

Imaging Outside The Box:
Enhancing the Performance and
Applicability of Time-of-Flight Sensors
through Computational Sensing

Dissertation
zur
Erlangung des Doktorgrades (Dr. rer. nat.)
der
Mathematisch-Naturwissenschaftlichen Fakultät
der
Rheinischen Friedrich-Wilhelms-Universität Bonn

von
Clara Callenberg
aus
Euskirchen

Bonn, April 2022

Angefertigt mit Genehmigung der Mathematisch-Naturwissenschaftlichen
Fakultät der Rheinischen Friedrich-Wilhelms-Universität Bonn

1. Gutachter: Prof. Dr. Matthias B. Hullin
2. Gutachter: Prof. Dr. Reinhard Klein

Tag der Promotion: 04.11.2022
Erscheinungsjahr: 2022

ABSTRACT

Time-of-Flight (ToF) imaging, which allows to measure the time that light has taken to travel from a light source to a camera, is an increasingly relevant sensing modality. ToF imagers are used, for example, in smartphones, cars, virtual and augmented reality hardware and robotics, where they usually serve as depth sensors that allow to capture the shape of a three-dimensional scene or environment. This thesis explores two types of ToF imagers – Amplitude-Modulated Continuous-Wave (AMCW) imagers and Single-Photon Avalanche Diode (SPAD) imagers – and their capabilities to go beyond their designated and conventional purpose or use cases.

Multiple general difference imaging applications are demonstrated using AMCW imaging systems by coupling different types of active light sources with the ToF sensor and employing specific modulation signal patterns. Those applications include spatial and temporal gradient imaging, direct-global separation, bipolar color-matching functions and depth edge detection. Due to the implementation in AMCW ToF hardware, the difference images can be recorded in a single shot, and the difference between the two signals is taken during the exposure, effectively eliminating alignment issues in dynamic scenes and improving noise characteristics.

The spatial resolution of time-resolved SPAD image data is increased by computational sensor fusion with a CCD camera. An experimental setup using a single objective lens and beam splitter to image the same scene on both types of sensors is employed and a computational light transport model is presented that allows to reconstruct a full high-spatial-resolution transient image for LIDAR and FLIM applications.

Finally, a very low-cost and low-resolution SPAD sensor device that is commonly used in smartphones as proximity sensors, is used to demonstrate challenging computer vision tasks: depth imaging, material classification, and tracking of an object's position that is not in the direct line of sight of the sensor.

CONTENTS

1	Introduction and Motivation	1
1.1	Thesis Outline and Contributions	2
2	Background	11
2.1	Light and its Physical Properties	11
2.1.1	Polarization	13
2.1.2	Energy and Wave-Particle Duality	16
2.2	Detecting Photons	18
2.2.1	Image Sensors	21
2.3	Time-of-Flight Sensing	22
2.3.1	Amplitude-Modulated Continuous-Wave ToF Sensing	23
2.3.2	SPAD Sensing	26
3	Snapshot Difference Imaging using Correlation Time-of-Flight Sensors	29
3.1	Introduction and Motivation	29
3.2	Related Work	31
3.3	Imaging Principle	34
3.3.1	Noise Model	35
3.4	Prototype Difference Imager	36
3.5	Applications and Evaluation	38
3.5.1	Polarization-Based Direct-Global Separation	38
3.5.2	Bipolar Color Matching Functions	40
3.5.3	Depth Edge and Directional Gradient Imaging	41
3.5.4	Spatio-Temporal Gradient Imaging	43
3.5.5	Quantitative Noise Analysis	46
3.6	Recovering Two Images From a Single Difference Image	51
3.7	Discussion	54
3.8	Minimal Setup	55
4	Super-Resolution Time-Resolved Imaging using Computational Sensor Fusion	57
4.1	Computational Fusion of Sensor Data	60

4.2	Experimental Results: LIDAR	61
4.3	Experimental Results: FLIM	63
4.4	Conclusion	64
4.5	Methods	65
4.5.1	Experimental Setup: LIDAR	65
4.5.2	Multiphoton Time-Domain Fluorescence Lifetime Imaging (FLIM)	66
4.5.3	Mammalian Cell Lines, Culturing Conditions and Transfections	66
4.5.4	Reconstruction Parameters	66
4.6	Supplementary Information	67
4.6.1	Light Transport and Image Formation	67
4.6.2	Forward Model and Reconstruction of High-Resolution Light-in-Flight Images	68
4.6.3	Forward Model Details	70
4.6.4	Numerical Simulations	73
4.6.5	Image Retrieval Runtimes	81
4.6.6	Upsampling Results on Other Data Sets	81
5	Low-Cost SPAD Sensing for Non-Line-of-Sight Tracking, Material Classification and Depth Imaging	87
5.1	Introduction	87
5.2	Related Work	90
5.3	System Description	92
5.4	Material Classification	94
5.5	Tracking Objects “Around the Corner”	99
5.5.1	Data Acquisition and Processing	101
5.5.2	Position Reconstruction	101
5.6	Depth Imaging	105
5.7	Comparison to AMCW ToF	111
5.8	Discussion and Future Work	111
6	Conclusion and Outlook	115
	Bibliography	119
	Acknowledgements	139

INTRODUCTION AND MOTIVATION

Imaging in its broadest sense denotes the formation of an image - a visual representation - of an object, a scene, or really any sort of (physical) reality. Throughout all of human history, images have played an important role as a means of expression, communication and documentation. While early human-made images were mostly drawings and paintings, even technologies that directly image the physical world have already been studied for centuries. The principle of a simple pinhole camera, that creates an image of a scene simply by restricting the angles under which light can travel to a screen, was already described more than 2000 years ago in the *Problemata Physica*, attributed to Aristotle [1]. Scientific and technological advances that include the understanding of optical refraction and the manufacturing of lenses, as well as the discovery of certain materials' light-sensitive properties, lead to the development of what we today call photography. Since then, a variety of imaging technologies and modalities has arisen, not only improving the visual and factual quality of image data, but also extending the range of what can be imaged. With better understanding of physics in general, and the nature of light in particular, new technologies have allowed the imaging of objects and realities that had previously been concealed from human perception, like elementary particles [2], black holes [3], the inside of a human body [4], or simply a room hidden around a corner [5].

Most of the recent advances in imaging technologies were enabled by the rapidly growing power of computers and their increasing prevalence and availability. Computational methods allow to compensate constraints in the physical measurement of an (optical) signal, like deficiencies in optical components or challenging measurement conditions, often by exploiting prior knowledge about the signal or its disturbances. They therefore extend the measurement process from a mere hardware task to a combination of physical signal collection and

computational treatment of the data. In many cases, the computational treatment of imaging data and the conjoined design of optical and sensing hardware and digital processing of the measured signal not only allows to improve the measurement quality, but also opens up new types and dimensions of retrievable information. For example, coupling an imaging sensor with a special pixel design with an adequate active light source can allow to measure the time that light takes to travel from the light source through a scene and to the camera for each image pixel individually. Combined with further computational treatment of such data, high-quality depth images can be created, and even three-dimensional reconstructions of objects outside the direct line of the imaging system's sight can be acquired.

Imaging systems of this kind are the core subject of this thesis. In particular, Amplitude-Modulated Continuous-Wave Time-of-Flight (AMCW ToF) imagers and Single-Photon Avalanche Diode (SPAD) imagers are studied with the aim to extend their capabilities beyond their designated purpose. In some cases, this aim is tantamount to - or at least intertwined with - achieving established levels and qualities of imaging results, but with less "cost" – i.e. by making the imaging setups cheaper, smaller, or faster. It also involves the demonstration of new imaging modalities and applications for these types of imagers.

The following section gives an overview of the research presented in this cumulative thesis, as well as my own contribution to the individual publications, and the overall structure of this document.

1.1 THESIS OUTLINE AND CONTRIBUTIONS

This thesis is a cumulative dissertation, meaning it is centered around three articles that have been published in peer-reviewed journals or conference proceedings. Each publication is presented in a separate chapter (a brief overview is given in Fig. 1.1). Since the publications are unaltered, each chapter is self-contained and features a summarizing abstract, an introduction, and an individual discussion on the presented work's background and positioning in the context of related work. In addition, the concepts and technologies used in these publications are described more comprehensively and in more detail in a dedicated chapter. The final, concluding chapter aims at evaluating the presented work with regard to the body of existing as well as possible future research. Thus, the structure of this thesis is as follows:


	Technology	Conventional Usage	Demonstrated Usage
Chapter 3	AMCW ToF	Depth images	<ul style="list-style-type: none"> • Direct-global separation • Bipolar color matching functions • Depth edges • Directional gradients • Spatio-temporal gradients
Chapter 4	SPAD	Low-resolution transient images	 High-resolution transient images
	emCCD	2D intensity images	
Chapter 5	(cheap) SPAD	Proximity sensing in consumer devices	<ul style="list-style-type: none"> • Depth images • Material classification • Non-line-of-sight tracking

Figure 1.1: Overview of the technologies used in the three main chapters, their conventional usage scenarios, and the types of applications and results demonstrated in this thesis.

CHAPTER 1: INTRODUCTION AND MOTIVATION This introductory chapter motivates and outlines the work presented in this thesis.

CHAPTER 2: BACKGROUND This chapter provides an overview of the foundation of the scientific work presented in this thesis. The nature of light is explained from a physical perspective with a special focus on the specific effects that are exploited in the experimental work presented in the later chapters. The process of light detection and measurement is discussed and the properties and operation principles of the types of sensors used in this thesis (AMCW ToF and SPAD sensors) are described.

CHAPTER 3: SNAPSHOT DIFFERENCE IMAGING This chapter contains the first publication presented in this thesis. It uses off-the-shelf AMCW ToF sensors and exploits their capability of collecting light in two different ‘buckets’ in every pixel, which can be switched between using an external modulation signal. In conventional operation of AMCW ToF cameras, this modulation signal is also used to modulate the intensity of an active light source that illuminates the scene

which then allows to infer the time of flight of the light measured at each pixel (for a more detailed description of AMCW ToF cameras see Section 2.3.1). This process involves calculating the difference between the signals measured in the two buckets, which happens on the hardware level of the sensor. Exploiting this functionality, the presented work demonstrates several alternative operating modes for this sensor type using various types of light sources and modulation signals. The capability of AMCW ToF sensors to switch between the two buckets – and, at the same time, two different light sources – with a very high frequency during the exposure allows to capture the light response of a scene under different illumination conditions virtually simultaneously and in a single snapshot. When compared to conventional difference images, where lighting conditions and images must be established and recorded in succession, this approach has several advantages. Consecutively recorded images suffer from alignment issues whenever either the camera or a scene element is moving, resulting in ghosting artefacts in difference images. In the snapshot approach, the modulation between the illumination scenarios is so much faster than the exposure time that the resulting difference image contains all relevant image data for both illuminations. Very fast moving scene elements therefore result in a natural motion blur in the difference image, instead of a misalignment artefact. Even ‘difference videos’ can be recorded using this approach, which feature a smooth rendition of moving objects. Another advantage of the proposed system is the relatively low noise level since no two images have to be read out before they can be subtracted, but the subtraction happens on the sensor before the readout, so that read noise is added only once.

As demonstrated in this chapter, the described functionality can be used with different kinds of light sources which allows a variety of imaging tasks. Multiple different setups and applications are described and experimentally demonstrated: Using two differently polarized but otherwise identical light sources, directly and indirectly (multiply) reflected light in a scene can be separated and imaged individually. By employing light sources with different emission spectra, the scene response difference between these two illumination spectra can be measured. Using identical light sources in specific locations relative to the camera and scene, depth edges and directional gradients can be imaged. A birefringent crystal separates light from differently polarized light sources spatially, allowing to image spatial gradients. Requiring no coupled active light sources at all, another operation modality shown in this chapter uses asymmetric modulation signals on the ToF sensor. This way, the resulting difference image shows temporal gradients of motions that occur during the exposure time in one single shot.

A detailed noise analysis of the proposed difference imaging method is performed,

and the reconstruction of two separate ‘pre-difference’ images of a scene under the two different lighting conditions purely from the Skellam noise characteristics [6] of the difference images is described and demonstrated.

The publication was authored by multiple people (see bibliographical information in the box below). My own contributions to the project include conceptualizing application scenarios, experiments and measurements, setting up experimental hardware and performing the measurements, analyzing and visualizing the measured data, as well as performing the necessary steps for the noise analysis and Skellam reconstruction and documenting the procedures and findings for the final publication.

The content of Chapter 3 has been published as []:

Clara Callenberg, Felix Heide, Gordon Wetzstein, and Matthias B. Hullin. 2017. *Snapshot difference imaging using correlation time-of-flight sensors*. *ACM Trans. Graph.* 36, 6, Article 220 (November 2017), 11 pages. DOI:<https://doi.org/10.1145/3130800.3130885>.

CHAPTER 4: SUPER-RESOLUTION TIME-RESOLVED IMAGING USING COMPUTATIONAL SENSOR FUSION In this chapter, the limitations of a different kind of time-of-flight sensor, consisting of an array of Single-Photon Avalanche Diodes, are challenged. In particular, their low spatial resolution is addressed, as this is a main disadvantage of current SPAD sensors. In order to gain more spatial information about the imaged scene, a CCD sensor with a higher resolution than the SPAD sensor is added to the imaging setup. The optical signal from a single objective lens is split using a beam splitter and projected on the SPAD as well the CCD sensor. The conventional intensity image from the CCD is then combined with the time-resolved, three-dimensional (two spatial, one temporal dimension) image from the SPAD sensor. To this end, another property of the SPAD sensor, that is generally a disadvantage – its low fill factor – is exploited. The active pixel area covers only about 2% of the whole sensor area meaning that most of the light (and signal information) is lost in inactive sensor area. The sensor is therefore moved slightly out of focus in order to blur the signal in such a way that each scene point is blurred over multiple active pixel areas, effectively sparsely sampling the blurred image. The light transport from the scene to the SPAD sensor is modeled as a linear transformation from the original high-resolution signal to the blurred, low-resolution SPAD measurement by modelling the three

transformation steps – blurring, masking, and downsampling – individually. This linear image formation model is used to reconstruct a ‘data-cube’ with the spatial resolution of the CCD sensor and the temporal resolution of the SPAD sensor by solving a minimization problem regularized with specifically tailored priors that ensure spatial and temporal coherence of the reconstructed transient image.

The image formation model is furthermore used to produce simulated measurements from synthetic transient images created using a transient renderer. This allows a detailed study of the reconstruction performance in dependence of different parameters like scene characteristics, data resolution, reconstruction parameters, and noise. Such an exploration of the proposed method is given and differences between scenes with mostly single-reflection signals and those with complex light response characteristics in each spatial pixel, for example due to multiple reflections inside the scene, are studied in particular.

Results of real-world measurements are presented for two different application scenarios:

- “lidar”-like images of objects and scenes of varying sizes ranging from centimeters to decimeters, from which depth maps of the object or scene can be constructed,
- and Fluorescence Lifetime Microscopy (FLIM) measurements of ovarian cancer cells in the order of micrometers.

Especially the latter highlights the ability of the method to not only upsample two-dimensional depth maps, but the whole three-dimensional ‘transient’ image, as the temporal light distribution in each pixel is used to infer fluorescence lifetimes. Upsampling factors of up to 4×4 are shown on real experimental data. In simulations, the effects and results of upsampling factors up to 12×12 as well as the influence of noise are analyzed.

The publication was authored by multiple people (see bibliographical information in the box below). While the experimental measurements were performed by my co-authors in Glasgow, my own contributions to the project consist in the mathematical construction of the image formation model and the minimization problem formulation including the specific regularization priors, implementing the optimization and tuning the parameters for the different data types, producing the simulated measurements and performing noise and resolution analyses on the different types of scenes, as well as coordinating the project and visualizing, documenting, and describing the reconstruction procedures and results.

The original publication consists of a main paper and a supplementary document, which are both contained in Chapter 4. References to the “main paper” or the “supplementary document” here therefore simply reference the preceding and

subsequent sections of this chapter, respectively.

The content of Chapter 4 has been published as []:

Clara Callenberg, Ashley Lyons, Dennis den Brok, Areeba Fatima, Alejandro Turpin, Vytautas Zickus, Laura Machesky, Jamie A. Whitelaw, Daniele Faccio, und Matthias B. Hullin. *Super-Resolution time-resolved imaging using computational sensor fusion*. Scientific Reports 11, Nr. 1 (18. Januar 2021): 1689. <https://doi.org/10.1038/s41598-021-81159-x>.

CHAPTER 5: LOW-COST SPAD SENSING FOR NON-LINE-OF-SIGHT TRACKING, MATERIAL CLASSIFICATION AND DEPTH IMAGING SPAD systems used for scientific applications are generally very expensive laboratory equipment. At the same time, they also exist in very cheap and coarse variants that are used as proximity sensors in electronic consumer devices. They feature an active light source that emits pulses of light that are then detected by the SPAD sensor after they have been reflected by an object. In smartphones, for instance, these sensors are often integrated at the front side to detect if something is covering the screen in order to switch it off, for example when it is held to the ear when making a phone call to avoid accidental touch input. The device used in this publication (STMicroelectronics VL53L1X) features a 16×16 SPAD array and an integrated infrared light source. Somewhat similarly to the generalized difference imaging approach to AMCW ToF sensors described in Chapter 3, this chapter contains the demonstration of multiple computational imaging applications performed with this type of cheap SPAD sensor device, partly with the assistance of additional (optical) equipment:

- The reconstruction of a target's position in a volume that is hidden from the direct line-of-sight of the sensor and light source by illuminating and observing the reflected signal on a diffuse wall ("non-line-of-sight (NLoS) tracking"),
- the classification of five different materials (in real time) by holding the device to the material surface and evaluating the sensor response,
- and the recording of depth images where each pixel corresponds to the distance of the imaged scene point to the imager.

