z_{i,j}/c$. Using this improved

$t_{i,j}^{ROM}$, our crosstalk censoring method was able to censor $> 99.9\%$ of background detections and $\sim 99\%$ of crosstalk detections, while retaining $\sim 98\%$ of signal detections.

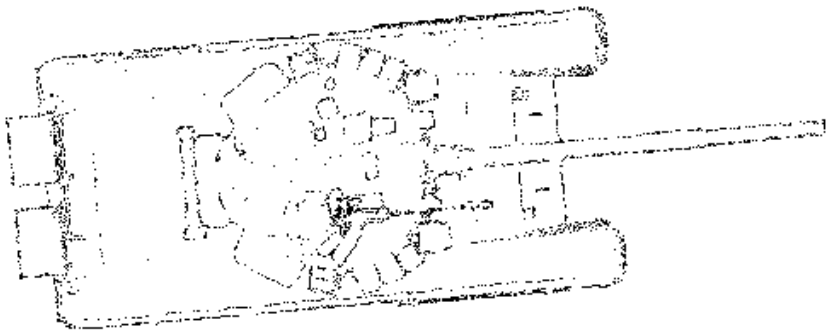


Figure 4-10: Black indicates pixels where signal detections are removed by crosstalk censoring.

Two other factors have a significant influence on our method's ability to censor crosstalk: (1) the width of the time bins as compared to the temporal delay of crosstalk, and (2) the scene's depth characteristics. If T_Δ is large, then crosstalk detections are most likely to occur at a delay of only a few time bins—although this makes crosstalk censoring more difficult, it also means that the small temporal delay of any uncensored crosstalk detections will not induce much error in depth estimation (compared to the depth resolution set by the time bin width). With $T_\Delta = 1$ ns for the same tank scene, our method was only able to censor $\sim 62\%$ of crosstalk detections without incurring an unacceptably high loss of signal detections. In this tank scene, the camera is looking straight down at the scene. It is more difficult to censor crosstalk when there is an overall depth gradient across the scene from the camera's viewing angle, as in Figure 4-11. When we rotated the camera such that there is an overall depth gradient across the scene, our method censored $\sim 92\%$ of crosstalk detections for $T_\Delta = 100$ ps, and $\sim 67\%$ of crosstalk detections for $T_\Delta = 1$ ns.

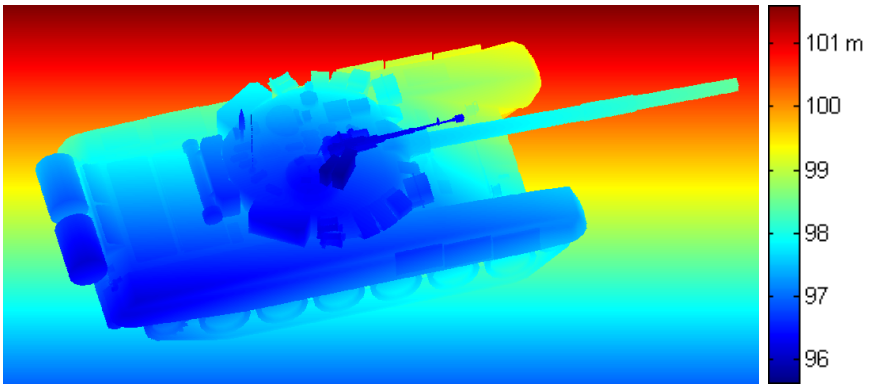


Figure 4-11: Ground truth depth for scene with overall depth gradient, for which crosstalk censoring loses accuracy.

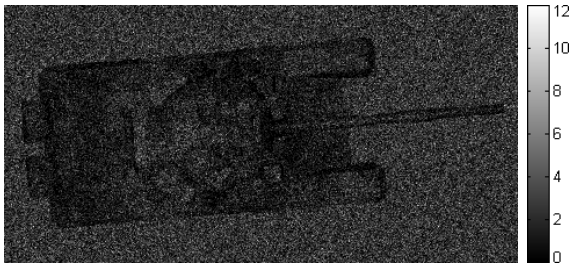
4.4.3 Effect of Background Noise on Reflectivity and Depth Estimation

Low levels of background noise can be suppressed by the spatial regularization term, and have little effect on our method's performance. We found that adding background noise at 10% of the signal level had no noticeable impact on our reflectivity and depth estimates.

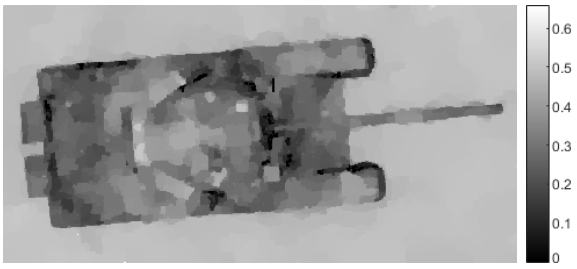
On the other hand, high levels of background noise (SBR = 1) can visibly degrade the results of reflectivity estimation, as we can see by comparing Figures 4-12, (b) and (d). In (b), we used a data set containing only signal detections, with an average of 2 ppp. In (d), we used the same set of signal detections and added background detections at an SBR of 1. With background noise, the reflectivity image loses some edge definition in detailed image features, and contains noisy artifacts in the scene's smooth background region.

