Computational Time-resolved Imaging

by

Ghulam A. Kirmani

S.M., Media Technology
Massachusetts Institute of Technology, 2010

Integrated M.Tech in Mathematics and Computing
Indian Institute of Technology Delhi, 2008

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

Doctor of Philosophy in Electrical Engineering and Computer Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2015

Signature redacted

Author .....
Department of Electrical Engineering and Computer Science
December 15, 2014

Certified by ..... Signature redacted .....
Jeffrey H. Shapiro
Julius A. Stratton Professor of Electrical Engineering
Massachusetts Institute of Technology
Thesis Supervisor

Signature redacted

Certified by .....
Vivek K Goyal
Assistant Professor of Electrical and Computer Engineering
Boston University
Thesis Supervisor

Accepted by ..... Signature redacted .....
Leslie A. Kolodziejski
Chair, Department Committee on Graduate Students

Computational Time-resolved Imaging

by

Ghulam A. Kirmani

Submitted to the Department of Electrical Engineering and Computer Science
on December 15, 2014, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Electrical Engineering and Computer Science

Abstract

Classical photography uses steady-state illumination and light sensing with focusing optics to capture scene reflectivity as images; temporal variations of the light field are not exploited. This thesis explores the use of time-varying optical illumination and time-resolved sensing along with signal modeling and computational reconstruction. Its purpose is to create new imaging modalities, and to demonstrate high-quality imaging in cases in which traditional techniques fail to even form degraded imagery. The principal contributions in this thesis are the derivation of physically-accurate signal models for the scene's response to time-varying illumination and the photodetection statistics of the sensor, and the combining of these models with computationally tractable signal recovery algorithms leading to image formation.

In active optical imaging setups, we use computational time-resolved imaging to experimentally demonstrate: non line-of-sight imaging or looking around corners, in which only diffusely scattered light was used to image a hidden plane which was completely occluded from both the light source and the sensor; single-pixel 3D imaging or compressive depth acquisition, in which accurate depth maps were obtained using a single, non-spatially resolving bucket detector in combination with a spatial light modulator; and high-photon efficiency imaging including first-photon imaging, in which high-quality 3D and reflectivity images were formed using only the first detected photon at each sensor pixel despite the presence of high levels of background light.

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

Thesis Supervisor: Vivek K Goyal
Title: Assistant Professor of Electrical and Computer Engineering
Boston University

Acknowledgments

The work in this thesis has been the culmination of the support and inspiration I derived from many people throughout my academic career. But my experience in the doctoral program itself has been very enriching in many dimensions which I credit to the stimulating environment at MIT, and in particular at the Research Laboratory of Electronics. I would like to express my gratitude to each person who made my time here so much more meaningful and interesting.

To my advisor Vivek Goyal, for accepting me to the doctoral program in EECS and welcoming me to his group even before I was officially a graduate student in the department. I am very grateful for his patience and strong support to new ideas, and for truly manifesting that the sky is the limit for any willing and eager graduate student and for fighting hurdles to support his students' interests even when it meant he had to go out of his way. As the leader of the Signal Transformation and Information Representation group, he created an environment complete with resources, collaborators, opportunities and encouragement there was no excuse for a STIR member to not try hard problems. I would like to thank him for this unique work environment he created and maintained.

To my advisor in the second half of my doctoral program, Jeff Shapiro, who has been a great mentor with very valuable critic at every step. Through his principled and rigorous approach to optical problems, I have become more a rigorous researcher myself.

In my graduate school career I had the unfortunate or fortunate experience of academic-style differences that escalated to unnecessary conflict. I am grateful to Jeff for helping me navigate these challenges and for teaching me the invaluable lesson of doing the right thing even when the situation is seemingly not in your favor. His judgment and wisdom will be an important part of my learning.

To my thesis reader Pablo Parrilo for his rigorous research and teaching style, and for teaching me that the *devil is in the little details* when it comes to doing good research. He was my inspiration to come to MIT for graduate school.

To my collaborators in the Optical and Quantum Communications Group, Dr. Franco Wong and Dheera Venkatraman for their insightful contributions to experiments in this

thesis. I am grateful to them for their willingness to share lab space and equipment which strongly supplemented all the theoretical work and provided concrete experimental results.

To the leaders in signal processing, optimization and information theory with whom I have had several opportunities to interact and learn from Professors Sanjoy Mitter, John Tsitsitklis, Devavrat Shah, George Verghese, Yury Polyanskiy, Moe Win, Greg Wornell.

To the STIR family of students Lav Varshney, Dan Weller, John Sun, Joong Bum Rhim for being great team members, for sharing their grad school experiences with me and brainstorming ideas with me when I was new to the group. To STIR fellow-student Dongeek Shin with whom I collaborated intensely for two years, who diligently found creative solutions to our research problems, and for many stimulating discussions on theoretical bounds to various problems we investigated together. I would also like to thank my collaborator Hye Soo Yang for all her help and expertise. I especially thank Andrea Colaço for embarking on many undefined problems and experiments with me when we first began work on 3D optical imaging in the STIR group, and for being a wise mentor on several key occasions. Several other friends at MIT have made my time really memorable and fun. I would like to thank Parikshit Shah, Vaibhav Rathi, Noah Stein, Raghavendra Hosur, Anirudh Sarkar, and Aditya Bhakta for the good times.

I dedicate this thesis to the memory of my late father who would have been very happy and proud today. I would also like to thank my mother for her infinite patience and support, my grandparents who raised me with great love and care and provided the best opportunities at every step, and my younger brother and sister who have been such wonderful supporters and lifelong friends.

This work in this thesis was supported by the U.S. NSF under grant numbers 1161413, 0643836 and 1422034, a Qualcomm Innovation Fellowship, a Microsoft Research Ph.D. Fellowship, and a Texas Instrument research grant.

Contents

Chapter 1

Introduction

The goal of imaging is to produce a representation in one-to-one spatial correspondence with an object or scene. For centuries, the primary technical meaning of image has been a visual representation formed through the interaction of light with mirrors and lenses, and recorded through a photochemical process. In digital photography, the photochemical process has been replaced by an electronic sensor array, but the use of optical elements is unchanged. Conventional imaging for capturing scene reflectivity, or *photography*, uses natural illumination and light sensing with focusing optics; variations of the light field with time are not exploited.

A brief history of time in optical imaging: Figure 1-1 shows a progression of key events which depict the increasing exploitation of temporal information contained in optical signals to form 3D and reflectivity images of a scene. The use of time in conventional photography is limited to the selection of a shutter speed (exposure time). The amount of light incident on the sensor (or film) is proportional to the exposure time, so it is selected to match the dynamic range of the input over which the sensor is most effective. If the scene is not static during the exposure time, motion blur results. Motion blur can be reduced by having a shorter exposure, with commensurate increase of the gain on the sensor (or film speed) to match the reduced light collection. Light to the sensor can also be increased by employing a larger aperture opening, at the expense of depth of field.

Moving away from conventional photography, the use of high-speed sensing in imaging

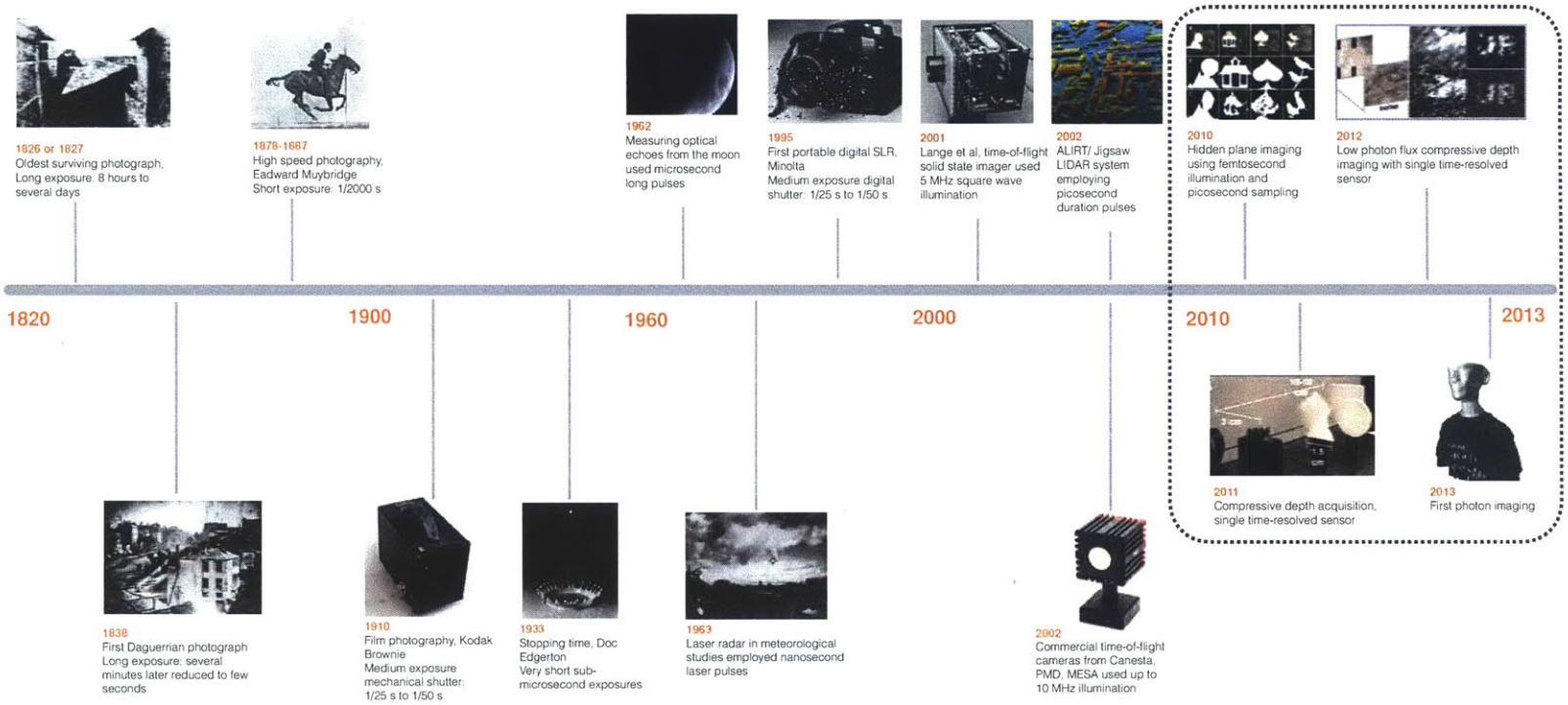


Figure 1 1: **A brief history of the use of time in optical imaging.** The thesis contributions are shown in the dotted box. Photo credits include Wikimedia Creative Commons, Edgerton photos are © 2010 MIT Courtesy of MIT Museum, © MIT Lincoln Laboratory, and related publications [1–8]

is associated with stopping motion. Very short exposures or flash illuminations are used to effectively "stop time" [4]. These methods, however, could still be called *atemporal* because even a microsecond flash is long enough to combine light from a large range of transport paths involving many possible reflections. No temporal variations are present at this time scale (nor does one attempt to capture them), so no interesting inferences can be drawn.

Optical range imaging is one of the first applications that employed time-varying illumination and time-resolved detection. Range measurement by detection of weak optical echoes from the moon was reported more than half a century ago [1]. Time-of-flight range measurement systems [9, 10] exploit the fact that the speed of light is finite. By measuring the time-delay introduced by roundtrip propagation of light from the imager to the scene and back, range can be accurately estimated. Various time-of-flight imagers exist today, and differ in how they modulate their illumination and perform time-resolved photodetection.

In this thesis, we present a radical departure from both high speed photography and time-of-flight systems. We develop computational imaging frameworks that are fundamentally rooted in principled signal modeling of the temporal variations in the light signals in order to solve challenging inverse problems and achieve accurate imaging for cases in which traditional methods fail to form even degraded imagery. In addition, by conducting proof-of-concept experiments, we have corroborated the advantages and demonstrated the practical feasibility of the proposed approaches.

Active optical imaging: In this thesis we are primarily interested in acquiring the scene's 3D structure (or scene depth) and reflectivity using an active imager one that supplies its own illumination. In such an imager, the traditional light source is replaced with a time-varying illumination that induces time-dependent light transport between the scene and sensor. The back-reflected light is sensed using a time-resolved detector instead of a conventional detector, which lacks sufficient temporal resolution. For mathematical modeling, we assume illumination with a pulsed light source and time-resolved sensing with a square-law detector whose output is linearly proportional to intensity of incident light. Figure 1-2 shows and discusses the various optoelectronic elements of an active optical imaging system used in this thesis. In this thesis we will focus on three specific imaging scenarios to demonstrate

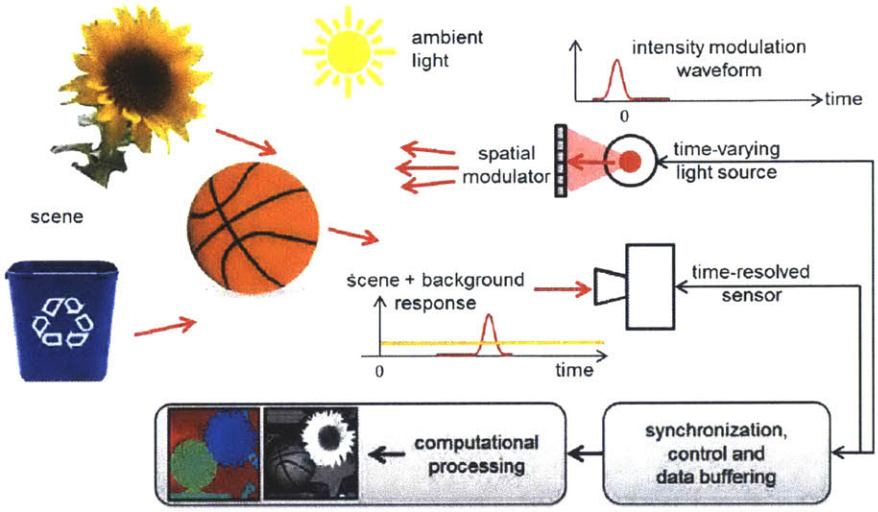


Figure 1 2: **Computational time-resolved imaging setup** The illumination is a periodically pulsed light source. The light from the source may also be spatially modulated. The scene of interest comprises life size objects at room scale distances. The light incident on the time resolved detector is a sum of the ambient light flux and the back reflected signal. The light source and detector are time synchronized. The collected data comprises time samples of the photocurrent produced by the incident optical flux in response to the scene illumination. The data is computationally processed using the appropriate signal models in order to form reflectivity or depth maps of the scene, or both.

the computational time-resolved imaging framework and its benefits over existing methods. These scenarios are described in the next section.

1.1 Summary of Main Contributions

This thesis introduces three imaging scenarios in which time-dependent illumination and sensing combined with physically-accurate signal modeling and computational reconstruction play central roles in acquiring the scene parameters of interest. As demonstrated through proof-of-concept experiments, traditional imaging methods can at best form degraded imagery in the scenarios considered in this thesis. These imaging scenarios are:

1. Looking around corners using pulsed illumination and time-resolved detection of diffusely scattered light (Chapter 2).
2. Compressive depth acquisition, which is developed in two distinct imaging setups:
   2a. One that uses a single bucket photodetector and a spatial light modulator to project pulsed illumination patterns on to the scene (Chapter 3) and,
   2b. a low-light imaging variant of compressive depth acquisition, which employs floodlight illumination, a single-photon counting detector and a digital micro-mirror device for spatial patterning (Chapter 4).
3. Photon-efficient active optical imaging, which is also developed in two different low-light imaging configurations:
   3a. The first-photon computational imager, which acquires scene reflectivity and depth using the first detected photon at each sensor pixel (Chapter 5) and,
   3b. an extension of the first-photon imaging framework for implementation with sensor arrays (Chapter 6).

The imaging setups and problem statements for each of the above scenarios are shown and discussed in Figs. 1-3-1-5. A brief overview and comparison of thesis contributions is presented in Table 1.1. The rest of the thesis is organized into chapters corresponding to the aforementioned scenarios. Within each section we discuss:

1. Overview of the problem.
2. Prior art and comparison with proposed computational imager.
3. Imaging setup, data acquisition, and measurement models.
4. Novel image formation algorithms.
5. Experimental setup and results.
6. Discussion of limitations and extensions.

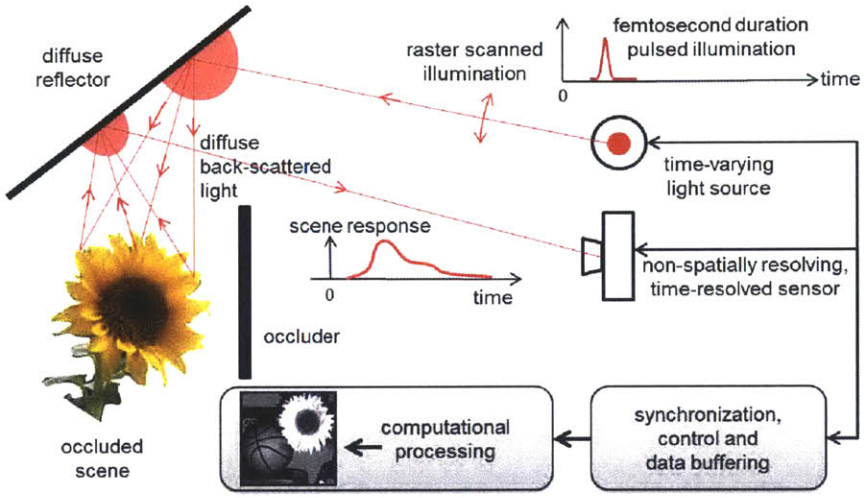


Figure 1 3: **Looking around corners using ultrashort pulsed illumination and picosecond-accurate time-sampling of diffusely scattered light:** Here the goal is to reconstruct the position and reflectivity of an object that is occluded from the light source and the sensor. We achieve this chal lenging task by raster scanning a diffuser with ultrashort pulsed illumination and by time sampling the backscattered light with picosecond accuracy. The computational processing of this time resolved data using elliptical Radon transform inversion reveals the hidden object's 3D position and reflectivity.

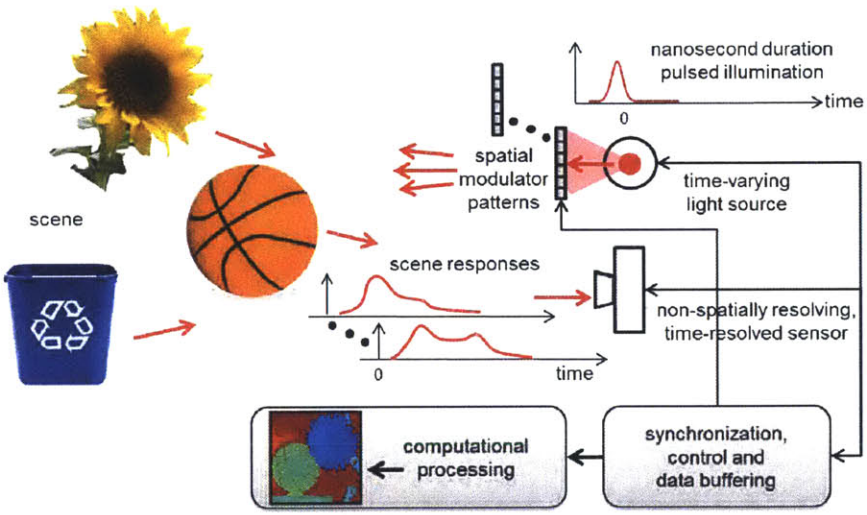


Figure 1 4: **Compressive depth acquisition using structured random illumination and a single time-resolved sensor:** Here, we are interested in compressively acquiring scene depth using structured illumination and single time resolved sensor. This problem may seem analogous to the single pixel camera [11, 12], but unlike reflectivity, it is not possible to measure linear projections of scene depth. We solve this problem for planar and fronto parallel scenes using parametric signal modeling of the scene impulse response and using finite rate of innovation methods to reconstruct the scene's impulse responses using time resolved illumination and detection.

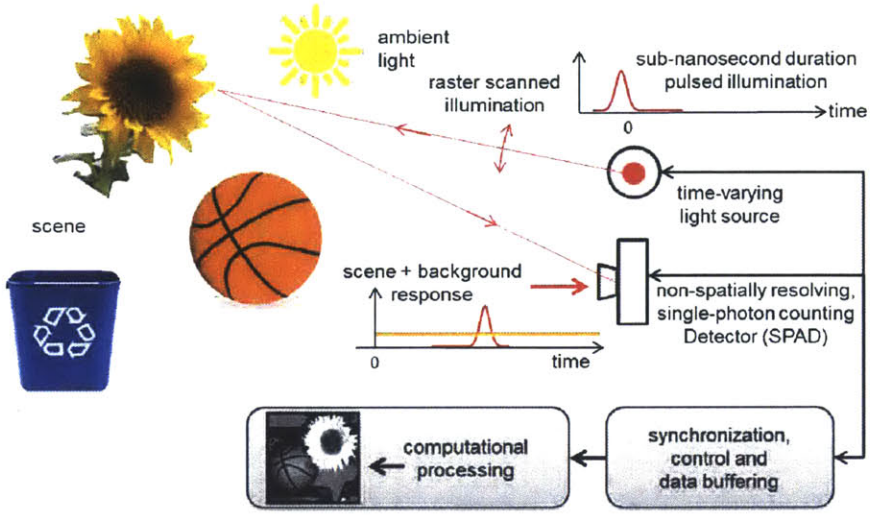


Figure 1 5: **3D and reflectivity imaging using one detected photon per pixel:** Here, we are interested in reconstructing accurate 3D and reflectivity images of the scene using only one detected photon per pixel, even in the presence of strong background light. We achieve this using physically accurate modeling of single photon photodetection statistics and exploitation of spatial correlations present in real world scenes.

1.2 Key Publications

Most of the material presented in this thesis has previously appeared in the following publications and manuscripts listed below:

Looking Around Corners or Hidden-Plane Imaging

1. A. Kirmani, A. Velten, T. Hutchison, M. E. Lawson, V. K. Goyal, M. G. Bawendi, and R. Raskar, *Reconstructing an image on a hidden plane using ultrafast imaging of diffuse reflections*, submitted, May 2011.

2. A. Kirmani, H. Jeelani, V. Montazerhodjat, and V. K. Goyal, *Diffuse imaging: Creating optical images with unfocused time-resolved illumination and sensing*, IEEE Signal Processing Letters, **19** (1), pp. 31-34, October 2011.

Compressive Depth Acquisition

1. A. Kirmani, A. Colaço, F. N. C. Wong, and V. K. Goyal, *Exploiting sparsity in time-of-flight range acquisition using a single time-resolved sensor*, Optics express **19** (22), pp. 21485-21507, October 2011.

2. A. Colaço, A. Kirmani, G. A. Howland, J. C. Howell, and V. K. Goyal, *Compressive depth map acquisition using a single photon-counting detector: Parametric signal processing meets sparsity*, In IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), pp. 96-102, June 2012.

Photon-efficient Active Optical Imaging

1. A. Kirmani, D. Venkatraman, D. Shin, A. Colaço, F. N. C. Wong, J. H. Shapiro, and V. K. Goyal, *First-Photon Imaging*, Science **343** (6166), pp. 58-61, January 2014.

2. D. Shin, A. Kirmani, V. K. Goyal, and J. H. Shapiro, *Photon-Efficient Computational 3D and Reflectivity Imaging with Single-Photon Detectors*, arXiv preprint arXiv:1406.1761, June 2014.

Thesis contribution	Temporal modulation	Spatial modulation	Sensor	Signal model	Inversion algorithm	Comparison to closest prior art
Looking around corners using time resolved detection of diffusely scattered light	Femtosecond pulsed laser	Raster scanning of the visible diffuser	Picosecond accurate streak camera	Elliptical Radon transform (ERT)	Linearized ERT inversion	Existing diffuse imaging methods require phase based speckle imaging [13]. In dual photography [14], the hidden object is visible to the light source
Compressive depth acquisition	Nanosecond pulsed laser illumination	Binary valued spatial light modulation	One omnidirectional photodetector	Parametric modeling of scene impulse response	Finite rate of innovation and wavelet domain sparsity	Single pixel camera only captures reflectivity images [11]. Recently proposed compressive depth acquisition methods work only for fronto parallel scenes [15]
First photon imaging	Picosecond pulsed laser illumination	Pixelwise raster scanning of the scene	Single photon avalanche photodiode (50 picosecond timing jitter)	Time inhomogeneous Poisson process	Penalized maximum likelihood estimation	Direct detection laser radar systems requires tens to hundreds of photons at each sensor pixel [16]

Table 1.1: **Summary of thesis contributions, highlighting the key differences and comparison with prior art.**

Chapter 2

Looking Around Corners

2.1 Overview

Conventional imaging involves direct line-of-sight light transport from the light source to the scene, and from the scene back to the camera sensor. Thus, opaque occlusions make it impossible to capture a conventional image or photograph of a hidden scene without the aid of a view mirror to provide an alternate path for unoccluded light transport between the light source, scene and the sensor. Also, the image of the reflectivity pattern on scene surfaces is acquired by using a lens to focus the reflected light on a two-dimensional (2D) array of light sensors. If the mirror in the aforementioned *looking around corners* scenario was replaced with a Lambertian diffuser, such as a piece of white matte paper, then it becomes impossible to capture the image of the hidden-scene using a traditional light source and camera sensor. Lambertian scattering causes loss of the angle-of-incidence information and results in mixing of reflectivity information before it reaches the camera sensor. This mixing is irreversible using conventional optical imaging methods.

In this chapter we introduce a computational imager for constructing an image of a static hidden-plane that is completely occluded from both the light source and the camera, using only a Lambertian surface as a substitute for a view mirror. Our technique uses knowledge of the orientation and position of a hidden-plane relative to an unoccluded (visible) Lambertian diffuser, along with multiple short-pulsed illuminations of this visible diffuser and time-

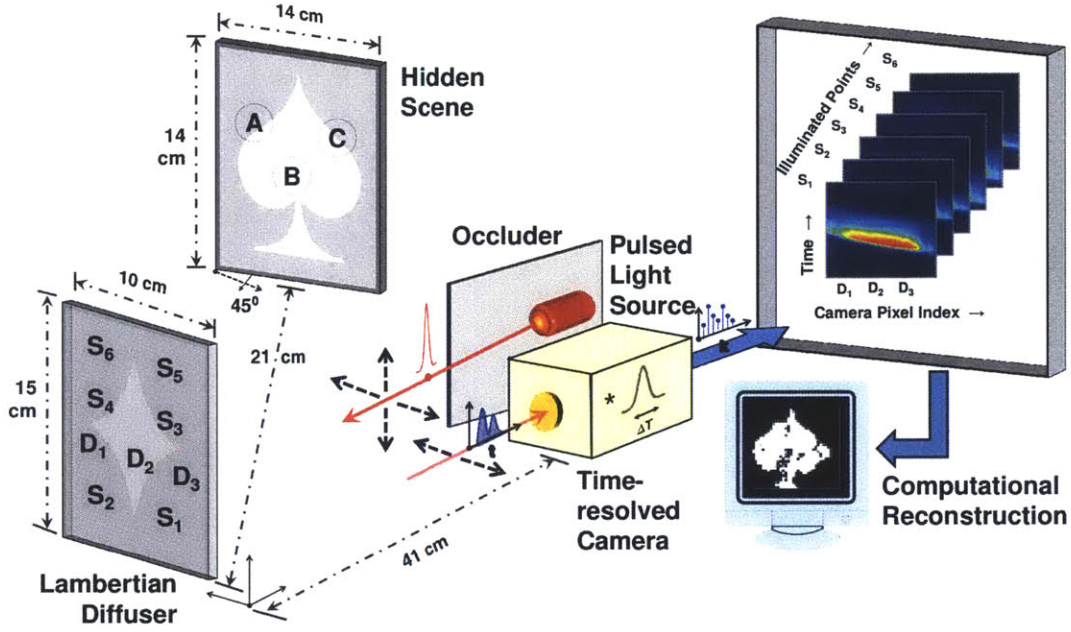


Figure 2-1: **Looking around corners imaging setup for hidden-plane reflectivity construction.**

resolved detection of the backscattered light, to computationally construct a black-and-white image of the hidden-plane.

The setup for looking around corners to image a static hidden-plane is shown in Fig. 2-1. A pulsed light source serially illuminates selected points, S_i, on a static unoccluded Lambertian diffuser. As shown, light transport from the illumination source to the hidden-scene and back to the camera is indirect, possible only through scattering off the unoccluded Lambertian diffuser. The light pulses undergo three diffuse reflections, diffuser $\rightarrow$ hidden-plane $\rightarrow$ diffuser, after which they return to a camera co-located alongside the light source. The optical path differences introduced by the multipath propagation of the scattered light generates a time profile that contains the unknown hidden-plane reflectivity (the spade). A time-resolved camera that is focused at the visible diffuser and accurately time-synchronized with the light source time-samples the light incident from observable points on the diffuser, D_j. Given the knowledge of the complete scene geometry, including the hidden-plane orientation and position, the time-sample data collected by interrogating several distinct source-detector pairs, (S_i, D_j), is computationally processed using a linear inversion framework (see Figs. 2-2 and 2-8) to construct an image of the hidden-plane (see Fig. 2-9).

The remainder of this chapter is organized as follows. Prior art is discussed in Section 2.2. Then Section 2.3 introduces the imaging setup and signal model for the proposed looking around corners computational imager. Next, Section 2.4 presents the scene response model. The measurement model and data acquisition pipeline are described in Section 2.5. These models form the basis for the novel hidden-plane reflectivity estimation algorithm developed in Section 2.6. The experimental setup and results are presented in Section 2.7 and finally Section 2.8 provides additional discussion of limitations and extensions.

2.2 Prior Art and Challenge

Non line-of-sight imaging: Capturing images despite occlusions is a hard inverse imaging problem for which several methods have been proposed. Techniques that use whatever light passes through the occluder have been developed. With partial light transmission through the occluder, time-gated imaging [17,18] or the estimation and inversion of the mesoscopic transmission matrix [13] enables image formation. While some occluding materials, such as biological tissue, are sufficiently transmissive to allow light to pass through them [19,20], high scattering or absorption coefficients make image recovery extremely challenging. There are also numerous attempts to utilize light diffusely-scattered by tissue to reconstruct embedded objects [13, 21]. Our proposed computational imager is based on the pulsed illumination of the hidden-scene by reflecting off the visible diffuser followed by time-sampling of the backscattered light, rather than on transmission through the occluder itself. As opposed to prior work on non line-of-sight imaging [13,19 21] we image around the occluder rather than through it, by exploiting the finite speed of light.

Illumination wavelengths may be altered to facilitate imaging through occluders, as in X-ray tomography [22] or RADAR [23, 24]. However, such imaging techniques do not yield images in the visible spectrum that are readily interpreted by a human observer.

Hidden-scene image capture has been demonstrated, provided there is direct line-of-sight at least between the light source and the scene, when the scene is occluded with respect to the sensor [14]. This was accomplished by raster scanning the scene by a light source followed by computational processing that exploits Helmholtz reciprocity of light transport.

Looking around corners using diffuse reflections: The closest precedent to the methods proposed in this chapter experimentally demonstrated the recovery of hidden-scene structure for one-dimensional scenes, entirely using diffusely-scattered light [25 27]. Selected visible scene points were illuminated with directional short pulses of light using one-dimensional femtosecond laser scanning. The diffuse, scattered light was time-sampled using a single-pixel detector with picosecond resolution. The resulting two-dimensional dataset was processed to compute the time-delay information corresponding to the different patches in the hidden-scene. Then, standard triangulation algorithms were used to estimate the geometric parameters of a simple occluded scene. The geometry reconstructions were noted to be highly susceptible to failure in the presence of noise and timing jitter. Also, the experimental validation of these methods was limited to a very simple one-dimensional scene comprising only three mirrors. While [25 27] used time-of-arrival measurements to enable estimation of hidden-scene geometry, estimation of hidden-scene reflectivity, which is also contained in the light signal, has not been accomplished previously.

Synthetic aperture radar: The central contribution in this chapter is to demonstrate the use of temporal information contained in the scattered light, in conjunction with appropriate post-measurement signal processing, to form images that could not have been captured using a traditional camera, which employs focusing optics but does not possess high temporal resolution. In this connection it is germane to compare our proposed computational imager to synthetic aperture radar (SAR), which is a well-known microwave approach for using time-domain information plus post-measurement signal processing to form high spatial-resolution images [28 30]. In stripmap mode, an airborne radar transmits a sequence of high-bandwidth pulses on a fixed slant angle toward the ground. Pulse-compression reception of individual pulses provides across-track spatial resolution superior to that of the radar's antenna pattern as the range response of the compressed pulse sweeps across the ground plane. Coherent integration over many pulses provides along-track spatial resolution by forming a synthetic aperture whose diffraction limit is much smaller than that of the radar's antenna pattern.

SAR differs from the proposed computational imager in two general ways. First, SAR requires the radar to be in motion to scan the scene, whereas the proposed imager does

not require sensor motion because it relies on raster scanning the beam on the visible diffuser. Second, SAR is primarily a microwave technique, and most real-world objects have a strongly specular bidirectional reflectance distribution function [31] (BRDF) at microwave wavelengths. With specular reflections, an object is directly visible only when the angle of illumination and angle of observation satisfy the law of reflection. Multiple reflections which are not accounted for in first-order SAR models can then be strong and create spurious images. On the other hand, most objects are Lambertian at optical wavelengths, so our imager operating at near-infrared wavelengths avoids these sources of difficulty.

Time-resolved imaging of natural scenes: Traditional imaging involves illumination of the scene with a light source whose light output does not vary with time. The use of high-speed sensing in photography is associated with stopping motion [4]. In this thesis, we use time-resolved sensing differently: to differentiate among paths of different lengths from light source to scene to sensor. Thus, as in time-of-flight depth cameras [9,10] that are covered in detail in Chapter 3, we are exploiting the speed of light being finite. However, rather than using the duration of delay to infer distances, we use temporal variations in the backreflected light to infer scene reflectivity. This is achieved through a computational unmixing of the reflectivity information that is linearly combined at the sensor because distinct optical paths may have equal path lengths (see Fig. 2-2).

Some previous imagers, like light detection and ranging (LIDAR) [9], systems have employed pulsed illumination and highly sensitive single-photon counting time-resolved detectors to more limited effect. Elapsed time from the pulsed illumination to the backreflected light is proportional to target distance and is measured by constructing a histogram of the photon-detection times. The operation of LIDAR systems is discussed in detail in Chapter 5. In the context of LIDAR systems, time-based range gating has been used to reject unwanted direct reflections in favor of the later-arriving light from the desired scene [32 34]. Capturing multiple such time-gated images allows for construction of a three-dimensional model of the scene [17,18], as well as for imaging through dense scattering media, such as fog or foliage [33,35]. As opposed to the methods proposed in this chapter, however, the aforementioned LIDAR-based techniques also rely on using temporal gating to separate the direct,

unscattered component of the backreflected light from the diffusely-scattered component.

Challenge in Hidden-Plane Image Formation

To illustrate the challenge in recovering the reflectivity pattern on the hidden-scene using only the backscattered diffuse light, we pick three sample points A, B, and C on the hidden-plane. Suppose we illuminate the point S_1 on the visible Lambertian diffuser with a very short laser pulse (Fig. 2-2A), and assume that this laser impulse is scattered uniformly in a hemisphere around S_1 and propagates toward the hidden-plane (Fig. 2-2B). Since the hidden points A, B, and C may be at different distances from S_1, the scattered light reaches them at different times. The light incident at A,B and C undergoes a second Lambertian scattering and travels back towards the visible diffuser with attenuations that are directly proportional to the unknown reflectivity values at these points (see Fig. 2-2C).

The visible diffuser is in sharp focus of the camera optics, so each sensor pixel is observing a unique point on the diffuser. We pick a sample point D_1 that functions as a bucket detector collecting backscattered light from the hidden-plane without angular (directional) sensitivity and reflecting a portion of it towards the camera. The hidden points A, B, and C may be at different distances from D_1 as well. Hence, light scattered from these points reaches the camera at different times. The times-of-arrival of the light corresponding to the distinct optical paths, light source$\rightarrow$ S_1 $\rightarrow$(A, B, C)$\rightarrow$ D_1 $\rightarrow$camera, depend on the scene geometry (see Fig. 2-2D), which we assume to be known.

Diffuse scattering makes it impossible to recover the hidden-plane reflectivity using traditional optical imaging. An ordinary camera with poor temporal resolution records an image using an exposure time that is at least several microseconds long. All of the scattered light, which includes the distinct optical paths corresponding to A, B, and C, arrives at the sensor pixel observing D_1 during this exposure time and is summed into a single intensity value (see Fig. 2-2E), making it impossible to recover the hidden-plane reflectivity.

We now discuss how temporal sampling of the backscattered light allows at least partial recovery of hidden-plane reflectivity. Suppose now that the camera observing the scattered light is precisely time-synchronized with the outgoing laser pulse, and possesses a very high

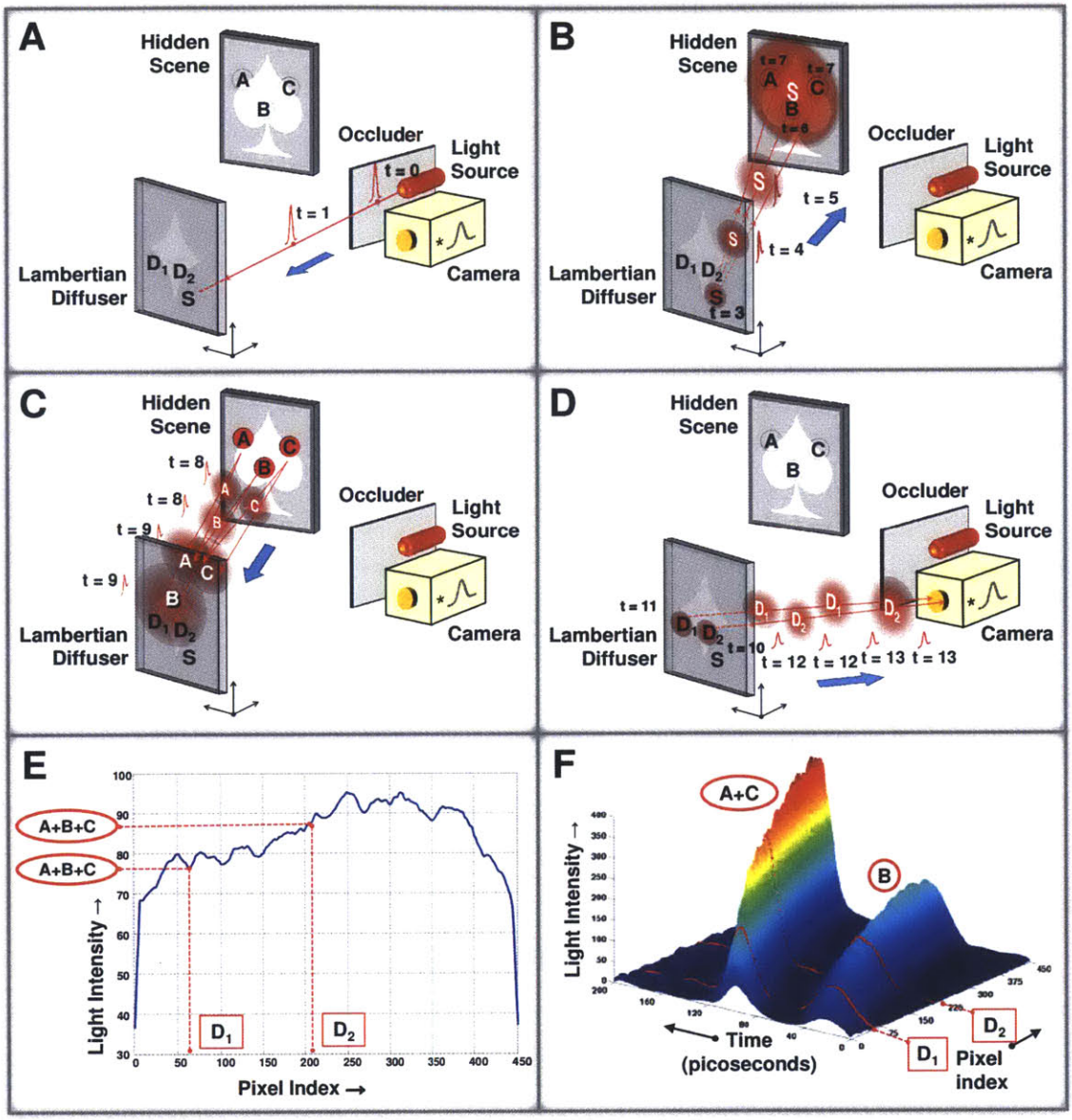


Figure 2-2: **Challenge in hidden-plane image formation.** (A)-(D) Time-dependent light transport diagrams showing diffuse scattering of the short pulse illumination through the imaging setup and back to the camera. The hidden point B has the shortest total path length, $S_1 \rightarrow \mathrm{B} \rightarrow D_1$, and the points A and C have equal total path lengths. (E) An ordinary camera with microsecond long exposure integrates the entire scattered light time profile into a single pixel value that is proportional to the sum of all hidden plane reflectivity values. (F) Exploiting the differences in total path lengths allows partial recovery of hidden plane reflectivity. For example, reflectivity at B is easily distinguishable from that at points A and C via temporal sampling of the scattered light. Several points on the hidden plane, such as A and C, have equal total path lengths and even using a time resolved camera, we may only record a sum of their reflectivity values.

time-sampling resolution. As shown in Fig. 2-2(F), the measured time-samples can potentially be used to estimate the reflectivity of the hidden points, A, B and C. However, there

may be multiple points on the hidden-plane such that the light rays corresponding to these points have equal optical path lengths and therefore have the same time-of-arrival at the camera sensor. We call this phenomena multipath mixing: the summation of the hidden-plane reflectivity information due to equal total optical path lengths. Although undesirable, this mixing is reversible.

As described in Section 2.4, the hidden-plane reflectivity values that are added together belong to a small set of points on the hidden-plane that lie within a known elliptical shell. As shown in Section 2.6, spatio-temporal sampling followed by a linear inversion allows the complete recovery of hidden-plane reflectivity from time-resolved measurements.

2.3 Imaging Setup

Our imaging setup (see Fig. 2-1) comprises a pulsed collimated light source serving as our directional illumination source which is co-located with a time-resolved camera that time-samples the backscattered light. The hidden-scene consists of a single occluded hidden-plane composed of Lambertian material with an unknown reflectivity pattern. We assume that the 3D orientation, position and dimensions of the hidden-plane are known *a priori*. If these geometric quantities are unknown, then they may be first estimated using the hidden-plane geometry estimation method described in [25 27]. We also assume that the reflectivity pattern, 3D position, and orientation of the visible Lambertian diffuser relative to the camera and light source are known. These quantities can be easily estimated using traditional imaging techniques. The complete knowledge of 3D scene geometry allows us to formulate hidden-plane reflectivity estimation as a linear inversion problem. This formulation is described next, but we first make a simplification.

Simplified illumination and sensing model: In our data acquisition setup the points of interest on the visible diffuser, $\{S_i\}_{i=1}^{6}$, are illuminated one at a time with a collimated beam of pulsed light. Since the visible diffuser is assumed to be perfectly Lambertian, it scatters the incident light uniformly in all directions. We assume that at least a part of this scattered light illuminates the entire hidden-plane for each $\{S_i\}_{i=1}^{6}$. Given this assumption,

it is possible to abstract the combination of the pulsed light source and the chosen visible diffuser points, $\{S_i\}$, into six omnidirectional pulsed light sources whose illumination profiles are computed using the surface normal at these points, the angle of incidence of the collimated beam and the reflectivity at these points.

Similarly, each ray of backreflected light from the hidden-plane is incident at the visible diffuser points, $\{D_j\}_{j=1}^{3}$, where it undergoes another Lambertian scattering and a portion of this scattered light reaches the time-resolved camera optics. The points, $\{D_j\}_{j=1}^{3}$, are in sharp focus relative to the time-resolved camera, which comprises of a horizontally oriented, linear array of sensors that time-sample the incident light profile relative to the transmitted laser pulse. Since all the orientations, distances and dimensions of the scene are assumed to be known, it is possible to abstract the combination of the time-resolved camera and the chosen visible diffuser points, $\{D_j\}$, into three time-resolved bucket detectors that are time-synchronized with the omnidirectional light sources $\{S_i\}_{i=1}^{6}$. The sensitivity profiles of these bucket detectors are known functions of the reflectivity and surface normals at these points, and their orientation relative to the hidden-plane that determines the angle of incidence of backscattered light.

In order to further simplify our mathematical notation, we assume that each of the omnidirectional light sources, $\{S_i\}_{i=1}^{6}$, transmits equal light power in all directions and that each of the bucket detectors, $\{D_j\}_{j=1}^{3}$, has a uniform angular sensitivity profile.

In effect, as shown in Fig. 2-3, we abstract the entire hidden plane imaging setup into a set of six omnidirectional pulsed light sources illuminating the hidden-plane with a common pulse shape $s(t)$, and three time-resolved bucket detectors with a common impulse response $h(t)$ that time-sample the backscattered light with sampling interval, T_s. The 3D positions of the (source, detector) pairs $\{S_i, D_j\}_{i=1,j=1}^{i=6,j=3}$ are known and they are assumed to be accurately time-synchronized to enable time-of-arrival measurements. We also assume that there is no background light and that the imager operates at single wavelength.

