High Resolution 2D Imaging and 3D Scanning with Line Sensors

Submitted in partial fulfillment of the requirements for
the degree of
Doctor of Philosophy
in
Department of Electrical and Computer Engineering

Jian Wang

B.S., Automation, Xi'an Jiaotong University
M.S., Automation, Univ. of Sci. and Tech. of China
M.S., Elec. & Comp. Eng., Carnegie Mellon University

Carnegie Mellon University
Pittsburgh, PA

August, 2018

Acknowledgments

First and foremost, I would like to express my utmost gratitude to my advisors, Aswin Sankaranarayanan and Srinivasa Narasimhan. Without their guidance, support and encouragement, this dissertation would not exist. It is the constant discussion with them and their high demands that make the work presented in this dissertation high-quality and have the opportunity of oral presentation in international conferences. I feel tremendously lucky and could not have asked for better advisors.

Furthermore, I feel grateful to my committee members, Vijayakumar Bhagavatula and Mohit Gupta, for going through the entire dissertation and giving me critical comments and suggestions, which make the dissertation more complete and better. I am especially grateful to Mohit Gupta for his great ideas and advices during our collaborations.

During the internships in Microsoft Research, Huawei Research, and Google Research, I also had the pleasure working and discussing with Yasuyuki Matsushita, Jinwei Gu, Jiawen Chen and Jon Barron. I learned a lot from their distinctive points of view in solving research problems. I would also like to thank Ioannis Gkioulekas for spending time sitting with me for many interesting discussions.

I feel lucky to have had terrific colleagues in Image Science Lab and Illumination and Imaging Lab. Exchanging views on how to tie camera pieces together with Joe and Supreeth and how to turn screws with Vishwa, Yi and Rick, making up jokes with Kuldeep and Robert, gossip with Anqi and Byeongjoo, practicing dialect with Tiancheng, Shumian and Canyi, facing occasional difficulties of PhD life together with Chia-Yin, Harry, Suren, Minh and Chao, and enjoying Thorlabs snacks made for a unique doctoral study experience that I will miss always.

Finally, I would like to thank my parents Yousheng and Yonglan, my sister Ping and especially my girlfriend Yang. Without their support and encouragement, I could not have finished this work.

This dissertation was supported in part by an NSF grant CCF-1117939, an ONR grant N00014-15-1-2358, an ONR DURIP Award N00014-16-1-2906, DARPA REVEAL Co-operative Agreement HR0011-16-2-0021, CIT Deans Tuition Fellowship and Liang Ji-Dian Fellowship.

Abstract

In the past few decades, imaging technology has made great strides. From high resolution sensors for photography to 3D scanners in autonomous driving, imaging has become one of the key drivers of the modern society. However, there are still many scenarios where the traditional methods of imaging are woefully inadequate. Examples include high-resolution non-visible light imaging, 3D scanning in the presence of strong ambient light, and imaging through scattering media. In these scenarios, the two classical solutions of single-shot imaging using 2D sensors and point scanning using photodiodes have severe shortcomings in terms of cost, measurement rate and robustness to non-idealities in the imaging process.

The goal of this dissertation is the design of computational imagers that work under traditionally difficult conditions by providing the robustness and economy of point scanning systems along with the speed and resolution of conventional cameras. In order to achieve this goal, we use line sensors or 1D sensors and make three contributions in this dissertation. The first contribution is the design of a line sensor based compressive camera (LiSens) which uses a line sensor and a spatial light modulator for 2D imaging. It can provide a measurement rate that is equal to that of a 2D sensor but with only a fraction of the number of pixels. The second contribution is the design of a dual structured light (DualSL) system which uses a 1D sensor and a 2D projector to achieve 3D scanning with same resolution and performance as traditional structured light system. The third contribution is the design of programmable triangulation light curtains (TriLC) for proximity detection by rotating a 1D sensor and a 1D light source in synchrony. This device detects the presence of objects that intersect a programmable virtual shell around itself. The shape of this virtual shell can be changed during operation and the device can perform under strong sunlight as well as in foggy and smoky environments. We believe that the camera architectures proposed in this dissertation can be used in a wide range of applications, such as autonomous driving cars, field robotics, and underwater exploration.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

"... to walk a thin line."

Over the last century, imaging has come of age. There have been many successes in diverse domains that rely fundamentally on imaging, including astronomy (the Hubble telescope), microscopy (confocal imaging [55], super-resolution microscopy [60]), consumer photography (cellphone photography), entertainment systems (Microsoft Kinect [105], Intel RealSense [46]) and autonomous driving (LIDAR systems). It is not an exaggeration to suggest that imaging is inarguably one of the key drivers of the modern society.

Imaging can be broadly divided into two categories: passive and active. Passive imaging uses an ambient light source to image a scene. The most common design of a passive imager or a camera is a combination of a lens and a 2D sensor, such as that found in an SLR or a cellphone camera. The material of choice for sensing in visible wavebands of light is silicon which is both abundant and endowed with efficient manufacturing technologies. As a consequence, not just has the resolution of these cameras gotten higher and higher over the last few decades, their cost has constantly been decreasing. In contrast, high-resolution imaging of non-visible light is not practical because non-silicon based 2D sensors with large pixel count are often extremely expensive [25]. For example, a megapixel short-wave infrared (SWIR) 2D sensor, made of InGaAs, can be as expensive as hundreds of thousands of dollars.

Active imaging, often used for 3D scanning and seeing through scattering media, works by controlling the light source in the scene. Active control of the light source is very desirable in estimating correspondences in the context of triangulation as well as direct estimation of depth via time-of-flight (TOF) measurements. However, even in the context of active imaging, it is not always possible to control the scene's illumination in its entirety. Existing sources of light in the scene, referred to as ambient light, reduces the contrast of the active illuminant. Even in the absence of ambient light, the captured image could still be deteriorated by undesirable effects like inter-reflection, sub-surface scattering and volumetric scattering by the transmitting media;

these effects are referred to as global light. Strong ambient light sources, such as direct sunlight, are often orders of magnitude stronger than the active source and can easily submerge it. This is why neither the 1st generation Kinect (based on structured light) nor the 2nd generation (based on continuous-wave TOF (CW-TOF)) can work outdoors [30]. An effective method is to concentrate all the power of the active light source into a point, and perform point-by-point scanning. This is the operating principle underlying LIDAR which is effective even under strong sunlight. LIDAR works by emitting a short pulse of laser light to a scene point and a single detector measures its time of flight. Each scene point is sequentially scanned by spinning the laser and the detector in synchrony. Such methods that rely on point scanning are often too slow with measurement rates (defined as the number of measurements that a device can obtain in unit time) that are orders of magnitude smaller than systems that use 2D sensors. Global light, on the other hand, leads to loss of contrast due to volumetric scattering in medium like fog, smoke or turbid water. An example is in the field of microscopy where it is very difficult to image parts of the sample surrounded by organic tissue. One promising method for imaging through scattering media is confocal imaging [55]. It uses a single point light source and a single photodiode which are both focused at the same point with very shallow depth of field by the use of pinholes. The illumination intensity rapidly falls off above and below the point of focus as the beam converges and diverges and the aperture of the photodiode blocks any light emanating from regions away from the illuminated point which provides yet further attenuation of out-of-focus light, i.e. scattered light. Hence, the contrast is significantly improved. However, much like LIDAR, confocal imaging sequentially scans the sample one point at a time, and hence the scanning time is very long.

As we have seen earlier, the use of 2D sensors as well as single detectors have their relative strengths and weaknesses. 2D sensors have high measurement throughput, but are inefficient in global and ambient light as well as expensive in non-visible wavelengths. Single detector-based imaging is inexpensive and can significantly alleviate the effects of global and ambient light, but has very low measurement throughput. Clearly, there is a need for a middle ground that provides a graceful tradeoff between these capabilities. This dissertation provides such a tradeoff by envisioning a space of designs that rely on line sensors or 1D sensors that consist of an array of photodiodes arranged in a linear configuration. For passive imaging in non-visible wavebands, 2D sensors are expensive, but 1D sensors can be low-cost. For active imaging, devices that spread out light into the entire scene such as Kinect cannot be used outdoors, and while point scanning can work outdoors it is too slow. By building a device based on a 1D sensor and line scanning, we can ensure that it can be used outdoors as well as be operated with high measurement throughput. Likewise, for seeing through scattering media, line scanning can achieve selective imaging much like point scanning, and greatly speed up the scan time. Armed with these insights, we propose three devices:

- **LiSens** [96], for passive imaging, uses a 1D sensor and a spatial light modulator for 2D imaging with high resolution and frame rate. Under ideal conditions, LiSens can provide measurement rate equal to a 2D sensor, but with a fraction of the number of pixels, thereby providing an inexpensive alternative.
- **DualSL** [97], for active imaging via structured light, uses a 1D sensor and a 2D projector to achieve 3D scans with high spatial and temporal resolutions. Under ideal conditions, DualSL is no worse than a traditional structured light system in the ability to provide high scan rate and its ability to tolerate ambient and global light.
- **TriLC** [98], for active imaging for proximity detection, uses a rotating 1D sensor and 1D light source to detect pedestrians and vehicles. It can be used under strong sunlight, and can see through scattering media like fog or smoke.

Line sensors can also provide numerous other benefits, including a high fill factor, high speed readout, high spatial resolution, large pixel, and frame transfer. For special modality sensors, although 2D versions are expensive, their 1D versions are generally much more affordable. Because these proposed devices use 1D sensor, there is a wide range of sensor types to choose from. For example, LiSens can use 1D SWIR for SWIR imaging. DualSL can use 1D SWIR to measure 3D outdoors and through scattering media; it can use dynamic vision sensor (DVS) [51] to realize outdoor 3D high-speed imaging. TriLC can use 1D SWIR, CW-TOF, DVS or single-photon avalanche diode (SPAD) to achieve more advanced performance.

To better understand the contributions of this dissertation, it is worth taking a deep dive into the challenges faced in imaging and the designs of classical systems that address them.

1.1 Challenges

The challenges underlying imaging in passive and active regimes are fundamentally different and manifold. We discuss each of them separately.

1.1.1 Passive imaging in non-visible wavebands

The fundamental nature of our world can often be better revealed when we image in wavelengths beyond the visible light. For example, short-wave infrared (SWIR) the waveband of light between 900nm $\sim 3\mu$m is used for imaging in scattering media and at night. Similarly, mid-wave infrared (MWIR) the waveband between 3μm $\sim 5\mu$m is used for high contrast thermal imaging and Terahertz wave the waveband between 100μm $\sim$ 1mm is used for applications spanning surveillance, astronomy and microscopy (see Figure 1.1). However, imaging in these wavebands is prohibitively expensive due to the need for specialized sensors that have high cost-per-pixel

(a) SWIR for seeing through fog and night vision

(b) MWIR for high contrast thermal imaging

(c) THz for surveillance

Figure 1.1: **Applications of imaging in non-visible wavebands.** *Image source: Sensors Unlimited, Infiniti Electro-Optics, and dailymail.co.uk*

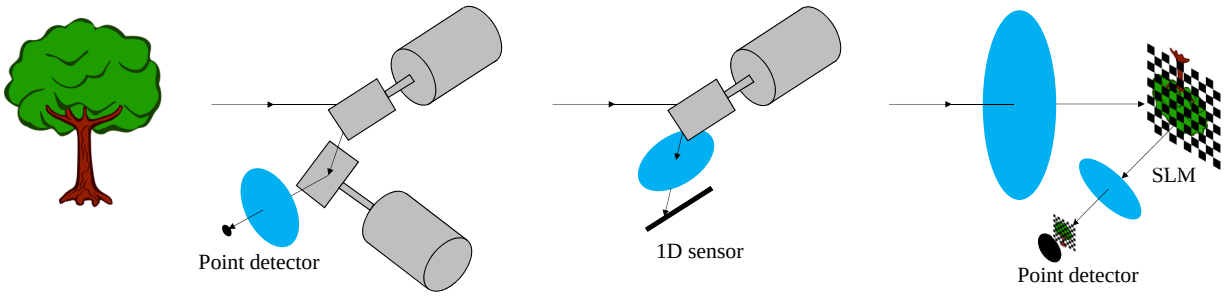
(see Table 1.1). For example, the typical cost-per-pixel for SWIR is \$0.1, MWIR \$0.1 and Terahertz $\$10^2 \sim 10^3$ [25], as opposed to visible light which is less than $\$10^{-6}$.

Let us take a look at the SWIR example to see why these sensors are expensive. Much like the silicon-based visible light sensor, the SWIR sensor is based on the photoelectric effect which transforms photons into electrons. Since the upper limit of the sensitivity of silicon to light is 1μm, other materials are needed to sense SWIR light, such as mercury cadmium telluride (MCT) or indium gallium arsenide (InGaAs) [74]. MCT requires strong cooling to raise the signal-to-noise ratio (SNR) to a suitable level, which is cumbersome and expensive. Although InGaAs does not require cooling, it is still expensive. The InGaAs photosensitive array, grown on an indium phosphide (InP) substrate, needs to be combined with a silicon-based CMOS readout circuit, which is complicated, time consuming as well as endowed with poor yields.

Single photodiode or line sensor-based scanning. An alternative to 2D sensors is to use a single photodiode along with a 2D scanning system, such as a galvomirror system as shown in Figure 1.2 (a). The galvomirror typically takes between 200μs and 1000μs to rotate a fraction of a degree to reach its next position and stabilize [92]. Even if the exposure time of the photodiode is short, less than a hundred microseconds, a 2D scanning system can only scan a few thousand points per second. A megapixel image therefore takes several hundred seconds which is unacceptably slow. Another alternative is to use a 1D sensor along with a 1D scanning system (see Figure 1.2 (b)) so that the entire system can scan thousands of lines per second which is marginally acceptable, but only achieves a few frames per second. Note that there are three aspects that affect the amount of light received regardless of the use of a photodiode or a 1D sensor. First, to ensure fast frame rate, the exposure time must be reduced to the microsecond time scale, while the typical exposure time of a 2D sensor in visible light is often in the millisecond

Spectral band	**Detector technology**	**Per-pixel price ($)**
mmW/THz	Multiple	$10^2 \sim 10^4$
LWIR	HgCdTe	< 10
	Bolometer	10^{-2}
MWIR	InSb/PbSe	10^{-1}
SWIR	MCT/InGaAs	10^{-1}
NIR/VIS/NUV	Si	$< 10^{-6}$
MUV	Si (thinned)	$< 10^{-3}$
EUV	Si-PIN/CdTe	$10^2 \sim 10^3$
Soft x ray	Si (thinned)	10^{-2}
	Si-PIN/CdTe	$10^2 \sim 10^3$
Hard x ray	Multiple	$10^2 \sim 10^4$

Table 1.1: **Per-pixel price for detectors of various wavebands.** *Table from [25]*



(a) Point detector with 2D scanning (b) 1D sensor with 1D scanning (c) Point detector with multiplexing

Figure 1.2: **Inexpensive alternative designs for 2D imaging in non-visible wavebands** (schematics not to scale).

time scale. Second, the focusing objective lens is generally placed behind the galvomirror. The advantage is that no matter how the galvomirror rotates, the single photodiode or line sensor is in focus. The disadvantage is that the light entrance pupil is limited by the size of the galvomirror which is at best a few millimeters in diameter [92]. Increasing the size of the galvomirror also increases its mass and inertia, which in turn makes it slower. Third, at any given moment, only one pixel or one row of pixels is collecting light, and light elsewhere is lost.

Single photodiode-based multiplexing. The single pixel camera (SPC) [21] uses a single photodiode and a spatial light modulator (SLM) for 2D multiplexing, which effectively solves the problem of light throughput. Figure 1.2 (c) shows schematic of SPC and in the following we explain how it works. The scene is first focused on the SLM through the objective lens, and then the light passing through the SLM is converged onto a single pixel through a convex lens. With changes of the mask on the SLM, the single pixel can get different measurements of the scene. After the single pixel obtains enough measurements, the scene can be reconstructed. The SLM can be a digital micromirror device (DMD), a liquid-crystal display (LCD), or a liquid crystal on silicon (LCoS). Since the fastest SLMs today can only change mask at tens of thousands times per second, we can only obtain megapixel images once every tens of seconds. If we want to speed up the acquisition, we would need to recover signals with fewer measurements than unknowns, or equivalently solve an under-determined linear system. Compressive sensing [8] provides this capability by exploiting the sparse nature of the sensed signal. The use of compressive sensing can reduce the number of required masks to a fraction of its original, but the reduction is often associated with loss of spatial detail especially at high compression. $10\times$ compression is a typical value for natural scenes as reported in [21, 76, 78]. This helps in improving the frame rate, but still several seconds per frame.

	Devices	LT	Frame rate	Strengths	Weaknesses
Photodiode	Point scanning	1	$\frac{5000}{RC}$	$	too slow
	SPC [21]	$\frac{RC}{2}$	$\frac{10^5}{RC}$	$, high LT	
1D sensor	Line scanning	R	$\frac{5000}{C}$	$$	slow, low LT
	LiSens (proposed)	$\frac{RC}{2}$	$\frac{10^5}{C}$	*$$, fast, high LT*	*slightly complicated hardware*
2D sensor	Staring array camera	RC	30	fast, high LT	$$$$$$

Table 1.2: **Camera architectures for passive imaging in non-visible wavebands.** LT refers to light throughput. R and C are the scene's row and column resolutions, respectively. Frame rate column are typical values when the galvomirror for the scanning systems can scan 5000 positions per second, SLM can change 10^4 masks per second, $10\times$ compressive sensing is used if applicable, and a 2D sensor can read out data at 30 frames per second (fps). Although 2D sensors have high bandwidth acquisition, they are very expensive. Single pixel-based 2D scanning system and 1D sensor-based 1D scanning system are inexpensive, but the light throughput and acquisition speed is too low. Compressive sensing-based single pixel camera is also inexpensive, and while the light throughput has been greatly improved, due to low rate of the SLM, it is still very slow.

The traditional designs for passive imaging are summarized in Table 1.2. In the table, R and C are the scene's row and column resolutions, respectively, and they can each be set equal to 1000 pixels for a megapixel image. The third column in the table is the amount of light throughput (LT). It is assumed here that the aperture size of the objective lens is the same for all devices (ignoring the fact that lens of the scanning systems is actually much smaller) and it considers the light within the field of view (FOV) (RC blocks) that each design can capture for a single exposure of a fixed time. Suppose here SPC uses Hadamard multiplexing [32], the single photodiode can receive half of the scene's light each time. The fourth column is the frame rate. For the scanning systems, we assume that the galvomirror used can scan 5000 positions per second [92]. For systems using SLM, we assume that SLM can change 10k masks per second, and the use of compressive sensing can accelerate the system 10 times. We use actual numbers here and in the following sections to ease the comparison of devices based on different mechanisms. The last two columns specify the advantages and disadvantages of each design.

1.1.2 Active imaging under ambient light

There are generally two approaches to use active illumination for measuring depth. The first approach is based on measurement of time of flight (TOF). Since the speed of light is known, as long as the time of light traveling from the source to the scene and back to the detector is known, depth can be measured. This can be further divided into two categories - impulse TOF (used in LIDAR) and continuous-wave TOF (used in Kinect). The second approach is triangulation, based on the fact that light travels in a straight line. By finding the directions from two perspectives of the same scene point, we can calculate the distance by finding the intersection. Examples include structured light [27] and active stereo. Here, active light is used to find the correspondence between the two perspectives robustly.

In sunny days, the total amount of power received at the surface of the Earth from the Sun can be as high as 1120W per square meter [63]. Compared to sunlight, the light source used in active light systems is many orders of magnitude weaker. For example, a large projector can output at most 10 watts of light, while a small portable laser projector has output power of only tens of milliwatts.

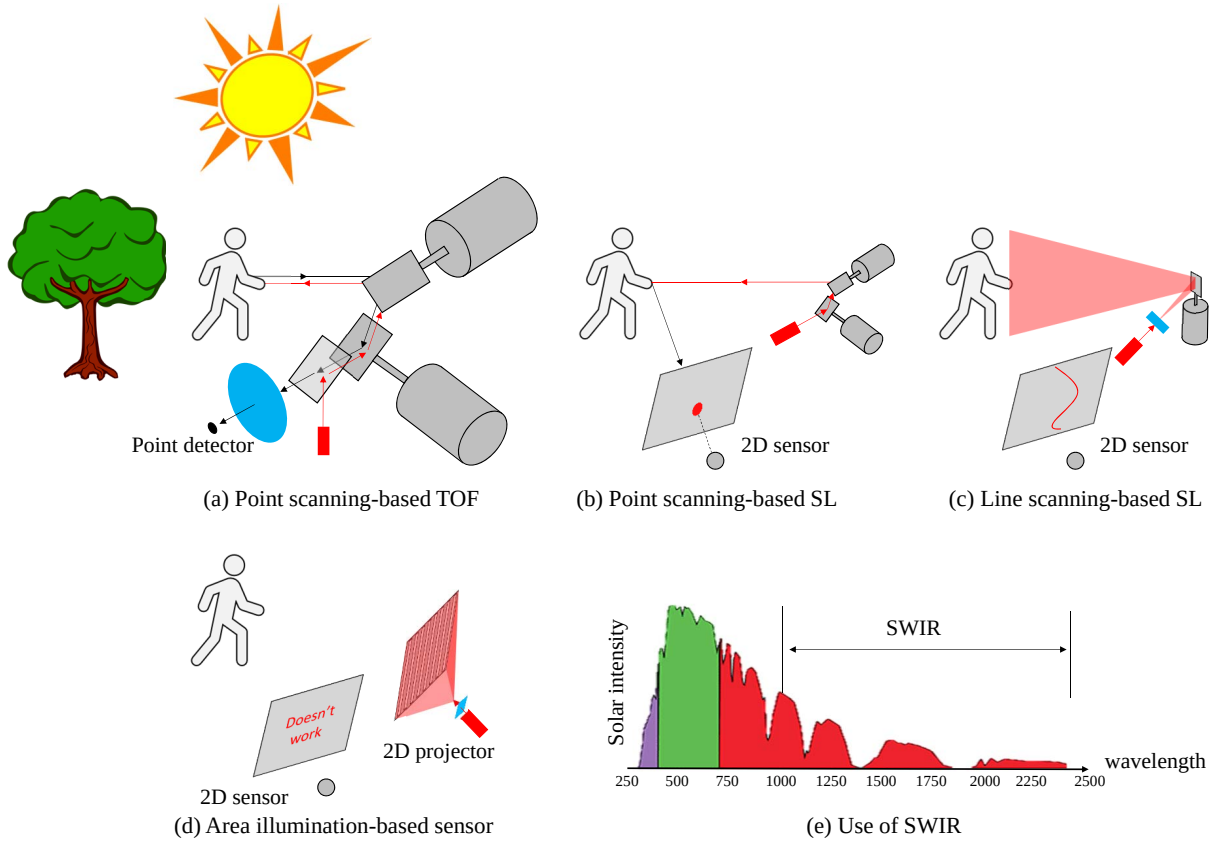
While a large amount of sunlight can be filtered out with use of a narrowband active source and a matching filter at the camera side, the leftover sunlight is still very strong. There are generally two categories of filters: absorption-based and interference-based. Absorption-based filters are sensitive to heat, have wide transition bands and low peak transmittance, which are all undesirable. In contrast, interference-based filters are much more efficient. However, they are very angle sensitive. Interference-based filters are designed with the assumption that the incident light is perpendicular to the filter. If the angle of the incident light is no longer perpendicular, the wavelength passing through the filter shifts. This means when selecting the filter, there is a tradeoff between the field of view and the bandwidth. The larger the field of view, the wider the bandwidth of the filter should have. Suppose the bandwidth of the filter used is 50nm, then nearly 90% of the sunlight is filtered out, and the remaining 10% of the sunlight still has about 100W per square meter.

Can sunlight simply be subtracted from the measurements since it is of constant intensity? The answer is no because all light sources are inherently random and the randomness is referred to as photon noise. The process of photons reaching the sensor can be modeled as a Poisson process. In any unit of time, the number of photons reaching the sensor will be different and will fluctuate around the mean. The stronger the light intensity, the greater the fluctuation. If the average brightness of the light source in the sensor is A, its standard deviation is $\sqrt{A}$. Thus if we take two images with and without active light source, and subtract the two, we get the signal caused by the active light source alone corrupted with twice the photon noise of the sun. Remember that the sunlight is significantly more powerful than the active light source and hence

the photon noise due to sunlight will be comparable to the active light source. This is further complicated due to $1/r^2$ fall off for the active light source. Even if the ambient light is not strong enough to let the active lighting system fail completely, it will also deteriorate its performance. Decoding failures can occur in structured light systems, and phase estimation can be very noisy in TOF systems. For these reasons, common 3D sensors like Kinect, whether it be based on structured light or based on TOF, do not work outdoors.

Concentrate and scan. One effective way to deal with sunlight is to concentrate the light from the active light source to a point and then perform point scanning. Suppose that active light source, with overall intensity P units, illuminates $R \times C$ blocks (the illuminated scene is captured by a $R \times C$ resolution camera), the brightness of the scene point (or block) due to the active light source is $\frac{P}{RC}$ and the brightness due to ambient light is A which contributes to photon noise. The SNR of the intensity measurement is approximately $\frac{P}{RC\sqrt{A}}$ [30]. After the active light source converges to illuminate only one block, the brightness due to active light source becomes P and the SNR increases to $\frac{P}{\sqrt{A}}$ which is RC times greater than before and hence the system works much more robustly. The weakness of point scannning, as stated before, is that it is very slow. If the system uses the TOF principle like LIDAR, uses galvomirror as scanning mechanism as shown in Figure 1.3 (a), and aims to obtain depth of RC points (with no loss of resolution), the frame rate is only $\frac{5000}{RC}$. If the system is structured light with point scanning as shown in Figure 1.3 (b) and uses a 2D camera which reads 30 frames per second, its frame rate is even less, only $\frac{30}{RC}$. The middle ground between full frame illumination and point scanning is line scanning. We concentrate the light previously scattered on the $R \times C$ blocks into a column. The brightness due to the active light source becomes $\frac{P}{R}$, and the SNR is $\frac{P}{R\sqrt{A}}$ which increases by R times. It can be guaranteed to work within a certain distance under sunlight, and each time the system can obtain depth of points of one column which is much faster than point scanning. Structured light line scanning is such a method (see Figure 1.3 (c)), but since the entire 2D image is read out every time, the frame rate is still slow, only $\frac{30}{C}$.

