THE UNIVERSITY
of EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e. g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

Guided Direct Time-of-Flight Lidar for Self-Driving Vehicles

FILIP TANESKI

Doctor of Philosophy
UNIVERSITY OF EDINBURGH
2023

CONTENTS

ABSTRACT

Self-driving vehicles demand efficient and reliable depth-sensing technologies. Lidar, with its capacity for long-distance, high-precision measurement, is a crucial component in this pursuit. However, conventional mechanical scanning implementations suffer from reliability, cost, and frame rate limitations. Solid-state lidar solutions have emerged as a promising alternative, but the vast amount of photon data processed and stored using conventional direct time-of-flight (dToF) prevents long-distance sensing unless power-intensive partial histogram approaches are used.

This research introduces a pioneering 'guided' dToF approach, harnessing external guidance from other onboard sensors to narrow down the depth search space for a power and data-efficient solution. This approach centres around a dToF sensor in which the exposed time widow of independent pixels can be dynamically adjusted. A pair of vision cameras are used in this demonstrator to provide the guiding depth estimates.

The implemented guided dToF demonstrator successfully captures a dynamic outdoor scene at 3 fps with distances up to 75 m. Compared to a conventional full histogram approach, on-chip data is reduced by over 25 times, while the total laser cycles in each frame are reduced by at least 6 times compared to any partial histogram approach. The capability of guided dToF to mitigate multipath reflections is also demonstrated.

For self-driving vehicles where a wealth of sensor data is already available, guided dToF opens new possibilities for efficient solid-state lidar.

ACKNOWLEDGEMENTS

This work would not have been possible without the help of many kind people. In particular, I would like to thank the following:

DECLARATION OF ORIGINALITY

I hereby declare that the research recorded in this thesis (excluding the exceptions stated below) and the thesis itself originated with and was composed by myself.

1. The QuantIC4×4 sensor was developed by the CMOS Sensors and System group at the Univeristy of Edinburgh headed by Prof. Robert Henderson. Guided dToF functionality was designed and incorporated into the revised sensor by myself.
2. The HSLIDAR sensor was developed by Dr Istvan Gyongy and Dr Ahmet Erdogan, both part of the CMOS Sensors and System group at the University of Edinburgh.

Other people who have contributed to this work are acknowledged and/or referenced appropriately within.

This work has not been submitted for any other degree or professional qualification except as specified.

Filip Taneski

LIST OF PUBLICATIONS

[1] F. Taneski, T. A. Abbas, and R. K. Henderson, "Laser Power Efficiency of Partial Histogram Direct Time-of-Flight LiDAR Sensors," *Journal of Lightwave Technology,* vol. 40, no. 17, pp. 5884-5893, 2022, doi: https://doi.org/10.1109/JLT.2022.3187293.

[2] F. Taneski, I. Gyongy, T. A. Abbas, and R. Henderson, "Guided Flash Lidar: A Laser Power Efficient Approach for Long-Range Lidar," presented at the International Image Sensor Workshop, Crieff, Scotland, 2023. [Online]. Available: https://web.archive.org/web/20230913082706/https://imagesensors.org/2023-papers/.

[3] F. Taneski, I. Gyongy, T. Al Abbas, and R. K. Henderson, "Guided Direct Time-of-Flight Lidar Using Stereo Cameras for Enhanced Laser Power Efficiency," *Sensors,* vol. 23, no. 21, p. 8943, 2023. [Online]. Available: https://doi.org/10.3390/s23218943.

[4] F. Taneski, T. A. Abbas, and R. Henderson, "Extending Lidar Depth Range using Stereo Depth Estimation on Intensity Data," presented at the Electronic Imaging, 2024. [Online]. Available: https://doi.org/10.2352/EI.2024.36.18.3DIA-102.

LIST OF TABLES

LIST OF FIGURES

Abbreviations

ALHAT Autonomous landing and hazard avoidance technology

ALV Autonomous land vehicle

AMCW Amplitude-modulated continuous wave

BSI Backside illuminated

CIS CMOS image sensor

CMOS Complementary metal-oxide-semiconductor

CNN Convolutional neural network

CRB Cramér Rao bound

DARPA Defense Advanced Research Projects Agency

DFF D-Type flip-flop

DToF Direct time-of-flight

DL Delay line

EEL Edge-emitting laser

ERIM Environmental Research Institute of Michigan

FE Front end

FMCW Frequency modulated continuous eave

FSI Frontside illuminated

FOV Field of view

GPU Graphics processing unit

GRO Gated ring oscillator

HRR Histogram reduction ratio

IToF Indirect time-of-flight

MEMS Micro-electromechanical systems

METDC Multievent time-to-digital converter

MPE Maximum permissible exposure

NASA National Aeronautical and Space Administration

OPA Optical phased arrays

PDE Photon detection efficiency

PVT Process, voltage and temperature

ROI Region-of-interest

SAE Society of Automotive Engineers

SPAD Single photon avalanche diode

SNR Signal-to-noise ratio

SRAM Static random-access memory

TAC Time-to-analogue converter

TDC Time-to-digital converter

ToF Time-of-flight

UAV Unmanned aerial vehicle

VCSEL Vertical cavity edge emitting laser

1 INTRODUCTION

1.1 Self-Driving Vehicles

1.1.1 Motivation

In 1939, General Motors captured the public's imagination with its highly acclaimed Futurama exhibit at the New York World's Fair, presenting a glimpse into the future of automotive transport [5]. The exhibit featured revolutionary concepts of autonomous vehicles driving continuously along automated highways and intersections [6]. However, beyond improving transport efficiency within a rapidly evolving society, the development of self-driving cars carries a noble mission: to save human lives. Every year, over a million lives are lost due to road accidents [7], of which over 90% are due to driver error [8].

While the widespread self-driving concept of 1939 is not yet realised, technological advancements and more permissive road regulations make such visions increasingly viable [9, 10]. Developers of self-driving vehicles continue to reach significant milestones including driving over a million miles on public roads with no human behind the wheel [11]. With a market pull for safer ways to commute and a market push from governments in the form of regulations [12], a driverless car future is imminent.

1.1.2 Levels of Autonomy

It could be argued that autonomous vehicles are already on the road today. The ability of vehicles to maintain a separation distance, or to drive within designated lane boundaries, is already commonplace.

Automation Level	0	1	2	3	4	5
Automation Features	Driver Support Features			Automated Driving Features		
Monitors Driving	Human		Shared		Vehicle	
Driver Required	Yes	Yes	Yes	Yes	No – most conditions	No – all conditions
Engagement	All driving	Hands OR feet off	Hands AND feet off	Eyes off	All off	

Fig. 1: Levels of driving autonomy as defined by J3016. Adapted from [13].

13

To adequately describe the full spectrum of autonomous driving abilities, the Society of Automotive Engineers (SAE) has set out a taxonomy for automated driving solutions. In keeping with the engineering stereotype, it was given the eloquent name of J3016 [14]. This landmark document in the field of self-driving vehicles defines 6 levels of driving autonomy which are summarised in Fig. 1.

The prestigious title of "fully autonomous" can only be given to those vehicles which achieve level 4 automation or above. At the time of writing, no commercial vehicle is yet to achieve this. A key enabler is the need for reliable and affordable long-range depth-sensing.

1.1.3 Depth-Sensing for Fully Autonomous Vehicles

Table 1 combines the guidelines outlined in J3016, along with the requirements of vehicle manufacturers [15-20] to summarise the performance of depth perception needed for commercial, fully autonomous vehicles. These specifications set the benchmark automotive depth-sensing performance targeted by research and industry today.

Specification	Description	Target
Maximum range	Furthest measurable distance	> 200 m
Field of view	Observable horizontal (H) and vertical (V) window	H120° x V25°
Angular resolution	Horizontal/vertical step size	< 0.1°
Frame rate	Image update rate	> 25 Hz
Distance error	Measured depth compared to actual depth	< 0.1 m
Target reflectivity	Minimum reflectivity of target that can be measured	> 10%
Ambient intensity	Maximum brightness of environment	100 klux
Reliability	Qualified robustness of mechanics after stress test	AEC-Q100
Cost	Price per unit	< $1,000
Power	Average power consumed	< 10 W

Table 1: Depth-sensing performance required for commercial, fully autonomous vehicles.

1.2 Depth-Sensing Technologies

1.2.1 Ultrasound

Ultrasound sensors measure the return time of emitted high-frequency sound waves to determine distance. These provide a cheap and compact method of measuring distance. However, the large angular beamwidth of sound waves as they travel through air limits ultrasound to low-range, low-resolution tasks such as parking assist [21].

1.2.2 Radar

Using a similar principle to ultrasound, radar uses radio waves which can propagate much further than sound waves. This allows it to achieve ranging over much longer distances. First developed over 100 years ago, radar is a proven and reliable method of depth-sensing [22], robust under poor weather conditions such as fog [23]. Automotive radar typically uses frequencies between 10-100 GHz to provide some angular resolution, however this is still limited to tens of centimetres [24].

1.2.3 Camera

Unlike the other sensors described in this section, cameras sense the environment without the use of an active emitter. This enables a low-power (few watts) method for sensing, detecting and classifying objects in the field of view.

A single camera (monocular) can estimate depth by classifying objects and inferring their relative distance, or recently using machine learning techniques. More accurate depth-sensing can be achieved through the use of two (stereo) or more (multi-baseline) imaging cameras by identifying the relative displacement of points between each camera image. This displacement is termed disparity and is typically given in units of pixels. The disparity value determined for each pixel in the image from one camera can then be directly converted into distance.

Objects close to the cameras experience a smaller displacement than those further away, as illustrated in Fig. 2. Therefore, the measurement accuracy of stereo depth reduces with distance, as the change in disparity becomes increasingly less pronounced. In addition, the challenge of matching points between

images on each camera, known as the *correspondence problem* is not trivial, particularly for large featureless objects [25]. Unlike radar and ultrasound, stereo depth is also susceptible to visibility conditions which further limit accuracy [26].
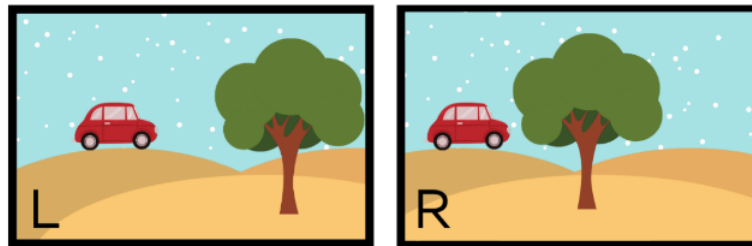
Fig. 2: Illustration of stereo disparity between a left (L) and right (R) camera. The tree in the foreground shifts more than the car in the background.

Nevertheless, depth ranging further than 200 m using vehicle-mounted stereo cameras has been readily demonstrated [27, 28]. Stereo cameras also inherently provide colour and brightness information, both valuable for self-driving perception. Finally, the ability to determine depth without the need for an active, power-consuming emitter has led to a continued high level of interest in this technology as the industry transitions to fully electric vehicles.

1.2.4 Lidar

Lidar sensors measure distance by determining the roundtrip time of an emitted laser. By mounting one or more point lidar sensors onto a rotating platform, as illustrated in Fig. 3, mechanical scanning lidar can provide high-resolution 3D depth-sensing over hundreds of metres with centimetre precision [29]. The short wavelength of light allows for a narrow beam divergence compared to sound and radio waves, providing superior angular resolution over ultrasound and radar. Lidar also enables standalone object detection [30], although this is made more robust through fusion with camera data [31].
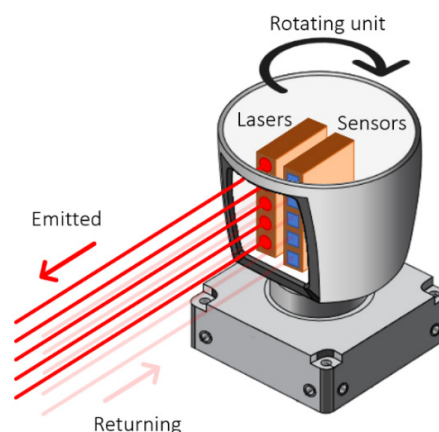
Fig. 3: Traditional mechanical spinning lidar.

However, traditional mechanical scanning lidar using motors and/or rotating mirrors results in an expensive solution with a large form factor [32], sensitivity to shock/vibrations [33], high power [33, 34] and requiring frequent recalibration [16]. The scanning mechanism also limits the achievable frame rate which is an important characteristic for avoiding collisions. For example, three frames are typically used to detect and confirm the trajectory of an object [35]. Therefore, a typical highway vehicle, traveling at around 30 m/s, would have moved by almost 100 m before a system sensing at 1 fps can make a decision.

Within the context of self-driving vehicles, the DARPA Grand Challenge has consistently shown lidar to be essential for safe and reliable autonomous vehicles. Lidar-equipped vehicles have consistently dominated this competition [36]. However, scaling this up to large scale deployment of self-driving vehicles brings inevitable challenges from mutual interference between lidars. While some initial research into solving interference issues have been made [37, 38], the full scale of the issue is yet to be realised.

1.2.5 Multisensory Approach

The performance and commercial viability of each prior-described sensor is summarised in Table 2. It shows that no single sensor achieves the level of performance required for safe autonomous vehicles. The resulting consensus from both market analysts and the industry is to combine these sensors in a multi-sensory approach [39]. This is reflected in the automated driving safety frameworks of vehicle manufacturers including Audi, BMW and Volkswagen [40]. Crucial self-driving tasks such as simultaneous location and mapping (SLAM) rely on further sensor data from a combination of satellite-based positioning and inertial measurement units (IMUs) containing accelerometers and gyroscopes [41]. Combining this diverse range of sensors allows the vehicle to harness the advantages of each while providing redundancy. The development of other technologies including unmanned aerial vehicles (UAVs) have similarly come to rely on fusing multiple sensors with lidar for more robust navigation [42].

Despite the commercial viability limitations of mechanical lidar, it remains a critical component for safe and reliable self-driving vehicles. This has resulted in a significant growth projection for the automotive lidar market. According to forecasts by Yole Group (Fig. 4), the market size is expected to rise from $26 million in 2020 to $2.3 billion in 2026, establishing automotive as the leading application for lidar technology. However, for self-driving vehicles to be widely adopted, the barriers to commercial viability associated with mechanical lidar must be addressed. A new generation of lidar, *solid-state* lidar, aims to provide this solution.

		Ultrasound	**Radar**	**Camera**	**Mechanical Lidar**
Performance	Range	< 10 m	> 200 m	> 200 m	> 200 m
	Angular resolution	> 10 cm	> 10 cm	< 1 cm	< 1 cm
	Accuracy (short-range)	< 1 cm	< 1 cm	< 1 cm	< 1 cm
	Accuracy (long-range)	N/A	< 10 cm	> 1 m	< 10 cm
	Object classification	No	No	Yes	Medium
	Robustness in low ambient light	High	High	Low	High
	Robustness in high ambient light	High	High	High	Medium
	Robustness in rain/snow/fog/dust	Medium	High	Low	Medium
Commercial Viability	Mechanical robustness	Small	Small	Small	Large
	Cost	< $100	< $100	< $100	> $10k
	Power consumption	~ 1 W	< 10 W	< 10 W	~ 100 W
	Form factor	Small	Medium	Small	Large

Table 2: Capabilities of the various sensor technologies integrated into self-driving vehicles, combining data from [23, 43, 44].
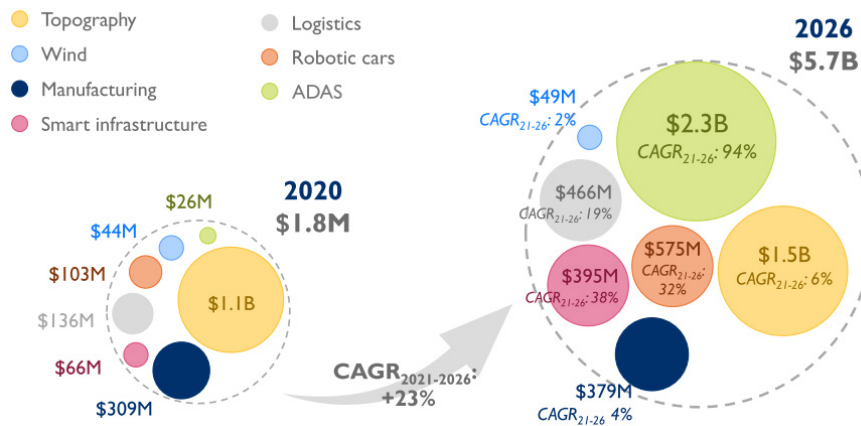
Fig. 4: Lidar market forecast as given by Yole Group [45].

1.3 Solid-State Lidar

Solid-state lidar aims to reduce expensive and bulky mechanical lidar down to a single (or few) chip(s). Currently, the two approaches closest to entering the automotive market are "flash" and "MEMS" (micro-electromechanical systems). The MEMS approach involves fabricating small micromirrors onto a chip to scan the laser around the scene, while flash removes scanning altogether by illuminating the entire field of view in a single flash. While the MEMS approach addresses the cost and form factor considerations for commercially viable lidar, it still relies on moving parts which deteriorate over time in an automotive environment. On the other hand, flash lidar is free from moving parts but requires a high laser power to illuminate the entire scene at once, leading to high power consumption and limited range. While technologies such as optical phased arrays (OPAs) may one day offer a complete solution [46], flash and MEMS are forecast to play a crucial role in the immediate future of automotive lidar, as shown in Fig. 5.
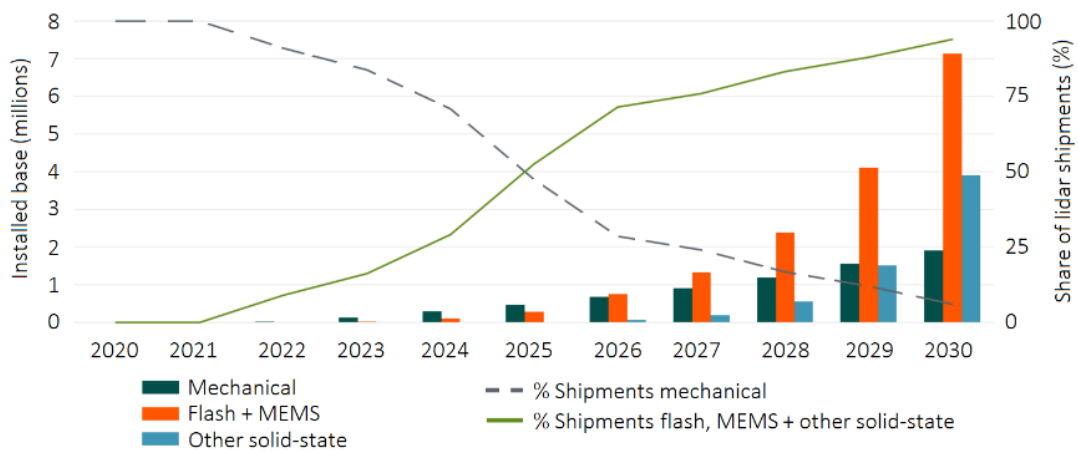
Fig. 5: Forecast worldwide lidar sensors within the automotive market as of 2021 [47].

1.4 Problem Statement

The most established lidar technique for measuring long distances is direct time-of-flight (dToF). This works on the principle of emitting a short laser pulse and directly measuring the roundtrip time of the returning laser to calculate distance (Fig. 6(a)). In practice, ambient background light is also detected by the lidar sensor, so the laser is cycled many times to average out the background (Fig. 6(b)).

Measure laser return time (Δt) to determine distance (d).

$$d = \frac{\Delta t \times c}{2}$$

1.33 µs ≡ 200 m

(a)

Laser
Background

Accumulate photon counts over multiple laser cycles

(b)

Photon count data compressed into a histogram

Interpolated peak

(c)

On-board sensor data

Every sensor pixel is guided to a coarse time window of interest

Background photons ignored

(d)
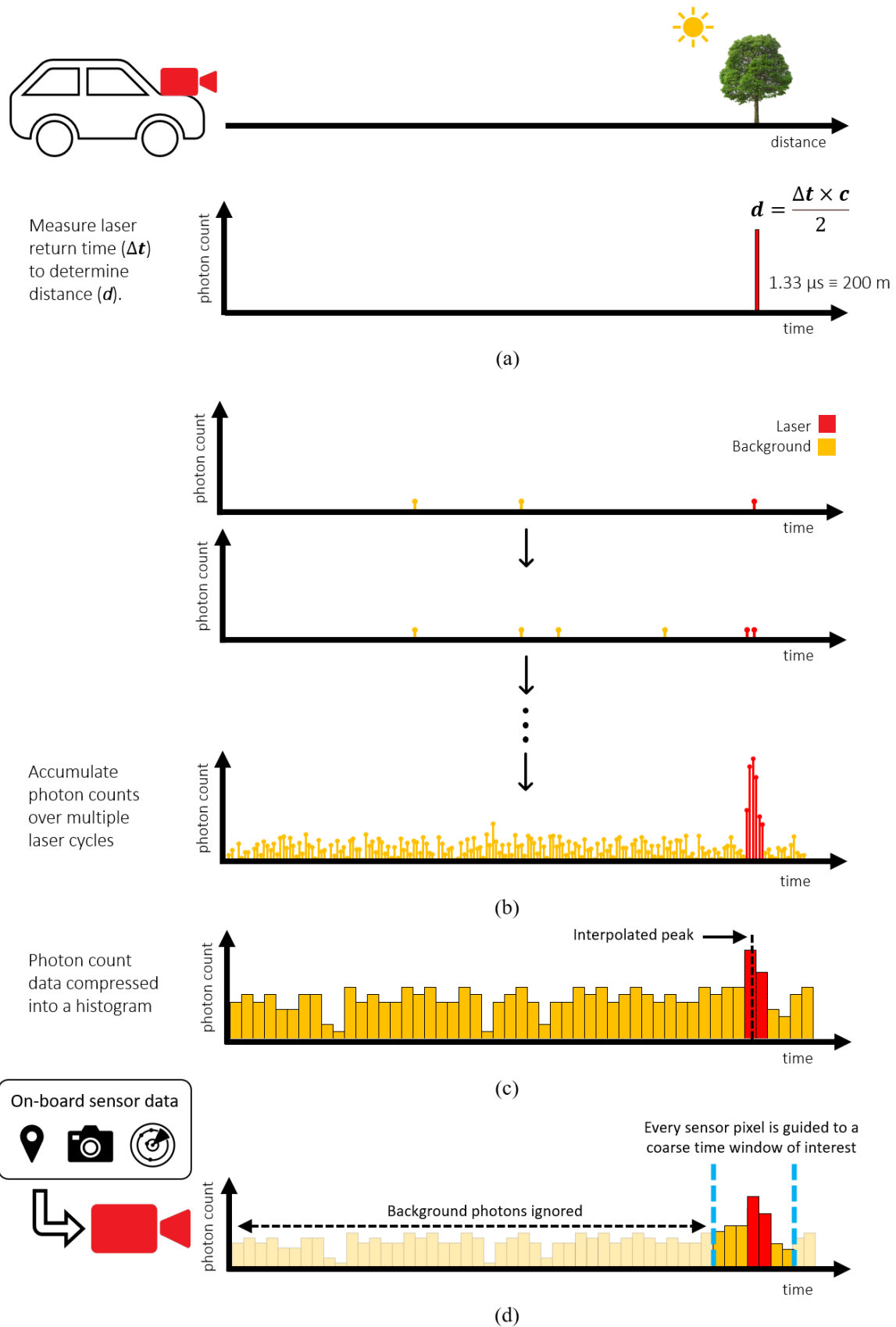
Fig. 6: Direct time-of-flight (a) ideal case with no ambient background photons (b) integrating multiple laser cycles to average out the background (c) compressing photon data into a histogram (d) proposed "guided" approach. Adapted from [3].

In most solid-state lidar implementations, particularly flash lidar, the sensor consists of a chip with a 2D array of detection pixels. For dToF, each pixel typically consists of one or more single photon avalanche diode (SPAD) for detecting photons, alongside processing to digitise and time the detected photons. As a result, every pixel needs to accommodate area to detect, process, time and store the captured photon data. However, for long-range measurements, the longer exposure time results in a greater number of detected photons. Therefore, storing fine precision photon arrival data over multiple laser cycles is no longer practical. An established method to address this is to store the data into coarse time intervals to form a histogram [48, 49], as illustrated in Fig. 6(c). 3D stacking of CMOS chips into separate sensing and processing tiers [50, 51] further increases the amount of photon data that can be stored in each pixel. Nevertheless, on-chip capacity inevitably limits the photon data required to meet the maximum range and precision for self-driving vehicles.

Many novel "partial histogram" solutions have recently been developed [52-54] to increase the maximum sensing range and/or precision in the face of limited on-chip capacity. However, these all require many additional laser cycles to complete a frame which increases sensor activity, limits the achievable frame rate and, above all, increases laser power consumption. This is particularly detrimental for flash lidar where the laser dominates system power consumption [54, 55]. For practical automotive lidar, a solution which makes optimal use of the emitted laser power is required.

1.5 Objective and Hypothesis

This work explores a new direct time-of-flight approach to enable long-range, high-precision lidar while making optimal use of the emitted laser power and collected photon data: *guided* direct time-of-flight. Self-driving vehicles already collect an excessive wealth of data [56] from sensors such as cameras, lidar, ultrasound and satellite navigation to name a few. Instead of treating the lidar as a closed system, inefficiently consuming laser cycles and collecting excessive data, this work proposes opening up the lidar to use the onboard data already available.

The proposed guided time-of-flight sensor allows each of its pixels to be externally and independently set to only count photons within a reduced time window, as illustrated in Fig. 6(d). This would enable the sensor to efficiently collect the returning laser photons; making optimal use of the laser power while minimising the data imposed by the sensor. Furthermore, by relying on fewer laser cycles per frame, laser eye-safety and interference challenges as the number of self-driving vehicles on the road increases

[57] are simplified. This thesis presents a pioneering exploration into the newly proposed concept of a guided direct time-of-flight sensor and its feasibility as a depth-sensing solution for self-driving vehicles.

1.6 Thesis Outline

With the foundational overview in place, Chapter 2 offers a comprehensive literature review. It serves to provide the necessary background information and evaluate the scope of related work within the field of solid-state direct time-of-flight lidar.

Chapter 3 reinforces the value of the proposed guided dToF system through the application of quantitative models. The limitations of both traditional full histogram and novel partial histogram approaches are explored.

Chapter 4 presents the design and evaluation of the dToF sensor equipped with integrated guided functionality, developed for this project. The contingency sensor option upon which this project ultimately relied is also outlined.

Chapter 5 explores different techniques for estimating depth using cameras, given their abundance onboard self-driving vehicles. The merits of various depth estimation techniques are evaluated to establish the most practical option for guiding the system for this project.

Chapter 6 brings together the prior chapters, combining the guided sensor and the depth estimation source to showcase the first operational guided dToF system. The system is evaluated in a variety of challenging conditions to determine the feasibility of guided dToF for self-driving vehicles.

Finally, Chapter 7 summarises and draws conclusions from the work presented in this thesis. The chapter finished by presenting areas of future work to further the development of guided dToF.

2 SOLID-STATE DTOF

2.1 A Brief History of Lidar

Following the development of the very first lasers in the late 1950s, the first lidar was invented in 1960 by Hughes Aircraft Company [58]. Initial designs functioned as a single-point rangefinder and were used in military applications and space exploration. In 1962, MIT Lincoln Lab used lidar to measure the distance to the moon [59] and in 1971 the first lidar-assisted moon landing was completed by Apollo 15 [60]. Meanwhile, applications in terrain (topography) and shallow water (bathygraphy) mapping were being explored as far back as 1965 [61] by mounting lidar onto aircraft.

Improvements in laser technology soon allowed for more compact lasers. Bulky single-point rangefinders evolved into multi-point mechanical scanning lidars. Early examples include the first general-purpose mobile robot "Shakey" developed by Stanford Research International and funded by DARPA (1968-1972) using lidar mounted on a rotating mechanical stepper motor [62]. This was followed shortly-after by NASA's Scanning Laser Rangefinder in 1977 which used a gimballed mirror driven by a stepper motor in both horizontal and vertical directions [63]. Notably, a succession of leading-edge innovations at the Environmental Research Institute of Michigan (ERIM) between 1984-1987 culminated in a scanning lidar onboard the Autonomous Land Vehicle (ALV) that achieved a frame rate of 2 Hz covering a FOV of 80° × 30° (256 × 64 pixels) [64]. Once again, these innovations were funded by DARPA.

Flash lidar was first conceived and demonstrated at Advanced Scientific Concepts [65] towards the end of the 1990s. By using a sensor containing a 2D array of pixels and illuminating the entire field of view with the laser, each pixel on the sensor can independently measure the laser return time. This made 3D ranging possible without the need for mechanical scanning. Once again, NASA was quick to adopt this new lidar technology, this time for lunar landing as part of its Autonomous Landing and Hazard Avoidance Technology (ALHAT) project in 2008 [66]. The flash architecture achieved a resolution of 128 x 128 pixels at a frame rate of 1 Hz and importantly eliminated motion artefacts that would otherwise have been present through a scanning lidar approach.

The next evolution in lidar was galvanised by the 2004 DARPA "Grand Challenge", a competition to accelerate the development of autonomous vehicles. This highlighted lidar as a crucial component for self-driving vehicles and prompted participant David Hall to develop the first scanning lidar with multiple laser points. By 2007, five of the six teams to successfully complete the challenge used Hall's scanning lidar, which prompted its commercialisation through his company Velodyne [67].

The DARPA Grand Challenge kick-started the development of self-driving vehicles for the mass market which continues to this day. However, the high cost of multiple-point scanning lidar, coupled with advancements in CMOS fabrication has regenerated interest in flash lidar as a solution for self-driving vehicles. In 2003, the fabrication of single-photon avalanche diodes (SPADs) using standard CMOS processes was achieved [68]. This enabled arrays of highly sensitive pixels for detecting the returning laser to be integrated alongside auxiliary circuitry on a single silicon chip. By 2013, the integrated processing circuits evolved to include photon event timing and memory, producing the first array of direct time-of-flight pixels on a single chip [69]. The adoption of 3D chip stacking in SPAD sensor fabrication processes [50] has further increased the achievable pixel resolution of flash lidar using a single sensor chip. The culmination of all this development was the world's first solid-state lidar in a commercial application: a stacked SPAD lidar sensor developed by Sony in 2020 for the Apple iPhone 11 Pro [70]. A timeline illustrating the evolution of lidar as described here is given in Fig. 7.
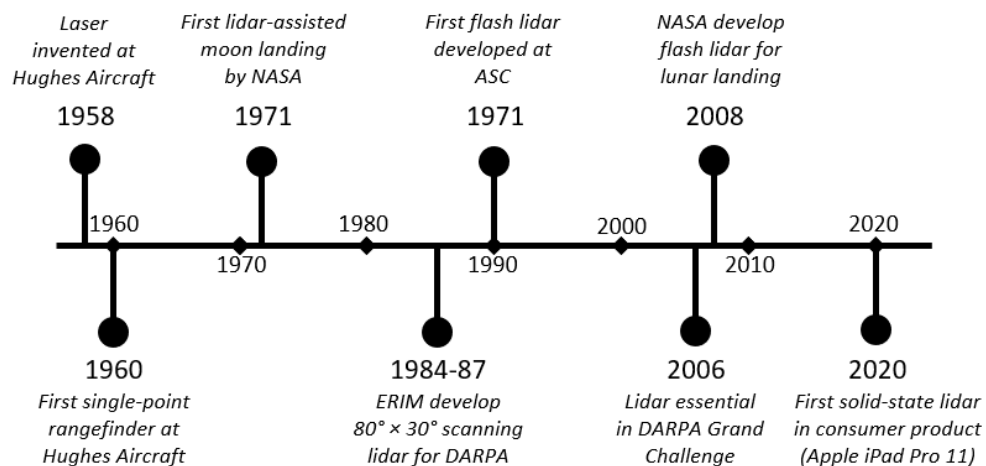
Fig. 7: A timeline of key events in the history of lidar.

2.2 Solid-State Lidar Technologies

Although all lidar sensors measure depth, the underlying technologies used to achieve this vary widely, each with its advantages. The key aspects which define the operation of a lidar are the measurement principle, imaging system and wavelength. The main approaches adopted in solid-state lidar in each of these categories are summarised in Fig. 8.
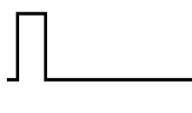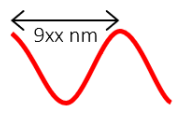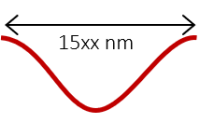
Measurement Principle	Imaging System	Wavelength
DToF	MEMS	NIR
IToF	OPA	9xx nm
	Sequential Switching	SWIR
FMCW	Scanning	15xx nm
	Flash	

Fig. 8: The key aspects which define the operation of solid-state lidar.

2.2.1 Measurement Principle

Direct Time-of-Flight (DToF)

The most intuitive measurement approach is to directly measure the roundtrip time of an emitted laser pulse. This method is known as direct time-of-flight (dToF) and is illustrated in Fig. 9(a). The target distance d is simply given as a function of the laser return time t and speed of light c by (1):

$$d = \frac{\Delta t \times c}{2} \tag{1}$$

While the principle of dToF is simple, implementing it requires the ability to detect photon arrivals with sub-nanosecond (for centimetre) precision. Early dToF sensors combined avalanche photodiodes (APDs) with analogue front-end circuitry to determine the peak return [71]. This progressed to using single-photon avalanche diodes (SPADs) with digital timing circuitry [72] for higher sensitivity in a more compact footprint, once the fabrication of SPADs in standard CMOS processes became established.

Since the returning laser power reduces over longer distances following the inverse square law, the signal becomes increasingly more difficult to detect over background noise. By utilising high peak, narrow pulse width laser signals, the returning signal can remain detectable over longer distances compared to indirect approaches which use a continuous laser signal (below). Shorter laser pulses also allow for higher peak laser power before eye-safety limits are breached [73]. As a result, dToF has been the dominant measurement principle for long-range automotive lidar since the earliest Velodyne HDL-64E automotive lidar in 2005 [74].
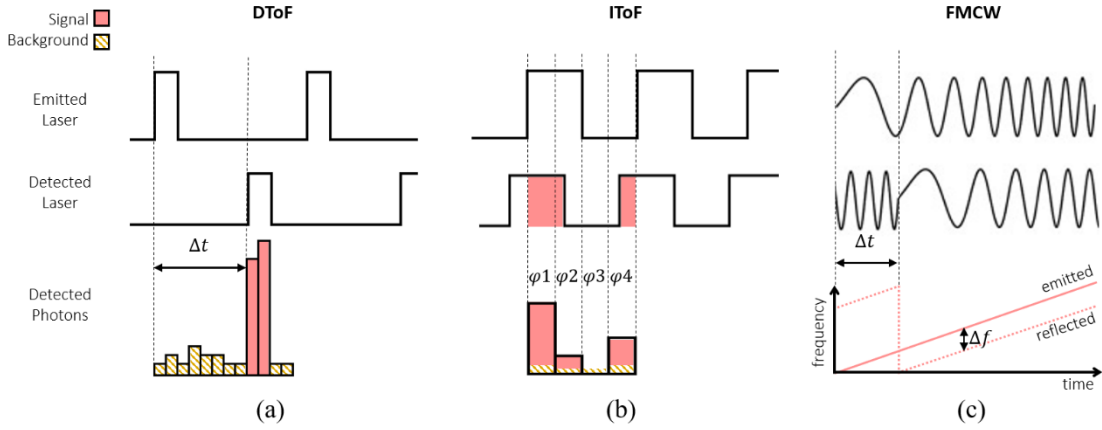
Fig. 9: Illustrations of lidar measurement principles: dToF, iToF and FMCW.

Indirect Time-of-Flight

By emitting an amplitude-modulated continuous wave (AMCW) and measuring the difference in phase of the returning pulse, the roundtrip time can be indirectly measured. This method is known as indirect time-of-flight (iToF). For a conventional four-phase iToF approach, illustrated in Fig. 9(b), the target distance d can be determined as a function of the emitted wave period T and the total signal in each integration phase φ using (2).

$$d = \frac{T \times c}{4\pi} \times tan^{-1}\left(\frac{\varphi_3 - \varphi_4}{\varphi_1 - \varphi_2}\right) \tag{2}$$

Since iToF requires only the ability to integrate the incoming light over coarse time periods, rather than accurate event timing, it can be implemented using established CMOS image sensing technology. Early implementation, known as "lock-in ToF cameras", could be created from charge-coupled devices (CCDs) as far back as the early 1990s [75]. Early commercial products soon followed, starting with the 176×144 resolution SR-3000 by Mesa Imaging. By 2012, Microsoft had replaced the active triangulation approach in favour of iToF [76]. Nowadays, modern iToF sensors such as Sony's IMX556 take advantage of state-

of-the-art CMOS image sensor (CIS) CIS fabrication such as backside-illuminate technology to achieve a pixel resolution of 640×480 and the latest generation Microsoft Kinect Azure achieves even megapixel resolution [77]. Nevertheless, optical power constraints and range ambiguity due to the laser pulse period wrapping around limits iToF to short-range indoor applications [78]. At the time of writing, no known iToF solutions have been developed for automotive applications.

Frequency Modulated Continuous Wave (FMCW)

Frequency modulated continuous wave (FMCW) lidar works by emitting a laser beam with a continuously changing frequency. The reflected signal is compared with the original to measure the frequency shift Δf which is directly proportional to the roundtrip time Δt, as illustrated in (c).

Since FMCW measures changes in signal frequency rather than amplitude, it is less sensitive to background light and could therefore operate at a lower peak laser power than dToF. The signature of the frequency-modulated signal also adds inherent robustness against interference from nearby lidar. In practice, integrating the complex laser and optics required for long-range FMCW lidar into a cost-effective solution for self-driving vehicles remains a major obstacle [15].

Nevertheless, the potential advantages of FMCW lidar have attracted considerable effort to address these challenges. Advances in optical phased arrays (discussed further in Section 2.2.3) have provided a key technology for realising the FMCW scanning mechanism on-chip [79, 80]. More recently, a collaboration between Photodigm and Lumentum took this a step further by developing the first FMCW with all transmit and receiver optics integrated into a single chip [81].
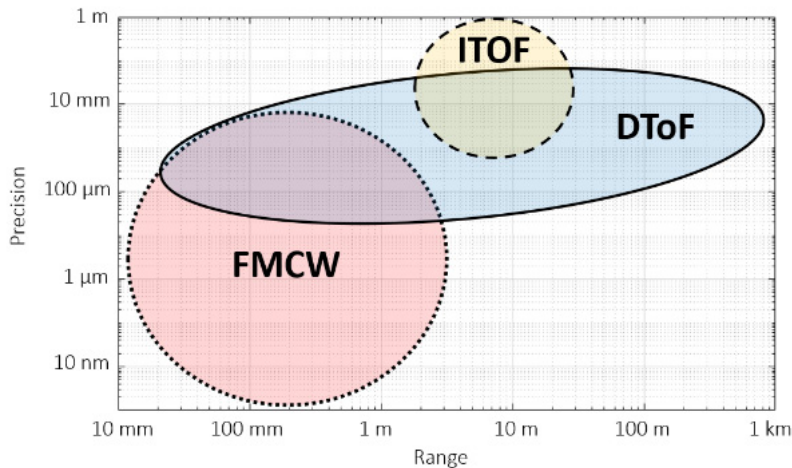
Fig. 10: Summary of published lidar performance for each measurement principle. Adapted from [82].

Fig. 10 summarizes the achieved ranging performance of published lidar research. While iToF remains suited for short-range applications and dToF is the dominant measurement principle for long-range lidar, the full potential of FMCW is yet to be realized and continues to improve in both range and precision. Commercial development from Insight [45], Intel Mobileye [47], and Aeva [83] shows strong interest in FMCW lidar, meaning that it could soon become a practical alternative for long-range lidar.

2.2.2 Wavelength

Automotive lidar commonly operates at a laser wavelength within either the near-infrared (NIR) or shortwave infrared (SWIR) regions. Due to established laser technology from telecommunication applications [84], the most common wavelength of operation are within the 850 nm to 940 nm band for NIR (referred to as 9xx) and 1550 nm for SWIR. Choosing the operating wavelength of a lidar within these values is influenced by many factors which are explored here.

Laser Eye Safety

For long-range dToF, operating at a high laser power is critical. However, this poses a danger to eye safety. The International Standard IEC 60825-1 defines four classes of laser safety, with Class 1 certifying safe operation under all conditions of normal use. In general, the maximum permitted laser power for a system increases as the laser wavelength moves away from the visible spectrum (between 380 nm and 700 nm). While exact values depend on multiple factors such as laser pulse shape, divergence angle and exposure time, operating within the SWIR over NIR typically permits a 10× increase in laser power. This is illustrated in Fig. 11 where the accessible emission limit from NIR to SWIR for a 100 s exposure of a collimated laser (i.e. angular extension less than 1.5 mrad) has been calculated.
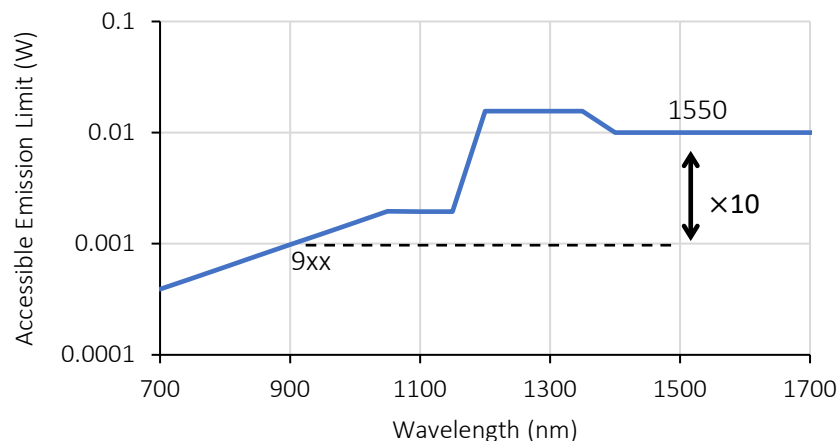
Fig. 11: IEC 60825-1 class 1 eye safety limit for a 100 s exposure of a collimated laser.

Solar Radiation Spectrum

The majority of unwanted background photons detected by the lidar sensor come from solar radiation. However, the solar irradiance on the surface of the earth depends on the wavelength being observed. ASTM International provide a reference for the solar spectral irradiance at each wavelength [85]. This has been reproduced in Fig. 12 and reveals two crucial characteristics which are relevant to lidar.
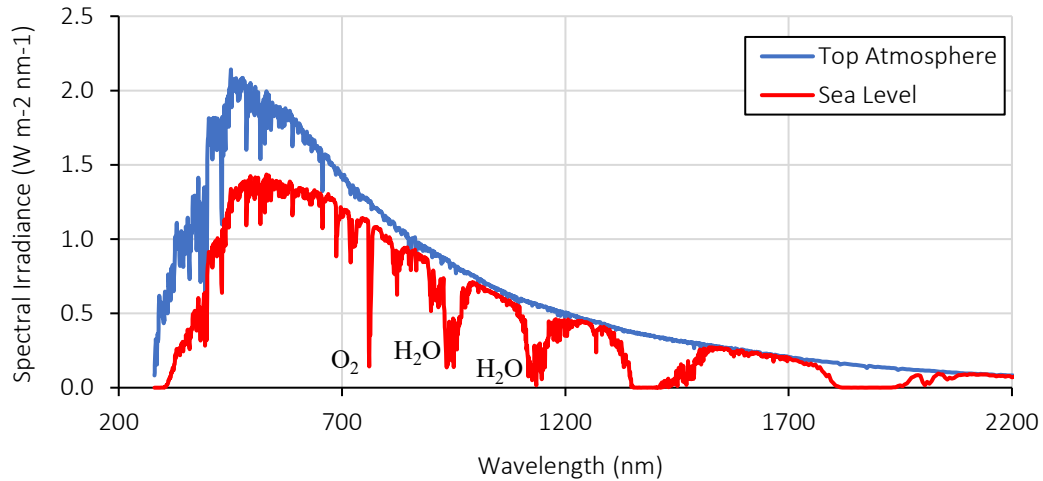
Fig. 12: ASTM G173 - 03(2020) solar spectral irradiance reference data.

Firstly, solar irradiance peaks in the visible range and continues to reduce for increasing wavelengths, allowing SWIR lidar to operate with lower background noise from ambient light compared to NIR. The second notable feature in the solar irradiance spectrum are various notches caused by absorption of different atmospheric constituents such as oxygen (O_2), and water vapour (H_2O). While operating within these notches such as 940 nm or in the SWIR band around 1400 nm may seem appropriate, the same atmospheric effect which attenuates solar irradiance at these wavelengths also attenuates the laser [86]. This effect is particularly problematic in wet and foggy conditions [23]. Furthermore, the irradiance spectrum provided at sea level does not precisely replicate solar absorption across the range of altitudes in which a vehicle is likely to operate, adding further complications. Nevertheless, many lidar developers choose to operate within these solar irradiance notches [87].

Detection Efficiency

While silicon-based image sensors benefit from well-established mass-market manufacturing processes, their peak sensitivity is within the visible spectrum and often extends out to the NIR. Beyond these wavelengths, the sensitivity of silicon-based falls to zero. An example of the photon detection efficiency (PDE) for a commercially available CMOS SPAD sensor [88] is presented in Fig. 13.

To achieve SWIR-sensitive photon detection, compound materials are required. Commercialisation of indium gallium arsenide (InGaAs) detectors is now realised but these require expensive fabrication processes such as molecular beam epitaxy (MBE). Furthermore, elaborate cooling mechanisms are typically required to mitigate high dark count rates (i.e., noise) [89]. State-of-the-art research now focuses on Group IV-based materials such as germanium-tin (GeSn) [90] where more easily tuneable bandgaps and larger wafers could one day make SWIR photon detectors more viable. Silicon-based detectors therefore remain the most common choice for consumer self-driving vehicle lidar, while manufacturers focus on employing cutting-edge fabrication techniques to enhance photon detection efficiency (PDE) at in the NIR band.

The mass-market manufacturability of silicon-based photons detectors mean that they remain the most common choice for consumer self-driving vehicle lidar. Working within the practical limits of silicon-based detectors, manufacturers focus on employing cutting-edge fabrication techniques to enhance photon detection efficiency (PDE) at in the NIR band for improved lidar performance.
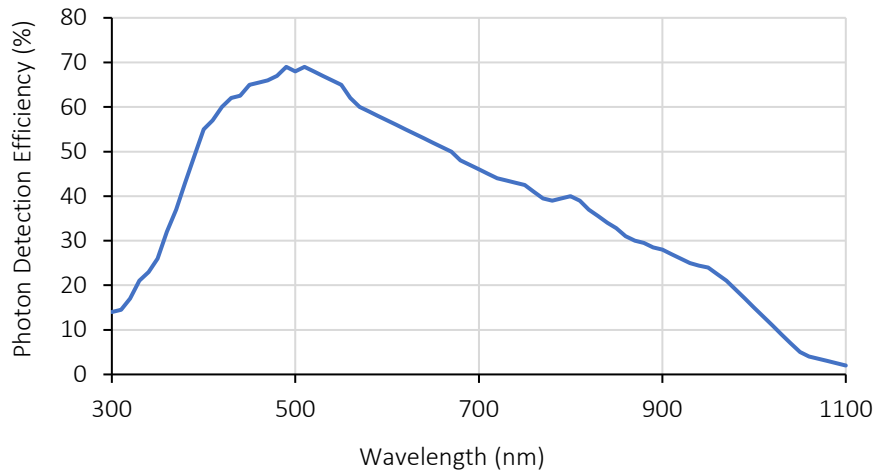
Fig. 13: Sensitivity of SPADs in a currently available process (at 2.5 V excess bias) [88].

Laser Technology

While both NIR and SWIR lasers are well established, NIR lasers can take advantage of low-cost technology such as vertical-cavity surface-emitting lasers (VCSELs) and edge-emitting lasers (EELs). Although some recent progress has been made towards VCSELs and EELs operating at SWIR wavelengths [91], current implementations rely on fibre lasers using complex optical components, resulting in expensive and bulky solutions [92]. In addition, the reduced peak power of SWIR lasers limits them to short-range iToF / FMCW applications.

Summary

A summary of automotive lidar wavelength considerations is given in Table 3. While SWIR lidar offers the potential for higher laser power before breaching eye safety limits, the availability of low-cost silicon detectors and lasers makes NIR the dominant choice for automotive lidar developers at present [45].

Characteristic	NIR (9xx nm)	SWIR (1550 nm)
Permissible laser power	Low	High
Performance under high solar irradiance	Low	High
Performance under adverse weather	High	Low
Detector affordability	High	Low
Laser affordability	High	Low

Table 3: Summary of wavelength considerations for automotive lidar.

2.2.3 Imaging System

Within the context of the lidar imaging system, the first design choice is whether or not to implement a scanning-based system. This can typically involve scanning both the sensor and laser (Fig. 14(a)) or only the laser (Fig. 14(b)). In the latter approach, a lens focuses the laser returning from each point in the scene onto a chip containing a 2D array of detectors. This can be taken a step further by using a laser which illuminates the entire field of view (Fig. 14(c)) and removes the need to scan altogether, known as flash lidar.
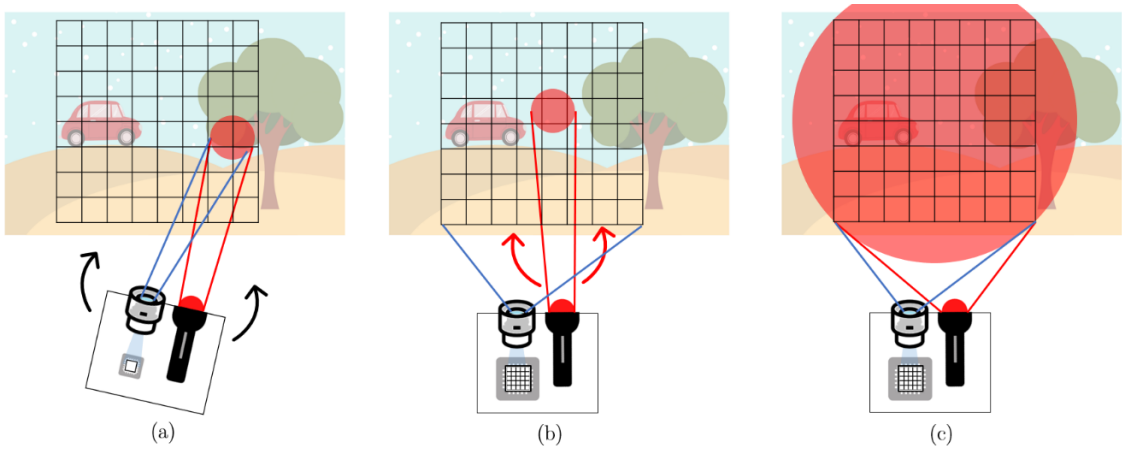
Fig. 14: Illustrating of lidar imaging systems (a) scanning both laser and sensor (b) laser only (c) no scanning (flash).