Hidden-plane scene setup: We have assumed that the position, orientation, and dimensions (W-by-W) of the hidden Lambertian plane are known or estimated *a priori*. Formation of an ideal grayscale image therefore is tantamount to the recovery of the reflectivity pattern

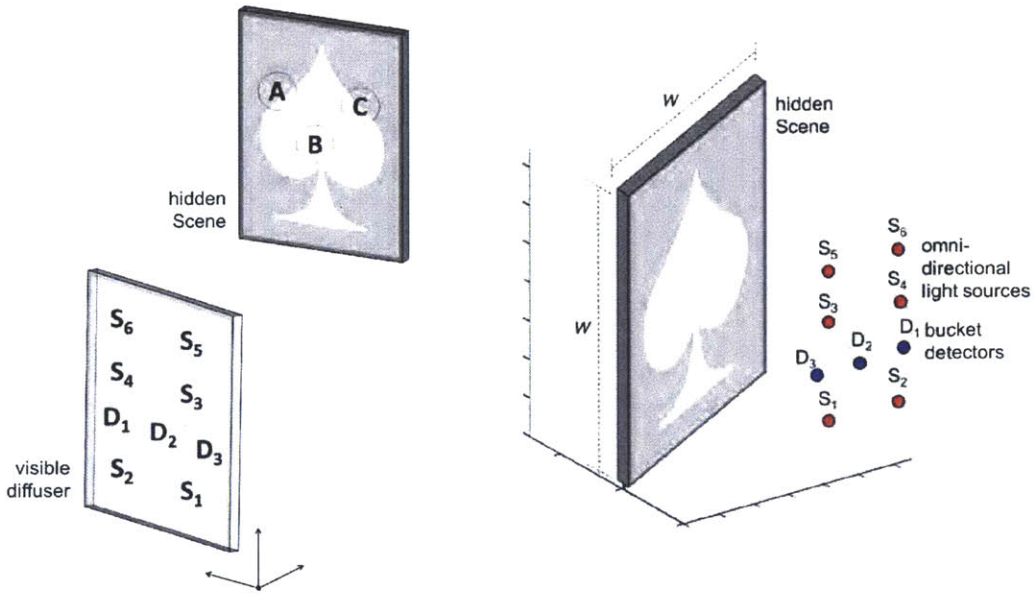
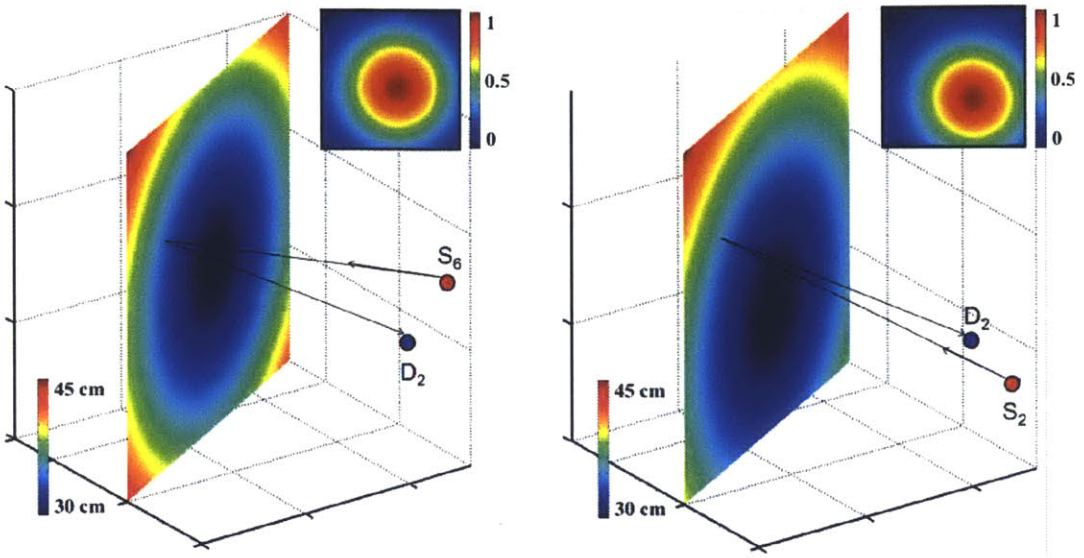


Figure 2 3: **Simplified illumination and detection model.** (left) Portion of the hidden plane imaging setup in Fig. 2 1 comprising the visible diffuser, the hidden plane, and the chosen points of illumination and detection. (right) The simplified imaging setup, under the assumptions stated in Section 2.3, is equivalent to the setup in the left sub figure.

on the hidden-plane. As is usual in optics and photometry, reflectivity is defined as the fraction of incident radiation reflected by a surface. Reflectivity is thus bounded between 0 and 1. Therefore, in this chapter, hidden-plane reflectivity is modeled as a 2D function $\boldsymbol{\alpha} : [0, W]^2 \rightarrow [0, 1]$. In general, the surface reflectivity must be treated as a directional property that is a function of the reflected direction, the incident direction, and the incident wavelength [36]. We assume, however, that the hidden-plane is Lambertian, so that its perceived brightness is invariant to the angle of observation [31]; incorporation of any known BRDF would not add insight.

2.4 Scene Response Modeling

The backreflected light incident at sensor D_j in response to hidden-scene illumination by source S_i is a combination of the time-delayed reflections from all points on the hidden-plane. For any point $\mathbf{x} = (x, y) \in [0, W]^2$, let $z_i^S(\mathbf{x})$ denote the distance from illumination source S_i to $\mathbf{x}$, and let $z_j^D(\mathbf{x})$ denote the distance from $\mathbf{x}$ to sensor D_j. Then $z_{ij}^{SD}(\mathbf{x}) = z_i^S(\mathbf{x}) + z_j^D(\mathbf{x})$



(a) source S_6 and sensor D_2 (b) source S_2 and sensor D_2

Figure 2 4: Main plots: Time delay from source to scene to sensor is a continuous function, $z_{ij}^{SD}(\mathbf{x})$, of the position in the scene. The range of $z_{ij}^{SD}(\mathbf{x})$ values shown in this figure corresponds to the scene geometry in Fig. 2 1. Insets: The normalized geometric attenuation $a_{ij}(\mathbf{x})$ of the light is also a continuous function of the position in the scene (see Equation (2.1)).

is the total distance traveled by the contribution from $\mathbf{x}$. This contribution is attenuated by the reflectivity $\boldsymbol{\alpha}(\mathbf{x})$, square-law radial fall-off, and $\cos(\theta(\mathbf{x}))$ to account for foreshortening of the hidden-plane with respect to the illumination, where $\theta(\mathbf{x})$ is the angle between the surface normal at $\mathbf{x}$ and a vector from $\mathbf{x}$ to the illumination source. Thus, the backreflected waveform from the hidden-scene point $\mathbf{x}$ is

$$a_{ij}(\mathbf{x})\,\boldsymbol{\alpha}(\mathbf{x})\,s(t - z_{ij}^{SD}(\mathbf{x})/c),$$

where c is the speed of light and

$$a_{ij}(\mathbf{x}) = \frac{\cos(\theta(\mathbf{x}))}{\left(z_i^S(\mathbf{x})\,z_j^D(\mathbf{x})\right)^2}. \tag{2.1}$$

Examples of distance functions and geometric attenuation factors are shown in Figure 2-4.

Combining contributions over the plane, the total light incident at detector D_j in response to hidden-scene illumination by source S_i is,

$$r_{ij}(t) = \int_0^W \int_0^W a_{ij}(\mathbf{x})\, \boldsymbol{\alpha}(\mathbf{x})\, s(t - z_{ij}^{SD}(\mathbf{x})/c)\, \mathrm{d}x\, \mathrm{d}y.$$
$$\stackrel{(1)}{=} \left[\int_0^W \int_0^W a_{ij}(\mathbf{x})\, \boldsymbol{\alpha}(\mathbf{x})\, \delta(t - z_{ij}^{SD}(\mathbf{x})/c)\, \mathrm{d}x\, \mathrm{d}y \right] * s(t).$$

where $*$ denotes continuous-time convolution, $\delta(.)$ is the Dirac delta function, and equality $^{(1)}$ follows from the Dirac sifting property.

Putting aside for now the effect of the pulse waveform, $s(t)$, we define the scene impulse response function corresponding to the (source, detector) pair, (S_i, D_j), as

$$p_{ij}(t) \triangleq \int_0^W \int_0^W a_{ij}(\mathbf{x})\, \boldsymbol{\alpha}(\mathbf{x})\, \delta(t - z_{ij}^{SD}(\mathbf{x})/c)\, \mathrm{d}x\, \mathrm{d}y. \tag{2.2}$$

Thus, evaluating $p_{ij}(t)$ at a fixed time t amounts to integrating over $\mathbf{x} \in [0, W]^2$ with $t = z_{ij}^{SD}(\mathbf{x})/c$. Define the isochronal curve $C_{ij}^t = \{\mathbf{x} : z_{ij}^{SD}(\mathbf{x}) = ct\}$. Then

$$p_{ij}(t) = \int_{C_{ij}^t} a_{ij}(\mathbf{x})\boldsymbol{\alpha}(\mathbf{x})\, \mathrm{d}s = \int\int a_{ij}(\mathbf{x}(k,u))\boldsymbol{\alpha}(\mathbf{x}(k,u))\, \mathrm{d}u\, \mathrm{d}k \tag{2.3}$$

where $\mathbf{x}(k,u)$ is a parameterization of $C_{ij}^t \cap [0,W]^2$. The scene impulse response $p_{ij}(t)$ thus contains the contour integrals over C_{ij}^t's of the desired function $\boldsymbol{\alpha}$. Each C_{ij}^t is a level curve of $z_{ij}^{SD}(\mathbf{x})$; as illustrated in Figure 2-4, these are ellipses.

2.5 Measurement Model

A digital system can use only samples of $r_{ij}(t)$ rather than the continuous-time function itself. We now show how uniform sampling of $r_{ij}(t)$ with a linear time-invariant (LTI) prefilter relates to linear functional measurements of $\boldsymbol{\alpha}$. This establishes the foundation for a Hilbert space view of our computational imaging formulation.

At the detector the backreflected light signal, $r_{ij}(t)$, is convolved with the sensor impulse response filter, $h(t)$, prior to discrete time-sampling with a sampling interval T_s in order to produce time-samples denoted,

$$r_{ij}[n] = \big(r_{ij}(t) * h(t)\big)\big|_{t=nT_s}, \qquad n = 0,\, 1,\, \ldots,\, (N-1).$$

The sample data length, N, is chosen sufficiently high so that NT_s is slightly greater than the time-domain support of the signal $r_{ij}(t)$. Also, in this chapter we use the *box*-detector impulse response to match the operation of the streak camera that was employed in our experimental setup in Section 2-7, i.e.,

$$h(t) = \begin{cases} 1, & \text{for } 0 \leq t \leq T_s; \\ 0, & \text{otherwise,} \end{cases} \tag{2.4}$$

corresponding to integrate-and-dump sampling in which the continuous signal value is first integrated for duration T_s and sampled immediately thereafter. Also, note that by the associativity of the convolution operator,

$$r_{ij}[n] = \big(p_{ij}(t) * s(t) * h(t)\big)\big|_{t=nT_s}, \qquad n = 0,\, 1,\, \ldots,\, (N-1).$$

Defining the combined impulse response of the source-detector pair, $g(t) \triangleq s(t) * h(t)$, we obtain

$$r_{ij}[n] = \big(p_{ij}(t) * g(t)\big)\big|_{t=nT_s}, \qquad n = 0,\, 1,\, \ldots,\, (N-1).$$

A time-sample $r_{ij}[n]$ can be seen as a standard $\mathcal{L}^2(\mathbb{R})$ inner product between $p_{ij}(t)$ and a time-reversed and shifted system impulse response, $g(t)$ [37], i.e.,

$$r_{ij}[n] = \langle p_{ij}(t),\, g(nT_s - t) \rangle. \tag{2.5}$$

Using Equation (2.2), we can express Equation (2.5) in terms of $\boldsymbol{\alpha}$ using the standard

$\mathcal{L}^2([0,W]^2)$ inner product:

$$r_{ij}[n] = \langle \boldsymbol{\alpha},\, \varphi_{i,j,n} \rangle \qquad \text{where} \tag{2.6}$$

$$\varphi_{i,j,n}(\mathbf{x}) = a_{ij}(\mathbf{x})\, g(nT_s - z_{ij}^{SD}(\mathbf{x})/c). \tag{2.7}$$

Over a set of sensors and sample times, $\{\varphi_{i,j,n}\}$ will span a subspace of $\mathcal{L}^2([0,W]^2)$, and a sensible goal is to form a good approximation of $\boldsymbol{\alpha}$ in that subspace.

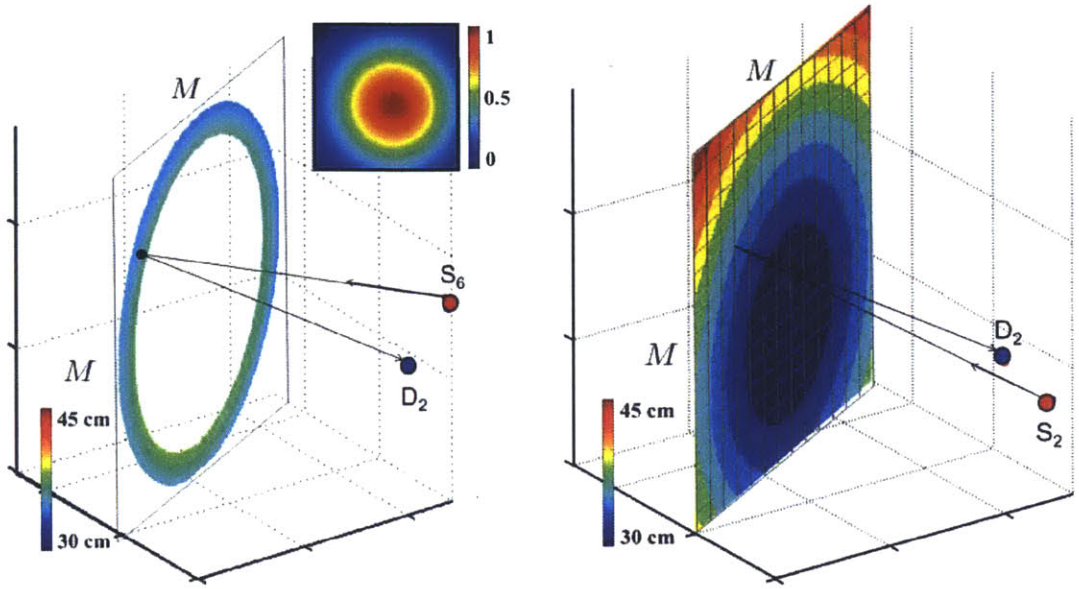
Now, since $h(t)$ is nonzero only for $t \in [0, T_s]$, by Equation (2.5), the time-sample $r_{ij}[n]$ is the integral of $r_{ij}(t)$ over $t \in [(n-1)T_s, nT_s]$. Thus, by Equation (2.3), $r_{ij}[n]$ is an a_{ij}-weighted integral of $\boldsymbol{\alpha}$ between the contours $C_{ij}^{(n-1)T_s}$ and $C_{ij}^{nT_s}$. To interpret this as an inner product with $\boldsymbol{\alpha}$, as in Equations (2.6) and (2.7), we note that $\varphi_{i,j,n}(\mathbf{x})$ is $a_{ij}(\mathbf{x})$ between $C_{ij}^{(n-1)T_s}$ and $C_{ij}^{nT_s}$ and zero otherwise. Figure 2-5(a) shows a single representative $\varphi_{i,j,n}$. For the case of the box-detector impulse response (see Equation (2.4)), the functions $\{\varphi_{i,j,n}\}_{n\in\mathbb{Z}}$ for a single sensor have disjoint supports; their partitioning of the domain $[0,W]^2$ is illustrated in Figure 2-5(b).

Data acquisition: For each time-synchronized source-detector pair, (S_i, D_j), N discrete time-samples were collected as follows: the light sources were turned on one at a time, and for each light source all the detectors were time-sampled simultaneously. The use of the resulting dataset, $\{r_{ij}[n]\}_{i=1,\,j=1,\,n=0}^{i=6,\,j=3,\,n=N-1}$, in recovering the hidden-scene reflectivity, $\boldsymbol{\alpha}$, is described next.

2.6 Image Recovery using Linear Backprojection

Piecewise-constant Model for Hidden Plane Reflectivity

To express an estimate $\hat{\boldsymbol{\alpha}}$ of the reflectivity $\boldsymbol{\alpha}$, it is convenient to fix an orthonormal basis for a subspace of $\mathcal{L}^2([0,W]^2)$ and estimate the expansion coefficients in that basis. For an



(a) source S_6 and sensor D_2 (b) source S_2 and sensor D_2

Figure 2 5: (a) A single measurement function $\varphi_{i,j,n}(\mathbf{x})$ when $h(t)$ is the box function defined in Equation (2.4). Inset is unchanged from Figure 2 4(a). (b) When $h(t)$ is the box function defined in Equation (2.4), the measurement functions for a single sensor $\{\varphi_{i,j,n}(\mathbf{x})\}_{n=1}^{N}$ partition the plane into elliptical annuli. Coloring is by discretized delay using the colormap of Figure 2 4. Also overlaid is the discretization of the plane of interest into an M by M pixel array.

M-by-M pixel representation, let

$$\psi_{m_x,m_y}(\mathbf{x}) = \begin{cases} M/W, \text{ for } (m_x-1)W/M \le x < m_xW/M, \\ \qquad\qquad (m_y-1)W/M \le y < m_yW/M; \qquad m_x, m_y = 1,\ldots,M \\ 0, \text{ otherwise} \end{cases} \tag{2.8}$$

so that the hidden-plane reflectivity estimate has the following representation

$$\hat{\boldsymbol{\alpha}} = \sum_{m_x=1}^{M}\sum_{m_y=1}^{M} \hat{\alpha}_{\psi}(m_x,m_y)\,\psi_{m_x,m_y}, \tag{2.9}$$

which lies in the span of the vector space generated by the orthonormal basis vector collection,

$$\left\{\psi_{m_x,m_y} \subset \mathcal{L}^2([0,W]^2)\right\}_{m_x=1,m_y=1}^{m_x=M,m_y=M}$$

and is constant on patches of size $(W/M)\times(W/M)$.

Linear System Formulation of Measurement Model

We will now form a system of linear equations to find the basis representation coefficients $\{\hat{\alpha}_\psi(m_x, m_y)\}_{m_x=1,m_y=1}^{m_x=M,m_y=M}$. For $\hat{\boldsymbol{\alpha}}$ to be consistent with the value measured by the detector D_j in response to hidden scene illumination by source S_j at time $t = nT_s$, we must have

$$r_{ij}[n] = \langle \hat{\boldsymbol{\alpha}},\, \varphi_{i,j,n} \rangle = \sum_{m_x=1}^{M} \sum_{m_y=1}^{M} \hat{\alpha}_\psi(m_x, m_y) \langle \psi_{m_x,m_y},\, \varphi_{i,j,n} \rangle. \tag{2.10}$$

Note that the inner products $\{\langle \psi_{m_x,m_y},\, \varphi_{i,j,n} \rangle\}$ exclusively depend on M, W, the positions of illumination sources and detectors, $\{S_i, D_j\}_{i=1,j=1}^{i=6,j=3}$, the hidden-plane geometry, the combined source-detector impulse response $g(t)$, and the sampling intervals T_s not on the unknown reflectivity of interest $\boldsymbol{\alpha}$. Hence, we have a system of linear equations to solve for the basis coefficients $\hat{\alpha}_\psi(m_x, m_y)$. (In the case of orthonormal basis defined in Equation (2.8), these coefficients are the pixel values multiplied by M/W.)

When we specialize to the box sensor impulse response defined in Equation (2.4) and the orthonormal basis defined in Equation (2.8), many inner products $\langle \psi_{m_x,m_y},\, \varphi_{i,j,n} \rangle$ are zero, so the linear system is sparse. The inner product $\langle \psi_{m_x,m_y},\, \varphi_{i,j,n} \rangle$ is nonzero when the reflection from the pixel (m_x, m_y) affects the light intensity at detector D_j in response to hidden-plane illumination by light source S_i within time interval $[(n-1)T_s, nT_s]$. Thus, for a nonzero inner product the pixel (m_x, m_y) must intersect the elliptical annulus between $C_{ij}^{(n-1)T_s}$ and $C_{ij}^{nT_s}$. With reference to Figure 2-5(a), this occurs for a small fraction of (m_x, m_y) pairs unless M is small or T_s is large. The value of a nonzero inner product depends on the fraction of the square pixel that overlaps with the elliptical annulus and the geometric attenuation factor.

To express Equation (2.10) as a vector-matrix multiplication, we replace double indices with single indices (i.e., vectorize, or reshape) to get

$$\mathbf{y} = \boldsymbol{\Theta}\, \hat{\boldsymbol{\alpha}}_\psi \tag{2.11}$$

where $\mathbf{y} \in \mathbb{R}^{18\times N}$ contains the data samples $\{r_{ij}[n]\}$, the first N from source-detector pair (S_1, D_1), the next N from source-detector pair (S_1, D_2), etc.; and $\hat{\boldsymbol{\alpha}}_\psi \in \mathbb{R}^{M^2}$ contains the coefficients $\hat{\alpha}_\psi(m_x, m_y)$, varying i first and then j. Then the inner product

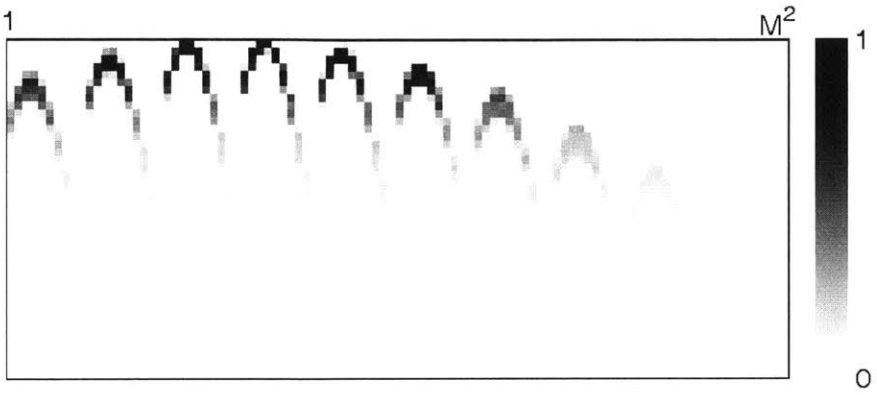


Figure 2 6: Visualizing linear system representation $\boldsymbol{\Theta}$ in (2.11). Contribution from source detector pair (S_1, D_1) is shown. To generate this system matrix we choose $M = 10$ and therefore the $\boldsymbol{\Theta}$ has 100 columns. Also we choose T_s such that $\boldsymbol{\Theta}$ has 40 rows. All zero rows arising from times prior to first arrival of light and after the last arrival of light are omitted.

$\langle \psi_{m_x,m_y}, \varphi_{i,j,n} \rangle$ appears in row $((i-1)\times 6 + (j-1)\times 3)N + n$, and column $(m_x - 1)M + m_y$ of $\boldsymbol{\Theta} \in \mathbb{R}^{18\times N\times M^2}$. Figure 2-6 illustrates an example of the portion of $\boldsymbol{\Theta}$ corresponding to source-detector pair (S_1, D_1) for the scene in Figure 2-3.

Assuming that $\boldsymbol{\Theta}$ has a left inverse (i.e., $\text{rank}(\boldsymbol{\Theta}) = M^2$), one can form an image by solving Equation (2.11). The portion of $\boldsymbol{\Theta}$ from one source-detector pair cannot have full column rank, because of the collapse of information along elliptical annuli depicted in Figure 2-5(a). Full rank and good matrix conditioning [38] are achieved with an adequate number of source-detector pairs, noting that such pairs must differ significantly in their spatial locations to increase rank and improve conditioning. As a rule of thumb, greater spatial disparity between (source,detector) pair positions improves the conditioning of the inverse problem. In this chapter, we use 18 well-separated (source, detector) pairs to satisfy the full-rank condition and to achieve good matrix conditioning.

Hidden-Plane Reflectivity Estimation

The *unconstrained* least-squares estimate for the hidden-plane reflectivity under the piecewise planar reflectivity model is [38]:

$$\hat{\boldsymbol{\alpha}}_{\psi}^{ULS} = \left(\boldsymbol{\Theta}^T\boldsymbol{\Theta}\right)^{-1}\boldsymbol{\Theta}^T\mathbf{y}.$$

The $M \times M$ pixel image is obtained after reshaping the solution vector back to an array. Since reflectivity is constrained to lie within the interval $[0, 1]$, the *constrained* least-squares estimate that satisfies this bounded reflectivity condition is obtained by the pixelwise thresholding of the unconstrained reflectivity estimate [38], i.e.,

$$\hat{\alpha}_{\psi}^{CLS}(m_x, m_y) = \begin{cases} 1 & , \text{ for } \hat{\alpha}_{\psi}^{ULS}(m_x, m_y) \geq 1; \\ 0 & , \text{ for } \hat{\alpha}_{\psi}^{ULS}(m_x, m_y) \leq 0; \quad m_x, m_y = 1, \ldots, M \\ \hat{\alpha}_{\psi}^{ULS}(m_x, m_y), & \text{otherwise.} \end{cases}$$

Mitigating the effect of measurement noise: We assume that the time samples, $r_{ij}[n]$, are corrupted with signal-independent, zero-mean, white Gaussian noise of a known variance, σ^2, which depends on the choice of the detector and time-sampling assembly.

We employ two main methods to reduce the effect of noise: First we reduced noise in the time-samples by repeatedly illuminating the static scene and averaging the measured time-samples. Second, we employed constrained Tikhonov regularization [38] to obtain the following hidden-plane reflectivity estimate,

$$\hat{\alpha}_{\psi}^{C-TIKH}(m_x, m_y) = \begin{cases} 1 & , \text{ for } \hat{\alpha}_{\psi}^{U-TIKH}(m_x, m_y) \geq 1; \\ 0 & , \text{ for } \hat{\alpha}_{\psi}^{U-TIKH}(m_x, m_y) \leq 0; \quad m_x, m_y = 1, \ldots, M \\ \hat{\alpha}_{\psi}^{U-TIKH}(m_x, m_y), & \text{otherwise,} \end{cases} \tag{2.12}$$

where

$$\hat{\boldsymbol{\alpha}}_{\psi}^{U-TIKH} = \left(\boldsymbol{\Theta}^T\boldsymbol{\Theta} + \beta\,\Phi^T\Phi\right)^{-1}\boldsymbol{\Theta}^T\mathbf{y},$$

Φ is the discrete wavelet transform (DWT) matrix derived from Daubechies's 2-tap filter [39] ($[1, -1]/\sqrt{2}$), and $\beta \in [0, 1]$ is a weight parameter to control the degree of Tikhonov regularization. A high value of the weight parameter, β, forces the reflectivity image to be overly smooth while a low-value of β leads to noisy reflectivity estimates. Therefore, an optimal value of β needs to be chosen. In this chapter, we selected this optimal β-value by constructing the reflectivity image for $\beta = \{0.1, 0.2, \ldots, 0.9\}$ and then choosing the one which minimized the squared error $\| \mathbf{y} - \boldsymbol{\Theta}\,\hat{\boldsymbol{\alpha}}_{\psi}^{U-TIKH} \|_2^2$.

In the next section we describe the experimental setup and results obtained through the use of the aforementioned scene response modeling and hidden-scene image construction methods.

2.7 Experimental Setup and Results

The physical setup corresponding to Fig. 2-1 is shown in Fig. 2-7 (also see [5, 40, 41]). The complete scene geometry, including all positions and orientations, was measured prior to data collection.

The pulsed light source illuminating the visible Lambertian diffuser is a 795-nm-wavelength femtosecond Ti:Sapphire laser operating at 75 MHz repetition rate. Its ultra-short laser pulses were Gaussian-shaped ($s(t)$) with 50-ps full-width half-max (FWHM). For data acquisition, we were only interested in time-sampling the backreflected light in response to illumination by a single laser pulse, but as stated before we leveraged the high repetition rate of the laser by repeating the same scene response measurement over a dwell time of 60 seconds and averaging the time-samples to mitigate noise.

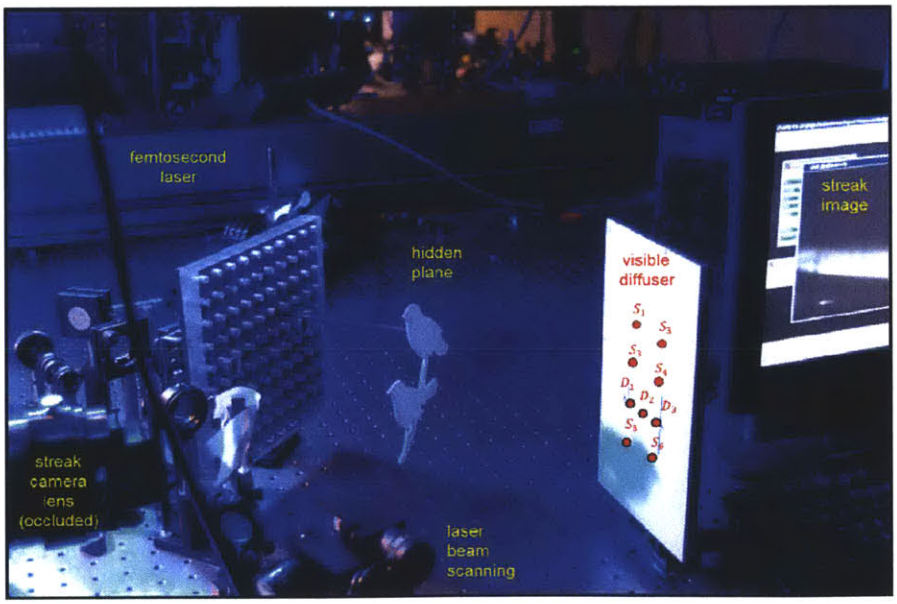


Figure 2 7: **Physical setup for looking around corners.** The various components for hidden plane imaging are shown.

Using a glass slide, about 5% of the laser beam light was split off to provide a time-synchronization signal. The time-resolved sensor was a Hamamatsu streak camera imaging along a horizontal line on the visible Lambertian diffuser. The streak camera comprised of a charge coupled device (CCD) sensor with a gating interval of $T_s = 2$ picoseconds, and a cathode ray tube with a 20 ps FWHM Gaussian-shaped impulse response. Time-samples of the impulse response of the overall imaging system, $\{g[n]\}_{n=0}^{N-1}$, which is the convolution of the instrument response and the laser pulse shape was measured by illuminating a heavily attenuated laser beam directly into to the streak camera opening. This system impulse response measurement also included the effect of timing jitter in the synchronization electronics.

In our experiment, we sampled the incident light signal for a 1 ns time window (NT_s). The total path lengths in our imaging setup, $S_i \rightarrow$ hidden-plane $\rightarrow D_j$, ranged from 35.1 cm to 41.2 cm corresponding to a maximum path difference of 6.1 cm or equivalently 204 ps time-duration of the light signal of interest. Given these optical path lengths we obtained between 100 to 200 time-samples for each source-detector pair.

The diffuser and hidden-plane's shapes were cut-out acrylic, painted with Lambertian white reflectivity coating (see Fig. 2-9(top row)). Given the different shapes and sizes of the scene cut-outs, we modeled the hidden-plane based on the smallest bounding square that enclosed all the scenes. The hidden reflectivity was modeled as piecewise-constant over square patches of dimensions 2.8 mm-by-2.8 mm ($W/M = 2.8$ mm) arranged in a planar 50-by-50 grid ($M = 50$).

The experiment was laid out such that scattering from mounts and other equipment did not contribute to the collected data. Also, by appropriate placement of occluders, it was ensured that no reflected light from the hidden-scene reached the camera optics without first reflecting off the visible diffuser. The same was ensured for the laser light, which could reach the hidden-scene only after reflecting off the visible diffuser. Thus, there was no direct line-of-sight between the hidden-scene and the co-located laser-camera setup. As a result, each laser pulse underwent three diffuse scatterings before reaching the streak camera. To make up for the resulting ~ 90 dB attenuation, we used large collecting optics and low-noise light intensification within the streak camera.

Data from a single source-detector pair did not provide enough observations to uniquely

solve for the hidden-plane's reflectivity. To resolve this problem, we illuminated six source locations, S_1 through S_6, and simultaneously time-sampled the backscattered light incident from three detector locations, D_1 through D_3. Each (S_i, D_j) pair provided 100 to 200 non-zero time samples, and we recorded a total of 2,849 streak camera samples using these 18 source-detector pairs (see Fig. 2-8).

As derived in Section 2.6, the measurement matrix, $\boldsymbol{\Theta}$ is a linear mapping from the 2,500 unknown reflectivity values to the 2,849 measured time-samples. The measurement matrix was constructed entirely in software, using the knowledge of scene geometry and the various optoelectronic parameters associated with the experiment (see Fig. 2-8). Finally as described in Section 2.6, we solved the resulting system of linear equations using Tikhonov regularization [38] to construct the 50-by-50 pixel images of the various hidden scenes. (see Fig. 2-9).

2.8 Discussion and Conclusion

In this chapter, we demonstrated hidden-plane image formation in the presence of occluding and scattering objects using a Lambertian diffuser in lieu of a lateral view mirror. Lambertian scattering makes it impossible to recover the hidden-plane image using an ordinary camera. Scattered light traverses different optical paths within the scene and differences in propagation path length introduce differences in the times-of-arrival at the sensor. We developed a mathematical model for this time-dependent scattering of a light impulse from the diffuse surfaces in the scene. Using femtosecond pulsed illumination of the Lambertian diffuser and ultra-fast time-resolved sensing of the diffuse reflections, we computationally constructed the hidden-plane image by solving a linear inversion problem. Understanding the trade-offs and limitations of the theoretical framework and proof-of-concept experiments presented in this chapter could lead to possibilities for further investigation.

The spatial resolution of the constructed image is a function of the signal-to-noise ratio (SNR) of the camera, its temporal resolution, the scene geometry, and the spatial sampling pattern on the Lambertian diffuser ($\{S_i, D_j\}$ locations). Even for the simplest case of imaging a single hidden-plane, there are several open questions in relation to these parameters,

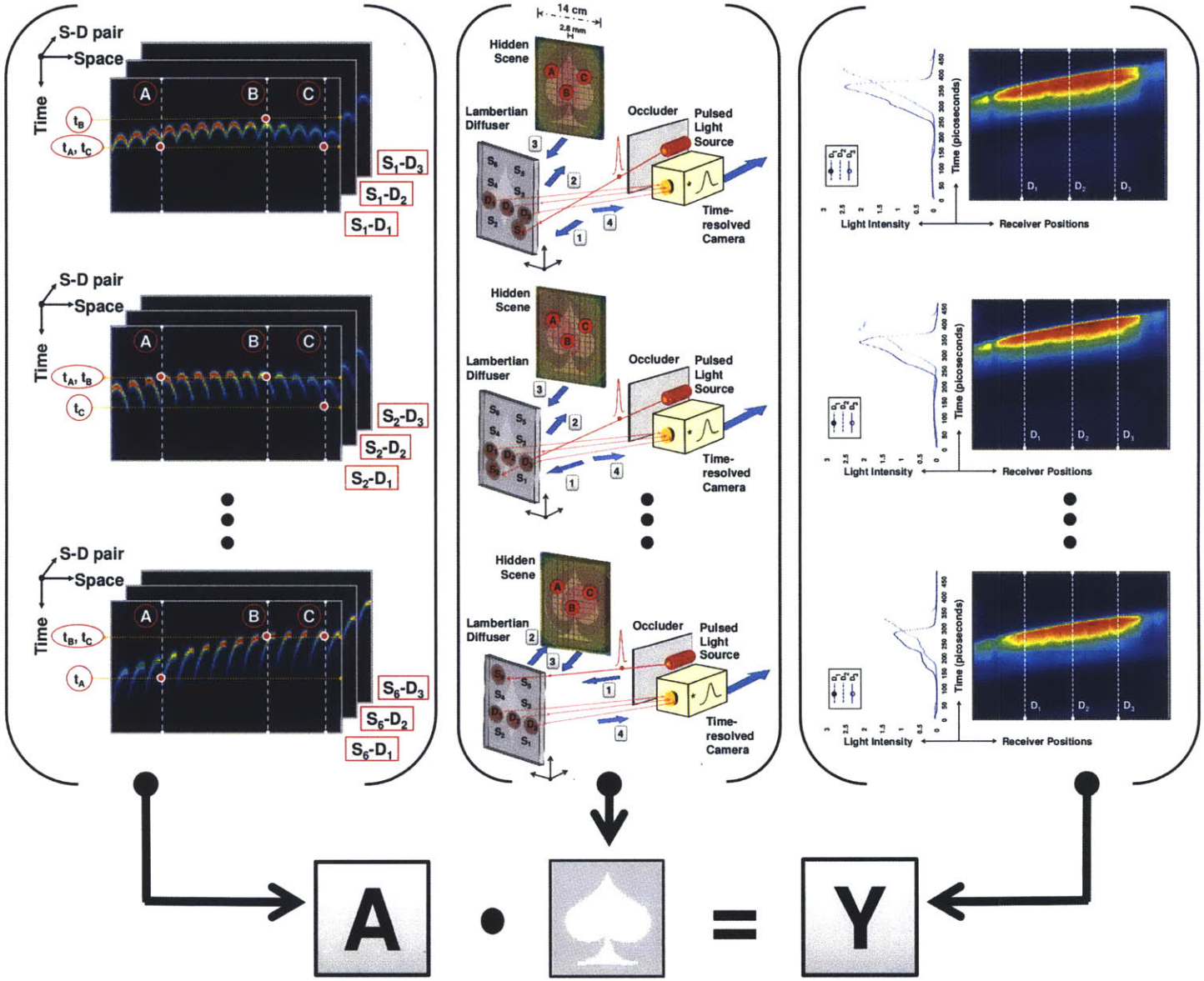


Figure 2-8: **Computational recovery of hidden-plane reflectivity from time-resolved measurements.** Consider pair (S_1, D_1): point B has shortest path length and A, C are equidistant so that their return light intensities get added linearly. Now consider pair (S_2, D_1): A, B are equidistant and light from C arrives later. Using a third pair, (S_6, D_1), we see that reflectivity values of B and C get added while A is separable. Solving these three independent linear equations in 3 unknowns allows estimation of the reflectivity values at A, B and C.

Figure 2 9: **Hidden-plane image construction results using the experimental setup.** (A) The hidden planar scenes used in our experiments were black and white cut out shapes. (B) Using an ordinary lateral view mirror instead of the Lambertian diffuser, the image of the hidden cut out is easily captured using a conventional camera. (C) Computational constructions of the hidden plane cut outs obtained by processing the streak images recorded in response to the ultra short pulse illumination of the Lambertian diffuser at several distinct locations (see Fig. 2 8). Images have a pixelated appearance because of the piecewise constant reflectivity assumption. Using a finer grid would allow sharper edges and higher pixel resolution images, but would also requires spatial sampling of a larger number of disparate source sensor locations and faster temporal sampling.

for example, how many source-sensor locations are necessary to ensure high quality image formation? What is optimal source-sensor geometry for hidden-scene imaging? What is the interplay between time-sampling and source-sensor geometry?

The answers to the aforementioned questions lie in understanding how each of the key system parameters affects the numerical conditioning of the linear transformation that relates hidden-plane reflectivity data to the measured time-samples. This task may be too difficult to characterize analytically, but a simulation-based approach may yield some key insights. For

example, it intuitively seems that acquiring data using a large number of spatially-separated source-sensor pairs, and the use of fine-temporal sampling generally result in higher quality hidden-plane image formation.

The framework and results presented in this chapter have several limitations. The constructed black-and-white images are very low-resolution, although the object outlines are clearly visible and one may recognize the object from its appearance in the constructed image. It is also unclear how robust the imaging framework is, for example, do small errors in the measured scene geometry have a significant effect on the image construction. Another parameter that will have a significant impact is the accuracy of time-synchronization between the laser and the streak camera. Although it was not attempted, gray-scale image capture may not be possible due to the significant SNR loss and poor numerical conditioning of the inverse imaging problem.

Overall, the computational imaging framework and the preliminary experimental results presented in this chapter by themselves do not point towards the design of a practical imaging system to look around corners. This chapter, however, serves as a powerful demonstration of the use of computational time-resolved imaging to tackle inverse imaging problems that cannot be solved using traditional imaging methods.

Chapter 3

Compressive Depth Acquisition

3.1 Overview

Sensing 3D scene structure is an integral part of applications ranging from 3D microscopy [42, 43] to geographical surveying [44]. Humans perceive depth using both monocular cues, such as motion parallax, and binocular cues, such as stereo disparity. Camera-based stereo vision techniques [45], however, suffer from poor depth resolution and high sensitivity to noise [46,47]. Computer vision techniques including structured-light scanning, depth-from-focus, depth-from-shape, and depth-from-motion [45, 48] are computation intensive, and the depth measurement using these methods is highly prone to errors from miscalibration, absence of sufficient scene texture, and low signal-to-noise ratio (SNR) [46 48].

In comparison, active depth acquisition systems, such as medium-range light detection and ranging (LIDAR) systems [9] and time-of-flight cameras [3, 10], are more robust against noise due to background light [47], capture depth data at video frame rates, only need a single viewpoint, and have little dependence on scene reflectivity or object texture. Typically, LIDAR systems consist of a pulsed illumination source such as a laser, a mechanical 2D laser scanning unit, and a single time-resolved photodetector or avalanche photodiode [9,49]. Some non-commercial LIDAR systems [50] employ floodlight laser illumination and avalanche photodiode arrays instead of raster scanned illumination and a single detector. The time-of-flight camera illumination unit is composed of an array of omnidirectional, modulated,

infrared light emitting diodes [3, 10, 51]. Light reflected from the scene with time delay proportional to distance is focused on a 2D array of time-of-flight depth sensing pixels. The detailed theory of operation of time-of-flight cameras and LIDAR systems is covered in Sections 3.2 and Chapter 5 respectively.

While high quality 2D imaging is now a mature commercial off-the-shelf technology, 3D acquisition techniques have room for significant improvements in spatial resolution, depth accuracy, and cost effectiveness. State-of-the-art laser ranging systems such as ALIRT [2] operate at kilometer range, but are limited to decimeter accuracy, expensive to build and difficult to mass produce. Moreover, such airborne raster-scanning systems typically require several passes over a region to collect data and are not effective in capturing fast moving objects. Commercial room-scale time-of-flight imagers capture real-time, centimeter-accurate depth data but have significantly poorer spatial resolution (less than one-megapixel) compared with the several megapixel RGB camera sensors in today's smart devices.

As the use of 3D imaging in consumer and commercial applications continues to increase, LIDAR systems and time-of-flight cameras with low spatial resolution are unable to sharply resolve spatial features, such as small depth discontinuities, that are needed for some applications. Low-resolution sensor arrays (for e.g., 128×128 pixel avalanche detector arrays) can be used to perform coarse tasks such as object segmentation. Upcoming applications, such as 3D biometrics and skeletal tracking for motion control, demand much higher spatial resolution. Due to limitations in the 2D time-of-flight sensor array fabrication process and readout rates, the number of pixels in the state-of-the-art time-of-flight camera (Microsoft's Kinect sensor [52]) is currently limited to a maximum of 424×512 pixels. Consequently, it is desirable to develop novel depth sensors that possess high spatial resolution without increasing the device cost and complexity.

In this chapter we introduce a framework for compressively acquiring the depth map of a piecewise-planar scene, with high depth and spatial resolution, using only a single omnidirectionally collecting photodetector as the sensing element and a spatially-modulated pulsed light source that produces patterned illumination (see Fig. 3-3).

In Chapter 4 we present the *dual* of this imaging configuration in a low-light level setting by employing floodlight illumination with a pulsed source and spatial patterning the light

that gets detected by using a digital micromirror device (DMD) in front of a single-photon counting detector (see Fig. 4-1). Despite the differences in imaging configuration and the employed sensing element, both computational imagers operate on the signal models and depth map construction methods that are developed in this chapter.

The remainder of this chapter is organized as follows. Prior art and challenges in compressive depth acquisition are discussed in Section 3.2. Then Section 3.3 introduces the imaging setup and signal model for compressive depth acquisition. The measurement model and data acquisition pipeline are described in Section 3.4. These models form the basis for the novel depth map construction developed in Section 3.5. Experimental setup and results are presented in Section 3.6, and finally Section 3.7 provides additional discussion of limitations and extensions.

3.2 Prior Art and Challenges

Time-of-flight camera operation: There are two classes of time-of-flight cameras, ones that operate using short-pulsed illumination [53] and others that use a continuously modulated light source [3, 10, 54 57]. In this section, we only consider the operation of amplitude modulated continuous wave (AMCW) homodyne time-of-flight cameras, which are now mass produced and commercially available as Microsoft Kinect 2.0 [52]. Figure 3-1 shows a signal-processing abstraction of an AMCW time-of-flight camera pixel; in the Microsoft Kinect, the camera sensor is a collection of 424×512 such AMCW time-of-flight measuring pixels. The radiant power of transmitted light is temporally modulated using a non-negative sinusoidal signal

$$s(t) = 1 + \cos(2\pi f_0 t), \tag{3.1}$$

with modulation frequency f_0 and modulation period $T_0 = 1/f_0$. In the absence of multipath interference, the reflected light from the scene is well modeled as

$$r(t) = \alpha \cos(2\pi f_0 (t - \tau)) + (\alpha + b_\lambda). \tag{3.2}$$

Here: α is the target reflectivity, $\tau = 2z/c$ is the time-delay due to roundtrip propagation of light at speed c between camera and a scene point at distance z; and b_λ is the time-invariant

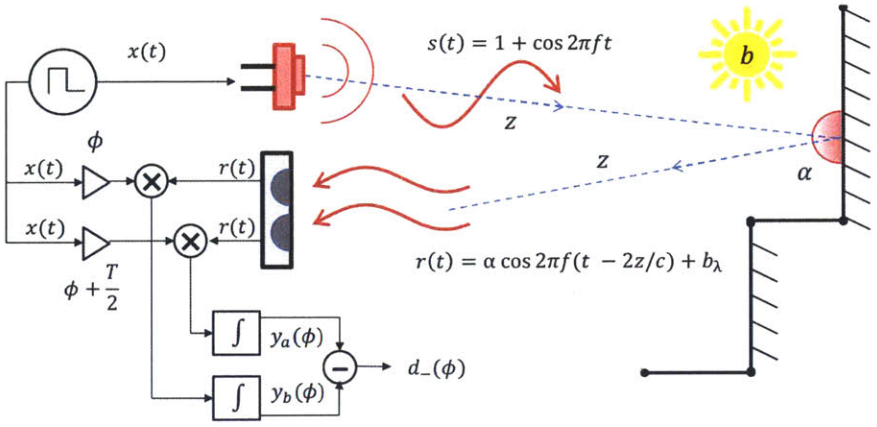


Figure 3-1: **Signal processing abstraction of a time-of-flight camera pixel.**

background or ambient light contribution from the scene at the operating wavelength λ. Our goal is to estimate the reflectivity α and the distance z under the assumption that there is no depth aliasing, which can be ensured by choosing T_0 such that the maximum scene depth $z_{max} < cT_0/2$.