Because the noise censoring step is consistently able to remove >99% of noise detections, the remaining detections approach the ideal signal-only case discussed in Section 4.2, and we can expect similar accuracy when re-estimating reflectivity after censoring. One artifact of our noise censoring method is that the estimated reflectivity values are likely to be too low where signal detections are incorrectly censored, namely, in low-SBR regions and near depth boundaries, as we can see in Figure 4-12(f). For this reason, reflectivity re-estimation performs better with a higher censoring threshold that retains more signal detections. The

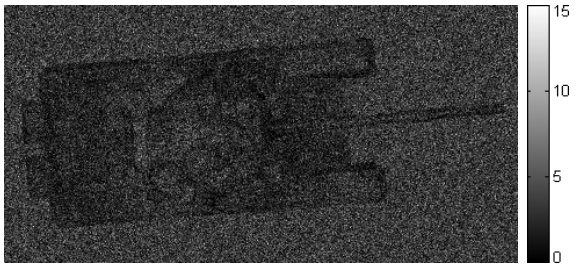
spatial regularization is effective at suppressing noise due to the small number of background detections that remain after censoring.



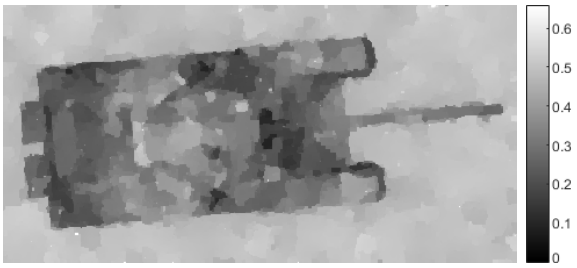
(a) Photon counts: mean 2 ppp, signal only.



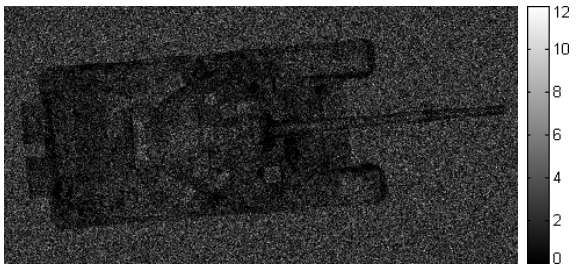
(b) Reflectivity estimate with only signal detections. MSE = -27.7 dB.



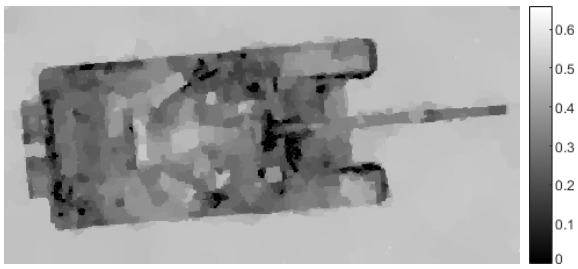
(c) Photon counts: mean 4 ppp, SBR = 1.



(d) Reflectivity estimate before noise censoring. MSE = -26.5 dB.



(e) Photon counts after censoring.



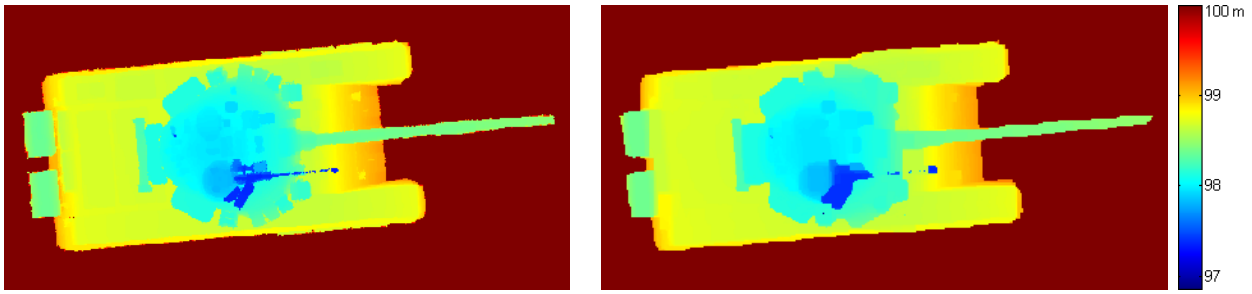
(f) Reflectivity estimate after censoring, with high censoring threshold. MSE = -27.4 dB.

Figure 4-12: Effect of background noise on reflectivity estimation.

