1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Computational Imaging for Improving Vehicle Safety		**5. Report Date** July 31, 2024 **6. Performing Organization Code**
7. Author(s) Aswin Sankaranarayanan, https://orcid.org/ 0000-0003-0906-4046 Vijayakumar Bhagavatula, https://orcid.org/ 0000-0001-7126-6381		**8. Performing Organization Report No.**
9. Performing Organization Name and Address Carnegie Mellon University, 5000Forbes Avenue, Pittsburgh, PA 15213, USA.		**10. Work Unit No.** **11. Contract or Grant No.** Federal Grant No. 69A35523244811
12. Sponsoring Agency Name and Address Safety21 University Transportation Center Carnegie Mellon University 5000 Forbes Avenue Pittsburgh, PA 15213		**13. Type of Report and Period Covered** Final Report (July 1, 2023-June 30, 2024) **14. Sponsoring Agency Code** USDOT
15. Supplementary Notes		

16. Abstract

This project investigates the design and deployment of sensors and associated algorithms for handling harsh imaging conditions. We are interested in depth perception in rain, snow and fog. By expanding the depth range at which objects can be reliably detected, especially in dense fog, the project will facilitate a higher level of safety for vulnerable road users.

17. Key Words Depth perception; bad weather	18. Distribution Statement		
19. Security Classif. (of this report)	**20. Security Classif. (of this page)**	**21. No. of Pages** 16	**22. Price**

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

Safety21

INNOVATING SAFETY FOR ALL

US DOT National
University Transportation Center for Safety

Carnegie Mellon University

Penn · The Ohio State University · CCAC · Community College of Philadelphia · The University of Texas Rio Grande Valley · Morgan State University

Computational Imaging for Improving Vehicle Safety

PI: Aswin Sankaranarayanan
OrcID: 0000-0003-0906-4046

Co-PI: Vijayakumar Bhagavatula
OrcID: 0000-0001-7126-6381

Contract: #69A35523244811

Final Project Report - July 31, 2024

Computational Imaging for Improved Vehicle Safety

Contents

List of Figures

1 Introduction

The last decade has seen increased adoption of technologies that enable assisted and autonomous driving systems.. This spans simple aids in driving like parking assists, automatic speed limit detection, and lane monitoring to more complicated features like collision avoidance and lane centering, with the eventual promise of fully automated driving. A key enabler for many of these features is the ability to maintain awareness of the scene around the vehicle using a suite of visual and non-visual sensors, including cameras, and LIDARs. In many ways, much of today's assisted systems rely critically on being able robust sense an image and depth map of the world in realtime, and processing it to enable the desired feature.

Adverse weather conditions like fog and rain present immediate challenges to the operation of assisted and autonomous driving systems. Rain, snow and fog present challenging operating scenarios for camera and LIDAR-based depth perception. In these scenarios, the medium between the camera/sensor and the target (be it a pedestrian, a vehicle or a road feature) can no longer be assumed to be "free space". In particular, the presence of fog or rain or snow results in the medium scattering light, which has undesirable effects on measurements made by a sensing systems.

For passive cameras, typically used in stereo-based depth estimation, imagery in adverse weather results in a loss of contrast. This loss of contrast arises from two sources: first, the presence of particulate matter in the medium (be it fog, rain or snow) scatters light from the sun or from the headlights, which adds a haze on the measurements; and second, the light from targets of interest are progressively scattering en route to the cameras, which reduces their intensity. Together, this results in a haze in the measurements that makes details harder to identify, with objects that are further away affected significantly more (see Figure 1). Since stereo-based depth estimation relies on matching textured regions across views, this loss of contrast makes correspondence across views hard to establish. In turn, this reduces the effectiveness of stereo-based depth estimation.

LIDAR, on the other hand, works on the principle of time of flight, measured by pulsing a laser and using a single-photon avalanche diode (SPAD) to measure the arrival time of the first returning photon. When the medium between the laser and the target is free space, we can expect the first-returning photon to be from the target. However, fog and rain generate spurious photon arrivals so that the first-returning photon is from the

Figure 1: Fog leads to loss of contrast in image sensor measurements. Shown above are simulated images of a scene in clear and foggy conditions. Such loss of contrast affects subsequent downstream processing algorithms like correspondence estimation.

medium and not the target. This severely compromise the quality of depth measurements. There are some LIDAR-based depth systems that look at the distribution of time of arrivals at each pixel to identify the photons from the target; such systems are capable of handling moderate amount of rain or fog. However, the electronics bandwidth and storage requirements of such techniques are formidable when we want a large field of view and high frame rates.