In a homodyne time-of-flight camera, the light signal, $r(t)$, incident at the sensor pixel is correlated with a reference signal, $x(t+\phi) = x(t+\phi+T_0)$ and its time-shifted copy, $x(t+\phi+T_0/2)$. The role of the shift parameter, ϕ, will become clear below. In practice, the reference signal, $x(t)$, is typically chosen to be a periodic, unit-amplitude, non-negative square wave with the following Fourier series representation:

$$x(t) = x(t+T_0) = 1 + \frac{4}{\pi}\left[\sin(2\pi f_0 t) + \frac{1}{3}\sin(6\pi f_0 t) + \frac{1}{3}\sin(6\pi f_0 t) + \ldots\right]. \tag{3.3}$$

Under the aforementioned assumption, the cross-correlation functions, denoted using y_a and y_b are given by

$$\begin{aligned} y_a(\phi+\tau) &= \int_0^{T_0} r(t)\, x(t+\phi)\, \mathrm{d}t \quad = \frac{2\alpha T_0}{\pi}\sin\left(2\pi f_0(\phi+\tau)\right) + (\alpha + b_\lambda)T_0 \\ y_b(\phi+\tau) &= \int_0^{T_0} r(t)\, x(t+\phi+T_0/2)\, \mathrm{d}t = -\frac{2\alpha T_0}{\pi}\sin\left(2\pi f_0(\phi+\tau)\right) + (\alpha + b_\lambda)T_0. \end{aligned} \tag{3.4}$$

For a fixed value of ϕ, the following values are measured at the output of a time-of-flight

sensor pixel,

$$\begin{aligned} d_+(\phi;\tau) &= [y_a + y_b]/2 = (\alpha + b_\lambda)T_0 \\ d_-(\phi;\tau) &= [y_a - y_b]/2 = \frac{2\alpha T_0}{\pi} \sin\left(2\pi f_0(\phi+\tau)\right). \end{aligned} \tag{3.5}$$

Note $d_+(\phi;\tau) = d_+$ is a constant and provides an estimate of the total incident radiant power. The function $d_-(\phi;\tau)$ is a sinusoid with no background component. To estimate α and τ, we sample $d_-(\phi;\tau)$ at $N \geq 2$ uniformly spaced values of ϕ. The n-th sample is

$$d_-[n;\tau] = \frac{2\alpha T_0}{\pi} \sin\left(2\pi f_0\left(\frac{nT_0}{N} + \tau\right)\right) \quad \text{for } n = 0, \ldots, N-1.$$

For modulation frequency, f_0, the amplitude estimate is

$$\hat{\alpha}(f_0) = \frac{\pi}{2T_0}\left\{\frac{1}{N}\sqrt{d_-[0;\tau]^2 + \ldots + d_-[N-1;\tau]^2}\right\}. \tag{3.6}$$

and an estimate of the wrapped or aliased distance is,

$$\hat{d}(f_0) = \frac{1}{2\pi}\left(\frac{cT_0}{2}\right) \arg\left\{\sum_{n=0}^{N-1} d_-[n;\tau] e^{-j\,2\pi n/N}\right\}. \tag{3.7}$$

Time-of-flight 3D cameras capture pixelwise depth information by focusing the back-reflected time-shifted and attenuated light signal onto an array of sensor elements using focusing optics identical to those found in traditional 2D cameras that capture scene reflectivity as photographs. Neither the traditional image sensors nor the time-of-flight cameras attempt to exploit the spatial correlations present in the reflectivity and depth of real-world objects to reduce the various costs and complexities associated with acquiring these natural scene characteristics. The next three sections discuss state-of-the art methods that were recently introduced to compressively acquire scene reflectivity and depth information by exploiting such spatial correlations.

Compressive acquisition of scene reflectivity: Many natural signals, including scene reflectivity, can be represented or approximated well using a small number of non-zero pa-

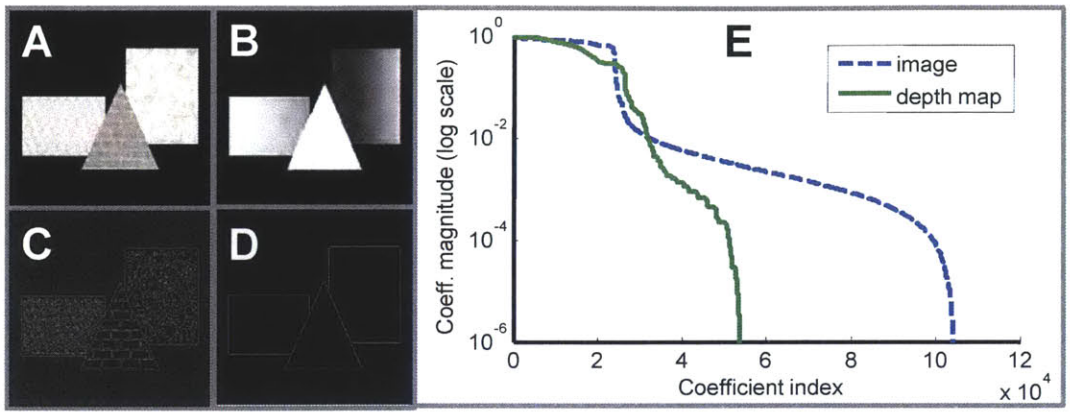


Figure 3 2: *Sparsity* of a signal (having a basis expansion or similar representation with a small number of coefficients significantly different from zero) is widely exploited for signal estimation and compression [58]. An $M \times M$ pixel reflectivity image (shown in **A**) or depth map (shown in **B**) of a scene requires M^2 pixel values for representation in the spatial domain. As illustrated with the output of an edge detection method, the Laplacian of a depth map (shown in **D**) typically has fewer significant coefficients than the Laplacian of the reflectivity image of the same scene (shown in **C**). This sparsity structure of natural scenes is also reflected in discrete wavelet transform (DWT) coefficients sorted by magnitude: the reflectivity image has a much slower decay of DWT coefficients and has more nonzero coefficients (shown in **E** blue, dashed line) as compared to the corresponding depth map (shown in **E** green, solid line). We exploit this sparsity of depth maps in our compressive depth acquisition framework.

rameters. This property, known as sparsity, has been widely exploited for signal estimation and compression [58]. Making changes in signal acquisition architectures often including some form of randomization inspired by the ability to effectively exploit sparsity in estimation has been termed compressed sensing [59, 60]. Compressed sensing provides techniques to estimate a signal vector $\boldsymbol{\alpha}$ from linear measurements of the form $\mathbf{y} = \mathbf{A}\,\boldsymbol{\alpha} + \mathbf{n}$, where $\mathbf{n}$ is additive noise and measurement vector $\mathbf{y}$ has significantly *fewer* entries than $\boldsymbol{\alpha}$. The compressed sensing estimation methods [61] exploit there being a linear transformation Φ such that $\Phi(\boldsymbol{\alpha})$ is approximately sparse and that the measurement matrix $\mathbf{A}$ satisfies the restricted isometry property [62].

An early instantiation of compressed sensing in an imaging context was the *single-pixel camera* [11, 12]. This single-pixel reflectivity imaging architecture employed a DMD to optically measure the linear projections of the scene reflectivity image onto pseudorandom binary patterns. It demonstrated, through the use of compressed sensing reconstruction algorithms, that it was possible to construct the scene reflectivity image using far fewer measurements than the number of pixels in that image.

Challenges in exploiting sparsity in depth map acquisition: The majority of depth sensing techniques make time-of-flight measurements either through raster scanning every point of interest in the field-of-view or by using focusing optics to establish a one-to-one correspondence between each spatial location (or patch) in the scene and an element in an array of sensors. The signal of interest depth in these cases is naturally sparse in a wavelet domain or has sparse gradient or Laplacian. Furthermore, the depth map of a scene is generally more compressible or sparse than its reflectivity image (see Fig. 3-2). In the context of depth map compression, the work in [63,64] has exploited this transform domain sparsity using pseudorandom projections in order to efficiently compress natural scene depth maps. Thus, we expect a smaller number of measurements to suffice; as expounded in Section 3.6, our number of measurements is 5% of the number of pixels as compared to 40% for reflectivity imaging [11, 12]. However, compressive depth map acquisition is far more challenging than depth map compression.

Compressively acquiring depth information using only a single detector poses a major challenge. The quantity of interest depth is embedded in the reflected signal as a time shift. The measured signal at the photodetector is a sum of all reflected returns and while all the incident signal amplitudes containing reflectivity information add linearly, the corresponding time-shifts in these signals containing time-of-flight (and depth information) *do not* add linearly. They add nonlinearly, and this nonlinearity worsens with the number of time-shifted waveforms that are combined at the photodetector. Nonlinear mixing makes compressive depth acquisition more challenging compared to compressive reflectivity acquisition discussed before. In particular, because it is not possible to obtain linear measurements of the scene depth by simply combining the backreflected light at the detector the compressive image formation pipeline employed in the single-pixel camera is useless for depth map acquistion.

Compressive LIDAR: In a preliminary application of the compressed sensing framework to LIDAR depth acquisition [15], a simple room-scale scene was flood illuminated with a short-pulsed laser. Similar to the single-pixel camera, the reflected light was focused onto a DMD that implemented a linear projection of the incident spatial signal with a

pseudorandom binary-valued pattern. All of the light from the DMD was focused on a photon-counting detector and gated to collect photons arriving from an *a priori* chosen depth interval. Then, conventional compressed sensing reconstruction was applied to recover an image of the objects within the selected depth interval. The use of range gating in this setup makes it a conventional compressed sensing imager in that the quantities of interest (reflectivity as a function of spatial position, within a depth range) are combined linearly in the measurements.

Hence, while this approach unmixes spatial correspondences, it does not directly solve the aforementioned challenge of resolving nonlinearly-embedded depth information. The need for accurate range intervals of interest prior to reconstruction is one of the major disadvantages of this system. It also follows that there is no method to distinguish between objects at different depths within a chosen range interval. Moreover, acquiring a complete scene depth map requires a full range sweep. The proof-of-concept system [15] had 60 cm depth resolution and 64×64 pixel resolution.

Next, we develop the theoretical framework that allows us to address the challenging problem of nonlinear mixing of scene depth values in order to compressively reconstruct the scene depth map.

3.3 Imaging Setup and Signal Modeling

In this chapter we restrict ourselves to the case of layered scenes comprised of planar objects, all with the same reflectivity placed at K distinct depths, $z^1 < \ldots < z^K$ (see Fig. 3-3). The case of scenes comprising inclined planar objects with uniform reflectivity was discussed in [6] (also see [65, 65 70]). The depth map construction algorithm employed therein, however, assumes that an inclined plane is composed of several smaller fronto-parallel facets placed at uniformly spaced depths. Thus, the computational imager described in [6] is a special case of the techniques developed in this chapter.

In our imaging setup, light from a pulsed illumination source (pulse shape $s(t)$) is modulated with a spatial light modulator (SLM) before it illuminates the layered scene of interest. The SLM pattern has $M \times M$ pixels and each pixel's opacity is denoted with $p(x, y)$. We

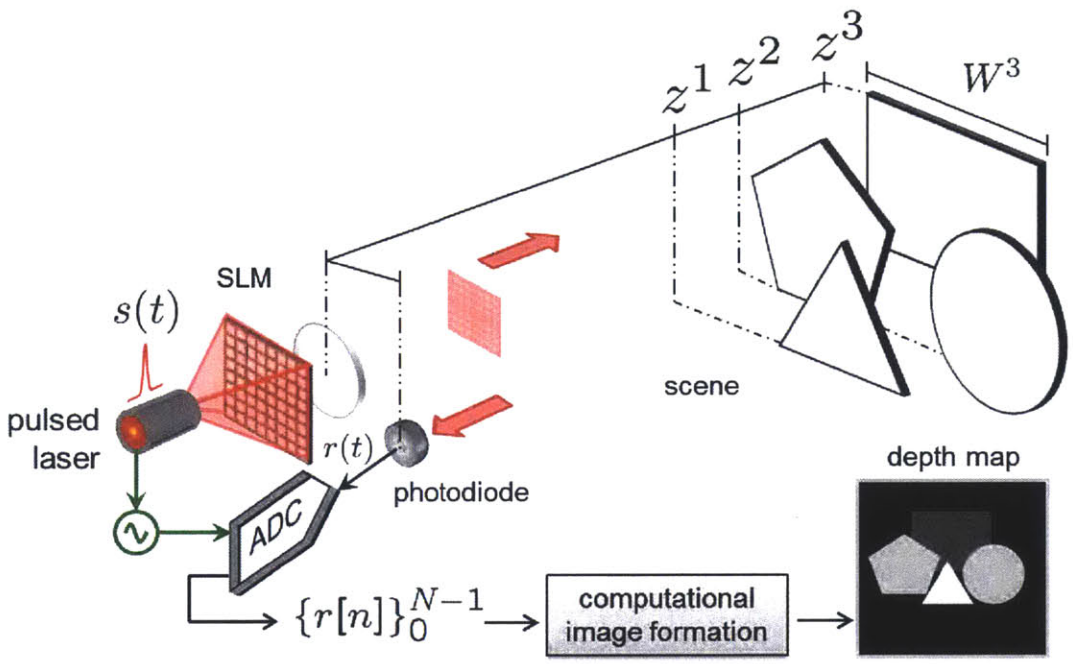


Figure 3 3: **Compressive depth imaging of layered scenes** The objects are placed at K distinct depths, the SLM pixel resolution is $M \times M$, and the scene is illuminated with $L \ll M^2$ patterns. For each illumination pattern, N digital time samples of the photocurrent are recorded.

consider binary-valued SLM patterns, i.e., each SLM pixel is either completely transparent ($p(x, y) = 1$) or opaque ($p(x, y) = 0$). Denote the scene reflectivity with α, which includes the effects of object surface albedo, light propagation losses due to Lambertian scattering, radial fall-off, and optoelectronic conversion using the experimental setup described in Section 3.6.

In our imaging formulation we assume that the number of distinct depths, K, and reflectivity, α, are known *a priori*. The omnidirectional detector collecting the backscattered light has an impulse response $h(t)$. We assume that there is no background light. The detector and the light source are precisely synchronized so that time-of-flight measurements are possible. The baseline between the detector and pulsed light source is assumed to be negligible. We also assume that the scene is far enough from imaging system that the backreflected light from different objects placed at depth z^k arrives at the detector within a small time interval centered at time $t^k = 2z^k/c$, where c is the speed of light.

The photocurrent, $r(t)$, generated in response to scene illumination is sampled using an analog-to-digital converter (ADC), which operates well above the Nyquist sampling rate

determined by the system transfer function, $g(t) = s(t) * h(t)$. The root mean square duration of the system transfer function is denoted by T_p and it includes the effect of the light pulse duration as well as the detector response time. Denote the ADC sampling interval by T_s. A total of N digital time-samples, $\{r[n]\}_{n=0}^{N-1}$, are recorded for every transmitted pulse, where $r[n] = r(nT_s)$, and N is chosen such that $N >> 2K + 1$ and $NT_s >> t^K + T_p$.

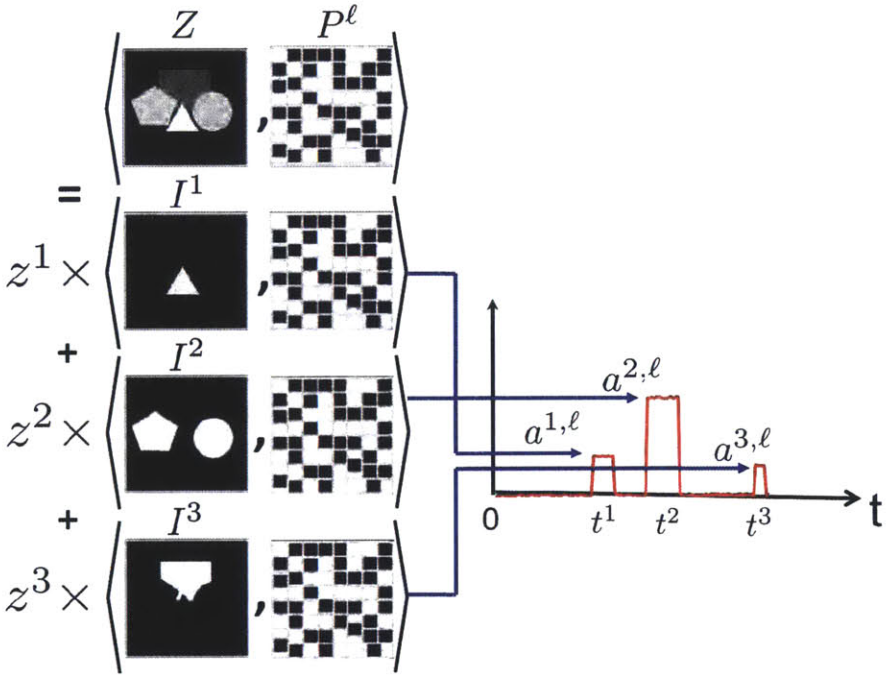


Figure 3 4: **Depth-map representation for layered scene shown in Fig. 3-3.** Also shown is a pictorial representation of Equation (3.10).

Depth Map Representation

The constructed depth map has the same pixel resolution as the SLM. For the layered scenes in consideration the $M \times M$ pixel depth map, $\boldsymbol{z} = \{z(x,y)\}_{x,y=1}^{M}$, can be expressed as a sum of K *index maps*, with $\boldsymbol{I}^k = \{i(x,y)\}_{x,y=1}^{M}$ of pixel size $M \times M$, i.e., $\boldsymbol{z} = \sum_{k=1}^{K} z^k \boldsymbol{I}^k$ (see Fig. 3-4). In our setup we only consider opaque objects hence our index maps are binary-valued. The index maps are defined as:

$$i^k(x,y) = \begin{cases} 1 & \text{if an object is present and visible at} \\ & \text{pixel } (x,y) \text{ and depth } z^k \\ 0 & \text{otherwise.} \end{cases}$$

Given this depth map representation, we note that estimating $\boldsymbol{z}$ from digital time-samples of the photocurrent is tantamount to estimation of the distinct depth values $\{z^1, \ldots, z^K\}$ and the associated index maps, $\{\boldsymbol{I}^1, \ldots, \boldsymbol{I}^K\}$.

We also introduce an additional index map, $\boldsymbol{I}^0$, to model the pixels at which there is no object present in the scene, i.e.,

$$i^0(x,y) = \begin{cases} 0 & \text{if an object is present at pixel } (x,y) \\ & \text{at any of the } K \text{ depths} \\ 1 & \text{otherwise.} \end{cases}$$

Note that (see Fig. 3-5),

$$i^0(x,y) + \sum_{k=1}^{K} i^k(x,y) = 1, \quad x,y = 1, \ldots, M. \tag{3.8}$$

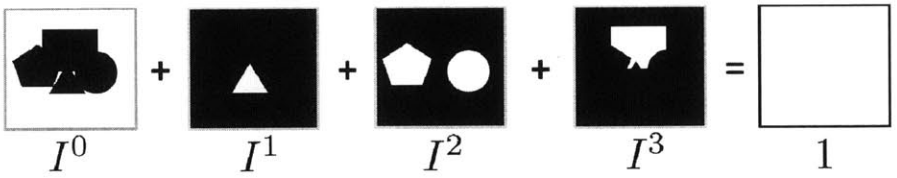


Figure 3 5: **Opacity constraint for the depth map of the layered scene shown in Fig. 3-3.**

3.4 Measurement Model

Data Acquisition

The scene is illuminated with a pre-selected sequence of L binary-valued SLM patterns. Each SLM pattern is denoted with $\boldsymbol{P}^\ell = \{p^\ell(x,y)\}_{x,y=1}^{M}$. Except for the first SLM pattern, all $p^\ell(x,y)$ values are chosen at random from a Bernoulli distribution with parameter 0.5. The first SLM pattern is chosen to be all-ones, i.e., $\{p^1(x,y) = 1\}_{x,y=1}^{M}$. For each illumination pattern we record N digital time-samples $\{r^\ell[n]\}_{n=0}^{N-1}$. In our experiments, we improve SNR by repeatedly illuminating the scene with light pulses for a fixed SLM pattern and averaging the digital time-samples. We assume that our scene is static during data acquisition.

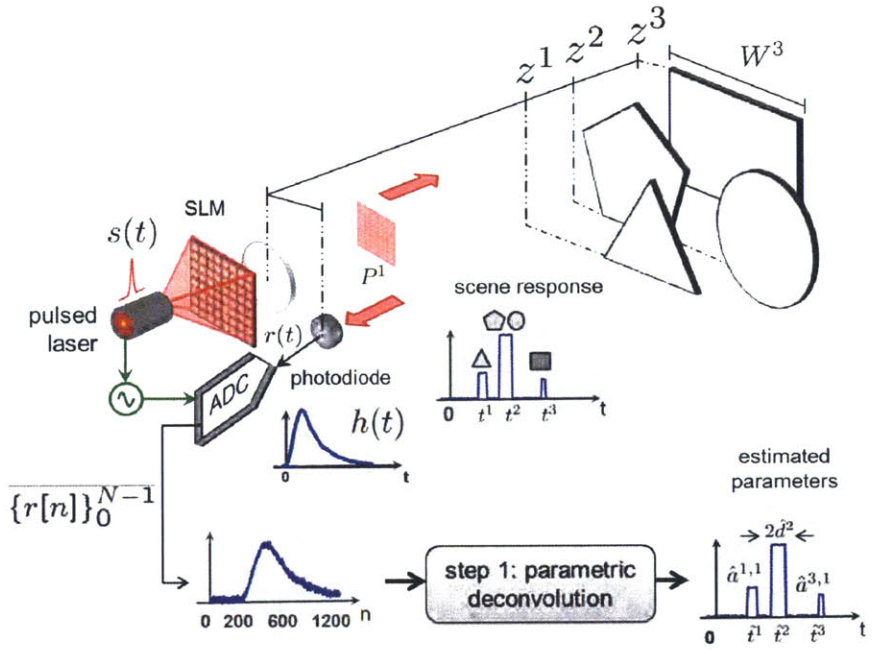


Figure 3 6: **Parametric signal modeling of the scene-response**. The figure shows the various parametric waveforms and received signals when the scene shown in Fig. 3 3 is illuminated with an all ones pattern.

Parametric Scene-response Model

Under the parametric signal model for the scene impulse response, the backreflected light signal incident on the omnidirectional detector is a combination of K non-overlapping waveforms, each corresponding to a distinct scene depth. Additionally, we model the individual waveforms as *boxcar* functions with unknown time duration, $2d^k$, unknown amplitude, $a^{k,\ell}$, and which are approximately centered around $t^k = 2z^k/c$. (see Fig. 3-6). This waveform modeling is based on the assumption that for each object in the scene, the largest object dimension, W^k, is much smaller than the distance of the object from the imaging setup, z^k.

As an example, assume that the object number 3 in Fig. 3-6 had a perfect square-shape, and was placed fronto-parallel to the imaging setup in such a way that the geometric center of the object was in perfect lateral alignment with the optical center of the imager. Using simple Cartesian geometry it can be shown that the duration of the boxcar function corresponding to object number 3 is, $2d^3 = 2(\sqrt{[z^3]^2 + [W^3]^2/4} - z^3)/c$, and the temporal center of the boxcar function is, $t^3 = 2\,(\sqrt{[z^3]^2 + [W^3]^2/4} + z^3)/c$. Additionally, our assumption that

$W^k \ll z^k$ for $k = 1, \ldots, K$ implies that $2d^k \ll t^k$ and ensures that the boxcar waveforms corresponding to the objects at different depth layers do not overlap with each other, i.e., $(t^{k-1} + d^{k-1}) < (t^k - d^k)$ for $k = 1, \ldots, K$.

Note that conservation of total optical flux implies that the area under the k-th boxcar function, which represents the total backreflected light corresponding to object number k, must be equal to the product of object reflectivity, α, and the total surface area of object number k that is illuminated by the light source. For example, the amplitudes of the boxcar functions, $a^{k,1}$, when the scene is illuminated with an all-ones illumination pattern are such that,

$$a^{k,1} \times 2\,d^k = \sum_{x=1}^{M}\sum_{y=1}^{M} \alpha\, i^k(x,y)\, p^1(x,y) = \sum_{x=1}^{M}\sum_{y=1}^{M} \alpha\, i^k(x,y), \quad \text{for } k = 1, \ldots, K. \tag{3.9}$$

We will assume that the temporal position, t^k, and the time duration, $2d^k$, of each boxcar function is independent of the choice of SLM pattern. Thus, because we have assumed that $2d^k \ll t^k$, there is negligible change in the waveform's boxcar shape when the scene is illuminated with different SLM patterns.

Given all the aforementioned assumptions, the photocurrent measured in response to the illumination with SLM pattern $\boldsymbol{P}^\ell$ is,

$$\begin{aligned} r^\ell(t) &= s(t) * h(t) * \left(\sum_{k=1}^{K} a^{k,\ell} \left[u(t - t^k + d^k) - u(t - t^k - d^k) \right] \right) \\ &= g(t) * u(t) * \left(\sum_{k=1}^{K} a^{k,\ell} \left[\delta(t - t^k + d^k) - \delta(t - t^k - d^k) \right] \right). \end{aligned}$$

Here, $u(t)$ is the Heaviside step function, and $\delta(t)$ is the Dirac delta function.

Note that similar to Equation (3.9), the conservation of optical flux implies that the waveform amplitudes, $\{a^{k,\ell}\}_{k=1}^{K}$, are dependent on the SLM patterns. In more precise terms, $(a^{k,\ell} \times 2\,d^k)$ is equal to the standard inner product between the k-th index map, $\boldsymbol{I}^k$, and the

ℓ-th SLM pattern, $\boldsymbol{P}^{\ell}$, i.e., (also see Fig. 3-4)

$$a^{k,\ell} \times 2\,d^{k} = \alpha \sum_{x=1}^{M}\sum_{y=1}^{M} i^{k}(x,y)\,p^{\ell}(x,y), \quad \text{for } k = 1,\ldots,K, \quad \ell = 1,\ldots,L. \tag{3.10}$$

3.5 Novel Image Formation

The scene depth map construction is a two-part algorithm that requires the following inputs:

1. SLM patterns, $\{\boldsymbol{P}^{\ell}\}_{\ell=1}^{L}$, and pixel resolution, $M \times M$,
2. digital time-samples, $\{r^{\ell}[n]\}_{n=0}^{N-1}$, $\ell = 1,\ldots,L$,
3. number of distinct depths, K, and
4. digital time samples of the system transfer function $\{g[n] = g(nT_s)\}_{n=0}^{N-1}$.

Step 1 Estimation of Distinct Depth Values

The first step in depth map formation is to estimate $\{t^{k} = 2z^{k}/c\}_{k=1}^{K}$ using the digital time-samples, $\{r^{1}[n]\}_{n=0}^{N-1}$, recorded in response to the first SLM pattern. In order to accomplish this task we analyze the signals in the Fourier domain,

$$\mathfrak{F}\{r^{1}(t)\} \;=\; \mathfrak{F}\left\{ g(t) * u(t) * \left(\sum_{k=1}^{K} a^{k,1}\left[\delta(t - t^{k} + d^{k}) - \delta(t - t^{k} - d^{k})\right] \right) \right\}$$

where $\mathfrak{F}\{.\}$ denotes the continuous-time Fourier transform (CTFT) operator, By deconvolving $g(t) * u(t)$ from $r^{1}(t)$, we see that it is possible to re-formulate our problem of estimating, $\{a^{k,1}, t^{k}, d^{k}\}_{k=1}^{K}$ as a traditional line spectrum estimation problem [71], or the problem of estimating the frequencies and weights that constitute a mixture of $2K$ sinusoidal signals. In our case the $2K$ time-delays we are seeking act like *pseudofrequencies*, i.e.,

$$\{f^{1} = (t^{1} - d^{1}), \ldots, f_{K} = (t^{K} - d^{K}), f^{K+1} = (t^{1} + d^{1}), \ldots, f^{2K} = (t^{K} + d^{K})\}.$$

Since we only have access to the digital time-samples, $\{r^{1}[n]\}_{n=0}^{N-1}$, it is necessary to formulate the discrete-time line spectrum estimation problem. Since $NT_s >> t^{K} + T_p$ we use the stan-

dard digital signal processing trick of sampling the CTFT, $R(j\omega)$, assuming a periodization of the photocurrent signal, i.e., assuming that $r^1(t) = r^1(t + NT_s)$. In our experiments, we accomplished this by setting the repetition rate of the laser equal to NT_s. Denote the discrete-time Fourier transform samples with $R^1[n]$. It follows from Nyquist sampling theory that $R^1[n] = R^1(j\,2\pi n/N)/T_s, n = 0, \ldots, (N-1)$. Then,

$$\frac{T_s\, R^1[n]}{G[n]\, U[n]} = \sum_{k=1}^{K} a^{k,1}\, e^{-j2\pi f^k n/N} - \sum_{k=K+1}^{2K} a^{k,1}\, e^{-j2\pi f^k n/N}, \quad n = 0, \ldots, (N-1). \tag{3.11}$$

The values $\{R^1[n], G[n], U[n]\}_{n-0}^{N-1}$ are computed using the discrete Fourier Transform (DFT) of the digital samples, $\{r^1[n], g[n], u[n] = u(nT_s)\}_{n=0}^{N-1}$, obtained during the calibration and data acquisition.

Estimation of pseudofrequencies, $\{f^k\}_{k=1}^{2K}$, using the model described in Equation (3.11) is accomplished using standard parametric line spectral estimation methods [71]. Due to noise in the acquisition pipeline, the samples $\{r^1[n]\}_{n=0}^{N-1}$ are corrupted. In our experiments, we reduce noise to inconsequential levels by averaging time-samples over multiple periodic illuminations. Among the various available algorithms, such as Prony's method [72], ESPRIT [71] etc., we found that the Matrix Pencil algorithm [73] achieves the best practical performance for our application in the presence of noise.

It is worth mentioning that this procedure does not require the knowledge of the weight parameters, $\{a^{k,1}\}_{k=1}^{K}$. The pseudofrequency estimates produced by the Matrix Pencil algorithm are sorted in a descending order and labeled appropriately to obtain the pseudofrequency estimates denoted by $\{\hat{f}^k\}_{k=1}^{2K}$. Then the parameters $\{t^k, d^k\}_{k=1}^{K}$ are estimated as follows:

$$\hat{t}^k = \frac{\hat{f}^k + \hat{f}^{K+k}}{2}, \quad \hat{d}^k = \frac{\hat{f}^{K+k} - \hat{f}^k}{2}, \quad k = 1, \ldots, K.$$

The estimates for the K distinct depth values are $\{\hat{z}^k = c\hat{t}^k/2\}_{k=1}^{K}$. The estimates for $\{a^{k,1}\}_{k=1}^{K}$ are obtained by solving the following least-squares problem,

$$\underset{\{x^1,\ldots,x^K\}}{\arg\min} \left\| \underbrace{\begin{bmatrix} \frac{T_s\, R^1[0]}{G[0]U[0]} \\ \vdots \\ \frac{T_s\, R^1[N-1]}{G[N-1]U[N-1]} \end{bmatrix}}_{N\times 1} - \underbrace{\begin{bmatrix} 1 & \cdots & 1 \\ \vdots & & \vdots \\ (e^{-j2\pi\hat{f}^1(N-1)/N} & \cdots & (e^{-j2\pi\hat{f}^K(N-1)/N} \\ -e^{-j2\pi\hat{f}^{K+1}(N-1)/N}) & & -e^{-j2\pi\hat{f}^{2K}(N-1)/N}) \end{bmatrix}}_{N\times K} \underbrace{\begin{bmatrix} x^1 \\ \vdots \\ x^K \end{bmatrix}}_{K\times 1} \right\|_2 \quad (3.12)$$

where $\|.\|_2$ denotes the ℓ_2-norm of a complex vector and $\{x^k\}_{k=1}^K$ are optimization variables.

Step 2 Estimation of Index Maps

The second step in depth map construction is to compute the index maps $\{\boldsymbol{I}^k\}_{k=1}^K$. This is accomplished using the inner product model from Equation (3.10). The time-samples, $\{r^\ell[n]\}_{n=0}^{N-1}$ recorded in response to SLM pattern, $\boldsymbol{P}^\ell$, are used to compute amplitudes estimates, $\{\hat{a}^{k,\ell}\}_{k=1}^K$, by employing the procedure outlined in step 1 (specifically Equation (3.12)). The depth parameters, $\{t^k, d^k\}_{k=1}^K$, are already estimated in step 1 and there is no need to recompute them since these parameters do not depend on the SLM pattern. Once the $K \times L$ amplitude estimates are computed, the optimal solution of the following optimization problem can be used to estimate the K binary-valued index maps,

$$\begin{aligned} \textbf{OPT1}: \quad & \underset{\boldsymbol{X}^0,\boldsymbol{X}^1,\ldots,\boldsymbol{X}^K}{\arg\min} \quad (1-\beta)\sum_{k=1}^{K}\sum_{\ell=1}^{L}\left[\hat{a}^{k,\ell} - \frac{\langle \boldsymbol{X}^k, \boldsymbol{P}^\ell\rangle}{2\alpha d^k}\right]^2 + \beta\sum_{k=0}^{K}\left\|\Phi_I\, \boldsymbol{X}^k\right\|_1 \\ & \text{subject to} \\ & \textbf{C1}: \quad \boldsymbol{X}^0(x,y) + \sum_{k=1}^{K}\boldsymbol{X}^k(x,y) \;=\; 1, \quad x,y = 1,\ldots,M \\ & \textbf{C2}: \quad \boldsymbol{X}^k(x,y) \in \{0,1\}, \quad k = 0,\ldots,K, \quad x,y = 1,\ldots,M. \end{aligned}$$

Here, the first term of the cost function corresponds to model described by Equation (3.10), $\{\boldsymbol{X}^0,\ldots,\boldsymbol{X}^K\}$ are the optimization variables corresponding to the binary-valued index maps,

and the inner product is defined as

$$\langle \boldsymbol{X}^k, \boldsymbol{P}^\ell \rangle \doteq \sum_{x=1}^{M} \sum_{y=1}^{M} \boldsymbol{X}^k(x,y) \boldsymbol{P}^\ell(x,y).$$

The operator Φ_I is the matrix representation of a sparsifying transform for the binary-valued index maps (we use the discrete wavelet transform derived from the 2-tap Daubechies filter), $\|.\|_1$ is the ℓ_1-norm of a real-valued vector and the weight parameter $\beta \in (0,1)$ is introduced to control the degree of spatial regularization.

The constraint **C1** corresponds to Equation (3.8). The index map estimates are the optimal solutions of **OPT1**, however this optimization problem is computationally intractable because of integer constraint **C2**. To make the index map estimation problem tractable and solve it using standard optimization packages [74], we relax the binary constraint **C2** and solve the following convex optimization problem instead:

$$\textbf{R-OPT1:} \quad \underset{\boldsymbol{X}^0, \boldsymbol{X}^1, \ldots, \boldsymbol{X}^K}{\arg\min} \quad (1-\beta) \sum_{k=1}^{K} \sum_{\ell=1}^{L} \left[\hat{a}^{k,\ell} - \frac{\langle \boldsymbol{X}^k, \boldsymbol{P}^\ell \rangle}{2\alpha d^k} \right]^2 + \beta \sum_{k=0}^{K} \left\| \Phi_I \, \boldsymbol{X}^k \right\|_1$$

$$\text{subject to}$$

$$\boldsymbol{X}^0(x,y) + \sum_{k=1}^{K} \boldsymbol{X}^k(x,y) = 1, \quad x,y = 1, \ldots, M$$

$$\boldsymbol{X}^k(x,y) \in [0,1], \quad k = 0, \ldots, K, \quad x,y = 1, \ldots, M.$$

Denote the optimal solutions of **R-OPT1** with $\{\hat{\boldsymbol{X}}^k\}_{k=0}^K$. The index map estimates are computed as follows:

$$\hat{\boldsymbol{I}}^k(x,y) = \begin{cases} 1 & \text{if } \hat{\boldsymbol{X}}^k(x,y) > \hat{\boldsymbol{X}}^{k'}(x,y) \quad k' \in \{0, \ldots, K\} \backslash k \\ 0 & \text{otherwise} \end{cases} \qquad x,y = 1, \ldots, M.$$

A high value of the weight parameter, β, enforces the index maps to be overly smooth while a low-value of β leads to noisy index map estimates. Therefore, an optimal value of β needs to be chosen. In this chapter, we selected this optimal β-value by solving the optimization problem **R-OPT1** for $\beta = \{0.1, 0.2, \ldots, 0.9\}$ and then choosing the one that minimized the

objective function defined in **R-OPT1**.

Finally the depth map estimate, $\hat{\boldsymbol{z}}$, is computed by combining the estimates, $\{\hat{z}^k, \hat{\boldsymbol{I}}^k\}$, i.e., $\hat{\boldsymbol{z}} = \sum_{k=1}^{K} \hat{z}^k \, \hat{\boldsymbol{I}}^k$.

In order to validate the proposed signal models and depth map construction technique, we conduced proof-of-concept experiments whose results are described in the next section.

3.6 Experimental Setup and Results

The light source was a mode-locked Ti:Sapphire femtosecond laser with a pulse width of 100 fs and a repetition rate of 80 MHz operating at a wavelength of 790 nm. The average power in an illumination pattern was about 50 mW.

The laser illuminated a MATLAB-controlled Boulder Nonlinear Systems liquid-crystal SLM with 512×512 pixels, each 15×15 μm. Pixels were grouped in blocks of 8×8 and each block phase-modulated the incident light The phase-modulated beam was passed through a sequence of wave plates and polarizers to obtain the desired binary intensity pattern. At a distance of 10 cm from the detector, the size of each SLM pixel was about 0.1 mm^2. Each SLM pattern in our experiment was randomly chosen and had about half of the 4096 SLM blocks corresponding to zero intensity.

A total of $L = 205$ binary patterns of 64×64 block-pixel resolution ($M = 64$) were used for illumination and construction of a sub-centimeter accurate depth map, implying a measurement compression ratio of $L/M^2 \approx 5\%$ when compared with depth map acquisition using raster-scanning in LIDAR and a sensor array in time-of-flight cameras, both of which require M^2 measurements.

The binary SLM patterns were serially projected onto a scene comprised of four Lambertian planar shapes (see Fig. 3-7A) at different distances. Our piecewise-planar scene consisted of 4 different objects at $K = 3$ distinct depths. The objects were placed at distances between 15 cm and 18 cm from the imaging setup. These planar objects were acrylic cut-outs of various geometric shapes to which Edmund Optics NT83-889 white reflectivity coating had been applied.

For each pattern, the light reflected from all the illuminated portions of the scene was

focused on a ThorLabs DET10A Si PIN diode with a rise time of 0.7 ns and an active area of 0.8 mm^2. A transparent glass slide was used to direct a small portion of the transmitted light into a second photodetector to trigger a 20 GHz bandwidth oscilloscope and obtain the time origin for all received signals. As per step 1, the scene was first illuminated with an all-ones pattern. The time sampling interval, T_s, of the oscilloscope was set such that $N = 1311$ samples of the photocurrent signal were obtained for every transmitted laser pulse. Sensor noise was reduced to inconsequential levels by averaging time-samples over 1000 repeated illuminations for each SLM pattern.

Prior to data processing a simple calibration was performed to obtain the system impulse response samples, $\{g[n]\}_{n=0}^{N-1}$. This was achieved by illuminating a single point on a bright diffuse reflector at 1 cm distance from the imaging setup and measuring the backreflected waveform. The time-samples of this backreflected waveform after normalization to unity peak amplitude were used as the system impulse response. In a separate calibration to obtain object reflectivity, α, a single spot on a bright diffuse reflector at 15 cm range was illuminated and time-samples of the backreflected waveform were obtained. The ratio of the amplitudes of the photocurrents measured at the 15 cm range and at 1 cm range was used as an estimate of the target reflectivity, α.

Figure 3-7(A) shows the relative positions and approximate distances between the SLM focusing lens, the photodetector, and the layered scene. The dimensions of the planar facets are about 10 times smaller than the separation between SLM/photodetector and scene. Thus, there is little variation in the times-of-arrival of reflections from points on any single planar facet, as evidenced by the three concentrated rectangular pulses in the estimated parametric signal using step 1 (see Fig. 3-7(C)). The time delays correspond to the three distinct depths (15 cm, 16 cm and 18 cm). Note that the depth-axis is appropriately scaled to account for ADC sampling frequency and the factor of 2 introduced due to light going to the object and back to the detector. Figure 3-7(D) shows the amplitudes recovered in the case of the first patterned illumination for the scene. Figure 3-7(B) shows the 64×64-pixel depth map reconstructed using time-samples from all of the patterned binary illuminations of the scene. The distinct depth values are rendered in gray scale with closest depth shown in white and farthest depth value shown in dark gray; black is used to denote the scene portions from

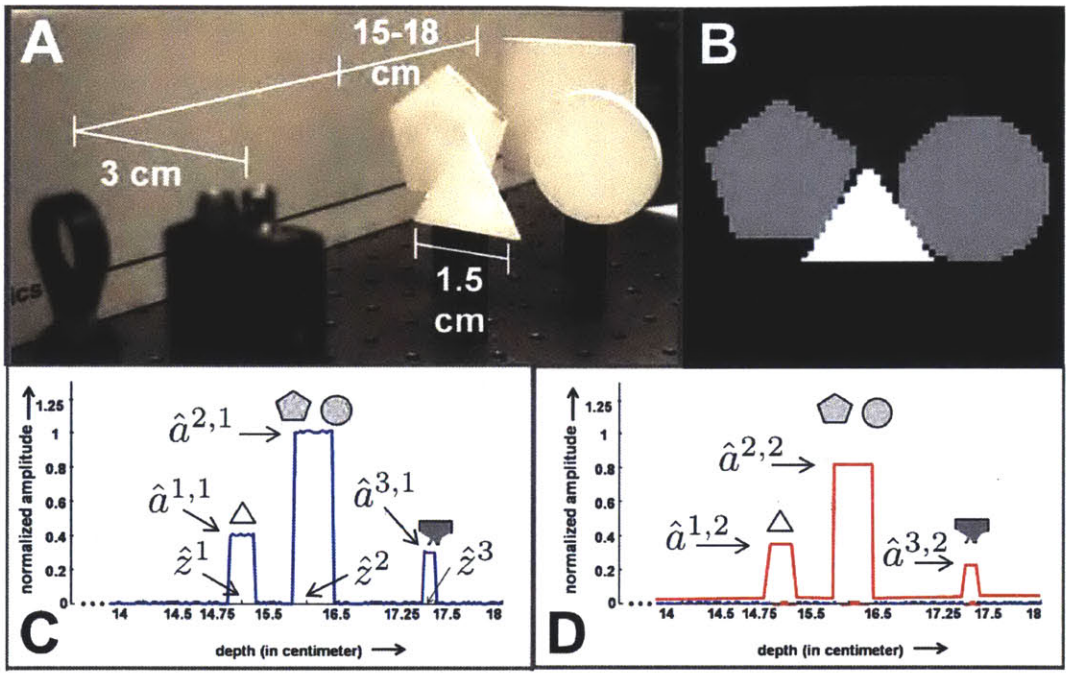


Figure 3 7: **Experimental setup and results** (A). Photograph of experimental setup. (B). Estimated depth map. (C). Parametric signal estimate in response to first (all ones) SLM illumination. (D). Parametric signal estimate in response to second (pseudorandomly chosen) SLM illumination.

which no light was collected.

Our technique yielded accurate sub-cm resolution depth-maps with sharp edges. The depth resolution of our acquisition method the ability to resolve objects that are closely separated in depth depends on the pulse duration of the temporal light modulation, the photodetector response time, and ADC sampling rate. The spatial resolution of our output depth map is a function of the number of distinct patterned scene illuminations; a complex scene with a large number of sharp features requires a larger number of SLM illuminations.

3.7 Discussion and Limitations

In this chapter, we presented a method for acquiring depth maps of fronto-parallel scenes using time samples of backreflected light measured using a single photodetector in response to a series of spatiotemporally-modulated scene illuminations. The two central novelties of our work, relative to standard LIDAR systems and time-of-flight cameras, is our mechanism for

obtaining spatial resolution through spatially-patterned illumination, and the exploitation of spatial correlations present in real-world objects to reduce the number of measurements required for accurate depth map acquisition.

In principle, the techniques proposed in this chapter may reduce data acquisition time relative to a LIDAR system because an SLM pattern can be changed more quickly than a laser position, and the number of acquisition cycles M is far fewer than the number of pixels in the constructed depth map N^2. The savings relative to a time-of-flight camera may be in the number of sensors. For a fair comparison, however, we must consider several other factors that govern the total acquisition time.

For example, consider a LIDAR system which employs point-by-point acquisition using raster-scanned illumination and a single omnidirectional avalanche photodiode. The total data acquisition time is equal to the product of the dwell time/pixel and the number of pixels in the depth map. In the absence of background light, tens of detected photons at each pixel, which can be acquired using a very short dwell time, are sufficient to form a high quality depth map obtained by simply averaging the photon detection times measured at each spatial location.

In contrast, compressive depth acquisition requires fewer illumination patterns but a long dwell time/pattern may be necessary to capture the backreflected scene response. If the scene comprises of several depth layers, then the total acquisition time for compressive depth acquisition may exceed that of a raster-scanned LIDAR system. Also, the computational depth map construction employed in this chapter is algorithmically complex and computationally time consuming relative to the simpler processing used in LIDAR systems. A more detailed analysis of such trade-offs is necessary to conclusively determine the imaging scenarios in which compressive depth acquisition outperforms traditional active imagers.

Beyond the preceding considerations, there are several limitations of the imaging framework presented in this chapter that should be mentioned. The first is the inapplicability of the compressive depth acquisition framework to construct depth maps of scenes comprising curvilinear objects. In this chapter, we restricted our attention to layered scenes comprised of planar objects, all with the same reflectivity, placed at distinct depths. A slightly more general case of scenes comprising inclined planar objects is discussed in [6]. Beyond this

generalization, adapting the compressive depth acquisition framework to curvilinear objects requires significant extensions of the current mathematical model, which could be the subject of future work.

Another limitation is in relation to spatially-varying scene reflectivity. If the objects in the scene have high reflectivity variations, then the measurements deviate significantly from our model, which causes depth map construction errors. In [6], we presented a simple solution to accommodate for known scene reflectivity at each SLM pixel position (such as obtained through a reflectivity image). The proposed solution assumes an SLM capable of performing grayscale light modulation (the experimental results in this chapter used a binary light modulation SLM), and the basic idea is to attempt to compensate for varying scene reflectivity by illuminating the scene with an inverse illumination pattern, i.e., by illuminating low reflectivity scene pixels with a proportionately more transparent SLM pattern and high reflectivity scene pixels with a proportionately more blocking SLM pattern.

The framework presented in this chapter also assumes that the number of depth layers, K, is known. This is a hard assumption to satisfy in practice. However, note that K is the model order in the line spectrum estimation problem formulated in Equation (3.11). It is theoretically and given a good signal-to-noise ratio in the time-samples practically feasible to estimate the model order K using the data, $\{r^1[n]\}_{n=0}^{(N-1)}$. There is well developed literature around this topic [71, 75 77].

