Light Sheet Depth Imaging

Joseph Bartels

CMU-RI-TR-19-05
April 2019

The Robotics Institute
Carnegie Mellon University
Pittsburgh, PA 15213

Thesis Committee:
William “Red” Whittaker, Robotics Institute (Co-Chair)
Srinivasa Narasimhan, Robotics Institute (Co-Chair)
Simon Lucey, Robotics Institute
Matthew Johnson-Roberson, University of Michigan

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy.

Keywords: Depth Imaging, Epipolar Imaging, Light Curtain, Adaptive Sensing

Abstract

As robots are developed for large-scale applications in autonomous driving, package delivery, and agriculture, there is a growing need for affordable and reliable depth sensing. Robots use active illumination sensors like scanning LIDAR and depth cameras to perceive their worlds. Scanning LIDAR is prevalent because it offers long-range, robust sensing, but it is expensive and only captures sparse point measurements. Consumer depth cameras, on the other hand, are inexpensive and produce high-rate, dense depth measurements but fail outdoors in bright light.

This thesis developed active illumination depth cameras and sensing methodologies that combine the robustness of scanning LIDAR with the speed, sampling density, and economy of consumer depth cameras. Rather than sample the entire scene at once like consumer depth cameras or with points like LIDAR, the key approach uses sheets of projected light and imaging to rapidly sample the scene along a single line at a time.

Using this approach, four contributions have been made. The first is a contribution to the development of a light sheet depth imaging device that applies the concept of epipolar imaging to continuous-wave time-of-flight cameras. The resulting depth camera can see up to 15 meters in bright sunlight and is robust to global illumination and motion. The next contribution developed a second generation of this camera that demonstrated sensing ranges up to 50 meters. The third contribution uses the projected sheets of light and imaging to triangulate and sense along a 3D line. By sweeping this line through the volume with galvomirrors, a programmable light curtain is formed that detects objects along its surface at five frames per second. Finally, rapid imaging of programmable light curtains at 60 frames per second was enabled with the custom development of a device that uses the rolling shutter of a camera to steer the imaging plane instead of a galvomirror. The speed and selectivity provided by this device enabled applications in agile depth sensing where scenes are adaptively sampled based on detected regions of interest.

The research developed in this thesis contributes methods and hardware for high-resolution depth imaging that works in challenging conditions, provides methods for computationally inexpensive agile depth sensing, and has the economics that could enable next generation and wide-scale applications in mobile robotics, agriculture, and industrial manufacturing.

Funding: This work was funded in part by NASA Space Technology Research Fellowship NNX14AM53H, National Aeronautics and Space Administration ESI Grant NNX16AD98G, the Office of Naval Research Grants N000141512358 and DURIP N000141612906, the Defense Advanced Research Projects Agency REVEAL Grant HR00111620021, and National Science Foundation Grants CNS-1446601 and IIS-1317749.

Acknowledgments

First, I thank my advisors, Red Whittaker and Srinivasa Narasimhan, for instructing and mentoring me as I navigated my graduate education and research. I am extremely grateful for all of the guidance and support they have provided during my doctoral studies. Especially, I thank them for enabling and encouraging me to develop hardware. When I arrived, I was told that CMU didn't do a lot of hardware, but luckily I managed to find two advisors that care about developing cool hardware as much as I do. Without them I wouldn't be where I am today.

I would also like to express my gratitude to the rest of my thesis committee, Simon Lucey and Matthew Johnson-Roberson. Thank you for making the time to be a part of my committee as well as your perspective and guidance.

With two advisors, I've had the privilege of interacting with two fantastic groups of individuals. Uland, thank you for giving a MechE a chance and getting me started. Heather, Chris, Wennie, Curt, and Eugene, thanks for your friendship and chats over the years. Jon, thanks for pushing me to make better slides and being a great friend. Robert, Jian, Chao, Shumian, Tiancheng, Suren, and Supreeth, thank you for being great labmates and always helping with experiments on short notice. Supreeth, thanks for your early guidance and the groundwork you laid for this research.

Facilitators of my work include Chuck Whittaker, Nora Kazour, Jim Teza, and Tim Angert. Thank you for all your help. A special thank you to Chuck. From last second machining, wiring, ordering, and the many many experiments, I can't count the number of times he's saved the day. Chuck is always there to lend a helping hand and has become a great friend. Thank you, Chuck, for everything you did to help me get here.

In addition to my research colleagues, I'm grateful for all of my friends in the Robotics Institute; especially the Nebraska crew: Eric and Stacie, Nate and Caitlin, and Nate and Alissa. It was nice to have a little home away from home, particularly on Husker Football game days.

My family deserves a lot of credit. Thank you first to my parents, for they taught me the value of education and so much more. They are the foundation of who I am today and I am grateful for everything they have done. Thanks to Chris, Kaylee, Cassie, Colten, Caleb, and Kiersten, as well as Katie's parents for their support and encouragement.

Most importantly, I thank my wife, Katie. Thank you for your endless love, encouragement, and support. Thank you for the sacrifices you made so I could follow my dreams. Thank you for always being there and always believing in me. This would not have been possible without you.

For Katie and Baby

Contents

List of Figures

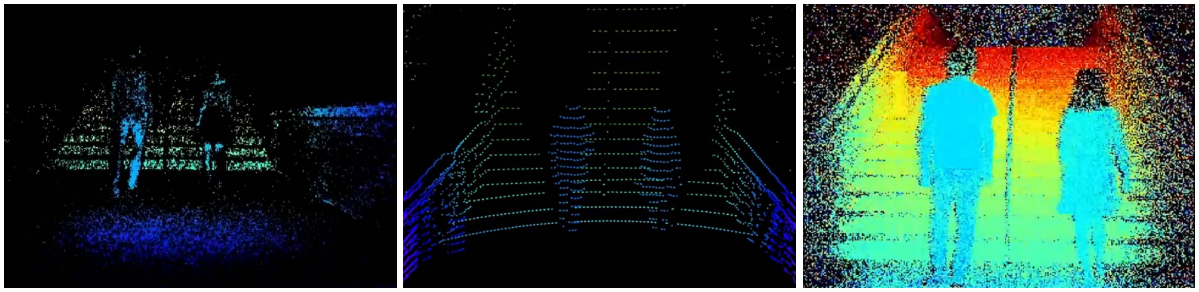
List of Tables

Chapter 1

Introduction

Once confined to industrial manufacturing facilities and research labs, robots are increasingly entering everyday life. As specialized robots are developed for tasks such as autonomous driving, package delivery, and aerial videography, there is a growing need for affordable depth sensing technology. Robots use sensors like scanning LIDAR, depth cameras, and passive stereo cameras to navigate the world. Challenges for depth sensors include conditions such as bright light, darkness, global light transport, and scattering media. Scanning LIDAR devices are the most robust to these conditions, but capture only sparse measurements, have poor coverage at close range, and are expensive. Alternatively, active illumination consumer depth cameras, such as the Microsoft Kinect, and Intel Realsense™ devices, are inexpensive and produce dense, high-rate depth measurements, but fail in bright ambient light, and are susceptible to effects of global light transport. These differences are caused by how each device concentrates and images their emitted light.

Scanning LIDAR devices emit a short pulse of highly concentrated light that when reflected is detected by a co-located receiver to record a measurement. This laser point is then repeatedly scanned to measure the rest of the scene. Consumer depth cameras, instead, emit a broad flash of light that covers the entire scene and exposes the camera for a few milliseconds to capture the image. Both of these devices may receive the same amount of energy for each pixel measurement, but due to the reduced light power per area, consumer depth cameras must expose several orders of magnitude longer to capture an acceptable measurement. This measurement difference has important implications on the ability to image in bright light. In general, for a given amount of energy collected by a camera, shorter exposures and greater concentrations of light result in higher quality imaging. This thesis addresses many of the challenges with current depth sensing technologies by using sheets of projected light and imaging to rapidly sample the scene along a single line at a time, which provides much of the robustness of LIDAR, but maintains the frame rate and resolution of consumer depth cameras, as shown in Figure 1.1.



(a) Consumer Depth Camera **(b)** Scanning LIDAR **(c)** Light Sheet Depth Imaging

Figure 1.1: Stair scene for comparison of depth sensors. (a) Consumer grade depth cameras are inexpensive and provide high-resolution data, but have very limited range outdoors in bright sunlight. (b) LIDAR systems work well outdoors in bright sunlight, but have poor resolution and are expensive. (c) Light sheet depth imaging devices can improve the robustness of consumer grade depth cameras technology so that it works outdoors in bright sunlight.

1.1 Motivation

Although scanning LIDAR devices are more robust than consumer depth cameras and can sense at long ranges, many robotics applications do not require the range provided by scanning LIDAR and the shorter sensing range, increased vertical field-of-view, and higher resolution sampling of outdoor depth cameras would be preferred. The motivation for this research is to increase the performance of depth camera technology such that it can be used to reliably image outdoors in bright sunlight and provide an economical and preferable alternative to scanning LIDAR for short to medium range applications.

By utilizing techniques presented in this thesis that trade some of the robustness of scanning LIDAR systems for the speed, resolution, and economics of depth cameras, a new class of depth sensor is established that offers unprecedented performance compared to existing consumer depth cameras. These devices use low cost, easy to fabricate sensor technologies and are well-placed for wide-scale adoption in a large number of applications.

1.2 Related Depth Sensing Methods

The most common depth sensors include LIDAR, Continuous Wave Time-of-Flight (CW-ToF), triangulation, SONAR, and RADAR devices. Each of these devices has their own benefits as well as limitations. This section briefly discusses these popular technologies.

LIDAR. LIDAR devices emit light in short and highly directional bursts. The light sensors are also highly directional and are co-located to the light source. The concentration of the light both temporally and spatially enables the pulses to be detected even in bright ambient light. Since only a single point of light is being illuminated at a time, very little global light reaches the sensor. Although this configuration offers great performance, the source and sensor must be actuated to change the scan angle and image a different point in the scene, which slows down the system. 2D line scan devices are severely limited in this respect, because to generate 3D measurements the entire LIDAR unit must also be actuated in some way. These line scan LIDARs have been used in many applications and actuated with LIDAR spinners [5], hand-held springs [19], and push-broomed [11] to create dense point-clouds. Other LIDAR devices increase field-of-view and sampling rate by using arrays of lasers and receivers that spin together to create full 360° models of the environment [117]. Most LIDAR devices operate in the near-infrared (NIR) spectrum because NIR light is invisible to the human eye and offers increased reflectance of low-albedo objects. Due to eye-safety concerns, these devices have a maximum rated range of 120 m [117].

Several other LIDAR systems use a short-wave infrared (SWIR) wavelength, of 1550 nm, to increase range to over 200 m [78]. These devices are able to use greater optical power because eye-safety concerns at this wavelength are minimal [7]. Besides being able to use more power, SWIR wavelengths are much less susceptible to scattering media such as smoke and even dust. SWIR devices have the disadvantage of requiring exotic InGaAs materials and difficult manufacturing processes to make the sensors, which, currently, causes these devices to be prohibitively expensive and limits their use in most applications.

Due to the low resolution of LIDAR devices, object recognition can be difficult with only a single or few laser/sensor pairs. Handling dynamic scenes can also be challenging due to the sparse data, and often a separate sensing modality such as a camera or inertial measurement unit is necessary to track motion of the scanner so that point measurements can be registered together from a moving platform.

Flash LIDAR devices address both the speed limitation and the need for moving parts. They emit short, diffuse flashes of light and measure the round-trip time of the reflected light with a two-dimensional imaging array [110]. These devices work well in ambient light, but performance can be degraded by global light transport. These devices are difficult to make and very expensive due to the need for an array of highly sensitive photodetectors with the required temporal resolution for time-of-flight measurements.

Continuous Wave Time-of-Flight Depth Cameras. CW-ToF cameras use undirected continuous wave modulated light rather than pulsed light like LIDAR. These devices typically measure

distance by detecting a phase shift between emitted modulated light and received light, and thus have a maximum unambiguous range dependent on the modulation frequency [70]. These devices are capable of generating high-resolution depth maps at high-frame rates, but are highly susceptible to global light transport and ambient light. CW-ToF imagers based on Photonic Mixer Device (PMD) technology have built-in background light suppression technology [67] which prevents saturation from ambient light but does not remove shot-noise. Several depth cameras have been based on this technology. Examples include, the Microsoft Kinect2 [38], PMDTech sensors [25], and MESA Swissranger devices [89]. All of these cameras are meant for indoor use. The high-resolution data provided by these sensors is useful for many of the same applications as regular visual images and is often paired with visual cameras to provide registered depth information to images. These RGB-D frames are then used in object detection, segmentation, and many other applications where LIDAR is not suitable.

Triangulation-Based Depth Sensors. Triangulation-based depth sensors use a broad class of depth imaging techniques that rely on the triangulation of corresponding points to determine depth. One of the most prolific is stereo-vision. Stereo-vision uses two rectified cameras separated by a baseline distance to image the scene. Corresponding points in the image are matched to compute their disparity and then depth. By matching windows around features, dense depth maps can be reconstructed [87]. The crux of stereo-vision and other triangulation based techniques is that the scene needs to have texture so corresponding points can be identified. If the scene is uniform, there are no uniquely identifiable points to match and stereo-vision will fail. Stereo is usually implemented without a light source and is thus called passive-stereo. To increase texture and stereo reliability, texture is sometimes projected onto the scene with a light source [66]. This method is called projected-texture stereo and is sometimes referred to as active stereo.

When one of the cameras in stereo-vision is replaced with a projector, traditional structured light systems are formed. Instead of imaging corresponding points with two cameras, structured light systems emit light with a projector and then detect it with a camera. If the projector emits light in a single direction and it is detected with the camera, the distance to the scene can be measured by triangulation, assuming the baseline between the camera and light source is known. Line striping systems triangulate depths one line at a time by sweeping this line across the scene to form the image [4]. Line striping has the benefits of trivial line detection and correspondence matching as well as reduction of global light transport effects such as scattering. To increase the rate of line striping, multiple light stripes are projected at time to create a structured light pattern [41, 104]. The scene geometry distorts these patterns and the imaged patterns are decoded

to establish correspondences. Structured light devices can generate depth maps at a few frames per second, but are sensitive to ambient illumination and can be affected by global light transport. Structured light and line striping methods are often used for detailed modeling due to their high precision.

Sonar and Radar. Sonar and radar are most similar to scanning LIDAR in that they emit a pulse of high energy in a specific direction and measure the time it takes for the signal to get back to determine the distance to the object using the speed of either sound or light in whatever medium the light traveled through. Sonar has been used for robot navigation in air [35, 64], but is most often used for imaging underwater due to its long range and predictable performance [71, 122]. Sonar has shown to improve localization results by tracking features in 2D images from imaging sonars [59], but it has too low of resolution for most inspection tasks. Radar is similar to sonar in that its long-wavelength is robust to scattering media and that it has low resolution. Many adaptive driver assistance systems (ADAS) on vehicles use RADAR and SONAR for detection of cars and other objects on and near roadways due to their reliable performance in all conditions. However, radar and sonar both only provide localized measurements and require either arrays of sensors [86] or scanning [37] to create maps of the scene. Sonar and radar have the large advantage of not being affected by scattering media but provide limited resolution that is often not adequate for tasks that require precise 3D information.

1.3 Challenges

Active illumination depth sensors face a number of challenges, many of which relate to the complex interaction between the emitted light and the scene. For example, bright ambient light sources, such as the Sun, can overwhelm the sensor's low-power illumination source which elevates noise and reduces range. Another example is scattering media, like fog or smoke, which attenuates and scatters emitted light from the sensor and corrupts measurements. Another challenge for these devices is sampling efficiency. Depth sensors have a fixed number of points they can image per second and many depth sensors do not use these samples very effectively. In the uniform sampling methods that most depth sensors use, low and high detail areas are sampled at the same resolution even when the low detail area could be defined by much fewer samples. This thesis addresses these three challenges and discusses them here in more detail.

Ambient Light The irradiance of the Sun that reaches the surface of the Earth can reach up to 1120 $\mathrm{W\,m^{-2}}$ on a sunny day [8], whereas light sources used by active illumination systems

are usually much less powerful. Shot noise is the fundamental limitation of any light sensor to detect a weak light signal in the presence of a much stronger source. The r^2 fall-off of effective brightness of a point source with distance heightens this limitation. Even when ambient light isn't strong enough to cause failure of active illumination systems, it adversely effects the signal-to-noise ratio and degrades performance.

Increasing the power of an active light source provides diminishing returns as four times the amount of light is needed to double a given working range. The only way active illumination systems can work in bright sunlight is to block as much of it as possible through design. Since the Sun is a broadband source that emits many wavelengths of light, a large fraction of light can be blocked by bandpass filtering the sensor with an optical filter matched to the wavelength of the emitting source. Most active illumination systems that operate in bright ambient light, and especially those that operate outdoors, use a narrow-band light source and matching optical filters to reduce the amount of light received from the ambient light.

A complimentary approach to reducing the effects of ambient light is by temporally and spatially concentrating the emitted light. LIDAR devices use this approach by emitting very short pulses of high power light, but the circuitry needed to accurately time these pulses for precision range calculation is expensive. Spatial and temporal concentration of light for increased robustness to ambient light is used in many types of other active illumination systems. An example in projector-camera based depth imaging systems include [85], where a raster scanning laser projector was used for line striping outdoors. An epipolar imaging system described in [93] followed a similar approach and paired a raster scanning laser projector with a synchronized rolling shutter camera to concentrate projected light and imaging to a single scanning line. This technique enabled the device to image structured light patterns outdoors for reconstruction in bright sunlight.

Scattering Media Scattering media is debilitating to active illumination imaging devices. Light that is emitted into a scattering media is absorbed and scattered throughout the volume. Images that are captured in scattering media suffer from loss of contrast and brightness. These effects are amplified as distance through the volume increases. When photons are emitted into the scattering media most of them are scattered, but a few will propagate in straight lines, and these are termed ballistic photons. As distance increases, the number of ballistic photons decreases exponentially as a function of the media's scattering and absorption coefficients, otherwise known as optical thickness. Sensing the ballistic photons is difficult especially when the sensor also captures scattered light. For any significant depths, the signal from ballistic photons will be overwhelmed by the scattered photons which saturate conventional sensors and obscure the image.

Most techniques to see through scattering media try to block or reduce the scattered photons in some way or another. Confocal imaging [81] and time gating [83] are highly effective methods of blocking scattered photons from reaching the sensor. Several structured light methods for imaging through scattering media include using thin laser lines [56] and sweeping light planes [88]. Disparity gating [93] is a triangulation based method that reduces scattering by only capturing light from a certain disparity or depth at a time. Polarization and stereo-vision have also been used together to descatter stereo images for reconstruction of dynamic scenes [102]. A larger overview of methods that selectively image the ballistic light is given in [33].

Sampling Efficiency The sampling efficiency of depth sensors is the amount of information they acquire from the scene per sampled point. Most depth sensors sample their field of view at a fixed uniform resolution. Although this sampling method is simple to implement with traditional depth sensors it is usually not efficient, because many sampled points fall on parts of the scene that are very similar and could have been defined by the other sampled points around them (e.g., the ground, walls, large nearby objects, etc). Once captured, this data is used for tasks like obstacle detection and recognition [128] or ground plane removal [24] where the data is heavily processed and most of the data is discarded. Instead of sampling at these locations and then discarding the data, these samples would be better utilized by sampling more entropic parts of the scene, like moving pedestrians or cyclists.

Adaptively sampling the scene and only imaging regions at the necessary resolution results in much higher sampling efficiency. Adaptive sampling based on scene complexity can be used to capture the fine details of complex objects, while sampling simpler objects in lower detail [113]. Similar work has made its way into commercial LIDAR platforms where companies claim their devices can make the perception system 10 times faster by intelligently selecting and sampling regions of interest [3, 52].

1.4 Approach

This thesis develops light sheet depth imaging cameras and sensing methodologies for high-resolution and adaptive depth imaging in challenging conditions such as bright light and scattering media. Light sheet depth imaging uses lines of imaging and lines of illumination to sample a scene. A line of illumination, like a line laser, creates a planar sheet of light when projected onto a scene. Likewise, a line of imaging captures only along a single plane. Geometrically, there are only three ways these two planes can interact. They can be co-planar, they can intersect, or they can be parallel. Each of these configurations results in a different type of light sheet depth

(a) Outdoor Depth Imaging **(b)** Sensing through Smoke **(c)** Adaptive Depth Sampling

Figure 1.2: Light sheet depth imaging improves performance in ambient light, scattering media, and enables high-resolution adaptive depth sampling. (a) Epipolar ToF imaging can see out to 50 m outdoors. (b) Triangulation range gating can be used to see through scattering media. (c) Adaptive sampling with triangulation light curtains provides high-resolution depth sensing out to 20 m and provides much better sampling efficiency than scanning LIDAR (shown as white points).

imaging. The co-planar case takes the form of epipolar depth imaging and the intersecting case is the principle behind triangulation range-gating. These two cases are a form of direct imaging, where the first reflection of the light is captured. The case of parallel planes on the other hand is a form of indirect imaging, where the light is not directly imaged and is captured after multiple reflections. By changing the distance between the parallel planes the captured indirect light can be separated into short and long-range indirect light [69].

This thesis uses epipolar-depth imaging and triangulation range-gating to increase the performance of active illumination depth cameras. These two light sheet depth imaging methods improve performance by concentrating the light and imaging into a single line that is then quickly swept through the volume to generate dense measurements. This compromises some of the robustness of scanning LIDAR for the speed of traditional depth cameras. The resulting imaging methods have less working range than scanning LIDAR, but have higher performance than traditional depth cameras. Specific benefits of light sheet depth imaging include dense outdoor depth imaging in bright light, seeing through smoke, and improved sampling efficiency, as shown in Figure 1.2.

Epipolar Depth Imaging In epipolar imaging, a line of pixels in a camera and a line emitted from a projector are colinear along a common epipolar plane. When the projector and camera are physically aligned so that they are in a rectified stereo configuration, each row of projector pixels corresponds to a row of pixels in the camera image. This geometric arrangement provides several benefits. First, it enables the concentration of light and imaging into a line, which as previously indicated, increases the robustness to ambient light and scattering media. This geometry also is important because the direct component of illumination, or the first bounce of light, always obeys the epipolar geometry constraint between the light source and imager, whereas scattered or multi-bounce light generally does not. This insight from [92], means that by implementing epipolar-imaging, most multi-bounce scattered light can be blocked and the majority of the light captured is direct light. Some multi-bounce light will end up on the epipolar-plane but it is only a small fraction of the light collected. Epipolar depth imaging was demonstrated in [93], with a custom-designed structured light system that showed impressive performance in ambient light and scattering media.

Triangulation Range-Gating Range-gating is the concept that light is *only* received by a camera from a specific distance in the scene, rather than the entire volume. Pulsed range-gating devices do this by emitting a very short (pico to nano-second range) pulse from a light source and then briefly opening the shutter of a time-gated camera for an instance after a certain time has elapsed. Light from a given distance is imaged by delaying the camera for the amount of time it would take for the light to travel from the source to the target plane and then back to the camera. This concept is used in some scanning LIDARs [84] and full-frame range-gated cameras [27, 109] to see through scattering media. With sufficient illumination power these devices can measure depths of greater than 250 m [27] which offers incredible performance in challenging conditions.

An alternative method of range-gating that does not rely on time, but rather on geometry is a technique termed here as triangulation gating. Triangulation gating is the concept that in two-dimensions a ray of illumination and a ray of imaging separated by a baseline and angled such that they are not parallel will intersect only at single point of known depth. At this intersection, light will only be reflected back to, and imaged by, the camera if there is an object at the intersection of the two rays. To image light at different depths, this intersection of light and imaging is steered by the relative angle between the emitted light ray and the imaging ray. This technique is similar to the disparity gating method demonstrated in [93] and works well in scattering media and bright light. Since range-gating methods only sample at a specific depth at a time, multiple measurements must be taken to sample the entire volume like a LIDAR or depth camera.

	Epipolar Imaging		Triangulation Range Gating	
	Epipolar ToF Imaging	EpiToF 2.0	Programmable Triangulation Light Curtains	Agile Depth Sensing with Light Curtains
Ambient				
Scattering Media				
Speed				
Resolution				
Agility				
Reliability				

Table 1.1: Thesis Contributions. Comparison of techniques presented in this thesis based on their characteristics. The red, yellow, and green circles indicate the performance in a given category. Colors are as follows: Poor (Red), Fair (Yellow), and Good (Green).

1.5 Thesis Statement & Contributions

This thesis asserts that 1) imaging a sheet of light with an aligned plane of imaging enables robust, high-rate, and agile depth sensing; 2) when these planes are aligned and scanned in an epipolar configuration, the depth of the entire scene can be captured out to a maximum working range; 3) when the rotation axes of these scanned planes are parallel and separated by a baseline they can triangulate to capture only the depths of objects along a defined surface in the volume; and 4) that these planes can be steered to select and adaptively change which regions of the scene are sampled and at what resolution.

The following contributions have been made in this thesis:

- **Epipolar Time-of-Flight Imaging** [2] (Chapter 2): In this work, the concept of epipolar imaging, described and implemented with rolling shutter cameras in [93], was extended to continuous wave time-of-flight imaging. By imaging lines containing only the direct components of reflected light, epipolar time-of-flight imaging solved many of the common problems that limit time-of-flight cameras, including multi-path interference, ambient light, and camera motion. This work was a collaboration with Supreeth Achar in which the author contributed to prototype development, experimentation, and dissemination.
- **Extending the Range of Epipolar Time-of-Flight Cameras** (Chapter 3): Simulations based on design considerations of various system parameters were used to guide the development of a second generation epipolar time-of-flight camera to improve the working range of the device. This new prototype demonstrated imaging of outdoor scenes at 50 m ranges and pushed the limits of existing hardware.

- **Programmable Triangulation Light Curtains** [119] (Chapter 4): By rotating a light sheet with respect to an imaging plane such that their rotation axes are parallel and separated by a baseline, a device was developed that senses depths only along the line formed by the intersection of the planes. This line can then be rapidly swept, by steering the planes, to form programmable light curtains. The constructed prototype uses a line sensor and line laser both steered by galvomirrors to generate light curtains along any ruled surface and can image curtains in most environments including in smoke and at over 25 m outdoors in bright sunlight. This work was a close collaboration with Jian Wang. The author contributed to prototype development, software, experimentation, and dissemination.
- **Agile Depth Sensing using Triangulation Light Curtains** (Chapter 5): A triangulation light curtain prototype was developed that uses a rolling shutter camera to steer the imaging plane and when synchronized with a light sheet projector can image 60 different curtains per second. This capability enables applications in agile depth sensing as well as adaptive sampling where regions of interest could be sampled on-demand with specified resolution. These capabilities were used for robotic mapping of environments including a cluttered high-bay environment and a smoke-filled tunnel.

1.6 Organization

The remainder of this dissertation is organized in order of the stated contributions. An overview of the related research is first provided within Chapter 1. Chapter 2 introduces epipolar time-of-flight imaging and describes results acquired with an initial hardware prototype. Design considerations for epipolar time-of-flight imaging systems and an advanced prototype are then detailed in Chapter 3. The concept and first prototype for light sheet depth imaging using programmable triangulation light curtains is then discussed in Chapter 4. Based on this concept, Chapter 5 introduces the idea of agile depth sensing using triangulation light curtains. The dissertation is concluded with a discussion on the future outlook of light sheet depth imaging and potential applications of the technology.

Chapter 2

Epipolar Time-of-Flight Imaging

2.1 Introduction

Time-of-flight (ToF) depth sensors have become the technology of choice in diverse applications today, from automotive and aviation to robotics, gaming and consumer electronics. These sensors come in two general types: LIDAR-based systems that rely on extremely brief pulses of light to sense depth, and continuous-wave systems that emit a modulated light signal over much longer duration. The former can acquire centimeter-accurate depth maps up to a kilometer away in broad daylight but they have low measurement rates and their cost per pixel is orders of magnitude higher than the latter—whose range, outdoor operation and robustness are extremely limited. Since low cost, large-scale production and high measurement rate often trump other considerations, continuous-wave ToF (CW-ToF) sensors continue to dominate the consumer electronics and low-end robotics space despite their shortcomings.

In this chapter we present a first attempt to significantly reduce these shortcomings by energy-efficient epipolar imaging. The idea is to project a continuously-modulated sheet of laser light onto a sequence of epipolar planes that is chosen carefully and that spans the field of view. At the same time, only the strip of CW-ToF pixels that belong to each epipolar plane is exposed, as shown in Figure 2.1. Our prototype implementation couples a custom-built projection system to an off-the-shelf CW-ToF sensor that has a controllable region of interest. It outputs live 320×240 3D video at 7.5 frames per second, with the frame rate only limited by the sensor's API.

Epipolar imaging was first proposed for acquiring live direct-only or global-only video with a conventional (non-ToF) video sensor [92]. This approach was then extended to the ToF domain [91] but its energy efficiency was very low and it involved capturing more than 500 images to calculate a single "direct-only" ToF image. In the context of triangulation-based 3D imaging, O'Toole *et al.* [93] showed that significant improvements in energy efficiency and robustness can

be achieved with a 2D scanning-laser projector and a rolling shutter camera. Our approach can be thought of as extending this idea to the ToF domain; as such, it inherits all the advantages of non-ToF energy-efficient epipolar imaging while also tackling major challenges that are specific to CW-ToF.

First and foremost, the range of CW-ToF sensors is severely limited by power consumption and eye safety considerations. Although most sensors electronically subtract the DC component of incident light [90], photon noise from strong ambient sources such as sunlight easily overwhelms the CW-ToF signal at distances more than a few meters outdoors [26, 42, 85]. By concentrating the light source's energy into a single sheet, epipolar ToF boosts this range to 10 meters and acquires useful—albeit noisier—depth signal at over 15 m outdoors.

Second, the depth accuracy of CW-ToF sensors is strongly affected by global illumination effects such as inter-reflections. These effects produce longer light paths and thus show up as a source of structured additive noise. They are extremely common indoors (e.g. corners between walls, shiny surfaces of tables and floors, mirrors, etc.). Methods that aim to cancel the effects of global illumination *a posteriori* [31, 43, 48, 61] require extra image measurements and make strong assumptions in how they model transient responses. In contrast, epipolar CW-ToF optically blocks almost all global light paths prior to acquisition. This provides significant robustness to all forms of global light transport without having to capture additional images.

Last but not least, CW-ToF sensors must acquire two or more frames with a different phase of emitted light in order to compute a single depth map. This makes them highly sensitive to camera shake: unlike conventional cameras where this shake merely blurs the image [36], camera shake in CW-ToF causes the static-scene assumption to be violated [107]. This leads to depth maps that are both blurry and corrupted by motion artifacts. Epipolar ToF makes it possible to address both of these problems: motion blur is minimized because only a very short exposure is used for each epipolar plane; motion artifacts and depth errors are minimized by acquiring multiple phase measurements per epipolar plane rather than per frame. The rolling-shutter-like distortions [10, 63] due to the sequential nature of epipolar-plane ToF can be reduced by scheduling the sequence of epipolar planes so that post-acquisition distortion correction becomes easier.

2.2 Continuous Wave Time-of-Flight

The operating principles of CW-ToF cameras are discussed thoroughly in [70], but a brief summary is provided here. CW-ToF cameras use a temporally-modulated light source and a sensor where the exposure is also modulated during integration. If the illumination modulation function is $f_\omega(t) = \cos(\omega t)$ and the sensor modulation function is $g_{\omega,\phi}(t) = \cos(\omega t + \phi)$ where ω

is the modulation frequency in rad/s and ϕ is the phase offset between the source and sensor modulation functions, then the measurement integrated over an exposure time of t_{exp} at a pixel x is

$$I_{\omega,\phi}(x) = \int_0^{t_{\text{exp}}} f_\omega(t) * [h_x(t) + \beta_x] g_{\omega,\phi}(t) dt \tag{2.1}$$

$$= \tfrac{t_{\text{exp}}}{2} \int_0^{\infty} \cos(\omega\tau - \phi)\, h_x(\tau)\, d\tau \ , \tag{2.2}$$

where $*$ denotes convolution, $h_x(t)$ represents a pixel's transient response to the active light source and β_x is the response due to the DC component of the active light source as well as other ambient sources. In practice, $I_{\omega,\phi}(x)$ is measured by integrating incoming light to two different storage sites (called taps) depending on whether $g_{\omega,\phi}(t)$ is positive or negative and then taking the difference between the stored values. Thus even though β_x drops out of the integral, ambient light still adds to the measurement shot noise.

If there are no indirect light paths between the light source and sensor pixel x, then $h_x(t) \propto \delta\left(t - {}^{l(x)}/_{c}\right)$ where c is the speed of light and $l(x)$ is the length of the path from the light source to the scene point corresponding to x and back to the sensor. Assuming the scene is static, we can recover the path length $l(x)$ by capturing a pair of images at the same frequency but two different modulation phases $\phi = 0$ and $\phi = \pi/2$:

$$l(x) = \frac{c}{2\omega} \operatorname{atan2}\left(I_{\omega,\frac{\pi}{2}}(x), I_{\omega,0}(x)\right) \ . \tag{2.3}$$

The pixel depth $z(x)$ can be computed from $l(x)$ using the geometric calibration parameters of the light source and sensor. This relationship has an inherent 2π phase ambiguity, where any distance greater than the maximum distance is phase wrapped.

2.3 Epipolar Time-of-Flight

The geometry required for epipolar ToF imaging was realized using a 2D sensor with controllable region of interest and a light sheet projector comprised of a line laser source with a 1D-scanning mirror that projects a steerable light sheet onto the scene. The ROI of the 2D sensor is set to be one row tall and the sensor is aligned in a rectified stereo configuration with the light sheet projector to match the requirements of epipolar imaging, as shown in Figure 2.1.

To capture a scene using epipolar ToF, the active plane must be swept across the field of view in one of a variety of ways. Several of these ways are illustrated in Figure 2.2. For example, the ordering in Figure 2.2b illustrates the operation of a hypothetical rolling-shutter ToF cam-

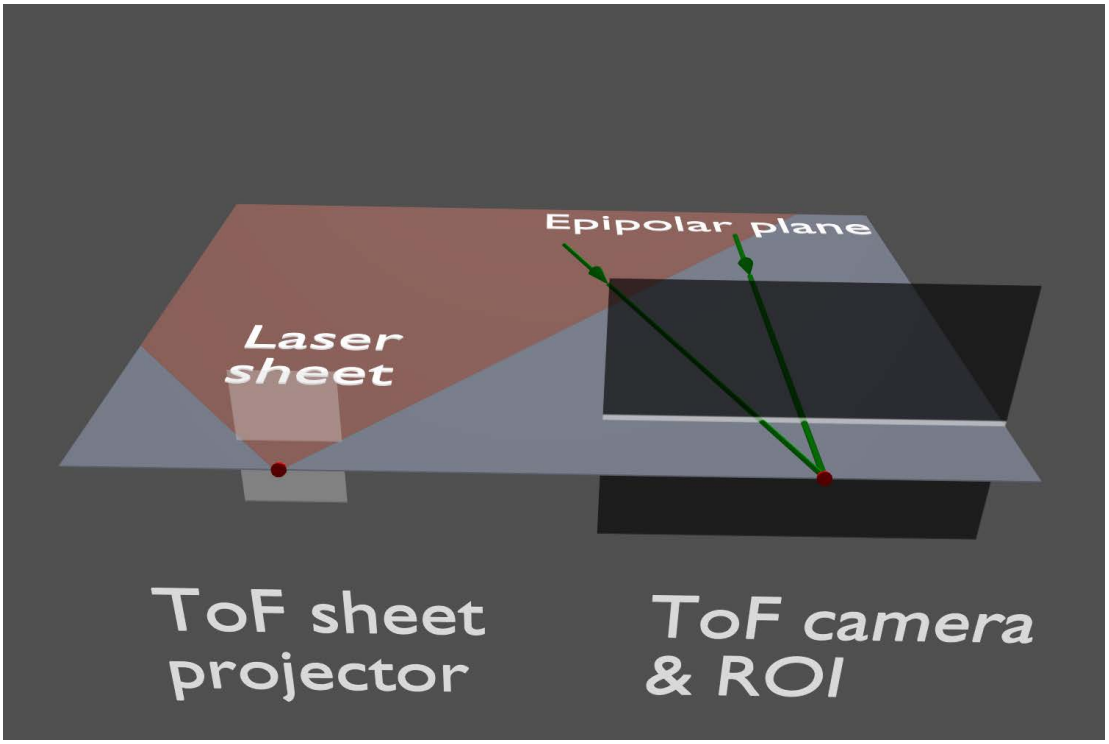


Figure 2.1: Epipolar time of flight. A projector that generates a steerable sheet of modulated laser light is combined with a ToF sensor whose rows can be exposed one at a time. The projector and camera are placed in a rectified stereo configuration so that the light sheet always lies on an epipolar plane between the projector and the camera. At any given instant, only the row of camera pixels on the epipolar plane is exposed to light.

era, where one complete image is acquired per modulation phase. This method is not desirable because if the scene or camera move during acquisition, the recovered depth map will contain errors that are difficult to correct.