Non-visual systems such as radar are inherently robust to such weather events; however, their angular resolution is often orders of magnitude smaller than that of cameras and LIDARs—a fundamental feature of this modality stemming from its large wavelength. As a consequence, there is an immediate need for enabling sensing systems that are inherently robust to adverse weather with the depth and angular resolutions typical to visual sensors.

Approach. The project aims to improve depth sensing in bad weather—fog, smoke and rain—where the performance of LIDAR and stereo based scanners suffers. Our solution to this was based on two ideas: first, eliminating light paths that are caused by scattering of light in the medium (i.e., fog, rain, smoke) improves contrast; and second, using single-photon detectors (SPADs) we can build a 3D structured light scanner that provides depth maps at hundreds of frames per second. Specifically, our main insight is that improved depth perception can be improved via careful imaging and algorithmic design that allows blocking of photons from the medium while preserving those from the scene of interest.

Improving perception in dense scattering media by blocking undesired photons requires imaging systems that can selectively choose between favorable light paths in the scene against unfavorable ones. Our specific proposition for this work is to build a structured light system with a high-speed projector/light scanner and an ultra-high speed SPAD array. With this setup, we will design patterns that will avoid single-bounce light paths off the medium. This approach is central to very successful microscopy techniques such as confocal imaging, and diffuse optical tomography, for imaging in highly scattering media like biological tissue. We will leverage this core intuition but expand it to macroscopic imaging in the real world.

The proposed system is a variant of a traditional structured light system consisting of a projector and a camera. For the projector, we use a laser-based line scanner that scans a line across the scene at a very high speed. For the camera, we replace the commonly-used CMOS image sensor with a single-photon avalanche diode (SPAD) sensor. This SPAD sensor is also an array of pixels, except in each image, each of its pixels records a binary value indicating whether or not a photon arrived during the exposure time. While this seems rather limiting, the sensor operates at extremely high speeds—upwards of 100 kHz. This allows us to scan the scene at a high speed and computationally identify photons from the target from those from the medium.

At its culmination, this project will enhance the range of scenarios where an autonomous vehicle can safely operate in. Specifically, it will lead to increased range in depth perception in fog and rain, which will increase safety in the use of assisted and autonomous driving.

2 Structured light under fog

We briefly describe structured light-based scanning and the influence of fog on its measurements. A detailed overview of structured light techniques can be found in prior art.

2.1 Line scanning in the absence of fog

Structured light (SL) [4] is one of the most popular techniques for 3D shape acquisition. An SL system uses active illumination, typically via a projector, to obtain robust correspondences between pixels on the projector and a camera, and subsequently, recovers

the scene depth via triangulation. In contrast to passive techniques like stereo, the use of active illumination enables SL systems to acquire depth even for textureless scenes at a low computational cost.

The simplest SL method is point scanning [3], where the light source illuminates a single scene point at a time, and the camera captures an image. Correspondence between camera and projector pixels is determined by associating the brightest pixel in each acquired image to the pixel illuminated by the projector. However, this approach requires a large number (N^2) of images to obtain a depth map with $N \times N$ pixels.

In order to reduce the acquisition time, stripe or line scanning technique was proposed where the light source emits a planar sheet of light [11, 1, 2]. In line scanning, the projector illuminates a single column at a time and sequentially scans its field of view through all of its columns. Since the pre-image of a line on the projector's image plane is a plane, this results in a plane in the real world being illuminated at a given time instant. This plane intersects with the 3D world in front of the imaging system, illuminating what is typically a curve or a collection of disjoint curves; the specifics of this depends on the geometry of the scene. An image sensor placed beside the projector observes the illuminated points, importantly from a different perspective from that of the projective; this allows for triangulation between the pre-image of each of the points observed in the image—a line in 3D space—with the plane illuminated by the projector. That is, consider a scene point that lies on the emitted light plane. Its depth can be estimated by finding the intersection between the light plane, and the ray joining the camera center and the camera pixel. This is illustrated in Figure 2(a).

