FOVEATED DEPTH SENSING

By

JUSTIN FOLDEN

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2024

Life Before Death
Strength Before Weakness
Journey Before Destination
- Brandon Sanderson, The Way of Kings

ACKNOWLEDGEMENTS

First and foremost, I would like to extend my deepest gratitude to Sanjeev (Dr. Sanjeev J. Koppal). Your other worldly patience, and understanding, have been the cornerstones of my journey through this PhD. The path was anything but smooth—strewn with rocks, debris, perilous obstacles, and more than one bottomless pit (∞)—yet , through it all, your support and insight never wavered. I couldn't have asked for a better advisor, and for that, I am genuinely grateful.

To my family, thank you for being my constant source of strength and encouragement. Your unwavering support gave me the fortitude to keep moving forward.

To my friends—Jaime, Ryan, Bekah, Chase, Max, and Mo—you were my anchors through this process. Your humor, friendship, and encouragement were invaluable, and I am forever grateful to each of you for being my rocks along this journey.

I would also like to express my heartfelt thanks to my friends, mentors, and peers in the FOCUS Lab. Each of you has left an indelible mark on my academic journey. Your brilliance and camaraderie made the experience all the more rewarding, and I am truly fortunate to have known and worked with you.

TABLE OF CONTENTS

page

LIST OF TABLES

Tables page

LIST OF FIGURES

Figures page

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

FOVEATED DEPTH SENSING

By

Justin Folden

December 2024

Chair: Dr. Sanjeev J. Koppal
Major: Electrical and Computer Engineering

Foveation, a key characteristic of biological vision systems, enables efficient use of sensing resources by concentrating attention on areas of interest. Drawing inspiration from this evolutionary trait, this dissertation presents novel approaches to foveated depth sensing for improving the efficiency and adaptability of modern sensing systems. Across three studies, we explore foveated algorithms and sensor designs that mimic this biological principle. First, we develop an adaptive LIDAR system that dynamically adjusts its sampling patterns using a MEMS mirror, optimizing resource through sensor fusion and improving depth granularity in regions of interest. Next, we apply foveated sensing to underwater environments, introducing a bistatic confocal LIDAR system that adapts to the challenges of light scattering in turbid water. Finally, we explore spatio-temporal foveation for SPAD-based depth sensing, demonstrating how depth priors can reduce memory and computational requirements without sacrificing accuracy. These contributions highlight the potential of foveated sensing as a transformative approach for the future design of depth sensors, improving efficiency and performance across diverse applications.

CHAPTER 1
INTRODUCTION

1.1 Foveation: Driving Principles and Motivation

Foveation is a key adaptation found in biological vision systems, evolving as a mechanism to optimize resource allocation. In many animals, including humans, the retina is designed with a high concentration of photoreceptors, at its center, while the peripheral regions have fewer of these receptors. This non-uniform distribution is an evolutionary feature rather than a flaw, providing enhanced visual acuity in the line of sight where it is needed most. This arrangement effectively grants the eye a form of super-resolution, maximizing clarity in regions of interest.

Faced with a restricted number of neurons to process visual information, biological systems have evolved to utilize these resources efficiently. By concentrating photoreceptors, and many of the processing units at the center of the visual field, the eye achieves a sharper and more detailed view in areas that are critical for tasks like hunting, communication, and navigating complex environments. This selective focus, along with the adaptation of refined motor control for the eyes, is what allows animals to perform a wide variety of visually demanding tasks without the need to dedicate an excesively large amount of energy demanding nuerons to a uniformly distributed high-resolution sensor.

Inspired by these natural systems, this dissertation explores how foveation can be applied to artificial depth sensing technologies, with the goal of enhancing performance while minimizing resource consumption. By drawing parallels between biological cones and the pixels or laser power in imaging systems such as cameras, LIDAR, and SPAD-based sensors, we show how foveation principles can be leveraged to optimize sensing. Not only does this approach offer a new paradigm for designing imaging sensors, but it also demonstrates how foveation can overcome challenges in areas like depth resolution, memory constraints, and power efficiency.

In this work, we aim to demonstrate how foveation can not only serve as a standalone enhancement but also be combined with other techniques to address specific challenges in depth sensing. From improving resolution in key regions to reducing computational load, the principles of foveation present a promising direction for developing more adaptive and efficient sensing

systems, applicable to fields as diverse as autonomous navigation, underwater exploration, and real-time 3D imaging.

1.2 Contributions and Organization

This dissertation is structured around three primary research contributions, each represented by a standalone paper. Collectively, these works explore foveated depth sensing from different perspectives, demonstrating how adaptive methods can improve sensor performance across a variety of environments.

- **RGB-Guided Foveated Depth Sensing (Chapter 2)**: The first contribution introduces an adaptive LIDAR system using a MEMS mirror to dynamically adjust sampling patterns based on the regions of interest. By integrating deep learning methods for depth completion, the system provides higher resolution in key areas while reducing the overall computational load. This work is particularly suited for small autonomous systems that need flexibility and low power consumption.

- **Foveated Depth Sensing for Challenging Underwater Environments (Chapter 3)**: The second contribution extends foveated sensing techniques into underwater environments. A bistatic confocal LIDAR system, modulating both the transmitter and receiver, is developed to overcome the challenges of light scattering in turbid water. This adaptive system adjusts its sampling to focus on regions of interest, and takes advantage of scattered light, making it highly effective in challenging underwater conditions.

- **Spatio-Temporal Foveated Depth Sensing (Chapter 4)**: The final contribution explores foveation in SPAD-based depth sensing systems. By leveraging depth priors such as monocular depth estimation, this approach dramatically reduces memory and data bottlenecks while maintaining high accuracy in key regions.

CHAPTER 2
RGB GUIDED FOVEATED DEPTH SENSING FOR IMPROVED MONOCULAR ESTIMATION

2.1 Introduction

The combination of active depth sensors with deep learning has impacted many fields, from video games to autonomous cars. With such deployment, vision researchers have started focusing on closing the loop between active sensing and inference—with methods for correcting deficiencies in incomplete and imperfect depth measurements[95, 105], as well as those that help the system decide where to sense next [59, 12].

However, such work is predicated on LIDAR systems that are flexible in the kind of measurements they make. But this capability does not exist in most existing LIDAR hardware, where sampling is done in a set of fixed angles, usually modulated by mechanical motors which do not allow fast changes in sensing direction without causing unacceptable wear-and-tear.

We present a proof-of-concept, adaptive LIDAR platform that can leverage modern vision algorithms. It permits making measurements with different sampling patterns—providing a speed advantage when fewer measurements are made—and is co-located with a color camera to fully realize the benefits of deep depth completion and guided sampling.

2.1.1 Why Adaptive LIDAR?

Unlike most artificial sensors, animal eyes foveate, or distribute resolution where it is needed. This is computationally efficient, since neuronal resources are concentrated on regions of interest. Similarly, we believe that an adaptive LIDAR would be useful on resource-constrained platforms, such as small robots, remote sensing nodes and UAV platforms.

Furthermore, our design uses a MEMS mirror as the scanning optics, which is compact and low-power. In addition, MEMS scanning is faster than mechanical motors, without similar wear-and-tear, and this allows for multiple fovea or regions-of-interest in a scene. Finally, the MEMS mirror is neither limited to coherent illumination, like phase arrays, nor constrained to specific light wavelengths, like photonics-based systems.

To demonstrate depth sensing flexibility, we first train a deep neural network for depth

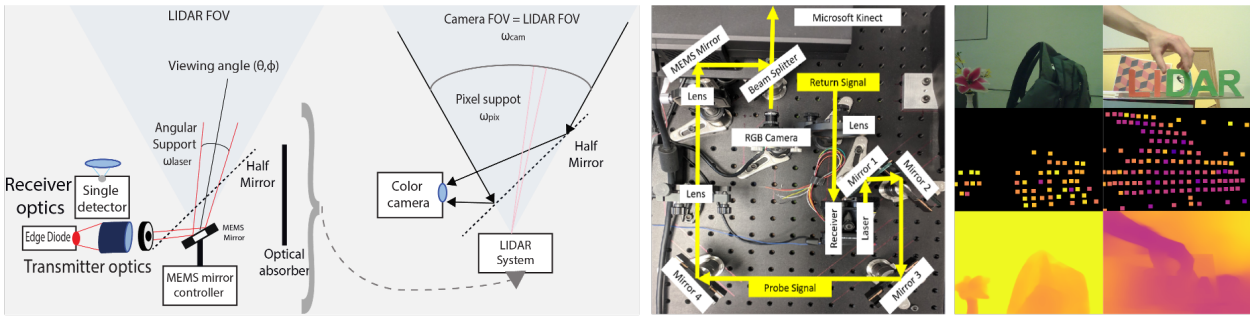


Figure 2-1. Experimental setup: We have designed a flexible MEMS mirror-modulated scanning LIDAR, as shown in (I). In (II), we co-locate this directionally controllable LIDAR with a color camera, allowing for deep depth completion of the sparse LIDAR measurements. In (III) we show a picture of the hardware setup corresponding to (I-II). The long optical path is simply an artifact of having a single circuit board for both the LIDAR receiver and transmitter. In (IV) we show adaptive sampling (middle) and deep depth completion (bottom) results captured with our Adaptive LIDAR Prototype.

completion and show that it delivers high quality scene geometry. We evaluate this with different sampling patterns, including those that are concentrated in a region of interest. Finally, we show vision-driven control of the sensing pattern. In summary, **our contributions** are:

- An adaptive LIDAR prototype (Fig. 2-1) that enables flexible deep depth completion (Fig. 2-4).
- Confirmation that LIDAR foveation improves sensing in areas of interest (Table 2-6).
- Analysis of receiver optics characteristics, particularly the issue of small aperture created by the MEMS mirror (Table A-1).
- First steps towards a working vision-based adaptive LIDAR, showing that real-time foveation is feasible (Fig. 2-5).

2.2 Related Work

Common depth modalities: Many high-quality depth sensors exist today. In Table 2-1 we show qualitative comparisons with these. Our sensor is the first proof-of-concept, real-time, adaptive LIDAR.

Sensor	Technology	Outdoors	Textureless	Adaptive
ELP-960P2CAM	Conventional Passive Stereo	✓	×	×
Kinect v2	Time-of-Flight (LED)	×	✓	×
Intel RealSense	Structured Light Stereo (LED)	✓	✓	×
Velodyne HDL-32E	Time-of-Flight (Laser)	✓	✓	×
Resonance MEMS / Intel L515	Time-of-Flight (Laser)	✓	✓	×
Robosense RS-LiDAR-M1	Solid State Time-of-Flight (Laser)	✓	✓	×
Programmable Light curtains	Adaptive Structured Light	✓	✓	✓
Our sensor	Adaptive LIDAR	✓	✓	✓

Table 2-1. Our Adaptive LIDAR vs. other common modalities: We compare common depth modalities such as stereo [61], Kinect [71], Velodyne [41], Robosense solid state LIDAR and Resonance MEMS sensors [28, 84, 56] such as the Intel L515. Our work is closest to programmable light curtains for flexible, structured light reconstruction [8, 98]. This paper is an alternate research direction with an adaptive LIDAR, rather than a structured light system.

MEMS/Galvo mirrors for vision and graphics: MEMS mirror modulation has been used for structured light [77], displays [53] and sensing [70]. In contrast to these methods, we propose to use angular control to increase sampling in regions of interest as in [90]. While MEMS mirrors have been used in scanning LIDARs, such as from NASA and ARL [28, 84, 56], *these are run at resonance with no control, while we show adaptive MEMS-based sensing.* Such MEMS control has only been shown [54] in toy examples for highly reflective fiducials in both fast 3D tracking and VR applications [65, 64], whereas we show results on real scenes. [80, 19] show a mirror modulated 3D sensor with the potential for flexibility, but without leveraging guided networks, and we discuss the advantages of our novel receiver optics compared to these types of methods. Galvo mirrors are used with active illumination for light-transport [44] and seeing around corners [72]. Our closest work is the use of light curtains for flexible, structured light reconstruction [8, 98]. In contrast, ours is a MEMS-mirror driven LIDAR system with an additional capability of increasing resolution in some region of interest. In this sense, we are the first to extend adaptive control [15, 89, 19], to LIDAR imaging of real dynamic scenes.

Adaptive Scanning Lidars: Commercially available systems from AEye and Robosense are designed to improve lidar-rgb fusion for large systems such as autonomous cars. In contrast, our goal is to impact small autonomous systems and our choice of MEMS mirror modulation and our optical innovations track these goals. [101] propose a progressive pedestrian scanning method

using an actively scanned LIDAR, but results are shown in simulation rather than on a hardware platform. [89] propose directionally controlling the LIDAR scan, but these adaptive results have been shown only for static scenes. In contrast, we show a real-time adaptive LIDAR that works for dynamic scenes.

Guided and Unguided Depth Completion: The impact of deep networks on upsampling and superresolution has been shown on images, disparity/depth maps, active sensor data etc. [9, 22, 62, 60, 95, 79, 49] with a benchmark on the KITTI depth completion dataset [95]. Upgrading from sparse depth samples has been shown [96], and guided upsampling has been used as a proxy for sensor fusion such as the work that has recently been done for single-photon imagers [58] and flash lidar [32]. In contrast, we measure sparse low-power LIDAR depth measurements and we seek to flexibly change the sensor capture characteristics in order to leverage adaptive neural networks such as [59, 12].

2.3 Sensor design

Fig. 2-1 shows our sensor design, which consists of a small aperture, large depth-of-field color camera, optically co-located with a MEMS-modulated LIDAR sensor. If the camera has a FOV of ω_{cam} steradians and a resolution of I pixels, then the average pixel support is $\omega_{pix} = (\omega_{cam}/I)$. If the LIDAR laser's beam divergence is ω_{laser} steradians, then the acuity increase from LIDAR to camera is $(\omega_{laser}/\omega_{pix})$.

The goal of guided depth completion is to extract this potential increase in acuity by using large datasets to complete or upgrade the existing measurements. A flexible LIDAR can leverage such techniques by, for example, measuring depths in regions of interest.

Next, in Sect. 2.3.1, we discuss the MEMS-modulated transmitter optics that enable compact, low-power, fast and flexible controlled scans. The cost, however, is that MEMS mirrors act as a small aperture that reduces the received radiance, when compared to large mirrors such as galvos. In the following Sect. 2.3.2 we model the receiver optical design space, comparing characteristics such as FOV, volume and received radiance.

Design	Volume	FOV	Received Radiance
Retroreflection	$\frac{\pi\ u\ w_o^2}{12}$	= MEMS FOV ω_{mirror}	$\frac{atan(\frac{w_O}{2Z})}{\omega_{\text{laser}} Z tan(\frac{\omega_{\text{laser}}}{2})}$
Receiver array	$u\ A^2$	$\min(2\ atan(\frac{A}{2u}), \omega_{mirror})$	$\frac{1}{2\ Z\ tan(\frac{\omega_{laser}}{2})}$
Single detector *Conventional* $(u \geq f)$ ***Ours*** $(u < f)$	$\frac{\pi\ u\ A^2}{12}$	$\min(2atan(\frac{A(Z-f)\|\|\frac{Zu-fu-fZ}{Z-f}\|\|}{2ufZ}), \omega_{mirror})$	$\frac{1}{4Zatan(\frac{A(Z-f)\|\|\frac{Zu-fu-fZ}{Z-f}\|\|}{2ufZ})tan(\frac{\omega_{\text{laser}}}{2})}$

Table 2-2. Receiver models (please see the appendix for derivations)

2.3.1 MEMS Mirror based Transmitter Optics

The transmitter optics consist of the pulsed light source and MEMS mirror. The LIDAR's beam is steered by the mirror, whose azimuth and elevation are given by changes in control voltages over time, $(\theta(V(t)), \phi(V(t))$ over the MEMS mirror FOV ω_{mirror}. The advantages of MEMS mirrors are compactness and speed, allowing the mirror's scan to cover the entire FOV quickly, or attend to a region of interest given by an adaptive algorithm. The challenge in transmitter optics is to provide a powerful, narrow laser with low beam divergence, given by

$$\omega_{laser} \approx \frac{M^2\ \lambda}{w_o\ \pi} \tag{2-1}$$

where M is a measure of laser beam quality and w_o refers to the radius at the beam waist, which we use a proxy for MEMS mirror size. Previous work has shown MEMS-mirror modulated LIDAR systems across this design space, from high-quality erbium fiber lasers with near-Gaussian profiles, used by [84] where M is almost unity, to low-cost edge-emitting diodes, such as [89] where $M \approx 300$ on the diode's major axis.

Our setup follows the low-cost diode route, with an additional two-lens Keplerian telescope to reduce the beam waist to $6mm$ and an iris to match the MEMS mirror aperture. This is an alternative to using an optical fiber [19].

2.3.2 Receiver Optics Design Tradeoffs

From the previous section, we can denote the transmitter optics design space as a combination of laser quality M and MEMS mirror size w_o, which we write as $\Pi_t = \{M, w_o\}$. Now, we add receiver optics to the design space, which we denote as $\Pi_r = \{n, A, u, f\}$, where n^2

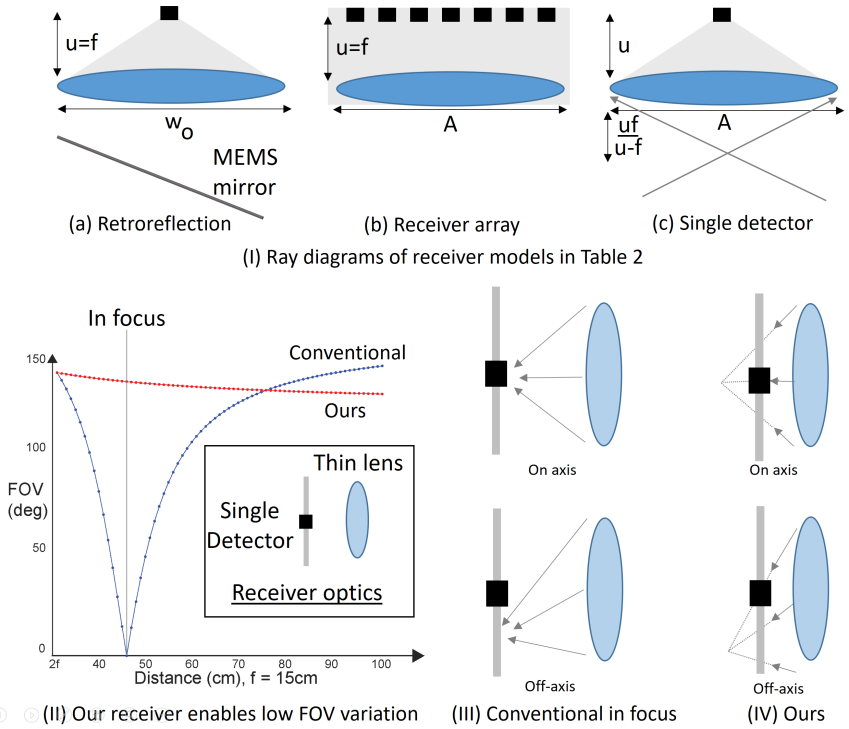


Figure 2-2. Our proposed design vs. other designs: In (I) we depict three common receiver designs, including retro-reflection (a), receiver array (b) and single detectors (c). Our design is a variant of (c), where we suggest a simple optical trick, such that the single detector is placed within the focal distance of the lens. This enables consistent FOV over range, as shown by the red curve in (II) and the designs in (III-IV). Simulations for a $f = 15mm$ unit diameter lens.

is the number of photodetectors in the receiver, A is the aperture, u is the distance between the photodetector array and the receiver optics, and f is the focal length of the receiver optics. Therefore the full design space consists of both receiver/transmitter optics, $\mathbf{\Pi} = \{\Pi_r, \Pi_t\}$.

We define the characterization of any instance within the design space $\mathbf{\Pi}$ as consisting of field-of-view Ω steradians, received radiance s and volume V denoted as $\Xi = \{\Omega, s, V\}$. The range Z is determined by the received radiance and the detector sensitivity. Computing these parameters depends on the design choices made, and we provide simulations comparing three designs: retro-reflective receivers [42] (Fig. 2-2I(a)), receiver arrays [19] (Fig. 2-2I(b)) and single-pixel detectors [89] (Fig. 2-2I(c)).

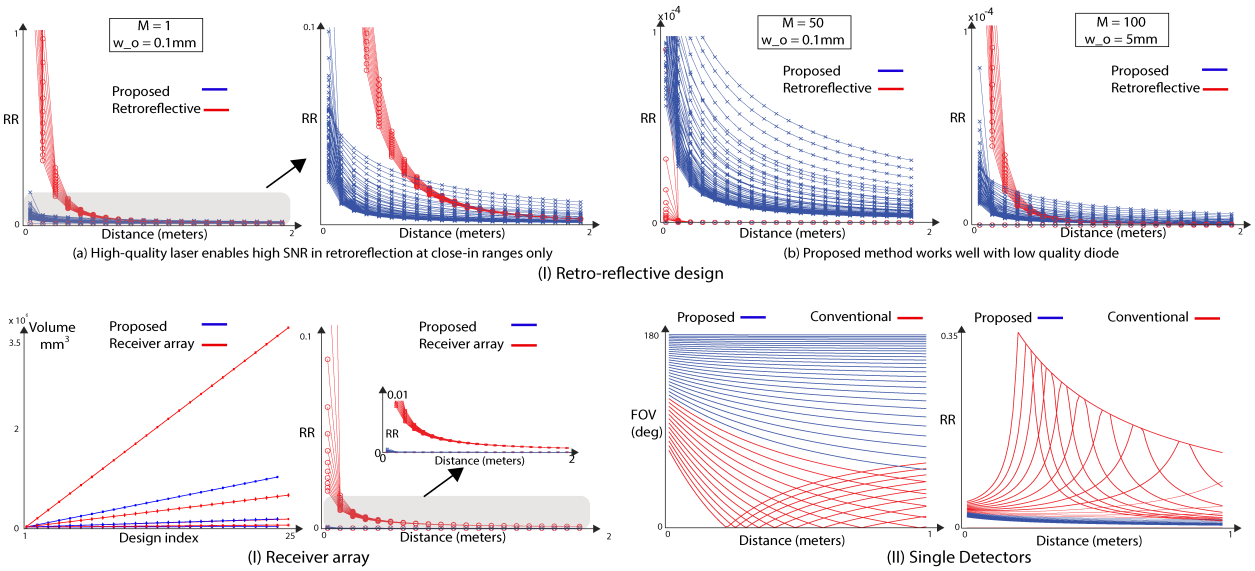


Figure 2-3. Noiseless simulations comparing proposed method with other designs. In (I) we compare the received radiance (RR) of proposed method with retroreflection for different laser qualities and mirror sizes. A high-quality laser (I)(a) enables higher RR for close-in scenes for retroreflective designs, but at large ranges, our method has higher RR. In (II)(a) we show that our proposed design has lower volume than a receiver array, across a wide range of focal lengths, but a receiver array has a higher RR (II)(b), even when compared to the best case for our sensor from (I). In (III) we compare our design with conventional single detectors, for a lens with $f = 15mm$. Although our sensor shows consistent FOV ((III) left), it is always defocused, and faces an RR cost ((III) right).