Because we censor noise before estimating depth, the degree to which noise affects our depth estimation depends on the accuracy of the noise censoring step. In depth estimation, spatial regularization can help to fill in regions that have no photon detections. However, it is less effective at suppressing noise due to uncensored background detections than in reflectivity re-estimation, as background detection times have a more significant impact on depth likelihood than on reflectivity likelihood. Thus, a lower censoring threshold is more

suitable for depth estimation because it is more negatively affected by uncensored background detections than by the loss of signal return. However, a low threshold also has the effect of diminishing or removing the signal return from small image features where there is a large difference in depth (e.g., the thin gun on top of the tank in Figure 4-13).

We found that as the SBR decreased past 2, it became more difficult to find a censoring threshold that gave good results for both reflectivity and depth images. The results shown in Figures 4-12(f) and 4-13(b) were obtained by using different thresholds. In practical applications with low SBR, the thresholds for reflectivity and depth estimation may each be independently optimized to achieve the desired results.



(a) Depth estimate (no censoring). Mean signal count = 2 ppp, no background. RMSE = 10 cm.

(b) Depth estimate (low censoring threshold). Mean signal count = 2 ppp, SBR = 1. RMSE = 15 cm.

Figure 4-13: Effect of background noise on depth estimation.

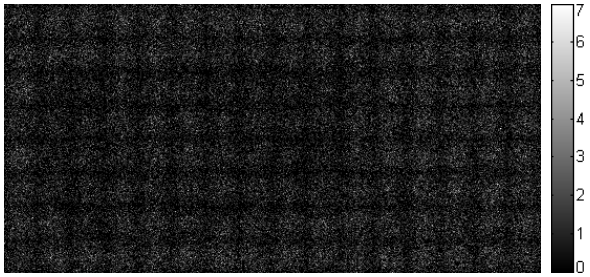
4.4.4 Reflectivity and Depth Estimation With Crosstalk

Crosstalk acts as a point spread function that blurs image features. Figure 4-14 shows that crosstalk also creates grid-like artifacts in reflectivity images, for which our method corrects by incorporating crosstalk into our statistical model of photon counts.

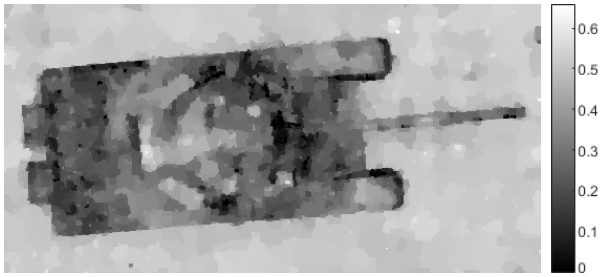
In Section 2.4, we presented Eqs. (2.8) and (2.10) as two possible ways to model the distribution of photon counts with crosstalk, where the former uses knowledge of detection events more precisely but is not computationally scalable. Figure 4-15 shows that the two models give very similar results, suggesting that the more computationally efficient Eq. (2.10) can be used without any substantial loss in performance. Although these statistical

models can correct for the grid-like artifacts, image features are still noticeably blurred as compared to Figure 4-12(d).

The results of reflectivity and depth estimation after censoring are shown in Figure 4-16(a) and (b). We see that the addition of crosstalk exacerbates the removal of signal detections along depth boundaries in this scene, because of the need for a very low threshold to censor crosstalk.

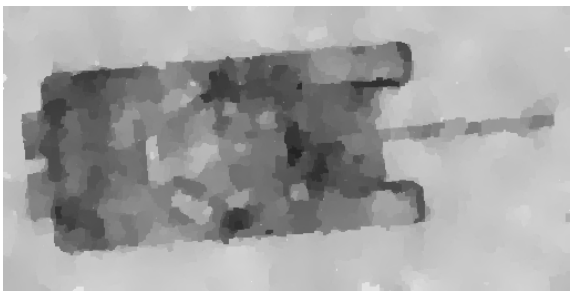


(a) Crosstalk counts only. Based on our spatial model for crosstalk generation, fewer secondary photons reach the edges and corners of the array.

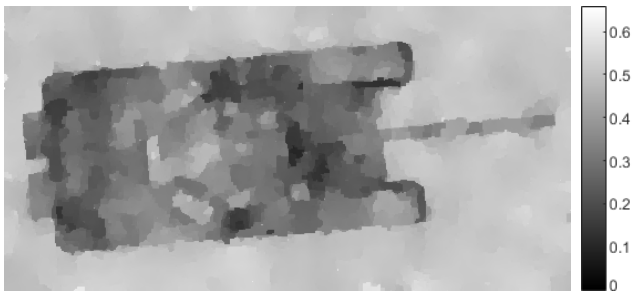


(b) Reflectivity estimate when crosstalk is modeled as a uniform background noise contribution, not using any knowledge of detection events. MSE = -26.0 dB.

Figure 4-14: Crosstalk causes grid-like artifacts in reflectivity images. Mean signal counts = 2 ppp, mean crosstalk counts = 0.4 ppp. No background detections added, to show effect of crosstalk more clearly.