The utilization of such cheap SPAD devices for the abovementioned tasks involves several challenges, mostly resulting from the properties that are required for their intended usage like low cost, small form factor, and eye-safe operation of the light

source. The low spatial and temporal resolution of the sensor poses limitations especially on complex tasks like the reconstruction of non-line-of-sight signals. This is further complicated by the low illumination intensity which makes the detection of light that has undergone multiple reflections (on the relay wall, the target, and then again the wall) almost infeasible. In the work presented in this chapter, these problems are addressed twofoldly. In order to increase the light efficiency and spatial resolution of the system, hardware additions in the form of collimating lenses and galvanometer mirrors are added to the system, which allow the collimation and exact orientation of the illumination and field of view. On the software side, artificial neural networks are trained to analyze the data in order to locate very low intensity reflection peaks and to identify material classes. The collimating lenses and galvanometer mirrors are also used to record data for the creation of depth images with a much higher resolution than the native spatial resolution of the VL53L1X's SPAD array (utilizing different methods, including fitting Gaussian functions to the temporal distribution of each pixel's data to determine the depth), while also the influence of different levels of ambient light on the depth data is analyzed. The material classification application, on the other hand, requires no additional hardware and could, in principle, be performed on a device which already features a SPAD-based proximity sensor, without any hardware modification.

This publication was authored by multiple people (see bibliographical information in the box below). My own contributions include the conceptualization of the demonstrated application scenarios, building and arranging the experimental hardware and setups, performing the measurements and recording the data, constructing and implementing the material classification, tuning the NLoS reconstruction algorithms, implementing the depth image construction, performing the ambient light study and distance calibration, as well as visualizing and describing the experimental procedures and findings and coordinating the project.

The content of Chapter 5 has been published as []:

Clara Callenberg, Zheng Shi, Felix Heide, and Matthias B. Hullin. 2021. *Low-Cost SPAD Sensing for Non-Line-of-Sight Tracking, Material Classification and Depth Imaging*. ACM Trans. Graph. 40, 4, Article 61 (August 2021), 12 pages. DOI:<https://doi.org/10.1145/3450626.3459824>.

CHAPTER 6: CONCLUSION AND OUTLOOK This last chapter examines how the presented research integrates into related work and the general context of ToF technology, especially subsequent publications and advances that have emerged recently. It also explores possible future developments and progressions.

BACKGROUND

Since the goal of scientific publications like the ones in Chapters 3-5 is the presentation of research in a concise manner, some of the concepts and technologies used in the presented work are introduced only briefly in the respective chapters. In order to give a thorough exposition of the treated topics, this chapter provides an introduction and explanation of some of the key aspects of the research projects that constitute this cumulative thesis. Some general physical and technical foundations are discussed, but with the concrete purpose of establishing the specific concepts used in the later chapters, like the exploitation of polarization angle conservation, or the technological basis and operating principle of the used ToF imagers.

2.1 LIGHT AND ITS PHYSICAL PROPERTIES

When an electric field changes over time, it induces a magnetic field in perpendicular direction, and vice versa. **Electromagnetic radiation** is this mutual induction of electric and magnetic fields as they propagate through space, creating an electromagnetic wave. In vacuum, an electromagnetic wave travels with the **speed of light** $c = 299,792,458 \frac{\text{m}}{\text{s}}$ and its wavelength λ is given by $\lambda = c/f$, where f is its oscillation frequency.

The human eye is capable of detecting electromagnetic radiation in the wavelength range from ~400-780 nanometers, which we therefore call the **visible spectrum** or simply **light**. Besides brightness, light is also associated with a perception of **color**, which depends on its wavelength and starts with a purple/blue impression at the lower end of the visible wavelength range and changes over green and yellow to red at the higher end (generally known as the colors of the rainbow). Electromagnetic radiation with higher wavelengths – from about

780 nm to 1 mm – is called **infrared radiation** and sometimes *infrared light*, even though it is invisible to the human eye. In imaging applications, infrared radiation often plays an important role, for example when invisibility is a desired feature for active illumination of a scene.

Electromagnetic phenomena, particularly the interaction between electric (\vec{E}) and magnetic (\vec{B}) fields, are mostly¹ described by Maxwell's equations:

$$\vec{\nabla} \times \vec{E} = -\frac{\partial \vec{B}}{\partial t}, \quad (2.1a)$$

$$\vec{\nabla} \times \vec{B} = \mu_0 \vec{j} + \frac{1}{c^2} \frac{\partial \vec{E}}{\partial t}, \quad (2.1b)$$

$$\vec{\nabla} \cdot \vec{E} = \frac{\rho}{\epsilon_0}, \quad (2.1c)$$

$$\vec{\nabla} \cdot \vec{B} = 0. \quad (2.1d)$$

Here, μ_0 is the vacuum permeability, ϵ_0 is the vacuum permittivity, ρ is the electric charge density, \vec{j} is the current density (of moving electric charges), and t is the time. $\vec{\nabla}$ is the three-dimensional gradient operator. In particular, Equations 2.1a-c describe the fact that temporally variable electric and magnetic fields induce each other, while electric fields are also induced by electric charges, and magnetic fields are also induced by *moving* electric charges. Equation 2.1d indicates that magnetic fields are solenoidal, meaning that they have no sources or sinks, and there is therefore no such thing as a single magnetic charge. [10]

In a vacuum free of charges and currents, Equations 2.1a and b become

$$\vec{\nabla} \times \vec{E} = -\frac{\partial \vec{B}}{\partial t}, \quad (2.2a)$$

$$\vec{\nabla} \times \vec{B} = \frac{1}{c^2} \frac{\partial \vec{E}}{\partial t}. \quad (2.2b)$$

¹ Together with the Lorentz force and Newton's second law of motion, the Maxwell equations actually describe all electromagnetic phenomena. [10]

Applying the curl operator $\vec{\nabla} \times$ to both equations results in

$$\vec{\nabla} \times \vec{\nabla} \times \vec{E} = -\vec{\nabla} \times \frac{\partial \vec{B}}{\partial t} = -\frac{\partial}{\partial t} (\vec{\nabla} \times \vec{B}) \stackrel{2.2b}{=} -\frac{\partial}{\partial t} \left(\frac{1}{c^2} \frac{\partial \vec{E}}{\partial t} \right) = -\frac{1}{c^2} \frac{\partial^2 \vec{E}}{\partial t^2}, \quad (2.3a)$$

$$\vec{\nabla} \times \vec{\nabla} \times \vec{B} = \vec{\nabla} \times \frac{1}{c^2} \frac{\partial \vec{E}}{\partial t} = \frac{1}{c^2} \frac{\partial}{\partial t} (\vec{\nabla} \times \vec{E}) \stackrel{2.2a}{=} \frac{1}{c^2} \frac{\partial}{\partial t} \left(-\frac{\partial \vec{B}}{\partial t} \right) = -\frac{1}{c^2} \frac{\partial^2 \vec{B}}{\partial t^2}. \quad (2.3b)$$

Using the identity $\vec{\nabla} \times (\vec{\nabla} \times \vec{a}) = \vec{\nabla}(\vec{\nabla} \cdot \vec{a}) - \vec{\nabla}^2 \vec{a}$ and the fact that according to Equation 2.1c, $\vec{\nabla} \cdot \vec{E} = 0$ in empty space (where $\rho = 0$), one obtains the **uncoupled, homogeneous wave equations**

$$\left(\frac{1}{c^2} \frac{\partial^2}{\partial t^2} - \vec{\nabla}^2 \right) \vec{E} = 0, \quad (2.4a)$$

$$\left(\frac{1}{c^2} \frac{\partial^2}{\partial t^2} - \vec{\nabla}^2 \right) \vec{B} = 0, \quad (2.4b)$$

which describe the propagation of electromagnetic waves, and therefore light, through vacuum. [11]

2.1.1 POLARIZATION

Electromagnetic waves (in vacuum) are *transverse waves* because the electric and magnetic fields are perpendicular to the direction of propagation. They can therefore be **polarized**.

A periodic plane wave can be described by sine or cosine functions, for example as

$$\vec{E} = \vec{E}_0 \cdot \cos(kz - \omega t) \quad (2.5a)$$

$$\vec{B} = \vec{B}_0 \cdot \cos(kz - \omega t) \quad (2.5b)$$

which is a possible solution to the wave equations 2.4 and describes the electric and magnetic fields of a wave that travels in z direction. Here, $k = 2\pi/\lambda$ is the wave number and $\omega = 2\pi f$ is the wave's angular frequency. If \vec{E}_0 always points in the same direction, the wave is called **linearly polarized**. It means that the x and y components of \vec{E} are in phase. If these two components are shifted against each other, the direction of \vec{E} rotates around the direction of propagation. If the phase shift is exactly 90° , and $E_{0,x} = E_{0,y}$ the wave is called **circularly polarized**

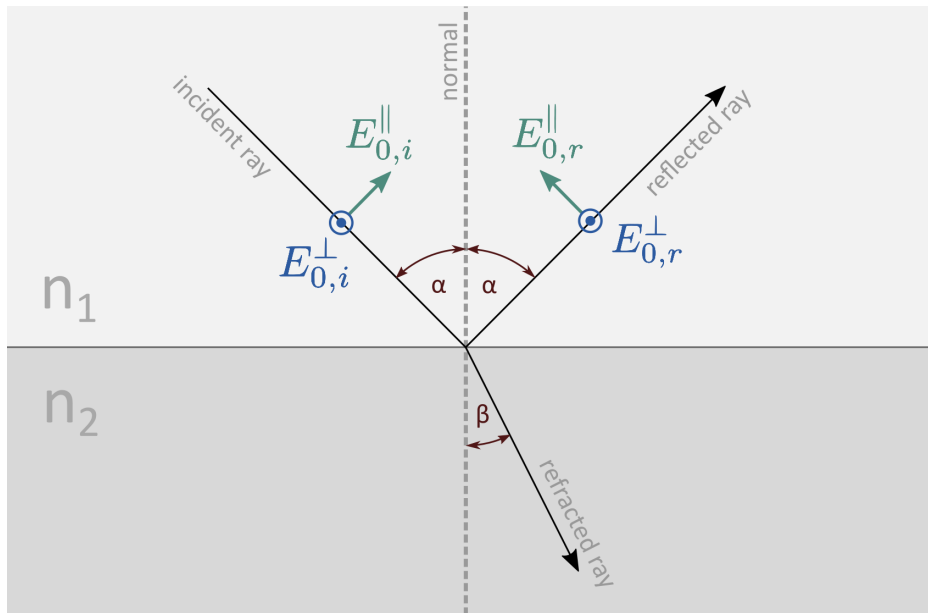


Figure 2.1: Reflection and refraction of light at the surface between two materials with different refractive indices n_1 and n_2 . The amplitudes \vec{E}_0 of the incident and reflected ray are decomposed into components parallel and perpendicular to the plane of incidence (the plane spanned by the surface normal and the incident ray).

because the tip of the vector \vec{E} describes a circle around the propagation axis. [10]

Generally, light waves are unpolarized because they are a superposition of waves created by statistically oriented oscillating dipoles (i.e. when they are emitted from excited atoms). Certain physical effects like reflection of light in certain angles or dichroism and birefringence in crystals, however, allows separation of a certain polarization direction by absorbing all other components or deflecting them in a different direction. A corresponding optical element acting as such a filter is called a **polarizer**.

When light falls onto a surface between two materials with different refraction indices n in an angle α , generally part of the light is **reflected** in the same angle $\alpha' = \alpha$ (**law of reflection**) and part is **refracted** as it is transmitted through the surface into the other medium under a different angle β . The relation between these angles is described by **Snell's law**:

$$n_1 \sin(\alpha) = n_2 \sin(\beta). \quad (2.6)$$

The amplitudes \vec{E}_0 of the incident, reflected and refracted waves can be decomposed into two components: one that is **parallel** to the plane of incidence (the plane that is spanned by the incident ray and the surface normal) and one that is **perpendicular**, as drawn in Fig. 2.1. The ratio between the perpendicular and parallel components of the incident (i) and reflected (r) wave is called the **reflection coefficient ρ** , described by the **Fresnel equations**

$$\rho_{\perp} = \frac{E_{0,r}^{\perp}}{E_{0,i}^{\perp}} = -\frac{\sin(\alpha - \beta)}{\sin(\alpha + \beta)} \quad (2.7a)$$

$$\rho_{\parallel} = \frac{E_{0,r}^{\parallel}}{E_{0,i}^{\parallel}} = \frac{\tan(\alpha - \beta)}{\tan(\alpha + \beta)} \quad (2.7b)$$

for non-magnetic dielectric materials. [10]

When the incident light is linearly polarized, there is a fixed angle γ_i between its electric field vector \vec{E}_i and the plane of incidence as illustrated in Fig. 2.2. One can therefore write:

$$\tan(\gamma_i) = \frac{E_{0,i}^{\perp}}{E_{0,i}^{\parallel}}. \quad (2.8)$$

Similarly, the angle γ_r between the reflected light's \vec{E} vector and the plane of incidence is

$$\tan(\gamma_r) = \frac{E_{0,r}^{\perp}}{E_{0,r}^{\parallel}}. \quad (2.9)$$

Using Equations 2.7, it follows that

$$\tan(\gamma_r) = -\frac{\cos(\alpha - \beta)}{\cos(\alpha + \beta)} \cdot \tan(\gamma_i) \quad (2.10)$$

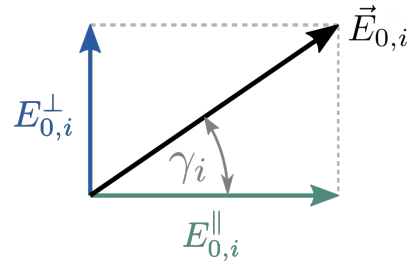


Figure 2.2: Illustration of the angle γ_i between the electric field vector and the plane of incidence. It can be calculated using the components of the incident ray's amplitude $\vec{E}_{0,i}$ which are parallel ($E_{0,i}^{\parallel}$) and perpendicular ($E_{0,i}^{\perp}$) to the plane of incidence.

which shows that at normal incidence, where $\alpha = 0^\circ$, the polarization angle is preserved². This fact is exploited in Section 3.5.1 by illuminating a scene with linearly polarized light in two different orientations and observing it with a camera equipped with an analyzing filter that is parallel to one and perpendicular to the other light sources' polarization direction. By placing the camera and light sources very closely together, light from the light sources that is reflected only once in the scene and ends up being recorded by the camera must be reflected approximately at a normal angle and therefore preserves the light source's polarization direction. This way, direct reflections can be excluded for any light that reaches the sensor from the light source which is polarized perpendicularly to the camera's analyzing filter allowing to separate direct and global components of the scene's light response.

2.1.2 ENERGY AND WAVE-PARTICLE DUALITY

Electromagnetic fields carry energy. This energy's flux (the amount of energy transported through a unit area per unit time) is described by the *Poynting vector*

$$\vec{S} = \mu_0^{-1} \vec{E} \times \vec{B} \quad (2.11)$$

which, in SI units, is measured in Watts per squaremeter (W/m^2). Since the fields in electromagnetic waves oscillate very fast (around 10^{15} Hz in the case of visible light waves), the wave's **irradiance** (often also referred to simply as **intensity**) I is given as the time-average of the Poynting vector $I = \langle S \rangle_T$. For a wave as given in Equations 2.5, this means

$$I = \langle S \rangle_T = \frac{1}{\mu_0} |\vec{E}_0 \times \vec{B}_0| \langle \cos^2(kz - \omega t) \rangle_T. \quad (2.12)$$

Plugging Equations 2.5 into the Maxwell equation 2.1a, it follows that

$$\vec{\nabla} \times \vec{E} = \frac{\partial \vec{E}}{\partial z} = -k \vec{E}_0 \sin(kz - \omega t) = -\frac{\partial \vec{B}}{\partial t} = -\omega \vec{B}_0 \sin(kz - \omega t) \quad (2.13)$$

$$\vec{E}_0 = \frac{\omega}{k} \vec{B}_0 = c \vec{B}_0. \quad (2.14)$$

² The equation holds for $\gamma_r = -\gamma_i$ when $\alpha = 0^\circ$, which describes the same polarization angle in external coordinates, considering the definition of the parallel component as shown in Fig. 2.1: When $\alpha \rightarrow 0$, the parallel component vectors of the incident and the reflected beam point in opposite directions.

Using this relation and the fact that $\langle \cos^2(kz - \omega t) \rangle_T = \frac{1}{2}$ when T is much greater than one oscillation period, it follows that

$$I = \frac{1}{c\mu_0} E_0^2, \quad (2.15)$$

meaning that the irradiance is proportional to the square of the electric field's amplitude. [10, 12]

In the case of a point light source which emits light isotropically in all directions, the outgoing light can be described by a spherical wave, which is a solution to the wave equations 2.4 where the wave's amplitude is inversely proportional to the distance r to the light source. The irradiance is therefore inversely proportional to r^2 which is known as the **inverse square law** [12]. This can easily be understood by the fact that the total energy flux emitted by the light source is spread over an increasingly large area (the surface of a sphere with increasing radius) as the wave travels away from the source. Since the surface area of a sphere is proportional to the square of its radius and the irradiance is the energy transported through this area (per time), it follows that $I \propto 1/r^2$, or

$$\frac{I_1}{I_2} = \frac{r_2^2}{r_1^2} \quad (2.16)$$

for intensities I_1, I_2 at distances r_1, r_2 from the light source. This fact is important to consider when comparing intensities of light that has travelled different distances (for example when analyzing light travel time histograms as done in Section 5.5).

While waves explain many properties of light and electromagnetic radiation in general, certain experimental observations can only be explained by the fact that electromagnetic energy transfer happens in multiples of a certain minimal amount of energy (that it is *quantized*) which is proportional to the radiations's frequency f :

$$E = hf \quad (2.17)$$

where $h \approx 6.626 \cdot 10^{-34}$ Js is the Planck constant [10]. Such a "package" of energy is called a **photon** and can be thought of as a *light particle*. The quantization of light is used in the detection and measurement of light in digital imaging sensors as described in Section 2.2.