In the next chapter, we present an implementation of the compressive depth acquisition framework in a low-light level imaging setting.

Chapter 4

Low-light Level Compressive Depth Acquisition

4.1 Overview

In this chapter we describe the extension of the compressive depth acquisition framework, developed in Chapter 3, to low-light imaging applications. The depth map construction in this chapter will rely on parametric signal modeling and processing to recover the set of distinct depths present in the scene (Section 3.3). As before, using a convex program that exploits the transform-domain sparsity of the depth map, we recover the spatial content at the estimated depths (Section 3.4).

One key difference here is that instead of using a light source which is spatially modulated by an SLM, we achieve spatial resolution through patterned sensing of the scene using a digital micromirror device (DMD) (see Fig. 4-1). In essence, the imaging configuration in this chapter is the *dual* of the imaging setup described in Chapter 3. The experimental results in Section 4.3 describe the acquisition of 64×64-pixel depth maps of fronto-parallel scenes at ranges up to 2.1 m using a pulsed laser, a DMD and a single photon-counting detector. These room-scale experimental results are in contrast with the near-range scenes (less than 0.2 m away) comprised of opaque objects that were imaged in Chapter 3. Additionally, we also experimentally demonstrate imaging in the presence of unknown partially-transmissive

occluders. The compressive depth acquisition prototype and experimental results presented here provide a potential directions for non-scanning, low-complexity depth acquisition devices for various practical imaging applications (also see [8]).

4.2 Experimental Setup

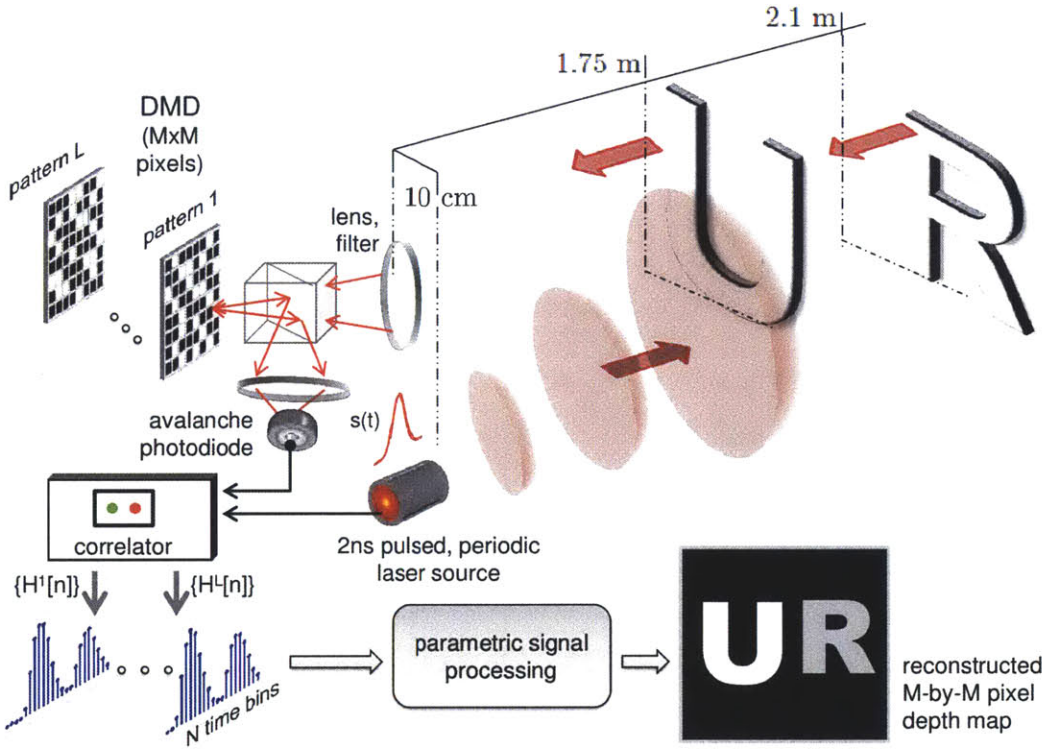


Figure 4 1: **Imaging setup for low-light level compressive depth acquisition.** Shown is a pulsed light source with pulse shape, $s(t)$, a DMD array with $M \times M$ pixel resolution, and a single photon counting detector (APD). For each sensing pattern, $\boldsymbol{P}^{\ell}$ the scene is repeatedly illuminated with light pulses and the photon detections are used to generate a photon count histogram with N time bins. This process is repeated for L pseudorandomly chosen binary patterns and the $N \times L$ histogram samples are processed using the computational framework outlined in Section 3.5 to construct an $M \times M$ pixel scene depth map.

Active illumination: The schematic imaging setup, shown in Fig. 4-1, and the corresponding physical setup, shown in Fig. 4-2, consist of an illumination unit and a single detector. The illumination unit comprises a function generator that produces 2 ns square pulses that drive a near-infrared (780 nm) laser diode to illuminate the scene with 2 ns Gaussian-shaped pulses with 50 mW peak power and a repetition rate of 10 MHz implying

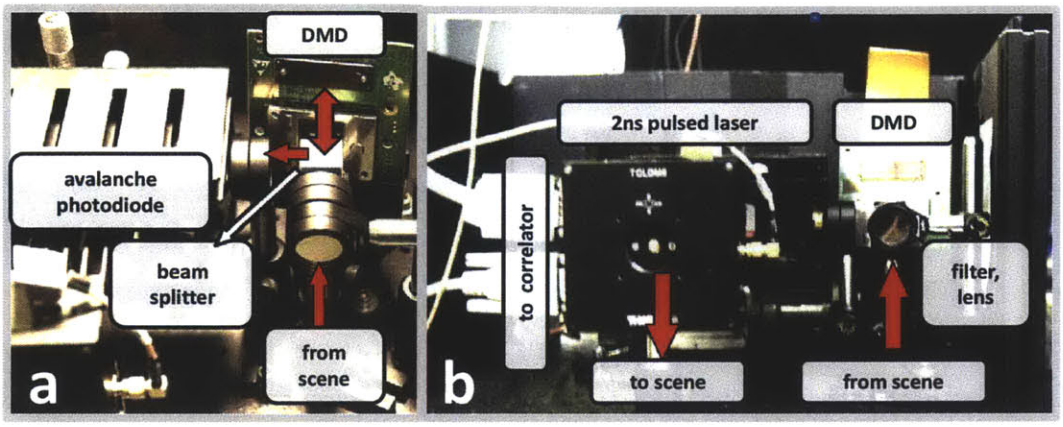


Figure 4 2: **Experimental setup for low-light level compressive depth acquisition.** (a) Close up of the sensing unit showing the optical path of backreflected light from the scene. (b) The complete imaging setup showing the pulsed source and the sensing unit.

a pulse repetition interval, $T_r = 100$ ns.

Details of the scene setup: Scenes are set up so that objects are placed fronto-parallel to the data acquisition setup. Objects are 30 cm-by-30 cm cardboard cut-outs of the letters **U** and **R** placed at distances $z^1 = 1.75$ m and $z^2 = 2.1$ m respectively. The two objects were painted with a white Lambertian coating of known reflectivity, α.

Detector-side spatial patterning: Backreflected light from the scene is imaged onto a DMD through a 10 nm filter centered at 780 nm with a 38 mm lens focused at infinity with respect to the DMD. We use a D4100 Texas Instruments DMD that has 1024×768 individually-addressable micromirrors. Each mirror can either be "ON", where it reflects the backreflected light into the detector ($p(x, y) = 1$), or "OFF", where it reflects light away from the detector ($p(x, y) = 0$). For the experiment we used only 64×64 block-pixels of the DMD to collect backreflected light. For each scene we recorded photon-arrival data for $L = 2000$ patterns in total. Each of the DMD patterns were pseudorandomly-chosen as described in Section 3.3. The pattern values are chosen uniformly at random to be either 0 or 1.

Low-light level detection: The detector is a cooled avalanche photodiode (APD) operating in Geiger mode. Upon photon detection, the detector outputs a TTL pulse about

10 ns in width, with edge timing resolution of about 300 ps. After a photon is detected, the detector then enters a dead time of about 30 ns during which it is unable to detect photons.

To measure the time-of-arrival of a detected photon relative to the transmitted pulse, we use a correlating device (Picoquant Timeharp) designed for time-correlated single-photon counting. The correlator has two inputs: *start* and *stop*. The output of the laser pulse generator is wired to start, and the APD output is wired to stop. Whenever an impinging photon causes the APD to generate an electrical impulse, the correlator outputs a time-of-arrival value that is accurate within a time-bin. In our experimental setup the duration of this time-bin, denoted by Δ, was 38 ps. In the low-light level formulation of compressive depth acquisition, this time-bin width is equivalent to the sampling period, T_s, defined in Section 3.3. Note that a photon arrival time cannot exceed the pulse repetition interval T_r.

The theory of operation of APDs and the Poisson statistics of low-light detection are described in the next chapter. The APD used in our experiment is not a number-resolving detector, so it can only detect one photon arrival per illumination pulse. Therefore it was important that in our experiment we maintained a low optical flux level (see Section 5.4). This assumption is perfectly aligned with our goal of imaging in low-light level scenarios.

Under the low optical flux assumption it is possible to repeatedly illuminate the scene with light pulses to build up a histogram of the arrival times of the detected photons, also referred to as a photon-count histogram [78, 79]. For pattern ℓ, denote this histogram with $\{H^{\ell}[n]\}_{n=0}^{N-1}$ where $N = T_r/\Delta$. In the low-light level formulation of compressive depth acquisition, this histogram is the equivalent of the discrete time-samples $\{r^{\ell}[n] = r(n\Delta)\}_{n-0}^{N-1}$ with the appropriate normalization. The photon counting mechanism and the process of measuring the timing histogram are illustrated in Fig. 4-3. In our experiment, we used a dwell time of 1 second per DMD pattern in order to build the photon-count histogram. Despite this long acquisition time, in which approximately 10^7 illumination pulses were transmitted toward the scene, the backreflected light levels were low enough that the effect of Poisson or shot noise [80, 81] was prominent in our histogram's bin values. Poisson noise and its effects are described in more detail in Section 5.3.

Prior to data processing a simple calibration was performed to obtain the photon-count histogram of the system impulse response, denoted by $\{G[n]\}_{n=0}^{N-1}$. This was achieved by

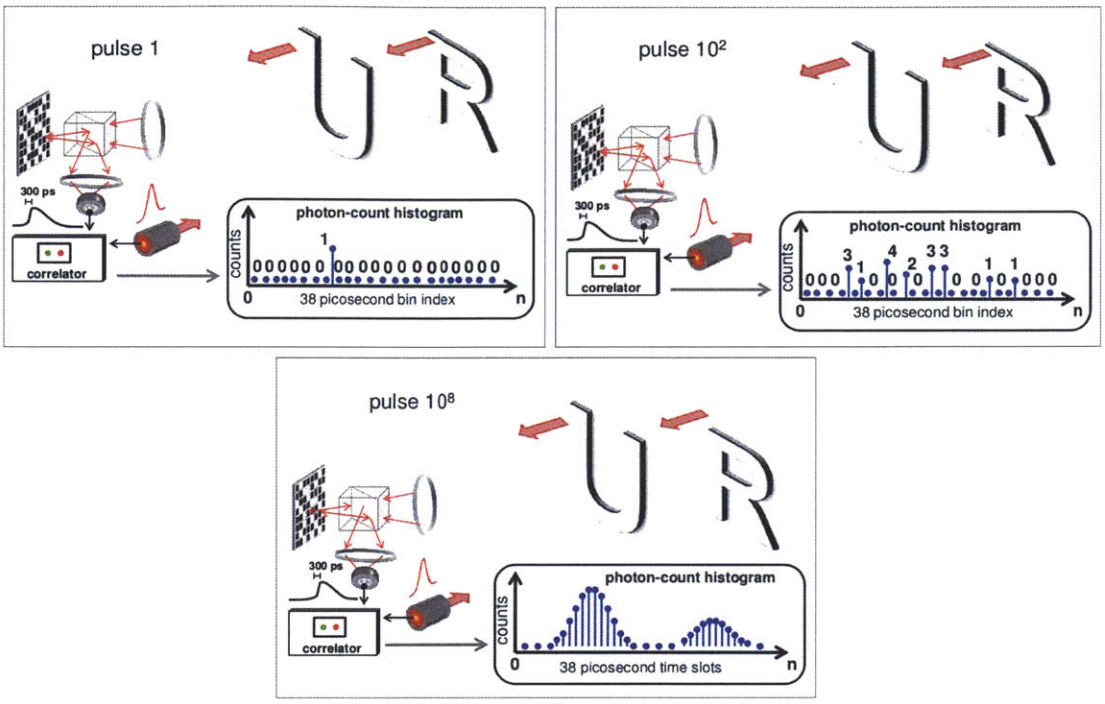


Figure 4 3: **Generating a photon-count histogram.** For a fixed DMD pattern, the scene is repeatedly illuminated with light pulses. The APD + time correlator device combination records the time of arrival of single photons with a 38 ps accuracy. Under our low optical flux assumption, for each illumination pulse at most one photon is detected. The build up of photon counts in the different time bins of the histogram, $\{H^{\ell}[n]\}_{n=0}^{N-1}$, in response to an increasing number of pulsed illuminations is shown.

illuminating a single point on a bright diffuse reflector at 5 cm range and building the histogram of photon counts. This histogram, after normalization to unity peak amplitude, was used as the system impulse response. In a separate calibration to obtain object reflectivity, α, a single spot on a bright diffuse reflector at 2 m range was illuminated and the histogram of photon counts was constructed. The amplitude of the histogram was used as an estimate of the target reflectivity, α.

4.3 Data Processing and Experimental Results

In this section we discuss depth map constructions for two scenes using varying number of measurements, L. The first scene (see Fig. 4-5) has cardboard cut-outs of the letters **U**, **R** placed at 1.75 m and 2.1 m respectively from the imaging setup. The depth values and

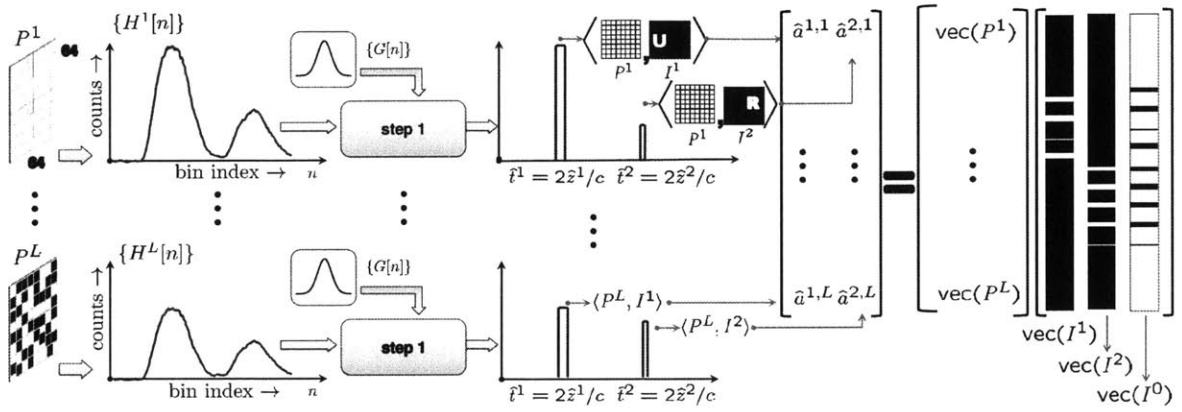


Figure 4 4: **Depth map construction algorithm.** The photon count histogram values for the first DMD pattern $\{H^1[n]\}_{n=0}^{N-1}$, along with the photon count histogram of the system impulse response, $\{G[n]\}_{n=0}^{N-1}$ are processed using step 1 to recover scene depth estimates, $\hat{z}^1$ and $\hat{z}^2$, which along with the knowledge of the DMD patterns, and the photon count histograms for the remaining DMD patterns, are further employed in step 2 processing to compute the inner product estimates $\{\hat{a}^{k,\ell}\}_{k=1,\ell=1}^{K,L}$. Finally the convex optimization program **R-OPT1** is used to reconstruct the index maps $\hat{\boldsymbol{I}}^1$, $\hat{\boldsymbol{I}}^2$ and $\hat{\boldsymbol{I}}^0$.

corresponding time-shifts are identified from the histogram samples per step 1 processing (see Section 3.5, also see Fig. 4-4). The recovered values for scene depths were, $\hat{z}^1 = 1.747$ m and $\hat{z}^2 = 2.108$ m, which correspond to approximately 1 cm depth accuracy. In our experiment, only one dataset corresponding to the all-ones illumination pattern was used to compute these depth estimates, the standard deviation or error bias could not be calculated.

Recovery of spatial correspondences through index maps proceeds as described in step 2 processing (Section 3.5). In Fig. 4-5 we see index maps for our first experimental scene (Fig. 4-5 (a)) for two different numbers of total patterns used. At $L = 500$ patterns (12% of the total number of pixels), we can clearly identify the objects in index maps $\hat{\boldsymbol{I}}^1$ and $\hat{\boldsymbol{I}}^2$ with some distortions due to shot noise in photon-count histogram measurement; we also see the background index map, $\hat{\boldsymbol{I}}^0$, corresponding to scene regions that do not contribute any reflected returns. Using $L = 2000$ patterns (48.8% of the total number of pixels) the proposed method reduces the noise level further, providing even more accurate index maps.

Imaging scenes with unknown transmissive occluders. In a second scene we considered a combination of transmissive and opaque objects and attempted to recover a depth map. The scene is shown in Fig. 4-6(a). Note that the burlap placed at 1.4 m from the

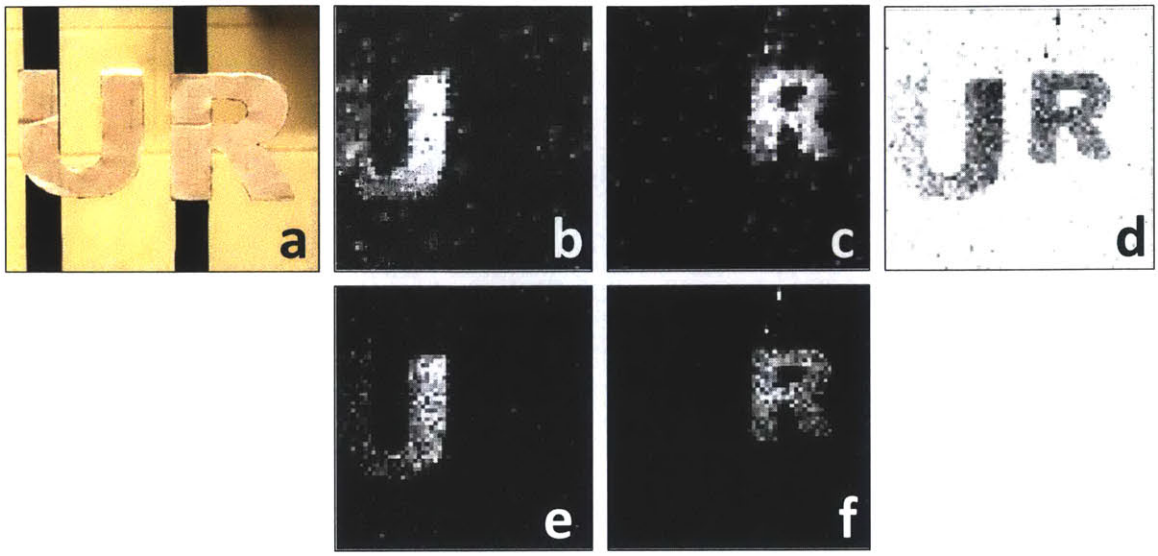


Figure 4 5: **Depth map construction results.** Scene layout is shown in (a) (the **U** and **R** cut outs are at different distances per Fig. 4 2). The index masks recovered using $L = 500$ patterns, $\hat{\boldsymbol{I}}^1$, $\hat{\boldsymbol{I}}^2$ and $\hat{\boldsymbol{I}}^0$ are shown in (b), (c) and (d) respectively. The index masks $\hat{\boldsymbol{I}}^1$, $\hat{\boldsymbol{I}}^2$, recovered using $L = 2000$ patterns are shown in (e) and (f) respectively.

imaging device completely fills the field of view. A 2D photograph of the scene would reveal only the burlap. However, located at 2.1 m from the imaging device are cardboard cut-outs of U and R both at the same depth. These objects would be completely occluded in a 2D reflectivity image. Also seen in Fig. 4-6 is a photon-count histogram acquired with a longer dwell time of 4 seconds per DMD pattern. The histogram shows that the burlap contributes a much larger reflected signal (approximately 12 times stronger) than the contribution of the occluded scene. Figure 4-6 shows depth masks $\hat{\boldsymbol{I}}^1$, $\hat{\boldsymbol{I}}^2$ for the burlap and occluded objects respectively for $L = 500$ and $L = 2000$ patterns.

Note that in this depth map construction, we did not assume the presence of a partially transmissive occluder. This is in contrast with the imaging methods which rely on range-gating to look through foliage [17, 82]. In our case, the presence of the burlap and its shape and position are revealed entirely through the depth map construction step.

Achieving high depth resolution with slower response systems. Consider the un-occluded imaging setup shown in Fig. 4-5. When the planar facets are separated by distances

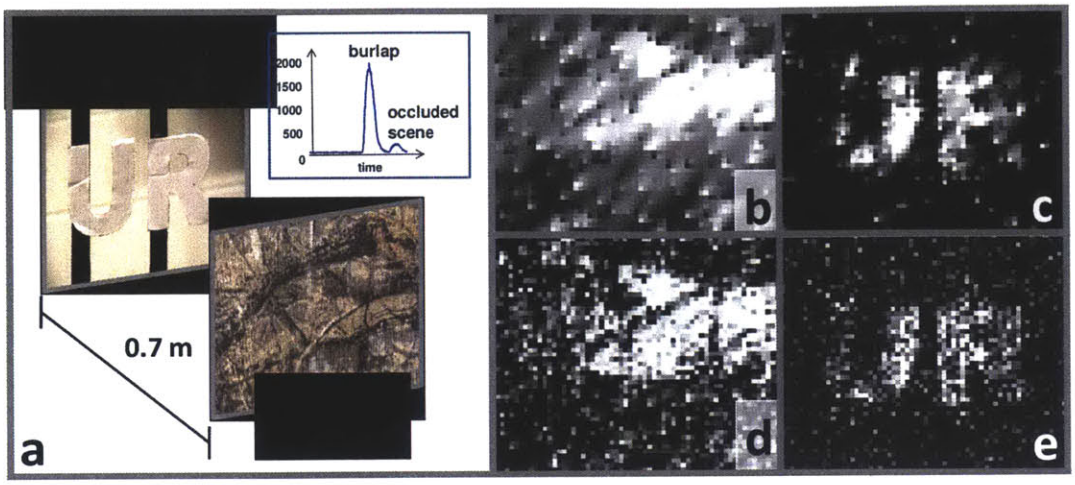


Figure 4 6: **Occluded scene imaging.** (a) Setup for the scene occluded with a partially transmissive burlap occluder; the shapes U and R are at the same depth. (b) and (c) Reconstructed depth maps for burlap and scene using $L = 500$ patterns; (d) and (e) using $L = 2000$ patterns. Note that no prior knowledge about the occluder was required for these reconstructions.

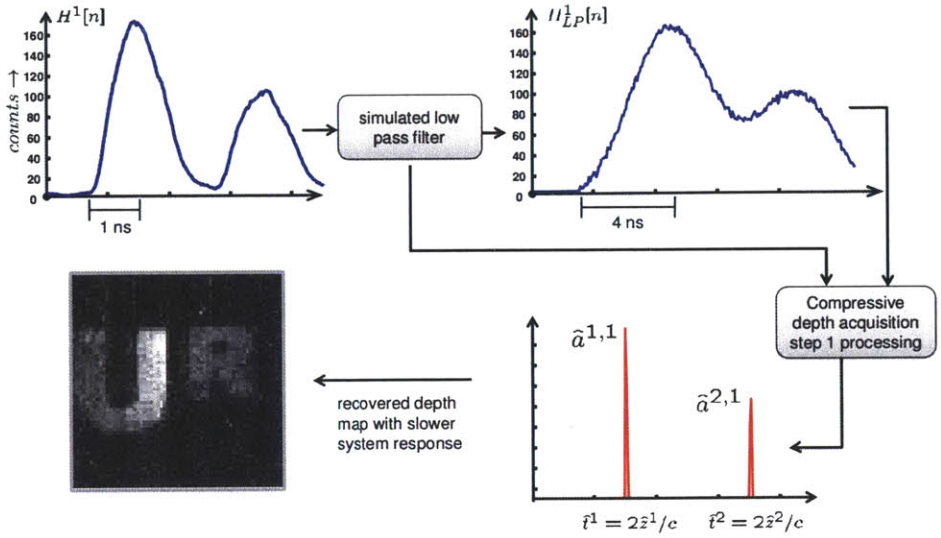


Figure 4 7: **Depth map reconstruction with simulated scene response to longer source pulse width.** We simulated poor temporal resolution by low pass filtering the captured photon count histograms so that the waveform corresponding to the two depth layers have significant overlap which perturbs the true peak values. The knowledge of the pulse shape, together with the parametric deconvolution framework proposed in step 1 of our computational imager allows accurate recovery of the two distinct depth values from the low pass filtered histogram profiles. These accurate depth values further enable accurate depth map construction using step 2 processing.

that correspond to time differences greater than the pulse width of the source, the time shift information can be trivially separated. The more challenging case is when the objects are

placed closed enough in depth, such that the backreflected pulses corresponding to distinct objects overlap. Depending on the object separation, the waveforms corresponding to the objects at different depths overlap with varying degree. In an extreme case, the waveform peaks may no longer be distinctly visible, or in a moderate case the overlap causes the observed peak positions to shift away from their accurate values. In both these cases, it is difficult to estimate the true depth values from simply observing the peak positions.

In the case of fronto-parallel scenes, the use of a parametric signal modeling and recovery framework (step 1) outlined in Section 3.5 may enable us to achieve high depth resolution relative to the overall system's temporal response. As an example, we demonstrate accurate depth map construction for the scene shown in Fig. 4-5, through simulating slower system response time by convolving the photon-count histograms with a Gaussian to correspond to a time-resolved system that has a four times slower rise time when compared to the system used in our experiments. These filtered histogram samples were processed through the compressive depth map construction steps 1 and 2 and the results corresponding to $L = 2000$ patterns is shown in Fig. 4-7. Note that in the absence of well-separated waveforms, techniques such as those implemented in [15] fail to resolve the two depth layers clearly and result in depth map construction errors. Despite our method's improved performance for scenes composed of fronto-parallel objects, the proposed techniques fail for scenes comprising curvilinear objects.

4.4 Discussion and Conclusions

The experimental setup and results presented in this chapter are a step towards practical compressive depth acquisition at low-light levels such as will be the case of standoff (long-distance) applications. A key practical shortcoming of the framework developed in Chapter 3 lies in the projection of the binary patterns on to the scene with an SLM. When the object ranges exceed a couple of decimeters, the SLM patterns suffer significant distortions that lead to severe reconstruction artifacts due to model mismatch. The experimental setup described in this chapter uses flood illumination of the scene followed by spatial patterning at the detector end, thereby implicitly resolving the aforementioned challenge with using an SLM for compressive depth acquisition.

Performance analysis for our depth acquisition technique entails analysis of the dependence of accuracy of depth recovery and spatial resolution on the number of patterns, scene complexity and temporal bandwidth. While the optimization problems introduced in our paper bear some similarity to standard compressed sensing problems, existing theory does not apply directly. This is because the amplitude data for spatial recovery is obtained after the scene depths are estimated in step 1, which is a nonlinear estimation step. The behavior of this nonlinear step in presence of noise is an open question even in the signal processing community. Moreover, quantifying the relationship between the scene complexity and the number of patterns needed for accurate depth map formation is a challenging problem.

Analogous problems in the compressed sensing literature are addressed without taking into account the dependence on acquisition parameters, whereas in our active acquisition system, illumination levels certainly influence the spatial reconstruction quality as a function of the number of measurements. Analysis of trade-offs between acquisition time involved with multiple spatial patterns for the single-sensor architecture and parallel capture using a 2D array of sensors (as in time-of-flight cameras) is a question for future investigation. The main potential advantage of the compressive depth mapping systems proposed in Chapter 3 and Chapter 4 lies in exploiting sparsity of natural scene depth to reduce the acquisition time, hardware cost, and hardware complexity without sacrificing depth accuracy and resolution.

Chapter 5

First-photon Imaging

5.1 Overview

Acquisition of three-dimensional (3D) structure and reflectivity of objects using active imagers that employ their own illumination, such as the Microsoft Kinect [52], typically requires millions of detected photons at each sensor pixel. Low-light level active imagers that employ photon-counting detectors, such as Geiger-mode avalanche photodiodes (APDs) [83], can acquire 3D and reflectivity images at extremely low photon fluxes. For example, in 3D light detection and ranging (LIDAR) systems [9, 84], the scene is illuminated with a stream of laser pulses, the backreflected light is detected with a Geiger-mode APD, pixel-by-pixel range information is obtained from histograms of the time delays between transmitted and detected pulses [79], and pixel-by-pixel relative reflectivity is found from the number of photons detected in a fixed dwell time.

Despite the use of highly sensitive photodetectors, hundreds of photon detections per pixel are still required for accurate range and reflectivity imaging, because photon-counting detectors are limited by Poisson noise, including that generated by ambient (background) light [80, 85].

We introduce *first-photon imaging*, a framework which allows us to capture accurately and simultaneously 3D spatial structure and reflectivity using only the first photon detection at each pixel (see Fig. 5-1). It is a new computational paradigm for low-flux imaging

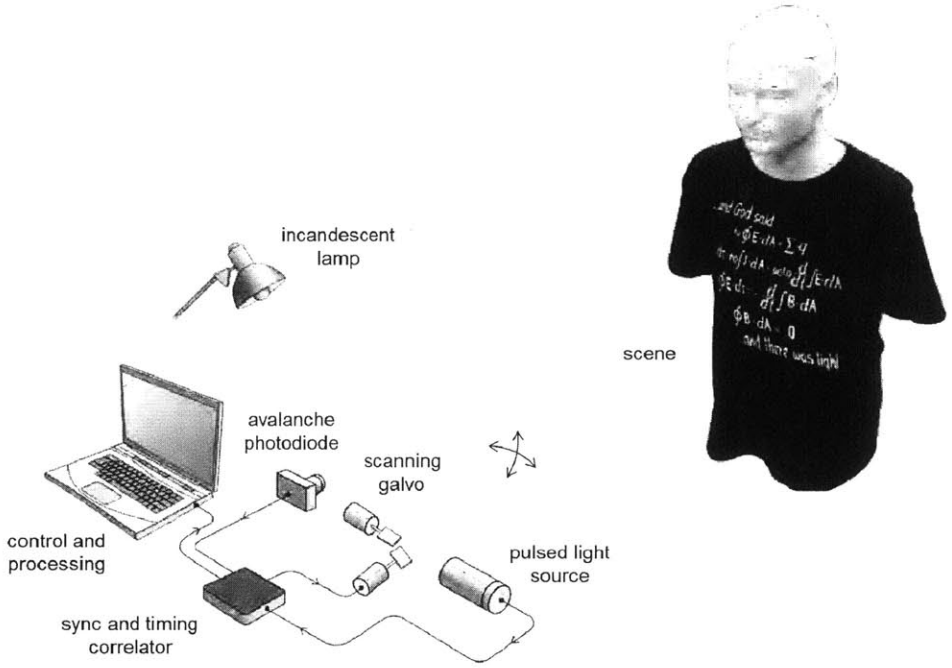


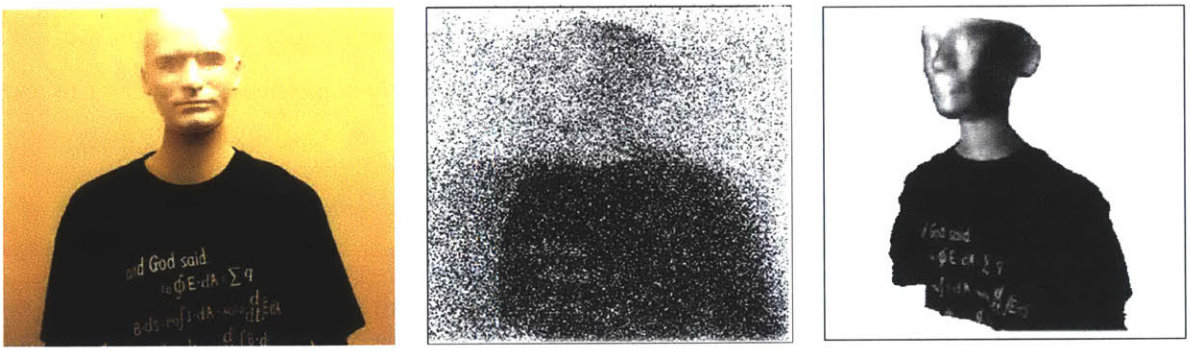
Figure 5 1: **First-photon imaging setup:** A pulsed light source illuminates the scene in a raster scan pattern. An incandescent lamp injects background light that corrupts the information bearing signal. The incident light is collected by a time resolved single photon detector. Each spatial location is repeatedly pulse illuminated until the first photon is detected. The photon's arrival time relative to the most recent transmitted pulse and the number of elapsed illumination pulses prior to first detection are recorded. This dataset is used to computationally reconstruct 3D structure and reflectivity.

that produces high-quality range images, despite the presence of high background noise, and high-quality reflectivity images, when a conventional reflectivity image built from one photon detection per pixel would be featureless. These surprising results derive from exploiting the spatial correlations present in real-world scenes within a computational framework that is matched to the physics of low-flux measurements.

For each pixel our computational imager uses the number of illumination pulses prior to the first photon detection as an initial reflectivity estimate. Poisson noise precludes these pixel-by-pixel estimates from providing a high-quality reflectivity image. So we suppress that noise by exploiting the high degree of spatial correlation present in real-world scenes, i.e., that neighboring pixels have strong distance and reflectivity correlations punctuated by sharp boundaries. Such correlations are captured through sparsity in the scene's discrete wavelet transform (DWT) coefficients [86, 87]. Thus we suppress Poisson noise in the reflectivity

image by means of a DWT-based regularization that does not sacrifice transverse spatial resolution. We also exploit spatial correlations to censor background-generated (anomalous) range values from an initial pixel-by-pixel range image.

By forming high quality images using far fewer detected photons (see Fig. 5-2), our method dramatically broadens the applicability of imaging techniques, for example to situations involving active imagers with very low optical output power, or to imaging applications where little backreflected light reaches the detector [79, 84, 88 90].



A B C

Figure 5 2: **Processing the first-photon experimental data. A**. A color photograph of the mannequin used in our experiments. For sub figures **B** and **C**, 3D estimates are rendered as point clouds and overlaid with reflectivity estimates, with no additional post processing applied. **B**. Reconstruction using traditional maximum likelihood estimation (notice the grainy appearance and low image contrast due to Poisson noise from the signal and background light). **C**. Our first photon imaging method recovers a great amount of detail (specular highlights, text and facial structure) that is heavily obscured in **B**.

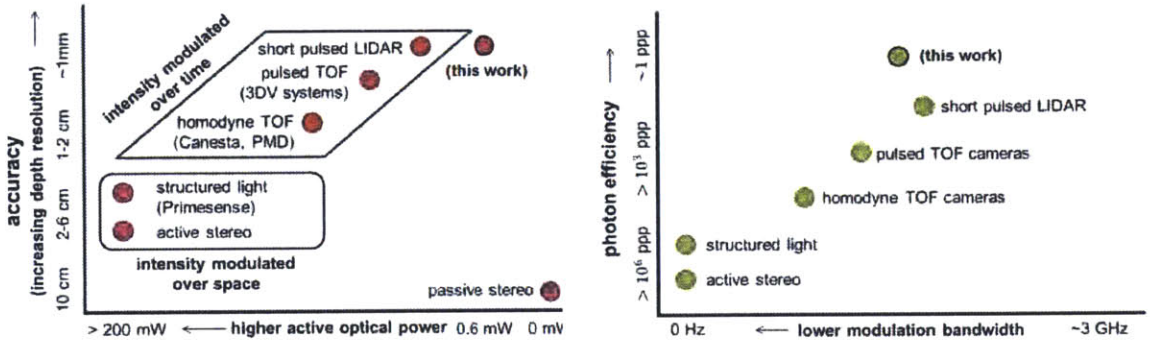
The rest of this chapter is organized as follows. In Section 5.2 we review state-of-the-art methods in active optical imaging and low-light level sensing. In Sections 5.3 and 5.4 we introduce the first-photon imaging image acquisition setup, notation and signal models. Section 5.5 describes image formation using conventional maximum-likelihood estimation, and Section 5.6 introduces the 3D and reflectivity image formation methods based on first-photon data. Finally Sections 5.7 and 5.8 discuss the experimental setup and results, ending with a short discussion of the limitations of the first-photon imaging framework in Section 5.9.

Note that in this chapter we use the terms scene depth map and scene 3D structure interchangeably, since they are in one-to-one correspondence with each other, related by a linear transformation which depends on the optical parameters of the system [91] (also see

Section 5.7, which describes the relation between 3D scene geometry and depth map data).

5.2 Comparison with Prior Art

Photon Efficiency of active optical 3D imagers: Active optical 3D imaging systems differ in how they modulate their transmitted power, leading to a variety of trade-offs in accuracy, modulation frequency, optical power, and photon efficiency; see Figure 5-3 for a qualitative summary. The ordering of time-of-flight sensors by increasing modulation bandwidth (decreasing pulse duration) is: homodyne time-of-flight cameras [3], pulsed time-of-flight cameras [53, 92], and picosecond laser radar systems [24, 90]. Compared with time-of-flight sensors, active imagers employing spatial modulation [45, 52, 55, 93] have low photon efficiencies because they use an always-on optical source. Additionally, the systems using temporal modulation typically have better depth accuracy and resolution than those using spatial modulation. The advantage of spatial modulation tends to be cheaper sensing hardware, since high-speed sampling is not required.



A. Accuracy vs. power trade offs.

B. Photon efficiency vs. modulation bandwidth trade offs.

Figure 5 3: Qualitative comparison of state of the art active optical 3D sensing technologies. Photon effi ciency is defined as photons per pixel (ppp) necessary for centimeter accurate depth imaging.

The most photon-efficient time-of-flight imagers those requiring the fewest photons for accurate imaging use single-photon avalanche diode (SPAD) detectors [94]. Earlier efforts in SPAD-based 3D imaging from on the order of 1 detected photon/pixel are reported in [95

99]. The framework presented here improves upon these works in part due to the use of estimated reflectivity.

Optoelectronic techniques for low light level sensing: In low light-level scenarios, a variety of optoelectronic techniques are employed for robust imaging. Active imagers use lasers with narrow spectral bandwidths and spectral filters to suppress background light and minimize the Poisson noise it creates. However, optical filtering alone cannot completely eliminate background light, and it also causes signal attenuation. Range-gated imaging [17] is another common technique, but this method requires a priori knowledge of object location. Furthermore, a SPAD may be replaced with a superconducting nanowire single-photon detector (SNSPD) [100], which is much faster, has lower timing jitter, and has lower dark-count rate than a SPAD. However, SNSPDs have much smaller active areas and hence have narrower fields of view than SPAD-based systems with the same optics.

Reflectivity inference: Under a continuous model of integrating light intensity, the time to reach a fixed threshold is inversely proportional to the intensity, and this principle has been employed previously for imaging under bright lighting conditions [101]. Our reflectivity imaging for the first time pushes this concept to the extreme of inferring reflectivity from the detection of a single photon.

Image denoising: For depth imaging using SPAD data, it is typical to first find a *pixelwise* or *pointwise* maximum likelihood (ML) estimate of scene depth using a time-inhomogeneous Poisson process model for photon detection times [78,80] and then apply a denoising method. The ML estimate is obtained independently at each pixel, and the denoising is able to exploit the scene's spatial correlations. This two-step approach commonly assumes a Gaussian noise model, which is befitting because of the optimal behavior of ML with large numbers of data samples [102]. At low light levels, however, performing denoising well is more challenging due to the signal-dependent nature of Poisson noise. In Section 5.8, we compare our technique with the state-of-the-art denoising methods that use sparsity-promoting regularization. Our superior performance is due in part to our novel method for classifying detection events as

being due to signal (backscattered light) or noise (background light and dark counts).

5.3 Imaging Setup and Signal Modeling

Figure 5-1 shows the imaging setup. Scene patches and the corresponding image pixels are indexed with (x, y). Distance to scene patch (x, y) is denoted by $z(x, y) \geq 0$. Scene patch reflectivity is denoted by $\alpha(x, y) \geq 0$ and it includes the effect of radial fall-off, view angle, and material properties. Our goal is to form an $M \times M$ pixel reflectivity image $\boldsymbol{\alpha} = \{\alpha(x, y)\}_{x,y=1}^{M} \in \mathbb{R}_{+}^{M \times M}$ and an $M \times M$ pixel depth map $\boldsymbol{z} = \{z(x, y)\}_{x,y=1}^{M} \in \mathbb{R}_{+}^{M \times M}$ of the scene. Our imaging setup uses 2D raster scanning, however our models and algorithms also apply a 2D SPAD array with floodlight illumination as described in Chapter 6.

Active illumination: We use an intensity-modulated light source with pulse shape $s(t)$ and repetition interval T_r seconds. Optically, $s(t)$ is the photon-flux waveform of the laser pulse emitted at $t = 0$ and is measured in counts/sec (cps). We assume $T_r > 2\, z_{\max}/c$, where $z_{\max}$ is the maximum scene depth and c is the speed of light, to avoid distance aliasing. With conventional processing, the pulse width T_p (the root mean square (RMS) pulse duration) governs the achievable depth resolution [80, 103]. As typically done in depth imaging, we assume that $T_p \ll 2\, z_{\max}/c < T_r$.

Background noise: An incandescent lamp pointed toward the detector injects ambient light with flux approximately as strong as the backreflected signal averaged over the scene. Therefore, each photon detection is due to noise with probability approximately 0.5. In Section 5.9, as well as in Section 6.6, we investigate the effect of signal-to-background ratio (SBR) on the performance of the computational imager.

Detection: A SPAD detector provides time-resolved single-photon detections [94], called *clicks*. Its quantum efficiency η is the fraction of photons passing through the pre-detection optical filter that are detected. Each detected photon is time stamped within a time bin of duration Δ, measuring a few picoseconds, that is much shorter than T_p. For mathematical

modeling in this chapter we assume that the exact photon arrival time is available at each pixel. In Chapter 6 (Section 6.3) we include the effect of time bin width, Δ, in our derivations.

Data acquisition: Each spatial location (patch), (x, y), is illuminated with a periodic stream of light pulses until the first photon is detected. We record the first detected photon's arrival time, $t(x, y)$, relative to the most recently transmitted pulse, along with the number of light pulses, $n(x, y)$, that were transmitted prior to the first detection (see Fig. 5-4). Also, we do not employ range gating [17]. The result of the data acquisition process is the dataset $\{t(x, y), n(x, y)\}_{x,y=1}^{M}$ (see Fig. 5-5).

The first-photon data $t(x, y)$ and $n(x, y)$ are outcomes or realizations of the random variables $T(x, y)$ and $N(x, y)$, respectively.

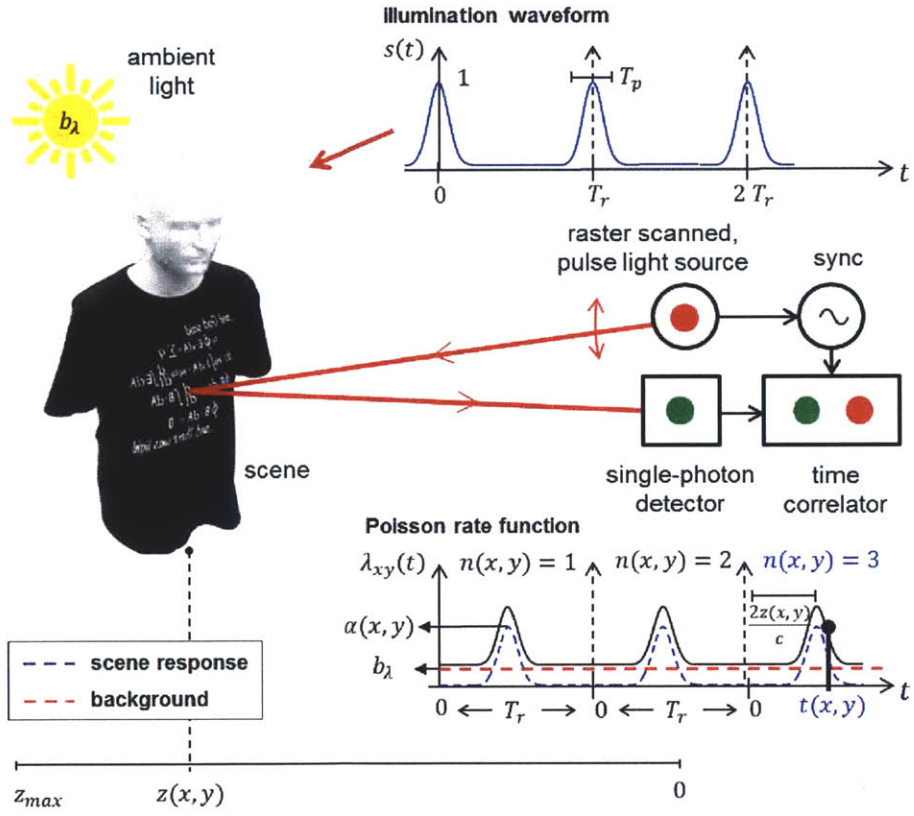


Figure 5 4: **First-photon data acquisition:** Rate function of inhomogeneous Poisson process combining desired scene response and noise sources is shown. The time origin is reset after every pulse. Here, a photon was detected in response to the third illumination pulse ($n(x, y) = 3$) with arrival time $t(x, y)$. Since the photon detections are modeled as arrivals in a merged Poisson process, it is not possible to determine whether the detected photon was generated due to the backreflected signal or background light.