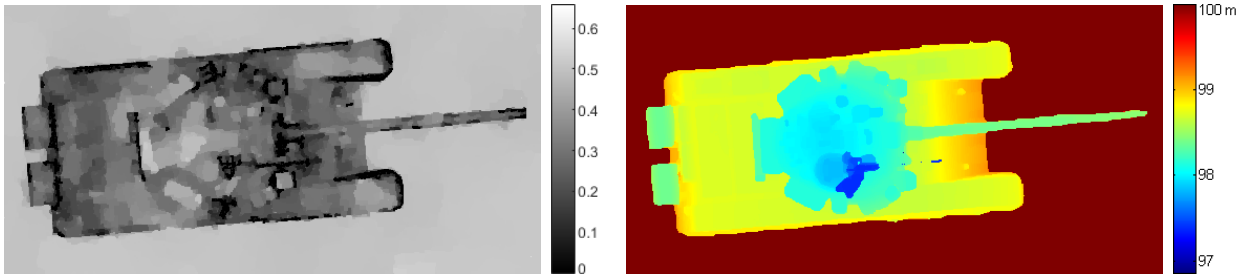


(a) $P_0(i,j,f)$ defined as in Eq. (2.8). MSE = -26.1 dB.



(b) $P_0(i,j)$ defined as in Eq. (2.10). MSE = -26.0 dB.

Figure 4-15: Comparison of statistical models for reflectivity estimation with crosstalk. Mean signal counts = 2 ppp, mean background counts = 2 ppp, mean crosstalk counts = 0.75 ppp.



(a) Reflectivity estimate after censoring. MSE = -25.6 dB.

(b) Depth estimate after censoring. RMSE = 24 cm.

Figure 4-16: Results of crosstalk censoring. Mean signal count = 2 ppp, mean background count = 2 ppp, mean crosstalk count = 0.75 ppp.

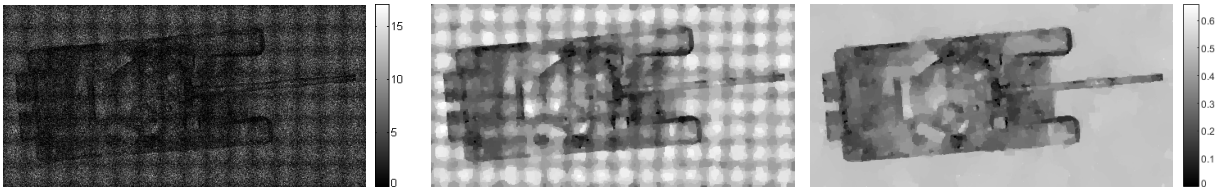
4.5 Correction of Non-Uniform Array Parameters

4.5.1 Non-Uniform Beam Intensity Profile

The non-uniformity of the laser beam's spatial intensity profile primarily affects reflectivity estimation. In Figure 4-17, we show the improvement in reflectivity estimation achieved by incorporating the non-uniformity of the beam profile into the statistical model. We simulated the photon detections using the same laser power needed to obtain an average of 1 signal detection per pixel when the entire array is within a $1 - 1/e^2$ beam radius. At this laser power, the optimal beam size ($\sigma = M/2$; see Appendix B for derivation) produced a mean of 3.2 signal detections per pixel, with a minimum of 1.5 expected signal detections per pixel at the corners of the array, and a maximum of 4 at the center of the array. Our method's ability to correct for non-uniform beam intensity significantly increases photon efficiency by allowing imaging systems to make use of a larger portion of the total expended laser power.

4.5.2 Non-Uniform Quantum Efficiency and Dead Pixels

We simulated detections from an array with pixel-wise quantum efficiencies normally distributed with a mean of 0.35 and standard deviation of 0.04, truncated to 0.35 ± 0.1. We also randomly assigned 1% of detector pixels to be dead. For a mean photon count of 1.1 ppp



(a) Photon counts.

(b) Reflectivity estimate assuming uniform beam profile. MSE = -22.4 dB.

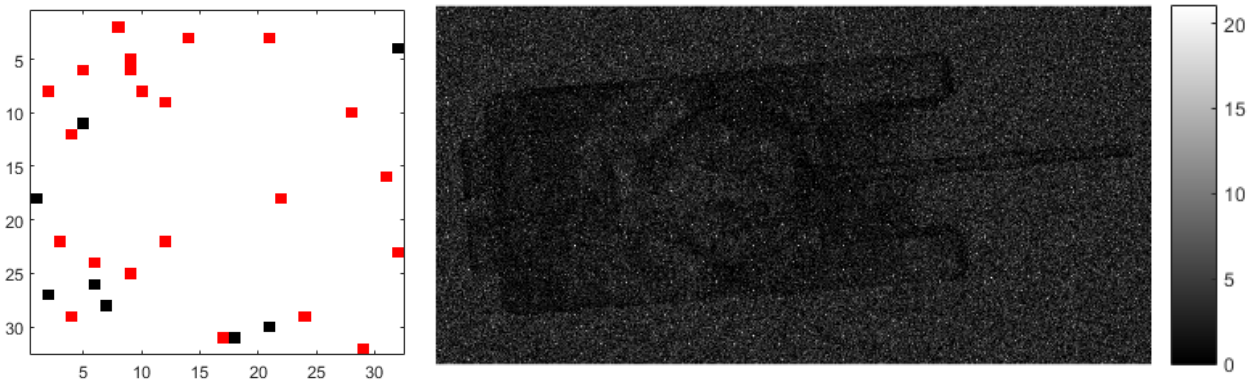
(c) Reflectivity estimate accounting for non-uniform beam profile. MSE = -28.4 dB.