2.3.3 Simulation Setup and Conclusions

Full derivations for the three receiver designs, shown in Table A-1, are in the appendix. The table refers to receiver sensor volume, field-of-view, and received radiance (normalized for a white Lambertian plane). The *volume* is the convex hull of the opaque baffles that must contain the receiving transducer electronics and is either a cone or cuboid. The *FOV* is the range of angles that the receiver is sensitive to, and is obtained from the defocus kernel, upper-bounded by the MEMS FOV ω_{mirror}. For simplicity, trignometric functions are written to act on steradian quantities, but in actuality act on the apex angle of the equivalent cone. Our definition of received radiance is the area-solid angle product used in optics [66] for a canonical LIDAR transducer, which can be loosely understood as loss of LIDAR laser dot intensity due to beam divergence and receiver aperture size when imaging a fronto-parallel, white Lambertian plane.

In our noiseless simulations, we assumed a geometric model of light. To illustrate the trade-offs, we vary the laser quality between $M = 1$ to $M = 100$, representing an ideal Gaussian beam vs. a cheap laser diode. For the same reason, we vary the MEMS mirror size w_o from $0.1mm$ (10 times larger than the TI DMD [46]) to $5mm$ (a large size for a swiveling MEMS mirror). The range of dimensions over which we explore the receiver design space are of the order of a small camera, with apertures $0cm \leq A \leq 10cm$, focal lengths $0mm \leq f \leq 50mm$ and image plane-lens distances $0mm \leq u \leq 50mm$. In Fig. 2-2(II)-(IV) we describe our proposed, simple modification to the conventional single-pixel receiver, where photodetector is placed on the optical axis, at a distance v larger than the focal length f.

Conclusions: In the next few pages, we discuss, at a high-level, simulations that compare our modification to retro-reflective receivers, receiver arrays and conventional single-pixel receivers. Full derivations are in the appendix. The **conclusion** from these simulations is that our design modification provides a new option for receiver design space tradeoffs. In contrast to existing work on defocusing received radiances for FOV adjustment and amplitude compensation (e.g. [66]), we do not require special optics (e.g. split lens) and we have large off-axis FOV since the MEMS is not the aperture for the receiver. This gives advantages when compared to alternate designs. For example, in volume, our design is smaller than receiver arrays but larger than retro-reflective designs. On the other hand, for received radiance from low-cost laser emitters, this situation is reversed. Here, our design does better than retro-reflective designs but worse than arrays.

2.3.4 Analysis of Sensor Design Tradeoffs

Retro-reflective receivers: If high-quality lasers such as erbium fiber lasers [84] are used, where M is near-unity, then these can be coupled with a co-located receiver and a beamsplitter, as shown in Fig. 2-2I(a), where the detector lens distance is equal to the focal length $u = f$. Consider the second column from Table A-1. The ratio of retro-reflective volume to our sensor's volume is $\frac{w_o}{A}$, which is usually less than one, since MEMS mirrors are small.

In other words, retro-reflective designs are smaller than ours. The small retroreflective design also has the optimal FOV of the MEMS, due to co-location. Our design does have a

received radiance advantage, since retroreflection requires the MEMS mirror to be the aperture for both receiver/transmitter. Fig. 2-3(Ia) shows how this advantage eventually trumps other factors such as laser quality ($M = 1$) or large mirrors. In the extreme case of low-cost diodes, Fig. 2-3(Ib), our sensor has higher received radiance at close ranges too.

Receivers arrays: If cost and size are not issues, the receiver can be made large, such as a custom-built, large SPAD array [19] or a parabolic concentrator for $1.5mm$ detectors [84]. Comparing such arrays' volume, in Table A-1's second column, we can easily see the cuboid-cone ratio of $12/\pi$ favors our design, and is unsurprisingly shown in Fig. 2-3(II) (left) across multiple focal lengths.

On the other hand, it is clear that a large receiver array would have higher received radiance, due to having a bigger effective aperture, when compared with our MEMS mirror. This is demonstrated in Fig. 2-3(II) (right) for the particular case of $M = 100, w_o = 5mm$, favoring our design. Despite this, large arrays have higher received radiance at all depths.

Conventional Single detector: Our approach is close to the conventional single pixel receiver, which can allow for detection over a non-degenerate FOV if it is defocused, as shown for a scanning LIDAR by [89]. When the laser dot is out of focus, some part of it activates the single photodetector. If the laser dot is in focus, the activation area available is smaller, but more concentrated. Next we describe and analyze our modification to the conventional single detector.

2.3.5 Proposed optical modification

Our approach is based on a simple observation; placing the image plane between the lens and the focus, i.e. $v < f$, will guarantee that the laser dot will never be in focus. For imaging photographs, this is not desirable, but for detecting the LIDAR system's received pulse, amplitude can be traded down, up to a point, as long as the peak pulse can be detected. Further, this optical setup ensures that the angular extent of the dot is nearly constant over a large set of ranges. This is further explained in the appendix and supported by simulations (red curve) in Fig. 2-3(III) (left) and explained in the ray diagrams of Fig. 2-3(III) (right).

For the conventional approach, when $u = f$, the FOV degenerates to a small value, where

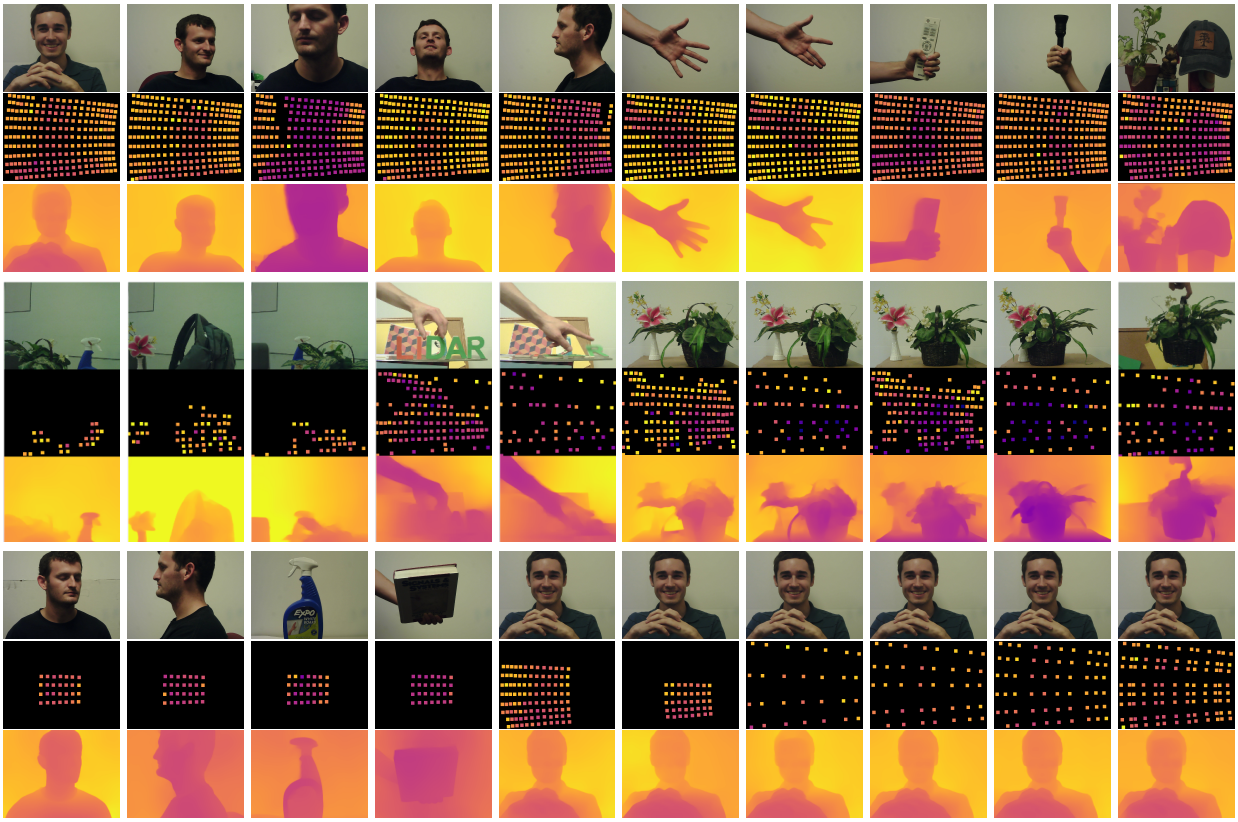


Figure 2-4. **Adaptive Lidar Sampling.** This figure qualitatively demonstrates the flexibility of our adaptive LIDAR by showing a range of scan patterns. In row 1, a fixed, equi-angular full FOV scan pattern was used. In row 2, the density of the scan pattern was automatically adapted according to the RGB image's entropy. In row 3, columns 1-6, constant sampling density was applied on a rectangular ROI with maximal scene entropy. In row 3, columns 7-10, the FOV of the scan pattern was kept fixed and a sweep of the sampling density was performed. Note, with no depth samples, our depth completion model defaults to monocular depth estimation from the colocated camera, since we randomly sparsified the input depth maps during training to encourage robustness to a range of sampling densities (including zero samples).

received radiance is also the highest. Our design does not suffer this depth-dependent FOV variation and is consistent across the range. As shown in the right, however, this results in a low received radiance since the system is always defocused. Simulations support this, in Fig. 2-3(III), for $u > f$, shown in red, for settings of $f = 15mm$, $A = 100mm$, over a range of sensor sizes and ranges. In practice, we find consistent FOV to be more valuable than received radiance.

2.4 Towards Adaptive LIDAR

In contrast to other MEMS modulated LIDARs we do not run the MEMS mirror at resonance [28, 84, 56], but instead trace a specific scan pattern. The Mirrorcle mirror [65, 64] that we use is capable of tens of KHz of scanning frequency for custom patterns, which is enough to sense most common dynamic objects.

Many adaptive methods exist to find good scan patterns, represented as voltage-dependent mirror angles over time, $(\theta(V(t)), \phi(V(t))$. These include open loop [15, 27, 89, 19] real-time estimation of regions of interest (ROIs) as well as end-to-end learning to help decide where to sense next [59, 12].

We term using such adaptive algorithms with our flexible platform as *foveating LIDAR*, since it increases resolution, similar to how our eyes' fovea control which scene region is imaged in detail. **Our contribution** here is to demonstrate LIDAR foveation for dynamic scenes with an open-loop algorithm based on motion detection [27].

Data	MRE (%)	RMSE (m)	$\log_{10}$ (m)	$\boldsymbol{\delta_1}$ (%)	$\boldsymbol{\delta_2}$ (%)	$\boldsymbol{\delta_3}$ (%)
Real	10.16	.1659	.0410	89.80	95.88	98.63

Table 2-3. LIDAR Evaluation. The table reports the mean relative error (MRE), root mean squared error (RMSE), average ($\log_{10}$) error, and threshold accuracy (δ_i) of the calibrated depth measurements, relative to the "ground-truth" Kinect V2 depths, over all 75 scenes of our real dataset. The Kinect V2 has an accuracy of 0.5% of the measured range [7].

Data	Method	MRE (%)	RMSE (m)	$\log_{10}$ (m)	δ_1 (%)	δ_2 (%)	δ_3 (%)
NYU	Mono	8.55	.3800	.0361	90.56	98.08	98.56
	Ours	5.89	.2488	.0245	97.69	99.68	99.92
Real	Mono	28.26	.3711	.1090	50.14	87.38	96.00
	Ours	12.29	.1668	.0395	85.86	95.89	99.18

Table 2-4. Base Comparison to Monocular Depth Estimation. As a baseline, we compare to state-of-the-art monocular depth estimation [5] (Mono) to our depth (Ours) completion method on a sub-sampled version of the NYUv2 Depth [68] (NYU) dataset and on our real dataset (Real). Both the monocular depth estimation and depth completion methods were trained only on NYUv2 data. To account for this, monocular depth estimates were scaled by the ground-truth median, as in [5]. Such scaling was not performed for depth completion predictions because the sparse input depth samples from the LIDAR already provide a reference absolute depth.

Experimental setup: Our LIDAR engine is a single beam Lightware SF30/C with an average power of $0.6mW$ and a pulse frequency of $36KHz$. This device is designed for outdoor use and can produce 1600 depth measurements per second at 100m. Data is captured as a stream of depth measurements, and each are time-stamped by the MEMS direction, given by the voltage $V(t)$. We modulate the single beam with a $3.6mm$ Mirrorcle MEMS mirror. Our current prototype has a range is $3m$ (due to optical losses that can be optimized closer to the $100m$ max in newer versions) and a field-of-view of $\approx 25°$. The laser dot, in steradians, is $6 \times 10^{-4}\Omega$ and this angular support is consistent over change in MEMS mirror angle.

Calibration and validation: Even with our novel optical system, the raw sensor measurements still provide depth discrimination. Since our sensor response is linear, we apply a 1D calibration to convert the LIDAR voltages into distances. We evaluate the quality of our sensor measurements and our calibration by computing the mean relative error (MRE), root mean squared error (RMSE), average ($\log_{10}$) error, and threshold accuracy (δ_i) of the calibrated depth measurements from our LIDAR. We do this relative to the "ground-truth" Kinect V2 depths, over all 75 scenes of our real dataset, and these are reported in Table 2-3. The Kinect V2 has an accuracy of 0.5% of the measured range [7]. Finally, we also captured 10 fronto-planar scenes (at ranges .5m-3m) and computed the RMSE of the depth measurements along the plane using the SVD method. The

resulting average RSME over all 10 scenes was 0.06918m.

Data	FPS	MRE (%)	RMSE (m)	$\log_{10}$ (m)	δ_1 (%)	δ_2 (%)	δ_3 (%)
NYU	30	5.89	.2488	.0245	97.69	99.68	99.92
	24	5.88	.2430	.0244	97.97	99.70	99.92
	18	5.59	.2261	.0233	98.50	99.77	99.94
	12	5.65	.2255	.0236	98.52	99.77	99.94
	6	5.15	.1879	.0217	99.32	99.91	99.98
Real	30	12.29	.1668	.0395	85.86	95.89	99.18
	24	12.09	.1644	.0446	86.34	96.04	99.26
	18	11.57	.1578	.0430	87.27	96.61	99.30
	12	11.59	.1558	.0435	88.26	97.01	99.33
	6	11.19	.1537	.0422	88.10	97.19	99.26

Table 2-5. Evaluation of Depth Completion. This table conveys three key features of our system: (1) It highlights, the trade-off between frame rate and depth uncertainty, which impacts real-time applications; (2) it provides a quantitative evaluation of the robustness of our depth completion algorithm to varying sampling densities; and (3) provides an illustrative example of our system flexibility, which can be leveraged for a range of applications. For frame rates of 30, 24, 18, 12 and 6, the samples per frame were 28, 40, 60, 104 and 231 respectively.

2.4.1 Depth completion

We now describe depth completion for the foveated measurements of our LIDAR. This builds on existing work [95, 105] where the sparse depth measurements are captured by our flexible LIDAR sensor and the "guide" image is captured by a RGB camera that is co-located with the sensor. We train a DenseNet-inspired [48] encoder-decoder network to perform RGB-guided depth completion of sparse measurements.

Architecture. We adopt [5]'s encoder-decoder network architecture, except that our network has 4 input channels, as it expects a sparse depth map concatenated with an RGB image. The encoder component of our network is the same as DenseNet 169 minus the classification layer. The decoder component consists of a three convolutional blocks followed by a final 3×3 convolutional layer. Each bilinear upsampling block consists of two 3×3 convolutional layers (with a leaky ReLU), and 2×2 max-pooling.

Optimization. We adopt [5]'s loss as a weighted sum of three terms:

$$L(y, \hat{y}) = \lambda L_{depth}(y, \hat{y}) + L_{grad}(y, \hat{y}) + L_{SSIM}(y, \hat{y}) \tag{2-2}$$

where y and $\hat{y}$ denote the ground-truth and estimated depth maps respectively, and λ denotes a weighting parameter, which we set to 0.1. The remaining terms are defined as in [5] which has the full expressions.

Datasets and Implementation: We perform our evaluations using two datasets: a real dataset captured with our LIDAR system and a simulated Flexible LIDAR dataset generated by sub-sampling the NYUv2 Depth dataset [68]. The real dataset consists of pairs of RGB images and sparse depth measurements of 75 different scenes captured with our LIDAR system. For each of the 75 scenes, we also capture a dense "ground-truth" depth map using a Kinect V2 depth sensor that is stereo calibrated with our LIDAR system. All real dataset images are used exclusively for testing. The simulated dataset is split into non-overlapping train, test, and validation scenes.

We train the model described in section 2.4.1 on a simulated Flexible LIDAR dataset generated by sub-sampling the NYUv2 Depth dataset. During training, depths were randomly scaled to prevent the network from overfitting to the color camera used to capture the RGB images in the NYUv2 dataset. For optimization, we used Adam [55] with $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 1\text{e} - 8$, a learning rate of 0.0001, and a batch size of 4. The learning rate was dropped to 0.00001 after $94k$ iterations. We used Xavier initialization for the first layer of our network. All other layers were initialized with the pre-trained weights from [5] for monocular depth estimation on NYUv2. To augment the data, we first randomly resize the input images such that the smallest dimension varies between $640, 832$, or 1024. We then apply a random crop to the reduce the size to 640×480. In addition, the RGB channels were randomly shuffled.

Basic 'sanity-check' validation against monocular estimation: As a minimum baseline, we confirm that guided depth completion outperforms monocular depth estimation using a state-of-the-art monocular depth estimation network [5] in Table 2-4.

Data	Method	MRE (%)	RMSE (m)	$\log_{10}$ (m)	δ_1 (%)	δ_2 (%)	δ_3 (%)
NYU	Full FOV	5.52	.2392	.0231	98.24	99.86	99.98
	Foveated	4.81	.1845	.0202	99.50	99.97	100
Real	Full FOV	15.72	0.1925	.0566	80.30	99.79	99.36
	Foveated	13.36	.1589	.0497	83.24	97.80	99.46

Table 2-6. Depth Completion on Foveated Lidar Data. "Foveated" means that the scan pattern was automatically adapted to densely sample a region of interest in the scene. "Full FOV" means that a scene independent equi-angular scanning pattern was utilized. In all cases, the "Foveated" and "Full FOV" scan patterns contain the same number of samples (hence, the equivalent frame rates). Results are evaluated at 30 FPS. Both Full FOV and Foveated errors are computed only in identical regions of interest, showing foveation increases accuracy.

2.4.2 Motion-based Foveated Depth Sampling

Our flexible platform allows us to ask if foveated LIDAR sampling improves depth measurements. We evaluate our guided depth completion network on LIDAR data captured with two different sampling regimes, full field-of-view sampling and foveated sampling in regions of interest, at various frame rates.

Table 2-5 shows our evaluation for full field-of-view sampling. Table 2-6 demonstrates that foveation improves reconstruction in a region of interest, with qualitative results in Figures 2-4 and 2-5.

We also perform foveated sampling in real-time, using an open-loop motion-based system to determine the scan patterns. For a dynamic scene, a foveating LIDAR can have fewer samples in the right places, decreasing latency and improving frame-rate. In Fig. 2-5, we show objects moving across the scene. At each instance, the system performs background subtraction to segment a motion mask. This mask drives the LIDAR sampling, which has less points than a full dense scan would have, and therefore has higher sampling rate. In each result, ROI sampling density was identically dense, and the rest-of-the-scene density was different and sparser. The amortized frame rates for the real-time foveated sequence in rows 1, 2, 3 and 4 of Figure 5 are 20 FPS, 13 FPS, 9 FPS and 24 FPS. Without foveation, dense sampling over the entire scene would result in a frame rate of 6FPS, which is much lower. Note that as the object changes position, the

ROI changes and the LIDAR senses a different area. If temporal sampling is not the focus, then the method can instead densely sample the points onto the region of interest, increasing the angular resolution (i.e. zooming). Finally, we note that all results include depth completion of the measurements, showing high-quality results.

Note that the scenes have a simple, planar background, but this is only to improve the specific vision algorithm we use, i.e. background subtraction. Any other computer vision technique could be used to obtain regions of interest, and our LIDAR would work accordingly. We have shown geometric reconstructions of everyday complex objects in Figures 2-1, 2-4 and 2-5.

2.5 Limitations and Conclusions

Our LIDAR engine has a $3m$ range, which enables initial feasibility tests and is appropriate for certain tasks such as gesture recognition. Range extension is achievable, since the Lightware LIDAR electronics engine has a 100m outdoor range and is only reduced by unnecessary optical losses. For future prototypes we wish to remove these losses with a GRIN lens, as done by [28]. In conclusion, our prototype enables the kind of flexibility that has, so far, only been seen in simulated experiments and follows a recent trend in computational photography to use data-driven approaches inside the sensor [18, 21, 20, 83].