SWIR. SWIR waveband active illumination is potential to be used for 3D sensing under sunlight. Figure 1.3 (e) shows the spectral distribution of the solar energy. It reaches a peak in the visible light region, is also relatively strong in the near-infrared region which many commercial 3D cameras use and becomes much weaker in the longer wavelength region. In the SWIR band, solar energy of some wavelengths is close to zero due to absorption by the atmosphere as they reach the Earth's surface. In addition, the light in the SWIR band is much safer than visible and near-infrared light and thus the device can use more power to increase SNR. SWIR light cannot effectively focus on the retina due to the strong absorption by the water content in the eye [87].



(a) Point scanning-based TOF (b) Point scanning-based SL (c) Line scanning-based SL

(d) Area illumination-based sensor (e) Use of SWIR

Figure 1.3: **Active 3D sensors under strong sunlight** (schematics not to scale).

Its maximum eye safety power can be 100 to 1000 times that of near-infrared light [37]. From the two points mentioned above, it can be expected that the active imaging system with use of SWIR light can perform reliably, but the SWIR sensor is too expensive and the spatial resolution is much lower than their visible light counterpart.

The above designs are summarized in Table 1.3. We define light spread ratio as the light power at a single scene point from the light source with power 1. The table also shows SNR of intensity measurement and frame rate of each design.

1.1.3 Active imaging under global light

Observing the objects through scattering medium is essential in many scenarios. For example, seeing the road signs or other cars in foggy or smoggy days is very important for safe transportation, seeing in muddy water is helpful for undersea exploration, and seeing through skin can help diagnose many diseases. When the scattering medium is illuminated, the camera not only receives the light reflected by the object but receives a large amount of light scattered by the

	LSR	SNR	Work outdoors?	Frame rate	Devices
Point scanning	1	$\frac{P}{\sqrt{A}}$	Yes	PD: $\frac{5000}{RC}$	LIDAR
				2D sensor: $\frac{30}{RC}$	SL with point scanning
Line scanning	$\frac{1}{R}$	$\frac{P}{R\sqrt{A}}$	Yes	1D sensor: $\frac{5000}{C}$	*TriLC (proposed)*
				2D sensor: $\frac{30}{C}$	SL with line scanning
Full frame	$\frac{1}{RC}$	$\frac{P}{RC\sqrt{A}}$	No	$\frac{30}{\log C} \sim 30$	SL with full frame patterns
				30	Kinect based on TOF
SWIR	-	$P\uparrow, A\downarrow$	Yes	-	*DualSL (proposed)*

Table 1.3: **Performance of active 3D sensors under strong sunlight.** LSR refers to light spread ratio. PD refers to point detector. SL referes to structured light. R and C are the scene's row and column resolutions, respectively. Frame rate column are typical values when the galvomirror for the scanning systems can scan 5000 positions per second and a 2D sensor can read out images at 30 fps. For 3D sensing under strong sunlight, full frame illumination methods which spread out light into the entire scene do not work. Point scanning systems that concentrate the light do work but their scanning speed is too slow. Line scanning is a middle ground solution, but structured light line scanning is still slow. Another potential method is to use SWIR because the SWIR content in sunlight is low, but SWIR-related sensors are very expensive.

medium itself as shown in Figure 1.4 (a). This greatly reduces the contrast of the image. This problem is exacerbated by the exponential drop of the reflected light from the object on the way to the camera.

Selective imaging. An effective way to see through scattering medium is selective imaging, which is to design the system so that most of the scattered light is blocked while most of the light from the scene can pass in image formation. A prominent example in the field of microscopy is scanning confocal microscopy [55] (see Figure 1.4 (b)). Confocal imaging uses a point light source and a photodiode with small apertures. Both of the apertures are focused on the same sample point with very shallow depth of field. The illumination intensity falls off rapidly to the left and right of the point of focus as the beam converges and diverges, thus reducing scattered light which happens out of the point being examined. In addition, the aperture of the photodiode blocks any light emanating from regions away from the illuminated point which provides yet further attenuation of out-of-focus light, i.e. scattered light. This method is very effective, but due to point scanning, its speed is greatly limited.

SWIR. SWIR is a potentially effective waveband to image through scattering media, especially when 3D measurements are needed. Its wavelength is longer than visible light and thus it has less scattering.

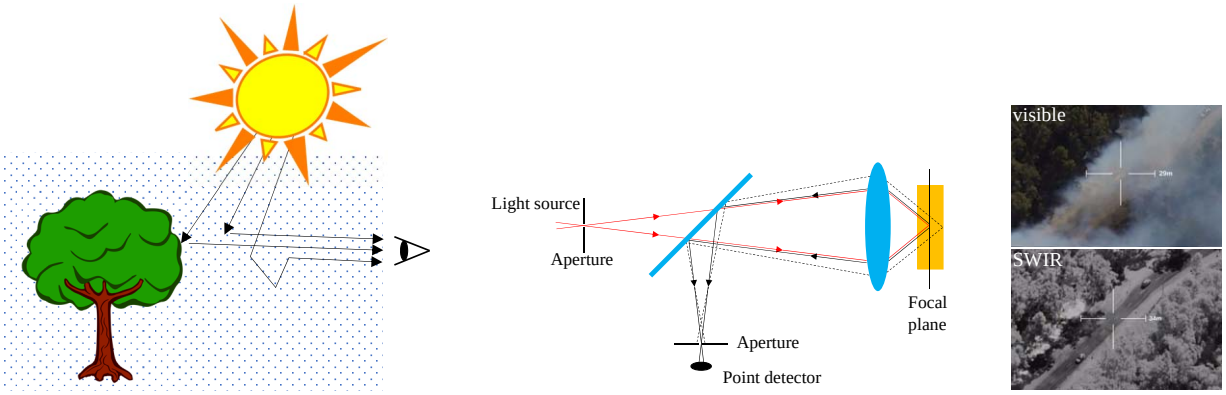
Scattering is a function of the wavelength of light and the size of a particle which scatters it. The brightness of a scene point after passing through scattering media of thickness d is

$$E(d) = E_0 e^{-\beta d} \tag{1.1}$$

where E_0 is the starting value, β is the scattering coefficient, and its value is

$$\beta = \frac{k}{\lambda^\gamma}, \tag{1.2}$$

where k is a constant, γ is $1 \sim 3$ when the wavelength of light λ is much greater than the scattering particles, named as Rayleigh scattering, and γ is 0 when the wavelength of light is comparable to the scattering particles, named as Mie scattering [12]. Dust, smoke and haze are typically composed of particles whose radii range from 0.1μm to 10μm and, hence, SWIR, the waveband ranging from 0.9μm to 3μm, is capable of imaging at higher contrast because it can pass some of the particles (see Figure 1.1 (a) and Figure 1.4 (c)). However, the sensors associated with it are too expensive. The above strategies are summarized in Table 1.4.



(a) Problem with imaging in scattering medium

(b) Point scanning based selective imaging: confocal imaging

(c) Use of SWIR

Figure 1.4: **Active imaging to see through scattering medium.** (b) is adapted from [28].

1.2 Contributions

The goal of this dissertation is the co-design of computations and imaging architectures for novel sensors that work under traditionally difficult conditions with both the robustness and the economy of point scanning systems and the speed and the resolution of common cameras. In order to achieve this goal, we make the following contributions in this dissertation:

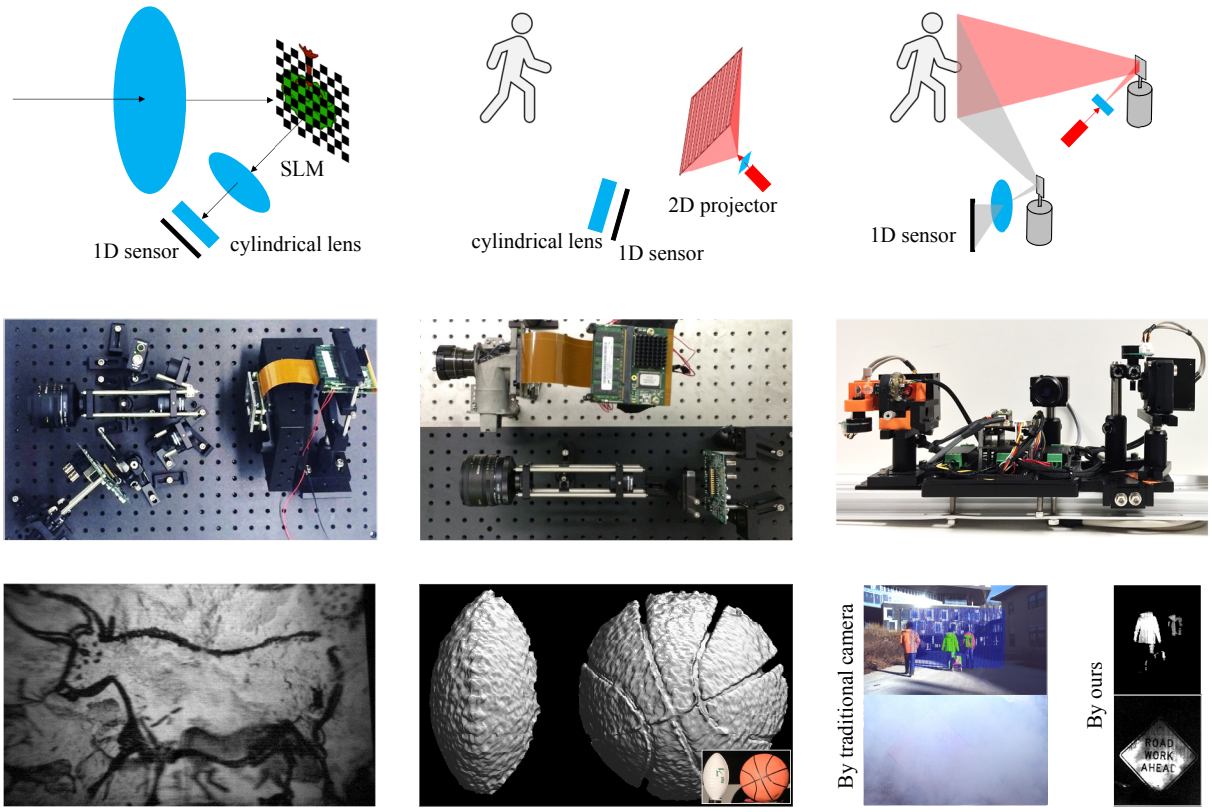
- **Line-Sensor-based compressive camera (LiSens)** [96] (Chapter 2): Traditional passive imaging devices in the non-visible wavebands are expensive, slow, and have low light throughput (see Table 1.2). We propose the LiSens camera, a novel spatial multiplexing device, that replaces the photodetector in the SPC with a line-sensor. We study the factors that determine the measurement rate for spatial multiplexing cameras (SMC) and show that increasing the number of pixels in the device increases the measurement rate, but there is an optimum number of pixels (typically, few thousands) beyond which the measurement rate does not increase. A key result that we derive is that the measurement rate of LiSens camera is equal to that of a 2D full frame sensor and this is achieved with a tiny fraction of the number of pixels; hence, we achieve an operating point in terms of spatial and temporal resolutions that is identical to a full frame sensor but with a fraction of cost. We illustrate the optical architecture underlying LiSens, build a prototype, and demonstrate results of a range of indoor and outdoor scenes (see Figure 1.5 (a)). LiSens delivers on the promise of SMCs: imaging at a megapixel resolution, at video rate, using an inexpensive low-resolution sensor.
- **Dual Structured Light 3D using a 1D Sensor (DualSL)** [97] (Chapter 3): Structured light-based 3D reconstruction methods often illuminate a scene using patterns with 1D translational symmetry such as stripes, Gray codes or sinusoidal phase shifting patterns. These patterns are

	Selective imaging	Frame rate	Devices
Point scanning	Yes	$\frac{5000}{RC}$	Confocal imaging
Line scanning	Yes	$\frac{5000}{C}$	*TriLC (proposed)*
Full frame	No	-	-
SWIR	-	$\frac{30}{\log C} \sim 30$	*DualSL (proposed)*

Table 1.4: **Strategies of active imaging to see through scattering medium.** R and C are the scene's row and column resolutions, respectively. Frame rate column are typical values when the galvomirror for the scanning systems can scan 5000 positions per second and a 2D sensor can read data at 30 fps. Selective imaging such as point scanning-based confocal imaging can achieve very good results, but it is too slow. SWIR has the potential, but the related sensors are expensive.

decoded using images captured by a traditional 2D sensor. In this work, we present a novel structured light approach that uses a 1D sensor with simple optics and no moving parts to reconstruct scenes with the same acquisition speed as a traditional 2D sensor. While traditional methods compute correspondences between columns of the projector and 2D camera pixels, our approach computes correspondences between columns of the 1D camera and 2D projector pixels. Such a device is the optical dual of traditional structured light system hence the name DualSL. DualSL has performance that is similar to traditional structured light system when imaging in the presence of ambient light and global light. Further, the use of a 1D sensor provides significant advantages in many applications that operate in SWIR (see Table 1.3 and 1.4) or require dynamic vision sensors (DVS), where a 2D sensor is prohibitively expensive and difficult to manufacture. Figure 1.5 (b) shows our hardware prototype and 3D structure of two balls acquired by it.

- **Programmable triangulation light curtains (TriLC)** [98] (Chapter 4): There are many scenarios that do not require a full-blown 3D depth sensor. For example, a vehicle on a road or a robot in the field only need to detect potential collisions or monitor its blind spot which can be achieved by monitoring if any object comes within its near proximity — a task that is easier than full depth scanning. We introduce a novel device that monitors the presence of objects on a virtual shell near the device, which we refer to as a light curtain. Our prototype for generating light curtains works by rapidly rotating a line sensor and a line laser, in synchrony (see schematic in Figure 1.5 (c)). Table 1.3 shows the comparison among various devices. TriLC works robustly under sunlight due to concentrating light into a line and has fast frame rate due to use of a line sensor and act as a proximity sensor. The device is capable of generating light



(a) LiSens (b) DualSL (c) TriLC

Figure 1.5: **Contributions of this dissertation**: LiSens [96], DualSL [97] and TriLC [98]. The first row shows schematics (not to scale), second rows shows our hardware prototype, and third row shows results of each.

curtains of various shapes with a range of 20∼30m in sunlight, 40m under cloudy skies and 50m indoors. It also has additional benefit in terms of improving visibility in fog because it can perform selective imaging much like point scanning (see Table 1.4). We build the hardware prototype and showcase the potential of light curtains using a range of real-world scenarios. Figure 1.5 (c) shows two results. The first result is detection of people passing a planar curtain under strong sunlight. The light curtain is rendered in blue and detections are rendered in green in traditional camera's image. The second result is seeing through smoke. Our device can see the road sign even when the smoke is very dense and the sign is invisible in the regular camera.

Chapter 2

LiSens — A Line Sensor Based Compressive Camera

"Sometimes you need to defocus to achieve what you want."

Many computational cameras achieve novel capabilities by using spatial light modulators (SLMs). Examples include light field cameras [56, 91], lensless cameras [38], and compressive cameras [21, 34, 52, 73]. The measurement rate of these cameras is often limited by the switching speed of the SLM used. This is especially true for the compressive single-pixel camera (SPC) [21] where a photodetector is super-resolved by a digital micromirror device (DMD). The SPC inherits both the spatial resolution of the DMD as well as its operating rate. On one hand, this allows capturing images at a high resolution in spite of the sensor being a single pixel. This is especially beneficial while imaging in non-visible wavebands, where the combination of a single photodetector coupled with a high-resolution DMD often provides an inexpensive alternative to full-frame sensors.[1] On the other hand, the SPC also inherits the operating rate of the DMD which, for commercially-available units, is in tens of kHz. At this measurement rate, acquiring images and videos at the spatial resolution of the DMD is feasible only at a very low temporal resolution.

It is instructional to compare the two widely differing models of sensing (see Figure 2.1) in terms of their cost, determined by the number of pixels in the camera, and measurement rate, defined as the number of measurements that the device obtains in unit-time. Full-frame 2D sensors that rely on Nyquist sampling are capable of achieving measurement rates in tens of MHz, but are prohibitively expensive in many non-visible wavebands. The SPC, which uses multiplexing as opposed to sampling, has low measurement rates, but is inexpensive for exotic wavebands. However, these imaging models are only the two extremes of a continuous camera design space with varying number of pixels. Our main observation is that the design that achieves

[1] Megapixel sensors in short-wave infrared, typically constructed using InGaAs, cost more than USD 100k.

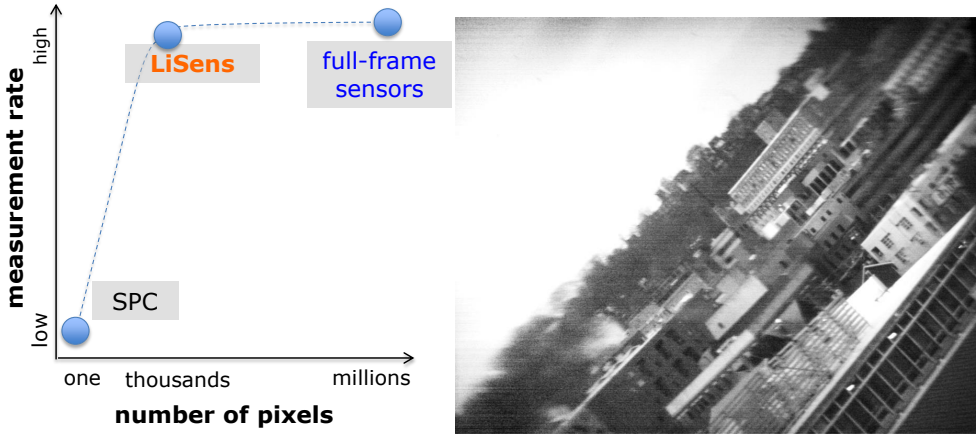


Figure 2.1: **Motivation.** Varying the resolution of the sensor in a multiplexing camera provides interesting tradeoffs. A single-pixel camera is inexpensive but has limited measurement rate. Full-frame sensors deliver high measurement rates but often at a steep price especially in non-visible wavebands. The proposed camera, LiSens, uses a 1024 pixel line-sensor and achieves a measurement rate in MHz. (right) A 1024×768 pixel image of an outdoor scene using our prototype.

the maximum measurement rate at low cost lies between the two extremes.

Based on this observation, we propose a novel architecture for video compressive sensing (CS) that is capable of delivering measurement rates in MHz. The proposed camera, which we call *LIne-SENSor-based compressive camera (LiSens)*, uses a 1D array of pixels to observe a scene via a DMD. Each pixel on the sensor is optically mapped to a row of micromirrors on the DMD and hence, the camera obtains a coded line-integrals of the image formed on the DMD. The high measurement rate of LiSens enables CS of images and videos at high spatial resolutions and reasonable frame-rates. We provide the optical schematic for building such a camera and illustrate its novel capabilities with a lab prototype.

2.1 Prior work

Multiplexed imaging. This chapter relies heavily on the concept of optical multiplexing. Suppose that we seek to sense a signal $\mathbf{x} \in \mathbb{R}^N$. A conventional camera samples this signal directly and has an image formation model given by

$$\mathbf{y} = \mathbf{x} + \mathbf{e},$$

where $\mathbf{e}$ is the measurement noise. Multiplexing expands the scope of the camera by allowing it to obtain any arbitrary linear transformation of the signal:

$$\mathbf{y} = A\mathbf{x} + \mathbf{e}.$$

Assuming that $A \in \mathbb{R}^{N\times N}$ is invertible, the signal can be estimated as $\widehat{\mathbf{x}} = A^{-1}\mathbf{y}$. It is well known that, when the entries of $A = [a_{ij}]$ are restricted to $|a_{ij}| \leq 1$, using a Hadamard measurement matrix is optimal provided the noise is signal independent [32]. Note that in the absence of any priors, we need at least N measurements to recover an N-dimensional signal.

Compressive sensing (CS). CS aims to sense a signal $\mathbf{x} \in \mathbb{R}^N$ from an under-determined linear system, i.e., measurement of the form

$$\mathbf{y} = \Phi\mathbf{x} + \mathbf{e},$$

where $\Phi \in \mathbb{R}^{M\times N}$ with $M < N$. For an arbitrary signal in $\mathbb{R}^N$ this is impossible since the map $\Phi : \mathbb{R}^N \mapsto \mathbb{R}^M$ is many-to-one and non-invertible. CS handles this by restricting the signal $\mathbf{x}$ to belong to a distinguished class; for example, sparse signals in a transform basis. The main results of CS state that when the measurement matrix Φ has a special structure and $\mathbf{x}$ is K-sparse in a transform basis, then we can robustly recover $\mathbf{x}$, provided $M = O(K\log(N/K))$ [8].

Video models for CS. There have been many models proposed to handle CS of videos. Wakin et al. [95] proposed the use of 3D wavelets as a sparsifying transform for videos. This model is further refined in Park and Wakin [66] where a lifting scheme is used to tune a wavelet to better approximate the temporal variations. Dictionary-based models were used in Hitomi et al. [34] for temporal resolution of videos; it is shown that a highly overcomplete dictionary provides high quality reconstructions. Inspired by the use of motion-flow models in video compression, Reddy et al. [73] and subsequently, Sankaranarayanan et al. [77] employed an iterative strategy where optical flow derived from an initial estimate is used to further constrain the video recovery problem. This provides a significant improvement in reconstruction quality, albeit at a high computational cost. Gaussian mixture models were used in [103] for video CS; a hallmark of this approach is that the mixture model is learned directly from the compressive measurements providing a model that is tailored to the specifics of the scene being reconstructed.

Compressive imaging hardware prototypes. The original prototype for the SPC used a DMD as the programmable SLM [21]. In [38], a variant of the SPC is proposed where an LCD panel is used in place of the DMD; the use of a transmissive light modulator enables a lensless architecture. Sen and Darabi [82] use a camera-projector system to construct an SPC exploiting a concept called dual photography [83]; the hallmark of this system is its use of active illumination.

There have been many multi-pixel extensions to the SPC. The simplest approach [44] is to map the DMD to a low-resolution sensor array, as opposed to a single photodetector, such that each pixel on the sensor observes a non-overlapping "patch" or a block of micromirrors on the DMD.[2] SMCs based on this design have been proposed for sensing in mid-wave infrared camera [54] and short-wave infrared [17]. Measurement matrix designs for such block-based compressive imaging architectures are presented in Kerviche et al. [45] and Ke and Lam [43].

While we focus on spatial multiplexing, many architectures have been proposed for multiplexing other attributes of a scene including temporal [24, 34, 35, 53, 73], angular [56, 91], and spectral [52, 94]. All of these architectures use a high-resolution sensor and sacrifice this spatial resolution partially to obtain higher resolution in time, spectrum and/or angle. In contrast, the imaging architecture proposed in this chapter takes a device with a high temporal resolution and sacrifices the temporal resolution in part to obtain higher spatial resolution.

2.2 Measurement rate of a camera

The measurement rate of a camera is given by the product of its spatial resolution, in pixels, and the temporal resolution, in frames per second. A spatial multiplexing camera (SMC) captures one or more coded linear measurements of a scene via a spatial light modulator (e.g., DMD). Multiple measurements are taken sequentially by changing the code displayed on the DMD. Thus, the measurement rate of an SMC is limited by the resolution and the frame-rate of the sensor, as well as the resolution and switching speed of the DMD.[3] The operating rate of the DMD is determined by the rate at which the micromirrors can be switched from one code into another. Denoting this rate as R_{DMD} Hz, we note that the frame rate of the SMC cannot be greater than R_{DMD} since the specific linear mapping from the scene to sensor is determined by the micromirror code. The frame rate of a sensor is also limited by its readout speed — typically, determined by the operating rate of the analog-to-digital converter (ADC) in the readout circuit. Suppose that we have an ADC that can perform readout at a rate of R_{ADC} Hz. Then, given a frame with F pixels, the ADC limits the sensor to R_{ADC}/F frames per second. Hence, the measurement rate of the SMC is given as

$$F \times \min\left(R_{DMD}, \frac{R_{ADC}}{F}\right) \quad \text{meas. per sec.}$$

Operating speeds of commercially-available DMDs are 2-3 orders of magnitude lower than that of ADCs in sensors. As a consequence, for small values of F, the measurement rate

[2]The interested reader is referred to earlier work by Nayar et al. [62] where a scene is observed by a sensor via a DMD to enable programmable imaging. However, this device was not used for spatial multiplexing or compressive sensing.

[3]Compressive cameras can further increase the spatial or temporal resolution by using scene priors.