2.2 DETECTING PHOTONS

The energy carried by photons can be converted into other forms of energy and thus measured. The relevant technology for computational imaging is based on **semiconductors**. A semiconductor is a material with a special electrical conduction behaviour. In electrical conductors, like metals, electrons can move freely inside the lattice structure of the material, which is why an applied voltage induces an electric current of moving electrons in, for example, a cable. Electrical insulators, on the other hand, do not allow electrons to move easily inside the material. The electrical conductivity of a material is determined by the energy levels that electrons can occupy inside it, which are in turn determined by the atomic structure of the material. Ranges³ of possible energy states for electrons in a material are called *bands*. The bands relevant for electrical conduction are called **valence band** and **conduction band** (illustrated in Fig. 2.3). Electrons that reside in the valence band are bound to the atoms making up the material lattice while in the conduction band, electrons can move freely inside the lattice. In conductors, both these bands overlap or are very close, so electrons can easily move from the valence band to the conduction band. In insulators, there is a large energy gap, called **band gap** between the valence and the conduction band. Moving electrons from the valence band to the conduction band would thus require large amounts of energy, for example in the form of heat, or very high voltages, that would generally destroy the material. In semiconductors, the valence and conduction band do not overlap, but the band gap is small enough to be passed by electrons under certain circumstances that provide them with a sufficient amount of energy, usually in the order of a few electronvolts. For example silicon, a very commonly used semiconductor material in imaging sensors, has a band gap of $E_g^{\text{Si}} = 1.1 \text{ eV}$ at room temperature ($T = 300 \text{ K}$). [13]

In order to use a semiconductor material to build a light-sensitive sensor, the material is **doped** which means that impurity atoms are introduced into the lattice structure of the material (see Fig. 2.4). For this doping, elements which have one more or one less valence electrons in the outer shell than the semiconductor element are used. For example in silicon, which has four valence electrons, atoms of an element with five valence atoms, e.g. arsenic or lithium, is used to create a so-called **n-doping**. These foreign atoms in the silicon lattice are called **donors** because they provide an additional electron compared to the silicon. Due to the surrounding lattice structure, this electron is not strongly bound to the atom

³ In reality, all possible energy states are discrete, but for practical purposes closely adjacent levels can be treated as a continuous band. [13]

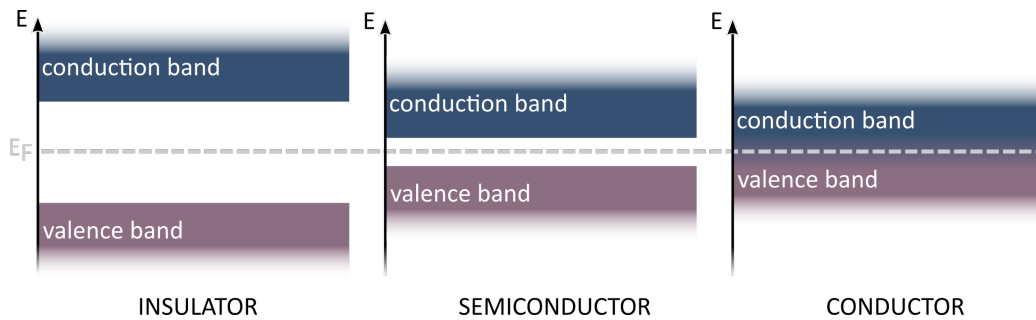


Figure 2.3: Band model for insulators, semiconductors and conductors. E_F is called the Fermi level and denotes the (hypothetical) energy level which is occupied with a probability of 50% at thermal equilibrium [14].

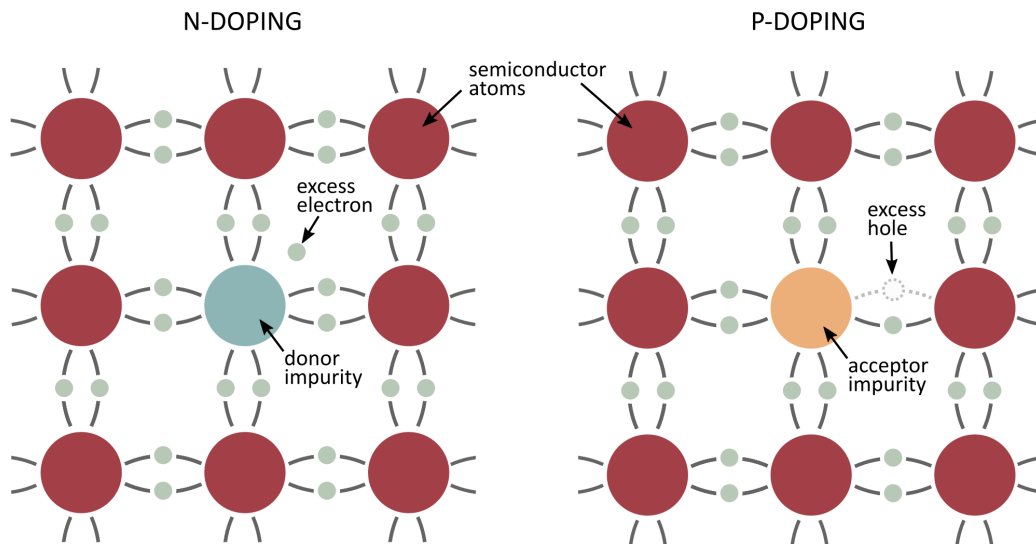


Figure 2.4: Doping of a semiconductor by introducing impurity atoms into the lattice. Left: n-doping where donors provide excess electrons. Right: p-doping where acceptors provide excess holes in the lattice structure. The illustration is based on [15].

and introduces an additional energy level slightly below the conduction band. Analogously, foreign atoms with only three valence electrons can be used for **p-doping**. They are called **acceptors** because they leave a “hole” in the lattice structure which creates an energy level slightly above the valence band. Doped semiconductors with a very low or very high doping concentration are usually denoted p^-/n^- and p^+/n^+ , respectively.

When an n-doped and a p-doped semiconductor are brought into contact, they form a **p-n junction**, as illustrated in Fig. 2.5. Around this junction, the

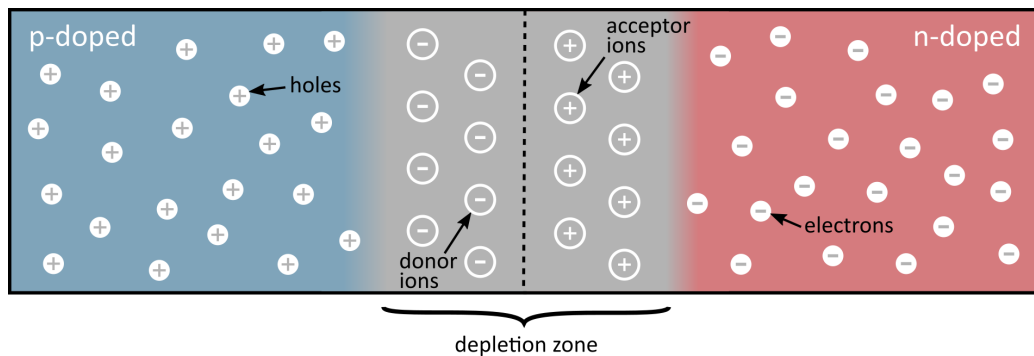


Figure 2.5: Schematic of a p-n junction with the p-doped semiconductor on the left and the n-doped semiconductor on the right. Indicated in grey is the depletion zone which contains the ionized impurity atoms. The density of holes and electrons drops exponentially in the depletion zone [13].

donor electrons from the n-doped region combine with the holes in the p-doped region, creating an insulating **depletion zone** that lacks free charge carriers.⁴ In order for electrons to pass this layer, sufficient energy needs to be provided by an applied voltage. The conduction behaviour of the p-n junction depends on the polarity of the voltage though. While a **forward bias** voltage narrows the depletion zone and allows charge carriers to pass through, a **reverse bias** voltage pulls electrons from the n-doped region even further into the p-doped region, enlarging the depletion zone. No current flows through the junction in this case – unless the voltage is high enough for charge carriers to overcome the insulating layer. The minimum voltage in reverse direction necessary to cause such a breakdown current is called the **breakdown voltage**.

Photodiodes are light detecting devices that are based on a p-n junction. When light hits the depletion region of a p-n junction, a photon can transfer its energy to a valence electron and lift it into the conduction band – as long as it carries enough energy to overcome the band gap. In the valence band, the electron leaves a hole, which is why the resulting state is called an **electron-hole pair**. Due to the space charge in the depletion region, electrons and holes are separated immediately, so instead of recombining they drift to opposite sides of the diode, effectively creating a measurable **photo current**. In an ideal photodiode which

⁴ Not all free electrons combine with all holes, because the doping atoms become ionized and form a space charge that induces an electric field in the depletion zone. The width of the depletion zone is determined by the equilibrium between the attracting forces of the free charge carriers and the electric field created by the doping ions. The density of charge carriers drops exponentially inside the depletion zone. [13]

converts all incident photons into electron-hole pairs (meaning it has a **quantum efficiency** of 100%), this current is proportional to the number of photons and therefore the light intensity. Photodiodes used to measure light intensity are usually operated in reverse bias mode in order to increase the size of the depletion zone and to reduce saturation effects. They constitute the basic element that the imaging technologies described in the following sections are built upon. [14, 16]

2.2.1 IMAGE SENSORS

Conventional image sensors used to measure two-dimensional optical signals are essentially arrays of photodiodes, where each photodiode is called a **pixel**. In classical image sensors, these diodes are operated in such a way that photoelectrons are not immediately measured, but collected in a potential well inside the pixel. This well is formed by potential barriers that can be shaped and adjusted by applying an external voltage. During the exposure time, photoelectrons are produced in the pixel as a result of the incoming photon flux and accumulate in the potential well. The maximum number of electrons that can be stored in each pixel, the *full well capacity*, determines the sensor's **dynamic range**, which describes the ratio between the largest and the smallest signal intensity that can be measured. The more photoelectrons can fit into the potential well of each pixel, the more light can be detected before the pixel is saturated. Two main categories of conventional image sensors can be defined by the circuitry design used to read out the accumulated charges from the pixels: **charge-coupled devices (CCD)** and **complementary metal-oxide-semiconductor (CMOS)** technology.

In CCDs, the sensor architecture allows to pass collected charges from one pixel to the next. This way, the charges are shifted into a transfer register row by row and from there, pixel by pixel, into a charge amplifier that outputs a voltage proportional to the number of measured photoelectrons that can then be converted into digital pixel values using an analog-to-digital converter (ADC). A special type of CCD sensor is the **emCCD (electron multiplying charge-coupled device)** which features an additional high-voltage electron-multiplication register right before the readout, providing a high signal gain that allows imaging at very low light intensity. [14]

Such a device is used in Chapter 4 as the temporally integrating image sensor whose signal is combined with that of a time-resolved SPAD sensor in order to obtain high-spatial-resolution time-resolved images.

CMOS sensors avoid the time-consuming charge-shifting readout process by providing individual readout electronics at every pixel, which is why it is also called an *active pixel sensor*. This increase in readout speed, however, comes at the

cost of slight gain and noise variations between the pixels and usually a smaller fill factor due to the required electronics taking up space on the sensor area. The fill factor is defined as the ratio of photosensitive area to the whole sensor area and is usually below 1. Some sensors employ microlens arrays where a tiny lens in front of each pixel collects light and focuses it onto the sensitive region in order to increase the quantum efficiency. [16]

Special image sensors whose capabilities exceed the simple integration of light are described in the following section.

2.3 TIME-OF-FLIGHT SENSING

A conventional camera, as described above, collects photoelectrons produced from light that is reflected or emitted by a scene and that falls onto the sensor during a certain exposure time. This time is usually much longer than a photon travels inside the scene, so the recorded image is an integration over all light transport that happens during the exposure. Time-of-Flight cameras, instead of simply integrating temporally over the light intensity, feature a light source that illuminates the scene and are capable of measuring the time that the light took to travel from the source through the scene and back to the sensor. There are multiple techniques to achieve this high level of temporal resolution which vary significantly in cost, flexibility, and resolution.

ToF imaging principles can be divided into two different categories: *indirect* and *direct* ToF imaging (*I-ToF* and *D-ToF*). While indirect ToF cameras use a continuous, amplitude-modulated light signal, direct ToF systems use a light source that sends out ultra short pulses of light. I-ToF systems then measure a phase difference between the emitted and received signal, similar to the concept of LIDAR⁵, while D-ToF systems directly measure the elapsed time between the emission and detection of light. An important difference between the two technologies is that I-ToF systems yield a single flight time value per pixel, hence the resulting measurement data is a two-dimensional image with a flight time (usually corresponding to a depth via $d = c \cdot t$) value per spatial pixel. D-ToF systems on the other hand are generally capable of measuring the full light response profile of a point or scene: the light source sends out pulses of photons and the time between their emission and detection at the sensor is measured. This process is repeated multiple times and a histogram of measured flight times is collected. Its shape characterizes the light response of the observed scene – a completely opaque

⁵ Light Detection And Ranging

object, for instance, would result in a sharp peak in the histogram because all light is reflected at the same point. When the object is translucent, though, light is also transmitted inside the material and scattered there, hence travelling longer paths before it reaches the sensor, which is reflected in the measurement histogram as non-zero values after the initial surface reflectance peak. The same holds for inter-reflections in a scene or light that passes through semi-transparent objects. Using D-ToF technologies, the full light response of a scene can be recorded as a three-dimensional data tensor (*data cube*) with two spatial dimensions and one temporal dimension, also called a **transient image** or **light-in-flight image**.

The work presented in this thesis uses two types of ToF imagers: **Amplitude-Modulated Continuous-Wave (AMCW) ToF** imagers (I-ToF) and **Single-Photon Avalanche Diode (SPAD)** imagers (D-ToF), which are therefore both explained in more detail in the following sections.⁶

Another important technology suitable for recording transient images with a very high temporal resolution of up to 180 femtoseconds are **streak cameras** [20] which work by accelerating a line of photoelectrons in a tube and deflecting them by an amount that is proportional to their time of emission. This way, a two-dimensional image is formed which features one spatial and one temporal dimension, depicting the temporal light response of a narrow line of the observed scene. In order to record full images with two spatial dimensions though, the scene needs to be scanned line by line and the resulting images stacked to form a three-dimensional “transient” data tensor, or the optical signal needs to be spatially encoded in order to allow for a later reconstruction of the two spatial dimensions [21, 22]. Additionally, streak cameras are generally expensive and bulky, which is why SPADs are a preferred technology in many application scenarios despite their lower temporal resolution, which is in the order of tens of picoseconds [19].

2.3.1 AMPLITUDE-MODULATED CONTINUOUS-WAVE TOF SENSING

The idea of modulation-based ToF (also called **correlation ToF** or **lock-in ToF**) imaging is to illuminate the scene with an amplitude-modulated signal with a fixed modulation frequency. The special camera sensor recording the scene is

⁶ Heide et al. [17] have shown that continuous-wave ToF systems can also be used to record transient images, although the achieved temporal resolution is much lower than that of designated D-ToF systems. Conversely, SPAD sensors, which are generally considered a D-ToF technology, can be used in I-ToF mode by sampling a modulated illumination signal similar to AMCW systems (Single-Photon Synchronous Detection - SPSD) [18, 19].

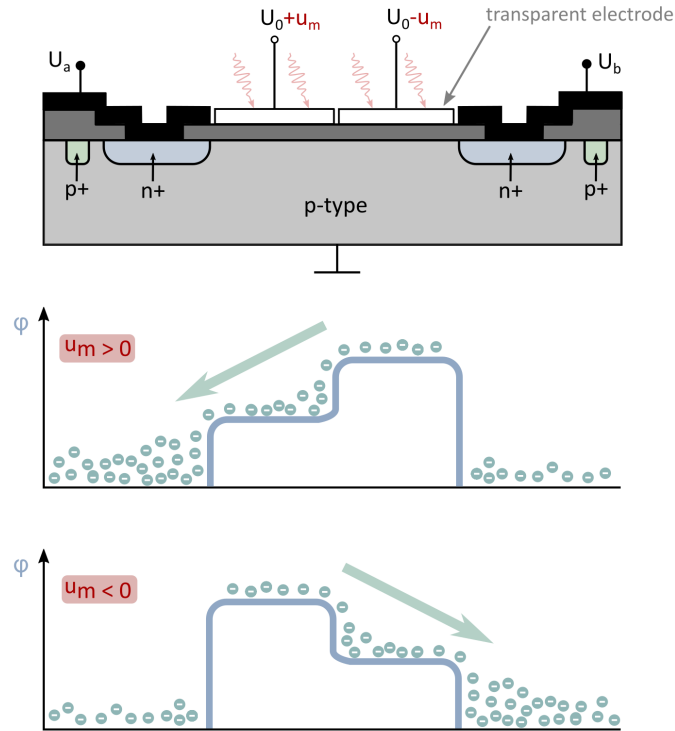


Figure 2.6: Schematic sketch of the structure and photo-electron collection mechanism of a PMD pixel: Depending on the polarity of the modulation signal u_m , the potential inside the pixel follows an asymmetric distribution that allows electrons to drift either to the left or to the right side where they're collected in a deeper potential well at the readout diodes a and b. The illustration is based on [23].

operated with the same modulation signal and can detect the phase difference $\Delta\varphi = 2\pi f_{\text{mod}}\Delta t$ between the emitted and the received signal, where f_{mod} is the modulation frequency and Δt is the time that has passed between emission and detection of the light. From this, the travelled light path can be easily calculated as $s = c \cdot \Delta t$; and the resulting depth, when light source and camera are located close to each other, is then simply $d = s/2$. Since different periods of the illumination signal can not be distinguished, the maximum range that can be measured unambiguously with this setup is given by one wavelength of the illumination signal. For a typical modulation frequency of 20 MHz, this corresponds to $\lambda_{\text{mod}} = c/f_{\text{mod}} = 15 \text{ m}$, the maximum depth range is therefore 7.5 m.

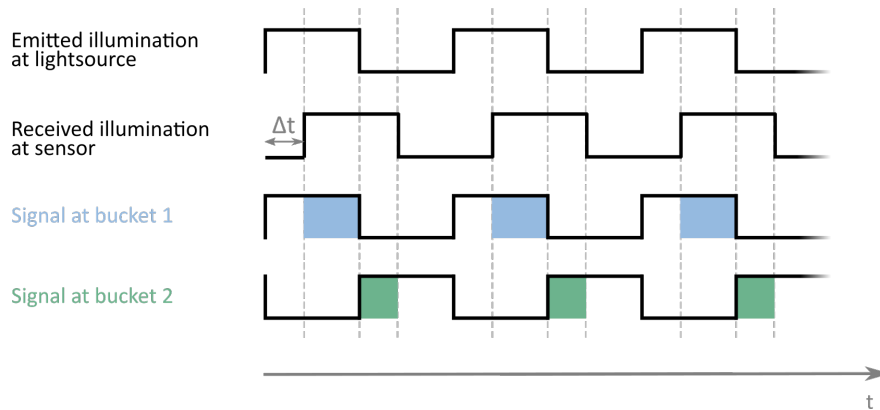


Figure 2.7: Measurement principle of a correlation ToF system: A light source emits an amplitude-modulated signal which undergoes a phase shift during the time Δt that the light takes to travel between light source, scene and sensor. Two ‘buckets’ inside the pixel are operated with inverse modulation signals which allows to determine the phase shift from the ratio of the two signals.