A better ordering strategy is to loop through the set of modulation phases at one epipolar plane before going onto the next row (Figure 2.2c). Since the exposure of each row is very short, all phases for a single row can be acquired quickly to minimize depth and motion blur artifacts from scene/camera motion.

Using this strategy, each row is captured at a slightly different time which introduces a rolling-shutter-like distortion effect to the acquired depth map. The individual depth values will be blur- and artifact-free and can be combined into a consistent model by post-processing using continuous time pose-estimation techniques [6, 63]. To make this post-processing easier, we order the plane sampling into a sawtooth pattern (Figure 2.2d), which essentially provides depth maps covering the full field-of-view at twice the frame rate but half the vertical resolution. This makes depth correction easier for fast camera shake and/or scene motions.

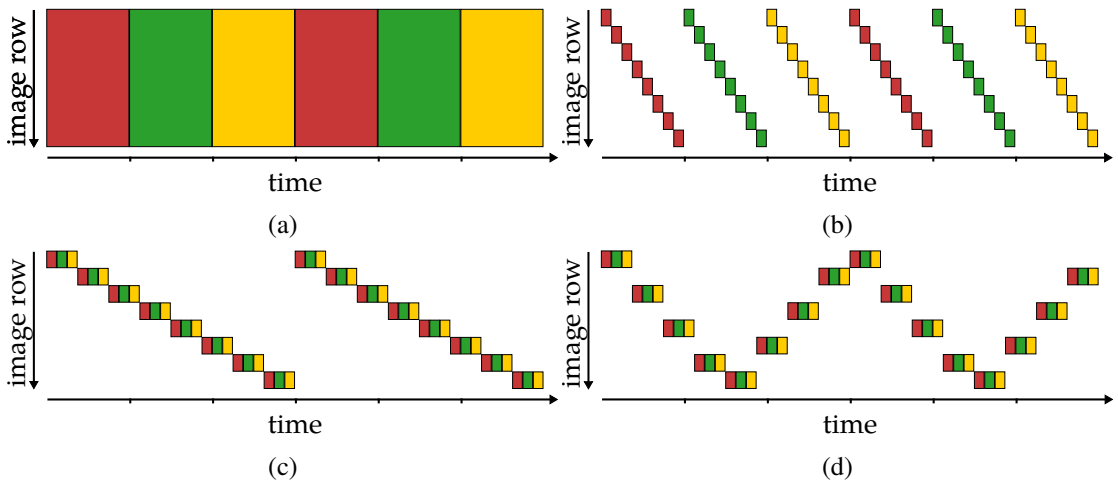


Figure 2.2: Epipolar plane sampling schemes and row exposures in ToF imaging. (a) In conventional CW-ToF all epipolar planes are illuminated simultaneously and all camera rows are exposed at the same time. This requires long exposures and leads to severe artifacts due to motion, ambient light, and global light transport. (b) Ordering the epipolar ToF planes similarly to a rolling-shutter camera confers robustness to ambient light, global illumination and motion blur. Sensitivity to motion remains, however, because of the significant delay between the phase measurements acquired for each row. (c) Interleaving those measurements plane by plane minimizes such artifacts. (d) Scanning the entire field of view twice within the same total exposure time yields higher temporal sampling of the scene and makes consistent merging of individual depth map rows simpler.

2.4 Hardware Prototype

We developed an epipolar time-of-flight hardware prototype that uses a galvomirror-based light sheet projector for illumination and a ToF sensor with adjustable region of interest (ROI) for imaging. The ToF sensor is the Espros Photonics EPC660, which has a resolution of 320×240 and pixels that prevent ambient saturation. The sensor is integrated and accessed with the camera development kit (DME660) from the manufacturer. The camera is fitted with an 8 mm F1.6 low-distortion lens and an optical bandpass filter (640 nm, 20 nm bandpass) to reduce ambient light. The sensor allows the ROI to be changed with every readout which we use to quickly select which row to image.

The light sheet projector we developed uses a 638 nm laser diode with a peak power of 700 mW. The laser light is first collimated and then passed through a Powell lens that produces a uniform, diverging line at a 45° fanout angle. This light sheet is then directed at and steered with a 1D scanning galvomirror. The optical scan angle of the galvomirror provides the projector a 40° field-of-view.

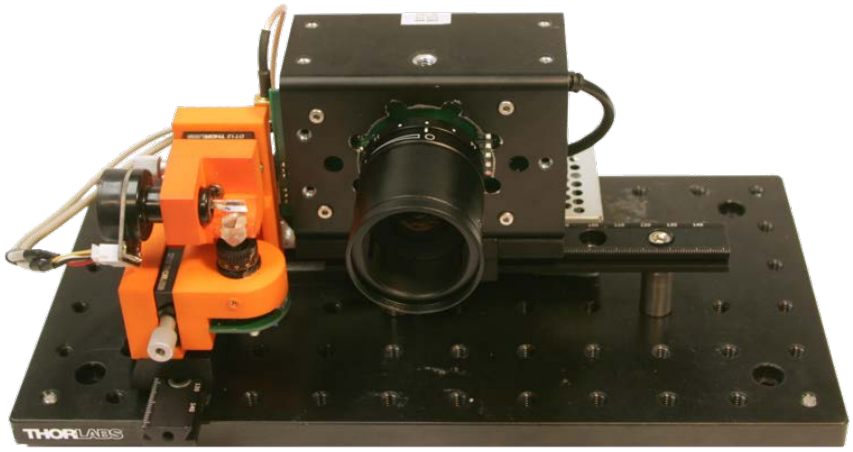


Figure 2.3: Our hardware prototype uses a custom light sheet projector (orange assembly) and a DME660 camera with fast ROI control to capture arbitrary rows of pixels.

A micro-controller (Teensy 3.2) is used to synchronize and command the projector and sensor. It communicates with the sensor over an I^2C bus to set exposure time, modulation frequency, and ROI as well as to trigger the capture. The micro-controller also commands the setpoint of the projector's galvomirror and can read the rotational velocity of the camera from an attached MEMS inertial measurement unit. A frequency generator circuit enables us to select a modulation frequency between $1\ \mathrm{MHz}$ and $24\ \mathrm{MHz}$.

The camera and projector are aligned side-by-side in a rectified stereo configuration as required for epipolar imaging. When precisely aligned, the projected light sheet illuminates a single row of pixels on the camera that is independent of depth. The galvomirror is calibrated to the camera to provide a known mapping between the galvomirror angle and illuminated camera row.

2.4.1 Sensor Calibration

Due to fixed-pattern noise, non-uniform pixel sensitivity, crosstalk between the taps, and small variations in the phase of the exposure modulation function at each pixel, the observed measurements from the sensors do not match their expected values. We model the relation between the expected sensor measurements $I_\omega(x)$ and the actual measurements $\hat{I}_\omega(x)$ using a 3×3 calibration matrix $H_\omega(x)$ at each pixel.

$$\begin{bmatrix} I_{\omega,0} \\ I_{\omega,\frac{\pi}{2}} \\ 1 \end{bmatrix} = H_{\omega}(x) \begin{bmatrix} \hat{I}_{\omega,0} \\ \hat{I}_{\omega,\frac{\pi}{2}} \\ 1 \end{bmatrix} \tag{2.4}$$

$H_{\omega}(x)$ is found by imaging a fronto-parallel surface at a set of known distances z_k, $k = 1, .., K$. At each position of the plane, sensor measurements at different aperture settings $(s = 1, .., S)$ are captured to simulate various scene albedos. For each plane position, k, the path length $l_k(x)$ and expected phase $\frac{2\omega l_k(x)}{c}$ at the pixel are computed. The calibration matrix that best explains the sensor measurements, $I_{w,k,s}(x)$, is then found by minimizing the least-squares error between the corrected measurements and expected phase. This calibration matrix is dependent on both the modulation frequency and the exposure time, so it is found for each setting.

2.4.2 Timing

The time needed for the prototype to image one row is a function of the number of readouts per row n, the exposure time t_{exp}, the readout time for a row t_{read}, and the time t_{mirror} taken by the galvomirror to move to the next rows position in the sampling sequence:

$$t_{row} = nt_{\text{exp}} + (n-1)\, t_{\text{read}} + \max\left(t_{\text{read}}, t_{\text{mirror}}\right). \tag{2.5}$$

The image sensor in the prototype has two-taps, so at least $n = 2$ readouts are required to measure the depth using a single modulation frequency. A timing example is shown in Figure 2.4, where t_{read} is 175 µs and for most experiments t_{exp} was 100 µs. In our row sampling sequence, the mirror rotates through two rows per step and, for this step size, t_{mirror} is approximately 100 µs. Combined, t_{row} is 550 µs when $n = 2$, which provides a frame rate of 7.5 fps (or 3.8 fps when $n = 4$).

2.4.3 Limitations

The frame rate is currently limited by the read-out time enforced by the development kit. Our approach requires only one sensor row to be read out, but the EPC660 sensor supports a minimum region of interest of 4 rows, which forces us to read out three extra rows that are unused. The development kit also limits the sensor readout bus to 20 MHz when the sensor actually supports rates up to 80 MHz. The frame rate is also limited by the minimum exposure time necessary to collect sufficient light at a desired range, which depends primarily on peak laser power. Our prototype uses a 700 mW laser source while most other experimental time-of-flight imaging

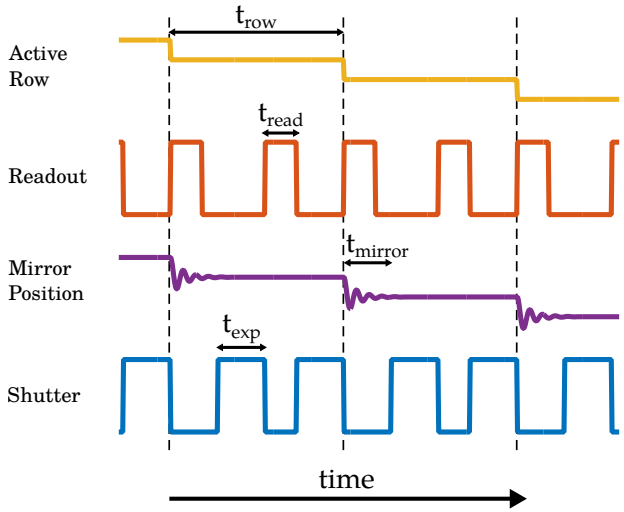


Figure 2.4: Timing diagrams for camera exposure, readout, and mirror position for a particular sequencing of the rows. First, the scanning mirror is moved to the new active row and takes t_{mirror} to settle in the position. When the previous row readout is complete (which takes t_{read}) and the the mirror is in position, the camera is triggered. Each exposure lasts for t_{exp} and at the end of each exposure the row is read out.

systems have peak light power in the 3 W to 10 W range. With a more powerful light source, much shorter exposure times could be used without loss of range. Lastly, the settling time of the low-cost galvomirror we used could be replaced with a higher-grade galvomirror with lower settling times, or even with a faster 1D MEMS mirror. With these improvements, a system based on this prototype would operate at video frame rates.

The sensor used for this prototype supports a maximum modulation frequency of only 24 MHz whereas most other time-of-flight sensors can operate in the 50 MHz to 100 MHz range. This low modulation frequency limits the depth precision and the prototype's ability to accurately scan smaller objects. The EPC660 datasheet specified that the sensor ADC returns 12-bit values but the preliminary version of the sensor we are using only returns 10-bit values, which affects the range and makes output depth maps noisier.

2.5 Results

We demonstrate the benefits of epipolar ToF imaging by comparing to regular ToF imaging in different scenes and conditions. For accurate comparison we implemented regular ToF imaging with our prototype sensor. The first way we implemented regular ToF was by removing the light sheet projector and replacing it with a diffused light module using the same laser and peak power. The second method we implemented was to still use the light sheet projector but keep

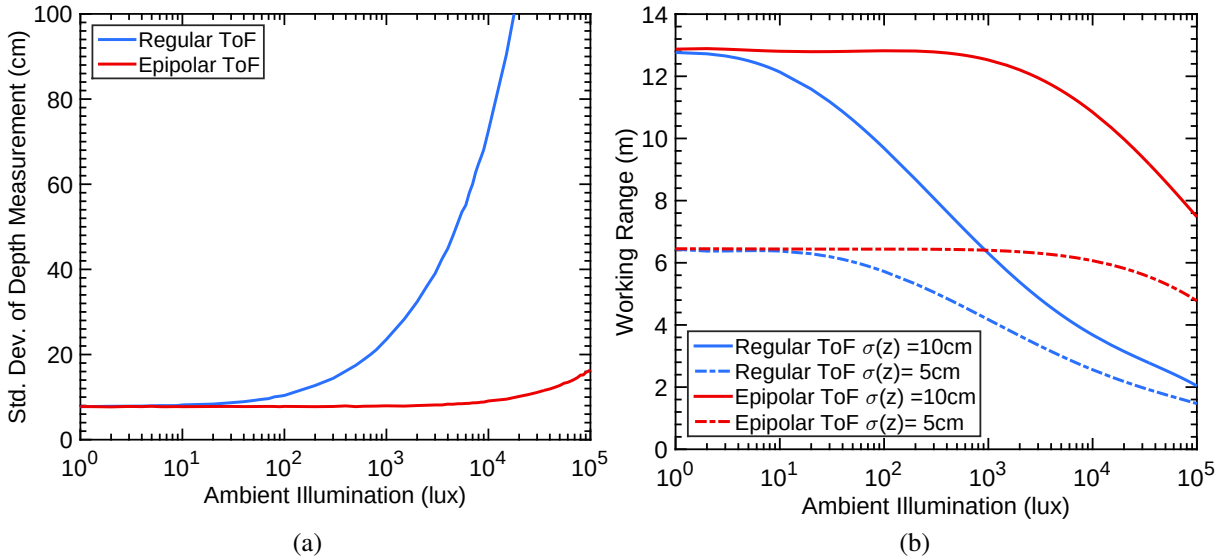


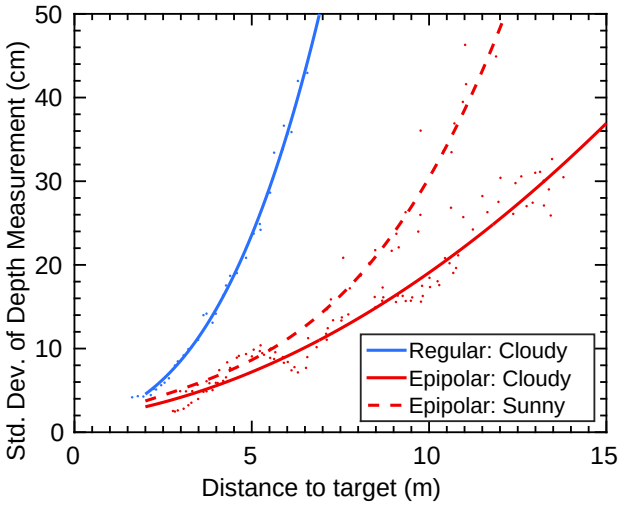
Figure 2.5: A simulation of the standard deviation in depth measurements obtained using regular and epipolar ToF imaging (15 MHz modulation frequency) for a target 10 m from the camera as a function of ambient light level is shown in (a). For both cases, the peak light source power is 2 W and the total exposure time is the same (7.2 ms per image) but epipolar ToF is more robust to ambient light because it concentrates light source power and uses a short exposure for each row (30 $\mu\mathrm{s}$). (b) shows the working range of the same simulated camera at different levels of acceptable range accuracy. Note that simulated camera's parameters differ from prototype.

the entire sensor exposed until the projected sheet had been swept across the full field of view. For the camera motion experiments we used the diffused light module, but for the ambient light comparisons we used the full frame ROI approach to prevent light loss at the diffuser from affecting our comparisons.

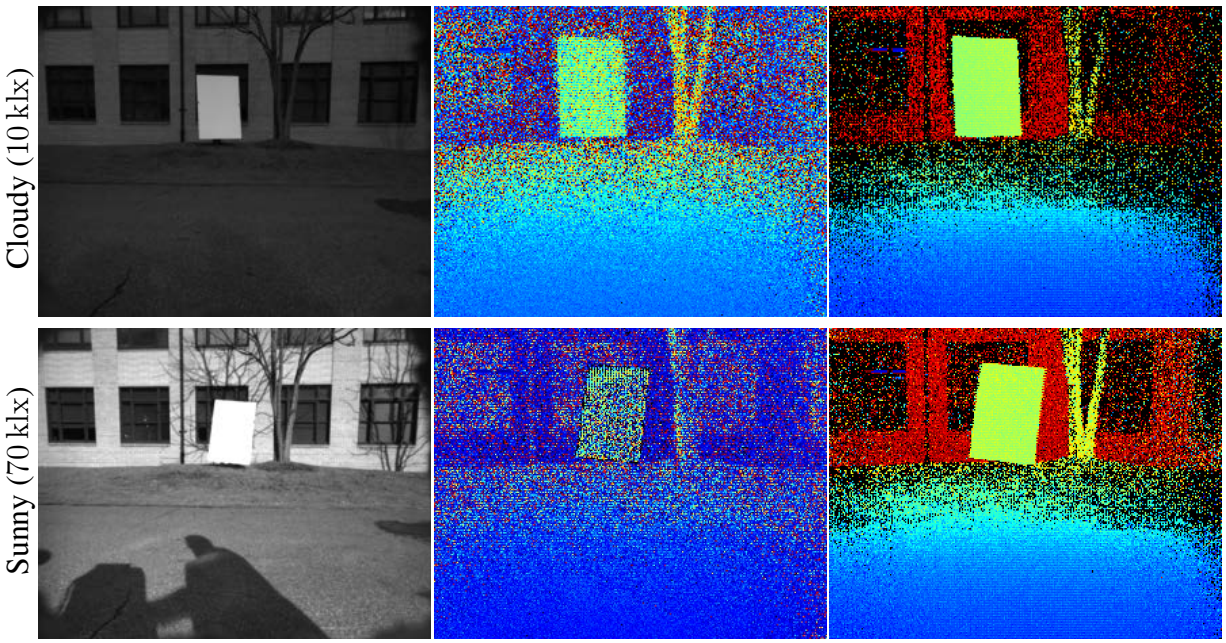
2.5.1 Ambient Light

CW-ToF cameras have been engineered to be unaffected by ambient light by implementing charge dumping, ambient light suppression, and by using bandpass filters on the imager. However, the fundamental limitation of these devices is that they cannot remove the shot noise from the system. By broadcasting the light over the entire image they must expose longer than an epipolar system which will increase the amount of shot noise they receive. Even in burst modes, where a higher-power illumination source is used with short exposures, the light is still not as concentrated as an epipolar system.

Figure 2.5 shows working range and depth error simulations of a regular and epipolar ToF system with the same illumination power and system parameters as the EpiToF prototype. For



(a) Measurement Errors



(b) scene (c) Regular ToF (d) Epipolar ToF

Figure 2.6: We placed a white planar target at a range of distances from the sensor in both cloudy weather and bright sunshine. Even under cloudy conditions, epipolar ToF imaging produced far less noisy depth measurements than regular ToF. Under bright sunlight, regular ToF failed completely whereas epipolar ToF still provided useful depth returns. Depth maps in column (c) and (d) range from 0 m to 15 m. The camera modulation frequency was set to 10 MHz. (a) shows standard deviation in depth estimates versus distance to target (slower rising curves are better). Our prototype has depth error of around 3% at 10 m in bright sunlight.

the epipolar simulations the exposure time of frame is divided by the number of rows (240) to get an exposure time of 100 µs per row. So the regular ToF system exposes the entire frame for 24 ms, whereas the epipolar system exposes each of the 240 rows for 100 µs to provide the same amount of exposure time on each frame. Range simulation is performed as described later in Chapter 3.

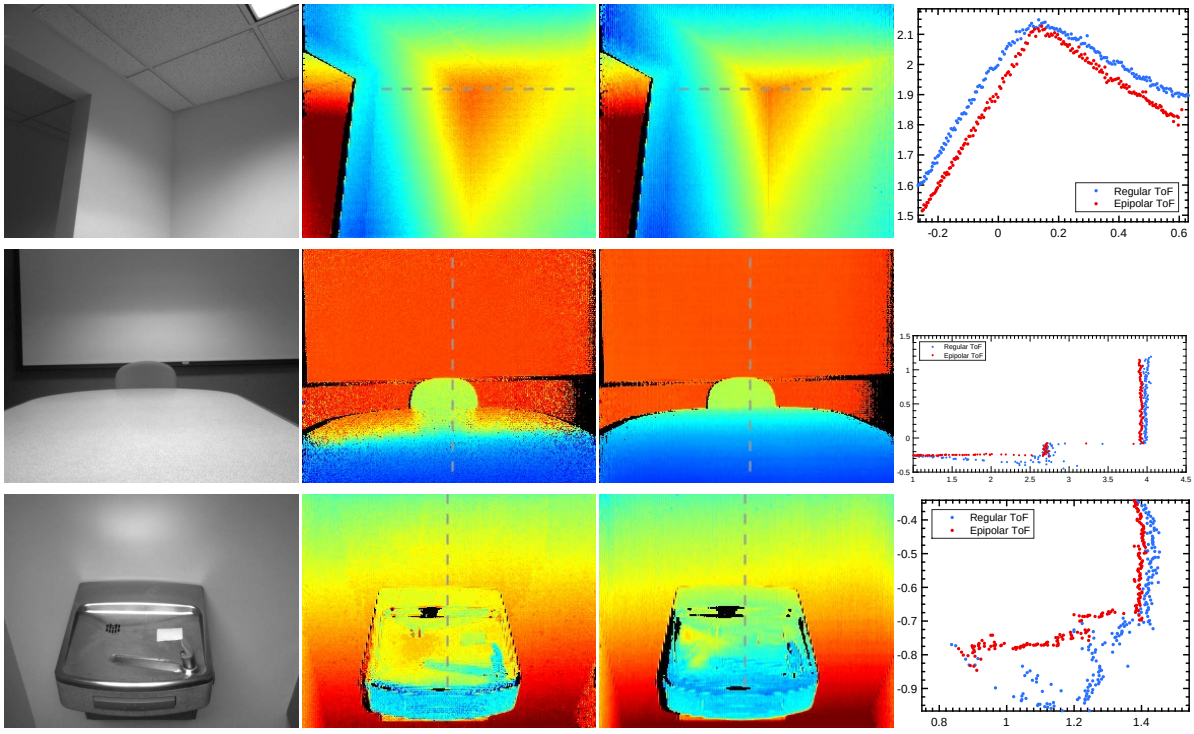
For the working range simulation, a maximum depth error (or resolution) of 5 cm and 10 cm was used. As expected and shown in Figure 2.5, the epipolar ToF imaging system performs better than the regular ToF system in ambient light. As the ambient illumination and distance increases, the working range decreases for both systems, but the EpiToF system provides much less error. In these simulations, it is assumed that the illumination lines up perfectly with the imaging to occupy only one row, which is a reasonable assumption given the size of ToF pixels. These simulations do not take into account any proprietary ambient suppression circuitry the imagers use, and when compared to empirical results, any differences seem negligible.

A simulation that illustrates the benefits of epipolar time-of-flight imaging in brightly lit environments is shown in Figure 2.5. For a given light source power, depth accuracy degrades rapidly with regular time-of-flight imaging as ambient light increases from 0 lux (complete darkness) to 100 klx (direct sunlight). Epipolar imaging degrades much more gradual.

A quantitative comparison of our prototype operating outdoors under both sunny and cloudy conditions in regular ToF and epipolar ToF imaging modes is shown in Figure 2.6. Regular ToF mode performs poorly in bright sunlight, while epipolar ToF is considerably more robust.

2.5.2 Global Illumination

The ability of epipolar ToF imaging to suppress the effects of global illumination is shown in Figure 2.7. These results were generated using a single modulation frequency of 24 MHz. In regular ToF mode, the diffuse interreflections between the walls and ceiling cause depths to be overestimated and the corner to be rounded. With epipolar imaging, the walls appear straight and meet at sharp right angles. The conference table in the second row appears specular at grazing angles and reflects light off the projector screen behind it. The water fountain is particularly challenging because the surface reflects mostly indirect light back to the sensor and the direct reflections are very weak. For epipolar imaging, we combine multiple exposures to recover a sufficient direct signal. Longer exposures do not help regular imaging because the interreflections saturate the sensor.



(a) scene (b) Regular ToF (c) Epipolar ToF (d) Profile

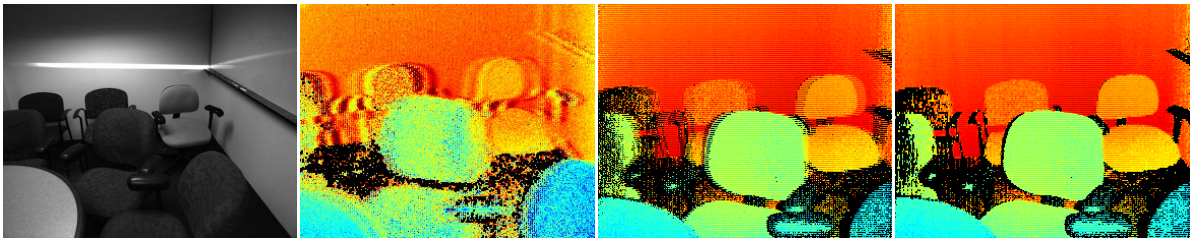
Figure 2.7: Comparing depth maps with epipolar and regular ToF imaging in the presence of global light transport: diffuse interreflections at the corner, glossy interreflection from projection screen onto a shiny conference table, reflections from in between the wall and the shiny water fountain. Epipolar ToF eliminates most of the global light transport resulting in depth maps that are significantly more accurate than regular ToF. All profile curves are in meters.

2.5.3 Camera Motion

Due to each row being captured at slightly different times, the captured frame will show effects from rolling-shutter-like distortion. Consider the case of a rotating camera with known rotational trajectory acquired from a MEMS gyroscope, shown in Figure 2.8. With regular imaging, motion blur and strong artifacts appear at depth discontinuities because the measurements are not aligned. With epipolar ToF imaging, motion blur has basically no effect on the depth values, but the depth map will have rolling shutter distortion. This distortion is corrected with a simple warp computed from the known rotation.

2.5.4 Outdoor Depth Imaging

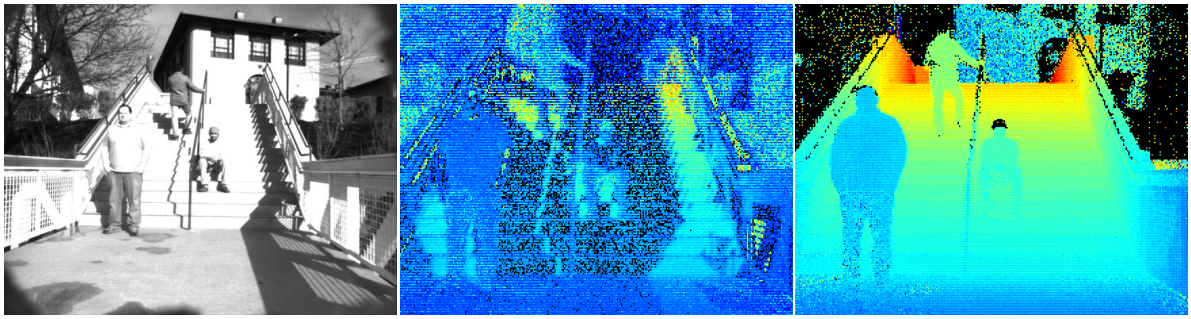
Epipolar ToF imaging has the capability of producing depth maps of scenes outdoors in bright sunlight (70 klx), as shown in Figure 2.9. While regular ToF imaging is only able to correctly



(a) scene (b) Regular ToF (c) Epipolar ToF (d) Corrected Epipolar

Figure 2.8: Fast camera motion causes blur and misalignment between images captured with regular ToF cameras (b). With epipolar ToF cameras, the motion causes rolling shutter distortion (c) that can be corrected with a simple image warp if the motion is known (d).

sense depths in the shadows of the scene, epipolar ToF imaging can reconstruct the entire scene. These results were generated with a modulation frequency of 10 MHz which provided a 15 m unambiguous range. Wrap-around is noticeable on the wall of the building beyond the staircase which is more than 15 m away.



(a) scene (b) Regular ToF (c) Epipolar ToF

Figure 2.9: Outdoor Depth Imaging. (a) An outdoor stair scene was captured with our ToF camera on a sunny day. (b) When the camera is operated in regular ToF mode the depth is only properly reconstructed in the shadows. (c) In epipolar mode the full scene is reconstructed.

2.6 Discussion

Epipolar imaging for continuous-wave time-of-flight depth cameras mitigates many of the problems commonly encountered with these devices. The line-by-line sampling mechanism employed by epipolar imaging mitigates problems of highly degraded performance in bright light, systematic errors from global illumination, and artifacts induced by sensor motion. Compared to depth cameras, systems like scanning LIDAR that illuminate and image a single point at a time are very robust to all of these effects but have a low measurement rate. Epipolar imaging

can be considered as a compromise between these two extremes of full-field capture and point-by-point capture. Because epipolar imaging illuminates and captures a single line at a time, it enables a depth camera to have much of the robustness of point scanning while still having a high measurement rate.

In our prototype, the scanning mirror follows a sawtooth pattern and captures rows in an ordered sequence. However with a faster mirror, row sampling strategies could be implemented that would enable adaptive sampling of the scene based on current scene knowledge. This would enable high-resolution and high-rate depth imaging of areas of interest and lower-resolution, lower rate sampling of other areas. Pseudo-random row sampling strategies could also be used with compressed sensing techniques to temporally super-resolve depth maps of fast-moving scenes.

Chapter 3

Extending the Range of Epipolar Time-of-Flight Cameras

3.1 Introduction

Epipolar time of flight imaging solves many common problems with continuous wave time-of-flight imaging, as described in Chapter 2. The first epipolar time-of-flight prototype, referred to from now on as EpiToF, demonstrated that pairing a modulated light sheet projector with a conventional 2D ToF camera and quick ROI adjustment was an effective implementation of epipolar time-of-flight. The EpiToF prototype demonstrated working ranges of up to 15 m in bright daylight and had a maximum frame rate of 7.5 fps.

Among its successes, the first EpiToF prototype also demonstrated the weak points of the design and left several areas for improvement. Improving many of these weak areas required a completely new design based on a ToF imager with characteristics that did not seem to exist at the time. The desired characteristics were higher modulation frequency, quicker row readout, and better modulation contrast, as well as the ability to quickly change the region of interest. Other areas of improvement such as working range and reliability of the device could be improved modestly with a new hardware design.

This chapter discusses the simulation-based procedure used to design a second generation epipolar time-of-flight imaging prototype, EpiToF 2.0, that has improved working range and higher reliability. Finally, results from EpiToF 2.0 are shown in a variety of scenes and then current limitations and the outlook for this technology is discussed.

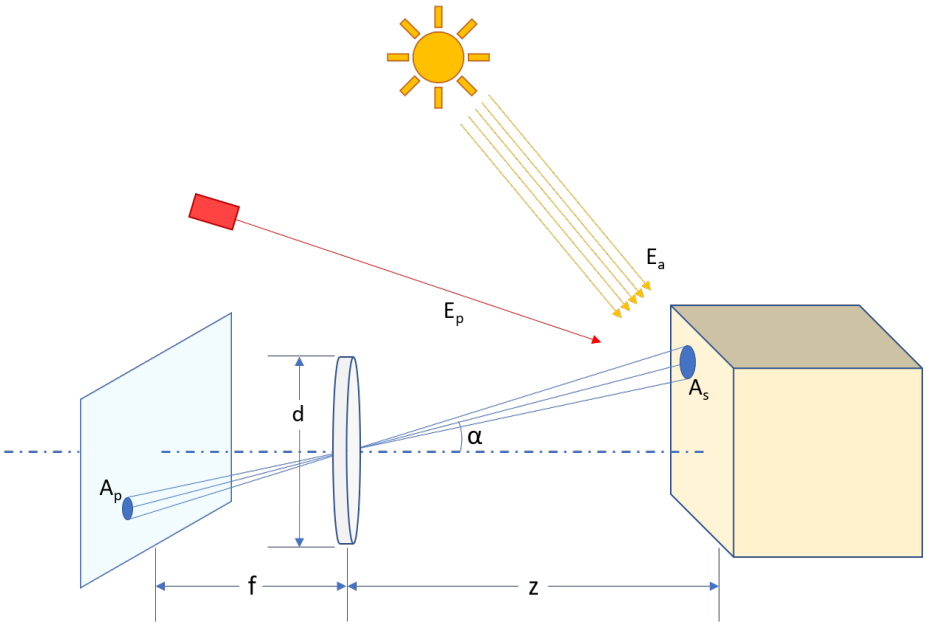


Figure 3.1: Diagram of image formation model, showing how the irradiance from the ambient light and a laser projector corresponds to the pixel on the image sensor surface.

3.2 Epipolar ToF Range Simulation

Range of active illumination systems are limited by noise. Contribution of ambient light to the image is essentially independent of distance, but light from the active illumination source has an inverse square fall off with increasing distance. When the ambient light reaching the camera is small compared to the light from the active illumination source, the range is limited by read noise. But as the relative contribution of ambient light to the image increases, the light from the active source is lost in the shot noise of the ambient component. When the ratio of the signal received to the noise received is less than a defined signal to noise ratio, then the device is defined to be at max working range. The amount of light received by the camera depends on many aspects of the system, including illumination power, camera optics, image sensor sensitivity, laser wavelength, ambient light, and scene albedo. The range of an active illumination camera system is analyzed here with the image formation model for such a system.

3.2.1 Image Formation

A camera with lens of aperture diameter, d, and focal length, f, viewing a scene patch at an angle, α, as shown in Figure 3.1, measures the irradiance of the scene, E_m, which is related to the radiance of the scene, L_s, by (3.1), where η_L is a loss factor to account for transmission losses in the camera optics [51].

$$E_m = \eta_L L_s \frac{\pi}{4} \left(\frac{d}{f}\right)^2 \cos^4 \alpha \tag{3.1}$$

Assuming point sources and a lambertian reflectance from the surface with albedo ρ_d, the radiance of the scene, L_s, is related to the incoming irradiance on the scene, E_s, at incident angle, θ_i by (3.2).

$$L_s = \frac{\rho_d}{\pi} E_s \cos\theta_i \tag{3.2}$$

For an active illumination imaging system, the irradiance on a scene patch is a combination of the irradiance from ambient light, E_a, and that emitted from the projection system, E_p. In these equations radiance and irradiance units are W·m^{-2}·sr^{-1} and W·m^{-2}, respectively.

$$E_s = E_a + E_p \tag{3.3}$$

Signal

The energy (in Joules) on a camera's pixel, $\mathcal{E}_p$, from the measured irradiance during a given exposure t_e (in seconds) is given by (3.4), where A_p is the area of the camera pixel in m^2.

$$\mathcal{E}_p = E_m A_p t_e \tag{3.4}$$

The number of photons to generate this energy is determined by dividing the energy on the pixel by the energy of a photon hitting the pixel. The energy of a photon, e_p, depends on its wavelength, λ and is calculated with (3.5), where h is Planck's constant in J·s, and c is the speed of light in m/s. Although each different wavelength in the bandpassed ambient light and the laser light have slightly different energies, it is assumed here that they have the average energy of the projector's wavelength.

$$e_p = \frac{hc}{\lambda} \tag{3.5}$$

The number of photons is then calculated with (3.6) and then converted into signal, e^-_{sig}, by finding the number of collected electrons with the quantum efficiency, η_{qe} of the image sensor, as shown in (3.7).

$$n_p = \frac{\mathcal{E}_p}{e_p} \tag{3.6}$$

$$e^-_{sig} = \eta_{qe} n_p \tag{3.7}$$

Noise

The noise of the camera is comprised of shot noise, readout noise, and dark current noise. Shot noise describes the inherent fluctuation in the arrival rate of photons on the imager and is a Poisson process. The shot noise, e^-_{shot}, is calculated as the standard deviation of the mean number of electrons generated on the imager, e^-_{sig}, which for a Poisson process is the square root of the mean,

$$e^-_{shot} = \sqrt{e^-_{sig}}. \tag{3.8}$$

RMS readout noise, e^-_{read}, is the noise associated with the electronic digitization of the signal and is constant with time. Readout noise can be determined by either calibration [97] or estimated from values on the datasheet $\left(e^-_{read} \approx Q_{sat}/10^{SNR/20}\right)$. The dark noise, e^-_{dark}, is due to thermal generation of electrons, e^-_{dcur}, in the imager and is highly dependent on temperature of the imager. Total dark signal noise depends on exposure time and the RMS value is calculated as

$$e^-_{dark} = \sqrt{e^-_{dcur} t_{exp}}. \tag{3.9}$$

The total noise e^-_{noise}, of the camera can be calculated as an RMS sum of the shot noise, readout noise, and dark signal noise, as shown in (3.10) [96].

$$e^-_{noise} = \sqrt{(e^-_{shot})^2 + (e^-_{dark})^2 + (e^-_{read})^2} \tag{3.10}$$

Projector Irradiance

The irradiance a scene receives from a laser illumination system is a function of the laser's power, P_L, the area of the scene imaged by a single camera pixel, A_s, and how many camera pixels the reflected irradiance of the laser covers, n_{pix}, during a single exposure. For a global shutter camera and illumination system, where a flash of diffused light is pulsed with the capture of an entire image of resolution $w \times h$, $n_p = w \times h$ and each pixel receives 1/($w \times h$) the power of the laser power. In epipolar imaging, where a single row is illuminated, $n_p = w$, and each pixel receives 1/(w) the power of the laser. To calculate the projector irradiance, E_p, (3.11) is used. The area of the scene, A_s imaged by a single camera pixel with area A_p, is calculated with (3.12), where z is the distance to the scene [51].

$$E_p = \frac{P_L}{n_{pix} A_s} \tag{3.11}$$

$$A_s = \frac{\cos\alpha}{\cos\theta_i} \left(\frac{z}{f}\right)^2 A_p \tag{3.12}$$

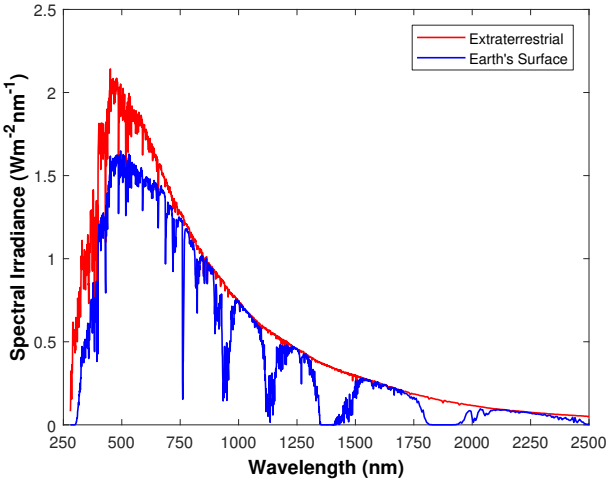


Figure 3.2: Solar irradiance on the atmosphere (extraterrestrial) and on the Earth's surface. The atmosphere absorbs some wavelengths of light and reduces Earth's exposure to them.