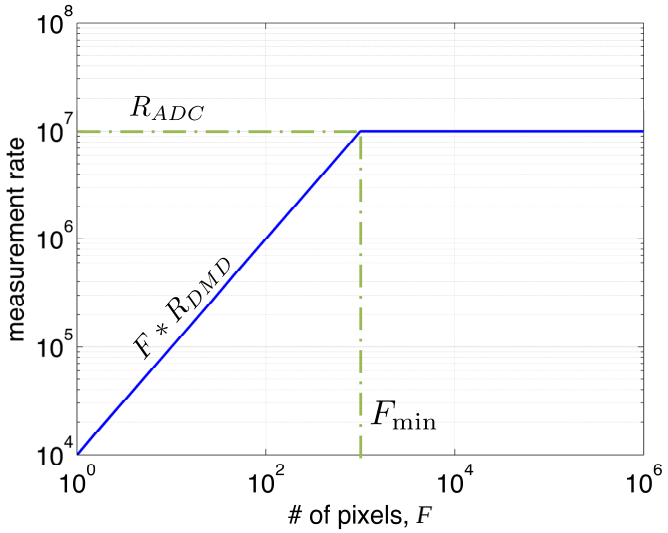


Figure 2.2: **Measurement rate and sensor resolution.** There are two regimes of operations: for $F < F_{\min}$, the bottleneck is the DMD and for $F > F_{\min}$, the bottleneck is the sensor readout. At $F = F_{\min}$ lies a design that obtains the highest measurement rate with the least sensor resolution.

of an SMC is dominated by the operating speed of the DMD (see Figure 2.2). Once $F \geq R_{ADC}/R_{DMD}$, then the bottleneck shifts from the DMD to the ADC. Further, the smallest sensor resolution (in terms of number of pixels) for which the measurement rate is maximized is

$$\text{(minimum number of pixels)} \quad F_{\min} = R_{ADC}/R_{DMD}.$$

In essence, at $F = F_{\min}$ we can obtain the measurement rate of a full-frame sensor but with a device with potentially a fraction of the number of pixels. This is invaluable for sensing in many wavebands, for example short-wave infrared.

As a case study, consider an SMC with a DMD operating rate $R_{DMD} = 10^4$ Hz and an ADC with an operating rate $R_{ADC} = 10^7$ Hz. Then, for a sensor with $F_{\min} = 10^3$ pixels, we can obtain 10^7 measurements per second. An SPC, in comparison, would only provide 10^4 measurements per second. This significant increase in measurement rate motivates the multi-pixel SMC design that we propose in the next section.

2.3 LiSens

The optical setup of LiSens is illustrated in Figure 2.3. The DMD splits the optical axis into two axes — one each for $\pm 11^\circ$ orientations of the micromirrors. Along one of these axes, the 2D image of the scene formed on the DMD plane is mapped onto the 1D line-sensor using a relay

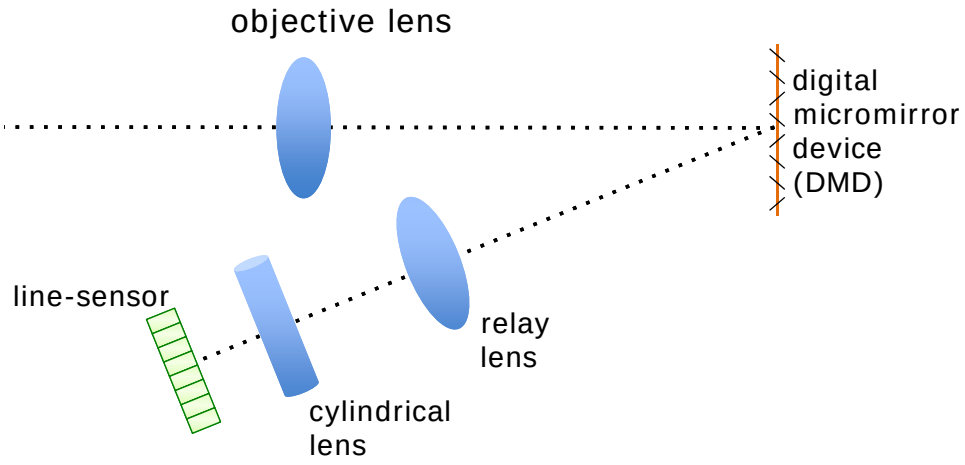


Figure 2.3: **Schematic of the LiSens imaging architecture.**

lens and a cylindrical lens. The image captured by the line-sensor can be represented as a 1D integral of the 2D image (along rows or columns) formed on the DMD plane. This is achieved by aligning the axis of the cylindrical lens with that of the line-sensor and placing it so that it optically mirrors the aperture plane of the relay lens onto the sensor plane (see Figure 2.4).

Suppose that the physical dimensions of the DMD and the line-sensor are (w_D, h_D) and (w_L, h_L), respectively, with $h_L \ll w_L$. The relay lens is selected to produce a magnification of w_L/w_D so that the DMD maps to the line-sensor along its width; this determines the ratio of the focal lengths used in the relay lens. The exact value of the focal lengths is optimized to minimize vignetting and needs to be determined along with the objective lens. If $w_D = w_L$, we can use a 1:1 relay lens with a focal length f_r and the distance between the relay lens and the line-sensor is $2f_r$. The parameters and the placement of the cylindrical lens are determined so that the aperture of the relay lens is magnified (or shrunk) within the line-sensor. This corresponds to a magnification of h_L/d_r, where d_r is the diameter of the relay lens. Let f_c be the focal length of the cylindrical lens, and suppose that it is placed at a distance of u_c from the line-sensor and v_c from the relay lens. Then,

$$u_c + v_c = 2f_r,\ \frac{1}{u_c} + \frac{1}{v_c} = \frac{1}{f_c},\ \frac{u_c}{v_c} = \frac{h_L}{d_r}$$

Hence,

$$f_c = 2f_r \frac{h_L}{d_r + h_L} \frac{d_r}{d_r + h_L} \approx 2f_r \frac{h_L}{d_r}. \tag{2.1}$$

In practice, we observe that, for a marginal loss of light throughput, we could obtain flexibility in both the choice of the cylindrical lens (focal length) as well as its positioning. Figure 2.5 shows the improvement in the vertical field-of-view when the cylindrical lens is introduced.

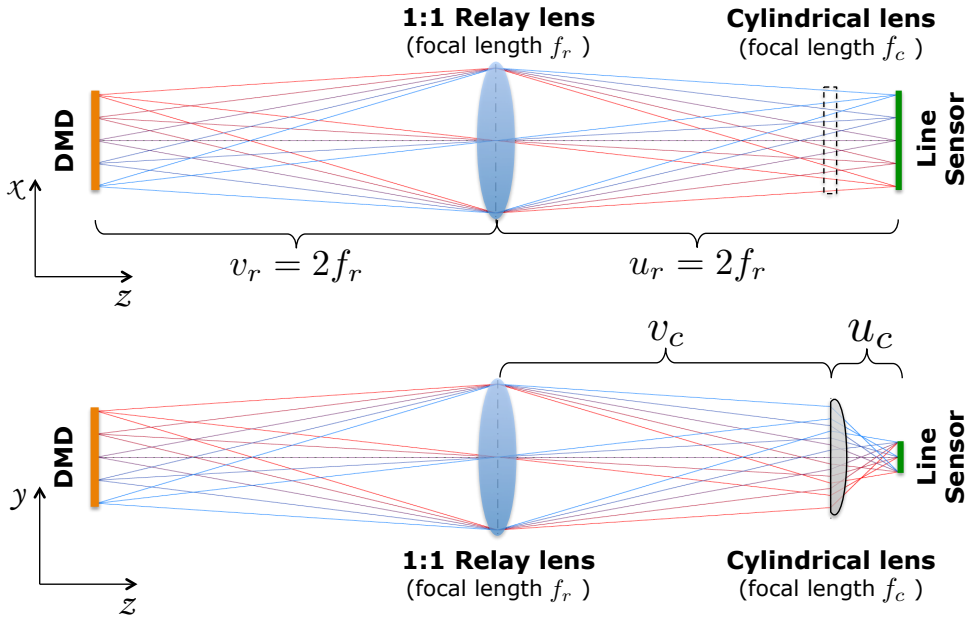


Figure 2.4: **The DMD-to-sensor mapping.** Shown are the ray diagrams for the mapping of the DMD to line-sensor along two orthogonal directions — (top) along and (bottom) orthogonal to the axis of cylindrical lens. The cylindrical lens focuses the aperture plane of the relay lens onto the sensor plane. Hence, the sensor plane is focused on the DMD along one-axis (top) but defocused along the other (bottom).

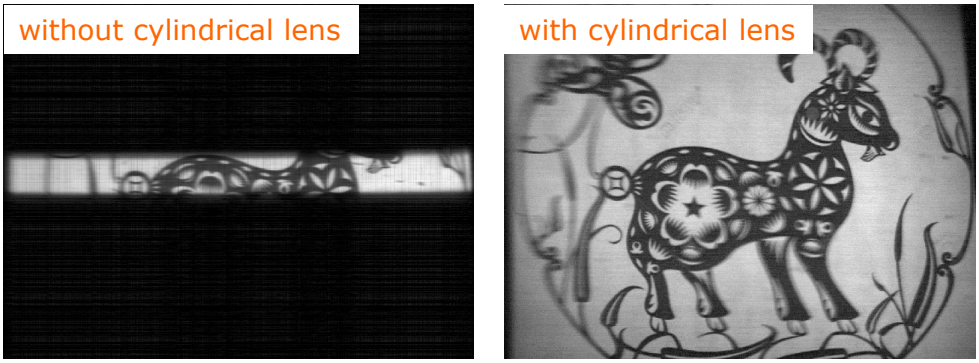
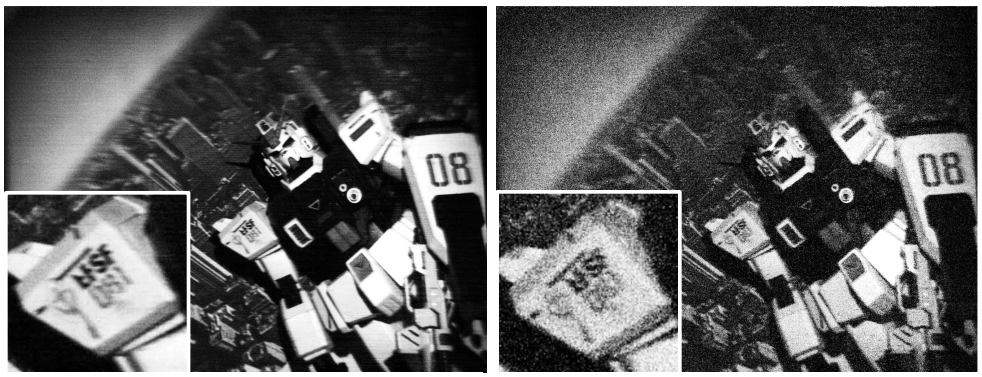


Figure 2.5: **Field-of-view enhancement by the cylindrical lens.** Shown are images acquired with our prototype.

2.3.1 Practical advantages

Before we proceed, it is worth pondering on alternative and potentially simpler designs for multi-pixel SMCs. An intuitive multi-pixel extension of the SPC would be using a low-resolution 2D sensor array [54**?**]. Having a 2D sensor array provides a simpler mapping from the DMD that



(a) simultaneous exposure and readout (b) sequential exposure and readout

Figure 2.6: **Benefits of frame transfer.** The DMD was operated with $400\mu s$ per pattern. The sensor had a readout of $350\mu s$ per frame. (a) Simultaneous readout allows for a $400\mu s$ exposure. (b) Sequential readout cut the exposure time down to $50\mu s$, resulting in a noisy image. Shown are images acquired with our prototype.

can be achieved using just relay lenses. Yet, there are important considerations that make the LiSens camera a powerful alternative to using 2D sensors.

Specifically, it is simpler and inexpensive to obtain frame transfer[4] on a line sensor without requiring complex circuitry or loss of light due to reduced fill-factor. Figure 2.6 demonstrates the benefits of frame transfer on our lab prototype. The 1D profile of the sensor also provides the possibility of having a per-pixel ADC; this provides a dramatic increase in the readout rate of the sensor. Finally, we benefit from the fact that line-sensor have long been manufactured for spectroscopy which requires very precise, low-noise sensors with high dynamic range and broad spectral response; these properties make them highly desirable for our application.

2.3.2 Imaging model

Suppose the DMD has a resolution of $N \times N$ micromirrors, and the line-sensor has N pixels such that each pixel maps to a row of micromirrors on the DMD. At time instant t, let $\mathbf{X}_t \in \mathbb{R}^{N \times N}$ be the scene image formed on the DMD, and let $\Phi_t \in \mathbb{R}^{N \times N}$ be the binary pattern displayed on the DMD. Then, the measurement obtained at the line-sensor, $\mathbf{y}_t \in \mathbb{R}^N$, is given as

$$\mathbf{y}_t = (\mathbf{X}_t \circ \Phi_t)\mathbf{1} + \mathbf{e}_t,$$

[4]Frame transfer is a technology used to enable simultaneous exposure and readout in a sensor. In a traditional camera, during the readout process, the sensor is not exposed; this reduces the duty cycle of the sensor and leads to inefficiencies in light collection. With frame transfer, there is a separate array of storage pixels beside the photo-sensitive array. After exposure, charge from the pixels are immediately transferred to the storage array allowing the sensor to be exposed again.

where $\circ$ denotes the entry-wise product, $\mathbf{1}$ is the vector of ones and $\mathbf{e}_t$ is the measurement noise.

For simplicity, we use DMD patterns where every column is the same, i.e., patterns of the form

$$\Phi_t = \mathbf{1}{\phi_t}^T,$$

where $\phi_t \in \mathbb{R}^N$ is a binary vector. This alleviates the need for extensive calibration. The sensor measurements are then given as

$$\mathbf{y}_t = \mathbf{X}_t\phi_t + \mathbf{e}_t.$$

For the experiments in the chapter, we use rows of a column-permuted Hadamard matrix for ϕ_t.

2.3.3 Recovery

Sensing images. When the scene is static, i.e., $\mathbf{X}_t = \mathbf{X}$ for $t \in \{1, \ldots, T\}$, we can write the imaging model as

$$\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_T] = \mathbf{X}\Phi + E,$$

where $\Phi = [\phi_1, \ldots, \phi_T]$ is the measurement matrix and $E = [\mathbf{e}_1, \ldots, \mathbf{e}_T]$ is the measurement noise.

If $T \geq N$ and Φ were well-conditioned, say for example, a Hadamard matrix, then we can obtain an estimate of the scene image as

$$\widehat{\mathbf{X}} = \mathbf{Y}\Phi^{\dagger}. \tag{2.2}$$

In practice, it is unreasonable to expect a scene to be static over a large duration and hence, we can expect $T < N$. To regularize the inverse problem, we enforce sparsity in the gradients of the recovered image using a minimum total-variation prior [16]. This leads to the following optimization problem.

$$\min_{\mathbf{X}} TV(\mathbf{X}), \quad \text{s.t.} \quad \|\mathbf{Y} - \mathbf{X}\Phi\|_F \leq \epsilon, \tag{2.3}$$

where $TV(\mathbf{X})$ is the total-variation norm that captures strength of the image gradients. We used the isotropic TV-norm defined as

$$TV(\mathbf{X}) = \sum_{i=1}^{N}\sum_{j=1}^{N}\sqrt{G_x^2(i,j) + G_y^2(i,j)},$$

where $G_x, G_y \in \mathbb{R}^{N\times N}$ are the spatial gradients of $\mathbf{X}$.

Sensing videos. We enforce sparse spatio-temporal gradients when recovering time-varying scenes. Specifically, suppose that we obtain measurements $\{\mathbf{y}_t, 1 \leq t \leq T\}$ for a time-varying scene $\{\mathbf{X}_t, 1 \leq t \leq T\}$. It is often an overkill to recover an image for every frame of the sensor since the problem becomes computationally overwhelming. Instead, we decide on a target frame-rate (say 10 frames per second) and group together successive measurements so as to obtain the desired frame rate. Let $\{\widetilde{\mathbf{X}}_k, k = 1, \ldots, Q\}$ be the frames associated with the scene at the desired frame-rate. Then, the imaging model reduces to

$$t = k\frac{T}{Q}, \quad [\mathbf{y}_t, \ldots, \mathbf{y}_{t+\frac{T}{Q}}] = \widetilde{\mathbf{X}}_k[\phi_t, \ldots, \phi_{t+\frac{T}{Q}}] + E_k$$

Grouping together measurements and associating them to a frame of a video also reduces the number of parameters to be estimated. Interested readers are referred to [67] for a frequency domain analysis on judicious selection of frame-rate for the SPC. Finally, similar to (2.3), we solve for the video by minimizing a spatio-temporal TV-norm constrained by the measurements.

2.4 Hardware prototype

Our proof-of-concept prototype was built using a Nikkor 50mm F/1.8 objective lens, a DLP7000 DMD, and a Hamamatsu S11156-2048-01 line-sensor. The DMD had a spatial resolution of 1024×768 with a micromirror pitch of $13.8\mu m$ and a maximum operating speed of 20 kHz. The line sensor had 2048 pixels with a pixel size of $14\mu m \times 1mm$; we were only able to use 1024 pixels due to our inability to find a cylindrical lens that was sufficiently long to span the entire line-sensor. Due to the pixel pitch and micromirror pitch being nearly the same, we used a 1:1 relay lens and oriented the line sensor so that the line sensor was aligned to the width of the DMD. Hence, each pixel on the line sensor summed up 768 micromirrors on the DMD.

Our line sensor provided access to simultaneous exposure and readout using frame transfer. We operated the prototype with an exposure of approximately $500\mu s$. Due to buffer limitations in the readout circuit (a Hamamatsu C11165-01 driver), we were able to obtain 100 frames at $500\mu s$ per frame, followed by a cool-down time of $60ms$. As a consequence, the device provided 900 frames per second and hence, a measurement rate of 9×10^5 measurements per second.

For comparison and validation, we built an SPC on the off-axis of the DMD (see Figure 4.4). The measurement rate of the SPC was 20 kHz, the speed of the DMD.

Figure 4.4 shows both the modified schematic of the camera as well as the optical layout. The subtle differences to the schematic shown in Figure 2.3 stem from a couple of practical constraints. Recall that the DMD reflects the optical axis by 22°. The flange distance of objective lenses (F-mount and C-mount) is insufficient to provide clearance for the cone of light reflected

from the DMD. To alleviate this, we optically mirror the image plane of the objective lens using a 1:1 relay lens, thereby providing ample space for the reflected cone of light. We used a 100mm:100mm achromatic doublets for the relay lens, thereby providing an $f_r = 50mm$ and a cylindrical lens with a focal length of $12.7mm$.[5]

Alignment. Misalignment of the optical components, in addition to introducing blur, results in a DMD column mapping to multiple pixels on the line detector. In our prototype, we minimized both effects by manually adjusting each component while observing and quantifying sharpness of the images in real-time. While this produces acceptable results, a calibration procedure is indeed required especially to resolve the image especially at the boundaries of the field of view.

Vignetting. The use of relay lenses to extend the optical axis causes significant vignetting. To alleviate this, we introduced a field lens at the DMD which dramatically reduces vignetting. A second field lens at the image plane of the objective lens further reduces vignetting but with an increase in spatial blur. The results in this chapter were produced with a single field lens at the DMD.

2.5 Experiments

We use two metrics to characterize the operating scenarios for our experiments:

(i) The under-sampling given as the ratio of the dimensionality of the acquired image to the number of measurements, i.e., if we seek to acquire an $N_1 \times N_2$ image with M measurements, then the under-sampling is $N_1 N_2 / M$. When the under-sampling is 1, the system is invertible and we use (2.2) to recover the image. For values greater than 1, we use the TV prior and solve (2.3).

(ii) Temporal resolution / capture duration per frame. Recall, from Section 2.3.3, that we pool together measurements and associate them with a single recovered frame. This determines the capture duration per frame and its reciprocal, the temporal resolution. The smaller the capture duration, the fewer the measurements that we acquire and hence, the greater the under-sampling.

These two metrics are linked by the measurement rate of the camera and the resolution at which we operate the DMD. For a capture duration of τ seconds, the under-sampling is given as

$$(N_1 N_2)/(\tau \times \text{measurement rate})$$

where the DMD resolution is $N_1 \times N_2$. Recall, that the measurement rate of our LiSens prototype is nearly 1 MHz while that of the SPC is 20 kHz.

[5]We used relay lenses with diameter $d_r = 25mm$. The value suggested in (2.1) is $f_c = 4mm$. However, to avoid custom designed optics, we went with a larger focal length.

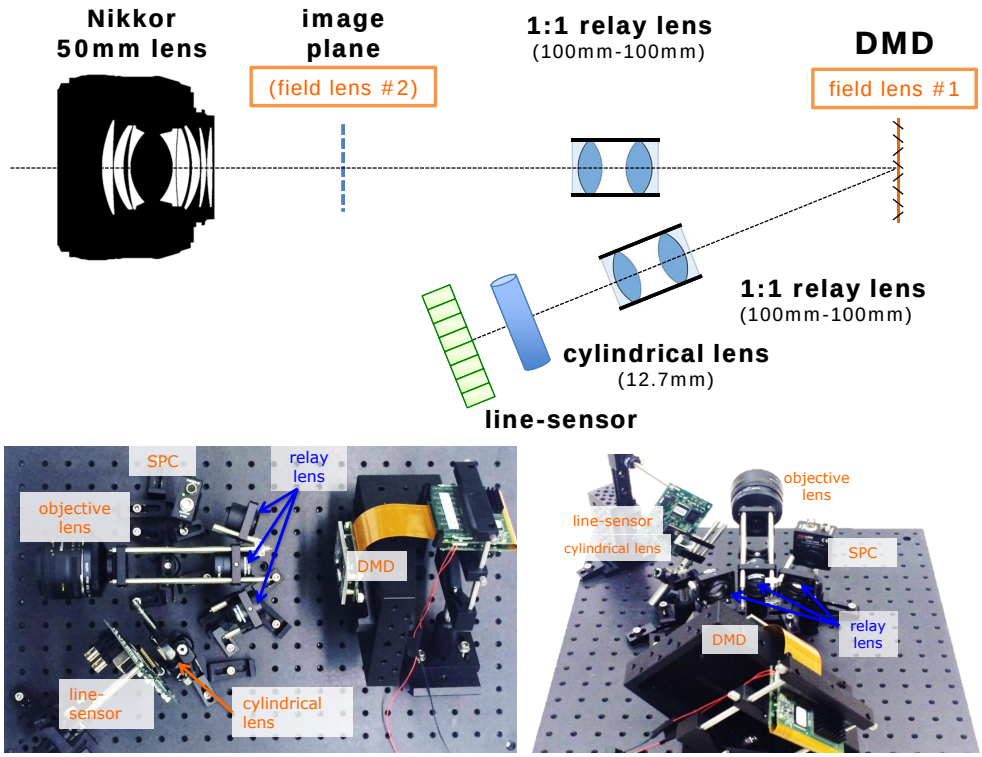


Figure 2.7: **Prototype of the LiSens camera.**

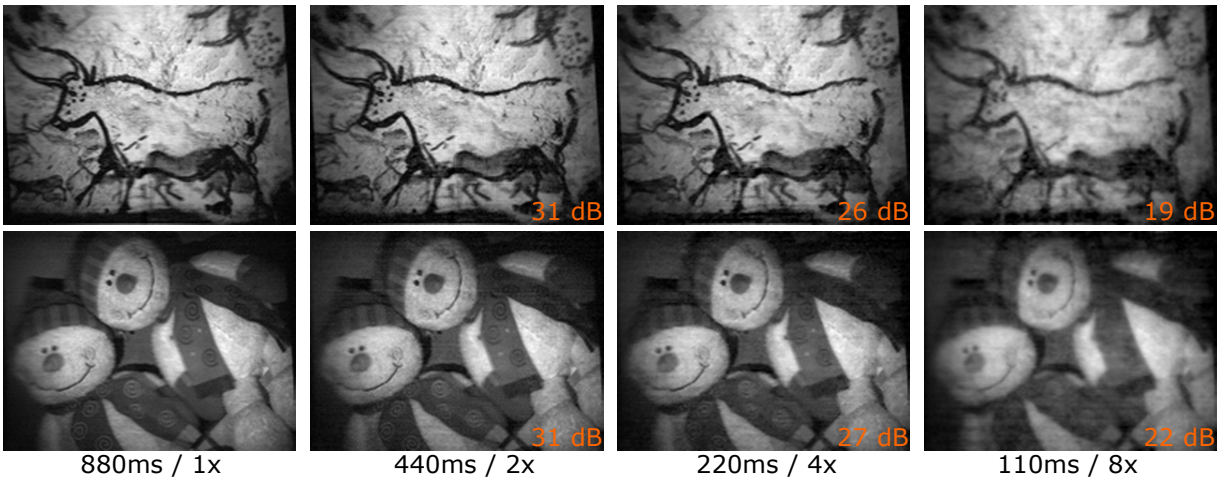


Figure 2.8: **LiSens reconstructions for a few scenes under varying capture durations / under-sampling.** The reconstructions suffer little loss in quality until a capture time of $110ms$; this corresponds to an under-sampling of $8\times$. Inset on each figure is the reconstruction SNR, in dB, using the $1\times$ reconstruction as the ground truth.