Figure 4-17: Reflectivity estimation with Gaussian beam profile. SBR = 10.

and SBR of 10, there was no noticeable difference in performance when assuming a uniform quantum efficiency versus accounting for the non-uniformity and dead pixels.

4.5.3 Non-Uniform Dark Counts and Hot Pixels

We simulated detections from an array with pixel-wise dark count rates normally distributed with a mean of 2 kHz and a standard deviation of 1 kHz, truncated to [0 kHz, 4 kHz]. We then randomly assigned 2% of detector pixels to be hot pixels, and increased the dark count rate at these pixels by a factor of 10. We also maintained the same non-uniformity of quantum efficiency and dead pixels that we previously described in Section 4.5.2. In Figure 4-18(a), we show the hot pixels in red and the dead pixels in black.

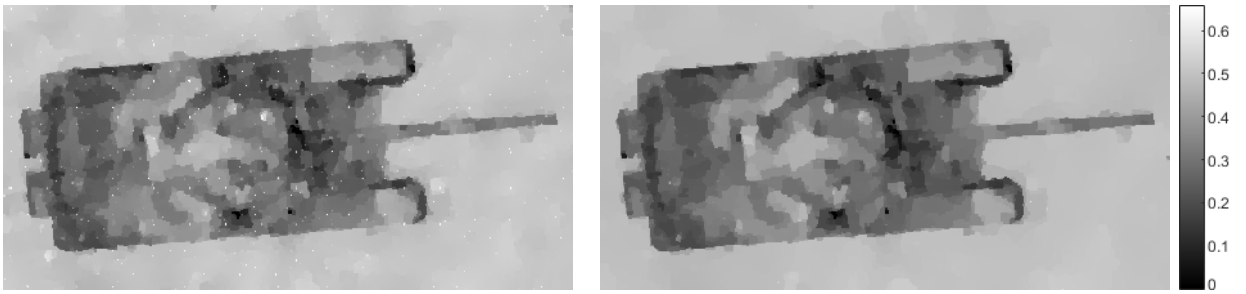


(a) Dead pixels (black) and hot pixels (red).

(b) Photon counts: mean signal count = 2 ppp, mean background count = 0.6 ppp (ignoring hot pixels).

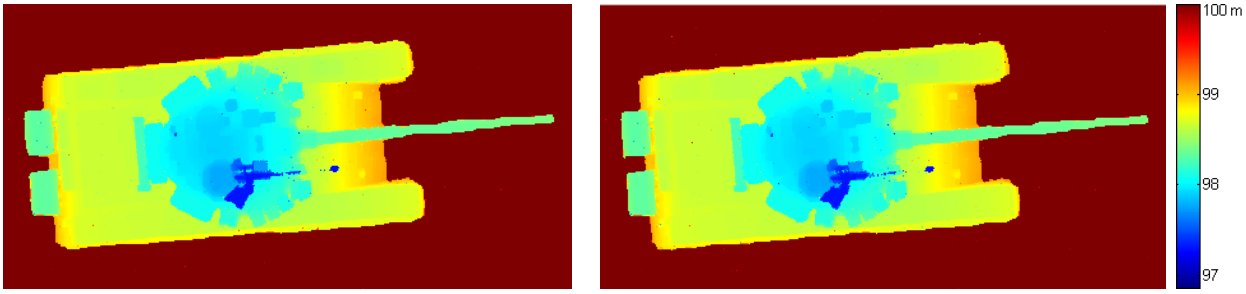
Figure 4-18: Non-uniform quantum efficiency and dark count rate with dead pixels and hot pixels.

Figure 4-19 shows a comparison of our results when we assumed a uniform quantum efficiency of 0.35 and a uniform dark count rate of 2.2 kHz, versus using our array method to account for the non-uniformities. In (a), we see that hot pixels in particular create visible artifacts in the reflectivity image, and in (b), we see that our method is able to correct for them. Figure 4-19(c) and (d) show that these non-uniformities have little effect on depth estimation.



(a) Reflectivity estimate assuming uniform parameters.

(b) Reflectivity estimate using our method.



(c) Depth estimate assuming uniform parameters.

(d) Depth estimate using our method.

Figure 4-19: Reflectivity and depth estimates for data in Figure 4-18.