Solar Irradiance

The sun is a 5800 K blackbody at a distance of 1 AU (1.496×10^{11} m) from the Earth. The total irradiance of the Sun on the atmosphere of the Earth is roughly 1367 $\mathrm{W\,m^{-2}}$, and is known as the solar constant [8]. Each wavelength of light has a different spectral irradiance, $\bar{E_\lambda}$, and each wavelength is absorbed differently in the atmosphere, as shown in Figure 3.2 to provide a total irradiance on the Earth's surface of 1000 $\mathrm{W\,m^{-2}}$. The total irradiance of a specific range of wavelengths, $E_{\lambda_1 \to \lambda_2}$ from wavelength λ_1 to wavelength λ_2 is found by integrating the spectral irradiance, $\bar{E_\lambda}$, of the wavelengths in the range as shown in (3.13). Their is no closed form solution, so this must be done numerically. Integrating over the entire distribution verifies the total irradiance values provided by [8].

$$E_{\lambda_1 \to \lambda_2} = \int_{\lambda_1}^{\lambda_2} \bar{E_\lambda} d_\lambda \tag{3.13}$$

3.3 Range Simulation for Design

In the design of new epipolar imaging devices the maximum working range as it relates to changes in system parameters is of key interest. The maximum working range, z^*, is determined by finding the range where the emitted light signal captured by the imager, S_p, is equal to the captured noise, N, by a specified signal-to-noise ratio, SNR.

$$S_p^* = SNR \cdot N^* \tag{3.14}$$

As detailed in [38, 45, 70, 115] and summarized in Chapter 2, CW-ToF systems use several images captured at different modulation phases to compute distance. The signal of a pixel, x, of a time of flight cameras is defined as the amplitude, α, of the measured signal, which is computed using

$$S_p(x) = \alpha(x) = \frac{\sqrt{\left(I_{\omega,\frac{\pi}{2}}(x) - I_{\omega,\frac{3\pi}{2}}(x)\right)^2 + \left(I_{\omega,0}(x) - I_{\pi}(x)\right)^2}}{2}, \tag{3.15}$$

where $I_{\omega,\phi}(x)$ are the number of electrons at the pixel captured at different modulation phases. Here, the working range is defined using measurements captured at four different modulation phases ($\phi = 0, \pi/2, \pi, 3\pi/2$) as is commonly used in practice, instead of the two measurements as described in Chapter 2. Using four measurements improves results by averaging out noise.

The noise of the measurements is comprised of shot noise from the ambient light and projector, as well as read noise, and dark noise. The shot noise from the ambient light first requires the ambient light be calculated as the DC offset coefficient,

$$\beta(x) = \frac{I_{\omega,0}(x) + I_{\omega,\frac{\pi}{2}}(x) + I_{\omega,\pi}(x) + I_{\omega,\frac{3\pi}{2}}(x)}{4}, \tag{3.16}$$

which is just the average of the four readings. Although this DC offset due to ambient light cancels out during the subtraction of measurements, its shot noise does not. The read noise is also subtracted out during the measurements since it is constant, and the dark noise is considered negligible at these short exposures. Therefore the noise, N, is just the sum of the shot noise from the ambient light and the projector,

$$N(x) = \sqrt{\alpha(x) + \beta(x)}, \tag{3.17}$$

where α and β are in electrons.

The signal-to-noise ratio is important to define the performance of the system, as it determines the depth resolution of the system. The depth resolution, Δd, is a function of the shot noise from each measurement and requires error propagation of (3.15). It is calculated and simplified from [70] as

$$\Delta d = \frac{d_{max}\sqrt{2}}{8} \cdot \frac{1}{SNR}, \tag{3.18}$$

where SNR is the measured signal to noise ratio. Rearranging (3.18), the required signal-noise-ratio can be found to achieve a required depth resolution at the maximum working distance.

Table 3.1: EpiToF System Parameters

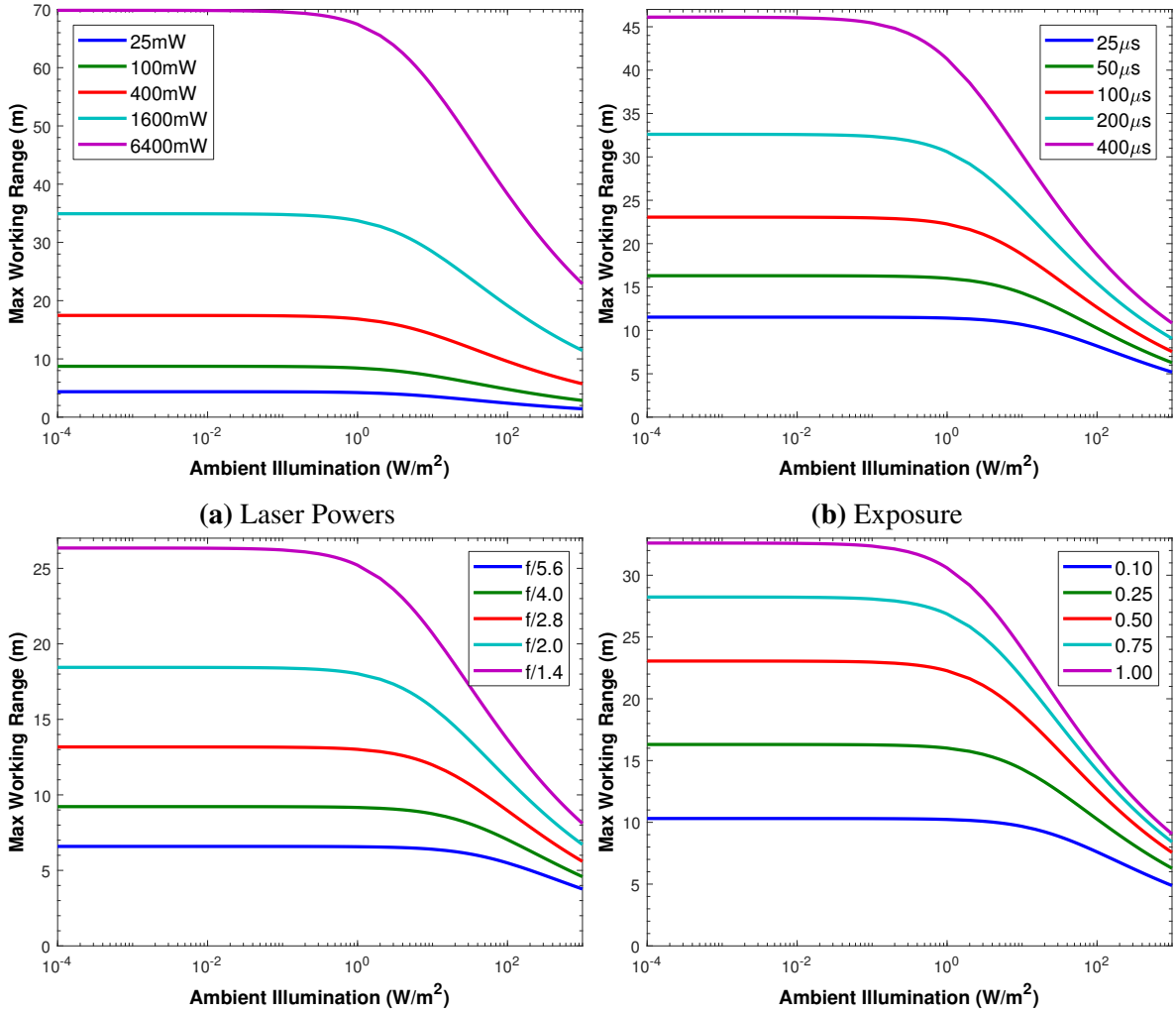
	Parameter	Value
Scene	Albedo	0.5
Camera	Model	ESPROS DME660
	Resolution (w x h)	320 x 240
	Pixel Pitch (μm)	20
	Exposure Time of Frame (ms)	24
	Quantum Efficiency	0.71
Filter	Center (nm)	638
	FWHM Passband (nm)	20
	Transmission	0.95
Lens	Focal Length (mm)	8
	Aperture (f-number)	1.6
	Lens Transmission	0.80
Projector	Optical Power (mW)	90
	Wavelength (nm)	638
	Field of View	45°H $\times$ 30°V

3.3.1 Comparison of EpiToF Systems

This section compares configurations of the EpiToF system detailed in Table 3.1 by varying laser power, row exposure, f-stop, and albedo as shown in Figure 3.3 and the modulation frequency as shown in Figure 3.4. All parameters are unchanged from the table unless they are the parameter varied in the given plot. All simulations were performed with a modulation frequency of 15 MHz, unless otherwise stated. As expected, these plots show that increasing laser power, lens aperture diameter, scene albedo, and row exposure time all increase working range. The effects of these changes are substantial without ambient light, but provide smaller yet significant increases in working range.

For laser power, these plots show the effect of the inverse-square law of light and that to double the working range, a four-fold increase in power is required. A reasonable powerful 400 mW laser doesn't enable imaging over even 10 m in bright sunlight. To increase range to a working distance of 25 W, a powerful 6.4 W laser is necessary with a depth error of 10 cm and 15 MHz modulation frequency.

The f-stop number of a lens, describes the aperture of the lens and is important in the performance of the system. The lens aperture is an opening in the lens through which light travels and determines the cone angle of light that focuses on the image plane. The f-number (N) (or f-stop) of an optical system is the ratio of the lens's focal length (f) to the diameter(D) of the entrance pupil ($N = f/D$). The amount of light captured by the lens is proportional to the area of



(a) Laser Powers **(b)** Exposure

(c) F-Stop **(d)** Albedo

Figure 3.3: Effects of system parameters on working range with a max depth error of 10 cm. (a) Maximum working range approximately doubles with the quadratic increase in laser power. (b) Increasing exposure has a large effect on performance in darkness, but has marginal effects in ambient light. (c) Decreased f-number results in a larger aperture and more light being let in. (d) The higher the scene albedo the more light that is reflected.

the aperture and a smaller f-stop number increases the aperture area and more light is collected, which increases range.

Albedo of the scene has large implications on the working range, as well. The lower the albedo of the scene, the less light that is reflected and the shorter the working range. Albedo ranges from 0.0–1.0, where 0.0 is a completely light absorbing material, and 1.0 is a completely reflective material (similar to a retro-reflector). Changing the albedo of the scene is not usually possible, and most common objects are in the albedo range of 0.05–0.5. For example, a white

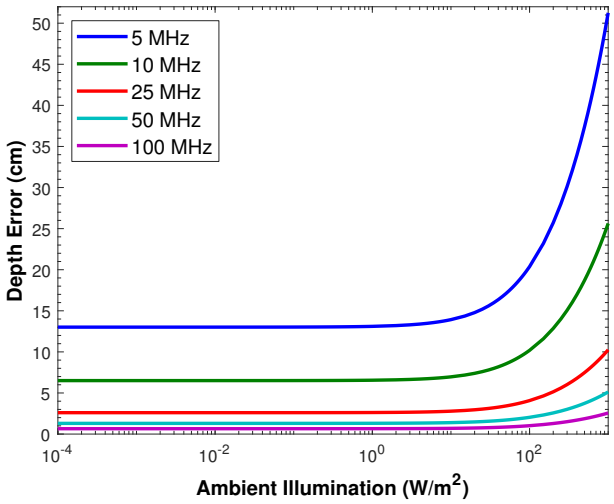


Figure 3.4: Effects of modulation frequency on depth error in ambient light. The modulation frequency directly controls the depth error, thus with a higher modulation frequency, the depth error is smaller for a given range, but the unambiguous range is shorter. With a high modulation frequency a depth error of a few centimeters can be achieved in bright ambient light.

piece of paper has an albedo of 0.6, grass has an albedo of 0.25, and black paint has an albedo of 0.05. The albedo plot shows that there is a significant increase in range with albedos under 0.5, but a much smaller increase in the albedo range of 0.5–1, when viewed in full ambient light.

When in darkness, increasing row exposure produces significant increases in range due to the increased time to collect the emitted light. However, when not in darkness, the longer exposure also increases the amount of ambient light (and shot noise) captured, which quickly reduces the benefits to working range.

As indicated in the simulation calculations, the modulation frequency has a large effect on the depth resolution of the system. The smaller the modulation frequency, the better depth resolution, but lower the unambiguous range. With systems capable of modulation frequencies of 50–100 MHz, the precision can be just a few centimeters in bright light, but the unambiguous range is 3 m and 1.5 m, respectively.

3.3.2 Simulation Insights

The developed range simulation methods can be used to design new devices with increased performance. Given a certain image sensor, several system parameters are easily adjusted to improve performance. The laser power is shown to have the greatest effect on performance, so the higher the laser power, the better the range. Increasing the lens aperture has a modest but beneficial effect on range. Matching the laser wavelength with the optimal transmissions of the lens, filter,

and quantum efficiency of the image sensor can have large benefits when done together. One of the largest benefits to the device is to switch to near infrared. Near infrared light has higher reflectance off many low albedo objects like grass, dark clothing, and some dark paints. Increasing the effective albedo of the scene can have large implications on detection range depending on the material.

3.4 Hardware Prototype

A new hardware prototype was designed based on an iterative multi-parameter analysis of different system configurations. Besides optimizing the working range, the prototype had to abide by a key component constraint; the ToF image sensor. At the time of design, the ESPROS EPC660 was the only ToF sensor identified that could change its ROI during the readout of every row. As this feature is critical to the operation of our epipolar camera, we were required to use this image sensor. Instead of pursuing the custom design of read-out electronics and interface to this sensor we opted to keep using the development kit, which sped up the development time, but maintained the limited frame-rate of the camera.

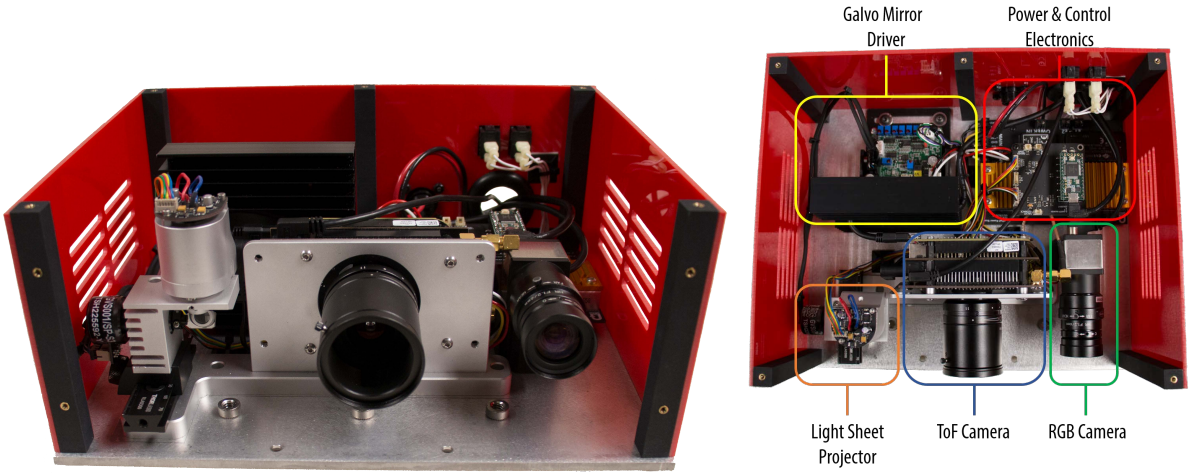
The completed EpiToF 2.0 prototype we built is shown in Figure 3.5 and detailed in Table 3.2. The light sheet projector, ToF camera, and other components are housed in an enclosure to help protect the system. The front and top panels of the enclosure are removed in these images to provide a full view of the system. A color helper camera was added to this system for color visualization of the scene and could be used for creating RGB-D imagery similar to that of other depth camera systems.

Light Sheet Projector. Many of the light sheet projector and imaging components were chosen based on the optimal wavelength of the system. A near-infrared wavelength of 830 nm provided the longest working range and also satisfied component availability. Other common wavelengths such as 808 nm and 850 nm were considered but, due to increased lens transmission losses or reduced image sensitivities, were not selected. The laser selected was a 830 nm laser diode with 1 W of peak output power (Thorlabs LD830-MA1). The light sheet projector assembly was redesigned based on the one used in the EpiToF. In addition to a different laser diode, the light sheet projector uses a more reliable galvomirror (Thorlabs GVS001) with a 50° scan angle. A collimation tube (Thorlabs LTN330-B) and 45° Powell line lens (Thorlabs PL0145) are used to form the laser line. The new projector mounting was custom designed and machined out of aluminum for simpler and more precise alignment as well as better heat-sinking ability.

Camera Optics. The Kowa LMVZ4411 vari-focal lens was selected for the camera optics as it provided an excellent combination of low-distortion (<0.35%) and light throughput with its f/1.6 aperture. A bandpass filter was then carefully selected to match the system's operating wavelength and 45° field-of-view. Bandpass filters are a very important aspect of the system for ambient light suppression and must be carefully selected to match the laser's wavelength. These filters are interference based and are sensitive to angle-of-incidence (AOI), so when the AOI is increased from 0° the features of the filter's spectrum shift to shorter wavelengths. This means that the center pass frequency of the filter moves to a shorter wavelength and light entering the filter at non-normal angles could be blocked if the shift is greater than the bandpass. Although, a tight bandpass filter blocks out more ambient light it will also block out more of the light entering at the edges of the lens due to the increased AOI. Therefore, the filter must have a correct center frequency and wide enough passband to avoid this light loss. For our system, we selected a nominal 56 nm bandpass filter centered at 842 nm. When at a 25° AOI (half the maximum field-of-view), this filter shifts the center-frequency to 825 nm with a 25 nm FWHM passband and at the nominal 0° incident angle the filter will still transmit light at 830 nm. Although this filter lets in a little more ambient light than ideal, it also ensures reception of all the projected light. This filter is placed between the camera lens and image sensor to minimize the size of the filter

Table 3.2: EpiToF 2.0 System Components

		EpiToF 2.0
Camera	Model	ESPROS DME660
	Resolution (w × h)	320×240
	Pixel Pitch (μm)	20
	Number of Exposed Rows	1
	Quantum Efficiency	0.7
Filter	Center (nm)	830
	FWHM Passband (nm)	56
	Transmission	0.95
Lens	Model	Kowa LMVZ4411
	Focal Length (mm)	8mm
	Aperture (f-number)	1.6
	Lens Transmission	0.63
	Field of View	45° H × 45° V
Projector	Optical Power (mW)	1000
	Wavelength (nm)	830
	Field of View	45° H × 45° V



(a) Front View **(b)** Detailed Top View

Figure 3.5: The EpiToF 2.0 prototype is comprised of a light sheet projector, ToF camera, RGB helper camera, a galvo mirror controller, and power and control electronics.

and for protection. Ideally, a custom filter would be designed specifically for this application that would optimally match the AOI and wavelength requirements.

Mounting & Alignment. A large limitation of the original EpiToF prototype was its reliability and ease of alignment. Issues with misalignment due to bumps and vibrations were common and most incidents required a tedious process of realignment and calibration. Precisely aligning the projector and camera systems was difficult but was required for epipolar imaging. Fixing these issues was a priority in the redesigned system. To enable quicker, easier, and more reliable alignment, the structures of all mechanical mounting components were redesigned to include easy-access fine-pitch adjustment screws where necessary. Many of the alignment issues were remedied by precise measurement and low-tolerance design and machining. Precisely designing the location and alignment of the centers of imaging and projection and only providing the required degrees of freedom enabled much simpler and more reliable alignment.

Electronics. Many of the electronics on the new prototypes were custom-designed. Although bench-top electronics could have been used to drive the galvomirrors and lasers, a smaller solution was necessary for a mobile prototype. The electronics designed include a laser driver, galvomirror driver, and system controller. The system controller interfaces with a host computer that initiates frames and then commands the laser driver and galvomirror driver to generate the necessary signals to modulate the laser and move the galvomirror. Reliable design of these electronic systems and their embedded software was necessary for a robust system.

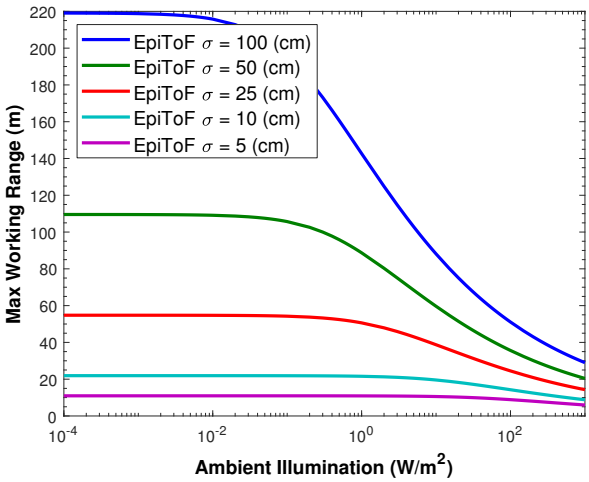


Figure 3.6: Simulated working range of EpiToF 2.0 at varying depth errors operating at a modulation frequency of 10 MHz and scene albedo of 0.5

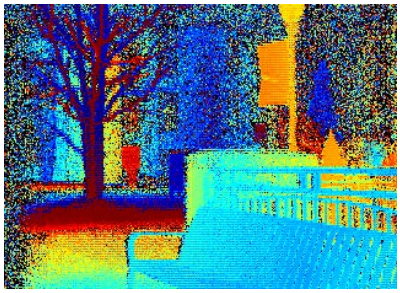
3.5 Results

The redesigned EpiToF was evaluated in simulation and experimentally. The prototype was used to capture data outdoors in both bright sunlight and under cloud cover. It was then compared to regular ToF, similar to comparisons in Chapter 2. Unambiguous ranging out to 50 m was performed using a single low-frequency and compared to using the same low-frequency to unwrap a data captured at higher-frequencies. The device was then used for long-range imaging of static and dynamic scenes.

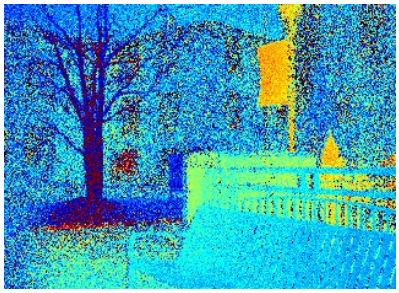
3.5.1 Simulated Results

Simulation results for the redesigned system, as detailed in Table 3.2, are shown in Figure 3.6. These simulations show the working range of the EpiToF as a function of ambient light with scene albedo of 0.5. The working range of the EpiToF system (operating at f_{mod} =10 MHz) is shown at different acceptable depth errors. This data was simulated with nominal exposures of 100 $\mu\mathrm{s}$ per capture. The maximum working range (ambiguous) of the sensor could be increased if a higher modulation frequency was used. For example, a modulation frequency of 20 MHz would double the working range at an error of 5 cm, but it would also have a maximum unambiguous range of only 7.5 m. The 10 MHz modulation frequency allows for a 15 m unambiguous range and is the primary frequency used on the device. For example, on a cloudy day (10 $\mathrm{W\,m^{-2}}$), the device can see 60 m, and the error percentage is then less than 1%. Although the device is able to see at these distances, the range is outside its unambiguous range and will phase-wrap several

(a) Scene



(b) EpiToF 2.0



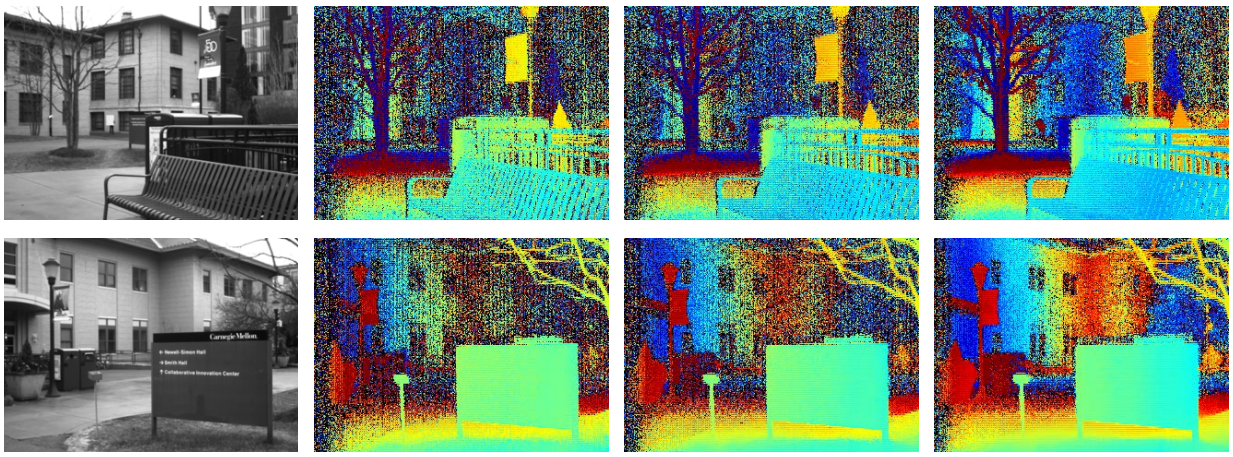
(c) Regular ToF

Figure 3.7: Comparison between EpiToF 2.0 and regular ToF imaging on a cloudy (10 $\mathrm{W\,m^{-2}}$) day. Even under cloudy conditions regular ToF is much noisier than epipolar time-of-flight. The white target is barely visible in the regular ToF image, while it is clearly visible in the EpiToF 2.0 image. The white target was at a range of 35 m, and the exposure time per row was 400 µs.

times. At the EpiToF's unambiguous range of 15 m, when operating at 10 MHz, it has a working error of 25 cm in full daylight, which is approximately 1.5%.

3.5.2 Initial Results

The first results from the EpiToF 2.0 were collected outdoors on a cloudy (10 $\mathrm{W\,m^{-2}}$) day. Before testing the device was calibrated as described in Chapter 2. The device imaged a white target until loss of detection in the depth images. Results comparing the EpiToF 2.0 to regular ToF (using the same system, but with regular imaging method) are shown in Figure 3.7. These results show that, even on a cloudy day, regular ToF is much noisier than epipolar time-of-flight. The white target is very noisy and barely recognizable in the regular ToF image at 35 m, while it is clearly visible in the epipolar captured images. Results from two scenes with varying exposure are shown in Figure 3.8. In the first row of images, a white target is set at 50 $\mathrm{m}$and is identifiable in the depth image captured with a 400μs exposure. These depth maps show phase-wrapping as the detectable ranges exceed the unambiguous range of the device. These images were captured with a modulation frequency of 10 MHz, so the unambiguous range occurs at 15 m. The captured images show several phase-wraps as indicated by the coloring from blue to red and then red to blue. The depth images shown in these figures are the raw data from the sensor and have not been thresholded with a signal-to-noise ratio to identify maximum range. With this type of filtering, the noise in these images could be significantly reduced.



(a) Scene (b) 100 µs exposure (c) 200 µs exposure (d) 400 µs exposure

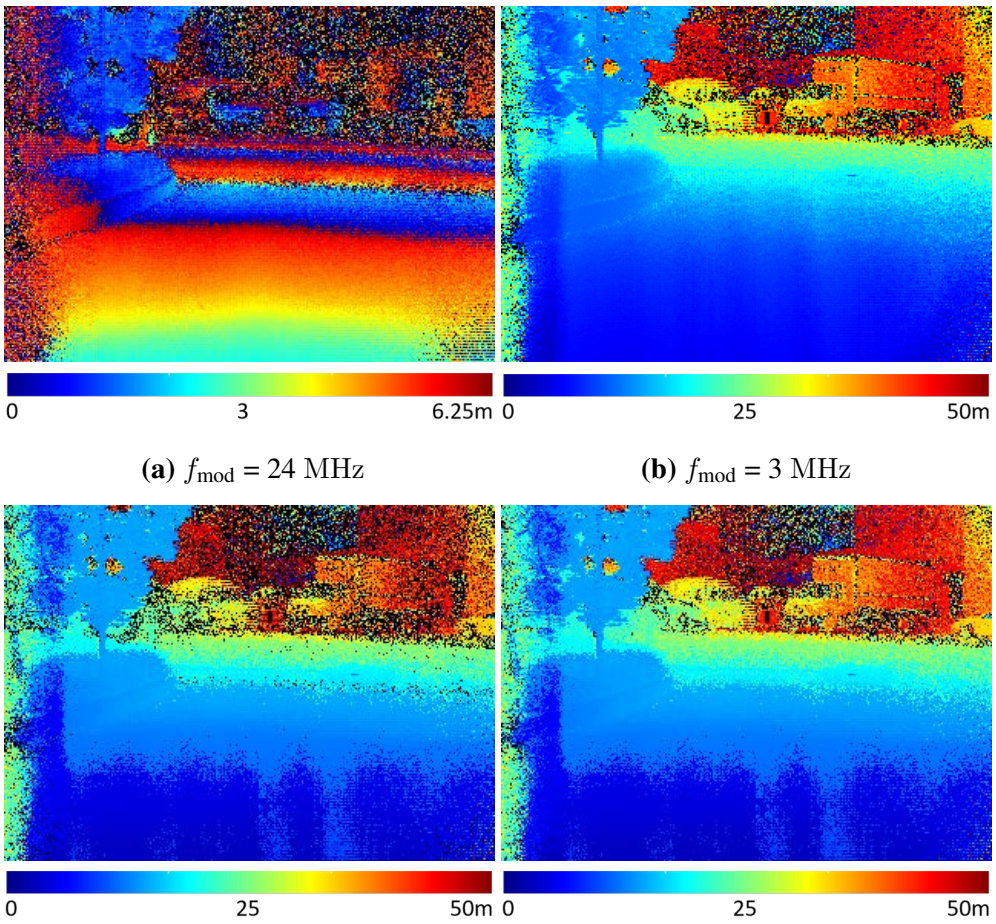
Figure 3.8: Initial unprocessed results from EpiToF 2.0 outdoor testing on a cloudy day (10 $\mathrm{W\,m^{-2}}$). Two different scenes were captured at several exposure times. As the exposure increases, so does the working range. Maximum range in these images is approximately 50 m. At these ranges phase wrapping occurs several times and parts of the scene are out of the 15 m unambiguous range. The white target towards the middle of the scene was at 50 m in the first row of images and is visible in the depth image captured with 400 µs exposure. The inner corner of the building in the bottom row has a range of 40 m and is identifiable with all exposures but has high noise levels. Depth error reduces as exposure increases.

3.5.3 Phase Unwrapping

Although these results show that the device is capable of imaging at ranges greater than the original EpiToF, the results don't have much significance due to phase ambiguities. To enable applications at these ranges, we must be able to sense unambiguously at these ranges. The maximum unambiguous range of a time of flight camera with modulation frequency f_{mod} is

$$d_{\text{max}} = \frac{c}{2f_{\text{mod}}}, \tag{3.19}$$

where c is the speed of light. To avoid phase wrapping a modulation frequency with large enough wavelength can be used so that the entire range of interest is resolvable in one period. Although this increases the unambiguous range, it sacrifices depth precision as more range is forced to be resolved with fewer bits. This effect also increases depth error because each bit of shot noise accounts for a larger depth error. To avoid this and still resolve the full-range of interest, two different modulation frequencies can be used together [32, 43, 45, 60]; one with high depth resolution and short ambiguous range and one with low depth resolution but long unambiguous range. A simple way of combining these is to use the depth from the lower modulation frequency, d_{low}



(a) f_{mod} = 24 MHz **(b)** f_{mod} = 3 MHz

(c) Combined (Raw) **(d)** Combined (SNR)

Figure 3.9: Phase unwrapping of ambiguous ToF depth maps using two frequencies. (a) A depth map was captured with a 24 MHz modulation frequency, which has a maximum unambiguous range of 6.25 m but low depth error. (b) Another depth map was captured with a 3 MHz modulation frequency, which has a maximum unambiguous range of 50 m, but higher depth error. (c) A full-range depth map is produced but with higher-precision by guiding the period of the phase-wrapped image in (a) with the depths from (b). (d) Combined full-range depth map where if the signal to noise ratio of the pixels in (a) are larger than a specified value the high-frequency image is unwrapped, but if the SNR is lower than the threshold, the depth value from the lower-frequency image is used. This results with the front of the image having higher-precision than the back of the image.

as a guide to find which period the waveform is in by solving (3.20) for n_p.

$$d_{\text{low}} = d_{\text{high}} + n_p d_{\text{max,high}}, \tag{3.20}$$

where d_{high} is the depth from the higher frequency, and $d_{\text{max,high}}$ is the maximum depth for that frequency. By using the low precision but full-range depth map to guide the period of the waveform

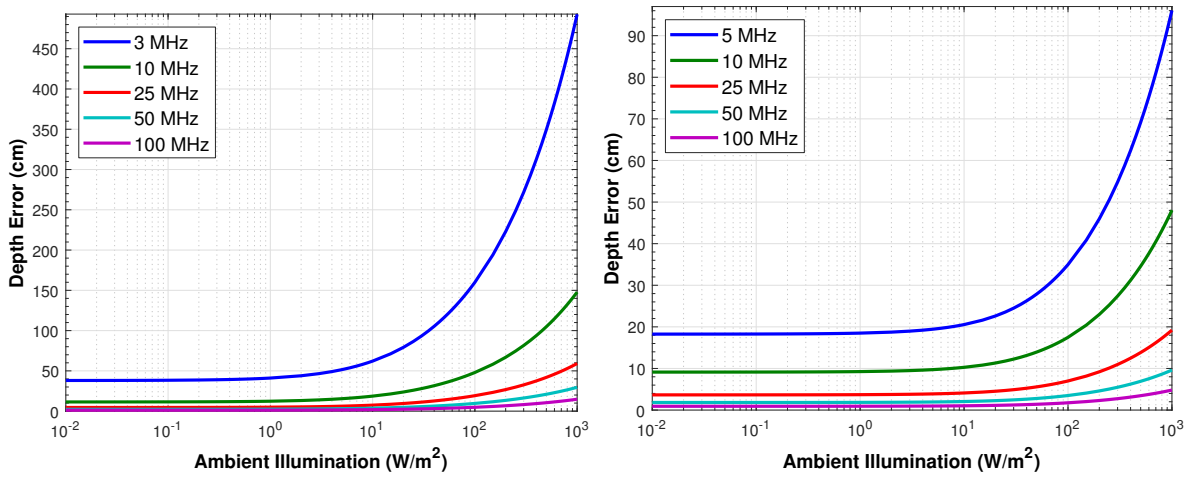
a more precise full-range depth map can be generated.