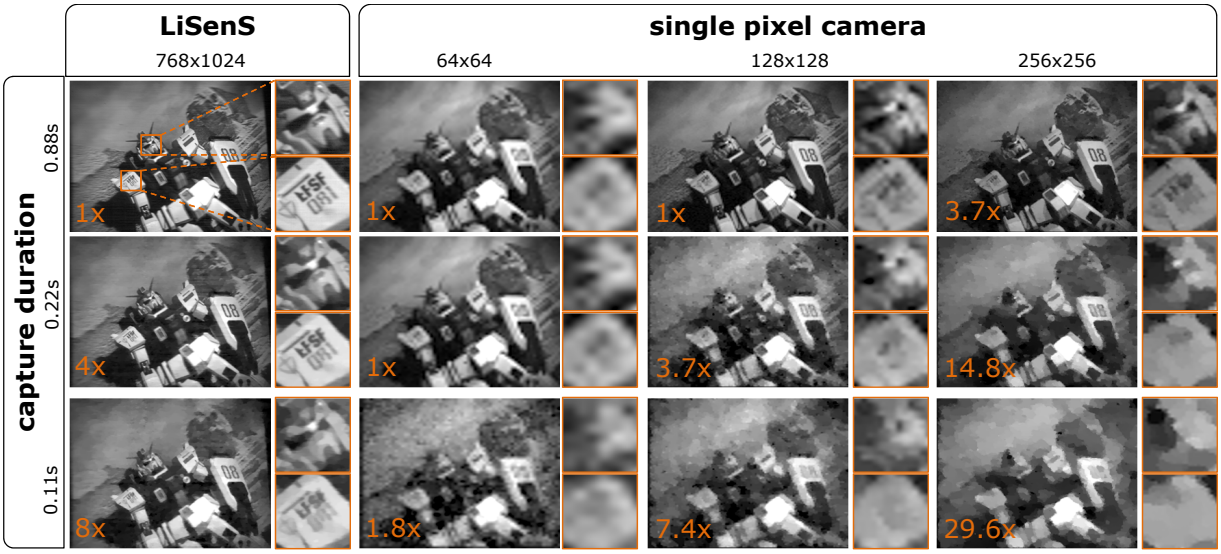


Figure 2.9: **Performance of LiSens and SPC.** We look at recovered images for various capture durations and, in the case of the SPC, various spatial resolution as well. The remarkable gap in the performance of the devices can be attributed to the difference in their measurement rates. Inset in orange are the under-sampling factors at which images were reconstructed.

Finally, we map the ± 1 entries of the Hadamard matrix to $0/1$ by mapping the '-1's to '0's; in post-processing, the measurements corresponding to the ± 1 matrix are obtained simply by subtracting the mean scene intensity. For time-varying scenes, we repeat the all-ones measurement in the Hadamard matrix once every 100 measurements to track the mean intensity value.

Gallery. We present the images recovered by the LiSens camera on two indoor static scenes for various capture durations in Figure 2.8. We also quantize the loss in performance with increased under-sampling using the reconstruction SNR. Specifically, given a ground truth image $\mathbf{x}_{gr}$ and a reconstructed image $\widehat{\mathbf{x}}$, the reconstruction SNR in dB is given as

$$-20 \log_{10} \left(\frac{\|\mathbf{x}_{gr} - \widehat{\mathbf{x}}\|_2}{\|\mathbf{x}_{gr}\|_2} \right).$$

We used the Nyquist rate reconstruction, i.e., under-sampling of $1\times$, as the ground truth image. We observe that there is little loss in performance until a capture duration of $220ms$ (under-sampling of $4\times$) and with a small drop in performance at $110ms$ ($8\times$). As we will see later, using a spatio-temporal prior improves the reconstruction quality even at an under-sampling of $8\times$.

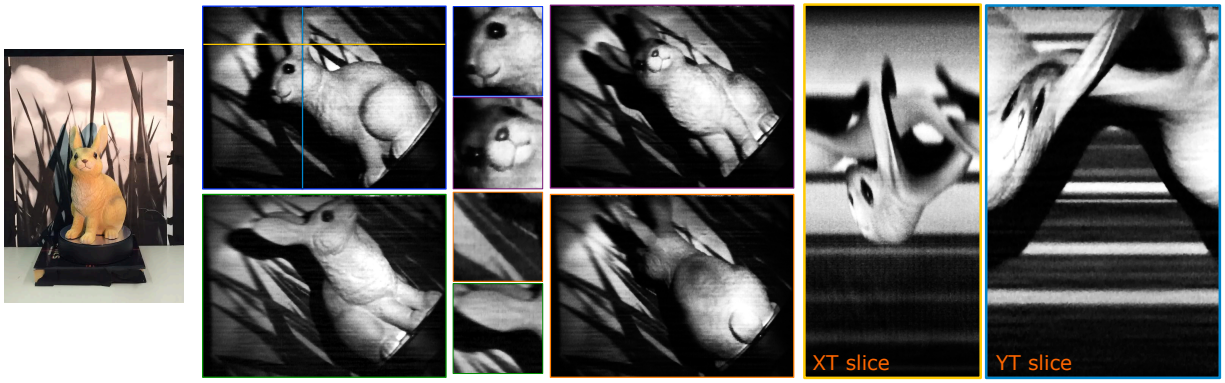


Figure 2.10: **Video CS using LiSens**. We image a bunny on a turntable (left) and recover the video at a capture duration of $110ms$ per frame (under-sampling of $8\times$) using a 3D TV prior. Shown are a few recovered frames and spatio-temporal slices.

Comparison to SPC. We compare the performance of LiSens and an SPC built on the off-axis of the DMD. We image a static scene using both cameras and compare the reconstructed image for varying capture durations (see Figure 2.9). Recall that the measurement rate of the SPC is limited to 20 kHz while the LiSens has nearly 1 MHz. This $50\times$ gap enables the LiSens camera to sense images at a resolution of 1024×768 at $5 - 10$ frames per second. As we will see next, using simple spatio-temporal priors like a 3D TV norm will enable higher-quality reconstructions at an increased temporal resolution. We note that the photodetector in the SPC had a lower dynamic range than the line-sensor. For a fairer comparison, we took multiple runs with the SPC and averaged the measurements to artificially boost its dynamic range.

Figure 2.9 also indicates that LiSens operating at $8\times$ under-sampling produces significantly higher quality results than the SPC operating at $7.4\times$ under-sampling. This is a consequence of the sparsity of images scaling sub-linearly with image resolution. That is, if we looked at sparsity of images in a wavelet basis as a function of image resolution, the ratio of sparsity level to image resolution decreases monotonically. This can also be attributed to images having a lot of energy in low-freq components. Hence, if we attempt to recover a scene at 128×128 and 1024×1024, each at say $8\times$ under-sampling, then the quality would be significantly better for the high-resolution reconstruction.

Dynamic scenes. Next, we sensed a time-varying scene — a bunny on a turntable spinning at one revolution every 15 seconds. We fix a capture duration of $110ms$ per frame or equivalently a temporal-resolution of 9 frames per second, and recover the frames of the video under a 3D total variation prior. Due to high levels of measurement noise, the recovered videos were subject to a

Figure 2.11: **Sensing outdoors with the LiSens prototype.**

$3 \times 3 \times 3$ median-filter. As seen in Figure 2.10, the recovered video preserves fine spatial detail. This can be attributed to the use of video priors, which capture inter-frame redundancies, as well as the high measurement rate enabled to our design.

This validates the potential of the LiSens architecture to acquire scenes at high spatial resolution (1024×768) and with a reasonable temporal resolution.

Finally, Figure 2.11 shows frames of a video of an outdoor scene reconstructed with the same parameters as the indoor bunny scene. The data was significantly noisier which contributed to the reduced reconstruction quality. The use of stronger priors such as motion flow models as in [73, 77] could help in such scenarios.

2.6 Discussion

We present a multiplexing camera that is capable of delivering very high measurement rates, comparable to that of a full-frame sensor, but with a number of pixels that is a small fraction. At its core, the proposed camera moves the bottleneck in the frame rate of the device from the spatial light modulator to the readout of the sensor. The ensuing boost in measurement rate enables sensing of scenes at video rate and at the full resolution of the light modulator. Finally, while our prototype was built for visible wavebands, the underlying principles transfer mutatis mutandis to other wavebands of interest.

Chapter 3

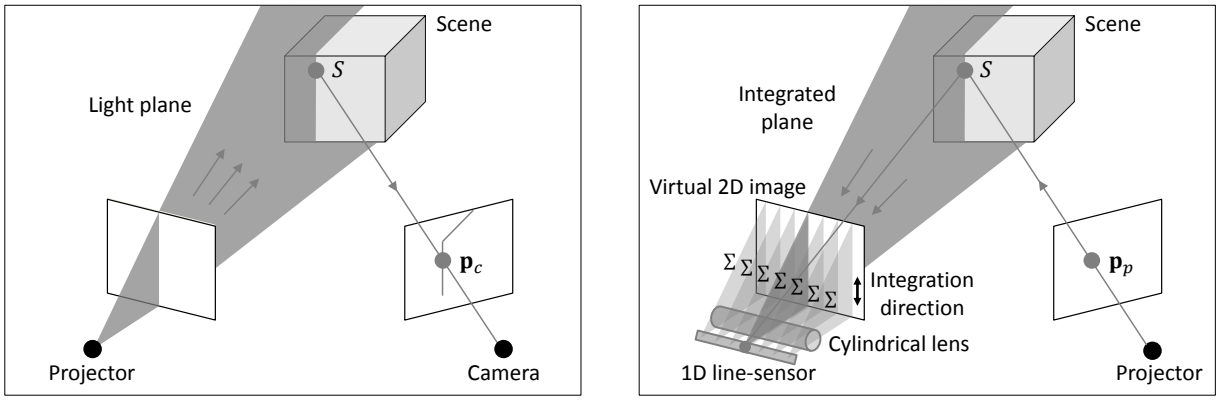
DualSL — Dual Structured Light Using a 1D Sensor

"Solving in the dual is easier."

Structured light (SL) [27] is one of the most popular techniques for 3D shape acquisition. An SL system uses active illumination, typically via a projector, to obtain robust correspondences between pixels on the projector and a camera, and subsequently, recovers the scene depth via triangulation. In contrast to passive techniques like stereo, the use of active illumination enables SL systems to acquire depth even for textureless scenes at a low computational cost.

The simplest SL method is point scanning [22], where the light source illuminates a single scene point at a time, and the camera captures an image. Correspondence between camera and projector pixels is determined by associating the brightest pixel in each acquired image to the pixel illuminated by the projector. However, this approach requires a large number (N^2) of images to obtain a depth map with $N \times N$ pixels. In order to reduce the acquisition time, stripe scanning technique was proposed where the light source emits a planar sheet of light [5, 19, 85]. Consider a scene point that lies on the emitted light plane. Its depth can be estimated by finding the intersection between the light plane, and the ray joining the camera center and the camera pixel. This is illustrated in Figure 3.1(a). We can further reduce the acquisition time by using more sophisticated temporal coding techniques; for example, binary codes [70], Gray codes [40, 81] and sinusoidal phase shifting [88].

Underlying all these methods is the idea that, for a calibrated camera-projector pair, we only need to measure disparity, i.e., a 1D displacement map. Thus, we need to perform coding along only one dimension of the projector image plane, thereby achieving significant speed-ups over point-scanning systems. For example, several structured light patterns have a 1D translational symmetry, i.e., in the projected patterns, all the pixels within a column (or a row) have the



(a) Conventional structured light (b) Dual structured light (DualSL)

Figure 3.1: **DualSL compared with traditional SL.** Depth from SL can be obtained by performing ray-plane triangulation. For traditional SL, the ray is from camera pixel, and plane is formed by center of projection and a column of the projector. In DualSL, ray is from projector pixel, and plane is formed by a line sensor pixel with cylindrical optics.

same intensities.[1] This is illustrated in Figure 3.1(a). *For such patterns with 1D translational symmetry, conventional structured light systems can be thought of as using a 1D projector, and a 2D sensor.*

In this chapter, we present a novel SL design called *DualSL* (Dual Structured Light) that uses *a 2D projector and a 1D sensor, or a line-sensor*. DualSL comprises of a novel optical setup where pixels on the line-sensor integrates light along columns of the image focused by the objective lens, as shown in Figure 3.1(b). As a consequence, the DualSL design can be interpreted as the optical dual [83] of a traditional SL system, i.e., we *find correspondences between columns of the camera and pixels on the projector.* In contrast, in conventional SL, we find correspondences between pixels of the camera and columns of the projector.

Why use a 1D sensor for structured light? The use of a line-sensor, instead of a 2D sensor, can provide significant advantages in many applications where a 2D sensor is either expensive or difficult to obtain. For example, the typical costs for sensors in shortwave infrared (SWIR; 900nm-2.5μm) is $\$0.10$ per pixel [25]; hence, a high-resolution 2D sensor can be prohibitively expensive. In this context, a system built using a 1D line-sensor, with just a few thousand pixels, can have a significantly lower cost. A second application of DualSL is in the context of dynamic vision sensors (DVS) [51], where each pixel has the capability of detecting temporal intensity

[1] An exception is 'single-shot' structured light techniques that use patterns with 2D intensity variations, for example, sparse 2D grid of lines [71], 2D color encoded grids [75], 2D pseudo-random binary code [93], and 2D random dots (used in the first generation Microsoft Kinect depth sensing cameras [58]).

changes in an asynchronous manner. It has been shown that the use of a DVS with asynchronous pixels can reduce the acquisition time of line striping based structured light by up to an order of magnitude [14, 57]. However, the additional circuit at each pixel for detecting temporal intensity changes and enabling asynchronous readout leads to sensors that are inherently complex and have a poor fill-factor (around 8.1% for commercially available units [20, 39]), and low resolution (e.g., 128×128). In contrast, a 1D DVS sensor [69] can have a larger fill-factor (80% for the design in [69]), and thus, a significantly higher 1D resolution (e.g., 2048 pixels), by moving the per-pixel processing circuit to the additional space available both above and below the 1D sensor array.

Our contributions are as follows:

- **SL using a line-sensor.** We propose a novel SL design that utilizes a line-sensor and simple optics with no moving parts to obtain the depth map of the scene. This can have significant benefits for sensing in wavelength regimes where sensors are expensive as well as sensing modalities where 2D sensors have low fill-factor, and thus poor resolution (e.g., dynamic vision sensors).
- **Analysis.** We analyze the performance of DualSL and show that its performance in terms of temporal resolution is the same as a traditional SL system.
- **Validation via hardware prototyping.** We realize a proof-of-concept hardware prototype for visible light to showcase DualSL, propose a procedure to calibrate the device, and characterize its performance.

3.1 DualSL

In this section, we describe the principle underlying DualSL, and analyze its performance in terms of the temporal resolution of obtaining depth maps.

3.1.1 Design of the sensing architecture

The optical design of sensing architecture, adapted from LiSens, is shown in Figure 3.2. The setup consists of an objective lens, a cylindrical lens, and a line-sensor. The line-sensor is placed on the image plane of the objective lens, so that the scene is perfectly in focus along the axis of the line-sensor. A cylindrical lens is placed in between the objective lens and the sensor such that its axis is aligned with that of the line-sensor. The cylindrical lens does not perturb light rays along the x-axis (axis parallel to its length). This results in the scene being in focus along the x-axis. Along the y-axis (perpendicular to the length of the cylindrical lens), the position and

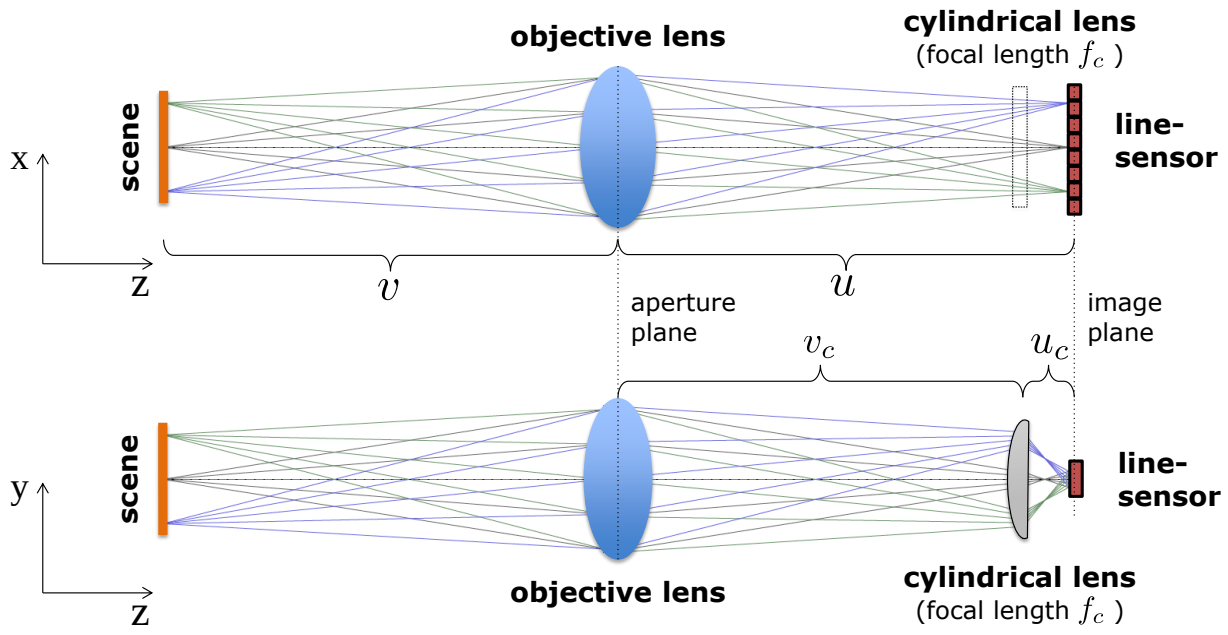


Figure 3.2: **Design of the sensing architecture visualized as ray diagrams along two orthogonal axes.** The line-sensor is placed at the image plane of the objective lens. A cylindrical lens is placed in front of the line-sensor such that its axis is aligned with that of the line-sensor. The cylindrical lens does not perturb light rays along the x-axis (top-row); this results in the scene being in focus along the x-axis. Along the y-axis (bottom-row), the cylindrical lens brings the aperture plane into focus at the image plane. Hence, the scene is completely defocused along the y-axis, i.e., each line-sensor pixel integrates light along the y-axis.

focal length of the cylindrical lens are chosen to ensure that its aperture plane is focused at the image plane. Hence, the scene is completely defocused along the y-axis, i.e., each line-sensor pixel integrates light along the y-axis. This is illustrated in Figure 3.2(bottom-row). Further, for maximum efficiency in gathering light, it is desirable that the aperture of the objective lens is magnified/shrunk to the height of the line-sensor.

Determining the parameters of the cylindrical lens. The focal length, f_c, of the cylindrical lens and its distance to the line-sensor, u_c, can be derived from the desiderata listed above. Given the aperture diameter of the objective lens D, the sensor-lens distance u, the height H of the line-sensor pixels, and the length of the line-sensor L, we require the following constraints to be satisfied:

$$\frac{1}{u_c} + \frac{1}{u - u_c} = \frac{1}{f_c} \qquad \text{(focusing aperture plane onto image plane)}$$

$$\frac{D}{H} = \frac{u - u_c}{u_c} \qquad \text{(magnification constraints)}$$

Putting them together, we can obtain the following expressions for u_c and f_c:

$$u_c = \frac{H}{D+H}u, \quad f_c = \frac{HD}{(D+H)^2}u. \tag{3.1}$$

The last parameter to determine is the height of the cylindrical lens which determines the field-of-view along the axis perpendicular to the line-sensor. For a symmetric field-of-view, we would require the height of the cylindrical lens to be greater than $\frac{D}{D+H}(L+H)$.

Remark. It is worth noting here that line-sensors are often available in form-factors where the height of the pixels H is significantly greater than the pixel pitch. For example, for the prototype used in this chapter, the pixel height is 1mm while the pixel pitch is 14μm. This highly-skewed pixel aspect ratio allows us to collect large amounts of light at each pixel with little or no loss of resolution along the x-axis. Further, such tall pixels are critical to ensure that the parameters' values defined in (3.1) are meaningful. For example, if $D \approx H$, then $u_c = u/2$ and $f_c = u/4$. Noting that typical values of flange distances are 17.5mm for C-mount lenses and 47mm for Nikkor lenses, it is easily seen that the resulting values for the position and the focal length are reasonable.

Scene to sensor mapping. The sensor architecture achieves the following scene-to-sensor mapping. First, the objective lens forms a virtual 2D image of the scene, $I(m,n)$.[2] Second, the effect of the cylindrical lens is to completely defocus the virtual image along one direction. Hence, the measurements on the line-sensor are obtained by projecting the virtual image along a direction perpendicular to the axis of the line-sensor. Specifically, each pixel of the line-sensor integrates the virtual image along a line, i.e., the measurement made by a pixel x on the line sensor is the integration of intensities observed along a line with a slope b/a:

$$i(x) = \int_{\alpha} I(x + a\alpha, b\alpha) d\alpha.$$

The slope b/a is controlled by the axis of the line-sensor/cylindrical lens. An important feature of this design is that the *pre-image* of a line-sensor pixel is a plane. Here pre-image of a pixel is defined as the set of 3D world points that are imaged at that pixel; for example, in a conventional perspective camera, the pre-image of a pixel is a ray. As we will demonstrate shortly, this property will be used by the DualSL system for acquiring 3D scans of a scene.

[2]For simplicity, the image formation is described while ignoring the effects of pixelation and quantization.

3.1.2 3D scanning using the DualSL setup

We obtain 3D scans of the scene by obtaining correspondences between projector and line-sensor pixels. Suppose that pixel (m, n) of the projector corresponds to pixel x on the line-sensor. We can obtain the 3D scene point underlying this correspondence by intersecting the pre-image of the projector pixel, which is a line in 3D, with the pre-image of the line-sensor pixel — a plane in 3D. As long as the line and the plane are not parallel, we are bound to get a valid intersection, which is the 3D location of the scene point. For simplicity, we assume that the projector and line-sensor are placed in a rectified left-right configuration; hence, we can choose $a = 0$ and $b = 1$ to integrate vertically (along "columns") on the virtual 2D image.

Obtaining projector-camera correspondences. The simplest approach for obtaining correspondences is to illuminate each pixel on the projector sequentially, and capturing an image with the 1D sensor for each projector pixel location. For each projector pixel, we can determine its corresponding line-sensor pixel by finding the pixel with the largest intensity. Assuming a projector with a resolution of $N \times N$ pixels, we would need to capture N^2 images from the line-sensor. Assuming a line-sensor with N pixels, this approach requires N^3 pixels to be read out at the sensor.

We can reduce the acquisition time significantly by using temporal coding techniques, similar to the use of binary/Gray codes in traditional SL (see Figure 3.3). In the DualSL setup, this is achieved as follows. We project each row of every SL pattern sequentially.[3] Given that the row has N pixels, we can find projector-camera correspondences using a binary/Gray code with $\log_2(N)$ projector patterns. For this scanning approach, we would require to read out $\log_2(N)$ frames for each row. Given that the projector has N rows, and the sensor has N pixels, a total of $N^2 \log_2(N)$ pixels will need to be read out at the sensor.

3.1.3 Analysis of temporal resolution

We now show that the temporal resolution of DualSL is the same as that of a traditional SL setup for synchronous-readout sensor, i.e. a conventional sensor. For this analysis, we assume that the goal is to obtain an $N \times N$-pixels depth map. The temporal resolution is defined in terms of the time required to acquire a single depth map. We assume that in a sensor, all pixels share one analog-to-digital converter (ADC). We also assume that the bottleneck in all cases is the ADC rate of the camera. This assumption is justified due to the operating speed of projectors

[3]It would be desirable to scan along epipolar lines of the projector-camera setup, since this avoids illuminating multiple scene points that lie on the pre-image of the same sensor pixel. However, this would be hard in non-rectified setups and hence, row-scanning is a simpler and effective alternative.

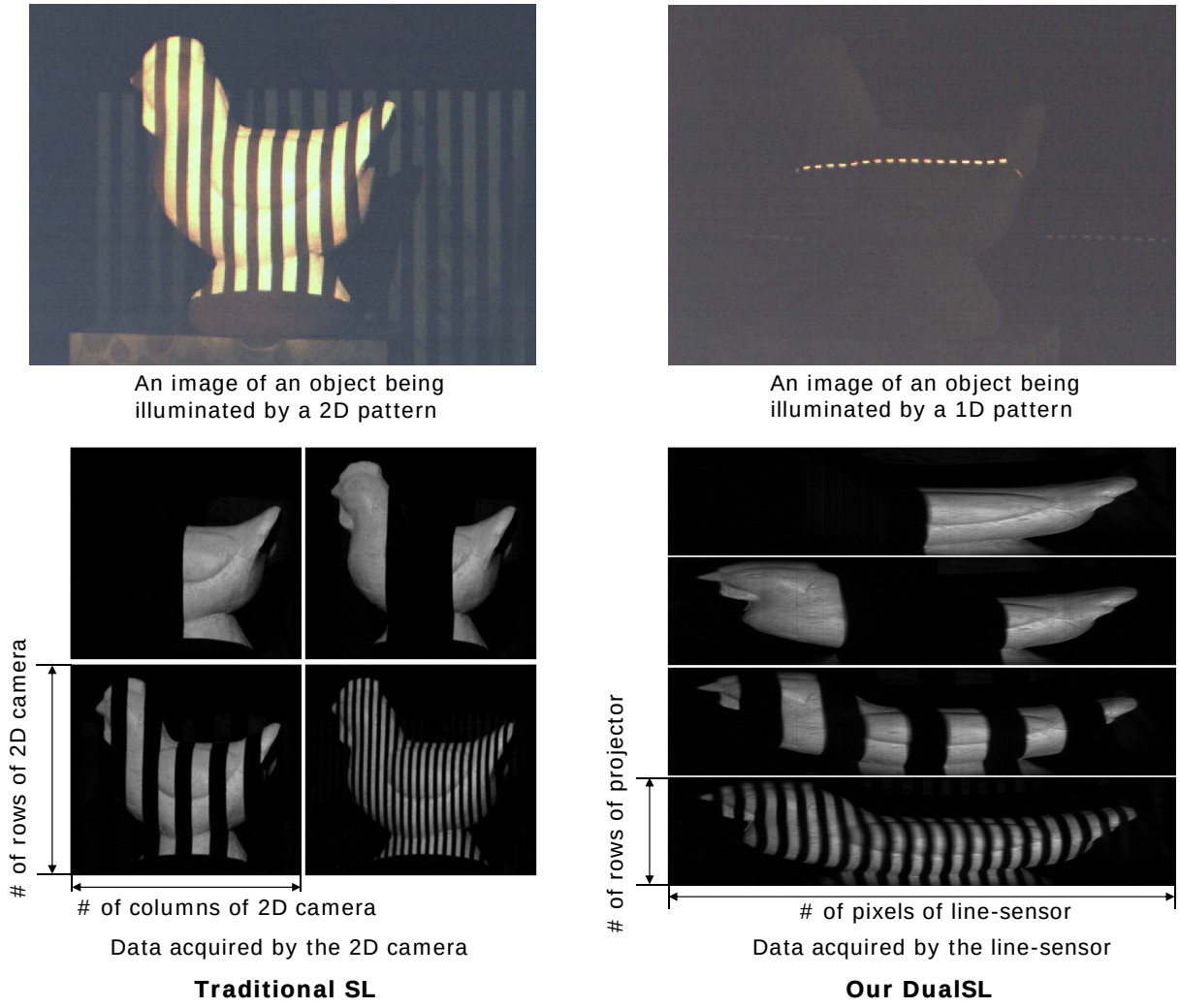


Figure 3.3: **3D scanning using (left) traditional SL and (right) DualSL.** For each setup, we show (top-row) the scene being illuminated by the projector, observed using an auxiliary camera, as well as (bottom-row) measurements made by the cameras. Here the auxiliary camera is not required for triangulation; it is used only for visualization purposes. For the traditional SL setup, these are simply images acquired by the 2D sensor. For DualSL, we stack together 1D measurements made for the same Gray code into individual image.