The scanning pattern used is primarily dictated by the characteristics of the chosen imaging system. Mechanical scanning lidar typically scan multiple vertical points in a 360° motion, while many MEMS-based system adopt a back-and-forth pattern referred to as raster scanning. Smarter scanning patterns such as foveated prioritizes high-resolution data capture in the region of interest (similar to the human eye fovea) while reducing resolution in peripheral areas. These are more novel but have seen recent development including commercial products by Baraja [93].

While scanning enables higher signal-to-background noise compared to flash [94], transitioning from problematic mechanical-based solutions (Section 1.2.4) to solid-state implementations has proved challenging. Additionally, scanning inherently imposes a constraint on the system frame rate contingent upon the scanning FOV. Furthermore, the narrow beam of a scanning lidar is more susceptible to blockage from dust and dirt particles [95]. In contrast, flash lidar removes the need to implement laser scanning and its associated complexities. However, illuminating the full FOV with a single flash places a high demand on the laser power. A review of the most established solid-state lidar approaches follows.

Micro-Electromechanical Systems

Micro-electromechanical system (MEMS) solutions attempt to address the manufacturability and cost issues of traditional mechanical scanning lidar by fabricating micromirrors to steer the laser. Movement is typically controlled using either electrostatic, electromagnetic, electrothermal or piezoelectric actuation and can be either resonant or non-resonant. Resonant scanning MEMS mirrors are easier to implement and offer a wider scanning angle. However, properties exhibit a strong dependence on environmental factors like temperature. The final aspect of a MEMS-based lidar to consider for 2D scanning is the mirror configuration. This can be configured using either single- or dual-axis mirrors. Using two single-axis mirrors (Fig. 15(a)) increases the number of moving parts, hence further deteriorating the robustness of the system. However, single-mirror solutions require either a more complex mirror or laser solution, illustrated in Fig. 15(b) and Fig. 15(c) respectively. An extensive review of MEMS mirror development for lidar is presented by Wang et al. [96].
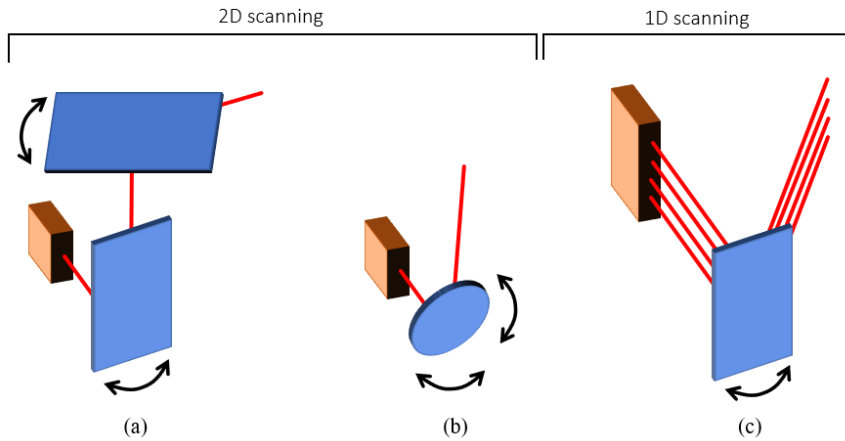
Fig. 15 MEMs scanning configurations (a) single laser, two single-axis mirrors (b) single-laser, one dual-axis mirror (c) laser array, one single-axis mirror.

While MEMS lidar offers a step towards affordable lidar compared to a mechanical approach, the high scanning frequency requirement for automotive lidar inevitably requires expensive mirrors with a high mechanical quality factor (Q) value [96]. The mirrors also introduce inefficiencies due to reflection losses and place constraints on the allowable peak laser power to prevent damage [36]. Nevertheless, MEMS lidar is currently the most realisable solid-state scanning approach and has therefore seen significant development from key industry players including Innoviz, Blickfeld and Robosense [97]. While the robustness of MEMS lidar for automotive remains uncertain, Robosense achieved a significant milestone in 2023 with the first MEMS lidar to qualify for AEC-Q100 automotive grade reliability [98].

Optical Phased Array

By directing the laser beam through an optical phase array (OPA), the phase of the laser can be individually controlled in each channel. This allows for precise manipulation and steering of the laser beam without any moving parts.

Due to the complexity of implementing a 2D array of phase shifters, 2D beam steering is commonly realised using a 1D array of phase shifters to steer in one direction (φ) and by varying the laser wavelength to change its diffraction angle through a grating in the second direction (θ), illustrated in Fig. 16(a). 2D OPA scanning in this way was first demonstrated on a silicon chip by Van Acoleyen et al. in 2019, achieving a maximum phase shift steering angle (φ) of 2.3° and wavelength steering angle (θ) of 14.1° [99]. For independent phase and amplitude control in both directions using a constant laser wavelength, a more complex 2D OPA array is required (Fig. 16(b)). The first integration of this on a single silicon chip was eventually demonstrated by Abediasl and Hashemi [100], albeit with a maximum steering angle of 1.6°.
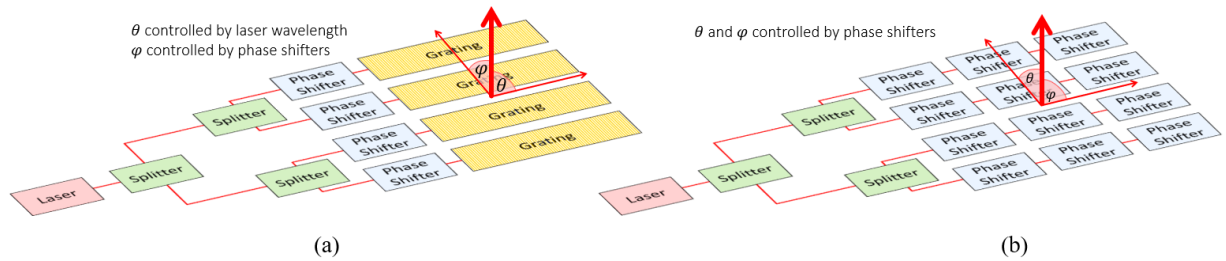
Fig. 16: 2D OPA beam steering using (a) combined phase shift and wavelength tuning or (b) a 2D array of phase shifters.

Achieving a steering angle wide enough to cover the required FOV for automotive continues to be an area of active research [46]. Equally as critical is the limited maximum laser power that can be achieved using OPAs [33]. Despite these challenges, solid-state scanning lidar with no moving parts is an attractive prospect, attracting companies such as Analog Photonics to pursue the commercial development of OPA-based lidar [101].

Sequential Switching

Scanning can also be achieved by discretely switching between emitter elements within a matrix. This can be done either electronically by controlling the current through separate laser elements (Fig. 17(a)),

or optically (Fig. 17(b)) by sequentially routing the laser path using optical switches for example. A lens can then angle the illumination produced from each point in the array onto the scene [102].
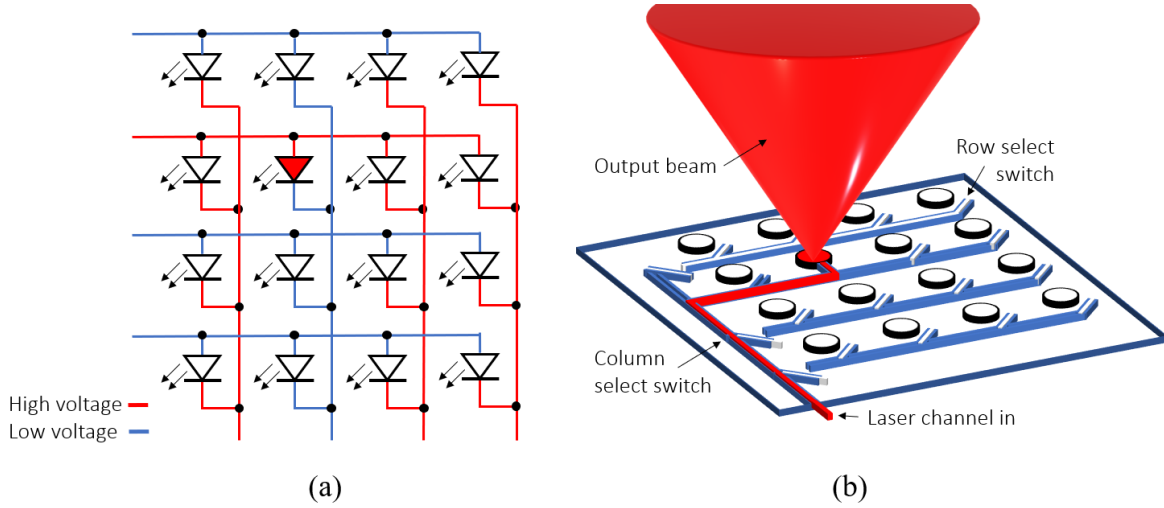
Fig. 17: Sequential switching (a) electronically and (b) optically.

Electronic sequential switching is achieved by integrating a matrix of lasers on a substrate and sequentially addressing the row and column line of individual or groups of lasers. For automotive lidar, this approach is being pursued most prominently by Ibeo (now Microvision) [103]. However, a similar scheme is also adopted by Apple [104] for indoor smartphone lidar applications. While electronic sequential switching presents a simple solution for solid-state laser scanning with no moving parts, the peak laser power in each step is limited by the maximum power of individual (or small groups of) lasers.

On the other hand, optical switching controls the path of laser illumination to be channelled out of a single antenna within a matrix of antennas, as illustrated in (Fig. 17(b)). A variety of state-of-the-art nanophotonics techniques have emerged to implement optical switching for lidar and are well-reviewed in [105]. While this approach is still in the early phases of development, the potential of this approach for power-efficient solid-state scanning lidar with no moving parts has gathered a lot of research momentum. Early developments in 2020 achieved a 10×10 array of elements [106]. By 2022, a 128×128 switch array had already been demonstrated and integrated into a lidar setup achieving 10 m range with a FOV of 16°×16° [107]. At present, the slow switching time achieved (currently many microseconds), among other properties, makes optical switching unsuitable for long-range dToF lidar. But with continued development, this early technology has the potential to soon become a viable option for long-range scanning lidar.

Flash

Flash lidar is performed by illuminating the entire field of view in a single flash, and a chip containing a 2D array of detectors times the laser returning from each point in the scene. This achieves a solid-state lidar solution without the need for any scanning mechanism. Not only does this avoid all the prior challenges of implementing scanning, but it also avoids image artefacts introduced by various scanning patterns [97].

The feasibility of high-resolution iToF sensors, coupled with a low laser power requirement, makes flash the principal choice for short-range indoor lidar. Examples include the 1-megapixel Microsoft Kinect Azure with a maximum range of 4.2 m [77]. However, for long-range dToF, the principal challenges are:

1. high peak laser power required to illuminate the entire field of view over long distances
2. detecting and precisely timing photons simultaneously across all pixels

State-of-the-art laser developments (see Section 2.3.1) and single-photon sensitivity detection (see Section 2.4.1) significantly increase the maximum range capabilities of dToF. While compression of photon data into histograms (see Section 2.4) and 3D chip stacking (see Section 2.5.1) allows more photon data to be collected simultaneously across all pixels. In 2019, flash dToF ranging up to 50 m in daylight conditions with a pixel resolution of 64×64 was demonstrated using a 3D stacked sensor [108]. By 2021, flash lidar operating outdoors at 100 m was demonstrated, adopting on-chip background light suppression to reduce the data load on the sensor [109].

While high laser power and data volume remain a limiting factor for long-range automotive flash lidar, the ease of achieving a solid-state solution with no moving parts brings numerous commercial development interests. Companies developing flash lidar for automotive applications include Fastree 3D [110], PreAct [111] and Ouster [112].

2.2.4 This Work

Given the current state of lidar technology reviewed here, the lidar developed in this work will use the **dToF** measurement principle, operating in **NIR** wavelengths to achieve a long-range solution for automotive applications. While a flash system will be implemented, this work applies to **any lidar imaging system using a 2D dToF sensor array**.

2.3 Overview of DToF Systems

Independent of the imaging system, the main components of a dToF lidar system are the laser, sensor, control logic, lens and optical filter. These are illustrated in Fig. 18(a).
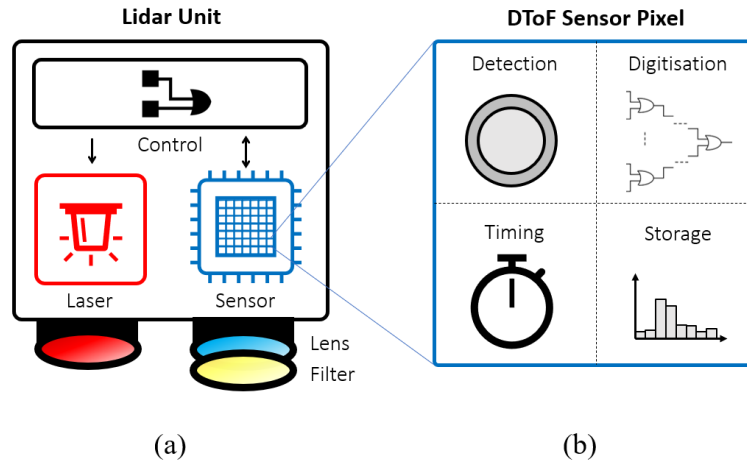
Fig. 18: Main components of (a) a dToF lidar system and (b) a dToF sensor pixel.

2.3.1 Laser

The spatial coherence and narrow spectrum of lasers make them ideal illumination sources for lidar. While the first gas-based lasers made lidar possible, modern solid-state lidar requires a more compact solution in the form of semiconductor lasers, most notably edge-emitting lasers (EELs) and vertical cavity surface-emitting lasers (VCSELs). The typical characteristics of both VCSELs and EELs as described in [113, 114] are summarised in Table 4.

Wall-plug efficiency (WPE) describes the efficiency of converting electrical to optical power. While this depends on temperature and drive current, EELs are generally considered to have a better WPE than VCSELs, with up to 100 times greater achievable brightness being their most attractive characteristic [113]. On the other hand, to make optimal use of the laser power, it is crucial to have an evenly distributed laser beam that matches the FOV of the lidar sensor and VCSELs produce a native circular beam that is well-suited to this requirement [115]. Additionally, matching the laser's bandwidth to that of the background filter is necessary to minimize filtering of the laser signal, and the wavelength stability of VCSELs makes them more suitable for this reason. Finally, the ability to test individual VCSELs at the wafer stage before dicing makes them more suitable for large-scale manufacturing [116].

Characteristic	EELs	VCSELs
WPE @ 60 °C (at nominal operating current)	45%	40%
Brightness	1 MW/mm²-sr	10 kW/mm²-sr
Wavelength stability	~0.3 nm/°C	~0.08 nm/°C
Beam shaping	Complex	Simple
Manufacturability	Complex	Simple

Table 4: Comparison of typical EEL vs. VCSEL characteristics. [113, 114]

While EELs have their advantages, the choice of laser type for a specific lidar application ultimately depends on a range of factors including range, sensitivity, field of view and cost. It should be noted that both VCSELs and EELs are rapidly evolving technologies (e.g. recent innovations in multi-junction high-power VCSELs [117]) and the relative advantages of each laser type continue to shift.

2.3.2 Lens

The function of the lens is to focus light returning from the scene onto the lidar sensor. The main characteristics of a lens are *focal length*, *diameter*, and *aberration*. The relationship between these characteristics and the trade-offs they induce are illustrated in Fig. 19.
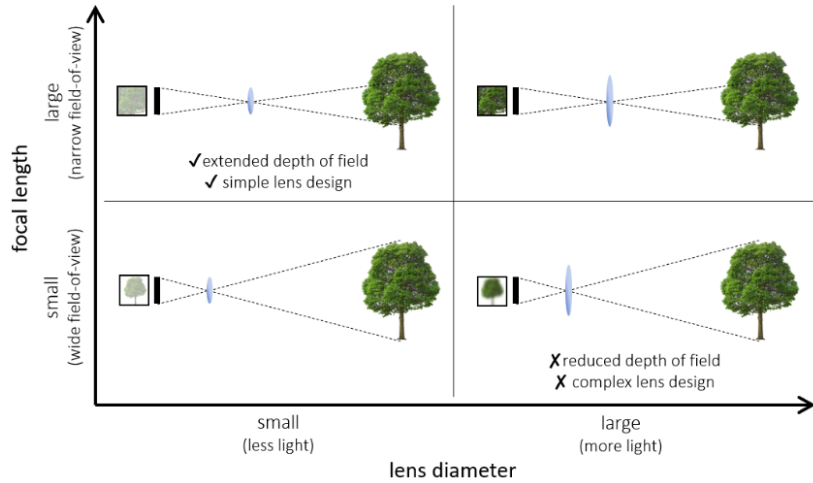
Fig. 19: Lens design trade-offs.

The focal length is the distance between the lens and the point where the light rays converge, also known as the focal point. A lens with a shorter focal length will have a wider FOV. The ratio of focal length (f) to lens diameter (D) is referred to as the f-number ($f\#$).

$$f\# = \frac{f}{D} \tag{3}$$

A larger lens diameter (lower $f\#$) allows more light onto the sensor. In lidar, where maximizing the number of returning laser photons is critical, this is a highly desirable feature. However, lenses with lower $f\#$ result in a narrow depth of field, meaning that only objects within a narrow distance region will appear in focus. Furthermore, the complexity of designing lenses with low $f\#$ results in high cost and poorer optical quality. The optical quality of a lens is determined by its ability to refract light without distortion or aberration. High-quality lenses will produce clear, sharp images with minimal distortion, while lower-quality lenses may produce blurred or distorted images.

2.3.3 Filter

An optical bandpass filter is used to remove any ambient light outside of the chosen system (laser) wavelength. While a narrow bandwidth reduces the number of unwanted background photon detections, a filter with too narrow of a bandwidth may result in the laser wavelength falling outside of the passband which varies due to process and environmental factors such as temperature. While VCSELs offer improved temperature stability, the broad operating range of 40 °C to 105 °C required for automotive qualification [118] means that even a small temperature coefficient result in a large variation. This is illustrated in Fig. 20 where a typical VCSEL temperature of ~0.08 nm/°C is assumed, along with +/- 10 nm process variation, resulting in over 30 nm of wavelength variability. Not only does the optical filter need to cover this large laser wavelength variability but the environmental sensitivity of the filter itself also needs to be accounted for.
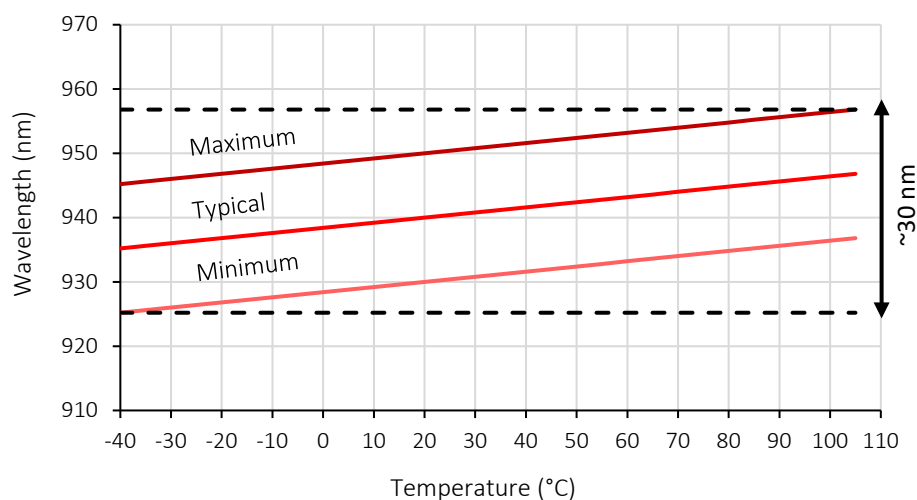
Fig. 20: Example of VCSEL wavelength variability over process and temperature.

2.4 DToF Sensors Overview

With an overview of all the other key lidar components illustrated in Fig. 18(a), the final section of this discussion focuses on the sensor. The four key components of a dToF sensor, as shown in Fig. 18(b), consist of detection, digitisation, timing, and storage.

Much of the early work regarding CMOS dToF sensors is attributed to MIT Lincoln Laboratory, with the first published work in 2001 [119]. This was made possible by two enabling technologies: solid-state lasers capable of producing short pulses and the use of avalanche photodiode (APD) sensors to produce a digital pulse from even a single photon arrival. The fabrication of APD arrays in CMOS had not yet been achieved and, as such, an array of APDs was stacked on top of a separate chip containing an array of CMOS timing circuitry. Over the years, this would develop to stack a third chip layer, allowing for more complex and precise timing circuitry to be integrated [120].

In 2003 however, the fabrication of single photon avalanche diodes (SPADs) in CMOS was achieved [68]. This breakthrough made it possible to integrate highly sensitive photon detectors and timing circuitry on the same chip, without the complexity and cost of stacking different chips. It was followed shortly by the first CMOS SPAD ToF sensor in 2004 [72]. Over time, CMOS SPADs developed to achieve greater sensitivity [121-123], lower noise [124, 125] and timing circuitry evolved to process multiple events per laser cycle (multievent TDCs) [126].

2.4.1 Detection

A photodiode serves as the fundamental building block for CMOS light sensors. When the photodiode is reverse-biased, it allows incident light to generate a current flow within the device. The extent of reverse biasing determines the operational state of the photodiode. At a low reverse, the current increases linearly with the amount of light. Operating in this way, the low light levels must be integrated for long periods of time to detect the resulting low current. A SPAD is formed when the reverse bias exceeds the breakdown voltage (V_{BD}) of the diode by an excess voltage (V_{EX}). Under these conditions, a single photon is enough to cause an avalanche of current. This provides the instant signal required to accurately measure the return time of the laser pulse even when the returning laser signal is small at long distances.

Upon the SPAD avalanche current being induced, it must be quenched by reducing the SPAD bias voltage below breakdown and recharged by restabilising the overdrive voltage (Fig. 21(a)). The most basic method of achieving this is through a passive quench series resistor (Fig. 21(b)), which reduces the voltage across the SPAD when the avalanche current flows. This is typically followed by a logic gate or comparator which converts the signal to a low-voltage domain. The combination of quench and voltage conversion is referred to as the SPAD front-end (FE). While numerous quench/recharge topologies exist, these are beyond the scope of this thesis. The key takeaway is that, between a photon arriving and the SPAD recharge, a dead time (t_d) exists during which additional photon arrivals are not counted (Fig. 21(c)). For this reason, SPADs can be grouped in parallel to form a macropixel, allowing SPADs within the group to fire while others recharge. Various techniques are used to combine and process the signals from macropixels containing multiple SPADs as will now be discussed.

Fig. 21: (a) SPAD avalanche, quench, and recharge phases (b) passive quench configuration (c) dead time causing photon arrivals to be missed.

2.4.2 Digitisation

A simple method of digitising and combining photon activity from multiple SPADs is using a combination of pulse-shaping (PS) followed by a network of OR gates [127] as illustrated in Fig. 22(a). Each PS shortens the period of electrical pulse signal that arises from the FE to decouple it from the SPAD's deadtime period. This increases the maximum achievable photon count rate of all SPADs in the macropixel when combined at the OR gate which changes state to indicate one or more SPADs have fired. Nevertheless, even with pulse shaping the OR gate eventually saturates under high enough photon rates, resulting in no change at the output and producing ambiguity in the number of SPADs that fired. The toggle flip-flop + XOR tree technique (Fig. 22(b)) removes the dependence on pulse width altogether, increasing the maximum count rate of the macropixel [128]. This technique eventually reaches its limit when the photon rate exceeds the time required for the XOR gates to switch state.

Under even higher photon return rates, coincidence detection (Fig. 22(c)) further extends the maximum count rate of the macropixel [48] by summing photon arrival events and only outputting the result if the sum exceeds a predetermined threshold. This is an effective technique for increasing the maximum count rate of the macropixel but requires prior knowledge of the expected arrival rate for each pixel to ensure the threshold does not result in missed photon arrivals.

The high photon rates experienced by a lidar sensor operating outdoors demand extremely high throughput from each macropixel. Even though the toggle flip-flop XOR approach improves throughput over the OR technique, the resulting lidar depth image is still severely degraded using this technique in an outdoor environment [129]. To address this, the synchronous summation technique (SST) enables all detected SPAD events to be combined without missing a count. The implementation is illustrated in (Fig. 22(d)) [130]. By using a distributed clock (CLK) and summing SPAD events within each clock cycle, SST trades off photon arrival timing precision for dynamic range. SST exceeds the count rate achieved by OR/XOR techniques without relying on prior knowledge of the scene (as required for coincidence detection) [129].

Fig. 22: SPAD SiPM digitization techniques (a) pulse-shaping + OR tree (b) toggle + XOR tree (c) coincidence detection (d) synchronous summation technique (SST). Adapted from [129].

2.4.3 Timing

An integral part of a time-of-flight sensor is the ability to count the elapsed time between the laser firing and photons arriving. This is achieved using a time-to-digital converter (TDC). The most common types of TDC are counter, gated ring oscillator (GRO) and delay-line (DL).

Counter

Of these, counter TDCs are the simplest, achieved by distributing a clock to increment a local counter in each ToF pixel. A globally distributed START or STOP signal indicates the start/end time intervals. While a counter TDC achieves timing with high linearity, the precision that can be achieved from the period of a globally distributed clock is typically limited to a few nanoseconds. In addition, driving and rebuffering the clock signal over increasing numbers of pixels leads to high power consumption.

Fig. 23: (a) Gated ring oscillator and (b) delay line TDC architectures.

Gated Ring Oscillator (GRO)

A faster counter clock can be achieved by generating the clock locally to each TDC using a GRO. A GRO is a chain of an odd number of inverters, with the output of the first fed back to the input of the first as shown in Fig. 23(a). This feedback creates an oscillation when the inverters are enabled, clocking the local counter. The faster clock period that can be achieved by removing the need for a globally distributed clock allows for much finer precision, with the achievable period t_{clk} determined by the time delay of a single inverter t_{inv} and the number of inverters in the chain n_{inv}.

$$t_{clk} = 2t_{inv}n_{inv} \tag{4}$$

Typical inverter delays lie in the range of few to tens of picoseconds, leading to sub-nanosecond time resolution. Further precision can be extracted by reading the state of the inverters when the TDC is stopped, given by OUTF in Fig. 23(a). Implementing a GRO in every pixel leads to significant power

43

consumption. The poorly defined delay time of each inverter also creates ambiguity in the clock period and varies with process, supply voltage, and temperature (PVT). Furthermore, the clock period of each GRO is likely to differ between pixels, requiring each pixel to be independently calibrated. However, it has been demonstrated that coupling neighbouring GROs can effectively synchronize the phase and frequency of adjacent oscillators, offering a potential solution to this issue [131]. This is achieved by either resistively or capacitively coupling between matching nodes in neighbouring GRO. Given enough time the coupled GROs will converge and "lock" to the same frequency and phase. This technique is known as injection locking.

Delay Line (DL)

A delay line TDC, shown in Fig. 23(b), measures time by triggering a chain of delay elements and observing the number of elements that the signal has propagated through once the STOP signal is raised. As with GROs, these remove the need for a globally distributed clock, allowing a finer timing resolution limited by the minimum inverter delay time (roughly 6 ps for a 40 nm CMOS process) [132]. Fewer toggling elements compared to GRO result in reduced power consumption and for solutions requiring only a few counting intervals, DL TDCs also provide a more compact solution.

The main disadvantage of DL TDCs is the non-linearity arising from different delay times through each step in the chain. While this can be partially compensated for by calibration, these also have a strong dependence on process, voltage and temperature.

Time-to-Analog Converter (TAC)

A review of TDCs would not be complete without an overview of time-to-analogue converters (TAC). This typically involves sinking a current into a capacitor, producing a voltage proportional to time which can then be digitised using an analogue-to-digital converter (ADC). As CMOS (and lidar sensors) scale to use smaller technology nodes with lower supply voltages, this approach has been largely abandoned in favour of a fully digital TDC approach such as those described above.

2.4.4 Storage

An established technique for measuring the laser return time is time-correlated single photon counting (TCSPC). In this approach, the laser is pulse and a TDC measures the first photon arrival with high precision. The data is read from the sensor and the cycle is repeated multiple times to establish the peak signal return time over background noise events, as illustrated in Fig. 24(a).
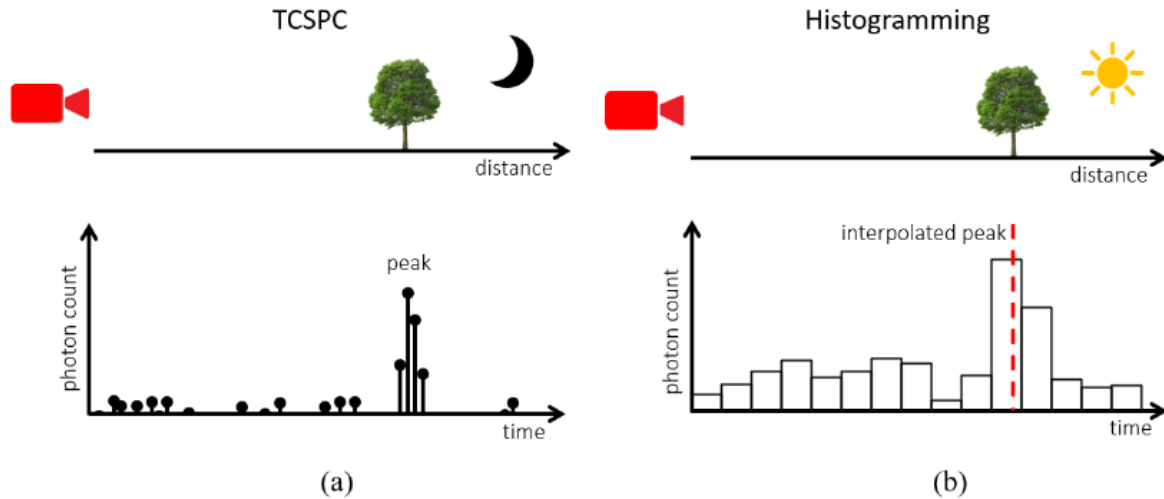
Fig. 24: (a) Fine precision TCSPC over multiple laser cycles under low background light (b) compression of photon data into a histogram under high background light.

For low-resolution imaging with low background noise such as fluorescent lifetime imaging, this technique provides an adequate solution [133, 134]. However, for long-range (i.e. long exposure time) lidar under high ambient background conditions and high image resolution, timing each photon event in every pixel with high precision and reading out all this data between laser cycles is not feasible.

As a result, CMOS dToF sensors evolved to not only detect and time photon arrivals but also to store and process arrivals from multiple laser cycles. A key technique which allows all this functionality to be integrated on-chip is histogramming. By sorting photon arrival times into coarse time bin intervals, the amount of photon data required to be stored is reduced, as shown in Fig. 24(b). Precision can then be recovered by interpolation between peak bins. This was first realised by Niclass et al. in 2013 in which ranging up to 100 m was demonstrated with 16 dToF pixels on a single CMOS chip [69].
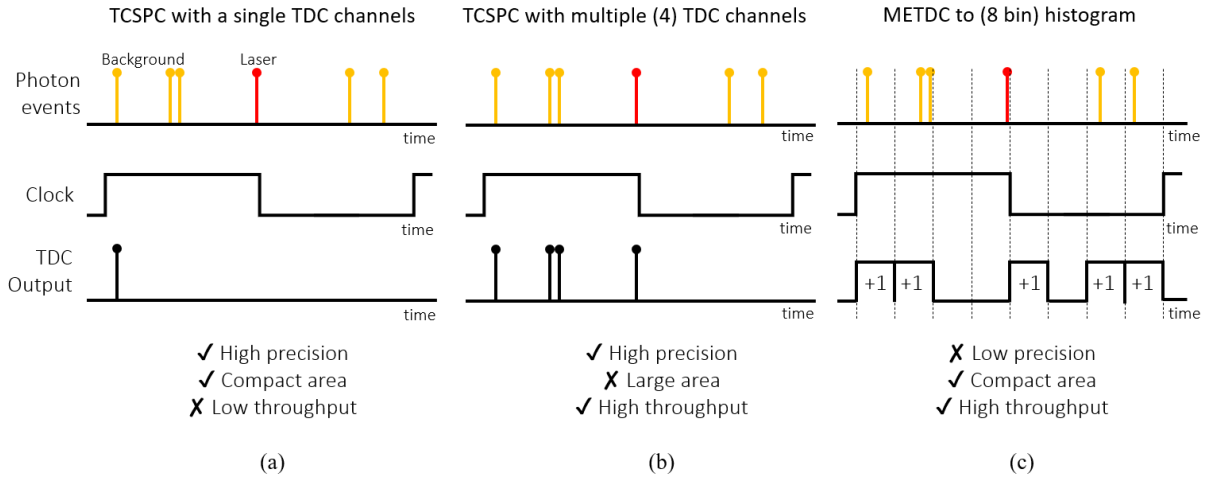
Fig. 25: Timing photon events using (a) TCSPC with a single channel TDC (b) TCSPC with multiple TDC channels and (c) METDC to histogram.

Sensor architectures adopting histogramming can also exploit the use of multi-event TDCs (METDCs) [126, 135], illustrated in Fig. 25. In the presence of high background noise, counting only the first photon event per laser cycle results in missed signal events. While using multiple time-interleaved TDC channels increases the maximum throughput of events per laser cycle, it results in a proportional increase in TDC area. An METDC architecture on the other hand, allows multiple photon events to be timed. While a multi-channel TDC approach can achieve timing precision independent of the final histogram resolution, the matched resolution of an METDC with the final histogram is a well-suited solution.

46

2.5 State-of-the-Art DToF Sensors

While histogramming plays a vital role in enabling high pixel resolution dToF lidar, additional advancements have been necessary to meet the long-range outdoor depth-sensing performance required for self-driving vehicles. Two extensively adopted techniques are *3D chip stacking* and *partial histogramming*. These have found widespread application in the current state-of-the-art dToF CMOS sensors, either individually or in combination.

2.5.1 3D Chip Stacking

Traditionally, CMOS image sensors (CIS) were designed as single-tier chips where the detection components and the supporting circuitry were integrated together (Fig. 26(a)). However, as the demand for higher resolution, improved image quality, and enhanced performance grew in the early 2000s, CIS manufacturers started exploring new approaches. They began developing 3D stacked sensors, which involved fabricating the detection components and support circuitry on separate wafers and vertically stacking them (Fig. 26(b)). This stacking technique allowed for greater pixel density while creating space for more advanced circuit processing. The development of stacked CIS began in the mid-2000s, and it reached a significant milestone in 2012 when Sony introduced Exmor RS, the first commercially available stacked CIS process [136]. Early stacked implementation such as these used through silicon vias (TSVs) to make electrical connections between the top and bottom tier [137] while modern stacked sensors employ hybrid bonding to further increase the total silicon area available for detection [138].
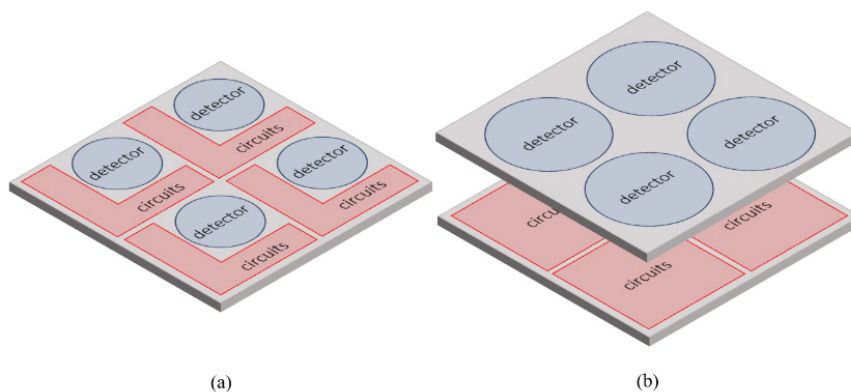
Fig. 26: Arrangement of photon detectors and supporting circuitry using (a) single-tier (b) 3D stacked chips.

With 3D chip stacking established for commercial image sensors, interest in this technology re-emerged for time-of-flight sensors, recognising the opportunities for increased photon detection and/or processing. In 2016, the first 3D-stacked chip containing a 2D array of SPAD detectors was published by **Al Abbas et al. [50]**, a collaboration between STMicroelectronics and the University of Edinburgh, paving the way for 3D stacked time-of-flight SPAD sensors.

The first 3D stacked direct time-of-flight sensor making use of 3D stacking was developed in 2018 by **Ximenes et. al [139]** from École Polytechnique Fédérale de Lausanne (EPFL) and fabricated by Taiwan Semiconductor Manufacturing Company (TSMC). It features 16×8 macropixels in groups of 8×8 SPADs, a 64×21-bit memory in each macropixel and a ring-oscillator TDC shared between pairs of macropixels. Rather than using a histogram, the sensor employs a digital infinite impulse response (IIR) filter, taking advantage of increased processing area. This filter averages successive photon arrival times to reach an estimated peak arrival time in the presence of ambient photons (noise) without the need to store all photo events. The filter can also be weighted towards either new photon events or the average of previous arrival times to allow trading off filtering for response time. Each filtered value is stored in the corresponding SPAD memory address to maintaining ToF pixel resolution equal to the total number of SPADs. The option of using a laser signature to scramble interfering laser signals is also available. In a follow-up publication [51] the sensor demonstrates ranging up to 300 m with 0.5 m error, although this is only given for a single-point (2D) measurement and neither the frame rate nor ambient background conditions are provided. Given the sensor's reliance on an IIR to produce the laser return time, ranging under the influence of high ambient background is likely to severely limit precision and/or increase acquisition time.

In 2019, **Henderson et al. [108]** (University of Edinburgh) presented QuantIC4×4, a stacked SPAD sensor featuring 64×64 macropixels with 4×4 SPADs in each. This sensor utilised the additional processing area to implement reconfigurable pixels capable of performing various functions including high dynamic range imaging, TCSPC and direct time-of-flight. Each macropixel is capable of storing 16×14-bit histogram bins and contains a METDC which can be configured to use either a global clock counter or a local GRO. In a follow-up publication [49], the sensor is demonstrated to range at 50 m while running at 30 fps under daylight conditions.

New 3D stacked SPAD dToF sensors were not published again until 2021. The first of which was published at the International Solid-State Circuit Conference (ISSCC). **Kumagai et al. [55]** (Sony)

presented a 189×600 stacked SPAD array which can be grouped into 63×200 macropixels of 3×3 SPADs, or 31×100 macropixels of 6×6 SPADs. Processing area is dedicated to SRAM banks for storing histogram data (amount undisclosed) as well as a filter to interpolate the peak return time on-chip. The sensor achieved the highest SPAD PDE performance in the NIR region at the time, 22% at 905 nm. The features combined to achieve a maximum sensing distance of 200 m with 30 cm accuracy running at 20 fps under maximum daylight conditions. In 2023 Sony would go on to announce the release of their first commercial dToF sensor, IMX611, based on this work [140]. In the same conference, **Preethi et al. [141]** (EPFL) showcased Jatayu, a stacked SPAD dToF sensor of 16×8 macropixels containing 16×16 SPADs each. As with EPFL's first stacked dToF sensor, histogramming was not adopted. Instead, a TCSPC-like approach was adopted, capturing the first photon event in each half-macropixel with fine (14-bit) precision, and the next 8 events with coarse (2-bit) precision. This is combined with coincidence detection and a programmable time (distance) window for each macropixel, reducing the effect of background photons which otherwise severely limits TCSPC-based lidar in long-range outdoor applications (see Section 2.4.4). Ranging up to 100 m with 0.7 m error is demonstrated under low outdoor lighting conditions (10 klux), although this is only given for a single-point (2D) measurement and at an undisclosed frame rate.

Continuing the publications of stacked SPAD dToF in 2021, **Stoppa et al. [54]** (AMS-Osram) presented a stacked SPAD dToF sensor aimed at indoor smartphone applications at the International SPAD Sensor Workshop (ISSW). It is comprised of 80×60 macropixels, each containing 4×4 SPADs; capacity for 32×12-bit histogram bins; and a configurable delay line TDC. The sensor also adopts a novel sliding partial histogram method to maximise the utilisation of limited histogram bins, discussed further in Section 2.5.2. The combination of these attributes results in 4.4 m ranging within 5% accuracy running at 30 fps under outdoor conditions of 50 klux. Note another stacked SPAD sensor with dToF functionality was presented at the same conference [142]; however, since no performance metrics were disclosed, it is not further discussed here.

Zhang et al. [143] closed off 2021 with a 240×160 stacked SPAD dToF, grouped into macropixels containing 4×4 SPADs. Macropixels are time multiplexed (rolling shutter) in groups of 4, sharing a delay line TDC and SRAM memory of 32×8-bit histogram bins for each SPAD. Each macropixel stores both photon return time and SPAD address, allowing the full 240×160 image resolution to be preserved. The sensor is designed for indoor smartphone applications and achieves a suitable 9.5 m maximum range with 1 cm accuracy. Finally, in 2022 **Taloud et al. [144]** (AMS-Osram) built upon

the previously described work of Stoppa et al. [54]. By switching from 40 nm technology to 22 nm on the bottom tier, the sensor accommodates an exceptional 59×12-bit bins per macropixel with a resolution of 42×32 ToF pixels. The delay line TDC is also upgraded to adopt dynamic element matching, allowing the nonlinearity in this style of TDC (see 2.4.3) to be averaged out. The result is a maximum range of 8.2 m indoor/5 m outdoor (60 klux) with 30 cm accuracy for a target with 92% reflectivity.

2.5.2 Partial Histogramming

Even with the use of 3D stacking, the total area available for storing dToF photon histogram data reaches a limit, constraining the achievable maximum range and/or precision. To address this, various novel *partial histogram* approaches have recently been adopted. These aim to resolve the photon return time with the same maximum range and precision of a conventional full histogram approach but with only a limited number of histogram bins.

The amount of reduction achieved by a given partial histogram implementation is termed the histogram reduction ratio (HRR). To date, all partial histogram approaches can be classified into two main categories, zooming and sliding, illustrated in Fig. 27.
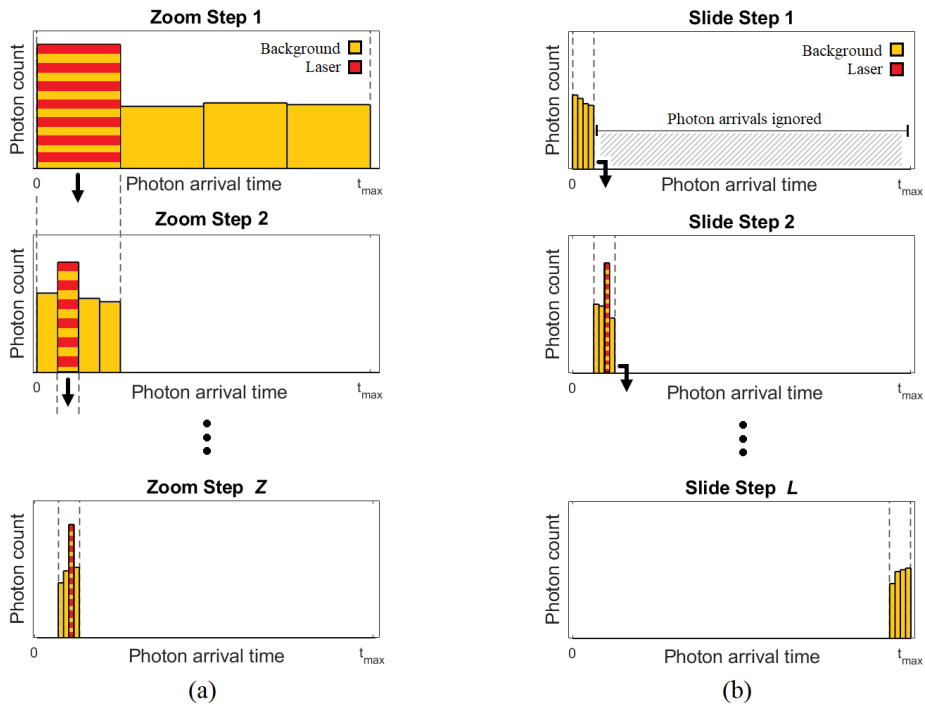
Fig. 27 Illustration of partial histogram methods (a) zooming over "Z" steps and (b) sliding over "L" steps.

Zooming

Fig. 27(a) illustrates the principle of zooming. The reduced set of histogram bins is initially allocated to cover the full distance range and, after multiple laser cycles, a peak is identified. The histogram range is then readjusted to a narrower time interval centred around the identified peak and another iteration of lidar acquisition is performed. Multiple zooming steps Z can be carried out until the desired resolution is achieved. A zooming histogram architecture using M bins at each step achieves a histogram reduction ratio of:

$$HRR = \frac{M}{M^Z} \tag{5}$$

In 2019, **Zhang et al.** [**52**] (EPFL) published the first dToF sensor capable of independent per-pixel histogram zooming. The sensor contains 252×144 SPAD pixels, each with SRAM memory for 8×10-bit histogram bins. Each frame is composed of 3 zooming steps to achieve a HRR of 1/64. The resulting performance is a maximum range of 50 m with 8.8 cm accuracy using a 60% reflective target. However, this is a result of a single-point (2D) measurement and, while the sensor can achieve a frame rate of 30 fps, the frame rate for this measurement is not disclosed. By moving to a stacked process, a revised sensor by the same author was able to increase the histogram bin capacity from 8 to 32 bins [**143**], allowing zooming to be reduced to a 2-step approach.

In 2021, **Kim et al.** [**53**] published an ambitious effort using only 2 histogram bins for each dToF pixel. By zooming 8 steps per frame (HRR of 1/128), this arrangement effectively trades off histogram bin capacity for frame rate. Reducing the photon timing to allocation into one of two bins removes the need for a TDC altogether; relying on a simple up-down counter. The result is a single-tier dToF sensor with 48×40 macropixel and 6 SPADs per macropixel, achieving a maximum range of 45 m with 4 cm accuracy. Relying on a high number of zoom steps undoubtedly has implications on the achievable frame rate, which is not disclosed as a result. This was noted in a follow-up publication by **Park et al.** [**145**] and addressed by dividing each zoom step into 2 halves, with separate up-down counters for each half. Equivalent This effectively covers 4 histogram bins at each zoom step and reduces the number of zooming steps from 8 to 4. This enables the sensor to range up to 9 m at 30 fps and 33 m at 1.5 fps, although the frame rate at the maximum quoted range of 45 m is not disclosed.

Sliding

In a sliding approach, the pixel starts by spreading its reduced histogram bins across only a subset of the full distance range. After sufficient laser cycles have been integrated, the time window is slid to a new time range and the process repeats until the full distance range has been covered. Assuming non-overlapping time windows, a sliding achieves a histogram reduction ratio equal to the number of slide steps L.

The sensor presented by **Stoppa et al.** [54], described earlier in Section 2.5.1, was the first sliding published sensor to adopt the sliding histogram approach. Each pixel uses 32 histogram bins to slide across 16 time windows with 6 bins of overlap between windows to cover edge cases. The next iteration of this sensor by **Taloud et al. [144]**, also already described in Section 2.5.1, maintained a sliding histogram approach. While the number of slide steps is not disclosed, the nearly twofold increase in the maximum range from 4.4 m to 8.2 m for the equivalent increase in histogram bins suggests that a similar number of steps was adopted.

Finally, a sensor by **Gyongy et al.** [146], first presented in 2021 at ISSW, can be programmed into a continuous sliding histogram mode with up to 128 steps. Each of the 32×16 macropixels accommodates 8×12-bit histogram bins. While operating in continuous sliding is not the primary mode of this sensor, outdoor performance running in continuous sliding over a limited 16 steps has been demonstrated at 6.25 fps in a follow-up publication [147].

Laser Power Penalty

Partial histogram techniques allow dToF lidar sensors to increase the maximum range and/or precision of a dToF sensor in the face of limited on-chip area, they impose a severe laser power penalty. In zooming, many laser cycles need to be used to determine the time window of interest, often over multiple zoom steps. In sliding, most time windows being observed at each step do not contain the returning laser signal while it is being pulsed. While additional laser power is generally undesirable, this is particularly problematic for flash implementation which already require a high laser power to illuminate the entire field of view at once.

Author	Ref	Year	Resolution (ToF pixels)	Maximum Distance (m)	Ambient Intensity (klux)	Precision / Accuracy (m)	Frame Rate (Hz)	Histogram bins/ToF pixel	Stacked	Partial Histogram
Ximenes	[139]	2018	64 × 128	300 [1]	-	0.47 / 0.8	20 [2]	-	Yes	No
Henderson	[108]	2019	64 × 64	50	-	- / 0.17	30	16	Yes	No
Zhang	[52]	2019	144 × 252	50 [1]	-	0.0014 / 0.88	30 [2]	8	No	Zooming
Okino	[148]	2020	900 × 1200	250	-	1.5 / -	-	-	No	No
Kim	[53]	2021	40 × 48	45	-	0.014 / 0.023	-	2	No	Zooming
Kumagai	[55]	2021	63 × 168	200	117	- / 0.3	20	-	Yes	No
Preethi	[109]	2021	128 × 256	100 [1]	10	- / 0.07	-	-	Yes	No
Gyongy	[146]	2021	32 × 64	50	20	0.13/0.05	50 [3]	8	64	No
Zhang	[143]	2021	160 × 240	9.5	10	0.01 / 0.02	20	32	Yes	Zooming
Park	[145]	2022	60 × 80	45	30	0.015 / 0.025	1.5 [2]	4	No	Zooming
Taloud	[144]	2022	32 × 42	8.2	1	0.007 / 0.03	30	59	Yes	Sliding
Zhang	[143]	2021	160 × 240	9.5	10	0.01 / 0.02	20	32	Yes	Zooming

Table 5: A summary of state-of-the-art CMOS dToF publications.