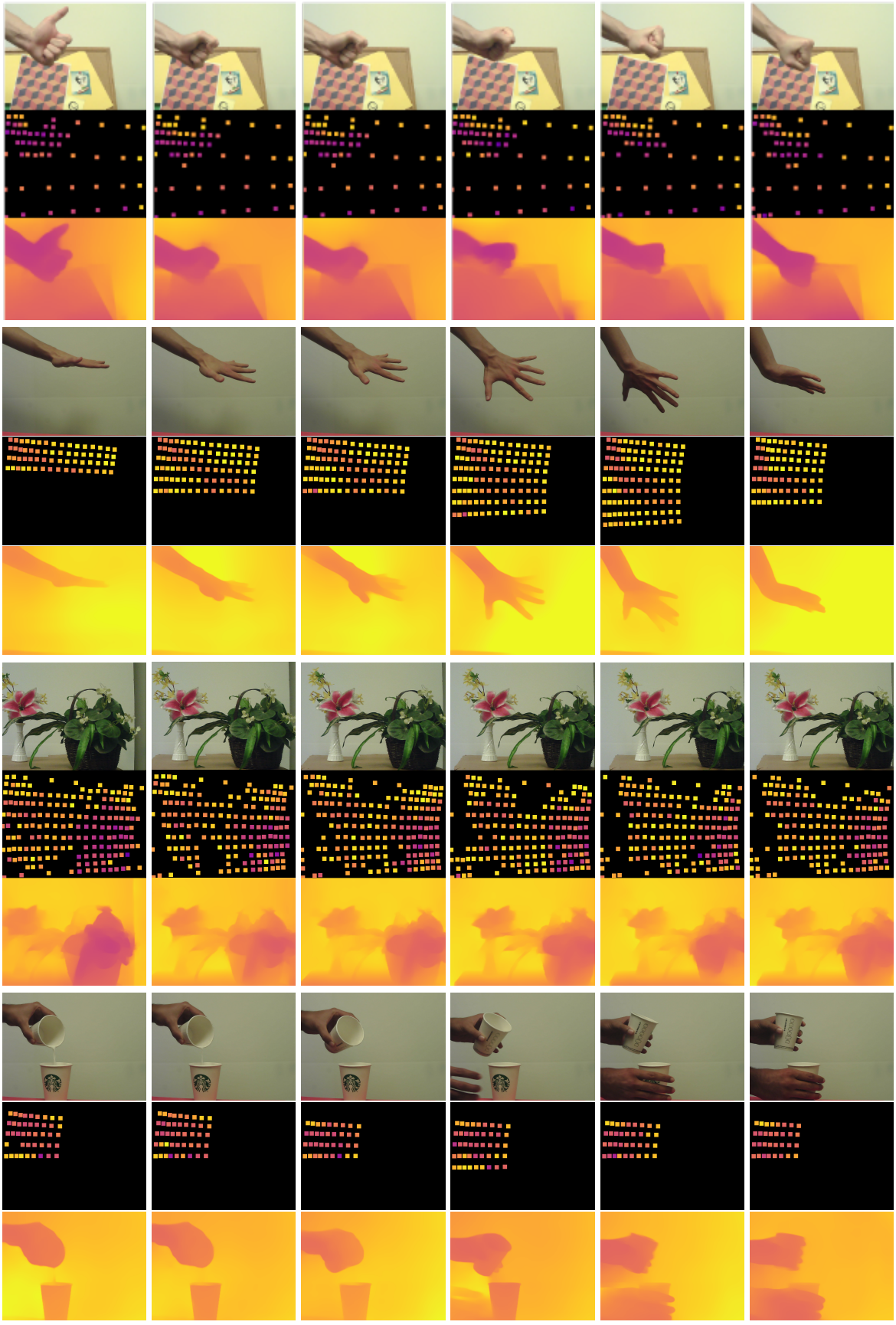


Figure 2-5. **Motion-based adaptive sensing.** As the object moves, we use background subtraction to detect the region of interest and the MEMS-modulated LIDAR puts the samples where the object is located.

CHAPTER 3
FOVEATED DEPTH SENSING FOR CHALLENGING UNDERWATER ENVIRONMENTS

3.1 Introduction

LiDAR sensors are critical in diverse depth sensing applications but face challenges in turbid underwater environments due to scattering effects. To address this, our work introduces a foveating confocal bistatic LiDAR system, inspired by automobile fog lights.

We present two major contributions:

- Introduction of a novel underwater bistatic LiDAR system employing MEMS mirror modulation for both transmitter and receiver. Our optical setup separates the illumination source from the receiver to diminish backscatter, a concept reflected in [36].
- Development and testing of a prototype in a laboratory tank setting, validating our design and models with real-world data. This enables a technique to control sensor sampling density at different depths, we term as *foveation* of the LIDAR.

3.2 Related Work

[0] **Time-of-Flight (TOF) Imaging and Adaptive Optics:** In contrast to using fast adaptive optics for atmospheric turbulence [10, 94] our research focuses on adaptive sampling within turbid water for depth measurement . Previous studies have explored TOF reconstruction in scattering media through phase frequency encoding [36] or efficient probing techniques [73]. However, such methods often rely on assumptions of minimal global illumination in the epipolar plane, an assumption that does not hold in heavily turbid underwater environments. Our approach utilizes a bistatic system, enabling effective reduction of global illumination through confocal imaging, a critical improvement for underwater applications.

MEMS Mirrors for Computer Vision: MEMS mirrors have found applications in diverse fields, from office automation to 360-degree displays [77, 52]. While recent innovations have incorporated MEMS mirrors for controlling LiDAR transmitters in adaptive depth sensing [89],

0 NAVAIR Public Release SPR-2024-0202—Distribution Statement A: Approved for public release; distribution is unlimited.

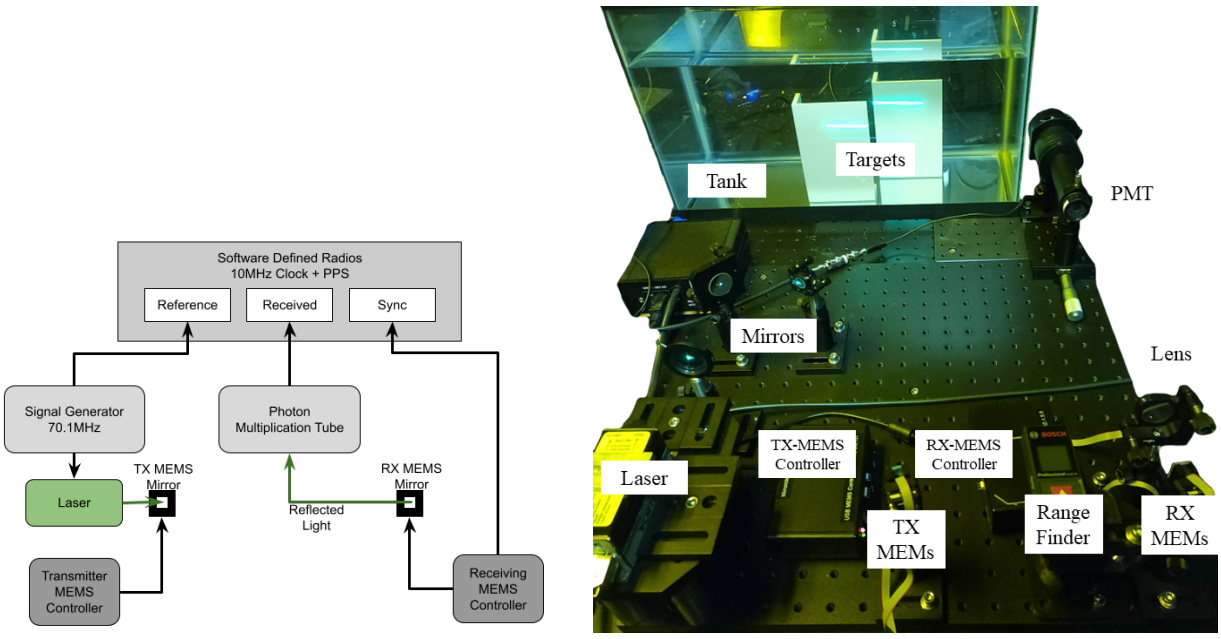


Figure 3-1. On the left, we show a block diagram of our prototype. The components that make up the prototype are a 514nm laser, a photomultiplier tube (PMT), three N200 software-defined radios, and two 3.6mm Mirrorcle MEMS mirrors. On the right, we show a labeled photo of our optical setup.

our work extends this concept by controlling both the transmitter and receiver, adaptive principles from structured light systems to LiDARs [67] [2].

Underwater LiDAR: While blue-green lasers have been used effectively in underwater LiDAR systems for shallow waters, turbidity significantly limits their efficacy [3, 13]. Our bistatic confocal design aims to overcome these limitations, enhancing LiDAR's impact in underwater environments with scattering effects.

Scanning LiDAR: MEMS-modulated LiDAR systems have been deployed in various scenarios, including depth sensing and robotics [28, 84, 56, 65, 64]. Unique to our project is the placement of MEMS mirrors on both the transmitter and receiver, creating a bistatic system optimized for turbid environments and enabling foveated sensing.

3.3 Methodology

In our setup shown in Fig. 3-1 we use two micro-electro-mechanical (MEMS) mirrors to control the transmitter and receiver directions for a lidar sensing through scattering media. There are pairs of voltages for each MEMS mirror, that physically rotate the mirror position into a

desired angle. Let the azimuth and elevation angles for the receiver mirror be (ϕ_r, θ_r) and for the transmitter mirror be (ϕ_t, θ_t).

3.3.1 Bistatic Confocal MEMS-modulation

In this paper, by "confocal" we mean that the optical axes of the mirrors (i.e., the unit vector perpendicular to the mirror surface) always lie in the epipolar plane [43] (see Fig. 3-2). Therefore, one necessary (but not sufficient) constraint for mirror control is that they must satisfy the fundamental matrix $\mathbf{F}$. To use the fundamental matrix, consider a virtual image plane in front of each mirror, where the "pixel" corresponding to each mirror position is given by a corrected angle. For example, for the transmitter mirror we have pixels $x_t = [f\tan(\phi_t), f\tan(\theta_t)]$, where f is the virtual focal length of the virtual image plane such that the maximum and minimum extent of the pixels are $(\pm 1, \pm 1)$ respectively. Similarly, we also define x_r. The fundamental matrix constraint is $x_t^T \mathbf{F} x_r = 0$, which we recast in terms of MEMS mirror angles as:

$$[f\tan(\phi_t), f\tan(\theta_t)]^T \mathbf{F} [f\tan(\phi_r), f\tan(\theta_r)] = 0 \tag{3-1}$$

Successful signal reception in a clear medium aligns with the epipolar constraint point on the surface of the scene. However, in scattering media, received signals can occur at unintended locations along the epipolar line, potentially leading to depth misestimation. Our method addresses these challenges by detecting discontinuities along the ray and analyzing amplitude changes to accurately estimate depth even in dense scattering media.

3.3.2 Modulated Continuous-wave LiDAR

Our Continuous-wave (CW) LiDAR operates by modulating a laser with a reference signal $A_{ref}\cos(\omega + \chi_{ref})$ at the sensor, which is further modulated by the target's reflectance or scattering properties after reflecting off the first MEMS mirror. The received signal $A_{meas}\cos(\omega + \chi_{meas})$ is analyzed for the phase difference $\delta\chi = \|\chi_{meas} - \chi_{ref}\|$. Through phase unwrapping F_{unwrap}, this enables depth measurement as $Z = \frac{cF_{unwrap}(\delta\chi,\omega)}{2}$. Our focus also extends to the amplitude A_{meas} of the received signal, particularly in relation to scattering media.

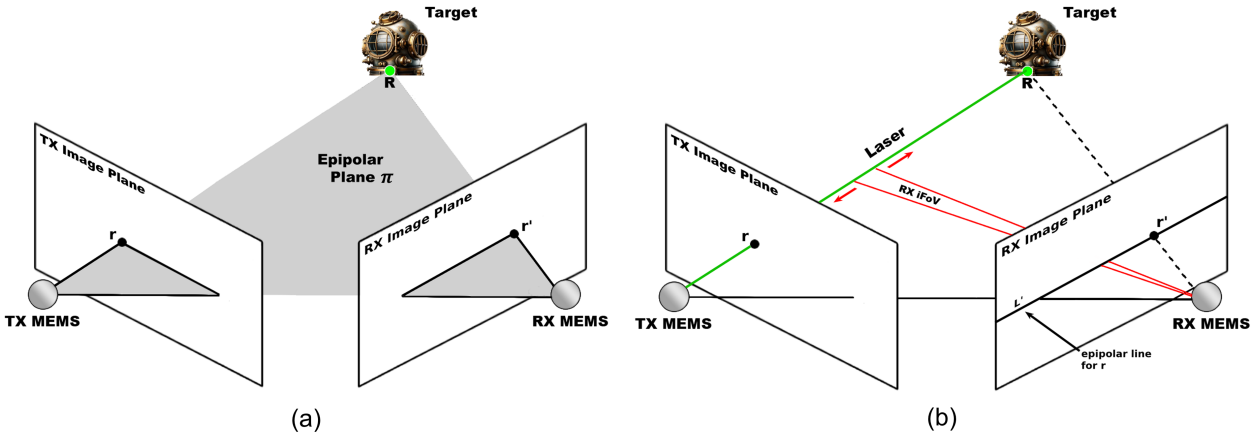


Figure 3-2. **Multi-view Geometry** The transmitter and receiver are indicated by the TX and RX MEMS, with their corresponding image planes. *(a)* The MEMS mirrors, 3D point **R**, and its images **r** and **r′** lie in a common plane π. *(b)* The ray defined by TX and **r**, ie. the laser, is imaged as a line L' in the RX image plane. The 3D point **R**, which projects to **r**, must lie along the ray, thus it must also lie on L'. We use the Epipolar line L' to scan the RX iFoV along the ray, capturing the reflected irradiance off the target and any backscatter off the ray. Figure inspired by a similar diagram in [43]

3.3.3 Relative Phase to Absolute Depth Calibration

Utilizing software-defined radios, we compare sinusoidal reference and received signals to calculate their relative phase $\delta\chi = \|\chi_{meas} - \chi_{ref}\|$. This phase difference, influenced by the oscillators' frequencies, allows us to measure relative depth, the variation in depth between measurements. However, this does not provide absolute depth, defined as the distance between the receiving MEMS and the laser target.

To convert relative phase measurements into absolute depth, we calibrate ground truth depth measurements with the phase differences. This process involves three planar experiments, using consistent transmitter sampling and depth measurements from the receiver MEMS with an adjacent hand-held laser rangefinder. By correlating these depth readings with phase differences, we create a linear regression model, effectively mapping relative phase to absolute depth.

3.4 Results

In this section, we explore the advancements facilitated by our innovative system, as shown in Figure 3-3. This figure illustrates the system's impact on enhancing underwater imaging through dynamic depth sampling, where the LiDAR can dynamically adjust the iFOV in response

to areas of interest. Through a series of comparative visual assessments, we reveal how the system's foveated sampling technique—powered by dual MEMS mirror-modulated components—substantially improves spatial detail in turbid water conditions. The visual evidence provided offers a clear narrative of the system's capabilities, underscoring the tangible benefits of our approach in underwater imaging.

3.4.1 Algorithmic Sampling and Foveation

Employing the fundamental matrix constraint (eq. 3-1), we generate receiver mirror angles for an in-depth amplitude profile creation. The measurements span an angular support volume ω_{rt}, defined by the azimuth and elevation intersections $(\phi_r, \theta_r) \cap (\phi_t, \theta_t)$. MEMS modulation allows comprehensive angular coverage within the sensor's field of view, allowing the receiver to traverse the epipolar line to compile a detailed amplitude profile $A_{meas}(\phi_r, \theta_r) \forall (\phi_t, \theta_t)$. This profile captures both scattered irradiance and target reflections, marked by discontinuities, thus enriching scene comprehension and range finding even under low visibility, as depicted in Fig. 3-3.

The significance of the foveation feature our system offers is clearly illustrated in Figure 3-3, where we elevate beyond the standard equiangular sampling approach of traditional LIDAR, i.e. sampling in all directions at the same density. By judiciously controlling the MEMS mirrors to slow down over areas of interest, we enhance sampling resolution where it matters. The figure portrays this by intensifying the sample density on targeted objects—evident in the middle and right columns where the left and right objects, respectively, are examined with higher spatial resolution. This tailored sampling is critical, as it allows for detailed detection and analysis of objects in varying visibility conditions, significantly optimizing the performance of LIDAR in precision-critical applications such as autonomous navigation and detailed environmental mapping.

3.5 Conclusions and Limitations

Conclusions: Our innovative bistatic adaptive LiDAR system demonstrates exceptional capability in confocal sampling and precise depth reconstruction, even under highly turbid conditions. By utilizing adaptive control coupled with a narrow iFOV on the receiver, our system

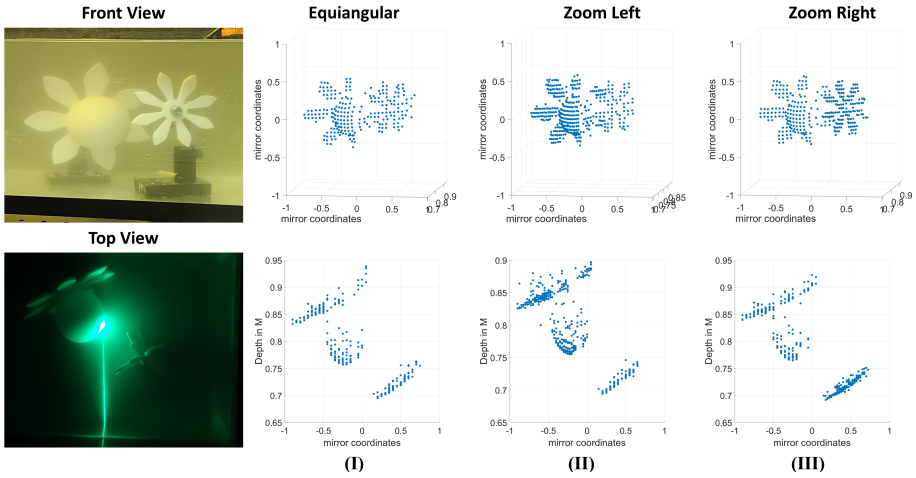


Figure 3-3. Demonstration of algorithmic sampling and foveation by our system, showcasing its ”zoom” functionality on various objects within a turbid scene. Column (I) displays equiangular scanning across the entire scene. Columns (II) and (III) illustrate the system's ability to selectively zoom in: Column (II) focuses on the object to the left, while column (III) zooms in on the object to the right, highlighting the system's precise and adaptable foveation capabilities. The range of the scene is from 65cm to 1 meter.

adeptly captures both scene irradiance and scattered light. This dual capture, facilitated by the small iFOV and adaptive control, effectively creates a spatially-gated LiDAR engine. This engine can be utilized both during and after data acquisition to yield accurate depth reconstructions, demonstrating its versatility and effectiveness in challenging environments.

Limitations: The current prototype of our bistatic LiDAR system operates within a limited working volume, suitable for preliminary testing in controlled tank environments. This constraint is primarily due to the MEMS mirrors' maximum scanning angle of ± 7 degrees. One potential solution to expand the working volume is the incorporation of wide-angle lenses in front of the transmitter and receiver MEMS. Additionally, the limited reflective capacity of the 3.6mm MEMS mirror used as our receiver poses a challenge. The mirror's small size acts as a restrictive aperture, limiting the light reflected from the scene.

CHAPTER 4
SPATIO-TEMPORAL FOVEATED DEPTH SENSING FOR BANDWIDTH LIMITATIONS IN SPAD CAMERAS

4.1 Introduction

Biological vision systems have the remarkable ability to *foveate* — i.e. redistribute cognitive resources towards "salient" features or objects in a scene, depending on context. Unfortunately, most conventional cameras and computer vision systems today capture scene information in a non-adaptive fashion, spending power and bandwidth on sensing scene components that may not help the overall imaging task. In fact, the current framework for deep learning-based systems assumes uniform sampling of the scene and overcomes these limitations through data-driven pipelines that focus on interesting regions of the scene [97] [81] in the input RGB images.

While this inefficient but popular framework for conventional RGB sensors may be difficult to change, our proposed method, called FoveaSPAD, can impact the next wave of single-photon avalanche diode (SPAD) sensor technology. SPADs can capture scene information at the granularity of individual photons, at timescales as small as 10's of picoseconds. Recent advances in CMOS-compatible SPAD pixel designs has enabled real-time in-pixel processing of these photon timestamp streams. Thus, SPADs are a natural candidate for designing efficient depth cameras — individual pixels can be reprogrammed on-the-fly to adaptively accept or reject a spatio-temporal subset of the photon stream.

Our FoveaSPAD algorithms enable capturing scene information at higher granularity in regions that are most relevant to a downstream vision task. *In this sense, we generalize the term "foveation" in the context of adaptive SPAD spatio-temporal sampling to allow both depth and memory efficiencies.* For robots, remote sensor nodes, and other resource-constrained systems, foveation for SPAD sensors can allow accurate depth sensing under constraints on power and bandwidth (see Fig. 4-1).

The raw data captured by an array of SPAD pixels can be thought of as a spatio-temporal photon stream. Each photon detection is represented as an (x, y, t) coordinate, where the $x - y$ coordinates denote the pixel location and the t coordinate denotes the photon detection timestamp.

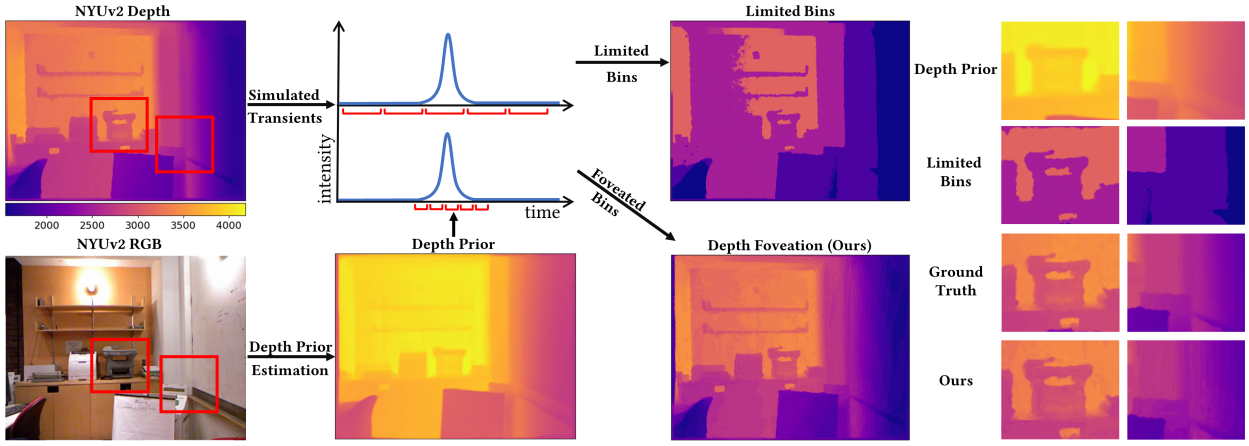


Figure 4-1. **Depth Prior Driven SPAD Depth Foveation:** SPAD sensors suffer from a data bottleneck, since thousands of histogram bins are used to generate depth as shown in the top left. If fewer bins are used, this reduces depth resolution, as shown in the limited bins depth result. Our idea is to use additional information, such as a color image (Sec. 4.4, 4.7) or optical flow (Sec. 4.6), to foveate the SPAD bins. Therefore, for the same memory cost we can place the bins near where the histogram peak should be, results in accurate depth, as shown in the depth foveation result. The insets show that our method achieves the accuracy and resolution of ground truth, with fewer bins. They also show that the depth prior, in this case monocular estimation, by itself cannot provide the correct depth, and foveation is required.