An example of this is shown in Figure 3.9 where d_{high} and d_{low} are 24 MHz and 3 MHz, respectively. At these frequencies their maximum unambiguous distances are 6.25 m and 50 m, respectively. The repeating depths in Figure 3.9a show the ambiguity caused by the short range compared to the full-range depth. It is also noticeable that the depth map captured with the higher 24 MHz frequency is noisier than the one captured at the 3 MHz. This seems to occur from decreased modulation contrast at the higher frequencies. This means that the sensor is not able to sense the changes in received light as well as it can at lower frequencies. This could either be caused from the sensor's modulation contrast or poor electrical design on the modulation circuitry. Regardless, it drastically reduces the effectiveness of using higher modulation frequencies and phase-unwrapping for reduced depth error with this prototype.

When the measurements are combined instead of fusing the noisy parts of the image, they are fused based on a simple signal to noise threshold. If the higher frequency image at a certain pixel is above a certain signal-to-noise ratio then (3.20) is used to calculate the depth of the pixel. Otherwise the depth of the pixel in the 3 MHz image is used. These results show that we can phase unwrap using two frequencies, but that imaging at the higher frequency in addition to the low frequency has little advantages with this prototype, especially due to the increased time it takes to capture the additional measurements at that frequency. For frame-rate considerations it is beneficial to only image at the lower frequency than take the minor near-range improvements from the dual sampling.

3.5.4 Long-Range Depth Imaging

Long-range depth imaging with traditional CW-ToF cameras is not possible outdoors in sunlight but it is possible with epipolar time-of-flight cameras. To evaluate this capability of the new prototype, several scenes were imaged with the device. Sampling the scene using a lower frequency is preferred due to decreased sensitivity and range at higher frequencies, so the long range scenes were captured at the 3 MHz modulation frequency and with an exposure of 400 $\mu\mathrm{s}$. The ambient light level of these scenes was around 500 $\mathrm{W\,m^{-2}}$ and the 50 m unambiguous range provided by the 3 MHz modulation frequency was adequate for the maximum range of the device in these conditions. Figures 3.11 shows the various scenes captured with the prototype. The second row of this figure shows the unprocessed raw depth data from the camera for a single frame. This frame shows significant noise at ranges that are outside the devices range. A mean depth image was created from the captured frames for each scene. This mean image shows the smoothing and filling effects of averaging 50 frames. The standard deviation of these frames per pixel is shown in the fourth row. Most areas of the scene that are within the working range have less than a 3 m,



(a) Error at 50 m Range

(b) Error at 20 m Range

Figure 3.10: Simulated depth error of the EpiToF 2.0 system at (a) 50 m and (b) 20 m at varying ambient light conditions and modulation frequencies. As the modulation frequency increases, the error levels decrease significantly and even approach LIDAR levels of noise (∼5 cm) under bright sunlight at ranges under 20 m with a 100 MHz modulation frequency.

or 6%, depth error as indicated by the standard deviation. Parts of the scene outside the working range have depth errors up to and sometimes greater than 10 m. A filtered version of the mean depth image is shown in the last row. Here, points outside the working range of the device were filtered by removing the ones that had too low of an amplitude or signal to noise ratio.

The calculated depth error of 3 m is approximately the same as the 3.5 m simulated depth error of our system in these conditions (500 $\mathrm{W\,m^{-2}}$) as shown in Figure 3.10a. To put this amount of error into perspective, 3 m error is approximately the same amount of error a small baseline (∼5 cm) stereo system like the Intel Realsense™ D435 would theoretically have at the same 50 m range. At 20 m range, the error is significantly less and can even reach LIDAR error levels of a few centimeters at high modulation frequencies of 100 MHz or greater as shown in Figure 3.10b. Although the image sensor used in the EpiToF 2.0 prototype has a maximum modulation frequency of 24 MHz, CW-ToF cameras capable of imaging at these high modulation frequencies are common. Phase unwrapping becomes necessary at these frequencies and is commonly done in consumer depth cameras like the Microsoft Kinect 2 [95, 126].

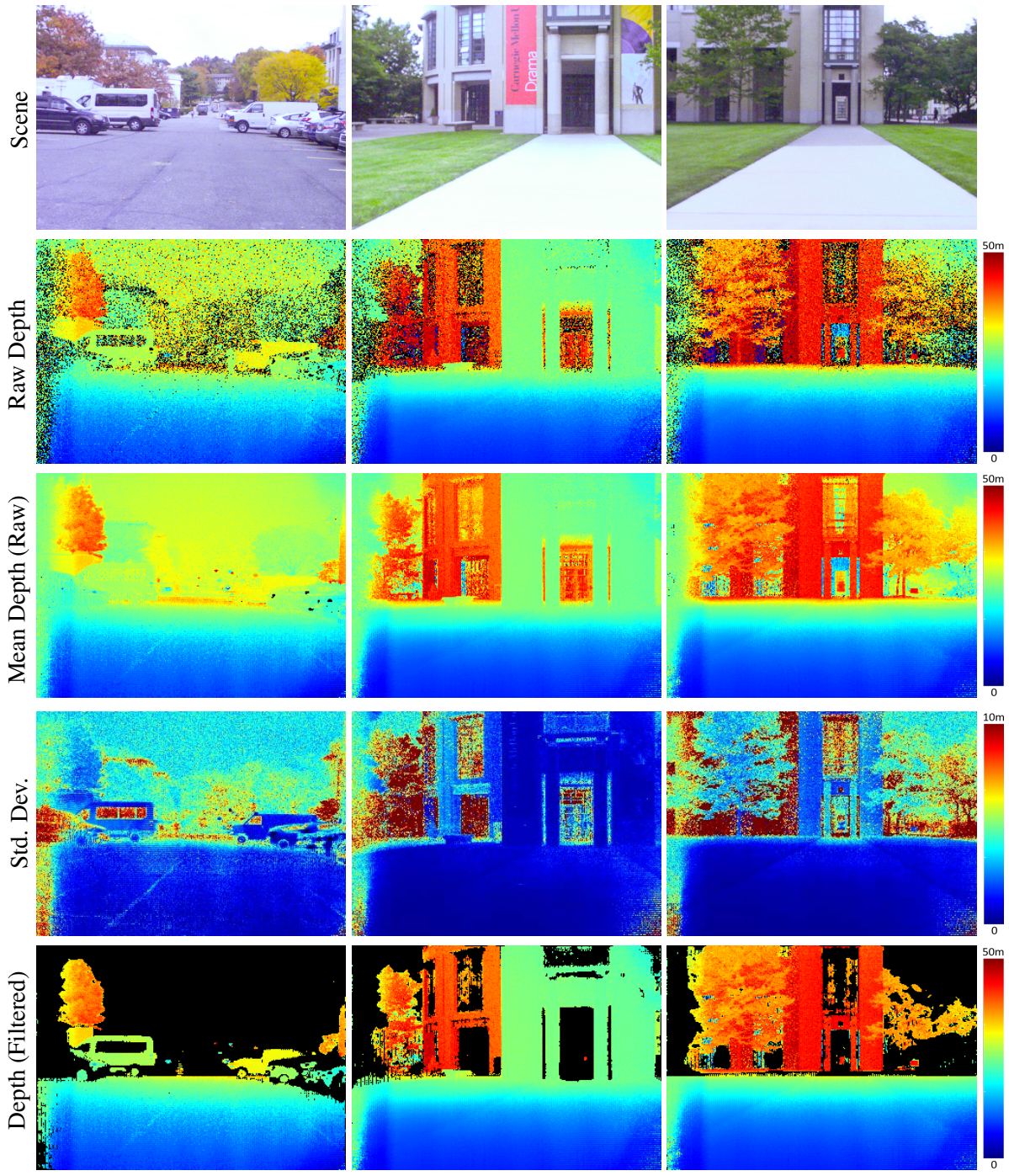


Figure 3.11: Multiple frames of several long-range scenes were captured at 3 MHz modulation frequency (shown in the second row). To reduce noise this set of images was then averaged to find the mean unprocessed depth image (third row). The standard deviation of the measurements is shown in the fourth row. A filtered version of the mean depth map is shown in the bottom row.

Long-range epipolar ToF imaging also works for moving scenes. Figure 3.12 shows results from moving the EpiToF 2.0 through a parking lot scene. In this figure the raw and processed depth maps for a few frames are shown. Unlike the previous experiment with a static scene where a filtered depth map of the averaged image was shown, these images are for each captured frame. Filtering the raw depth information by amplitude and signal-to-noise thresholding as before produces adequate depth images even of single frames. One limitation of the device with movement of the scene is that moving objects will appear distorted because of the rolling shutter like progression of the imaged line as discussed in Chapter 2. Minor effects from this appear in the third column of the image with the two people walking through the scene. This can be minimized by using a faster frame-rate but will always appear as a small blur if the object is moving too fast.

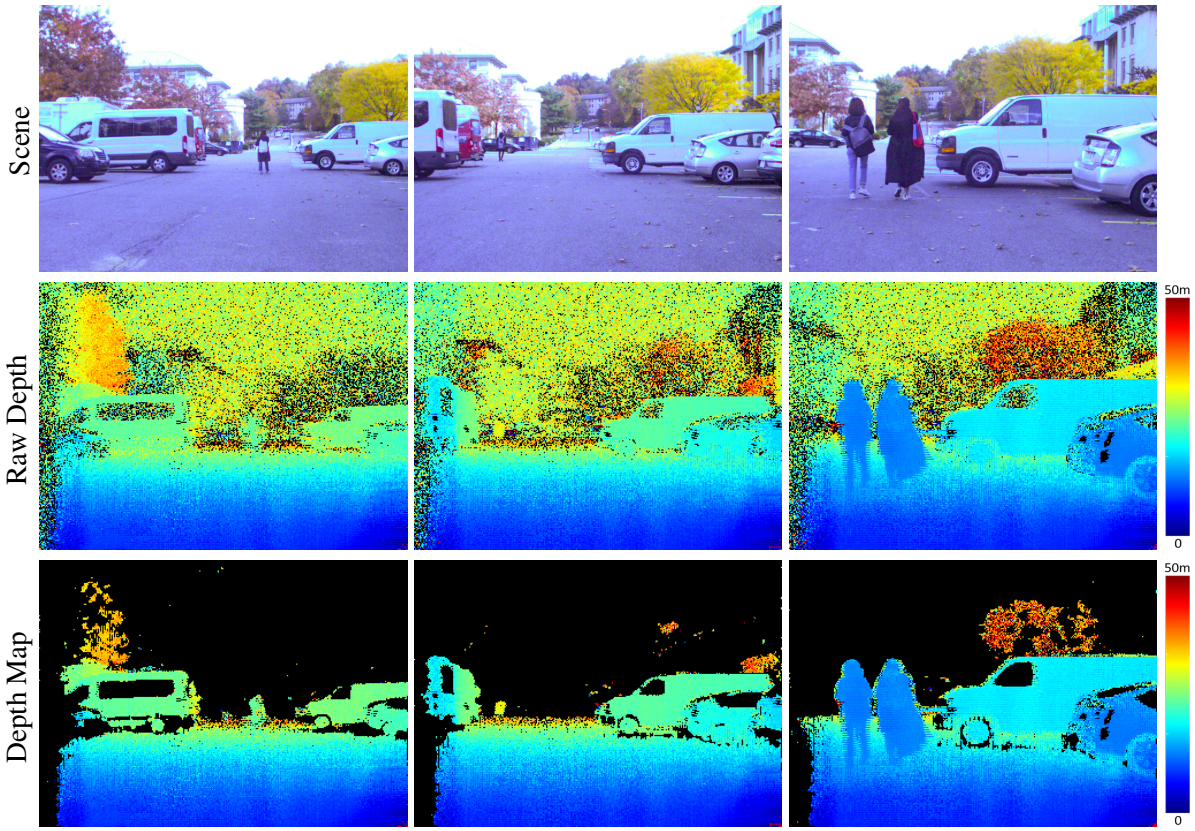


Figure 3.12: The EpiToF 2.0 was moved through a parking lot scene while capturing data. Raw depth data from the device was thresholded by amplitude and signal-to-noise to remove most of the noisy data and produce a depth map for each captured frame. Data was captured with a 3 MHz modulation frequency.

3.5.5 Eye-Safety

Eye-safety requirements place a limit on the power that can be emitted by the light source of an active illumination device. This has implications for accuracy, range, and frame rate. The quantity of interest in determining eye safety for a laser source is the Maximum Permissible Exposure or MPE. MPE is expressed in terms of energy of power per unit area [7] and is a function of light source wavelength and exposure time among other factors. In our light sheet projector, light spreads out from a point so the power density drops as the distance from the source increases. While the EpiToF is imaging a single line it captures four images, each with a given exposure, before moving onto the next row. Due to the modulation of the emitted signal, the duty cycle of each exposure is 50%. These four exposures are each separated by a constant 175 µs readout. At the nominal 100 µs exposure time, the energy density is safe at a distance of at least 40 cm from the source. This minimum safety distance is termed the Nominal Optical Safety Hazard Distance, or NOSHD. For modest increases in range, the exposure time can be increased to 200 µs or 400 µs, in which the NOSHD is increased to 54 cm and 70 cm. As the EpiToF exposure time increases, the on-time of the laser becomes greater than the off time of the laser during readouts and the NOSHD inscreases as shown in Figure 3.13. The discontinuities in the plot are due to the different conditions affecting eye-safety. These conditions are that: (1) a single pulse must be eye safe for the duration of the pulse. (2) a train of pulses that lasts one exposure time must be eye-safe. (3) the sequence of consecutive exposures and readouts performed on a single row before moving to the next must be eye-safe. (4) if more than one consecutive row can illuminate the pupil, the total light received must be eye-safe. (5) in continuous operation (8-hours), the total light received must be eye-safe. The fourth case is the limiting factor for most of the examined exposures. For the detailed eye-safety calculation procedure refer to [2, 7].

3.6 Discussion

The EpiToF 2.0 device is capable of depth imaging at ranges over 50 m in ideal conditions, which is well beyond the limits of traditional CW-ToF cameras. At this long range, detection is no longer the limiting factor but now precision and phase wrapping are the limiting factors for broad use. Although the device can detect objects at this range, its ability to resolve depth accurately and precisely at this range is limited by the modulation frequency and phase-wrapping. Phase unwrapping can be solved in many ways, one of which was shown in Figure 3.9, but the precision or depth error of the measurements is determined by the signal to noise ratio which ultimately limits the maximum range. Although the signal to noise ratio cannot be changed directly, the amount of depth error produced by the noise can be reduced by imaging with higher

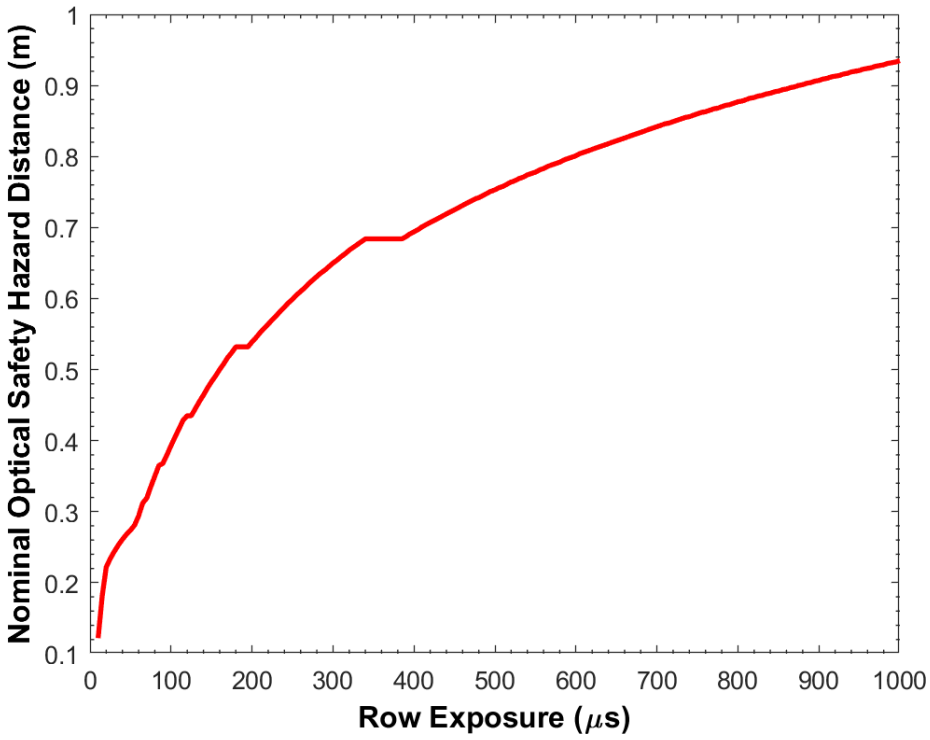


Figure 3.13: The nominal optical safety hazard distance of the EpiToF 2.0 device operating with a 100 µs row exposure time is 40 cm. Increased exposure time results in increased safety hazard distance.

frequencies. The higher the frequency the higher the precision. Unfortunately, as the modulation frequency increases, the modulation contrast and ability for the imager to resolve the modulated signal decreases. The CW-ToF imager we are using has a maximum modulation frequency of 24 MHz, which is already much lower than other devices on the market [53, 95]. In addition, the modulation contrast at this frequency seems to be much lower than that at the lower frequencies as shown by the amount of noise captured at 24 MHz in Figure 3.9. Although the EpiToF 2.0 can range out to 50 m, the level of noise at those ranges unfortunately limits the device's use at those ranges.

The frame rate of the device also limits its applicability. The speed of the current device is limited by the speed of the development board's readout clock and the image sensor's requirement to read out three more rows than required for each imaged row. Before this technology will be broadly applicable to computer vision, these limitations must be surpassed. This should be possible by using other time of flight imagers that do not have these limitations. An ideal CW-ToF imager for epipolar imaging would have a maximum modulation frequency of at least 100 MHz, high modulation-contrast at all frequencies, sufficient resolution, and the ability to rapidly change and read-out the imaged row. CW-ToF imagers are emerging now that have

most of these capabilities, but since most applications of CW-ToF imagers do not require them to rapidly change the region of interest, most of them are not specifically designed to do this. To solve this problem, a custom imager may need to be designed specifically for this type of imaging.

One of the other factors limiting the applicability of these devices is laser power and eye-safety. For a device to be considered eye-safe, it has to be ensured that nothing can get closer than the NOSHD to the device. Currently, this standoff distance of the EpiToF 2.0 is 40 cm at best, which is greater than most reasonable enclosures for depth imaging devices. This standoff distance is determined by the maximum laser power collected per unit area on the eye. Since our laser originates from a point, the eye can refocus it back into a point of light the same size as emitted. To reduce the MPE per unit area and thus NOSHD, an extended laser source such as a VCSEL array or multiple point lasers could be used to illuminate the scene. Since these types of sources cannot be focused into a single point, the allowed laser power can be higher, which would increase the maximum range.

Chapter 4

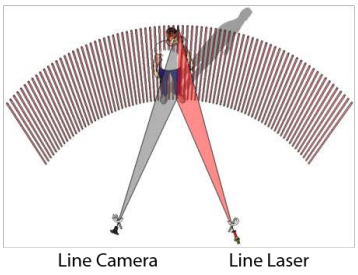
Programmable Triangulation Light Curtains

4.1 Introduction

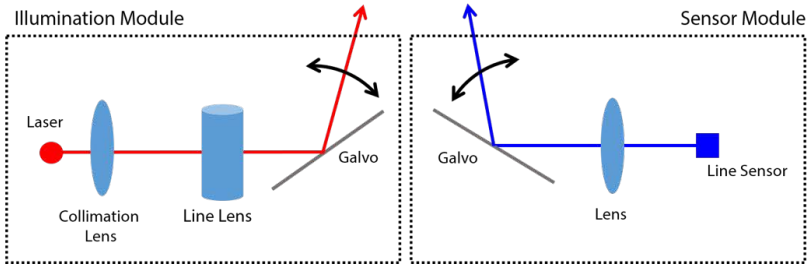
3D sensors are critical in the deployment of autonomous systems such as field robots, unmanned aerial vehicles (UAVs), and self-driving cars. However, there are many tasks which fully capable 3D scanners, such as LIDAR or ToF cameras, are not necessary and can even complicate tasks. For example, consider a self-driving car's collision avoidance system. This system must continually monitor the vehicle's surrounding for obstacles entering its path. In this system, full 3D perception is important for long-term path planning, but it is less useful for time-critical tasks like obstacle detection and avoidance. For these tasks, a proximity sensor requiring less energy and computational power is sufficient and arguably better. The notion of proximity sensing here is generalized by proposing an optical system to detect the presence of objects that intersect a virtual shell around the system. By detecting only the objects that intersect the virtual shell, many tasks pertaining to collision avoidance and situational awareness can be solved with little or no computational cost. This virtual shell is referred to as a light curtain.

A light curtain is a safety device that detects nearby obstacles (e.g., a human) to stop the operation of a machine. Light curtains are ubiquitous. They are used in garage doors and elevators to stop the doors from closing when an object blocks them. They are used on factory floors around dangerous machinery. Light curtains operate on a simple principle that an object is detected when it obstructs the line of sight between a source and a sensor. Even though they are simple, they are very reliable. However, light curtain systems are specifically customized for each machine and configured for each task, which hinders their broad use for vision and robotics.

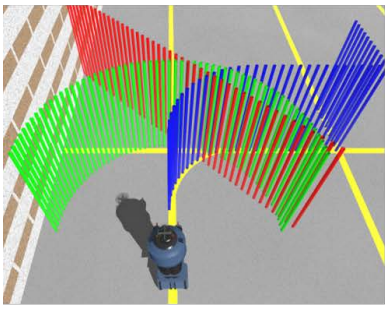
Programmable triangulation light curtains extend this principle to generate light curtains of



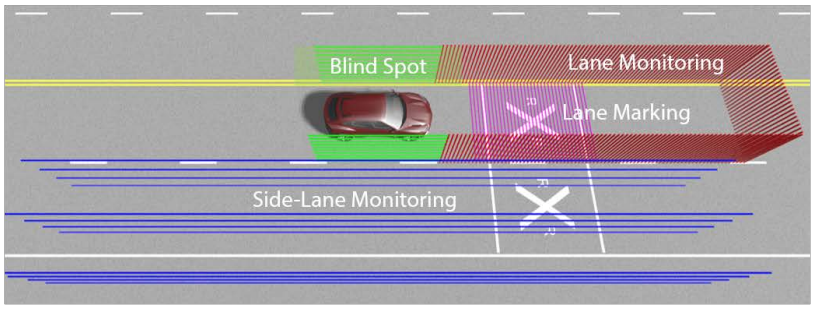
(a) Working principle



(b) Optical schematic (top view)



(c) Light curtains in a robot



(d) Light curtains in a car

Figure 4.1: A programmable light curtain monitors for the entrance of objects into a virtual shell around the device. (a, b) This is implemented by intersecting a light plane emitted from a line laser and a plane imaged by a line scan camera. The two planes are rapidly rotated in relation to each other to generate light curtains of varying shapes as required by the application. Example curtains are shown for use on a robot (c) and a car (d). The device detects objects on the virtual curtains with little computational overhead, making it useful for collision detection and avoidance.

any shape using a line sensor and a line source. Here, an obstacle is detected when it intersects both the plane of illumination generated from the line source and the imaging plane captured by the line sensor (see Figure 4.1a). If there is an object at the intersection of these planes, light will be reflected to the camera, but if there is *not* an object at the intersection of the planes, then no light is reflected. This provides a binary result of whether there is an obstacle in contact with the line or not. The full light curtain is then created by synchronously moving the illumination and imaging planes to sweep this line through the scene. This idea can be interpreted as a generalization of pushbroom stereo [12] to active illumination for determining the presence of an object that intersects an arbitrary ruled surface in 3D. The shape of a light curtain is programmable and can be configured dynamically to suit the demands of the immediate task. For example, light curtains can be used to determine whether a neighboring vehicle is changing lanes, whether a pedestrian is in the crosswalk, or whether there are vehicles in the adjacent lanes. Similarly, a robot might use a light curtain that follows its planned motion trajectory. Figure 4.1c-d shows various light curtains for use in robots and cars.

4.2 Triangulation Light Curtains

Fundamentally, a light curtain is generated by intersecting a ray of light with a ray of imaging. In the absence of ambient and indirect illumination, the sensor measures only light reflected by an object at this point. In general, an imaged point $\mathbf{s}$ on the light curtain surface in the camera's frame at time t is defined by

$$\mathbf{s}(t) = \hat{\mathbf{r}}_c(t)\, u_c(t), \tag{4.1}$$

where $\hat{\mathbf{r}}_c$ and u_c are the unit vector of the camera ray and its magnitude at time t, respectively. The light curtain point $\mathbf{s}$ occurs at the intersection of the camera ray and the light ray, so it is also defined by

$$\mathbf{s}(t) = \mathbf{p}_{op} + \hat{\mathbf{r}}_p(t)\, u_p(t), \tag{4.2}$$

where $\mathbf{p}_{op}$ is the origin of the light ray in the camera frame and $\hat{\mathbf{r}}_p$ and u_p are the unit vector of the laser ray and its magnitude at time t, respectively.

By changing the point of intersection of the camera ray and associated light ray, the surface of the light curtain can be formed. Programmable light curtains can be generated from a set of any of these intersecting rays, and although very flexible, scanning a single intersection point for full coverage of a surface is very time consuming. Here, we consider the case where an entire line is scanned at once which results in much quicker coverage of a curtain.

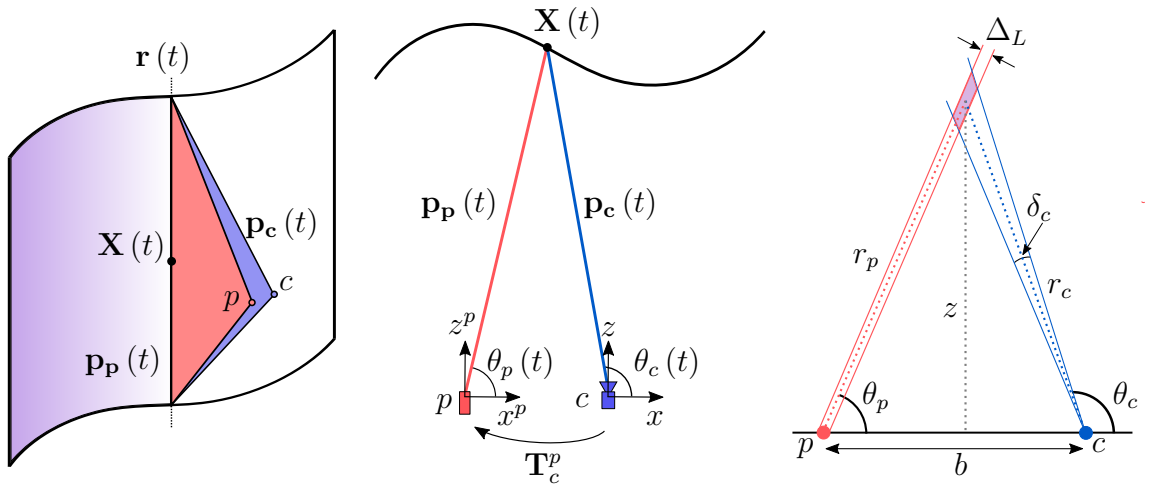
4.2.1 Light Curtains with Planes

A light curtain line is formed by the intersection of an imaging plane and a plane of light. The equation of this intersection line is

$$\mathbf{r}(t) = \mathbf{r_0} + \lambda \mathbf{a}, \tag{4.3}$$

where $\mathbf{r_0}$ is a point that is on both the camera and light planes and $\mathbf{a}$ is the direction vector of the line at time t. This line will be perpendicular to both the camera plane, $\mathbf{p_c}$, and projected light plane, $\mathbf{p_p}$ at time t, and therefore must be parallel to the line defined by the cross product of the unit normal vectors, $\hat{\mathbf{n_c}}$ and $\hat{\mathbf{n_p}}$, of the planes, where both planes are defined with respect to the camera frame. Therefore,

$$\mathbf{a} = \hat{\mathbf{n_c}} \times \hat{\mathbf{n_p}}. \tag{4.4}$$



(a) Light Curtain 3D Geometry (b) 2D Profile (Top-Down View) (c) Curtain Thickness

Figure 4.2: Light Curtain Geometry. (a) The projected light plane, $\mathbf{p_p}(t)$, and camera plane, $\mathbf{p_c}(t)$, intersect at a line, $\mathbf{r}(t)$, on the curtain surface. (b) Top-view of the planar-parallel geometry of a triangulation light curtain. Given that a 3D point, $\mathbf{X}(t)$, is on line $\mathbf{r}(t)$, the required camera plane angle, $\theta_c(t)$, and projector angle, $\theta_p(t)$, to image that point can be found through simple geometry. (c) The finite sizes of camera pixels and the laser sheet produce a light curtain with certain thickness.

The common point can be found by solving (4.5) for $\mathbf{r_0}$, where p_c and p_p are the distances of the camera and light planes from the origin, as defined by the Hessian normal form of the planes.

$$\begin{bmatrix} \hat{\mathbf{n_c}} & \hat{\mathbf{n_p}} \end{bmatrix}^\top \mathbf{r_0} = - \begin{bmatrix} p_c \\ p_p \end{bmatrix} \tag{4.5}$$

Although light curtains can be generated by sweeping the intersection of a plane of light and imaging through any 6-DoF motion profile (as discussed in [119]), here we only consider the case where both the camera and projector each have a single degree of motion around their y-axes.

Given that a 3D point in the camera frame, $\mathbf{X}(t)$, lies on the light curtain surface, the camera plane, $\mathbf{p_c}(t)$, going through this point is found by creating a plane from $\mathbf{X}(t)$ and two points that lie on the camera plane's rotation axis (e.g. $(0,0,0)$ and $(0,1,0)$). The required illumination plane is then found in a similar manner by first projecting $\mathbf{X}(t)$ into the frame of the projector to get

$$\mathbf{X}^p(t) = \mathbf{T}_c^p \mathbf{X}(t), \tag{4.6}$$

where $\mathbf{T}_c^p$ is the transformation matrix that converts points in the camera frame to the projector frame found through calibration. The point $\mathbf{X}^p(t)$ is then used with two points on the projector

rotation axis to find the projected light plane, $\mathbf{p}_{\mathbf{p}}^{p}(t)$, which is in the frame of the projector. This plane in the camera frame is then found by

$$\mathbf{p}_{\mathbf{p}}(t) = \mathbf{T}_{c}^{p\top}\mathbf{p}_{\mathbf{p}}^{p}(t). \tag{4.7}$$

Now, the equation for the entire line containing this point can be found using (4.3-4.5). To generate these curtains, the required angle about the y-axis of the camera, $\theta_c(t)$, and projector, $\theta_p(t)$, can be found using the x and z components of the points by

$$\theta_c(t) = \operatorname{atan2}(x_z(t), x_x(t)) \tag{4.8}$$

$$\theta_p(t) = \operatorname{atan2}(x_z^p(t), x_x^p(t)). \tag{4.9}$$

With these equations, light curtains can be generated from any organized path of points. Figure 4.1(c, d) shows different types of light curtains for use on robots and cars and Figure 4.3 explains each in detail. For each curtain, we show the rendered scene with the light curtain, a 2D cross section of the curtain, and the corresponding rotation angle profiles $\theta_c(t)$ and $\theta_p(t)$, computed using (4.8) and (4.9).

4.2.2 Curtain Thickness

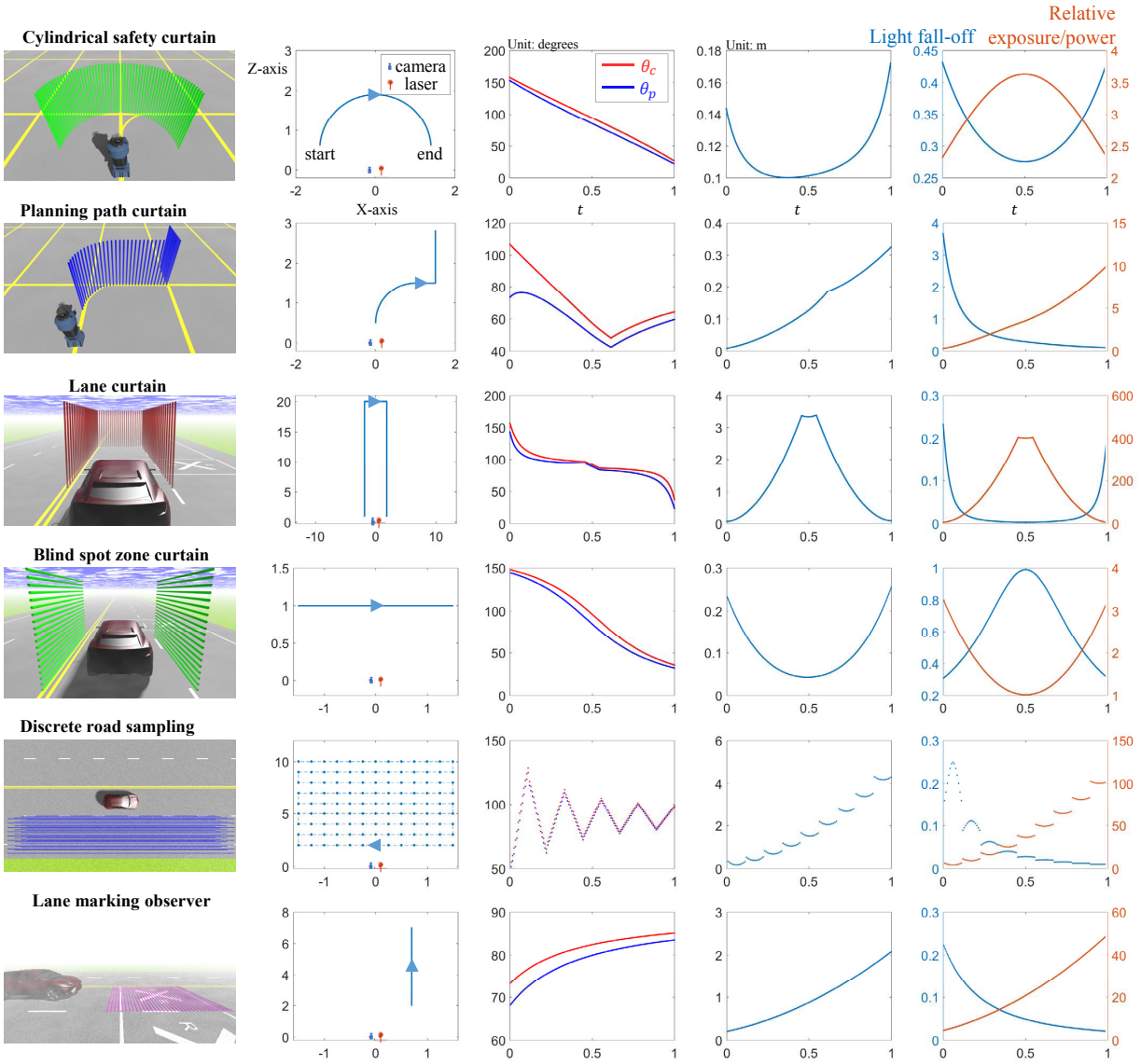
Triangulation light curtains have a finite thickness due to the finite sizes of the sensor pixels and the laser illumination [18]. Suppose that the laser plane has a thickness of Δ_L meters and each pixel has an angular spread of δ_c radians. Given a device with a baseline of length b meters and imaging a point at depth $z(t) = z$, the thickness of the light curtain is derived (see [119]) as an area of a parallelogram shaded in Figure 4.2c, which evaluates to

$$A = \frac{r_c^2 r_p}{z}\frac{\delta_c \Delta_L}{b} \tag{4.10}$$

where r_c and r_p is the distance between the intersected point and the camera and laser, respectively. Since different light curtain geometries can produce curtains of the same area, a more intuitive and meaningful metric for characterizing the thickness is the length of this parallelogram,

$$U = \frac{A}{\Delta_L} = \frac{r_c^2 r_p}{z}\frac{\delta_c}{b}. \tag{4.11}$$

In any given system, changing the laser thickness Δ_L requires changing the optics of the illumination module. Similarly, changing δ_c requires either changing the pixel width or the focal length of the camera. In contrast, varying the baseline provides an easier alternative to changing



(a) Application scenarios (b) 2D cross section (c) $\theta_c(t)$ and $\theta_p(t)$ (d) Uncertainty U (e) Light fall-off

Figure 4.3: Different types of light curtains used by a robot and a car. (a) Envisioned application scenarios visualized using 3D renderings. (b) 2D cross section (all units in meters) of the light curtain and placement of the camera and laser. The arrow on the curtain indicates the scanning direction. (c) Rotation angle profiles of camera and laser to achieve desired light curtain in each scan. (d) Thickness of the light curtain for a camera with 50µm pixel width and focal length of 6mm. (e) Light fall-off and corresponding adaptation of exposure or laser power to compensate for it.

the thickness of the curtain that involves a single translation. This is important since different applications often have specific requirements for curtain thickness. A larger baseline helps in achieving very thin curtains which is important when there is a critical need to avoid false alarm.