(especially laser projectors) being many orders of magnitude greater than cameras. Hence, the number of pixels to be read out divided by ADC rate is temporal resolution of the system.

Line striping. The simplest instance of a traditional SL setup is to illuminate projector columns, one at a time. For each projector column, we read out an $N \times N$-pixels image at the camera.

Hence, for a total of N projector columns, we read out N^3 pixels at the ADC. DualSL has an identical acquisition time, equivalent to the readout of N^3 pixels per depth map, when we sequentially scan each projector pixel.

Binary/Gray codes. As mentioned earlier, by scanning one projector-row at a time and using binary/Gray temporal codes, we can reduce the acquisition time to the readout of $N^2 \log_2(N)$ pixels per depth map. This readout time is identical to the amount required for a traditional SL system when using binary/Gray coding of the projector columns, where $\log_2(N)$ images, each with $N \times N = N^2$ pixels, are captured.

In essence, with appropriate choice of temporal coding at the projector, the acquisition time and hence, the temporal resolution, of the DualSL is identical to that of a traditional SL system.

For asynchronous readout using a DVS sensor, the temporal resolution is determined by the minimum time between two readouts. Current 1D DVS sensors typically support approximately one million readouts per second [10]. This would be the achievable limit with a DualSL system using DVS.

3.1.4 DualSL as the optical dual of traditional SL

Consider a traditional SL system involving a 1D projector and a 2D image sensor. Recall that all pixels in a single projector columns are illuminated simultaneously. Let us consider the light transport matrix $\mathcal{L}$ associated with the columns of the projector and pixels on the 2D sensor. Next, let us consider the optical setup whose light transport is $\mathcal{L}^\top$. Under principles of Helmholtz reciprocity, this corresponds to the dual projector-camera system where the dual-projector has the same optical properties as the camera in the traditional SL system, and the dual-camera has the same optical properties as the projector. Hence, the dual-camera integrates light along the planes originally illuminated by the 1D projector in the traditional setup. This dual architecture is the same as that of the DualSL setup and enables estimation of the depth map (and the intensity image) as seen by a camera with the same specifications as the projector.

3.1.5 Performance of DualSL under ambient illumination

Performance of SL systems often suffers in the presence of ambient illumination, in part due to potentially strong photon noise. We can measure the effect of ambient illumination using signal-to-noise ratio (SNR), the ratio of the intensity observed at a camera pixel when a scene point is directly illuminated by the projector to the photon noise caused by ambient illumination. The larger the value of SNR, the less is the effect of ambient illumination on the performance since we

can more reliably provide thresholds to identify the presence/absence of the direct component. We ignore the presence of global components for this analysis.

The hardware prototype used in this chapter uses a DMD-based projector which attenuates a light source, spatially, using a spatial light modulation to create a binary projected pattern. In a traditional SL system, since we read out each camera pixel in isolation, SNR can be approximated as $\frac{P}{\sqrt{A}}$ where P and A are the brightness of the scene point due to the projector and ambient illumination, respectively [30]. Unfortunately, due to integration of light at each line-sensor pixel, SNR of DualSL drops to $\frac{P}{\sqrt{NA}} = \frac{1}{\sqrt{N}}\frac{P}{\sqrt{A}}$ where N is the number of pixels that we sum over. This implies that DualSL is significantly more susceptible to the presence of ambient illumination when we use attenuation-type projectors. This is a significant limitation of our prototype. An approach to address this limitation is to use a scanning laser projector which concentrates all of its light onto a single row of the projector. As a consequence, SNR becomes $\frac{NP}{\sqrt{NA}} = \sqrt{N}\frac{P}{\sqrt{A}}$. In contrast, traditional SL has no gain from using scanning laser projector because it needs projector to illuminate the entire scene.

A more powerful approach is to avoid integrating light and instead use a mirror to scan through scene points in synchrony with a scanning laser projector. Here, we are optically aligning the line-sensor and the illuminated projector pixels to be on epipolar line pairs and is similar to the primal-dual coding system of [65]. This enables acquisition of 3D scans that are highly robust to global and ambient illumination.

3.1.6 Performance of DualSL under global illumination

Global illumination is often a problem when dealing with scenes that have inter-reflections, subsurface- and volumetric-scattering. Similar to ambient illumination, global illumination also leads to loss in performance in decoding at the camera. In a traditional SL system, it is typical that half-the-scene points are illuminated and hence, at any camera pixel, we can expect to receive contributions to the global component from half the scene element. Suppose projector's resolution is $N \times N$, then each camera pixel receives $\frac{N^2}{2}$ projector pixels' global light. In DualSL, even though we only illuminate one projector-row at a time (and hence, fewer illuminated scene points), each camera pixel integrates light along a scene line. Suppose line camera's virtual 2D image has resolution $N \times N$, then each camera pixel receives $\frac{N}{2} \times N = \frac{N^2}{2}$ projector pixels' global light. The global light noise of both SL systems is at the similar level, hence we can expect DualSL has no inferior performance compared to traditional SL.

3.2 Hardware prototype

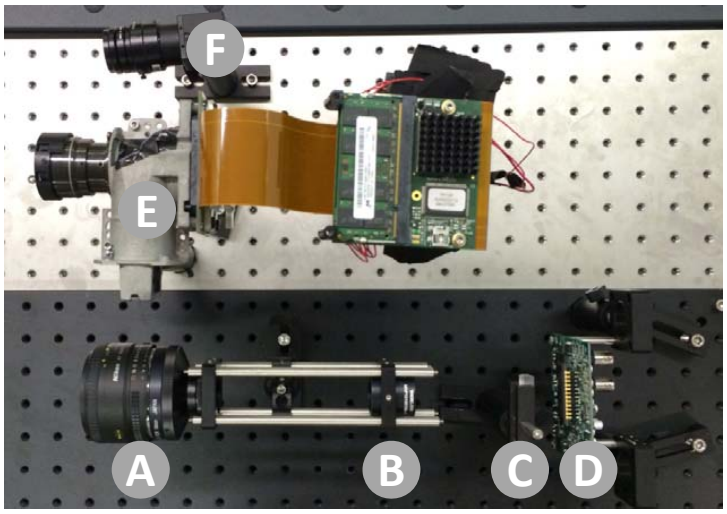
In this section, we present the specifications of our DualSL hardware prototype, shown in Figure 3.4(a). The hardware prototype consists of a 50mm F/1.8 objective lens, a 15mm cylindrical lens, a Hamamatsu S11156-2048-01 line-sensor, and a DMD-based projector built using DLP7000 development kit and ViALUX STAR CORE optics. The line-sensor has 2048 pixels, each of size 14μm $\times 1$mm. The projector resolution is 768×1024 pixels.

We implemented a slightly different optical design from the ray diagrams shown in Figure 3.2. To accommodate the cylindrical lens in the tight spacing, we used a 1:1 relay lens to optically mirror the image plane of the objective lens. This provided sufficient spacing to introduce the cylindrical lens and translational mounts to place it precisely. The resulting schematic is shown in Figure 3.4(b). Zemax analysis of this design shows that the spot-size has a RMS-width of 25μm along the line-sensor and 3.7mm perpendicular to the line-detector. Given that the height of the line-detector pixel is 1mm, our prototype loses 73% of the light. This light loss is mainly due to sub-optimal choice of optical components.

Calibration. The calibration procedure is a multi-step process with the eventual goal of characterizing the light ray associated with each projector pixel, and the plane associated with each line-sensor pixel.

- We introduce a helper 2D camera whose intrinsic parameters are obtained using the MATLAB Camera Calibration Toolbox [13].
- The projector is calibrated using a traditional projector-camera calibration method [48]. In particular, we estimate the intrinsic parameters of the projector as well as the extrinsic parameters (rotation and translation) with respect to the helper-camera's coordinate system.
- To estimate the plane corresponding to each line-sensor pixel, we introduce a white planar board in the scene with fiducial markers at corners of a rectangle of known dimensions. The helper-camera provides the depth map of the planar board by observing the fiducial markers.
- The projector illuminates a pixel to the board which is observed at the line-sensor, thereby providing one 3D-1D correspondence, where the 3D location is computed by intersection of projector's ray and the board.
- This process is repeated multiple times by placing the board in different poses and depths to obtain more 3D-1D correspondences. Once we obtain sufficiently many correspondences, we fit a plane to the 3D points associated with each pixel to estimate its pre-image.

As a by-product of the calibration procedure, we also measure deviations of the computed depth from the ground truth obtained using the helper-camera. The root-mean-square error



A **Objective lens**
Nikkor 50mm F/1.8

B **Relay lens**
Thorlabs MAP10100100-A

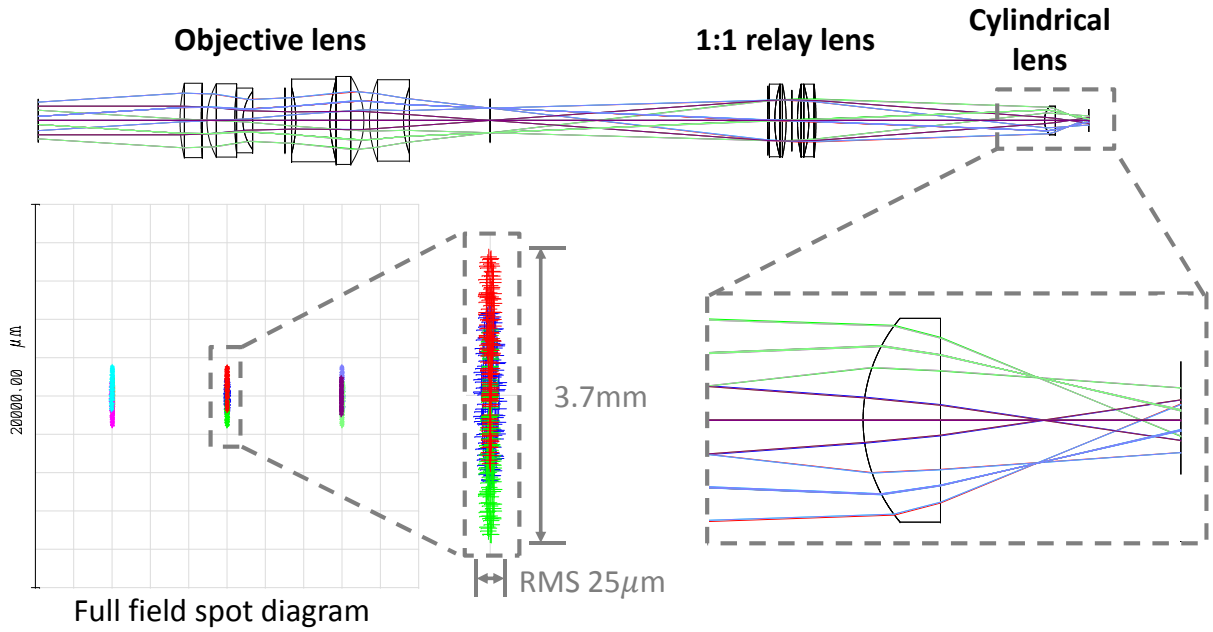
C **Cylindrical lens**
Thorlabs LJ1636L2-A f15mm

D **Line-sensor**
Hamamatsu S11156-2048-01

E **Projector**
DLP7000 DMD and
ViALUX STAR CORE optics

F **2D camera helper**
Point Grey FL3-U3-13E4C-C (not required for triangulation in DualSL)

(a) Hardware prototype with list of components



(b) Analysis of spot size

Figure 3.4: **Our hardware prototype.** We used a 1:1 relay lens to mirror the image plane of the objective lens to provide more space to position the cylindrical lens. The spot-size of the resulting setup has a RMS-width of 25μm along the axis of the line-sensor and 3.7mm across.

(RMSE) over 2.5 million points with depth ranges from 950mm to 1300mm (target is out-of-focus beyond this range) was 2.27mm.

3.3 Experiments

We showcase the performance of DualSL using different scenes. The scenes were chosen to encompass inter-reflections due to non-convex shapes, materials that produce diffuse and specular reflectances as well as subsurface scattering. Figure 3.5 shows 3D scans obtained using traditional SL as well as our DualSL prototype. The traditional SL was formed using the helper-camera used for calibration. We used Gray codes for both systems. To facilitate a comparison of the 3D scans obtained from the two SL setups, we represented the depth map as seen in the projector's view, since both systems shared the same projector. Depth maps from both traditional SL and DualSL were smoothened using a 3×3 median filter. We computed the RMSE between the two depth maps for a quantitative characterization of the difference of the two depth maps. Note that, due to differences in view-points, each depth map might have missing depth values at different locations. For a robust comparison, we compute the RMSE only over points where the depth values in both maps were between 500mm and 1500mm.

For the chicken and ball scenes (rows 1 and 2 in Figure 3.5), both systems get good results and the average difference is smaller than 2mm. For the box scene (row 3), the average difference is only slightly larger in spite of complex geometry of the scene. We fit planes to four different planar surfaces, and the mean deviation from the fitted planes was 0.45mm with the average distance to camera being 1050mm. The porcelain bowl scene (row 4), which has strong inter-reflections, and wax scene (row 5), which exhibits subsurface scattering, have strong global components. The depth maps generated by DualSL in both cases are significantly better than that of the traditional SL. This is because traditional SL illuminates the entire scene, and in contrast, DualSL illuminates a line at a time, thereby reducing the amount of global light. The exact strength of global illumination for a general scene, however, depends on the light transport. For instance, it may be possible to construct scenes where the amount of global light is smaller for conventional SL over DualSL, and vice-versa. However, a formal analysis is difficult because global illumination is scene dependent. Here, the depth map recovered by traditional SL has many "holes" because of missing projector-camera correspondences and the removal of depth values beyond the range of (500, 1500)mm.

Figure 3.6 shows the 3D scans of the five scenes in Figure 3.5 visualized using MeshLab. We observe that DualSL can capture fine details on the shape of the object. We can thus conclude that DualSL has a similar performance as traditional SL for a wide range of scenes, which is immensely satisfying given the use of a 1D sensor.

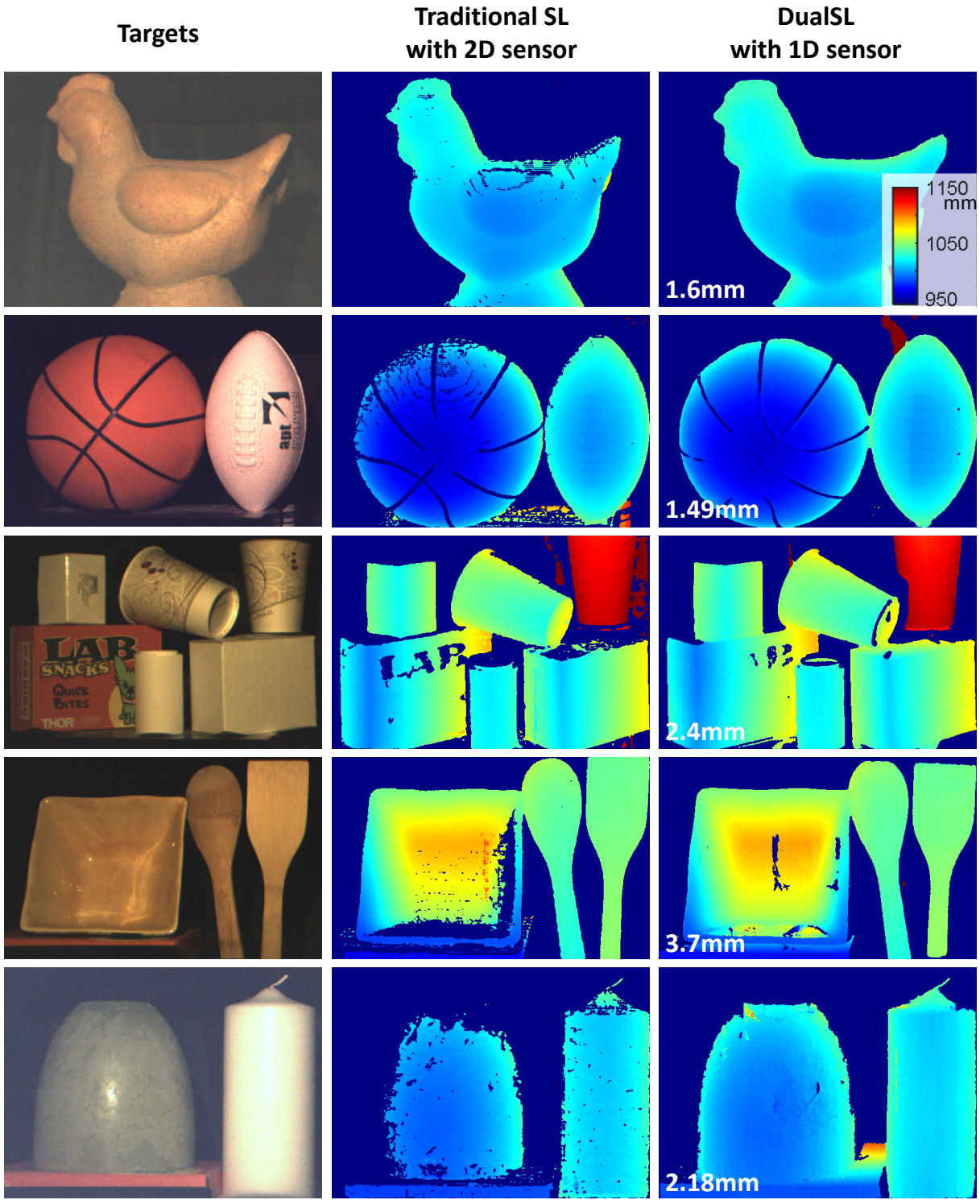


Figure 3.5: **Depth maps of five scenes obtained using traditional SL and DualSL.**

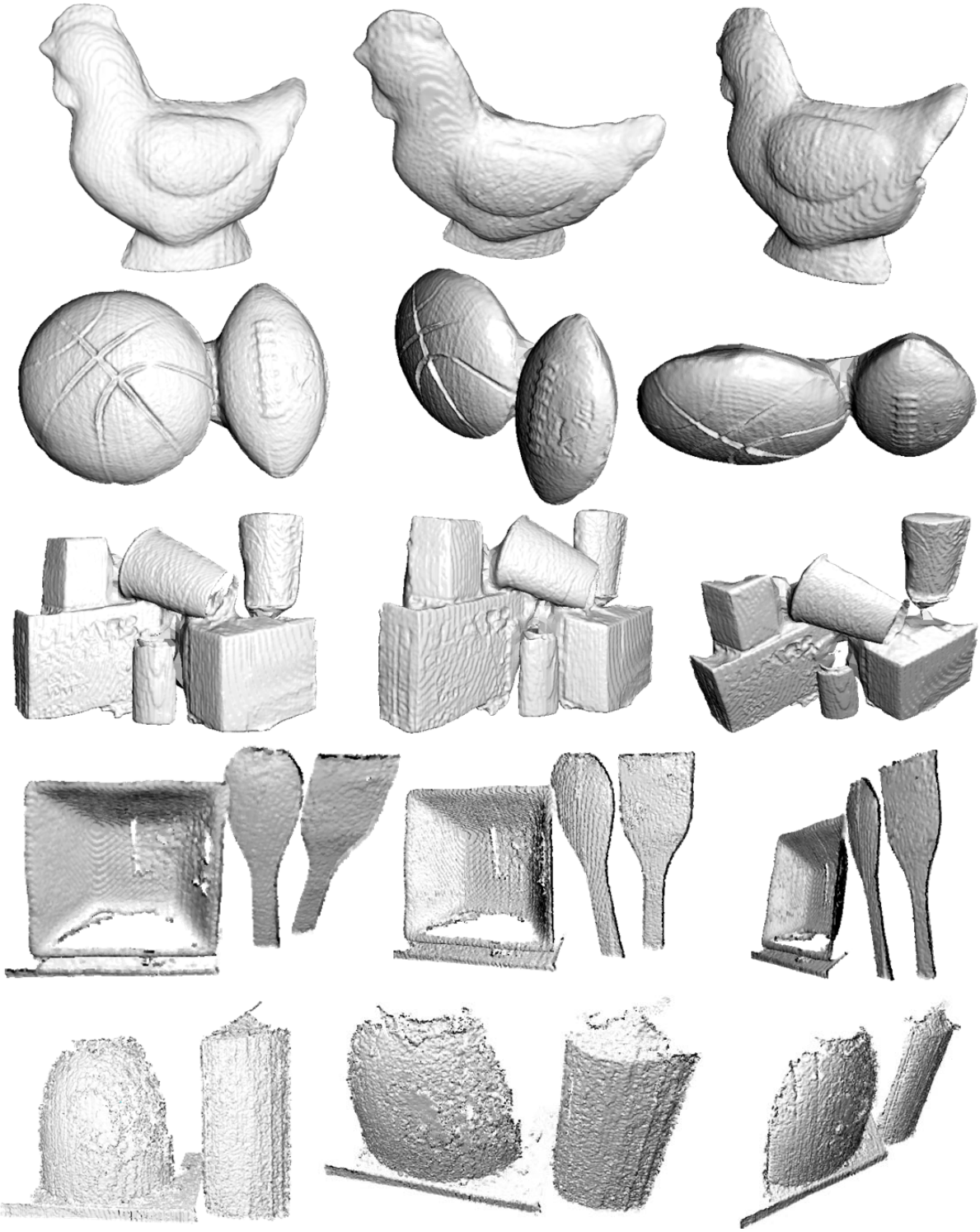


Figure 3.6: **3D reconstructions of scenes scanned by DualSL.**

3.4 Discussion

DualSL is a novel SL system that uses a 1D line-sensor to obtain 3D scans with simple optics and no moving components, while delivering a temporal resolution identical to traditional setups. The benefits of DualSL are most compelling in scenarios where sensors are inherently costly. To this end, we briefly discuss potential applications of DualSL.

Applications of SWIR DualSL. Imaging through volumetric scattering media often benefits via the use of longer wavelengths. SWIR cameras, which operate in the range of 900nm to 2.5μm, are often used in such scenarios (see [84]). The DualSL system design for SWIR can provide an inexpensive alternative to otherwise-costly high-resolution 2D sensors. This would be invaluable for applications such as autonomous driving and fire-fighting operations, where depth sensitivity can be enhanced in spite of fog, smog, or smoke.

High-speed depth imaging using DVS line-sensors. The asynchronous readout underlying DVSs allows us to circumvent the limitations imposed by the readout speed of traditional sensors. Further, the change detection circuitry in DVSs provides a large dynamic range ($\sim$120 dB) that is capable of detecting very small changes in intensity even for scenes under direct sunlight. This is especially effective for SL systems where the goal is simply detecting changes in intensity at the sensor. The MC3D system [57] exploits this property to enable high-speed depth recovery even under high ambient illumination; in particular, the system demonstrated in [57] produces depth maps at a resolution of 128×128 pixels, due to lack of commercial availability of higher resolution sensors. In contrast, a DualSL system using the line-sensor in [10] would produce depth maps of 1024×1024 pixels at real-time video rates. Further, the DualSL system would also benefit from higher fill-factor at the sensor pixels (80% versus 8%).

Active stereo using DualSL. Another interesting modification to the DualSL setup is to enable active stereo-based 3D reconstruction. The envisioned system would have two line-sensors, with its associated optics, and a 1D projector that illuminates one scene plane at a time. By establishing correspondences across the line-sensors, we can enable 3D reconstructions by intersecting the two pre-images of sensor-pixels, which are both planes, with the plane illuminated by the projector. Such a device would provide very high-resolution depth maps (limited by the resolution of the sensors and not the projector) and would be an effective solution for highly-textured scenes.