Chapter 5

Conclusions

In this thesis, we have presented statistical models that accurately reflect the photon detection process in a LIDAR system using a SPAD array, and we have described a photon-efficient image formation method that uses these models to correct for the adverse effects of the non-idealities and non-uniformities that are present in an array-based system. By adapting Shin *et al.*'s imaging method in [7] for use with detector arrays, we have made it more practical for imaging applications that benefit from the ability to collect data at thousands of pixels in parallel. While the non-idealities of the imaging parameters would otherwise require that a larger number of photons be collected to obtain good results, we have demonstrated that our method can successfully correct for the undesirable effects of these non-idealities at very low photon counts. Through our analysis in Chapter 4, we have seen that certain imaging parameters affect the data more than others. We have shown that our method's ability to correct for the radial falloff of laser beam intensity allows a LIDAR system to focus a larger fraction of the transmitted light within the array's field of view, thereby increasing photon efficiency by at least 5 times. Our method's ability to censor background and crosstalk detections, as well as correct for hot pixels, may also allow SPAD arrays to be operated at a higher quantum efficiency, which generally has the caveat of higher dark count and crosstalk levels. One of the next steps in our future work would be to test our method with a real imaging system to confirm the results we obtained with simulated data.

Our experiments showed that our censoring method can remove detections due to background light and dark counts with very high accuracy, giving us a data set that approaches ideal signal-only conditions. The accuracy of our noise censoring method suffers near depth boundaries and in low-SBR conditions, where signal detections are likely to be incorrectly labeled as noise. The effects of this phenomenon are visible in reflectivity and depth images, but we have shown that they can be minimized by independently optimizing the censoring thresholds for reflectivity and depth estimation, combined with spatial regularization. We expect that our noise censoring method would benefit from further progress in processing multi-range pixels, which may help to achieve better performance at depth boundaries. Multi-range processing also has the potential for sub-pixel depth resolution and for the reconstruction of scene objects that are partially occluded.

We have also shown that the accuracy of our crosstalk censoring method is highly dependent on the time bin resolution and the scene's depth characteristics. From this behavior, we conclude that our method may perform well for specific applications where the SPAD array has small time bins and the scenes of interest have no overall depth gradient. For more general imaging applications, however, statistical crosstalk censoring may not be a reliable method. For this reason, photon-efficient LIDAR systems would still benefit from hardware improvements to decrease crosstalk in SPAD arrays.

Our method's high photon efficiency has promising implications for imaging applications where laser power or data acquisition time is limited, whether by hardware constraints or external factors. Compared to conventional image formation methods, our method relaxes the requirement for high-quality, high-volume data, thus shifting much of the burden of image quality from data collection to data processing. Despite our method's strengths in processing photon-limited, noisy data, it is not without its own limitations. As we have discussed, various imaging and algorithmic parameters play a role in determining the particular behavior at each step of our image formation method. Our analysis not only suggests directions for further progress in our computational framework, but may also help to guide the future development of SPAD arrays and LIDAR systems that use them. As SPAD array

hardware continues to improve, our hope is that our computational method will open new avenues for imaging where it was previously infeasible.

Appendix A

Proof of Convexity of Reflectivity Likelihood Equation

The SPIRAL optimization algorithm makes the assumption that the negative log-likelihood function $L(\boldsymbol{\theta})$ is proper convex—that is, the negative log-likelihood $L(\boldsymbol{\theta})$ is convex and $L(\boldsymbol{\theta}) > -\infty$ for every x. We show here that our likelihood equations satisfy this assumption and are therefore appropriate for the SPIRAL algorithm.

From Eq. (2.3), we have the negative log-likelihood of reflectivity:

$$\begin{aligned} L_{\boldsymbol{\alpha}}(\alpha_{i,j}; k_{i,j}) &= -\log \binom{N_P}{k_{i,j}} + (N_P - k_{i,j})(\eta_{i,j}\alpha_{i,j}S_{i,j} + B_{i,j}) - k_{i,j}\log(1 - e^{-(\eta_{i,j}\alpha_{i,j}S_{i,j}+B_{i,j})}) \\ &\approx (N_P - k_{i,j})(\eta_{i,j}\alpha_{i,j}S_{i,j} + B_{i,j}) - k_{i,j}\log(\eta_{i,j}\alpha_{i,j}S_{i,j} + B_{i,j}) \end{aligned}$$

First derivative:

$$\frac{\partial L_{\boldsymbol{\alpha}}}{\partial \alpha_{i,j}} = (N_P - k_{i,j})\eta_{i,j}S_{i,j} - \frac{k_{i,j}\eta_{i,j}S_{i,j}}{\eta_{i,j}\alpha_{i,j}S_{i,j} + B_{i,j}} \tag{A.1}$$

Second derivative:

$$\frac{\partial^2 L_{\boldsymbol{\alpha}}}{\partial \alpha_{i,j}^2} = \frac{k_{i,j}(\eta_{i,j}S_{i,j})^2}{(\eta_{i,j}\alpha_{i,j}S_{i,j} + B_{i,j})^2} \geq 0 \tag{A.2}$$

The second derivative of $L_{\boldsymbol{\alpha}}$ with respect to $\alpha_{i,j}$ is always nonnegative, so $L_{\boldsymbol{\alpha}}$ is convex and can be optimized efficiently.