The detection of the phase shift $\Delta\phi$ at the sensor is based on the photonic mixer device (PMD) [23]. It is a special kind of photo-diode illustrated in Fig. 2.6: instead of collecting the produced photo-electrons in a single potential well during the exposure that is then later read-out, it features two transparent electrodes at the photosensitive area, whose voltage is modulated with a modulation signal u_m . Depending on the polarity of this modulation signal, the potential inside the pixel forms a slope so that photoelectrons drift either to the left or the right side of the pixel where they are stored in separate potential wells (‘buckets’) that can be read out individually. Figure 2.7 illustrates the measurement principle: The amount by which the illumination signal has been shifted due to the travel time Δt from the light source (to the scene and then) to the sensor can be determined by the ratio of signals collected in the first and the second bucket.

In real scenarios, however, ambient light produces an intensity offset in the measured signals. This can either be compensated by acquiring a measurement without active illumination and subtracting it from each bucket’s signal, or by employing a more complex measurement scheme where the difference of the two buckets’ signals is calculated (effectively removing the offset) and the signal is sampled four times, each measurement with a different fixed phase shift (0° , 90° , 180° , and 270°) to the original modulation signal [24]. From these samples, the phase, amplitude and offset of the measured signal can be calculated [25]. This method allows to also compensate for different gains at the readout of the two

buckets [26] and is explained in more detail in [24] and [27]. The ToF sensor's capability to calculate the difference of the two buckets in hardware, before the read-out, is exploited in Chapter 3 where correlation ToF sensors are used as general difference imagers.

2.3.2 SPAD SENSING

Avalanche photodiodes (APDs) are special photodiodes that are operated with a strong reverse-bias voltage that accelerates photo-electrons in the depletion zone after they have been produced by incident photons. This way they gain enough energy (more than the width of the band gap) to excite new electrons in the semi-conductor by impact ionization. These are again accelerated and excite even more charge carriers, effectively creating an avalanche that produces a strong measurable signal. Due to this multiplication process, similar to a photomultiplier, even very low light intensities can be measured.

A special type of APD is the Single-Photon Avalanche Diode (SPAD) whose bias voltage is so high that even a single electron-hole pair is sufficient to initiate a self-sustaining avalanche that creates a large current pulse. Due to this binary photon detection mode, where each detected photon creates a single countable pulse, it is also called **Geiger-mode APD** because the working principle is similar to that of a Geiger counter for ionizing radiation. Using a light source that emits very short pulses of light, the time between the emission of a photon and its detection at a SPAD sensor can be measured. Repeating this measurement, a histogram of photon flight times can be formed, yielding a characteristic light response distribution. Therefore, while correlation ToF systems only yield a single time (or depth) value per pixel, a SPAD system yields a temporally resolved light response signal as illustrated in Fig. 2.8. In a two-dimensional array of SPADs (a *SPAD sensor*), each pixel yields such a light response histogram, so the full SPAD image is a three-dimensional data cube with two spatial dimensions and one temporal dimension.

This time-resolved light response data enables a wide variety of applications. Besides direct depth imaging by finding the reflection peak, SPADs can, for instance, also be used to record fluorescence decays in Fluorescence Lifetime Imaging Microscopy (FLIM), where a pulsed light source excites a fluorescent material and the SPAD measures the decay profile from which the fluorescence lifetime can be extracted. The information contained in a SPAD histogram can also be used to distinguish multiple reflections, enabling imaging of objects outside the direct line of sight by observing the light signal reflected by a diffuse wall that is in view of the target scene.

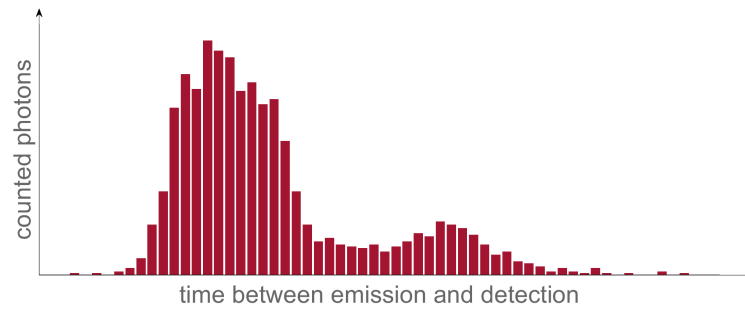


Figure 2.8: Example of a light response distribution as it could be measured with a SPAD setup. The shape of the histogram depends on the light response profile of the imaged scene, object, or point.

Like a Geiger counter, a SPAD needs to be quenched after each avalanche [28]. Therefore, after the detection of a photon, the SPAD is insensitive for a certain **dead time** which limits the maximum photon count rate of the detector and can create a bias towards lower photon flight times in ToF histograms. Additional to this dead time, other relevant parameters for the performance of a SPAD are its sensitivity (photon detection probability), dark count rate, timing jitter and afterpulsing probability (a time-correlated type of noise caused by fabrication defects in the diode) [19]. Arrays of SPADs can generally be implemented in CMOS technology, which however introduces additional performance parameters, mainly concerning the uniformity of the aforementioned properties, as well as crosstalk between pixels [19].

Chapter 4 of this thesis aims at increasing the spatial resolution of data measured with current SPAD sensors, and Chapter 5 explores the capabilities of very cheap SPAD devices to perform challenging imaging and vision tasks.

The content of the following chapter (Chapter) has been published as:

Clara Callenberg, Felix Heide, Gordon Wetzstein, and Matthias B. Hulin. 2017. *Snapshot difference imaging using correlation time-of-flight sensors*. ACM Trans. Graph. 36, 6, Article 220 (November 2017), 11 pages. DOI:<https://doi.org/10.1145/3130800.3130885>.

SNAPSHOT DIFFERENCE IMAGING USING CORRELATION TIME-OF-FLIGHT SENSORS

ABSTRACT Computational photography encompasses a diversity of imaging techniques, but one of the core operations performed by many of them is to compute image differences. An intuitive approach to computing such differences is to capture several images sequentially and then process them jointly. In this paper, we introduce a snapshot difference imaging approach that is directly implemented in the sensor hardware of emerging time-of-flight cameras. With a variety of examples, we demonstrate that the proposed snapshot difference imaging technique is useful for direct-global illumination separation, for direct imaging of spatial and temporal image gradients, for direct depth edge imaging, and more.

This work was supported by the German Research Foundation (HU-2273/2-1), the X-Rite Chair for Digital Material Appearance, a National Science Foundation CAREER award (IIS 1553333), a Terman Faculty Fellowship, and the KAUST Office of Sponsored Research through the Visual Computing Center CCF grant. We thank Nick Maggio for his help on early experiments.

3.1 INTRODUCTION AND MOTIVATION

Over the last two decades, research in computational photography has been striving to overcome limitations of conventional imagers via a co-design of optics, sensors, algorithms, and illumination. Using this paradigm, unprecedented imaging modalities have been unlocked, such as direct-global light transport separation [29], gradient imaging [30], temporal contrast imaging [31], and direct depth edge imaging via multi-flash photography [32]. A common operation for many of these techniques is to record two or more images and then compute the

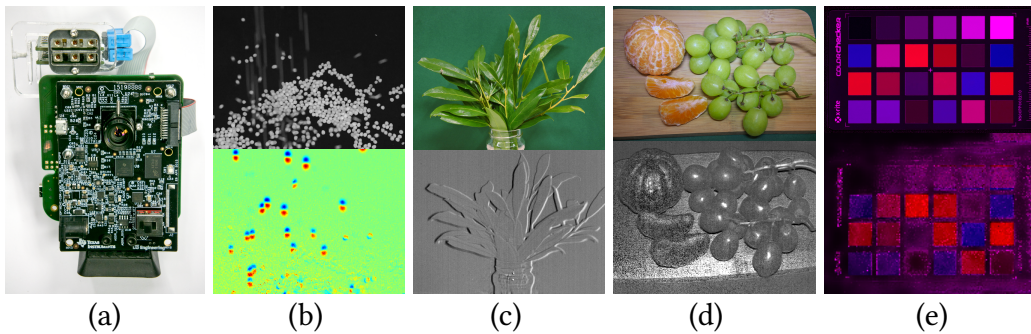


Figure 3.1: We introduce a method that uses time-of-flight (ToF) imagers not for measuring scene depth, but rather as analog computational imagers that can directly measure difference values at a pixel level. We demonstrate this principle with a slightly modified ToF camera system (a), where simple reconfigurations in the optical path enable a wide range of imaging modalities. For instance, our system can directly sense temporal gradients (b), depth edges (c), direct-light-only images (d) and spatial gradients (Fig. 3.10) – each in a single exposure and without any additional decoding steps. We further show that the remarkable noise statistics of such imagers can be exploited to extract two color channels (here: red and blue) from a single snapshot taken under red *and* blue illumination (e). The top images in columns (b)–(e) are reference photographs of the respective scenes; the bottom ones visualize the output of our system.

difference between them. Unfortunately, difference imaging is challenging for dynamic scenes, because motion creates misalignment between successively captured photographs which is in many cases difficult to mitigate in post-processing. In this paper, we explore a new approach to capturing difference images in a single exposure and generalize difference imaging to a variety of applications.

We propose to re-purpose time-of-flight (ToF) sensors to facilitate instantaneous difference imaging. The usual application for these sensors is depth imaging. In that context, they are operated in conjunction with a periodically modulated light source. Light that has been reflected by the scene is demodulated by the sensor, reconstructing the shift in modulation phase and thereby the depth estimate per pixel. This functionality is achieved by a pixel architecture that employs two potential wells for photoelectrons to be stored in during the exposure, and that subtracts the charges accumulated in these two wells (Fig. 3.2). In other words, the core functionality of time-of-flight sensors is based on being able to take the difference of two incident signals before analog-to-digital (A/D) conversion.

Rather than computing scene depth, we demonstrate how ToF sensing technology can be used to conveniently implement a range of computational photo-

graphy techniques, including direct-global separation, direct depth edge imaging, spatio-temporal gradient imaging, and more. The capabilities unlocked with snapshot difference imaging are particularly interesting for applications that require low-power, low-latency on-board processing with low bandwidth communication channels, such as internet-of-things devices. With this paper, we take first steps towards these directions.

Specifically, our contributions are the following:

- We introduce the concept of generalized difference imaging and develop an image formation and a noise model for this principle.
- We construct a prototype difference imager using a modified time-of-flight camera combined with multiple, spatio-temporally coded light sources.
- We evaluate the proposed imaging concept with several practical applications, including direct-global separation, direct depth edge as well as spatio-temporal gradient imaging.
- We demonstrate that two images can be recovered from a single difference image by exploiting characteristics of the proposed image formation model.

OVERVIEW OF BENEFITS AND LIMITATIONS The proposed method has two primary benefits. First, capturing a difference image within a single exposure allows for faster time scales to be recorded than capturing two separate images and subtracting them digitally. Second, the noise properties of difference imaging before A/D conversion are shown to be favorable over digital subtraction post A/D conversion. A limitation of the proposed technique is that it relies on ToF sensors, which currently provide much lower resolution and signal quality than well-established CMOS or CCD sensors. Thus, comparing digital difference imaging with CMOS sensors and analog difference imaging with ToF sensors may not be beneficial for the latter approach. Yet, we demonstrate that our method yields superior noise performance for sensors with comparable characteristics.

3.2 RELATED WORK

COMPUTATIONAL TOF IMAGING This work presents a method for difference imaging by re-purposing two-bucket sensors usually used for depth imaging in lock-in ToF cameras. Lock-in time-of-flight sensors are a rapidly emerging sensor technology, with Microsoft’s Kinect for XBOX One as the highest-resolution sensor available on the market at 512×424 pixels [33]. For technical details on

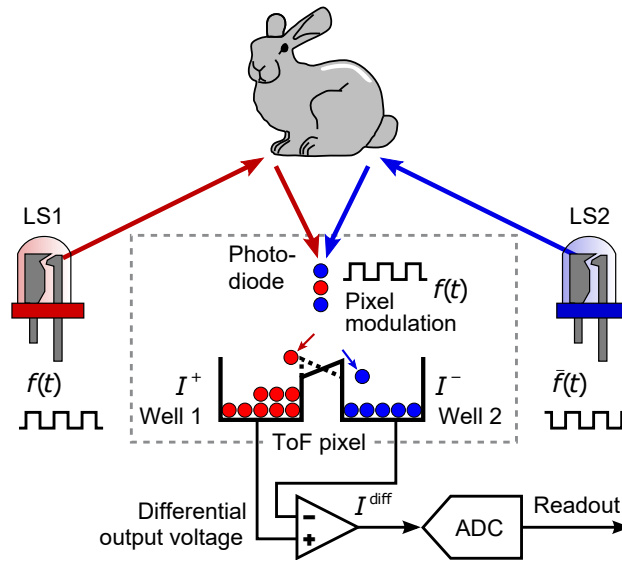


Figure 3.2: Principle of operation of a time-of-flight (ToF) pixel. A light source is temporally modulated, its emitted light is reflected by the scene, and then demodulated in the pixel. To demodulate the coded illumination in the detector, two wells in each pixel collect charge carriers and an electric field oscillates at the demodulation frequency to direct incident photoelectrons into one or the other well. The sensor circuit measures the voltage difference before digitizing it by an analog-to-digital converter (ADC). Here, we illustrate how the difference between two modulated light sources can be directly measured with such a pixel architecture.

lock-in ToF sensors we refer the reader to Lange et al. [34] and Hansard et al. [35]. A growing body of literature re-purposes these emerging sensors, in combination with computational methods, to address a variety of challenging problems in imaging and vision. Kadambi et al. [36] reduce multi-path interference by coding the modulation profiles, Heide et al. [17] recover temporal profiles of light transport by measuring frequency sweeps, which allows for improved imaging in scattering media [37]. Recently, Tadano et al. [38] designed depth-selective modulation functions enabling virtual-blue screening and selective back-scatter removal as applications.

DIFFERENTIAL-PIXEL SENSORS The proposed difference imaging method subtracts two signals before A/D conversion by “piggybacking” on two-bucket ToF sensor technology. Wang et al. [39] have previously proposed a custom sensor design that also performs pre-ADC subtraction for the purpose of optoelectronic filtering and

light field capture. Specifically, the authors use programmable gain operational amplifiers to compute the sum and difference of pixel pairs, which is then passed on to the A/D converter. In combination with local diffractive gratings as optics on every pixel, this allows to realize filtering with positive and negative filter coefficients. In contrast to a conventional sequential capture approach, these differential-pixel sensors offer reduced bandwidth [40], at the cost of spatial resolution. Compared to the proposed method, the optical filters are static and prohibit the flexible modes of operation demonstrated in this work. Changing from one difference-imaging task to another would require placing a different mosaicking pattern on the sensor.

DIFFERENTIAL OPTICAL SYSTEMS Instead of this optoelectronic approach to difference imaging, one could also imagine cameras that perform the signal subtraction purely optically, plus a DC offset to ensure positivity. Building on Zomet and Nayar’s work [41], Koppal et al. [42] present an optical design consisting of a micro-aperture mask in combination with lenslets allowing to design custom optical template filters for a variety of computer vision tasks. This approach may be adopted to design optical filters that perform spatial gradient or other filter differences in a single-shot, by designing difference filters with a DC offset to ensure non-negative coefficients. In theory, this approach would require variable high-resolution aperture patterns [41]. Note also, that the proposed approach would be a natural choice for suppressing the DC in such a setup by relying on the adaptive background suppression of recent ToF sensors.

EVENT-BASED SENSORS A further sensor design for differential measurements are event-based dynamic vision sensors [31, 43], which have been demonstrated for applications in vision and robotics, such as tracking [44] and simultaneous localization and mapping (SLAM) [45]. Each pixel in such sensors asynchronously measures temporal intensity changes and generates spike events for temporal differences with magnitude above a given activation threshold. This event-driven behavior is achieved by augmenting each pixel with its self-timed switched-capacitor differencing circuit. By reading out pixels asynchronously, the core benefit of this sensor design is the the low bandwidth requirement, enabling high frame rates and low power consumption [31, 43]. However, similar to the differential-pixel sensors, this comes at the cost of reduced spatial resolution, when compared to conventional sequential capture. While the proposed solution, based on ToF sensors, shares limitations in resolution, temporal differencing sensors do not support the very flexible modes of operation shown in this work.

For example, capturing intensity or depth images requires solving ill-posed inverse problems [44].

SPLIT-PIXEL HDR SENSORS Backside-illuminated split-pixel architectures have become the dominant choice for high-dynamic-range (HDR) imaging in high-speed vision cameras [46]. Single-shot HDR capture is essential for vision-based autonomous or assisted driving systems where reacting to fast moving objects over a wide dynamic range is critical [47]. A variety of HDR sensor designs for high frame rates have been proposed in the past. Skimming HDR sensors perform partial resets (draining) of the accumulated charges during integration, allowing repeated partial integration with successively shorter resets [48]. The repeated integration can cause motion artefacts if partial saturation are reached quickly. Split-pixel architectures eliminate this issue by dividing each pixel into multiple buckets [49, 50]. Multiple exposures are captured simultaneously by implementing different-sized photosensitive areas (OmniVision OV10640, OV10650). Given the emerge of split-pixel architectures as a key vision sensor technology, we believe that the proposed two-bucket difference imaging method may have broad applications even beyond the ones in this work.

3.3 IMAGING PRINCIPLE

Of all technologies that can be used for time-of-flight imaging, correlation sensors are the most widespread and affordable. This is also the type of imager we are using in this work; throughout the paper, we use the term “time-of-flight (ToF) sensor” synonymously for this particular technology.