We can further reduce the acquisition time by using more sophisticated temporal coding techniques; for example, binary codes [7], Gray codes [10, 5] and sinusoidal phase shifting [12]. Underlying all these methods is the idea that, for a calibrated camera-projector pair, we only need to measure disparity, i.e., a 1D displacement map. Thus, we need to perform coding along only one dimension of the projector image plane, thereby achieving significant speed-ups over point-scanning systems. For example, several structured light patterns have a 1D translational symmetry, i.e., in the projected patterns, all the pixels within a column (or a row) have the same intensities. An exception is 'single-shot' structured light techniques that use patterns with 2D intensity variations, for example, sparse 2D grid of lines [8], 2D color encoded grids [9], 2D pseudo-random

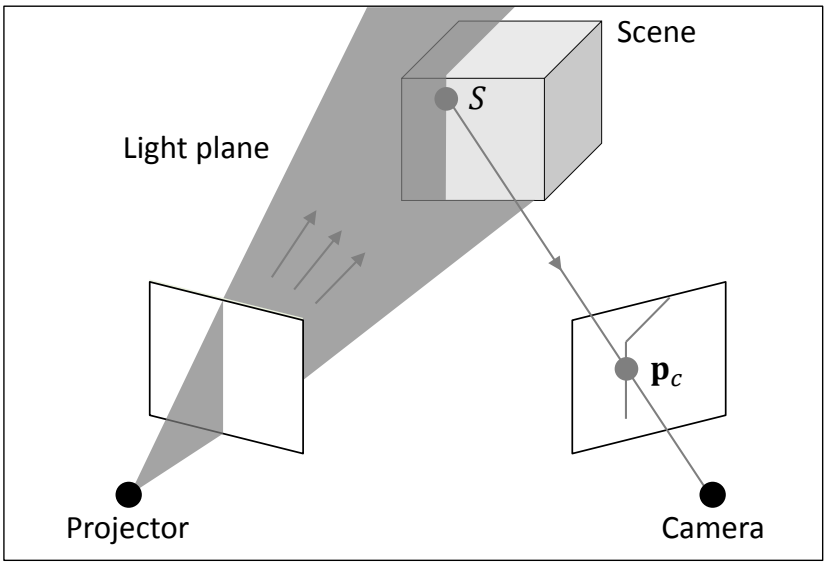


Figure 2: In a line-scanning SL system, the ray is from camera pixel, and plane is formed by center of projection and a column of the projector.

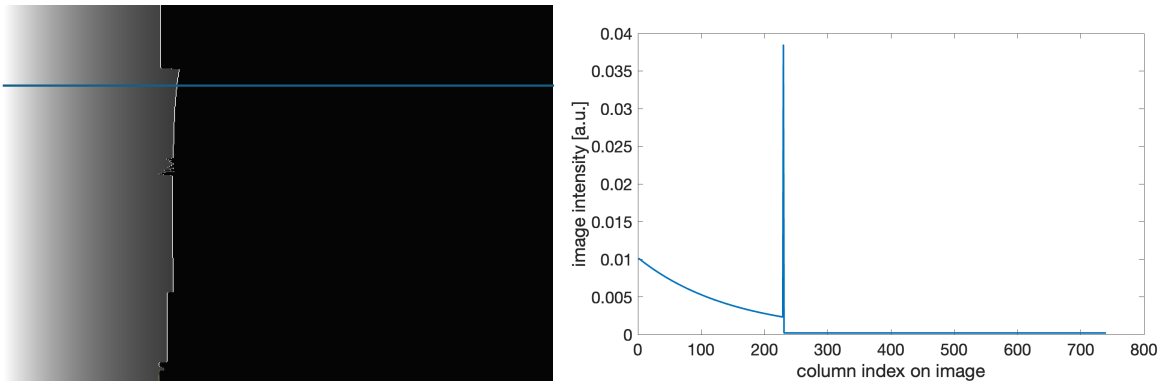
binary code [15], and 2D random dots (used in the first generation Microsoft Kinect depth sensing cameras [6]).

Note that, since the medium between the imaging system and the scene is considered to be free space, i.e., allows light to freely propagate without any additional effects, the camera only sees points that are illuminated on (potential) objects of interest. A caveat here is that we are not considering inter-reflections or sub-surface scattering effects induced by the object itself.

2.2 Measurements under fog

In the presence of fog, the medium scatters light and so the image formed on the camera is no longer that of the single stripe on the scene. Instead we see scattered light from the fog as well. Figure 3(a) shows a simulation of such an image for a scene in the presence of fog. Here, we observe that in addition to the illuminated stripe there is a strong haze to its left arising from the fog itself due to the projector being placed to the left of the camera.