Appendix B

Optimal Beam Size

In this appendix, we explain our derivation of the optimal beam size referred to in Section 2.1.1. For our discussion here, we will define the 2-D Gaussian used to generate the beam profile as:

$$h_g(x,y) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2}, \tag{B.1}$$

where x and y are in pixel units. With this definition, we place the center of both the beam and the array at $(0,0)$, with the array bounds from $-M/2 \leq x, y \leq M/2$. This function is normalized such that $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h_g(x,y)dxdy = 1$.

The power expended by the laser over some area is proportional to the volume beneath the 2-D surface $h_g(x,y)$ within that area. We can thus write the fraction of the total expended laser power that falls within the array's field of view as a function of σ:

$$\begin{aligned}
\frac{P_{ARR}(\sigma)}{P_{TOT}} &= \frac{\int_{-M/2}^{M/2}\int_{-M/2}^{M/2} h_g(x,y)dxdy}{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h_g(x,y)dxdy} \\
&= \frac{1}{2\pi\sigma^2}\int_{-M/2}^{M/2}\int_{-M/2}^{M/2} e^{-(x^2+y^2)/2\sigma^2}dxdy \\
&= \left[\mathrm{erf}\left(\frac{M}{2\sqrt{2}\sigma}\right)\right]^2.
\end{aligned} \tag{B.2}$$

We will first calculate the beam size such that the intensity profile is fairly uniform over the entire array. Let σ_u be the value of σ such that the entire array is within a $1-1/e^2$ beam radius. We can calculate σ_u by setting $h_g(\frac{M}{2}, \frac{M}{2})$ equal to $1 - 1/e^2$ times the peak value of $h_g(x, y)$, which occurs at $(x, y) = (0, 0)$:

$$h_g(\frac{M}{2}, \frac{M}{2}) = \left(1 - \frac{1}{e^2}\right) h_g(0, 0)$$

$$\frac{1}{2\pi\sigma_u^2} e^{-((\frac{M}{2})^2 + (\frac{M}{2})^2)/2\sigma_u^2} = \left(1 - \frac{1}{e^2}\right) \frac{1}{2\pi\sigma_u^2}$$

$$e^{-M^2/4\sigma_u^2} = 1 - \frac{1}{e^2}$$

$$-\frac{M^2}{4\sigma_u^2} = \ln\left(1 - \frac{1}{e^2}\right)$$

$$\sigma_u^2 = -\frac{M^2}{4\ln\left(1 - \frac{1}{e^2}\right)}$$

$$\sigma_u = \frac{M}{2\sqrt{-\ln\left(1 - \frac{1}{e^2}\right)}} \tag{B.3}$$

$$\approx 1.3112M. \tag{B.4}$$

At this beam size, the fraction of the laser's power that is utilized is:

$$\frac{P_{ARR}(\sigma_u)}{P_{TOT}} = \left[\text{erf}\left(\sqrt{-\frac{1}{2}\ln(1 - \frac{1}{e^2})}\right)\right]^2 \approx 0.09. \tag{B.5}$$

Now we want to find the beam size that maximizes the beam intensity at the corners of the array. To do this, we first write $h_g\left(\frac{M}{2}, \frac{M}{2}\right)$ and its derivative with respect to σ.

$$h_g\left(\frac{M}{2}, \frac{M}{2}\right) = \frac{1}{2\pi\sigma^2} e^{-M^2/4\sigma^2} \tag{B.6}$$

$$\frac{\partial h_g\left(\frac{M}{2}, \frac{M}{2}\right)}{\partial \sigma} = \frac{\sigma^2\left(\frac{M^2}{2\sigma^3} e^{-M^2/4\sigma^2}\right) - 2\sigma e^{-M/4\sigma^2}}{2\pi\sigma^4} \tag{B.7}$$

$$= \frac{(M^2 - 4\sigma^2) e^{-M^2/4\sigma^2}}{4\pi\sigma^5}. \tag{B.8}$$

We then set Eq. (B.8) equal to 0, and find the value σ_* that maximizes $h_g(\frac{M}{2}, \frac{M}{2})$.

$$\frac{(M^2 - 4\sigma_*^2) e^{-M^2/4\sigma_*^2}}{8\pi^2\sigma_*^5} = 0 \tag{B.9}$$

$$M^2 - 4\sigma_*^2 = 0 \tag{B.10}$$

$$\sigma_* = \frac{M}{2} \tag{B.11}$$

At this optimal beam size, the fraction of power utilized is:

$$\frac{P_{ARR}(\sigma_*)}{P_{TOT}} = \left[\text{erf}\left(\frac{1}{\sqrt{2}}\right)\right]^2 \approx 0.47. \tag{B.12}$$