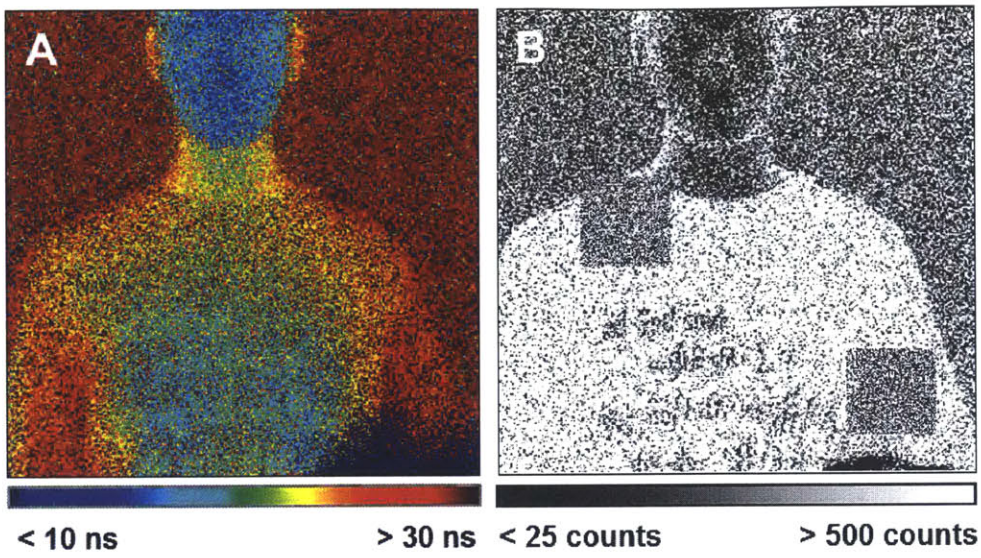


Figure 5 5: **First-photon data:** (A) Photon times of arrival $\{t(x,y)\}$. (B) Elapsed pulse counts $\{n(x,y)\}$ (scene patches with higher reflectivity result in smaller $n(x,y)$ values; see Fig. 5 4 and 5 6(B)). Zooming into a high quality PDF reveals that in the portion of the image corresponding to the back wall (see Fig. 5 2(A)), there are approximately equal regions of black and white pixels in sub figure (B). This is empirically consistent with the fact that averaged over the scene the photon detections due to signal and background light were approximately equal.

Noise sources: Measurement uncertainty results from:

- *Dark counts*: Dark counts are detections that are not due to light incident on the detector. They are generated by inherent detector characteristics.
- *Background light*: Ambient light at the operating wavelength causes photon detections unrelated to the scene.

In addition to the noise sources listed above, certain characteristics of the data acquisition system add uncertainity to the photodetection process. Two such key characteristics are:

- *Pulse width*: The timing of a detected photon could correspond to the leading edge of the pulse, the trailing edge, or anywhere in between. This uncertainty translates to error in depth estimation.
- *Optical loss and detector efficiency*: The light from the illumination source undergoes diffuse scattering at the quasi-Lambertian object. As a result, only a small fraction of

the transmitted photons are backreflected toward the detector. Moreover, not every photon incident on the sensor's active area is detected by the SPAD. Because of these optical losses, a target pixel must be repeatedly illuminated with laser pulses until the first photon is detected. In our work, we show that the number of elapsed pulses prior to first photon detection contain reflectivity information.

Accounting for these characteristics is central to our contribution, as described in the following section.

5.4 Measurement Model

Illuminating scene patch (x, y) with intensity-modulated light pulse $s(t)$ results in a backreflected light signal which is incident at the detector along with the background light. This total photon flux is denoted by $r_{x,y}(t) = \alpha(x, y)\, s(t - 2z(x, y)/c) + b_\lambda$, where b_λ denotes the time-invariant background light flux at the operating wavelength. This is only the incident light-field; the detection model must account for the quantum nature of photodetection.

Poisson photodetection statistics: The photodetections produced by the SPAD in response to the backreflected light from the scene is an inhomogeneous Poisson process whose time-varying rate function is $\eta\, r_{x,y}(t)$, where η is the detection efficiency. It is also necessary to add the detector dark counts, which are modeled as an independent homogeneous Poisson process with rate d. The observed inhomogeneous Poisson process thus has rate (see Fig. 5-4):

$$\lambda_{x,y}(t) = \eta\, r_{x,y}(t) + d = \eta\, \alpha(x, y)\, s(t - 2z(x, y)/c) + (\eta b_\lambda + d). \tag{5.1}$$

Define $S \triangleq \int_0^{T_r} \eta s(t)\, \mathrm{dt}$ (units: counts) as the average counts from perfectly-reflecting pixel; and $B \triangleq (\eta b_\lambda + d)T_r$ (units: counts) as the average background counts per pulse-repetition period, where we have used and will use in all that follows background counts to include dark counts as well as counts arising from ambient light. We assume that both S and B are known, since it is straightforward to measure these physical quantities before we begin data acquisition (see Section 5.7).

Low-rate flux assumption: The derivations that follow assume that the total photon flux incident at the detector from each scene patch is low, i.e., $\alpha(x,y)S + B \ll 1$, as would be the case in low-light imaging, where photon efficiency is important.

Signal vs. noise photons: A detected photon could originate from the backreflected light signal or noise (ambient light and dark counts). The arrival statistics observed at the detector result from the merging of the Poisson processes corresponding to these two sources [104]. Estimates of following probabilities are used in our framework to censor detections that are generated due to background light and dark counts.

$$\Pr[\text{detected photon is due to background light}] = \frac{B}{\alpha(x,y)\,S + B}, \tag{5.2}$$

$$\Pr[\text{detected photon is due to backreflected signal }] = \frac{\alpha(x,y)S}{\alpha(x,y)\,S + B}. \tag{5.3}$$

Model for number of elapsed pulses $N(x,y)$: Novel to this work, we exploit the fact that the reflectivity of patch (x,y) is encoded in the number of pulses, $N(x,y)$, transmitted prior to first photon detection. The probability of *not* detecting a photon when pixel (x,y) is illuminated by a single laser pulse,

$$P_0(x,y) \;=\; e^{-(\,\alpha(x,y)\,S+B)}, \tag{5.4}$$

follows from Equation (5.1) and elementary Poisson statistics. Because each transmitted pulse gives rise to independent Poisson noise, $n(x,y)$ has the geometric distribution with parameter $P_0(x,y)$, i.e.,

$$\Pr[N(x,y) = n(x,y)] \;=\; P_0(x,y)^{n(x,y)-1}\big(1 - P_0(x,y)\big), \quad n(x,y) = 1, 2, \ldots. \tag{5.5}$$

As suggested by this reflectivity model, scene patches with lower reflectivity, $\alpha(x,y)$, on average result in higher $n(x,y)$ values, compared with brighter scene patches (see Fig. 5-5(B)).

Model for photon detection time $T(x,y)$: The distribution of the first detection time for patch (x,y) depends on whether the detection is due to signal or background noise. The time of a photon detection originating from backreflected signal is characterized by the normalized time-shifted pulse shape. The time of a detection due to background noise is uniformly distributed over the pulse repetition period. The conditional probability density functions for the first photon's arrival time $t(x,y)$ given our low optical flux assumption that $\alpha(x,y)S + B \ll 1$, are

$$f_{T(x,y)|\text{signal}}(t(x,y)) = \eta\, s(t(x,y) - 2\,z(x,y)/c)/S, \text{ for } t(x,y) \in [0, T_r), \tag{5.6a}$$

$$f_{T(x,y)|\text{background}}(t(x,y)) = 1/T_r, \quad \text{for } t(x,y) \in [0, T_r). \tag{5.6b}$$

The detailed derivations are included in Appendix A. As a consequence of the low optical flux assumption, Equation (5.6)(a) has no dependence on scene patch reflectivity, $\alpha(x,y)$. Also note that for ease of theoretical derivations, we assume that the transmitted pulse is centered at 0 and that $3T_p < 2z(x,y)/c < z_{\max} < T_r - 3T_p, \forall(x,y)$. Since we assume a zero-centered light pulse, this assumption ensures that the backscattered pulses lies entirely within the repetition intervals as shown in Fig. 5-4. In practice, the time-delay is computed relative to the mode of the measured pulse shape (see Section 5.7).

5.5 Conventional Image Formation

In the limit of large sample size or high signal-to-noise ratio (SNR), ML estimation converges to the true parameter value [102]. However, when the SNR is low such as in our problem pointwise or pixelwise ML processing yields inaccurate estimates. Given the first-photon data, $\{t(x,y), n(x,y)\}_{x,y=1}^{M}$, the pointwise estimates for scene reflectivity and depth based on the models described by Equations (5.5) and (5.6) are:

$$\hat{\alpha}_{geo}^{\text{CML}}(x,y) = \max\left\{\frac{1}{(n(x,y)-1)\,S} - \frac{B}{S}, 0\right\} \qquad \hat{z}(x,y) = \frac{c\,(t(x,y) - T_m)}{2} \tag{5.7}$$

where the reflectivity estimate has been constrained to be non-negative and $T_m = \arg\max_t s(t)$ is the mode of the normalized pulse shape. The detailed derivations for these estimate expressions are included in Appendix A.

In the absence of background light and dark counts, the pointwise ML reflectivity estimate, $\hat{\alpha}_{\mathrm{ML}}(x,y)$, is proportional to $1/n(x,y)$ for $n(x,y) \gg 1$, and the depth estimation error is governed by the pulse duration. However, in the presence of background illumination these pointwise estimates are extremely noisy as shown in Fig. 5-6(A-C) and through derivations in Appendix A.

5.6 Novel Image Formation

In the proposed approach we use the first-photon detection statistics derived in Section 5.4 along with the spatial correlation present in real-world objects, in order to construct high quality reflectivity and depth images. We accomplish this through regularized ML estimation incorporating regularization by transform-domain sparsity within the ML framework. Our computational reconstruction proceeds in three steps.

Step 1: Computational Reconstruction of Object Reflectivity

As seen from Equation (5.5), the negative of the log-likelihood function, $\mathcal{L}_\alpha(\alpha(x,y); n(x,y)) = -\log \Pr[N(x,y) = n(x,y)]$, relating the pulse count data, $n(x,y)$, to the reflectivity parameter, $\alpha(x,y)$, at each transverse location is

$$\mathcal{L}_\alpha(\alpha(x,y); n(x,y)) = [\alpha(x,y)\, S + B]\, [n(x,y) - 1] - \log[(\alpha(x,y)S + B)].$$

This likelihood is a strictly convex function of $\alpha(x,y)$ (see Appendix A for derivation).

Spatial correlations present in the object reflectivity are captured by wavelet-domain sparsity. For this work we used the discrete wavelet transform (DWT) derived from Daubechies's 4-tap filters [39]. This transform $\Phi(\cdot)$ is implemented as a matrix multiplication. Let $\Phi\left(\{\alpha(x,y)\}_{x,y=1}^{M}\right) = \Phi(\boldsymbol{\alpha})$ denote the collection of DWT coefficients, $\{w_\alpha(x,y)\}$, of the reflectivity image. A standard measure of sparsity used in image processing is the sum of

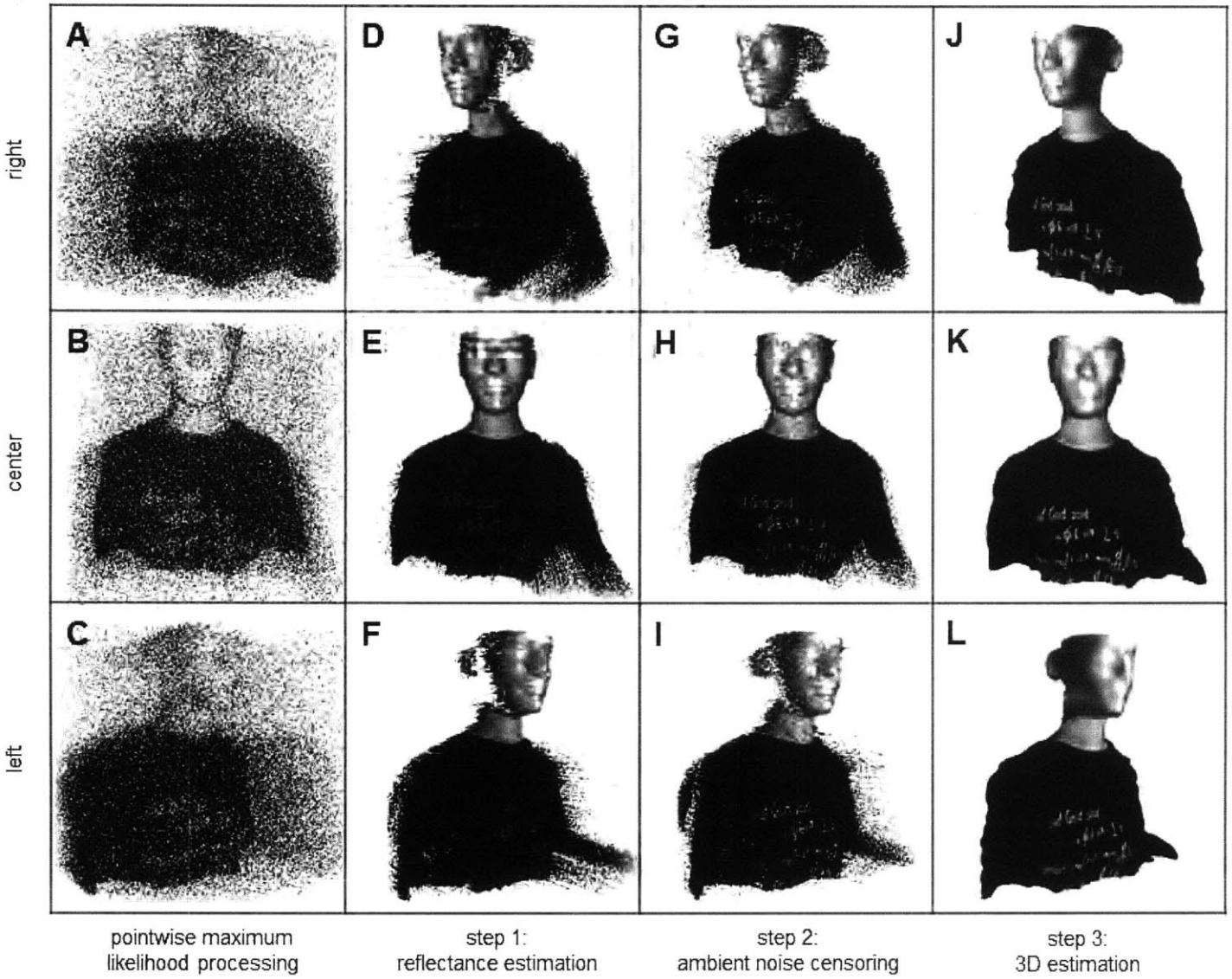


Figure 5 6: **Depth and reflectivity reconstruction:** First photon data (Fig. 5 5) is used to compute depth estimates that are rendered here as point clouds and overlaid with reflectivity estimates.

absolute values of DWT coefficients, denoted here using the ℓ_1-norm:

$$\|\Phi(\boldsymbol{\alpha})\|_1 = \sum_{x=1}^{M}\sum_{y=1}^{M}|w_\alpha(x,y)|. \tag{5.8}$$

It is well known that $\|\Phi(\boldsymbol{\alpha})\|_1$ is a strictly convex function of the reflectivity image $\boldsymbol{\alpha}$ [105].

The first-photon imaging reflectivity estimate, denoted by $\hat{\boldsymbol{\alpha}}_{\text{fpi}}$, is computed by minimizing over the set of all possible images the sum of negative log-likelihood functions over all transverse locations and the sparsity measuring function. A weight parameter, $\beta_\alpha \in [0.1, 0.9]$, needs to be introduced to trade-off between the likelihood and sparsity terms in the objective function. For a fixed value of β_α, the reflectivity reconstruction problem is the following convex optimization program:

$$\begin{aligned} \underset{\boldsymbol{\alpha}}{\text{minimize}} \quad & (1-\beta_\alpha)\sum_{x=1}^{M}\sum_{y=1}^{M}\mathcal{L}_\alpha(\alpha(x,y);n(x,y)) + \beta_\alpha\,\|\Phi(\boldsymbol{\alpha})\|_1. \qquad \text{(OPT-1)} \\ \text{subject to} \quad & \alpha(x,y) \geq 0, \quad \text{for all } x,y. \end{aligned}$$

OPT-1 is a strictly convex optimization program, because it is the nonnegative weighted sum of individually convex functions. It is solved using standard numerical methods described in [74] and the initial solution point for OPT-1 is chosen to be the pointwise constrained ML estimate (from Equation (5.7)),

$$\hat{\alpha}_{geo}^{\text{CML}}(x,y) = \max\left\{\frac{1}{(n(x,y)-1)\,S} - \frac{B}{S},\, 0\right\}, \tag{5.9}$$

where the constraint $\alpha(x,y) \geq 0$ captures the fact that reflectivity is nonnegative. Each β_α value produces a candidate reflectivity image. We selected a β_α value by solving the optimization problem OPT-1 for $\beta_\alpha = \{0.1, 0.2, \ldots, 0.9\}$ and then choosing the one that minimized the objective function defined in OPT-1.

As seen in Fig. 5-6(D-F), step 1 processing yields a considerable improvement in reflectivity estimation over pointwise ML estimates (Fig. 5-6(A-C)). As discussed in Section 5.8, our proposed reflectivity reconstruction method shows significant improvement over state-of-the-art denoising methods as well.

The pointwise range estimate from a first-photon detection is $\hat{z}(x,y) = c\,(t(x,y) - T_m)/2$ for a transmitted pulse whose peak is at time $t = T_m$, where c is the speed of light. Its root mean square (RMS) estimation error is $(c/2)\sqrt{(T_p^2 + T_r^2/12)/2}$ in the presence of background light, where $T_p \ll T_r$ is the laser pulse's RMS time duration (see Appendix A for complete derivation). Direct application of a spatial-correlation regularization to maximizing time-of-arrival likelihoods is infeasible, because background light makes the optimization objective function multimodal. Background light also causes pointwise methods to fail (see Fig. 5-6(A-C)), so we use spatial correlations to censor $t(x,y)$ values that are range anomalies. This censoring constitutes step 2 of our proposed computational imager.

Step 2: Background Noise Censoring

Anomalous detections have arrival times that are uniformly distributed over the time interval $[0, T_r]$ and mutually independent over spatial locations, so that they have high variance $(T_r^2/12)$ relative to that of signal (backreflected laser pulse) detections, which are temporally concentrated and spatially correlated. As shown in Appendix A, under the low optical flux assumption, i.e., when $(\alpha(x,y)S + B) \ll 1$, *if* the first photon detected from location (x,y) is a signal photon, then the probability density function for $T(x,y)$ is:

$$f_{T(x,y)}(t(x,y)) = \eta\, s(t(x,y) - 2z(x,y)/c) \,/\, S, \qquad 0 \le t(x,y) < T_r, \tag{5.10}$$

which has variance $T_p^2 \ll T_r^2/12$, regardless of the reflectivity at (x,y). The high degree of spatial correlation in the scene's 3D structure then implies that signal-generated photon arrival times have much smaller conditional variance, given data from neighboring locations, than do anomalous detections.

Step 2 of the computational imager uses this statistical separation to censor background-generated arrival times from an initial pixel-by-pixel range image. At each pixel location, (x,y), the rank-ordered absolute difference (ROAD) statistic [106, 107] is first computed using the time-of-arrival measurements of the eight nearest transverse neighbors, denoted by

$(x_1, y_1), \ldots, (x_8, y_8)$. Except at the boundaries, these neighbors are

$$(x-1, y-1), (x-1, y), (x-1, y+1), (x, y-1), (x, y+1), (x+1, y-1), (x+1, y), (x+1, y+1).$$

The eight absolute time-of-arrival differences

$$|t(x_1, y_1) - t(x, y)|, \ldots, |t(x_8, y_8) - t(x, y)|$$

are sorted in ascending order, and the ROAD statistic, ROAD(x, y), is the sum of the first four absolute differences from this sorted collection.

Then, a binary hypothesis test is applied to classify the photon detection at (x, y) as signal or background noise. To apply this test, we require an accurate reflectivity estimate, $\hat{\boldsymbol{\alpha}}_{\text{fpi}} = \{\hat{\alpha}_{\text{fpi}}(x, y)\}_{x,y=1}^{M}$, which we obtained in step 1. The probabilities defined in Equation (5.3) are estimated, using $\hat{\boldsymbol{\alpha}}_{\text{fpi}}$, via

$$\Pr[\text{detected photon is due to background light}] = \frac{B}{\hat{\alpha}_{\text{fpi}}(x, y)\, S + B},$$

$$\Pr[\text{detected photon is due to backreflected signal }] = \frac{\hat{\alpha}_{\text{fpi}}(x, y)\, S}{\hat{\alpha}_{\text{fpi}}(x, y)\, S + B}.$$

These estimated probabilities are used to generate thresholds for the following binary hypothesis test based on the computed ROAD statistic:

if $\text{ROAD}(x, y) \geq 4T_p \dfrac{B}{\hat{\alpha}_{\text{fpi}}(x, y)\, S + B}$, **then** the detected photon is censored;

if $\text{ROAD}(x, y) < 4T_p \dfrac{B}{\hat{\alpha}_{\text{fpi}}(x, y)\, S + B}$, **then** the detected photon is *not* censored.

Once background detections have been rejected, depth map estimation using photon arrival times, $\{t(x, y)\}_{x,y=1}^{M}$ becomes tractable. Our final processing step is thus to compute the regularized ML depth map estimate by maximizing the product of data likelihoods (Equation (5.10)), over the uncensored spatial locations, combined with a DWT-based penalty function that exploits spatial correlations present in a scene's 3D structure. This is described in the next section.

Step 3: Computational Reconstruction of 3D structure:

From Equation (5.6) and the knowledge fitted pulse shape (see Section 5.7), the negative of the log-likelihood function relating the signal photon's arrival time, $t(x,y)$, to the distance, $z(x,y)$, at each *uncensored* spatial location is

$$\mathcal{L}_z(z(x,y);t(x,y)) = -\log\left[s\left(t(x,y) - \frac{2\,z(x,y)}{c}\right)\right].$$

Our framework allows the use of arbitrary pulse shapes, but many practical pulse shapes (including Gaussian and lopsided pulses) are well approximated as $s(t) \propto \exp[-v(t)]$, where $v(t)$ is a convex function in t. In such cases, $\mathcal{L}_z(z(x,y);t(x,y)) = v(t(x,y) - 2z(x,y)/c)$ is a convex function in $z(x,y)$.

As was done for reflectivity estimation in step 1, spatial correlations present in the depth map are captured by wavelet-domain sparsity. The first-photon imaging depth map estimate, denoted by $\hat{\boldsymbol{z}}_{\text{fpi}}$ is computed by minimizing over the set of all possible depth maps the sum of negative likelihood functions over the uncensored spatial locations and the sparsity measuring function. As before, a weight parameter, $\beta_z \in [0.1, 0.9]$, needs to be introduced. For a fixed β_z, the 3D reconstruction problem is the following optimization program:

$$\begin{aligned} \underset{\boldsymbol{z}}{\text{minimize}} \quad & (1-\beta_z) \sum_{\text{uncensored }(x,y)} \sum \mathcal{L}_z(z(x,y);t(x,y)) + \beta_z \left\|\Phi(\boldsymbol{z})\right\|_1 \qquad \text{(OPT-2)} \\ \text{subject to} \quad & 0 \le z(x,y) \le z_{\max}, \quad \text{for all } x,y. \end{aligned}$$

The constraint $0 \le z(x,y) \le z_{\max}$ captures the fact that distance is always nonnegative, and within the scope of our problem formulation it is bounded as well. The starting points for optimization problem OPT-2 is chosen to be the pointwise estimate,

$$\begin{aligned} \hat{z}(x,y) &= \frac{c\,(t(x,y) - T_m)}{2}, \text{ for uncensored spatial locations} \\ \hat{z}(x,y) &= \quad 0, \quad \text{for censored spatial locations,} \end{aligned}$$

In our case, the signal photons' detection times serve as excellent starting points for numerical

optimization because of their proximity to the true depth. Also note that $\Phi(\boldsymbol{z})$ is a function of the entire depth map, i.e., depth values are also assigned to (x, y) locations whose arrival times were censored because they were predicted to originate from background light. As was done for the reflectivity estimates, we solve the optimization problem OPT-2 for $\beta_z = \{0.1, 0.2, \ldots, 0.9\}$, and the final depth map construction is chosen to be the one whose β_z value minimized the objective function in OPT-2.

Baseline Estimates using Many Photons per Pixel:

Given a large number, N_{baseline}, of photon detections at each pixel in the absence of background noise, yielding dataset $\{t^\ell(x,y), n^\ell(x,y)\}_{\ell=1}^{N_{\text{baseline}}}$, the pointwise ML estimates for scene depth and reflectivity are

$$z_{\text{baseline}}(x,y) = \underset{z\in[0,z_{\max}]}{\arg\max} \sum_{\ell=1}^{N_{\text{baseline}}} \log s\left(t^\ell(x,y) - \frac{2z}{c}\right) \tag{5.11}$$

$$\alpha_{\text{baseline}}(x,y) = \frac{N_{\text{baseline}}/S}{\sum_{\ell=1}^{N_{\text{baseline}}} n^\ell(x,y) - N_{\text{baseline}}}. \tag{5.12}$$

Detailed derivations of these expressions are available in Section 6.4 in the context of sensor array imaging. It is well known that when N_{baseline} is large, these estimates converge to the true parameter values [102]. We used this method to generate *ground truth* data for comparison. The baseline reflectivity image and depth map are denoted using $\boldsymbol{\alpha}_{\text{baseline}} = \{\alpha_{\text{baseline}}(x,y)\}_{x,y=1}^{M}$ and $\boldsymbol{z}_{\text{baseline}} = \{z_{\text{baseline}}(x,y)\}_{x,y=1}^{M}$ respectively.

5.7 Experimental Setup

Our experimental setup follows Fig. 5-1 and is shown in Fig. 5-7.

Equipment details: The active illumination source was a 640 nm wavelength, 0.6 mW average power, pulsed laser diode that produced $T_p = 226$ ps RMS duration pulses at a 10 MHz repetition rate. An incandescent lamp illuminated the detector to create background

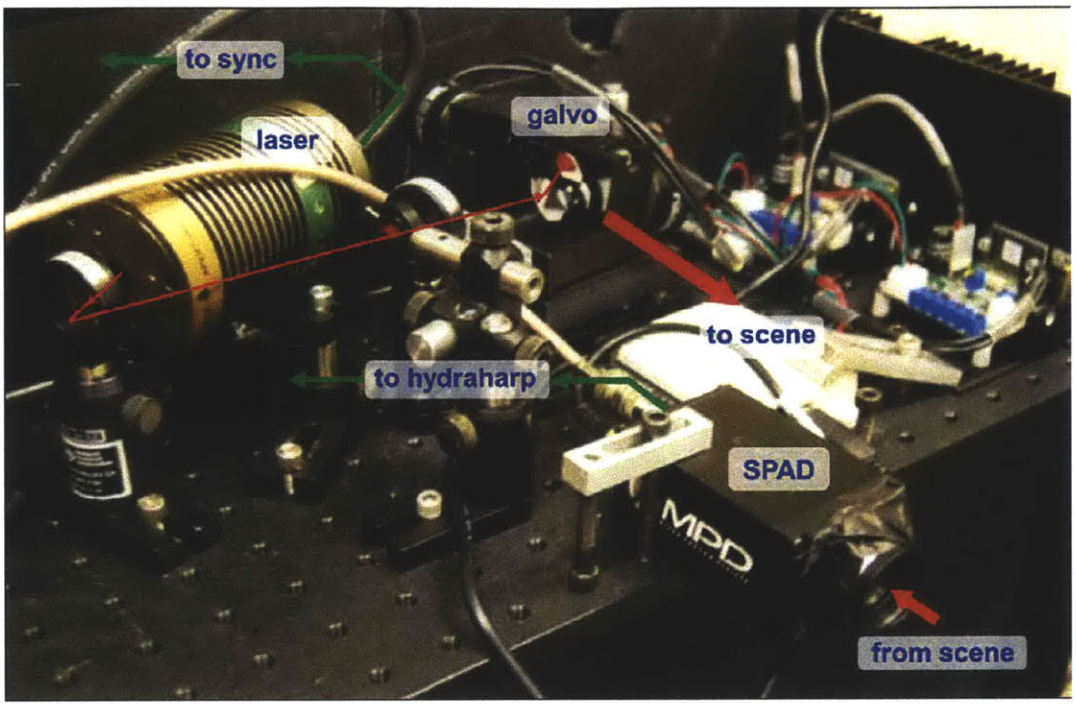


Figure 5 7: **Experimental setup:** Physical setup indicating the optical paths.

noise from extraneous photon detections. A two-axis scanning galvo (maximum scan angle $\pm 20°$) was used to raster scan a room-scale scene consisting of life-size objects. The laser spot size at 2 m distance was measured to be 1.5 mm. Prior to detection, the light was filtered using a 2 nm bandwidth free-space interference filter centered at 640 nm wavelength whose peak transmission was 49%. The Geiger-mode APD was a Micro Photon Devices PDM series detector with 100 μm $\times$ 100 μm active area, 35% quantum efficiency, less than 50 ps timing jitter, and less than 2×10^4 dark counts per second. The photon detection events were time stamped relative to the laser pulse with 8 ps resolution using a PicoQuant HydraHarp TCSPC module. The objects to be imaged were placed between 1.25 m to 1.75 m distance from the optical setup. The laser and the single-photon detector were placed in the same horizontal plane, at a separation of 7 cm, making our imaging setup effectively monostatic.

Radiometric calibration: The detection efficiency is the product of the interference filter's transmission and the detector's quantum efficiency, $\eta = 0.49 \times 0.35 = 0.17$. A reference calibration for S, the average photon number in the backreflected signal from a unity reflectiv-

ity pixel received from a single laser pulse, was obtained as follows. All sources of background light were turned off, and the laser was used to illuminate a transverse location $(x_{\text{ref}}, y_{\text{ref}})$ on a reflective Lambertian surface at a distance of 2 m. In our experiment, this reference target was a point on the white wall behind our scene setup (see Fig. 5-2)(A)). The average number of transmitted pulses before a photon detection was found to be $\langle n(x_{\text{ref}}, y_{\text{ref}})\rangle = 65$. Using Equation (5.4), with $\alpha(x_{\text{ref}}, y_{\text{ref}}) = 1$ and $B = 0$, we find

$$\langle n(x_{\text{ref}}, y_{\text{ref}})\rangle = \frac{1}{1 - P_0(x_{\text{ref}}, y_{\text{ref}})} = \frac{1}{1 - \exp(-S)},$$

from which $\langle n(x_{\text{ref}}, y_{\text{ref}})\rangle = 65$ results in $S = \int_0^{T_r} \eta s(t)\text{dt} = 0.09$.

For adjusting background illumination power, the laser was first turned off and all objects were removed from the scene. Then the incandescent lamp's optical power was adjusted such that the average number of background photons reaching the detector in a pulse repetition period was $B = 0.1 \approx S$.

We used a reference point on the white wall in the background to set the signal-to-background ratio, $S/B \approx 0.5$, in a scene-independent manner. Upon data collection, however, we also noticed that for the mannequin dataset, $B \approx \bar{\alpha}S$, where $\bar{\alpha}$ was the average scene reflectivity. This relationship was partly due to the fact that the reference white wall was at a 2 m distance from the imaging setup and radial fall-off played a significant role in weakening the backreflected signal incident at the photodiode. In contrast, the mannequin was set closer at approximately 1.25 m and contained both high and low reflectivity surfaces, averaging out the overall scene reflectivity to a moderate value.

Transverse optical calibration: A geometric calibration is necessary to relate the reconstructed depth map to the actual 3D scene patches. This was accomplished by first capturing high-quality images (using the baseline methods described above) of three different views of a planar checkerboard. Then a standard computer vision package[1] was used to compute the optical center coordinates, (c_x, c_y), and the perspective projection matrix containing the effective focal lengths, f_x and f_y. Also, at each pixel location (x, y), the estimate $z(x, y)$ is a

[1] OpenCV: Open Source Computer Vision Library. `http://opencv.org/`

measurement of the radial distance to the object from the optical center. For visualization with a 3D graphics package, we convert the depth map data, $[x, y, z(x, y)]$, to a 3D point cloud comprising the 3D real-world Cartesian coordinates of the scene patches, $[\mathrm{x}, \mathrm{y}, z(x, y)]$. This type of coordinate transformation is standard practice in computer vision, but we describe it next for the sake of completeness.

Extraction of 3D scene geometry from depth map data: The perspective projection camera model [91] relates the desired 3D point cloud data to the constructed 2D depth map by the following linear transformation expressed in homogeneous coordinates:

$$z(x,y)\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} f_x & 0 & c_x & 0 \\ 0 & f_y & c_y & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} \mathrm{x} \\ \mathrm{y} \\ z(x,y) \\ 1 \end{pmatrix} = \begin{pmatrix} f_x\,\mathrm{x} + c_x z(x,y) \\ f_y\,\mathrm{y} + c_y z(x,y) \\ z(x,y) \end{pmatrix}$$

The left hand side of this equation is obtained from the depth map data, and the intrinsic camera parameters, (c_x, c_y, f_x, f_y), are obtained from transverse optical calibration. Thus, the unknown 3D point cloud coordinates are readily computed by comparing the corresponding matrix entries, i.e.,

$$\mathrm{x} = \frac{(x - c_x)\, z(x,y)}{f_x}, \quad \text{and} \quad \mathrm{y} = \frac{(y - c_y)\, z(x,y)}{f_y}$$

First-photon data acquisition: To generate one complete data set, we raster scan over $(M = 1000) \times (M = 1000)$ pixels with the two-axis galvo. For transverse location (x, y), only two values are used for constructing first-photon images: $n(x, y)$, the number of laser pulses transmitted prior to the first detection event; and $t(x, y)$ the timing of the detection relative to the pulse that immediately preceded it. (Because our entire scene was contained within a few meters of the imaging setup, our 100 ns pulse repetition period guaranteed that each non-background photon detection came from the immediately preceding laser pulse.) We collected data to generate our reference images by averaging over $N_0 = 1000$ photon detections at every pixel, the first-photon imager reads only the first detection recorded

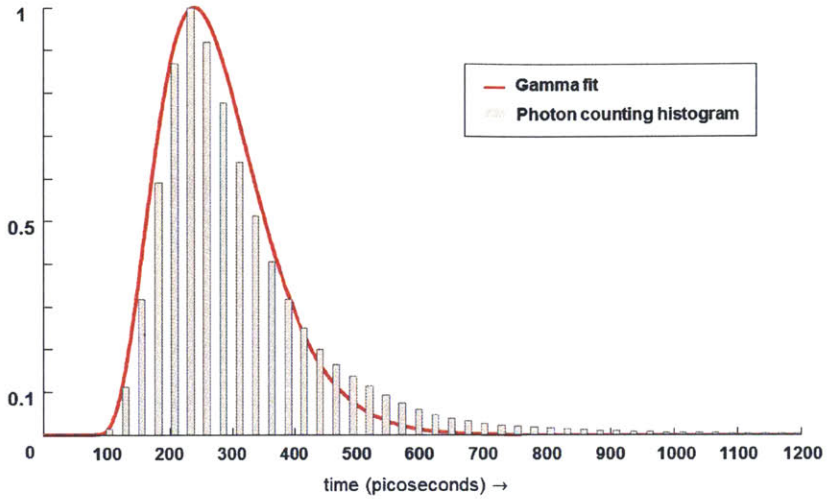


Figure 5 8: **Pulse shape measurement.** Plot of photon count histogram overlaid with the fitted pulse shape. The histogram and the fitted pulse shape have been normalized to have unity maximum values.

at every pixel, ignoring the rest of the data in the file. If we did not need to acquire a reference image for comparison, it would be possible to acquire a megapixel first-photon scan in approximately 20 minutes using our current equipment. This time estimate was limited only by the mechanical speed of our two-axis galvo; our photon flux is high enough to permit much faster acquisition.

Photon-flux waveform measurement: For range estimation, our computational imager requires knowledge of the laser pulse's photon-flux waveform, $s(t)$, launched at $t = 0$. This pulse shape was measured by directly illuminating the detector with highly attenuated laser pulses and binning the photon detection times to generate a histogram of photon counts. Fitting a skewed Gamma function to this histogram yielded

$$s(t) \propto (t - T_\gamma)^4 \exp\left(-\frac{(t - T_\gamma)}{T_c}\right), \tag{5.13}$$

where $T_\gamma = 80$ ps and $T_c = 40$ ps (see Fig. 5-8). The RMS pulse duration is defined as

$$T_p = \sqrt{\frac{\int_0^{T_r} (t - \bar{T})^2 \, s(t) \, \mathrm{d}t}{\int_0^{T_r} s(t) \, \mathrm{d}t}},$$

where

$$\bar{T} = \frac{\int_0^{T_r} t\, s(t)\, \mathrm{d}t}{\int_0^{T_r} s(t)\, \mathrm{d}t},$$

and T_r is the pulse repetition period. Note that the fitted pulse shape, $s(t)$, is not zero-centered. This implies that the estimated depth value at each pixel was uniformly offset by approximately 6.3 cm corresponding to $\bar{T} \approx 210$ ps. This systematic bias uniformly affects both the baseline and first-photon depth estimation algorithms. Thus, the stated depth resolution results and depth reconstruction error values remain unchanged.

Numerical details and image formation time: For a one-megapixel reflectivity and 3D reconstruction, the total run time required by the 9 instances of OPT-1 followed by ROAD filtering and 9 instances of OPT-2 was less than 3 minutes for each of the processed datasets. The convex optimization solver reached the desired precision and terminated after 4 or 5 iterations. The computation was carried out on a standard desktop computer with 4 GB memory and Intel Core 2 Duo processor (2.7 GHz).

5.8 Experimental Results

After calibration, we tested our first-photon computational imager on scenes with both geometric and reflectivity complexity. Experiments were conducted with real-world scenes as well as with resolution charts to evaluate 3D and reflectivity imaging using our framework.

For comparison, we also processed the corrupted pointwise ML estimates (Equation (5.7)) with a state-of-the-art nonlinear denoising method that uses spatial correlations to mitigate Poisson noise called BM3D with Anscombe transform [108]. We also investigated the performance of a well known and standard image filtering method called median filtering to mitigate impulsive noise.

For studying the reconstruction error, we compared the reconstructed reflectivity images and depth maps with the ground truth (baseline reconstruction) using absolute difference images. We quantified the performance of reflectivity estimation, $\hat{\boldsymbol{\alpha}}_{\text{fpi}} = \{\hat{\alpha}_{\text{fpi}}(x,y)\}_{x,y=1}^{M}$, relative to the baseline, $\boldsymbol{\alpha}_{\text{baseline}} = \{\alpha_{\text{baseline}}(x,y)\}_{x,y=1}^{M}$, (see Equation (5.12)) using peak

signal-to-noise ratio (PSNR):

$$\mathrm{PSNR}(\boldsymbol{\alpha}_{\mathrm{baseline}}, \hat{\boldsymbol{\alpha}}_{\mathrm{fpi}}) = 10\log_{10}\left(\frac{\left(\max_{(x,y)}\ \{\alpha_{\mathrm{baseline}}(x,y)\}\right)^2}{\sum_{x=1}^{M}\sum_{y=1}^{M}\{\alpha_{\mathrm{baseline}}(x,y) - \hat{\alpha}_{\mathrm{fpi}}(x,y)\}^2/M^2}\right). \tag{5.14}$$

We quantified the performance of a depth map estimation, $\hat{\boldsymbol{z}}_{\mathrm{fpi}}$, relative to the baseline depth map, $\boldsymbol{z}_{\mathrm{baseline}}$ using root mean-square error (RMSE):

$$\mathrm{RMSE}(\boldsymbol{z}_{\mathrm{baseline}}, \hat{\boldsymbol{z}}_{\mathrm{fpi}}) = \sqrt{\frac{1}{M^2}\sum_{x=1}^{M}\sum_{y=1}^{M}\{z_{\mathrm{baseline}}(x,y) - \hat{z}_{\mathrm{fpi}}(x,y)\}^2}. \tag{5.15}$$

The first target object that we investigated was a life-sized mannequin with white polystyrene head and torso donned with a black cotton shirt imprinted with white text (see Fig. 5-4). The approximate height × width × depth dimensions of the head were 20 cm × 16.5 cm × 24 cm, while those of the torso were 102 cm × 42 cm × 29 cm. The mannequin was placed at a range of 1.25 m from the imaging setup. Using the first-photon data, we estimated the object reflectivity and 3D spatial form as described above. A standard graphics package was then used to visualize the object's 3D profile, overlaid with reflectivity data, after each processing step (see Fig. 5-6).

Our computational first-photon imager recovers a great deal of object detail. Figure 5-6(J-L) and Fig. 5-9 show recovery of reflectivity information including text, facial features, surface reflectivity variations, and specular highlights that are heavily obscured in pointwise maximum-likelihood estimates. As shown in Fig. 5-6(G-I), background-detection censoring affords significant improvement in range estimation, so that the reconstructed 3D form reveals fine structural features such as the nose, eyes, and lips (Fig. 5-10(A-C)).

To quantify the reconstruction error of our approach, we compared the 3D reconstruction of the mannequin head with a 3D image captured using our imaging setup operating as a direct detection LIDAR. For this reference capture, background light was first reduced to

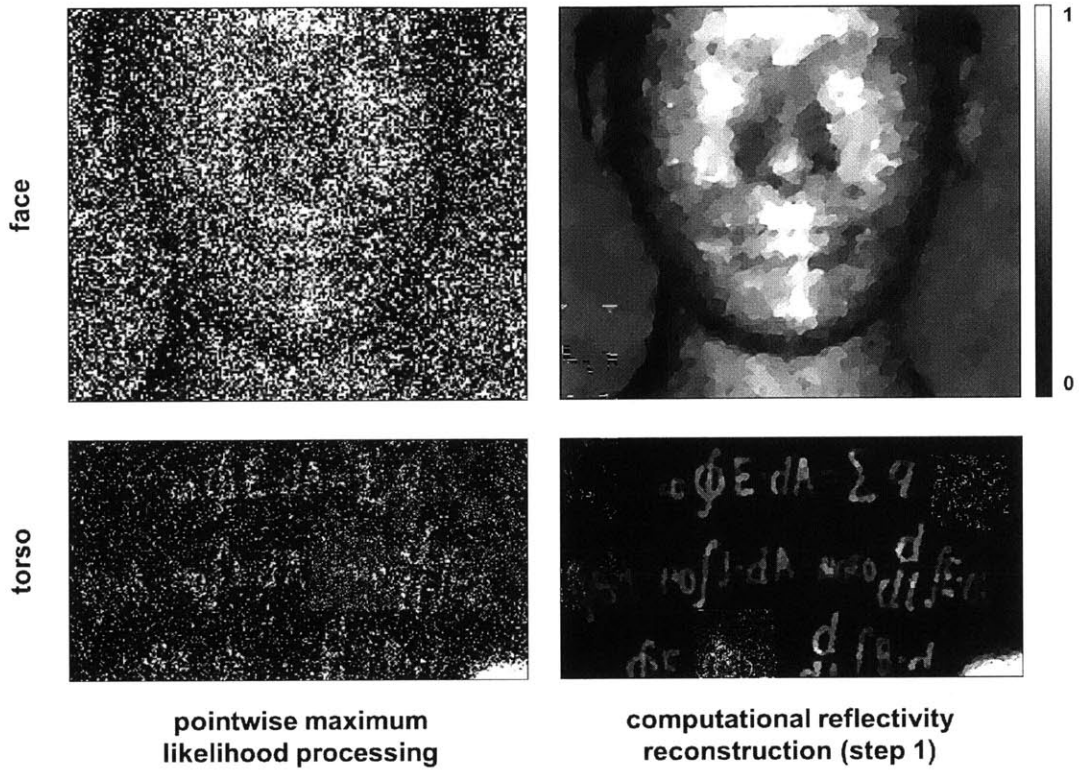


Figure 5 9: **Reflectivity reconstruction from first-photon detections:** The scale quantifies reflectivity relative to that of a high reflectivity calibration point, $\alpha(x_{\text{ref}}, y_{\text{ref}})$, on a 0 to 1 scale.

negligible levels, after which ~1000 photon detections were recorded at each spatial location. Pointwise 3D estimates were then obtained using baseline methods (see Equation (5.12)). This data-intensive baseline technique allows sub-mm accuracy 3D reconstruction for our $T_p = 226$ ps value [16].

Figure 5-10(D-F) shows superimposed facial profiles from the two methods. Both 3D forms were measured using the same imaging setup, obviating the need for registration or scaling. The root mean square (RMS) error of our computational imager was slightly lower than 3.5 mm, with higher values near the edge of the mannequin and around the sides of the nose and the face. These locations have surface normals that are nearly perpendicular to the line of sight, which dramatically reduces their backreflected signal strength relative to background light. Consequently, they incur more anomalous detections than do the rest of the pixels. Although our method censors range anomalies near edges, it estimates the missing ranges using spatial correlations, leading to loss of subtle range details at these edges.

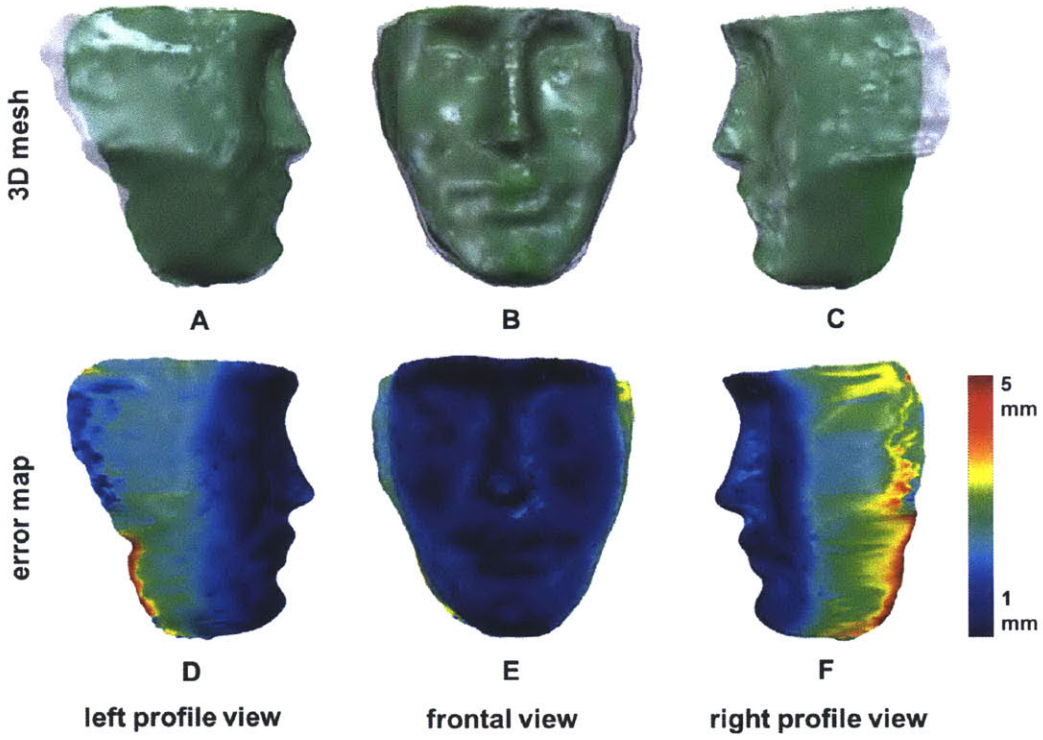


Figure 5 10: **Comparison between computational first-photon 3D imager and LIDAR system:** Rendered views of the facial surfaces reconstructed with computational imaging (gray) and 3D LIDAR (green) are shown in frontal (B) and lateral (A) and (C) profiles. Color coded absolute pointwise differences between the two surfaces, overlaid on the LIDAR reconstruction, are shown in (D) to (F).