Digging deeper, if we look along a horizontal slice of the image—or an epipolar line—we observe the profile seen in Figure 3(b), which shows that an exponentially decaying curve followed by an impulse. The impulse marks the stripe or the intersection of the



(a) Image observed on a CMOS sensor

(b) Intensity along an epipolar line

Figure 3: Line scanning in the presence of fog results in a strong scattering in the captured image. Shown on the left is a simulated image captured by the sensor in the presence of scattering where we observe the line/stripe off the target of interest but also light scattered off the medium. The projector is to the left of the camera. (right) Image intensity observed along an epipolar line.

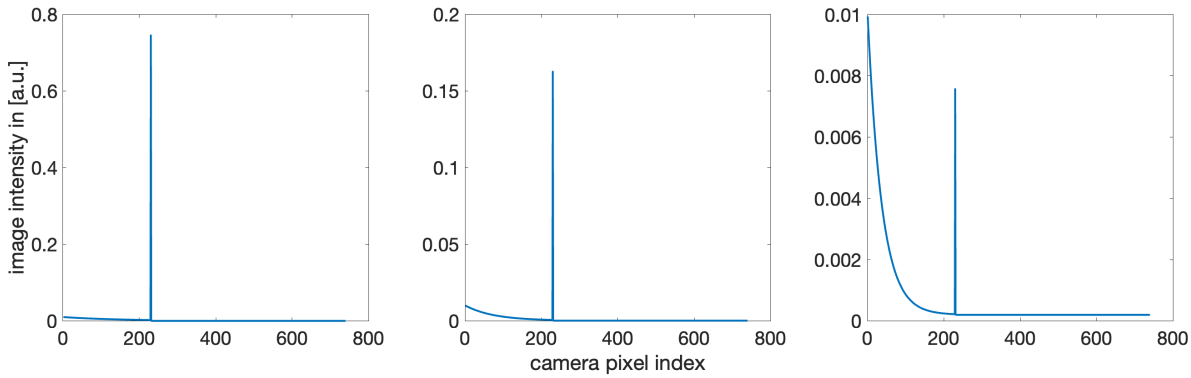


Figure 4: Profile of line scan measurements along an epipolar line with increasing amount of scattering from left to right. This results in the impulse associated with the target to reduce in intensity.

object of interest and the projector illuminated plane. The exponential decay is the contribution of scattering from the fog and its specific shape is explained from Beer-Lambert's law for isotropic scatterers.

The nature of the scattering also affects the profiles we can expect to see. Thicker mediums scatter more light, causing increased haze while reducing the intensity on the target. Figure 4 shows the effect of this with increasing scattering parameters.

One of the challenges here is that, in the presence of read noise, we will need sufficiently

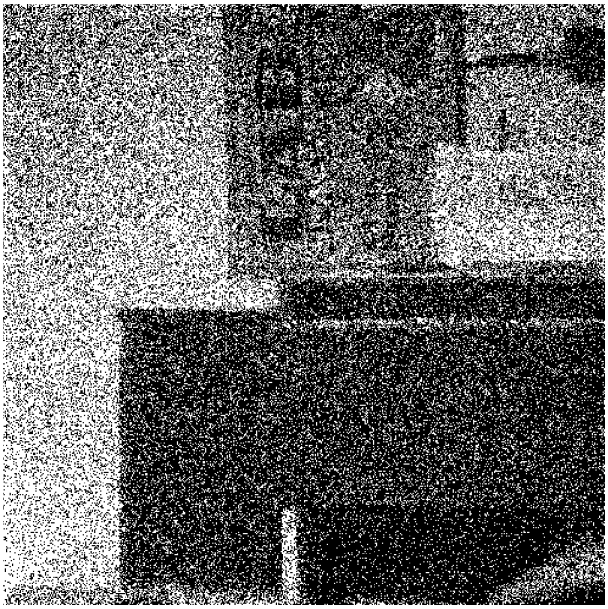


Figure 5: An image from a SPAD sensor. Each pixel reports a binary value indicating whether or not photons arrived at the pixel during the exposure time.

large exposure to estimate the location of the impulse, which essentially slows down line scanning techniques. We next show that moving to a SPAD image sensor—a device that counts photons without any read noise penalty—can effectively address this.