At this optimal beam size, the beam's intensity profile is highly non-uniform over the array's field of view. The minimum beam intensity, at the corners of the array's field of view, is:

$$h_g\left(\frac{M}{2}, \frac{M}{2}\right) = \frac{1}{2\pi\left(\frac{M}{2}\right)^2} e^{-\frac{M^2}{4\left(\frac{M}{2}\right)^2}} = \frac{2}{\pi M^2} e^{-1}. \tag{B.13}$$

The maximum beam intensity, at the center of the beam, is:

$$h_g\left(0, 0\right) = \frac{1}{2\pi\left(\frac{M}{2}\right)^2} = \frac{2}{\pi M^2}. \tag{B.14}$$

Thus, we can expect the signal return at the center of the array to be $e \approx 2.7$ times higher than the signal return at the corners.

In practice, the optimal choice of σ will depend on the specific imaging application. In these calculations, we chose to maximize the beam intensity at the corners of the array's field of view (i.e., maximize the minimum beam intensity within the array's field of view). Another option might be to find the smallest value of σ such that the SBR at the corners is above some minimum threshold—this choice of σ would maximize the utilization of the total amount of power expended by the laser, and therefore maximize the system's photon efficiency, given the choice of the minimum allowable SBR.

Bibliography

[1] J. J. Degnan, "Photon-counting multikilohertz microlaser altimeters for airborne and spaceborne topographic measurements," *Journal of Geodynamics*, vol. 34, no. 3, pp. 503–549, 2002.

[2] M. E. O'Brien and D. G. Fouche, "Simulation of 3D laser radar systems," *Lincoln Laboratory Journal*, vol. 15, no. 1, pp. 37–60, 2005.

[3] E. Grigoriev, A. Akindinov, M. Breitenmoser, S. Buono, E. Charbon, C. Niclass, I. Desforges, and R. Rocca, "Silicon photomultipliers and their bio-medical applications," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 571, no. 1-2, pp. 130–133, 2007.

[4] D. Renker, "Geiger-mode avalanche photodiodes, history, properties and problems," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 567, no. 1, pp. 48 – 56, 2006.

[5] A. McCarthy, X. Ren, A. Della Frera, N. R. Gemmell, N. J. Krichel, C. Scarcella, A. Ruggeri, A. Tosi, and G. S. Buller, "Kilometer-range depth imaging at 1550 nm wavelength using an InGaAs/InP single-photon avalanche diode detector," *Optics Express*, vol. 21, no. 19, pp. 22 098–22 113, 2013.

[6] 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.

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

[8] F. Zappa, S. Tisa, A. Tosi, and S. Cova, "Principles and features of single-photon avalanche diode arrays," *Sensors and Actuators A: Physical*, vol. 140, no. 1, pp. 103–112, 2007.

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

[10] E. Abreu and S. Mitra, "A signal-dependent rank ordered mean (SD-ROM) filter-a new approach for removal of impulses from highly corrupted images," in *1995 International Conference on Acoustics, Speech, and Signal Processing*, vol. 4. IEEE, 1995, pp. 2371–2374.

[11] C. Veerappan, J. Richardson, R. Walker, D. Li, M. Fishburn, D. Stoppa, F. Borghetti, Y. Maruyama, M. Gersbach, R. Henderson *et al.*, "Characterization of large-scale non-uniformities in a 20k TDC/SPAD array integrated in a 130nm CMOS process," in *2011 Proceedings of the European Solid-State Device Research Conference (ESSDERC)*. IEEE, 2011, pp. 331–334.

[12] M. Itzler, U. Krishnamachari, M. Entwistle, X. Jiang, M. Owens, K. Slomkowski *et al.*, "Dark count statistics in Geiger-mode avalanche photodiode cameras for 3-D imaging LADAR," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 20, no. 6, pp. 318–328, 2014.

[13] 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.

[14] R. D. Younger, K. A. McIntosh, J. W. Chludzinski, D. C. Oakley, L. J. Mahoney, J. E. Funk, J. P. Donnelly, and S. Verghese, "Crosstalk analysis of integrated Geiger-mode avalanche photodiode focal plane arrays," in *Proc. SPIE*, vol. 7320, 2009, pp. 73 200Q–1.

[15] I. Rech, A. Ingargiola, R. Spinelli, I. Labanca, S. Marangoni, M. Ghioni, and S. Cova, "Optical crosstalk in single photon avalanche diode arrays: a new complete model," *Optics Express*, vol. 16, no. 12, pp. 8381–8394, 2008.

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

[17] D. Shin, J. H. Shapiro, and V. K Goyal, "Single-photon depth imaging using a union-of-subspaces model," *arXiv preprint arXiv:1507.06985*, 2015.

[18] A. N. Vasile, L. J. Skelly, M. E. O'Brien, D. G. Fouche, R. M. Marino, R. Knowlton, M. J. Khan, and R. M. Heinrichs, "Advanced coincidence processing of 3d laser radar data," in *Advances in Visual Computing*. Springer, 2012, pp. 382–393.