Each SPAD pixel captures the round trip time of a laser pulse to and from a given scene point, constructing a photon timing *histogram* which records the number of photons captured at various time delays with respect to the time the laser pulse was transmitted. Each pixel must construct one such histogram, typically with 1000's of bins, which causes a severe data bottleneck for today's SPAD cameras. To illustrate the severity of the bottleneck, consider a 1-megapixel SPAD array with a 1000-bin histogram per pixel, storing 1 byte per bin. At 30 frames per second, this setup generates a staggering 30 GB of data every second.

Our algorithms foveate across the spatio-temporal histogram space to efficiently recover the peak, providing the time-delay t for depth computation. We adaptively capture subranges to locate laser photons, rejecting ambient photons. Note that our proposed algorithms are not exhaustive; rather, we aim to define a class of algorithms that rely on a depth prior. In this work, we propose three methods for acquiring priors, though many other methods exist, such as depth from stereo, depth from defocus [100], or non-vision-based methods such as sonar. Each method has

trade-offs, and it is up to the user to determine which method best suits their use case. Our contributions in this work are as follows:

- We present a theoretical model for expected gains (in terms of increased signal-to-noise-ratio and depth resolution) from foveation with SPADs.
- We explore the question of how to foveate in space and time at a single time instant by leveraging monocular depth estimates, which can come either from the SPAD-generated image or a cheap, external color camera. We propose different flavors of practical FoveaSPAD designs that optimize for memory/bandwidth and depth resolution.
- For images of moving scenes, we demonstrate how to use optical flow cues to direct SPAD foveation.
- We show results both in simulation and using recently available real SPAD datasets.

4.1.1 Hardware Emulation

Time-correlated single photon counting is the technique that enables SPAD cameras to build histograms and control binning on-sensor. Our work is limited to simulation experiments and hardware emulation of existing SPAD LiDAR data. SPAD sensor arrays with native support for foveation are currently not available off-the-shelf. However, we believe it is possible to implement our proposed techniques in-pixel or in-camera, thanks to several recent proofs-of-concept of kilopixel resolution reconfigurable SPAD pixel arrays with in-pixel timestamping, gating, and histogramming capabilities [50, 102]. We expect that this work will influence future hardware implementations and designs, resulting in more efficient and versatile SPAD sensor arrays.

4.1.2 Scope: Simulation and Emulations

In this work, we anticipate future hardware advancements that will enhance SPAD-based depth sensing. Our simulations and emulations are intended to project the performance of emerging SPAD sensor technologies, focusing on adaptive and efficient bin sampling to mitigate memory bottlenecks with minimal loss of accuracy. Potential future implementations could feature a shared "macropixel" architecture and a dynamic gating system, allowing pixel groups to

adjust to appropriate gating signals in real time. We explore these ideas further in Sec. 4.9 and present a speculative "macropixel" array design in Figure 4-11, which includes a variable-resolution TDC—a key component for one of the proposed methods. These simulations play a critical role in validating our algorithms and highlighting their potential impact on future sensor designs, even in the absence of current hardware.

4.2 Related Work

Our research takes inspiration from biology, since many animals have a region of high spatial acuity, i.e. the fovea, which they scan over the scene. In this sense, we are allied with foveated imaging research in computer vision and computational photography, and we now outline these related efforts:

Efficiency in Single-photon 3D Cameras: The data bottleneck issue in SPADs due to high-resolution sampling in histograms is well-known. Research that attempts to mitigate this issue include novel statistical representations [45] as well as compressive histograms [40, 30, 39] that use a small number of bins at maximum resolution to recover the entire scene. In contrast, our approach works and scales easily with a large number of SPAD pixels. Other efforts include partial histogram methods such as using sliding windows for sub-range gating has been investigated [78, 24, 26, 25] which have linear efficiency and two-stage coarse-to-fine resolution scaling [103] which provide logarithmic efficiency. Our method uses context from cues such as optical flow to provide $\approx O(1)$ near-constant time efficiency. Finally, other work has used external sensors for guided upsampling or upscaling, [58, 88], but these are post-capture processes. In contrast, we perform foveation during capture and this gives us SNR and compute efficiencies that we have theoretically analyzed. A complementary approach to foveation is to use adaptive "equi-depth" histogramming approach for the signal peak [51]. Our approach is also complementary to adaptive gating approaches for SPAD LiDARs [76], with adaptive gating and exposure techniques working with or without a prior.

Foveated Depth Sensors: Our work is related to post-capture methods for upsampling and superresolution shown on data from many modes, such as depth images, color photographs etc.

[9, 22, 62, 60, 95] and many of these have blended deep learning algorithms into the process of deciding where to sample [79, 49, 95, 96, 32, 91]. In fact, some of these algorithms are mature enough that commercial depth and LIDAR sensors allow post-capture foveation of the 3D point cloud through, for example, LIDAR-RGB fusion. In contrast, FoveaSPAD adapts during capture, and the efficiencies can impact small autonomous systems with power constraints. Directionally controlled LIDAR systems foveate spatially [101, 89, 12, 75]. These results complement our work on temporal foveation of SPAD sensors, including spatio-temporal foveation results (Sec. 4.5).

Foveation in Display Graphics: Foveation is an important research topic in computer graphics, where data displayed to a viewer on AR/VR glasses, for example, is rendered in a way that reduces bandwidth [33]. Most of the work in this area does not focus on data capture but only on data visualization post-capture [4, 93]. Foveated light-field optics have been proposed [47] and these can be integrated with algorithms that foveate which portions of the scene to render at high resolution to reduce rendering resource consumption. Algorithms include perceptually guided foveation [85, 74] and hardware-optimized rendering [63]. Unlike our depth sensor, these use passive displays and cameras to optimize bandwidth, storage, and compute

SPAD Histogram Techniques: Various techniques have been developed to address the issue by optimizing how histogram data is captured, processed, and stored, thereby reducing memory usage and computational overhead while maintaining depth accuracy [38]. These approaches focus on retaining essential depth information and compressing the data footprint, leading to more efficient systems that can scale with higher resolutions and larger arrays.

Zhang et al. proposed First Arrival Differential (FAD) LiDAR, which reduces per-pixel data throughput by capturing temporal differences between adjacent SPAD pixels, achieving up to a 100x reduction in data while preserving depth resolution [104]. Similarly, Tontini et al. introduced a histogram-less SPAD system, using a simple averaging method to extract depth information directly from photon timestamps, bypassing the need for memory-intensive histograms [92]. Sheehan et al.'s sketching framework compresses photon arrival distributions into "sketches," achieving high compression ratios without loss of accuracy [82]. This

complements White et al.'s differential SPAD architecture, which mitigates saturation and reduces the need for large counters and TDCs by recording relative photon arrival times between pixels [99]. Additionally, Sun et al.'s optical coding and super-resolution techniques leverage a phase plate and deep learning to achieve super-resolved images with minimal photon counts, further optimizing SPAD-based imaging [86].

These advanced methods can work synergistically with our foveated capture approach, collectively reducing data transfer and computational demands. By integrating differential measurements, compression techniques, and adaptive processing, these innovations enhance efficiency in SPAD systems.

4.3 Imaging Model and the Foveation Advantage

In this section, we explore the imaging model and the concept of foveation, specifically focusing on how foveation can enhance the efficiency and effectiveness of (SPAD) LiDAR systems. We will delve into the specifics of how the imaging model is constructed, including assumptions about the behavior of laser pulses and photon detection, and how these factors influence the design and performance of SPAD sensors. Furthermore, the impact of ambient light on signal-to-noise and signal-to-background ratios will be examined, demonstrating how foveation can mitigate these effects. The theoretical foundations laid out in this section will serve as the basis for the foveation techniques proposed in the subsequent sections, where we will develop and analyze algorithms to optimize the selection of foveated bins in SPAD imaging.

4.3.1 Foveation and Scene Priors

We propose two methods of foveation, specifically memory foveation and depth foveation, are designed to optimize the efficiency of SPAD LiDAR systems by leveraging a priori knowledge about the scene's depth. Both methods require adaptive per-pixel gating, for which the hardware has yet to be developed.

Memory foveation focuses on reducing the amount of data that needs to be stored and processed by concentrating on a subset of histogram bins where the depth information is most likely to reside. Depth foveation, on the other hand, aims to improve depth resolution by

reallocating histogram bins into a smaller, more focused region around the expected depth. The strategies proposed are fundamentally dependent on the accuracy and reliability of the scene depth prior, which guide the allocation of sensor resources.

Depth priors may be derived from any variety of means, including coarse initial scans, external sensors, or deep learning models. In this paper, we explore a few options, namely monocular estimation in Sect. 4.4, optical flow warping in Sect. 4.6, and coarse initial scans in Sect. 4.7. The quality of the prior directly impacts the success of foveation, with inaccurate priors potentially misallocating memory resources into incorrect regions. This dependence implies a trade-space between depth prior accuracy, and the amount of resources foveation stands to reduce. Exploring this trade-space is out of the scope of this paper, rather, we focus on using priors that are prone to error or are otherwise lower quality.

In the following subsections, we will define the image formation model, detailing the assumptions and mechanics of photon detection. We will then explore the effects of ambient light on SPAD histogram formation and discuss how the proposed foveation techniques provide an advantage.

4.3.2 Image Formation Model

We assume that each pixel in the SPAD sensor array is co-located with a pulsed laser illumination source with a Gaussian pulse shape. Assuming no multi-path or sub-surface scattering effects, the photon flux incident on each pixel consists of a superposition of laser photons (that arrive in a short time window corresponding to the round-trip time-of-flight to and from the scene point) and background photons due to ambient light (that arrive uniformly randomly distributed throughout the capture duration). The laser repetition period (T) determines the maximum depth range of the SPAD LiDAR. We assume that this period is discretized into N bins (N is often on the order of 1000's of bins in conventional SPAD cameras). The number of photons captured by the SPAD pixel in the n^{th} bin ($1 \leq n \leq N$) is Poisson distributed with a mean of $\Phi_{\text{sig}}\mathbf{1}(n = i) + \Phi_{\text{bkg}}$ where i is the bin location corresponding to the true scene depth. Various sources of noise such as dark counts and afterpulsing are assumed to be absorbed in the Φ_{bkg}

Table 4-1. Mathematical symbols used in this paper to study the foveated SPAD imaging model.

Symbol	Meaning
N	Number of bins across full histogram
M	Number of bins across foveated histogram
i	Bin location of corresponding to true scene depth
Z	Working volume of the sensor
T	Temporal volume calculated from Z and speed of light
SNR	Signal-to-noise ratio
SBR	Signal-to-background ratio
C	Number of cycles to create histogram
Φ_{sig}	Mean number of signal photons received per bin
Φ_{bkg}	Mean number of background photons received per bin
p_{gt}	Probability that a detected photon originated from the laser
$p_{\text{multipath}}$	Probability that a detected photon experienced multipath bounces
p_{floor}	Probability of a low noise floor
S	Number of pixels in the camera

term. A complete histogram captured by this SPAD pixel over C laser cycles is given by a Poisson random vector with mean $C\Phi_{\text{sig}}\mathbf{1}(n = i) + C\Phi_{\text{bkg}}$ for $1 \leq i \leq N$.

The simplified imaging model assumes all laser photons arrive in a single bin i. In practice, the laser pulse spans several bins "smearing" the signal photons over more than one bin. The laser peak is often modeled as a Gaussian shaped pulse; we use a 1 nanosecond full width at half maximum (FWHM) in our simulation results. Since the peak can span more than one histogram bin location, the defined Gaussian pulse may be used to estimate depth through match filtering. It is also possible to obtain a pseudo-intensity image by aggregating photon counts across histograms for each pixel which can be used in lieu of a co-located RGB or monochrome camera image for monocular depth cues.

4.3.3 Effects of Ambient Light

The integration time taken for all experiments is consistent. In this scenario, we show how foveation saves memory or improves depth resolution, and how the signal-to-noise ratio changes

depending on ambient light, bin width, and the number of laser cycles or exposure time.

Consider a SPAD pixel imaging a scene point illuminated by a pulsed laser. Initially, let us assume there are no multi-bounce effects and no ambient light, although we address these issues later on.

Photon detections from the SPAD pixel generate a histogram of arrival times. A conventional approach would use all N bins across the full histogram, whereas we propose methods to foveate attention onto a subset $M \leq N$ of these bins, where M is a window or gate with a user defined width (number of bins). Therefore, it is not surprising that, in the SNR analysis of our system, the ratio $\frac{M}{N}$ appears since this represents the advantage due to foveation.

In the analysis below, we will not make any assumption as to how the foveated bins M were obtained and instead just characterize the advantage of these, given that the desired histogram peak is captured by these bins. The analysis is not specific to any one method of acquiring a depth prior. In Sections 4.4, 4.5, 4.6, and 4.7 we propose algorithms to drive the selection of the foveated bins M and in Sect. 4.9 we provide a worst case analysis for whether the foveated M bins capture the histogram peak or not.

4.3.3.1 Low Ambient Light (No Pileup)

Now consider the conventional imaging case, where the SPAD sensor detects time-of-arrival of photons and accumulates into a photon timing histogram to find the time that corresponds to the true depth of the scene point.

We assume that the histogram has a full scale range of T seconds which is related to the maximum unambiguous depth range Z as $T = \frac{2Z}{c}$ where c is the speed of light. Consider N histogram bins that are uniformly distributed across the full scale range T. The width of each bin is $\frac{T}{N}$. Since narrower bins produce fewer photons, the SNR for each bin is proportional to the width of that time bin:

$$\text{SNR} \propto C \sqrt{\frac{T}{N}}, \tag{4-1}$$

where C denotes the number of laser cycles (i.e., the total exposure time) that was used to capture the histogram.

We now consider two types of foveation. In *memory foveation*, only a limited number of bytes in memory can be dedicated to the task of finding the histogram peak, and therefore placing these at the peak is most efficient. In *depth foveation*, memory allocation remains fixed but is concentrated in the foveated region, bringing the bins closer together near the histogram peak, thereby improving depth resolution.

Memory foveation: In memory foveation, we identify M bins $M \ll N$ where the true depth exists. The width of the bins remains the same $\frac{T}{N}$, and therefore the SNR is also identical to the conventional case:

$$\text{SNR} \propto \sqrt{\frac{M\frac{T}{N}}{M}} \propto \sqrt{\frac{T}{N}} \tag{4-2}$$

Depth foveation: In depth foveation, we concentrate the N bins that would have been distributed over the entire depth range, into a small region. The region is the same region used in memory foveation, and is given by multiplying the number of memory foveation bins M with the original bin width to give $M\ \frac{T}{N}$. This region is divided into N bins, and therefore the new bin width is $\frac{MT}{N^2}$. As before, the SNR is proportional to the bin width, and therefore much lower,

$$\text{SNR} \propto C\ \sqrt{\frac{MT}{N^2}} = \sqrt{\frac{M}{N}\frac{T}{N}} \tag{4-3}$$

Therefore, we have improved depth resolution but at the cost of SNR. To increase the SNR of the foveated depth we can increase C, the number of cycles the laser pulses through to create the histogram. The new cycle number must be equal to or greater than $\frac{C_{\text{new}}}{C} \geq \frac{N^2}{M^2}$, then,

$$\text{SNR}_{\text{new}} \propto C_{\text{new}}\ \sqrt{\frac{MT}{N^2}} = C\ \sqrt{\frac{T}{N}}. \tag{4-4}$$

In summary, memory foveation reduces memory usage with no change in SNR. Depth foveation increases depth resolution but with reduced SNR that can be compensated by more laser photons (i.e. longer exposure).

Below, in alg. 4-1, we define the general algorithm for memory and depth foveation. Note that the algorithms are independent of depth prior, and the spatio-temporal step, which we show in sec. 4.5, is optional.

Require: Total histogram bins N, Temporal Volume T, Number of foveated bins M, Total histogram bins for depth foveation N'

Calculate bin widths
$\Delta t = \frac{T}{N}$, $\Delta t_{depth} = \frac{T}{N'}$
Acquire a depth prior:
Monocular Sec. 4.4, Optical-Flow Sec. 4.6, Low-Resolution Super-Pixel Sampling Sec. 4.7
for $(x, y) \in S$ **do**
  Utilize the depth prior to find $\hat{d}(x, y)$
  Center foveation window M around $\hat{d}(x, y)$
  Memory Foveation:
    Capture histogram in the foveated window with bin width Δt and M number of bins
  Depth Foveation:
    Capture histogram in the foveated window with bin width Δt_{depth} and N' number of bins
end for
return Histogram image H
Decode depth image D. $H \rightarrow D$

Optional Spatio-Temporal steps:
  Quantization Based Sampling Sec. 4.5
    Quantize depth prior into discrete buckets B
    Select several pixels in each bucket at random. $S \rightarrow \hat{S}$
    Complete steps 3-10 with $\hat{S}$
    Quantize sparse depth map. $D(B) = \min(D(\hat{S}) \in B)$
  SuperPixel Based Sampling Sec. 4.7
    Acquire a pseudo-intensity map through photon counting
    Apply the superpixel algorithm to segment the pseud-intensity map
    Sample the centroid of each superpixel segment at full histogram resolution. $\hat{d}_{\mathrm{SP}}$
    Complete steps 3-10 with S and $\hat{d}_{\mathrm{SP}}$

[1]

Object 4-1. Memory and Depth Foveation

4.3.3.2 Strong Ambient Light (Pileup)

With strong ambient light, we now focus on the signal-to-background ratio (SBR), defined in [34] for SPADs as the ratio of the total number of signal photons to the total number of background photons received over each laser cycle. W.l.o.g, here we note that the SBR is

proportional to the probability of receiving signal photons divided by the probability of receiving background photons.

With ambient light, photons from both the laser source and the ambient illumination may be measured by the SPAD. Each time a photon is detected, the SPAD sensor resets creating a pause. It is this pause that creates a binomial model for image capture in SPADs [35, 34].

Therefore, the SBR analysis cannot simply compare the photon bin widths as in the prior section for the full resolution (N bins) and the foveated resolution (M bins). Instead, SBR calculations must include the *probability* of photons from the source vs. the background.

Conventional scenario: Let us first consider the SBR in the conventional case, with no foveation. From [35], using the Poisson model for photon distribution, we can write the probability of a photon from the laser incident on the bin corresponding to the correct depth as $p_{\text{laser}} = (1 - e^{-\Phi_{\text{sig}}})$. Correct depth detection will happen even if an ambient photon is detected at the correct depth, so the probability of correct depth detection is $p_{\text{correct}} = (1 - e^{-(\Phi_{\text{sig}}+\Phi_{\text{bkg}})})$.

Let i be the location of the bin corresponding to the correct depth of the scene point. This photon is only detected at i if, in addition, no photon from the laser is detected at any prior bin. Since the laser photons only show up at bin i, constrained by depth, the probability of the photon showing up at any other bin is zero. However, in this conventional scenario, photons from ambient light could show up at any prior bin to i, pausing detection at bin i. Therefore, the probability that the photon from the laser is detected at the correct depth is $p_{\text{sig}} = (1 - e^{-(\Phi_{\text{sig}}+\Phi_{\text{bkg}})})\ e^{-\Sigma_1^{i-1}\Phi_{\text{bkg}}}$.

The situation is different for ambient photons, which can arrive at any time instant before photons from the i^{th} bin arrive. We can write the probability that an ambient photon is detected at location q as $p_{\text{bkg}}^{q} = (1 - e^{-\Phi_{\text{bkg}}})\ e^{-\Sigma_1^{q-1}\Phi_{\text{bkg}}}$. We can therefore write the SBR proportionality for the conventional imaging case as:

$$\text{SBR} \propto \frac{p_{\text{sig}}}{p_{\text{bkg}}} \propto \frac{(1 - e^{-(\Phi_{\text{sig}}+\Phi_{\text{bkg}})}))\ e^{-\Sigma_1^{i-1}\Phi_{\text{bkg}}}}{\Sigma_{q=1}^{i} p_{\text{bkg}}^{q}}. \tag{4-5}$$

FoveaSPAD with Ambient Light: We now consider both memory foveation and depth foveation

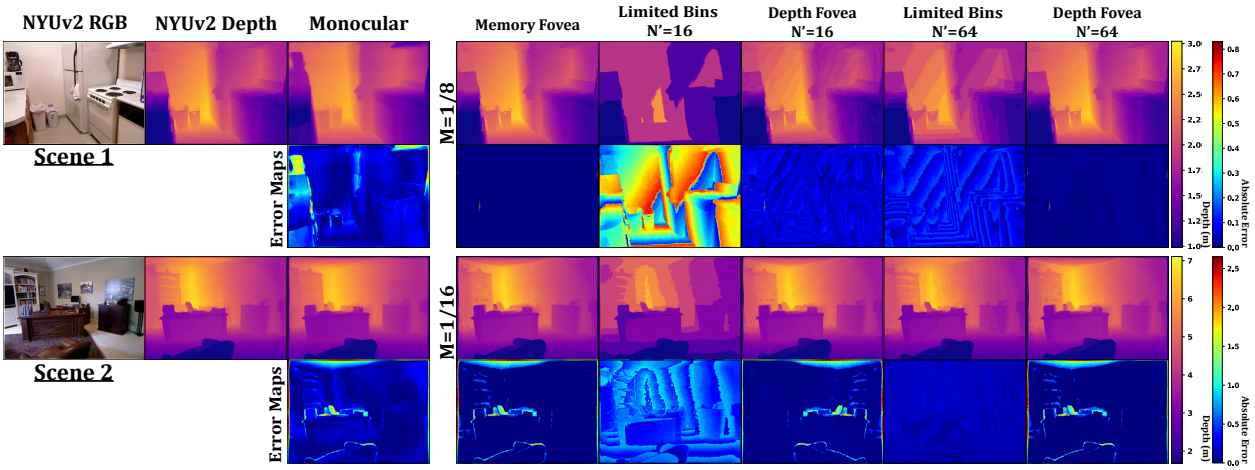


Figure 4-2. **Qualitative Comparison on NYUv2** Our memory and depth foveation techniques produce quality depth reconstructions with a fraction of the memory usage. Each row consists of the NYUv2 ground truth images, the monocular depth output from ZoeDepth, a simulated SPAD output with N′ bins, and our foveation techniques. The rows show different combinations of M and N′, where M is the number of bins in the foveated histograms, and N′ is the limited number of bins used for depth foveation. Monocular estimation is just one method of obtaining a depth prior in a class of methods, in sec. 4.6 and sec. 4.7 we show two more methods.

where the foveated bins N are given to us. In both these scenarios, we model the arrival of photons from both ambient and laser sources.