Depth Resolution Test:

To test depth resolution, we created a resolution chart with 5 cm $\times$ 5 cm square plates mounted on a flat board to present our imager with targets of varying heights from that board (see Fig. 5-11). This test target was placed at a distance of 3 m and first-photon data was acquired. The depth map was then constructed using the computational imager. For comparison, the pointwise ML 3D estimate based on first-photon arrivals was computed, and the photon-count histogram method using 115 photons at each pixel was also implemented.

The smallest resolvable plate height is an indicator of achievable depth resolution. As shown in Fig. 5-12, our method achieves a depth resolution slightly better than 4 mm using only first-photon detections. In contrast, the pointwise maximum-likelihood 3D estimates are extremely noisy, and the pointwise photon-count histogram method requires 115 photons

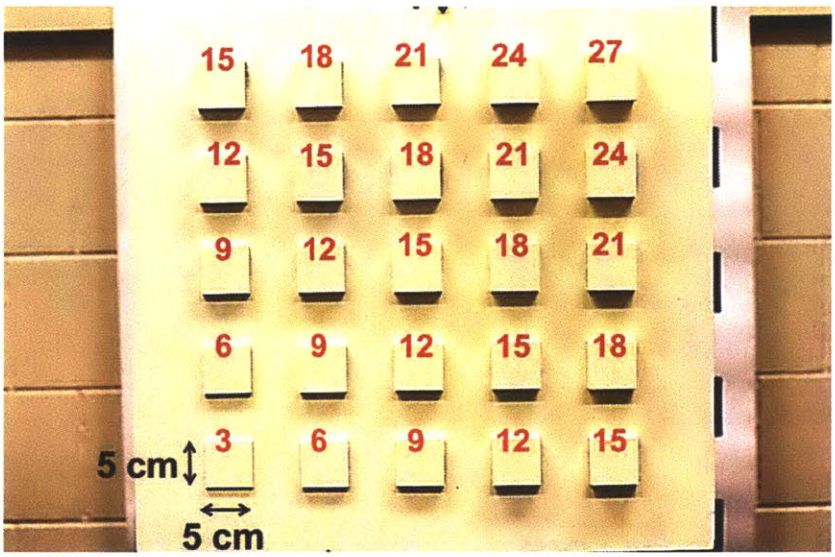


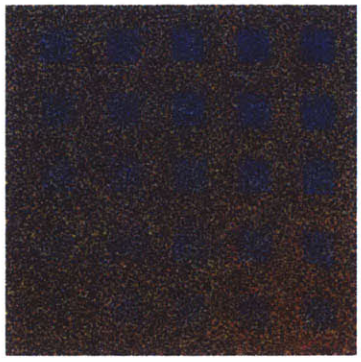
Figure 5 11: **Photograph of the depth-resolution test target.** The height of the squares (in millimeters) relative to the flat surface is overlaid in red in this image. Each square has dimensions 5 cm × 5 cm (marked in black).

detection at each transverse location to achieve ∼4 mm depth resolution under identical imaging conditions.

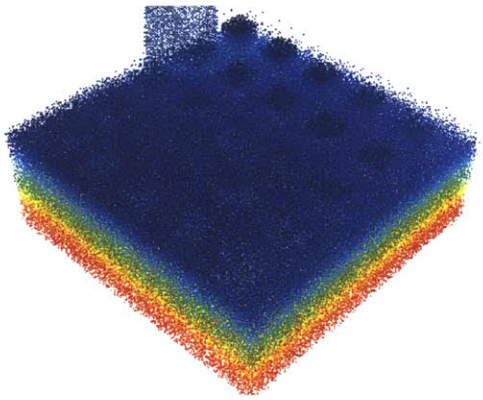
Given a single signal-photon detection time, the RMS pulse duration of 226 ps corresponds to $cT_p/2 = 34$ mm depth uncertainty. At the background level in our experiment, the theoretical value of RMS error in the first-photon pointwise estimate is equal to $(c/2)\sqrt{(T_p^2 + T_r^2/12)/2} = 3.06$ m (see Appendix A for derivation). In comparison, the 4 mm depth resolution achieved by our computational imager is 8.5 times smaller than the RMS pulse duration and 765 times smaller than the RMS depth error of the first-photon pointwise estimate.

Statistical analysis of depth resolution test

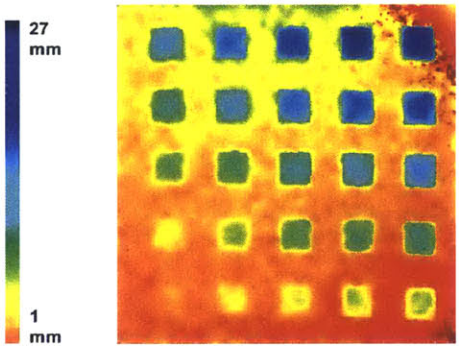
We used a histogram-based analysis to quantify the achievable depth resolution, which in our test corresponds to the the height of the square plate of least height that is distinguishable from the flat board. To quantify this distinguishability, we plot a histogram of the reconstructed depth values in the neighborhood of an edge of the square plates (see Figs. 5-13 and 5-14).



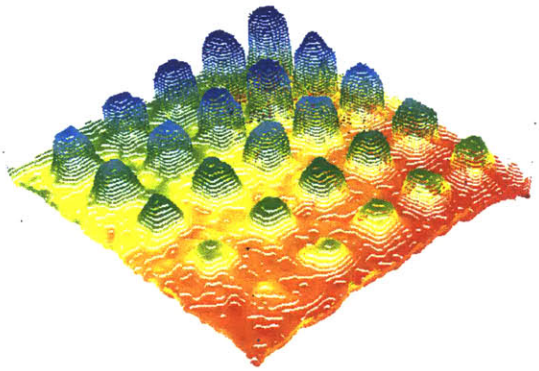
pointwise ML estimate



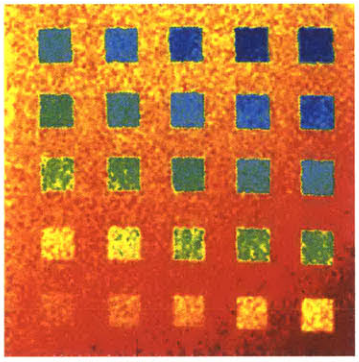
side view



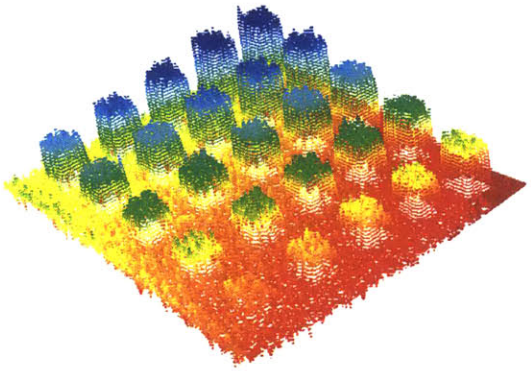
our computational imaging



side view



pointwise photon-count histogram
method using 115 detections/pixel



side view

Figure 5 12: **Depth resolution test.** The color bar indicates the height of the squares above the flat surface. The first photon pointwise ML estimate is noisy. With our computational imager and the photon count histogram method, the square in the bottom left (3 mm above the surface) is barely visible, but the other squares are clearly distinguishable. Histogram analysis in Fig. 5 13 indicates ~ 4 mm depth resolution for our computational imager.

As demonstrated in Fig. 5-13, the depth histograms generated using the proposed computational imager have two clearly distinguishable components in the edge region B: one corresponding to the flat board and the other centered around the height of the square (6 mm). In edge regions A and D, both of which correspond to squares of height 6 mm, the components are also readily distinguishable although some overlap exists. Finally, in the edge region C corresponding to a square of height 2 mm, it is not possible to resolve the different depth components and identify a depth edge.

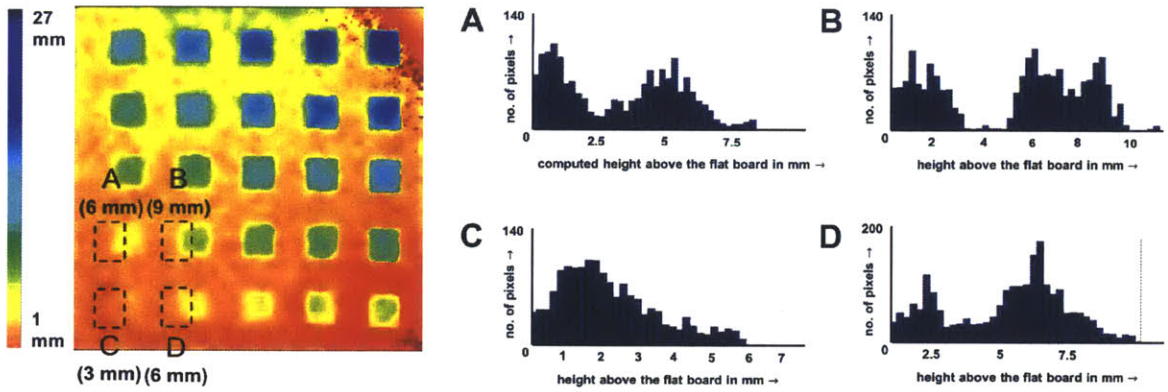


Figure 5 13: **Statistical analysis of depth resolution test for First-photon imaging method.** To quantify the achievable 4 mm depth resolution of our computational imager we histogram the reconstructed pixel heights in the neighborhoods of depth edges (A D). The height of a pixel is the difference of its reconstructed depth value and the depth of the bottom left most corner of the flat board. The difference in the mean values of clearly separated histogram components in the edge regions A and D is ~ 4 mm, which is also the depth resolution of the first photon imager.

Similar histogram analysis was conducted for the pointwise photon-counting histogram method which uses 115 photon detections/pixel. As demonstrated in Fig. 5-14, the histogram-based analysis yields similar resolvability of depth components when compared with our proposed first-photon imager. However, the variances of the resolvable histogram components are smaller than those of the first-photon computational imager.

Reflectivity Resolution Test

To test the reflectivity resolution achievable with our computational imager, we printed a linear gray-scale chart on matte paper. (see Fig. 5-15). This test target was placed at a range of 2 m and a reference image was acquired with a 1 msec dwell time at each pixel.

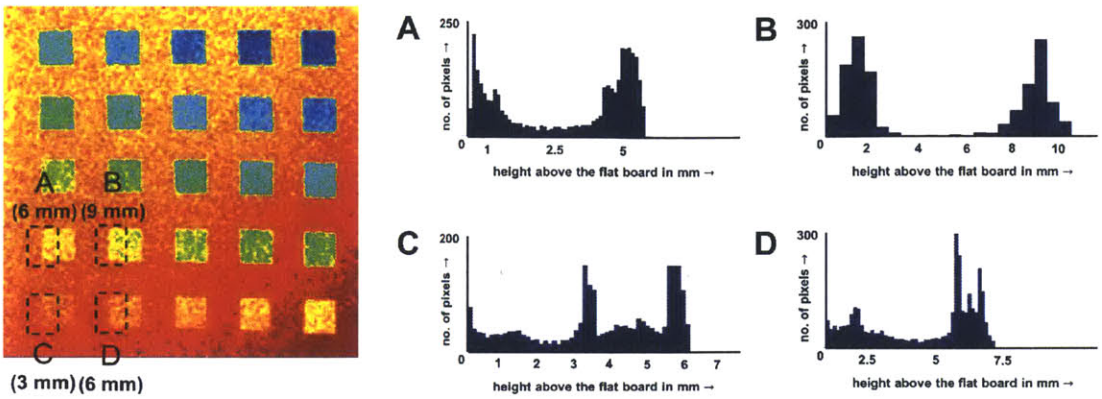


Figure 5 14: **Statistical analysis of depth resolution test for photon-counting histogram method using** 115 **photons/pixel.** Using the same analysis as in Fig. 5 13, we conclude that the pointwise photon counting method also achieves a 4 mm depth resolution. We note that the depth value histogram in Fig. 5 13(C) appears to be smooth and concentrated around one value. This is because the first photon computa tional imager employs spatial regularization which has the effect of blurring neighboring depth values which are too close to one another. In contrast, the depth histogram shown in sub figure (C) is formed using the pointwise photon counting histogram method, and it seemingly has well separated peaks. These peaks, however, are concentrated around incorrect depth values.

We then used our computational imager to reconstruct the reflectivity using only the first photon detections at each pixel, ignoring the rest of the data. For comparison, pointwise maximum-likelihood estimates based on first-photon data were also computed.

The number of distinguishable gray levels is an indicator of the reflectivity resolution achieved by the imager. As shown in Fig. 5-17, our method is able to discriminate 16 gray levels, which implies a reflectivity resolution of at least 4 bits using only first-photon detections. In comparison, the pointwise maximum-likelihood estimation allows visual discrimination of about 3 gray levels. Our computational imager achieves a reflectivity resolution similar to that of the baseline measurement, which required at least 900 photon detections per pixel.

Statistical analysis of reflectivity resolution test:

Similar to the histogram-based analysis of depth resolution, we plot the histogram of reconstructed reflectivity values in the image regions corresponding to the different linear grayscales (highlighted in red) on the reflectivity test chart as shown in Fig. 5-18. These his-

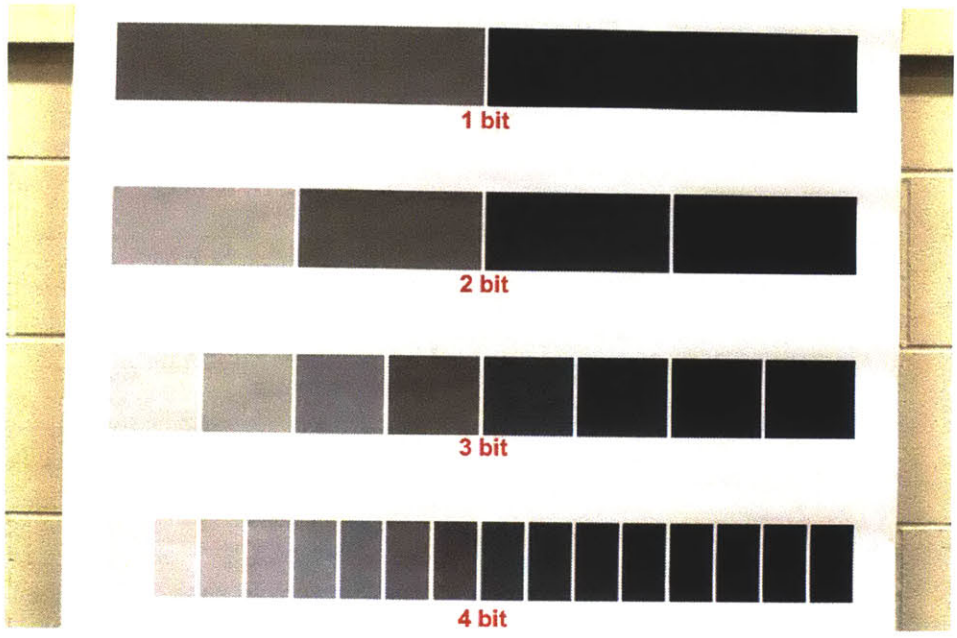


Figure 5 15: **Photograph of the reflectivity-resolution test target.** The text describes the reflectivity resolution achievable by distinguishing the linear gray scales.

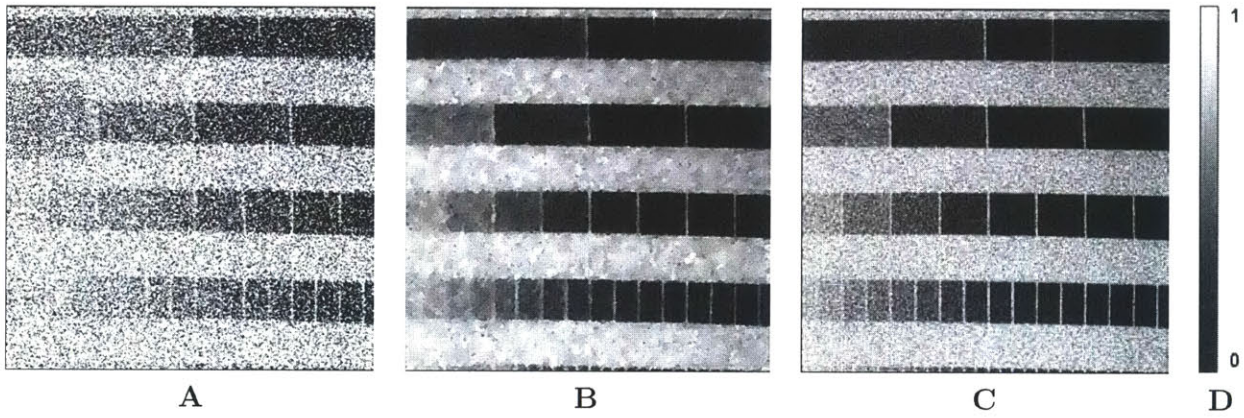


Figure 5 16: **Reflectivity resolution test.** **A**. Pointwise ML estimate. **B**. First photon Imaging. **C**. Baseline reflectivity estimate with at least 900 photons/pixel. **D**. Reflectivity scale. No postprocessing has been applied to the images.

tograms and cumulative distribution plots are generated for each of the following methods: pointwise ML estimation, first-photon computational imager and the baseline reflectivity measurement requiring at least 900 photons/pixel.

As demonstrated in Fig. 5-17, pointwise ML estimation fails to clearly identify beyond 3

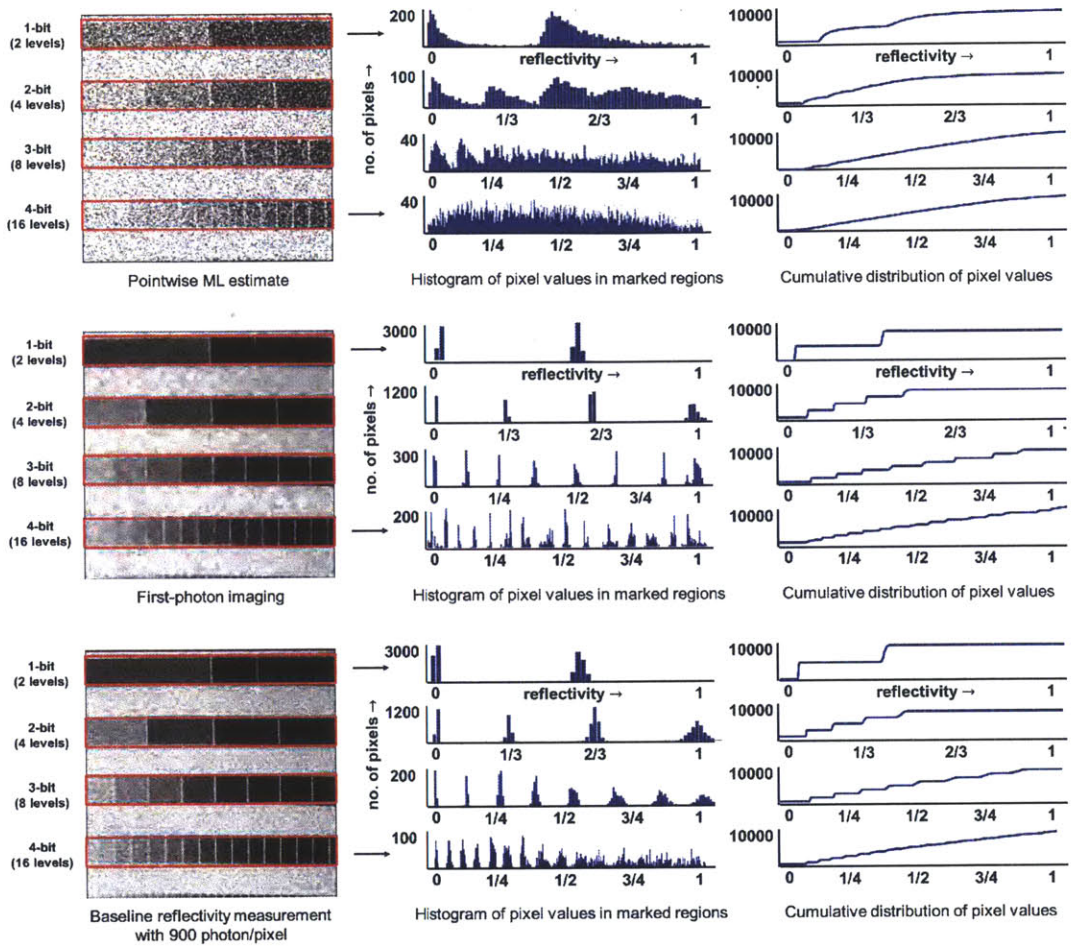


Figure 5 17: **Histogram-based analysis of reflectivity resolution.**

gray levels, and in comparison first-photon imaging has 16 different histogram features with clearly resolvable modes, each of which has a low-variance (also shown in Fig. 5-18).

The baseline reflectivity estimation method, which uses at least 900 photons/pixel, gives slightly worse performance than our method because of the high variance in the histogram components due to Poisson noise. The use of spatial correlations in our technique mitigates this Poisson noise and results in clearly resolvable reflectivity modes.

Imaging of Natural Scenes

Three additional scenes, consisting of real-world objects, were imaged. For each dataset, the pointwise maximum-likelihood estimate and the computationally reconstructed images

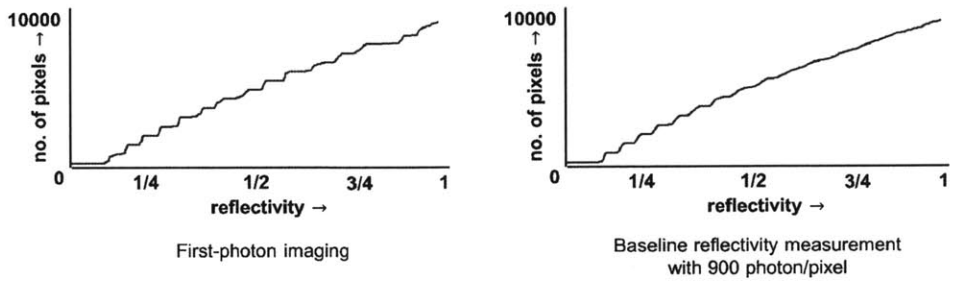


Figure 5 18: **Cumulative distribution plots showing 16 steps corresponding to the reflectivity modes in Fig. 5-17.**

are shown in Fig. 5-19, Fig. 5-20, and Fig. 5-21. In each of these figures 3D estimates are rendered as point clouds and overlaid with reflectivity estimates, with no additional post processing applied.

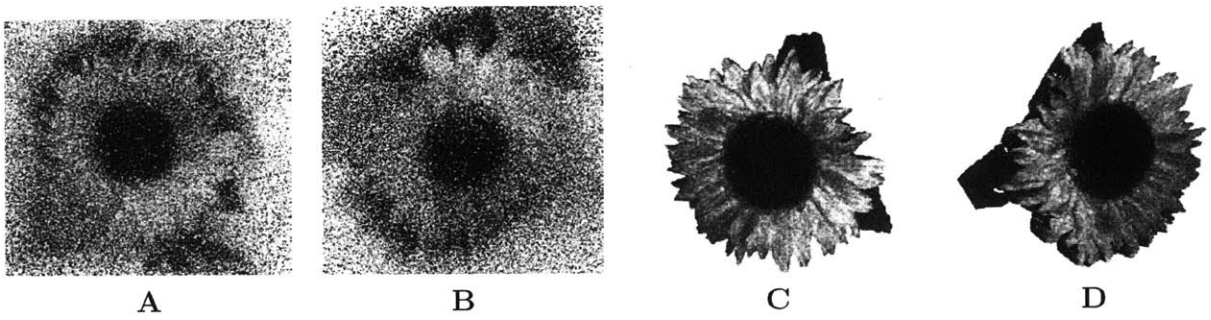


Figure 5 19: **Sunflower reconstruction.** Pointwise ML estimate (**A**. front view and **B**. right profile). First photon Imaging (**C**. front view and **D**. right profile). The lateral dimensions of the sunflower were 20 cm × 20 cm. The depth variation of the petals and the flower center was approximately 7 cm.

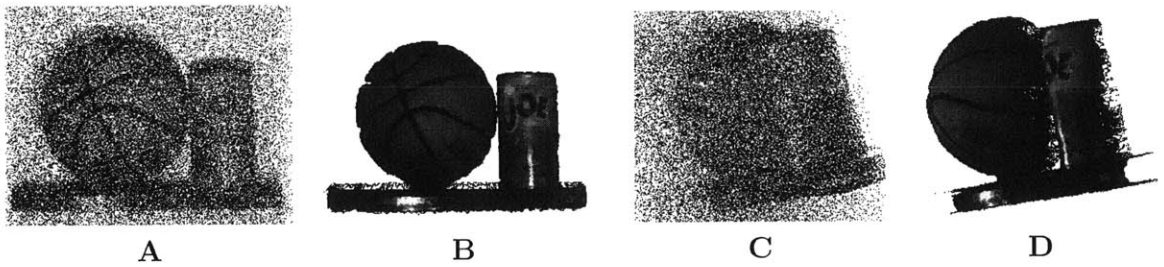


Figure 5 20: **Basketball + Can dataset reconstruction.** Pointwise ML estimate (**A**. front view and **C**. left profile). First photon Imaging (**B**. front view and **D**. left profile). The basketball diameter was 40 cm.

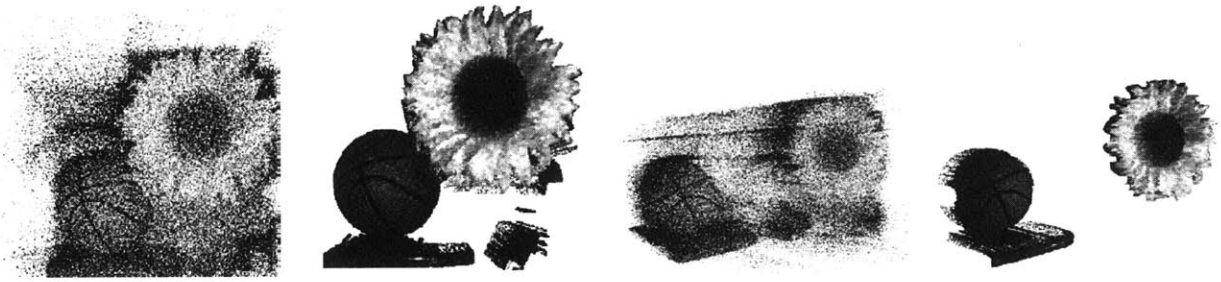


Figure 5 21: **Reconstruction of a layered scene with two life-sized objects.** Pointwise ML estimate (**A**. front view and **C**. left profile). First photon Imaging(**B**. front view and **D**. left profile).

Performance of ROAD Filtering in Step 2

In order to demonstrate the efficacy of the ROAD statistic-based noise rejection in Step 2, we obtained ground truth noisy photon labels by first computing the pixelwise absolute difference between ground truth depth profiles and the pointwise ML estimate computed using the first-photon data. All pixels at which this absolute depth difference was higher than cT_p, where T_p is the pulse-width in seconds, were labeled as noisy pixels. The ground truth noisy pixel labels were then compared with the labels produced by the noise censoring algorithm employed in Step 2 of our computational imager.

A pixelwise XOR of the two label images yielded the pixels at which there were discrepancies. These are pixels that were either classified wrongly as noise pixels (false-positives) or noisy pixels that our noise censoring algorithm failed to identify (false-negatives). Figure 5-22 shows that our proposed noise-censoring algorithm is successful at identifying the majority of noisy pixels while retaining the signal photon data for Step 3 depth reconstruction.

Comparison with Other Denoising Methods

For comparison, we processed the corrupted pointwise ML estimates (Equation (5.7)) with state-of-the-art nonlinear denoising methods that use spatial correlations to mitigate high Poisson noise (BM3D with Anscombe transform [108]) as well as high levels of impulsive noise (median filtering). We observed that BM3D achieved better PSNR than median filtering for reflectivity reconstruction, whereas median filtering with a 9×9 window achieved lower

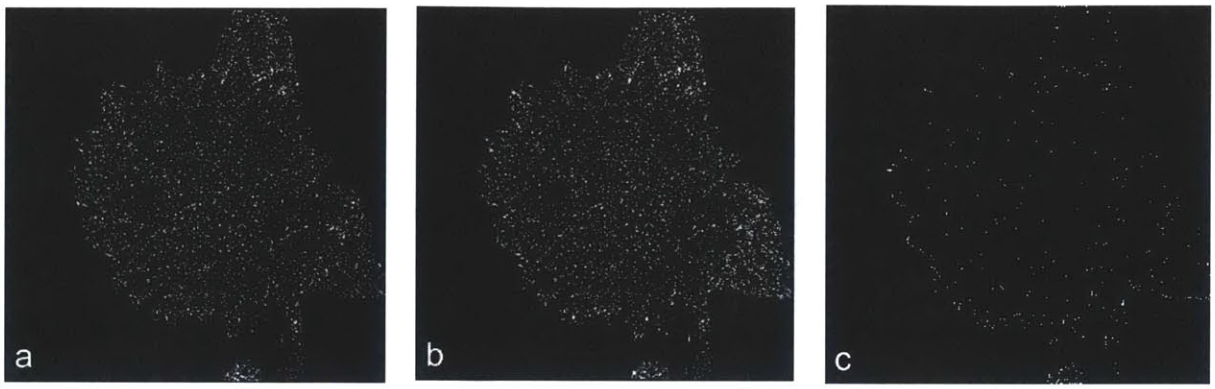


Figure 5 22: **Performance analysis of ROAD statistic-based filtering on sunflower dataset. A.** Ground truth for noise pixels. **B.** First photon imaging step 2 noise pixels. **C.** Pixelwise XOR of label images **A** and **B**.

RMSE than BM3D for depth reconstruction. First-photon imaging performs significantly better than both BM3D and median filtering and is able to recover reflectivity and structural details that are heavily obscured in images denoised using these two methods. (see Figures 5-23 5-30). For all processing methods used, the parameters were chosen to minimize RMSE for depth maps and maximize PSNR for reflectivity reconstruction. Table 5.1 summarizes the PSNR and RMSE results for the three different natural scenes for the various processing methods employed in this chapter.

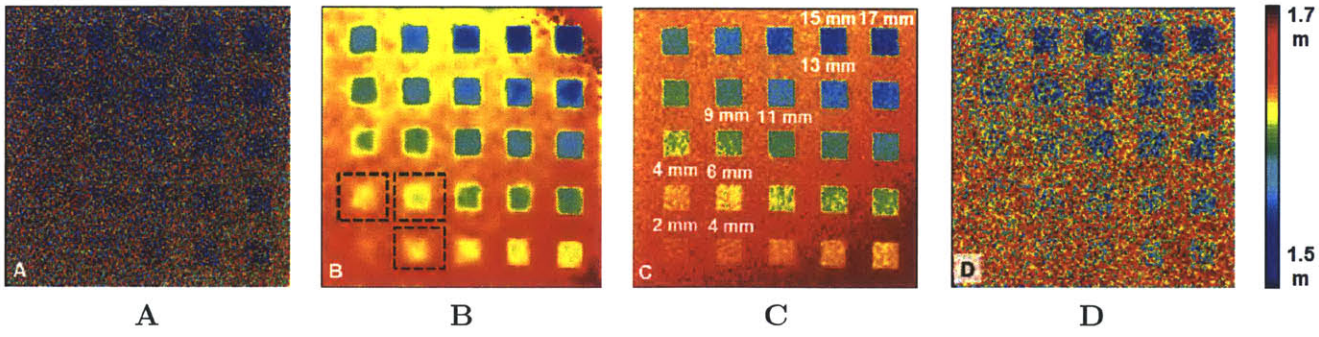
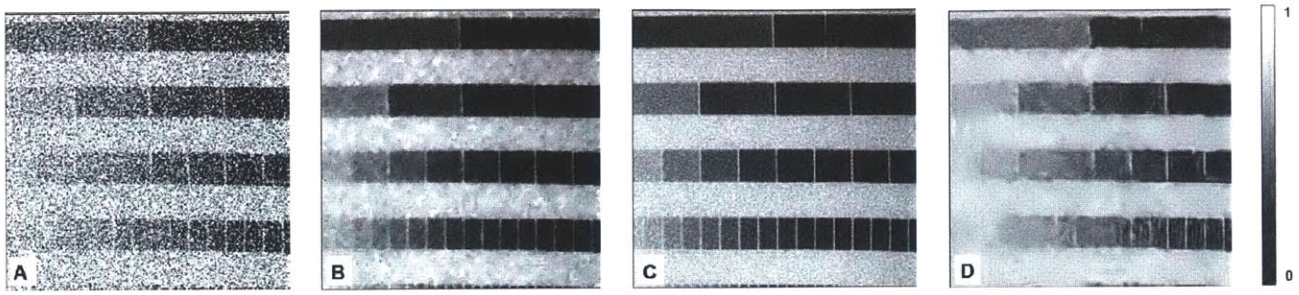
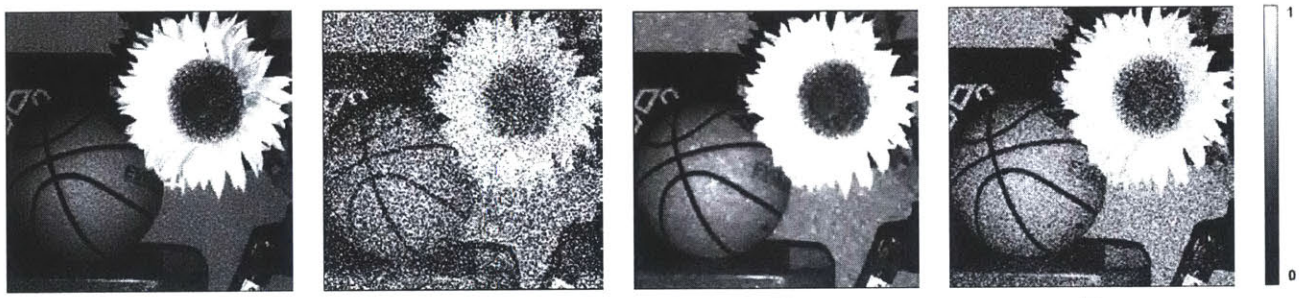


Figure 5 23: **Depth resolution test chart comparison. A.** Pointwise ML estimate. **B.** First photon imaging. **C.** Ground truth. **D.** Median filtering (lower RMSE than BM3D).



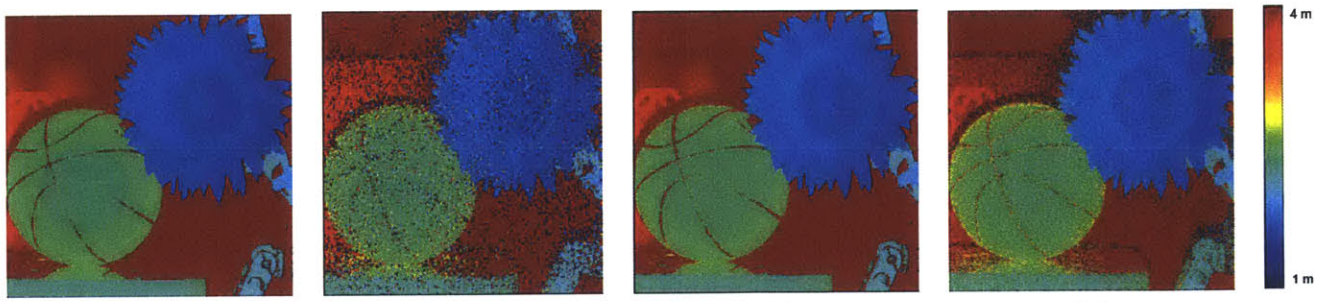
A B C D

Figure 5 24: **Reflectivity resolution test chart comparison: A.** Pointwise ML estimate. **B.** First photon imaging. **C.** Ground truth. **D.** BM3D (higher PSNR than median filtering).



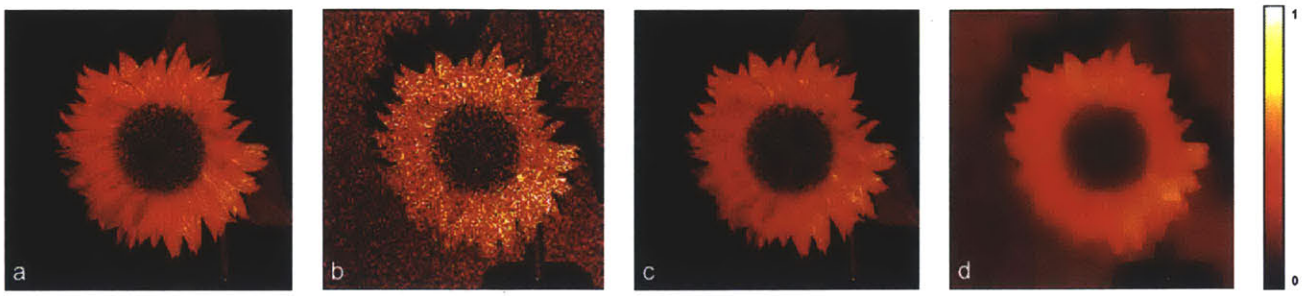
A B C D

Figure 5 25: **Layered scene dataset reflectivity reconstruction comparison: A**. Ground truth. **B**. Pointwise ML estimate. **C**. First photon imaging **D**. BM3D with Anscombe transformation.



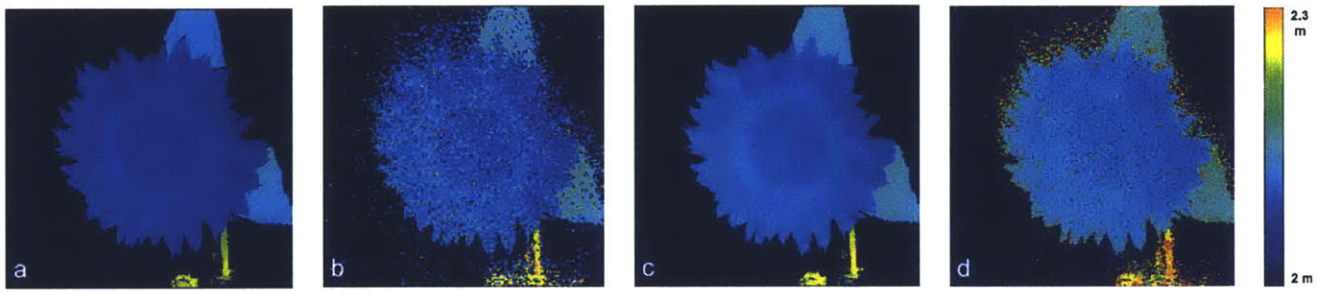
A B C D

Figure 5 26: **Layered scene dataset depth reconstruction comparison: A**. Ground truth. **B**. Pointwise ML estimate. **C**. First photon imaging **D**. Median filtering.



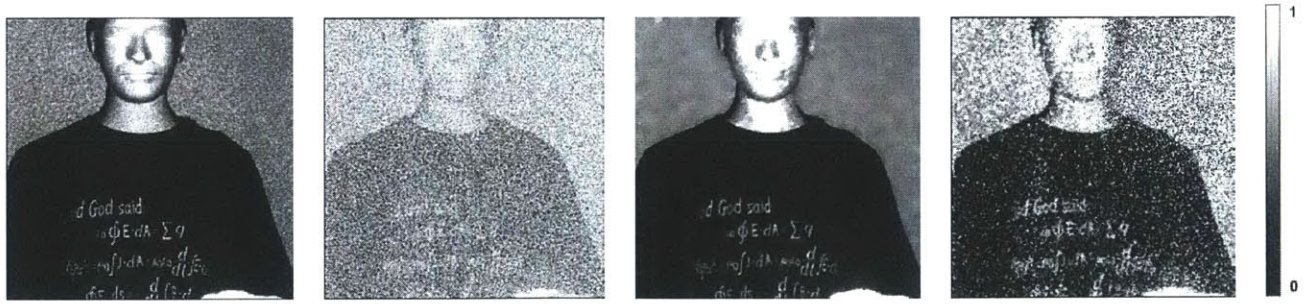
A B C D

Figure 5 27: **Sunflower dataset reflectivity reconstruction comparison: A**. Ground truth. **B**. Point wise ML estimate. **C**. First photon imaging **D**. BM3D with Anscombe transformation. First photon imaging rejects background and increases image contrast while retaining fine spatial features like flower petals. In comparison, BM3D reduces errors at the expense of over smoothing and losing spatial features.



A B C D

Figure 5 28: **Sunflower scene dataset depth reconstruction comparison: A**. Ground truth. **B**. Pointwise ML estimate. **C**. First photon imaging **D**. Median filtering. Our method rejects background and denoises while retaining fine spatial features like flower petals.



A B C D

Figure 5 29: **Mannequin dataset reflectivity reconstruction comparison: A**. Ground truth. **B**. Pointwise ML estimate. **C**. First photon imaging **D**. BM3D with Anscombe transformation.

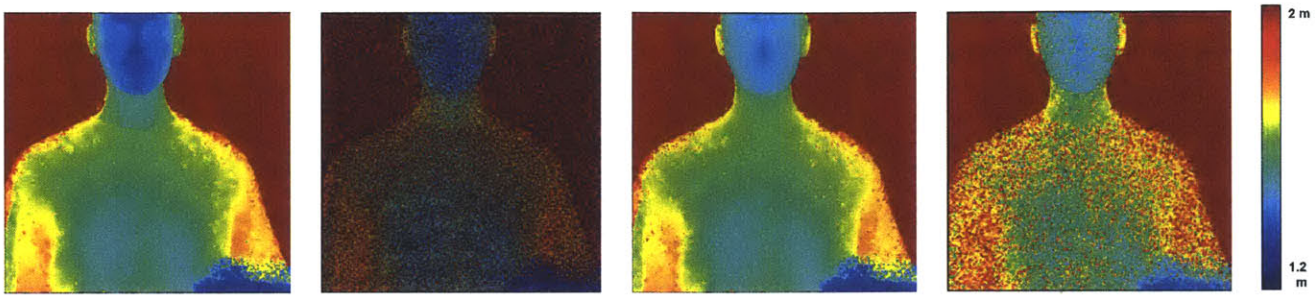


Figure 5 30: **Mannequin dataset depth reconstruction comparison: A**. Ground truth. **B**. Pointwise ML estimate. **C**. First photon imaging **D**. Median filtering.

		Pointwise ML	First-photon Imaging	BM3D	Median Filtering
Mannequin	PSNR	11 dB	35 dB	18 dB	11.5 dB
	RMSE	212 cm	2.4 cm	27.3 cm	14.7 cm
Sunflower	PSNR	10 dB	35 dB	19 dB	15 dB
	RMSE	135 cm	5.3 cm	21.3 cm	10.6 cm
Basketball and can	PSNR	8 dB	44 dB	20 dB	17 dB
	RMSE	157 cm	6.8 cm	19.4 cm	11.8 cm
Reflectivity chart	PSNR	15 dB	54 dB	25 dB	18 dB
Depth chart	RMSE	240 cm	0.4 cm	15.7 cm	12.9 cm

Table 5.1: Performance comparison of the various data processing methods proposed in this chapter using the natural scene datasets.

Repeatability analysis

For testing the repeatability performance of the first-photon computational imager, 500 independent first-photon datasets for the layered scene were collected and processed using a fixed set of numerical and optical parameters. As an example, the pixelwise standard deviations of the reconstructed scene depth using pointwise ML estimation, first-photon

imager and median filtering are shown in Fig. 5-31. The first-photon computational imager achieves a low standard deviation ($4 - 6$ mm) throughout the image, which is consistent with its depth resolution discussed earlier. This indicates that our proposed computational imager is robust and consistently improves estimation performance across independent trials. The SNR at object edges and in low-reflectivity regions is low and therefore the estimation quality is poorer in these regions. Both pointwise ML estimation and median-filtering have high standard deviations throughout the image, indicating that these denoising methods fail due to a mismatch in signal modeling. For repeatability tests conducted with all other datasets as well, the first-photon computational imager consistently outperformed BM3D and median filtering for both depth and reflectivity reconstruction.

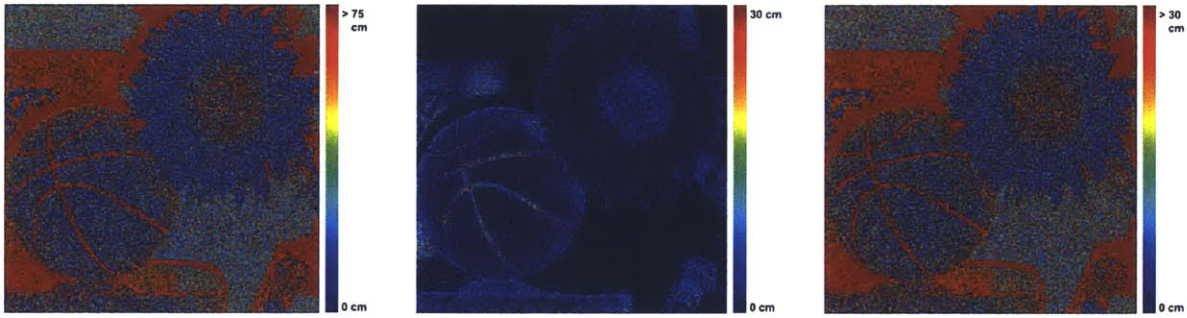


Figure 5 31: **Repeatability analysis for depth reconstruction.** Pixelwise standard deviation of depth estimates computed by processing 500 first photon data trials processed using: **A.** Pointwise ML estimate. **B.** First photon imaging. **C.** Median filtering (which performed better than BM3D).

5.9 Discussion and Limitations

As shown in Fig. 5-10, our method incurs the highest depth error near edges. The surface normals at these locations are nearly perpendicular to the line of sight, which dramatically reduces SNR. Consequently, these regions incur more noisy detections than do the rest of the pixels. Although our method censors depth anomalies near edges, it estimates the missing depth values using spatial correlations, leading to loss of subtle depth details.