3 Structured light scanning with Single-Photon Detectors

Our work relies on a *Single-Photon Structured Light* system that are based on single-photon detectors, such as Single Photon Avalanche Diodes (SPADs). SPADs can be operated at very high speeds when detecting photons and not their time-of-arrivals. In this 'photon detection' mode, the measurements are binary-valued indicating whether or not a photon arrival occurred during a given acquisition time. For instance, a recently developed SPAD array [14] can capture $\sim 10^5$ binary frames at 1/8-th megapixel resolution. Our key observation is that the binary measurements, normally considered a limitation due to limited information, are sufficient for a large family of SL techniques where the coding is binary as well.

3.1 Model for SPAD measurements

Consider a SPAD pixel array observing a scene. The number of photons N arriving at a pixel $\mathbf{x}$ during an exposure time t_{exp} is modelled as a Poisson random variable:

$$\Pr\{N = k\} = \frac{(\Phi(\mathbf{x})\, t_{\text{exp}})^k \; e^{-\Phi(\mathbf{x})\, t_{\text{exp}}}}{k!}, \tag{1}$$

where $\Phi(\mathbf{x})$ is the flux; for simplicity, we assume a 100% quantum efficiency and use the term "flux" interchangeably with the arrival rate of photo-electrons. During each exposure, a pixel detects at most one photon, returning a binary value $B(\mathbf{x})$ such that $B(\mathbf{x}) = 0$ if the pixel receives no photons; otherwise, $B(\mathbf{x}) = 1$. Hence, $B(\mathbf{x})$ is a Bernoulli random variable [16] with

$$\Pr\{B(\mathbf{x}) = 0\} = e^{-(\Phi(\mathbf{x}) + r_q) t_{\text{exp}}}, \tag{2}$$

where r_q is the dark current rate—the rate of spurious counts unrelated to incident photons. Figure 5 shows an example of a SPAD binary measurement.

Recently, we developed a technique for SPAD-based SL [13] to enable 3D scanning at high-frame rates and low-light levels. This technique, called "Single-Photon SL", works by sensing binary images that indicates the presence or absence of photon arrivals during each exposure; the SPAD array is used in conjunction with a high-speed binary projector, with both devices operated at speeds as high as 20 kHz. In a typical SL scan, the scene is illuminated with a sequence of 2D binary patterns from a projector. The SPAD captures a binary frame for each pattern. Each SPAD pixel receives a binary code over time, from which we estimate the projector column observed at the pixel—an operation that is critical for the success of any SL technique. However, the binary images sensed by the SPAD array are heavily influenced by photon noise and are easily corrupted by ambient sources of light (as seen in Figure 5). To address this, we developed novel temporal sequences using error correction codes, and further, design the codes to be robust to short-range effects like projector/camera defocus and resolution mismatch between the two devices. Single-Photon SL is capable of 3D imaging in challenging scenarios involving low-albedo objects, strong ambient illumination as well as fast-moving objects. However, it is not geared to handle scattering by the medium as is the case with imaging in fog or rain.