Memory foveation: Consider the foveated bins N, which we assume contain the bin with the histogram peak. Suppose the closest index for these bins is j. Then, the SBR increases, since the histogram sensitivity is unaffected by photons that impact the sensor before bin j.

$$\text{SBR} \propto \frac{(1 - e^{-(\Phi_{\text{sig}}+\Phi_{\text{bkg}})})\, e^{-\Sigma_j^{i-1}\Phi_{\text{bkg}}}}{\Sigma_{q=j}^{i} p_{\text{bkg}}^{q}}. \tag{4-6}$$

In the extreme case, where we have perfect foveation, and $i = j$, then the terms for ambient light before bin i become 1,

$$\text{SBR} \propto (1 - e^{-(\Phi_{\text{sig}}+\Phi_{\text{bkg}})}). \tag{4-7}$$

i.e. in other words, the effect of foveation is to remove the dependence on prior photon arrival for detection, since these no longer delay the measurement of photons at the ith bin. This "perfect

foveation" SBR term is dependent on the ratio of the strength of the laser and ambient signal directly and is not constrained by the binomial nature of SPAD photon capture.

Depth foveation: Since we concentrate all N bins into the foveation window, we are again susceptible to the binomial nature of SPAD photon capture. In addition, the bins are smaller to fit within the window, and as described in the non-ambient light section, the bin width is reduced as $\frac{M}{N}$.

We can write the probability that an ambient photon is detected at location q as $p^q_{\text{bkg}} = (1 - e^{-\frac{M}{N}\Phi_{\text{bkg}}})\ e^{-\Sigma_1^{q-1}\frac{M}{N}\Phi_{\text{bkg}}}$. The SBR proportionality also shows the effect of reduced signal strength as:

$$\text{SBR} \propto \frac{p_{\text{sig}}}{p_{\text{bkg}}} \propto \frac{(1 - e^{-(\frac{M}{N}(\Phi_{\text{sig}}+\Phi_{\text{bkg}}))})\ e^{-\Sigma_1^{i-1}\frac{M}{N}\Phi_{\text{bkg}}}}{\Sigma_{q=1}^{i} p^q_{\text{bkg}}}. \tag{4-8}$$

In summary, memory foveation increases SBR. While depth foveation has the same SBR as conventional capture, it improves depth resolution. It is this theory that motivates the remaining simulation results in the paper, where we explore different ways of creating depth and memory foveation for SPAD sensors.

4.4 SPAD Foveation from Monocular Depths

With the imaging model defined, we proceed to our first experiment, demonstrating how our memory and depth foveation techniques can effectively work with a monocular depth prior.

Monocular depth estimation is inherently brittle due to biases in training datasets, whereas SPADs provide high-accuracy sensor measurements. In this section, we leverage the less accurate monocular depth to reduce the number of SPAD bins needed for capturing data, thereby saving memory and improving depth resolution.

Simulation Details: We conducted our simulations using the SPAD simulation framework provided in Gutierrez-Barragan et al. [37, 38], utilizing the code available on GitHub. While the simulations are initialized with RGBD datasets, all "ground truth" depth images presented in this paper result from SPAD simulation on full high-resolution histograms.

Monocular depth estimation algorithms use visual cues from 2D images to infer depth information and are trained on annotated datasets such as NYU Depth v2 [69] and KITTI [31]. We employed ZoeDepth [14], a monocular depth estimator chosen for its performance and ability to produce metric depth estimates. The monocular depth is used to guide a foveation window consisting of M bins in the histogram. The window size is a hyper-parameter, with larger sizes offering better accuracy at the cost of reduced efficiency.

For effective use of the monocular estimate as a prior, it must provide metric depth, and to enhance foveation performance, it needs to be scaled to match the scene. ZoeDepth fulfills the metric depth requirement, and we ensure compatibility with the dataset through appropriate scaling and bounding.

We chose a polynomial fit for scaling, observing that a majority of points in a randomly selected subset of the monocular output for the NYUv2 dataset exhibited a linear relationship. This scaling can be performed either locally, fitting the data to a specific scene, or generally across the dataset. In both cases, a small set of pixels is sampled at full histogram resolution, and the relationship between the monocular estimate and the SPAD estimate at these pixels is modeled. The fit is then applied to the entire monocular estimate, with bounds enforced for the minimum and maximum values across the dataset, which are 0m and 10m for NYUv2.

We now describe our results shown in Fig. 4-2 and evaluated in Table 4-2 which are calibrated locally. The first two columns in the figure show the ground truth from the NYUv2 dataset. The depth is not simply the depth from the NYUv2 dataset, but the output of full-resolution SPAD simulation followed by the detection of the histogram peak. The third column shows the **scaled** monocular output.

Memory Foveation: The fifth column in Fig. 4-2 shows our memory foveation results. Here, most bins are not used, saving memory for the same SNR. The foveated window is given at the right of the figure as a fraction of the original number of bins N, with N set to 1000 bins for all experiments. The results are visually indistinguishable from ground truth, in some cases with a $\frac{1}{16}$ save in memory. In Table 4-2 we show the change in accuracy with these memory savings.

Table 4-2. Memory and Depth Foveation Evaluation - Local Scale This table shows a quantitative comparison of RMSE and depth inlier metrics for different depth and memory foveation strategies for the NYUv2 dataset and a monocular estimation prior. For each memory foveation fraction, we vary the number of histogram bins in the foveated sub-window to achieve depth foveation. Metrics used from left to right: Root-mean-squared errror, Absolute log_{10} error, Absolute Relative Error, $\delta < 1.25$, $\delta < 1.25^2$, $\delta < 1.25^3$

M (Fraction)	RMSE↓ (m)	$\log_{10}$↓ (m)	REL↓	δ_1↑ (%)	δ_2↑ (%)	δ_3↑ (%)	N′ (Num. Bins)	RMSE↓ (m)	Lim. Bins↓ RMSE (m)	$\log_{10}$↓ (m)	REL↓	δ_1↑ (%)	δ_2↑ (%)	δ_3↑ (%)
1/16	0.211	0.0106	0.0211	97.07	99.13	99.55	16	0.235	0.504	0.0173	0.0360	96.55	98.96	99.48
							32	0.211	0.250	0.0119	0.0241	97.1	99.14	99.55
							64	0.211	0.121	0.012	0.0242	96.44	99.01	99.54
1/8	0.151	0.005	0.0109	98.36	99.42	99.79	16	0.201	0.509	0.018	0.0418	97.87	99.26	99.71
							32	0.184	0.250	0.011	0.0254	98.1	99.38	99.77
							64	0.152	0.121	0.0064	0.0141	98.36	99.45	99.81
1/4	0.117	0.0032	0.00686	99.24	99.57	99.79	16	0.221	0.501	0.0326	0.0714	98.77	99.6	99.82
							32	0.166	0.2497	0.015	0.0355	99.15	99.59	99.82
							64	0.145	0.123	0.0087	0.0195	99.01	99.52	99.78

Unsurprisingly, there is an inverse relationship between memory usage and depth error.

Depth Foveation: In Fig. 4-2 the foveated window around the estimated monocular depth is packed with a limited number of bins. With no foveation, as in the fourth column, a limited number of bins $N^{'}$ are distributed over the entire SPAD volume. The depth foveation in the last column shows what happens when these limited number of bins are packed into the foveated window. Note that the depth resolution has increased from the limited bins case because the samples are placed within a foveated window where we expect to find the histogram peak. In Table 4-2, entries with the same memory usage demonstrate the effects of depth foveation, where higher depth resolution consistently produces better results. These depth foveation outcomes are directly dependent on the memory foveation results, as both algorithms place fovea windows based on the same depth prior, with the depth foveation experiments having a lower depth resolution. Meaning, the memory foveation results establish a lower bound for the depth foveation error. Additionally, the limited bins case, which is not confined to a foveated window and thus reliant on a depth prior, shows that the error continues to decrease as depth resolution increases.

4.5 Spatio-Temporal SPAD Foveation

The previous section seeks to reduce the SPAD histogram bottleneck by reducing the number of bins to examine per-pixel with a monocular estimate prior. This section aims to

Table 4-3. Spatio-Temporal Foveation Evaluation - Local Scale Here we look at a quantitative comparison between the size of the foveation window (memory usage), the number of bins in depth foveation, and the number of total samples per the spatio-temporal algorithm.

Sparsity = 0.26 %	M (Fraction)	RMSE↓ (m)	$\log_{10}$↓ (m)	REL↓	δ_1↑ (%)	δ_2↑ (%)	δ_3↑ (%)	N′ (Num. Bins)	RMSE↓ (m)	Lim. Bins↓ RMSE (m)	$\log_{10}$↓ (m)	REL↓	δ_1↑ (%)	δ_2↑ (%)	δ_3↑ (%)
	1/16	0.39	0.06	0.124	84.901	97.054	99.429	16	0.649	0.509	0.102	0.15	83.788	95.189	96.514
								32	0.687	0.251	0.103	0.151	81.556	95.272	96.972
	1/8	0.392	0.068	0.137	80.154	94.812	99.046	16	0.738	0.502	0.129	0.19	71.23	91.362	95.817
								32	1.055	0.269	0.17	0.202	69.595	89.694	92.852
	1/4	0.355	0.054	0.10	88.244	98.114	99.186	16	0.756	0.497	0.131	0.199	67.472	92.431	96.184
								32	0.837	0.25	0.137	0.202	68.609	86.771	93.232
Sparsity = 0.52 %	M (Fraction)	RMSE↓ (m)	$\log_{10}$↓ (m)	REL↓	δ_1↑ (%)	δ_2↑ (%)	δ_3↑ (%)	N′ (Num. Bins)	RMSE↓ (m)	Lim. Bins↓ RMSE (m)	$\log_{10}$↓ (m)	REL↓	δ_1↑ (%)	δ_2↑ (%)	δ_3↑ (%)
	1/16	0.414	0.07	0.12	87.672	97.008	98.139	16	0.582	0.505	0.092	0.134	86.543	96.111	97.068
								32	0.484	0.25	0.07	0.119	87.664	96.492	98.158
	1/8	0.387	0.051	0.108	87.292	99.162	99.919	16	0.518	0.519	0.071	0.136	84.049	98.177	99.255
								32	0.587	0.248	0.074	0.142	78.714	94.945	97.969
	1/4	0.38	0.049	0.0996	90.254	96.965	98.365	16	0.734	0.518	0.121	0.184	74.702	93.679	96.225
								32	0.553	0.256	0.068	0.127	85.968	96.942	98.09
Sparsity = 1.04 %	M (Fraction)	RMSE↓ (m)	$\log_{10}$↓ (m)	REL↓	δ_1↑ (%)	δ_2↑ (%)	δ_3↑ (%)	N′ (Num. Bins)	RMSE↓ (m)	Lim. Bins↓ RMSE (m)	$\log_{10}$↓ (m)	REL↓	δ_1↑ (%)	δ_2↑ (%)	δ_3↑ (%)
	1/16	0.288	0.039	0.0855	94.214	99.582	99.935	16	0.364	0.508	0.048	0.0959	93.693	99.248	99.646
								32	0.412	0.254	0.051	0.0933	93.048	98.179	99.145
	1/8	0.313	0.04	0.0881	91.782	99.443	99.841	16	0.386	0.495	0.056	0.111	90.719	99.057	99.474
								32	0.432	0.257	0.053	0.106	89.662	98.472	99.276
	1/4	0.274	0.035	0.0786	94.264	99.045	99.875	16	0.471	0.503	0.072	0.148	82.311	97.104	98.821
								32	0.399	0.25	0.063	0.111	91.482	97.966	98.643

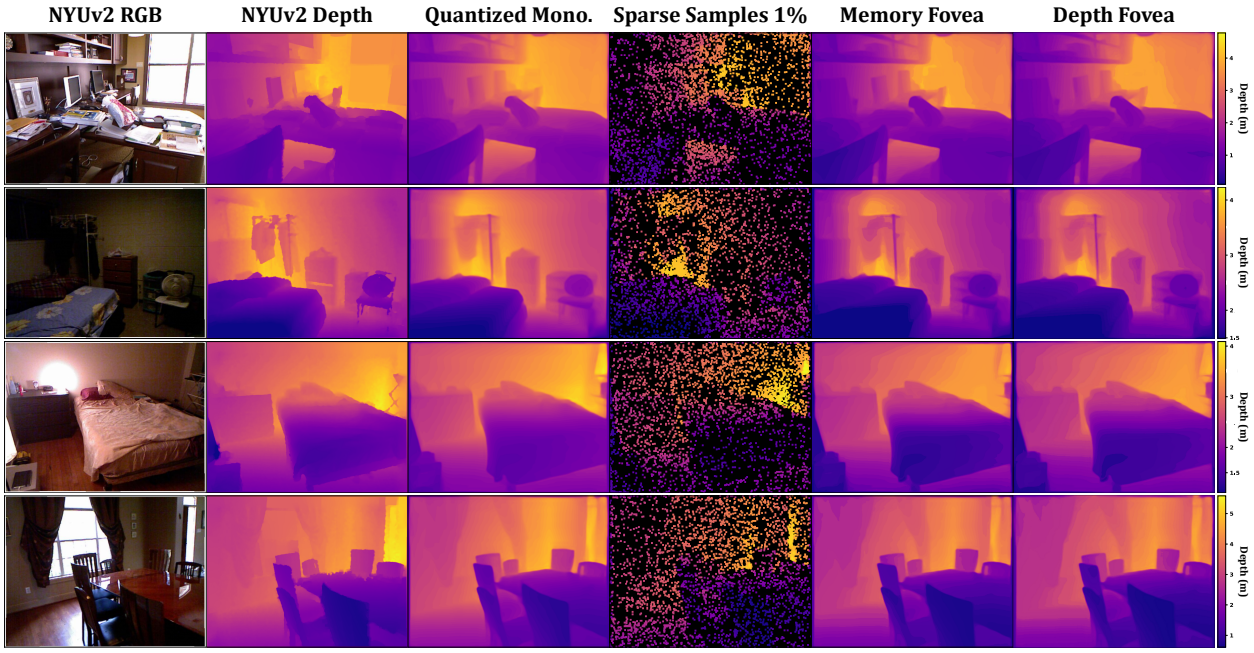


Figure 4-3. **Spatio-temporal foveation** The first two columns display the scene's color and ground truth depth. Using the quantized monocular depth in the third column, we select certain pixels in the fourth column. Processing only histograms at these locations with foveated windows generates results in the last column, indicating a 1548-fold reduction in memory usage. This is calculated by measuring memory allocation for full-res and spatio-temporal histograms. The results shown are with **M=1/16N** and **N′ = 16**

improve these savings by incorporating spatial foveation. By exploiting depth coherencies and applying foveated windows to a small selection of pixels we show an order of magnitude increased bandwidth savings.

Foveated LiDAR systems [89, 75, 12] can place samples onto depth edges and recover the rest of the scene, post-capture, through algorithmic estimation such as deep guided upsampling or gradient-based reconstruction. Similarly, here, we place samples *across* depth edges and, rather than use an algorithm, we use the SPAD measurement to provide correct depths in redundant areas.

Quantized Sampling: Our approach to spatial sampling begins by quantizing the prior through thresholding, resulting in digitized regions that we refer to as 'buckets.' We make the *assumption* that the values within each quantized bucket are redundant. From each bucket, we randomly select

pixels and use the SPAD to measure these points in the scene, applying memory foveation in the process. These measurements provide a sparse depth map, which we subsequently sort and quantize based on the buckets defined by the depth prior.

In Fig. 4-3 we show examples of our approach, where the first two columns show the scene and ground truth depths. The third column is a quantized version of the monocular depth estimation, where the number of quantized buckets is 64. For each of these buckets, we picked 50 points at random and recovered the SPAD depths of these points. Note that these transients were also foveated in time, using the method described in the previous section. The fourth column in Fig. 4-3 depicts exactly those points in the SPAD camera that were sampled, with the number of bins sampled at $\frac{1}{16}$ of the original histogram. This is a factor of 1548 memory savings, compared to the ground truth measurement, with depth results in the last column. These efficiencies are evaluated in Table 4-3.

4.6 Optical Flow Driven SPAD Foveation

In previous sections, we focused on static scenes. However, one of the key advantages of using SPAD arrays is their fast capture speed, making them ideal for dynamic environments, such as when mounted on a vehicle. In this section, we demonstrate how our techniques can be applied to moving scenes by utilizing optical flow to guide the foveation process.

Consider a SPAD sensor on a moving platform, say an autonomous vehicle, where high-frame rate and efficient depth capture are important [57, 11]. The foveation algorithm described in the previous section analyses pixels in each frame, reducing the bins in the histogram that need to be processed. Here we consider an approach to reduce the computation even further, using temporal information by transferring foveation information from previous frames to subsequent frames.

Consider a sequence of frames containing both depth and reflectance information from a scene. Assume that the depth in the first frame is reconstructed at high quality, such as from full-resolution SPAD histograms. Now, for a subsequent frame, we can calculate optical flow between the frames (color or grayscale), producing a vector (u, v) for each pixel at a given time t.

These vectors satisfy the brightness consistency principle, meaning that $I(x + u \cdot \delta t, y + v \cdot \delta t, t + \delta t) = I(x, y, t)$. holds true. We use the depth information from the previous frame to guide the positioning of the foveating window in the current frame, by warping the previous frame based on the vector (u, v). Although the object may move and the histogram peak will shift from frame to frame, it will remain within a nearby range, allowing a window of pixels to recover the histogram peak in the current frame.

However, optical flow is never perfect, often having errors at the edges of a frame. Further, these propagate incorrect depths through time, since our optical flow method only considers the depths in the previous frame. To remove this error, we compare the distribution of the photons under a foveated region to that from a noise floor. If they match, we ignore the erroneous optical flow, and recompute depth from the full histogram. In practice this is done by thresholding the values in the foveated window.

In Fig. 4-4, we show some optical flow results. These were created on the CARLA simulator [23] and the results show two street scenes with ground truth depths. We found the native optical flow in CARLA to be noisy, and so we used OpenCV's in-built optical flow estimator. The third and fourth columns show first the incorrect results from optical flow, and our method to detect these regions, shown in red. The optical flow driven depth foveation results are shown in the last column. Calculating errors using a running average across all video frames reveals compounding errors over time. In the first scene, at $\frac{1}{10}N$, RMSE and SSIM are 101.9m and 0.530, and at $\frac{1}{4}N$, 38.6m and 0.884. In the second scene, RMSE and SSIM are 0.164m and 0.87 for both $\frac{1}{10}N$ and $\frac{1}{4}N$.

4.7 Hardware Emulation Results

In this section, we present hardware emulation results for depth and memory foveation using SPAD data captured using real hardware. The goal of hardware emulation study is to de-risk future in-pixel implementations of foveation algorithms. We use datasets by Lindell et al. [58] and Gutierrez-Barragan et al.[38] from prior sources [34, 35].

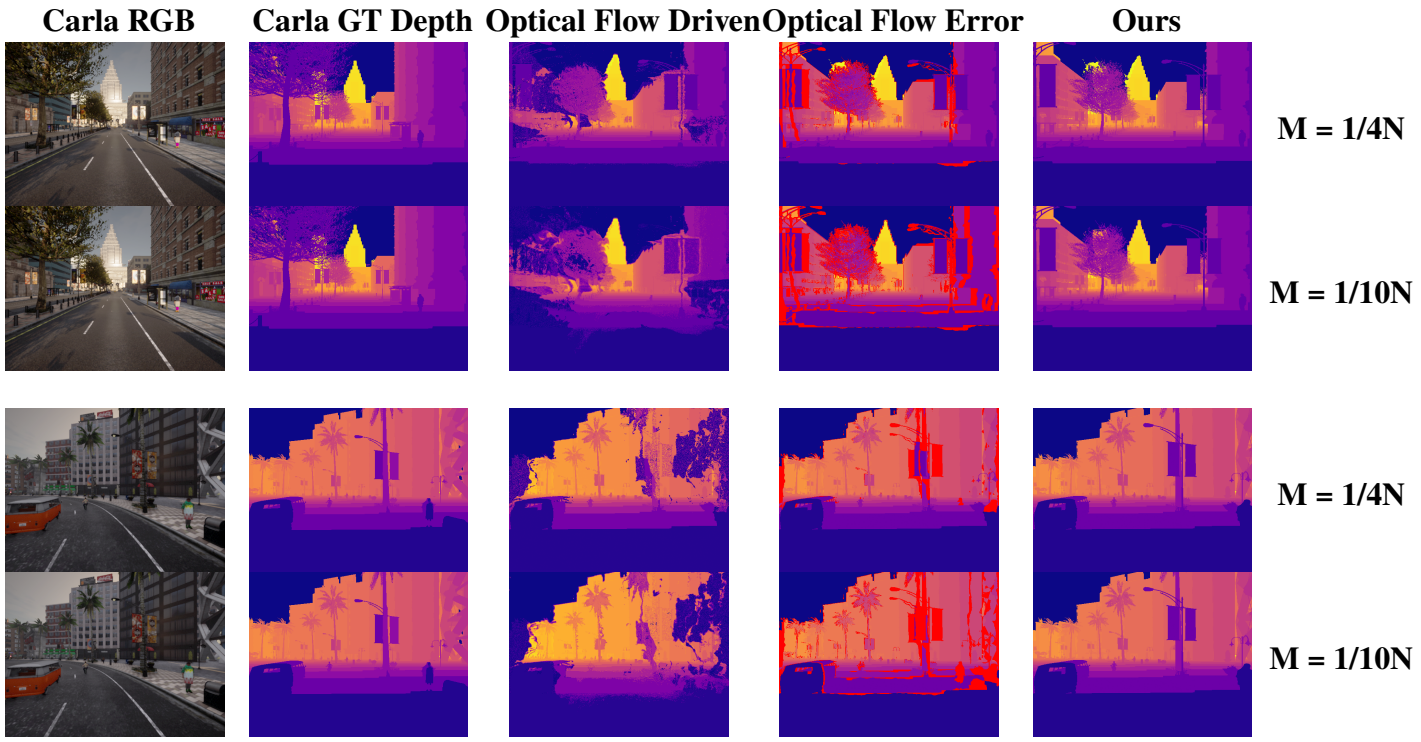


Figure 4-4. **Optical Flow Driven Foveation** Here we see our optical flow driven SPAD foveation using the Carla simulator whose color and ground-truth depth are shown in the first two columns. Directly using optical flow, as shown in the third column, creates errors that propagate over time. We correct for the optical flow error by detecting those pixels whose foveated windows are close to the noise floor. The last column shows the final optical flow driven foveated depth at different window sizes.