Alternatively, thick curtains can be achieved by having a smaller baseline and are important in scenarios where missed detections, especially those arising from the discreteness of the curtain, are to be avoided.

The thicknesses of various light curtains are shown in Figure 4.3(d). The camera's pixel width is set to 50 μm with a lens of focal length $f = 6$ mm, thereby giving a value for $\delta_c = \frac{50\mu\text{m}}{f} = 0.008$ radians. The baseline b was set to 300 mm for the first two rows, 2 m for the third row, and 200 mm for the last three rows.

4.2.3 Combining with Time-of-Flight Sensors

The analysis in (4.11) indicates that $U \approx \frac{z^2\delta_c}{b}$ when $r_c, r_p \approx z$ and that the light curtain is expected to get thicker, quadratically, with depth. Increasing baseline and other parameters of the system can only alleviate this effect in part due to the physical constraints on sensor size, laser spot thickness as well as the baseline. We show that replacing the line intensity sensor with a 1D continuous-wave time-of-flight (CW-ToF) sensor [74] alleviates the quadratic dependence of thickness with depth.

CW-ToF sensors measure phase to obtain depth. A CW-ToF sensor works by illuminating the scene with an amplitude modulated wave, typically a periodic signal with frequency, f_m, and measuring the phase difference between the illumination and the light received at each pixel. The phase difference ϕ and the depth d of the scene point are related as

$$\phi = \text{mod}\left(\frac{f_m d}{c}, 2\pi\right). \tag{4.12}$$

As a consequence, the depth resolution of a ToF sensor $\Delta d = \frac{c\Delta\phi}{f_m}$ (ignoring the phase wrapping) is constant and independent of depth. Further, the depth resolution increases with the frequency of the amplitude wave. However, ToF-based depth recovery has a phase wrapping problem due to the presence of the mod$(\cdot)$ operator; this implies that the depth estimate has an ambiguity problem and this problem gets worse at higher frequencies. In contrast, traditional triangulation-based depth estimation has no ambiguity problem, but at the cost of quadratic depth uncertainty.

We can leverage the complementary strengths of traditional triangulation and CW-ToF to enable light curtains with near-constant thickness over a large range. This is achieved as follows. First, the phase and intensity of the triangulated region are measured by the CW-ToF sensor; examples of this is shown in Figure 4.8(iii, iv). Second, knowing the depth of the curtain, we can calculate the appropriate phase to retain and discard pixels with phase values that are significantly different. An alternative approach to achieving this is to perform phase-based depth gating

using appropriate codes at illumination and sensing [112]. The use of triangulation automatically eliminates the depth ambiguity of phase-based gating *provided* the thickness of the triangulation is smaller than half of the wavelength of the amplitude wave. With this, it is possible to create thinner light curtains over a larger depth range.

4.2.4 Implementation

One implementation of a light curtain device consists of a line scan camera and a line laser, as shown in Figure 4.1 (a, b). A Powell lens fans a laser beam out into a planar sheet of light and the line camera senses light from a single plane. In the general configuration, the two planes intersect at a line in 3D and, in the absence of ambient and indirect illumination, the sensor measures light reflected by any object on the line. By rotating both the camera plane and the laser plane at a high speed, we can sweep the intersecting line to form any ruled surface. We refer to this ruled surface, on which we detect presence of objects, as the light curtain. The resulting device enables programmable imaging of any ruled light curtain profile.

4.3 Hardware Prototype

A hardware prototype of the light curtain device was developed that has a sensing module and illumination module as shown in Figure 4.4. For the sensor, an Alkeria NECTA N2K2-7 line scan intensity sensor with a 6 mm $f/2$ S-mount lens whose diagonal field of view is 45° and has a 7 mm diameter image circle. The line camera has a resolution of 2048×2 with square pixels and pitch of 7 $\mu\mathrm{m}$. Only the center 1000 pixels are used due to the limited image circle of the lens. The line sensor is capable of reading out 95,000 lines/second. To suppress ambient light, a 50 nm optical bandpass filter, centered at 630 nm, is placed directly between the lens and the imager. A low cost 1D galvomirror is used to rotate the camera's viewing angle. The 2D helper camera (shown in the middle) is used for a one time calibration and then solely for visualizing the light curtains in the scene of interest by projecting the curtain to its view.

The illumination system uses a custom made light sheet projector, similar to those used in Chapters 2- 3, comprised of a custom 1D laser module and a galvomirror. The 1D laser module contains a 700 mW, 638 nm laser diode (Thorlabs L638P700M) that is collimated and then emitted through a 45° Powell lens to create a uniform light sheet. The light sheet is then projected onto its own 1D galvomirror that sweeps the light sheet across the scene. The galvomirrors chosen for the modules each use a 11 mm × 7 mm mirror and have a 45° field of view. The galvomirrors need 500 $\mu\mathrm{s}$ for a 0.2° step change in optical angle. A microcontroller (Teensy 3.2)

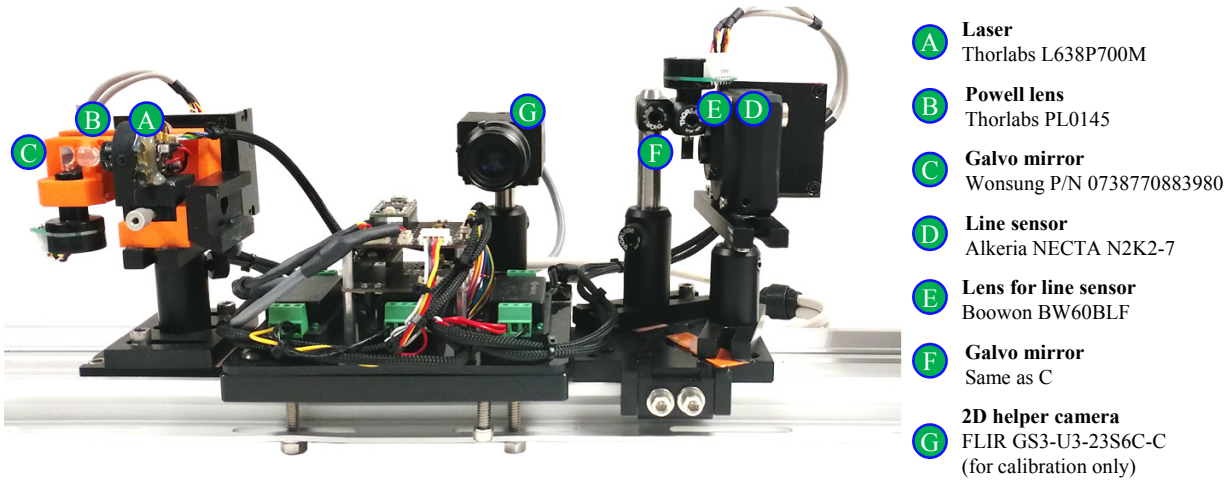


Figure 4.4: Hardware prototype with components marked. The prototype implements the schematic shown in Figure 4.1(b). The illumination module is on the left and the sensing module is on the right. The 2D helper camera marked as (G) is used for a one-time calibration and for visualization of the light curtain.

interfaces to custom-designed galvomirror and laser driver circuitry for synchronization of the camera and laser to the galvomirror. The illumination and imaging modules are then arranged such that the rotation axes of the galvomirrors are parallel with a baseline of 300mm. The light curtain device's field of view is approximately $45° \times 45°$.

Ambient Light Suppression. In darkness, the only light received is that from the intersection of the imaging and illumination planes, but in daylight, the camera also measures the contribution of ambient light to the entire scene. This ambient light hinders detection of objects in the camera image because now the sensor receives light even when the intersection of the sensing plane and illumination plane is clear. This light is suppressed by capturing two images at each position of the light curtain — one image with the laser illumination (both ambient and laser light are collected) and one image without (only ambient light is collected). Subtracting the ambient-only image from the combined image provides just the light from the illumination system. This is critical for improving the performance of the device in ambient light. The synchronization and design of the system provides the ability to capture these images rapidly with 100μs exposures for each image and enables real time performance even in dynamic environments.

Scan Rate. The prototype is implemented with low quality galvomirrors that take approximately 500 µs to stabilize before a line can be captured, which limits the overall scan rate of the device. With each line requiring two 100µs exposures and the stabilization time, the system is limited to capturing approximately 1400 lines per second. For experiments, the system imaged

curtains containing 200 lines, which with a small readout time for each line enabled a curtain refresh rate of 5.6 Hz. This rate can be increased by using higher quality galvomirrors with lower stabilization time or MEMs mirrors with much lower stabilization time.

Calibration. Successful deployment of light curtains requires precise calibration of the light sheet projector to the virtual 2D camera formed by the line sensor and galvomirror. Calibration is performed by identifying the plane in the real world associated with each position of the light sheet projector and the line associated with each position of the 1D camera's galvomirror and the camera pixels. The helper camera and a 2D projector are used to perform calibration, following steps largely adapted from prior work in calibration [73, 127].

4.4 Results

Evaluating Light Curtains. Figure 4.5 shows the results of implementing various light curtain shapes both indoors and outdoors. When nothing contacts the light curtain, the image is dark; when a person or other intruders contact the light curtain, they are immediately detected. The small insets show the actual images captured by the line sensor (after ambient subtraction). The light curtain and the detection are geometrically mapped to the 2D helper camera's view for visualization. Our prototype uses a visible spectrum (red) laser and switching to near infrared can improve performance when the visible albedo is low (e.g. dark clothing).

Light curtains under sunlight. Figure 4.6(a) shows the detections of a white board with curtains of various depths in bright sunlight ($\sim 1000\ \mathrm{W/m^2}$). The ambient light suppression works even at 25 m range, but does not work at 35 m. Under cloudy skies, the range increases to more than 40 m, and indoors the range is approximately 50 m. All of these ranges use the same refresh rate (and exposure time) for the curtains. In Figure 4.6(b), we create a depth map by sweeping the curtain over a dense set of depths and accumulate the resulting detections.

Performance under volumetric scattering. Figure 4.7 shows the benefits of using light curtains in scattering media. A fronto-planar light curtain is set at a fixed distance of about 5 m to image the sign in the scene. Thick fog is then produced to reduce the visibility of the scene. Images from the visual camera show that visibility to the sign is highly reduced, but the light curtain suppresses multi-path light scattering significantly and is capable of seeing much clearer and deeper. The sign text is still readable when the sign is not even visible in the visual image.

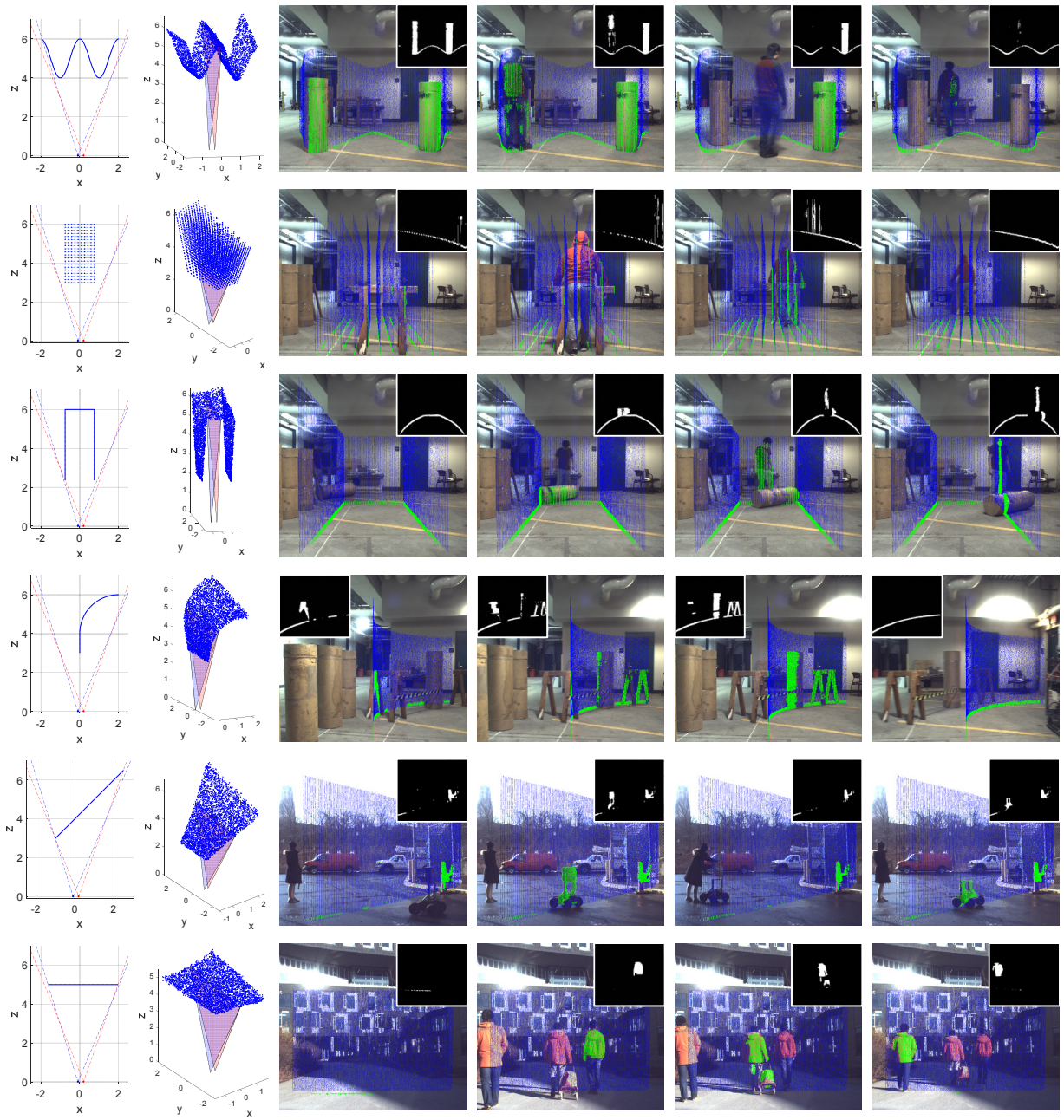
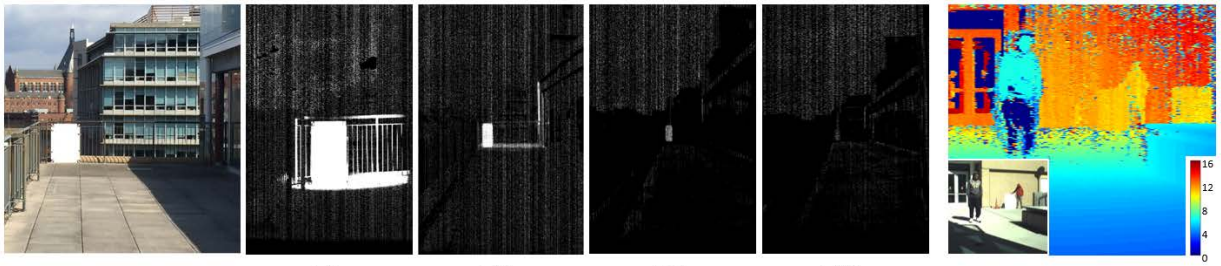


Figure 4.5: Light curtains of many shapes including a sinusoidal, curved path, cuboidal and tilted plane. The curtain in the second row detects objects by sampling a volume with a discrete set of lines. The images shown are from the 2D helper camera's view with the light curtain rendered in blue and detections rendered in green. The curtains are shown both indoors and outdoors in sunlight. The insets are images captured by the line sensor as people/objects intersect the curtains. The curtains have 200 lines with a refresh rate of 5.6 fps.

Reducing curtain thickness using a ToF sensor. We use our device with a line ToF sensor to form a fronto-parallel light curtain at a fixed depth. The results are shown in Figure 4.8. Because



(a) Fronto-parallel planar light curtain at different depths

(b) Depth map by sweeping light curtains

Figure 4.6: Performance under bright sunlight ($\sim$ 1000 $\mathrm{W/m^2}$) for two different scenes. (a) Raw light curtain data is shown at various depths from a white board (no noise reduction or image enhancement is applied). Notice the small speck of the board in the middle visible even at $35\mathrm{m}$. (b) The light curtain is swept over the entire scene to generate a depth map. Depth maps are used only to visualize performance at various depths and not necessarily as a substitute to a depth sensor.

Figure 4.7: Seeing through volumetric scattering media. A fronto-planar light curtain is set at a fixed distance of $5\mathrm{m}$ from the targets in the scene and thick fog is then introduced to the scene. Visibility is highly reduced, but the light curtains suppress multi-path light significantly and can see much clearer.

of triangulation uncertainty, the camera could see a wide depth range as shown in (a.iii) and (b.iii). However, phase data, (a.iv) and (b.iv) helps to decrease the uncertainty as shown in (a.v) and (b.v). Note that in (b.iv), there is phase wrapping which is mitigated using triangulation.

Adapting laser power and exposure. Finally, we showcase the flexibility of our device in combating light fall-off by adapting the exposure and/or the power of the laser associated with each line in the curtain. We show this using depth maps sensed by sweeping fronto-parallel curtains with various depth settings. For each pixel we assign the depth value of the planar curtain at which its intensity value is the highest. We use an intensity map to save this highest intensity

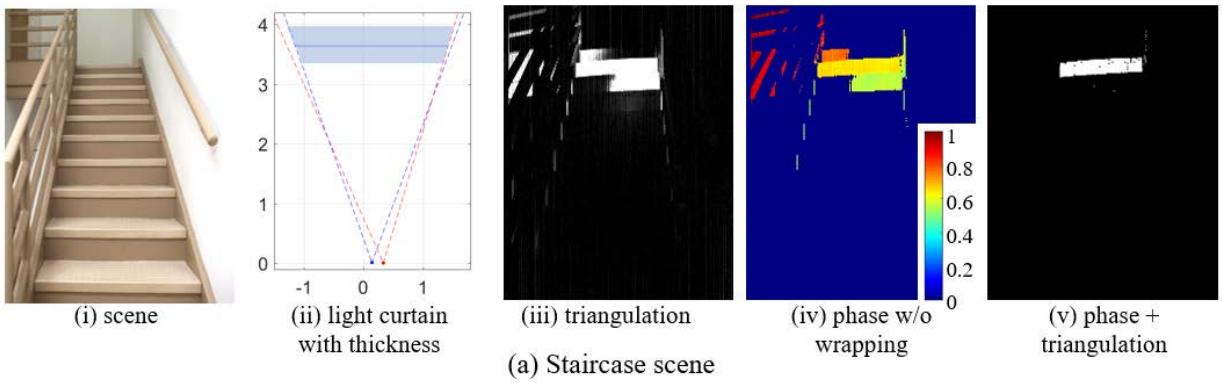
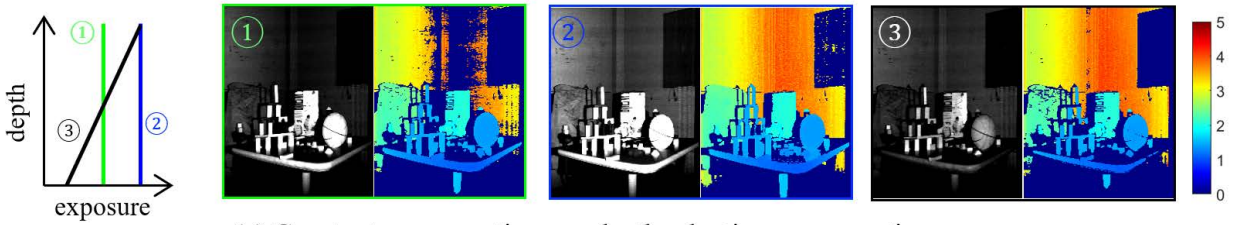
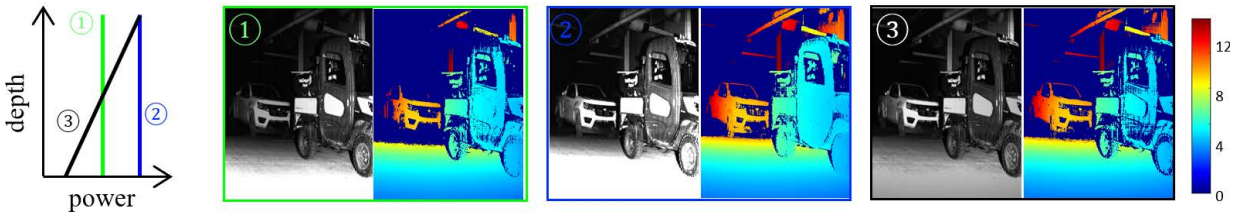


Figure 4.8: CW-ToF Light Curtains. For both scenes, we show that fusing (iii) triangulation gating with (iv) phase information leads to (v) thinner light curtains without any phase wrapping ambiguity.

value. In Figure 4.9(a), we sweep 120 depth planes in an indoor scene. We performed three strategies: two constant exposures per intersecting line and one that is depth-adaptive such that exposure is linear in depth. We show an intensity map and depth map for each strategy. Notice the saturation and darkness in the intensity maps with the constant exposure strategies and uniform brightness with the adaptive strategy. The performance of the depth-adaptive exposure is similar to that of a constant exposure mode whose total exposure time is twice as much. Figure 4.9(b) shows a result from an outdoor scene with curtains at 40 depths, but here the power is adapted linearly with depth. As before, a depth-adaptive budgeting of laser power produces depth maps that are similar to those of a constant power mode with $2\times$ the total power. Strictly speaking, depth-adaptive budgeting should be quadratic though we use a linear approximation for ease of comparison.



(a) Constant exposure time vs. depth-adaptive exposure time



(b) Constant power vs. depth-adaptive power

Figure 4.9: We use depth adaptive budgeting of (a) exposure and (b) power to construct high-quality depth maps by sweeping a fronto-parallel curtain. In each case, we show the results of three strategies: (1) constant exposure/power at a low value, (2) constant exposure/power at a high value, and (3) depth-adaptive allocation of exposure/power such that the average matches the value used in (1). We observe that (3) achieves the same quality of depth map as (2), but using the same time/power budget as in (1).

4.5 Discussion

Benefits of Programmable Light Curtains Besides being able to dynamically change the shape of the light curtain, another benefit is the ability to dynamically adapt illumination power, exposure time, and light curtain refresh rate to compensate for the inverse square light fall-off with distance. This enables the allocation of higher power and longer exposures to lines of the curtain that are further away and is a significant advantage over traditional depth sensors which typically expend max energy in all directions. Programmable light curtains also have benefits in dynamic range and computational efficiency.

Two of the greatest benefits are the performance of these light curtains in scattering media and ambient light. The optical design of the light curtain shares similarities with confocal imaging in that only a small region is selectively illuminated and sensed [81]. This has inherent advantages when imaging in scattering media, such as smoke and murky water, because many multi-bounce light paths are optically avoided thereby providing images with increased contrast. The key advantage in ambient light is that the illumination and imaging are concentrated to a thin region which together with power and exposure adaptability, enables significantly better performance under strong ambient illumination at maximum working ranges of $\backsim$20-30 m. In the absence of ambient light indoors or under cloudy skies this range increases to 40-50 m.

Another benefit is that it provides a large dynamic range. At any time instant, the sensor only captures a single line of the light curtain that often has small depth variations and consequently, little variation in intensity fall-off. Thus, the dynamic range of the measured brightness is often low. Hence, even a one-bit sensor with a programmable threshold would be ample for the envisioned tasks.

Light curtains are not just limited to intensity and CW-ToF sensors. They can also use exotic sensors like InGaAs [46], SPAD [20], or neuromorphic sensors (DVS) [76]. The fact that we scan lines from 1D imagers enables the creation of 2D images with these exotic and expensive sensors without purchasing a 2D array.

Limitations of Programmable Light Curtains A fundamental limitation of light curtains is that they require the laser and sensor to have line-of-sight to the intersection of the light curtain. When it does not, the intersection of the sensing plane and illumination plane is inside an object and cannot be seen by the camera. This can be partially resolved by sampling multiple lines in the scene to maybe carve out the volume or at least detect the object at another location.

These devices could also be limited by interference in the case of several sensors operating a time. This could be resolved by synchronization, adding a second camera, or using a time of flight sensor operating at different frequencies. Objects that move at high speeds could also possibly avoid detection by crossing the light curtain in a vacant region between two successive scans. This is unlikely though for large targets as it would need to be highly maneuverable given the high scan rate. This is likely for small targets however as the scan rate of the device is not fast enough to detect high-speed small objects.

In the realm of 3D sensing, the fact that light curtains only image along a single surface and not the entire volume can be thought of as a limitation of the device in some applications. However, having the ability to sample only around objects of interest can be a great advantage. For example, many applications with UAVs and mobile robots would prefer to have high-resolution depth sensing just around a few objects and then lower resolution in other areas. With conventional depth cameras and most LIDAR systems this is not possible. Usually, uniform high-resolution data is collected and then down-sampled to the desired resolution. In this regard, programmable light curtains can be thought of more generally as on-demand depth sensors.

Future Work. These results show that programmable light curtains have promise as a sensor for autonomous systems with applications in path planning and sensing in bright light and scattering media. Further areas of research for these systems include how to use light curtains for obstacle avoidance with robots, how these devices could be used to enable robots to navi-

gate in challenging conditions, and how they may be used in underwater environments to enable depth imaging for AUVs. With improved hardware that enables higher frame rates they could be used for agile and adaptive depth sensing that image only what is required rather than the entire volume.

Chapter 5

Agile Depth Sensing using Triangulation Light Curtains

5.1 Introduction

In this chapter, we build upon the idea of triangulation light curtains and develop a general framework for agile depth sensing for vision and robotics. Since triangulation light curtains intersect the scene along 3D lines, instead of capturing 3D data of the scene in the entire volume of interest (using say, a LIDAR [106] or Kinect [126]), this framework allows us to flexibly sample the depths along 3D lines in a scene over time. The depth sampling detects the presence or absence of obstacles (or scene points) at these locations in real-time without any additional computation. The depth sampling could be sparse, non-uniform (including random), or dense only at specified 3D surfaces. The sparse sampling can be used to adaptively increase the spatial density of the depths only in the regions of interest as specified by an application. Or objects in the scene can be discovered quickly by initial random sampling followed by adaptive sampling of depths. The depth sampling can be rapidly varied over time depending on the task at hand.

Our agile depth sensing framework has several advantages over traditional depth sensing that uses a fixed acquisition strategy independent of the scene. First, we show that it is possible to capture small, thin, and fast moving objects that is typically difficult for low frequency, uniform angular resolution LIDAR, or low spatial resolution RGB-D sensors. Example objects include thin wires or meshes, or balls thrown at high speed. We also demonstrate fast and dense 3D capture of objects (e.g. pedestrians) far away from the sensor when they are barely visible in a LIDAR point cloud, allowing for better detection, recognition and tracking of such objects. Our framework also allows a robot to explore a region of interest based on initial sparse depth estimates. By continuously and sparsely sampling the scene as the robot moves, it is possible to

simultaneously detect obstacles and map the 3D scene.

To achieve these results, we present a novel design for triangulation light curtains that uses a 2D camera and a laser line. The rolling shutter of the 2D camera and the rotating laser line triangulate at a set of 3D lines in the scene forming a light curtain. By controlling the pixel clock and steering the source mirror, it is possible to generate arbitrary ruled surfaces as in Chapter 4. However, because we use a rapid 2D rolling shutter sensor, we achieve significant advantages over the line sensor design: (a) our light curtains have a refresh rate of 60 fps (more than a 10x speedup) allowing us to change the curtains rapidly and adaptively for the first time, (b) our system is more light efficient and achieves similar range with less collected light because the optics in front of a 2D sensor are not limited by the size of a steering mirror, and (c) the system has fewer moving parts and is more reliable. Our system of triangulation light curtains and depth sampling works outdoors at ranges up to 20-30 m and indoors at up to 50 m (something that the Kinect [126] cannot do). For this, we propose a new method to suppress strong ambient illumination. Due to sampling only a single line at a time our device even works in smoke by blocking most scattered light.

5.2 Related Work

Safety light curtains are used for ensuring safe operation of automated machinery near humans but since they typically only protect a planar region, 2D scanning LIDAR units are used for more complex settings. Safety laser scanners detect objects radially out to a maximum range from the device and can be configured to trigger a safety event if something enters a programmed 2D safety zone [108].

The ability of triangulation light curtains to image light from a single depth in a scene is a geometric form of depth gating. Temporal depth gating [9, 40] also images light from a specific depth but does so by imaging a pulsed laser with a synchronized high-speed gated shutter camera. By emitting a short laser pulse and then briefly opening the shutter once the time has elapsed for the light to travel to the depth of interest and back (usually pico- to nano-seconds), the camera will only receive light from this depth of interest. Another depth selective imaging method uses time-of-flight cameras and on-chip modulation to reject light outside a specified depth range [112]. Push-broom stereo selectively images depths by only processing pixel disparity pairs that have matching features [12], and enables high-speed obstacle detection and avoidance on UAVs.

Existing work in robust depth scanning has shown that imaging points [81] and lines of light [42, 88] perform much better in the presence of global and ambient light than full frame methods. Synchronized imaging of epipolar-aligned planes of light with planes of imaging enable video

frame rate depth imaging of scenes with high levels of robustness to ambient light and scattering [2, 93].

To reduce the amount of data processed from 3D LIDAR scanners and depth cameras, advanced pointcloud filtering methods have been developed that intelligently and adaptively filter entire pointclouds and process only virtual regions of interest around objects [24]. Recent advances in beam steering technology and MEMS mirrors have enabled the adaptive sampling of scenes with LIDAR point scanners by "zooming in" on certain regions of interest while sampling the rest of the scene at a lower resolution [62, 113]. This type of adaptive sampling can reduce pointcloud processing and enable more intelligent sampling of the scene so that fewer points are sensed and immediately filtered out.

5.3 Light Curtains with 2D Cameras

Chapter 4 showed that light curtains could be imaged by steering a line imager with a galvomirror, but the nature of that design limited the frame rate to 5.6 fps and used a small lens to fit the line sensor's field of view onto the rotating galvomirror, which reduced its light efficiency and range. Both of these issues can be improved by using a 2D camera and larger aperture lens to image the light curtains.

Instead of the continuous scanning of a single line sensor plane over time, a 2D camera has a discrete set of imaging planes that are defined by the optics and pixel array. To design a light curtain for a 2D camera, the intersection of the pixels on these planes and the desired light curtain surface must be found. In other words, only points at the intersection of the camera rays and a surface can be imaged. This is in contrast to being able to image any desired point on the light curtain surface as with the dual-galvomirror solution. This has some implications that will be discussed later, such as the inability to sample curtains uniformly.

For simplicity of design, we assume that the pixel rays of a given camera column are coplanar and that the light sheet projector emits a true plane of light and that its rotation axis is parallel to columns of the camera. These assumptions can be generally enforced by using a low-distortion lens, careful design of the light sheet projector optics, and precise alignment of the projector axis to the camera columns. With these assumptions, light curtains can now be designed in two-dimensions by looking at the rays on the xz-plane, as shown in Figure 5.1. This is the same design approach as in Chapter 4, but now with known discrete camera rays. The main difference here is that instead of being able to select any point, $\mathbf{X}(t)$, to image at time t, the points now have to lie on the discrete rays of the camera, which will originate from one of the pixels on column i of the image sensor. So now the point being imaged is written as $\mathbf{X}_i(t)$.

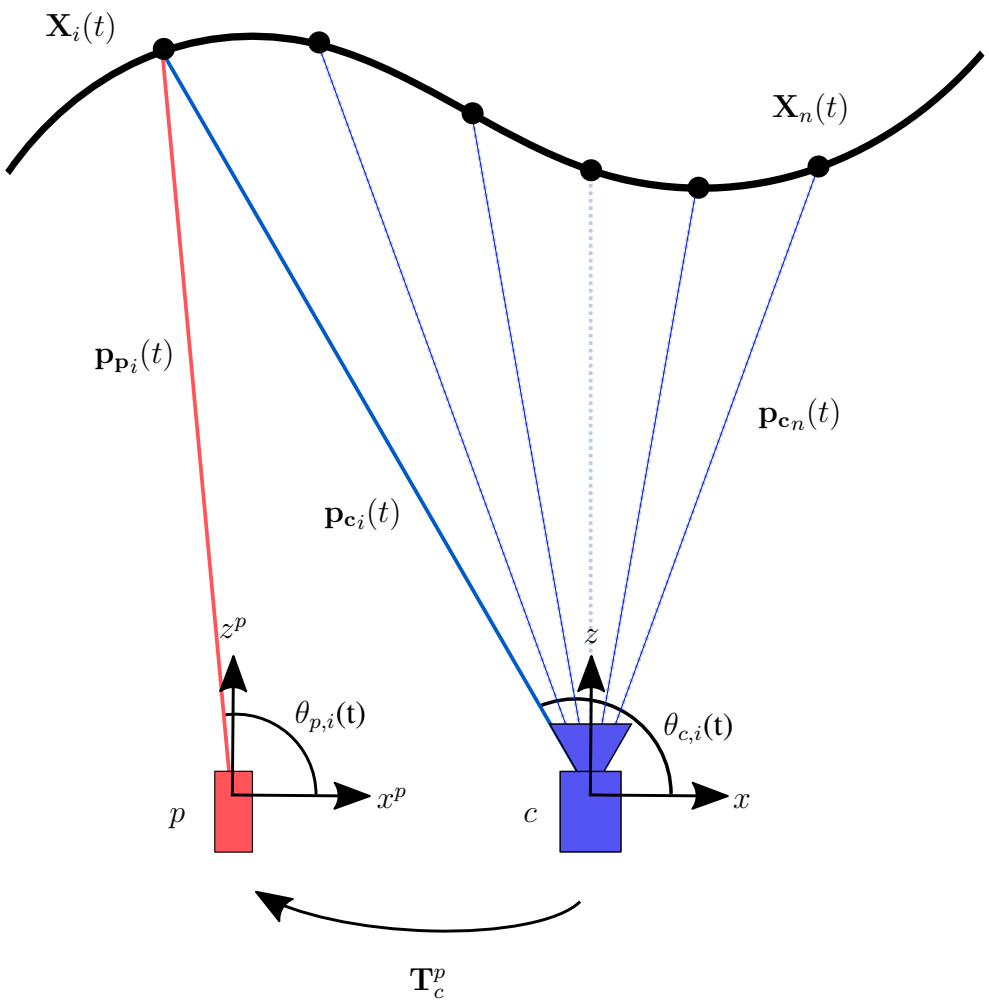


Figure 5.1: Top-view of the planar-parallel geometry of a triangulation light curtain using 2D cameras. Given that a 3D point, $\mathbf{X}_i(t)$, is on the intersection of the camera plane, $\mathbf{p}_{\mathbf{c}_i}(t)$, and the desired curtain surface, the required projected light plane, $\mathbf{p}_{\mathbf{p}_i}(t)$, and angle, $\theta_{p,i}(t)$, to image that point can be found through simple geometry. As the camera plane changes (as indicated by the faint lines), the required light planes are calculated for each new point.