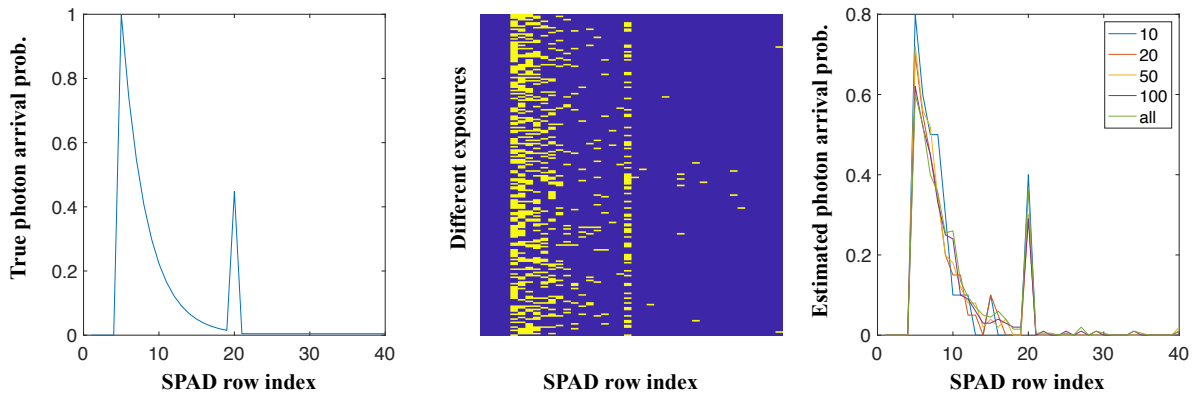
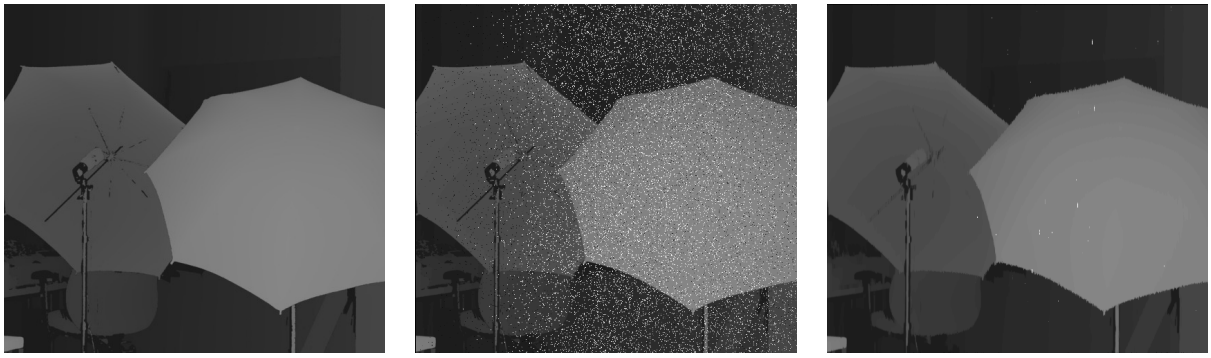


Figure 6: Profile of line scan measurements along an epipolar line with SPAD measurements. (left) True intensities (or photon arrival rates) on an epipolar line on the SPAD. (center) SPAD measurements corresponding to this epipolar line across multiple exposures. Note that the location of the peak associated with the stripe is clearly visible. (c) Estimates of the photon arrivals with different number of SPAD exposures. With as little as ten exposures we can robustly estimate the peak associated with the stripe.

3.2 Measurements under fog

To build a robust SL scanner under fog, we revert to stripe/line scanning, where the projector illuminates a single column. As elaborated in Figures 3 and 4 this results in each epipolar line observing the stripe as well as haze due to scattering. Given the image formation model underlying SPADs, each measurement of the SPAD sensor would simply provide a binary image indicating photon arrival events; critically, the probability of detecting a photon is still proportional to the relative intensities. If we take multiple measurements, then due to photon noise, we get different realizations of the underlying Poisson random variable. Aggregating measurements and averaging them provides an estimate of the true photon arrival rates (or the image sensed by a traditional image sensor).

Figure 6 shows this principle in action. We observe that while an individual binary frame looks random, aggregating multiple measurements by averaging them provides a robust estimate of the peak. The reason for this stems from the sharp transition at the peak that is easily observable from the averaged measurements. In fact, our experiments show that with as little as ten exposures, our approach obtains the peak accurately.



(a) Ground truth disparity (b) Raw estimated disparity (c) Median filtered disparity

Figure 7: Estimated disparities from simulations of the proposed SPAD-based SL setup. (a) True disparity for the scene. (b) Raw estimated disparities from ten exposures per projector column. (c) Median filtered disparity output.

3.3 Performance of SPAD-based structured light

We can now characterize the time resolutions we can obtain with commercially-available SPAD sensors and projectors. Laser projectors can easily scan scenes in tens to hundreds of hertz, and as such as not bottlenecks for our envisioned system. The SPAD prototype we plan to use operates at 100 kHz. If we average 10 measurements for each projector column, and have 500 projector columns to scan (matching the resolution of the 512 $\times$ 512 pixel array of the SPAD), then our system is still capable of delivering dense depth maps at 20 fps.

4 Results

We simulated a structured light-based SPAD imaging system similar to [13] for our proposed stripe/line-scanning system. We simulated these on a scene from the Middlebury dataset.

Figure 7 shows reconstructed depth maps with ten SPAD exposures per projector column. The raw estimated disparity values are corrupted due to the inherent stochasticity of photon noise. However a simple median filter applied to the estimated disparity significantly suppresses these artifacts.

Figure 8 characterizes disparity estimation errors as a function of number of SPAD exposures. As we expect, a single SPAD measurement does provide a robust estimate of the stripe given that it is impossible to classify a photon from the fog versus the target.