---

[1] Single point measurement.

[2] Frame rate not specified at maximum distance.

[3] Frame rate not given in sliding partial histogram mode.

2.5.3 Summary

A performance summary of published state-of-the-art CMOS dToF sensors discussed in this chapter is presented in Table 5. It shows that, for state-of-the-art dToF performance, almost all sensors have adopted chip stacking and/or partial histogram. To illustrate the importance of 3D stacking as a key enabling technology for dToF, the histogram capacity per ToF pixel of sensors in Table 5 has been plotted in Fig. 28. It shows the growing development to push the boundaries of 3D stacking to increase the per-pixel histogram bin capacity. Sensors which cannot utilise this technology are limited to low bin count (8 or fewer) and resort to partial histogram approaches at the cost of a severe laser power penalty.
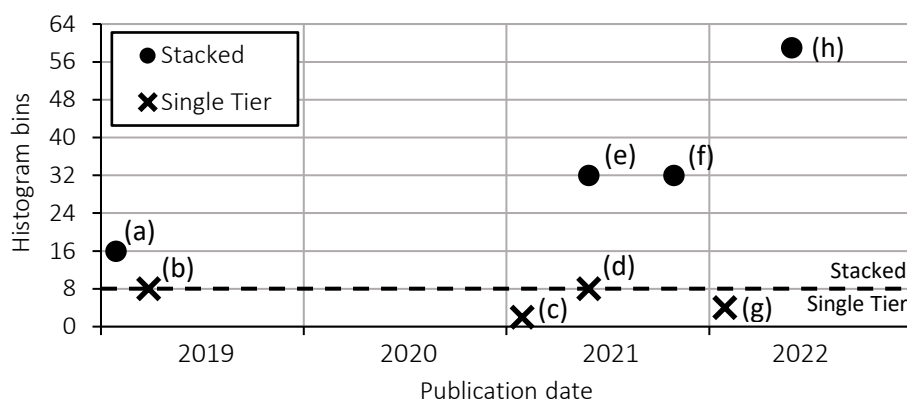
Fig. 28: Histogram bin capacity per pixel of published dToF sensor arrays.
(a) Henderson 2019 [108] (b) Zhang 2019 [52] (c) Kim 2021 [53] (d) Gyongy 2021 [146]
(e) Stoppa 2021 [54] (f) Zhang 2021 [143] (g) Park 2022 [145] (h) Taloud (2022) [144].

2.6 Conclusion

Direct time-of-flight (dToF) lidar is currently the most established lidar measurement principle to achieve the long-range performance required for self-driving vehicles. Of all solid-state implementations being developed for a low-cost scalable alternative to mechanical lidar, flash lidar is the most immediately feasible implementation which does not use moving parts. However, flash and even many scanning lidar solutions, require a 2D array of dToF pixels, each capable of detecting, processing, timing, and storing photon arrivals. This makes it difficult to achieve long-range without resorting to expensive 3D chip stacking or partial histogram techniques which incur a severe laser power penalty. This is particularly detrimental for flash solutions, which already consume a high amount of laser power to illuminate the entire field of view at once. Therefore, a 2D dToF sensor solution that makes optimal use of the returning laser power would be a significant breakthrough in enabling solid-state lidar that can deliver the depth-sensing performance required for self-driving vehicles.

3 MODELLING DToF SYSTEMS

3.1 Introduction

The requirement of every pixel to accommodate detection, processing and storage of photon arrivals limits the ranging capabilities of a dToF sensor. This chapter quantitively explores these limits for a conventional full histogram approach, explaining the growing trend of partial histogram approaches. The laser power penalty imposed by adopting a partial histogram approach is then quantified to establish the advantage of a guided dToF approach, as proposed in this thesis.

The chapter begins by introducing fundamental concepts and models related to dToF lidar. These are then applied to a realistic set of automotive lidar system parameters. The models and methodology developed here not only enable the analysis carried out in the chapter but also allow the developed guided lidar system to be benchmarked in subsequent chapters.

3.2 Models

3.2.1 Photon Budget

A photon budget is fundamental to modelling a lidar system. It quantifies the number of detected signal and background photons for a given system. Tontini et al. [149] provide a valuable resource for calculating the photon budget of SPAD-based dToF lidar systems. As such, this model is utilised here, with some minor modifications to simplify dependencies. An overview of the model is provided here for completeness, but for further details the reader is encouraged to refer to the original paper.

The aspects of a lidar system which contribute to the photon budget are illustrated in Fig. 29. For each point incident on the target, only a proportion of the reflected light radiates back to the lens (i.e., that contained within the angle α). Assuming a Lambertian target (i.e., one whose brightness appears equal from any angle) this proportion of returning light R is given by:

$$R = \sin^2 \alpha = \frac{d_{lens}^2}{4z^2 + d_{lens}^2} \tag{6}$$
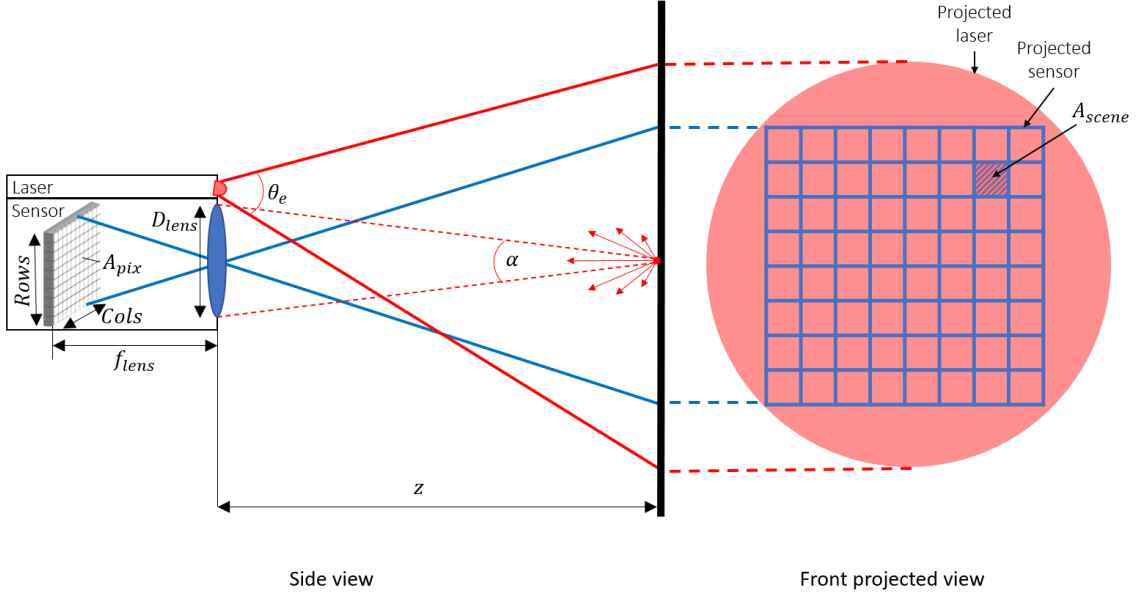
Fig. 29: Diagram of parameters which contribute to the photon budget of a lidar system.

The total laser power incident on each point on the target, known as radiance exitance, is given by the emitted laser power P_{tx} over the total area covered by the projection. For a rectangle spot optimised around the projected sensor area, as illustrated in Fig. 30(a), the radiance exitance is simply given as:

$$M_{target,source} = \frac{f_{lens}^2 \, P_{tx}}{z^2 \, Rows \, Cols \, A_{pix}} \qquad \text{(rectangle spot)} \qquad (7)$$

For an optimised round laser spot, the laser power spills over a larger area, with a radiant exitance of:

$$M_{target,source} = \frac{4 \, f_{lens}^2 \, P_{tx}}{z^2 \, \pi \, (Rows^2 + Cols^2) \, A_{pix}} \qquad \text{(round spot)} \qquad (8)$$

While a rectangle spot can be more challenging to implement than a circle spot, the advantage of less wasted laser power is clear. The minimum wasted laser power for an optimised round spot and a sensor with equal rows and columns is $\pi/2 \approx 1.6$ times greater than the equivalent optimised rectangle spot. The remainder of this derivation will assume an optimised rectangle spot.

A portion of light incident on a target is absorbed. This is represented by the reflectivity ρ_{target}. The sensor lens also presents a medium for absorbing power, characterised by the lens transmittance τ_{opt}. Finally, assuming a Lambertian surface (i.e. uniform radiance) a factor of $1/\pi$ gives the projected area over which the radiance is emitted [150]. The overall irradiance L_{target} is therefore:

$$L_{target,source} = \frac{\rho_{target} \, \tau_{opt} \, f_{lens}^2 \, P_{tx}}{z^2 \, Rows \, Cols \, A_{pix} \, \pi} \qquad (9)$$
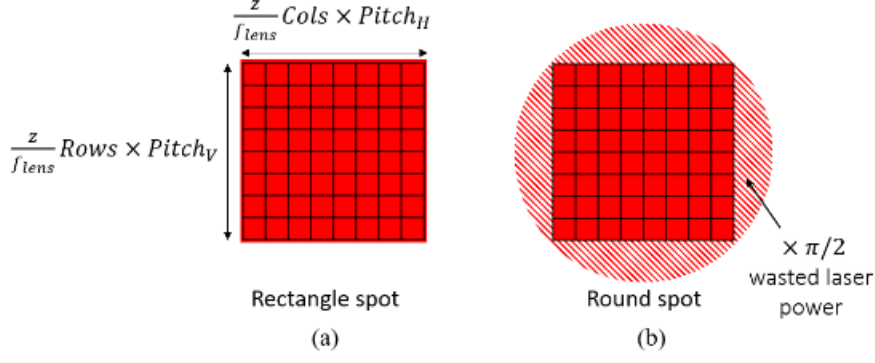
Fig. 30: (a) Projected dimensions of a square laser spot (b) wasted laser power from a circle laser spot.

The solar background radiant exitance on the target $M_{target,bg}$ is independent of the laser source and therefore of the system parameters too. It can be obtained at maximum brightness (100 klux) by integrating the solar irradiance spectrum [85] over the wavelengths of interest. The observed solar background irradiance for any pixel is therefore given by:

$$L_{target,bg} = \frac{M_{target,bg} \, \rho_{target} \, \tau_{opt}}{\pi} \tag{10}$$

The total power (laser or background) observed by any pixel is a product of the observed irradiance across the target and the overlapping projected pixel area given by:

$$A_{scene} = A_{pix} \left(\frac{z}{f_{lens}}\right)^2 \tag{11}$$

The observed laser $P_{pix,source}$ and background $P_{pix,bg}$ power for each pixel is therefore expressed as:

$$P_{pix,source} = R \, L_{target,source} \, A_{scene}$$

$$= \frac{d_{lens}^2}{4z^2 + d_{lens}^2} \times \frac{\rho_{target} \, \tau_{opt} \, f_{lens}^2 \, P_{tx}}{z^2 \, Rows \, Cols \, A_{pix} \, \pi} \times A_{pix} \left(\frac{z}{f_{lens}}\right)^2$$

$$= \frac{\rho_{target} \, \tau_{opt} \, P_{tx} \, d_{lens}^2}{4z^2 \, \pi \, Rows \, Cols} \qquad \text{(assuming } 4z^2 \gg d_{lens}^2\text{)} \tag{12}$$

$$P_{pix,bg} = \frac{d_{lens}^2}{4z^2 + d_{lens}^2} \times \frac{\rho_{target} \, \tau_{opt} \, M_{target,bg}}{\pi} \times A_{pix} \left(\frac{z}{f_{lens}}\right)^2$$

$$= \frac{\rho_{target} \, \tau_{opt} \, M_{target,bg} \, A_{pix} \, d_{lens}^2}{4 \, \pi \, f_{lens}^2} \qquad \text{(assuming } 4z^2 \gg d_{lens}^2\text{)} \tag{13}$$

Two key observations can be drawn from Equation (12) and Equation (13). First, assuming the projected laser and sensor FOV are always matched, the received laser power is independent of pixel area. As an example, if pixel area is increased the total projected FOV increases and hence the same laser power must be projected over a wider area. The increase projected pixel area is therefore directly countered by a reduced laser power per unit area. Second, the received background power is independent of the target distance because the projected pixel area, and hence total solar power within this area increases with the square of the target. However, this is directly countered by the returning power from the target decreasing with the square of the target distance.

Using Planck's relation, the number of photons γ present in the received power of wavelength λ and energy E over an interval t is given by Equation (14) where h is Planck's constant and c is the speed of light:

$$\gamma = \frac{\lambda}{h\,c} \int P\,dt \tag{14}$$

Since the laser power P_{tx} in a dToF system is not continuous, it is more useful to relate the returning laser photons as a function of the total emitted energy E_{tx} of each laser pulse. Accounting for SPAD photon detection efficiency (PDE) which relates the ratio of received to detected photons, the total number of detected photons per laser pulse Ph_s and background photon arrival rate Ph_b/t is given by:

$$\gamma_s = \frac{PDE\,\lambda\,\rho_{target}\,\tau_{opt}\,E_{tx}\,d_{lens}^2}{h\,c\,4z^2\,\pi\,Rows\,Cols} \qquad \text{(Laser photons per pulse)} \tag{15}$$

$$\frac{\gamma_b}{t} = \frac{PDE\,\lambda\,\rho_{target}\,\tau_{opt}\,M_{target,bg}\,A_{pix}\,d_{lens}^2}{h\,c\,4\,\pi\,f_{lens}^2} \qquad \text{(Background photon rate)} \tag{16}$$

3.2.2 Peak Detection

In a dToF system, the returning laser signal is identified by a peak in the resulting histogram. However, under high background/low signal conditions, the highest bin value may correspond to a non-signal bin if sufficient laser cycles have not been integrated, as illustrated in Fig. 31. After completing a frame of N laser cycles, the probability that the peak bin is correctly identified as the signal bin is known as the probability of detection P_D and depends on (i) the expected signal photon count per laser pulse; (ii) the ambient background photon arrival rate; (iii) the number of laser integration cycles. The model provided

here aims to determine the minimum number of laser cycles required to meet a set probability of detection for any given dToF system.
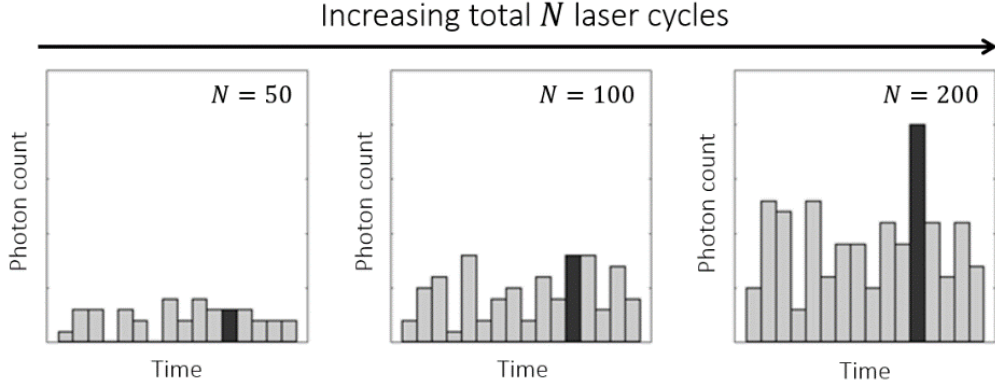
Fig. 31: The histogram bin containing the laser signal becomes increasingly more distinguished as more laser cycles are integrated. Adapted from [1].

Using a computationally intensive Monte Carlo method is a common method for characterising P_D for a given dToF system. In this approach, a Poisson-distributed background photon count is generated over M histogram bins based on the expected count per bin after N laser cycles. A Poisson-distributed signal photon count is then added to a single pre-determined bin based on the average expected signal photon after N laser cycles. By observing the ratio of runs in which the signal bin contains the most photon counts, the probability of detection can be characterized.

Here, a derivation of P_D is introduced to provide an alternative deterministic method over the common Monte Carlo approach method. This model is based on the following simplifying assumptions:

1. The laser signal falls entirely within a single time bin.
2. No dead time associated with photon detection, thereby negating any pileup effects.
3. All photon arrivals per cycle are counted.

Let μ_B represent the average background photon arrivals per bin after N cycles, and μ_S the average signal photon arrivals after N cycles. For a system with M bins, P_D can be described as: the probability that the signal bin takes on a given photon count value x "and" all $M-1$ non-signal bins take on a value lower than x, for all possible photon count values of x:

$$P_D = \sum_{x=0}^{\infty} f_S(x) \times F_B(x)^{M-1} \tag{17}$$

Here, $f_S(x)$ is the probability density function (PDF) of the signal bin and $F_B(x)$ is the cumulative density (CDF) of any non-signal bin. These are formally expressed using a Poisson distribution. However, since a typical dTof system involves integrating thousands of laser cycles, it is practical to approximate these to a Gaussian form to further simplify computation:

$$f_S(x) \approx \frac{1}{(\mu_S + \mu_B)\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x - (\mu_S + \mu_B)}{(\mu_S + \mu_B)}\right)^2} \qquad \text{(Gaussian PDF with mean and standard deviation of } (\mu_S + \mu_B)) \qquad (18)$$

$$F_B(x) \approx \Phi\left(\frac{x - \mu_B}{\mu_B}\right) \qquad \text{(Gaussian CDF with mean and standard deviation of } \mu_B) \qquad (19)$$

These can be solved using standard mathematic computing software (e.g. MATLAB) after applying appropriate limits. Recognizing that the photon count value x cannot take on negative values binds the lower limit to zero. The upper limit can be set by recognizing that values of x multiple times greater than $\mu_S + \mu_B$ result in negligible accuracy improvement.

To examine the accuracy of the derived expression for probability of detection P_D, it is tested on a system composed of the following parameters:

- average signal photon count per laser cycle is $\gamma_s = 0.1$ counts
- average background photons in each M bin per laser cycle is $\gamma_b = 1/M$ counts
- total of 2, 4, or 8 histogram bins

The probability of detection for the test system is evaluated for different values of total laser cycles N. A Monte Carlo approach is used to evaluate the probability of detection of the system at 20 different values of N using 5000 runs at each point. A deterministic approach by solving Equation (17) at each point is also used to evaluate the system. The results of both approaches are plotted in Fig. 32.
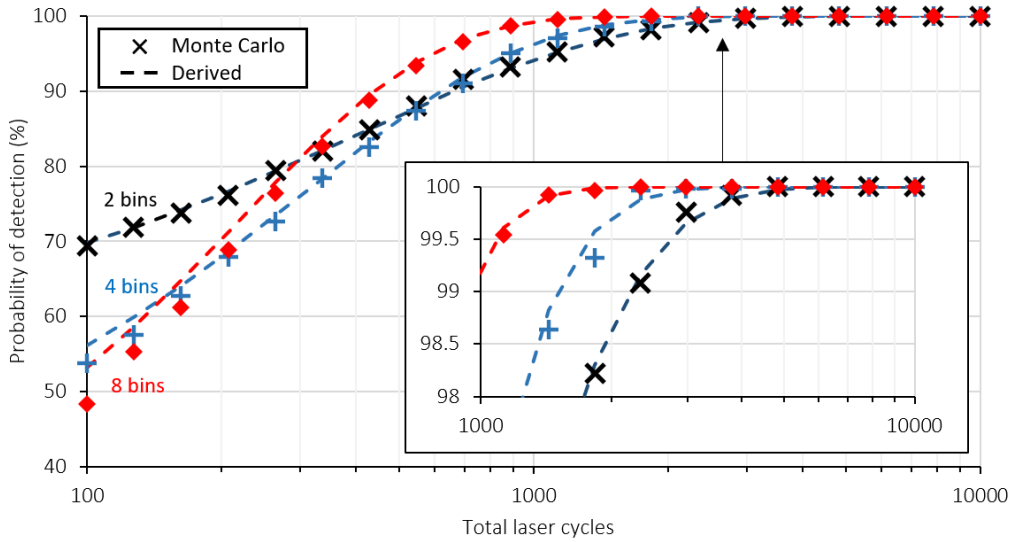
Fig. 32: Evaluating the probability of detection on an example system with is $\gamma_s = 0.1$ and is $\gamma_b = 1/M$ using a Monte Carlo approach. This is shown to match closely with the derived deterministic approach Equation (17).

The first significant observation is how closely the deterministic approach matches the computationally intensive Monte Carlo approach. The matching deteriorates for low values of total laser cycles as the assumption of Gaussian distributed photon arrivals deteriorates, however thousands of laser cycles are typically conducted in dToF.

The second observation worth noting is the inflexion point below-which a lower bin value exhibits a higher probability of detection. Under these conditions, too few laser cycles have been integrated to provide a sufficient laser photon count above the evenly distributed ambient background level. As a result, every bin has an equal chance of being the peak bin after N laser cycles and so the probability of detection converges to $1/M$. Note this behaviour is also successfully captured by the deterministic approach (although not fully captured within the scale bounds of Fig. 32).

Arguably more useful, Equation (17) can be used to determine the minimum number of laser cycles required for a dToF system to achieve a given P_D. This can be solved as an optimization problem, again using simple mathematic computing software, by searching for the value N which minimises the difference between Equation (17) and the desired value of P_D. An example of one such function is MATLAB's "fminsearch" function [151]. This provides a useful tool in the design of lidar systems as will be seen in subsequent sections.

61

3.2.3 Peak Precision

While the previous section focused on the minimum number of laser cycles required to detect the presence of a signal, multiple more cycles must be integrated to achieve a degree of precision. This is thoroughly examined by Korner [152] where the following 4 parameters are identified as determining the achieved precision (δ) of a dToF lidar system:

- histogram bin interval a
- signal standard deviation (primarily laser pulse width) σ
- average background photons per bin after N cycles μ_B
- average signal photons after N cycles μ_S

The two key models described are (1) Thompson and (2) Cramér Rao bound (CRB). The Thompson model provides an intuitive relation between the system parameters affecting precision and is given as:

$$\delta = \sqrt{\frac{\sigma^2 + a^2/12}{\mu_S} + \frac{4\sqrt{\pi}\,\sigma^3\,\mu_B}{a\,\mu_S{}^2}} \tag{20}$$

The first term of the Thompson model attributes the effect of quantisation due to the finite dToF histogram bin width, while the second term attributes the effect of background noise. Under high signal-to-noise ratio (SNR), the first term dominates and reducing the histogram bin or laser width produces the most significant improvement in precision (lower δ is better). Under low SNR conditions, the returning laser power becomes an increasingly more significant factor and has a greater effect on precision than an opposite change in background power by the same amount.

While the Thompson model provides valuable insight into tuning system precision, its accuracy degrades when modelling systems with low SNR. The error in matching Monte Carlo results under these conditions can be as high as 20%. For such cases, the alternative CRB model is shown to provide better matching of system precision. While the CRB model does not take the form of an identity equation like the Thompson model, Koerner provides a Python toolbox to execute the model [153].

The deviation of the Thompson model is illustrated in Fig. 33 for a system with the following parameters:

- histogram bin interval $a = 1$ ns
- laser pulse $\sigma = 1.5$ ns
- average background photons per bin $\mu_B = 100$
- average signal photons $\mu_S = 100$ to 100k

As with the probability of detection model (Equation (17)), the CRB model can also be solved using optimization to estimate the minimum number of laser cycles for a given precision requirement. This is included in Koerner's Python toolbox under the function name "exp_search".
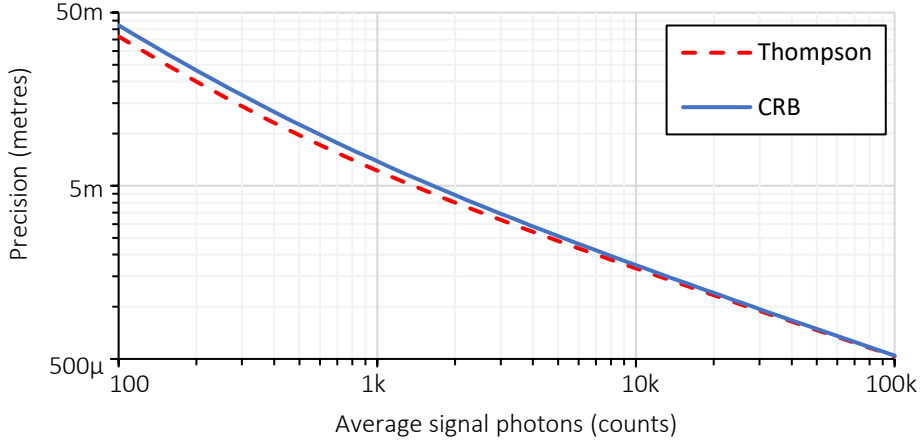
Fig. 33: Deviation of Thompson model in approximating dToF precision compared to the more accurate CRB as system SNR decreases. Example model uses $a = 1$, $\sigma = 1.5$ ns and $\mu_B = 100$.

One final key takeaway from the work by Koerner on dToF precision is a rule-of-thumb for the optimal histogram bin interval relative to the laser pulse width. Using excessively wide histogram bin intervals leads to severe precision degradation irrespective of the SNR, while excessively narrow intervals result in an overengineered solution. A practical ratio of histogram bin interval to pulse width is therefore suggested and reproduced in Equation (21). For a Gaussian-shaped laser pulse, this is equivalent to Equation (22) as a function of the laser full-width half maximum ($FWHM$).

$$\frac{a}{\sigma} \approx 1.5 \tag{21}$$

$$\frac{a}{FWHM} \approx 0.64 \tag{22}$$

3.3 Analysis

3.3.1 Design Specification

To evaluate both full and partial histogram approaches, a lidar design specification typical of long-range automotive lidar is outlined in Table 6.

Component	Parameter	Value
Target	Maximum distance	200 m
	Minimum reflectivity	10%
Ambience	Maximum intensity	100 klux
	Irradiance reference	ASTM G173
System	Probability of detection	99.7%
	Precision	0.1 m
	Frame rate	30 fps
	Field of view	26° × 8°
Laser	Pulse energy	50 µJ
	Pulse shape	Gaussian
	Pulse width (FWHM)	8 ns
	Wavelength	905 nm
	Spot shape	Rectangle
Sensor (achievable SPAD characteristics based on [55])	Photon detection efficiency	22%
	Pixel pitch	30 µm
	SPADs per pixel	3 × 3
	Pixel resolution	200 × 63
Lens	Diameter	6 mm
	F#	2.2
	Filter bandwidth	20 nm
	Transmission	100%

Table 6: A typical set of system requirements for long-range lidar.

3.3.2 Photon Budget

Applying the system parameters in Table 6 to Equation (**15**), the average signal photons per laser pulse (γ_s) as a function of distance (z) comes to $\gamma_s = 1140z^{-2}$. This is equivalent to 0.03 photon counts per laser pulse at the maximum distance of 200 m.

To determine the solar background photon budget, the ASTM solar irradiance spectrum [85] must first be integrated within 905 nm $\pm$ 10 nm, resulting in a calculated solar background radiance across the target of $M_{target,bg} = 13.8$ W/sr¹/m² at maximum ambient brightness (100 klux). Combining this with Equation (16), the average background photon arrival rate comes to 20 Mcounts/s.

The resulting photon budget of the system as a function of distance is plotted in Fig. 34. While the optimal bin interval for the laser in this design specification is around 5 ns following Equation (22), a range of values are plotted to illustrate how this influences the distance beyond which $\mu_S/\mu_B < 1$.
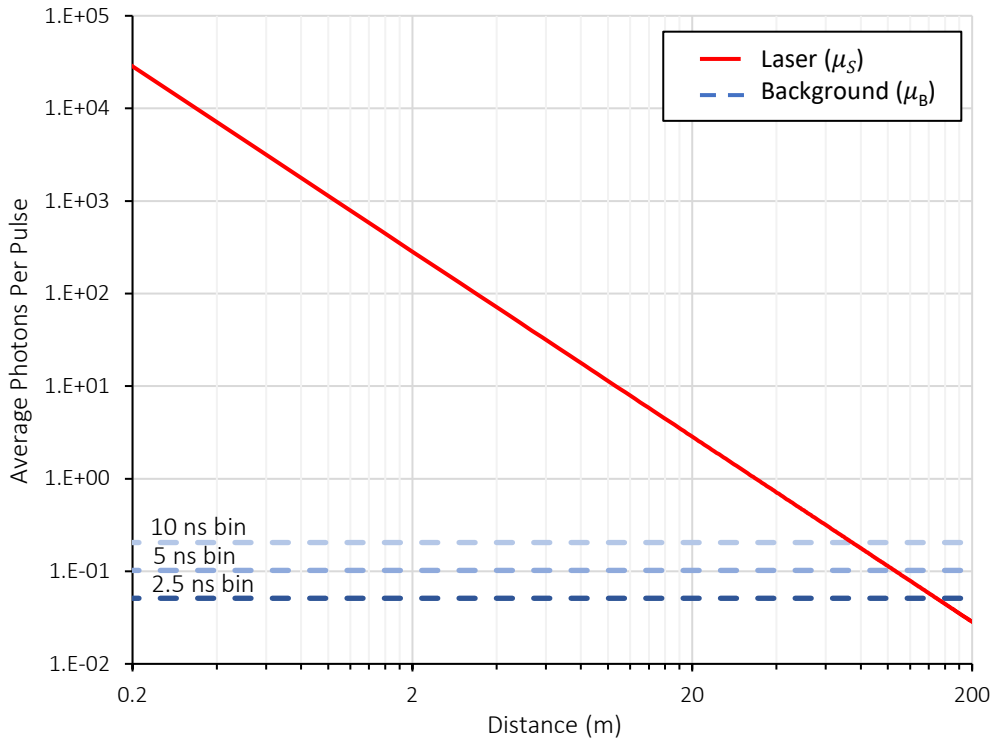
Fig. 34: Modelled photon budget for the lidar systems described in Table 6. Average background photon count per bin plotted for various bin interval values.

3.3.3 Laser Cycle Requirement

Having established the expected signal and background photon return rates, the CRB model can now be used to find the minimum number of laser cycles required to achieve the target system precision of 0.1 m. This is plotted as a function of distance in Fig. 35 assuming a bin interval of 5 ns (0.8 m) optimised for precision following Equation (22). The roundtrip time of the laser, in combination with the frame rate specification (30 fps), also place a limit on the maximum number of laser cycles that can fit in a frame.

For this system, the minimum required to meet the target precision, and the maximum required to meet the target frame rate, converge to around 25,000 laser cycles at 200 m. Any requirement to further increase the total laser cycles per frame will breach the frame rate specification for this design.
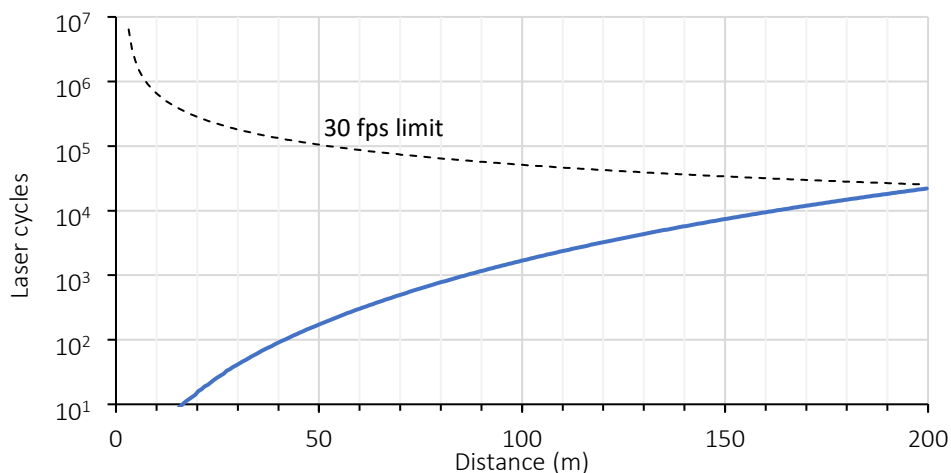
Fig. 35: Applying the calculated photon budget in Fig. 34 to find the minimum number of laser cycles for the lidar systems in Table 6 to achieve 0.1 m precision assuming 5 ns (0.8 m) bin interval size as determined by Equation (21).

3.3.4 Histogram Size Requirements

The histogram stored on each pixel needs to accommodate both (a) the maximum expected bin count and (b) the total number of required bins. For the former, the chosen histogram bin interval sets the average background photon count per bin. To have capacity for storing signal counts, the histogram in each dToF pixel (containing 3×3 SPADs in the example case) needs to at least accommodate the baseline expected background count.

The scope of bin count values for this system is shown in Fig. 36, along with the full design space of possibilities that a given dToF system might exhibit. For long-distance outdoor dToF lidar, 12-bit bins are considered a bare minimum, which is consistent with state-of-the-art implementations (see Section 2.5).
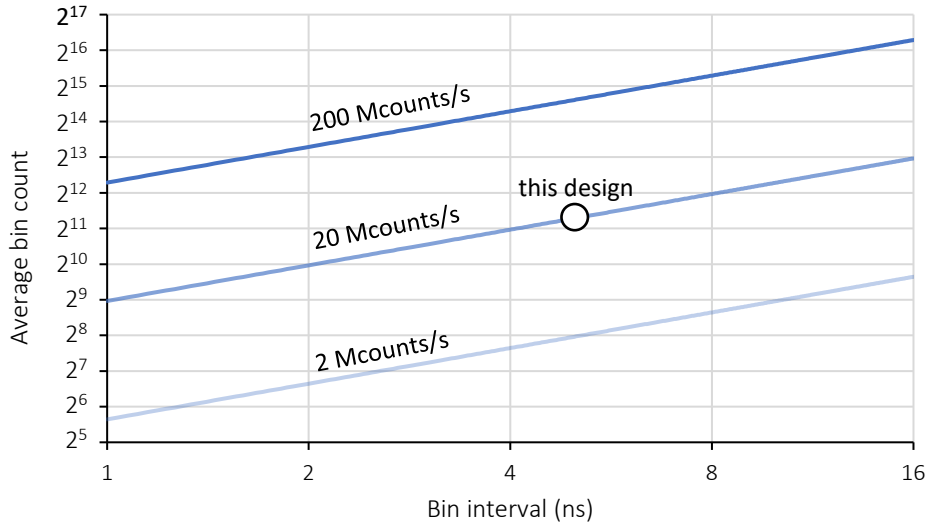
Fig. 36: Minimum bin count values required for a dToF sensor to achieve the system specifications in Table 6. Expanded to show a broad range of background counts and bin interval sizes to illustrate the range of possible values for a generic dToF system.

If a full histogram approach is adopted, the total number of histogram bins required in each pixel is simply a function of the maximum sensing distance and the chosen bin interval. This is plotted in Fig. 37, showing that 256 bins are required using the optimal bin interval of 5 ns (0.8 m). The full design space of possible bin interval values and maximum distance is presented for a more complete picture.
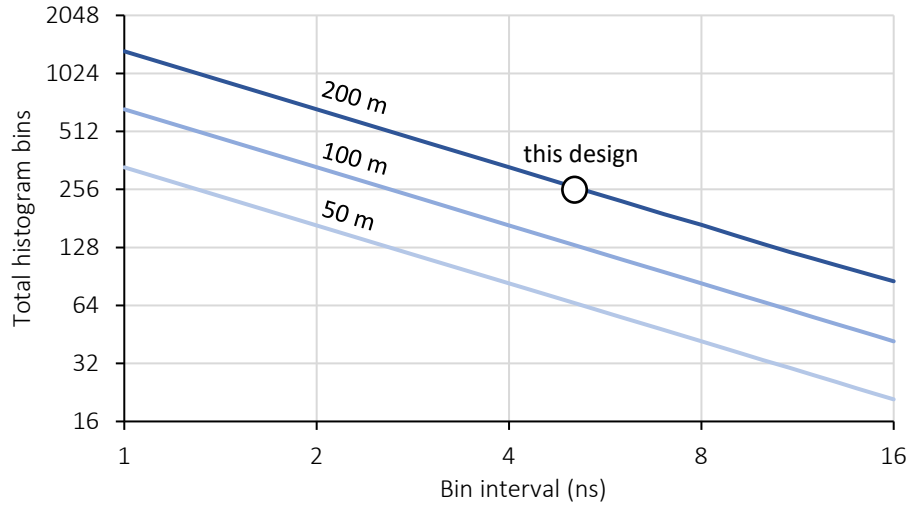
Fig. 37: Total histogram bins required for a dToF sensor to achieve the system specifications in Table 6 using a full histogram approach. Expanded to a broad range of maximum sensing distances and bin interval sizes to show the potential design space.

Note that, while increasing the histogram bin interval of the system increases the minimum required bin count value, it decreases the total number of required bins. In terms of overall storage, the trend remains that increasing the histogram bin interval reduces the total number of required bits per pixel because the minimum count value only increases the total bits required for each bin as a function of $\log_2$. This is exemplified in Fig. 38 which combines both the minimum capacity per bin (in bits) with the total number of bins for a full histogram approach. For this design, where 256 bins with a minimum of 12 bits per bin are required, each dToF pixel would need to accommodate over 3 kbits of storage.
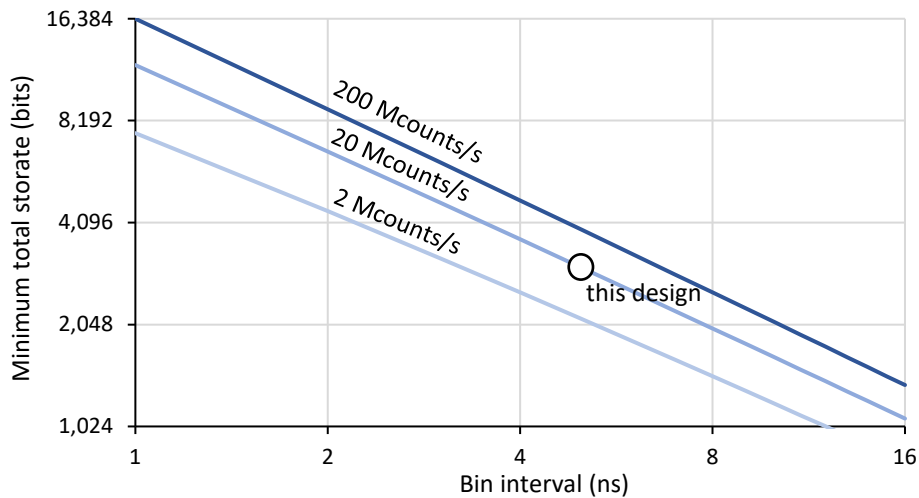
Fig. 38: Minimum storage capacity required for a dToF pixel using a full histogram approach ranging at 200 m. Plotted for various background photon rates.

3.3.5 Pixel Capacity

Having established the full histogram storage required (3 kbits) to implement the design specification in Table 6, an examination of pixel capacity for storing this data follows. Static random-access memory (SRAM) is increasingly adopted for storing histogram data on dToF sensors, as reviewed in Section 2.5. It provides a compact solution while being easily integrated in standard CMOS processes.
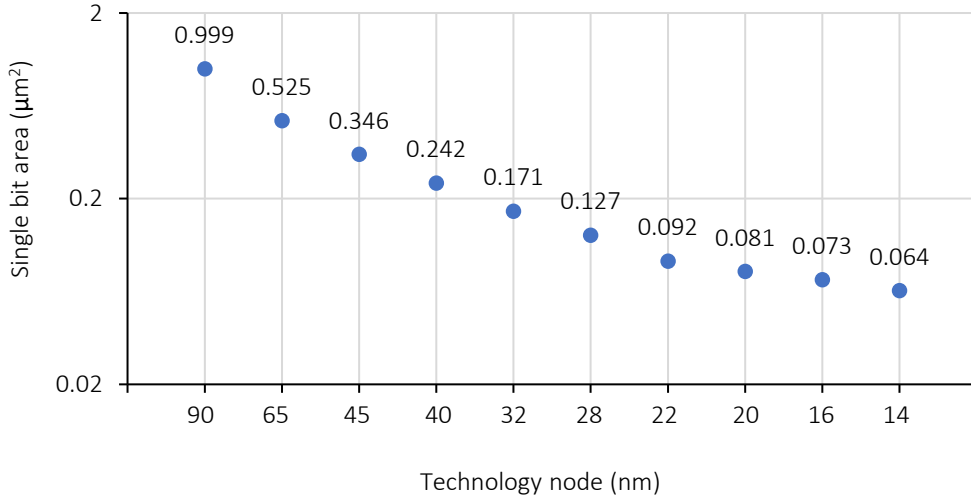
Fig. 39: Typical area of a single-bit SRAM cell at each technology node. Reproduced from [154].

Fig. 39 shows the typical on-chip area occupied by a single bit of SRAM memory in different technology nodes. Following this, the total required full histogram area for the studied design can be compared relative to the dToF macropixel area, as illustrated in Fig. 40. To implement a full histogram, this design would not only require an expensive 3D stacked technology, but also adopt at least a 40 nm process for the circuit-level tier. In reality, an even smaller (more expensive) technology node would need to be used to accommodate circuitry processing and timing of photon events at the very least.

A more general overview of the constraints posed by limited in-pixel storage area is possible by decoupling bin capacity from the number of histogram bins. Assigning a realistic ¼ of the available pixel area (225 μm²) all achievable combinations of histogram dimensions can be produced (Fig. 41).

Both Fig. 40 and Fig. 41 highlight the impracticalities of using a full histogram approach for outdoor and long-distance dToF. Even if such area constraints are overcome through exotic and expensive semiconductor processes, transferring many kbits of data from potentially millions of pixels between every frame (Tbits/s) poses yet another challenge which is yet to be solved.
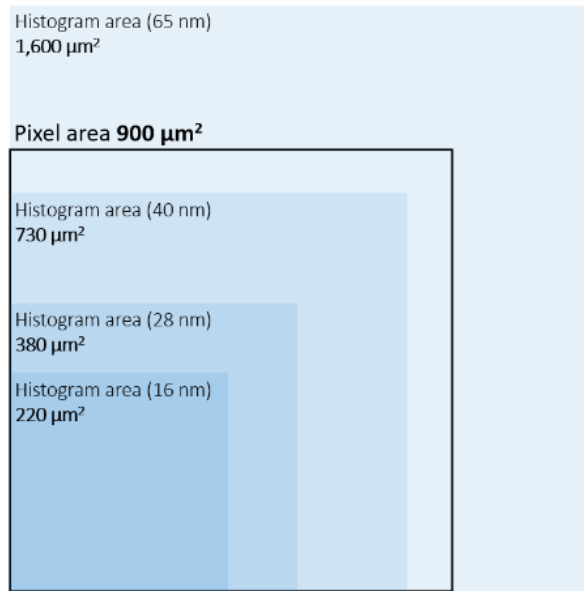
Fig. 40: Histogram area required for a dToF macropixel at various technology nodes to achieve the design specification in Table 6 using a full histogram approach.

While implementing a full histogram approach with this design specification would require using 16 nm technology at the very least, Fig. 41 indicates that a reduced set of 16 histogram bins would allow for a significantly more economical technology (90 nm in this example). This necessity to reduce the dToF histogram and make long-range solid-state lidar practical has resulted in the development of multiple novel "partial histogram" approaches (see Section 2.5.2). However, these impose a laser power penalty on the lidar, which will now be quantified.
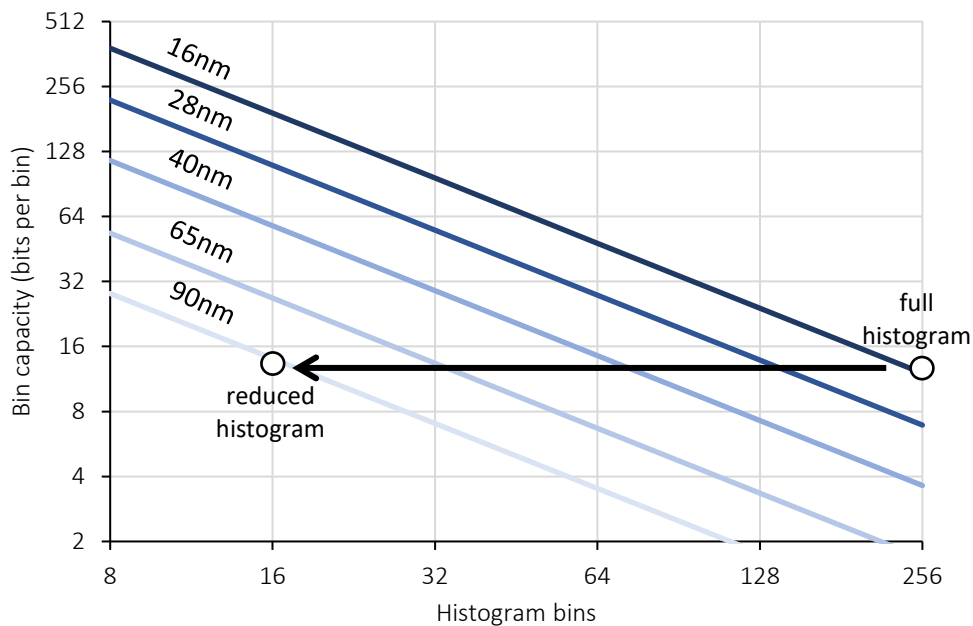
Fig. 41: Design space of possible histogram bit/bin configurations at each technology node that can fit into ¼ of the allocated dToF macropixel area (225 µm²).

70

3.4 Laser Power Efficiency of Partial Histogram Approaches

A more practical 16-bin zooming or sliding partial histogram solution with equivalent precision to the 256-bin full histogram approach can be configured as presented in Table 7.

Parameter	Zooming	Sliding
Bins	16	16
Steps	2	16
Bin interval	12.5 m (step 1 "detection") 0.8 m (step 2 "precision")	0.8 m

Table 7: Partial histogram configurations equivalent to a 256-bin full histogram.

3.4.1 Zooming

At each step of an M-bin zooming approach, the bin interval is $M\times$ wider than in the next step. However, the number of laser cycles required in all steps, except the final step, only needs be enough to detect the peak signal bin, rather than to precisely interpolate the return time. The zooming configuration in this design is composed of one detection step (12.5 m bins) and one precision step (0.8 m bins) to result in equivalent bin intervals as the prior 256-bin full histogram design.

The minimum number of laser cycles required in the detection step is a function of the maximum sensing distance. This can be evaluated using the probability of detection (P_D) Equation (17) for a given photon budget. Below 12.5 m the detection step can be skipped altogether. Up to 25 m the signal location must be identified between one of two bins assuming a photon budget at the maximum distance of 25 m. Up to 37.5 m the signal location must be identified between one of three bins assuming a photon budget at the maximum distance of 37.5 m. This is evaluated right up to the maximum sensing distance of 200 m and shown in Fig. 42. Since the minimum number of laser cycles required for the targeted precision using a bin interval of 0.8 m has already been evaluated for the full histogram case (Fig. 35), the sum of these two plots gives the total minimum required laser cycles to complete all required zooming steps.
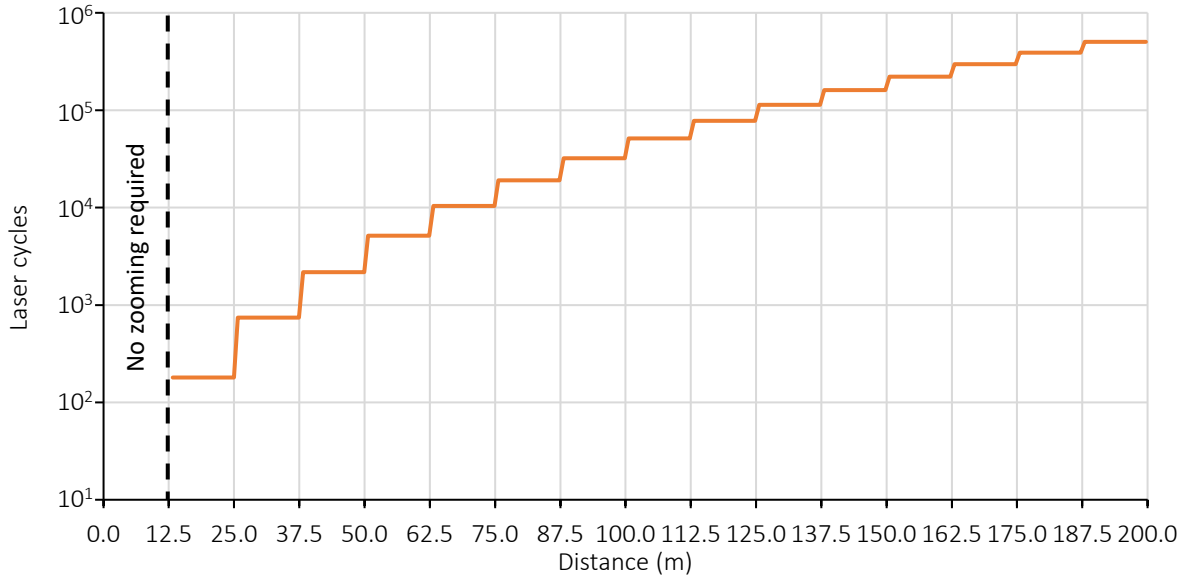
Fig. 42: Minimum laser cycles required by a 16-bin zooming approach during the "detection" step to achieve the design specification in Table 6

3.4.2 Sliding

Sliding a reduced set of 16×0.8 m histogram bins across 16 window steps gives a final bin interval equivalent to the 256-bin full histogram design. For any distance up to the end of the first slide window, the minimum number of laser cycles required is the number required to achieve the target precision at the maximum distance in that window (12.5 m). For any distance up to the end of the second window, the minimum number of laser cycles is the value needed to achieve the targeted precision at the maximum distance in the second window (25 m) and the first window. In this way, the total laser cycles required up to any slide step distance can be evaluated, as illustrated in Fig. 43.
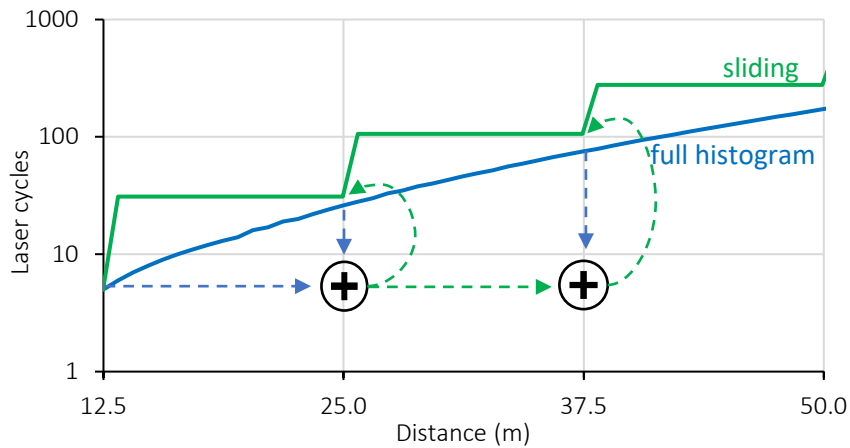
Fig. 43: Process for calculating the minimum number of laser cycles at each slide step.