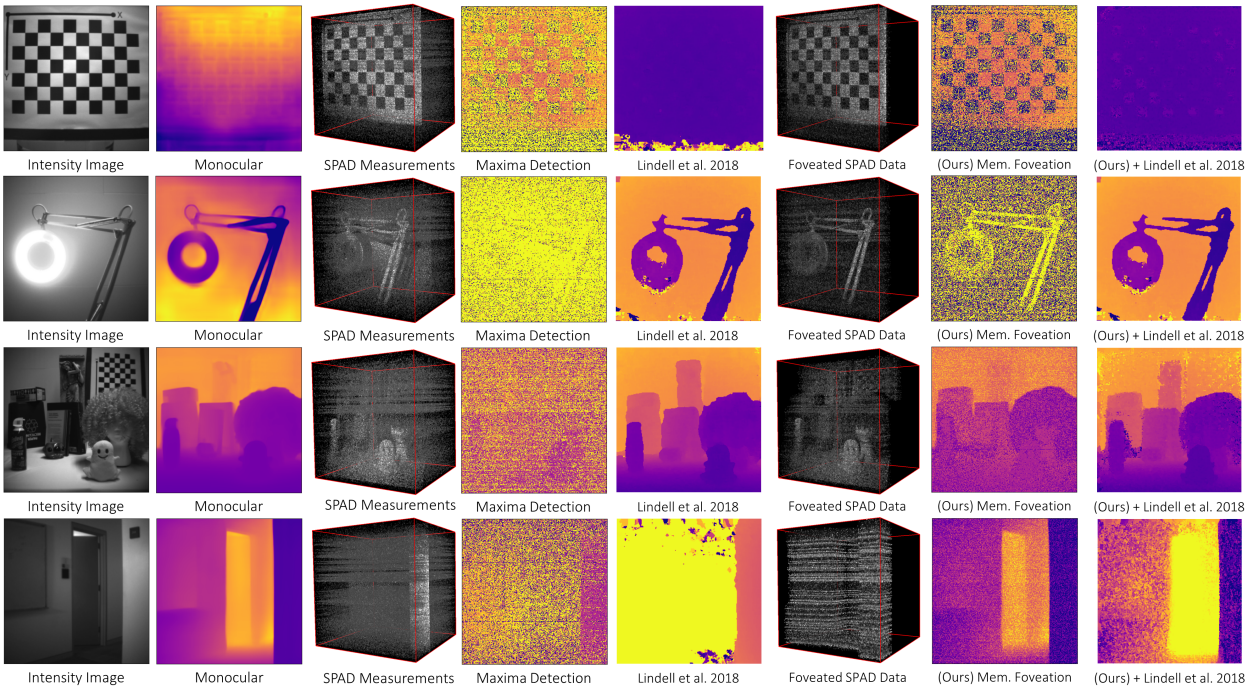


Figure 4-5. **Hardware emulation results for scenes from Lindell et al. [58].** (Column 1) The Lindell dataset consists of monochrome images captured by a camera co-aligned with the SPAD sensor that captures photon data cubes. (Column 2) We obtain monocular depth maps from these monochrome images. (Column 3) Raw photon data cube without foveation shows a "cloud" of background photon detections. (Column 4) Maxima detection on low SBR photon clouds leads to unusable depth maps. (Column 5) The CNN-based algorithm of Lindell et al. improves depth map reconstruction. (Column 6) Our approach relies on memory foveation in a 1/4th size sub-window around an estimate of the true depth obtained from monocular depth maps. Observe that the photon data cubes are less noisy. (Column 7) Even a simple max-estimator provides better depth map estimates after foveation. (Column 8) Providing foveated clouds to the CNN denoiser of Lindell et al. further improves reconstructions.

4.7.1 Using Monocular for Memory Foveation

We'll start by showcasing how our memory foveation technique works on the dataset by Lindell et al. [58] by using monocular as a prior. The Lindell dataset consists of scenes under different ambient illumination conditions captured using a linear SPAD pixel array [16] co-aligned with a monochrome camera that captures intensity images.

We use these intensity images to obtain a monocular depth prior. Because the performance of monocular estimation networks is dependent on the dataset, we perform a calibration step by using the "elephant" scene in the dataset to define a global scaling function. We place foveation

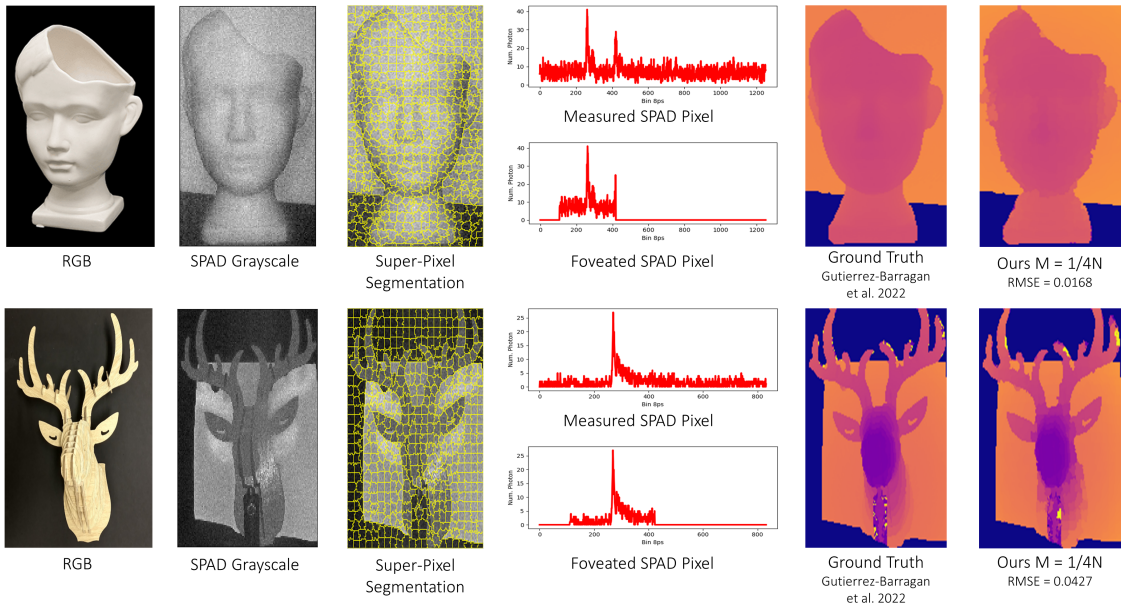


Figure 4-6. **Hardware emulation results for scenes without co-aligned monochrome camera [38].** (Column 1) RGB images of the "face-vase" and "reindeer" scenes shown for visualization. (Column 2) A pseudo-intensity image is estimated by accumulating photon counts for each pixel. (Column 3) Pseudo intensity maps are converted into superpixel representations, and a single pixel in each superpixel is used for measuring complete histograms. (Column 4) The peak location of the chosen pixel is used to apply foveation windows of 1/4th the total temporal extent for the remaining pixels in each superpixel. (Column 5) Ground truth depth maps obtained using matched filtering. (Column 6) Our result requires 64× less memory per pixel for > 99% of the pixels in these scenes.

windows of 1/4th the total temporal extent of the full histograms centered around these scaled monocular depth estimates for each pixel.

Memory foveation improves the overall SBR, in a scene-adaptive manner, by focusing on regions of the spatio-temporal photon cube where signal photons arrive. Comparing columns 3 and 6 in Fig. 4-5, foveated SPAD measurement cubes show fewer background photon detections, with clear 3D object structure in the photon cubes. Depth estimates are improved even with a simple maxima-detection approach — observe that the lamp is barely visible in the non-foveated maxima-detection-based depth map in column 5, but is visible after memory foveation in column 7. Running memory foveated measurements through the denoising algorithm of Lindell et al. further improves the depth map, as seen in the last column of Fig. 4-5.

4.7.2 A Different Approach to Spatio-Temporal Foveation

To illustrate the flexibility of our foveation techniques and their independence from external sensors as a prior, we propose an alternative spatio-temporal method, which we apply to two scenes from the Gutierrez-Barragan et al. dataset [38], for which there is no co-located camera. The dataset is captured using a single-pixel point scanned SPAD detector co-aligned with a pulsed laser. Fig 4-6 shows the results of the alternate approach for the single object "face-vase" and "reindeer" scenes, with the RGB images shown in column 1 for visualization purposes.

SuperPixels: Because there is no intensity map captured in the dataset, we instead obtain a pseudo-intensity map by summing the raw photon data cubes along the temporal axis for each pixel. In a real hardware implementation, this process would be achieved by utilizing a counter in each SPAD pixel, a feature commonly available in existing commercial SPAD arrays. We then run a superpixel algorithm [1] on the pseudo-intensity maps to obtain coarse segmentations of the scene, as shown in column 3. For each superpixel segment, we capture a complete (non-foveated) histogram of the centroid pixel. By identifying the true peak location in this histogram, we can then foveate within a 1/4th sub-window centered around this peak for all remaining pixels in the superpixel segment, reducing the overall bandwidth requirement per pixel by a factor of 64.

In the "face-vase" scene, with a spatial resolution of 174×154 pixels, the segmentation reduces the data to 473 superpixels. Similarly, the "deer" scene, originally at 204×116 pixels, is reduced to 515 superpixels. This reduction translates to a 3/4 reduction in memory requirement for approximately 99.98% pixels in both scenes. Examples of foveated histograms in column 4 show that the laser impulse response function has a non-ideal shape which departs significantly from the commonly assumed Gaussian shape used in simulation studies. (The second peak is likely due to optical inter-reflections in the hardware setup). Yet, our method is able to produce reliable depth maps (columns 5 and 6).

We also examine the impact of reconstruction error under increasing background noise for the "deer" scene. As shown in Fig. 4-9, foveation allows for the accurate selection of the correct depth peak, even in the presence of strong background illumination, thereby expanding the

operable SBR range in practice.

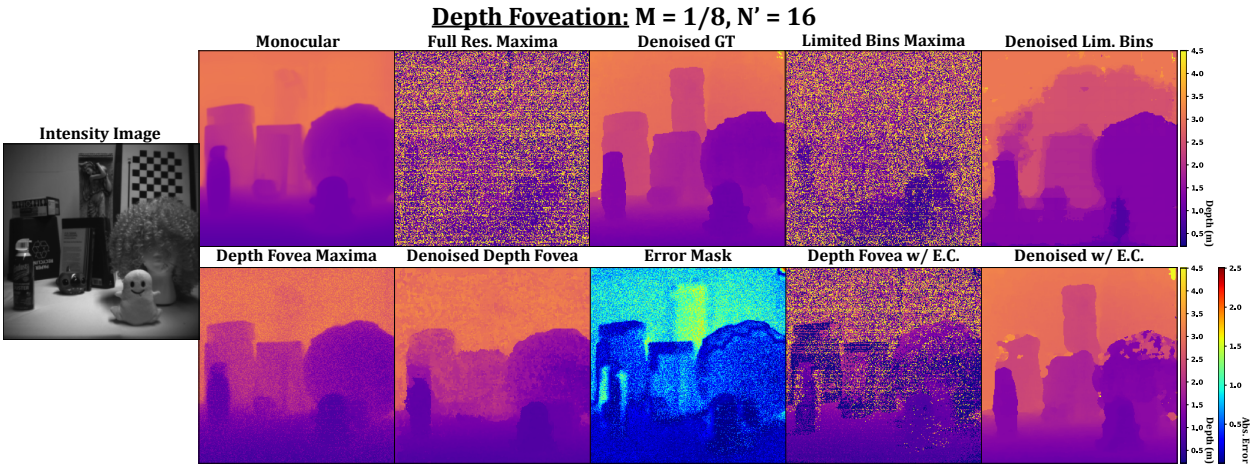


Figure 4-7. **Additional Results: Depth Fovea**This figure demonstrates the application of the depth foveation technique described in Sec. 4.4 to the Lindel dataset, along with the error correction technique presented in the appendix material. A window size of M = 1/8 and a bin count of N' = 16 were used. The results were subsequently processed using the sensor fusion denoising network [58].

4.8 Worst Case Stochastic Limits

In this section, we characterize the worst case scenario where depth is wrongly detected by a foveated SPAD pixels. This lower bound helps us understand the limits of the approach. However, it is different from a best or average case analysis, which would be useful for deployment, and we leave such analysis to future work.

Foveation errors in our framework may be due to monocular depth calibration errors, ambient light, and global effects such as multi-bounce inter-reflections. In these scenarios, the window predicted by foveation may not overlap with the expected transient peak. To characterize these errors, we use an analysis method described in Gupta et al. to find the probability that a peak will be detected in the set of foveated bins.

Consider the initial foveation window of M bins for all the S pixels in the camera. We define p_{gt} as the probability that a detected photon originated from the laser dot that illuminates the scene point of interest. We also defined $p_{\text{multipath}}$ as the probability that a detected photon experienced multipath bounces and p_{floor} as the probability that the sensor noise does not create spurious peaks.

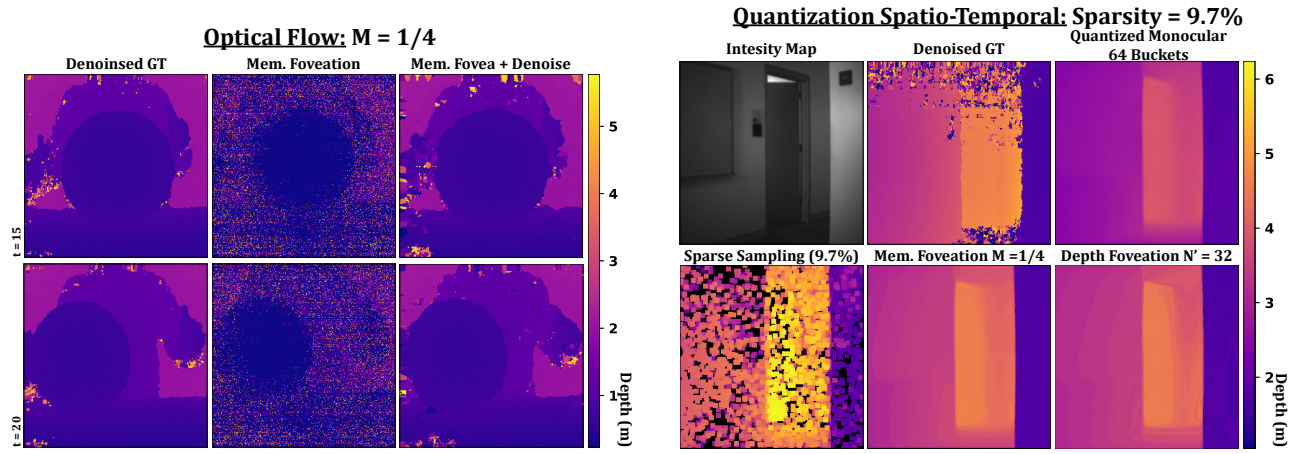


Figure 4-8. **Additional Results: Optical Flow and Quantization Spatio-Temporal** This figure illustrates the application of the techniques described in Sec. 4.6 and Sec. 4.5 to the Lindel dataset. The left portion showcases our optical flow algorithm on the "roll" scene. The first column displays the denoised ground truth, followed by the optical-flow-driven memory foveation result using maxima detection, and finally the denoised memory foveation result. The right portion of the figure presents our quantization spatio-temporal foveation technique, utilizing 9.7% sampling to mitigate the high levels of noise and the abundance of pixels with no photon counts in the scene.

First, we consider the probability that the direct, single bounce photon from the laser to the scene point was detected in the M binned foveation window — this is the definition of p_{gt}. We also consider photons from the laser that experience multipath effects, which we model as $p_{\text{gt}}p_{\text{multipath}}$. The foveation window must have none of these multipath photons from any of the other $M-1$ bins that originate at the laser. Further, the noise floor must be low for this detection. In other words, the probability of peak detection is

$$p_{\text{gt}}(1-p_{\text{gt}}p_{\text{multipath}})^{M-1}p_{\text{floor}}. \tag{4-9}$$

We now model the worst case scenario, where none of the S pixels get the correct foveated depth. The chances that this happens are:

$$p_{\text{worst}} = (1-p_{\text{gt}}(1-p_{\text{gt}}p_{\text{multipath}})^{M-1}p_{\text{floor}})^{S}. \tag{4-10}$$

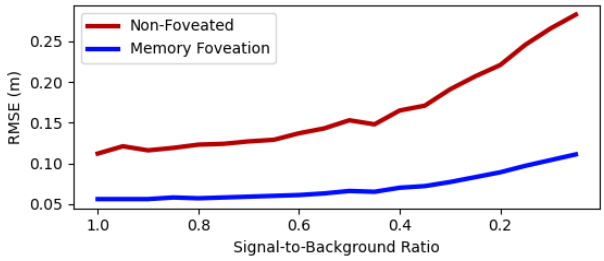


Figure 4-9. **Effect of increasing background illumination.** The conventional (non-foveated) depth map quality degrades more rapidly as background illumination increases. Using memory foveation allows reliable depth map recovery for the "deer" scene for a wider range of SBR levels.

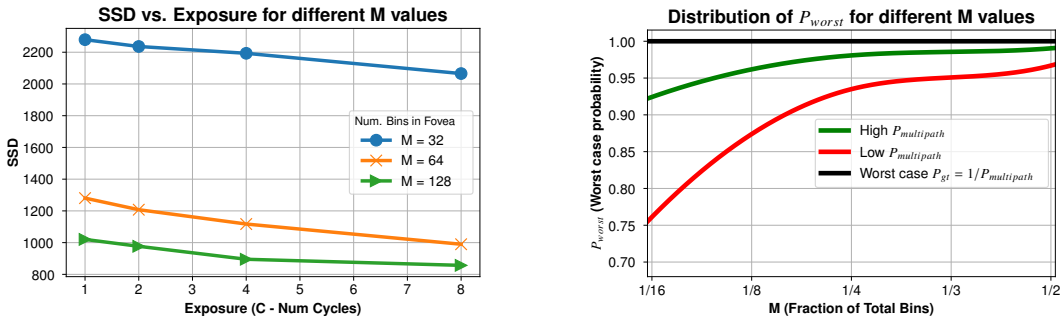
As in Gupta et al. , we set $\frac{\delta p_{\text{worst}}}{\delta p_{\text{gt}}} = 0$ to analyze when this worst case probability is maximized. As we show in appendix A, this simplifies to the following:

$$\begin{aligned} &S(1 - p_{\text{gt}}(1 - p_{\text{gt}}p_{\text{multipath}})^{M-1}p_{\text{floor}})^{S-1} \cdot \\ &(-p_{\text{floor}})((1 - p_{\text{gt}}p_{\text{multipath}})^{M-2} \cdot \\ &((1 - Mp_{\text{gt}}(p_{\text{multipath}})) = 0 \end{aligned} \tag{4-11}$$

We now explain how this relation can be used in practice. Recall that p_{gt} is the probability that laser photons are detected, i.e. the chances that accurate depth recovery occurs. Only two values of p_{gt} make the above worst case relation zero. The first term to zero out the relation is that $p_{\text{gt}} = \frac{1}{p_{\text{multipath}}}$. From the definition of probability, this is only possible if the probability is 1 for every bin to have both photons from the laser and have multipath effects — i.e. the scene is degenerate, such as made entirely from mirror BRDFs.

The second possibility happens when $p_{\text{gt}} = \frac{1}{M\ p_{\text{multipath}}}$, where the number of bins M and the probability of multipath effects $p_{\text{multipath}}$ vary under the condition that $0 \leq p_{\text{gt}} \leq 1$. This suggests that heuristics to avoid the worst case, where ideal bands of foveated windows M can be used for scenes with particular global illumination characteristics denoted by $p_{\text{multipath}}$.

As an example, consider a set of bins $M = 1000$. Consider a situation where multipath effects are very low, and $p_{\text{multipath}} = 0.001$. In this scenario, the probability of accurate depth



(a) (b)

Figure 4-10. **Eq. 4 and 10 validation:** (a) Depth foveation reduces bin width, reducing SNR. Increasing exposure can compensate for this SNR decrease (and improve the sum-squared difference SSD). (b) The red and green curves show the upper bound on p_{worst} from Eq. 10. These are generated based on nominal and worst case distributions of $p_{\text{multipath}}$, with $p_{\text{gt}} = \frac{1}{M\ p_{\text{multipath}}}$.

recovery is $p_{\text{gt}} = 1$, which is the case in our simulated results where there are no multi-path effects. However, if the multipath effects are, say one in ten, then $p_{\text{gt}} = 0.1$ then the probability of depth recovery falls, in the worst case to $p_{\text{gt}} = 0.01$. Attempting to improve the probability of detecting laser photons p_{gt} by varying the number of bins cannot be done without reducing depth resolution.