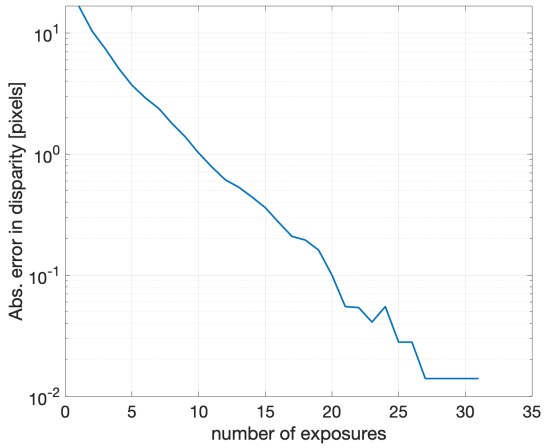


Figure 8: Disparity error as a function of number of SPAD exposures.

As we gradually increase the number of measurements, the estimation error in log-scale decreases linearly with the number of measurements, indicating exponential reduction in the accuracy of the technique. With ten measurements, the absolute disparity error drops below a single pixel and to a tenth of a pixel with twenty measurements. This suggests that high-speed robust line striping is indeed possible in the presence of fog.

We validate this further with by repeating the results of Figure 7 with different number of measurements. Our results validate the increased robustness of disparity estimation from a small set of measurements.

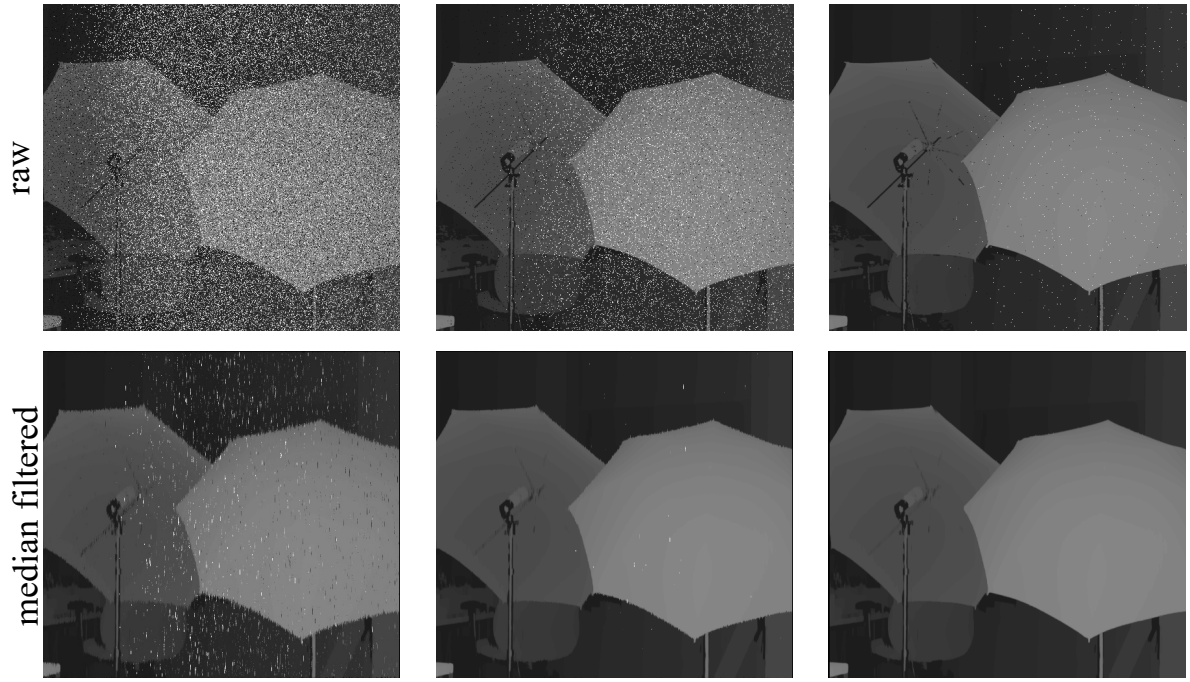


Figure 9: Estimated disparities with (left) five, (center) ten, and (right) twenty SPAD measurements with top row showing raw estimates and bottom row showing filtered ones.