72
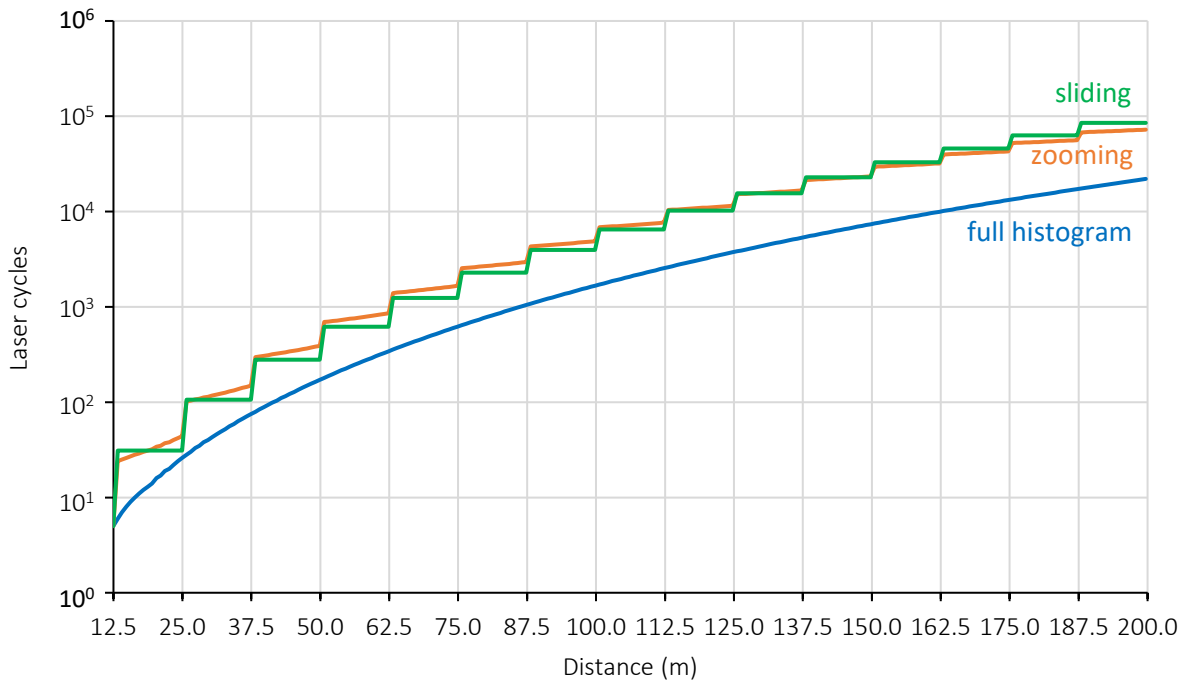
Fig. 44: Minimum laser cycles required for a full histogram (256-bin) approach and the equivalent sliding and zooming partial histogram (16-bin) approaches.
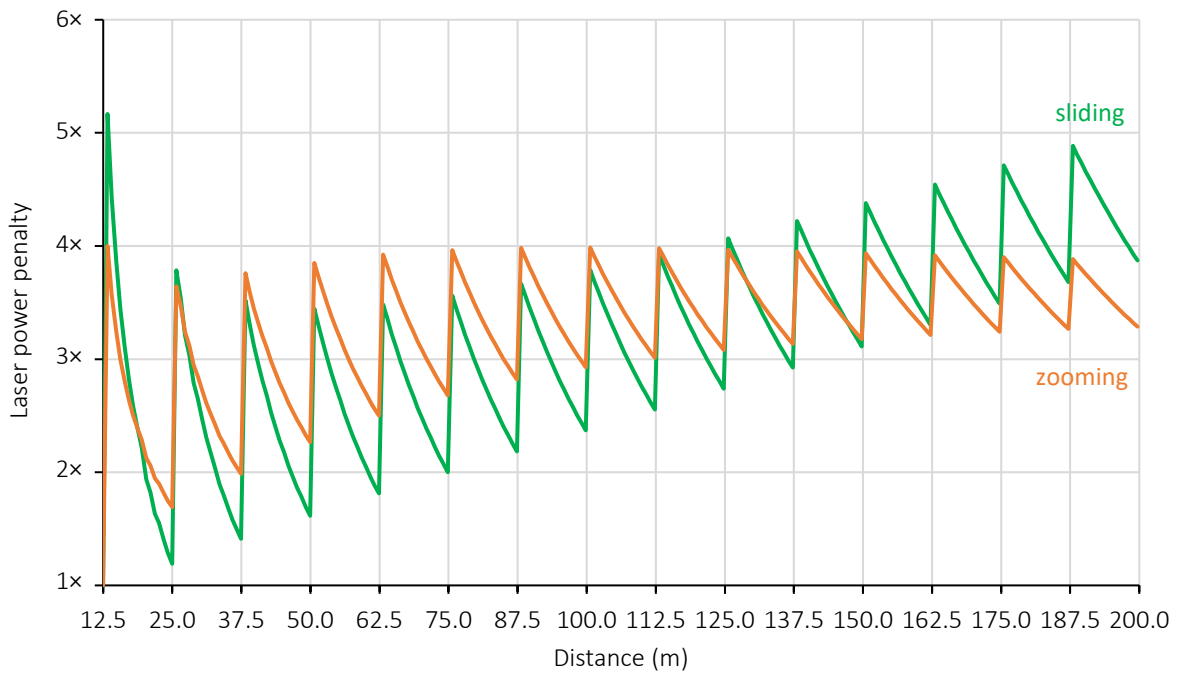
Fig. 45: Laser power penalty of equivalent 16-bin zooming and sliding configurations for the design to achieve the design specification in Table 6 compared to a 256-bin full histogram approach.

3.4.3 Comparison

The minimum number of required laser cycles for both partial histogram approaches, along with the full histogram approach are plotted in Fig. 44. The increase in total laser cycles required through the equivalent zooming or sliding approach infers a laser power penalty, plotted in Fig. 45. This analysis indicates that achieving the target specification in Table 6 using a partial histogram approach imposes a laser power penalty approach 4× compared to a conventional full histogram approach.

While the result in Fig. 45 may suggest that zooming is preferred over sliding at longer distances, it is worth noting that the configuration presented here would require an additional zooming step above 200 m, greatly increasing the laser power penalty.

3.5 Conclusion

To gain an intuition into the expected photon arrival rates for a dToF sensor used in automotive lidar, a realistic design specification was evaluated based on the achievable performance of a state-of-the-art sensor. For a solid-state dToF sensor, the expected photon arrival rates indicated that the total amount of in-pixel storage capacity required vastly exceeds the available area. Unless an expensive stacked sensor process on a 16 nm technology node (or smaller) is used, the histogram size must be reduced.

While zooming and sliding partial histogram approaches allow reducing the histogram to a more realisable size, the equivalent configurations to the full histogram solution studies here are shown to impose a power penalty approaching 4×. The guided lidar approach proposed in this thesis, if realised, would therefore be of significant value, allowing for a practical dToF sensor implementation with no laser power penalty.

The models outlined and developed in this chapter provide a valuable tool for benchmarking the final realised guided lidar setup against partial histogram approaches.

4 GUIDED DTOF SENSOR DESIGN

4.1 Introduction

The key feature of a guided dToF lidar sensor that distinguishes it from regular dToF sensors is the ability for each pixel to independently observe a different exposure time window. This chapter details the integration of guided dToF functionality into an existing dToF CMOS QuantIC4×4 using a programmable down counter (P-counter). The various P-counter design considerations are explored, along with the final implementation optimised for minimal pixel area impact. Finally, a contingency sensor option is presented in the event that the revised QuantIC4×4 sensor is unavailable, along with the repercussions of conceding to this contingency option.

4.2 Primary Sensor "QuantIC4×4"

The CMOS lidar sensor into which guided lidar functionality has been specifically integrated for this project is a revised version of the chip published in [49] known as QuantIC4×4. It was fabricated in STMicroelectronics 3D-stacked backside illuminated (BSI) SPAD process, allowing for photon processing, timing and storage on the bottom tier without impacting the SPAD detector fill-factor on the top tier.
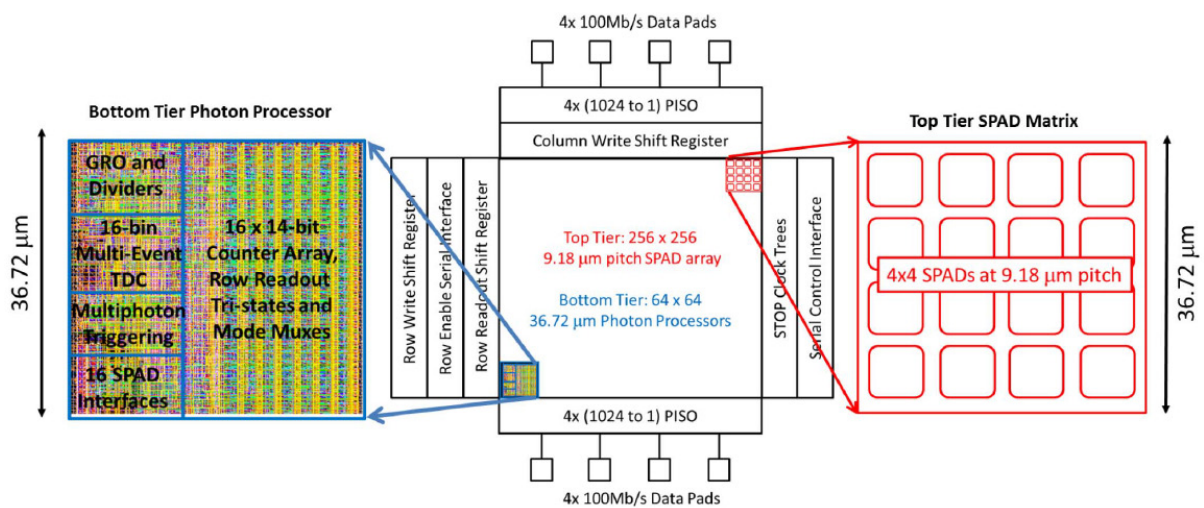
Fig. 46: Overview of the original QuantIC4×4 sensor into which guided dToF functionality has been integrated. Reproduced from [49] with author's permission.

The sensor is to be revised to use an upgraded state-of-the-art fabrication process to achieve a SPAD photon detection efficiency (PDE) of 18.5% at 940 nm [123]. This presents the opportunity to integrate guided dToF functionality into the sensor during the design revision.

An overview of the original pre-revised QuantIC4×4 sensor is shown in Fig. 46. The sensor comprises of 256×256 SPAD detectors on the top layer, grouped into macropixels of 4×4. Each macropixel has a dedicated processing block on the bottom tier containing a multievent time-to-digital converter (METDC) and storage 16×14-bit histogram bins.

A functional overview of the sensor's key macropixel sub-blocks when configured for dToF is shown in Fig. 47. Upon a successful photon detection, the raw SPAD pulses *SPAD_RAW<15:0>* are retimed to a clock CLK which is globally distributed to every macropixel on the sensor. The retimed SPAD events *SPAD_TIMED<15:0>* are then passed onto a correlation block. This consolidates the 16 SPAD detector signals into a single SPAD_EVENT signal, dependent on a programmable *TYPE* (less than, equal or greater than) and *THRESHOLD* (integer) condition. In the case of Fig. 47, the example condition is set to trigger whenever more than zero (i.e., one or more) SPADs fire within the same clock period. Finally, an METDC updates the in-pixel counters based on the detected SPAD events in the last 16 CLK periods preceding the rising edge of a globally distributed STOP signal. There are a total of 16 counters (i.e., histogram bins) per macropixel, hence if the STOP signal has a period longer than 16 CLK periods, then only SPAD events within the final 16 periods before each rising STOP edge are counted. Table 8 provides a convenient summary of these signals for the reader's reference.

Signal	Type	Description
CLK	Input	A clock that defines the histogram bin width. Raw SPAD events are retimed to this signal and allocated into their corresponding bin.
STOP	Input	A clock that defines the start of each laser pulse period. The 16 prior periods of CLK define the full histogram time window.
SPAD_RAW	Internal	Raw SPAD events directly from the front-end output.
SPAD_TIMED	Internal	RAW_SPAD retimed to the clock edge of CLK
SPAD_EVENT	Internal	Indicates if the total SPAD events within a clock period has met the pre-programmed threshold.

Table 8: Summary of signals which relate to timing and histogramming SPAD events on the original QuantIC4×4 sensor chip.
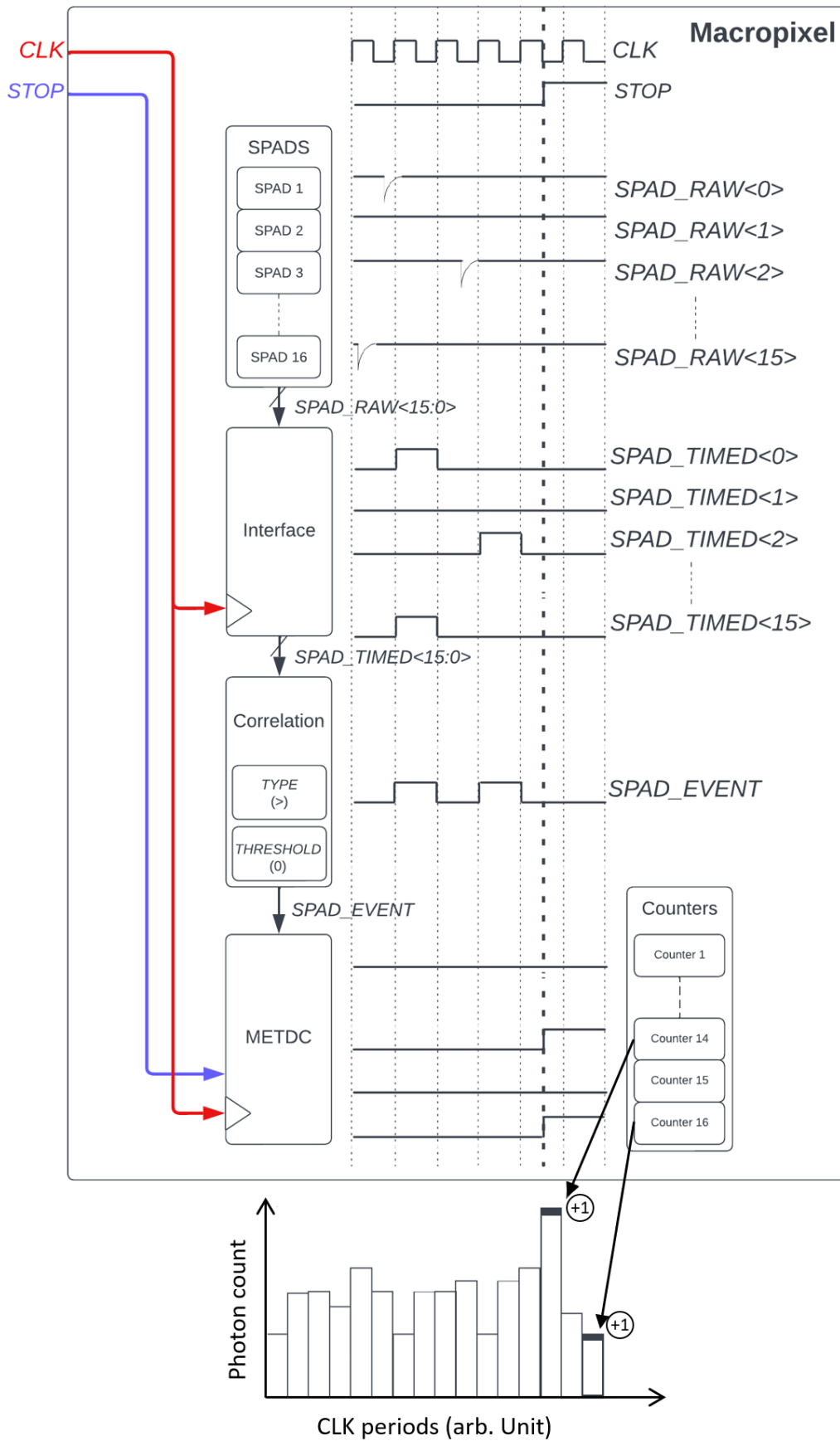
Fig. 47: Functional overview of a macropixel on the original QuantIC4×4 sensor chip when configured for conventional (i.e., non-guidable) dToF.

4.3 Integrating DToF Guiding

To implement guided dToF functionality, each macropixel is given independent control of its own local STOP signal (LOCAL_STOP) triggered by a newly integrated programmable counter (P-counter). For every rising edge of the global STOP signal (i.e. the start of each laser pulse), each macropixel P-counter resets to an independent pre-programmed value and for every rising edge of the global CLK signal, all P-counters decrease by one count. Any P-counter that rolls over past zero raises the local STOP signal for that macropixel and triggers its METDC to update its histogram counters based on photon events within the previous 16 CLK periods. This new functionality is illustrated in Fig. 48, where "Macropixel 1" counts for 52 CLK periods after the global STOP signal before raising its local STOP signal while "Macropixel 2", programmed to count down from 49, triggers its METDC 3 CLK periods earlier.
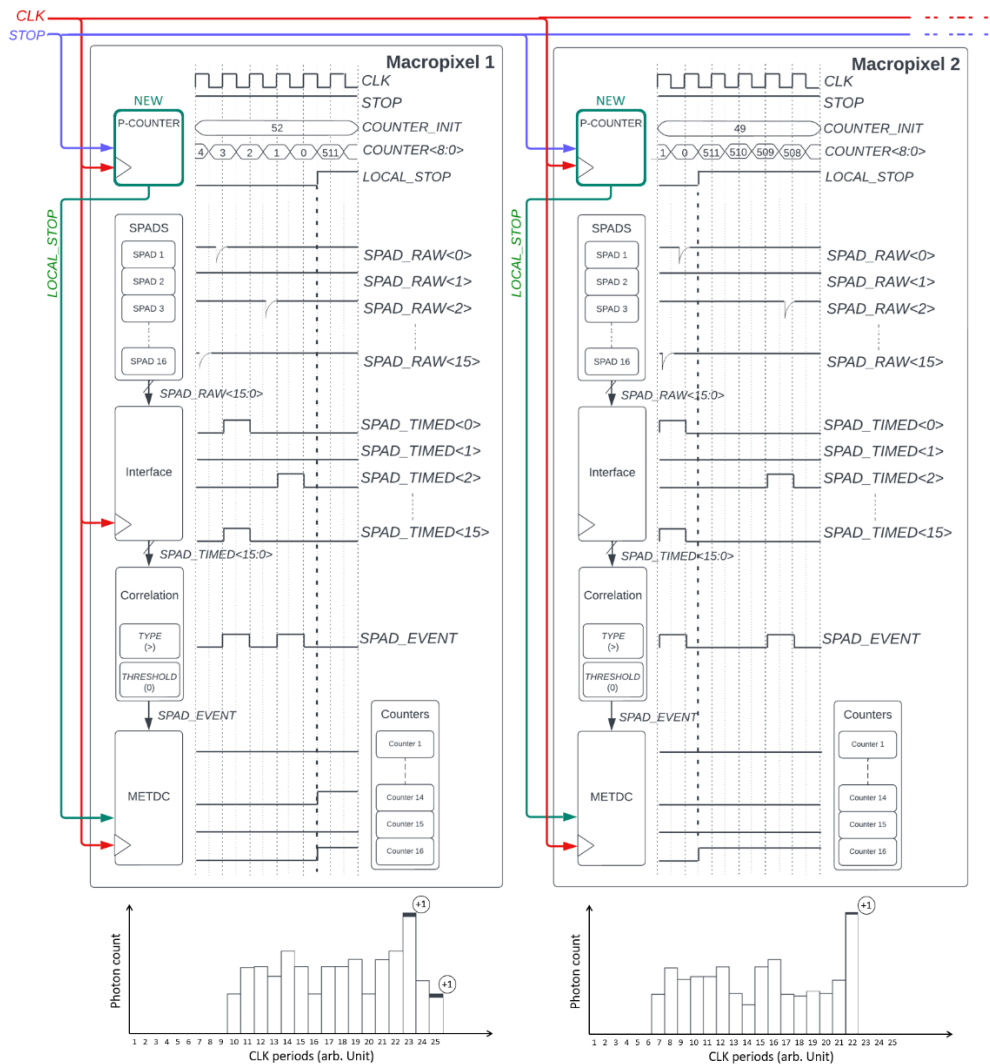
Fig. 48: Functional overview of QuantIC4×4 macropixels using the new "Local STOP" mode to create a guidable lidar sensor. In this example Macropixel 2 is programmed to update the METDC 3 CLK periods before Macropixel 1.

4.4 P-Counter Design

For guided dToF to be a practical solid-state lidar solution, the addition of the required programmable counter (P-counter) should consume minimal pixel area to maximise capacity for detecting, timing and storing photon arrival times. With this in mind, the design decisions leading to the final implemented P-counter design are presented here.

4.4.1 Comparator vs. Loadable Counter

To determine when the programmed number of global CLK periods has passed, the most intuitive solution is to store the value in memory and use a comparator to flag when the count value has reached the programmed value, as illustrated in Fig. 49(a). However, the logic to implement a digital comparison between two multi-bit values results in an area-inefficient solution.
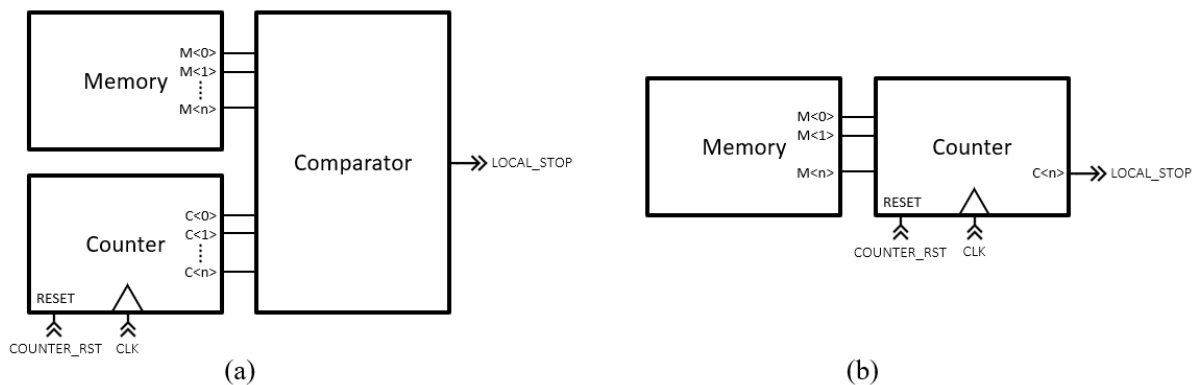
Fig. 49: Triggering LOCAL_STOP using (a) a comparator to detect when the counter reaches the programmed value (b) a loadable counter with LOCAL_STOP connected to the counter's most significant bit (MSB).

Instead, the P-counter is configured such that it loads the programmed value from memory each time it is reset by a COUNTER_RST pulse (derived from the global STOP edge). The counter is decremented by one on every CLK edge and when it rolls over from zero to maximum, LOCAL_STOP connected to the counter's most significant bit (MSB) is raised.

4.4.2 Synchronous vs. Asynchronous

In a conventional synchronous binary counter, each bit is updated simultaneously when the counter is clocked. A typical n-bit synchronous counter implementation is shown in Fig. 50(a). Each bit consists of a D-type flip-flop (DFF), XOR gate and an AND gate. An asynchronous counter on the other hand, pictured in Fig. 50(b), requires only a single DFF per counter bit. This provides a more compact solution at the expense of introducing a ripple delay between the clock triggering and the MSB updating. For the P-counter, this delay does not pose an issue provided it does not cause the LOCAL_STOP signal to trigger the METDC within the sample and hold time of the global CLK. In the event that this does happen, the phase of the externally provided global CLK can simply be adjusted. As a result, the more compact asynchronous counter topology is an appropriate choice for the P-counter.

(a)   (b)
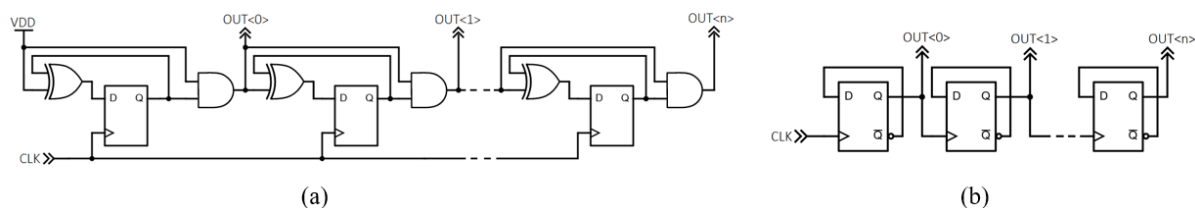
Fig. 50: Circuit implementation of a (a) synchronous and (b) asynchronous binary down counter.

4.4.3 Memory vs Memory-less

In order to reset every P-counter to a pre-programmed value after each global STOP period, without requiring to reprogram all pixels after every laser cycle, a programmable memory must be integrated into every P-counter. An alternative memory-less approach using an n-bit P-counter can be implemented by constraining the global STOP clock to a value of 2^n global CLK periods and simply letting every P-counter wraparound at the same rate as the global CLK (laser cycle period). The limited combination of STOP period (laser cycle rate) and CLK period (histogram bin interval) introduced by a wraparound approach would greatly degrade the functionality of the sensor. As a result, the P-counter is designed with a programmable memory at the expense of additional pixel area.

4.4.4 Counter Unit Implementation

To implement a loadable counter, the output value COUNTER_OUT of each ripple counter unit should reset to the value held by its corresponding memory bit MEMORY_IN, as illustrated in Fig. 51(a).
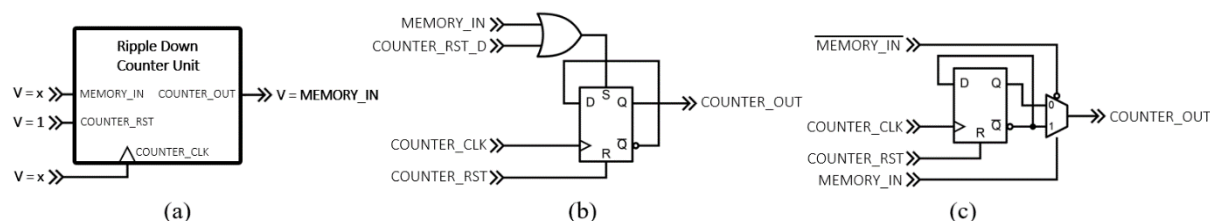
Fig. 51: (a) required reset behaviour of each ripple down counter unit (b) implementation using a set/reset DFF (c) implementation using a basic DFF.

One way of implementing the counter unit with this functionality is by using a DFF which features both a set and reset pin, as illustrated in Fig. 51(b). In this configuration, the COUNTER_RST pulse (derived from global STOP) resets all DFFs to logic low, followed shortly by a delayed pulse COUNTER_RST_D which sets any DFFs to logic high if their MEMORY_IN value is high. However, to keep the pixel layout compact, a custom verified DFF cell is used throughout the sensor design over the more elaborate library cells available in this technology. Fig. 52 shows the difference in layout area between a library cell DFF containing both set and reset, compared to the custom DFF which is around $3\times$ smaller but contains a reset pin.

Fig. 52: D-type flip flop layout (a) library cell from process design kit with set and reset (b) custom cell with reset only.

To take advantage of the custom DFF available, the implementation in Fig. 51(c) was adopted to create a down counter unit which can be reset to either logic high or low. Not only does this utilise the more compact DFF, but it only requires a 4-transistor multiplexor (mux) instead of a 6-transistor OR gate and removes the need to generate a delayed reset pulse.

4.5 P-Counter Top-Level

The top-level P-counter schematic incorporating all the prior-described design decisions is shown in Fig. 53. The core of the P-counter is a ripple down-counter, with an additional memory DFF for each bit to store the reset value loaded at the start of every laser cycle. By rerouting each ripple counter unit to connect to the next through either its Q or $\bar{Q}$ terminal based on the stored memory value in each unit, each bit of the counter is reset to either 1 or 0, allowing counting to begin from any chosen integer value. The compact design of each P-counter unit permitted a maximum 9-bit counter to be accommodated into each pixel of this sensor.
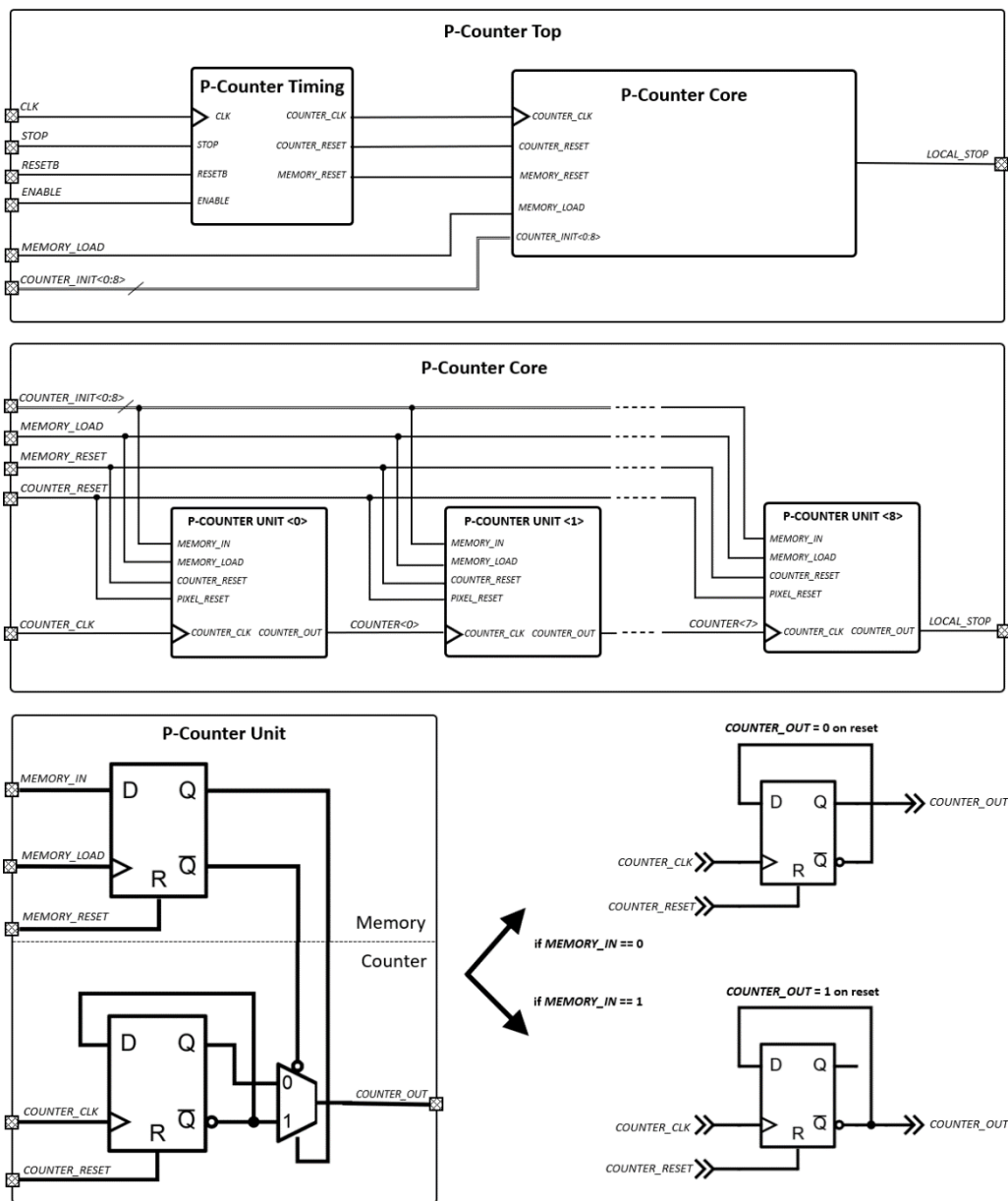
Fig. 53: Circuit diagram of the implemented P-Counter and associated sub-blocks.

An auxiliary P-Counter Timing block enables gating of signals when the P-Counter is disabled, as well as deriving the required local signals for each P-counter from the global sensor signals:

- COUNTER_RESET: a short pulse on every rising STOP edge to reset the counters.
- COUNTER_CLK: compliment of CLK to maintain LOCAL_STOP rising on the negative edge of CLK (required for the successive METDC).
- MEMORY_RESET: compliment of the sensor reset signal RESETB.

An example timing diagram is given in Fig. 54. In this example, a value of 5 is programmed into the counter on the rising edge of MEMORY_LOAD. On the first rising STOP edge, all the counters are reset (remaining at a value of 5) and then proceed to count down on each rising COUNTER_CLK edge. Once the counter reaches 0, LOCAL_STOP tied to the MSB of the counter transitions high on the next COUNTER_CLK edge. The next rising edge of the global STOP signal resets the counter and the cycle continues until either a new counter value is programmed or the block is disabled.
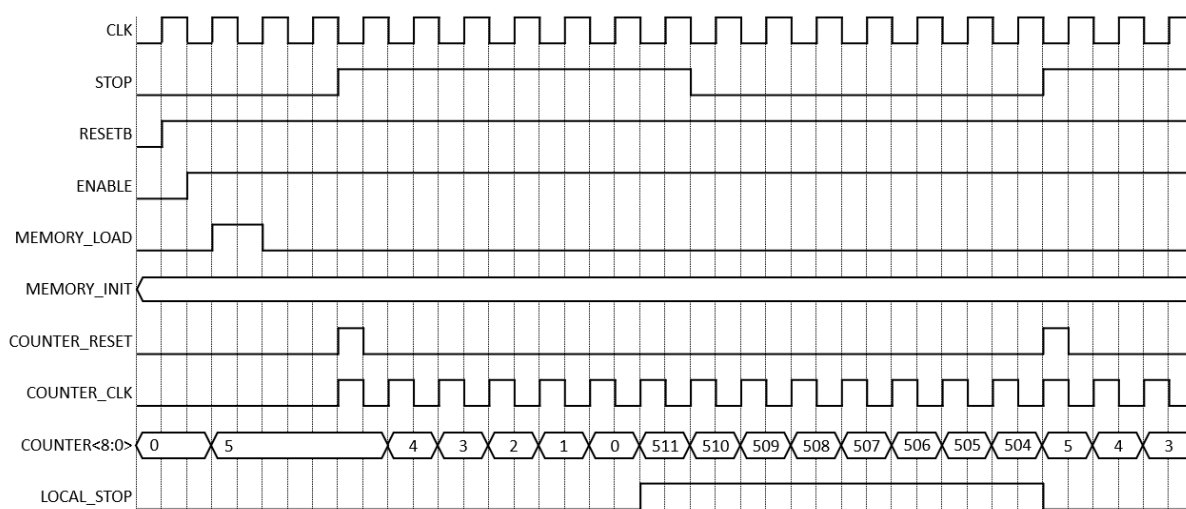
Fig. 54: Example timing of P-Counter programmed to enable LOCAL_STOP after every 5×CLK cycles of the global STOP signal.

The designed programmable down counter, requiring only 2 DFFs and 4 transistors per stage, allows for a compact solution to integrate guided dToF functionality into an existing dToF sensor. The final integrated circuit layout of a single macropixel circuit is given in Fig. 55, and shows the functionality required to implement guided lidar on this sensor consumes less than 5% of the total macropixel area.
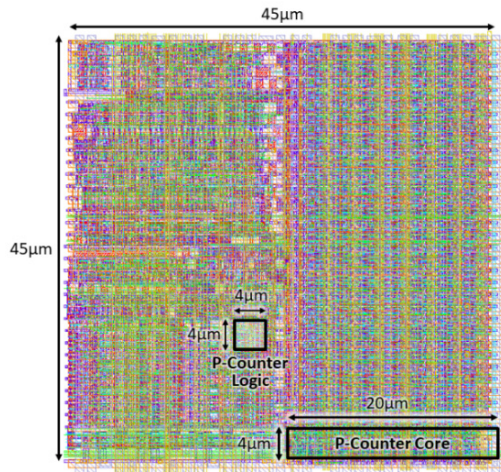
Fig. 55: Integrated circuit layout of a single QuantIC4×4 macropixel and the circuit area added to enable guided dToF.

4.6 Ripple Delay

The disadvantage of this compact ripple counter implementation is the associated ripple delay, which increases with the number of stages (maximum count value). This results in LOCAL_STOP becoming out of sync with the global CLK signal. Since the METDC increments the macropixel counters corresponding to the photon events in the 16 CLK periods before the rising edge of LOCAL_STOP, this delay places constraints on the minimal clock period (i.e. histogram bin interval).
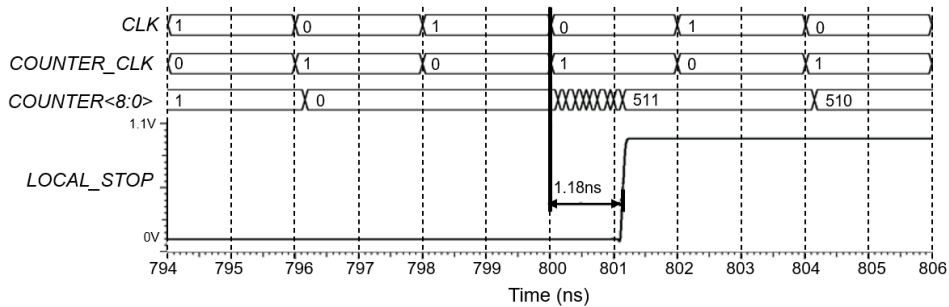
Fig. 56: Simulated worst-case P-counter ripple delay.

The post-layout ripple delay simulated at 50 °C in the typical process corner including all parasitic capacitances is shown in Fig. 56. If the global STOP signal is applied to the chip such that the rising edge is synchronised to the falling edge of CLK, then the worst-case ripple delay of 1.2 ns sets the minimum period of CLK to be around 2.4 ns (~400 MHz). However, if finer adjustment of global STOP with respect to CLK is achievable then the ripple can be compensated for, allowing for even smaller CLK period values. This issue could be avoided altogether by implementing a synchronous counter instead of a ripple counter at the cost of additional on-chip area should this be an acceptable trade-off.

4.7 Configuring Pixel Counters

An essential aspect of a guided dToF sensor involves updating the initial value of every pixel's P-counter at the start of every frame. The circuit diagram in Fig. 57 shows the blocks and signals responsible for configuring the pixel counters. To keep the sensor's pin count to a minimum, the data containing the initial value (COUNTER_INITS) of every pixel's counter is routed through a single pin. A deserialiser then converts the serial stream of data into the full 9-bit counter value (COUNTER_INIT_ROW) for every pixel in a given row. Once the initial counter values for a given row have been loaded, a row scanner triggers MEMORY_LOAD on all pixels in that row. The process repeats until all rows have been updated, as shown by the timing diagram in Fig. 57.
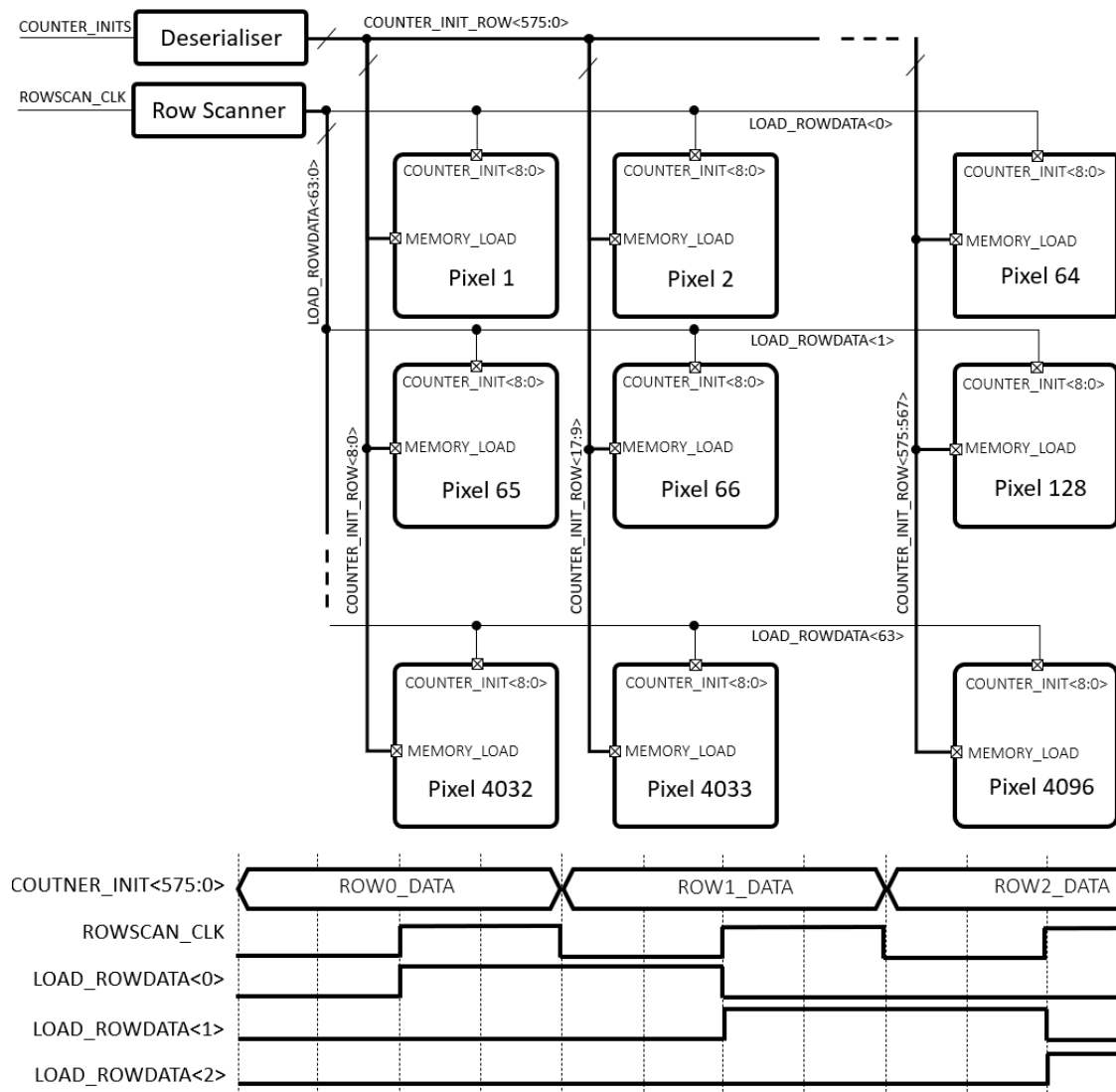
Fig. 57: The blocks and signals relevant to configuring the pixel counters are shown in the circuit diagram.

In this implementation, the time taken to reconfigure all pixel counters is limited by the deserialiser. Routing typically limits the deserialiser clock to a few hundred megahertz, making the total time to write to all pixel counters (36.9 kbits) in the order of a few hundred microseconds. However, the deserialiser piping data into a sensor is typically a separate pipeline to the serialiser piping data out and can therefore be executed at the same time. Once the acquisition period is complete, both reading out the sensor data and writing in the new depth windows for the next acquisition period can happen simultaneously. Since the total depth window data (9-bits per macropixel for this sensor) is unlikely to ever exceed the total histogram data (16×14-bits for this sensor), adding guided functionality is not expected to ever impact the time required to capture a dToF frame.

4.8 Design Limits

The available on-chip area limited this P-counter implementation to a maximum of 9 bits. The impact of this on the performance of the guided dToF sensor is explored here.

4.8.1 Maximum Range/Precision

The maximum guided lidar sensing distance z_{max} using an n-bit P-counter can be related to either the TDC clock frequency f_{CLK} or the equivalent histogram bin interval distance z_{bin} as:

$$z_{max} = \frac{2^n \times c}{2 \times f_{CLK}} = 2^n \times z_{bin} \tag{23}$$

The 9-bit P-counter integrated into this dToF sensor enables each guidable time window to be shifted by up to 512 TDC clock cycles. To achieve a maximum sensing distance of 200 m, as per the self-driving depth-sensing requirements in Table 1, the maximum TDC clock frequency that permits this distance is 384 MHz, equivalent to a minimum histogram bin interval of 0.4 m. This in turn sets a limit on the achievable measured distance precision as related by Equation (20) but still makes it feasible to reach the targeted maximum measurement error specification of 0.1 m.

In contrast, a conventional full histogram approach using the limited 16 bins would require a minimum interval of 12.5 m per bin to cover a maximum sensing distance of 200 m. This would make achieving 0.1 m precision greatly impractical.

4.8.2 Depth Estimate Error Tolerance

A second system design consideration is the maximum error that can be tolerated from the depth estimation source guiding the lidar sensor. For a sensor with M bins per macropixel, the resulting histogram will span a distance z_{hist} of:

$$z_{hist} = \frac{M \times c}{2 \times f_{CLK}} = M \times z_{bin} \tag{24}$$

Assuming a guiding depth estimate source with normally distributed error σ_{source}, the following conditions should be satisfied for a pixel to be guided to the correct time window with 99.7% ($\pm 3\sigma$) success rate:

$$\sigma_{source} < \frac{M \times c}{12 \times f_{CLK}} \quad \equiv \quad \sigma_{souce} < \frac{M \times z_{bin}}{6} \tag{25}$$

Continuing the same TDC clock example (384 MHz), this gives a maximum allowable distance error of standard deviation of 1 m for the depth estimates source guiding this guided dToF sensor up to 200 m. A reduction in TDC clock frequency would allow relaxing the guided depth error tolerance at the expense of coarser histogram bin intervals.

A look-up table of various TDC clock frequency configurations and the corresponding maximum distance and allowable guiding depth error tolerance for this guided dToF sensor is provided in Table 9.

TDC clock f_{CLK} (MHz)	Maximum distance z_{max} (m)	Histogram bin interval z_{bin} (m)	Maximum guiding depth error σ_{source} (m)
100	768	1.50	4
150	512	1.00	2.7
200	384	0.75	2.0
250	307	0.60	1.6
300	256	0.50	1.3
350	219	0.43	1.1
400	192	0.38	1.0

Table 9: Design parameter combinations using QuantIC4×4 as a guided dToF sensor.

4.9 Contingency Sensor "HSLIDAR"

In the event of the revised QuantIC4×4 sensor not being available for this project, a recently developed dToF sensor developed within the same research group provides a contingency option. The silicon-proven dToF sensor, known as HSLIDAR, was first published in [146] with further details provided in [147] and is fabricated in STMicroelectronics' single-tier front-side illuminated SPAD process [122]. A micrograph of the sensor is shown in Fig. 58. Each dToF macropixel comprises of a 4×4 array of SPADs, an METDC and storage of 8×12-bit histogram bins.
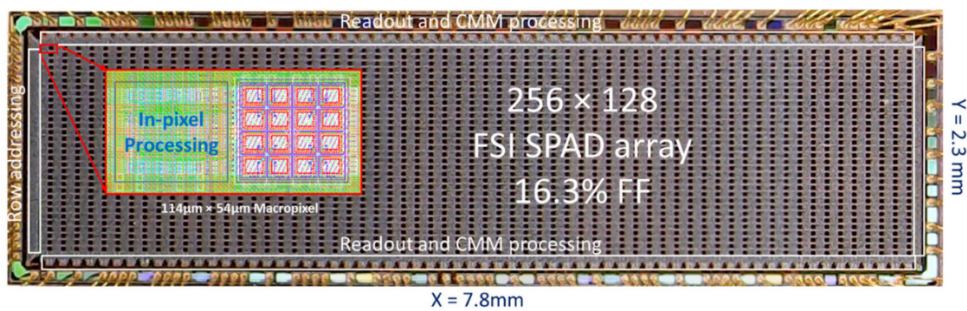
Fig. 58: Micrograph of the contingency sensor HSLIDAR. Reproduced from [146] with author's permission

The unique feature of HSLIDAR is that each self-contained dToF pixel can independently step its histogram time window over successive distance intervals and then lock on to a peak when detected. The peak laser signal is then tracked if it moves across adjacent time windows. While HSLIDAR was not designed specifically for use as a guided dToF sensor, the ability to manually set every dToF pixel to an independent time window is available as a validation feature. Therefore, this sensor presents a valuable backup solution in the event of any issues related to the fabrication of the primary QuantIC4×4, albeit with some important differences which are now explored.

4.9.1 Histogram Capacity

The contingency sensor has half the number of histogram bins per macropixel, 8 instead of 16, which reduces the depth estimate error tolerance. It also has a fourfold reduction in the capacity of each bin, 12 bits instead of 14. When targeting long distances, a greater number of laser cycles must be integrated to detect the reduced signal power. This raises the noise floor of the histogram making it more likely to saturate. The smaller bin capacity of this sensor therefore results in a reduced maximum measurable distance under high ambient background conditions.

4.9.2 P-Counter Bit Depth

To implement programmable time windows into each dToF pixel, HSLIDAR uses a similar P-Counter architecture to the one designed into the revised QuantIC4×4. However, the P-counter unit implemented on HSLIDAR uses the larger OR-gate design with standard cell DFFs (Fig. 51(b)), limiting the P-Counter to 7-bits.