Given that a 3D point in the camera frame, $\mathbf{X}_i(t)$, lies on the light curtain surface and along a discrete camera ray, the camera plane, $\mathbf{p}_{\mathbf{c}_i}(t)$, going through this point is found by creating a plane from $\mathbf{X}_i(t)$ and two points that lie on the camera plane's rotation axis (e.g. $(0, 0, 0)$ and $(0, 1, 0)$). The required laser plane is then found in the same manner by first projecting $\mathbf{X}_i(t)$ into the frame of the projector to get

$$\mathbf{X}_i^p(t) = \mathbf{T}_c^p \mathbf{X}_i(t), \tag{5.1}$$

where $\mathbf{T}_c^p$ is the transformation matrix that converts points in the camera frame to the projector frame found through calibration. The point $\mathbf{X}_i^p(t)$ is then used with two points on the projector rotation axis to find the projected light plane, $\mathbf{p}^p_{\mathbf{p}_i(t)}$, which is in the frame of the projector. This plane in the camera frame is then found by

$$\mathbf{p}_{\mathbf{p}_i}(t) = \mathbf{T}_c^{p\top} \mathbf{p}_{\mathbf{p}_i^p}(t). \tag{5.2}$$

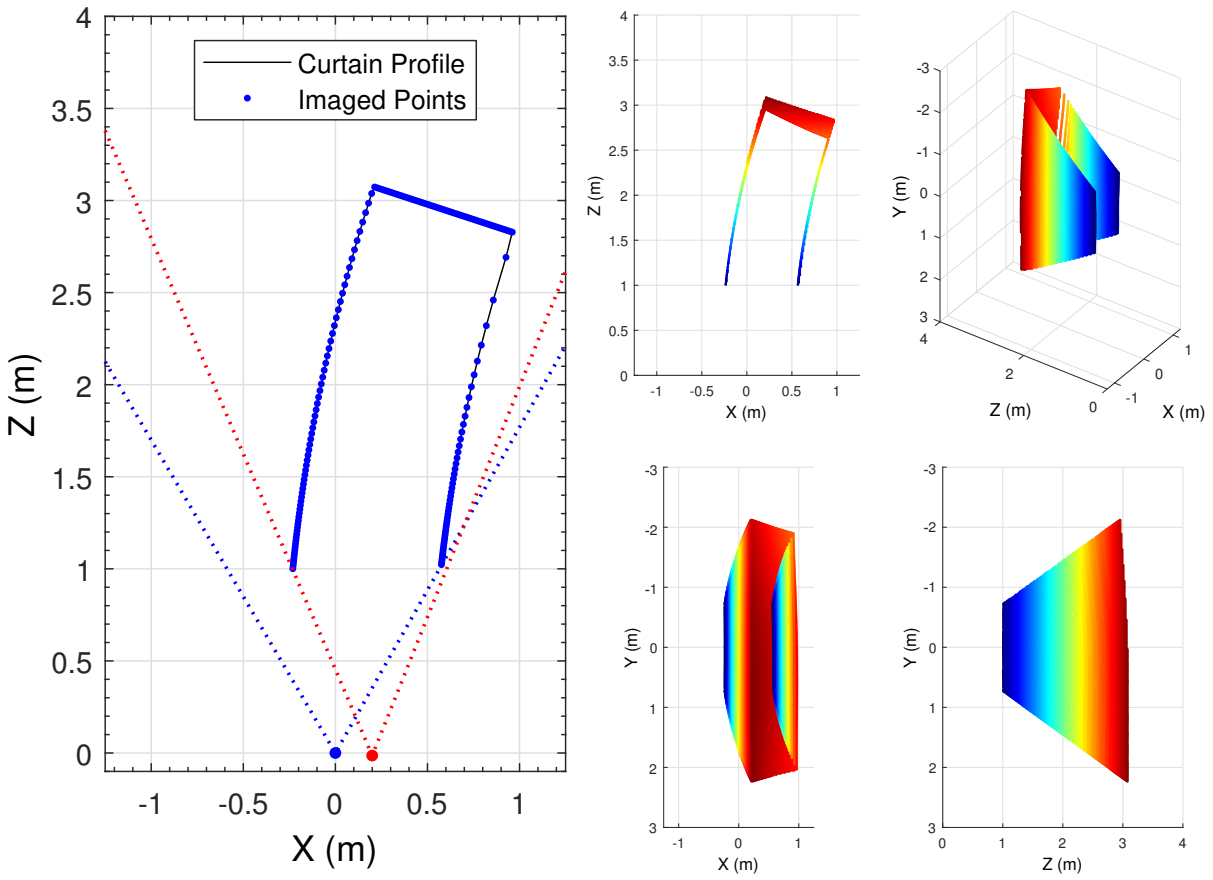
To image a curtain profile, the desired profile is first discretized into a series of m points uniformly distributed along the curtain profile, where m is approximately the number of columns of the camera. Line segments are then formed between adjacent points. The ray representing each column is then checked for intersection with the line segments following the approach in [98] to produce a series of points, $\mathbf{X}_i(t)\ ..\ \mathbf{X}_n(t)$, on the column rays that lie on the light curtain profile, as shown in Figure 5.2. It is possible that some of the camera rays will not intersect the desired curtain profile or that the design points will be outside the field of view of the light sheet projector or camera. Any points that lie outside the field of view of either module are marked invalid and not used to image the light curtain. If the design point is valid, it is transformed into the frame of the light sheet projector using (5.1) and the galvomirror angle needed to create a light sheet that will travel through the design point is calculated using

$$\theta_{p,i}(t) = \text{atan2}\left(x^p_{z,i}(t), x^p_{x,i}(t)\right). \tag{5.3}$$

For an ideal system with true planes of illumination and imaging, the 3D points of the surface are defined by the intersection line of the two planes. However, for a real system, the lens has a small amount of distortion and there may be some small alignment error between the camera and light sheet projector, so the true points of intersection with the light sheet plane for pixels in a given column will not be coplanar and can be found by calculating the intersection of each pixel ray with the light sheet plane, using known calculation methods for ray-plane intersection. This means that the actual 3D surface of the light curtain can vary from the designed profile throughout the camera field of view as shown in Figure 5.2. However, depending on the severity of the distortion and misalignment, the curtain profile is fairly consistent towards the middle of the field-of-view and only changes significantly towards the edges of the field-of-view.

5.4 Rapid Curtain Imaging with Rolling Shutter Cameras

Light curtain imaging requires the imaging plane to change and intersect the light curtain surface where desired. For a 2D camera, this can be done by capturing an entire frame and only using a



(a) 2D Profile **(b)** 3D Profile

Figure 5.2: Light curtain design with 2D cameras. (a) Light curtains are designed with a 2D camera by finding the points at the intersection of the camera rays with the desired curtain profile that is within the field-of-view of the both the camera (blue dotted lines) and the projector (red dotted lines). This produces a non-uniformly spaced set of points on the curtain surface. (b) Due to small optical distortion and alignment error, the actual 3D profile of the curtain is not a perfect extrusion of the designed curtain into 3D, as shown by the difference in curtain shape along each axis. For example, note the varying depth of the front of the curtain in the z-axis.

given column or by only imaging a select region-of-interest on the imager. Although simple to implement, both of these methods are too slow to enable agile light curtain imaging.

A much quicker method of imaging the curtain is to use the rolling shutter of a 2D CMOS imager to move the imaging plane. The rolling shutter rapidly moves the imaging plane at a uniform speed by changing the actively exposed line on the imager. This characteristic enables light curtain imaging at the full frame rate of the rolling shutter camera.

Imaging a curtain is as simple as commanding the light sheet projector to the angle necessary to project the light sheet at the point defined by the intersection of the active line of the 2D camera

and the light curtain profile. For a rolling shutter camera, the active line of pixels is defined by the speed of the rolling shutter and the time since the start of the frame. The speed of the rolling shutter is determined by the readout rate of the pixels, known as the pixel clock.

Assuming no trigger delay, the active line, l_i, of the camera at time t since the camera trigger is found by

$$l_i = \left\lfloor t \frac{p_{clk}}{n_{pix}} \right\rfloor, \tag{5.4}$$

where p_{clk} is the pixel clock of the camera in pixels per second and n_{pix} is the number of pixels in a given line. The equation of the imaging plane (at a given 3D design point on this plane) for this active line is known through calibration. The maximum time that a given line is active and exposed is found by dividing the number of pixels on the line by the pixel clock.

$$t_{exp,max} = \frac{n_{pix}}{p_{clk}} \tag{5.5}$$

By synchronizing the motion of the rolling shutter with the motion of the light sheet, light curtains forming any ruled surface can be imaged.

For each frame of the rolling shutter camera, an image of the captured curtain is produced. While imaging light curtains, the camera captures both laser light and ambient light. If the ambient light is low enough, (e.g., indoor imaging), the image from the camera can be directly thresholded to produce a mask indicating the detected points. However, in many circumstances the captured ambient light is much greater than the captured laser light and the curtain cannot be detected (i.e. outdoors in sunlight). A narrow band-pass filter significantly reduces the captured ambient light, but for maximum performance we want to detect the laser signal at as few bits as possible.

5.4.1 Ambient Subtraction

To increase the performance in the presence of ambient light, we follow a similar approach to Chapter 4, where ambient light was subtracted by capturing both an ambient image and laser+ambient image at each light curtain position. This enabled great ambient performance but required that the same camera plane be imaged twice. This can be done with a 2D camera if the entire image or a selectable ROI is being imaged, but not with a single rolling shutter frame capture. To solve this, we developed a method using adjacent columns of a captured image to perform ambient subtraction.

Our method sacrifices curtain resolution and captures a raw image where even columns are captured with the laser on and odd columns with the laser off to get an ambient only image and

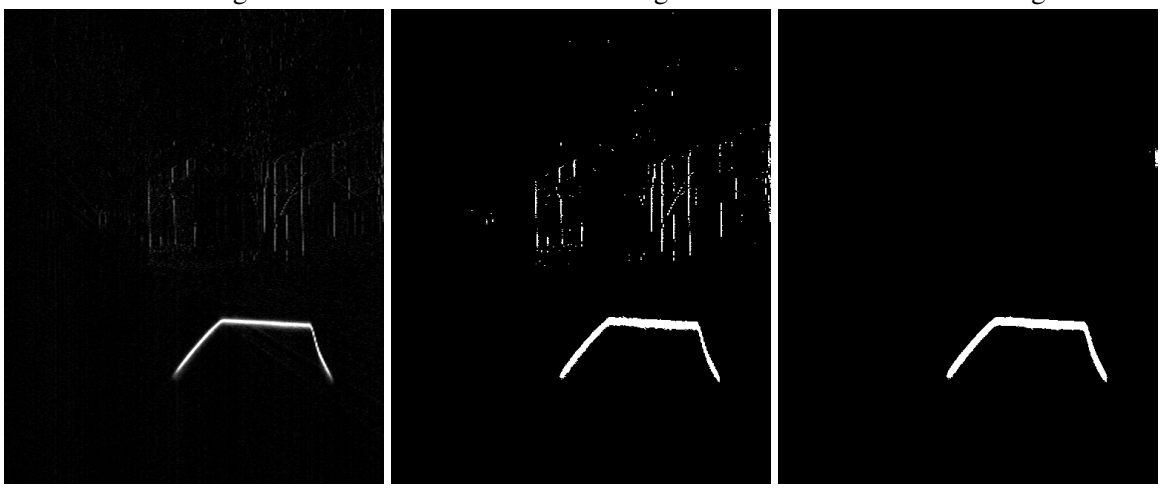
a combined image, as shown in Figure 5.3. We then interpolate the two images to form full resolution images and subtract the ambient image from the combined image to get an image with just the laser light, which we call the curtain image. This image is then thresholded to provide a mask indicating the detected points. In areas of high ambient light, this technique may still produce errors at locations of high intensity gradient as shown by the faint edges in the laser image and vertical thin lines in the thresholded mask image in Figure 5.3. We remove these and any upsampling artifacts by filtering the mask with a thin horizontal erosion/dilation filter and/or by weighting the laser image with the gradient of the ambient image.

Depending on the synchronization precision of the rolling shutter motion with the motion and timing of the light sheet projector, there may be slight bleed through of the laser light onto neighboring columns which shows up in the ambient image. Since this light appears in the ambient image, it is subtracted from the combined image and reduces the measured signal from the laser light and can reduce the performance of the device. With a precise synchronization, the amount of this light can easily be limited to a few bits or removed altogether, so that it does not greatly affect the performance of the system.

5.4.2 Limitations

In exchange for the imaging speed and enhanced light efficiency of using a rolling shutter camera, the uniform sampling of the light curtain profile that dual galvomirrors provided in Chapter 4 is surrendered due to the discrete nature of the camera pixels. This leads to a situation where there may be a gap in the curtain profile that the light curtain device cannot image and occurs when the rays of the camera are similar in direction to the curtain surface. This effect is shown in the top-right portion of the curtain profile in Figure 5.2. One other disadvantage of using a rolling shutter camera is that each plane of the camera can only be imaged once in a given frame. If a camera plane intersects multiple curtain segments (e.g., a zig-zag) one of the segments must be chosen for the imaged curtain and sequential curtains must be imaged to capture all the points along a given ray. A limitation of the galvomirror is that if it is commanded to move very quickly (greater than a 100 Hz step function) then it will lag behind the commanded position and the intersection of the planes will be at a point different than designed, which will causes errors in the locations of the detected points. This constrains which curtains can be imaged and in practice requires random and adaptive curtains to be smoothed with a spline before imaging. This error can be accounted for by using the angle measured from the closed-loop feedback of the mirror to calculate the position of the detected points, rather than trusting open-loop control.

Raw Image | Combined Image | Ambient Image



Laser Image | Thresholded Mask | Filtered Mask

Figure 5.3: Our method of ambient light suppression captures an image where the even columns of the image are captured with the laser on and the odd columns captured with the laser off. Separating these columns and upsampling them forms a full resolution ambient-only image and a combined image. Subtracting the ambient image from the combined image produces a laser-only image, which can then be thresholded to find the light curtain detections. Filtering this mask provides improved robustness to large intensity gradients and upsampling artifacts.

5.5 Depth Sensing with Light Curtains

Sensing depth with light curtains does not provide depth of the entire volume at once like LIDAR or traditional RGB-D sensors, but rather enables the selective depth sampling of the volume. Only lines of points in the volume that are of interest can be imaged, rather than sensing the entire volume and then filtering out what is not needed. This capability can not only be used to sample a scene uniformly, but also in many non-uniform and more efficient ways. Here we

discuss several methods of depth sensing that can be used to sample a volume.

Plane Sweeping A volume can be sampled by simply sweeping a planar light curtain over its extents. This has the advantage of providing a uniform sampling of the volume and is useful if a fully predictable and defined sampling pattern is desired. This type of depth sensing provides the most similar results to LIDAR and RGB-D sensors, but takes longer to fully sample the scene. Since most scenes are not uniformly distributed with points of interest, this type of sampling is not efficient.

Random Sampling A possibly more efficient but non-deterministic method is random sampling of the volume. Random sampling of a volume is done by fitting a curtain to randomly generated points within the specified volume. Ideally, the randomly generated points should be uniformly distributed over the volume to produce a light curtain that covers the extents of the scene. By rapidly changing the random curtains, different parts of the scene are imaged and an estimate of the full volume is produced quicker than a uniform plane sweep. As more and more random curtains are imaged, the scene is iteratively refined to ultimately produce a high-resolution estimate of the scene.

Adaptive Sampling The previous methods sample the scene naively and do not adapt their sampling strategy based on their current knowledge of the scene. Provided an estimate of the current scene, light curtains can be designed based on specific regions or objects of interest that have been identified. One interesting example is scene discovery. When discovering a scene, the light curtain initially knows nothing about its environment, and first uses random or swept planar curtains to generate a sparse map of the scene. This sparse map provides guidance on where to sample next and a curtain profile can be generated to fit the detected points in the scene. To discover more of the scene, this curtain can be dithered back and forth to scan nearby areas at high-resolution and build up the map in these regions. These curtains can be tuned to precisely scan an area of interest at high-resolution, or to scan the area at courser resolution to discover more of the scene. In order to monitor other areas of the scene and dynamically adapt to new objects entering the scene in other locations, the entire volume can continue to be sampled at low-resolution by interleaving random curtains or another type of coverage curtains in with the adaptive curtains. In this way, the scene can be adaptively sampled to generate high-resolution 3D information at areas of high interest and low-resolution information in other areas.

On-Demand Sampling Triangulation light curtains also enable on-demand depth sampling based on information from other sensors. For example, ultrasonic or RADAR sensors that pro-

vide robust but low-resolution proximity detection can be paired with light curtains to perform on-demand high-resolution sampling of a certain region of interest when required.

Another example is providing 3D information about detected objects in a monocular camera. A monocular camera provides information about the entire volume and can be used to guide the light curtain to sample depths around identified objects of interest such as people or cars. One specific application here is people tracking, where a person detection algorithm is used to find and track people in the scene of a monocular camera. By using the light curtain to first probe along the directions of the detected people from the monocular camera, the depth and therefore 3D pose and size of the detected people can be found. Then the light curtain can be used to rapidly sample around these people and adaptively adjust to track their 3D location in the scene. During this, the monocular camera can be providing information on new objects entering into the scene for tracking. This method of on-demand sampling is powerful as it provides high-resolution depth information in areas indicated by a sensor that can see the entire scene, but only in two dimensions. Ideally these methods would work in symbiosis to provide just the right amount of depth information for each given area of the scene.

Each of these sampling methods can be interleaved to create a powerful adaptive depth sampling mechanism as needed by each unique application and scene.

5.6 Hardware Prototype

The hardware prototype is comprised of a light sheet projector and a rolling shutter camera. The light sheet projector contains a custom-designed line laser module using a 1 W 830 nm laser diode (Thorlabs LD830-MA1W) that is then collimated and shaped into a line with a 45° Powell Line Lens (Thorlabs PL0145). This laser line is then projected onto and steered with a galvomirror (Thorlabs GVS001). The line laser module is aligned and mounted to the galvomirror in an aluminum mount that enables the precise collinear alignment of the laser line with the galvomirror's axis of rotation. The mirror has dimensions of 14.5 $\mathrm{mm} \times 8.5\ \mathrm{mm}$ and has a 50° optical scan angle.

The rolling shutter camera (IDS UI-3240CP-NIR-GL) was fitted with a low distortion C-mount lens (Kowa LMVZ4411) with 70°(h) $\times$ 60°(v) field of view. This camera provides 10-bit precision and has a native resolution of 1280 $\times$ 1024 pixels with 5.3 $\mu\mathrm{m}$ square pixels. We operate the camera in 2x binned mode for a resolution of 640 $\times$ 512 to increase the signal of the received light and reduce the noise. We used a low-distortion lens to ensure that the intersection of the light sheet and the camera rays along a given column formed as close to a line as possible. We placed a 12 nm bandpass filter centered at 830 nm between the lens and the image sensor to

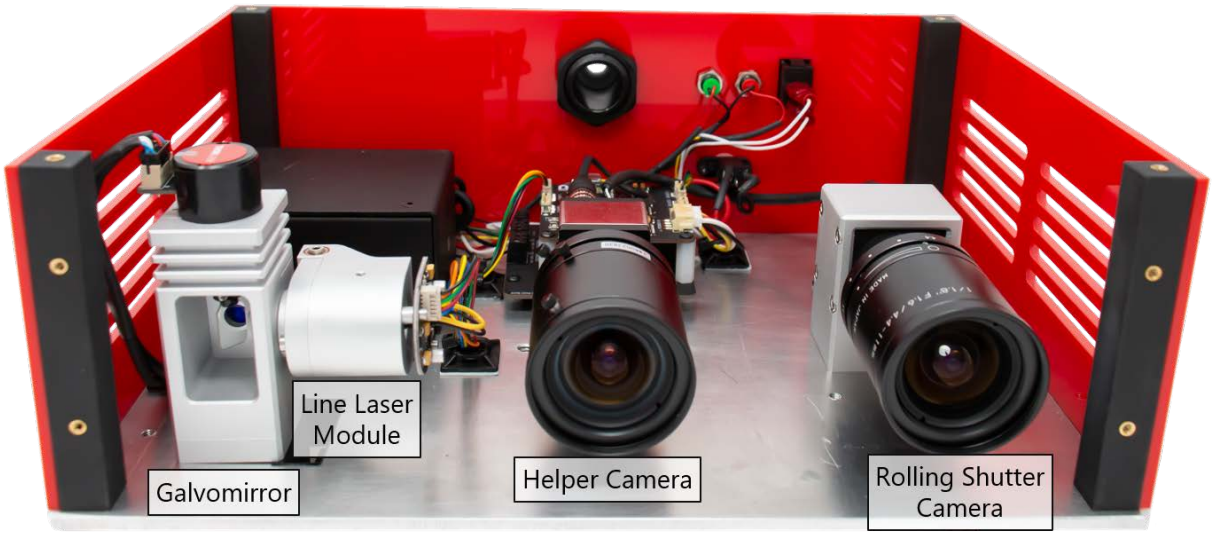


Figure 5.4: The hardware prototype is comprised of a light sheet projector, rolling shutter camera, and helper camera. The light sheet projector uses a custom line laser module to project a laser line onto a galvomirror which then steers the light sheet onto the light curtain surface. The rolling shutter camera is aligned and rotated 90° so that the rolling shutter and rows of the camera are parallel with the galvomirror axis, with a baseline of 200 mm. The helper camera is used for visualization of the light curtains only.

reduce the amount of collected ambient light. The camera is aligned to the galvomirror such that the rows of the camera are parallel with the galvomirror's axis of rotation. The rotated camera was then placed at a fixed baseline of 200 mm from the galvomirror's rotation axis. A microcontroller (Teensy 3.2) is used to synchronize the camera, the laser, and the galvomirror. A color 2D helper camera is used for visualizing the light curtains and detected results in the scene by projecting the light curtain to its view.

5.6.1 Calibration

The performance of light curtains depends on precise calibration of the camera and light sheet projector. First, the camera intrinsics were determined by following the traditional calibration approach of [127]. Next, the extrinsic calibration of the light sheet projector and camera was determined by imaging a set of light planes projected by the light sheet projector onto a planar wall. A checkerboard target of known dimensions was attached to the wall and imaged with the calibrated camera to get the known 3D coordinates of points on each imaged laser line. This was repeated with the same set of planes at several depths to fully define each plane of light. Best fit plane equations were then found for each set of points using weighted least squares where the weights were the normalized intensity values. Then, given the equations for the planes, the

location of the galvomirror axis with respect to the camera was then found by a least squares fit of a line to the intersection of all the planes. We then fit a continuous function to the relationship of the plane angle (with respect to the galvomirror axis) and the commanded light sheet projector angle, which is then used to determine the commanded galvomirror position needed for any specified angle of a given design point.

5.6.2 Capture Process

To image a light curtain, a host computer transmits the required galvomirror positions to the micro-controller which then triggers the camera to start the frame capture. The micro-controller then sequentially commands the galvomirror positions and laser power in lock step with the timed progression of the rolling shutter. This process then repeats for each successive light curtain. Sending the galvomirror positions for every frame enables a different light curtain to be imaged every frame at the full frame rate of the camera.

5.6.3 Working Range

The working range of of our device depends on the maximum exposure time for a given row. At the maximum 60 frames per second capture rate of the camera in our prototype, the maximum active exposure time of a line is approximately 15 $\mu\mathrm{s}$. With this exposure, the rolling shutter light curtain device has a maximum working range of 15 m while imaging a whiteboard in 100 klx ($\approx$ 1000 $\mathrm{W/m^2}$) of ambient light and a working range of 50 m indoors. As shown in Figure 5.5, this is similar to the dual galvomirror device, but with only 15% of the 100 $\mu\mathrm{s}$ exposure time used in that device. When the rolling shutter device is configured for an exposure of 100 $\mu\mathrm{s}$ the device can image over 25 m outdoors in similar conditions, but at this range the detection ability is actually more limited by the resolution of the camera.

5.7 Results

Agile and Dynamic Light Curtains: Our device is capable of imaging 60 different light curtains per second. This speed and flexibility enables agile and dynamic light curtains that can be used to intelligently and adaptively sample the scene. This capability can be applied to many areas including, path planning, high-resolution safety curtains, and depth sensing. Figure 5.6 shows the results of imaging various types of light curtains both indoors and outdoors and are just a sample of the different types of curtains that can be imaged with our device. Images shown

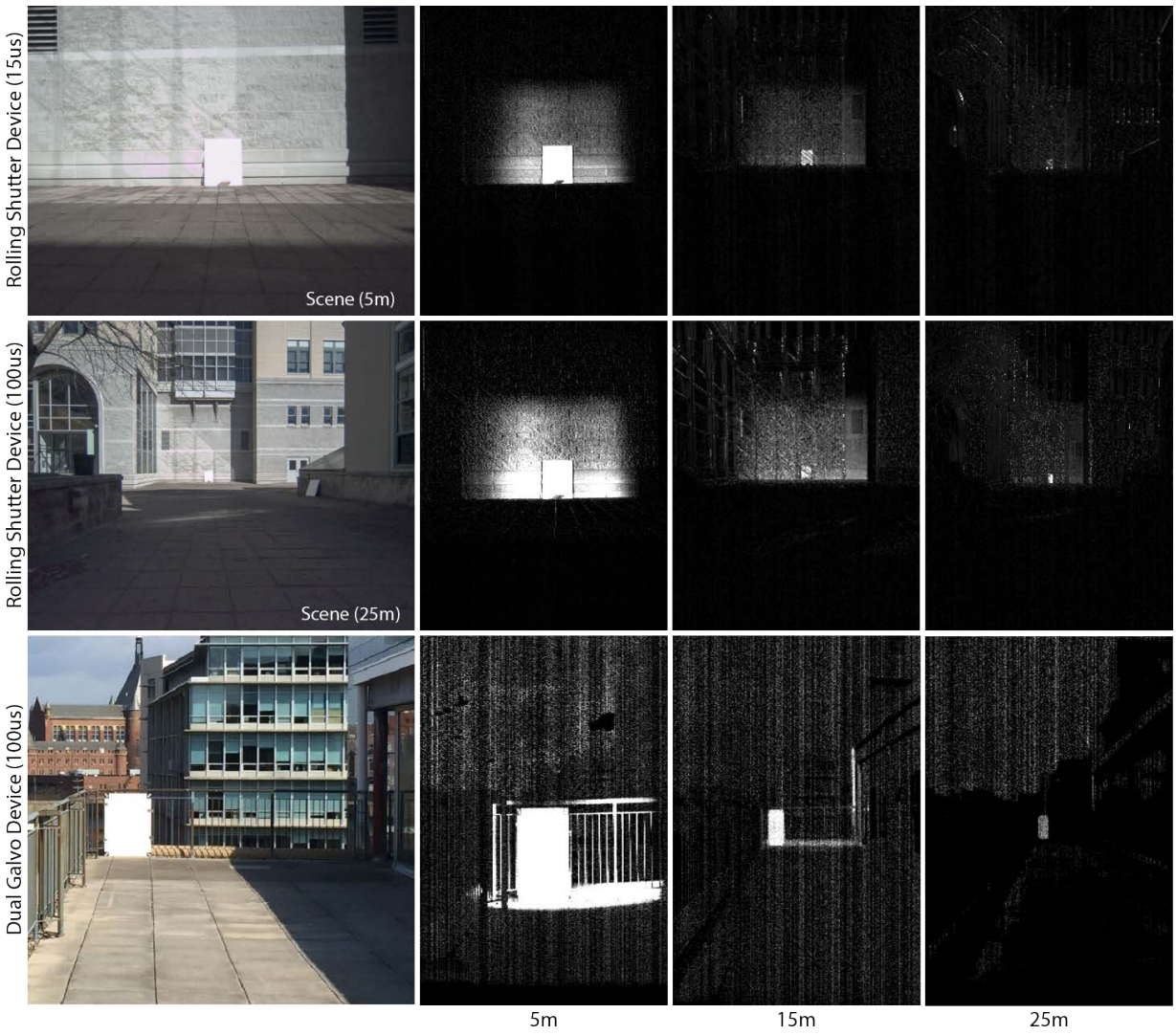
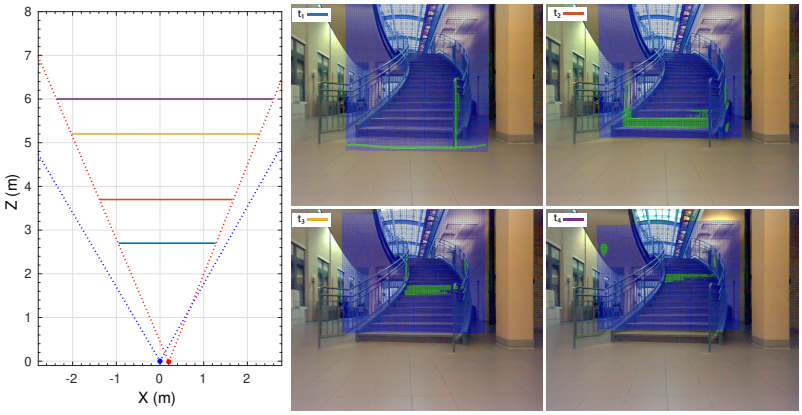
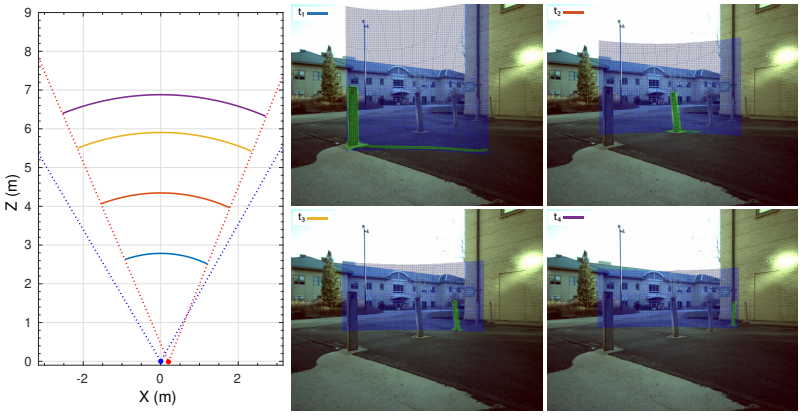


Figure 5.5: The rolling shutter light curtain device has a working range similar to that of the dual galvomirror device, but with better light efficiency. In bright sunlight with an exposure of 15 µs per line (top row), the rolling light curtain device has similar range to that of the dual galvomirror device with a 100 µs exposure (bottom row). When the rolling shutter device is configured with an exposure of 100 µs (middle row) the detection range is longer than the dual galvo device. However, at longer ranges, the resolution of the camera limits the detection ability of small obstacles.

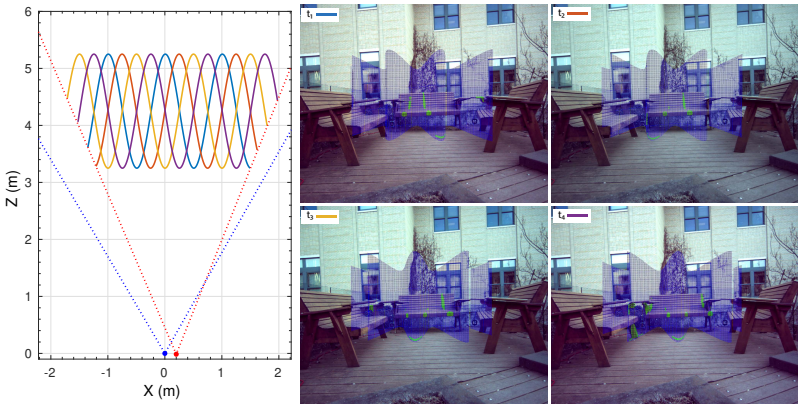
in the figure are the light curtain surface and detections projected into the helper camera's view for visualization on the scene.



(a) Plane Sweeping

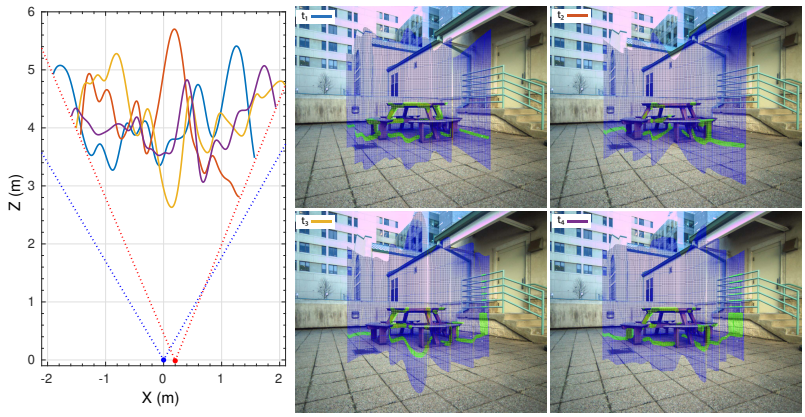


(b) Proximity Search

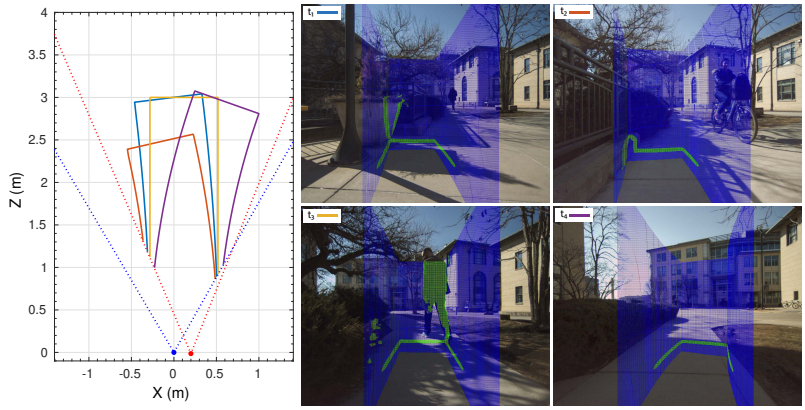


(c) Sine Sweep

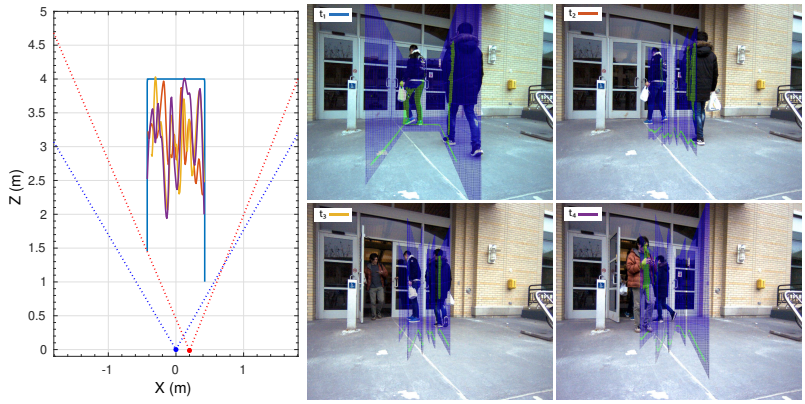
Figure 5.6: Agile and Dynamic Light Curtains. Light curtains captured with our prototype can be imaged and changed up to 60 times per second to quickly scan different curtains through a volume. The images shown here are from the 2D helper cameras view with the light curtain surface rendered in blue and detections rendered in green. The agile nature of these curtains enables many types of curtains to quickly be imaged such as plane sweeping (a), proximity searching (b), and sine sweep (c).



(d) Random Curtains



(e) Path Checking Curtains



(f) Safety-Zone Monitoring

Figure 5.6: Agile and Dynamic Light Curtains (cont). Random curtains randomly sample a defined volume with randomly generated splines (d). Curtains can be generated based on the planned path of a vehicle to ensure nothing enters its path (e). A safety zone (f) can be monitored by rapidly alternating between checking the border of the safety-zone and the area within it.

High-Resolution, High Speed Light Curtains: The rapid capture rate and resolution of our device enables the imaging of small and fast objects as they pass through light curtains, as shown in Figure 5.7a-5.7c. The resolution of light curtains provide increased detail over scanning LIDAR devices and can enable enhanced object recognition and critical detection of small objects (e.g. wires, branches, etc). This is especially noticeable when imaging thin structures or objects at a distance as shown in Figure 5.7d-5.7i.

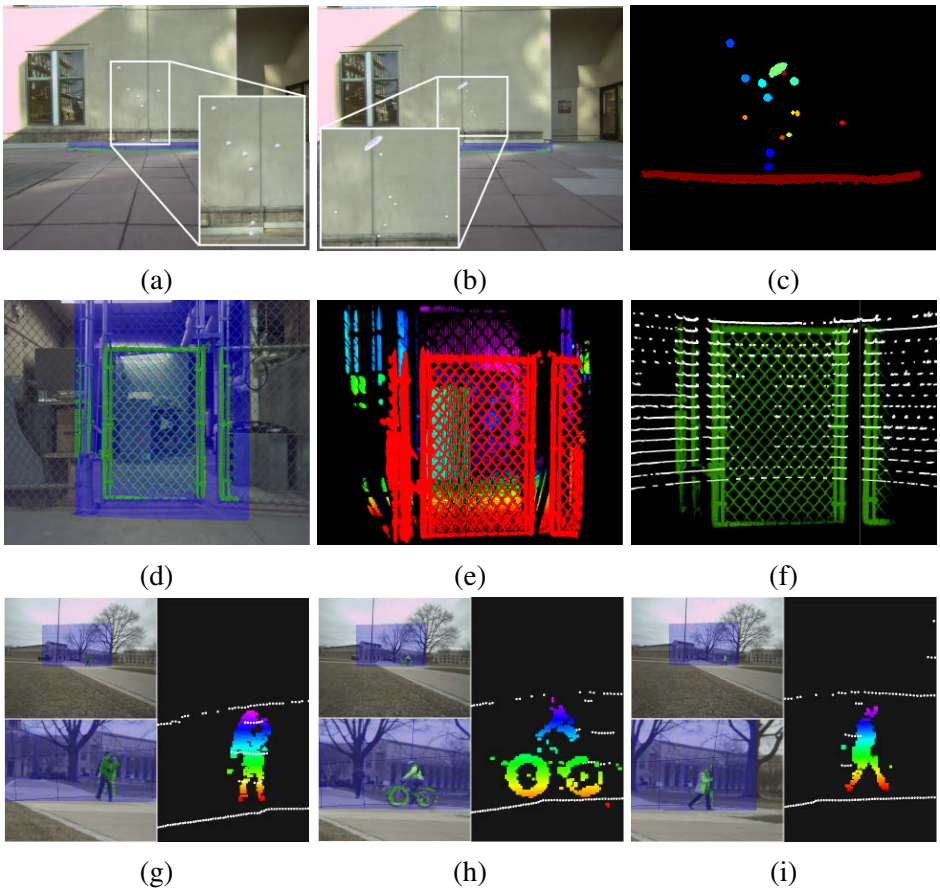


Figure 5.7: Our device can capture the depth of small, fast objects and provides more detail than scanning LIDAR devices. Top Row: Several composite images show the view and detected location of small 70 mm (a) and 30 mm diameter balls (b) and a 30 mm thick, 265 mm diameter Frisbee that were thrown through a planar curtain 5 m away. The detections of the objects are shown in (c). Middle Row: A planar curtain (d) was swept through a volume containing a thin wire fence to create a dense 3D pointcloud (e). The light curtain (green points) in (f) reconstructs the fence mesh at much higher resolution than a 16-beam Velodyne VLP-16 scanning LIDAR (white points) at a distance of 1.5 m away. Bottom Row: By imaging a planar curtain, the range and resolution of the light curtain enable it to create high resolution height maps of objects at ranges of 15 m outdoors (g-i). At this range, a static VLP-16 scanning LIDAR only senses the objects as a few points, shown in white.