A pixel in a ToF sensor measures the amount of correlation between the incoming temporally varying photon flux $g_i(t)$ and a sensor modulation signal $f(t) \in [0, 1]$ that also varies in time [34]. Unlike regular CCD or CMOS sensors where electrical charges generated in a photodiode are collected in a potential well, ToF sensors feature two such wells per pixel (Fig. 3.2). The sensor modulation $f(t)$ decides whether a charge generated at time t will tend to end up in one well or the other. At the end of the integration phase, the difference between the two wells is read out and digitized. Neglecting quantization from the A/D conversion, this results in the digital value

$$\hat{I}^{\text{diff}} = \rho \cdot \eta \cdot (\hat{I}^+ - \hat{I}^-), \quad (3.1)$$

where ρ is the conversion factor from electron counts to digital units, and η denotes the so-called *demodulation contrast* [51]. \hat{I}^+ and \hat{I}^- are the photoelectrons collected in the two wells over the integration period $[0, T]$:

$$\begin{bmatrix} \hat{I}^+ \\ \hat{I}^- \end{bmatrix} = \int_0^T \begin{bmatrix} f(t) \\ 1 - f(t) \end{bmatrix} g_i(t) dt \quad (3.2)$$

The incoming photon rate $g_i(t)$ is a function of the scene and the time-varying intensity $g(t)$ of an active light source that illuminates it. In ToF imaging, $f(t)$ and $g(t)$ are periodic functions of the same high frequency, typically 20–100 MHz, and the delay of light propagating from a source to the sensor results in a relative phase shift which is measured to recover depth. In snapshot difference imaging, we introduce two modifications to this scheme. Firstly, we reduce the modulation frequency to a point (1–5 MHz) where the propagation of light through near-range scenes can be assumed to be instantaneous and $f(t)$, typically generated by a digital circuit, only assumes the values 0 and 1. Secondly, we use two light sources, one (LS1) driven using the same function $f(t)$ and the other one (LS2) with its logical negation $\bar{f}(t)$. According to Eq. 3.2, the photocharges collected in \hat{I}^+ will record an image of the scene as illuminated by LS1, and LS2 will fill \hat{I}^- . The pixel value \hat{I}^{diff} thus measures the difference between two images taken under different illumination conditions, an insight that forms the foundation of this work.

3.3.1 NOISE MODEL

Time-of-flight imagers are complex photonic devices and as such suffer from noise of various different sources [51]. The differential measurement scheme, and in particular the multi-tap measurement schemes typically used in ToF operation, cancel out many of the systematic errors introduced by the hardware. None of these measures, however, are capable of removing shot noise, which is the uncertainty that occurs during the counting of photoelectrons.

If \hat{I}^\pm are the *expected* electron counts for the two wells, the *actual* number of collected electrons I^\pm in any image recorded is a Poisson-distributed random variable with mean μ^\pm and variance $(\sigma^\pm)^2$ that are both identical to the respective expected value:

$$\mu^\pm = (\sigma^\pm)^2 = \hat{I}^\pm \quad (3.3)$$

As the difference of two independent random variables, the final pixel value is also a random variable, and it follows a Skellam distribution [6, 52]. Mean μ_{diff} and variance σ_{diff}^2 relate to the means μ_{\pm} and variances σ_{\pm}^2 of I_{\pm} as

$$\mu_{\text{diff}} = \eta (\mu_+ - \mu_-) = \eta (\hat{I}_+ - \hat{I}_-) \quad (3.4)$$

$$\sigma_{\text{diff}}^2 = \eta^2 (\sigma_+^2 + \sigma_-^2) + \sigma_{\text{read}}^2 = \eta^2 (\hat{I}_+ + \hat{I}_-) + \sigma_{\text{read}}^2 \quad (3.5)$$

where σ_{read}^2 models additional noise sources (assumed to be zero-mean), and the device constant $\eta \in [0, 1]$ is the imager's contrast [51]. In terms of a matrix-vector product:

$$\begin{pmatrix} \mu_{\text{diff}} \\ \sigma_{\text{diff}}^2 - \sigma_{\text{read}}^2 \end{pmatrix} = \underbrace{\begin{pmatrix} \eta & -\eta \\ \eta^2 & \eta^2 \end{pmatrix}}_H \begin{pmatrix} \hat{I}_+ \\ \hat{I}_- \end{pmatrix}. \quad (3.6)$$

Note that the uncertainty σ_{diff}^2 of the measurement I^{diff} depends not primarily on the net difference value, but rather on the latent components \hat{I}^{\pm} that are subtracted from one another. Thus, even when there is zero signal ($\hat{I}^{\text{diff}} = 0$), the actual observation I^{diff} can suffer from significant noise. This is a principal property of difference imaging and holds for all sorts of settings, applying to in- as well as post-capture differencing techniques. We call the system-specific matrix H the *Skellam mixing matrix*. In Section 3.6, we show how it can be used to recover two color channels from a single exposure — an insight that, to our knowledge, has not been discovered before.

3.4 PROTOTYPE DIFFERENCE IMAGER

We constructed snapshot difference imagers based on two different time-of-flight sensing platforms. Our first prototype (not pictured) is a recreation of Heide et al.'s system [17] that is based on the discontinued PMD Technologies CamBoard nano. The second prototype (Figs. 3.1 and 3.3) combines the Texas Instruments (TI) OPT8241-CDK evaluation module with external modulation and light sources in a similar way to the system described by Shrestha et al. [53]. Both imagers have their infrared-pass filters removed so they can sense visible light. (For the TI sensor, we carefully polished the filter coating off using a Dremel 462 rotary tool.) To enable difference imaging, the external light sources are configured to operate in and out of phase, respectively. The systems are configured to capture at an exposure time of 2000 μs (PMD) and 1000 μs (TI) and 60 frames per second.

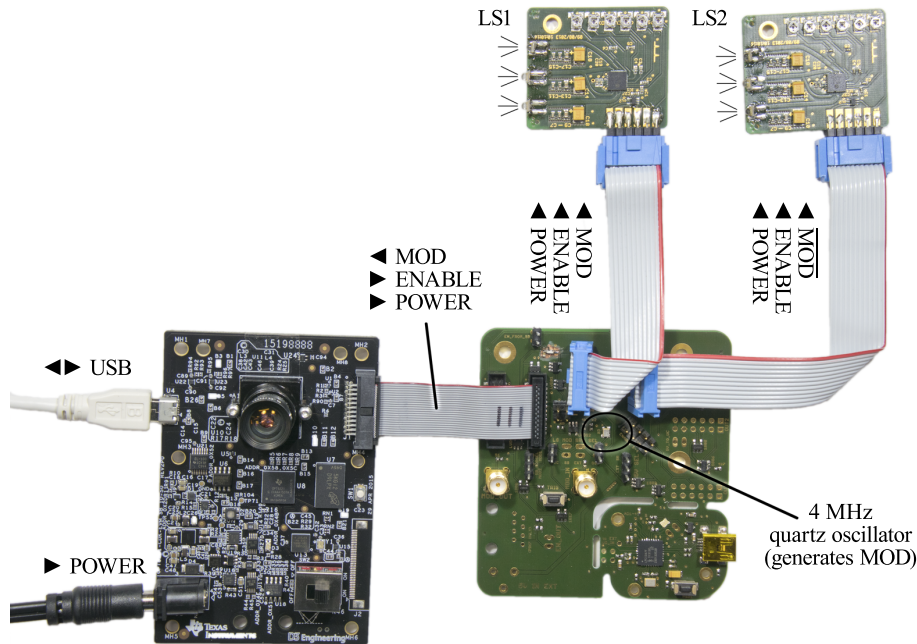


Figure 3.3: Components of our system based on the Texas Instruments OPT8241-CDK module (bottom left). An external function generator (bottom right) produces a relatively low-frequency square wave, MOD, and its negated version, $\overline{\text{MOD}}$, that correspond to $f(t)$ and $\tilde{f}(t)$ and modulate the sensor and two light sources (top right). The ENABLE signal is held high by the camera during integration.

Each of our light sources carries three OSRAM OSOLON LEDs that are switched using the same signal. To implement the different imaging modalities described in Section 3.5, we equipped the light sources with LEDs of different colors (red, blue, infrared), placed them in different positions and equipped LEDs and camera with polarization filters as required for each particular purpose.

In Appendix 3.8, to lower the entry barrier for the reader, we describe an alternative modification for stock OPT8241-CDK modules that does not require any custom hardware.

MEASUREMENT PROCEDURE To reduce fixed pattern noise, a black frame was recorded before the data acquisition with our setup, and later subtracted from each measured frame. As a result, the difference image pixels obtain negative and positive values, depending on the charge balance of the two potential wells. This applies to all results shown in this paper.

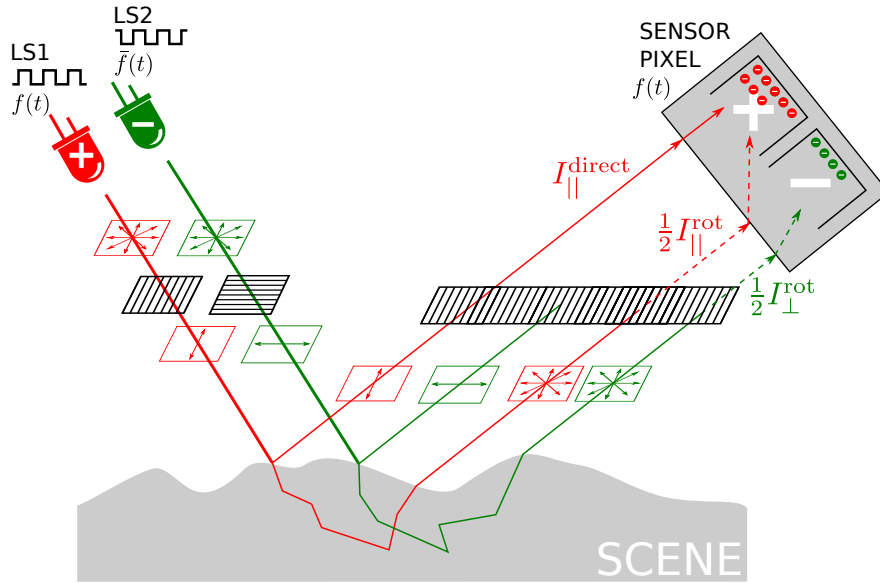


Figure 3.4: Polarization difference imaging principle: the light of two identical light sources is modulated in (+) and out (-) of phase, respectively, with the sensor. The light sources' polarization directions are mutually perpendicular, one them being aligned in parallel with the analyzer filter in front of the sensor. Light that is reflected directly off the surface of the scene preserves the polarization while the light that scatters multiple times in the scene becomes depolarized. The sensor measures the difference image $I^{\text{diff}} = I_{||}^{\text{direct}} + \frac{1}{2}I_{||}^{\text{rot}} - \frac{1}{2}I_{\perp}^{\text{rot}} = I_{||}^{\text{direct}}$.

3.5 APPLICATIONS AND EVALUATION

3.5.1 POLARIZATION-BASED DIRECT-GLOBAL SEPARATION

Many computational imaging techniques are based on a decomposition of light transport into a *direct* component, i.e., light that has undergone exactly one scattering event between light source and camera, and a multiply scattered *indirect*, or *global*, component. Being able to investigate these components separately has been shown to enable, for instance, more realistic digital models for human skin [54] or more robust 3D scans of objects made of challenging materials [55, 56]. While true separation into direct and indirect components is not within reach, researchers have used common observations about light transport to derive useful heuristics: indirect light tends to be spatially low-pass filtered [29], it generally does not fulfill the epipolar constraint [57] and does not preserve polarization [58]. Here, we use our setup to exemplarily implement the third of

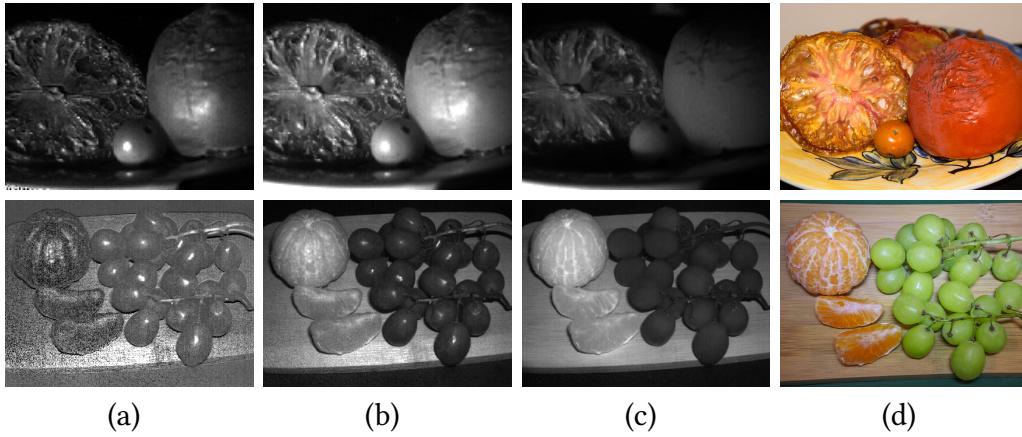


Figure 3.5: Polarization difference images of the two different scenes seen as RGB images in column (d). Column (a) shows the difference image I^{diff} containing only directly reflected light. The images in column (b) depict I^+ (parallel polarizers), the ones in column (c) I^- (crossed polarizers). The latter contains only light that has undergone multiple scattering events. The top dataset was captured with the PMD-based system, the bottom row with the TI sensor. The images have been rescaled for better visibility.

these heuristics in the form of a single-shot polarization-difference imager, and demonstrate its capability to isolate directly reflected light.

According to Fig. 3.4, illuminating the scene using two light sources with crossed linear polarization and an analyzing filter on the camera, one can consider four different components of the image:

$I_{\parallel}^{\text{direct}}$ light that initially passed the polarization filter parallel to the analyzer and that was reflected directly in the scene, hence preserving the orientation of the polarization,

$I_{\parallel}^{\text{rot}}$ light that initially passed the polarization filter parallel to the analyzer and that was scattered multiple times in the scene, thus not preserving the orientation of the polarization,

$I_{\perp}^{\text{direct}}$ light that initially passed the polarization filter perpendicular to the analyzer and that was reflected directly in the scene,

I_{\perp}^{rot} light that initially passed the polarization filter perpendicular to the analyzer and that was scattered multiple times in the scene.

Assuming that multiple scattering in the scene completely depolarizes the light for both initial directions of polarization, the amount of the light reaching the

camera after illumination with light polarized in parallel with the analyzer is the component that is in phase with $f(t)$:

$$I^+ = I_{\parallel}^{\text{direct}} + \frac{1}{2}I_{\parallel}^{\text{rot}},$$

while the amount of light reaching the sensor after illumination with light polarized perpendicularly to the analyzer is only the ratio of the incident light that has been depolarized and therefore passes the analyzing filter:

$$I^- = \frac{1}{2}I_{\perp}^{\text{rot}}.$$

The light sources are identical and we assume that the ratio between directly and indirectly scattered light is equal for both initial directions of polarization, hence $I_{\perp}^{\text{rot}} = I_{\parallel}^{\text{rot}}$. Therefore, the difference image that the PMD sensor acquires in lock-in operation with the light sources is

$$I^{\text{diff}} = I^+ - I^- = I_{\parallel}^{\text{direct}},$$

leaving an image containing only directly reflected light. Images of two sample scenes can be found in Fig. 3.5. Column (a) shows the difference image I_{diff} . Columns (b) and (c) show the parallel and perpendicular components $I_{\parallel}^{\text{meas}}$ (b) and I_{\perp}^{meas} (c), respectively, where the latter visibly contains only indirectly scattered light.

3.5.2 BIPOLAR COLOR MATCHING FUNCTIONS

Being able to characterize and classify materials is important in many applications. Liu and Gu [59] proposed to use discriminative illumination, or optimized pairs of spectro-angular illumination patterns to classify materials on a per-pixel basis. Here, we adopt the spectral aspect of this work, using our PMD setup to construct an active camera that discriminates between objects of red and blue reflectance in a single shot. By equipping L1 with red and L2 with blue LEDs, we obtain a *bipolar* color camera that measures a positive response for objects that are predominantly red, and a negative response for bluish objects. Fig. 3.6 shows an example measurement taken of the X-Rite ColorChecker, where the positive or negative response in the colored patches can clearly be seen. Patches that reflect red and blue to equal parts, like the greyscale, result in a response that is approximately zero.

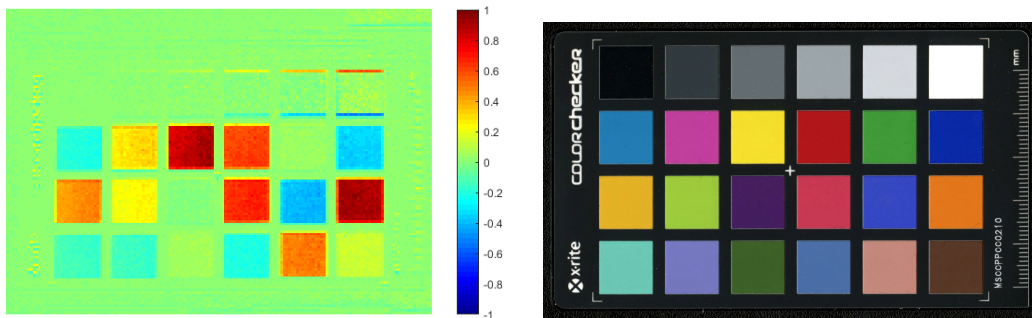


Figure 3.6: Left: Difference image of a color calibration chart, taken using PMD setup using alternating illumination of red and blue light. Reddish and bluish color patches obtain values on the opposite ends of the scale. Patches whose reflection spectra do not favor either red or blue obtain a value of approximately zero. Edges of some patches appear exaggerated due to partial shadowing of the light sources. Right: RGB scan taken of the color chart.

This example demonstrates the applicability of snapshot difference imaging to discriminative color imaging. The red–blue illumination we used is merely a simple example case. Multiple different light sources can be combined to obtain complex bipolar illumination spectra that can be individually adapted to a variety of applications. We also envision this capability to facilitate interesting and novel approaches to image segmentation and classification or to enable direct sensing of primary colors with bipolar matching functions, like the red primary $\bar{r}(\lambda)$ in the CIE 1931 RGB color space¹.

3.5.3 DEPTH EDGE AND DIRECTIONAL GRADIENT IMAGING

It is often hard to deduce the structure and shape of three-dimensional objects from conventional photographs, as they may show low contrast between spatially distinct features of the object. Illuminating the object from two different angles, however, can unveil the depth structure of a scene and facilitate, for instance, a segmentation of the image. Similarly to Raskar et al. [32], our setup can be used to produce directional gradient images of a scene, visualizing depth continuities as shown in Fig. 3.7. In this mode of operation, two identical light sources of opposite polarity are placed on opposite sides of the sensor. Whenever a depth discontinuity shadows one of the light sources, the resulting image displays

¹ https://commons.wikimedia.org/wiki/File:CIE1931_RGBCMF.svg

positive or negative values. All other pixels obtain a value around zero. By varying the distance between the light sources, different edge widths are obtained. As the light source separation approaches the distance between scene and camera, the system records shading images like Fig. 3.8. Similarly to Woodham’s photometric stereo method [60], they could be used to estimate the surface orientation of an object.