4.9.3 Histogram Time Window Control

In the QuantIC4×4 design, both the histogram bin interval (a) and time window step size (s), illustrated in Fig. 59, are defined by the global CLK period. In HSLIDAR, these are configured differently and also depend on what mode the sensor is operated under. In the context of this project, the two relevant modes can be referred to as "clocked" and "delay line". In clocked mode, the timing of photons is performed by a counter TDC (Section 2.4.3) clocked by a global CLK signal. The bin interval is a single CLK period, while the time window steps in intervals of 4 CLK periods, equivalent to a 50% overlap between successive window steps. In delay line mode, a delay line TDC is used instead which sets the width of each histogram bin to the unit delay T_{DL}. In this mode, the window step size is decoupled from the histogram bin interval and is set by the global CLK signal. These modes are summarised in Table 10.
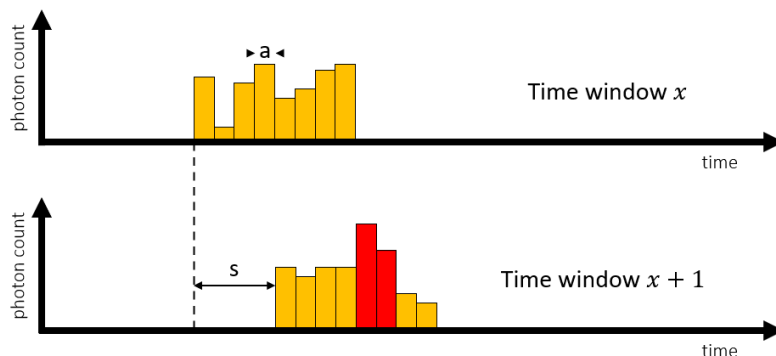
Fig. 59: Parameters which define the controllable time window in HSLIDAR.

Mode	a	s
Clocked	1/CLK	4/CLK
Delay Line	T_{DL}	1/CLK

Table 10: Configuring the histogram bin interval and window step size under the two relevant HSLIDAR modes.

4.9.4 SPAD Performance

The growing interest in SPADs for applications such as solid-state lidar has greatly boosted the development and performance of CMOS SPADs in a few short years. The SPADs on the revised QuantIC4×4 sensor, fabricated in a state-of-the-art 3D stacked BSI technology, are expected to achieve a typical PDE of 18.5% at a wavelength of 940 nm [123]. On the other hand, the typical probability of detection (PDP) of SPADs on HSLIDAR is characterised as 5% at a wavelength of 850 nm, equivalent to a PDE of 3.5% at 70% fill-factor [122], and the sensitivity continues to drop for higher wavelengths. The result is a near tenfold decrease in NIR photon sensitivity in the contingency sensor, and hence a near threefold ($\sqrt{10}$) reduction in the maximum measurable distance for a given exposure period.

4.9.5 Updating Time Windows

In order to write new configurations to HSLIDAR (i.e., to update macropixel time windows), the sensor needs to be reset each time. This is not an issue when operating the sensor in the way it was originally designed. However, when used as a guided dToF sensor, resetting the sensor between frames adds additional process steps to each frame, reducing the maximum achievable frame rate.

4.9.6 Summary

A summary of the above discussion regarding the drawbacks of resorting to the contingency HSLIDAR sensor over the revised QuantIC4×4 is provided in Table 11.

Parameter	Primary "QuantIC4×4"	Contingency "HSLIDAR"
Histogram bins per macropixel	16	8
Histogram bin capacity	14-bit	12-bit
Macropixel resolution	64×64	64×32
P-counter size	9-bit	7-bit
SPAD PDE	18.5% @ 940 nm [123]	3.5% @ 840 nm [122]
Reset sensor to update time windows?	No	Yes

Table 11: Comparison of the primary and contingency sensor

4.10 Conclusion

This chapter presents a comprehensive overview of integrating direct time-of-flight functionality into an existing dToF sensor. A design upgrade of a sensor known as QuantIC4×4 to a state-of-the-art fabrication process presents the opportunity to integrate the proposed guided dToF functionality for this thesis.

Guided dToF functionality is achieved through the addition of a compact 9-bit programmable down counter (P-counter) design with a minimal overhead of 5% of the pixel area. Along with the 16 histogram bins available in each macropixel, this allows for a guided dToF configuration to range 200 m using narrow 0.4 m bin intervals. In this configuration, the allowable error tolerance of the guiding depth estimate source to achieve 99.7% probability of detection is 1 m (single standard deviation).

In the event that the upgraded QuantIC4×4 is not available to be physically integrated into the guided dToF demonstrator, a contingency sensor HSLIDAR is identified. The mechanism by which HSLIDAR can be configured into a reconfigured for use as a guided dToF sensor is described, along with the repercussions of using this contingency option. Most notably, the reduced number of bins from 16 to 8 tightens the guiding estimate error tolerance requirement, while the lower SPAD sensitivity greatly reduces the maximum range for a given exposure time.

As well as the sensor, the other key component of a guided dToF system is the source of guiding depth estimates. The next chapter explores the imaging sensors available on a self-driving vehicle for providing depth estimates in a guided dToF system, along with their merits and limitations.

5 DEPTH ESTIMATION FROM IMAGES

5.1 Introduction

A key component of the proposed guided dToF system is the source of depth estimates for guiding the dToF sensor. Cameras are essential for autonomous vehicles and are capable of providing depth estimates after image processing, making them highly suited to guided dToF for self-driving vehicles. This chapter examines various methods of extracting depth information from images. These span from traditional techniques such as stereo depth estimation to novel methods involving the extraction of depth information from lidar intensity images. In each case, an imaging system is constructed and performance is evaluated within real-world environments. This allows a suitable depth estimation approach to be identified for integration into the guided dToF system.

5.2 Stereo Depth Estimation

The most established method of extracting depth information from camera images is stereo depth estimation using two (or more cameras). The principle is illustrated in Fig. 60 and involves matching pixels in the image of one (e.g., left) camera O_l to its matching image pixel in the image of another (e.g., right) camera O_r. The number of pixels any point has shifted by $(d_l - d_r)$ between camera images is termed disparity d and can be directly related to distance. The left camera is typically chosen to be the principle with each pixel assigned a disparity with respect to the secondary camera.
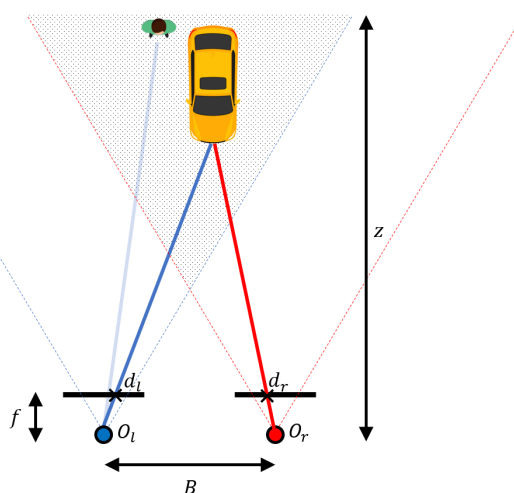
Fig. 60: Illustration of the parameters related to stereo depth estimation.

For a pair of cameras separated by a baseline distance B with focal length f, the distance z to the world point observed by a pixel is related to the disparity d of that pixel by Equation (26).

$$z = \frac{f\,B}{d} \qquad \text{where } d = d_l - d_r \tag{26}$$

The method by which depth is estimated using this technique presents some inherent limitations which are explored in the subsequent section.

5.2.1 Limitations of Stereo Depth Estimation

Disparity Error

The distance accuracy Δz of stereo depth estimation is derived from Equation (26) to give Equation (27) below. This reveals the characteristic squared increase in error associated with stereo depth estimation.

$$\frac{\Delta z}{\Delta d} = -\frac{f\,B}{d^2} \;\; \Rightarrow \;\; |\Delta z| = \frac{z^2\,\Delta d}{f\,B} \tag{27}$$

Since each image is composed of discrete pixels, this can lead to quantised values of disparity uncertainty Δd which inevitably limits the achievable depth accuracy. However, many disparity correspondence algorithms can achieve sub-pixel matching using techniques such as interpolation [155].

The matching of pixels between the principal camera image and the reference camera image is performed by the stereo-matching algorithm and is the foundational element for reliable depth estimation. In reality, the stereo depth estimation accuracy is limited by the stereo-matching ability of the chosen algorithm [26]. Stereo-matching algorithms can be broadly categorized into two main groups (i) traditional approaches use classic computer vision methods like block matching to find disparities between left-to-right camera images and (ii) Machine learning-based approaches employ convolutional neural networks (CNNs) to directly learn disparity estimation from large publicly available datasets.

Mechanical Stability

Any movement in the translation or rotation of either camera after calibration introduces a measurement error [156], this is particularly problematic for self-driving vehicles.

Occlusions

Another limitation is the challenge of resolving objects which are occluded from the view of the secondary cameras. This is illustrated in Fig. 60 where the pedestrian is only visible to the left camera, hence there is no corresponding pixel in the right camera image to match to for estimating disparity.

Visibility

The captured signal quality of any image sensor working in the visible spectrum is susceptible to adverse weather and low light conditions. This in turn degrades depth estimation performance. Although in the context of self-driving vehicles, headlamps help mitigate this issue.

Range

Fig. 60 also illustrates how stereo depth estimation is limited to the region where both camera field-of-views overlap. While the camera baseline, image sensor size and focal length physically limit this space and hence the minimum measurable distance, in reality this is limited by the maximum resolvable disparity. Allocating a wide disparity search range to the stereo-matching algorithms greatly increases the runtime. As a result, this is often limited to a maximum value of 128 pixels. The maximum practical range is limited by the squared increase in error at small disparity values. Setting a minimum disparity value of 10 pixels limits the distance error for $\Delta d = \pm 1$ pixels to 10%. Combining these limits, the practical measurable distance range for a stereo depth rig with a 1 m baseline as a function of focal length is summarised in Fig. 61.
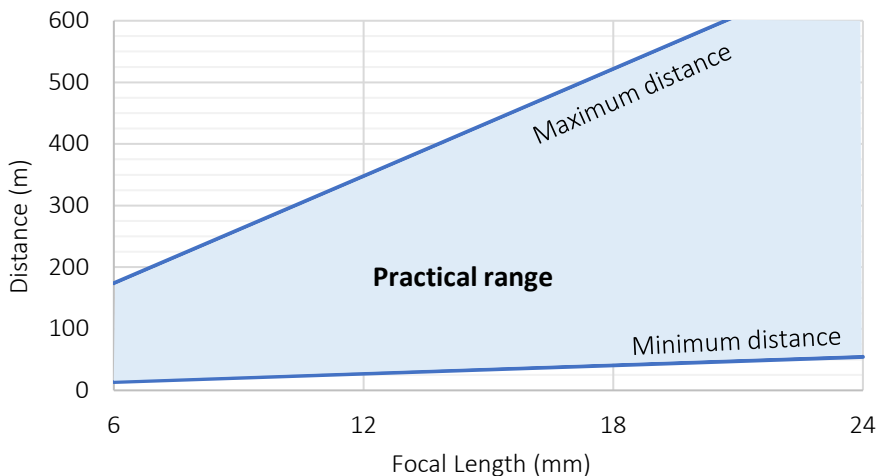
Fig. 61: Practical distance limits for a stereo camera rig separated by a 1 m baseline.

5.2.2 Practical Setup

A summary of the stereo camera setup built for this investigation is provided in Table 12. While basic camera models are used, the fast frame rate of these cameras should ensure that camera readout is not a bottleneck within the implemented guided dToF system. The 12 mm focal length lenses provide a balance between optical distortion and stereo depth estimation error which are inversely proportional to focal length, while having the potential to range at the automotive maximum distance target of 200 m as set out in Fig. 61. The resulting FOV of each camera is 23° × 18°. Although state-of-the-art machine learning algorithms now outperform traditional computer vision algorithms for stereo depth estimation [157], the primary aim of this work is to explore the feasibility and practicability of guided dToF. Therefore, the established semi-global matching (SGM) algorithm is adopted for simplicity [158]. The 1 m rail provides a practical stereo camera baseline for vehicle integration. A photograph of the stereo camera rig for depth estimation is shown in Fig. 62.

Part	Parameter	Value
Cameras	Model	FLIR Blackfly BFS-U3-162M-CS
	Sensor	Sony IMX273
	Chroma	Mono
	Maximum frame rate	226 fps
	Image sensor resolution	1440 × 1080
	Image sensor pixel pitch	3.45 µm
Lens	Model	Computar M1224-MPW2-R
	Aperture (f-number)	2.4
	Lens focal length	12 mm
Processing	Processor	1.9 GHz Intel Core i7 8th generation
	Stereo-matching algorithm	Semi-global matching (SGM)
Rail	Length	1 m

Table 12: Itemised summary of the implemented stereo camera rig

Fig. 62: Stereo camera rig used in this work.

5.2.3 Process Flow

The full pipeline from acquiring camera images to producing a stereo depth estimate image is shown in Fig. 63. Before stereo-matching can be conducted, the cameras must be carefully calibrated by taking images of a checkerboard in various poses using both cameras [26]. This allows camera parameters to be extracted in the form intrinsic and extrinsic camera matrices using the MATLAB Stereo Camera Calibrator [159]. The intrinsic camera matrix is a 3×3 matrix that describes the internal characteristics of a camera affecting image formation such as focal length and optical centre, while extrinsic camera matrix is a 3×4 defining the camera positions and orientations relative to each other. This enables (i) image rectification which aligns all points in both images along the same horizontal plane for simplified stereo-matching and (ii) conversion of disparity to distance.
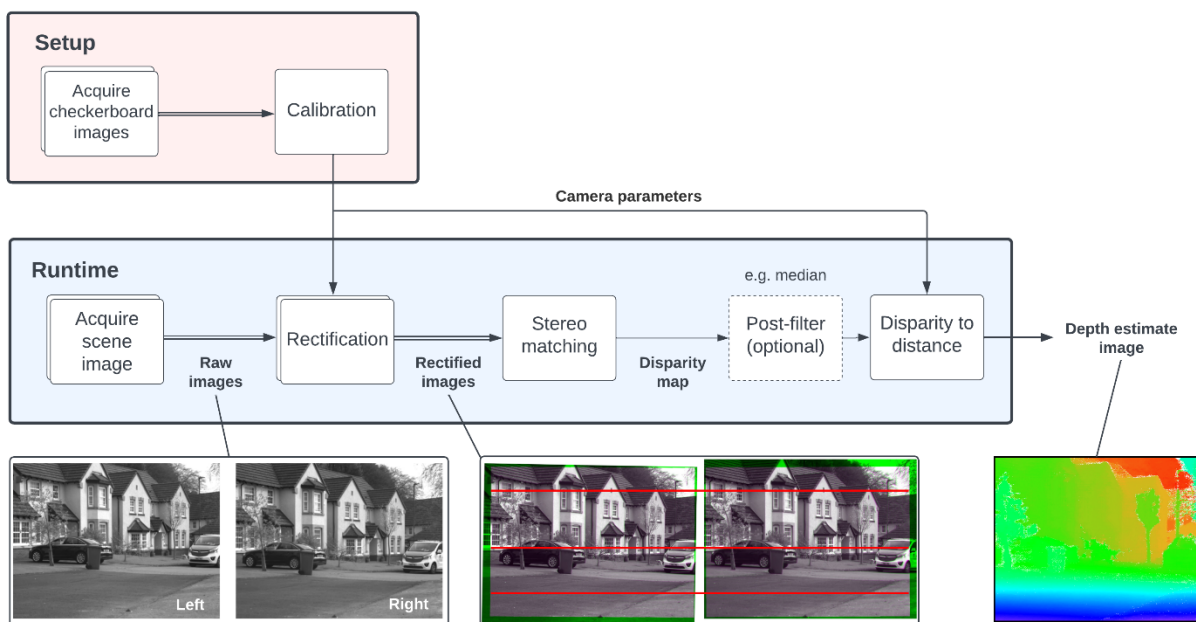
Fig. 63: Process flow diagram for estimating depth from stereo cameras.

97

5.2.4 Performance

To assess the standalone stereo depth estimation performance of the implemented camera rig, a distance sweep is performed. A target is moved at various distance intervals up to a maximum distance of 200 m in a real-world outdoor environment. At each new target position, the distance to the target is estimated and compared to the ground truth value established using a Bosch GLM250VF laser rangefinder.
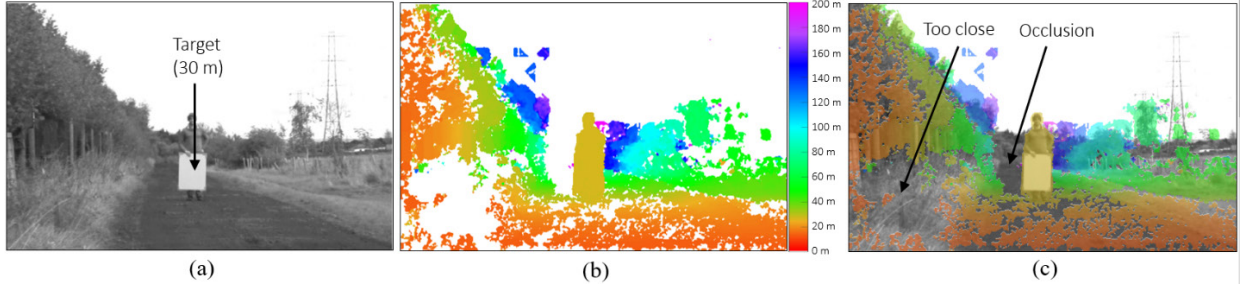
Fig. 64: Sample capture during the distance sweep experiment (a) left camera image (b) estimated stereo depth image (c) both images overlaid.

A sample image from this experiment is shown in Fig. 64, with further samples at various target distances shown in Fig. 65. While the target continues to be resolved as far away as 200 m, a couple of the prior discussed challenges of stereo depth estimation are evident in these images. Most notably, the region to the left of the target at 30 m is occluded from the perspective of the right camera, leading to unresolved stereo-matching. However, this effect becomes less pronounced as the target moves further from the cameras. Also evident is the limitation of resolving parts of the scene which are too close to the cameras, as established in Fig. 61. For this configuration of 12 mm focal length cameras with a baseline separation of 1 m, this minimum distance limit occurs at a distance of around 15 m.

To quantify the accuracy and precision of the camera rig's stereo depth estimates in an outdoor environment, a 5×5 pixel window of the target is sampled at each distance. The average estimated distance in each sample window, along with the spread of values (given as 3 standard deviations) is plotted in Fig. 66. The single standard deviation value (σ) is also plotted separately for each distance interval in Fig. 67 along with the estimated $\pm 3\sigma$ interval based on fitting a second order polynomial to the data. Finally, the root-mean-square (RMS) error at each distance interval is shown in Fig. 68, exemplifying the squared increase in distance error characterised by Equation (27). The corresponding error given by Equation (27) for an alogrithm achieving a sub-pixel disparity of ⅓ pixels with perfect stereo-matching is also overlaid for reference.
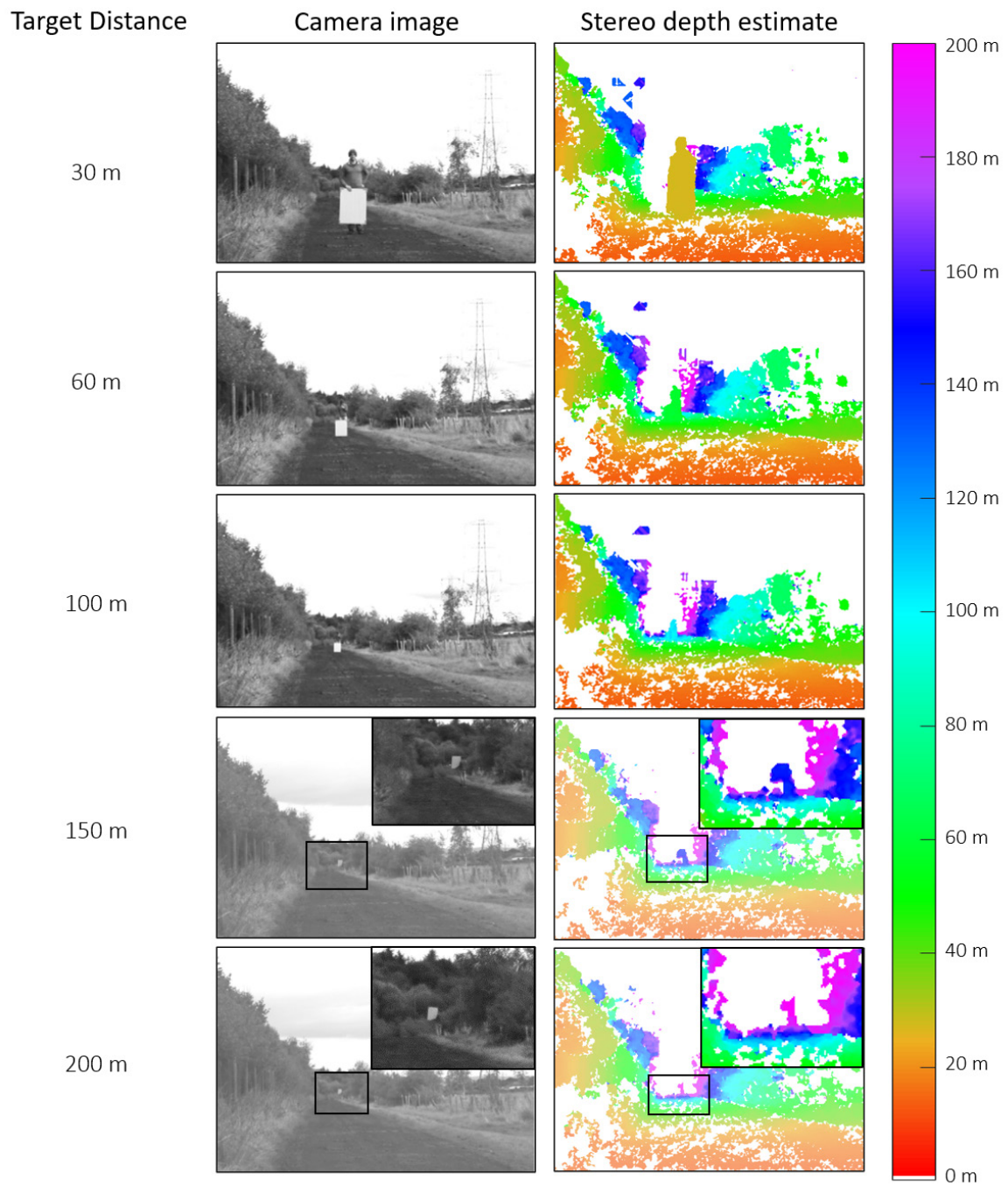
Fig. 65: Sample capture during the distance sweep experiment at various target distances and the corresponding depth image from stereo depth estimates.
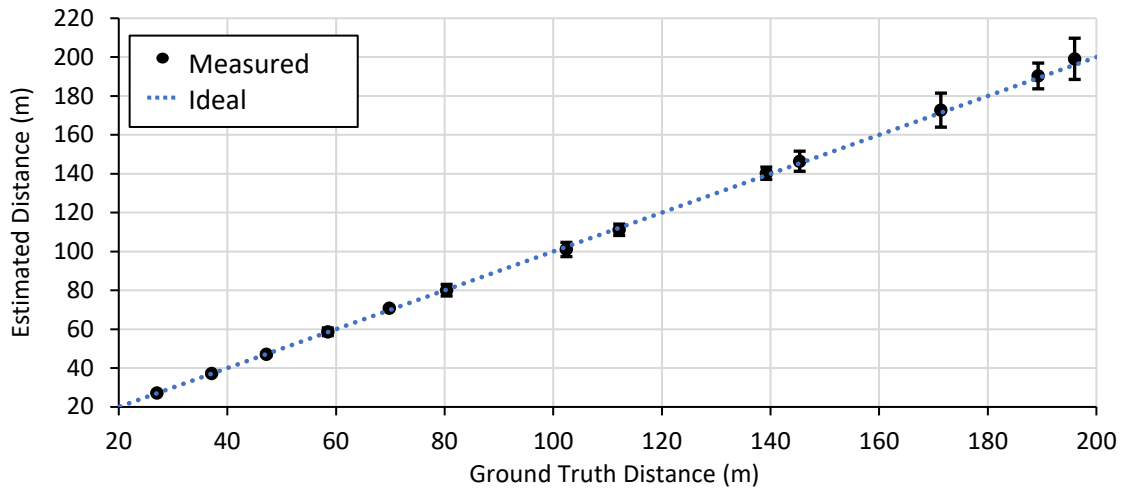
Fig. 66: Estimated stereo depth mean and variance (3 standard deviations) at each step in the distance sweep test over a sample window of 5×5 pixels.
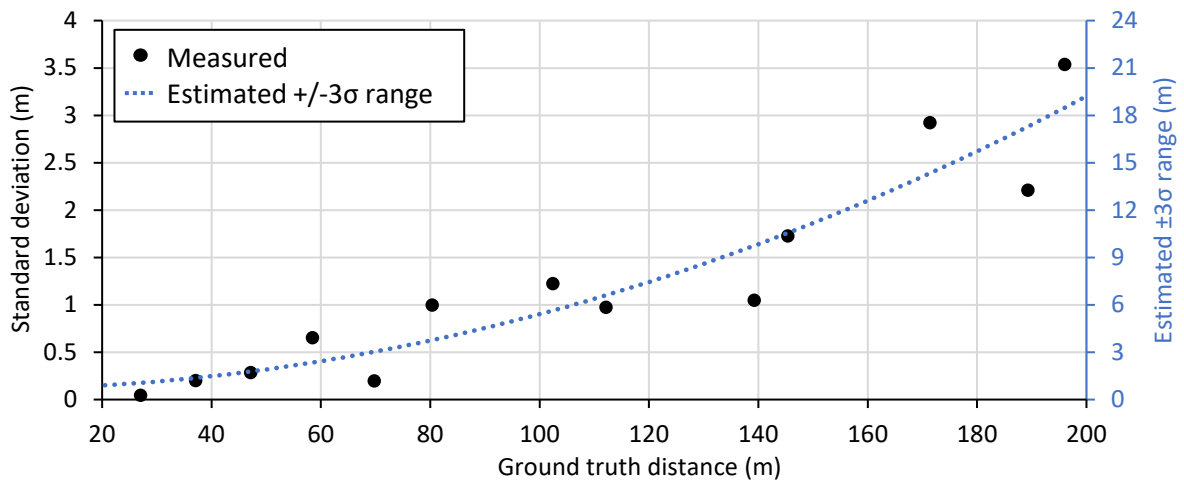
Fig. 67: Measured standard deviation (σ) at each distance interval and estimated $\pm 3\,\sigma$ range based on a squared polynomial best fit.
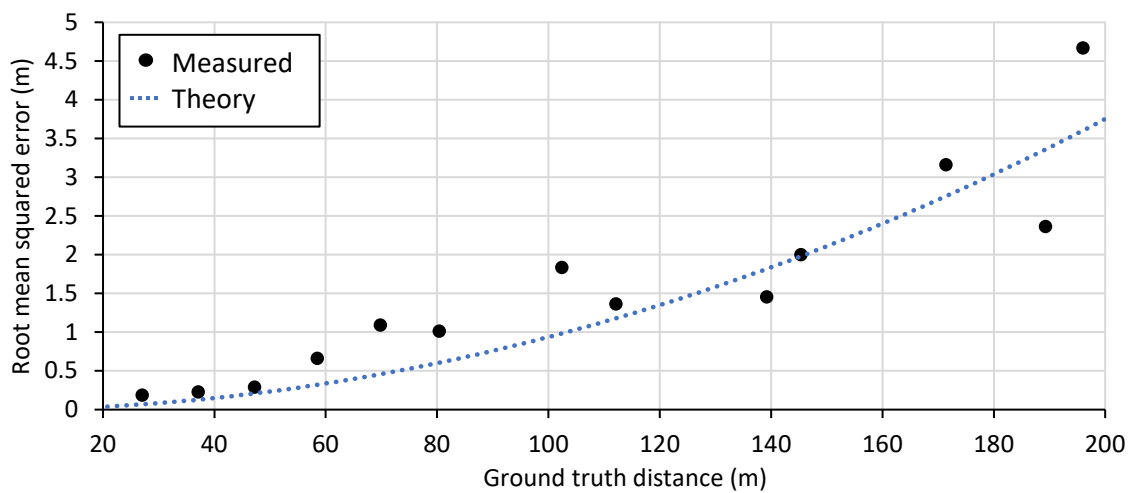
Fig. 68: RMS error of stereo depth estimates at each distance step, alongside theoretical accuracy given by Equation (26) for sub-pixel disparity of ⅓ pixels.

As well as the quality of depth estimates, another factor which impacts the guided dToF system is the time taken to produce the guiding depth estimates. Fig. 69 shows the time taken for the full process chain (Fig. 63) from image acquisition to depth estimation.
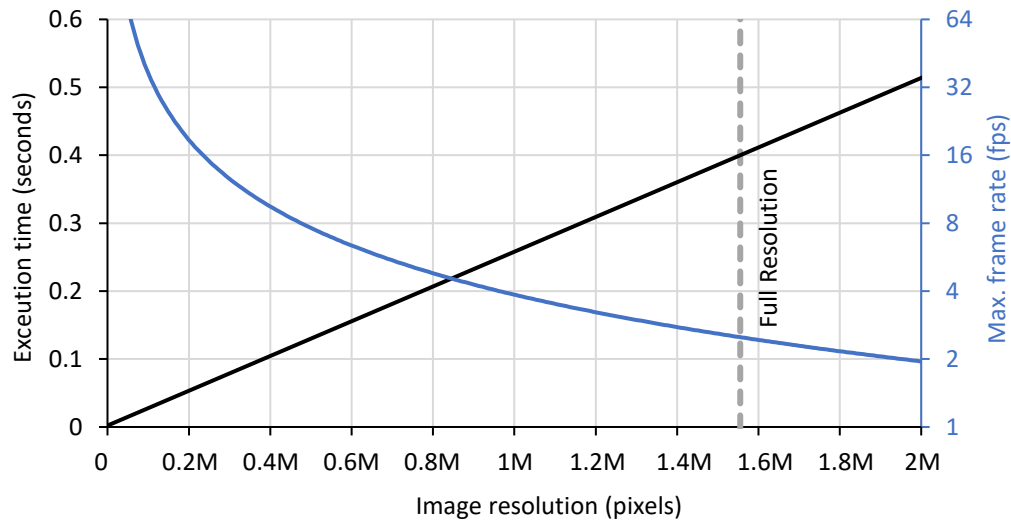
Fig. 69: Elapsed time for the implemented stereo camera rig to produce a depth image.

This setup achieves a frame rate of 3 fps using full-resolution images, with the rectification and stereo-matching steps equally dominating the frame time. As a result, the frame rate can be significantly improved by reducing the resolution of captured images, either through binning or cropping.

Finally, the sensitivity of the rig to mechanical variations is investigated. Fig. 70 shows the change in measured distance for a target at a distance of 40 m when the camera angle is adjusted without performing recalibration. Even sub-degree movement is shown to exhibit substantial measurement error.
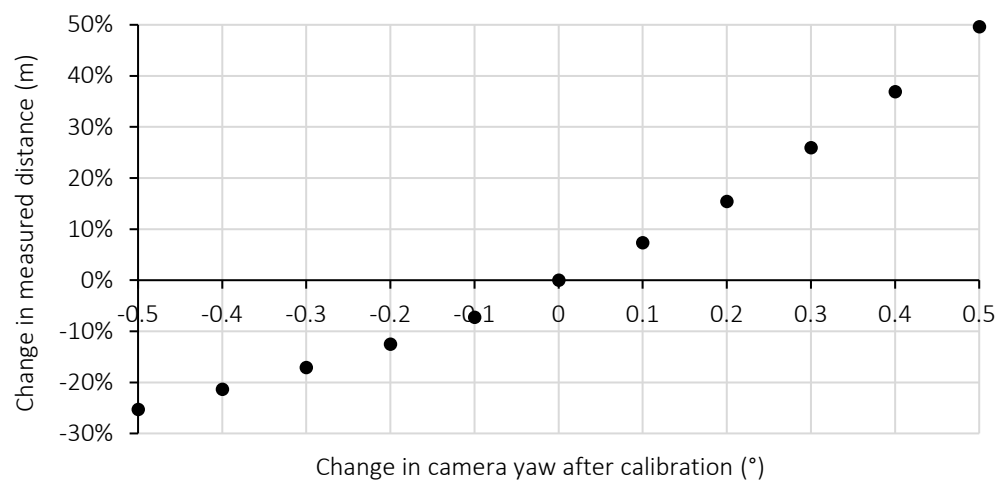
Fig. 70: Effect of changing the camera yaw after calibration at a 40 m target. Positive angles represent the cameras verging away from each other.

5.2.5 Discussion

The implemented stereo camera rig is able to provide depth estimates as far as the maximum distance of 200 m required for fully autonomous vehicles (Table 1), albeit with increasing error over longer distances. In Chapter 4 the relationship between the guided depth sensor's histogram width and the error from the depth estimation source required to successfully guide a pixel 99.7% ($\pm3\sigma$) of the time was outlined in Equation (25). Based on the measured $\pm3\sigma$ depth estimation spread achieved by the implemented stereo camera rig shown in Fig. 67, a guideline for the equivalent minimum histogram bin interval for both sensor options presented in Chapter 4 is shown in Fig. 71. To recap, "primary" refers to the better suited revised QuantIC 4×4 sensor should it be available and "contingency" refers to the lesser suited HSLidar sensor which is already available. The pros and cons of each are summarised in Table 11. While adopting a smaller bin interval than that advised in this figure may be desirable for better lidar accuracy and precision, this would be at the expense of fewer than $\pm3\sigma$ (99.7%) of stereo depth estimates landing in the correct guiding window.
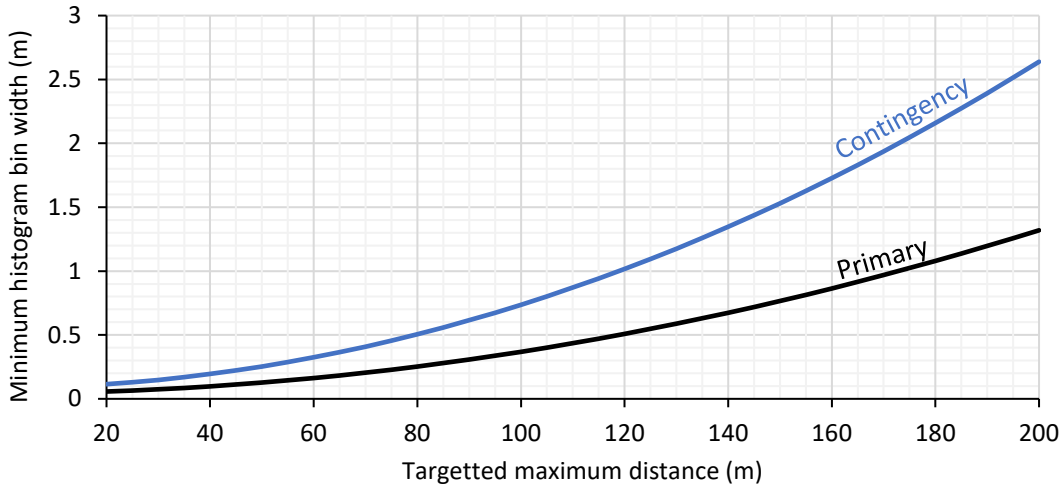
Fig. 71: The minimum histogram bin interval required for each sensor option from Chapter 4 to achieve 99.7% of depth estimates landing in the correct guiding window.

The minimum resolvable 15 m distance of the implemented stereo camera rig will place a limitation on the guided dToF system. While this could be resolved by implementing a three-camera (trinocular) rig, with various baselines providing a mixture of both long- and short-distance depth estimation, the added complexity of a trinocular system is superfluous for the purpose a guided dToF demonstrator. In addition, self-driving vehicle designs typically adopt a variety of short and long-range lidar [38, 160], as illustrated in Fig. 72. The limited short-range depth-sensing performance of a long-range guided dToF sensor using this stereo camera rig would therefore complement such a lidar configuration.
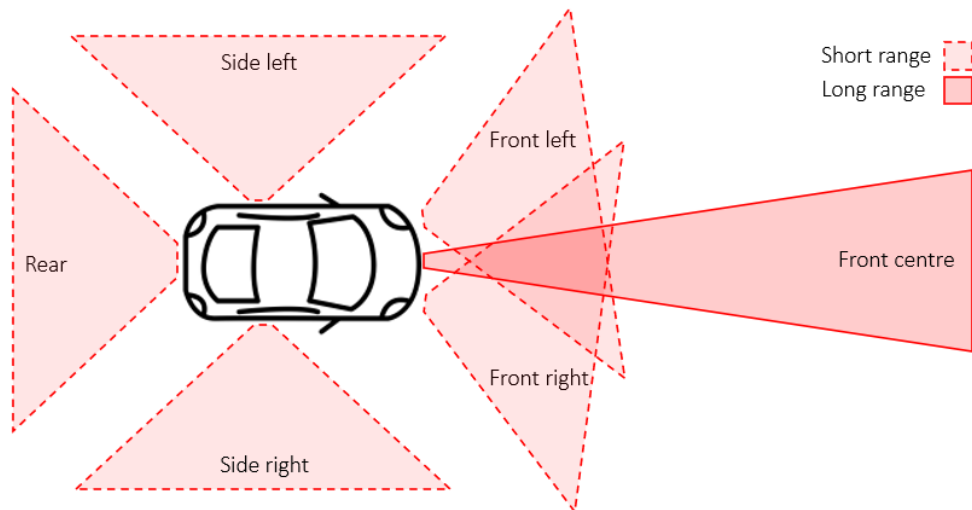
Fig. 72: Prospective lidar arrangement for fully autonomous self-driving vehicles.

While faster and more accurate depth estimation may be possible through the adoption of graphics processing units (GPUs), better cameras or advanced learning-based algorithms, the exhibited performance of the implemented stereo camera rig is suitable to demonstrate the concept of guided dToF.

5.3 Monocular Depth Estimation

While stereo depth estimation is a well-established technique, methods of extracting depth information from a standalone camera also exist. This method is known as monocular depth estimation. As with stereo depth estimation, these techniques fall into two categories. Traditional methods use visual cues such as perspective projection to infer depth based on the apparent sizes and positions of recognised objects [161], motion parallax [162] and camera focus [163]. Challenges occurs when scenes lack any distinct features where these methods can be applied on. With the extensive development of deep learning, more modern approaches now use CNNs trained on large datasets to directly predict depth maps from single images [164]. While these can provide exceptionally dense depth maps [165], their performance relies on the availability of ground truth datasets of which are currently lacking in sufficient quantity.

Although traditional stereo depth techniques such as SGM can produce accurate depth maps, traditional monocular depth techniques are less robust, particularly in scenes lacking reliable cues [164]. As a result, learning-based approaches have quickly assumed a more favourable choice for resolving monocular depth. In 2020, a pre-trained monocular depth neural network developed by Intel called MiDaS which

achieves state-of-the-art results on several benchmarking platforms [166] was made open-source. It is therefore also worthwhile exploring this as a means of guiding a guided dToF system.

5.3.1 Process Flow

Since there is no need to align images from multiple cameras via rectification, the preliminary step of calibration is not required for monocular depth estimation using MiDaS. However, image transformation is still a necessary step during runtime in order to scale images to the same pixel resolution in which the network is originally trained on. Once scaled, the image is passed through the trained MiDaS network to produce a disparity image which can then be transformed back to the original resolution.

While the notion of disparity values is clear for stereo depth estimation (i.e., the number of pixels a point has shifted by between the left and right camera) it is not as trivial in the use if monocular depth estimation and as such requires further explanation which is provided here. Many large datasets of synthetic (e.g. Middlebury [25]) and real (e.g. KITTI [26]) scene images along with per-pixel ground truth distance have been gathered. These were established to accelerate stereo depth estimation given its long-established research efforts. As a result, these datasets provide ground truth in the form of disparity. Learning-based monocular depth algorithms such as MiDaS, which depend on large volumes of training data, are therefore trained on these stereo datasets and consequently output depth in values of disparity. However, if a reference point of known distance is available in the image, as will be the case in the distant sweep tests to follow, then converting from disparity to distance for that pixel and subsequently all pixels in the image can be performed.
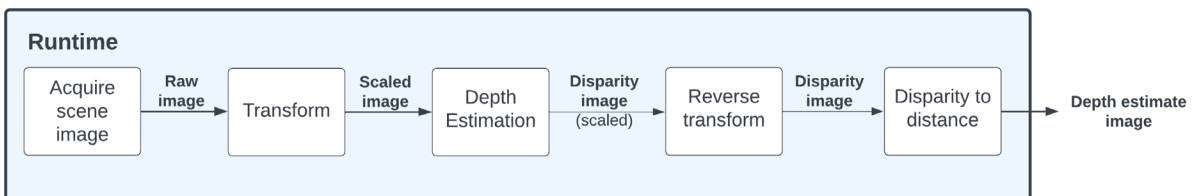
Fig. 73: Process flow diagram of learning-based monocular depth estimation.

5.3.2 Performance

To provide a direct comparison to the stereo depth estimation established earlier in the chapter, the same dataset of images is used for assessing monocular depth estimation using MiDaS. The resulting processed depth image using MiDaS on the same sample image of Fig. 64 is shown in Fig. 74.

The high density of the resulting depth image compared to the stereo depth estimate using SGM is immediately apparent. By eliminating the reliance on a second camera, both the inability to resolve distances close to the camera and the effect of occlusions are circumvented. However, the quality of a depth image hinges not on the abundance of depth estimates, but on their precision and accuracy.
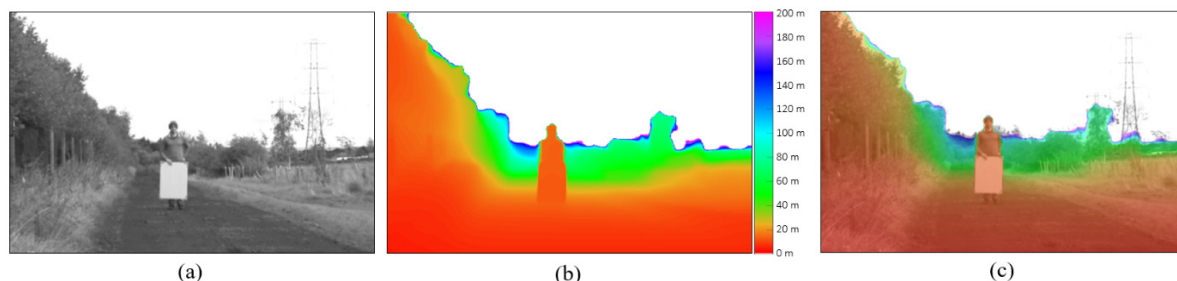
Fig. 74: Depth estimation using MiDaS (a) original camera image (b) estimated depth image (c) both images overlaid.

Observing the depth image as the target moves further from the camera in Fig. 75, the limitation of this technique in its current state becomes apparent. Unlike the results using stereo depth estimation (Fig. 65), the target quickly disappears from the depth image as it moves further from the camera.

In Fig. 76 the monocular depth estimate to the target at each distance step is benchmarked against the measured ground truth, the resulting RMS error is summarised in Fig. 77. The data demonstrates that the accuracy of this approach significantly lags behind that of the implemented stereo camera approach (Fig. 68). This discrepancy is evident not only at longer distances (e.g., beyond 120 m) where the accuracy substantially degrades but also at shorter distances where the RMS error is more than ten times greater than the equivalent stereo depth estimate.

The last crucial aspect to take into account, especially when dealing with neural network-based methods, pertains to processing time. When implemented on the same system as the stereo camera setup, each frame of the monocular depth process requires approximately 15 seconds for execution, resulting in a frame rate of under 0.1 fps. As this is dominated by the depth estimation network which uses images transformed to a set resolution, the execution time is largely independent of the original image resolution, unlike the execution time using traditional stereo depth estimation (Fig. 69).
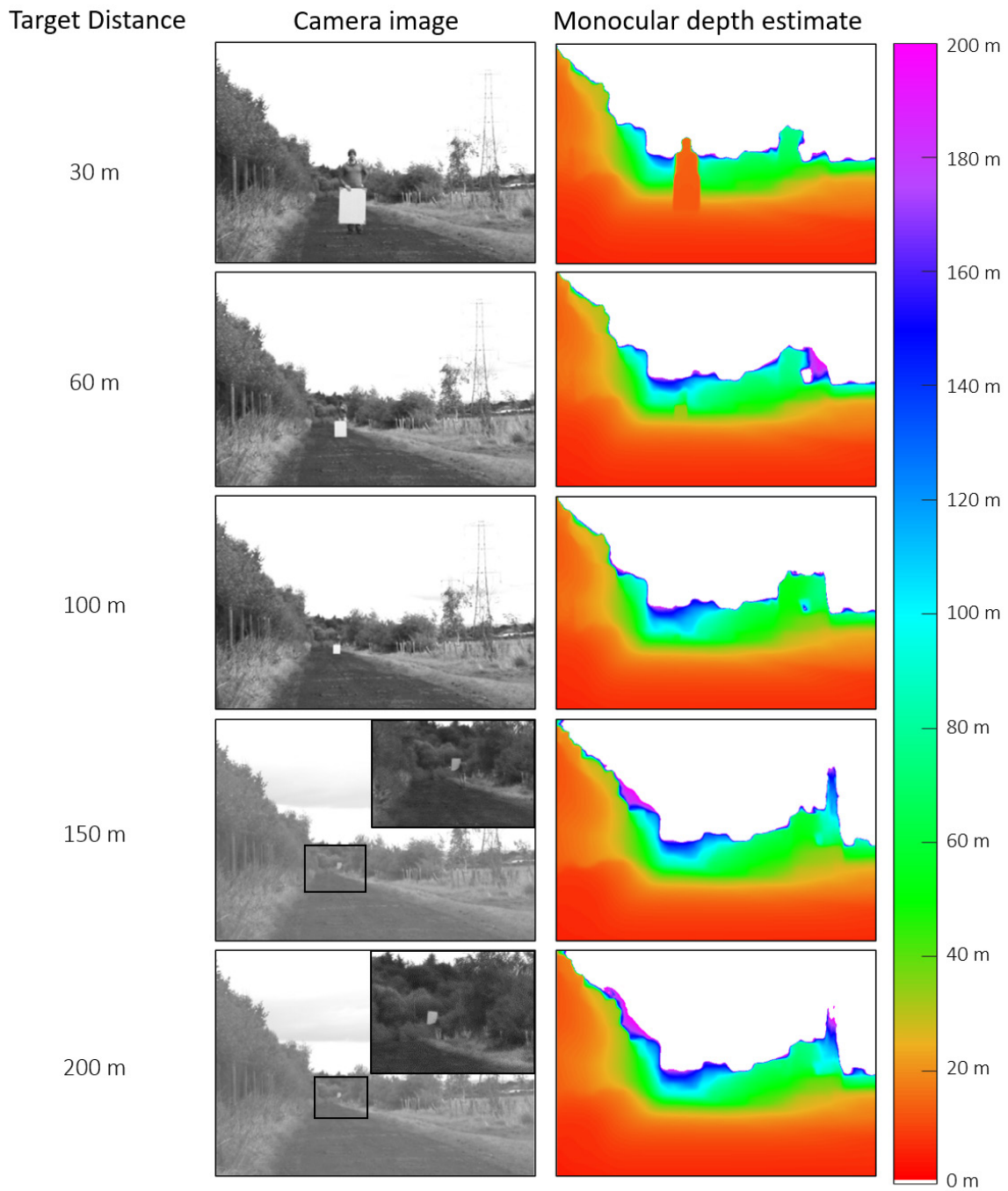
Fig. 75: Sample capture during the distance sweep experiment at various target distances and the corresponding depth image from stereo depth estimates.
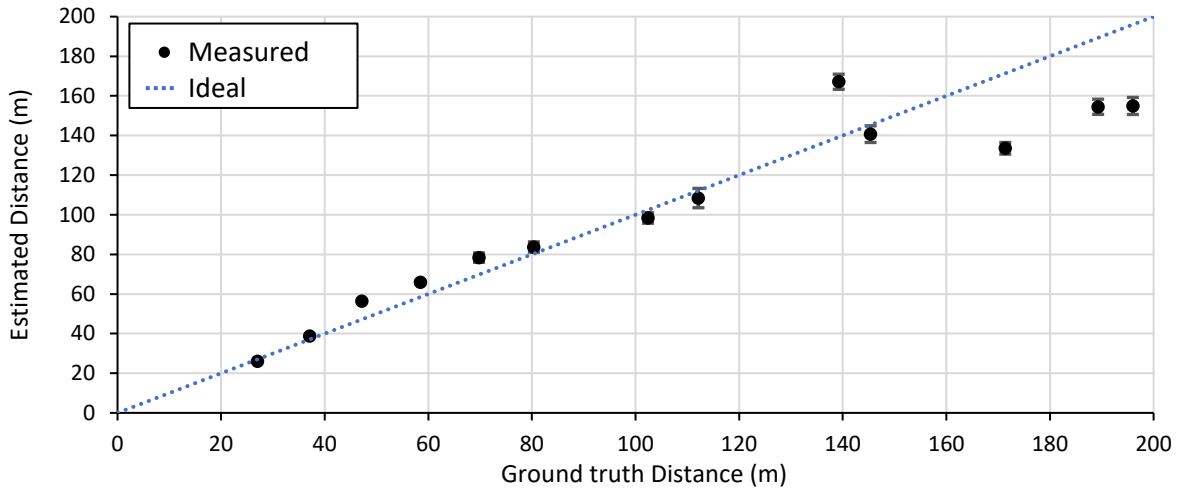
Fig. 76: Average estimated monocular depth and variance (3 standard deviations) at each step in the distance sweep test over a sample window of 5×5 pixels.
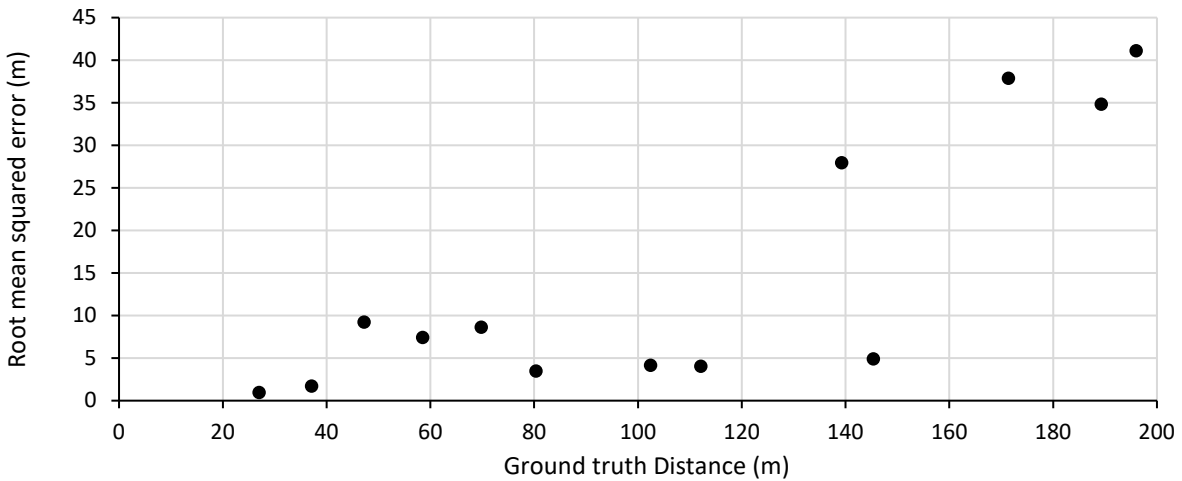
Fig. 77: RMS error of monocular depth estimates at each distance step.