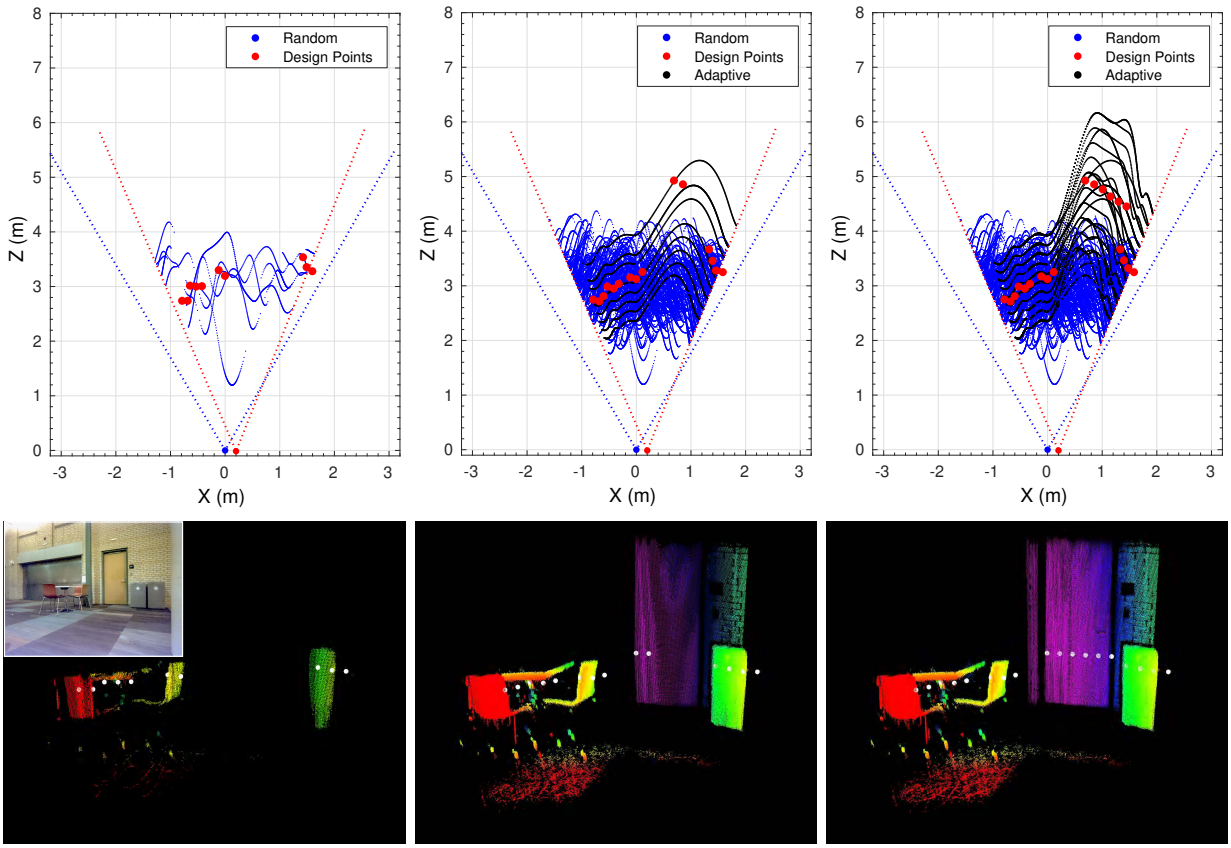


Figure 5.8: Adaptive depth imaging of a table and chairs scene is performed by first randomly sampling a volume of interest to find objects. Then, the detected points at the front of these objects (indicated by red dots in plots and white dots in bottom row) are used to design a set of light curtains which is then imaged to refine and discover more of the scene. As time continues, the curtains eventually discover and image the back wall of the scene.

Adaptive Depth Imaging: The ability to specify depths of interest at high-rates enables intelligent depth imaging of an environment based on the current knowledge of the scene. For example, when a device first enters into the scene, it has no knowledge of it's environment, but by quickly scanning a volume of interest with light curtains it can generate a coarse estimate of the locations of objects in the scene, and a new curtain can be designed to image around these points of interest. This process is then rapidly repeated to form an accurate map of the objects in the scene. Figure 5.8 shows an example where random curtains were used to initialize the curtain by randomly sampling a volume of interest within 3 m of the light curtain. These curtains detected several objects, which the device then used as design points to fit a new curtain to the front surfaces of the objects. In less than a few milliseconds, a set of 10 curtains were designed that were scaled versions of this spline to cover the area directly in front of and behind the detected scene points. These new curtains were then imaged and used to refine and discover more

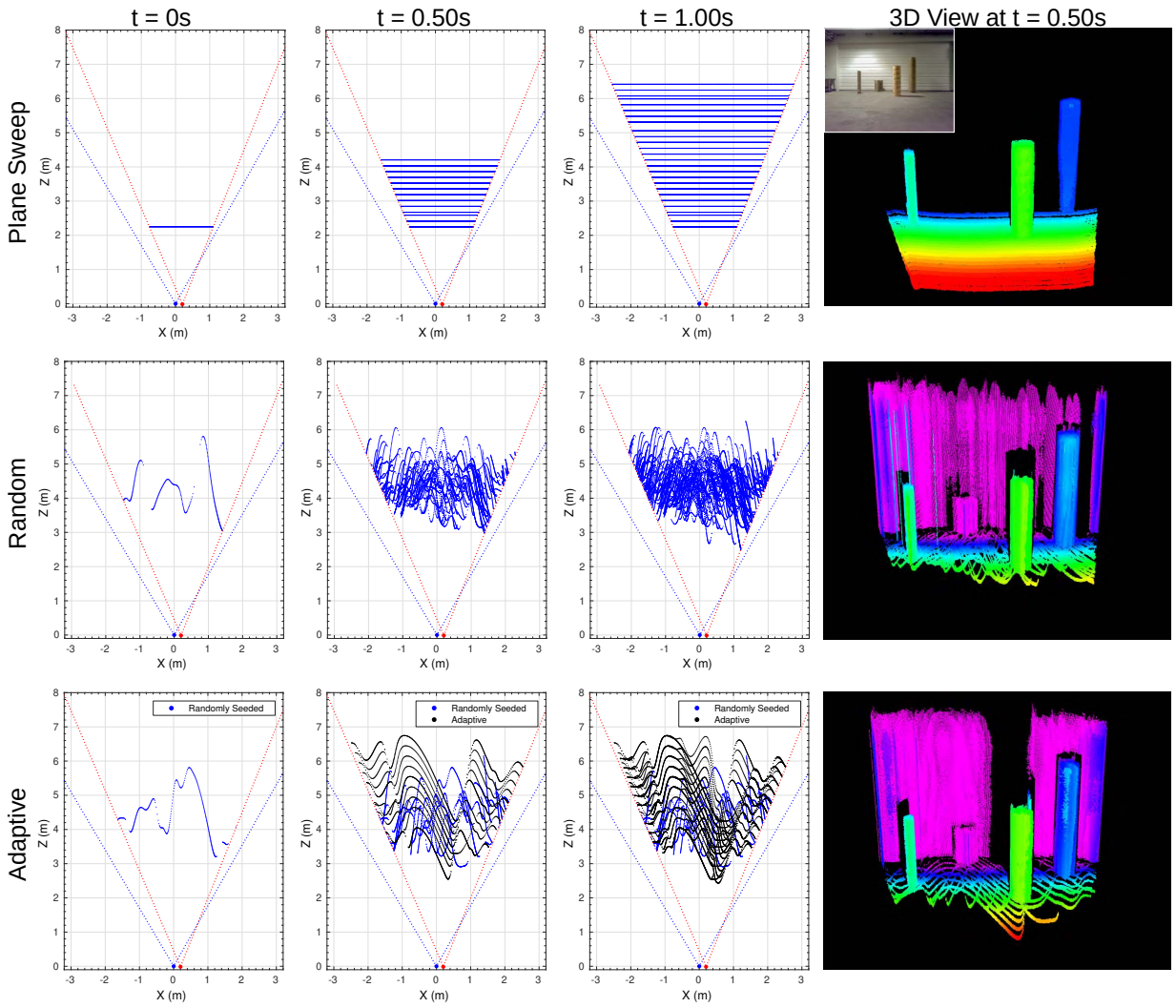


Figure 5.9: Depth imaging of a scene using planar curtains, random curtains, and adaptive curtains, shows that adaptive curtains intelligently cover more of the scene in less time than sweeping a plane or random sampling of the scene. After initialization with random sampling, the adaptive curtains discovered the interesting parts of the scene in 0.25 s and then continued to refine the map.

of the scene. By interleaving a few random curtains with the adaptive curtains, the device can continue checking the scene for any changes and sample the rest of the scene at low resolution. The design process for the experiment in Figure 5.8 projected all of the detected points to the xz-plane and used the closest point within an angular region as the design point for that region. By splitting the entire plane into these uniformly spaced angular regions, a set of design points were determined.

Depth Imaging Comparison: A comparison of depth imaging methods using different light curtain types was performed by capturing the same scene with each curtain type and comparing the coverage of the scene at specified times. The methods included plane sweeping, random sampling, and adaptive depth imaging. For plane sweeping, the planes were designed to fully image the $4.25\ \mathrm{m}$ scene in $1.0\ \mathrm{s}$ with a depth resolution of $0.2\ \mathrm{m}$. The adaptive curtains were tuned to sense the front surfaces of detected objects at a high resolution and the rest of the scene at a low resolution. Figure 5.9, shows the results of this comparison. Once initialized with $0.25\ \mathrm{s}$ of random detections, the discovery nature of the adaptive curtain enabled it to sense near the detected objects and not waste time sensing empty areas at the front of the scene and it was able to quickly cover the interesting parts of the scene in less than $0.25\ \mathrm{s}$. Given the time, plane sweep curtains can provide high resolution and complete coverage, but with limited time random and adaptive curtains can image more of the scene in less time. For example, the plane sweep curtains could have been configured to image the entire scene in $0.5\ \mathrm{s}$ but at $0.4\ \mathrm{m}$ depth resolution, which is much less than the other methods.

Discovery and Mapping using Adaptive Light Curtains: Adaptive light curtains can be used to discover and map a scene from a moving platform. We used a small robot with on-board localization to move the light curtain device through a cluttered highbay scene. As the robot progressed through the scene, adaptive light curtains discovered the scene structure and continuously adapted to image the newly detected objects from each frame. Rather than using a fixed depth sensing strategy, the light curtains intelligently sampled regions around the detected objects at high-resolution and sampled the rest of the scene with random curtains at a lower resolution. For our mapping experiments, a set of 5 uniquely random curtains were interleaved with every 10 adaptive curtains. Figure 5.10 shows a few instances of the curtains fitting to objects in the scene.

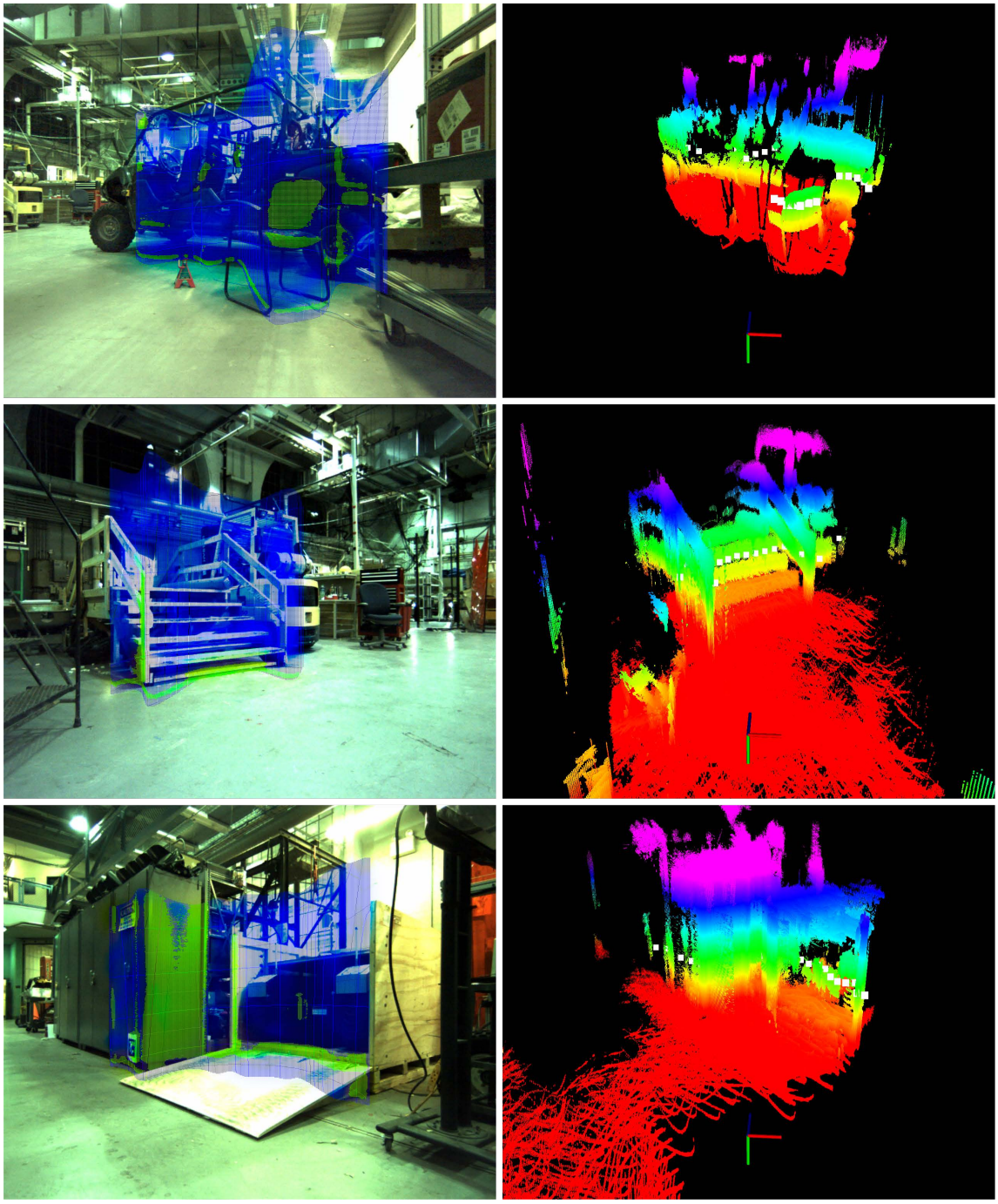


Figure 5.10: As the light curtain device was moved through a highbay scene, adaptive light curtains discovered the scene and adapted to image the surfaces in the scene. Notice that the curtains (shown in blue) fit tightly around the objects in the scene. When moved throughout the environment, the detected points on the curtains (shown in green) also mapped the scene (right column). The white points on the 3D maps show the design points for the set of adaptive curtains.

(a) Robot Inside of Tunnel **(b)** Outside View of Tunnel

Figure 5.11: Tunnel environment for smoke experiments. (a) A small robot was equipped with the light curtain device and a scanning LIDAR and was then used to map a 25 m long tunnel. The beams in the tunnel are made of wood. (b) An outside view shows that the wood structure is covered with a dark tarp material to form the enclosed tunnel. The robot in (a) is just inside the white doors in (b) and looking at the end of the tunnel (right side of b).

5.8 Mapping in Smoke using Light Curtains

As demonstrated in Chapter 4, light curtains work well to see through smoke. The key insight here is that since light curtains only receive light from a single depth at a time they can block most of the light that is scattered by the smoke through the volume since it is not at the same depth, which in turn reduces the signal to noise ratio and extends the visibility range. We demonstrated the enhanced ability of light curtains to see through smoke by mapping a smoke-filled tunnel with the light curtain prototype. The light curtain was moved through a 1.5 m tall, 1.5 m wide, by 25 m long tunnel by a tele-operated robot shown in Figure 5.11. A 2.5 m planar curtain was pushed through the volume and a map was created using the robot's on-board localization system. Figure 5.12 shows the results of these tests. A comparison of the light curtain to a Velodyne scanning LIDAR (VLP-16) is shown in Figure 5.13. In many locations, the maximum range of the LIDAR is roughly 2.5 m; the same distance the light curtain was set to. Several longer range curtains were tried between 3-5 m, and although they worked where the smoke was thinner they were not adequate for areas with thicker smoke. The LIDAR returns more points on the side of the tunnel than the light curtain for two reasons. First, is that the side of the tunnel was a black material which has a low albedo that the 905 nm operating wavelength of the LIDAR can see better than the 830 nm wavelength of the light curtain device. The second reason is that since the light curtain is only looking at a single depth at its maximum range in smoke, the returns from the side are even lower than they would be at closer ranges. The LIDAR

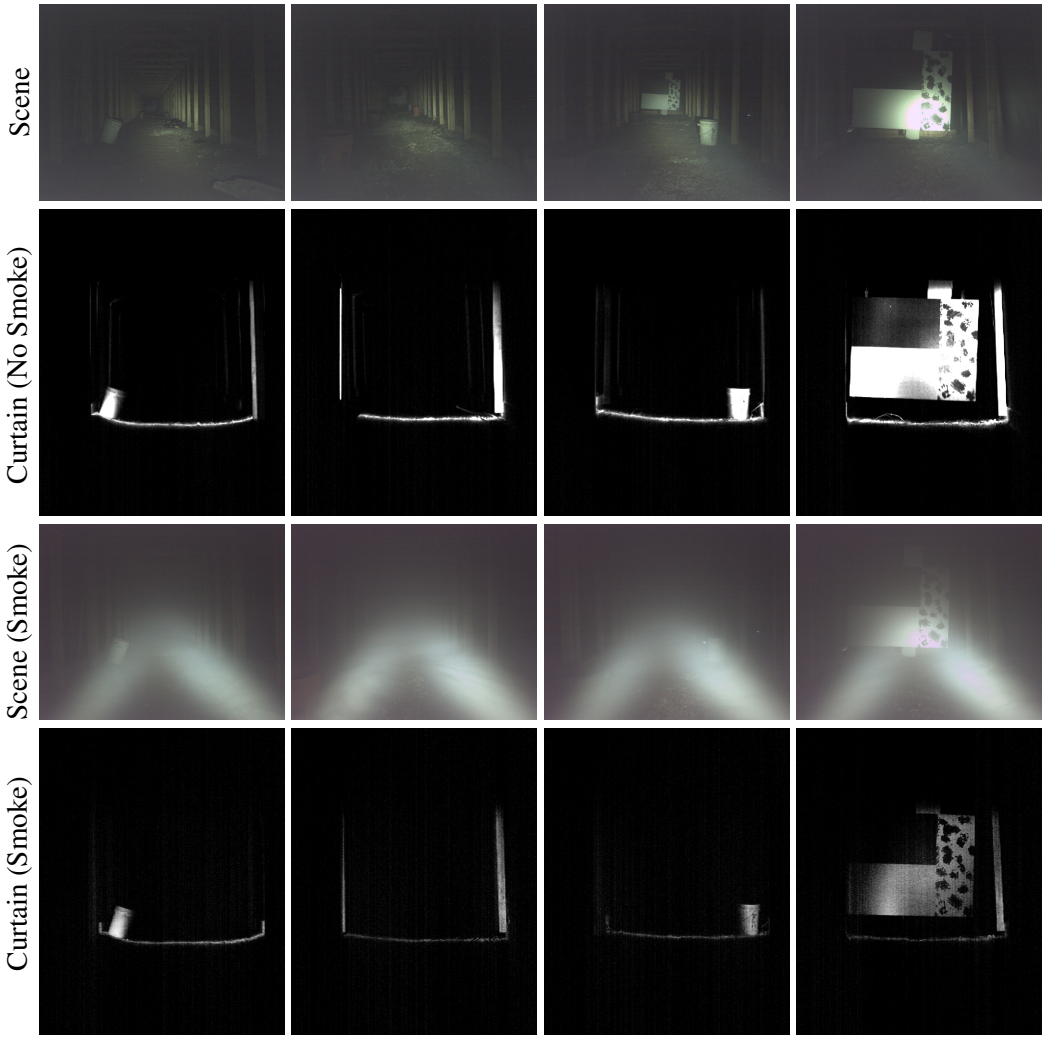


Figure 5.12: Light curtains in smoke. A planar light curtain set at 2.5 m was pushed through a tunnel by a tele-operated robot. Data was collected with and without smoke in the tunnel for comparison. Although, the light curtain intensity is not as strong in the smoke images, it is strong enough to be thresholded adequately. The reflectivity target at the end of the tunnel (right column) shows the difference in return intensity for low and high albedo targets with and without smoke.

senses the side of the tunnel at all depths and often it doesn't see the side until it is closer than 2.5 m away. A generated map from the light curtain data is shown in Figure 5.14. This figure shows that there is little difference between the maps of the tunnel with and without smoke. The map generated with smoke shows a reduced pointcloud thickness at some regions and along the walls of the tunnel, but the floor, objects, and vertical beams of the tunnel are detected. One thing worth noting here is that although the smoke attenuates the strength of the returned light, the light curtain images show very little effects of scattering (i.e. there is very little noise in the

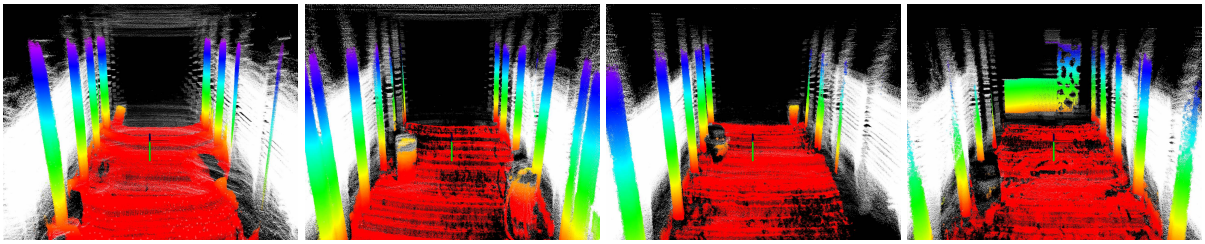
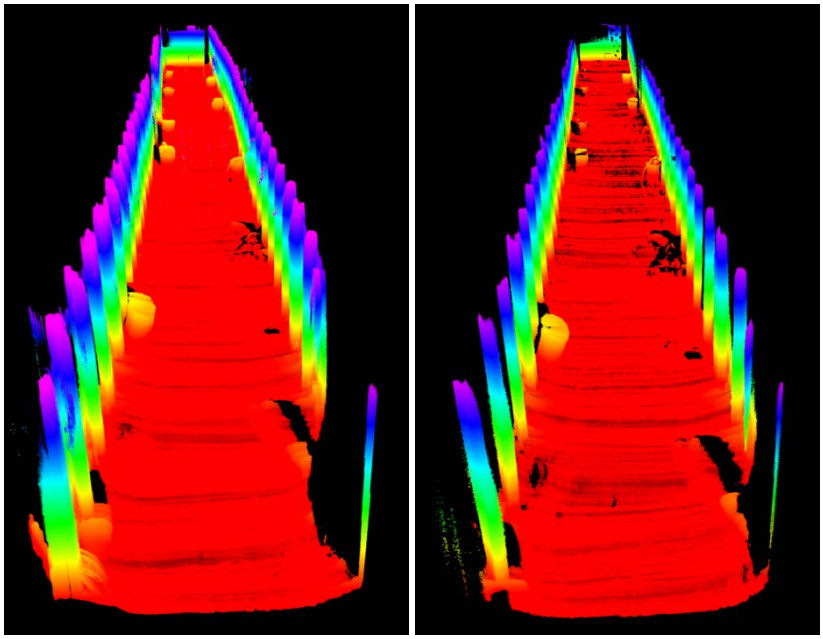


Figure 5.13: Comparison of LIDAR to light curtains in smoke. Columns in this figure are the associated pointclouds for the curtains in the columns of Figure 5.12. In thinner smoke the LIDAR (white points) can see farther than the set 2.5 m distance of the light curtain, but as the smoke gets thicker, the 2.5 m planar light curtain senses as far as the LIDAR. The LIDAR detects the low albedo walls better because of its higher NIR operating wavelength and because it gets multiple (closer) views of them, whereas the light curtain only has a single view at 2.5 m away. The tunnel is approximately 1.5 m tall.



(a) 3D Map (No Smoke) **(b)** 3D Map (Smoke)

Figure 5.14: Comparison of map generated by light curtain in tunnel with and without smoke. There are only slight differences between the maps generated in smoke and without. There is slight reduction in pointcloud thickness in some areas where the smoke was very thick, but the light curtain still detected the floor, objects, and vertical beams in the tunnel.

images). This means that only a small amount of scattered light is being captured and that the light curtain device could see farther through the smoke with increased projected light power.

5.9 Discussion

In summary, our rolling shutter triangulation light curtain prototype enables rapid on-demand depth sensing of a scene. Our system can be used for uniform depth sensing by sweeping well-defined curtains over the scene, which offers predictable but slow coverage of the scene. For quicker but non-deterministic coverage, randomly generated curtains can be used. Both of these methods, like traditional LIDAR and depth cameras, sample the scene without regard for its structure. Adaptive light curtains can sample the scene more efficiently based on its structure. We demonstrated a simple method that showed this by first detecting objects in the scene with low resolution random sampling and then adapting the curtains for high-resolution scanning of the detected objects while continuing a low resolution scan of the rest of the scene. More advanced sampling algorithms in the future may use machine learning to determine the best places to scan a given scene. For example, in self-driving car applications, an algorithm could learn where to sample based on the eye-movements that human drivers use to safely navigate roads [57, 72].

This work was the first to show real-time adaptive depth-sampling of a scene with triangulation light curtains and offers many benefits to applications that require computationally efficient depth imaging. Our system can be used in robotics to discover and map scenes; including scenes in bright ambient light, darkness, and in smoke. This work enables real-time adaptation for agile depth sensing tasks ranging from human robot interaction [28], to robot manipulation [68], path planning [30], and navigation [16]. Future work in this area could look at how to fuse the on-demand depth sensing of light sheet depth imaging devices with learning algorithms for in the loop sensing for applications in monocular depth reconstruction [34, 103], person tracking [49, 100], and obstacle recognition [21].

Chapter 6

Conclusion

6.1 Summary

Depth imaging systems that sample only at a line have distinct advantages to systems that image points or entire fields of light. Specifically, these systems offer the resolution and frame-rate of consumer depth cameras but work outside in bright light. Although, they do not have the range of LIDAR devices, they offer higher-resolution and more economical depth sensing at short to moderate ranges. Similar to LIDAR, the sampled line is scanned to cover the scene. But unlike LIDAR and consumer grade depth cameras our systems are not constrained to capturing the whole volume uniformly and can adaptively sample the scene. Adaptive sampling of the scene can be used to only capture the interesting and necessary regions of the scene. Unlike traditional depth imaging sensors which capture the entire volume to produce dense pointclouds that require extensive processing to find the interesting regions, adaptive sampling can provide these regions directly—especially when paired with triangulation light curtains. Triangulation light curtains offer the advantage that they can detect the presence of objects along a given surface very efficiently with just simple image thresholding, but can only do so along a single defined surface in the volume per frame. Epipolar time-of-flight imaging on the other hand captures the entire volume in a single frame, but requires complex pointcloud processing for obstacle detection.

This thesis asserts that 1) imaging a sheet of light with an aligned plane of imaging enables robust, high-rate, and agile depth sensing; 2) when these planes are aligned and scanned in an epipolar configuration, the depth of the entire scene can be captured out to a maximum working range; 3) when the rotation axes of these scanned planes are parallel and separated by a baseline they can triangulate to capture only the depths of objects along a defined surface in the volume; and 4) that these planes can be steered to select and adaptively change which regions of the scene

are sampled and at what resolution.

Chapter 2 discussed the development of an epipolar continuous-wave time-of-flight depth imaging camera using a light sheet projector and a 2D ToF camera. This first prototype demonstrated robustness to multi-path interference caused by global light transport as well as demonstrated a working range of 15 m outdoors in bright sunlight.

Chapter 3 considered design parameters for epipolar time-of-flight imaging. A simulation frame-work was developed that was then used to guide the design of a new prototype which demonstrated a working range of 50 m outdoors.

Chapter 4 introduced the concept of programmable triangulation light curtains. The key insight here was that by rotating the light sheet with respect to the imaging plane such that their rotation axes were parallel and separated by a baseline the device would sense depths only along the line at the intersection of the planes. This line can then be rapidly swept by steering the planes. The constructed prototype used a line sensor and line laser both steered by galvomirrors to generate light curtains along any ruled surface. The prototype could image curtains over 25 m away outdoors in bright sunlight and had a frame rate of 5.6 fps.

Chapter 5 detailed development of a new programmable light curtain prototype that uses a rolling shutter camera to steer the imaging plane in sync with the light sheet projector to image 60 different curtains per second rather than one curtain at 5.6 fps. This increased frame rate and improved prototype enabled curtains to be dynamically adjusted to adaptively sample the scene based on its current knowledge of the scene structure. This improved frame rate also enabled applications for robot mapping, path planning, and navigation in smoke.

6.2 Future Work

6.2.1 Motion Estimation

One of the limitations of epipolar imaging and triangulation light curtains is that they capture measurements at different times for each line sample in the image. When the sensor and scene are static, this is not an issue, but when either are moving, the time offset between each measurement causes distortion in the resulting image similar to that of rolling shutter cameras. The problem with regard to rolling shutter cameras has been solved with several methods that essentially all estimate the motion of the camera during the capture and then unwarp the data to create a corrected image. For large movements during capture, many of these methods fail due to assumptions about small motion, photometric consistency, and occlusions. When motion is large between rows, as it can be for these devices at slow frame-rate, the motion estimation problem

is more similar to a LIDAR SLAM problem, and similar methods can be used. Precise pose estimation for devices with large motion between measurements requires modelling the pose of each measurement. This is most efficiently done with a continuous-time, spline-based representation. Many state-of-the-art LIDAR algorithms model each point as being captured at a different time, but use linear interpolations for motion during the scan [125]. Adaptive sampling methods where rows are exposed for different times or another non-uniform sampling method could violate this linear assumption. In addition, rapidly changing motion faster than the sampling rate, would violate this assumption. A continuous-time, cubic B-Spline based approach [77] that uses feature points similar to [125], instead of dense image representations, could be used to model the motion of the device and generate pose and map estimates from epipolar depth cameras.

6.2.2 Towards Economical, High-Rate, and Precise Range Sensing

The prototype devices developed in this thesis demonstrated reliable depth imaging outdoors out to 25 m in bright sunlight. At this range however, the noise of the time of flight camera was high and light curtains were several meters thick. In addition, the frame rate of the time of flight camera was limited to approximately 5 fps. Since these devices were research prototypes they were also not economical. Two paths of future work exist to both increase the frame rate and improve the range precision of these devices, as well as make them more economical.

First, is by improving the hardware. As detailed in Chapter 3 using a CW-ToF image sensor capable of imaging at modulation frequencies in the 100 MHz range combined with phase unwrapping should enable depth errors of just a few centimeters at extended ranges in bright sunlight. As shown in Chapter 4, these ToF sensors can then be used in triangulation light curtains to greatly reduce the thickness of even short base-line light curtain devices. Both the range and speed of the devices can be improved by using more light power. Using more light power can directly improve range but using shorter exposures with more light power can increase the frame rate while maintaining range. However, improving the range by increasing laser power has implications on the eye-safety of the device [1] which must be carefully considered. Although rolling shutter imagers enable rapid imaging, the region of interest control used to steer the plane of imaging with time-of-flight imagers is not as fast. Ideally a rolling shutter time-of-flight sensor would be used, but this is usually not desired for ToF cameras and they are exclusively global shutter. Since most applications do not require rapid region of interest switching, this may be a limitation that can only be removed with the design of a custom image sensor. If not for economic reasons, at some point the galvomirror used to steer the light sheet will limit the speed of the device. MEMS mirrors offer much higher speeds, are less expensive, and could be used to steer the light sheet in a faster, more compact, and more economical manner. The hardware

improvements mentioned here all either currently exist in some form or are straightforward extensions of current technology. Since there are no complicated electronics or exotic imaging sensors, these devices can be made inexpensively at scale.

In addition to improving the hardware, several algorithmic methods can improve both the range and frame rate. A large improvement for the signal to noise ratio of these devices would be to use a more optimal time-of-flight modulation method. Continuous sine-wave modulation offers little robustness to noise and can be improved with coding of the modulation and demodulation. One promising method is Hamiltonian coding which improves the depth resolution of ToF cameras by an order of magnitude without changing the light source power or acquisition time [44]. Deep learning methods can be used in addition to this to estimate the correct depths from data with low signal to noise ratio [22, 111]. To improve frame-rate, adaptive sensing can be used to image only the necessary objects. LIDAR depth completion methods[23, 79] and fusion with a color camera could then be used to improve the sampling rate and resolution of the devices by intelligently upsampling the captured points.

Although, these devices will likely never have the range of LIDAR systems, these hardware changes coupled with advanced algorithms should provide high-rate, precise, and economical depth sensing outdoors at short to moderate ranges useful for many applications in robotics.

6.3 Outlook

The depth camera prototypes developed in this thesis demonstrated capabilities including robust outdoor depth imaging, programmable light curtains, and adaptive depth sensing. Each of these has their own merit and application, but robust outdoor imaging is likely the most applicable today.

Programmable light curtains have specific applications in industrial manufacturing and for safety sensors around vehicles, but since there are not any directly similar technologies it will take time for them to catch on—especially, since their main applications are safety related. Proving out a new safety sensor will not be easy and will require significant research, development, and testing. However, in time, their capabilities could revolutionize safety sensing and be enabling technology for advanced applications in human and machine interactions that require adaptive sensing.

Adaptive depth sensing, especially when paired with programmable light curtains, offers significant benefits to computational constrained systems by imaging different parts of the scene only at the resolution necessary, removing much of the heavy processing. LIDAR hardware capable of pointing the beam and imaging areas of interest at different resolutions has been

demonstrated [62, 113] and companies are starting to produce hardware capable of this [3, 52]. Intelligent adaptive depth sensing is an emerging area, with promises to drastically improve sampling efficiency of scenes and reduce the computational load for many applications. As self-driving cars and robotics develop, adaptive depth sensing will be an enabling technology.

Robust outdoor depth imaging with epipolar time-of-flight cameras has the most immediate future as it is similar to sensors currently used for these tasks, but with increased capabilities. Since consumer depth cameras do not work outdoors, the sensor of choice for many outdoor applications is scanning LIDAR, simply because it offers precise ranging at long ranges in bright sunlight. Many short-mid range applications that could use passive stereo do not use it because of its unreliability in areas of low texture and its high-processing requirement. Slow moving robots or self-driving cars are both examples of applications for short-mid range depth sensing that customarily use LIDAR to see at these ranges. Many self-driving cars will go so far as to have multiple small LIDAR units on their roofs angled downward just to sense the area within a few meters of the car. These are prime examples of where an economical, short to mid range, outdoor depth camera would be useful. For self-driving cars and robots in general to ever become common place and successful they must have an economical method for sensing their environment. Although LIDAR companies are increasing production to meet demand of these new applications, its not clear if the increased scale will be enough to reduce the price of these devices to that necessary for their target applications.