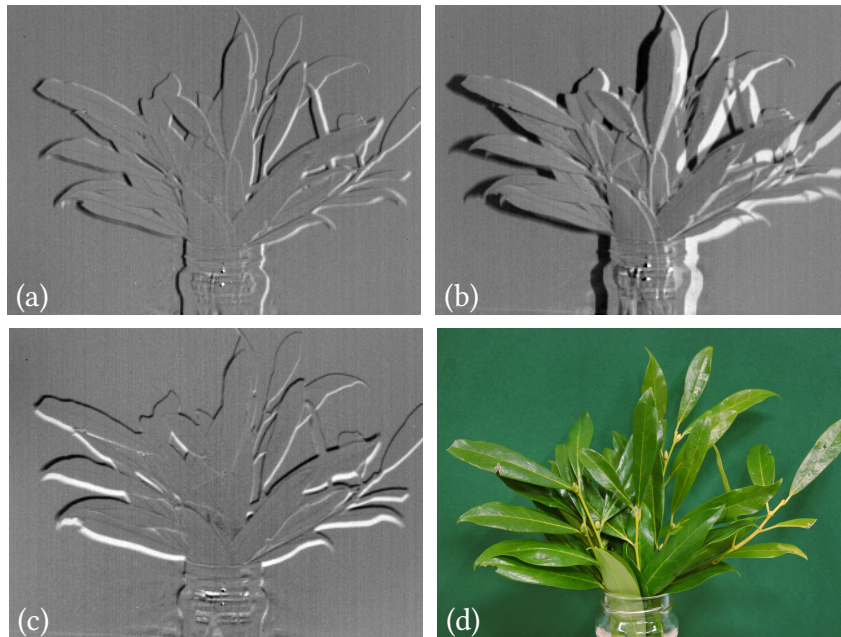


Figure 3.7: Horizontal (a)/(b) and vertical (c) alignment of the light sources creates corresponding gradient images (TI camera). The distance between the light sources determines the width of the edges in the image: short distance (a) vs. larger distance (b).

COMPARISON OF SINGLE- AND TWO-SHOT EDGE IMAGING One of the key advantages of snapshot difference imaging is that it is immune to scene motion, whereas multi-shot techniques typically suffer from alignment issues when objects are rapidly moving. To illustrate this, we recorded two image sequences of a moving scene (bonsai tree shaken by wind) at the same frame rate of 60 frames per second. In Sequence 1, we used snapshot difference imaging with both light sources active; for Sequence 2, we alternated between LS1 and LS2, and digitally computed difference images between successive frames. As the results (provided as supplemental video) show, the single-shot difference images are significantly clearer with more consistent leaf shapes than the two-shot ones, and virtually

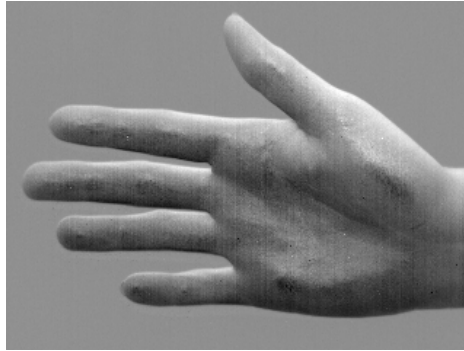


Figure 3.8: Directional difference image acquired with two light sources in wide vertical spacing (TI camera).

free of ghosting artifacts. On the other hand, the single-shot images show a slight increase in fixed-pattern noise.

3.5.4 SPATIO-TEMPORAL GRADIENT IMAGING

A feature of difference imaging is its capability to extract essential information from heavy streams of image data. Here, we use our setup to implement cameras that selectively sense spatial or temporal changes in the input, opening use cases such as machine vision and data compression [31].

SPATIAL GRADIENT. We devise an optical setup that, in combination with a snapshot difference imager, performs edge detection in analog hardware. The key is to introduce a small spatial displacement between the images I^+ and I^- , so the net image becomes the difference between two shifted copies of the scene. While this could in principle also be done through a mechanical element in the optical path (similar to active image stabilization in photography), we only add optically passive components to our setup. In particular, we use oppositely polarized light sources as in Section 3.5.1. Instead of the analyzing filter on the lens, we place a birefringent crystal immediately on top of the sensor, behind the camera lens. We determined that an undoped YVO4 crystal, 1 mm thick and inclined by 20° with respect to the optical axis, causes a displacement between light of different polarization directions by about $15\ \mu\text{m}$, or one pixel of the TI sensor. For a polarization-preserving scene, this setup produces two identical images on the sensor area, displaced by one pixel and with opposite polarity. Uniform areas in the image cancel out in this difference image, while edges are detected as non-zero response (positive or negative depending on the direction). Figure 3.10

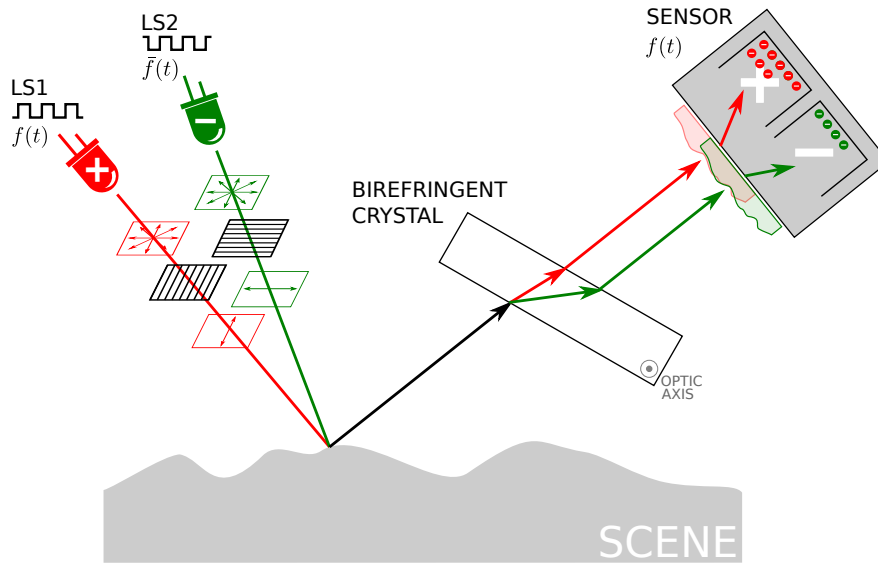


Figure 3.9: A birefringent crystal is placed between the sensor and the scene; the scene is illuminated using identical light sources that are polarized in perpendicular angles. One of them is operated in phase (+) and the other in opposite phase (-) with the sensor. In direct reflections which preserve the polarization, light from the light sources will be refracted in different angles inside the birefringent crystal and hence undergo a relative shift.

shows a gradient image of a planar aluminum resolution chart, recorded in a single shot using the TI setup.

TEMPORAL GRADIENT. We conclude with an example for our difference imaging approach that can even be used without active illumination. So far, we modulated the sensor with a high-frequency square wave at 50% duty cycle, which effectively made the sensor insensitive to ambient light. We now introduce a bias by choosing an asymmetric modulation pattern (Fig. 3.11). Light that arrives at the beginning of the exposure will now contribute to I^- , and light that arrives near the end will contribute more to I^+ . In doing so, we make the camera sense temporal changes of intensity, similar to the works by Wei et al. [61] and Hontani et al. [62]: pixels that receive more light during the second half of the exposure than during the first half appear as positive pixel values and vice versa. Fig. 3.12 shows an image thus captured of a rotating fan. From the temporal gradient image, the direction of rotation can be identified by the black and white edges of the blades.

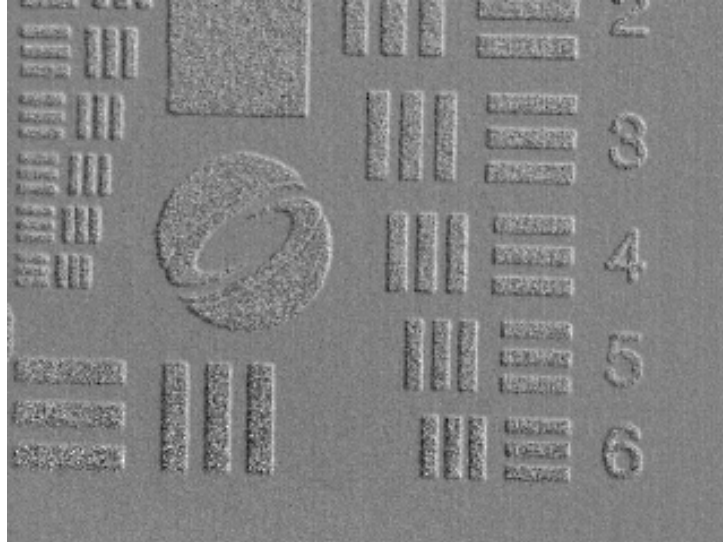


Figure 3.10: Spatial gradient image of an aluminum resolution chart (TI camera). In horizontal direction, the markers show black and white edges. The markers reflect more light than the surrounding area, which is why the noise in the difference image is higher inside the markers than outside.

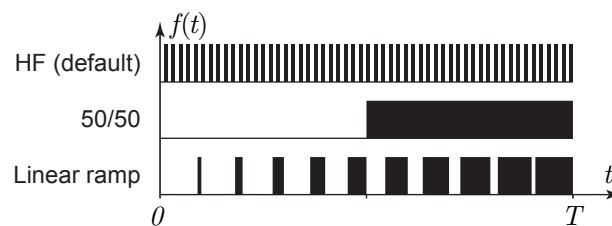


Figure 3.11: Example filters (modulation patterns) for use with our system. By default, the sensor is modulated at high frequency (top row). For the analog computation of temporal gradients, we use the 50/50 pattern with only one transition per exposure interval. This is implemented by modulating the sensor with a delayed version of the camera's ENABLE signal (cf. Fig. 3.3).

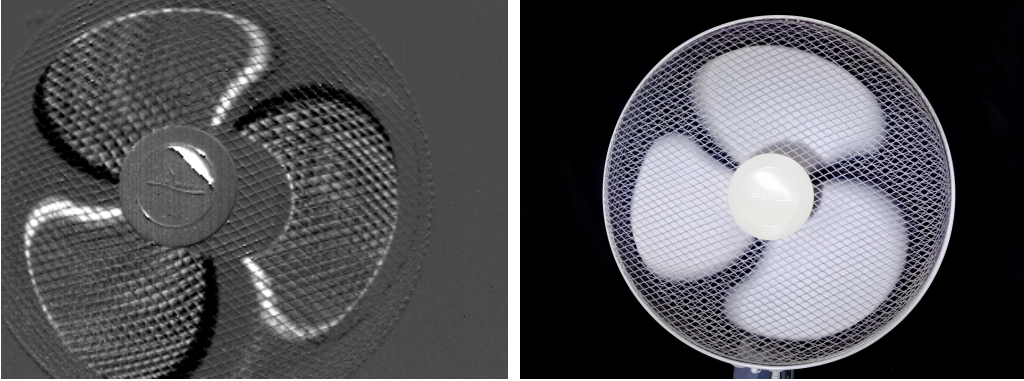


Figure 3.12: TI camera temporal gradient image (left) of a rotating fan (RGB image on the right). From the color gradient of blade edges the rotation direction of the fan is identifiable as clockwise.

Another example is shown in Fig. 3.1(b), where white pellets are shown falling on the ground. The direction of motion is visible in the temporal gradient image: Pellets falling to the ground feature positive values (red) on the bottom end and negative values (blue) on the top end. Those that have bounced off the ground and fly back up (as seen in the right part of the image) have reversed shading. Pellets lying still on the ground are barely visible. Please see the real-time motion in the supplementary video.

The exposure time used for this method is 1 ms, thus transferring this method to a conventional camera would correspond to a required frame rate of 2000 fps.

3.5.5 QUANTITATIVE NOISE ANALYSIS

In contrast to conventional cameras, our snapshot difference imaging approach performs only one read-out operation in the process of obtaining a difference image, since the differencing operation is performed before the readout. Hence, assuming shot noise and read noise as the main contributions to the measurement uncertainty, a pre- and post-ADC difference image are expected to suffer from different noise levels:

$$\begin{aligned} \left(\sigma_{\text{diff}}^{\text{post}}\right)^2 &= \eta^2 \sigma_+^2 + \sigma_{\text{read}}^2 + \eta^2 \sigma_-^2 + \sigma_{\text{read}}^2, \\ \left(\sigma_{\text{diff}}^{\text{pre}}\right)^2 &= \eta^2 \sigma_+^2 + \eta^2 \sigma_-^2 + \sigma_{\text{read}}^2. \end{aligned} \quad (3.7)$$

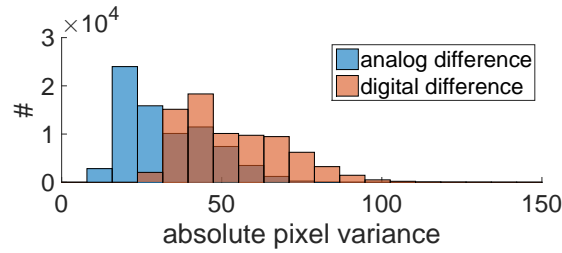


Figure 3.13: Histograms of the absolute pixel variance in the analog and digital difference images of the scene depicted in the bottom row of Fig. 3.5.

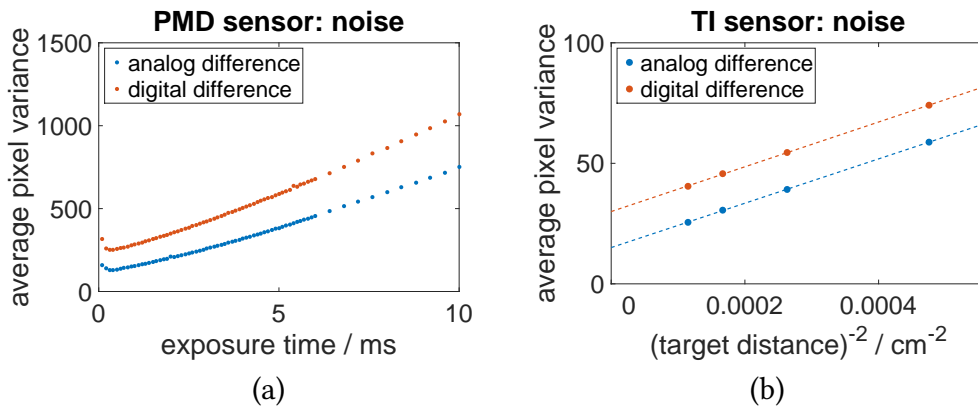


Figure 3.14: Variance of each pixel, averaged over the whole difference image, for the PMD sensor (a) and the TI sensor (b). The values were obtained from 400 frames each. The lines in (b) are linear fits to the data.

To compare the relative performance of snapshot difference imaging with two-shot, post-capture difference imaging under otherwise identical conditions, we acquired three image sequences of a still scene, each N frames long: one sequence with both light sources activated (also shown in Fig. 3.5(a)), and two more with only LS1 or LS2 turned on, respectively (Figs. 3.5(b) and 3.5(c)). We then used the data acquired with separate light sources to compute another set of difference frames. As a measure for the signal quality of the difference frames, we computed for each pixel the variance across the N recorded frames, and plotted the values for all pixels in a histogram. As Fig. 3.13 illustrates for a case with $N = 100$ frames, the noise in the snapshot difference image is significantly lower than in the post-capture difference image.

Fig. 3.14 shows the pixel variance (averaged over all image pixels) in dependence of the intensity of the incident light. For the PMD camera, we varied the exposure time of otherwise identical shots using red and blue illumination. Since

the TI sensor does not allow adjustment of the exposure time, we placed a white, homogeneous target in different distances from the camera and light sources in order to obtain different intensities. In both cases, the post-ADC difference image consistently shows higher noise. For low light intensities (short exposure times or large target distances, respectively), the shot noise of both pixel buckets tends to zero, so in Eq. 3.7, only the read noise terms remain and one has

$$\left(\sigma_{\text{diff}}^{\text{post}}\right)^2 = 2 \left(\sigma_{\text{diff}}^{\text{pre}}\right)^2 .$$

The data depicted in Fig. 3.14 supports this expectation as the lowest measured variance values for the PMD sensor are

$$\left(\sigma_{\text{diff}}^{\text{post}}\right)_{\text{min}}^2 = 251.37, \quad \left(\sigma_{\text{diff}}^{\text{pre}}\right)_{\text{min}}^2 = 127.90$$

(ratio 1.97) and the extrapolated (via a linear fit) lowest variance values for the TI sensor are

$$\left(\sigma_{\text{diff}}^{\text{post}}\right)_{\text{min}}^2 = 29.99 \pm 0.30, \quad \left(\sigma_{\text{diff}}^{\text{pre}}\right)_{\text{min}}^2 = 15.05 \pm 0.22$$

(ratio 1.99 ± 0.37), which is in good agreement with a ratio of 2.²

In order to embed our setup into the context of existing camera hardware, we compared the noise level of the TI ToF camera to a conventional camera of type Point Grey Flea3 (“PG”). Fig. 3.15 shows the variance of each pixel in a series of difference images of a color chart taken with both cameras. For the PG camera, two sets of images have been recorded and subtracted digitally. In order to make this comparison as fair as possible, we used the same LED light sources and a shutter time of 1 ms for the TI camera and 0.5 ms for the two separate images taken with the PG camera. Evidently, the ToF camera shows higher noise and lower resolution than the conventional camera, which is expected due to the

² The variance values for pre- and post-ADC differencing in the PMD sensor diverge with higher exposure times while the difference between the two measurements appears constant in the TI sensor, which can be explained by dark current noise which is not explicitly accounted for in Eq. 3.7. Since dark current noise vanishes with the exposure time tending to zero, it does not affect the analysis of the PMD sensor, while its impact on the TI sensor analysis is unknown. Since it is not possible to read out the pixel buckets separately, we furthermore cannot exclude the possibility that what we model as σ_{read}^2 is partly constituted of noise that is introduced by the process of taking the difference voltage of both pixel buckets. This would reduce the factor between post- and pre-ADC-differencing in a setup with an otherwise identical conventional sensor to a value between 1 and 2.