5.3.3 Discussion

State-of-the-art learning-based monocular depth estimation algorithms such as MiDaS now allow notably dense depth data to be extracted from single camera images, without the artefacts present in stereo depth estimation such as occlusion. However, after benchmarking the accuracy of this technique on real-world distance sweep, it is clear that using MiDaS in its current state would not be suitable for a guided dToF system. Furthermore, the long execution time required for images to pass through the trained neural network would not only present a major bottleneck to the system frame rate but also cause motion artefacts by guiding pixels to regions where an object has already passed through.

5.4 Lidar Intensity Depth Estimation

While vision cameras are the most direct means of capturing photographic images, lidar is also capable of producing camera-like images. In measuring the laser return time, the total light accumulated by each ToF pixel within the exposure period is data that is also inherently collected. This intensity data can be used to produce a monochromatic of the scene, as illustrated in Fig. 78. Consequently, it follows that analogous image processing techniques to those employed in traditional camera systems, can also be applied to lidar intensity data to infer depth information. This work presents the first known application of this inherent lidar intensity data to extract additional depth information.
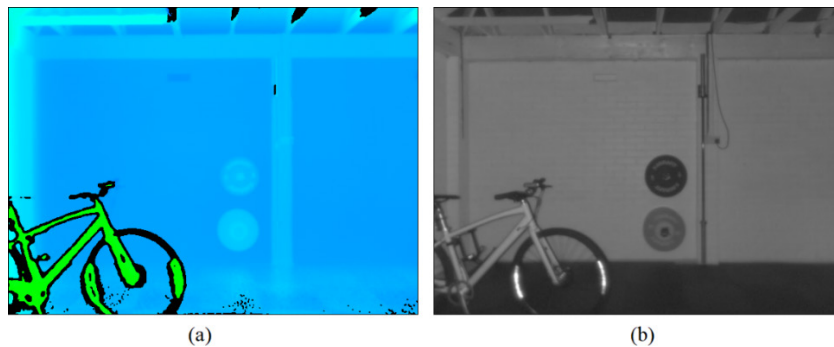
Fig. 78: (a) depth data from a lidar sensor (b) corresponding intensity data from the same scene.

Consider the autonomous vehicle lidar configuration presented in Fig. 72. In any setup such as this where there is scope for overlapping lidar field of views, the applicability of stereo depth estimation techniques is a worthwhile investigation.

While the work presented in this section specifically derive depth estimation from the lidar's intrinsic intensity image, an alternative and frequently explored approach involves augmenting the lidar depth image with additional vision cameras. These approaches can generally be classified into two categories (i) depth completion and (ii) error refinement. In depth completion, vision camera images are utilized to enhance the resolution of typically low-resolution lidar cameras. One of the earliest examples was introduced in [167], employing a Markov Random Fields (MRF) technique to match discontinuities in lidar depth with those in camera brightness. A comprehensive overview of similar methodologies stemming from this approach is provided in [168]. The MRF technique is further enhanced in [6] to better accommodate dynamic scenes. In error refinement, additional data from vision sensors is used to enhance the accuracy. A prominent examples is explored in [169] where a fusion algorithm is used to improve accuracy in scenes with varying texture.

Fig. 79 illustrates the proposed method of using lidar intensity data to enhance the final depth data of a pair of lidar sensors. By applying stereo-matching algorithms to the monochromatic intensity images, additional depth estimates are obtained which are then combined with the original lidar depth data to form a more complete depth image of the scene.
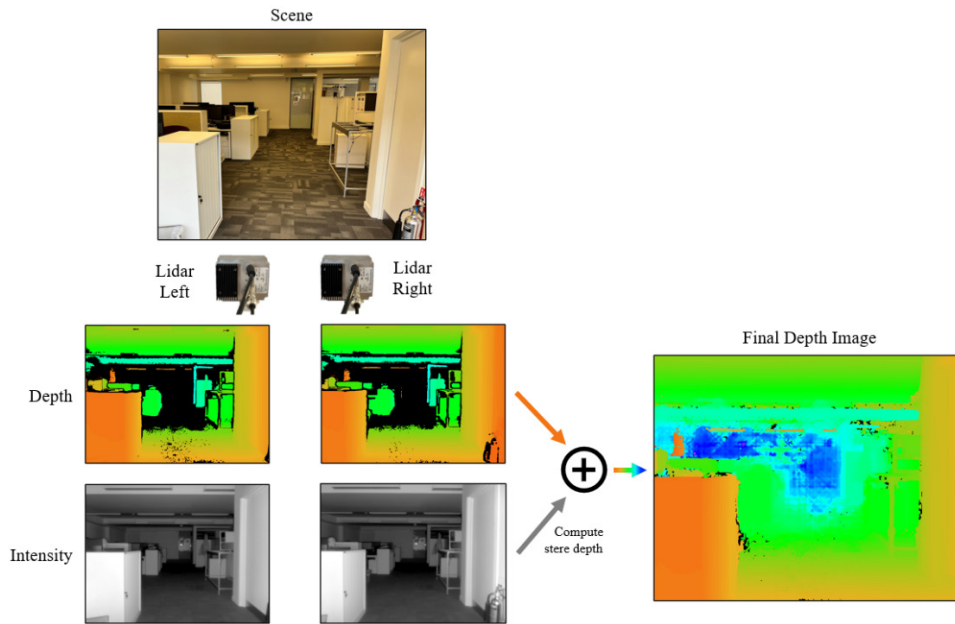
Fig. 79: Illustration of extending depth range of a pair of lidar cameras by processing intensity images. Reproduced from [4].

5.4.1 Practical Setup

A reasonable pixel resolution is required to estimate depth through image processing. As a result, the Basler Blaze-101 iToF solid-state flash lidar is used in this work to provide a pixel resolution of 640×480 at an affordable cost. This lidar has a maximum range of 10 m indoors and operates at a wavelength of 940 nm. Two units are mounted on a rigid dovetail rail to achieve an adjustable baseline of up to 1 m. The final setup is shown in Fig. 80. As with the earlier camera-based approach, the established semi-global matching (SGM) algorithm is adopted. A Bosch GLM 250 VF rangefinder is used to assess the ground truth accuracy of the processed depth estimates.
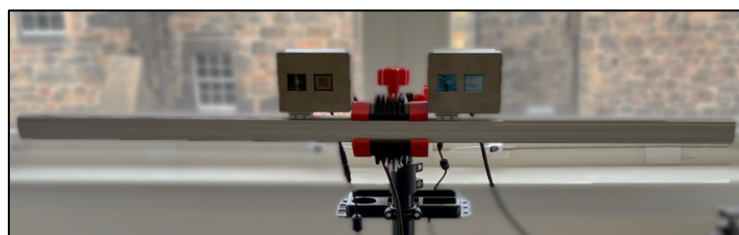
Fig. 80: Setup for lidar-based stereo depth estimation. Adapted from [4].

5.4.2 Process Flow

The process of enhancing the native lidar depth image using stereo depth estimates derived from the lidar intensity data is illustrated in Fig. 81.
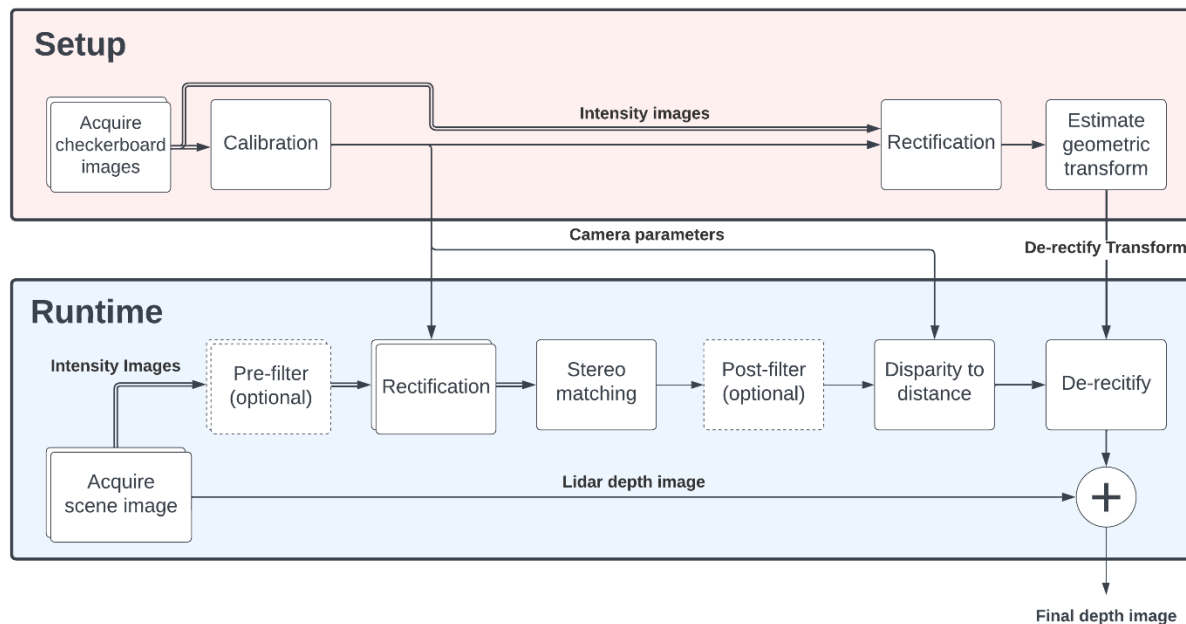
Fig. 81: Processing flow diagram used to extract depth from intensity images. Adapted from [4].

As with the process of estimating stereo depth using conventional imaging cameras (Section 5.2) checkerboard calibration is first performed. However, unlike imaging cameras, a unique challenge in checkerboard calibration using intensity data from a flash lidar is the presence of glare introduced by the flash, as shown in Fig. 82(a). This can be overcome by acquiring numerous checkerboard images at a variety of positions and angles (Fig. 82(b)) until a suitable number of glare-free images have been acquired. A calibration of less than 0.1 pixels of reprojection error was found to give sufficiently accurate camera parameters for stereo-matching.

While rectification is a necessary step in the stereo depth estimation process, the transformation it applies to the acquired images prevents it from being directly superimposed onto the native lidar depth image. To resolve this issue, an optimization function [170] is applied to a sample pair of pre- and post-rectification intensity images to estimate a geometric transform which converts the rectified image points back to the original coordinates. The value of this transform only needs to be computed once during setup and can then be reused during runtime for minimal processing overhead.

Fig. 82: (a) Glare from flash lidar intensity data which can degrade calibration and (b) use of various target angles to capture images without glare. Reproduced from [4].

During runtime, both depth and intensity images are acquired simultaneously. The first step in the process is an optional pre-filter (median) to improve the intensity image quality, degraded by the optical bandpass filter, to assist with stereo-matching. After rectification and stereo-matching (SGM), an optional post-filter can also be applied to replace small patches of missing pixel disparities. Pixel disparities are then converted to their corresponding distance by combining Equation (26) with the camera parameters extracted during calibration. Finally, the stereo depth estimate image is transformed using the pre-established geometric transform and combined with the original lidar depth image.

5.4.3 Performance

The first set of results is comprised of images taken in an indoor environment. Under these conditions, most of the light captured by the sensor is the reflected laser light. To assess the performance of the processed intensity depth within a real-world environment, images of a real human figure have been captured at varying distance away from the lidar cameras. The captured and processed images are shown in Fig. 83.

At the shorter distances, the benefit of the additional processed intensity depth data in situations that are challenging for lidar sensors is apparent in (i) surfaces that are near-perpendicular to the sensors such as floor, ceiling and tables (ii) narrow objects such as chair legs. As the human target approaches 8 m and beyond, the returning laser signal is no longer sufficient for the lidar to measure distance. However, the processed intensity images continue to resolve the distance to the human target.

111
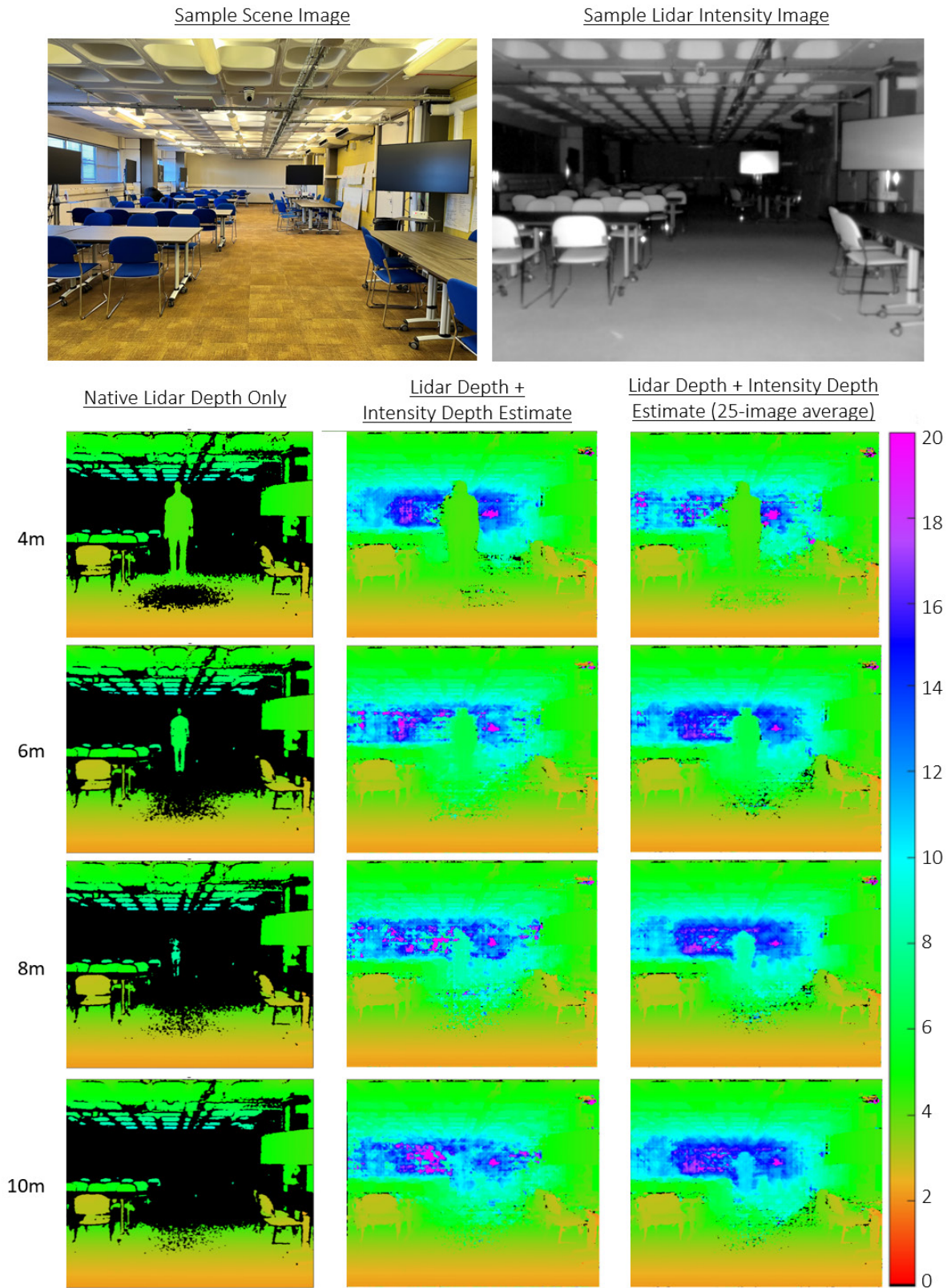
Fig. 83: Combined native lidar depth data and stereo depth estimates from intensity data in an indoor environment. Adapted from [4].

By trading off frame rate, multiple consecutive intensity images can be averaged to extract further detail. The resulting processed depth from averaging 25 intensity frames is shown in the final column of images in Fig. 83. Using this technique, the human figure is resolved as far as 18 m away, increasing the native lidar depth when ranging the human figure by over 2×. The processed intensity depth accuracy of the human figure at each distance step over a 5×5 window of pixels (25 samples) is summarized in Fig. 84. A degradation in accuracy with distance is observed with a root-mean-squared error of 1 m at the maximum distance.
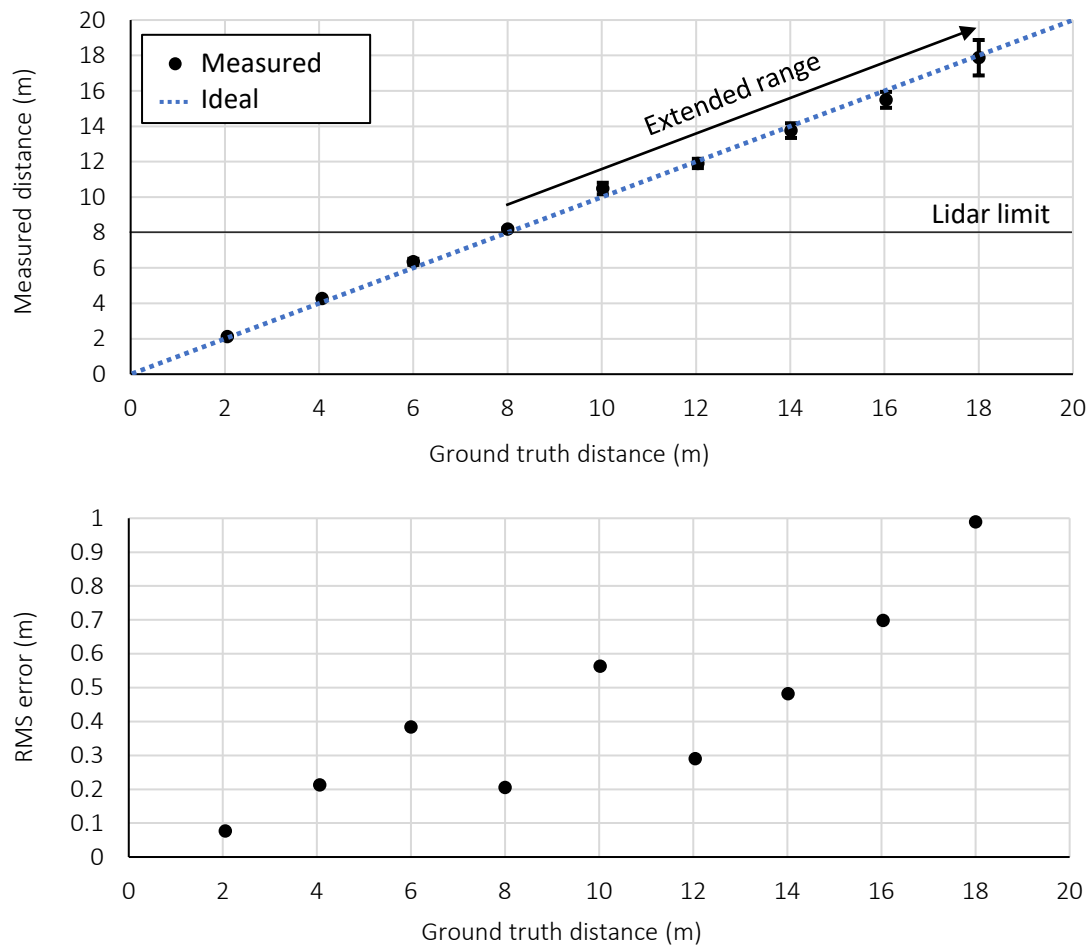
Fig. 84: Stereo depth estimation accuracy using an indoor dataset of lidar intensity data. Adapted from [4].

A second set of results is comprised of images taken in an outdoor environment. Under these conditions, most of the light captured by the lidar is ambient background light. The resulting processed images from this scene are shown in Fig. 85.
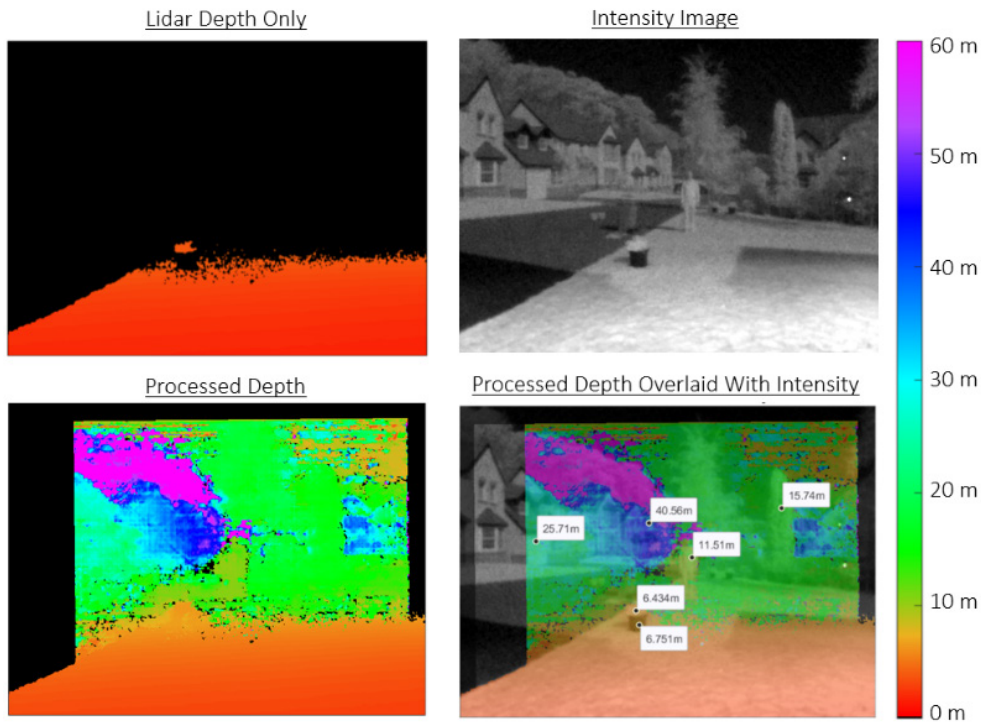
Fig. 85: Combined native lidar depth data and stereo depth estimates from intensity data in an outdoor environment. Adapted from [4].

The challenge of lidar ranging under these conditions compared to the indoor environment is evident by the lack of native lidar depth points beyond even 5 meters. The additional challenge presented to the lidar by dark surfaces is also evident in the lack of depth data points present on the plant pot in the foreground. In this scenario, the additional range data extracted from the intensity images not only fills in the missing plant pot image but also reveals the presence of a human figure at a distance of 11.5 m which is otherwise undetected by the native lidar depth data. Furthermore, the processed intensity data adds depth estimation to buildings as far as 40 m from the target, extending the maximum range in this outdoor scenario by nearly 10×.
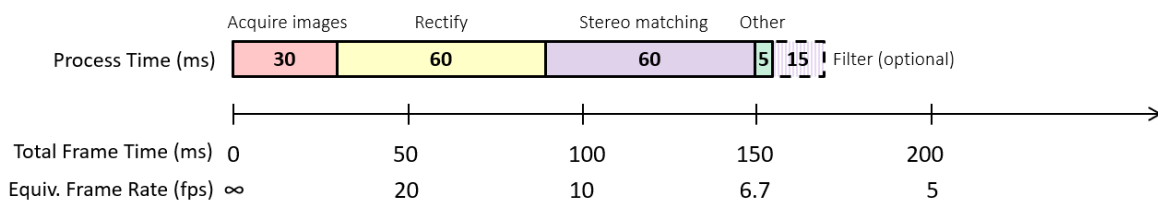
Fig. 86: Time to produce the final depth image with processed intensity data.

A breakdown of processing time for each step in the lidar-based depth estimation frame is provided in Fig. 86. The process adopted here allows additional depth to be extracted from the lidar sensors while achieving a minimum frame rate of 5 fps. While the image acquisition period is a fixed requirement of

the cameras used, further reduction in the total frame time could be achieved through implementing a more effective rectification or stereo matching algorithm (at the cost of potentially degraded stereo depth estimation).

5.4.4 Discussion

This work presents the first application of inherent lidar intensity data to extract additional depth information, increasing the maximum depth range with no additional power consumption or hardware modification. Missing depth information is also successfully extracted from features in the environment such as floors, ceilings and dark objects. By averaging multiple frames of intensity images, the native lidar maximum distance is increased by $2\times$ in an indoor environment and almost $10\times$ outdoors. Evidently, the amount of extra depth information that can be extracted using this technique diminishes with the amount of ambient light available within the wavelength of the lidar filter. However, this is complimentary to the native lidar depth performance which improves under low ambient conditions as seen in Fig. 83 (indoor) vs. Fig. 85 (outdoor).

In the context of guided dToF, the arrangement of two forward-facing short-range lidar and a single long-range lidar depicted in Fig. 72 could open up the possibility of guiding the long-range lidar from stereo depth estimation provided by the short-range lidar intensity data.

5.5 Future Work: Lidar Intensity Monocular Depth Estimation

In the prior sections of this chapter, the ability to extract depth estimation from stereo cameras was extended to using single cameras (monocular depth estimation). Therefore, it stands to reason that the same logic could be applied for lidar-based depth estimation from intensity data, opening up the possibility for self-guided dToF. Such an implementation would require a monocular depth estimation model that is trained on lidar intensity data. This is presently not available. While such training could be conducted, for example by using the native lidar depth data as ground truth, it is beyond the scope of this thesis which aims to develop a proof of concept for the principle of guided lidar. However, this is left as an interesting development for future work.

5.6 Conclusion

This chapter explores both established and novel depth estimation techniques based on imaging to establish a suitable method for guiding the dToF sensor.

To evaluate a camera-based approach, a stereo camera rig is built and tested in an outdoor environment. Leveraging the established SGM algorithm, this setup effectively resolved distances up to the 200 m range required for fully autonomous vehicles. Subsequently, a novel learning-based monocular depth estimation approach using the MiDaS model is explored. This enables depth estimation using only a single camera. However, both depth accuracy and processing time performance currently prevent the integration of this approach into a practical guided dToF system.

The final approach adopts a novel method of deriving depth information from lidar intensity data. Using stereo depth estimation techniques, it is shown that the native lidar maximum distance can be significantly enhanced. However, the optical bandpass filter in lidar reduces the total incoming light which necessitates long exposure times for useable images. As with the camera-based approach, it is proposed that monocular depth estimation may also be realisable with lidar intensity images, enabling self-guided lidar. However, a neural network trained on such data would need to be established.

A summary of the different approaches is provided in Table 13. In the pursuit of developing the first guided dToF demonstrator, a balance between the quality of depth estimation and the simplicity of implementation is paramount. As a result, the camera-based approach using traditional stereo depth estimation techniques is adopted for the guided dToF developed in this work.

Attribute	Stereo (traditional)	Monocular (learning)	Lidar intensity
Estimate density	Medium	High	Medium
Distance accuracy	Medium	Low	Low
Processing time	Fast	Slow	Medium

Table 13: Performance summary of explored image-based techniques for integrating into the guided dToF lidar demonstrator.

Having established both key parts of a guided dToF system (i) the guidable sensor and (ii) the source of depth estimates, integration and evaluation of the full guided dToF system can begin.

6 GUIDED DToF SYSTEM

6.1 Introduction

This chapter presents the integration and evaluation of the first guided dToF system; combining both the guided dToF sensor and stereo camera depth estimation developed in the preceding chapters. The setup of the lidar and stereo camera rig are first covered separately, before integrating the two together to implement the full guided dToF system. The operation of the system is then evaluated under a variety of practical test conditions to determine the feasibility of guided dToF for self-driving vehicles.

6.2 Preface

Chapter 4 outlined both the primary (QuantIC4×4) and contingency (HSLIDAR) choice of dToF sensor for this project. Due to unresolved technical hurdles, the primary sensor was not operational within this project's timeframe. As a result, the system implemented here uses the contingency sensor HSLIDAR.

6.3 Setup

6.3.1 Lidar

Imaging System

As discussed in Section 2.2.3, guided dToF can be applied to any type of lidar imaging system where the sensor is a 2D array of dToF pixels. To remove the added complexity of scanning, the guided dToF demonstrator developed in this work is built around a flash lidar system.

Laser

The laser inherited for this project has a wavelength centred around 940 nm. To ensure the dToF sensor is configured appropriately for the laser, the pulse shape is measured using a fast photodiode connected to an oscilloscope. The resulting pulse shape is plotted in Fig. 87 and measured to have a full width half maximum (FWHM) of around 4.5 ns.
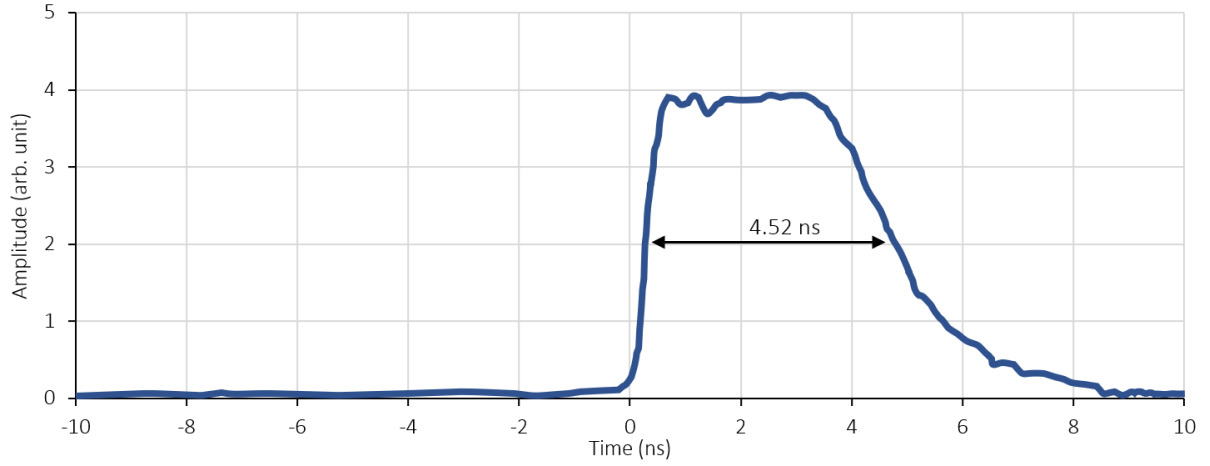
Fig. 87: Measured pulse profile of the 940 nm laser source integrated into the guided dToF system.

Sensor

Given the measured laser FWHM of around 4.5 ns, the optimal bin interval (a) to configure the sensor histogram as per Equation (22) is:

$$a = 0.64 \times FWHM = 2.9 \ ns$$

Recall from Section 4.9 that HSLIDAR can be operated in one of two TDC modes, as laid out in Table 10 (i) clocked (ii) delay line (DL). In the first mode the histogram bin time intervals are defined by the period of an external clock signal and in the second they are defined by the timing between a line of delay elements in each individual macropixel. While a clocked TDC achieves histogram bins with highly uniform intervals, the sensor is limited using a maximum external clock frequency of 200 MHz, limiting the minimum bin interval in this mode to 5 ns which exceeds the target interval of 2.9 ns. As a result, the sensor must be operated in DL mode.

Operating the sensor in DL mode presents two key challenges. The first is in setting the desired bin interval, since there is no analytical relation between the global control voltage which supplies the delay elements and the resulting delay. Secondly, delay elements are sensitive to mismatch, producing nonuniform delay times across each bin and every pixel, as illustrated in Fig. 88. Since the average desired bin interval is in the order of a couple of nanoseconds, even small variation between delay elements will significantly degrade the final measured laser return time. Therefore, thorough characterisation of every histogram bin interval in each pixel is undertaken to determine the appropriate control voltage for the required bin interval and calibrate out the nonuniformity.

118
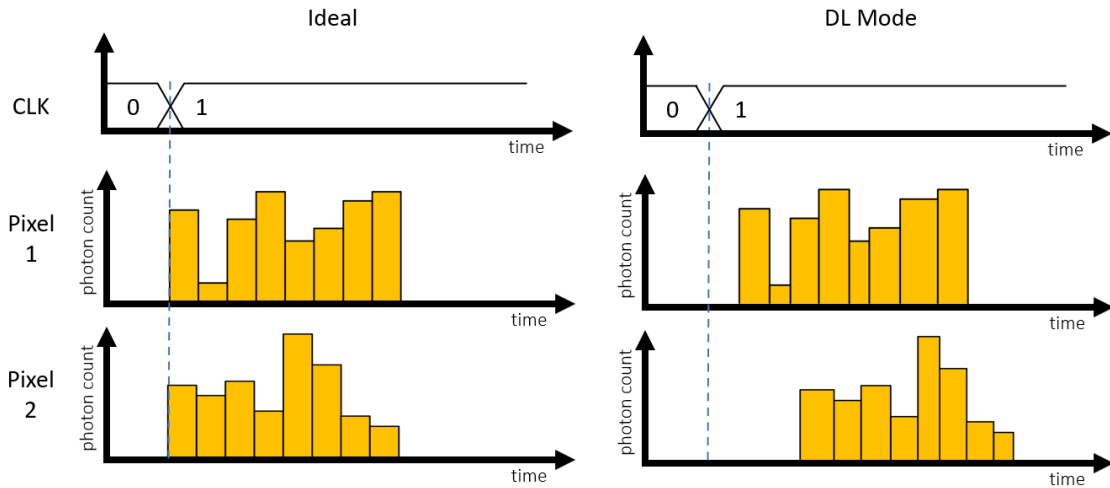
Fig. 88: Illustration of ideal vs. real histogram timing when configured using a delay line.

DL mode calibration was conducted using the setup pictured in Fig. 89. It consists of the lidar sensor, a delay generator (Stanford DG645) and a picosecond pulse laser driver (Picoquant Taiko PDL M1) with a 940 nm laser head (Picoquant LDH-IB-940-B) pointed directly at the lidar sensor. The setup is automated such that the sensor triggers the laser source via the delay generator; the resulting histogram for each pixel is logged; and the delay generator is incremented by 10 ps. The cycle is repeated until the full histogram window time for every pixel has been covered.
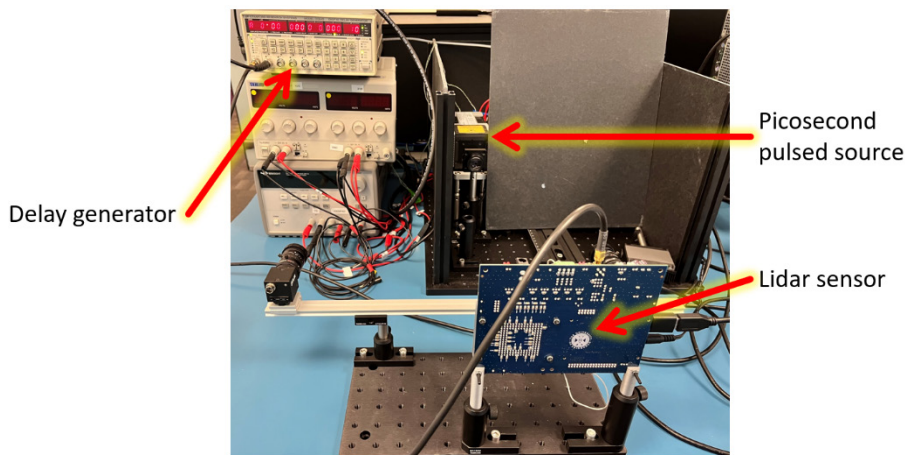
Fig. 89: Setup used to characterise every bin interval in each pixel by sweeping a picosecond pulsed laser source.

After conducting a few iterations to establish close to the targeted bin interval, the full calibration sweep was performed. The identified peak bin in every pixel at various intervals during the calibration is shown in Fig. 90. As well as showing the peak bin incrementing roughly at the expected interval, the sweep shows a notable delay across the sensor from left to right as a result of clock routing.
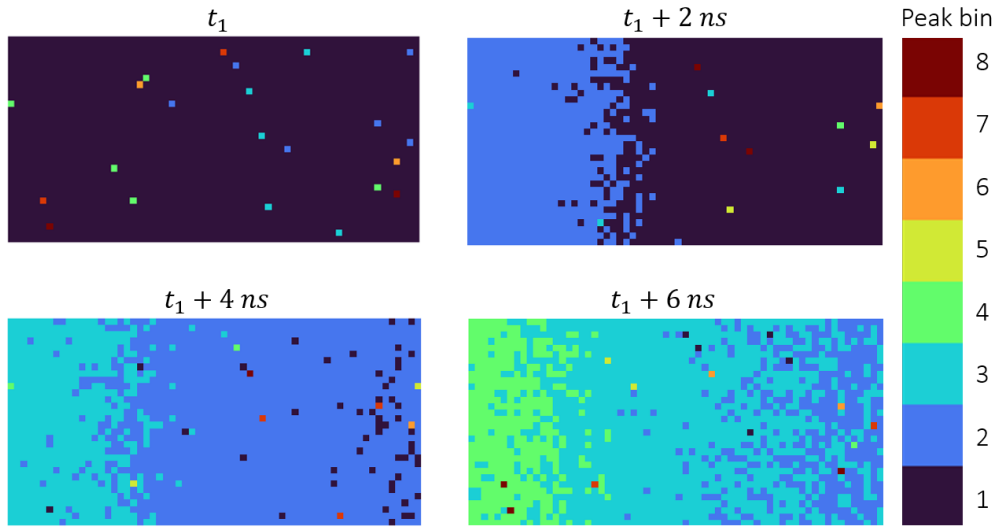
Fig. 90: Sweeping the laser delay and identifying the peak bin to characterise the bin interval of every bin in every pixel.

Once the sweep is performed, the time interval of every bin in each pixel is saved as a calibration profile to be used as a reference during histogram peak interpolation. Note that subtleties in the sensor chip design produce different delay profiles for even and odd-numbered time windows, as a result two calibration profiles are generated to cover both window types. An overview of all the bin interval values across every pixel in the sensor at the established delay line control voltage is shown in Fig. 91. The plot confirms the achieved average bin interval to be 2.6 ns value (equivalent to 0.39 m), while the spread of 0.4 ns single standard deviation highlights the importance of conducting calibration.
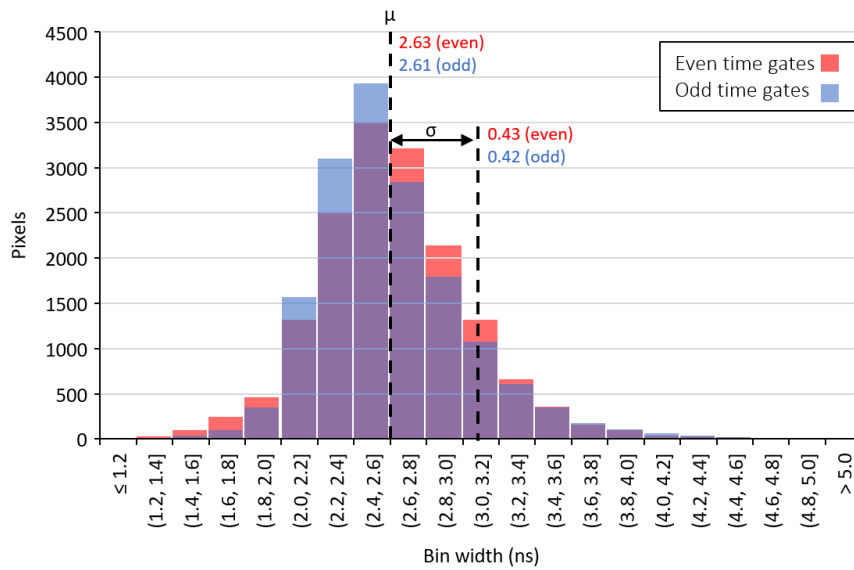
Fig. 91: Bin interval values across every pixel of the dToF sensor at the chosen delay line control voltage.

120

This calibration data (containing the delay time between the laser trigger and the mid-point time of each bin interval in every pixel) corrects the final measured depth during runtime by using it as a correction factor C_i in the histogram interpolation algorithm. The basic interpolation algorithm shown in Equation (28) interpolates the precise bin value for any M-bin histogram with a photon count B_i in each i bin. By adding a correction factor C_i containing the actual bin mid-point time vs. laser trigger for a given pixel, the calibrated peak photon arrival time can be obtained using Equation (29).

Uncalibrated peak arrival time (in bins) $\qquad t = \dfrac{\sum_{i=1}^{M} B_i \times i}{\sum_{i=1}^{M} B_i}$ (28)

Calibrated peak arrival time (in seconds) $\qquad t = \dfrac{\sum_{i=1}^{M} B_i \times C_i}{\sum_{i=1}^{M} i}$ (29)

Note that for simplicity the above equations consider all bins for a given histogram. In practice, only a pre-determined number bins either side the detected peak bin needs to be considered.

To illustrate the value of calibrating the sensor in DL mode based on each individual histogram bin interval across all pixels, the lidar is used to image a flat surface at a distance of 2.5 m. The resulting point cloud is shown in Fig. 92.
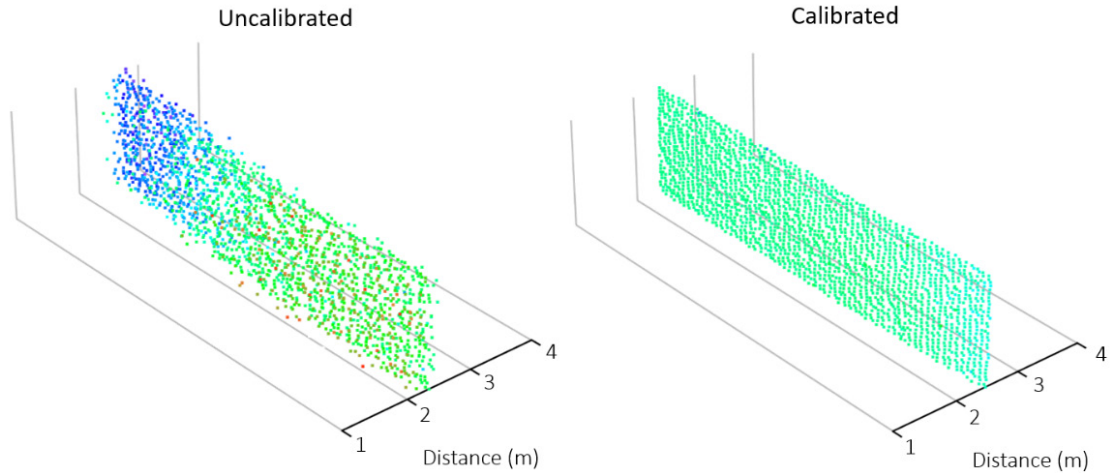
Fig. 92: The impact of calibrating the dToF sensor on the final depth image quality when used in DL mode.

To further illustrate the effectiveness of this calibration, the histogram of distance values measured by each pixel is shown in Fig. 93. Before calibration, the measured distance to the 2.5 m target exhibits a standard deviation of 0.175 m (7%), whereas after calibration this is reduced to only 0.029 m (1%).
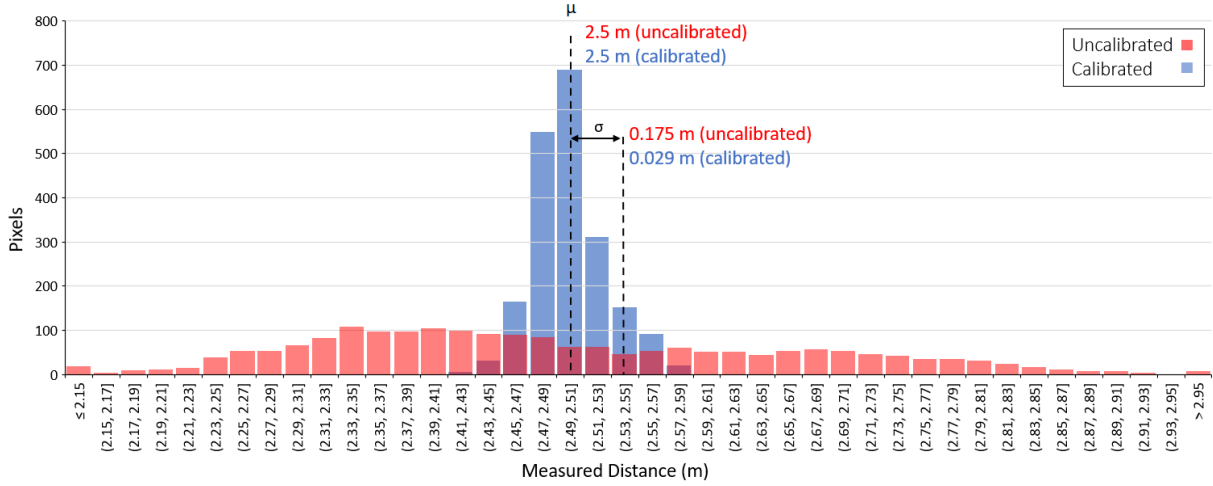
Fig. 93: Distance measured by each pixel to a 2.5 m target, before and after calibration.

Optics

The large pixel area (114 µm × 54 µm) combined with the limited pixel resolution (64 × 32) of the dToF sensor limits the practical focal length options for the optics in front of the sensor. While a short focal length would allow for a wider FOV, this would severely limit the size of objects that can be resolved. A focal length of 25 mm is chosen, equivalent to a 16° × 4° FOV.

Photon Budget Characterisation

Before its integration into the complete guided dToF system, standalone characterization of the flash lidar is first carried out to determine the achieved photon budget. This is performed using a 1 m^2 Lambertian target calibrated to 10% reflectivity as pictured in Fig. 94(a) under illuminance of 60 klux. A lidar acquisition performed with the laser turned off allows the background to be estimated for a given exposure time, while a follow-up acquisition with the laser on allows the signal count to be estimated when subtracted from the prior established background count. This is performed 20 times to determine an average value and assessed over a 3×3 pixel window at various target distances.

Given the sensor macropixel pitch of 54 µm (Section 4.9) and the chosen 25 mm focal length, the distance at which the large 1 m^2 Lambertian target becomes smaller than a projected macropixel area as given by Equation (11) is at over 400 m. Characterisation was conducted up to a distance of 30 m which ensured a wide enough range of measurements while leaving ample margin to keep the target within a projected macropixel area.

The results are summarised in Fig. 94(b). As per the photon budget model outlined in Section 3.2.1, the background return rate (4.8 Mcounts/s) remains independent of distance (z) and the signal photon return rate follows an inverse square law, with a trendline fitted to $89z^{-2}$ photons per laser cycle.

<div align="center">(a)   (b)</div>
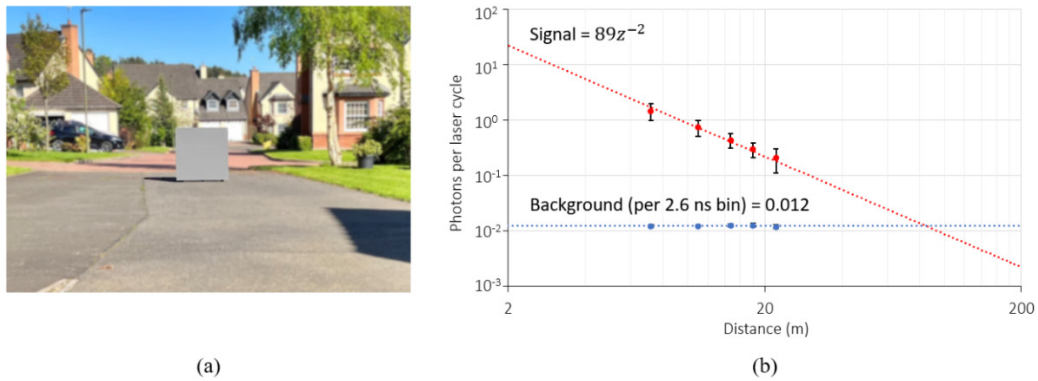
Fig. 94: (a) Setup for characterising the photon budget of the implemented flash lidar and (b) measured signal and background photon budget at 60 klux.

6.3.2 Stereo Cameras

Recall that the processing time to perform stereo depth estimation is proportional to the pixel resolution of the images used (Section 5.2). Since the FOV of the stereo cameras (23° × 18°) is wider than the lidar (16° × 4°), there is scope to optimise depth estimation by reducing the processed image area, as illustrated in Fig. 95. Furthermore, the larger area of the projected lidar pixels means that many stereo camera pixels accommodate the same projected area. By applying pixel binning (2 × 2), the resolution of processed images is further reduced. The final result is an 8× reduction in stereo depth estimation, boosting the maximum achievable frame rate to 16 fps. Moreover, this also reduces the total data and processing time for subsequent steps in the guided dToF process including pixel mapping which is discussed in the next section.