Chapter 4

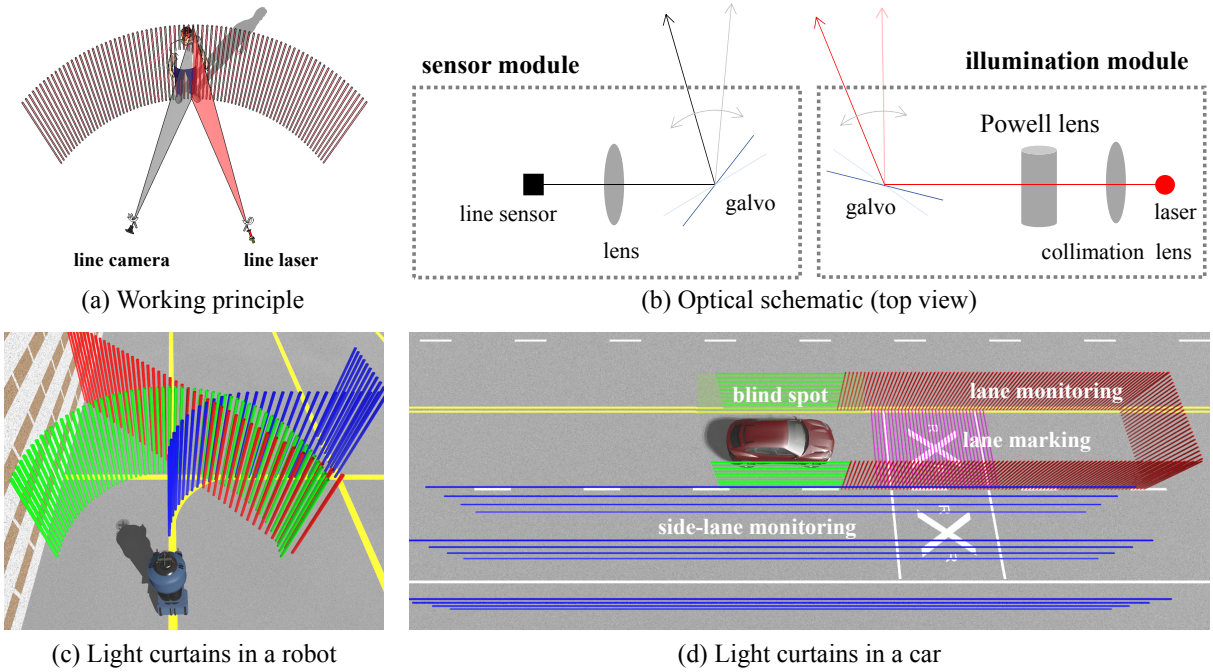
TriLC — Triangulation Light Curtain Proximity Sensor

"Pay no attention to that man behind the curtain!."
— The Wizard of Oz (1939)

3D sensors play an important role in the deployment of many autonomous systems including field robots and self-driving cars. However, there are many tasks for which a full-fledged 3D scanner is often unnecessary. For example, consider a robot that is maneuvering a dynamic terrain; here, while full 3D perception is important for long-term path planning, it is less useful for time-critical tasks like obstacle detection and avoidance. Similarly, in autonomous driving, collision avoidance — a task that must be continually performed — does not require full 3D perception of the scene. For such tasks, a proximity sensor with much reduced energy and computational footprint is sufficient.

We generalize the notion of proximity sensing 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 with the virtual shell, we can solve many tasks pertaining to collision avoidance and situational awareness with little or no computational overhead. We refer to this virtual shell as a *light curtain*. We implement light curtains by triangulating an illumination plane, created by fanning out a laser, with a sensing plane of a line sensor (see Fig. 4.1(a)). In the absence of ambient illumination, the camera senses light only from the intersection between these two planes — which is a line in the physical world. The light curtain is then created by sweeping the illumination and sensing planes in synchrony. This idea can be interpreted as a generalization of pushbroom stereo [9] to active illumination for determining the presence of an object that intersects an arbitrary ruled surface in 3D.

Benefits of triangulation light curtains. (1) *Shape programmability*: The shape of a light curtain is programmable and can be configured dynamically to suit the demands of the immediate



(a) Working principle

(b) Optical schematic (top view)

(c) Light curtains in a robot

(d) Light curtains in a car

Figure 4.1: **We introduce the concept of a programmable triangulation light curtain — a safety device that monitors object presence in 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 synchrony to generate light curtains of varying shapes as demanded by the specifics of an application. (c, d) Example curtains are shown for use on a robot and a car. The device detects objects on the virtual curtains with little computational overhead, making it useful for collision detection and avoidance.

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 adjacent lanes. Similarly, a robot might use a curtain that extrudes its planned (even curved) motion trajectory. Fig. 4.1(c, d) shows various light curtains for use in robots and cars.

(2) *Adaptability of power and exposure*: Given an energy budget, in terms of average laser power, exposure time, and refresh rate of the light curtain, we can allocate higher power and exposure to lines in the curtain that are further away to combat inverse-square light fall-off. This is a significant advantage over traditional depth sensors that typically expend the same high power in all directions to capture a 3D point cloud of the entire volume.

(3) *Performance in scattering media*: The optical design of the light curtain shares similarities with confocal imaging [55] in that we selectively illuminate and sense a small region. When imaging in scattering media, such as fog and murky waters, this has the implicit advantage that

back scattered photons are optically avoided, thereby providing images with increased contrast.
(4) *Performance in ambient light*: A key advantage of programmable light curtains is that we can concentrate the illumination and sensing to a thin region. Together with the power and exposure adaptability, this enables significantly better performance under strong ambient illumination, including direct sunlight, at large distances ($\sim$20-30m). The performance increases under cloudy skies and indoors to $40m$ and $50m$ respectively.
(5) *Dynamic range of the sensor*: 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.
(6) *Sensor types*: Any line sensor could be used with our design including intensity sensors (CMOS, CCD, InGaAs [31]), time-of-flight (ToF) sensors (correlation, SPAD [15]), and neuromorphic sensors (DVS) [51].

Limitations of triangulation light curtains. (1) *Line-of-sight to the light curtain.* Our technique requires that the laser and sensor have line-of-sight to the light curtain. When this is not the case, the intersection of line camera plane and laser sheet is inside the object and is not seen by the camera; so the technique will fail to detect objects intersecting with the light curtain. This can be resolved partially by determining if there is an object between the system and the desired curtain using a curtain whose "shape" is a sparse set of lines. (2) *Interference*: When simultaneously used, several devices can accidentally interfere with each other. This can be alleviated by using a time-of-flight sensor with different light amplitude frequencies or adding a second line camera or several line lasers. (3) *Fast motion*: Objects that move at high speeds might be able to avoid detection by crossing the light curtain in a region between two successive scans. However, the target would need to be highly maneuverable to accomplish this given our high scan rate.

4.1 Related work

Safety devices and proximity sensors. Devices such as "light curtains" [1], pressure-sensitive mats, and conductive rubber edges are common safeguards used in homes and factories. They are designed to stop the operation of a machine (e.g. a garage door) automatically upon detection of a nearby object. Such proximity sensing devices use many physical modalities to detect presence including capacitance, inductance, magnetism, infrared, ultrasound, Radar and LIDAR.

Depth gating. Temporal gating [7, 29] uses a combination of a pulsed laser and a camera with gated shutter, typically in the range of pico- to nano-seconds. By delaying the exposure of the camera with respect to the laser pulse, the device enables depth selection. An alternate

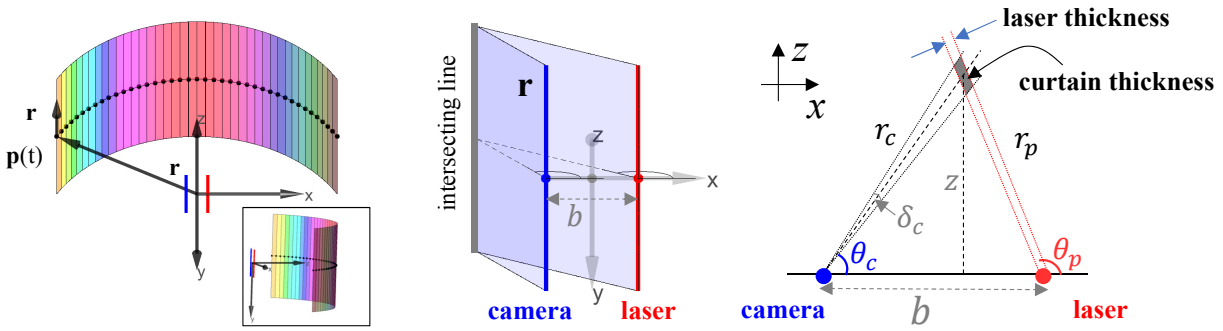
approach relies on on-chip implementations of temporal modulated light sources and sensors [89]. Our technique can be interpreted as a specialized instance of primal-dual coding [64] where simultaneous coding is performed at the camera and projector to probe the light transport matrix associated with a scene, including implementing depth gating.

The proposed technique is inspired from existing work on robust depth scanning in presence of global and ambient light. Early work on imaging through turbid includes scanning confocal microscopy [55] and light sheet microscopy [99] which both illuminate and image the same depth with very shallow depth of focus and block out-of-focus light i.e. scattered light during image formation. Recent work for 3D measuring in scattering media have line striping-based [61] and SPAD-based [80] methods through analyzing spatial or temporal distribution of photons, respectively. An effective method for handling ambient light is concentrating the active light source's output and scanning. Gupta et al. [30] adjust the light concentration level in a structured light system adaptively according to the ambient light level. Episcan3D [65] and EpiToF [4] scan a line laser and a line sensor in synchrony such that the illumination and the sensing planes are coplanar. Finally, there are many benefits in using special functionality sensors for imaging and depth scanning; examples include the DVS sensor for structured light point scanning at high speed and dynamic range [57] as well as the short-wave infrared sensor for enhanced eye-safe property and decreased scattering [97, 100]. Such sensors can be easily incorporated in our light curtains for additional robustness and capabilities.

4.2 Geometry of triangulation light curtains

The proposed device consists of a line scan camera and a line scan laser, as shown in Fig. 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 illuminations, the sensor measures light scattered by any object on the line. By rotating both the camera and the laser at a high speed, we can sweep the intersecting line to form any ruled surface [2]. We refer to this ruled surface, on which we detect presence of objects, as the light curtain. The resulting device is programmable, in terms of its light curtain shape, and flexible, in terms of being able to vary laser power and camera exposure time to suit the demands of an application. In this section, we present the mathematical model for a light curtain as a function of the camera and laser rotation axes. Then, we describe how to estimate the camera and laser rotation angles for a particular light curtain shape and show several examples.

We consider the case where the camera and laser can each be rotated about a single fixed parallel axis (see Fig. 4.2). This can be easily implemented by placing two 1D galvo mirrors,



(a) Parallel rotation axis (b) Coordinate system (c) Top view

Figure 4.2: **Geometry of triangulation light curtains.** (a) Viewing and illumination geometry of a triangulation light curtain generated by rotating the laser light plane and sensor plane about parallel axes $\mathbf{r}$. The intersection line is also parallel to the two rotation axes. (b, c) The coordinate frame and top view showing various parameters of interest. Note that changing θ_c and θ_p in synchrony generates light curtains with different shapes. (c) The finite sizes of camera pixels and finite thickness of laser sheet leads to a thick light curtain upon triangulation.

one each in the front of the line sensor and the laser, respectively. Let the camera and laser rotation axes be $\mathbf{r}$. We observe that intersecting line in the curtain will also be parallel and of the form

$$\mathbf{p}_0 + u\mathbf{r},$$

where $\mathbf{p}_0$ is any 3D point on the line and $u \in (-\alpha, \alpha)$ is the offset along the axis of rotation (see Fig. 4.2(a, b)). Then, the light curtain $s(t, u) \subset \mathbb{R}^3$ is obtained by sweeping the intersection line such that

$$s(t, u) = \mathbf{p}(t) + u\mathbf{r},$$

where $\mathbf{p}(t) \in \mathbb{R}^3$ is a 3D path that describes the points scanned by the center pixel on the line sensor and $t \in [0, 1]$ is the parameterization of the path.

Given a light curtain $s(t, u)$, we next show how to compute the rotation angles of the camera and laser respectively. Without loss of generality, we assume that the origin of our coordinate system is at the midpoint between the centers of the line camera and the laser. We further assume that the rotation axes are aligned along the y-axis and that the 3D path can be written as $\mathbf{p}(t) = [x(t), 0, z(t)]^\top$. To achieve this light curtain, suppose that the laser rotates about its axis with an angular profile of $\theta_p(t)$, where the angle is measured counter-clockwise with respect to the x-axis. Similarly, the line sensor rotates with an angular profile of $\theta_c(t)$. Let b be the baseline

between the laser and the line camera. We can derive $\theta_c(t)$ and $\theta_p(t)$ as

$$\begin{bmatrix}\theta_c(t)\\ \theta_p(t)\end{bmatrix} = \begin{bmatrix}\text{atan2}(z(t), x(t) + b/2)\\ \text{atan2}(z(t), x(t) - b/2)\end{bmatrix}. \tag{4.1}$$

Fig. 4.1(c, d) shows different types of light curtains for use on robots and cars and Fig. 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 (4.1). The first two light curtains — a cylindrical curtain for safety zone monitoring and a curved curtain for monitoring obstacles along a path — are envisioned for use on robots. The next four kinds of curtains are envisioned for use on cars: $\sqcap$-shape lane curtain to detect proximity vehicles in front as well as those that encroach on to the lane being used (row 3), side curtain to cover blind spot zones (row 4), and a discrete sampling of neighbor lane's condition to identify presence of a vehicle in that volume (row 5). As noted in the introduction, light curtains also offer improved contrast in the presence of scattering media and hence, a curtain observing lane markings (row 6) is especially useful on foggy days.

Our proposed device can also be configured with the line sensor and laser rotating over non-parallel axes or even with each of them enjoying full rotational degree of freedom. These configurations can generate other kinds of ruled surfaces including, for example, a Möbius strip. For applications that require light curtains of arbitrary shapes, we would need to use point sources and detectors; this, however, comes at the cost of large acquisition time that is required for two degrees of freedom in scanning. On the other hand, 2D imaging with a divergent source does not require any scanning but has poor energy efficiency and flexibility. In comparison, line sensors and line lasers provide a unique operating point with high acquisition speeds, high energy efficiency, and a wide range of light curtain shapes.

4.3 Optimizing triangulation light curtains

We now quantify parameters of interest in practical light curtains — for example their thickness and signal-to-noise ratio (SNR) of measured detections — and approaches to optimize them. In particular, we are interested in minimizing thickness of the curtain as well as optimizing exposure time and laser power for improved detection accuracy when the curtain spans a large range of depths.

Thickness of light curtain. The light curtain produced by our device has a finite thickness due to the finite sizes of the sensor pixels and the laser illumination. Suppose that the laser plane has a thickness of Δ_L meters and each pixel has an angular extent 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 given as an area of a parallelogram shaded in Fig. 4.2(c), which evaluates to

$$A = \frac{r_c^2 r_p}{z} \frac{\delta_c \Delta_L}{b} \tag{4.2}$$

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

$$U = \frac{A}{\Delta_L} = \frac{r_c^2 r_p}{z} \frac{\delta_c}{b}. \tag{4.3}$$

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 the thickness of the curtain that involves a single translation. This is important since different applications often have differing needs on the thickness of the curtain. A larger baseline helps in achieving very thin curtains which is important when there is a critical need to avoid false alarm. On the other hand, thick curtains that can be achieved by having a smaller baseline are important in scenarios where mis-detections, especially those arising from the discreteness of the curtain, are to be avoided.

In Fig. 4.3(d), we visualize the thickness of various light curtains. We set the camera's pixel width to 50μm with a lens of focal length $f = 6$mm, thereby giving us a value for $\delta_c = \frac{50\mu m}{f} = 0.008$ radians. The baseline b was set to $300mm$ for the first two rows, $2m$ for the third row, and $200mm$ for the last three rows. For example, for the lane marking observer (last row), a thickness of $2m$ at the furthest point is beneficial to deal with uneven roads.

Adapting laser power and exposure. Another key advantage of our light curtain device is that we can adapt the power of the laser or the exposure time for each intersecting line to compensate for light fall-off, which is inversely proportional to the square of the depth. In a traditional projector-camera system, it is common to increase the brightness of the projection to compensate for light fall-off, so that far-away scenes points can be well illuminated; however, this would imply that points close to the camera get saturated easily thereby requiring a high dynamic range camera. In contrast, our system has an additional degree of freedom where-in the laser's power and/or the camera's exposure time can be adjusted according to depth so that light fall-off is compensated to the extent possible under the device constraints and eye safety. Further, because our device only detects presence or absence of objects, in an ideal scenario where albedo is the same everywhere, the laser can send small amounts of light to just overcome the camera's readout noise or the photon noise of ambient light, and only a 1-bit camera is required. Fig. 4.3(e) shows light fall-off and depth-adaptive exposure time or laser power for all the light curtains.

Combining with time-of-flight sensors. The analysis in (4.3) indicates that $U \approx \frac{z^2\delta_c}{b}$ when $r_c, r_p \approx z$ and 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 [50] 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 of frequency f_m Hz, 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

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

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 Fig. 4.9(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 [89]. 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.

4.4 Hardware prototype

Our prototype device has two modules (see Fig. 4.4). At the sensor end, we use an Alkeria NECTA N2K2-7 line intensity sensor with a 6mm $f/2$ S-mount lens whose diagonal FOV is 45° and image circle has a diameter of 7mm. The line camera has 2048×2 pixels with pixel size $7\mu m \times 7\mu m$; we only use the central 1000×2 pixels due to the lens' limited circle of illumination. The line sensor is capable of reading out $95,057$ lines/second. We fit it with an optical bandpass

filter, $630nm$ center wavelength and $50nm$ bandwidth to suppress ambient light. For the ToF prototype, we used a Hamamatsu S11961-01CR sensor in place of the intensity sensor whose pixel size is $20\mu m \times 50\mu m$. We used a low cost 1D galvo mirror to rotate the camera's viewing angle. The 2D helper FLIR camera (shown in the middle) is used for one-time calibration of the device and visualizing the light curtains in the scene of interest by projecting the curtain to its view.

At the illumination side, we used a Thorlabs L638P700M 638nm laser diode with peak power of $700mW$. The laser is collimated and then stretched into a line with a $45°$ Powell lens. As with the sensor side, we used a 1D galvo mirror to steer the laser sheet. The galvo mirror has dimension of $11mm \times 7mm$, and has $22.5°$ mechanical angle and can give the camera and laser $45°$ FOV since optical angle is twice the mechanical angle. It needs $500\mu s$ to rotate $0.2°$ optical angle. A micro-controller (Teensy 3.2) is used to synchronize the camera, the laser and the galvos. Finally, we aligned the rotation axes to be parallel and fixed the baseline as $300mm$. The FOV is approximately $45° \times 45°$.

Ambient light. The camera also measures the contribution from the ambient light illuminating the entire scene. To suppress this, we capture two images at each setting of the galvos — one with and one without the laser illumination, each with exposure of $100\mu s$. Subtracting the two suppresses the ambient image.

Scan rate. Our prototype is implemented with galvo mirrors that take about 500 μs to stabilize, which limits the overall frame-rate of the device. Adding two $100\mu s$ exposures for laser on and off, respectively, allows us to display 1400 lines per second. In our experiments, we design the curtains with 200 lines, and we sweep the entire curtain at a refresh rate of 5.6 fps. To increase the refresh rate, we can use higher-quality galvo mirrors with lower settling time, a 2D rolling shutter camera as in [65] or spatially-multiplexed line camera by digital micromirror device (DMD) [96] in the imaging side and a MEMS mirror on the laser side so that 30 fps can be easily reached.

Calibration. Calibration is done by identifying the plane in the real world associated with each setting of laser's galvo and the line associated with each pair of camera's galvo and camera's pixel index.

To calibrate laser's galvo, we do the following:

- introduce a 2D helper camera whose intrinsic parameters are pre-calibrated [13];
- introduce a white planar board in the scene with fiducial markers at corners of a rectangle with known dimensions so that the board's plane equation can be known from the 2D helper camera;

- let the laser illuminate a line on the board which is observed by the 2D camera. 3D location of the dots in the line can be computed by intersection of 2D camera's ray and the board;
- move the board to another depth and get 3D location of another line;
- fit a plane based on these two lines, and repeat for the next galvo position.

  To calibrate line camera's galvo, we do the following
- introduce a 2D helper projector;
- let the helper projector project horizontal and vertical gray code patterns to the same white board which are observed by the scanning line camera and 2D helper camera;
- do the decoding on gray coded images;
- find the correspondence between $\{$galvo's position, line camera pixel$\}$ pair and helper camera's pixel; 3D location of the point corresponded to each $\{$galvo's position, line camera pixel$\}$ pair can be known by the helper camera;
- move the board to another depth and get another point;
- fit a line based on two points for each pair.

In reality, we usually place the board at more than two depths to make the calibration more accurate.

Laser safety. Laser safety, especially eye safety, puts a limit on the maximum energy that a laser system can emit to the scene, hence restricting the maximum working range. The laser's power density in W/m^2 at a distance smaller than Maximal Permissible Exposure (MPE) is considered eye safe. MPE is a function of laser wavelength, the subtended angle to the eye and duration [6].

Suppose our system scans the scene in the fashion that the laser is on for $100us$ and off for next $100us$, and waits for $500us$ for the mirror to rotate to next position and repeat, these three conditions must be below MPE limit: (a) a single pulse with $100us$, (b) pulses of consecutive rows which illuminate the pupil for short depth, and (c) continuous operation over long period like a few hours for stationary eyes.

The specification of our laser is the following: peak power $700mW$, wavelength $638nm$, spot diameter $2mm$, fanout angle $\pi/4$ radians, thickness divergence angle 0.0021 radians (in last section thickness of laser is assumed to be constant for ease of derivation), and 200 line scans uniformly distributed in $\pi/4$'s angle of view. Duty cycle for single pulse is 1, and for one row is $100us/700us = 1/7$. The actual power density is computed as $(\text{duty cycle} \times \text{peak power}/\text{illuminated area})$. Allowed power density MPE is calculated according to ANSI Z136.1 Table 5b in [6]. We plot MPE and actual power density in Fig. 4.5. It turns out requirement (a) gives the shortest safety distance which is $0.72m$ for our system. The power can be

reduced depending on the application, for example, indoors we usually use 30% of the maximum power. If our red laser is replaced by a near-infrared or short-wave infrared laser, the safety distance is even shorter.

We can also use multiple line lasers. Although most of the time the light curtain is set to be several meters away, it must be guaranteed to be eye-safe at very close range (usually half meter or one meter away) and there is a huge gap between the usage distance and safety distance. By distributing power to multiple line lasers which intersect or "focus" at the same line in 3D, it can be eye-safe easily at close range as well as focused distance. In other words, the device is more eye-safe or higher total power can be used. Interference between multiple devices is also relieved by using this method.

4.5 Results

Evaluating programmability of triangulation light curtains. Fig. 4.6 shows the results of implementing various light curtain shapes both indoors and outdoors. When nothing "touches" the light curtain, the image is dark; when a person or other intruders touch the light curtain, it is immediately detected. The small insets show the actual images captured by the line sensor (after ambient subtraction) mosaicked for visualization. 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. Fig. 4.7(a) shows the detections of a white board placed at different depths (verified using a laser range finder) in bright sunlight ($\sim 100klx$). The ambient light suppression is good even at 25m range. Under cloudy skies, the range increases to more than 40m. Indoors, the range is approximately 50m. These ranges assume the same refresh rate (and exposure time) for the curtains. In Fig. 4.7(b), we verify the accuracy of the planar curtains by sweeping a fronto-parallel curtain over a dense set of depths and visualizing the resulting detections as a depth map.

Performance under volumetric scattering. When imaging in scattering media, light curtains provide images with higher contrast by suppressing backscattered light from the medium. This can be observed in Fig. 4.8 where we show images of an outdoor road sign under heavy smoke from a planar light curtain as well as a conventional 2D camera. The 2D camera provides images with low contrast due to volumetric scattering of the sunlight, the ambient source in the scene.

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 Fig. 4.9. Because 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.