In Fig. 4-10(a) we show a verification of Eq 4 from the main text. We varied the exposure, foveation interval M, and computed SSD for one scene. These simulations show that depth quality does not increase linearly with increase in foveation bins, but does so with exposure, as predicted by Eq 4. In 4-10(b), we have also shown verification for Eq. 10 for the degenerate mode in black ($p_{\text{worst}} = 1$) and for the recommended mode $p_{\text{gt}} = \frac{1}{M\ p_{\text{multipath}}}$. p_{gt} and p_{floor} were modeled as Gaussians and $p_{\text{multipath}}$ is shown for two cases, high (green) and low (red). As the graph shows, with lower probabilities of $p_{\text{multipath}}$, tighter foveation intervals are possible even in these upper bounds of worst cases.

Given this worst-case analysis, we now tackle data from real SPAD sensors and scenes, which have noisy histogram floors, inter-reflections and other complex effects.

4.9 Limitations and Discussion

Worst Case Stochastic Limits: We explored the limitations of our approach by analyzing the worst-case scenario where depth is incorrectly detected due to various errors, such as monocular

depth calibration issues, ambient light interference, and global effects like multipath inter-reflections. We characterized these errors using a probabilistic framework. Specifically, we defined the probability p_{gt} as the chance that a detected photon originates from the laser i.e. single-bounce photons, $p_{\text{multipath}}$ as the probability of multipath photon detection, and p_{floor} as the probability of spurious peaks due to sensor noise. The overall probability of accurate depth detection is given by

$$p_{\text{gt}}(1 - p_{\text{gt}}p_{\text{multipath}})^{M-1}p_{\text{floor}}, \tag{4-12}$$

where M is the number of foveated bins. We further derived the probability p_{worst} for the worst-case scenario, where none of the S pixels detect the correct depth, expressed as

$$p_{\text{worst}} = (1 - p_{\text{gt}}(1 - p_{\text{gt}}p_{\text{multipath}})^{M-1}p_{\text{floor}})^{S}. \tag{4-13}$$

Through optimization, we identified two conditions that lead to this worst-case scenario, linked to specific relationships between p_{gt}, $p_{\text{multipath}}$, and M.

- The first condition occurs when $p_{\text{gt}} = \frac{1}{p_{\text{multipath}}}$. This situation arises when the probability is 1 for every bin to contain both direct photons from the laser and photons that have undergone multipath effects, indicating a degenerate scene, such as one made entirely of mirror-like surfaces.

- The second condition occurs when $p_{\text{gt}} = \frac{1}{M \cdot p_{\text{multipath}}}$. This scenario implies that the number of foveated bins M and the probability of multipath effects $p_{\text{multipath}}$ must satisfy this relationship, under the constraint that $0 \leq p_{\text{gt}} \leq 1$. This suggests that it is possible to avoid the worst-case scenario by adjusting the number of bins M for scenes with specific global illumination characteristics.

In order to illustrate the findings of this analysis, consider a toy example with a number of bins $M = 1000$ and pronounced multipath effects, such as $p_{\text{multipath}} = 0.1$. In the worst case, the

probability of depth recovery would be significantly hindered $p_{\text{gt}} = 0.01$, but can be improved by changing the number of bins M at the cost of depth resolution. The detailed derivations of these results are provided in the appendix.

Quality of depth priors: Our algorithms can enable memory-efficient SPAD sensing while maintaining depth accuracy. However, our method strongly relies on the accuracy of the depth prior. If the prior is incorrect, our algorithms may produce errors, highlighting the importance of robust error correction mechanisms. We can correct for such errors by trading off efficiency. We show one example error mask in the appendix which can be used to drive corrections, such as a larger foveation window (using the entire span of the transient in the extreme case).

Hardware complexity: A key limitation of our approach is the lack of available hardware that fully supports our algorithms, necessitating more complex pixel architectures and driving up costs. Each SPAD pixel in the 2D array requires a programmable gate, along with a variable TDC and histogrammer, which increases the complexity and expense of the hardware. This presents a significant challenge to the widespread adoption and practical implementation of our method. In Fig. 4-11, we propose a potential array design with per-pixel gating capability, where a global ramp generator provides individualized on/off thresholds for each pixel. To enhance the fill factor, the TDC and histogrammer are shared among groups of neighboring pixels, forming "macropixels".

We believe the next generation of programmable and software-defined SPAD cameras [6, 87] will be key enablers for in-pixel and on-chip implementation of memory- and energy-efficient foveated sensing schemes. As SPAD cameras become low-cost and widely available [17], the integration of in-pixel foveated sensing algorithm proposed here will reduce memory consumption while maintaining depth accuracy, or alternatively, provide more accurate depth estimates without increasing memory usage.

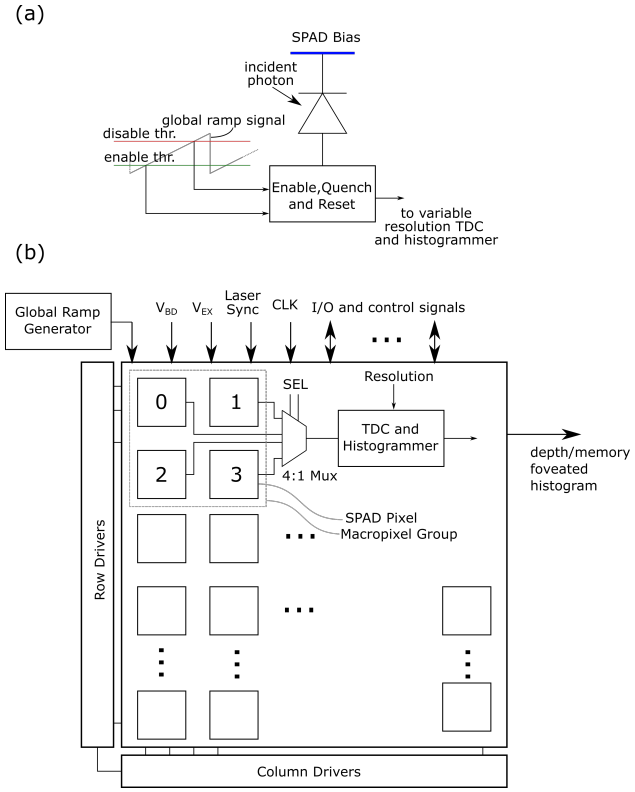


Figure 4-11. **Future pixel and array designs for foveated single-photon 3D imaging.** (a) A speculative pixel design where individual SPADs are gated on or off based on thresholds set with respect to a linear ramp signal. Pixels only need to store the thresholds; the ramp signal is generated externally. (b) A possible array of SPAD pixels with per-pixel gating. Observe that the ramp signal is generated globally, simplifying pixel design. Variable-resolution TDCs and histogrammers are shared by small pixel neighborhoods (e.g., 2×2 multiplexed "macropixels") to improve fill factor.

CHAPTER 5
SUMMARY AND CONCLUSION

Throughout this dissertation, the concept of optimizing depth sensing through principles inspired by biological foveation has been thoroughly explored. Each chapter provided unique insights and advancements that contribute to this overarching theme. Here is a synthesis of the topics discussed and a forward-looking perspective on their implications.

Our work on **RGB-Guided Foveated Depth Sensing** in Chapter 2 demonstrated the potential of dynamically adjusting LIDAR sampling patterns using a MEMS mirror. By integrating deep learning methods for depth completion, we showcased a system capable of achieving high-resolution depth sensing in regions of interest while reducing the overall computational load [75]. This approach is particularly relevant for applications like autonomous vehicles, where the balance between resource efficiency and accuracy is critical. However, there are still areas for future work, such as enhancing the real-time adaptability of the system and integrating more complex scene understanding algorithms to refine the foveation process. Expanding the system's robustness to diverse environmental conditions would also further its real-world applicability.

In **Chapter 3**, we extended the principles of foveated depth sensing to underwater environments, where traditional sensing methods often struggle. The **Bistatic Confocal LIDAR** system developed here, utilizing MEMS mirrors for both the transmitter and receiver, demonstrated how foveated sampling could mitigate the challenges posed by light scattering in turbid water [29]. This contribution has significant implications for underwater navigation and exploration, where accurate and efficient depth sensing is crucial. Future work could focus on improving the robustness of the system in more extreme underwater conditions and integrating real-time environmental adaptation algorithms to handle the dynamic nature of aquatic environments.

In **Chapter 4**, we explored **Spatio-Temporal Foveated Depth Sensing** for SPAD-based systems, leveraging depth priors such as monocular depth estimation to reduce memory and computational overhead while preserving accuracy in key regions. This work paves the way for

SPAD sensors to be used in real-time 3D imaging applications where resource constraints are a concern. However, pushing the limits of this approach in more complex and dynamic environments will require further exploration. Integrating more advanced machine learning models to predict regions of interest in real time could improve the adaptability of the system, making it even more suitable for real-world applications.

In conclusion, the adaptive, foveation-inspired depth sensing systems presented in this dissertation hold significant potential for revolutionizing a variety of fields, from autonomous vehicle navigation to underwater exploration and 3D imaging. By drawing on principles of biological vision, these systems offer a compelling blend of efficiency, adaptability, and precision. While substantial progress has been made, there remain exciting avenues for future work. This includes refining hardware systems, exploring new applications in fields such as robotics and augmented reality, and further integrating deep learning techniques to enhance adaptive sensing capabilities. The journey ahead promises further advancements and breakthroughs, ensuring that the next generation of depth sensing systems will be more intelligent, efficient, and responsive to the demands of their environments.

APPENDIX

APPENDIX: RGB GUIDED FOVEATED DEPTH SENSING FOR IMPROVED MONOCULAR ESTIMATION

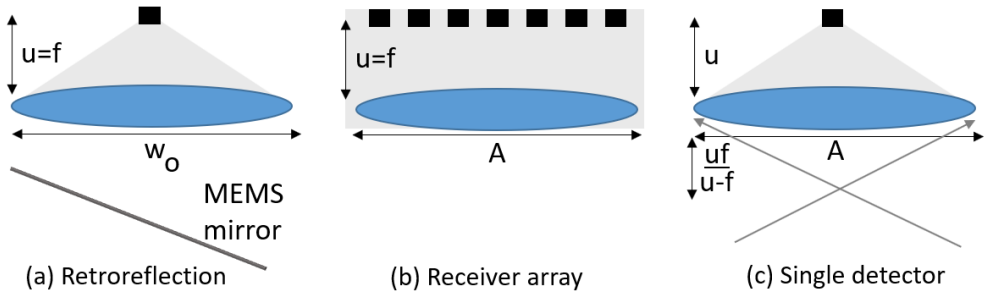


Figure A-1. Ray diagrams of designs

A.1 Derivations

Here we derive all the formulae in A-1 for the three designs. We have provided the ray diagrams of the designs in Fig. A-1 and have reproduced the table here.

A.1.1 Volume

For the retroreflection and single detector, the volume of the camera is a cone whose vertex is the location of the single detector. From the ray diagrams and from the equation of the volume of a cone, this is easily seen to be $\frac{\pi u w_o^2}{12}$ for the retroreflector and $\frac{\pi u A^2}{12}$ for the single detector. For the receiver array, the volume is the entire enclosure, given by the volume of a cuboid, $u * A * A$.

A.1.2 FOV

The retroreceiver has the exact same FOV as the mirror, by definition. From Fig. A-1(b), the FOV of the receiver array is given by the vertex angle of the cone at the central pixel, given by $2 \arctan\left(\frac{A}{2u}\right)$, bounded by the FOV of the mirror. This assumes the receiver and transmitter are close enough to ignore angular overlap issues.

To find the FOV of the single detector, consider the diagram in Fig. A-1(c), where the single detector is focused on the laser dot at distance Z from the sensor. From similar triangles, the kernel size is given by first finding the in-focus plane at u' from the lens equation:

$$\frac{1}{f} = \frac{1}{u'} + \frac{1}{Z} \tag{A-1}$$

Design	Volume	FOV	Received Radiance
Retroreflection	$\frac{\pi\ u\ w_o^2}{12}$	= MEMS FOV ω_{mirror}	$\frac{atan(\frac{w_O}{2Z})}{\omega_{\text{laser}} Z tan(\frac{\omega_{\text{laser}}}{2})}$
Receiver array	$u\ A^2$	$\min(2\ atan(\frac{A}{2u}), \omega_{mirror})$	$\frac{1}{2\ Z\ tan(\frac{\omega_{laser}}{2})}$
Single detector *Conventional* $(u \geq f)$ ***Ours*** $(u < f)$	$\frac{\pi\ u\ A^2}{12}$	$\min(2atan(\frac{A(Z-f)\|\|\frac{Zu-fu-fZ}{Z-f}\|\|}{2ufZ}), \omega_{mirror})$	$\frac{1}{4Zatan(\frac{A(Z-f)\|\|\frac{Zu-fu-fZ}{Z-f}\|\|}{2ufZ})tan(\frac{\omega_{\text{laser}}}{2})}$

Table A-1. Receiver models (please see the appendix for derivations)

and so $u^{'} = \frac{fZ}{(Z-f)}$. From the two vertex-shared similar triangles on the left of the lens, we now have an expression for the kernel size:

$$kersize = \left| u - u^{'} \right| \cdot \frac{A}{u^{'}} \tag{A-2}$$

Substituting the value of $u^{'}$, we get an expression for:

$$kersize = \left| u - \frac{fZ}{Z-f} \right| \cdot \frac{A}{\frac{fZ}{Z-f}} \tag{A-3}$$

$$= \frac{A(Z-f)\,|Zu - f(u+Z)|}{fZ\,|Z-f|} \tag{A-4}$$

$$= \frac{A(Z-f)\left|\frac{Zu-f(u+Z)}{Z-f}\right|}{fZ} \tag{A-5}$$

APPENDIX

APPENDIX: SPATIO-TEMPORAL FOVEATED DEPTH SENSING FOR BANDWIDTH LIMITATIONS IN SPAD CAMERAS

B.1 Worst-Case Analysis

We set $\frac{\delta p_{\text{worst}}}{\delta p_{\text{gt}}} = 0$ to analyze when this worst case probability is maximized:

$$\begin{aligned} &S(1 - p_{\text{gt}}(1 - p_{\text{gt}}p_{\text{multipath}})^{M-1}p_{\text{noise}})^{S-1} \cdot \\ &\frac{\delta}{\delta p_{\text{gt}}}(1 - p_{\text{gt}}(1 - p_{\text{gt}}p_{\text{multipath}})^{M-1}p_{\text{noise}}) = 0 \end{aligned} \tag{B-1}$$

$$\begin{aligned} &S(1 - p_{\text{gt}}(1 - p_{\text{gt}}p_{\text{multipath}})^{M-1}p_{\text{noise}})^{S-1} \cdot \\ &(0 - \frac{\delta}{\delta p_{\text{gt}}}(p_{\text{gt}}(1 - p_{\text{gt}}p_{\text{multipath}})^{M-1}p_{\text{noise}}) = 0 \end{aligned} \tag{B-2}$$

$$\begin{aligned} &S(1 - p_{\text{gt}}(1 - p_{\text{gt}}p_{\text{multipath}})^{M-1}p_{\text{noise}})^{S-1}(-p_{\text{noise}}) \cdot \\ &(1.(1 - p_{\text{gt}}p_{\text{multipath}})^{M-1} + p_{\text{gt}}\frac{\delta}{\delta p_{\text{gt}}}(1 - p_{\text{gt}}p_{\text{multipath}})^{M-1}) = 0 \end{aligned} \tag{B-3}$$

$$\begin{aligned} &S(1 - p_{\text{gt}}(1 - p_{\text{gt}}p_{\text{multipath}})^{M-1}p_{\text{noise}})^{S-1}(-p_{\text{noise}}) \cdot \\ &((1 - p_{\text{gt}}p_{\text{multipath}})^{M-1} + (M-1)p_{\text{gt}} \cdot \\ &(1 - p_{\text{gt}}p_{\text{multipath}})^{M-2}\frac{\delta}{\delta p_{\text{gt}}}(1 - p_{\text{gt}}p_{\text{multipath}})) = 0 \end{aligned} \tag{B-4}$$

$$\begin{aligned} &S(1 - p_{\text{gt}}(1 - p_{\text{gt}}p_{\text{multipath}})^{M-1}p_{\text{noise}})^{S-1}(-p_{\text{noise}}) \cdot \\ &((1 - p_{\text{gt}}p_{\text{multipath}})^{M-1} + (M-1)p_{\text{gt}} \cdot \\ &(1 - p_{\text{gt}}p_{\text{multipath}})^{M-2}(-p_{\text{multipath}})) = 0 \end{aligned} \tag{B-5}$$

$$
\begin{aligned}
&S(1-p_{\text{gt}}(1-p_{\text{gt}}p_{\text{multipath}})^{M-1}p_{\text{noise}})^{S-1}\cdot \\
&(-p_{\text{noise}})((1-p_{\text{gt}}p_{\text{multipath}})^{M-2}\cdot \\
&((1-p_{\text{gt}}p_{\text{multipath}})+(M-1)p_{\text{gt}}(-p_{\text{multipath}}))=0
\end{aligned}
\tag{B-6}
$$

$$
\begin{aligned}
&S(1-p_{\text{gt}}(1-p_{\text{gt}}p_{\text{multipath}})^{M-1}p_{\text{noise}})^{S-1}\cdot \\
&(-p_{\text{noise}})((1-p_{\text{gt}}p_{\text{multipath}})^{M-2}\cdot \\
&((1-Mp_{\text{gt}}(p_{\text{multipath}}))=0
\end{aligned}
\tag{B-7}
$$

B.2 Memory Usage

The memory usage experiments in the Append. Table B-1 and in all experiments throughout main text, full resolution indicates that the total number of bins $M = 1000$. Memory calculations were with simulation results using the NYUv2 dataset, the histogram images at full resolution being of size $(640, 480, 1000)$.

Append. Table B-2 shows the memory usage for the spatio-temporal algorithm in experiments varying the number of sampled pixels.

Table B-1. **Memory Usage:** Memory Foveation experiments.

Histogram Resolution	Memory (MB)
Full	2343.75
1/4	585.94
1/8	292.97
1/16	145.31

B.3 Error Masks for Memory Foveation

In Append. Fig. B-1 we show example error masks for two different scenes from NYUv2 dataset shown in Fig. 2 in the main text. We spatially downscale the true depth maps by a factor of 4× and compute depth errors in our memory-foveated results. Observe that this reveals some regions around boundaries and object edges where the error is large. This information can be used

Table B-2. **Memory Usage:** Spatio-Temporal experiments at 1/16 M

Num. Pixels	Num. Buckets	Memory (MB)	% of Total Pix.
50	32	0.76	0.52%
100	32	1.51	1.04%
500	32	7.57	5.21%
50	64	1.51	1.04%
100	64	3.03	2.08%
500	64	15.14	10.42%

in a feedback loop by changing the foveation strategy and drive the error down. For example, we can use a wider foveation window.

Figure B-1. **Error Masks.** The absolute distance errors for two scenes from the NYUv2 dataset show depth errors around object edges. Brighter pixels show higher absolute error for memory foveation.

LIST OF REFERENCES

[1] Radhakrishna Achanta and Sabine Susstrunk, *Superpixels and polygons using simple non-iterative clustering*, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.

[2] Supreeth Achar, Joseph R Bartels, William L Whittaker, Kiriakos N Kutulakos, and Srinivasa G Narasimhan, *Epipolar time-of-flight imaging*, ACM Transactions on Graphics (TOG) **36** (2017), no. 4, 37.

[3] Derya Akkaynak and Tali Treibitz, *Sea-thru: A method for removing water from underwater images*, Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2019, pp. 1682–1691.

[4] Rachel Albert, Anjul Patney, David Luebke, and Joohwan Kim, *Latency requirements for foveated rendering in virtual reality*, ACM Transactions on Applied Perception (TAP) **14** (2017), no. 4, 1–13.

[5] Ibraheem Alhashim and Peter Wonka, *High quality monocular depth estimation via transfer learning*, arXiv preprint arXiv:1812.11941 (2018).

[6] Andrei Ardelean, *Computational imaging spad cameras*, Ph.D. thesis, EPFL, 2023.

[7] C. S. Bamji, P. O'Connor, T. Elkhatib, S. Mehta, B. Thompson, L. A. Prather, D. Snow, O. C. Akkaya, A. Daniel, A. D. Payne, T. Perry, M. Fenton, and V. Chan, *A 0.13 um cmos system-on-chip for a 512 × 424 time-of-flight image sensor with multi-frequency photo-demodulation up to 130 mhz and 2 gs/s adc*, IEEE Journal of Solid-State Circuits **50** (2015), no. 1, 303–319.

[8] Joseph R Bartels, Jian Wang, William Whittaker, and Srinivasa G Narasimhan, *Agile depth sensing using triangulation light curtains*, Proceedings of the IEEE International Conference on Computer Vision, 2019, pp. 7900–7908.

[9] Ramy Battrawy, René Schuster, Oliver Wasenmüller, Qing Rao, and Didier Stricker, *Lidar-flow: Dense scene flow estimation from sparse lidar and stereo images*, 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE, 2019, pp. 7762–7769.

[10] Jacques M Beckers, *Adaptive optics for astronomy: principles, performance, and applications*, Annual review of astronomy and astrophysics **31** (1993), no. 1, 13–62.

[11] Maik Beer, Olaf M Schrey, Jan F Haase, Jennifer Ruskowski, Werner Brockherde, Bedrich J Hosticka, and Rainer Kokozinski, *Spad-based flash lidar sensor with high ambient light rejection for automotive applications*, Quantum Sensing and Nano Electronics and Photonics XV, vol. 10540, SPIE, 2018, pp. 320–327.

[12] A. Bergman, D. Lindell, and G. Wetzstein, *Deep adaptive lidar: End-to-end optimization of sampling and depth completion at low sampling rates*, ICCP (2020).