<div align="center">(a)   (b)</div>
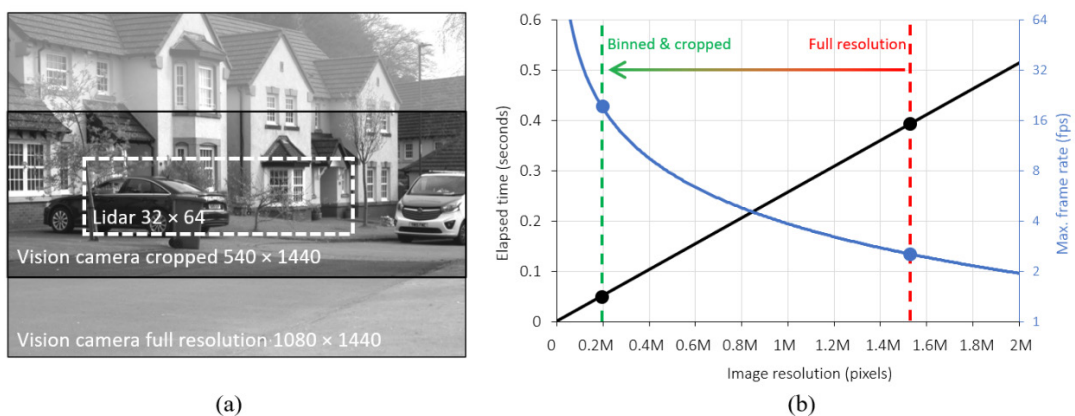
Fig. 95: Reducing the stereo camera pixel resolution to optimise processing time. Adapted from [3].

6.3.3 System

Pixel Mapping

Once depth estimates of the scene have been acquired from the stereo camera, they must be mapped onto each individual pixel of the lidar sensor to guide it to the appropriate depth window. This process is illustrated in Fig. 96.
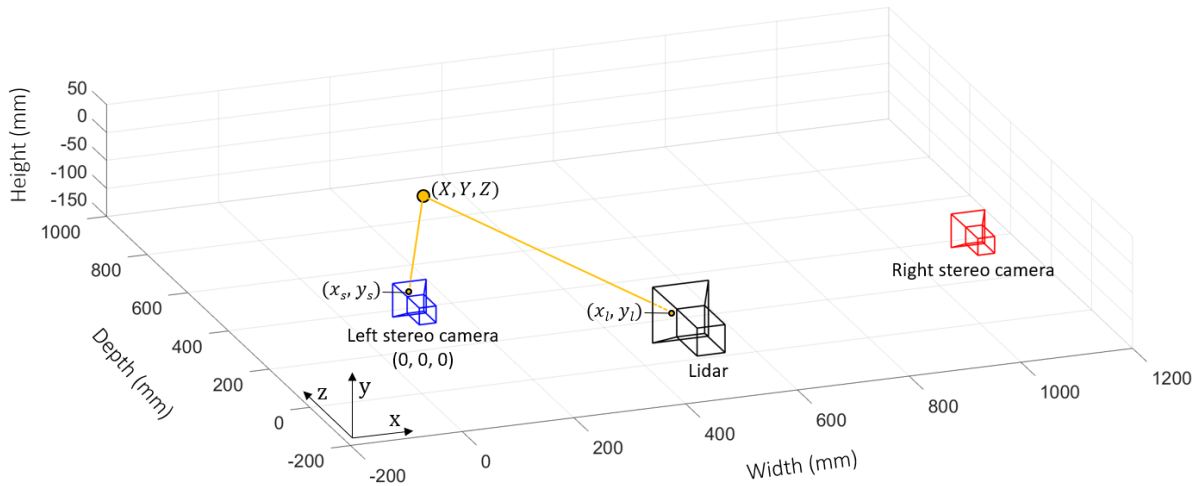
Fig. 96: Pixel mapping from one stereo camera to the corresponding lidar sensor pixel. Reproduced from [3].

Camera calibration is first used to determine the critical intrinsic and extrinsic camera parameter matrices which describe the physical attributes of the camera and their positions between one another, as described in Section 5.2.2. For the lidar sensor, the checkerboard image can be captured by configuring it for photon counting and using the resulting intensity images, as described in Section 5.4.1. Once calibrated, the following parameters are known:

- K_S 3×3 intrinsic camera matrix of the principal stereo camera
- K_l 3×3 intrinsic camera matrix of the lidar sensor
- R_l 3×3 camera rotation matrix of the lidar sensor with respect to the principal stereo camera
- T_l 3×1 camera translation matrix of the lidar sensor with respect to the principal stereo camera

The pixel mapping process is then performed in two steps: (i) map each pixel coordinate in the stereo depth image (x_s, y_s) to its corresponding world coordinate (X,Y,Z) and (ii) map each world coordinate to the corresponding lidar camera pixel coordinate (x_l, y_l). In the first step, the inverse intrinsic camera matrix of the principal stereo camera K_l is multiplied by the camera coordinate to give a normalized

124

world coordinate. The world coordinate can then be scaled appropriately by multiplying by the distance Z to that point which has already been estimated by the stereo depth algorithm.

$$K_s^{-1} \times \begin{pmatrix} x_s \\ y_s \\ 1 \end{pmatrix} = \begin{pmatrix} \tilde{X} \\ \tilde{Y} \\ 1 \end{pmatrix} \tag{30}$$

$$Z \cdot \begin{pmatrix} \tilde{X} \\ \tilde{Y} \\ 1 \end{pmatrix} = \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \tag{31}$$

In the second step, the lidar camera's extrinsic matrix P_l is multiplied by the prior calculated world coordinated.

$$P_l \times \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} = \begin{pmatrix} x_l \\ y_l \\ 1 \end{pmatrix} \tag{32}$$

$$\text{where } P_l = K_l \times [R_l \, T_l]$$

In the case where multiple camera pixels map to one lidar pixel, the modal pixel value can be taken. In this work, duplicates are discarded to optimise processing time.

Process

An overview of each key step in the guided dToF processes is illustrated in Fig. 97. These are:

1. Acquire images from the stereo camera rig
2. Rectify images and process stereo depth estimates
3. Pixel map each depth estimate to the corresponding dToF sensor pixel
4. Update the time window of each dToF pixel based on the estimated distance
5. Cycle the laser to build up a histogram in each pixel and interpolate the precise target distance
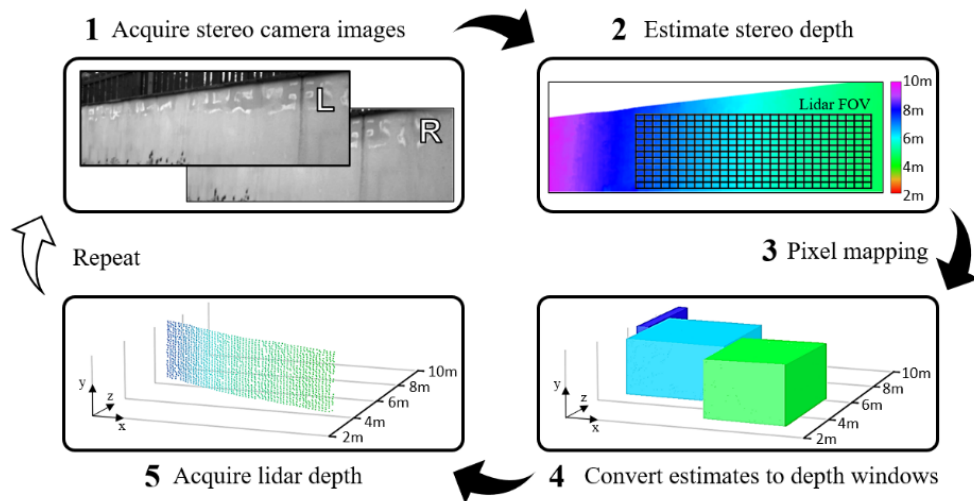
Fig. 97: Overview of key steps in the guided dToF implementation. Adapted from [2].

Final Assembly

The culmination of the research presented here is the first guided dToF demonstrator shown in Fig. 98. The setup runs on a 1.9 GHz Intel Core i7 8th generation laptop, with a Bosch GLM250VF rangefinder providing ground truth distance for benchmarking. A summary of the system is provided in Table 14.
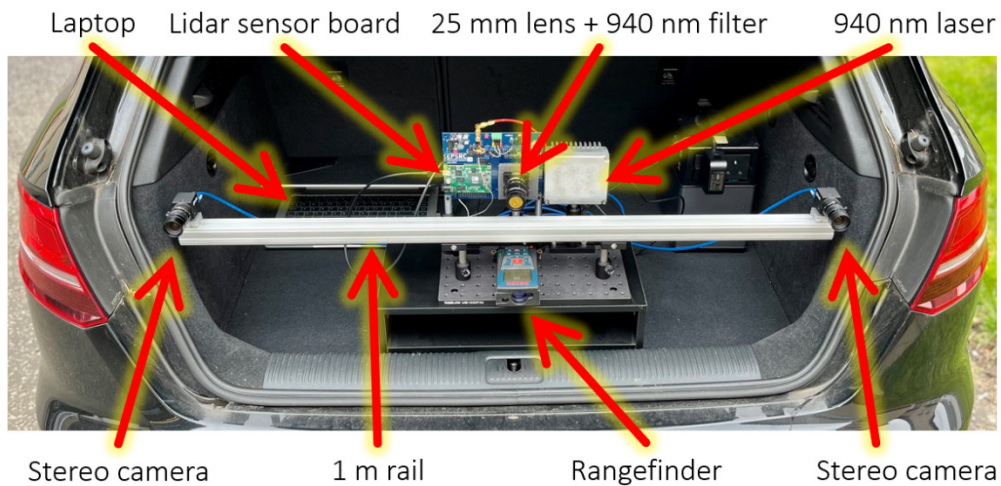
Fig. 98: The first working guided dToF demonstrator. Adapted from [2].

Component	Parameter	Value
Lidar	Laser pulse width	4.5 ns FWHM
	Laser repetition rate	80 kHz
	Wavelength	940 nm
	Filter bandwidth	10 nm FWHM
	Pixel resolution	64 × 32
	Focal length	25 mm
	Field of view (H×V)	16° × 4°
	Histogram bins	8×12-bit
	Histogram bin interval	0.39 m (2.6 ns)
	Histogram window step	1.875 m (1.25 ns)
Stereo Cameras	Baseline	1 m
	Camera type	BFS-U3-16S2M-CS
	Maximum resolution	1080×1440
	Focal length	12 mm
	Field of view (H×V)	23° × 18°

Table 14: Summary of components and values which make up the guided dToF system.

6.4 Results

To evaluate the guided lidar demonstrator, various dynamic scenarios are captured with the system running in real-time at 3 fps. Performance is then quantitively evaluated along with the resulting laser power reduction over equivalent partial histogram approaches.

6.4.1 Scenarios

Outdoor Clear Conditions

The first test scenario is conducted under clear daylight conditions where the illuminance levels reached 15 klux. This scene captures a van as it moves away from the lidar, and Fig. 99 showcases the various sub-processes within a single frame of guided lidar data from this environment. Configuring the lidar to a step size of 1.875 m, around half the full histogram width, ensures a seamless depth map across the van even as it spanned multiple time windows.
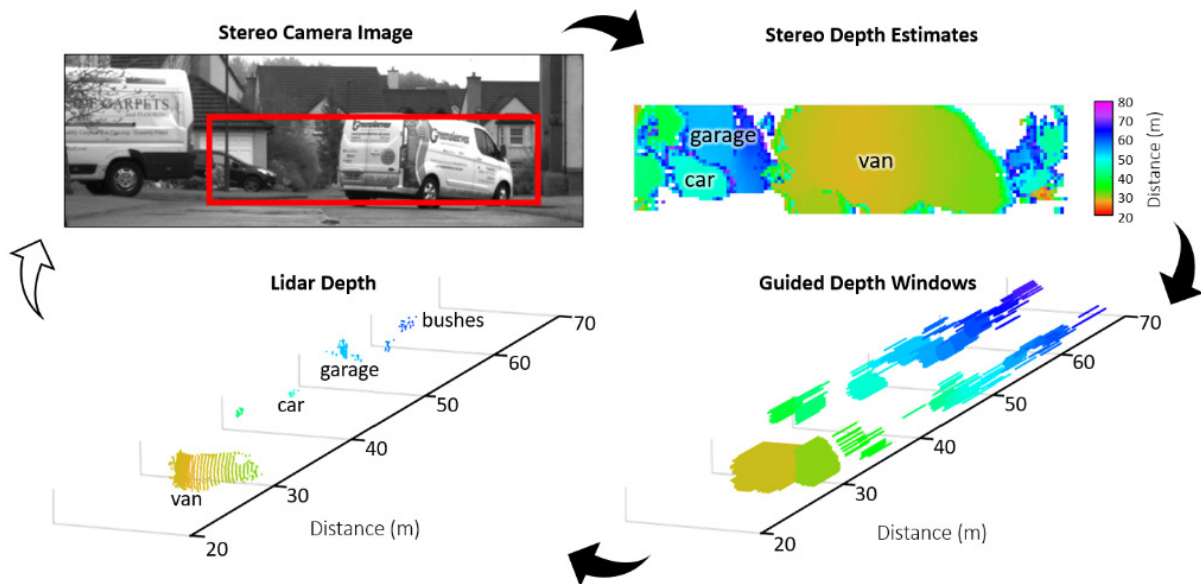
Fig. 99: Each sub-process within a single guided lidar frame. Adapted from [3].

Fig. 100 shows the subsequent frames captured from the same scene. In each frame, the illustration depicts the histogram and guided time window of a sample lidar pixel, demonstrating how the pixel continues to update accurately as the van gradually distances itself from the setup. The guided lidar system successfully tracked and resolved the distance to the van at distances extending up to 75 m.
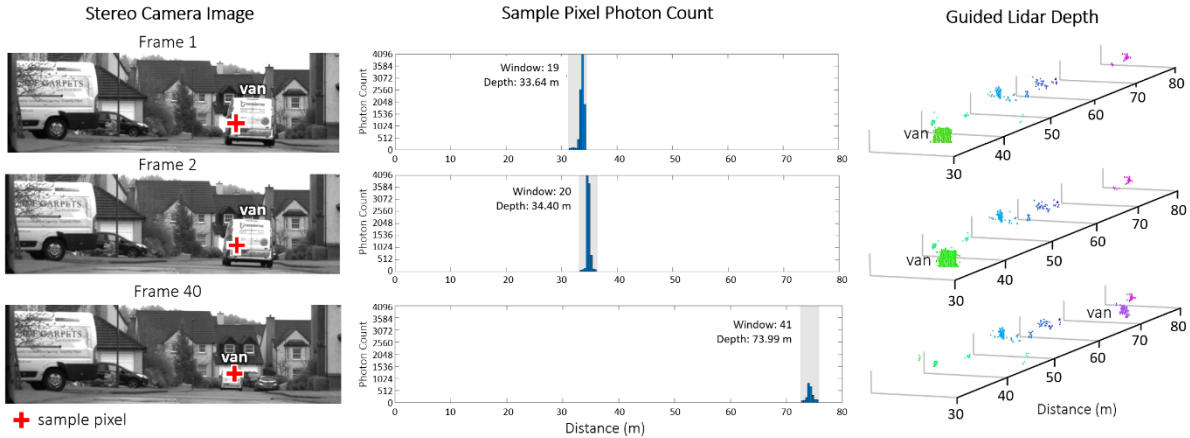
Fig. 100: Three further frames of the scene in Fig. 99 including the configured time window of a sample lidar pixel and the resulting histogram produced. Adapted from [2].

Outdoor Foggy Conditions

The presence of fog poses a significant challenge for lidar systems. It not only diminishes the intensity of returning laser signals but also generates early laser returns as a result of fog-induced reflections [92]. Fig. 101 presents a scene captured under foggy conditions, featuring both a moving pedestrian and a car. This scene showcases how the system copes with these adverse weather conditions. Notably, it depicts the time window of a lidar pixel accurately updating while observing the pedestrian and separately tracking the car's distance. Even in these challenging conditions, the guided lidar system managed to resolve the car's distance up to 60 m.
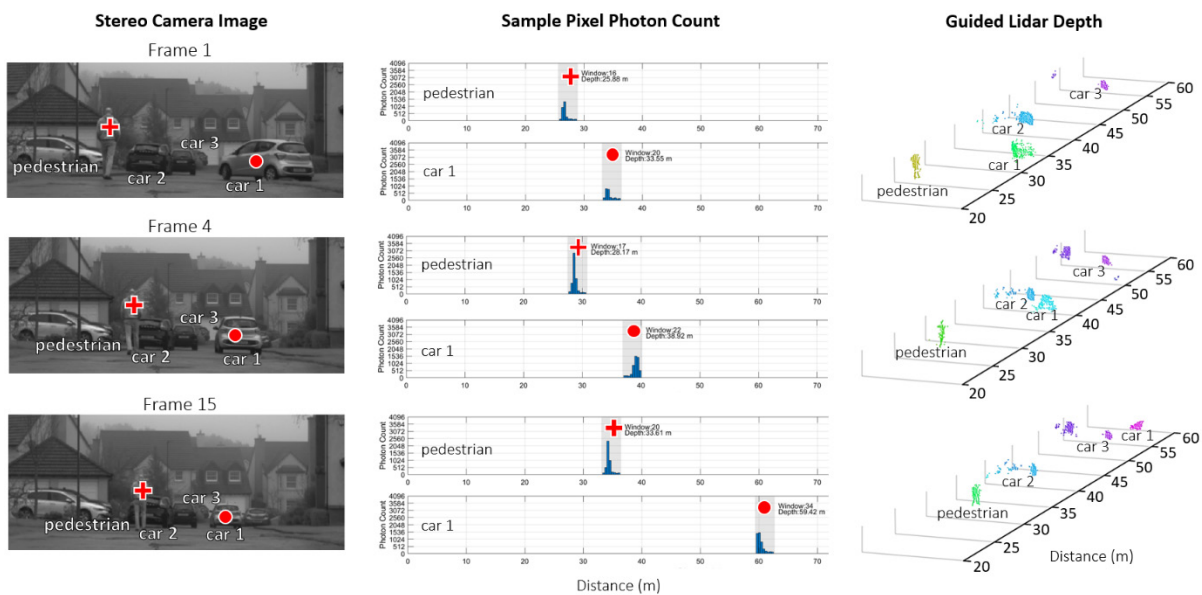
Fig. 101: Three frames from a scene on a foggy day. Sample lidar pixels show which time windows they are configured to and the resulting histogram produced. Adapted from [3].

128

Transparent Objects

Transparent objects such as glass, introduce unique challenges for lidar systems due to the multipath reflections they generate [171]. This issue can be particularly problematic for techniques like partial histogram zooming, which favour the first signal peak. In the scene presented in Fig. 102(a) featuring a human figure behind a glass door, the point cloud resulting from evaluating only the first peak is obscured by the door. Using guided dToF, Fig. 102(c) demonstrates how each lidar pixel is appropriately guided to resolve the presence of a human figure behind the glass door, highlighting the effectiveness of guided dToF in addressing scenarios involving transparent objects.
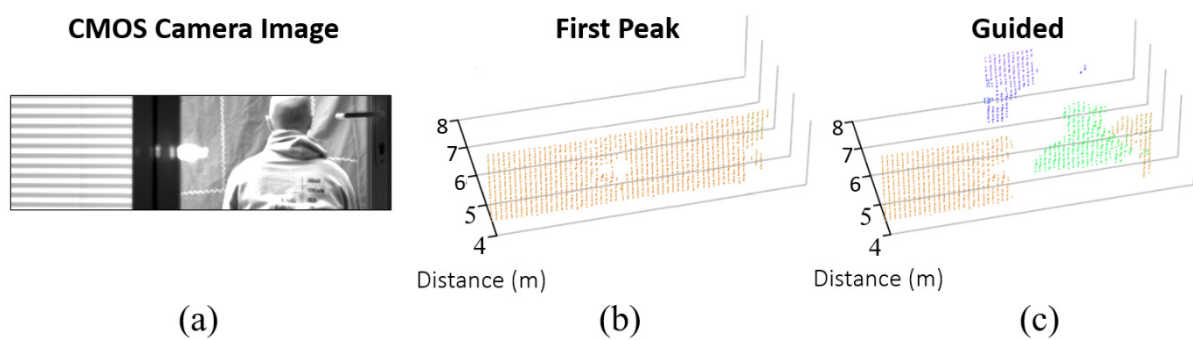
Fig. 102: (a) Scene of a human figure behind a glass door (b) obscured when using only
the first lidar peak (c) revealed using guided dToF. Adapted from [2]

6.4.2 Performance

Measurement Error

To quantitatively assess the performance of the guided dToF demonstrator, the measured distances to a human target are compared with the ground truth distance obtained from the rangefinder. This evaluation was conducted by analysing a window of 3×3 pixels across 9 frames, providing a total of 81 sample points at each distance step up to 50 m. The experiment took place outdoors under an illuminance level of 72 klux and the results presented represent the real system performance while running in real time at the nominal 3 fps.
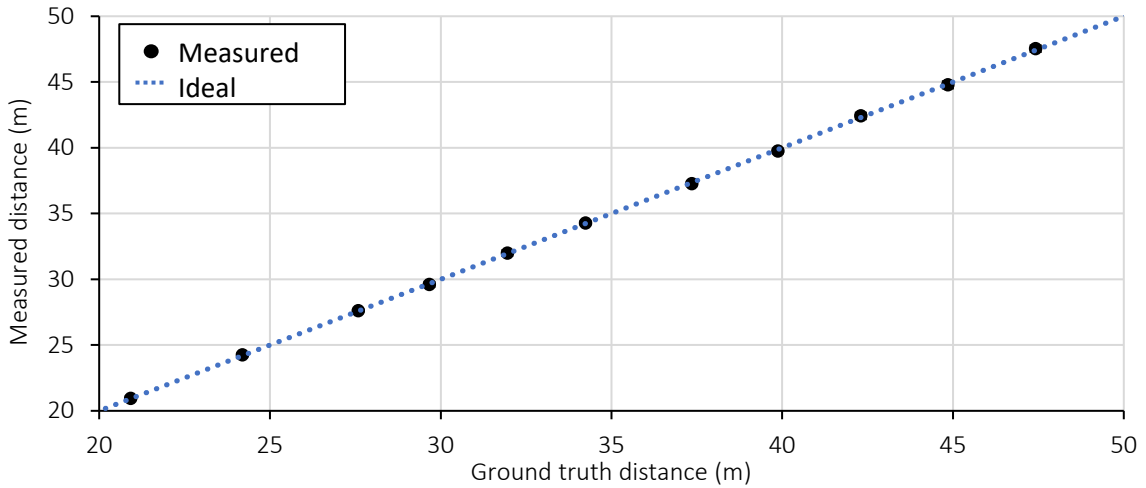
Fig. 103: Measured guided dToF distance mean and variance (3 standard deviations) at each step in the distance sweep test over a sample window of 3×3 pixels.
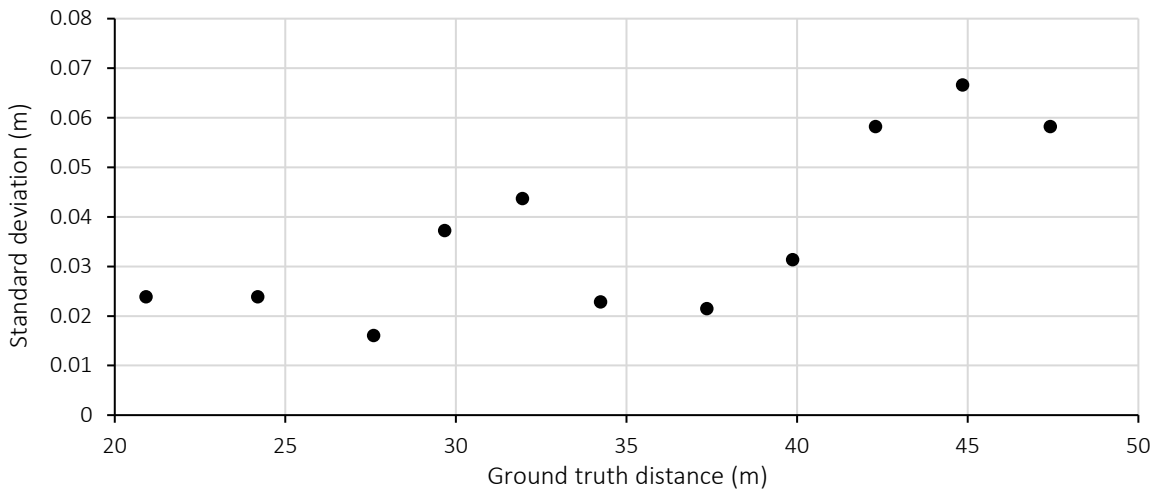
Fig. 104: Measured guided dToF standard deviation at each distance.
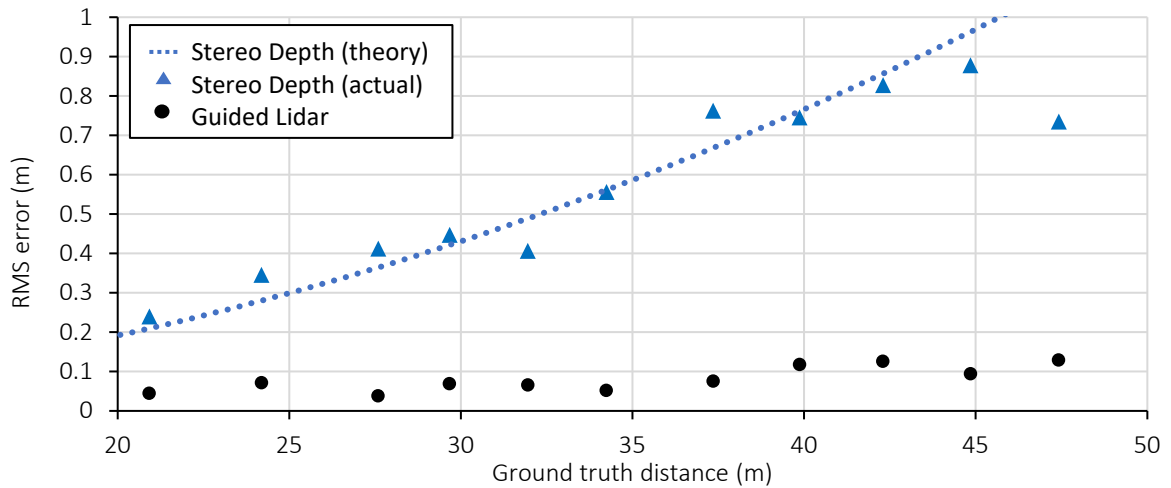
Fig. 105: Guided dToF and stereo depth estimations RMS error, along with theoretical stereo depth error based on to Equation (2) for sub-pixel disparity of ¼ pixels.

The mean dToF measurement of all sample points at each distance, along with ± 3 standard deviations are plotted in Fig. 103. The single standard deviation value is also plotted separately in Fig. 104 for clarity. The measurements across pixels and frames are narrowly distributed, exhibiting a standard deviation of around 0.1% of the measured distance at the maximum distance.

The RMS error of all dToF measurements at each distance are summarised in Fig. 105. The error of the guiding stereo depth estimates is also included, exhibiting the characteristic squared increase in error. However, the error is sufficient to guide the dToF sensor, resulting in no more than 0.15 m RMS error up to 50 m.

Processing Time

The execution required for each of the most time-intensive processes within one frame of this guided dToF implementation is outlined in Fig. 106. Before the lidar exposure period, the total execution time of all other processes is around 145 ms, equivalent to a maximum frame rate of 7 fps. Within this, there is scope to further reduce the execution time by vectorising some processes within the pixel mapping algorithm. Furthermore, there exists many more modern stereo depth estimation algorithms which may provide faster processing time than the traditional SGM algorithm used in this design. The subsequent lidar exposure period of 200 ms allows the system to achieve the long-range outdoor performance described above, reducing the frame rate to 3 fps.
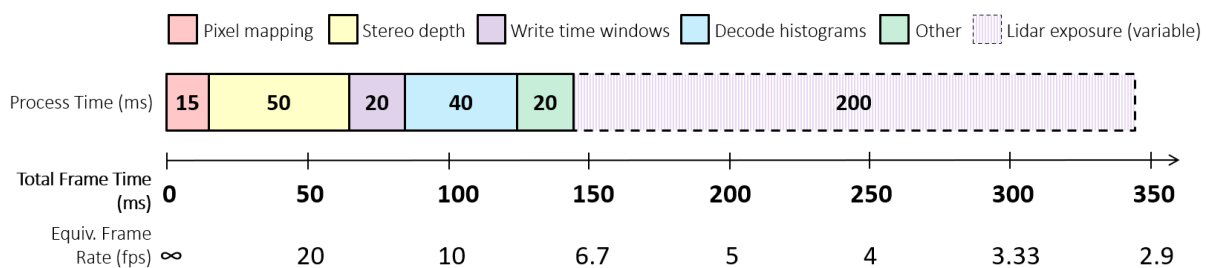
Fig. 106: Execution time of processes within one frame of this guided dToF implementation.

Motion Artefacts

When assessing the standalone performance of the implemented stereo camera rig in Section 5.2.4, it was shown that even sub-degree variations in the camera angles post-calibration result in significant depth estimation error. For a target at 40 m, Fig. 70 showed the rig to exhibit a 1-2 m error from even

a 0.1° mechanical deviation. This has been reproduced in Fig. 107, causing the camera rig to overestimate depth and a pedestrian moving between depth windows to be allocated the incorrect depth window. As a result, the pedestrian disappears between frames.
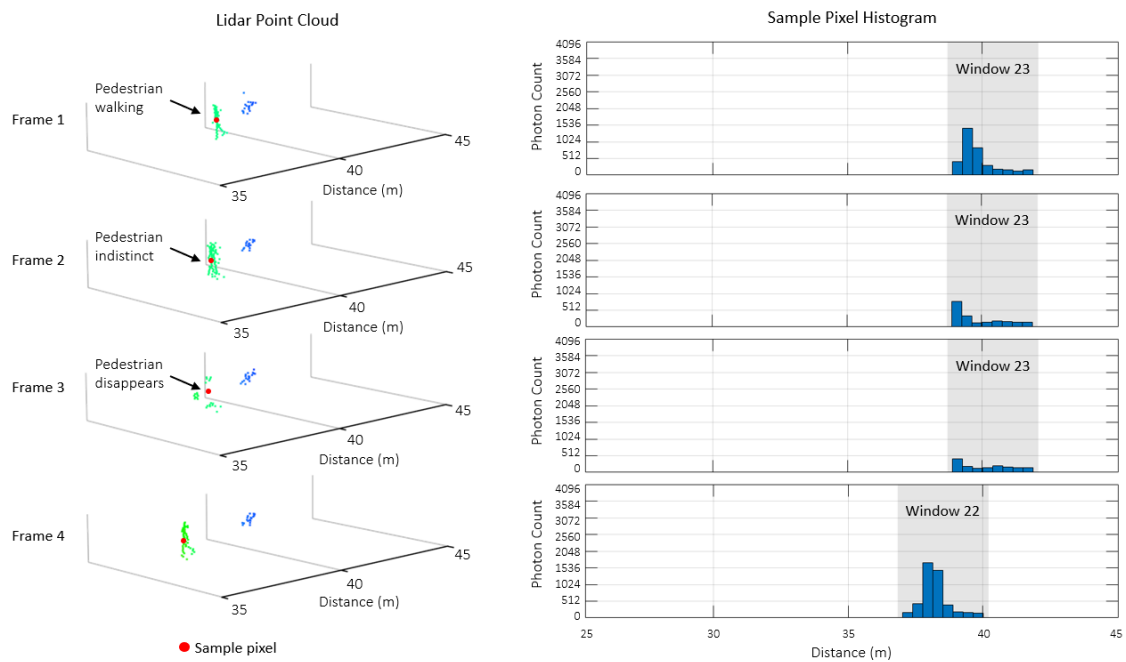
Fig. 107: Stereo cameras over-estimating depth due to sub-degree mechanical deviation in camera angle. This causes the dToF sensor to be guided to the incorrecr depth window as the pedestrian moves closer.

Laser Power Efficiency

In Section 3.4 of the modelling chapter, the laser power penalty incurred using a partial histogram approach instead of a full/guided histogram was evaluated on an example design specification. The same methodology can be applied using the actual photon budget of the implemented guided dToF lidar measured in Section 6.3.1. For convenience, the relevant parameters are collected in Table 15.

Parameter	Value
Target precision	10 cm
Target probability of detection	99.7%
Histogram bins	8
Histogram bin interval (full/guided/sliding)	2.6 ns
Laser pulse width (FWHM)	4.5 ns
Signal rate	$89z^{-2}$ counts/pulse
Background rate	8 Mcounts/s

Table 15: Parameters relevant for modelling zooming and sliding laser power penalty.

The Thompson model determines the minimum required number of laser cycles for a full histogram approach as given by the plot in Fig. 108. From this, the minimum number required to implement a sliding approach (assuming no overlap between windows) is directly related by using the same methodology outlined in the modelling chapter (Section 3.4). For a zooming approach, a second zoom step is required for distances over 3 m (using a bin interval of 20.8 ns) and a third zoom step is required for distances over 25 m (using a bin interval of 166 ns). Passing these through the probability of detection model and combining with the full histogram requirement results in the minimum number of laser cycles for an equivalent zooming approach as given in Fig. 108. By evaluating the increase in laser cycles compared to a full histogram approach, the laser power penalty of equivalent partial histogram approaches is plotted in Fig. 109. To operate at the 75 m distance achieved by the guided dToF system, an equivalent partial histogram approach is shown to require a minimum of 6× more laser cycles.
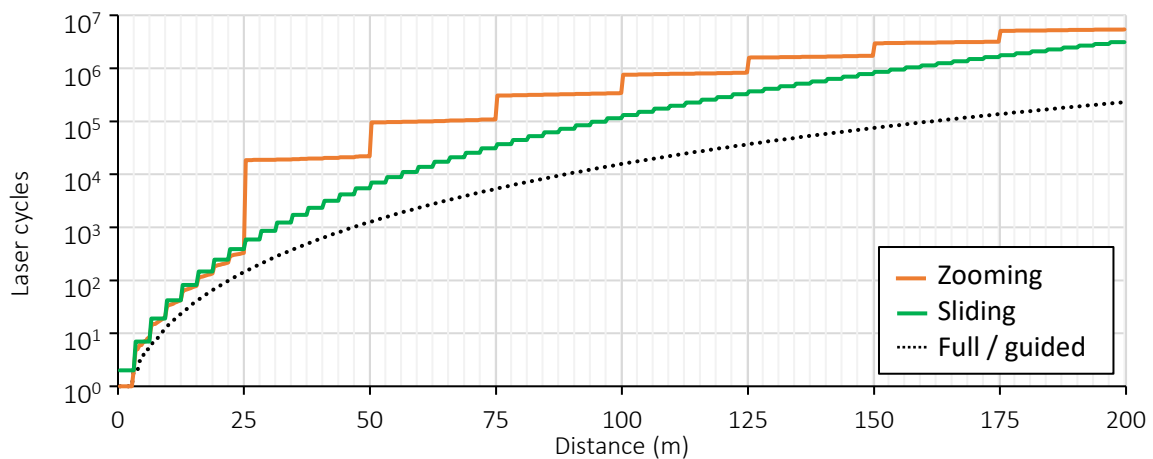
Fig. 108: Minimum number of laser cycles required to achieve 10 cm precision using each dToF approach. Reproduced from [3].
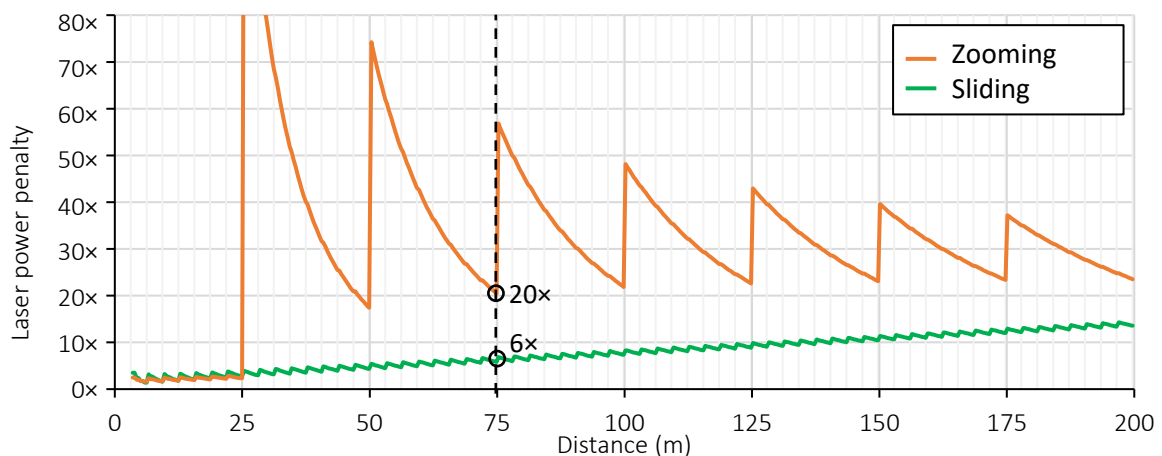
Fig. 109: Resulting laser power penalty using an equivalent partial histogram approach instead of the presented guided dToF approach. Reproduced from [3].

133

6.5 Discussion

6.5.1 Performance Assessment

The implemented guided dToF system achieves a maximum sensing distance of 75 m while running in full resolution at 3 fps under daylight conditions. This is achieved using a single-tier dToF sensor with storage for only 8 histogram bins per pixel. To achieve the same maximum distance, an equivalent full histogram approach would require each pixel to accommodate almost 200 histogram bins, a capacity far exceeding anything realised by the current state of the art (see Fig. 28). Even if such a sensor could be realized, the significant increase in data would only exacerbate the already problematic issue of excessive data loads on self-driving vehicles [56]. On the other hand, an equivalent partial histogram implementation would require at least a 6× increase in laser cycles per frame. Not only does a guided dToF solution provide a substantial power saving over the partial histogram approach, but it also enables a much higher frame rate. This is exemplified in the state-of-the-art dToF sensors summarised in Table 16, where no partial histogram or single-tier sensor is able to achieve the same long-distance sensing (75 m) in full resolution at the frame rate (3 fps) achieved by the implemented guided dToF system. The closest performing partial histogram or single-tier implementation (Park et al. [145]) achieves almost half the range (45 m) and frame rate (<1.5 fps).

In addition to these quantitative merits, the ability to guide the sensor under multipath conditions is advantageous. While the glass obstruction tested in Fig. 102 shows one such example, many other real-world conditions cause such multi-peak signatures including smoke [172] and reflective road signs [173]. The challenge of actually detecting transparent objects using lidar is an area of active research, particularly in mapping where the preference is to detect such objects rather than see through them [174, 175]. Under these preferences, guiding the DToF sensor using depth estimation from vision cameras in the visible spectrum may not be ideal.

Author	Ref	Year	Resolution (ToF pixels)	Maximum Distance (m)	Ambient Intensity (klux)	Precision / Accuracy (m)	Frame Rate (Hz)	Histogram bins / ToF pixel	Stacked	Partial Histogram
Ximenes	[139]	2018	64 × 128	300[1]	-	0.47 / 0.8	20[2]	-	Yes	No
Henderson	[108]	2019	64 × 64	50	-	- / 0.17	30	16	Yes	No
Zhang	[52]	2019	144 × 252	50[1]	-	0.0014 / 0.88	30[2]	8	No	Zooming
Okino	[148]	2020	900 × 1200	250	-	1.5 / -	-	-	No	No
Kim	[53]	2021	40 × 48	45	-	0.014 / 0.023	-	2	No	Zooming
Kumagai	[55]	2021	63 × 168	200	117	- / 0.3	20	-	Yes	No
Preethi	[109]	2021	128 × 256	100[1]	10	- / 0.07	-	-	Yes	No
Stoppa	[54]	2021	60 × 80	4.4	50	0.007 / 0.04	30	32	Yes	Sliding
Zhang	[143]	2021	160 × 240	9.5	10	0.01 / 0.02	20	32	Yes	Zooming
Park	[145]	2022	60 × 80	45	30	0.015 / 0.025	1.5[3]	4	No	Zooming
Taloud	[144]	2022	32 × 42	8.2	1	0.007 / 0.03	30	59	Yes	Sliding
This work	-	2023	32 × 64	75	70	0.2	3	8	No	No

Table 16: Performance overview of state-of-the-art direct time-of-flight lidar sensors

[1] Single point measurement.

[2] Frame rate not specified at maximum distance.

[3] Frame rate specified at 33 m, not maximum distance.

135

6.5.2 Practical Challenges

In Fig. 70, it was illustrated that even minor deviations in camera orientation can lead to substantial errors in stereo depth estimation. These errors can cause undesired effects in the guided dToF system, such as objects disappearing between depth windows, as observed in Fig. 107. The specific guiding mechanism implemented here through stereo depth estimation presents a particular challenge, as any error not only degrades stereo depth estimation accuracy but also affects the pixel mapping of each depth estimate to the lidar sensor. For a guided dToF system employing stereo depth to be capable of robust performance onboard self-driving vehicles, continuous self-calibration techniques such as those presented in [176] should be implemented.

6.5.3 Design Trade-Offs

While the frame rate achieved by the guided dToF system is competitive with the state-of-the-art, it still falls short of the 25 fps requirement for fully autonomous vehicles (Table 1). This could have additional consequences for guided dToF, where slow depth estimation on a fast-moving target might lead the lidar to be guided to an incorrect depth window. For instance, consider a vehicle traveling at 65 mph. At the current frame rate of 3 fps, the vehicle would cover a distance of 10 m per frame, exceeding the 3 m depth window size used in the implemented system.

While further development to implement key processes within each frame (e.g., depth estimation and lidar acquisition) in parallel will make 25 fps more attainable, some trade-offs may also be adopted. By configuring the dToF sensor to wider depth windows, the system can tolerate less accuracy from depth estimation. In the context of a stereo camera guided dToF system, this allows for the use of faster stereo-matching algorithms [177]. Wider depth windows can be implemented by either (i) increasing the bin interval of each dToF pixel histogram at the expense of increased measurement error or (ii) increasing pixel capacity for more histogram bins.

6.6 Conclusion

This chapter presents the implementation and characterisation of the first guided dToF system. The system is demonstrated under high ambient daylight conditions of 72 klux while operating at 3 fps and performs ranging up to 75 m using the full sensor resolution. This performance is unrivalled for any single-tier state-of-the-art sensor. Furthermore, this is achieved without relying on partial histogram approaches, of which an equivalent implementation would require at least 6× more laser cycles per frame. An overview comparing the proven merits of a guided dToF approach against conventional full or partial histogram approaches is provided in Table 17. Provided that camera alignment and frame rate limitations can be overcome, guided dToF presents a tangible solution for 3D depth-sensing onboard self-driving vehicles where a wealth of sensor data for guiding is readily available.

Parameter	Full Histogram	Partial Histogram		Guided
		Zooming	Sliding	
Laser power penalty	Low	High	High	Low
Area requirement	High	Low	Low	Low
Data volume	High	Low	High	Low
Multipath reflection artefacts	Low	Medium	Low	Low
Motion artefacts	Low	Medium	Medium	Medium
System complexity	Low	Low	Low	High

Table 17: Comparison of full, partial and guided histogram dToF. Reproduced from [3].

7 CONCLUSION

7.1 Summary

This thesis aims to contribute to a key piece of technology to enable the widespread commercialization of self-driving vehicles: 3D depth-sensing. Solid-state lidar offers a mass-market depth-sensing solution compared to traditional mechanical scanning approaches, and the direct time-of-flight (dToF) approach is currently the most suited to meet the depth-sensing requirement for autonomous vehicles. However, solid-state dToF requires a sensor chip containing a 2D array of pixels, each capable of detecting, timing and storing the laser return time. In the modelling conducted in Chapter 3, it was shown that a sensor would need to be manufactured in an expensive stacked process using no larger than a 16 nm technology node to meet typical automotive depth-sensing requirements. Even if such a sensor could be realised, the processing of all photon data would present a significant challenge. To address this problem, various partial histogram approaches have been developed which summarise into two distinct approaches: sliding and zooming. As also shown in Chapter 3, these impose a significant laser power penalty on the lidar system, approaching $4\times$ the laser power requirement compared to a full histogram approach when modelled within the context of a typical automotive lidar system. This thesis documents the design, construction and evaluation of a new approach "guided dToF", allowing auxiliary sensors to narrow down the depth search space for a guidable dToF sensor for a practical solution without power-hungry partial histogram techniques.

The design of a guided dToF sensor is outlined in Chapter 4, showing that guided functionality can be easily integrated into any dToF sensor chip with minimal area overhead. In the case of the upgraded QuantIC4$\times$4 sensor, the added guided functionality occupied only 5% of the total pixel area. The project ultimately relied on a contingency dToF sensor "HSLIDAR" which was not designed for guided dToF. However, the successful execution of the eventual guided dToF system despite these circumstances only further exemplifies the feasibility of guidable dToF sensors.

Stereo cameras were chosen for proving depth estimates to guide the dToF sensor. This is an appropriate choice given the indispensable role of cameras in self-driving vehicles. A stereo camera rig is constructed with a 1 m baseline and the established semi-global matching algorithm used to perform stereo-matching. The performance of the standalone camera rig is evaluated and, while the depth accuracy

degrades with distance, the designed guided dToF sensor would only need its histogram bin intervals configured to around 1 m for the system to operate at a maximum distance of 200 m.

Once the guidable dToF sensor and the guiding stereo camera rig were in place, the entire system was integrated. To eliminate any additional complexities that might arise from scanning, a flash lidar illumination was adopted. The guided dToF system was evaluated running real time under a variety of scenarios to assess its practical performance.

7.2 Key Findings and Results

This research contributes to the current state-of-the-art by providing a power- and data-efficient approach to dToF in the form of "guided dToF". The system is demonstrated under high ambient daylight conditions of 72 klux while operating at 3 fps and performs ranging up to 75 m using the full sensor resolution. This performance is unrivalled when compared against any state-of-the-art single-tier sensor. Furthermore, the system performs better than any published state-of-the-art partial histogram sensors, without the inherent laser power penalty associated with these approaches (calculated to be at least $6\times$ greater).

Guided dToF is also proven to be effective under conditions such as fog which typically prove challenging to lidar due to the presence of early reflections. The sensor can be guided to see past these problematic reflections. Similarly, transparent obstructions (e.g. glass) which cause multipath reflections can cause the scene behind it to be undetected in many lidar configurations, but a guided dToF system is shown to see through such obstructions.

7.3 Limitations

The primary limitations are identified surrounding this specific implementation of guided dToF are blind spots from stereo depth estimation, camera misalignment, maximum range and frame rate. These are each discussed separately here, along with proposed future solutions.

7.3.1 Stereo Blind Spots

Stereo depth estimation is only possible in the region where both camera field-of-views overlap. As a result, there is a blind spot region at short distances before the overlap region begins as illustrated in Fig. 60. As a result, guiding cannot be performed in this region. For the specific implementation presented here, this region extends to the first 15 m in front of the cameras. While exploring a trinocular (three-camera) arrangement would reduce the extent of this blind spot, the anticipated integration of both short- and long-range lidar sensors in self-driving vehicles would effectively address this issue.

7.3.2 Camera Misalignment

In Chapter 5, even minor deviations in camera orientation were shown to produce substantial errors in stereo depth estimation (Fig. 70). The repercussions of this were realised in Chapter 6, causing objects to disappear between depth windows for example (Fig. 107). Furthermore, the pixel mapping technique adopted, as described in Section 6.3.3, is also sensitive to errors in stereo depth estimation. To ensure reliable performance under the harsh shock and vibrations experienced by the vehicle, self-calibration techniques would need to be employed to continuously update the cameras' extrinsic properties.

7.3.3 Maximum Range & Frame Rate

While the achieved maximum 75 m outdoor ranging at 3 fps using the full sensor resolution surpasses published state-of-the-art dToF sensor performances, it falls short of the 200 m range at 25 fps targeted for automotive lidar. Once the primary QuantIC4×4 sensor developed for this project becomes available, the improved SPAD sensitivity from 3.5% to 18.5% PDE at the 940 nm laser wavelength would significantly boost the maximum achievable range of the setup. Alternatively, the required number of laser cycles may be reduced to boost frame rate without degrading the maximum range. However, as shown in the process breakdown in Fig. 106, even without the lidar acquisition period the maximum achievable frame rate of the presented implementation is 6 fps. On the other hand, adopting the primary QuantIC4×4 would double the available histogram bins per pixel, giving wider depth windows that can tolerate faster stereo depth estimation at the expense of lower accuracy. Further frame rate boost could be achieved through use of graphics processing units (GPUs) for accelerated processing and executing the main processes within the guided system in parallel rather than sequentially. By exploring these in combination there is nothing to suggest that the system couldn't achieve 200 m ranging at 25 fps.

7.4 Future Work

7.4.1 Evolution of Guided dToF

Further to addressing the current limitations of the system through the suggestions outlined above, there exist exciting possibilities in exploring entirely new methods of guiding in a guided dToF system. Depth estimation from single cameras, as explored in Section 5.3 brings significant benefits compared to a stereo camera guided system, including the elimination of occlusions and the short-range blind spot. While the monocular depth estimation techniques explored in this thesis did not perform to the level required for guided dToF, this is a rapidly evolving field and could soon present a viable option. As discussed in Section 5.5, this could even be evolved to extract depth information from the lidar sensor's own intensity image, creating a self-guided dToF lidar.

Beyond image sensors, guiding from other types of sensors onboard self-driving vehicles presents a valuable exploration. In particular, radar is a long-established technology which is more power-efficient than lidar over longer distances. While radar lacks the fine lateral resolution of lidar, it could point a guided dToF system to a broad area of interest, allowing the lidar to reveal the object with a dense and precise point cloud.

Taking this a step further, guiding may even be realised without any auxiliary sensors at all. Through data processing and prediction techniques such as Kalman filtering, the historical depth data previously captured by the lidar sensor could be leveraged to guide its own depth windows.

In practice, the most effective guided dToF solution would make use of the variety of data available onboard a self-driving vehicle. Exploring how to effectively combine different sources of data to provide the most appropriate guiding data for a given scenario would make for a robust guided dToF system.

7.4.2 Further Evaluation

The scenarios under which the guided dToF demonstrator was tested allowed it to be evaluated under real-world conditions. However, there is also value to be gained from testing under controlled environments. A large indoor test environment would allow quantitative evaluation of a guided lidar system for a variety of different target distances, object reflectivity, textures and lighting conditions, all

of which affect both the guiding source and the lidar. Using an automotive-grade lidar would also provide a greater amount of ground truth data for benchmarking in such tests, compared to the single-point rangefinder used in this work. Finally, if a guided dToF sensor can be configured into equivalent sliding and zooming partial histogram modes, experimentally evaluating the laser power penalty incurred in a controlled environment would be valuable.