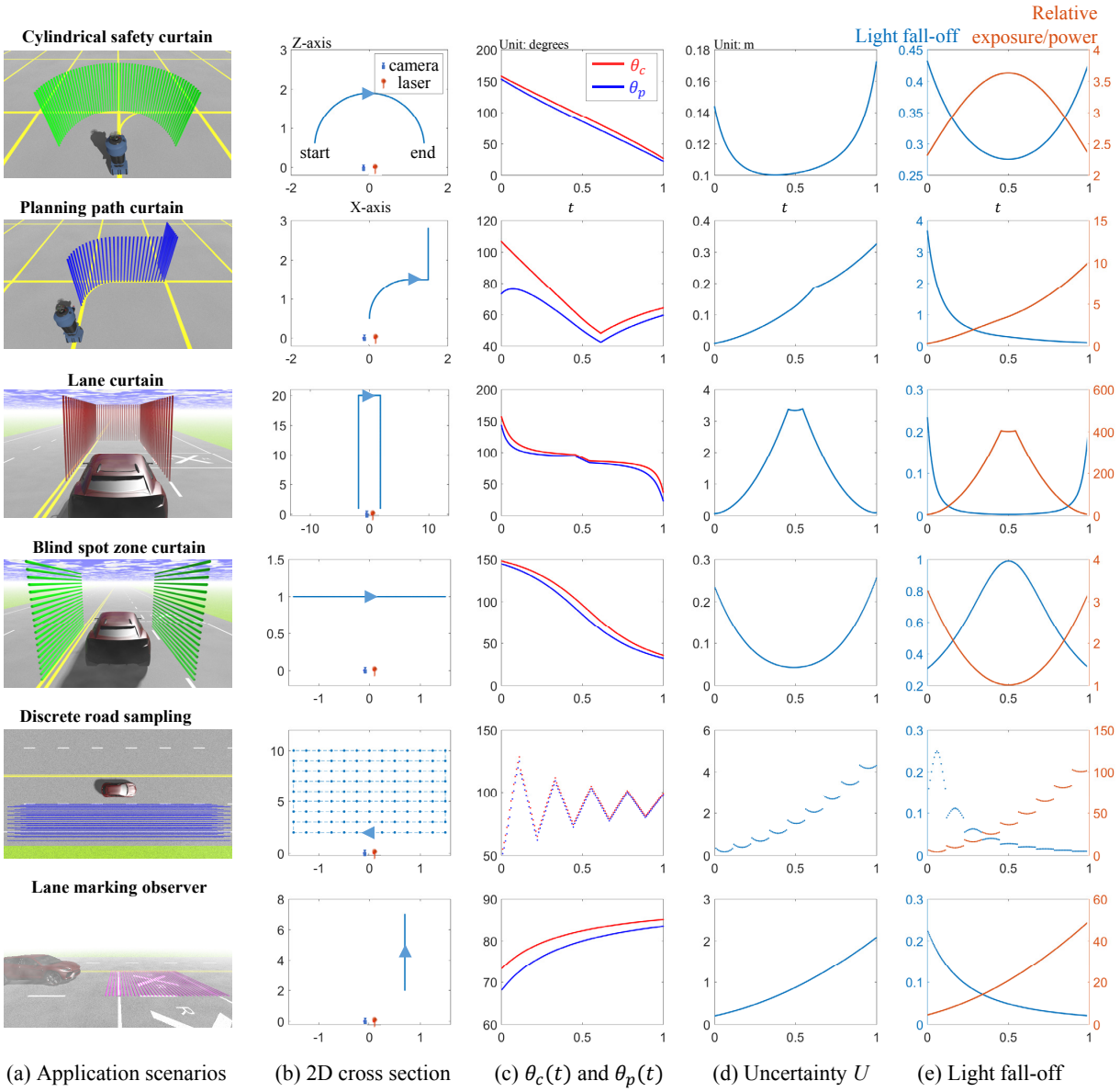
Adapting laser power and exposure. Finally, we showcase the flexibility of our device in combating light fall-off by adapting the exposure and/or the power of the laser associated with each line in the curtain. We show this using depth maps sensed by sweeping fronto-parallel curtains with various depth 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 value. In Fig. 4.10(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 intensity map and depth map for each strategy. Notice the saturation and darkness in intensity maps with the constant exposure strategies and even 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. Fig. 4.10(b) shows a result on 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.

4.6 Discussion

A 3D sensor like LIDAR is often used for road scene understanding today. Let us briefly see how it is used. The large captured point cloud is registered to a pre-built or accumulated 3D map with or without the help of GPS and IMU. Then, dynamic obstacles are detected, segmented, classified using complex, computationally heavy and memory intensive deep learning approaches [26, 106]. These approaches are proving to be more and more successful but perhaps all this computational machinery is not required to answer questions, such as: “Is an object cutting across into my lane?”, or “Is an object in a cross-walk?”.

This paper shows that our programmable triangulation light curtain can provide an alternative solution that needs little to no computational overhead, and yet has high energy efficiency and flexibility. We are by no means claiming that full 3D sensors are not useful but are asserting that light curtains are an effective addition when depths can be pre-specified. Given its inherent

advantages, we believe that light curtains will be of immense use in autonomous cars, robotics safety applications and human-robot collaborative manufacturing.



(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 (baselines in the first two rows was 300mm, third row was 2m, and the remaining rows was 200mm). 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 $50\mu m$ pixel width and focal length 6mm. (e) Light fall-off and the corresponding adaptation of exposure/power to compensate for it.

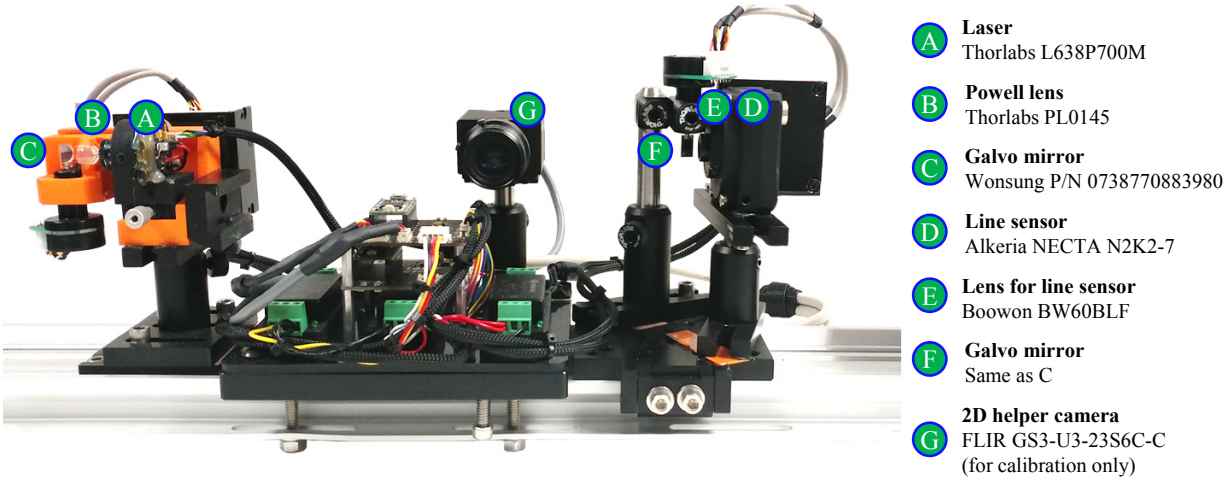
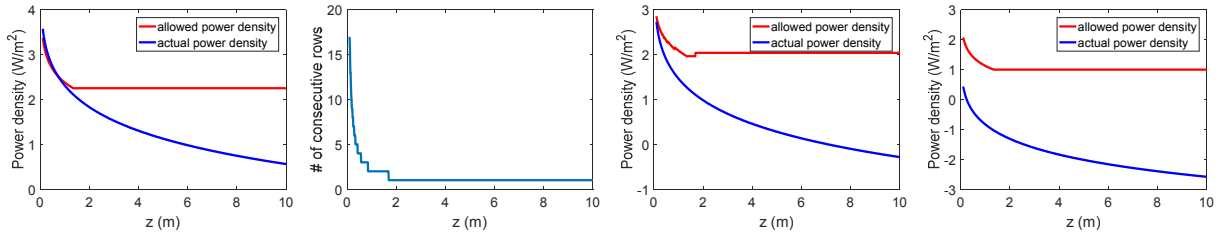


Figure 4.4: **Hardware prototype with components marked.** The prototype implements the schematic shown in Fig. 4.1(b). The 2D helper camera marked as (G) is used for a one-time calibration. The light curtain is mapped to the 2D helper camera's view to aid visualization.



(a) Single pulse test (b) Consecutive rows test (c) Long time run (8 hrs) test

Figure 4.5: **Laser safety calculation.** y-axis in the power density plots is in log-scale.

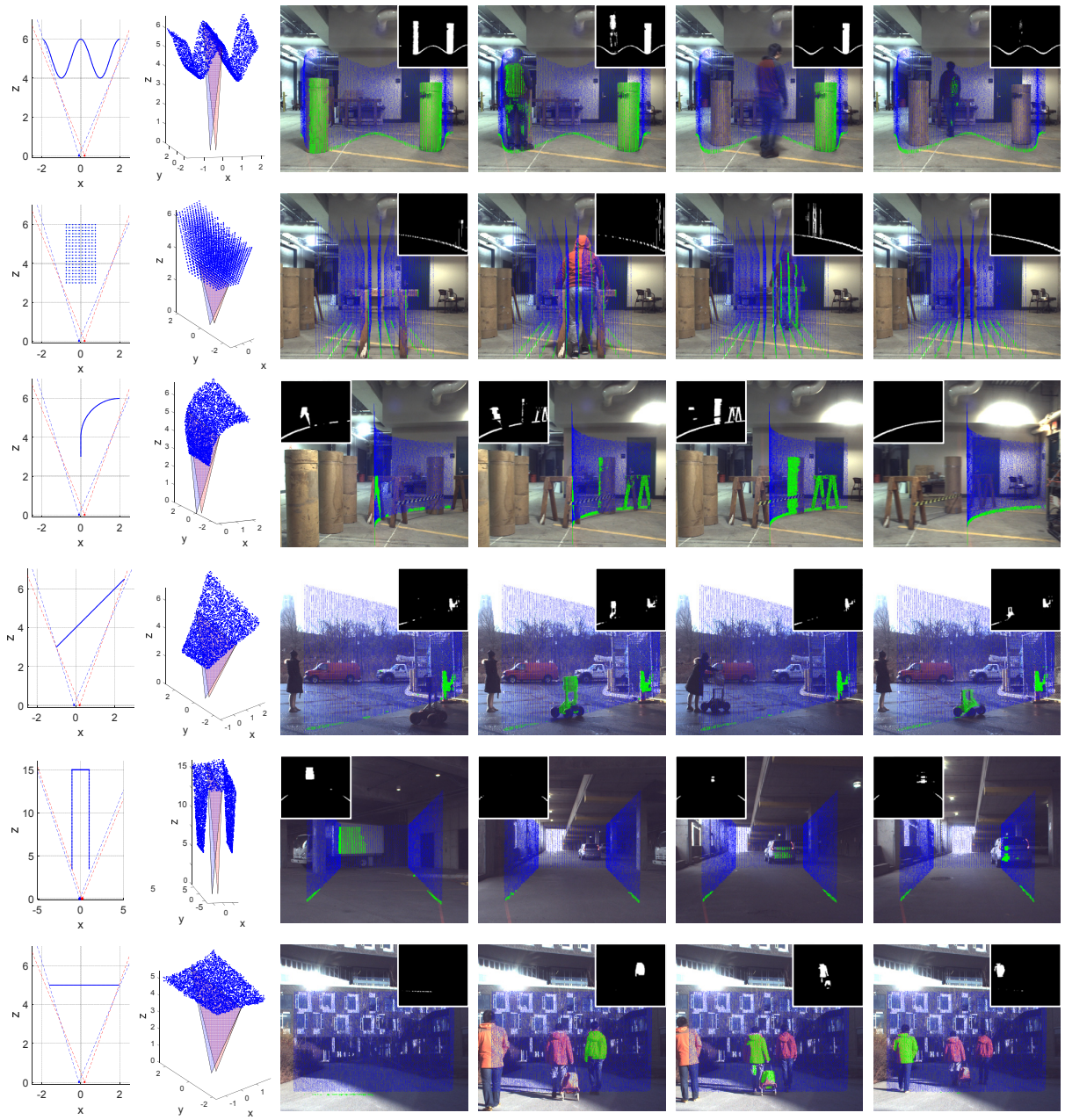
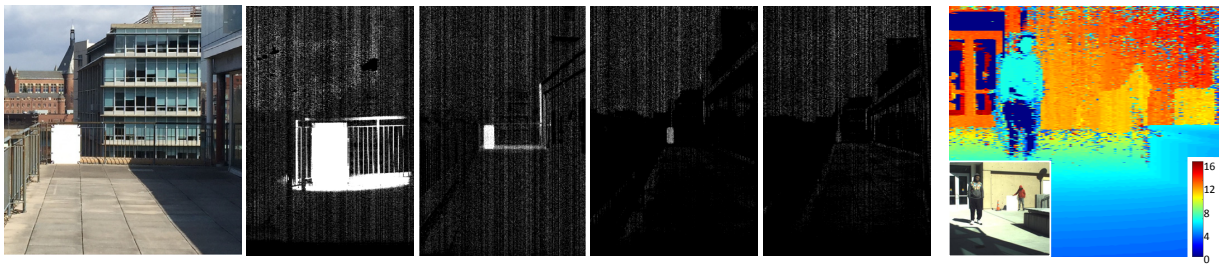


Figure 4.6: **Generating light curtains with different shapes.** The images shown are from the 2D helper camera's view with the light curtain rendered in blue and detections rendered in green. The curtains are shown both indoors and outdoors in sunlight. The insets are images captured by the line sensor as people/objects intersect the curtains. The curtains have 200 lines with a refresh rate of 5.6 fps. The curtain in the second row detects objects by sampling a volume with a discrete set of lines.

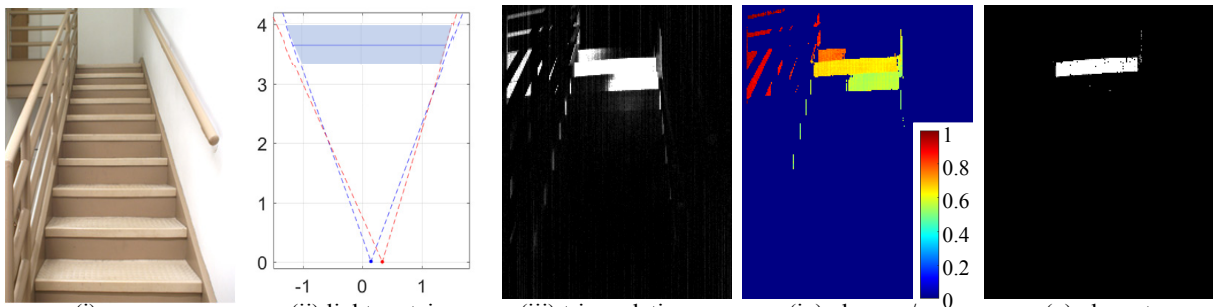


(a) Fronto-parallel planar light curtain at different depths

(b) Depth map by sweeping light curtains

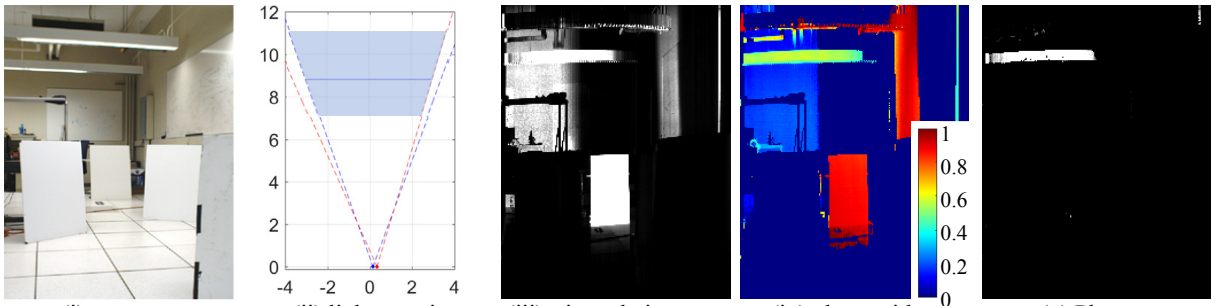
Figure 4.7: **Performance under bright sunlight** ($100klux$) **for two different scenes.** (a) We show raw light curtain data, without post-processing, at various depths from a white board. Notice the small speck of the board visible even at 35m. (b) We sweep the light curtain over the entire scene to generate a depth map.

Figure 4.8: **Seeing through volumetric scattering media.** Light curtains suppress backscattered light and provide images with improved contrast.



(i) scene (ii) light curtain with thickness (iii) triangulation (iv) phase w/o wrapping (v) phase + triangulation

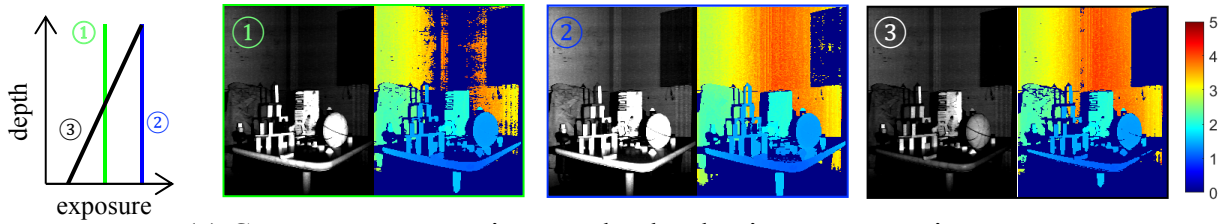
(a) Staircase scene



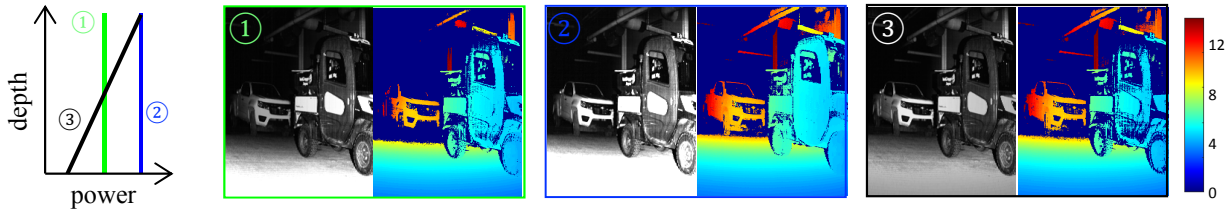
(i) scene (ii) light curtain with thickness (iii) triangulation (iv) phase with wrapping (v) Phase + triangulation

(b) Lab scene

Figure 4.9: **Using line CW-TOF sensor.** 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.



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



(b) Constant power vs. depth-adaptive power

Figure 4.10: **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).

Chapter 5

Conclusions

"Measure what is measurable and make measurable what is not so."
— Galileo Galilei

Throughout this dissertation, we are trying to solve some of the problems with imaging technology. For example, 2D high-resolution non-visible light imaging is difficult because of the high cost in sensor manufacturing process. 3D active illumination imaging may not work under intense ambient light which can submerge the active light. Imaging in scattering media has very low contrast, because the sensor receives much light scattered by the medium itself. Existing imaging architecture designs based on 2D sensors or based on photo-detectors / point scanning, have necessary drawbacks that limit their practicability. The 2D sensor-based designs are fast but expensive, and the point scanning-based designs are robust but the acquisition speed is slow.

We proposed 1D sensor-based designs that provide the robustness and economy of point scanning-based designs as well as the acquisition speed or density of 2D sensor-based designs. In Chapter 2, we proposed LiSens which can be used for non-visible light high-resolution imaging. It uses a 1D sensor and a 2D SLM to achieve 2D imaging by realizing multiplexing in the direction perpendicular to the 1D sensor. Since only a 1D sensor is needed, the cost is very low, and the resolution and frame rate of the entire device is the same as a full frame 2D sensor. In Chapter 3, we proposed DualSL. The traditional structured light can be deemed as using a 2D sensor and a 1D projector. Through the reversibility of light rays, called Helmholtz reciprocity, we can also do the same work with a 1D sensor and a 2D projector. DualSL uses this idea and its performance, including resolution, frame rate, and tolerance to ambient light and global light, is not inferior to traditional structured light systems. Because it uses a 1D sensor, it is possible to use some special functionality sensors, such as SWIR and DVS sensors. In Chapter 4, we proposed TriLC which forms a virtual shell in 3D by rotating a 1D sensor and a 1D laser in synchrony. Any object that touches this surface can be easily detected, and hence the device can be used for proximity detection. Since the device concentrates light into a line, compared to area

illumination devices, it works more reliably under strong sunlight. In the scattering medium, the device blocks a lot of backscattered light which happens at medium particles during the imaging process, thus greatly improving the contrast.

5.1 Comparisons

There are other methods proposed before and during the completion of this dissertation. We compare them with our devices in this section. An important taxonomy of imaging devices is based on the dimension of the sensors used (point, 1D, or 2D with low / high-resolution) and how the active light is distributed (point, line or area illumination). Based on such taxonomy, we summarize the traditional designs, the recently proposed designs and ours in Table 5.1.

Passive imaging. The alternatives to an extremely costly high-resolution 2D sensor can be designs based on low-resolution sensors with some super-resolution mechanisms. One is scanning (whether based on 2D / 1D galvo mirror, mobile platform like push broom scanner [101], or translation stage like MSRA gigapixel camera [11]), the other is multiplexing by an SLM, such as single pixel camera (SPC) [21], LiSens, and FPA-CS [17] which can be deemed as replacing photo-diode by a low-resolution 2D sensor in the SPC architecture. Among these designs, both LiSens and FPA-CS have high light throughput as well as high frame rate. Comparing LiSens with FPA-CS, LiSens has two advantages: it can take benefit from line sensors over 2D sensors including large pixels, high dynamic range, high fill factor, etc.; it has no stringent requirement on calibration of the mapping from the SLM to the sensor when the SLM uses same multiplexing code for each pixel. The disadvantage is that LiSens uses one more optics component than FPA-CS which is a cylindrical lens.

Active imaging. Most of the traditional designs lie in the upper-left and bottom-right corners of the design matrix in Table 5.1, point illumination with a photo-detector or area illumination with a 2D sensor. The proposed designs, DualSL and TriLC, are in the middle ground which use line illumination and line sensors. Achar et al. [3] introduces two devices, Episcan3D [65] and EpiTOF [4], which lie in the same design space as ours. In Episcan3D and EpiTOF, the line laser and line sensor are aligned such that the illumination plane and sensing plane, while being rotated, are always in the same plane, and compared to traditional designs with area illumination and 2D sensor, the sensor here receives much less global light and ambient light. We put these devices together in Figure 5.1 for easy comparison. They are similar in needed components, and are different in configuration and performance and hence application scenarios. Both DualSL and Episcan3D can get 3D depth map. DualSL is inferior to Episcan3D in ambient light and

global light because of the 1D integration which would include useless light, but has much less requirement on alignment and has gradual degradation in depth map, not drastic as Episcan3D, when the placement changes after calibration due to temperature or vibration. The difference between TriLC and Episcan3D/EpiTOF is only in the placement that pivots of the rotating line sensor and line laser are parallel in TriLC and are in the same line in Episcan3D/EpiTOF. Comparing both, TriLC cannot obtain full depth map and only monitors a pre-defined surface in 3D, hence being a proximity sensor, but its computation is much more straightforward, just thresholding over raw data, and has better ability in seeing through scattering media due to the block of backscattered light while Episcan3D/EpiTOF can not. It is worth mentioning that the configuration used by LeddarTech LIDAR [49] is very similar to DualSL, and the difference is that one using triangulation and the other time-of-flight.

In other design spaces, several methods have also been proposed for 3D measurement under strong sunlight and in scattering media. For example, MC3D [57] uses a 2D DVS sensor for structured light with point scanning. Since the DVS sensor only outputs the position of pixels which have changes in intensity instead of outputting all pixel values, it is especially efficient here and makes point scanning based structured light very fast. In [30] a structured light system with concentrated light whose concentration level is adaptive to ambient light level is proposed. In such a way, the overall speed of the system is increased. For seeing through scattering media, Narasimhan et al. [61] use line stripping and find objects through analyzing the spatial intensity curve. Similarly, Satat et al. [80] analyze temporal transient profile by a SPAD sensor to estimate the presence or absence of obstacles. These methods have implications for the design we propose, which we will elaborate in the future work section.

5.2 Implementation challenges

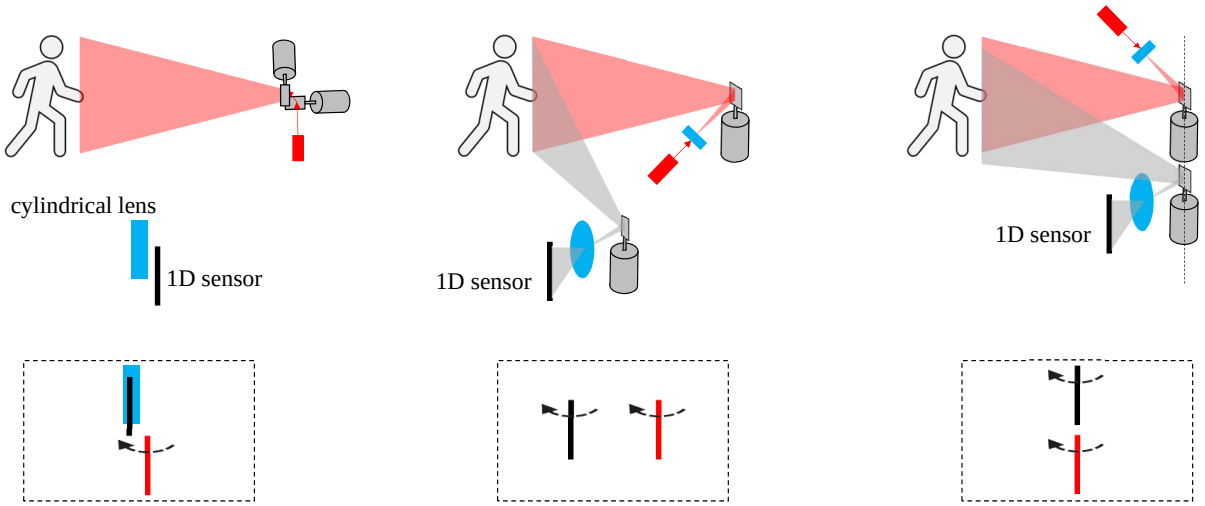
There are many aspects to consider when deploying our designs.

Frame rate of the device. The speed of the line camera depends on the slower one of the line sensor and readout circuits. Line sensors are generally fast, with tens of thousand to hundreds of thousand lines per second. However, the readout circuits that read and transmit the data to the host computer can be slow due to the small buffer size or the use of old protocols such as USB 2.0. The line cameras used in LiSens and DualSL hardware prototype can only provide 1K lines per second due to the slow readout circuits, and the line camera in TriLC prototype can reach 95K lines per second.