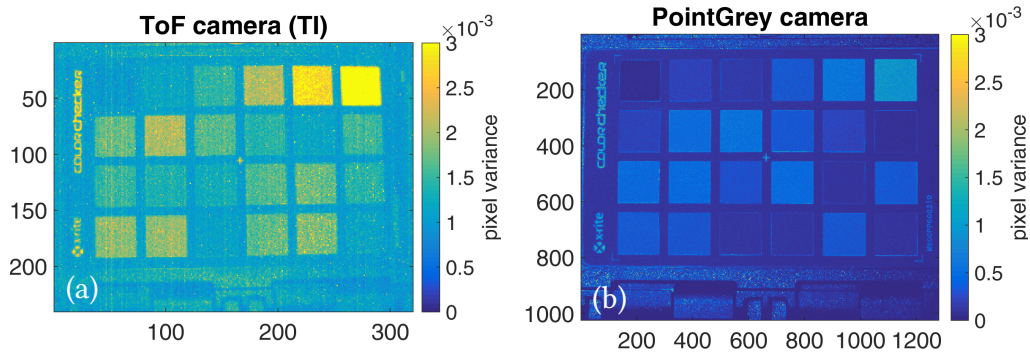


Figure 3.15: Pixel variance in difference images of a color chart recorded with a TI ToF camera (a) and a Point Grey Flea3 camera (b). The variances were calculated from 400 and 137 frames, respectively.

much longer development history of conventional image sensors compared to ToF sensors.

The special architecture of ToF sensors, however, allows to neglect photoelectric charges in the pixel buckets that have been produced by light that is not correlated with the sensor modulation, i.e. ambient light. Figure 3.16 shows difference images of a color chart obtained with a PMD ToF sensor and a PG camera with and without ambient illumination of (several magnitudes) higher intensity than the active illumination. As shown in the upper row of Fig. 3.16, the additional light decreases contrast and increases noise in the difference image of the PMD sensor. With the PG camera, the sensor's dynamic range does not allow to find a setting that captures both the target setup with and without ambient light correctly. Thus, in Subfigure 3.16(d), several patches of the color chart obtain pixel values of zero, because two saturated pixels have been subtracted from each other. If the camera parameters are adjusted such that the image with ambient illumination is properly exposed, the image without ambient light is too dark to show meaningful values (bottom row of Fig. 3.16). While in both cases the image quality for the properly exposed images is clearly better with the PG camera, the PMD camera shows higher variability and adaptability in terms of ambient illumination. Since the ability to suppress ambient illumination is unique to the ToF sensor type, we suppose that it has the potential of increasing the range of possible applications for the snapshot difference imaging approach in contrast to conventional camera setups in the future.

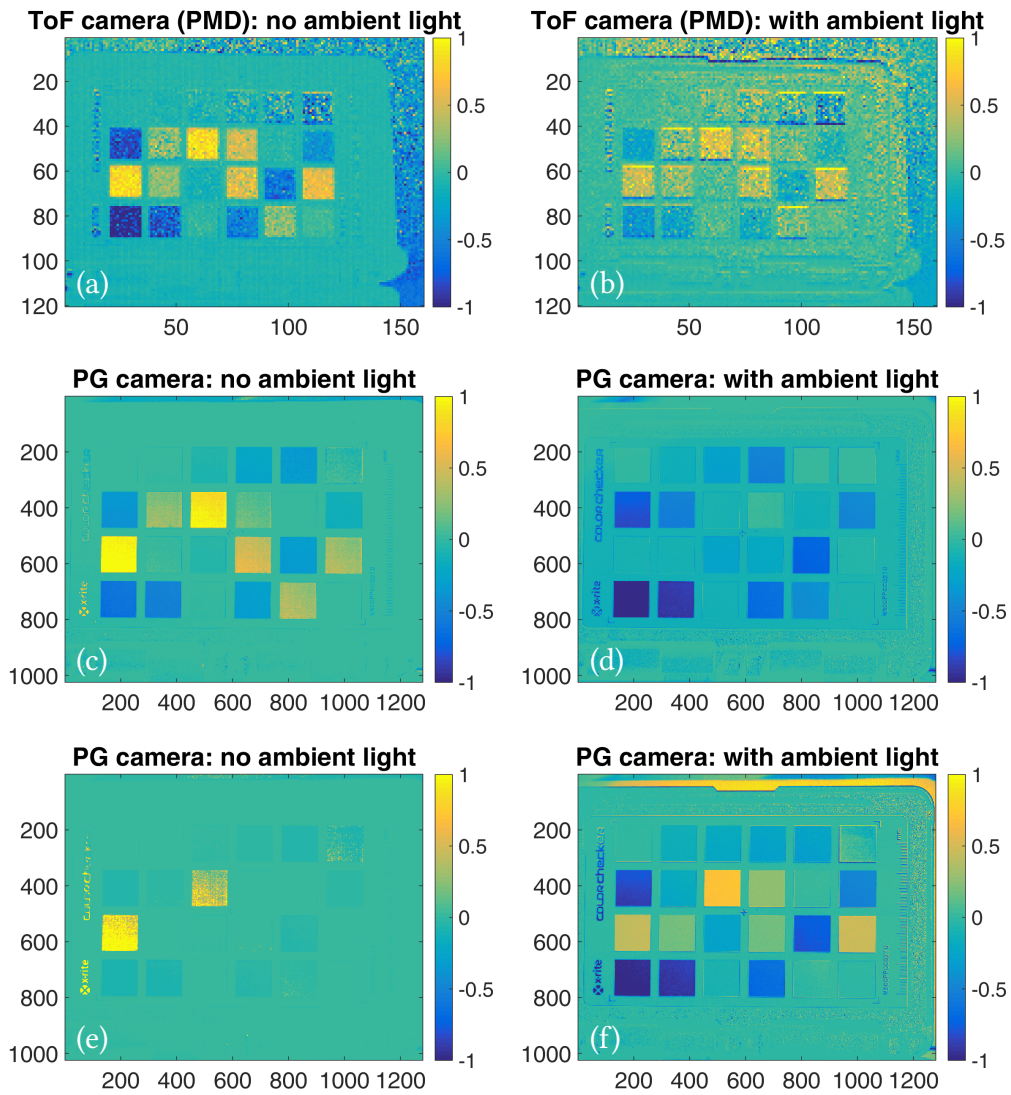


Figure 3.16: Difference images without (left) and with (right) ambient illumination, acquired with a PMD ToF camera (top row) and a conventional PointGrey Flea3 camera with larger (center row) and smaller (bottom row) aperture setting.

3.6 RECOVERING TWO IMAGES FROM A SINGLE DIFFERENCE IMAGE

In this section, we document an interesting side observation that falls directly out of the proposed image formation model for difference imaging. We can recover the two original images from a single difference image by exploiting the noise characteristics of both photon limited signals. According to Eq. 3.6, the noise in each pixel of a difference image is dependent on the amount of charges stored in the individual wells, rather than the resulting difference value. Therefore, we can calculate the separate values I^+ and I^- from the noise statistics (mean and variance of each pixel) of the difference image:

$$\begin{pmatrix} I^+ \\ I^- \end{pmatrix} = H^{-1} \begin{pmatrix} \mu \\ \sigma^2 \end{pmatrix}. \quad (3.8)$$

We propose three methods (M1–M3) to estimate these quantities:

M1. Analysis of a sequence of input image frames $I_{1..N}^{\vec{x}}$ taken under identical conditions:

$$\mu^{\vec{x}} = \frac{1}{N} \sum_{i=1}^N I_i^{\vec{x}} \quad \text{and} \quad (\sigma^2)^{\vec{x}} = \frac{1}{N-1} \sum_{i=1}^N (I_i^{\vec{x}} - \mu^{\vec{x}})^2,$$

where $I_i^{\vec{x}}$ denotes the pixel value at location $\vec{x} = (x, y)$ in the i^{th} frame.

M2. Patch-based analysis of a single pre-segmented image:

$$\mu^{\vec{x}} = \frac{1}{|P^{\vec{x}}|} \sum_{\vec{x}' \in P^{\vec{x}}} I^{\vec{x}'} \quad \text{and} \quad (\sigma^2)^{\vec{x}} = \frac{1}{|P^{\vec{x}}|-1} \sum_{\vec{x}' \in P^{\vec{x}}} (I^{\vec{x}'} - \mu^{\vec{x}})^2,$$

where $P^{\vec{x}}$ denotes the set of pixels belonging to the same image segment (patch) as pixel $\vec{x} = (x, y)$.

M3. Analysis of a single image using a bilateral filter:

$$\mu^{\vec{x}} = \frac{1}{\sum_{\vec{x}'} w} \sum_{\vec{x}'} I^{\vec{x}'} w \quad \text{and} \quad (\sigma^2)^{\vec{x}} = \frac{1}{\sum_{\vec{x}'} w} \sum_{\vec{x}'} (I^{\vec{x}'} - \mu^{\vec{x}})^2 w$$

with the bilateral weight w [63]

$$w(\vec{x}', \vec{x}, I^{\vec{x}'}, I^{\vec{x}}) = e^{-(I^{\vec{x}'} - I^{\vec{x}})^2 / 2\sigma_{\text{range}}^2} e^{-(\vec{x}' - \vec{x})^2 / 2\sigma_{\text{domain}}^2}.$$

Fig. 3.17 shows the reconstructions of the individual blue and red channels from the difference image shown in Fig. 3.6a (for acquisition details, see Section 3.5.2),

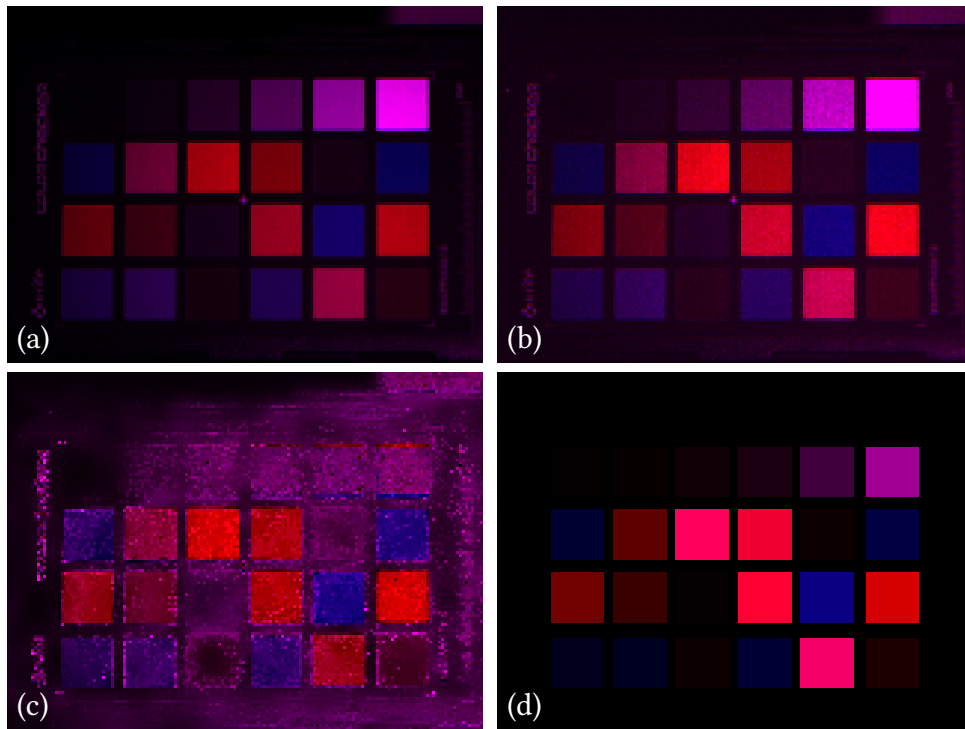


Figure 3.17: Reconstruction of two color channels from a single exposure by exploiting photon statistics. (a) Ground-truth image combined from isolated measurements of red and blue illumination; (b) Reconstruction from 1000 difference images (Method 1); (c) Reconstruction from one single difference image (Method 3); (d) Reconstruction from one manually segmented image (Method 2). All images were acquired with the PMD sensor.

obtained using M1–M3 without the read noise term ($\sigma_{\text{read}}^2 := 0$). 100 dark frames were acquired, averaged and subtracted from the difference images before performing the reconstruction. M1, here using $N = 1000$ frames, delivers the best result. M2 and M3 sacrifice quality to separate the sources from one single difference image, which makes them suitable for fast moving target scenes. M2 yields the next-best reconstruction regarding color quality and particularly the gray scale (top row of patches), but it requires flat homogeneous image regions (pre-segmented by hand). M3 uses a bilateral filter to weight down dissimilar pixels when computing mean and variance. This reduces the overall estimated variance and introduces bias; nevertheless, this algorithm would be simple enough for real-time applications.

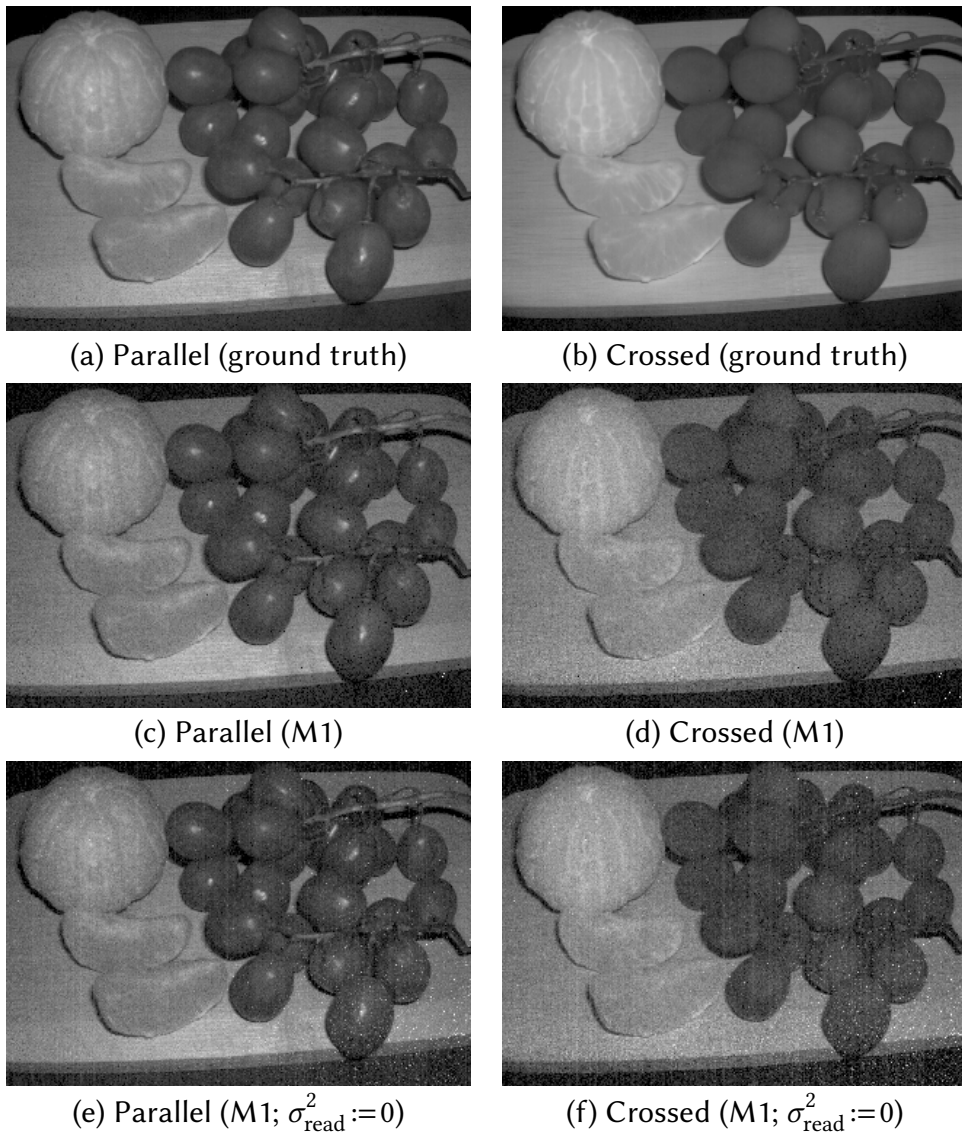


Figure 3.18: Source images of the polarization difference image shown in Fig. 3.5a (bottom), reconstructed from the statistics of 400 difference images, with (c,d) and without (e,f) pre-calibrated read noise term. Subfigures (a) and (b) show ground truth images for comparison.

As another example of this differential recovery method, Fig. 3.18 shows reconstructions of the source images via M1 from a series of $N = 400$ difference images from the direct-global separation application described in Section 3.5.1 with and without correction for read noise, as well as ground truth. As expected, polarization-preserving reflections such as specular highlights appear in the “parallel” channel only, while sub-surface scattered (depolarized) light contributes to both channels.

Since we exploit basic properties of the Skellam distribution, this method also enables source separation for traditional difference imaging. As this approach essentially enables high-speed spatial multiplexed capture without spatial separation on the sensor, we envision a variety of applications beyond the two presented above.

3.7 DISCUSSION

In summary, we propose a new imaging system for direct recording of image differences in a snapshot. The proposed technique directly maps to the emerging technology of time-of-flight sensors and will therefore continue to benefit from the ongoing technological development in that area. The primary benefits of snapshot difference imaging include high video framerates that are only limited by the readout interface as well as lower noise and reduced alignment artifacts as compared to sequential, digital difference imaging. Finally, we devise an algorithm that is capable of encoding and extracting two different images from the mean and variance of a single photograph captured with the proposed method.

LIMITATIONS Similar to range imaging, most of the demonstrated applications of snapshot difference imaging with time-of-flight sensors require active illumination. Joint coding and precise synchronization between the light sources and the sensor are required. The power of the employed light sources limits the range of distance within which the proposed method would function.

FUTURE WORK In the future, we would like to explore passive implementations of the proposed method, for example when using them with the natural flicker rates of existing indoor lighting. We would like to explore more sophisticated temporal coding strategies that may be able to separate direct and global illumination based on their temporal characteristics rather than their polarization properties. We would also like to explore spatio-temporal coding strategies that would allow the light sources to be used as temporally-coded projectors rather than isotropic

emitters [56]. Finally, we would like to extend the application to mitigating multi-path interference for time-of-flight cameras and other tasks that may benefit from gradient cameras, such as Visual SLAM [44] and 3D scanning [64].