6.3.1 Applications

The devices developed in this thesis have a large economical benefit over existing LIDAR sensors. Since epipolar imaging and light curtain devices use CMOS-based imagers, entire high-resolution 2D arrays are straight-forward and inexpensive to produce. This is in contrast to avalanche photodiodes that LIDAR utilizes which, although are much more sensitive and robust, are larger and more expensive. Although avalanche photodiode arrays are produced for flash LIDAR cameras, these cameras are exorbitantly expensive and inaccessible to most applications. The commodity that comes from being able to use normal and existing CMOS-based image sensors is important, as it enables our technology to be used in many existing, as well as up and coming, applications that require inexpensive depth sensing. Beyond economics, many applications would benefit from the low computational requirements provided by triangulation curtains and adaptive sampling.

Mobile Robots. As the role of mobile robots in society increases, the need for economical, short to medium range, and high resolution outdoor depth sensing will grow. Mobile robots are

one of the best use cases of these sensors, as their speed of travel reduces their required stopping distance and makes them amenable for using the short to medium range depth sensors proposed in this thesis as their sole depth sensing technology. The economics is key here as large-scale applications like delivery robots [58, 105], and robotic shuttles will require it. Companies are already springing up around these ideas and in their development phase are using a combination of expensive LIDAR units and consumer grade stereo cameras. Neither of which is optimal, for reasons discussed previously.

Self-Driving Cars. Although many applications in self-driving cars require long-range depth sensing, critically important applications in obstacle detection and recognition at close ranges prefer high-resolution, short-range sensing. Having high-resolution depth sensing at these close ranges is important because it helps the system to distinguish, for example, between a sack blowing with the wind and a fast moving person on a bicycle. One of these is safe to hit and the other is not. LIDAR systems are starting to be developed for these applications [118, 121], but the economics is likely not there even at large scale compared to CMOS-based sensors. Besides applications in obstacle recognition, self-driving cars do not move fast all the time (e.g., urban environments, dense traffic, construction zones, etc.) and many of these low speed circumstances could benefit from the proposed depth imaging systems.

Outdoor Depth Imaging. There are many applications for these sensors in the more general category of outdoor depth imaging. Applications include depth imaging for security systems, agriculture, and infrastructure modeling [50, 80]. Some advanced security systems use LIDAR for intelligent patrol robots [65] and imaging [99]. Applications in agriculture include plant imaging for yield estimation [120], plant phenotyping [75], and livestock monitoring [116, 123]. For many of these applications the range of LIDAR is not necessary and stereo-imaging is not reliable enough. A short to medium range depth camera would be ideal.

Underwater Imaging. A large majority of Earth's surface is covered by water, but detailed 3D maps of these environments are few and limited in resolution. Acoustic methods have previously been used for large-scale underwater mapping, but the maps are noisy and the resolution is low [15]. Inspection and mapping of ships [94], pipes [101], bridges [29], and other underwater infrastructure [124] requires more accurate and higher-resolution 3D modeling than acoustic methods can offer. The highest performing underwater depth imaging technology is range gated imaging [54, 82], but the requirement on highly precise timing makes these devices too costly for widespread use. More commonly laser line scanning systems are are used as first proposed

by [55]. These line scanning systems are very similar to epipolar imaging and triangulation light curtain devices in that they only image a single emitted line at a time and reduce scattering, but they do so with by capturing a full 2D image of the line and are slow to scan the full scene. The rapid scanning ability of our developed devices would increase the speed of these devices and offer improved robustness in many cases by optically blocking the scattered light.

Industrial Manufacturing. Applications for depth sensing in industrial manufacturing are vast. Many depth sensing technologies originated for sensing in these applications. Tasks in metrology [47], human safety [17], and automated material handling [13, 114] all heavily rely on different types of depth sensors. Laser line imaging systems are used for robotic part handling and inspection [14]. 2D scanning LIDAR sensors are used for robotic cart navigation [39]. Each of these specialized sensors are based on the same core technology of emitting and capturing lines of light with an imaging system, but each are configured differently for their specific application and require different hardware. The technology developed in this thesis could be adapted to perform many of these tasks with a single depth sensor.

6.4 Conclusions

Conclusions of this research are:

- Rather than sampling the scene with points, like scanning LIDAR, or all at once, like consumer depth cameras, robust depth imaging can be enabled by sampling lines.
- Concentrating the light and imaging into a single line reduces required exposure time and improves signal to noise ratio of received light.
- Depth sampling along lines can be implemented by imaging sheets of projected light with precisely aligned cameras.
- Rapid sampling of the scene is enabled by sweeping this line through the scene.
- When the sheet of light is imaged with an epipolar-aligned row of a time-of-flight camera, the resulting device can see at least 15 m outdoors in bright sunlight, and is robust to multi-path interference, and camera motion.
- Simulation-guided design and development of epipolar time-of-flight cameras results in a device that demonstrates depth imaging up to 50 m outdoors.
- When the rotation axis of a light sheet projector is separated by a baseline but parallel to the rotation axis of the sheet of imaging, sampled lines can be triangulated and swept to form a programmable triangulation light curtain.

- Programmable triangulation light curtains only see what is directly touching the curtain and in this way form a surface of obstacle detection requiring very low computation.
- The shape of light curtains can be programmed to form any ruled surface.
- When triangulation light curtains are implemented with a light sheet projector and a line imager steered by a galvomirror, the resulting device can image static curtains at 5.6 frames per second and can see up to 25 m outdoors in bright sunlight as well as through smoke.
- When the imaging plane of a triangulation light curtain is steered by the rolling shutter motion of a intensity imager, the resulting device can image curtains up to 60 frames per second.
- Rapid imaging of programmable light curtains can be used for agile depth sensing to adaptively sample scenes or only image specified areas of interest.
- Adaptive sampling of the scene with triangulation light curtains can selectively and adaptively image different parts of the scene at the required resolutions which can reduce unnecessary processing.
- Rolling shutter light curtain devices can be used to see through smoke and map the environment when navigated with a robot.
- Light sheet depth imaging devices like epipolar time-of-flight cameras and programmable light curtains provide most of the robustness of LIDAR with the speed and sampling density of consumer depth cameras and can be used for short to moderate range depth sensing outdoors.
- Since light sheet depth imaging devices use commodity image sensors and components, they can be produced economically for use in wide-spread applications and offer an alternative to scanning LIDAR for applications that prefer short to medium range sensing.

In summary, the work developed within this thesis contributed methods and hardware for high-resolution depth imaging that works in challenging conditions, provides methods for computationally inexpensive agile depth sensing, and has the economics that will enable next generation and wide-scale applications in robotics, agriculture, and industrial manufacturing.

Bibliography

[1] Supreeth Achar. *Active Illumination for the Real World*. Cmu-ri-tr-17-65, Carnegie Mellon University, 2017. 6.2.2

[2] Supreeth Achar, Joseph R. Bartels, William L. 'Red' Whittaker, Kiriakos N. Kutulakos, and Srinivasa G. Narasimhan. Epipolar time-of-flight imaging. *ACM Trans. Graph.*, 36 (4):37:1–37:8, July 2017. ISSN 0730-0301. doi: 10.1145/3072959.3073686. 1.5, 3.5.5, 5.2

[3] aEye. idar, 2019. URL `https://www.aeye.ai/idar/`. 1.3, 6.3

[4] G Agin. Calibration and use of a light stripe range sensor mounted on the hand of a robot. In *Robotics and Automation. Proceedings. 1985 IEEE International Conference on*, volume 2, pages 680–685. IEEE, 1985. 1.2

[5] Hatem Alismail and Brett Browning. Automatic calibration of spinning actuated lidar internal parameters. *Journal of Field Robotics*, 32(5):723–747, 2015. 1.2

[6] Hatem Alismail, L Douglas Baker, and Brett Browning. Continuous trajectory estimation for 3d slam from actuated lidar. In *Robotics and Automation (ICRA), 2014 IEEE International Conference on*, pages 6096–6101. IEEE, 2014. 2.3

[7] American National Standards Institute. *American National Standard for Safe Use of Lasers Z136.1*. Laser Institute of America, 2014. ISBN 9781940168005. 1.2, 3.5.5

[8] ASTM International. ASTM G173-03(2012) Standard Tables for Reference Solar Spectral Irradiances: Direct Normal and Hemispherical on 37 Tilted Surface, 2012. 1.3, 3.2.1

[9] Ian M Baker, Stuart S Duncan, and Jeremy W Copley. A low-noise laser-gated imaging system for long-range target identification. In *Infrared Technology and Applications XXX*, volume 5406, pages 133–145. International Society for Optics and Photonics, 2004. doi: 10.1117/12.541484. 5.2

[10] Simon Baker, Eric Bennett, Sing Bing Kang, and Richard Szeliski. Removing rolling shutter wobble. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference*

on, pages 2392–2399. IEEE, 2010. 2.1

[11] Ian Baldwin and Paul Newman. Road vehicle localization with 2d push-broom lidar and 3d priors. In *Robotics and automation (ICRA), 2012 IEEE international conference on*, pages 2611–2617. IEEE, 2012. 1.2

[12] Andrew J Barry and Russ Tedrake. Pushbroom stereo for high-speed navigation in cluttered environments. In *2015 IEEE International Conference on Robotics and Automation (ICRA)*, pages 3046–3052. IEEE, 2015. 4.1, 5.2

[13] N Bellomo, E Marcuzzi, L Baglivo, M Pertile, E Bertolazzi, and M De Cecco. Pallet pose estimation with lidar and vision for autonomous forklifts. *IFAC Proceedings Volumes*, 42 (4):612–617, 2009. 6.3.1

[14] Igor Bei, Nick Van Gestel, Jean-Pierre Kruth, Philip Bleys, and Janko Hodoli. Accuracy improvement of laser line scanning for feature measurements on cmm. *Optics and Lasers in Engineering*, 49(11):1274 – 1280, 2011. ISSN 0143-8166. 6.3.1

[15] Brian Bingham, Brendan Foley, Hanumant Singh, Richard Camilli, Katerina Delaporta, Ryan Eustice, Angelos Mallios, David Mindell, Christopher Roman, and Dimitris Sakellariou. Robotic tools for deep water archaeology: Surveying an ancient shipwreck with an autonomous underwater vehicle. *Journal of Field Robotics*, 27(6):702–717, Nov 2010. doi: 10.1002/rob.20350. 6.3.1

[16] J. Biswas and M. Veloso. Depth camera based indoor mobile robot localization and navigation. In *2012 IEEE International Conference on Robotics and Automation*, pages 1697–1702, May 2012. 5.9

[17] David Blau. Self-checking light curtain system and method of operation, 1995. 6.3.1

[18] Steven D Blostein and Thomas S Huang. Error analysis in stereo determination of 3-d point positions. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, PAMI-9(6):752–765, 1987. 4.2.2

[19] Michael Bosse, Robert Zlot, and Paul Flick. Zebedee: Design of a spring-mounted 3-d range sensor with application to mobile mapping. *IEEE Transactions on Robotics*, 28(5): 1104–1119, 2012. 1.2

[20] Samuel Burri, Harald Homulle, Claudio Bruschini, and Edoardo Charbon. Linospad: a time-resolved 256$\times$ 1 cmos spad line sensor system featuring 64 fpga-based tdc channels running at up to 8.5 giga-events per second. In *Optical Sensing and Detection IV*, volume 9899, page 98990D. International Society for Optics and Photonics, 2016. 4.5

[21] Xiaozhi Chen, Huimin Ma, Ji Wan, Bo Li, and Tian Xia. Multi-view 3d object detection

network for autonomous driving. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017. 5.9

[22] Yan Chen, Jimmy Ren, Xuanye Cheng, Keyuan Qian, and Jinwei Gu. Very power efficient neural time-of-flight. *arXiv preprint arXiv:1812.08125*, 2018. 6.2.2

[23] Nathaniel Chodosh, Chaoyang Wang, and Simon Lucey. Deep convolutional compressed sensing for lidar depth completion. *arXiv preprint arXiv:1803.08949*, 2018. 6.2.2

[24] J. Choi, S. Ulbrich, B. Lichte, and M. Maurer. Multi-target tracking using a 3d-lidar sensor for autonomous vehicles. In *16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013)*, pages 881–886, Oct 2013. 1.3, 5.2

[25] Jacky CK Chow and Derek D Lichti. A study of systematic errors in the pmd camboard nano. In *Videometrics, Range Imaging, and Applications XII; and Automated Visual Inspection*, volume 8791, page 87910X. International Society for Optics and Photonics, 2013. 1.2

[26] Oliver S Cossairt, Nathan Matsuda, and Mohit Gupta. Motion contrast 3d scanning. In *Computational Optical Sensing and Imaging*, pages CT2E–1. Optical Society of America, 2015. 2.1

[27] Ofer David, Norman S Kopeika, and Boaz Weizer. Range gated active night vision system for automobiles. *Applied optics*, 45(28):7248–7254, 2006. 1.4

[28] A. De Luca and F. Flacco. Integrated control for phri: Collision avoidance, detection, reaction and collaboration. In *2012 4th IEEE RAS EMBS International Conference on Biomedical Robotics and Biomechatronics (BioRob)*, pages 288–295, June 2012. doi: 10.1109/BioRob.2012.6290917. 5.9

[29] J.E. DeVault. Robotic system for underwater inspection of bridge piers. *IEEE Instrumentation & Measurement Magazine*, 3(3):32–37, 2000. doi: 10.1109/5289.863909. 6.3.1

[30] Dmitri Dolgov, Sebastian Thrun, Michael Montemerlo, and James Diebel. Path planning for autonomous vehicles in unknown semi-structured environments. *The International Journal of Robotics Research*, 29(5):485–501, 2010. doi: 10.1177/0278364909359210. 5.9

[31] Adrian A Dorrington, John P Godbaz, Michael J Cree, Andrew D Payne, and Lee V Streeter. Separating true range measurements from multi-path and scattering interference in commercial range cameras. In *Three-Dimensional Imaging, Interaction, and Measurement*, volume 7864, page 786404. International Society for Optics and Photonics, 2011. doi: 10.1117/12.876586. 2.1

[32] David Droeschel, Dirk Holz, and Sven Behnke. Multi-frequency Phase Unwrapping for Time-of-Flight cameras. In *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1463–1469. IEEE, Oct 2010. doi: 10.1109/IROS.2010.5649488. 3.5.3

[33] C Dunsby and PMW French. Techniques for depth-resolved imaging through turbid media including coherence-gated imaging. *Journal of Physics D: Applied Physics*, 36(14):R207, 2003. 1.3

[34] David Eigen, Christian Puhrsch, and Rob Fergus. Depth map prediction from a single image using a multi-scale deep network. In *Advances in neural information processing systems*, pages 2366–2374, 2014. 5.9

[35] Alberto Elfes. Sonar-based real-world mapping and navigation. *IEEE Journal on Robotics and Automation*, 3(3):249–265, 1987. 1.2

[36] Rob Fergus, Barun Singh, Aaron Hertzmann, Sam T. Roweis, and William T. Freeman. Removing camera shake from a single photograph. *ACM Trans. Graph.*, 25(3):787–794, July 2006. ISSN 0730-0301. doi: 10.1145/1141911.1141956. 2.1

[37] Alex Foessel-Bunting, John Bares, and WL Whittaker. Three-dimensional map building with mmw radar. In *Proceedings of the International Conference on Field and Service Robotics, RCA Halme and E. Prassler, Eds., Helsinki, Finland*, 2001. 1.2

[38] S.B. Gokturk, Hakan Yalcin, and Cyrus Bamji. A Time-Of-Flight Depth Sensor - System Description, Issues and Solutions. In *2004 Conference on Computer Vision and Pattern Recognition Workshop*, pages 35–35. IEEE, 2004. doi: 10.1109/CVPR.2004.291. 1.2, 3.3

[39] Hossein Golnabi. Role of laser sensor systems in automation and flexible manufacturing. *Robotics and Computer-Integrated Manufacturing*, 19(1):201 – 210, 2003. ISSN 0736-5845. 12th International Conference on Flexible Automation and Intelligent Manufacturing. 6.3.1

[40] Yoav Grauer and Ezri Sonn. Active gated imaging for automotive safety applications. In *Video Surveillance and Transportation Imaging Applications 2015*, volume 9407, page 94070F. International Society for Optics and Photonics, 2015. doi: 10.1117/12.2078169. 5.2

[41] Mohit Gupta and Shree K Nayar. Micro phase shifting. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 813–820. IEEE, 2012. 1.2

[42] Mohit Gupta, Qi Yin, and Shree K Nayar. Structured Light In Sunlight. *International Conference on Computer Vision*, 2013. 2.1, 5.2

[43] Mohit Gupta, Shree K. Nayar, Matthias B. Hullin, and Jaime Martin. Phasor Imaging. *ACM Transactions on Graphics*, 34(5):1–18, Nov 2015. doi: 10.1145/2735702. 2.1, 3.5.3

[44] Mohit Gupta, Andreas Velten, Shree K Nayar, and Eric Breitbach. What are optimal coding functions for time-of-flight imaging? *ACM Transactions on Graphics (TOG)*, 37 (2):13, 2018. 6.2.2

[45] Miles Hansard, Seungkyu Lee, Ouk Choi, and Radu Horaud. *Time-of-Flight Cameras: Principles, Methods and Applications*. Springer, 2013. ISBN 978-1447146575. 3.3, 3.5.3

[46] Marc P Hansen and Douglas S Malchow. Overview of swir detectors, cameras, and applications. In *Thermosense Xxx*, volume 6939, page 69390I. International Society for Optics and Photonics, 2008. 4.5

[47] Kevin Harding. *Handbook of optical dimensional metrology*. CRC Press, 2013. 6.3.1

[48] Felix Heide, Lei Xiao, Andreas Kolb, Matthias B. Hullin, and Wolfgang Heidrich. Imaging in scattering media using correlation image sensors and sparse convolutional coding. *Opt. Express*, 22(21):26338–26350, Oct 2014. doi: 10.1364/OE.22.026338. 2.1

[49] David Held, Sebastian Thrun, and Silvio Savarese. Learning to track at 100 fps with deep regression networks. In *European Conference on Computer Vision*, pages 749–765. Springer, 2016. 5.9

[50] Peter Henry, Michael Krainin, Evan Herbst, Xiaofeng Ren, and Dieter Fox. Rgb-d mapping: Using kinect-style depth cameras for dense 3d modeling of indoor environments. *The International Journal of Robotics Research*, 31(5):647–663, 2012. 6.3.1

[51] Berthold K P Horn. *Robot Vision*. MIT Press, 1986. 3.2.1, 3.2.1

[52] Innoviz. Innovizpro datasheet, 2019. 1.3, 6.3

[53] Texas Instruments. Opt8241 evaluation module user guide, 2017. 3.6

[54] Jules S. Jaffe. Underwater Optical Imaging: The Past, the Present, and the Prospects. *IEEE Journal of Oceanic Engineering*, 40(3):683–700, Jul 2015. doi: 10.1109/JOE.2014.2350751. 6.3.1

[55] Jules S. Jaffe and Chris Dunn. A Model-Based Comparison Of Underwater Imaging Systems. In Marvin A. Blizard, editor, *Proc. SPIE 0925, Ocean Optics IX, (12 August 1988)*, page 344, Aug 1988. doi: 10.1117/12.945742. 6.3.1

[56] Jules S Jaffe, Kad D Moore, John McLean, and Michael P Strand. Underwater optical imaging: status and prospects. *Oceanography*, 14(3):66–76, 2001. 1.3

[57] Ashesh Jain, Hema S. Koppula, Bharad Raghavan, Shane Soh, and Ashutosh Saxena. Car

that knows before you do: Anticipating maneuvers via learning temporal driving models. In *The IEEE International Conference on Computer Vision (ICCV)*, December 2015. 5.9

[58] Martin Joerss, Jürgen Schröder, Florian Neuhaus, Christoph Klink, and Florian Mann. Parcel delivery: The future of last mile. *McKinsey & Company*, 2016. 6.3.1

[59] Hordur Johannsson, Michael Kaess, Brendan Englot, Franz Hover, and John Leonard. Imaging sonar-aided navigation for autonomous underwater harbor surveillance. In *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*, pages 4396–4403. IEEE, 2010. 1.2

[60] Adrian P. P. Jongenelen, Donald G. Bailey, Andrew D. Payne, Adrian A. Dorrington, and Dale A. Carnegie. Analysis of Errors in ToF Range Imaging With Dual-Frequency Modulation. *IEEE Transactions on Instrumentation and Measurement*, 60(5):1861–1868, May 2011. doi: 10.1109/TIM.2010.2089190. 3.5.3

[61] Achuta Kadambi, Refael Whyte, Ayush Bhandari, Lee Streeter, Christopher Barsi, Adrian Dorrington, and Ramesh Raskar. Coded time of flight cameras: Sparse deconvolution to address multipath interference and recover time profiles. *ACM Trans. Graph.*, 32(6): 167:1–167:10, November 2013. ISSN 0730-0301. doi: 10.1145/2508363.2508428. 2.1

[62] Abhishek Kasturi, Veljko Milanovic, Bryan H Atwood, and James Yang. Uav-borne lidar with mems mirror-based scanning capability. In *Laser Radar Technology and Applications XXI*, volume 9832, page 98320M. International Society for Optics and Photonics, 2016. doi: 10.1117/12.2224285. 5.2, 6.3

[63] Christian Kerl, Jörg Stückler, and Daniel Cremers. Dense continuous-time tracking and mapping with rolling shutter rgb-d cameras. In *Computer Vision (ICCV), 2015 IEEE International Conference on*, pages 2264–2272. IEEE, 2015. 2.1, 2.3

[64] Lindsay Kleeman and Roman Kuc. Mobile robot sonar for target localization and classification. *The International Journal of Robotics Research*, 14(4):295–318, 1995. 1.2

[65] Knightscope. Knightscope, 2019. URL `https://www.knightscope.com/`. 6.3.1

[66] Kurt Konolige. Projected texture stereo. In *Robotics and Automation (ICRA), 2010 IEEE international conference on*, pages 148–155. IEEE, 2010. 1.2

[67] H Kraft, J Frey, T Moeller, M Albrecht, M Grothof, B Schink, H Hess, and B Buxbaum. 3d-camera of high 3d-frame rate, depth-resolution and background light elimination based on improved pmd (photonic mixer device)-technologies. *OPTO, Nuernberg, May*, 2004. 1.2

[68] M. Krainin, B. Curless, and D. Fox. Autonomous generation of complete 3d object mod-

els using next best view manipulation planning. In *2011 IEEE International Conference on Robotics and Automation*, pages 5031–5037, May 2011. doi: 10.1109/ICRA.2011.5980429. 5.9

[69] Hiroyuki Kubo, Suren Jayasuriya, Takafumi Iwaguchi, Takuya Funatomi, Yasuhiro Mukaigawa, and Srinivasa G Narasimhan. Acquiring and characterizing plane-to-ray indirect light transport. In *2018 IEEE International Conference on Computational Photography (ICCP)*, pages 1–10. IEEE, 2018. 1.4

[70] Robert Lange and Peter Seitz. Solid-state time-of-flight range camera. *IEEE Journal of Quantum Electronics*, 37(3):390–397, Mar 2001. doi: 10.1109/3.910448. 1.2, 2.2, 3.3, 3.3

[71] Dirk Langer and Martial Hebert. Building qualitative elevation maps from side scan sonar data for autonomous underwater navigation. In *Robotics and Automation, 1991. Proceedings., 1991 IEEE International Conference on*, pages 2478–2483. IEEE, 1991. 1.2

[72] T. Langner, D. Seifert, B. Fischer, D. Goehring, T. Ganjineh, and R. Rojas. Traffic awareness driver assistance based on stereovision, eye-tracking, and head-up display. In *2016 IEEE International Conference on Robotics and Automation (ICRA)*, pages 3167–3173, May 2016. doi: 10.1109/ICRA.2016.7487485. 5.9

[73] Douglas Lanman and Gabriel Taubin. Build your own 3d scanner: 3d photography for beginners. In *ACM SIGGRAPH 2009 Courses*, page 8. ACM, 2009. 4.3

[74] Larry Li. Time-of-flight camera–an introduction. *Technical white paper*, 1(SLOA190B), 2014. 4.2.3

[75] Lei Li, Qin Zhang, and Danfeng Huang. A review of imaging techniques for plant phenotyping. *Sensors*, 14(11):20078–20111, 2014. 6.3.1

[76] Patrick Lichtsteiner, Christoph Posch, and Tobi Delbruck. A 128x128 120 dB 15us latency asynchronous temporal contrast vision sensor. *IEEE journal of solid state circuits*, 43(2): 566–576, 2008. 4.5

[77] Steven Lovegrove, Alonso Patron-Perez, and Gabe Sibley. Spline fusion: A continuous-time representation for visual-inertial fusion with application to rolling shutter cameras. In *BMVC*, 2013. 6.2.1

[78] Luminar. Luminar Technology, 2018. URL `https://www.luminartech.com/technology/index.html`. 1.2

[79] Fangchang Ma, Guilherme Venturelli Cavalheiro, and Sertac Karaman. Self-supervised sparse-to-dense: self-supervised depth completion from lidar and monocular camera.

arXiv preprint arXiv:1807.00275, 2018. 6.2.2

[80] Hélène Macher, Tania Landes, and Pierre Grussenmeyer. From point clouds to building information models: 3d semi-automatic reconstruction of indoors of existing buildings. *Applied Sciences*, 7(10):1030, 2017. 6.3.1

[81] Minsky Marvin. Microscopy apparatus, December 19 1961. US Patent 3,013,467. 1.3, 4.5, 5.2

[82] Miquel Massot-Campos and Gabriel Oliver-Codina. Optical Sensors and Methods for Underwater 3D Reconstruction. *Sensors*, 15(12):31525–31557, Dec 2015. doi: 10.3390/s151229864. 6.3.1

[83] EA McLean, HR Burris, and MP Strand. Short-pulse range-gated optical imaging in turbid water. *Applied optics*, 34(21):4343–4351, 1995. 1.3

[84] Dan McLeod, John Jacobson, Mark Hardy, and Carl Embry. Autonomous inspection using an underwater 3d lidar. In *Oceans-San Diego, 2013*, pages 1–8. IEEE, 2013. 1.4

[85] Christoph Mertz, Sanjeev J Koppal, Solomon Sia, and Srinivasa Narasimhan. A low-power structured light sensor for outdoor scene reconstruction and dominant material identification. In *2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pages 15–22. IEEE, 2012. 1.3, 2.1

[86] Hans Moravec and Alberto Elfes. High resolution maps from wide angle sonar. In *Robotics and Automation. Proceedings. 1985 IEEE International Conference on*, volume 2, pages 116–121. IEEE, 1985. 1.2

[87] Don Murray and James J Little. Using real-time stereo vision for mobile robot navigation. *autonomous robots*, 8(2):161–171, 2000. 1.2

[88] Srinivasa G. Narasimhan, Shree K. Nayar, Bo Sun, and Sanjeev J. Koppal. Structured light in scattering media. In *Proceedings of the Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, ICCV '05, pages 420–427, Washington, DC, USA, 2005. IEEE Computer Society. ISBN 0-7695-2334-X-01. doi: 10.1109/ICCV.2005.232. 1.3, 5.2

[89] Thierry Oggier, Michael Lehmann, Rolf Kaufmann, Matthias Schweizer, Michael Richter, Peter Metzler, Graham Lang, Felix Lustenberger, and Nicolas Blanc. An all-solid-state optical range camera for 3d real-time imaging with sub-centimeter depth resolution (swissranger). In *Optical Design and Engineering*, volume 5249, pages 534–546. International Society for Optics and Photonics, 2004. 1.2

[90] Thierry Oggier, Rolf Kaufmann, Michael Lehmann, Bernhard Buttgen, Simon Neukom,

Michael Richter, Matthias Schweizer, Peter Metzler, Felix Lustenberger, and Nicolas Blanc. Novel pixel architecture with inherent background suppression for 3d time-of-flight imaging. In *Electronic Imaging 2005*, pages 1–8. International Society for Optics and Photonics, 2005. 2.1

[91] 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 Trans. Graph.*, 33(4):87:1–87:11, July 2014. ISSN 0730-0301. doi: 10.1145/2601097.2601103. 2.1

[92] Matthew O'Toole, John Mather, and Kiriakos N Kutulakos. 3d shape and indirect appearance by structured light transport. In *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*, pages 3246–3253. IEEE, 2014. 1.4, 2.1

[93] Matthew O'Toole, Supreeth Achar, Srinivasa G. Narasimhan, and Kiriakos N. Kutulakos. Homogeneous codes for energy-efficient illumination and imaging. *ACM Transactions on Graphics*, 34(4):35:1–35:13, Jul 2015. doi: 10.1145/2766897. 1.3, 1.3, 1.4, 1.4, 1.5, 2.1, 5.2

[94] Paul Ozog, Giancarlo Troni, Michael Kaess, Ryan M. Eustice, and Matthew Johnson-Roberson. Building 3D mosaics from an Autonomous Underwater Vehicle, Doppler velocity log, and 2D imaging sonar. *Proceedings - IEEE International Conference on Robotics and Automation*, 2015-June(June):1137–1143, 2015. doi: 10.1109/ICRA.2015.7139334. 6.3.1

[95] A. Payne, A. Daniel, A. Mehta, B. Thompson, C. S. Bamji, D. Snow, H. Oshima, L. Prather, M. Fenton, L. Kordus, P. O'Connor, R. McCauley, S. Nayak, S. Acharya, S. Mehta, T. Elkhatib, T. Meyer, T. O'Dwyer, T. Perry, V. Chan, V. Wong, V. Mogallapu, W. Qian, and Z. Xu. 7.6 a 512424 cmos 3d time-of-flight image sensor with multi-frequency photo-demodulation up to 130mhz and 2gs/s adc. In *2014 IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC)*, pages 134–135, Feb 2014. doi: 10.1109/ISSCC.2014.6757370. 3.5.4, 3.6

[96] Photometrics. Keep the Noise Down! Technical report, Photometrics, 2010. URL `https://www.photometrics.com/resources/technotes/pdfs/snr.pdf`. 3.2.1

[97] Photometrics. Read Noise Calculator. Technical report, Photometrics, 2011. URL `https://www.photometrics.com/resources/whitepapers/pdfs/read-noise-calculator.pdf`. 3.2.1

[98] William H Press, Saul A Teukolsky, William T Vetterling, and Brian P Flannery. *Numerical recipes in C++*, volume 3. Cambridge University Press, 2007. 5.3

[99] Quanergy. Q-guard: The ultimate lidar-based intrusion detection solution, 2019. 6.3.1

[100] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 779–788, 2016. 5.9

[101] P Rives and JJ Borrelly. Underwater pipe inspection task using visual servoing techniques. In *Proceedings of the 1997 IEEE/RSJ International Conference on Intelligent Robot and Systems. Innovative Robotics for Real-World Applications. IROS '97*, volume 1, pages 63–68. IEEE, 1997. doi: 10.1109/IROS.1997.648999. 6.3.1

[102] Amin Sarafraz, Shahriar Negahdaripour, and Yoav Y Schechner. Enhancing images in scattering media utilizing stereovision and polarization. In *Applications of Computer Vision (WACV), 2009 Workshop on*, pages 1–8. IEEE, 2009. 1.3

[103] Ashutosh Saxena, Sung H Chung, and Andrew Y Ng. Learning depth from single monocular images. In *Advances in neural information processing systems*, pages 1161–1168, 2006. 5.9

[104] Daniel Scharstein and Richard Szeliski. High-accuracy stereo depth maps using structured light. In *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, volume 1, pages I–I. IEEE, 2003. 1.2

[105] Jürgen Schröder, Bernd Heid, Florian Neuhaus, Matthias Kasser, Christoph Klink, and Simon Tatomir. Fast forwarding last-mile delivery - implications for the ecosystem. *McKinsey & Company*, 2018. 6.3.1

[106] Brent Schwarz. Lidar: Mapping the world in 3d. *Nature Photonics*, 4(7):429, 2010. 5.1

[107] S. Shrestha, F. Heide, W. Heidrich, and G. Wetzstein. Computational imaging with multi-camera time-of-flight systems. *ACM Trans. Graph. (SIGGRAPH)*, 2016. 2.1

[108] *microScan3 - The new generation of safety laser scanners*. SICK, 2 2018. 5.2

[109] Nick Spooren, Bert Geelen, Klaas Tack, Andy Lambrechts, Murali Jayapala, Ran Ginat, Yaara David, Eyal Levi, and Yoav Grauer. Rgb-nir active gated imaging. In *Electro-Optical and Infrared Systems: Technology and Applications XIII*, volume 9987, page 998704. International Society for Optics and Photonics, 2016. 1.4

[110] Roger Stettner. Compact 3d flash lidar video cameras and applications. In *Laser Radar Technology and Applications XV*, volume 7684, page 768405. International Society for Optics and Photonics, 2010. 1.2

[111] Shuochen Su, Felix Heide, Gordon Wetzstein, and Wolfgang Heidrich. Deep end-to-end time-of-flight imaging. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6383–6392, 2018. 6.2.2

[112] R. Tadano, A. K. Pediredla, and A. Veeraraghavan. Depth selective camera: A direct, on-chip, programmable technique for depth selectivity in photography. In *2015 IEEE International Conference on Computer Vision (ICCV)*, pages 3595–3603, Dec 2015. doi: 10.1109/ICCV.2015.410. 4.2.3, 5.2

[113] Zaid Tasneem, Dingkang Wang, Huikai Xie, and Koppal Sanjeev. Directionally controlled time-of-flight ranging for mobile sensing platforms. In *Proceedings of Robotics: Science and Systems*, Pittsburgh, Pennsylvania, June 2018. doi: 10.15607/RSS.2018.XIV.011. 1.3, 5.2, 6.3

[114] Seth Teller, Matthew R Walter, Matthew Antone, Andrew Correa, Randall Davis, Luke Fletcher, Emilio Frazzoli, Jim Glass, Jonathan P How, Albert S Huang, et al. A voice-commandable robotic forklift working alongside humans in minimally-prepared outdoor environments. In *2010 IEEE International Conference on Robotics and Automation*, pages 526–533. IEEE, 2010. 6.3.1

[115] Möller Tobias, Kraft Holger, Frey Jochen, Albrecht Martin, and Lange Robert. Robust 3D Measurement with PMD Sensors. *Range Imaging Day, Zürich*, 7(Section 5):8, 2005. doi: 10.1.1.132.5821. 3.3

[116] Manuel Vázquez-Arellano, Hans Griepentrog, David Reiser, and Dimitris Paraforos. 3-d imaging systems for agricultural applicationsa review. *Sensors*, 16(5):618, 2016. 6.3.1

[117] Velodyne. Velodyne HDL-64E Manual, 2018. URL `http://www.velodynelidar.com/lidar/products/manual/HDL-64EManual.pdf`. 1.2

[118] Velodyne. Velodyne veladome, 2019. URL `https://velodynelidar.com/veladome.html`. 6.3.1

[119] Jian Wang, Joseph Bartels, William Whittaker, Aswin C Sankaranarayanan, and Srinivasa G Narasimhan. Programmable triangulation light curtains. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 19–34, 2018. 1.5, 4.2.1, 4.2.2

[120] Qi Wang, Stephen Nuske, Marcel Bergerman, and Sanjiv Singh. Automated crop yield estimation for apple orchards. In *Experimental robotics*, pages 745–758. Springer, 2013. 6.3.1

[121] Waymo. Waymo laser bear honeycomb, 2019. URL `https://waymo.com/lidar/`. 6.3.1

[122] Stefan B Williams, Paul Newman, Gamini Dissanayake, and Hugh Durrant-Whyte. Autonomous underwater simultaneous localisation and map building. In *Robotics and Automation, 2000. Proceedings. ICRA'00. IEEE International Conference on*, volume 2, pages 1793–1798. IEEE, 2000. 1.2

[123] Jiahua Wu, Robin Tillett, Nigel McFarlane, Xiangyang Ju, J Paul Siebert, and Paddy Schofield. Extracting the three-dimensional shape of live pigs using stereo photogrammetry. *Computers and Electronics in Agriculture*, 44(3):203–222, 2004. 6.3.1

[124] J. Yuh, Giacomo Marani, and D. Richard Blidberg. Applications of marine robotic vehicles. *Intelligent Service Robotics*, 4(4):221, Jul 2011. doi: 10.1007/s11370-011-0096-5. 6.3.1

[125] Ji Zhang and Sanjiv Singh. LOAM: Lidar Odometry and Mapping in Real-time. In *Robotics: Science and Systems X*. Robotics: Science and Systems Foundation, jul 2014. ISBN 9780992374709. doi: 10.15607/RSS.2014.X.007. 6.2.1

[126] Z. Zhang. Microsoft kinect sensor and its effect. *IEEE MultiMedia*, 19(2):4–10, Feb 2012. ISSN 1070-986X. doi: 10.1109/MMUL.2012.24. 3.5.4, 5.1

[127] Zhengyou Zhang. A flexible new technique for camera calibration. *IEEE Transactions on pattern analysis and machine intelligence*, 22, 2000. 4.3, 5.6.1

[128] Yin Zhou and Oncel Tuzel. Voxelnet: End-to-end learning for point cloud based 3d object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4490–4499, 2018. 1.3