7.5 Final Remarks

The guided dToF system presented in this thesis marks the first implementation of this approach. The depth-sensing performance demonstrated, realised using a single-tier silicon chip sensor, while at the same time minimising laser and data usage already emphasises its substantial potential. As we look towards a future where self-driving vehicles become prevalent, not only does addressing excessive lidar laser usage save power, but also helps mitigate the looming problem of interference between vehicles. The versatility of guided dToF, coupled with limitless conceivable ways to guide the sensor, makes it an exciting forward-looking prospect for self-driving vehicles and a wide range of applications.

8 REFERENCES

[1] F. Taneski, T. A. Abbas, and R. K. Henderson, "Laser Power Efficiency of Partial Histogram Direct Time-of-Flight LiDAR Sensors," *Journal of Lightwave Technology,* vol. 40, no. 17, pp. 5884-5893, 2022, doi: https://doi.org/10.1109/JLT.2022.3187293.

[2] F. Taneski, I. Gyongy, T. A. Abbas, and R. Henderson, "Guided Flash Lidar: A Laser Power Efficient Approach for Long-Range Lidar," presented at the International Image Sensor Workshop, Crieff, Scotland, 2023. [Online]. Available: https://web.archive.org/web/20230913082706/https://imagesensors.org/2023-papers/.

[3] F. Taneski, I. Gyongy, T. Al Abbas, and R. K. Henderson, "Guided Direct Time-of-Flight Lidar Using Stereo Cameras for Enhanced Laser Power Efficiency," *Sensors,* vol. 23, no. 21, p. 8943, 2023. [Online]. Available: https://doi.org/10.3390/s23218943.

[4] F. Taneski, T. A. Abbas, and R. Henderson, "Extending Lidar Depth Range using Stereo Depth Estimation on Intensity Data," presented at the Electronic Imaging, 2024. [Online]. Available: https://doi.org/10.2352/EI.2024.36.18.3DIA-102.

[5] H. Petroski, "The Road Ahead" *American Scientist*, vol. Volume 104, Number 4, p. 236doi: 10.1511/2016.121.236.

[6] General Motors, "To New Horizons," ed. New York, NY, 1940, p. https://archive.org/details/ToNewHor1940.

[7] World Health Organisation (WHO). Disease burden and mortality estimates [Online] Available: https://www.who.int/data/gho/data/themes/mortality-and-global-health-estimates

[8] National Highway Traffic Safety Administration (NHTSA), "Critical Reasons for Crashes Investigated in the National Motor Vehicle Crash Causation Survey," in "Traffic Safety Facts," U.S. Department of Transportation, Washington, 2015. [Online]. Available: https://web.archive.org/web/20240214030530/https://crashstats.nhtsa.dot.gov/Api/Public/Publication/812506

[9] GOV.UK, "Code of Practice: automated vehicle trialling" *GOV.UK,* , 28 January 2022. [Online]. Available: https://web.archive.org/web/20230326054657/https://www.gov.uk/government/publications/trialling-automated-vehicle-technologies-in-public/code-of-practice-automated-vehicle-trialling

[10] R. Furchgott, "Public Streets Are the Lab for Self-Driving Experiments" *The New York Times, The New York Times,* , 24 December 2021. [Online]. Available: https://web.archive.org/web/20240220171721/https://www.nytimes.com/2021/12/23/business/tesla-self-driving-regulations.html

[11]   Waymo, "First Million Rider-Only Miles: How the Waymo Driver is Improving Road Safety"
         Blog,   28th   February   2023.   [Online].   Available:
         https://web.archive.org/web/20231028025044/https://waymo.com/blog/2023/02/first-
         million-rider-only-miles-how.html

[12]   L. Caudet. "Road safety: Commission welcomes agreement on new EU rules to help save lives."
         European   Commission.
         https://web.archive.org/web/20220708234954/https://ec.europa.eu/commission/presscorner/d
         etail/en/IP_19_1793 (accessed 19/01/2021.

[13]   B. Schoettle, "Sensor Fusion: A Comparison of Sensing Capabilities of Human Drivers and
         Highly Automated Vehicles," The University of Michigan, 2017. [Online]. Available:
         https://web.archive.org/web/20231109122134/http://websites.umich.edu/~umtriswt/PDF/S
         WT-2017-12.pdf

[14]   *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road
         Motor Vehicles (J3016)*, SAE International, 2018.

[15]   M. E. Warren, "Automotive LIDAR Technology," in *2019 Symposium on VLSI Circuits*, 9-14
         June 2019 2019, pp. C254-C255, doi: https://doi.org/10.23919/VLSIC.2019.8777993.

[16]   X. Z. Sun, L.; Zhang, Q.; Zhang, W., "Si Photonics for Practical LiDAR Solutions," *Applied
         Sciences,* vol. 9, no. 4225, 2019, doi: https://doi.org/10.3390/app9204225.

[17]   P. Wesling, *Heterogeneous Integration Roadmap*: IEEE, 2019. [Online]. Available:
         https://web.archive.org/web/20210625001909/https://eps.ieee.org/images/files/HIR_2019/HI
         R1_ch05_automotive.pdf.

[18]   C. Pulikkaseril and S. Lam, "Laser Eyes for Driverless Cars: The Road to Automotive LIDAR,"
         in *2019 Optical Fiber Communications Conference and Exhibition (OFC)*, 3-7 March 2019
         2019, pp. 1-4.

[19]   Velodyne, "Revolutionizing Driver Assistance Systems With Forward-looking Lidar," Velodyne
         Lidar,   White   Paper   2019.   [Online].   Available:
         https://web.archive.org/web/20210916153145/https://velodynelidar.com/wp-
         content/uploads/2019/08/Forward-Looking-Lidar-for-ADASver.5.2019.pdf

[20]   H. Holzhüter, J. Bödewadt, S. Bayesteh, A. Aschinger, and H. Blume, "Technical concepts of
         automotive LiDAR sensors: a review," *Optical Engineering,* vol. 62, no. 3, p. 031213, 2023.
         [Online]. Available: https://doi.org/10.1117/1.OE.62.3.031213.

[21]   L. Alonso, V. Milanés, C. Torre-Ferrero, J. Godoy, J. P. Oria, and T. de Pedro, "Ultrasonic
         sensors in urban traffic driving-aid systems," (in eng), *Sensors (Basel, Switzerland),* vol. 11, no.
         1, pp. 661-673, 2011, doi: https://doi.org/10.3390/s110100661.

[22]   G. Galati, "The Unfortunate Inventor and the Lucky One—The UR-RADAR, Early Apparatus," in *100 Years of Radar*, G. Galati Ed. Cham: Springer International Publishing, 2016, pp. 1-11.

[23]   J. Vargas, S. Alsweiss, O. Toker, R. Razdan, and J. Santos, "An Overview of Autonomous Vehicles Sensors and Their Vulnerability to Weather Conditions," *Sensors,* vol. 21, no. 16, p. 5397, 2021. [Online]. Available: https://doi.org/10.3390/s21165397.

[24]   T. Zhou, M. Yang, K. Jiang, H. Wong, and D. Yang, "MMW Radar-Based Technologies in Autonomous Driving: A Review," (in eng), *Sensors (Basel, Switzerland),* vol. 20, no. 24, p. 7283, 2020, doi: https://doi.org/10.3390/s20247283.

[25]   D. Scharstein, R. Szeliski, and R. Zabih, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," in *Proceedings IEEE Workshop on Stereo and Multi-Baseline Vision (SMBV 2001)*, 9-10 Dec. 2001 2001, pp. 131-140, doi: https://doi.org/10.1109/SMBV.2001.988771.

[26]   A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? The KITTI vision benchmark suite," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, 16-21 June 2012 2012, pp. 3354-3361, doi: 10.1109/CVPR.2012.6248074.

[27]   K. Zhang, J. Xie, N. Snavely, and Q. Chen, "Depth Sensing Beyond LiDAR Range," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 13-19 June 2020 2020, pp. 1689-1697, doi: 10.1109/CVPR42600.2020.00176.

[28]   Brad Templeton, "Light's 'Clarity' Depth Camera Could Be A Game Changer" *Forbes*, 29 Oct 2020. [Online]. Available: https://web.archive.org/web/20220211193653/https://www.forbes.com/sites/bradtempleton/2020/10/29/lights-clarity-depth-camera-could-be-a-game-changer/

[29]   J. Liu, Q. Sun, Z. Fan, and Y. Jia, "TOF Lidar Development in Autonomous Vehicle," in *2018 IEEE 3rd Optoelectronics Global Conference (OGC)*, 4-7 Sept. 2018 2018, pp. 185-190, doi: 10.1109/OGC.2018.8529992.

[30]   Y. Li *et al.*, "Deep Learning for LiDAR Point Clouds in Autonomous Driving: A Review," *IEEE Transactions on Neural Networks and Learning Systems,* vol. 32, no. 8, pp. 3412-3432, 2021, doi: 10.1109/TNNLS.2020.3015992.

[31]   D. Khan, M. Baek, M. Y. Kim, and D. S. Han, "Multimodal Object Detection and Ranging Based on Camera and Lidar Sensor Fusion for Autonomous Driving," in *2022 27th Asia Pacific Conference on Communications (APCC)*, 19-21 Oct. 2022 2022, pp. 342-343, doi: 10.1109/APCC55198.2022.9943618.

[32]   S. Shi *et al.*, "Pathway to a compact, fast, and low-cost LiDAR," in *2018 4th International Conference on Control, Automation and Robotics (ICCAR)*, 20-23 April 2018 2018, pp. 232-236, doi: 10.1109/ICCAR.2018.8384676.

[33]   S. Royo and M. Ballesta-Garcia, "An Overview of Lidar Imaging Systems for Autonomous Vehicles," *Applied Sciences,* vol. 9, no. 19, p. 4093, 2019. [Online]. Available: https://www.mdpi.com/2076-3417/9/19/4093.

[34]   V. Lidar, "How Lidar Sensor Choices Can Impact Electric Vehicle (EV) Power Efficiency" *Blog.* [Online].   Available: https://web.archive.org/web/20230607030309/https://velodynelidar.com/blog/lidar-and-power-consumption-electric-vehicles/

[35]   A. Petrovskaya and S. Thrun, "Model based vehicle detection and tracking for autonomous urban driving," *Autonomous Robots,* vol. 26, no. 2, pp. 123-139, 2009/04/01 2009, doi: 10.1007/s10514-009-9115-1.

[36]   R. Roriz, J. Cabral, and T. Gomes, "Automotive LiDAR Technology: A Survey," *IEEE Transactions on Intelligent Transportation Systems,* vol. PP, 06/15 2021, doi: 10.1109/TITS.2021.3086804.

[37]   G. Kim, J. Eom, S. Park, and Y. Park, "Occurrence and characteristics of mutual interference between LIDAR scanners," in *SPIE Optics + Optoelectronics*, Prague, 2015, vol. 9504: SPIE. [Online]. Available: https://doi.org/10.1117/12.2178502. [Online]. Available: https://doi.org/10.1117/12.2178502

[38]   G. Kim, I. Ashraf, J. Eom, and Y. Park, "Concurrent Firing Light Detection and Ranging System for Autonomous Vehicles," *Remote Sensing,* vol. 13, p. 1767, 05/01 2021, doi: 10.3390/rs13091767.

[39]   R. C. Limited, "Automotive LiDAR Report:Technology Deep-Dive, Trends, M&A, Future Predictions,"   2022.   [Online].   Available: https://web.archive.org/web/20220429090111/https://www.jmgraphicdesign.com/files/renevo/2022-RCL-Automotive-LiDAR-Report.pdf

[40]   Aptiv *et al.*, "Safety First For Automated Driving [White Paper]," 2019. [Online]. Available: https://group.mercedes-benz.com/documents/innovation/other/safety-first-for-automated-driving.pdf

[41]   D. V. Nam and K. Gon-Woo, "Solid-State LiDAR based-SLAM: A Concise Review and Application," in *2021 IEEE International Conference on Big Data and Smart Computing (BigComp)*, 17-20 Jan. 2021 2021, pp. 302-305, doi: 10.1109/BigComp51126.2021.00064.

[42]   Z. Zhang and L. Zhu, "A Review on Unmanned Aerial Vehicle Remote Sensing: Platforms, Sensors, Data Processing Methods, and Applications," *Drones,* vol. 7, no. 6, p. 398, 2023. [Online]. Available: https://www.mdpi.com/2504-446X/7/6/398.

[43]   E. Yurtsever, J. Lambert, A. Carballo, and K. Takeda, "A Survey of Autonomous Driving: Common Practices and Emerging Technologies," *IEEE Access,* vol. 8, pp. 58443-58469, 2020, doi: 10.1109/ACCESS.2020.2983149.

[44]   S. Abuelsamid, "Accurate Ranging Perception for Assisted and Automated Driving," Guidehouse Insights, 2021.

[45]   Y. Développement, "LiDAR for Automotive and Industrial Applications," 2021. [Online]. Available: https://web.archive.org/web/20221229085924/https://s3.i-micronews.com/uploads/2021/09/YINTR21174-LiDAR-for-Automotive-and-Industrial-Applications-2021-Sample.pdf

[46]   C. P. Hsu *et al.*, "A Review and Perspective on Optical Phased Array for Automotive LiDAR," *IEEE Journal of Selected Topics in Quantum Electronics,* vol. 27, no. 1, pp. 1-16, 2021, doi: 10.1109/JSTQE.2020.3022948.

[47]   D. Bonte, D. Shey, J. Hodgson, R. Whitton, and D. Pavlakis, "Demystifying LiDAR: IoT And Automotive Applications, Industries, And Business Models," ABI Research, 2021. [Online]. Available: https://web.archive.org/web/20230203205159/https://f.hubspotusercontent10.net/hubfs/6705264/Marketing/Whitepapers/Demystifying%20LiDAR%20-%20IoT%20and%20Automotive%20Applications,%20Industries,%20and%20Business%20Models/ABI_Research_Demystifying_LiDAR_IoT_and_Automotive_Business_Models.pdf?hsCtaTracking=b0888361-2bf0-456c-8ed9-b438c7b07547%7C98787b4a-60a6-4ed7-9e8c-d859a8d03313

[48]   C. Niclass, M. Soga, H. Matsubara, M. Ogawa, and M. Kagami, "A 0.18um CMOS SoC for a 100-m-Range 10-Frame/s 200x96-Pixel Time-of-Flight Depth Sensor," *IEEE Journal of Solid-State Circuits,* vol. 49, no. 1, pp. 315-330, 2014, doi: 10.1109/JSSC.2013.2284352.

[49]   S. W. Hutchings *et al.*, "A Reconfigurable 3-D-Stacked SPAD Imager With In-Pixel Histogramming for Flash LIDAR or High-Speed Time-of-Flight Imaging," *IEEE Journal of Solid-State Circuits,* vol. 54, no. 11, pp. 2947-2956, 2019, doi: 10.1109/JSSC.2019.2939083.

[50]   T. Al Abbas, N. Dutton, O. Almer, S. Pellegrini, Y. Henrion, and R. Henderson, "Backside illuminated SPAD image sensor with 7.83μm pitch in 3D-stacked CMOS technology," 2016/12/3 2016, doi: 10.1109/IEDM.2016.7838372.

[51]   A. R. Ximenes, P. Padmanabhan, M. Lee, Y. Yamashita, D. Yaung, and E. Charbon, "A Modular, Direct Time-of-Flight Depth Sensor in 45/65-nm 3-D-Stacked CMOS Technology," *IEEE Journal of Solid-State Circuits,* vol. 54, no. 11, pp. 3203-3214, 2019, doi: 10.1109/JSSC.2019.2938412.

[52]   C. Zhang, S. Lindner, I. M. Antolović, J. M. Pavia, M. Wolf, and E. Charbon, "A 30-frames/s, 252 x 144 SPAD Flash LiDAR With 1728 Dual-Clock 48.8-ps TDCs, and Pixel-Wise Integrated Histogramming," *IEEE Journal of Solid-State Circuits,* vol. 54, no. 4, pp. 1137-1151, 2019, doi: 10.1109/JSSC.2018.2883720.

[53]   B. Kim *et al.*, "7.2 A 48×40 13.5mm Depth Resolution Flash LiDAR Sensor with In-Pixel Zoom Histogramming Time-to-Digital Converter," in *2021 IEEE International Solid- State Circuits*

147

Conference (ISSCC), 13-22 Feb. 2021 2021, vol. 64, pp. 108-110, doi: 10.1109/ISSCC42613.2021.9366022.

[54]   D. Stoppa *et al.*, "A Reconfigurable QVGA/Q3VGA Direct Time-of-Flight 3D Imaging System with On-chip Depth-map Computation in 45/40nm 3D-stacked BSI SPAD CMOS," in *International Image Sensor Workshop 2021*, 2021.

[55]   O. Kumagai *et al.*, "7.3 A 189×600 Back-Illuminated Stacked SPAD Direct Time-of-Flight Depth Sensor for Automotive LiDAR Systems," in *2021 IEEE International Solid- State Circuits Conference (ISSCC)*, 13-22 Feb. 2021 2021, vol. 64, pp. 110-112, doi: 10.1109/ISSCC42613.2021.9365961.

[56]   S. Sudhakar, V. Sze, and S. Karaman, "Data Centers on Wheels: Emissions From Computing Onboard Autonomous Vehicles," *IEEE Micro,* vol. 43, no. 1, pp. 29-39, 2023, doi: 10.1109/MM.2022.3219803.

[57]   R. Burkard, R. Viga, J. Ruskowski, and A. Grabmaier, "Eye safety considerations and performance comparison of flash- and MEMS-based lidar systems," in *SPIE Photonics Europe*, 2020, vol. 11353: SPIE. [Online]. Available: https://doi.org/10.1117/12.2554726. [Online]. Available: https://doi.org/10.1117/12.2554726

[58]   G. Smith, "The early laser years at Hughes Aircraft Company," *IEEE Journal of Quantum Electronics,* vol. 20, no. 6, pp. 577-584, 1984, doi: 10.1109/JQE.1984.1072445.

[59]   L. D. Smullin and G. Fiocco, "Optical Echoes from the Moon," *Nature,* vol. 194, no. 4835, pp. 1267-1267, 1962/06/01 1962, doi: 10.1038/1941267a0.

[60]   J. Abshire, "NASA's Space Lidar Measurements of Earth and Planetary Surfaces," 10/24 2010, doi: 10.1364/FIO.2010.SMB1.

[61]   B. Miller, "Laser altimeter may aid photo mapping," *Aviation Week & Space Technology,* vol. March 29, pp. 60-64, 1965.

[62]   N. J. Nilsson, "Shakey The Robot" *SRI International*, vol. 1984. [Online]. Available: https://web.archive.org/web/20230624084036/https://ai.stanford.edu/~nilsson/OnlinePubs-Nils/shakey-the-robot.pdf

[63]   R. A. J. Lewis, A. R., "A scanning laser rangefinder for a robotic vehicle," *Technical Memorandum,* vol. 33-809, 1977.

[64]   K. Gleichman, L. Harmon, T. Miciek, C. Miller, and D. Zuk, "Sensor Support for the DARPA Autonomous Land Vehicle Program," Environmental Research Institute of Michigan (ERIM), 1988.

[65]   R. Stettner, H. Bailey, and R. Richmond, *Eye-safe laser radar 3D imaging* (Aerospace/Defense Sensing, Simulation, and Controls). SPIE, 2001.

[66]   F. Amzajerdian, M. Vanek, L. Petway, D. Pierrottet, G. Busch, and A. Bulyshev, "Utilization of 3-D Imaging Flash Lidar Technology for Autonomous Safe Landing on Planetary Bodies," *Proc SPIE,* 01/23 2010, doi: 10.1117/12.843324.

[67]   T. Neff, *The Laser That's Changing the World: The Amazing Stories behind Lidar, from 3D Mapping to Self-Driving Cars.* Prometheus Books, 2018.

[68]   A. Rochas *et al.*, "First fully integrated 2-D array of single-photon detectors in standard CMOS technology," *IEEE Photonics Technology Letters,* vol. 15, no. 7, pp. 963-965, 2003, doi: 10.1109/LPT.2003.813387.

[69]   C. Niclass, M. Soga, H. Matsubara, M. Ogawa, and M. Kagami, "A 0.18μm CMOS SoC for a 100m-range 10fps 200×96-pixel time-of-flight depth sensor," in *2013 IEEE International Solid-State Circuits Conference Digest of Technical Papers*, 17-21 Feb. 2013 2013, pp. 488-489, doi: 10.1109/ISSCC.2013.6487827.

[70]   J. Yoshida, "Breaking Down iPad Pro 11's LiDAR Scanner" *EE Times.* [Online]. Available: https://web.archive.org/web/20201217173019/https://www.eetimes.com/breaking-down-ipad-pro-11s-lidar-scanner/

[71]   D. C. Carmer and L. M. Peterson, "Laser radar in robotics," *Proceedings of the IEEE,* vol. 84, no. 2, pp. 299-320, 1996, doi: 10.1109/5.482232.

[72]   C. L. Niclass, A. Rochas, P. A. Besse, and E. Charbon, "A CMOS single photon avalanche diode array for 3D imaging," in *2004 IEEE International Solid-State Circuits Conference (IEEE Cat. No.04CH37519)*, 15-19 Feb. 2004 2004, pp. 120-517 Vol.1, doi: 10.1109/ISSCC.2004.1332623.

[73]   International Electrotechnical Commision (IEC), "IEC 60825-1," in "Safety of laser products," 2001.

[74]   Velodyne, "User's Manual and Programming Guide HDL-64E S2 and S2.1," 2011. [Online]. Available: https://web.archive.org/web/20180107044417/http://velodynelidar.com/docs/manuals/63-HDL64ES2h%20HDL-64E%20S2%20CD%20Users%20Manual.pdf

[75]   T. Spirig, P. Seitz, O. Vietze, and F. Heitger, "The lock-in CCD-two-dimensional synchronous detection of light," *IEEE Journal of Quantum Electronics,* vol. 31, no. 9, pp. 1705-1708, 1995, doi: 10.1109/3.406386.

[76]   V. Molebny, P. McManamon, O. Steinvall, T. Kobayashi, and W. Chen, "Laser radar: historical prospective—from the East to the West," *Optical Engineering,* vol. 56, no. 3, p. 031220, 2016. [Online]. Available: https://doi.org/10.1117/1.OE.56.3.031220.

[77]   C. S. Bamji *et al.*, "IMpixel 65nm BSI 320MHz demodulated TOF Image sensor with 3μm global shutter pixels and analog binning," in *2018 IEEE International Solid - State Circuits Conference - (ISSCC)*, 11-15 Feb. 2018 2018, pp. 94-96, doi: 10.1109/ISSCC.2018.8310200.

[78]   F. Piron, D. Morrison, M. R. Yuce, and J. M. Redouté, "A Review of Single-Photon Avalanche Diode Time-of-Flight Imaging Sensor Arrays," *IEEE Sensors Journal,* pp. 1-1, 2020, doi: 10.1109/JSEN.2020.3039362.

[79]   C. Poulton *et al.*, "Coherent solid-state LIDAR with silicon photonic optical phased arrays," *Optics Letters,* vol. 42, p. 4091, 10/15 2017, doi: 10.1364/OL.42.004091.

[80]   T. Kim *et al.*, "A Single-Chip Optical Phased Array in a Wafer-Scale Silicon Photonics/CMOS 3D-Integration Platform," *IEEE Journal of Solid-State Circuits,* vol. 54, no. 11, pp. 3061-3074, 2019, doi: 10.1109/JSSC.2019.2934601.

[81]   K. Sayyah *et al.*, "Fully Integrated FMCW LiDAR Optical Engine on a Single Silicon Chip," *Journal of Lightwave Technology,* vol. 40, no. 9, pp. 2763-2772, 2022, doi: 10.1109/JLT.2022.3145711.

[82]   B. Behroozpour, P. A. M. Sandborn, M. C. Wu, and B. E. Boser, "Lidar System Architectures and Circuits," *IEEE Communications Magazine,* vol. 55, no. 10, pp. 135-142, 2017, doi: 10.1109/MCOM.2017.1700030.

[83]   Brian J. Roxworthy, Pradeep Srinivasan, and A. I. Ashwin Samarao, "FMCW LIDAR using array waveguide receivers and optical frequency shifting," USA Patent US11372105B1, 28 June 2022, 2021.

[84]   H. Kaushal, V. K. Jain, and S. Kar, "1 Overview of Wireless Optical Communcation Systems," in *Free Space Optical Communication*: Springer India, 2017.

[85]   ASTM International, "ASTM G173-03(2020)," in "Standard Tables for Reference Solar Spectral Irradiances: Direct Normal and Hemispherical on 37° Tilted Surface," 2020.

[86]   I. Kim, B. McArthur, and E. Korevaar, "Comparison of laser beam propagation at 785 nm and 1550 nm in fog and haze for optical wireless communications," *Proc. SPIE,* vol. 4214, 02/01 2001, doi: 10.1117/12.417512.

[87]   S. Rangwala, "LiDAR - Now You See Me, Soon You Won't!" *Forbes*. [Online]. Available: https://web.archive.org/web/20220831084156/https://www.forbes.com/sites/sabbirrangwala/2020/07/17/lidarnow-you-see-me-soon-you-wont/

[88]   K. Morimoto *et al.*, "3.2 Megapixel 3D-Stacked Charge Focusing SPAD for Low-Light Imaging and Depth Sensing," in *2021 IEEE International Electron Devices Meeting (IEDM)*, 11-16 Dec. 2021 2021, pp. 20.2.1-20.2.4, doi: 10.1109/IEDM19574.2021.9720605.

[89] R. H. Hadfield *et al.*, "Single-photon detection for long-range imaging and sensing," *Optica,* vol. 10, no. 9, pp. 1124-1141, 2023/09/20 2023, doi: 10.1364/OPTICA.488853.

[90] Y. Miao *et al.*, "Review of Ge(GeSn) and InGaAs Avalanche Diodes Operating in the SWIR Spectral Region," *Nanomaterials,* vol. 13, no. 3, p. 606, 2023. [Online]. Available: https://www.mdpi.com/2079-4991/13/3/606.

[91] A. Babichev *et al.*, "Long-Wavelength VCSELs: Status and Prospects," *Photonics,* vol. 10, no. 3, p. 268, 2023. [Online]. Available: https://doi.org/10.3390/photonics10030268.

[92] M. Bijelic, T. Gruber, and W. Ritter, "A Benchmark for Lidar Sensors in Fog: Is Detection Breaking Down?," in *2018 IEEE Intelligent Vehicles Symposium (IV)*, 26-30 June 2018 2018, pp. 760-767, doi: 10.1109/IVS.2018.8500543.

[93] D. Tsai, J. S. Berrio, M. Shan, S. Worrall, and E. Nebot, "See Eye to Eye: A Lidar-Agnostic 3D Detection Framework for Unsupervised Multi-Target Domain Adaptation," *IEEE Robotics and Automation Letters,* vol. 7, no. 3, pp. 7904-7911, 2022, doi: 10.1109/LRA.2022.3185783.

[94] Augusto Ximenes, "3. Lidar Operation and Modeling," in *Modular time-of-flight image sensor for light detection and ranging. A digital approach to LIDAR*: Delft University of Technology, 2019.

[95] S. Son *et al.*, "Experimental Analysis of Various Blockage Performance for LiDAR Sensor Cleaning Evaluation," *Sensors,* vol. 23, no. 5, p. 2752, 2023. [Online]. Available: https://www.mdpi.com/1424-8220/23/5/2752.

[96] D. Wang, C. Watkins, and H. Xie, "MEMS Mirrors for LiDAR: A Review," *Micromachines,* vol. 11, no. 5, p. 456, 2020. [Online]. Available: https://www.mdpi.com/2072-666X/11/5/456.

[97] J. Schulte-Tigges *et al.*, "Benchmarking of Various LiDAR Sensors for Use in Self-Driving Vehicles in Real-World Environments," *Sensors,* vol. 22, no. 19, 2022, doi: 10.3390/s22197146.

[98] G. Ye, "RoboSense M Series LiDAR Achieved AEC-Q100 Certification Issued by SGS" *Business Wire.* [Online]. Available: https://web.archive.org/web/20230605093110/https://www.businesswire.com/news/home/20230328005618/en/RoboSense-M-Series-LiDAR-Achieved-AEC-Q100-Certification-Issued-by-SGS

[99] K. Van Acoleyen, W. Bogaerts, J. Jágerská, N. Le Thomas, R. Houdré, and R. Baets, "Off-chip beam steering with a one-dimensional optical phased array on silicon-on-insulator," *Optics Letters,* vol. 34, no. 9, pp. 1477-1479, 2009/05/01 2009, doi: 10.1364/OL.34.001477.

[100] H. Abediasl and H. Hashemi, "Monolithic optical phased-array transceiver in a standard SOI CMOS process," *Opt. Express,* vol. 23, no. 5, pp. 6509-6519, 2015/03/09 2015, doi: 10.1364/OE.23.006509.

[101] C. V. Poulton, M. J. Byrd, B. Moss, E. Timurdogan, R. Millman, and M. R. Watts, "8192-Element Optical Phased Array with 100° Steering Range and Flip-Chip CMOS," in *2020 Conference on Lasers and Electro-Optics (CLEO)*, 10-15 May 2020 2020, pp. 1-2.

[102] X. Cao, G. Qiu, K. Wu, C. Li, and J. Chen, "Lidar system based on lens assisted integrated beam steering," *Optics Letters,* vol. 45, no. 20, pp. 5816-5819, 2020/10/15 2020, doi: 10.1364/OL.401486.

[103] U. Kabuk, "4D solid state LIDAR – NEXT Generation NOW," presented at the International SPAD Sensor Workshop, 2020. [Online]. Available: https://web.archive.org/web/20230621143535/https://www.imagesensors.org/Past%20Workshops/2020%20ISSW/Unsal_Kabuk_ISSW_2020_Ibeo_Automotive%20Systems.pdf.

[104] M. Y, "Apple LIDAR Demystified: SPAD, VCSEL, and Fusion…" *3da medium.* [Online]. Available: https://web.archive.org/web/20230203170218/https://4sense.medium.com/apple-lidar-demystified-spad-vcsel-and-fusion-aa9c3519d4cb

[105] N. Li *et al.*, "A Progress Review on Solid-State LiDAR and Nanophotonics-Based LiDAR Sensors," *Laser & Photonics Reviews,* vol. 16, no. 11, p. 2100511, 2022, doi: https://doi.org/10.1002/lpor.202100511.

[106] E. Cook *et al.*, "Polysilicon Grating Switches for LiDAR," *Journal of Microelectromechanical Systems,* vol. PP, pp. 1-6, 06/30 2020, doi: 10.1109/JMEMS.2020.3004069.

[107] X. Zhang, K. Kwon, J. Henriksson, J. Luo, and M. C. Wu, "A large-scale microelectromechanical-systems-based silicon photonics LiDAR," *Nature,* vol. 603, no. 7900, pp. 253-258, 2022/03/01 2022, doi: 10.1038/s41586-022-04415-8.

[108] R. K. Henderson *et al.*, "5.7 A 256×256 40nm/90nm CMOS 3D-Stacked 120dB Dynamic-Range Reconfigurable Time-Resolved SPAD Imager," in *2019 IEEE International Solid- State Circuits Conference - (ISSCC)*, 17-21 Feb. 2019 2019, pp. 106-108, doi: 10.1109/ISSCC.2019.8662355.

[109] P. Padmanabhan *et al.*, "7.4 A 256×128 3D-Stacked (45nm) SPAD FLASH LiDAR with 7-Level Coincidence Detection and Progressive Gating for 100m Range and 10klux Background Light," in *2021 IEEE International Solid- State Circuits Conference (ISSCC)*, 13-22 Feb. 2021 2021, vol. 64, pp. 111-113, doi: 10.1109/ISSCC42613.2021.9366010.

[110] C. Florin, "Flash Imaging Miniaturization," 2021. [Online]. Available: https://web.archive.org/web/20231103141314/https://epic-assoc.com/wp-content/uploads/2021/12/Claude-Florin-Fastree3D.pdf.

[111] S. Rangwala, "Lidar Miniaturization" *ADAS & Autonomous Vehicle International*, vol. April, pp. 34-38. [Online]. Available: https://web.archive.org/web/20230528100056/https://www.ukimediaevents.com/publication/9cb6eeeb

[112] S. Rangwala, "Ouster Acquires Sense Photonics To Create Next Generation VCSEL Based LiDAR" *Forbes Transportation*. [Online]. Available: https://web.archive.org/web/20211031162004/https://www.forbes.com/sites/sabbirrangwala/2021/10/05/ouster-acquires-sense-photonics-to-create-next-generation-vcsel-based-lidar/

[113] J. Skidmore, "Semiconductor Lasers for 3-D Sensing" *Optica Optics & Photonics News*, vol. February.

[114] P. Staudinger *et al.*, *Multi-junction lasers for LiDAR applications* (SPIE LASE). SPIE, 2023.

[115] I. Kenichi, "Semiconductor Lasers and VCSEL History," in *VCSEL Industry: Communication and Sensing*: IEEE, 2022, pp. 1-27.

[116] J. Tatum, "Appendix D: Wafer Level Testing," in *VCSEL Industry*, 2021, pp. 245-253.

[117] Jean-Francois Seurin *et al.*, "Multi-junction vertical-cavity surface-emitting lasers in the 800-1100nm wavelength range," presented at the SPIE Vertical-Cavity Surface-Emitting Lasers XXV, 2021.

[118] A. E. Council, "AEC - Q100 - Rev-H. Failure Mechanism Based Stress Test Qualification for Intergrated Circuits," 2014. [Online]. Available: https://web.archive.org/web/20230530095904/http://www.aecouncil.com/Documents/AEC_Q100_Rev_H_Base_Document.pdf

[119] R. Heinrichs *et al.*, *Three-dimensional laser radar with APD arrays* (Aerospace/Defense Sensing, Simulation, and Controls). SPIE, 2001.

[120] B. Aull *et al.*, "Laser Radar Imager Based on 3D Integration of Geiger-Mode Avalanche Photodiodes with Two SOI Timing Circuit Layers," in *2006 IEEE International Solid State Circuits Conference - Digest of Technical Papers*, 6-9 Feb. 2006 2006, pp. 1179-1188, doi: 10.1109/ISSCC.2006.1696163.

[121] E. A. G. Webster, L. A. Grant, and R. K. Henderson, "A High-Performance Single-Photon Avalanche Diode in 130-nm CMOS Imaging Technology," *IEEE Electron Device Letters,* vol. 33, no. 11, pp. 1589-1591, 2012, doi: 10.1109/LED.2012.2214760.

[122] S. Pellegrini *et al.*, "Industrialised SPAD in 40 nm technology," in *2017 IEEE International Electron Devices Meeting (IEDM)*, 2-6 Dec. 2017 2017, pp. 16.5.1-16.5.4, doi: 10.1109/IEDM.2017.8268404.

[123] B. Mamdy *et al.*, "A high PDE and high maximum count rate and low power consumption 3D-stacked SPAD device for Lidar applications " presented at the International Image Sensor Worksop, Crieff, 2023.

[124] J. A. Richardson, L. A. Grant, and R. K. Henderson, "Low Dark Count Single-Photon Avalanche Diode Structure Compatible With Standard Nanometer Scale CMOS Technology,"

IEEE Photonics Technology Letters, vol. 21, no. 14, pp. 1020-1022, 2009, doi: 10.1109/LPT.2009.2022059.

[125]   M. Gersbach *et al.*, "A Time-Resolved, Low-Noise Single-Photon Image Sensor Fabricated in Deep-Submicron CMOS Technology," *IEEE Journal of Solid-State Circuits,* vol. 47, no. 6, pp. 1394-1407, 2012, doi: 10.1109/JSSC.2012.2188466.

[126]   N. Dutton *et al.*, "Multiple-event direct to histogram TDC in 65nm FPGA technology," in *2014 10th Conference on Ph.D. Research in Microelectronics and Electronics (PRIME)*, 30 June-3 July 2014 2014, pp. 1-5, doi: 10.1109/PRIME.2014.6872727.

[127]   L. H. C. Braga *et al.*, "A CMOS mini-SiPM detector with in-pixel data compression for PET applications," in *2011 IEEE Nuclear Science Symposium Conference Record*, 23-29 Oct. 2011 2011, pp. 548-552, doi: 10.1109/NSSMIC.2011.6154110.

[128]   S. Gnecchi *et al.*, "Digital Silicon Photomultipliers With OR/XOR Pulse Combining Techniques," *IEEE Transactions on Electron Devices,* vol. 63, no. 3, pp. 1105-1110, 2016, doi: 10.1109/TED.2016.2518301.

[129]   S. Patanwala *et al.*, "A High-Throughput Photon Processing Technique for Range Extension of SPAD-based LiDAR Receivers," *IEEE Open Journal of the Solid-State Circuits Society,* vol. PP, pp. 1-1, 10/11 2021, doi: 10.1109/OJSSCS.2021.3118987.

[130]   S. Patanwala, I. Gyongy, N. Dutton, B. R. Rae, and R. Henderson, "A Reconfigurable 40 nm CMOS SPAD Array for LiDAR Receiver Validation," in *International Image Sensor Workshop*, 2019/6/23 2019.

[131]   A. Ximenes, P. Padmanabhan, and E. Charbon, "Mutually Coupled Time-to-Digital Converters (TDCs) for Direct Time-of-Flight (dTOF) Image Sensors," *Sensors,* vol. 18, p. 3413, 10/11 2018, doi: 10.3390/s18103413.

[132]   J. Szyduczyński, D. Kościelnik, and M. Miśkowicz, "Time-to-digital conversion techniques: a survey of recent developments," *Measurement,* vol. 214, p. 112762, 2023/06/15/ 2023, doi: https://doi.org/10.1016/j.measurement.2023.112762.

[133]   C. Veerappan *et al.*, "A 160×128 single-photon image sensor with on-pixel 55ps 10b time-to-digital converter," in *2011 IEEE International Solid-State Circuits Conference*, 20-24 Feb. 2011 2011, pp. 312-314, doi: 10.1109/ISSCC.2011.5746333.

[134]   R. M. Field, S. Realov, and K. L. Shepard, "A 100 fps, Time-Correlated Single-Photon-Counting-Based Fluorescence-Lifetime Imager in 130 nm CMOS," *IEEE Journal of Solid-State Circuits,* vol. 49, no. 4, pp. 867-880, 2014, doi: 10.1109/JSSC.2013.2293777.

[135]   H. Seo *et al.*, "A 36-Channel SPAD-Integrated Scanning LiDAR Sensor with Multi-Event Histogramming TDC and Embedded Interference Filter," in *2020 IEEE Symposium on VLSI Circuits*, 16-19 June 2020 2020, pp. 1-2, doi: 10.1109/VLSICircuits18222.2020.9162807.

[136]   S. Corporation, "Sony Develops "Exmor RS," the World's First Stacked CMOS Image Sensor".
         [Online].   Available:
         https://web.archive.org/web/20230408082759/https://www.sony.com/en/SonyInfo/News/Pre
         ss/201208/12-107E/

[137]   S. Sukegawa *et al.*, "A 1/4-inch 8Mpixel back-illuminated stacked CMOS image sensor," in *2013
         IEEE International Solid-State Circuits Conference Digest of Technical Papers*, 17-21 Feb.
         2013 2013, pp. 484-485, doi: 10.1109/ISSCC.2013.6487825.

[138]   S. G. Wuu, H. L. Chen, H. C. Chien, P. Enquist, R. M. Guidash, and J. McCarten, "A Review
         of 3-Dimensional Wafer Level Stacked Backside Illuminated CMOS Image Sensor Process
         Technologies," *IEEE Transactions on Electron Devices,* vol. 69, no. 6, pp. 2766-2778, 2022, doi:
         10.1109/TED.2022.3152977.

[139]   A. R. Ximenes, P. Padmanabhan, M. J. Lee, Y. Yamashita, D. N. Yaung, and E. Charbon, "A
         256×256 45/65nm 3D-stacked SPAD-based direct TOF image sensor for LiDAR applications
         with optical polar modulation for up to 18.6dB interference suppression," in *2018 IEEE
         International Solid - State Circuits Conference - (ISSCC)*, 11-15 Feb. 2018 2018, pp. 96-98,
         doi: 10.1109/ISSCC.2018.8310201.

[140]   Sony, "Sony Semiconductor Solutions to Release SPAD Depth Sensor for Smartphones with
         High-Accuracy, Low-Power Distance Measurement Performance, Powered by the Industry's
         Highest   Photon   Detection   Efficiency".   [Online].   Available:
         https://web.archive.org/web/20230307171409/https://www.sony-
         semicon.com/en/news/2023/2023030601.html

[141]   P. Padmanabhan, "Direct time-of-flight SPAD image sensors for light detection and ranging,"
         EPFL, 8231, 2021.

[142]   F. Gramuglia *et al.*, "CMOS 3D-Stacked FSI Multi-Channel Digital SiPM for Time-of-Flight
         Vision Applications," presented at the International Image Sensor Workshop, 2021, R19.

[143]   C. Zhang *et al.*, "A 240 x 160 3D Stacked SPAD dToF Image Sensor with Rolling Shutter and
         In Pixel Histogram for Mobile Devices," *IEEE Open Journal of the Solid-State Circuits Society,*
         vol. 2, pp. 3-11, 2021, doi: 10.1109/OJSSCS.2021.3118332.

[144]   P.-Y. Taloud *et al.*, "A 1.2K dots dToF 3D Imaging System in 45/22nm 3D-stacked BSI SPAD
         CMOS," presented at the International SPAD Sensor Workshop, 2022. [Online]. Available:
         https://web.archive.org/web/20230725120711/https://www.imagesensors.org/Past%20Works
         hops/2022%20ISSW/1145%20-%201215_Pierre-Yves_Taloud.pdf.

[145]   S. Park *et al.*, "5.3 An 80×60 Flash LiDAR Sensor with In-Pixel Histogramming TDC Based
         on Quaternary Search and Time-Gated Δ-Intensity Phase Detection for 45m Detectable Range
         and Background Light Cancellation," in *2022 IEEE International Solid- State Circuits
         Conference (ISSCC)*, 2022.

[146] I. Gyongy, A. T. Erdogan, N. A. Dutton, H. Mai, F. M. D. Rocca, and R. K. Henderson, "A 200kFPS, 256×128 SPAD dToF sensor with peak tracking and smart readout," in *International Image Sensor Workshop 2021*, 2021.

[147] I. Gyongy *et al.*, "A Direct Time-of-flight Image Sensor with in-pixel Surface Detection and Dynamic Vision," *IEEE Journal of Selected Topics in Quantum Electronics,* pp. 1-12, 2023, doi: 10.1109/JSTQE.2023.3238520.

[148] T. Okino *et al.*, "5.2 A 1200×900 6μm 450fps Geiger-Mode Vertical Avalanche Photodiodes CMOS Image Sensor for a 250m Time-of-Flight Ranging System Using Direct-Indirect-Mixed Frame Synthesis with Configurable-Depth-Resolution Down to 10cm," in *2020 IEEE International Solid- State Circuits Conference - (ISSCC)*, 16-20 Feb. 2020 2020, pp. 96-98, doi: 10.1109/ISSCC19947.2020.9063045.

[149] A. Tontini, L. Gasparini, and M. Perenzoni, "Numerical Model of SPAD-Based Direct Time-of-Flight Flash LIDAR CMOS Image Sensors," (in eng), *Sensors (Basel),* vol. 20, no. 18, Sep 12 2020, doi: 10.3390/s20185203.

[150] D. Grainger, "Radiometric Basics," *A Primer on Atmospheric Radiative Transfer*: University of Oxford, 2020. [Online]. Available: https://web.archive.org/web/20230615213749/https://eodg.atm.ox.ac.uk/user/grainger/research/book/protected/Chapter3.pdf

[151] MathWorks, *MATLAB Function Reference*, 2022. [Online]. Available: https://web.archive.org/web/20220620044140/https://www.mathworks.com/help/pdf_doc/matlab/matlab_ref.pdf.

[152] L. J. Koerner, "Models of Direct Time-of-Flight Sensor Precision That Enable Optimal Design and Dynamic Configuration," *IEEE Transactions on Instrumentation and Measurement,* vol. 70, pp. 1-9, 2021, doi: 10.1109/TIM.2021.3073684.

[153] *dtof-sim.* (2021). [Online]. Available: https://web.archive.org/web/20220621130909/https://github.com/lucask07/dtof-sim

[154] B. Hoefflinger, "Memory," in *CHIPS 2020 VOL. 2: New Vistas in Nanoelectronics*, B. Höfflinger Ed. Cham: Springer International Publishing, 2016, pp. 181-187.

[155] R. Fisher, "Subpixel Estimation," 2021, pp. 1217-1220.

[156] X. Ding, L. Xu, H. Wang, X. Wang, and G. Lv, "Stereo depth estimation under different camera calibration and alignment errors," *Appl. Opt.,* vol. 50, no. 10, pp. 1289-1301, 2011/04/01 2011, doi: 10.1364/AO.50.001289.

[157] H. Laga, L. V. Jospin, F. Boussaid, and M. Bennamoun, "A Survey on Deep Learning Techniques for Stereo-Based Depth Estimation," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 44, no. 4, pp. 1738-1764, 2022, doi: 10.1109/TPAMI.2020.3032602.

[158] H. Hirschmuller, "Stereo Processing by Semiglobal Matching and Mutual Information," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 30, no. 2, pp. 328-341, 2008, doi: 10.1109/TPAMI.2007.1166.

[159] MathWorks, "MATLAB Computer Vision Toolbox Reference," *vol. R2022b, Stereo Camera Calibrator*, 2022, pp. 390-391.

[160] Y. Wang, Y. Lou, Y. Zhang, W. Song, F. Huang, and Z. Tu, "A Robust Framework for Simultaneous Localization and Mapping with Multiple Non-Repetitive Scanning Lidars," *Remote Sensing,* vol. 13, p. 2015, 05/20 2021, doi: 10.3390/rs13102015.

[161] A. Gupta, A. A. Efros, and M. Hebert, "Blocks World Revisited: Image Understanding Using Qualitative Geometry and Mechanics," in *Computer Vision – ECCV 2010*, Berlin, Heidelberg, K. Daniilidis, P. Maragos, and N. Paragios, Eds., 2010// 2010: Springer Berlin Heidelberg, pp. 482-496.

[162] B. Rogers and M. Graham, "Motion Parallax as an Independent Cue for Depth Perception," *Perception,* vol. 8, no. 2, pp. 125-134, 1979/04/01 1979, doi: 10.1068/p080125.

[163] J. Ens and P. Lawrence, "An investigation of methods for determining depth from focus," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 15, no. 2, pp. 97-108, 1993, doi: 10.1109/34.192482.

[164] A. Masoumian, H. A. Rashwan, J. Cristiano, M. S. Asif, and D. Puig, "Monocular Depth Estimation Using Deep Learning: A Review," *Sensors,* vol. 22, no. 14, p. 5353, 2022. [Online]. Available: https://doi.org/10.3390/s22145353.

[165] V. Arampatzakis, G. Pavlidis, N. Mitianoudis, and N. Papamarkos, "Monocular Depth Estimation: A Thorough Review," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* pp. 1-20, 2023, doi: 10.1109/TPAMI.2023.3330944.

[166] R. Ranftl, K. Lasinger, D. Hafner, K. Schindler, and V. Koltun, "Towards Robust Monocular Depth Estimation: Mixing Datasets for Zero-shot Cross-dataset Transfer," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 2020.

[167] J. Diebel and S. Thrun, *An Application of Markov Random Fields to Range Sensing.* 2005.

[168] H. Andreasson, R. Triebel, and A. Lilienthal, "Non-Iterative Vision-Based Interpolation of 3D Laser Scans," in *Autonomous Robots and Agents*, S. C. Mukhopadhyay and G. S. Gupta Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 83-90.

[169] J. Zhu, L. Wang, R. Yang, J. E. Davis, and Z. pan, "Reliability Fusion of Time-of-Flight Depth and Stereo Geometry for High Quality Depth Maps," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 33, no. 7, pp. 1400-1414, 2011, doi: 10.1109/TPAMI.2010.172.

[170]   MathWorks, "MATLAB Image Processing Toolbox Reference," *vol. R2022b, imregtform*, 2022, pp. 1900-1907.

[171]   I. Gyongy, N. A. Dutton, and R. K. Henderson, "Direct Time-of-Flight Single-Photon Imaging," *IEEE Transactions on Electron Devices,* pp. 1-12, 2022, doi: 10.1109/TED.2021.3131430.

[172]   A. M. Wallace, A. Halimi, and G. S. Buller, "Full Waveform LiDAR for Adverse Weather Conditions," *IEEE Transactions on Vehicular Technology,* vol. 69, no. 7, pp. 7064-7077, 2020, doi: 10.1109/TVT.2020.2989148.

[173]   A. Schönlieb, D. Lugitsch, C. Steger, G. Holweg, and N. Druml, "Multi-Depth Sensing for Applications With Indirect Solid-State LiDAR," in *2020 IEEE Intelligent Vehicles Symposium (IV)*, 19 Oct.-13 Nov. 2020 2020, pp. 919-925, doi: 10.1109/IV47402.2020.9304684.

[174]   H. Tibebu, J. Roche, V. De Silva, and A. Kondoz, "LiDAR-Based Glass Detection for Improved Occupancy Grid Mapping," *Sensors,* vol. 21, no. 7, p. 2263, 2021. [Online]. Available: https://doi.org/10.3390/s21072263.

[175]   P. Foster, S. Zhenghong, P. Jong Jin, and B. Kuipers, "VisAGGE: Visible angle grid for glass environments," in *2013 IEEE International Conference on Robotics and Automation*, 6-10 May 2013 2013, pp. 2213-2220, doi: 10.1109/ICRA.2013.6630875.

[176]   T. Dang, C. Hoffmann, and C. Stiller, "Continuous Stereo Self-Calibration by Camera Parameter Tracking," *IEEE Transactions on Image Processing,* vol. 18, no. 7, pp. 1536-1550, 2009, doi: 10.1109/TIP.2009.2017824.

[177]   Abhishek Badki, Alejandro Troccoli, Kihwan Kim, Jan Kautz, Pradeep Sen, and O. Gallo, "Bi3D: Stereo Depth Estimation via Binary Classifications," *Computer Vision and Pattern Recognition CVPR 2020,* 2020.