A detected photon may have originated from an indirect bounce, causing estimation inaccuracy. However, for quasi-Lambertian scenes, diffuse scattering causes the light multipath

bounces to be considerably weaker than the direct reflection. The objects used in our experiments were quasi-concave so there were no noticeable depth estimation errors that could be attributed to light multipath.

The reflectivity estimation step of the first-photon computational imager fails if background noise is sufficient to provide a detection in each pulse repetition period with high probability. Hence we employed a suitably narrowband spectral filter to hold background noise level to $B \approx \bar{\alpha}S$, where $\bar{\alpha}$ is the average scene reflectivity. Figure 5-32 visually demonstrates the effect of increasing background light levels on depth and reflectivity reconstruction. As shown, doubling the background light level increases its shot noise contribution significantly and degrades the imaging quality of the first-photon computational imager.

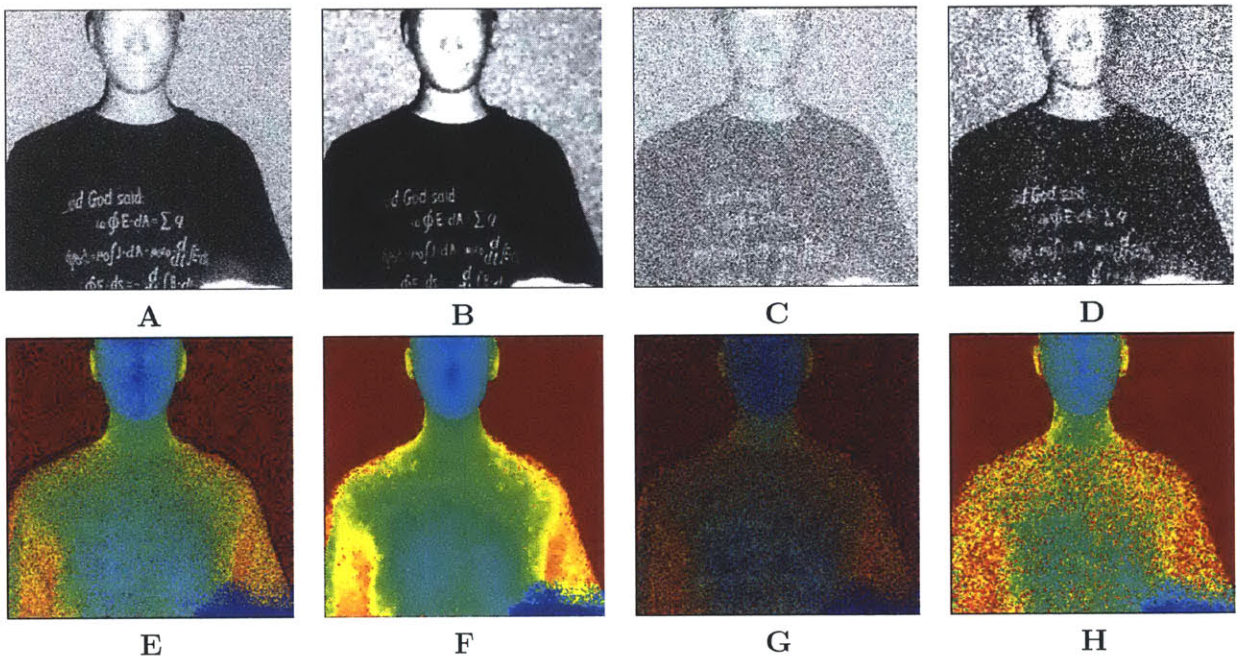


Figure 5 32: **Effect of increasing ambient noise.** Reflectivity reconstruction using **A.** Pointwise ML estimation and **B.** First photon computational imager with $B \approx \bar{\alpha}S$. Reflectivity reconstruction using **C.** Pointwise ML estimation and **D.** First photon computational imager with $B \approx 2\bar{\alpha}S$. Depth reconstruction using **E.** Pointwise ML estimation and **F.** First photon computational imager with $B \approx \bar{\alpha}S$. Depth reconstruction using **G.** Pointwise ML estimation and **H.** First photon computational imager with $B \approx 2\bar{\alpha}S$.

The first-photon computational imager assumes that average background photon counts, B, is spatially invariant, i.e., it is constant across the scene patches. This may not be true in practice, so achieving accurate imaging would require the estimation of B at each pixel

separately. Finally, the first-photon imager has a variable dwell time per pixel, since the time to first-photon detection is a random variable. This makes first-photon imaging unsuitable for implementation with sensor arrays because they typically have a fixed dwell time for each acquired image. The next chapter focuses on adapting the first-photon imaging principles to sensor arrays with fixed dwell times.

Another limitation of the first-photon computational imager presented in this chapter is its inability to resolve spatial features that are comparable in size to the illumination beam. Sub-pixel raster-scanning may potentially be used resolve object features that are smaller than the laser beam spot-size. Also, the first-photon imager requires direct line-of-sight for its operation. It is not possible to image scenes that are occluded from either the light source or the sensor. A potentially interesting possibility is to combine the low-light detection framework proposed in this chapter with the hidden-scene imaging framework outlined in Chapter 2. After all, one of the major limitations of the hidden-plane imaging setup is the SNR loss due to Lambertian scattering resulting in extremely low-light levels of backreflected signal light reaching the detector. Thus wedding our hidden plane imager to photon-efficient first-photon imaging could be quite beneficial.

Chapter 6

Photon Efficient Imaging with Sensor Arrays

6.1 Overview

In this chapter, we describe and demonstrate another active optical imaging framework that recovers accurate reflectivity and 3D images simultaneously using on the order of one detected photon per pixel averaged over the scene. Similar to the first-photon imaging method described in Chapter 5, the proposed computational imager avoids the use of a photon-count histogram to infer time delay and amplitude relative to the transmitted pulse's temporal profile [78, 79, 84]. Instead it combines probabilistic modeling at the level of individual detected photons with exploitation of the spatial correlations present in real-world scenes to achieve accurate 3D and reflectivity imaging when very little backreflected light reaches the detector, as will be the case with low optical-power active imagers [90].

First-photon imaging data acquisition involves a variable dwell time per pixel, i.e., the raster scanning pulsed light source continues to illuminate a scene patch with light pulses until the SPAD detector records a photon arrival in response to that illumination. Since low-light photodetection is governed by Poisson statistics [80, 81], the number of pulses transmitted prior to the first photon detection is a random variable (see Section 5.4). In the absence of background light, scene patches with low reflectivity will require on average, a longer

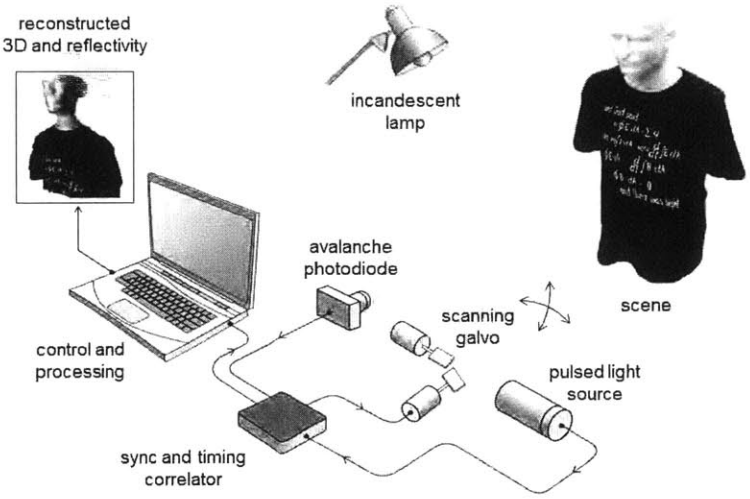


Figure 6 1: **Fixed dwell time imaging setup.** A pulsed light source illuminates the scene in a raster scan pattern. The backscattered light is collected by a time resolved single photon detector. Each spatial location is illuminated with *exactly* N light pulses (fixed dwell time). An incandescent lamp injects background light which corrupts the information bearing signal. The photon detection times and the total photon count are recorded at every image pixel. This dataset is used to estimate the 3D structure and reflectivity. The setup is analogous to having a floodlight illumination source and an array of single photon counting detectors operating at a fixed dwell time.

dwell time when compared with high reflectivity scene patches. Thus, first-photon imaging does not extend naturally to operation using SPAD arrays since simultaneous measurement implies equal dwell times thus precluding the dramatic speedup in image acquisition that such arrays enable.

The image reconstruction methods and statistical models that we develop in this chapter are analogous to first-photon imaging but are applicable to the case of fixed dwell time at each pixel. The use of deterministic dwell times is both more convenient for raster scanning and amenable to parallelization through the use of a detector array (see [109 111]).

Using the imaging setup described in Fig. 6-1, we demonstrate that the performance of the proposed computational imager is similar to or slightly better than the first-photon computational imager when compared for equal total acquisition time in raster-scanned operation. The main challenge with such a small but fixed dwell time is that there may not be any photons detected from some scene patches. As a result, at pixels observing such scene patches there is no available data from which to infer scene depth and reflectivity.

Despite this lack of data, the proposed computational imager exploits spatial correlations to accurately estimate scene depth and reflectivity at each image pixel.

In addition to the comparison with first-photon imaging, it is experimentally demonstrated in Section 6.6 that the proposed computational imager is able to accurately recover scene depth and reflectivity, while traditional maximum-likelihood based fixed dwell time imaging methods lead to estimates that are highly noisy. Furthermore, with an increase in illumination power and the number of pixels in the sensor array, the proposed computational imager achieves a proportional speed-up in acquisition time compared to a single-detector raster-scanned system.

The remainder of this chapter is organized as follows. Section 6.2 introduces the LIDAR-like imaging configuration that we consider. The key probabilistic models for the measured data are derived in Section 6.3. These models are related to conventional image formation in Section 6.4, and they are the basis for the novel image formation method in Section 6.5. Section 6.6 presents experimental results for the novel method, and Section 6.7 provides additional discussion and conclusions. Appendix B presents performance bounds for the proposed imaging framework.

6.2 Imaging Setup and Signal Modeling

The imaging setup comprising scene parameters, active illumination and single-photon detector specifications for the fixed dwell time configuration considered in this chapter is identical to the one described in Section 5.3. Also identical is the signal model for backreflected light and the background light, and the measurement model as well as the low-rate flux assumption described in Section 5.4. The main differences between the imaging framework described in this chapter and first-photon imaging lie in the data acquisition, modeling and processing. These topics are discussed next.

Data Acquisition: Each scene patch (x, y) is illuminated with N laser pulses. The total acquisition time (dwell time) is thus $T_a = NT_r$. We record the total number of photon detections $k(x, y)$, along with their detection times $\{t(x, y)^{(\ell)}\}_{\ell=1}^{k(x,y)}$, where the latter are

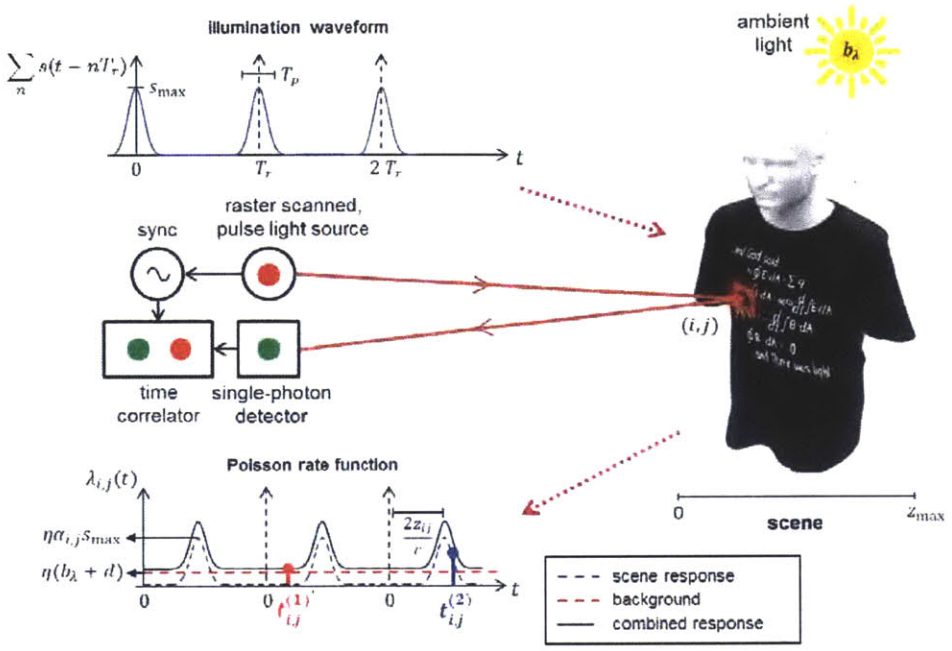


Figure 6 2: **Summary of data acquisition and observation model.** Rate function of inhomogeneous Poisson process combining desired scene response and noise sources is shown. Here, $N = 3$ and $k(x, y) = 2$. A background count (red) occurred after the second pulse was transmitted, and a signal count (blue) occurred after the third pulse was transmitted.

measured relative to the immediately-preceding transmitted pulse. We also note that the fixed dwell time data $\{k(x,y)\}_{x,y=1}^{M}$ and $\left\{\{t(x,y)^{(\ell)}\}_{\ell=1}^{k(x,y)}\right\}_{x,y=1}^{M}$ are outcomes or realizations of the random variables $K(x,y)$ and $T^{\ell}(x,y)$ respectively. For a fixed (x,y) location, the random variables $T^{\ell}(x,y)$ are independent and identically distributed since they are photon time-of-arrival observations in response to distinct light pulses. Thus to simplify the notation, we use the random variable $T(x,y)$ without the superscript ℓ.

6.3 Measurement Model

Model for the numbers of detected photons $K(x,y)$: A SPAD detector is not *number-resolving*, meaning that it reports at most one click from detection of a signal pulse. As derived in Equation (5.4) the probability of the SPAD detector *not* recording a detection at pixel (x,y) from one illumination trial is $P_0(x,y) = e^{-(\,\alpha(x,y)\,S+B)}$. Because we illuminate with a total of N pulses, and the low-flux condition ensures that multiple detections per repetition interval can be neglected, the number of detected photons $K(x,y)$ is binomially

distributed with probability mass function

$$\begin{aligned}&\Pr\left[K(x,y)=k(x,y);\alpha(x,y)\right]\\&\quad=\binom{N}{k(x,y)}P_0(x,y)^{N-k(x,y)}\left[1-P_0(x,y)\right]^{k(x,y)},\end{aligned}$$

for $k(x,y)=0,\,1,\,\ldots,\,N$ and where $\binom{\cdot}{\cdot}$ is the binomial coefficient.

In the ultimate low-flux limit in which $\alpha(x,y)S+B\to 0^+$ with $N\to\infty$ such that $N\{1-P_0(x,y)\}=C(x,y)$ is held constant, $K(x,y)$ converges to a Poisson random variable [104] with probability mass function

$$\Pr[K(x,y)=k(x,y);\,\alpha(x,y)]=\frac{C(x,y)^k}{k\,!}\,\exp[-C(x,y)].$$

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

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

where (a) uses the independent increments property of the Poisson process and (b) uses Equation (5.4). The probability density function of $T(x,y)\in[0,T_r)$, the continuous time-of-detection random variable, is then obtained by evaluating the preceding probability on a per unit time basis as $\Delta\to 0^+$:

$$
\begin{aligned}
& f_{T(x,y)}(t(x,y);\, \alpha(x,y), z(x,y)) \\
&= \frac{1}{1-\exp[-(\alpha(x,y)S+B)]} \times \\
&\quad \lim_{\Delta\to 0^+} \frac{1}{\Delta}\left\{ \exp\left[-\int_0^{t(x,y)} \left(\alpha(x,y) s\left(\tau - \frac{2z(x,y)}{c}\right) + \frac{B}{T_r}\right) d\tau \right] \right. \\
&\quad \left. - \exp\left[-\int_0^{t(x,y)+\Delta} \left(\alpha(x,y) s\left(\tau - \frac{2z(x,y)}{c}\right) + \frac{B}{T_r}\right) d\tau \right] \right\} \\
&= \frac{\alpha(x,y) s(t(x,y) - 2z(x,y)/c) + B/T_r}{1-\exp[-(\alpha(x,y)S+B)]} \\
&\quad \times \exp\left[-\int_0^{t(x,y)} \left(\alpha(x,y) s(\tau - 2z(x,y)/c) + \frac{B}{T_r}\right) d\tau \right] \\
&\overset{(a)}{=} \frac{\alpha(x,y) s(t(x,y) - 2z(x,y)/c) + B/T_r}{\int_0^{T_r} [\alpha(x,y) s(t(x,y) - 2z(x,y)/c) + B/T_r]\, dt} \\
&= \frac{\alpha(x,y)S}{\alpha(x,y)S+B}\left(\frac{s(t(x,y) - 2z(x,y)/c)}{S}\right) + \frac{B}{\alpha(x,y)S+B}\left(\frac{1}{T_r}\right),
\end{aligned}
\tag{6.1}
$$

where (a) follows from $\alpha(x,y)S + B \ll 1$. As discussed in Section 5.4, a photon detection could be generated due to the backreflected signal or due to background light. In this computational imager, we will make use the probabilities described in Equation (5.3) during the background censoring step.

6.4 Conventional Image Formation

Pointwise ML reflectivity estimation: Given the total observed photon count $k(x,y)$ at pixel (x,y), the constrained ML (CML) reflectivity estimate is

$$
\begin{aligned}
\hat{\alpha}(x,y)^{\mathrm{CML}} &= \arg\max_{\alpha(x,y)\geq 0} \Pr[\, K(x,y) = k(x,y); \alpha(x,y)\,] \\
&= \max\left\{ \frac{1}{S}\left[\log\left(\frac{N}{N-k(x,y)}\right) - B \right], 0 \right\}.
\end{aligned}
$$

where log is the natural logarithm. Traditionally, the normalized photon-count value is used as the reflectivity estimate [79],

$$
\tilde{\alpha}(x,y) = \frac{k(x,y)}{NS}. \tag{6.2}
$$

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

Pointwise ML depth estimation: Using the photon detection-time dataset $\{t(x,y)^{(\ell)}\}_{\ell=1}^{k(x,y)}$, the pixelwise or pointwise constrained ML depth estimate is

$$\begin{aligned}\hat{z}(x,y)^{\mathrm{CML}} &= \underset{z(x,y)\in[0,cT_r/2)}{\arg\max} \prod_{\ell=1}^{k(x,y)} f_{T(x,y)}(t(x,y)^{(\ell)};\, \alpha(x,y), z(x,y)) \\ &= \underset{z(x,y)\in[0,cT_r/2)}{\arg\max} \sum_{\ell=1}^{k(x,y)} \log\left[\alpha(x,y)s\left(t(x,y)^{(\ell)} - \frac{2z(x,y)}{c}\right) + \frac{B}{T_r}\right],\end{aligned}$$

assuming that $k(x,y) \geq 1$. If $B > 0$, then the ML depth estimate is obtained by solving a non-convex optimization problem. Moreover, ML estimation when $B > 0$ requires the knowledge of the true reflectivity $\alpha(x,y)$, which is not typically available. Thus, the log-matched filter [80] is instead traditionally used for estimating depth:

$$\tilde{z}(x,y) = \underset{z(x,y)\in[0,cT_r/2)}{\arg\max} \sum_{\ell=1}^{k(x,y)} \log\left[s\left(t(x,y)^{(\ell)} - 2z(x,y)/c\right)\right]. \tag{6.3}$$

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

6.5 Novel Image Formation

Similar to first-photon imaging our computational image formation proceeds in three steps.

Step 1: Reflectivity Estimation

The negative log-likelihood of scene reflectivity $\alpha(x,y)$ given count data $k(x,y)$ is

$$\begin{aligned}\mathcal{L}_\alpha(\alpha(x,y); k(x,y)) &\triangleq -\log \Pr[\, K(x,y) = k(x,y); \alpha(x,y)\,] = \\ &(N - k(x,y))S\alpha(x,y) - k(x,y)\log\{1 - \exp[-(\alpha(x,y)S + B)]\},\end{aligned} \tag{6.4}$$

after constants independent of $\alpha(x,y)$ are dropped. Since $\mathcal{L}_\alpha(\alpha(x,y);k(x,y))$ is a strictly convex function in $\alpha(x,y)$, it is amenable to global minimization using convex optimization, with or without the inclusion of sparsity-based regularization [112]. The penalized ML (PML) estimate for scene reflectivity is obtained from noisy data $\{k(x,y)\}_{x,y=1}^{M}$ by solving the following convex program:

$$\hat{\boldsymbol{\alpha}}^{\mathrm{PML}} = \underset{\boldsymbol{\alpha}:\alpha(x,y)\geq 0}{\arg\min}\,(1-\beta_\alpha)\sum_{x=1}^{M}\sum_{y=1}^{M}\mathcal{L}_\alpha(\alpha(x,y);k(x,y)) + \beta_\alpha\,\mathrm{pen}_\alpha(\boldsymbol{\alpha}),$$

where $\mathrm{pen}_\alpha(\cdot)$ is a convex function that penalizes the non-smoothness of the reflectivity estimate, and β_α controls the degree of penalization. We used the total variation seminorm [39] as the penalty function in our experiments in this chapter.

Step 2: Rejection of Background Detections

As in first-photon imaging, direct application of a similar regularized-ML approach to depth estimation using time-of-detection data is infeasible. This is because the background contribution to the likelihood function creates a non-convex cost function with locally-optimal solutions that are far from the global optimum. Hence, before estimating depth, a second processing step attempts to identify and censor the detections that are due to background. Our method to censor a noisy detection at transverse location (x,y) based on the photon arrival data is as follows:

1. Compute the rank-ordered mean (ROM) $t^{\mathrm{ROM}}(x,y)$ for each pixel, which is the median value of the detection times at the 8 neighboring pixels of (x,y) [107]. If $t^{\mathrm{ROM}}(x,y)$ cannot be computed due to missing data, then set $t^{\mathrm{ROM}}(x,y)=\infty$.

2. Estimate the set of uncensored detections, $U(x,y)$, i.e., those presumed to be signal detections, as follows:

$$\left\{\ell : |t(x,y)^{(\ell)} - t^{\mathrm{ROM}}(x,y)| < 2T_p\left(\frac{B}{\hat{\alpha}^{\mathrm{PML}}(x,y)S+B}\right), 1\leq \ell \leq k(x,y)\right\}.$$

If $k(x,y)=0$, then set $U(x,y)=\varnothing$.

It is demonstrated in [107] that the ROM image t^{ROM} is a good approximation of the true image when the true image is corrupted by high-variance impulse noise at every pixel. In our imaging setup, the background photon detections are uniformly distributed with high variance. Also, the variances of signal photon detections depend on the duration of the transmitted pulse. Thus, the quality of ROM approximation at pixel (x, y) deteriorates as the probability of detecting a background photon $B/(\alpha(x, y)S + B)$ increases or RMS pulse-width T_p increases. Because the condition for censoring photon detections must be relaxed for an unreliable ROM estimate at every pixel, we set the censoring threshold parameter to be linearly dependent on both the RMS pulse-width and our estimate of the background detection probability. We have found that removing dependence on $\hat{\alpha}^{\mathrm{PML}}(x, y)$ from the censoring rule results in significantly worse performance; this link between estimation of reflectance and depth is a feature common to this work and first-photon imaging but not seen in earlier methods for photon-efficient imaging.

Step 3: Depth Estimation

With background detections rejected, the negative log-likelihood function of depth $z(x, y)$, given uncensored data $\{t^{(\ell)}(x, y)\}_{\ell \in U(x,y)}$, is

$$\mathcal{L}_z\big(z(x,y);\, \{t(x,y)^{(\ell)}\}_{\ell \in U(x,y)}\big) = -\sum_{\ell \in U(x,y)} \log\big[s\big(t(x,y)^{(\ell)} - 2z(x,y)/c\big)\big]\,.$$

If $|U(x, y)| = 0$, then set $\mathcal{L}_z(z(x, y); \{t(x, y)^{(\ell)}\}_{\ell \in U(x,y)}) = 0$, so that it has no contribution to the scene's negative log-likelihood cost function.

Many practical pulse shapes, including the pulse shape employed in our experiments, are well approximated as $s(t) \propto \exp[-v(t)]$, where $v(t)$ is a convex function in t. Then, $\mathcal{L}_z(z(x, y); \{t(x, y)^{(\ell)}\}_{\ell \in U(x,y)}) = \sum_{\ell \in U(x,y)} v(t(x, y)^{(\ell)} - 2z(x, y)/c)$ is a convex function in $z(x, y)$. Our penalized ML estimate for the scene depth image is thus obtained using uncensored data and solving the following convex optimization problem:

$$\hat{\boldsymbol{z}}^{\mathrm{PML}} = \underset{z:z(x,y)\in[0,cT_r/2)}{\arg\min}\; (1-\beta_z) \sum_{x=1}^{M}\sum_{y=1}^{M} \mathcal{L}_z\big(z(x,y);\, \{t(x,y)^{(\ell)}\}_{\ell \in U(x,y)}\big) + \beta_z\, \mathrm{pen}_z(\boldsymbol{z}),$$

where $\text{pen}_z(\cdot)$ is a convex function that penalizes non-smoothness of the depth estimate, and $\beta_z > 0$ controls the degree of penalization. Similar to regularized reflectivity estimation in step 1, we used the total variation seminorm as the penalty function for the depth map construction experiments in this chapter.

6.6 Experimental Results

Using the experimental setup and performance metrics described in Sections 5.7 and 5.8 we evaluated the performance of our proposed fixed dwell time depth map and reflectivity reconstruction method.

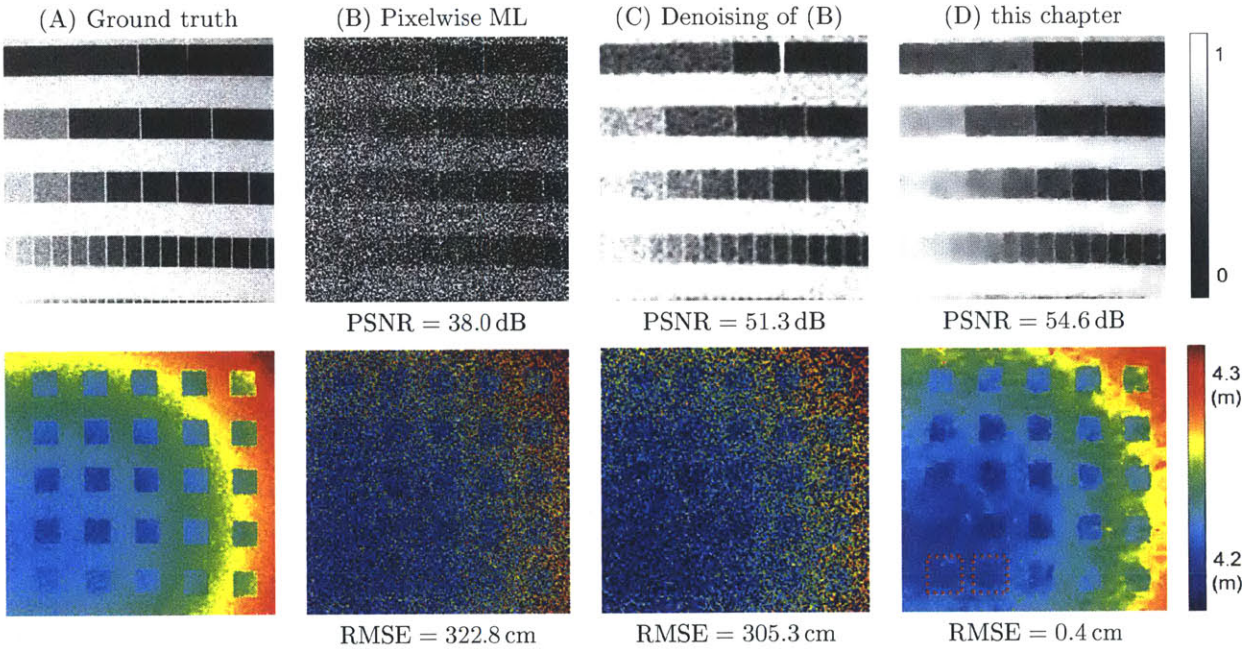


Figure 6 3: **Resolution test experiments**. Reflectivity chart imaging (top) was done using $T_a = 300\mu$s and had a mean count per pixel of 0.48. They were scaled to fill the reflectivity interval $[0, 1]$. Depth chart imaging (bottom) was done using $T_a = 6.2\mu$s and had a mean count per pixel of 1.1 with 33% of the pixels having missing data, i.e., no detections. For (d), bilateral and median filtering were used to denoise ML reflectivity and depth estimates, respectively.

Reflectivity Resolution Test

Reflectivity estimation was tested using the linear grayscale reflectivity chart shown in Fig. 5-15. Figure 6-3(D) shows that our method resolves 16 gray levels, performance similar to that of the ground-truth image from Fig. 6-3(A), which required about 1000 photon detections per pixel. We quantified the performance of a reflectivity estimator $\hat{\boldsymbol{\alpha}}$ of a true scene reflectivity $\boldsymbol{\alpha}$ using peak signal-to-noise ratio (PSNR) (see Equation (5.14) for definition) Figure 6-3(B) and (C) show that our method's PSNR exceeds that of pointwise ML (Equation (6.2)) by 16 dB, and it exceeds that of the bilateral-filtered [113] pointwise ML estimate by 3 dB.

Depth Resolution Test

Depth resolution was evaluated using the same test target and procedure used in first-photon imaging. Fig. 6-3(D) shows that our method achieves 4 mm depth resolution, which is comparable to that of the ground truth image (Fig. 6-3(A)), which required 100 detections per pixel, and far superior to the very noisy pointwise ML estimate (Equation (6.3)), and its median-filtered [114] version, which appear in Figures 6-3(B) and (C), respectively.

We quantified the performance of a depth estimator $\hat{\boldsymbol{z}}$ of a true scene depth $\boldsymbol{z}$ using root mean-square error (RMSE) metric defined in Equation (5.15). At the background level in our experiment, the pointwise ML estimates have an RMSE of at least 3 m. Because many pixels are missing photon detection-time observations, in order to denoise the pointwise ML estimate, we first perform bicubic interpolation and then apply median filtering, which is typically effective in eliminating moderate levels of impulse noise. The depth resolution of our method (4 mm) corresponds to 760-fold depth error reduction, compared to the denoised estimate. Due to the high levels of background noise the photon detection times have a variance that is much higher than what median filtering can mitigate.

Imaging of Natural Scenes

Reflectivity and depth images of two natural scenes a life-size mannequin, and a basketball next to a can are shown in Fig. 6-4. Ground-truth images, obtained using ML estimation from 200 detections at each pixel, appear in Fig. 6-4(a). The mannequin dataset for pointwise

ML estimation and for our method was generated using acquisition time $T_a = 100\,\mu$s. This dataset had 1.21 detections per pixel averaged over the entire scene with 54% of the pixels having no detections. The basketball-plus-can dataset for pointwise ML imaging and for our method also had $T_a = 100\,\mu$s, but its mean number of detections per pixel was 2.1, and 32% of its pixels had no detections. All reflectivity images were scaled to fill the interval $[0, 1]$.

Figure 6-3(b) shows that the pointwise ML estimation approach gives reflectivity and 3D estimates with low PSNR and high RMSE due to background-count shot noise at low light-levels. Pixels with missing data were imputed with the average of their neighboring 8 pointwise ML estimate values. Denoising the ML reflectivity estimate using bilateral filtering [113] and the ML depth estimate using median filtering [114] improves the image qualities (Fig. 6-3(c)). However, denoising the 3D structure of the mannequin shirt fails, because this region has very low reflectivity so that many of its pixels have missing data. On the other hand, our framework, which combines accurate photon-detection statistics with spatial prior information, constructs reflectivity and 3D images with 30.6 dB PSNR and 0.8 cm RMSE, respectively (Fig. 6-3(d)). We used the total variation semi-norm [115] as the penalty function in our method, and the penalty parameters were chosen to maximize PSNR for reflectivity imaging and minimize RMSE for 3D imaging.

Figure 6-5 shows how much photon efficiency we gain over traditional LIDAR systems that use the photon-count histogram approach. The photon-count histogram approach is a pixelwise or pointwise depth-estimation method that simply searches for the location of the peak in the photon-count histogram of the backreflected pulse. Whereas the log-matched filter is asymptotically ML as $B \to 0^+$, the photon-count histogram depth estimation method is asymptotically ML as $N \to \infty$. Thus, when T_a is long enough, as is the case in traditional LIDAR, it is effective to use the photon-count histogram depth estimation method. Based on PSNR and RMSE values, we see that our framework can allow more than $30\times$ speed-up in acquisition, while constructing the same high-quality 3D and reflectivity images that a traditional LIDAR system would have formed using long acquisition times.

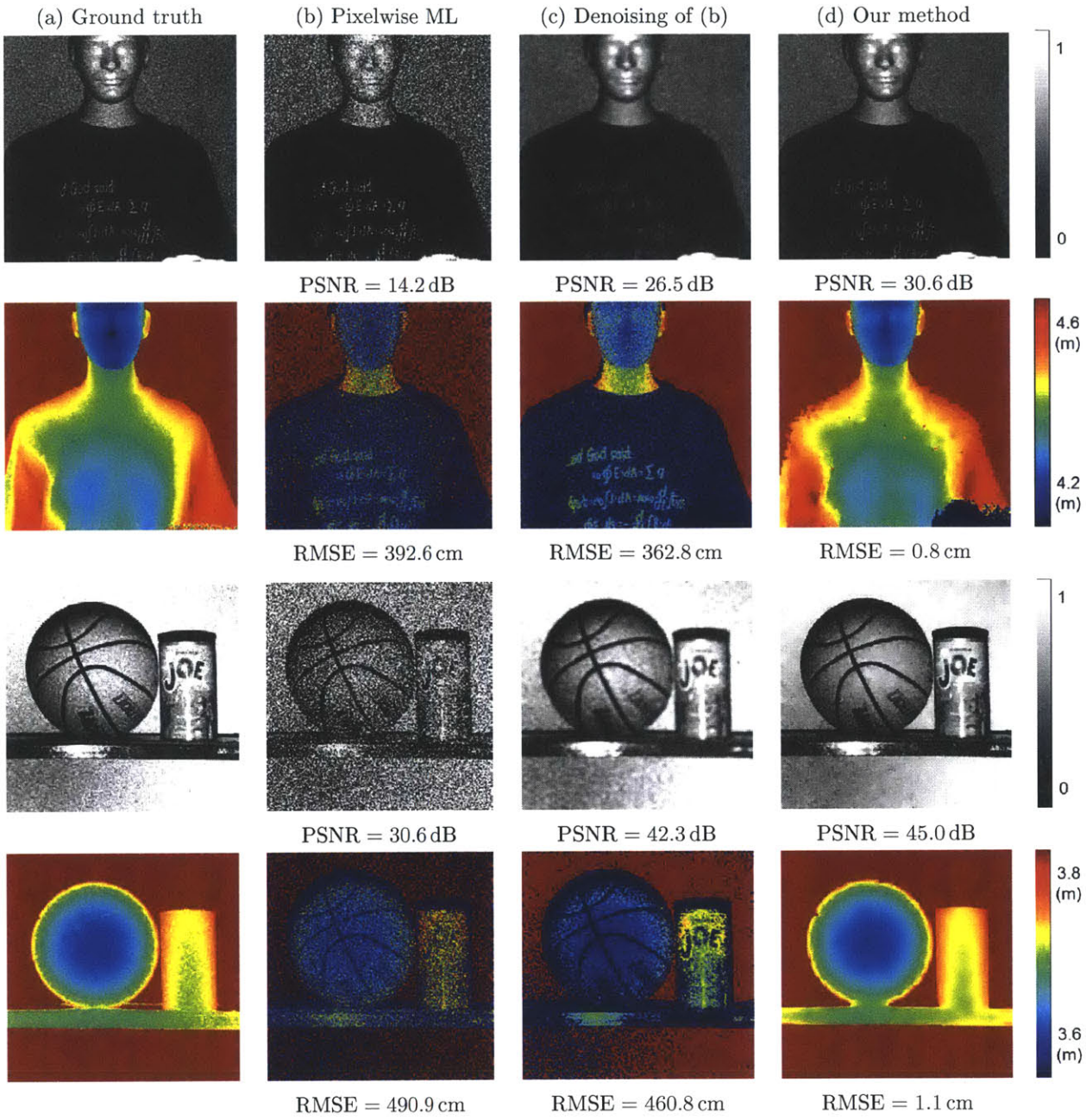


Figure 6 4: **Experimental results for reflectivity and 3D imaging of natural scenes.** We compare the reflectivity and depth images from our proposed method with those from pointwise ML estimation (see Section 6.4). For each method, the PSNR and RMSE values for the reconstructed reflectivity and 3D images are given. For the mannequin dataset (top), the mean per pixel count was 1.21 and 54% of the pixels were missing data. For the basketball plus can dataset (bottom), the mean per pixel count was 2.1 and 32% of the pixels were missing data. For (c), bilateral and median filtering were used to denoise ML reflectivity and depth estimates, respectively.

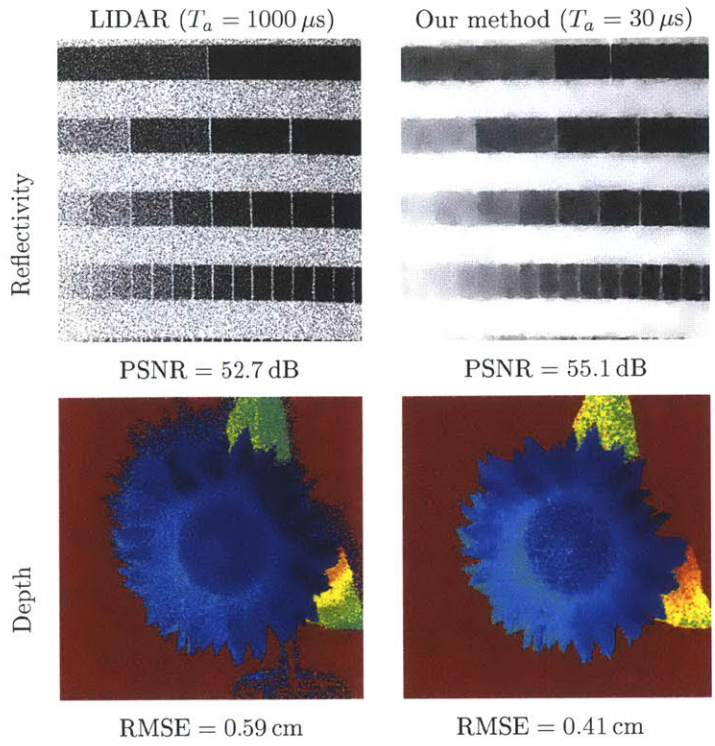


Figure 6 5: **Comparison between our framework and conventional LIDAR approach.**

Repeatability Test

For each scene, we processed 100 independent datasets and computed the sample RMSE images that approximate $\sqrt{\mathbb{E}[(\alpha(x,y) - \hat{\alpha}(x,y)^{\mathrm{PML}})^2]}$ and $\sqrt{\mathbb{E}[(z(x,y) - \hat{z}(x,y)^{\mathrm{PML}})^2]}$. The pixelwise RMSE images, provided in Fig. 6-6, corroborate the consistent accuracy and high resolution of our computational reflectivity and 3D imager.

Effect of System Parameters

Figure 6-7 shows how the performance of traditional ML and our image-formation methods are affected by changing the acquisition time T_a and the signal-to-background ratio (SBR), defined to be

$$\mathrm{SBR} = \frac{1}{M^2} \sum_{x=1}^{M} \sum_{y=1}^{M} \frac{\alpha(x,y)S}{B}.$$

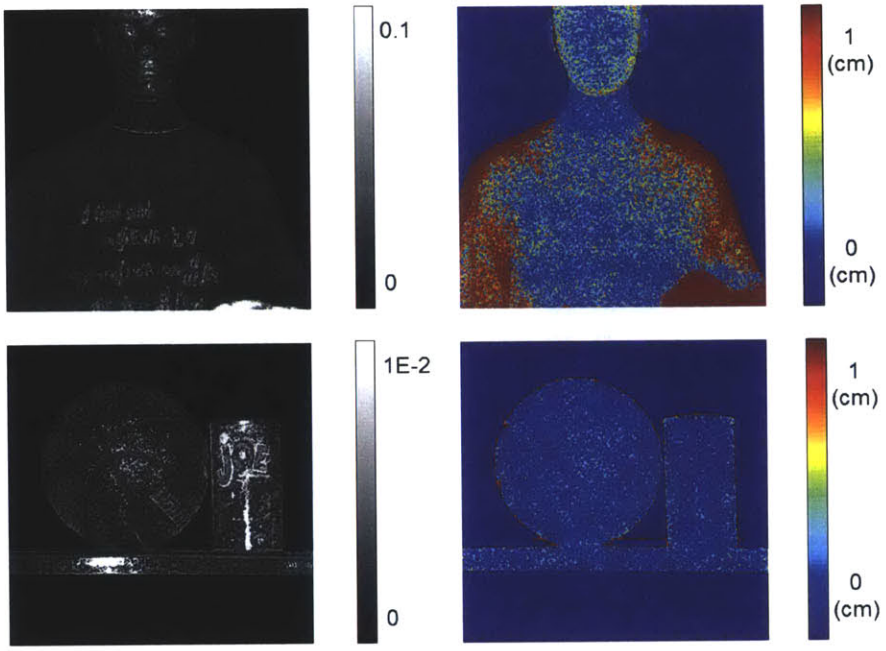
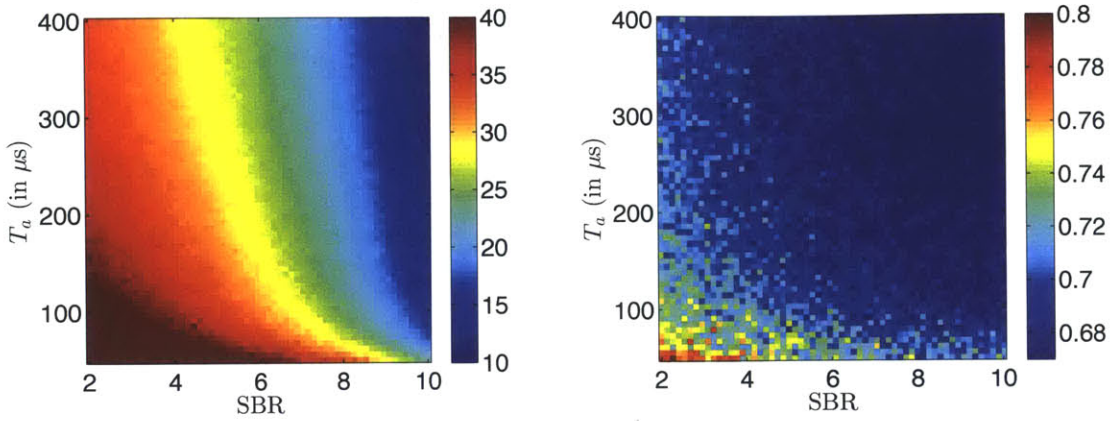


Figure 6-6: **Repeatability test results.** Pixelwise RMSEs for the reflectivity images and depth maps using our method were generated from 100 trials of the experiments.

In our experiment, SBR was modified by changing T_r such that $B = (\eta\, b_\lambda + d)T_r$ is varied at constant S. To obtain the results reported in Fig. 6-7, SBR was modified by simulating increases in B through the addition of pseudorandom detections at times uniformly distributed over $[0, T_r]$. Figure 6-8 provides additional evidence that our method's RMSE decreases monotonically with increasing T_a and SBR, as one would expect. More importantly, it demonstrates that the fixed dwell time 3D imaging method is robust under strong background noise and short acquisition times.

Comparison with First-Photon Imaging

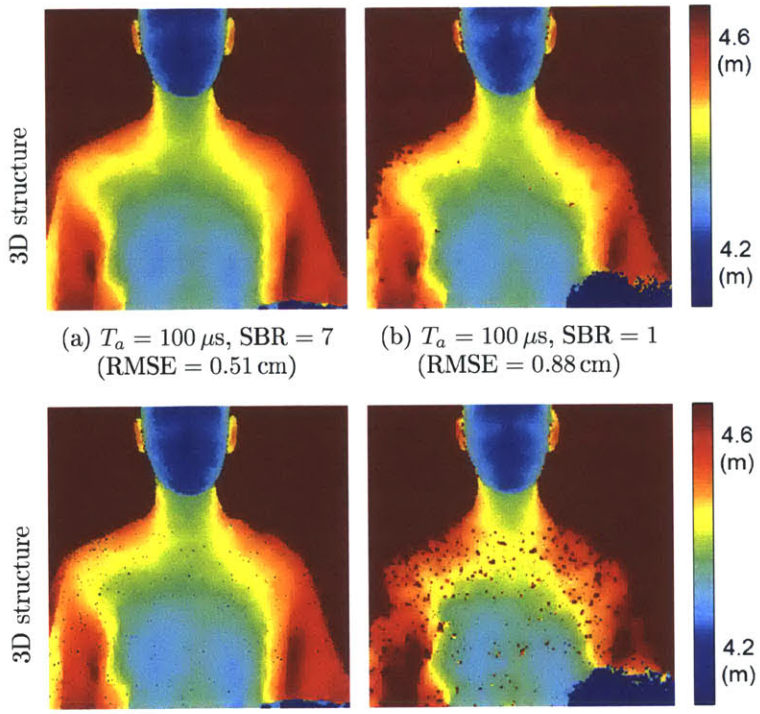
First-photon imaging [7] requires a single detection at each pixel, hence its dwell time on each pixel is a random variable. The method in this chapter requires a fixed dwell time on each pixel, hence its number of detections on each pixel is a random variable. So, to compare the performance of first-photon imaging with that of the fixed dwell time method, we set the average per pixel dwell time of the former equal to the fixed per pixel dwell time of the latter. That comparison, shown in Table 6.1, between the PSNRs of their reflectivity images and



(a) RMSE (in cm) of pointwise ML

(b) RMSE (in cm) of our method

Figure 6-7: **RMSE results for 3D imaging.** Signal-to-background ratio (SBR) was varied by simulating background levels on the ground-truth mannequin dataset. Note the differences in the colorbar scales.