5 Conclusions

In this work, we have devised a novel 3D scanner for robust depth map acquisition in the presence of rain, fog or snow. Specifically, we have implemented a simulation platform for the measurements made by a SPAD sensor under fog. We simulate the gross behavior of the scattering using the Beer-Lambert law, which provides the photon arrival rate at the sensor. Specifically, we assume that a projector is illuminating one or a few planes in the world. For each of these planes, we first estimate its intersection with the scene (described in terms of a depth map), and simulate the image formed on the sensor, including the effect of scattering in the medium (fog). The measurements of the SPAD sensor reports binary frames indicating the presence or absence of photon arrival. We have implemented an algorithm for estimating depth from a collection of these binary SPAD measurements, relying on the fact that the nature of scattered light is exponentially decaying and it ends with a discontinuity (in the form of a peak) when it intersects with the object. We use this to robustly estimate the peak along each epipolar scan line. With these simulations, we have the following result. Suppose we illuminate a single plane in the scene; with as little as ten binary SPAD images we can reliably estimate the depth of the scene points

illuminated. This suggests that high-speed 3D scanning is indeed possible with out setup. In terms of future directions, a key next step is to demonstrate performance of the system using a real-world prototype, which we hope to perform in the upcoming months.

References

[1] G. J. Agin and T. O. Binford. Computer description of curved objects. *IEEE Transactions on Computers*, 100(4):439–449, 1976. 3

[2] B. Curless and M. Levoy. Better optical triangulation through spacetime analysis. In *Proc. of International Conference on Computer Vision (ICCV)*, pages 987–994, 1995. 3

[3] G. E. Forsen. Processing visual data with an automaton eye. In *Pictoral Pattern Recognition*, 1968. 3

[4] J. Geng. Structured-light 3d surface imaging: a tutorial. *Advances in Optics and Photonics*, 3(2):128–160, 2011. 3

[5] S. Inokuchi, K. Sato, and F. Matsuda. Range imaging system for 3-d object recognition. In *Proc. of International Conference on Pattern Recognition (ICPR)*, volume 48, pages 806–808, 1984. 4

[6] Microsoft. Kinect for xbox 360. https://en.wikipedia.org/wiki/Kinect#Kinect_for_Xbox_360, 2010. 4

[7] J. Posdamer and M. Altschuler. Surface measurement by space-encoded projected beam systems. *Computer graphics and image processing*, 18(1):1–17, 1982. 4

[8] M. Proesmans, L. J. Van Gool, and A. J. Oosterlinck. One-shot active 3d shape acquisition. In *Proc. of International Conference on Pattern Recognition (ICPR)*, pages 336–340, 1996. 4

[9] R. Sagawa, Y. Ota, Y. Yagi, R. Furukawa, N. Asada, and H. Kawasaki. Dense 3d reconstruction method using a single pattern for fast moving object. In *Proc. of International Conference on Computer Vision (ICCV)*, pages 1779–1786, 2009. 4

[10] K. Sato and S. Inokuchi. Three-dimensional surface measurement by space encoding range imaging. *Journal of Robotic Systems*, 2:27–39, 1985. 4

[11] Y. Shirai and M. Suwa. Recognition of polyhedrons with a range finder. In *Proc. of International Joint Conference on Artificial Intelligence*, pages 80–87, 1971. 3

[12] V. Srinivasan, H. C. Liu, and M. Halioua. Automated phase-measuring profilometry: A phase mapping approach. *Applied Optics*, 24(2):185–188, 1985. 4

[13] V. Sundar, S. Ma, A. C. Sankaranarayanan, and M. Gupta. Single-photon structured light. In *IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2022. 7, 9

[14] A. C. Ulku, C. Bruschini, I. M. Antolovic, Y. Kuo, R. Ankri, S. Weiss, X. Michalet, and E. Charbon. A 512 × 512 SPAD Image Sensor With Integrated Gating for Widefield FLIM. *IEEE Journal of Selected Topics in Quantum Electronics*, 25(1):1–12, Jan. 2019. 6

[15] P. Vuylsteke and A. Oosterlinck. Range image acquisition with a single binary-encoded light pattern. *IEEE Trans. Pattern Anal. Mach. Intell.*, 12(2):148–164, 1990. 4

[16] F. Yang, Y. M. Lu, L. Sbaiz, and M. Vetterli. Bits from photons: Oversampled image acquisition using binary poisson statistics. *IEEE Transactions on Image Processing*, 21(4):1421–1436, 2012. 7