Our designs need scanning. In the prototype we use galvo mirrors whose scanning speed can only reach several thousand positions per second. In practice, we can use MEMS mirror [36]

Sensor \ Illum.	**Passive** None	Point	**Active** Line	Area
Point	Point scanning SPC[21]	Confocal imaging[55] Scanning LIDAR		
Line	Line scanning Push broom scanner[101] **LiSens**[96]		**Episcan3D**[65] **EpiTOF**[4] **DualSL**[97] LeddarTech LIDAR[49] **TriLC**[98]	
Low-res area	MSRA gigapixel camera[11] **FPA-CS**[17]			Flash LIDAR[80]
Area	FPA	MC3D[57]	Line stripping[61]	Adaptive[30] Structured light CW-TOF

Table 5.1: **Taxonomy of 2D imaging and 3D scanning devices** based on the dimension of the sensor used and how the light is distributed. The passive imagers which use low-resolution sensors for high-resolution imaging based on scanning, translation and multiplexing are listed in the leftmost column. For active 3D measuring under strong sunlight, people have proposed MC3D [57], Episcan3D [65], EpiTOF [4] and adaptive structured light [30] as alternatives to LIDAR. For 3D measuring in scattering media, Narasimhan et al. [61] and Satat et al. [80] achieve this by analyzing spatial and temporal distribution of photons, respectively.



(a) DualSL (b) TriLC (c) Episcan3D/EpiTOF

Figure 5.1: **Illustration of DualSL, TriLC and Episcan3D/EpiTOF.** They are similar in composed components but different in configuration.

to direct the light beam which is much faster, for example, the laser projector which uses a 2D MEMS mirror can have 60 fps [23]. In the imaging side, we can use a galvo mirror at resonant mode or a rotating polygonal mirror to direct the viewing angle, or use a rolling shutter camera, a camera with fast region of interest (ROI) setting or a recent camera with fast mask setting [79]. Whether they meet the requirements and how to do the design depend on the specific situations.

Calibration. We need to consider the following issues, how to do the calibration, how fast it can be (like in a factory), and how to deal with the out of calibration problem due to external factors like thermal heat or vibration.

LiSens can get good results without calibration when we use the same mask for each line camera pixel as shown in Chapter 2. We can get better results if calibrating the PSF of each DMD pixel on the line camera and doing deblurring, but with additional computational overhead. Hadamard multiplexing and very bright light sources can be used to obtain the PSFs with high SNR. Drift after calibration is not severe because the image content is preserved though tilted.

Both DualSL and TriLC require precise calibration, either to calculate depth or to design curtains. Both require a 2D helper camera and TriLC needs a 2D helper projector. They are based on basic camera calibration and can use robotic arm to speed up in the factory. If they are used in NIR light region, we can use a NIR 2D helper camera and build a 2D helper projector based on DMD. If used in SWIR region, the calibration is nontrivial without an additional camera.

To calibrate SWIR laser's position / SWIR line camera, we can swing a SWIR photo-detector / SWIR LED, respectively, and they are attached with black-white checkerboard or visible LEDs to know its 3D locations from a 2D helper camera. Another method is to add a visible laser and make it co-axial to SWIR laser. If out of calibration, DualSL still outputs correct disparity map but incorrect depth map and TriLC device can generate a curtain but with a different shape. The deviation depends on the drift level after calibration and on-the-fly calibration is needed.

Laser eye safety. Laser safety, especially eye safety, puts a limit on the maximum energy that a laser system can emit to the scene, hence restricting the maximum working range. The laser's power density in W/m^2 at a distance smaller than Maximal Permissible Exposure (MPE) is considered eye-safe. MPE is a function of laser wavelength, subtended angle of the laser with finite size to the eye and duration [6]. We must perform rigorous calculation processes to ensure that it is eye-safe even when people are very close to the device. Although we used visible light in the prototype, we should use NIR or SWIR in practice because they will not disturb or dazzle people's eyes and have higher eye-safety power.

Compactness. Compact devices are popular in the consumer market. LiSens needs space to accommodate DMD, TIR prism like the one used in [72] and other imaging lenses, and the compactest design may be down to one square inch. Embedding it into a mobile phone is still tricky unless the objective lens is facing the side, or part of the LCD screen is used as the SLM. Both DualSL and TriLC need baseline which limits the compactness. For individual component, both devices' light emission side can use MEMS mirror which is compact, DualSL's sensor side can use diffractive optics [68], and TriLC's sensor side can use a rolling shutter camera and is limited by the size of galvo mirror if used.

Dynamic range of the camera. All three devices need cameras with high dynamic range and fortunately line cameras usually have such property because their pixels can be rectangular and have a large size. When they are operated indoors, this requirement can be relaxed.

5.3 Future work

There are many ways to extend the ideas mentioned in this paper. In addition to using special functionality line sensors in our designs, such as SWIR, DVS, CW-TOF, SPAD and position sensitive device (PSD), we can also design different configurations as described below.

LiSens with multiple line sensors. The DMD can reflect light in two directions. We can build two LiSenses on both sides, one with row multiplexing and the other column multiplexing. In this way, there is no light loss and joint demultiplexing can be done. When blur kernel is wide, that is one DMD pixel is mapped to multiple line sensor pixels, deblurring result is bad due to much loss in high frequency. When we have two line sensors, we can obtain two images which are the underlying images blurred row-wisely and column-wisely. The joint deblurring based on two images should give a much better result as is expected by [41] which states that robust image deblurring can be achieved when two blur kernels are orthogonal to each other even though the individual blurs are large.

TriLC with multiple line lasers. Though most the time light curtain is set to be several meters away, it must be guaranteed to be eye-safe at very close range (usually half meter or one meter away) and there is a huge gap between the usage distance and safety distance. By distributing power to multiple line lasers which intersect or "focus" at the same line in 3D, it can be eye-safe easily at close range as well as focused distance. In other words, the device is more eye-safe now or higher total power can be used. Interference between multiple devices is also relieved by using this method.

Triangulation gating + temporal gating for seeing through scattering media. TriLC is an application of the concept of depth gating, wherein only light from a selective depth or a range of depths which satisfies triangulation geometry is allowed to enter the camera. Temporal gating [18] uses a combination of a pulsed laser and a camera with gated shutter, typically in the range of pico- to nano-seconds. By delaying the exposure of the camera with respect to the laser pulse, the device enables depth selection. Since the two gatings use completely different mechanisms, combining them by using a line temporal gated sensor in TriLC can isolate more scattered light and thus can see deeper. We can also use line SPAD sensor in our TriLC device. Since backscattered light (single-bounce scattered light) is blocked from the light path, the temporal jump caused by real objects in the transient histogram can be more prominent, so the method [80] can detect the temporal jump more easily or see deeper into the media. When we use a 2D SPAD sensor with line illumination, we can combine the methods proposed in [61] and [80]. Since [61] detects objects by finding spatial jump in spatial intensity profile, [80] detects objects by finding temporal jump in temporal transient profile, combining the two methods based on two cues is expected to have much better performance than every single method.

Outdoor real-time photometric stereo. Both DualSL and TriLC are depth sensing devices. However, for many objects, fine details like surface normal are also important. Photometric

stereo [102] is an active illumination technology that calculates the surface normal by shadings under different lightings. Photometric stereo usually uses several point light sources, but this is not feasible outdoors under strong sunlight because the power of the active light source is often limited. Previous methods use sunlight and skylight [42, 104] as the light source, but they cannot be operated in real time. The idea of the proposed designs in this dissertation to be able to work in bright sunlight is to concentrate the light into a line. Based on the same idea, we can use two line lasers and one rolling shutter camera where the two line lasers are aligned and synced with the being exposed row in the rolling shutter camera. Photometric stereo requires three non-collinear light sources, and here the third light source is the sun. The emission of two line lasers interleaves such that every first / second / third row is illuminated by the first laser and the sun / the second laser and the sun / sun only, respectively. A surface normal map can be computed from such a single image. We can also add a point light source to provide a third source for cloudy days. Line illumination by an LED may be needed if the line laser has strong speckles.

Incorporating machine learning. In the past decade, machine learning has made great strides, from face recognition [90] to general object recognition [33] to chess games [86]. Machine learning is a data-driven approach that uses a large amount of data to train a (complex) model and then use the model for prediction. In LiSens, we can use learning to reconstruct the image. In the training process, we can use a lot of SWIR images and use the acquired PSF to synthesize the data collected by LiSens to achieve end-to-end training. With learning, we expect to use a larger compression ratio [47] to reduce the amount of data collected, hence increase the frame rate, and obtain high-quality images. In DualSL, we can use the least amount of patterns through learning to increase the frame rate as the work already shown in [59]. In TriLC, we can assign the lines of the light curtain according to the view of a 2D traditional camera, and allocate more in the position where cars, robots or human may appear.

Bibliography

[1] Light curtain wikipedia. *https://en.wikipedia.org/wiki/Light_curtain*.

[2] Ruled surface wikipedia. *https://en.wikipedia.org/wiki/Ruled_surface*.

[3] S. Achar. Active illumination for the realworld. Doctoral dissertation. Carnegie Mellon University. 2017.

[4] S. Achar, J. R. Bartels, W. L. Whittaker, K. N. Kutulakos, and S. G. Narasimhan. Epipolar time-of-flight imaging. *ACM Transactions on Graphics (ToG)*, 36(4):37, 2017.

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

[6] American National Standards Institute. American national standard for safe use of lasers z136.1. 2014.

[7] I. M. Baker, S. S. Duncan, and J. W. Copley. A low-noise laser-gated imaging system for long-range target identification. In *Defense and Security*, pages 133–144. International Society for Optics and Photonics, 2004.

[8] R. G. Baraniuk. Compressive sensing [lecture notes]. *IEEE signal processing magazine*, 24(4):118–121, 2007.

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

[10] A. N. Belbachir, S. Schraml, M. Mayerhofer, and M. Hofstatter. A novel hdr depth camera for real-time 3d 360-degree panoramic vision. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 425–432, 2014.

[11] M. Ben-Ezra. High resolution large format tile-scan camera: Design, calibration, and extended depth of field. In *IEEE International Conference on Computational Photography (ICCP)*, pages 1–8. IEEE, 2010.

[12] C. F. Bohren and D. R. Huffman. *Absorption and scattering of light by small particles*. John Wiley & Sons, 2008.

[13] J.-Y. Bouguet. Matlab camera calibration toolbox. `http://www.vision.caltech.edu/bouguetj/calib_doc/`, 2000.

[14] C. Brandli, T. A. Mantel, M. Hutter, M. A. Höpflinger, R. Berner, R. Siegwart, and T. Delbruck. Adaptive pulsed laser line extraction for terrain reconstruction using a dynamic vision sensor. *Frontiers in neuroscience*, 7(EPFL-ARTICLE-200448), 2014.

[15] S. Burri, H. Homulle, C. Bruschini, and E. Charbon. Linospad: a time-resolved 256× 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.

[16] A. Chambolle. An algorithm for total variation minimization and applications. *Journal of Mathematical Imaging and Vision*, 20(1-2):89–97, 2004.

[17] H. Chen, M. Salman Asif, A. C. Sankaranarayanan, and A. Veeraraghavan. Fpa-cs: Focal plane array-based compressive imaging in short-wave infrared. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2358–2366. IEEE, 2015.

[18] F. Christnacher, S. Schertzer, N. Metzger, E. Bacher, M. Laurenzis, and R. Habermacher. Influence of gating and of the gate shape on the penetration capacity of range-gated active imaging in scattering environments. *Optics express*, 23(26):32897–32908, 2015.

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

[20] T. Delbruck. Frame-free dynamic digital vision. In *International Symposium on Secure-Life Electronics, Advanced Electronics for Quality Life and Society*, pages 21–26, 2008.

[21] M. F. Duarte, M. A. Davenport, D. Takhar, J. N. Laska, T. Sun, K. F. Kelly, and R. G. Baraniuk. Single-pixel imaging via compressive sampling. *IEEE signal processing magazine*, 25(2):83–91, 2008.

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

[23] M. Freeman, M. Champion, and S. Madhavan. Scanned laser pico-projectors: seeing the big picture (with a small device). *Optics and photonics news*, 20(5):28–34, 2009.

[24] L. Gao, J. Liang, C. Li, and L. V. Wang. Single-shot compressed ultrafast photography at one hundred billion frames per second. *Nature*, 516(7529):74–77, 2014.

[25] M. Gehm and D. Brady. Compressive sensing in the eo/ir. *Applied optics*, 54(8):C14–C22,

2015.

[26] A. Geiger, P. Lenz, and R. Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3354–3361. IEEE, 2012.

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

[28] B. I. Gramatikov. Modern technologies for retinal scanning and imaging: an introduction for the biomedical engineer. *Biomedical engineering online*, 13(1):52, 2014.

[29] Y. Grauer and E. 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.

[30] M. Gupta, Q. Yin, and S. K. Nayar. Structured light in sunlight. In *IEEE International Conference on Computer Vision (ICCV)*, pages 545–552, 2013.

[31] M. P. Hansen and D. S. Malchow. Overview of SWIR detectors, cameras, and applications. In *SPIE Defense and Security Symposium*, pages 69390I–69390I, 2008.

[32] M. Harwit and N. J. Sloane. *Hadamard transform optics*. New York: Academic Press, 1979.

[33] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *IEEE conference on computer vision and pattern recognition (CVPR)*, pages 770–778, 2016.

[34] Y. Hitomi, J. Gu, M. Gupta, T. Mitsunaga, and S. K. Nayar. Video from a single coded exposure photograph using a learned over-complete dictionary. In *IEEE International Conference on Computer Vision (ICCV)*, 2011.

[35] J. Holloway, A. C. Sankaranarayanan, A. Veeraraghavan, and S. Tambe. Flutter shutter video camera for compressive sensing of videos. In *IEEE International Conference on Computational Photography (ICCP)*, 2012.

[36] S. T. Holmstrom, U. Baran, and H. Urey. Mems laser scanners: a review. *Journal of Microelectromechanical Systems*, 23(2):259–275, 2014.

[37] H. Horvath. Atmospheric light absorptiona review. *Atmospheric Environment. Part A. General Topics*, 27(3):293–317, 1993.

[38] G. Huang, H. Jiang, K. Matthews, and P. Wilford. Lenless imaging by compressive sensing. In *IEEE International Conference on Image Processing (ICIP)*, 2013.

[39] iniLabs. Dvs128 specifications. `http://inilabs.com/`

products/dynamic-and-active-pixel-vision-sensor/davis-specifications/, 2015.

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

[41] A. Ito, A. C. Sankaranarayanan, A. Veeraraghavan, and R. G. Baraniuk. Blurburst: Removing blur due to camera shake using multiple images. *ACM Transactions on Graphics (TOG)*, 3(1), 2014.

[42] J. Jung, J.-Y. Lee, and I. So Kweon. One-day outdoor photometric stereo via skylight estimation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4521–4529. IEEE, 2015.

[43] J. Ke and E. Y. Lam. Object reconstruction in block-based compressive imaging. *Optics express*, 20(20):22102–22117, 2012.

[44] K. Kelly, R. Baraniuk, L. McMackin, R. Bridge, S. Chatterjee, and T. Weston. Decreasing image acquisition time for compressive imaging devices, Oct. 14 2014. URL https://www.google.com/patents/US8860835. US Patent 8,860,835.

[45] R. Kerviche, N. Zhu, and A. Ashok. Information-optimal scalable compressive imaging system. In *Computational Optical Sensing and Imaging (COSI)*, 2014.

[46] L. Keselman, J. I. Woodfill, A. Grunnet-Jepsen, and A. Bhowmik. Intel (r) realsense (tm) stereoscopic depth cameras. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1267–1276. IEEE, 2017.

[47] K. Kulkarni, S. Lohit, P. Turaga, R. Kerviche, and A. Ashok. Reconnet: Non-iterative reconstruction of images from compressively sensed measurements. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 449–458, 2016.

[48] D. Lanman and G. Taubin. Build your own 3D scanner: 3D photograhy for beginners. In *ACM SIGGRAPH 2009 courses*, pages 30–34, 2009.

[49] LeddarTech. 2d and 3d flash lidar fundamentals. *https://www.youtube.com/watch?v=Qbo9mGAFkqc/*, 2017.

[50] L. Li. Time-of-flight camera–an introduction. *Technical white paper*, (SLOA190B), 2014.

[51] P. Lichtsteiner, C. Posch, and T. Delbruck. A 128×128 120 db 15μ s latency asynchronous temporal contrast vision sensor. *IEEE journal of solid-state circuits*, 43(2):566–576, 2008.

[52] X. Lin, G. Wetzstein, Y. Liu, and Q. Dai. Dual-coded compressive hyperspectral imaging. *Optics letters*, 39(7):2044–2047, 2014.

[53] P. Llull, X. Liao, X. Yuan, J. Yang, D. Kittle, L. Carin, G. Sapiro, and D. Brady. Compressive sensing for video using a passive coding element. In *Computational Optical Sensing and Imaging (COSI)*, 2013.

[54] A. Mahalanobis, R. Shilling, R. Murphy, and R. Muise. Recent results of medium wave infrared compressive sensing. *Applied Optics*, 53(34):8060–8070, 2014.

[55] M. Marvin. Microscopy apparatus, Dec. 19 1961. US Patent 3,013,467.

[56] K. Marwah, G. Wetzstein, Y. Bando, and R. Raskar. Compressive light field photography using overcomplete dictionaries and optimized projections. *ACM Transactions Graphics (TOG)*, 32(4):46, 2013.

[57] N. Matsuda, O. Cossairt, and M. Gupta. Mc3d: Motion contrast 3d scanning. In *IEEE International Conference on Computational Photography (ICCP)*, pages 1–10. IEEE, 2015.

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

[59] P. Mirdehghan, W. Chen, and K. N. Kutulakos. Optimal structured light à la carte. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6248–6257, 2018.

[60] W. E. Moerner. Single-molecule spectroscopy, imaging, and photocontrol: Foundations for super-resolution microscopy (nobel lecture). *Angewandte Chemie International Edition*, 54(28):8067–8093, 2015.

[61] S. G. Narasimhan, S. K. Nayar, B. Sun, and S. J. Koppal. Structured light in scattering media. In *IEEE International Conference on Computer Vision (ICCV)*, volume 1, pages 420–427. IEEE, 2005.

[62] S. K. Nayar, V. Branzoi, and T. E. Boult. Programmable imaging using a digital micromirror array. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2004.

[63] Newport Corporation. Introduction to solar radiatio. *`https://www.newport.com/t/introduction-to-solar-radiation/`*, 2013.

[64] M. O'Toole, R. Raskar, and K. N. Kutulakos. Primal-dual coding to probe light transport. *ACM Transactions on Graphics (ToG)*, 31(4):39–1, 2012.

[65] M. O'Toole, S. Achar, S. G. Narasimhan, and K. N. Kutulakos. Homogeneous codes for energy-efficient illumination and imaging. *ACM Transactions on Graphics (ToG)*, 34(4): 35, 2015.

[66] J. Y. Park and M. B. Wakin. A multiscale framework for compressive sensing of video. In *Picture Coding Symposium*, 2009.

[67] J. Y. Park and M. B. Wakin. Multiscale algorithm for reconstructing videos from streaming compressive measurements. *Journal of Electronic Imaging*, 22(2), 2013.

[68] Y. Peng, Q. Fu, F. Heide, and W. Heidrich. The diffractive achromat full spectrum computational imaging with diffractive optics. In *ACM SIGGRAPH ASIA 2016 Virtual Reality meets Physical Reality: Modelling and Simulating Virtual Humans and Environments*, page 4. ACM, 2016.

[69] C. Posch, M. Hofstätter, D. Matolin, G. Vanstraelen, P. Schön, N. Donath, and M. Litzenberger. A dual-line optical transient sensor with on-chip precision time-stamp generation. In *International Solid-State Circuits Conference*, pages 500–618, 2007.

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

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

[72] Y. Qiao, X. Xu, T. Liu, and Y. Pan. Design of a high-numerical-aperture digital micromirror device camera with high dynamic range. *Applied optics*, 54(1):60–70, 2015.

[73] D. Reddy, A. Veeraraghavan, and R. Chellappa. P2C2: Programmable pixel compressive camera for high speed imaging. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2011.

[74] A. Rogalski. Infrared detectors: an overview. *Infrared Physics & Technology*, 43(3-5): 187–210, 2002.

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

[76] A. C. Sankaranarayanan and A. Veeraraghavan. Parallel compressive imaging. In *Computational Optical Sensing and Imaging (COSI)*, pages CTh3E–4, 2015.

[77] A. C. Sankaranarayanan, C. Studer, and R. G. Baraniuk. CS-MUVI: Video compressive sensing for spatial-multiplexing cameras. In *IEEE International Conference on Computational Photography (ICCP)*, 2012.

[78] A. C. Sankaranarayanan, J. Wang, and M. Gupta. Radon transform imaging: low-cost video compressive imaging at extreme resolutions. In *Sensing and Analysis Technologies for Biomedical and Cognitive Applications 2016*, volume 9871, page 98710Q. International Society for Optics and Photonics, 2016.

[79] N. Sarhangnejad, H. Lee, N. Katic, M. OToole, K. Kutulakos, and R. Genov. Cmos image sensor architecture for primal-dual coding. In *International Image Sensor Workshop*,

2017.

[80] G. Satat, M. Tancik, and R. Raskar. Towards photography through realistic fog. In *IEEE International Conference on Computational Photography (ICCP)*, pages 1–10. IEEE, 2018.

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

[82] P. Sen and S. Darabi. Compressive dual photography. *Computer Graphics Forum*, 28(2): 609–618, 2009.

[83] P. Sen, B. Chen, G. Garg, S. R. Marschner, M. Horowitz, M. Levoy, and H. Lensch. Dual photography. *ACM Transactions Graphics (TOG)*, 24(3):745–755, 2005.

[84] Sensors Unlimited Inc. SWIR image gallery. `http://www.sensorsinc.com/gallery/images`, 2016.

[85] Y. Shirai and M. Suwa. Recognition of polyhedrons with a range finder. In *International Joint Conferences on Artificial Intelligence (IJCAI)*, pages 80–87, 1971.

[86] D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, T. Hubert, L. Baker, M. Lai, A. Bolton, et al. Mastering the game of go without human knowledge. *Nature*, 550(7676):354, 2017.

[87] D. H. Sliney and J. Mellerio. *Safety with lasers and other optical sources: a comprehensive handbook*. Springer Science & Business Media, 2013.

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

[89] R. Tadano, A. Kumar Pediredla, and A. Veeraraghavan. Depth selective camera: A direct, on-chip, programmable technique for depth selectivity in photography. In *International Conference on Computer Vision (ICCV)*, pages 3595–3603. IEEE, 2015.

[90] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf. Deepface: Closing the gap to human-level performance in face verification. In *IEEE conference on computer vision and pattern recognition (CVPR)*, pages 1701–1708, 2014.

[91] S. Tambe, A. Veeraraghavan, and A. Agrawal. Towards motion aware light field video for dynamic scenes. In *IEEE International Conference on Computer Vision (ICCV)*, 2013.

[92] Thorlabs. Galvanometers. *https://www.thorlabs.com/navigation.cfm?guide_id=2269&gclid=CjwKCAjw0ujYBRBDEiwAn7BKtzKIqA-XZsCraXLYD__WZ_zKKSCAJKbdfyl247zudmwDhobbI6iuVRoC2RkQAvD_BwE/*, 2018.

[93] P. Vuylsteke and A. Oosterlinck. Range image acquisition with a single binary-encoded light pattern. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(2): 148–164, 1990.

[94] A. Wagadarikar, R. John, R. Willett, and D. Brady. Single disperser design for coded aperture snapshot spectral imaging. *Applied Optics*, 47(10):44–51, 2008.

[95] M. B. Wakin, J. N. Laska, M. F. Duarte, D. Baron, S. Sarvotham, D. Takhar, K. F. Kelly, and R. G. Baraniuk. Compressive imaging for video representation and coding. In *Picture Coding Symposium*, 2006.

[96] J. Wang, M. Gupta, and A. C. Sankaranarayanan. Lisens-a scalable architecture for video compressive sensing. In *IEEE International Conference on Computational Photography (ICCP)*, pages 1–9. IEEE, 2015.

[97] J. Wang, A. C. Sankaranarayanan, M. Gupta, and S. G. Narasimhan. Dual structured light 3d using a 1d sensor. In *European Conference on Computer Vision (ECCV)*, pages 383–398. Springer, 2016.

[98] J. Wang, J. Bartels, W. Whittaker, A. C. Sankaranarayanan, and S. G. Narasimhan. Programmable triangulation light curtains. In *European Conference on Computer Vision (ECCV)*, pages 1–9. Springer, 2018.

[99] M. Weber, M. Mickoleit, and J. Huisken. Light sheet microscopy. In *Methods in cell biology*, volume 123, pages 193–215. Elsevier, 2014.

[100] S. Wheaton, A. Bonakdar, I. H. Nia, C. L. Tan, V. Fathipour, and H. Mohseni. Open architecture time of fight 3d SWIR camera operating at 150 MHz modulation frequency. *Optics Express*, 25(16):19291–19297, 2017.

[101] Wikipedia. Push broom scanner. `https://en.wikipedia.org/wiki/Push_broom_scanner`, 2018.

[102] R. J. Woodham. Photometric method for determining surface orientation from multiple images. *Optical engineering*, 19(1):191139, 1980.

[103] J. Yang, X. Yuan, X. Liao, P. Llull, D. J. Brady, G. Sapiro, and L. Carin. Video compressive sensing using gaussian mixture models. *IEEE Transactions on Image Processing*, 23(11): 4863–4878, 2014.

[104] L.-F. Yu, S.-K. Yeung, Y.-W. Tai, D. Terzopoulos, and T. F. Chan. Outdoor photometric stereo. In *IEEE International Conference on Computational Photography (ICCP)*, pages 1–8. IEEE, 2013.

[105] Z. Zhang. Microsoft kinect sensor and its effect. *IEEE multimedia*, 19(2):4–10, 2012.

[106] Y. Zhou and O. Tuzel. Voxelnet: End-to-end learning for point cloud based 3d object detection. *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.