[13] Dana Berman, Deborah Levy, Shai Avidan, and Tali Treibitz, *Underwater single image color restoration using haze-lines and a new quantitative dataset*, IEEE transactions on pattern analysis and machine intelligence **43** (2020), no. 8, 2822–2837.

[14] Shariq Farooq Bhat, Reiner Birkl, Diana Wofk, Peter Wonka, and Matthias Müller, *Zoedepth: Zero-shot transfer by combining relative and metric depth*, 2023.

[15] Gøril M Breivik, Jens T Thielemann, Asbjørn Berge, Øystein Skotheim, and Trine Kirkhus, *A motion based real-time foveation control loop for rapid and relevant 3d laser scanning*, CVPR 2011 WORKSHOPS, IEEE, 2011, pp. 28–35.

[16] Samuel Burri, Claudio Bruschini, and Edoardo Charbon, *Linospad: a compact linear spad camera system with 64 fpga-based tdc modules for versatile 50 ps resolution time-resolved imaging*, Instruments **1** (2017), no. 1, 6.

[17] Clara Callenberg, Zheng Shi, Felix Heide, and Matthias B Hullin, *Low-cost spad sensing for non-line-of-sight tracking, material classification and depth imaging*, ACM Transactions on Graphics (TOG) **40** (2021), no. 4, 1–12.

[18] Ayan Chakrabarti, *Learning sensor multiplexing design through back-propagation*, Advances in Neural Information Processing Systems, 2016, pp. 3081–3089.

[19] Susan Chan, Abderrahim Halimi, Feng Zhu, Istvan Gyongy, Robert K Henderson, Richard Bowman, Stephen McLaughlin, Gerald S Buller, and Jonathan Leach, *Long-range depth imaging using a single-photon detector array and non-local data fusion*, Scientific reports **9** (2019), no. 1, 8075.

[20] Julie Chang, Vincent Sitzmann, Xiong Dun, Wolfgang Heidrich, and Gordon Wetzstein, *Hybrid optical-electronic convolutional neural networks with optimized diffractive optics for image classification*, Scientific reports **8** (2018), no. 1, 12324.

[21] Huaijin G Chen, Suren Jayasuriya, Jiyue Yang, Judy Stephen, Sriram Sivaramakrishnan, Ashok Veeraraghavan, and Alyosha Molnar, *Asp vision: Optically computing the first layer of convolutional neural networks using angle sensitive pixels*, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 903–912.

[22] Zhao Chen, Vijay Badrinarayanan, Gilad Drozdov, and Andrew Rabinovich, *Estimating depth from rgb and sparse sensing*, Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 167–182.

[23] Alexey Dosovitskiy, German Ros, Felipe Codevilla, Antonio Lopez, and Vladlen Koltun, *Carla: An open urban driving simulator*, Conference on robot learning, PMLR, 2017, pp. 1–16.

[24] Neale AW Dutton, Istvan Gyongy, Luca Parmesan, Salvatore Gnecchi, Neil Calder, Bruce R Rae, Sara Pellegrini, Lindsay A Grant, and Robert K Henderson, *A spad-based qvga image sensor for single-photon counting and quanta imaging*, IEEE Transactions on Electron Devices **63** (2015), no. 1, 189–196.

[25] Neale AW Dutton, Istvan Gyongy, Luca Parmesan, and Robert K Henderson, *Single photon counting performance and noise analysis of cmos spad-based image sensors*, Sensors **16** (2016), no. 7, 1122.

[26] Ahmet T Erdogan, Richard Walker, Neil Finlayson, Nikola Krstajić, Gareth Williams, John Girkin, and Robert Henderson, *A cmos spad line sensor with per-pixel histogramming tdc for time-resolved multispectral imaging*, IEEE Journal of Solid-State Circuits **54** (2019), no. 6, 1705–1719.

[27] Silvia Ferrari, *Track coverage in sensor networks*, 2006 American Control Conference, IEEE, 2006, pp. 7–pp.

[28] Thomas P Flatley, *Spacecube: A family of reconfigurable hybrid on-board science data processors*, (2015).

[29] Justin Folden, Derek Alley, David Illig, Linda Mullen, and Sanjeev J. Koppal, *Confocal Bistatic LIDAR in scattering media*, Ocean Sensing and Monitoring XVI (Weilin Hou and Linda J. Mullen, eds.), vol. 13061, International Society for Optics and Photonics, SPIE, 2024, p. 1306108.

[30] Genevieve Gariepy, Jonathan Leach, Ryan Warburton, Susan Chan, Robert Henderson, and Daniele Faccio, *Picosecond time-resolved imaging using spad cameras*, Emerging Imaging and Sensing Technologies, vol. 9992, SPIE, 2016, pp. 130–137.

[31] Andreas Geiger, Philip Lenz, and Raquel Urtasun, *Are we ready for autonomous driving? the kitti vision benchmark suite*, Conference on Computer Vision and Pattern Recognition (CVPR), 2012.

[32] Tobias Gruber, Frank Julca-Aguilar, Mario Bijelic, Werner Ritter, Klaus Dietmayer, and Felix Heide, *Gated2depth: Real-time dense lidar from gated images*, arXiv preprint arXiv:1902.04997 (2019).

[33] Brian Guenter, Mark Finch, Steven Drucker, Desney Tan, and John Snyder, *Foveated 3d graphics*, ACM transactions on Graphics (tOG) **31** (2012), no. 6, 1–10.

[34] Anant Gupta, Atul Ingle, and Mohit Gupta, *Asynchronous single-photon 3d imaging*, Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 7909–7918.

[35] Anant Gupta, Atul Ingle, Andreas Velten, and Mohit Gupta, *Photon-flooded single-photon 3d cameras*, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 6770–6779.

[36] Mohit Gupta, Shree K Nayar, Matthias B Hullin, and Jaime Martin, *Phasor imaging: A generalization of correlation-based time-of-flight imaging*, ACM Transactions on Graphics (ToG) **34** (2015), no. 5, 156.

[37] Felipe Gutierrez-Barragan, Huaijin Chen, Mohit Gupta, Andreas Velten, and Jinwei Gu, *itof2dtof: A robust and flexible representation for data-driven time-of-flight imaging*, IEEE Transactions on Computational Imaging **7** (2021), 1205–1214.

[38] Felipe Gutierrez-Barragan, Atul Ingle, Trevor Seets, Mohit Gupta, and Andreas Velten, *Compressive single-photon 3d cameras*, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 17854–17864.

[39] Felipe Gutierrez-Barragan, Fangzhou Mu, Andrei Ardelean, Atul Ingle, Claudio Bruschini, Edoardo Charbon, Yin Li, Mohit Gupta, and Andreas Velten, *Learned compressive representations for single-photon 3d imaging*, Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023, pp. 10756–10766.

[40] Istvan Gyongy, Sam W Hutchings, Abderrahim Halimi, Max Tyler, Susan Chan, Feng Zhu, Stephen McLaughlin, Robert K Henderson, and Jonathan Leach, *High-speed 3d sensing via hybrid-mode imaging and guided upsampling*, Optica **7** (2020), no. 10, 1253–1260.

[41] David S Hall, *High definition lidar system*, June 28 2011, US Patent 7,969,558.

[42] Ryan Halterman and Michael Bruch, *Velodyne hdl-64e lidar for unmanned surface vehicle obstacle detection*, Tech. report, SPACE AND NAVAL WARFARE SYSTEMS CENTER SAN DIEGO CA, 2010.

[43] Richard Hartley and Andrew Zisserman, *Multiple view geometry in computer vision*, Cambridge university press, 2003.

[44] Tim Hawkins, Per Einarsson, and Paul E Debevec, *A dual light stage.*, Rendering Techniques **5** (2005), 91–98.

[45] Felix Heide, Steven Diamond, David B Lindell, and Gordon Wetzstein, *Sub-picosecond photon-efficient 3d imaging using single-photon sensors*, Scientific reports **8** (2018), no. 1, 17726.

[46] Larry J Hornbeck, *Architecture and process for integrating dmd with control circuit substrates*, May 28 1991, US Patent 5,018,256.

[47] Fu-Chung Huang, David P Luebke, and Gordon Wetzstein, *The light field stereoscope.*, SIGGRAPH emerging technologies, 2015, pp. 24–1.

[48] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger, *Densely connected convolutional networks*, Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 4700–4708.

[49] Tak-Wai Hui, Chen Change Loy, and Xiaoou Tang, *Depth map super-resolution by deep multi-scale guidance*, Proceedings of European Conference on Computer Vision (ECCV), 2016.

[50] Sam W Hutchings, Nick Johnston, Istvan Gyongy, Tarek Al Abbas, Neale AW Dutton, Max Tyler, Susan Chan, Jonathan Leach, and Robert K Henderson, *A reconfigurable 3-d-stacked spad imager with in-pixel histogramming for flash lidar or high-speed time-of-flight imaging*, IEEE Journal of Solid-State Circuits **54** (2019), no. 11, 2947–2956.

[51] Atul Ingle and David Maier, *Count-free single-photon 3d imaging with race logic*, IEEE Transactions on Pattern Analysis and Machine Intelligence (2023).

[52] A. Jones, I. McDowall, H. Yamada, M. Bolas, and P. Debevec, *Rendering for an interactive 360 degree light field display*, SIGGRAPH, ACM, 2007.

[53] Andrew Jones, Ian McDowall, Hideshi Yamada, Mark Bolas, and Paul Debevec, *Rendering for an interactive 360 light field display*, ACM Transactions on Graphics (TOG) **26** (2007), no. 3, 40.

[54] Abhishek Kasturi, Veljko Milanovic, Bryan H Atwood, and James Yang, *Uav-borne lidar with mems mirror-based scanning capability*, Proc. SPIE, vol. 9832, 2016, p. 98320M.

[55] Diederik P Kingma and Jimmy Ba, *Adam: A method for stochastic optimization*, arXiv preprint arXiv:1412.6980 (2014).

[56] Krassimir T Krastev, Hendrikus WLAM Van Lierop, Herman MJ Soemers, Renatus Hendricus Maria Sanders, and Antonius Johannes Maria Nellissen, *Mems scanning micromirror*, September 3 2013, US Patent 8,526,089.

[57] Jongho Lee, Atul Ingle, Jenu V Chacko, Kevin W Eliceiri, and Mohit Gupta, *Caspi: collaborative photon processing for active single-photon imaging*, Nature Communications **14** (2023), no. 1, 3158.

[58] David B Lindell, Matthew O'Toole, and Gordon Wetzstein, *Single-photon 3d imaging with deep sensor fusion*, ACM Transactions on Graphics (ToG) **37** (2018), no. 4, 1–12.

[59] Chao Liu, Jinwei Gu, Kihwan Kim, Srinivasa Narasimhan, and Jan Kautz, *Neural rgb-to-d sensing: Depth and uncertainty from a video camera*, arXiv preprint arXiv:1901.02571 (2019).

[60] Jiajun Lu and David Forsyth, *Sparse depth super resolution*, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 2245–2253.

[61] Bruce D Lucas, Takeo Kanade, et al., *An iterative image registration technique with an application to stereo vision*, (1981).

[62] Fangchang Mal and Sertac Karaman, *Sparse-to-dense: Depth prediction from sparse depth samples and a single image*, 2018 IEEE International Conference on Robotics and Automation (ICRA), IEEE, 2018, pp. 1–8.

[63] Xiaoxu Meng, Ruofei Du, Joseph F JaJa, and Amitabh Varshney, *3d-kernel foveated rendering for light fields*, IEEE transactions on visualization and computer graphics **27** (2020), no. 8, 3350–3360.

[64] V Milanović, A Kasturi, N Siu, M Radojičić, and Y Su, *"memseye" for optical 3d tracking and imaging applications*, Solid-State Sensors, Actuators and Microsystems Conference (TRANSDUCERS), 2011 16th International, IEEE, 2011, pp. 1895–1898.

[65] Veljko Milanović, Abhishek Kasturi, James Yang, and Frank Hu, *A fast single-pixel laser imager for vr/ar headset tracking*, Proc. of SPIE Vol, vol. 10116, 2017, pp. 101160E–1.

[66] Jason Mudge, *Range-compensating lens for non-imaging active optical systems*, Applied optics **58** (2019), no. 28, 7921–7928.

[67] Srinivasa G Narasimhan, Shree K Nayar, Bo Sun, and Sanjeev J Koppal, *Structured light in scattering media*, Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1, vol. 1, IEEE, 2005, pp. 420–427.

[68] Pushmeet Kohli Nathan Silberman, Derek Hoiem and Rob Fergus, *Indoor segmentation and support inference from rgbd images*, ECCV, 2012.

[69] _____, *Indoor segmentation and support inference from rgbd images*, ECCV, 2012.

[70] Shree K Nayar, Vlad Branzoi, and Terrance E Boult, *Programmable imaging using a digital micromirror array*, Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004., vol. 1, IEEE, 2004, pp. I–I.

[71] Richard A Newcombe, Shahram Izadi, Otmar Hilliges, David Molyneaux, David Kim, Andrew J Davison, Pushmeet Kohi, Jamie Shotton, Steve Hodges, and Andrew Fitzgibbon, *Kinectfusion: Real-time dense surface mapping and tracking*, 2011 10th IEEE International Symposium on Mixed and Augmented Reality, IEEE, 2011, pp. 127–136.

[72] Matthew O ´ Toole, David B Lindell, and Gordon Wetzstein, *Confocal non-line-of-sight imaging based on the light-cone transform*, Nature **555** (2018), no. 7696, 338.

[73] Matthew O'Toole, Felix Heide, Lei Xiao, Matthias B Hullin, Wolfgang Heidrich, and Kiriakos N Kutulakos, *Temporal frequency probing for 5d transient analysis of global light transport*, ACM Transactions on Graphics (ToG) **33** (2014), no. 4, 87.

[74] Anjul Patney, Joohwan Kim, Marco Salvi, Anton Kaplanyan, Chris Wyman, Nir Benty, Aaron Lefohn, and David Luebke, *Perceptually-based foveated virtual reality*, ACM SIGGRAPH 2016 emerging technologies, 2016, pp. 1–2.

[75] Francesco Pittaluga, Zaid Tasneem, Justin Folden, Brevin Tilmon, Ayan Chakrabarti, and Sanjeev J Koppal, *Towards a mems-based adaptive lidar*, 2020 International Conference on 3D Vision (3DV), IEEE, 2020, pp. 1216–1226.

[76] Ryan Po, Adithya Pediredla, and Ioannis Gkioulekas, *Adaptive gating for single-photon 3d imaging*, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 16354–16363.

[77] Ramesh Raskar, Greg Welch, Matt Cutts, Adam Lake, Lev Stesin, and Henry Fuchs, *The office of the future: A unified approach to image-based modeling and spatially immersive displays*, Proceedings of the 25th annual conference on Computer graphics and interactive techniques, ACM, 1998, pp. 179–188.

[78] Ximing Ren, Peter WR Connolly, Abderrahim Halimi, Yoann Altmann, Stephen McLaughlin, Istvan Gyongy, Robert K Henderson, and Gerald S Buller, *High-resolution depth profiling using a range-gated cmos spad quanta image sensor*, Optics express **26** (2018), no. 5, 5541–5557.

[79] Gernot Riegler, Matthias Rüther, and Horst Bischof, *Atgv-net: Accurate depth super-resolution*, European Conference on Computer Vision, Springer, 2016, pp. 268–284.

[80] Thilo Sandner, Claudia Baulig, Thomas Grasshoff, Michael Wildenhain, Markus Schwarzenberg, Hans-Georg Dahlmann, and Stefan Schwarzer, *Hybrid assembled micro scanner array with large aperture and their system integration for a 3d tof laser camera*, MOEMS and Miniaturized Systems XIV, vol. 9375, International Society for Optics and Photonics, 2015, p. 937505.

[81] Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra, *Grad-cam: Visual explanations from deep networks via gradient-based localization*, Proceedings of the IEEE international conference on computer vision, 2017, pp. 618–626.

[82] Michael Sheehan, Julian Tachella, and Mike Davies, *A sketching framework for reduced data transfer in photon counting lidar*, IEEE Transactions on Computational Imaging **7** (2021), 989–1004.

[83] Vincent Sitzmann, Steven Diamond, Yifan Peng, Xiong Dun, Stephen Boyd, Wolfgang Heidrich, Felix Heide, and Gordon Wetzstein, *End-to-end optimization of optics and image processing for achromatic extended depth of field and super-resolution imaging*, ACM Transactions on Graphics (TOG) **37** (2018), no. 4, 114.

[84] Barry L Stann, Jeff F Dammann, Mark Del Giorno, Charles DiBerardino, Mark M Giza, Michael A Powers, and Nenad Uzunovic, *Integration and demonstration of mems–scanned ladar for robotic navigation*, Proc. SPIE, vol. 9084, 2014, p. 90840J.

[85] Qi Sun, Fu-Chung Huang, Joohwan Kim, Li-Yi Wei, David Luebke, and Arie Kaufman, *Perceptually-guided foveation for light field displays*, ACM Transactions on Graphics (TOG) **36** (2017), no. 6, 1–13.

[86] Qilin Sun, Jian Zhang, Xiong Dun, Bernard Ghanem, Yifan Peng, and Wolfgang Heidrich, *End-to-end learned, optically coded super-resolution spad camera*, ACM Trans. Graph. **39** (2020), no. 2.

[87] Varun Sundar, Andrei Ardelean, Tristan Swedish, Claudio Bruschini, Edoardo Charbon, and Mohit Gupta, *Sodacam: Software-defined cameras via single-photon imaging*, Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023, pp. 8165–8176.

[88] Filip Taneski, Istvan Gyongy, Tarek Al Abbas, and Robert K Henderson, *Guided direct time-of-flight lidar using stereo cameras for enhanced laser power efficiency*, Sensors **23** (2023), no. 21, 8943.

[89] Zaid Tasneem, Dingkang Wang, Huikai Xie, and Sanjeev J. Koppal, *Directionally controlled time-of-flight ranging for mobile sensing platforms*, Robotics: Science and Systems, 2018.

[90] Brevin Tilmon, Eakta Jain, Silvia Ferrari, and Sanjeev Koppal, *Foveacam: A mems mirror-enabled foveating camera*, 2020 IEEE International Conference on Computational Photography (ICCP), IEEE, 2020, pp. 1–11.

[91] Brevin Tilmon and Sanjeev J Koppal, *Saccadecam: Adaptive visual attention for monocular depth sensing*, Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 6009–6018.

[92] Alessandro Tontini, Sonia Mazzucchi, Roberto Passerone, Nicolò Broseghini, and Leonardo Gasparini, *Histogram-less lidar through spad response linearization*, IEEE Sensors Journal **PP** (2023), 1–1.

[93] Okan Tarhan Tursun, Elena Arabadzhiyska-Koleva, Marek Wernikowski, Rados law Mantiuk, Hans-Peter Seidel, Karol Myszkowski, and Piotr Didyk, *Luminance-contrast-aware foveated rendering*, ACM Transactions on Graphics (TOG) **38** (2019), no. 4, 1–14.

[94] Robert K Tyson, *Principles of adaptive optics*, CRC press, 2015.

[95] Jonas Uhrig, Nick Schneider, Lukas Schneider, Uwe Franke, Thomas Brox, and Andreas Geiger, *Sparsity invariant cnns*, 2017 International Conference on 3D Vision (3DV), IEEE, 2017, pp. 11–20.

[96] Wouter Van Gansbeke, Davy Neven, Bert De Brabandere, and Luc Van Gool, *Sparse and noisy lidar completion with rgb guidance and uncertainty*, arXiv preprint arXiv:1902.05356 (2019).

[97] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin, *Attention is all you need*, Advances in neural information processing systems **30** (2017).

[98] Jian Wang, Joseph Bartels, William Whittaker, Aswin C Sankaranarayanan, and Srinivasa G Narasimhan, *Programmable triangulation light curtains*, Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 19–34.

[99] Mel White, Shahaboddin Ghajari, Tianyi Zhang, Akshat Dave, Ashok Veeraraghavan, and Alyosha Molnar, *A differential spad array architecture in 0.18 um cmos for hdr imaging*, 2022 IEEE International Symposium on Circuits and Systems (ISCAS), 2022, pp. 292–296.

[100] Yalin Xiong and Steven Shafer, *Depth from focusing and defocusing*, Proceedings of (CVPR) Computer Vision and Pattern Recognition, June 1993, pp. 68 – 73.

[101] Taiki Yamamoto, Yasutomo Kawanishi, Ichiro Ide, Hiroshi Murase, Fumito Shinmura, and Daisuke Deguchi, *Efficient pedestrian scanning by active scan lidar*, Advanced Image Technology (IWAIT), 2018 International Workshop on, IEEE, 2018, pp. 1–4.

[102] Chao Zhang, Scott Lindner, Ivan Michel Antolović, Juan Mata Pavia, Martin Wolf, and Edoardo Charbon, *A 30-frames/s,* 252×144 *SPAD Flash LiDAR With 1728 Dual-Clock 48.8-ps TDCs, and Pixel-Wise Integrated Histogramming*, IEEE Journal of Solid-State Circuits **54** (2018), no. 4, 1137–1151.

[103] Chao Zhang, Ning Zhang, Zhijie Ma, Letian Wang, Yu Qin, Jieyang Jia, and Kai Zang, *A 240× 160 3d-stacked spad dtof image sensor with rolling shutter and in-pixel histogram for mobile devices*, IEEE Open Journal of the Solid-State Circuits Society **2** (2021), 3–11.

[104] Tianyi Zhang, Mel J. White, Akshat Dave, Shahaboddin Ghajari, Ankit Raghuram, Alyosha C. Molnar, and Ashok Veeraraghavan, *First arrival differential lidar*, 2022 IEEE International Conference on Computational Photography (ICCP), 2022, pp. 1–12.

[105] Yinda Zhang and Thomas Funkhouser, *Deep depth completion of a single rgb-d image*, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 175–185.

BIOGRAPHICAL SKETCH

Justin Folden recieved his Bachelor of Science in Electrical Engineering from Florida Polytechnic University, where he was part of the inaugural class and a founding member of the IEEE branch. In 2018, he began his PhD at the University of Florida, earning his master's degree in Spring 2024. His research focuses on foveated depth sensing technologies, including adaptive LIDAR and SPAD-based systems. Justin gained professional experience through internships at imec USA and NAWCAD with the U.S. Navy as part of the NREIP program.