(a) $T_a = 100\,\mu$s, SBR $= 7$ (RMSE $= 0.51$ cm)

(b) $T_a = 100\,\mu$s, SBR $= 1$ (RMSE $= 0.88$ cm)

(c) $T_a = 50\,\mu$s, SBR $= 7$ (RMSE $= 5.80$ cm)

(d) $T_a = 50\,\mu$s, SBR $= 1$ (RMSE $= 11.91$ cm)

Figure 6-8: **Effect of acquisition time T_a and signal-to-background ratio (SBR) on our 3D recovery method.** For acquisition times of $100\,\mu$s and $50\,\mu$s, we calculated the mean photon count $k(x, y)$ over all pixels to be 1.4 and 0.6, respectively.

		first-photon imaging	Ours
Mannequin	Mean T_a	244 μs	244 μs
	Mean $k(x,y)$	1 ppp	2.7 ppp
	Pixels missing data	0%	33%
	PSNR	35 dB	37 dB
	RMSE	0.4 cm	0.3 cm
Sunflower	Mean T_a	15 μs	15 μs
	Mean $k(x,y)$	1 ppp	8.7 ppp
	Pixels missing data	0%	18%
	PSNR	47 dB	47 dB
	RMSE	0.8 cm	0.5 cm
Basketball and can	Mean T_a	181 μs	181 μs
	Mean $k(x,y)$	1 ppp	1.7 ppp
	Pixels missing data	0%	24%
	PSNR	44 dB	45 dB
	RMSE	1.1 cm	1.1 cm
Reflectivity chart	Mean T_a	120 μs	120 μs
	Mean $k(x,y)$	1 ppp	1.7 ppp
	Pixels missing data	0%	27%
	PSNR	54 dB	56 dB
Depth chart	Mean T_a	6.2 μs	6.2 μs
	Mean $k(x,y)$	1 ppp	1.1 ppp
	Pixels missing data	0%	35%
	RMSE	0.4 cm	0.4 cm

Table 6.1: **Comparison between first-photon imaging and fixed dwell time imaging framework.** Note that $k(x,y)$ is fixed and T_a per pixel is a random variable for first photon imaging, whereas $k(x,y)$ is a random variable and T_a per pixel is fixed for the fixed dwell time imaging framework.

the RMSEs of their depth images, reveals several interesting characteristics. In particular, when the fixed dwell time method's image-acquisition time is matched to that of first-photon imaging, a substantial fraction of its pixels have missing data (no detections). Nevertheless, the fixed dwell time method successfully deals with this problem and yields performance similar to, or slightly better than, that of first-photon imaging for the five different scenes we have measured.

6.7 Discussion and Limitations

We have extended the first-photon imaging framework from [7] which has a random per-pixel dwell time, because it records exactly one detection for each pixel in the scene to one that has a fixed dwell time per pixel, but records a random number of detections for each pixel in the scene. Both systems combine physically accurate single-photon detection statistics with exploitation of the spatial correlations found in natural scenes. The new fixed dwell time method, unlike first-photon imaging, is compatible with detector arrays. Hence it is significant that we demonstrated its ability to produce accurate reflectivity and depth images using on the order of 1 detected photon per pixel averaged over the scene, even with significant background light and a substantial fraction of the pixels having no detections. This highly photon-efficient performance motivates the development of accurate and low-power SPAD array-based 3D and reflectivity imagers. Current commercial CMOS-based depth imagers, for example Kinect and TOF cameras, have significantly impacted research in 3D imaging. These sensors offer high depth resolution, but their use is limited due to poor spatial resolution and high power consumption. Our approach offers a potential route to solving these problems.

More generally, the fixed dwell time imaging framework can be used in a variety of low light-level imaging applications using photon-counting detectors, such as spatially-resolved fluorescence lifetime imaging (FLIM) [116] and high-resolution LIDAR [94]. It naturally extends to imaging at a variety of wavelengths, making it suitable for practical implementations. Furthermore, future advances in optoelectronic methods can improve the accuracy of this 3D and reflectivity imager. In particular, it can benefit from improved background suppression techniques [90] and range-gating methods [17].

The fixed dwell time computational imager inherits most of the limitations from the first-photon computational imager. These limitations include, errors due to light multipath, high reconstruction error around object boundaries and in regions of low-reflectivity, potential degradation in imaging quality due to spatially-varying background light, and the inability to resolve object features that are comparable to a pixel size.

In addition to the aforementioned limitations, several considerations need to be addressed

in order for the fixed dwell time imager to be useful in practical imaging scenarios. In the implementation of the proposed imager using a 2D SPAD array and floodlight illumination, the system would have to deal with optical non-idealities such as non-uniform illumination, lens vignetting and radial distortion.

Also, the performance of the fixed dwell time computational imager will also be affected by sensor non-idealities such as, long sensor reset-time or dead time which follows each photon detection, variations in quantum efficiency across the SPAD array, dead pixels, optoelectronic cross-talk between the elements of the SPAD array. These effects did not exist in the first-photon computational imaging framework since we were employing a single omnidirectional sensor and raster-scanned illumination to only detect the first photon.

With proper optoelectronic characterization of the SPAD array using careful calibration procedures, it may be possible to include the aforementioned effects in the theoretical models proposed in this chapter and improve the practical performance of the fixed dwell time computational imager.

Chapter 7

Closing Discussion

This thesis introduced three different computational imaging frameworks in which 3D and reflectivity images of a scene were obtained through computational processing of time-samples measured in response to time-varying illumination. Existing high speed cameras employ fast shutter speeds to capture fast moving objects, and time-of-flight sensors use time-resolved techniques to estimate scene depth by measuring the time delay due to roundtrip light propagation. In contrast to these existing imagers, the key contribution of this thesis is to use the temporal information contained in light signals to create new sensing modalities and form images in challenging scenarios in which traditional methods would fail to form even degraded imagery.

As described in Chapter 2, the first demonstration of the computational imaging framework outlined in Fig. 1-2 was to form images of hidden scenes that were completely occluded from the light source and the sensor, using only the time-resolved detection of light scattered by a Lambertian diffuser. Although this looking around corners framework required *a priori* knowledge of the complete scene geometry, the constructed hidden-plane images produced the seemingly-magical effect of using a Lambertian surface as if it were a mirror. Another challenging problem was to form a depth map of a fronto-planar scene in an unknown configuration, only using a small number of measurements obtained using a single time-resolved bucket detector. As discussed in Chapter 3, this was achieved using spatially-patterned pulse illumination of the scene combined with parametric signal modeling and estimation of the

scene response. In Chapter 5, we demonstrated that it is possible to form high quality scene 3D and reflectivity images from just one detected photon at each sensor pixel. This was accomplished by combining the physics of low-light level photodetection with the exploitation of spatial correlations present in real-world objects. Chapter 6 extended the overarching principles from Chapter 5's first-photon imager to produce similar high-quality 3D and reflectivity images using a fixed per-pixel dwell time and an average of ~ 1 detected photon per pixel. Such an approach makes high-photon efficiency imaging possible with detector arrays.

Computational time-resolved imaging is a rich area for both exploration of theoretical foundations and development of practical optical imaging systems and applications. Both theory and practice should incorporate key optophysical details such as diffraction, lens distortion, and device realities such as limitations on illumination source power and pulse width constraints, sensor non-idealities such as reset time in avalanche detectors, sensor speed and jitter, thermal noise and dark counts, and shot noise due to background light.

The mathematical frameworks and proof-of-concept experiments introduced in this thesis show radically new capabilities, but they are far from the end of the story. Rather, they motivate us to work toward a complete theory while they also demonstrate that computational time-resolved imaging can have practical technological impact. Of course, this thesis does not suggest that standard, inexpensive cameras will be displaced, but in the area of active optical 3D acquisition, in which sensor cost and optical power are two key trade-offs, computational time-resolved imaging may become part of competitive solutions.

Appendix A

Derivations: First Photon Imaging

A.1 Pointwise Maximum Likelihood Reflectivity Estimate

At each pixel, the maximum-likelihood estimate for that location's reflectivity, $\alpha(x, y)$, is obtained by finding the reflectivity value that maximizes the likelihood (Equation (5.5)) or, equivalently, the logarithm of the likelihood given the pulse count data $n(x, y)$, i.e.,

$$\begin{aligned}
\hat{\alpha}_{geo}^{\mathrm{CML}}(x,y) &= \underset{\alpha:\,\alpha\geq 0}{\arg\max} \quad \log\left\{e^{-(\alpha S+B)\,[n(x,y)-1]}\,[1-e^{-(\alpha S+B)}]\right\} \\
&\approx \underset{\alpha:\,\alpha\geq 0}{\arg\max} \quad \log\left\{e^{-(\alpha S+B)\,[n(x,y)-1]}\,[(\alpha S+B)]\right\} \qquad \text{(A.1)} \\
&= \underset{\alpha:\,\alpha\geq 0}{\arg\max} \quad -\alpha S\,[n(x,y)-1] + \log(\alpha S+B) \\
&= \underset{\alpha:\,\alpha\geq 0}{\arg\min} \quad \alpha S\,[n(x,y)-1] - \log(\alpha S+B) \qquad \text{(A.2)}
\end{aligned}$$

Equation (A.1) uses the leading term in its Taylor series to approximate $1-e^{-(\alpha S+B)}$, which is valid because $(\alpha S+B) \ll 1$. The objective function defined in (A.2) will be shown below to be strictly convex. The solution to the optimization problem is the $\alpha(x, y)$ value at which the objective function's derivative vanishes, unless that stationary-point value is negative. In the latter eventuality we set $\hat{\alpha}_{geo}^{\mathrm{CML}}(x, y) = 0$, because of the non-negativity constraint.

This computation yields

$$\hat{\alpha}_{\mathrm{ML}}(x,y) = \max\left\{\frac{1}{(n(x,y)-1)\,S} - \frac{B}{S},\, 0\right\}.$$

A.2 Derivation of Depth Estimation Error

The pointwise depth estimate is $\hat{z}(x,y) = c\,(t(x,y) - T_m)/2$, where T_m is the mode of the normalized pulse shape. In our experiments, each photon detection is either due to backreflected laser light (the signal) or to background light. The conditional probability density functions for the first-photon's arrival time, $t(x,y)$, are

$$\begin{aligned} f_{T(x,y)|\mathrm{signal}}(t(x,y)) &= \eta\, s(t(x,y) - 2\,z(x,y)/c)/S, \quad \text{for } 0 \le t(x,y) < T_r \\ f_{T(x,y)|\mathrm{background}}(t(x,y)) &= 1/T_r, \quad \text{for } 0 \le t(x,y) < T_r. \end{aligned}$$

The variance of the pointwise estimate $\hat{z}(x,y)$ is at least as large as the conditional variance of $\hat{z}(x,y)$ given knowledge of whether a detected photon is due to signal or background. Using the distribution of $t(x,y)$ for detection of signal photons, we get

$$\mathrm{var}\,(\hat{z}(x,y) \mid \text{ signal}) = \frac{c^2}{4}\,\mathrm{var}(t(x,y)) = \frac{c^2 T_p^2}{4},$$

where we have used $T_p \ll T_r$. Using the distribution of $t(x,y)$ for detection of background photons, we get

$$\mathrm{var}\,(\hat{z}(x,y) \mid \text{ background}) = \frac{1}{12}\left(\frac{cT_r}{2}\right)^2.$$

Because in our experiments detections that are due to signal and background occur with approximately equal probability, the unconditional variance of $\hat{z}(x,y)$ is

$$\frac{1}{2}\,\frac{c^2 T_p^2}{4} + \frac{1}{2}\,\frac{c^2 T_r^2}{48}.$$

Taking the square root gives the unconditional standard deviation, which is the RMS

error of pointwise range estimation because the estimator is approximately unbiased.

$$\frac{c}{2}\sqrt{\frac{1}{2}\left(T_p^2+\frac{T_r^2}{12}\right)}.$$

A.3 Derivation of Signal Photon Time-of-arrival Probability Distribution

Suppose that the laser pulse launched at $t = 0$ interrogates spatial location (x, y) and that the resulting backreflected laser light leads to a first-photon detection in the interval $0 \leq t < T_r$. Consider an incremental time interval of duration δt starting at time $\tau \in [0, T_r)$. Using time-inhomogeneous Poisson photon-counting statistics, we obtain the following probability:

$$\begin{aligned}
&\Pr[\text{first photon was detected at } t \in [\tau, \tau+\delta) \text{ and was due to signal}] \\
&= \Pr[\text{no photons detected in } t \in [0,\tau)] \times \Pr[\text{signal photon detected in } t \in [\tau, \tau+\delta t)] \\
&\quad \times \Pr[\text{no background photon detected in } t \in [\tau, \tau+\delta t)] \\
&= \left[\int_\tau^{\tau+\delta t} \eta\,\alpha(x,y)\,s(t-2z(x,y)/c)\,\mathrm{d}t\right] \exp\left[-\int_0^{\tau+\delta t} [\eta\,\alpha(x,y)\,s(t-2z(x,y)/c) + B/T_r]\,\mathrm{d}t\right],
\end{aligned}$$

where we used the fact that Poisson processes have at most one count in an incremental time interval. Defining

$$j(\tau) = \lim_{\delta t \to 0} \frac{\Pr[\text{first photon was detected at } t \in [\tau, \tau+\delta) \text{ and was due to signal}]}{\delta t},$$

and noticing that τ is a dummy variable and can be interchanged with $t(x, y)$, we obtain the desired conditional probability density as follows,

$$
\begin{aligned}
f_{T(x,y)|\text{signal}}(t(x,y)) &= \frac{j(t(x,y))}{\int_0^{T_r} j(t)\,\mathrm{d}t} \\
&= \frac{s(t(x,y) - 2z(x,y)/c)\,\exp\left[-\int_0^{t(x,y)} [\eta\,\alpha(x,y)\,s(\tau - 2z(x,y)/c) + B/T_r]\,\mathrm{d}\tau\right]}{\int_0^{T_r} s(t - 2z(x,y)/c)\,\exp\left[-\int_0^{t} [\eta\,\alpha(x,y)\,s(\tau - 2z(x,y)/c) + B/T_r]\,\mathrm{d}\tau\right]\,\mathrm{d}t} \\
&= \eta\,s(t(x,y) - 2z(x,y)/c)/S, \quad \text{for } 0 \le t(x,y) < T_r,
\end{aligned}
$$

where the approximation is valid under the low-flux condition, $(\alpha(x,y)S + B) \ll 1$. Our computational imager's censoring process (step 2) is sufficiently good that it is safe to process uncensored arrival times as though they were due to signal-photon detections.

A.4 Derivation of Background Photon Time-of-arrival Probability Distribution

Suppose that the laser pulse launched at $t = 0$ interrogates spatial location (x,y) but background light is responsible for the first-photon detection in the interval $0 \le t < T_r$. Consider an incremental time interval of duration δt starting at time $\tau \in [0, T_r)$. Using time-inhomogeneous Poisson photon-counting statistics, we obtain the following probability:

$$
\begin{aligned}
&\Pr[\text{first photon was detected at } t \in [\tau, \tau+\delta) \text{ and was due to background}] \\
&= \Pr[\text{no photons detected in } t \in [0,\tau)] \times \Pr[\text{background photon detected in } t \in [\tau, \tau+\delta t)] \\
&\quad \times \Pr[\text{no signal photon detected in } t \in [\tau, \tau+\delta t)] \\
&= \frac{B}{T_r}\,\delta t \exp\left[-\int_0^{\tau+\delta t} [\eta\,\alpha(x,y)\,s(t - 2z(x,y)/c) + B/T_r]\,\mathrm{d}t\right],
\end{aligned}
$$

where we used the fact that Poisson processes have at most one count in an incremental time interval. Defining

$$
j'(\tau) = \lim_{\delta t \to 0} \frac{\Pr[\text{first photon was detected at } t \in [\tau, \tau+\delta) \text{ and was due to background}]}{\delta t},
$$

the desired conditional probability density then follows from

$$
\begin{aligned}
f_{T(x,y)|\text{background}}(t(x,y)) &= \frac{j'(t(x,y))}{\int_0^{T_r} j'(t)\,\mathrm{d}t} \\
&= \frac{\exp\left[-\int_0^{t(x,y)} \left[\eta\,\alpha(x,y)\,s(\tau - 2z(x,y)/c) + B/T_r\right]\mathrm{d}\tau\right]}{\int_0^{T_r} \exp\left[-\int_0^{t} \left[\eta\,\alpha(x,y)\,s(\tau - 2z(x,y)/c) + B/T_r\right]\mathrm{d}\tau\right]\,\mathrm{d}t} \\
&= \frac{1}{T_r}, \quad \text{for } 0 \le t(x,y) < T_r,
\end{aligned}
$$

where the approximation is valid under the low-flux condition, $(\alpha(x,y)S + B) \ll 1$.

A.5 Proof of Convexity of Negative Log-likelihoods

The negative log-likelihood function for reflectivity estimation, $\mathcal{L}_\alpha(\alpha(x,y); n(x,y))$, is the objective function in Equation (A.2):

$$
\mathcal{L}(\alpha(x,y); n(x,y)) = [\alpha(x,y)\,S + B]\,[n(x,y) - 1] - \log[(\alpha(x,y)S + B)]
$$

The second derivative of the likelihood function $\mathcal{L}_\alpha(\alpha(x,y); n(x,y))$ with respect to the reflectivity $\alpha(x,y)$ is, $S^2/(\alpha(x,y)\,S+B)^2 > 0$, confirming the strict convexity of $\mathcal{L}_\alpha(\alpha(x,y); n(x,y))$ with respect to reflectivity. Figure A-1 shows how the negative log-likelihood function changes as the background illumination power B is varied.

The negative log-likelihood function for range estimation, $\mathcal{L}_z(z(x,y); t(x,y))$, is derived using Equations (5.6) and (5.13):

$$
\begin{aligned}
\mathcal{L}_z(z(x,y); t(x,y)) &= -\log\, f_{T(x,y)|\text{signal}}(t(x,y)) \\
&= -\log\,[\eta\,s(t(x,y) - 2z(x,y)/c)/S] \\
&= \frac{(\tau - T_s - 2z(x,y)/c)}{T_c} - 4\log(t(x,y) - T_s - 2z(x,y)/c).
\end{aligned}
$$

The second derivative of $\mathcal{L}_z(z(x,y); t(x,y))$ with respect to scene depth $z(x,y)$ is, $4/c^2(\tau -$

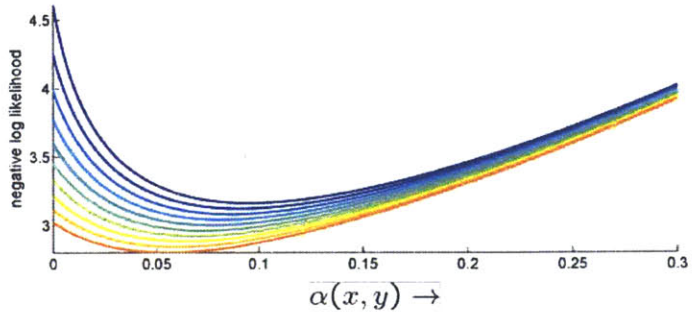


Figure A 1: $\mathcal{L}_\alpha(\alpha(x,y); n(x,y))$ vs. $\alpha(x,y)$ when $n_{ij} = 10$ and $S = 1$ for several values of B. B ranges from 0.01 (blue) to 0.05 (red). Note that the global minimum of negative log likelihood shifts as B changes.

$T_s - 2z(x,y)/c)^2 > 0$, for all values of $z(x,y)$. This fact confirms the strict convexity of $\mathcal{L}_z(z(x,y); t(x,y))$ with respect to range.

We additionally note that in general if the illumination waveform $s(t)$ is log-concave. then the negative log-likelihood function for range estimation is a convex function as well. For example, choosing the pulse to be in the family of generalized Gaussian distributions such that $s(t) \propto e^{-(|t|/a)^p}$, where $p > 1$ and $a > 0$, leads to a convex optimization problem for regularized maximum likelihood estimation. Figure A-2 shows the negative log-likelihood functions of generalized Gaussian distributions, which are log-concave, and the resulting negative log-likelihood functions which are convex.

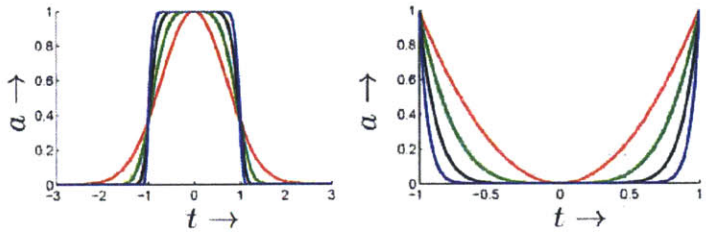


Figure A 2: **Left:** Plot of generalized Gaussian functions with $p = 2$ (red), 3 (green), 4 (black), 5 (blue) with fixed $a = 1$ and amplitude 1. The generalized Gaussian function includes the Gaussian function ($p = 2$) and the square function ($p \rightarrow \infty$). **Right:** Plot of negative log generalized Gaussian functions for the same p, a values.

Appendix B

Derivations: Photon Efficient Imaging with Sensor Arrays

This appendix provides performance analyses for pixelwise estimation. The Cramér-Rao lower bound (CRLB) sets the limit on the mean-square error (MSE) of an unbiased estimator of a parameter. Let x be a scalar continuous parameter in the probability density function $f_Y(y;x)$ of random variable Y. The CRLB for an unbiased estimator, $\hat{x}$, of the parameter x based on observation of Y is the inverse of the Fisher information $J(x)$ [102]:

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

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

B.1 Mean-square Error of Reflectivity Estimation

With some algebra, the CRLB for estimating the reflectivity, $\alpha(x,y)$, at pixel (x,y) can be shown to be

$$\begin{aligned}
\mathrm{CRLB}\left(\alpha(x,y)\right) &= \left\{\mathbb{E}\left[\frac{d^2}{d^2\alpha(x,y)}\left(-\log\Pr[K(x,y)=k;\alpha(x,y)]\right)\right]\right\}^{-1} \\
&= \left\{\mathbb{E}\left[\frac{k\eta^2S^2\exp\left[\eta\alpha(x,y)S+B\right]}{\left(\exp\left[\eta\alpha(x,y)S+B\right]-1\right)^2}\right]\right\}^{-1} \\
&= \frac{\exp\left[\eta\alpha(x,y)S+B\right]-1}{N\eta^2S^2} \\
&\approx \frac{\eta\alpha(x,y)S+B}{N\eta^2S^2},
\end{aligned} \tag{B.2}$$

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

Note, however, that we cannot directly use the CRLB result to lower bound the mean-square error of the unconstrained ML reflectivity estimate $\hat{\alpha}(x,y)^{\mathrm{ML}}$ given by

$$\hat{\alpha}(x,y)^{\mathrm{ML}} = \frac{1}{\eta S}\left[\log\left(\frac{N}{N-k(x,y)}\right)-B\right].$$

This is because the ML estimate is biased, $(\mathbb{E}[\hat{\alpha}^{\mathrm{ML}}(x,y)] \neq \alpha(x,y))$:

$$\begin{aligned}
\mathbb{E}\left[\hat{\alpha}^{\mathrm{ML}}(x,y)\right] &= \mathbb{E}\left[\frac{1}{\eta S}\log\left(\frac{N}{N-k(x,y)}\right)-\frac{B}{\eta S}\right] \\
&= \frac{1}{\eta S}\log N - \frac{1}{\eta S}\mathbb{E}\left[\log\left(N-K(x,y)\right)\right] - \frac{B}{\eta S} \\
&> \frac{1}{\eta S}\log N - \frac{1}{\eta S}\log\left(N-\mathbb{E}[K(x,y)]\right) - \frac{B}{\eta S} \\
&= \alpha(x,y),
\end{aligned}$$

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

When $N \to \infty$ and $\eta\alpha(x,y)S + B \to 0^+$ with $N[1-\exp(\eta\alpha(x,y)S+B)]$ equal to a

constant $C(\alpha(x,y))$, the ML reflectivity estimate is

$$\hat{\alpha}(x,y)^{\mathrm{ML}} = \frac{k}{N\eta S} - \frac{B}{\eta S}. \tag{B.3}$$

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

$$\mathrm{CRLB}(\alpha(x,y)) = \mathbb{E}\left[\left(\alpha(x,y) - \hat{\alpha}(x,y)^{\mathrm{ML}}\right)^2\right] = \frac{1}{N}\left(\frac{\alpha(x,y)}{\eta S} + \frac{B}{\eta^2 S^2}\right),$$

We see that the CRLB expression from the Poisson likelihood is equal to the first-order Taylor expansion of the CRLB expression of the exact binomial likelihood given by Equation (B.2).

Knowing that the ML solution for the limiting Poisson distribution is unbiased and efficient, we conclude that the ML reflectivity estimate $\hat{\alpha}(x,y)^{\mathrm{ML}}$ is efficient asymptotically as $(\eta\alpha(x,y)S + B) \to 0^+$ and $N \to \infty$, with $N[1 - \exp(-(\eta\alpha(x,y)S + B))]$ held constant.

B.2 Mean-Square Error of Depth Estimation

We again assume that $\eta\alpha(x,y)S + B \to 0^+$ and $N \to \infty$ such that $N[1 - \exp(-(\eta\alpha(x,y)S + B))]$ is a constant $C(\alpha(x,y))$. The CRLB for estimating the depth $z(x,y)$ is then

$$\begin{aligned}
\mathrm{CRLB}(z(x,y)) &= \left\{\mathbb{E}\left[\frac{d^2}{d^2 z(x,y)}\left(-\log f_{T(x,y)}(\{t(x,y)^{(\ell)}\}_{\ell=1}^{k(x,y)}; z(x,y))\right)\right]\right\}^{-1} \\
&= \left\{\mathbb{E}\left[-\sum_{\ell=1}^{k(x,y)} \frac{d^2}{d^2 z(x,y)} \log f_{T(x,y)}(t(x,y)^{(\ell)}; z(x,y))\right]\right\}^{-1} \\
&= \frac{1}{C(\alpha(x,y))}\left(\int_0^{T_r} \frac{\dot{p}(t; z(x,y))^2}{p(t; z(x,y))}\, dt\right)^{-1},
\end{aligned} \tag{B.4}$$

where

$$p(t; z(x,y)) = \frac{\lambda(x,y)(t)}{\int_0^{T_r} \lambda(x,y)(\tau)\, d\tau}$$

with $\lambda(x,y)(t)$ being the single-pulse rate from Equation (5.1), and $\dot{p}(t; z(x,y))$ the derivative of $p(t; z(x,y))$ with respect to time.

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

$$\hat{z}^{\mathrm{ML}}(x,y) = \underset{z(x,y)}{\arg\max} \sum_{\ell=1}^{k(x,y)} \log\left[s(t^{(\ell)}(x,y) - 2z(x,y)/c)\right] = \frac{c}{2}\left(\frac{\sum_{\ell=1}^{k(x,y)} t(x,y)^{(\ell)}}{k(x,y)}\right),$$

given $k(x,y) \geq 1$. If $k(x,y) = 0$, then a standard pixelwise data imputation is done by making a uniformly random guess over the interval $[0, cT_r/2)$. Assuming $B = 0$, the MSE expression can be written as

$$\begin{aligned}
\mathbb{E}[(z(x,y) - \hat{z}(x,y)^{\mathrm{ML}})^2] &= \mathbb{E}_{K(x,y)}\{\mathbb{E}[(z(x,y) - \hat{z}(x,y)^{\mathrm{ML}})^2 \mid K(x,y)]\} \\
&= \sum_{k=0}^{\infty} \frac{C^k(\alpha(x,y))e^{-C(\alpha(x,y))}}{k!} \mathbb{E}[(z(x,y) - \hat{z}(x,y)^{\mathrm{ML}})^2 \mid K(x,y) = k] \\
&= e^{-C(\alpha(x,y))}\left[\left(\frac{cT_r}{2}\right)^2 + \left(z(x,y) - \frac{cT_r}{4}\right)^2 \right. \\
&\qquad \left. + \sum_{k=1}^{\infty} \frac{C^k(\alpha(x,y))}{k!} \frac{1}{k}\left(\frac{cT_p}{2}\right)^2\right] \\
&= e^{-C(\alpha(x,y))}\left(\underbrace{\left(\frac{cT_r}{2}\right)^2 + \left(z(x,y) - \frac{cT_r}{4}\right)^2}_{\text{random guess error}} \right. \\
&\qquad \left. + \underbrace{\left(\frac{cT_p}{2}\right)^2 \int_0^{C(\alpha(x,y))} \frac{\exp[\tau] - 1}{\tau}\, d\tau}_{\text{pulse-width error}}\right).
\end{aligned} \tag{B.5}$$

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

Bibliography

[1] L. D. Smullin and G. Fiocco, "Optical echoes from the moon," *Nature*, vol. 194, p. 1267, 1962.

[2] R. Knowlton, "Airborne ladar imaging research testbed," Tech Notes: MIT Lincoln Laboratory, 2011.

[3] S. B. Gokturk, H. Yalcin, and C. Bamji, "A time-of-flight depth sensor system description, issues and solutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog. Workshop*, vol. 3, 2004, p. 35.

[4] H. E. Edgerton and J. R. Killian, Jr., *Flash! Seeing the Unseen by Ultra High-Speed Photography.* Hale, Cushman and Flint, 1939.

[5] A. Kirmani, A. Velten, T. Hutchison, M. E. Lawson, V. K. Goyal, M. Bawendi, and R. Raskar, "Reconstructing an image on a hidden plane using ultrafast imaging of diffuse reflections," 2011, submitted.

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

[7] 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.

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

[9] B. Schwarz, "Lidar: Mapping the world in 3d," *Nat. Photonics*, vol. 4, no. 7, pp. 429 430, 2010.

[10] S. Foix, G. Alenya, and C. Torras, "Lock-in time-of-flight cameras: A survey," *IEEE Sensors J.*, vol. 11, no. 9, pp. 1917 1926, 2011.

[11] M. B. Wakin, J. N. Laska, M. F. Duarte, D. Baron, S. Sarvotham, D. Takhar, K. F. Kelly, and R. G. Baraniuk, "An architecture for compressive imaging," in *Proc. IEEE Int. Conf. Image Process.*, 2006, pp. 1273 1276.

[12] M. F. Duarte, M. A. Davenport, D. Takhar, J. N. Laska, T. Sun, K. Kelly, and R. G. Baraniuk, "Single-pixel imaging via compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 83 91, 2008.

[13] S. Popoff, G. Lerosey, M. Fink, A. C. Boccara, and S. Gigan, "Image transmission through an opaque material," *Nat. Comm.*, vol. 1, p. 81, 2010.

[14] P. Sen, B. Chen, G. Garg, S. R. Marschner, M. Horowitz, M. Levoy, and H. Lensch, "Dual photography," in *ACM Trans. Graph.*, vol. 24, no. 3, 2005, pp. 745 755.

[15] G. Howland, P. Zerom, R. W. Boyd, and J. C. Howell, "Compressive sensing LIDAR for 3D imaging," in *Conf. Lasers and Electro-Optics*. OSA, 2011, p. CMG3.

[16] A. Wallace, G. Buller, and A. Walker, "3d imaging and ranging by time-correlated single photon counting," *IET J. Computing & Control Engineering*, vol. 12, no. 4, pp. 157 168, 2001.

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

[18] P. Andersson, "Long-range three-dimensional imaging using range-gated laser radar images," *Optical Engineering*, vol. 45, no. 3, pp. 034 301 034 301, 2006.

[19] J. C. Hebden, S. R. Arridge, and D. T. Delpy, "Optical imaging in medicine: Experimental techniques," *Physics in Medicine and Biology*, vol. 42, no. 5, p. 825, 1997.

[20] S. R. Arridge, "Optical tomography in medical imaging," *Inverse Problems*, vol. 15, no. 2, p. R41, 1999.

[21] S. Farsiu, J. Christofferson, B. Eriksson, P. Milanfar, B. Friedlander, A. Shakouri, and R. Nowak, "Statistical detection and imaging of objects hidden in turbid media using ballistic photons," *Applied Optics*, vol. 46, no. 23, pp. 5805 5822, 2007.

[22] W. A. Kalender, "X-ray computed tomography," *Physics in Medicine and Biology*, vol. 51, no. 13, p. R29, 2006.

[23] S. Nag, M. A. Barnes, T. Payment, and G. Holladay, "Ultrawideband through-wall radar for detecting the motion of people in real time," in *AeroSense 2002*. International Society for Optics and Photonics, 2002, pp. 48 57.

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

[25] A. Kirmani, T. Hutchison, J. Davis, and R. Raskar, "Looking around the corner using transient imaging," in *Proc. 12th Int. Conf. Comput. Vis.* IEEE, 2009, pp. 159 166.

[26] A. Kirmani, "Femtosecond transient imaging," Master's thesis, Massachusetts Institute of Technology, 2010.

[27] A. Kirmani, T. Hutchison, J. Davis, and R. Raskar, "Looking around the corner using ultrafast transient imaging," *International Journal of Computer Vision*, vol. 95, no. 1, pp. 13 28, 2011.

[28] L. J. Cutrona, "Synthetic aperture radar," in *Radar Handbook*, M. I. Skolnik, Ed. New York, NY: McGraw Hill, 1990, ch. 21.

[29] J. J. Kovaly, *Synthetic Aperture Radar.* Dedham, MA: Artech House, 1976.

[30] D. C. Munson, Jr., J. D. O'Brien, and W. K. Jenkins, "A topographic formulation of spotlight-mode synthetic aperture radar," *Proc. IEEE*, vol. 71, no. 8, pp. 917 925, Aug. 1985.

[31] M. Oren and S. K. Nayar, "Generalization of the Lambertian model and implications for machine vision," *International Journal of Computer Vision*, vol. 14, no. 3, pp. 227 251, Apr. 1995.

[32] U. Wandinger, M. McCormick, C. Weitkamp *et al.*, *Lidar: Range-Resolved Optical Remote Sensing of the Atmosphere.* Springer, 2005, vol. 1.

[33] E. Repasi, P. Lutzmann, O. Steinvall, M. Elmqvist, B. Göhler, and G. Anstett, "Advanced short-wavelength infrared range-gated imaging for ground applications in monostatic and bistatic configurations," *Applied Optics*, vol. 48, no. 31, pp. 5956 5969, 2009.

[34] E. McLean, H. Burris Jr, M. Strand *et al.*, "Short-pulse range-gated optical imaging in turbid water," *Applied Optics*, vol. 34, no. 21, pp. 4343 4351, 1995.

[35] B. E. Roth, K. C. Slatton, and M. J. Cohen, "On the potential for high-resolution lidar to improve rainfall interception estimates in forest ecosystems," *Frontiers in Ecology and the Environment*, vol. 5, no. 8, pp. 421 428, 2007.

[36] K. J. Dana, B. Van Ginneken, S. K. Nayar, and J. J. Koenderink, "Reflectance and texture of real-world surfaces," *ACM Transactions on Graphics*, vol. 18, no. 1, pp. 1 34, 1999.

[37] M. Unser, "Sampling 50 years after Shannon," *Proc. IEEE*, vol. 88, no. 4, pp. 569 587, Apr. 2000.

[38] P. C. Hansen, *Rank-deficient and Discrete Ill-posed Problems: Numerical Aspects of Linear Inversion.* SIAM, 1998, vol. 4.

[39] S. Mallat, *A Wavelet Tour of Signal Processing.* Acad. Press, 1999, vol. 1.

[40] A. Kirmani, H. Jeelani, V. Montazerhodjat, and V. K. Goyal, "Diffuse imaging: Creating optical images with unfocused time-resolved illumination and sensing," *IEEE Signal Process. Lett.*, vol. 19, no. 1, pp. 31 34, 2012.

[41] , "Diffuse imaging: Replacing lenses and mirrors with omnitemporal cameras," in *Proc. SPIE Wavelets & Sparsity XIV*, San Diego, CA, Aug. 2011.

[42] K. Carlsson, P. E. Danielsson, R. Lenz, A. Liljeborg, L. Majlöf, and N. Åslund, "Three-dimensional microscopy using a confocal laser scanning microscope," *Opt. Lett.*, vol. 10, no. 2, pp. 53 55, 1985.

[43] J. Sharpe, U. Ahlgren, P. Perry, B. Hill, A. Ross, J. Hecksher-Sørensen, R. Baldock, and D. Davidson, "Optical projection tomography as a tool for 3d microscopy and gene expression studies," *Science*, vol. 296, no. 5567, pp. 541 545, 2002.

[44] A. Wehr and U. Lohr, "Airborne laser scanning an introduction and overview," *ISPRC J. Photogrammetry & Remote Sensing*, vol. 54, no. 2 3, pp. 68 82, 1999.

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

[46] S. M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski, "A comparison and evaluation of multi-view stereo reconstruction algorithms," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, vol. 1, 2006, pp. 519 528.

[47] S. Hussmann, T. Ringbeck, and B. Hagebeuker, "A performance review of 3D TOF vision systems in comparison to stereo vision systems," in *Stereo Vision*, A. Bhatti, Ed. InTech, 2008, pp. 103 120.

[48] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *International Journal of Computer Vision*, vol. 47, no. 1 3, pp. 7 42, 2002.

[49] A. P. Cracknell and L. W. B. Hayes, *Introduction to Remote Sensing.* London, UK: Taylor & Francis, 1991, vol. 1.

[50] M. A. Albota, R. M. Heinrichs, D. G. Kocher, D. G. Fouche, B. E. Player, M. E. OBrien, B. F. Aull, J. J. Zayhowski, J. Mooney, B. C. Willard *et al.*, "Three-dimensional imaging laser radar with a photon-counting avalanche photodiode array and microchip laser," *Lincoln Laboratory Journal*, vol. 13, no. 2, pp. 351 370, 2002.

[51] A. Medina, F. Gayá, and F. Del Pozo, "Compact laser radar and three-dimensional camera," *J. Opt. Soc. Amer. A.*, vol. 23, no. 4, pp. 800 805, 2006.

[52] Kinect, "Microsoft Kinect Sensor," www.xbox.com/en-US/kinect.

[53] G. J. Iddan and G. Yahav, "Three-dimensional imaging in the studio and elsewhere," in *Photonics West 2001-Electronic Imaging.* SPIE, 2001, pp. 48 55.

[54] R. Lange and P. Seitz, "Solid-state time-of-flight range camera," *IEEE Journal of Quantum Electronics*, vol. 37, no. 3, pp. 390 397, 2001.

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

[56] J. Mei, A. Kirmani, A. Colaço, and V. K. Goyal, "Phase unwrapping and denoising for time-of-flight imaging using generalized approximate message passing," in *Proc. IEEE Int. Conf. Image Proc.*, 2013.

[57] A. Kirmani, A. Benedetti, and P. A. Chou, "Spumic: Simultaneous phase unwrapping and multipath interference cancellation in time-of-flight cameras using spectral methods," in *Proc. IEEE Int. Conf. on Multimedia and Expo (ICME)*, 2013.

[58] M. Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing.* Springer, 2010, vol. 1.

[59] E. J. Candès and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *IEEE Trans. Inform. Theory*, vol. 52, no. 12, pp. 5406 5425, 2006.

[60] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inform. Theory*, vol. 52, no. 4, pp. 1289 1306, 2006.

[61] E. Candès and J. Romberg, "l1-magic: Recovery of sparse signals via convex programming," 2005.

[62] E. J. Candès, "The restricted isometry property and its implications for compressed sensing," *Comptes Rendus Mathematique*, vol. 346, no. 9-10, pp. 589 592, 2008.

[63] M. Sarkis and K. Diepold, "Depth map compression via compressed sensing," in *Proc. IEEE Int. Conf. Image Process.*, 2009, pp. 737 740.

[64] I. Tošić, B. A. Olshausen, and B. J. Culpepper, "Learning sparse representations of depth," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 5, pp. 941 952, 2011.

[65] A. Kirmani, A. Colaço, F. N. C. Wong, and V. K. Goyal, "Codac: A compressive depth acquisition camera framework," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Proc.*, Kyoto, Japan, Mar. 2012, pp. 3809 3812.

[66] A. Kirmani, A. Colaço, F. N. C. Wong, and V. K. Goyal, "Codac: Compressive depth acquisition using a single time-resolved sensor," in *OSA Computational Optical Sensing and Imaging*, 2012, pp. JW3A 5.

[67] A. Kirmani, A. Colaço, and V. K. Goyal, "SFTI: Space-from-time imaging," in *Emerging Technologies for 3D Video: Creation, Coding, Transmission and Rendering*, F. Dufaux, B. Pesquet-Popescu, and M. Cagnazzo, Eds. Wiley, 2013, ch. 2, pp. 17 36.

[68] A. Colaço, "Compact and low-power computational 3d sensors for gestural input," Ph.D. dissertation, Massachusetts Institute of Technology, 2014.

[69] A. Colaço, A. Kirmani, H. S. Yang, N.-W. Gong, C. Schmandt, and V. K. Goyal, "Mime: compact, low power 3d gesture sensing for interaction with head mounted displays," in *Proceedings of the 26th annual ACM symposium on User interface software and technology (UIST).* ACM, 2013, pp. 227 236.

[70] A. Colaço, A. Kirmani, F. N. C. Wong, and V. K. Goyal, "Codac: compressive depth acquisition using a single time-resolved sensor," in *ACM SIGGRAPH 2012 Talks*, 2012.

[71] P. Stoica and R. L. Moses, *Spectral Analysis of Signals.* Pearson/Prentice Hall Upper Saddle River, NJ, 2005, vol. 2.

[72] G. C. M. R. de Prony, "Essai éxperimental et analytique: Sur les lois de la dilatabilité de fluides élastique et sur celles de la force expansive de la vapeur de l'alkool, à différentes températures," *J. de l'École Polytechnique*, vol. 1, no. 22, pp. 24 76, 1795.

[73] Y. Hua and T. K. Sarkar, "Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 38, no. 5, pp. 814 824, 1990.

[74] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 1.21," http://cvxr.com/cvx, 2011.

[75] J. A. Cadzow, "Spectral estimation: An overdetermined rational model equation approach," *Proc. IEEE*, vol. 70, no. 9, pp. 907 939, 1982.

[76] B. Friedlander, "A sensitivity analysis of the music algorithm," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 38, no. 10, pp. 1740 1751, 1990.

[77] V. F. Pisarenko, "The retrieval of harmonics from a covariance function," *Geophysical J. Royal Astronomical Soc.*, vol. 33, no. 3, pp. 347 366, 1973.

[78] G. S. Buller and A. Wallace, "Ranging and three-dimensional imaging using time-correlated single-photon counting and point-by-point acquisition," *IEEE J. Sel. Topics Quantum Electronics*, vol. 13, no. 4, pp. 1006 1015, 2007.

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

[80] D. L. Snyder, *Random Point Processes.* Wiley New York, 1975, vol. 1.

[81] B. Saleh, *Photoelectron Statistics: With Applications to Spectroscopy and Optical Communications.* Springer-Verlang, 1978, vol. 1.

[82] R. M. Marino, T. Stephens, R. E. Hatch, J. L. McLaughlin, J. G. Mooney, M. E. O'Brien, G. S. Rowe, J. S. Adams, L. Skelly, R. C. Knowlton *et al.*, "A compact 3d imaging laser radar system using geiger-mode apd arrays: System and measurements," in *AeroSense 2003.* International Society for Optics and Photonics, 2003, pp. 1 15.

[83] N. Savage, "Single-photon counting," *Nat. Photonics*, vol. 3, no. 12, pp. 738 739, 2009.

[84] A. McCarthy, N. J. Krichel, N. R. Gemmell, X. Ren, M. G. Tanner, S. N. Dorenbos, V. Zwiller, R. H. Hadfield, and G. S. Buller, "Kilometer-range, high resolution depth imaging via 1560 nm wavelength single-photon detection," *Opt. Expr.*, vol. 21, no. 7, pp. 8904 8915, Apr. 2013.

[85] 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.

[86] B. A. Olshausen and D. J. Field, "Natural image statistics and efficient coding," *Netw.: Comput. Neural Syst.*, vol. 7, pp. 333 339, 1996.

[87] A. Srivastava, A. B. Lee, E. P. Simoncelli, and S.-C. Zhu, "On advances in statistical modeling of natural images," *J. Math. Imaging Vision*, vol. 18, no. 1, pp. 17 33, Jan. 2003.

[88] J. C. Nicolas, "Applications of low-light imaging to life sciences," *J. Biolumin. Chemilumin.*, vol. 9, no. 3, pp. 139 144, May Jun. 1994.

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

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

[91] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision.* Cambridge university press, 2003, vol. 1.

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

[93] S. B. Kang, J. A. Webb, C. L. Zitnick, and T. Kanade, "A multibaseline stereo system with active illumination and real-time image acquisition," in *Proc. Fifth Int. Conf. Comput. Vis.* IEEE, 1995, pp. 88 93.

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

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

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

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

[98] D. Shin, A. Kirmani, and V. K. Goyal, "Low-rate Poisson intensity estimation using multiplexed imaging," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Proc.*, Vancouver, Canada, May 2013, pp. 1364 1368.

[99] D. Shin, A. Kirmani, V. K. Goyal, and J. H. Shapiro, "Information in a photon: Relating entropy and maximum-likelihood range estimation using single-photon counting detectors," in *Proc. IEEE Int. Conf. Image Process.*, 2013, pp. 83 87.

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

[101] J.-E. Eklund, C. Svensson, and A. Astrom, "VLSI implementation of a focal plane image processor-a realization of the near-sensor image processing concept," *IEEE Trans. Very Large Scale Integration Syst.*, vol. 4, no. 3, pp. 322 335, 1996.

[102] S. M. Kay, *Fundamentals of Statistical Signal Processing.* Prentice Hall, 1998, vol. 1.

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

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

[105] S. P. Boyd and L. Vandenberghe, *Convex Optimization.* Cambridge university press, 2004, vol. 1.

[106] R. Garnett, T. Huegerich, C. Chui, and W. He, "A universal noise removal algorithm with an impulse detector," *IEEE Trans. Image Process.*, vol. 14, no. 11, pp. 1747 1754, 2005.

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

[108] M. Makitalo and A. Foi, "Optimal inversion of the generalized anscombe transformation for poisson-gaussian noise," *IEEE Trans. Image Process.*, vol. 22, no. 1, pp. 91 103, 2013.

[109] D. Shin, "Computational 3d and reflectivity imaging with high photon efficiency," Master's thesis, Massachusetts Institute of Technology, 2014.

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

[111] , "Photon-efficient computational 3d and reflectivity imaging with single-photon detectors," arXiv:1406.1761 [stat.AP], Jun. 2014.

[112] Z. T. Harmany, R. F. Marcia, and R. M. Willett, "This is spiral-tap: Sparse poisson intensity reconstruction algorithms, theory and practice," *IEEE Trans. Image Process.*, vol. 21, no. 3, pp. 1084 1096, 2012.

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

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

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

[116] D. O'Connor and D. Phillips, *Time-correlated Single Photon Counting.* Academic Press, 1984.