3.8 MINIMAL SETUP

A basic implementation of a snapshot difference imager for infrared only can be obtained by connecting a TI OPT8241-CDK sensor board with two of its original infrared light sources via an extended ribbon cable (Fig. 3.19). The modulation polarity is reversed (wires #10 and #12 swapped) and wire #16 is cut for the second light source. Note, however, that the OPT8241-CDK board by itself cannot generate modulation signals below 10 MHz. Since the camera captures groups of four phase-shifted sub-frames (in 90° steps), the effective frame rate is reduced by a factor of 4 compared to our system.

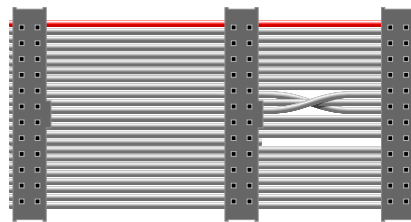


Figure 3.19: Modified ribbon cable with (top to bottom) connectors for sensor board, LS1 and LS2.

The content of the following chapter (Chapter) has been published as:

Clara Callenberg, Ashley Lyons, Dennis den Brok, Areeba Fatima, Alejandro Turpin, Vytautas Zickus, Laura Machesky, Jamie A. Whitelaw, Daniele Faccio, und Matthias B. Hullin. *Super-Resolution time-resolved imaging using computational sensor fusion*. Scientific Reports 11, Nr. 1 (18. Januar 2021): 1689. <https://doi.org/10.1038/s41598-021-81159-x>.

SUPER-RESOLUTION TIME-RESOLVED IMAGING USING COMPUTATIONAL SENSOR FUSION

ABSTRACT Imaging across both the full transverse spatial and temporal dimensions of a scene with high precision in all three coordinates is key to applications ranging from LIDAR to fluorescence lifetime imaging. However, compromises that sacrifice, for example, spatial resolution at the expense of temporal resolution are often required, in particular when the full 3-dimensional data cube is required in short acquisition times. We introduce a sensor fusion approach that combines data having low-spatial resolution but high temporal precision gathered with a single-photon-avalanche-diode (SPAD) array with set of data that has high spatial but no temporal resolution, such as that acquired with a standard CMOS camera. Our method, based on blurring the image on the SPAD array and computational sensor fusion, reconstructs time-resolved images at significantly higher spatial resolution than the SPAD input, upsampling numerical data by a factor 12×12 , and demonstrating up to 4×4 upsampling of experimental data. We demonstrate the technique for both LIDAR applications and FLIM of fluorescent cancer cells. This technique paves the way to high spatial resolution SPAD imaging or, equivalently, FLIM imaging with conventional microscopes at frame rates accelerated by more than an order of magnitude.

Conventional cameras produce images that show static illumination in a depicted scene, because exposure times are usually much longer than the photon transit time. Time-of-Flight (ToF) imaging systems, however, reach temporal resolutions of picoseconds or less and can therefore record the propagation of light in the scene. Obtaining both a high temporal and spatial resolution is particularly important for ToF imaging systems, which are often limited by the resolution in the time domain. A simple example are LIDAR-based systems where the meas-

urement resolution in the temporal domain directly equates to the spatial depth resolution. In more complex applications, such as non-line-of-sight imaging (NLOS), the resolution of the ToF of the photons is critical for determining an object's position in all three spatial dimensions [65–68]. Access to high-resolution temporal information is also pertinent to challenges such as imaging through complex media [69–71], and fluorescence lifetime imaging (FLIM), where the fluorophores of a target object are identified from their fluorescence lifetimes [72].

The challenge of time resolved imaging amounts to sampling a spatio-temporal impulse response $I(x, y, t)$, where x and y are image coordinates and t is the delay between emission and arrival of the light. Since the first capture of such data by Abramson in 1978 [73], different technologies and methods for recording light-in-flight images have emerged. Streak cameras accelerate and deflect photoelectrons in order to separate them depending on their time of production. This allows very high temporal resolution but is limited to imaging one line at a time. Therefore, for two-dimensional imaging, it requires either scanning of the scene [74] (which makes data acquisition time-costly and rules out single-shot imaging) or further modification of the set-up, like adding a digital micromirror device (DMD) in order to encode the signal spatially [21, 22]. To-date streak cameras provide the optimum temporal resolution with commercially available systems claiming resolutions of ~ 100 fs, but are also the most expensive of the available technologies.

Intensified charge-coupled devices (ICCD) provide high pixel counts and have recently been shown to be able to reach down to 10 ps temporal resolution for suitably prepared scenes [75]. This is, however, limited by various restrictions on the type of measured data, and requires bulky and costly hardware.

A cheap alternative is the use of photonic mixer devices (PMD) [17], which are based on intensity modulated illumination and a special sensor pixel design that allows measuring the phase shift between outgoing and incoming illumination. They are generally used as ToF sensors for depth imaging and provide high spatial, but low temporal resolution.

Arrays of single-photon avalanche diodes (SPAD) are rapidly becoming a leading technology for high temporal resolution imaging. This is due to the ability to manufacture time-correlated single-photon counting (TCSPC) electronics for each individual pixel directly onto the sensor chip allowing for timing resolutions on the order of tens of picoseconds [76, 77]. Currently SPAD arrays suffer from a relatively low pixel count and thus by themselves cannot be employed for many of the above imaging applications.

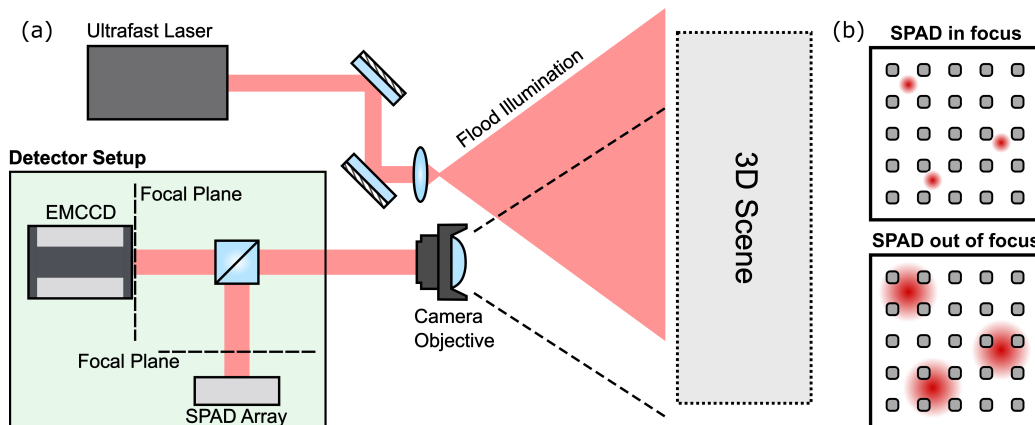


Figure 4.1: (a) Schematic sketch of the imaging set-up for the LIDAR experiment: The scene is uniformly illuminated by pulsed laser from the same direction as the camera setup. Light is collected and then imaged onto both the high-resolution CCD array and the low-resolution SPAD array by the same objective lens. The SPAD sensor is placed slightly out of the focal plane to ensure the temporal information from each point within the scene spreads across multiple pixels. (b) Effect of the optical blur: With the low fill-factor sensor in focus such that the imaging system’s point-spread-function PSF (coloured circles) is smaller than the pixel pitch, regions of the scene are not collected by the pixels (grey areas). Shifting the sensor out of the focal plane blurs the PSF and ensures collection of the temporal information from each point in the scene.

In this paper, we present a method to provide three-dimensional images with both high spatial and temporal resolution. In our approach, we fuse the time-resolved image of a SPAD detector, which is low in spatial resolution, with the image of a conventional CCD sensor which integrates the signal over the whole acquisition time but provides higher spatial resolution. Both sensors share the same objective lens and the acquired images are aligned. Our method uses an optical blur to ensure that temporal information from the entire field of view is captured by the SPAD detector despite the low fill-factor. In this respect the approach is similar to other methods that exploit blur for increased dynamic range, image restoration methods, or to otherwise avoid a loss of critical information [78, 79]. Point spread functions optimized by artificial neural networks have also been proposed, using the neural network for the image reconstruction and the SPAD data alone [80]. By using convex optimization, a data cube with the temporal resolution of the SPAD detector and the spatial resolution of the CCD camera is reconstructed. Furthermore, our approach is capable of compensating sensor flaws like dead pixels in the SPAD array. We first verify our method with

numerical simulations and assess its performance — details of this can be found in the Supplementary Information. We then demonstrate the method practically on two different temporal imaging schemes: namely multipath LIDAR and FLIM.

Our method upsamples the whole three-dimensional light-in-flight image. Similar to the works of O’Toole [81] and Lindell [82], our optimization acts on the data cube as a whole, not on a reduction to a two-dimensional depth map. This feature is key for applications where simple interpolation methods will yield an incorrect result, such as micron-scale changes in the fluorescence lifetime arising from the structure of single cells in FLIM. To test the robustness of our approach, we also demonstrate its potential using a separate publicly accessible dataset acquired with a similar experimental configuration [83] (see Supplementary Information). Our method retains high quality image reconstructions even in the presence of ambient light.

4.1 COMPUTATIONAL FUSION OF SENSOR DATA

The goal of our method is to construct a final dataset with the spatial resolution of a high pixel density CCD sensor and the temporal resolution of a SPAD array. For a SPAD dataset of spatial resolution $m \times n$ pixels and τ timebins, and a high pixel density dataset of $M \times N$ pixels, this results in a final datacube, $i_{\text{HR}}(x, y, t)$, of dimensions $M \times N \times \tau$.

SPAD arrays typically suffer from poor fill-factor (around 1% for the array in this work, see Methods), this results in a loss of information from light falling outside of the active areas. The image, therefore, needs to be optically filtered to prevent aliasing. We achieve this by moving the sensor slightly out of focus, such that the light from each point in the scene reaches at least one pixel’s active area and we therefore capture the temporal information from each point within the scene. The resulting blur is then accounted for during the data analysis (see Supplementary Information) such that the algorithm retrieves the full $i_{\text{HR}}(x, y, t)$ with the correct temporal information at each spatial coordinate.

The forward model is designed to encapsulate these features, we represent this with a matrix:

$$A = P \cdot S \cdot B \tag{4.1}$$

where B performs a blurring operation to account for the defocusing, S is a mask accounting for the sparse sampling of the SPAD array, and P performs a spatial downsizing of the higher $M \times N$ dimensions to the lower $m \times n$ dimensions (full

details in Supplementary Information). The final SPAD camera measurement, the low spatial-resolution time-of-flight image d , is then given by:

$$d = A_\tau \cdot i_{\text{HR}} \quad (4.2)$$

with A_τ applying A to all time bins and both i_{HR} and d being in vector form. The high-resolution transient image, i_{HR} , is reconstructed via

$$\begin{aligned} i_{\text{HR}} = \arg \min_{i \in \mathbb{R}^{MN\tau}} & \|A_\tau i - d\|_2 + \alpha \|Ti - c\|_2 + \beta \|K_h i - K_l d\|_2 + \gamma \|i\|_1 + \delta \|\nabla_{2D} i\|_1 \\ & \text{subject to } i \geq 0 \end{aligned} \quad (4.3)$$

where T performs a temporal integration over the data cube, c is the vectorized CCD image and K_h and K_l perform a spatial integration over the high resolution and low resolution data cube, respectively. The third term enforces a similarity between the temporal distribution of photon counts in the measured data and in the reconstruction, this proved to be an essential prior in the reconstruction. The fourth term promotes sparsity of the reconstructed data cube and, while not affecting the quality of the result significantly, it ensures stability of the reconstruction. The last term is a 2-dimensional total variation prior that acts on the spatial dimensions of each frame, which we found to significantly improve the results for scenes with large amounts of multiply scattered light.

The relative weights α , β , γ and δ have been tuned to yield the best results (see Numerical Simulations in Supplementary). The optimization is performed using CVX 2.1 [84, 85] and Gurobi 7.52 [86].

4.2 EXPERIMENTAL RESULTS: LIDAR

We verify our method with data from a LIDAR scene depicted in Fig. 4.1 and described in detail in the Methods section. The raw data is first denoised and adjusted as described in the Supplementary Information. The high spatial resolution data cube is then reconstructed according to Eq. 4.3. The parameters used for this and all other data shown in this paper and the Supplementary Information are listed in Tab. 4.1. To model the blur of the defocused image on the SPAD sensor, a standard deviation of 6 CCD pixel widths was used. This value was found empirically as the one yielding the best reconstruction results and its accordance with the data was verified using in- and out-of-focus data acquisitions.

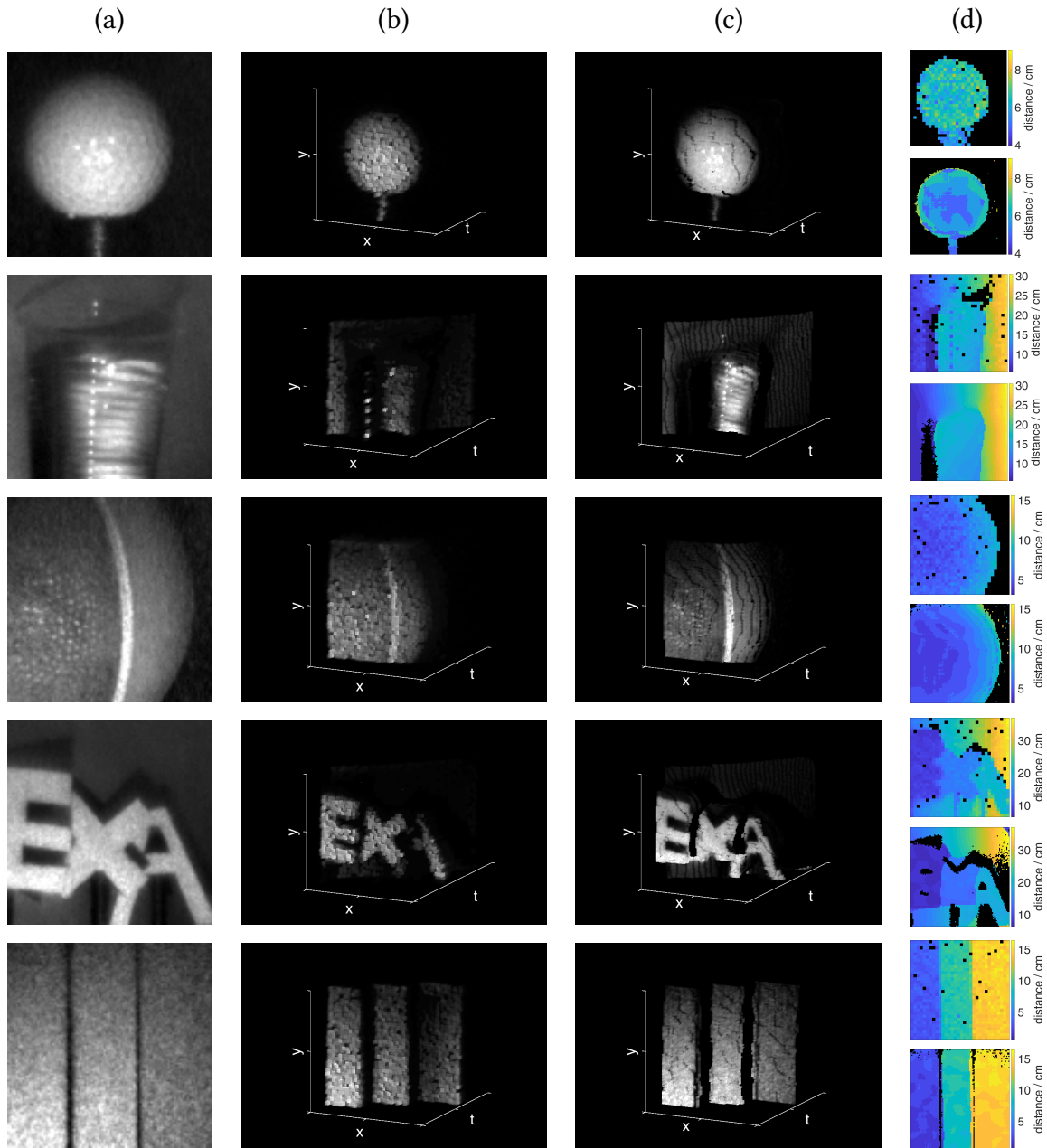


Figure 4.2: (a) CCD image, (b) $32 \times 32 \times t$ SPAD measurement, (c) $96 \times 96 \times t$ reconstructed light-in-flight image and (d) depth images extracted from the SPAD measurement (top) and the reconstruction (bottom) for different scenes. (a) & (b) are used as the inputs for our algorithm. From top to bottom the measurements show a golfball, a plastic cup filled with water in front of a slanted wall, detail of a basketball, three cardboard letters with a few centimetres distance between them in front of a slanted wall, three cardboard steps. Black areas in the depth images correspond to pixels with very low signal-to-noise ratio that therefore contain no meaningful depth information.

The reconstruction results for different scenes are shown in Fig. 4.2. Column (a) and (b) show the raw measurements, (c) shows the reconstructed data cubes. One can see that high-frequency textures that are well visible in the CCD image but not in the SPAD measurements have been transferred into the light-in-flight image. Surfaces are much smoother, less noisy, and sharpened in all dimensions.¹

In addition to the reconstructed data cubes, simple depth images of the raw SPAD measurement and reconstructed scene are shown in column (d), where the time bin with the highest photon count per spatial pixel was used as the depth value. Here, it is well visible that dead pixels from the SPAD array have been filled in, even though they have not been masked or otherwise specifically addressed before or during the reconstruction. This is possible because due to the blur, information from the dead regions is not lost, but spread over and mixed into surrounding pixels and can therefore be reconstructed. Noise has also been reduced in comparison to the raw SPAD data.

The raw data of our SPAD measurements, as well as the reconstructed high resolution light-in-flight images rendered as videos showing the light propagating through the scene, can be found in the Supplementary Information, along with run times for all datasets.

4.3 EXPERIMENTAL RESULTS: FLIM

We next show the potential of our method for FLIM using a commercial microscope, the details of which are described in the Methods section. The sample consists of ovarian cancer cells expressing Raichu-Rac clover-mCherry [87, 88] and images are acquired using a single point scanning approach in a 256×256 grid. The temporal information is acquired with TCSPC in 75 timebins of 160 ps duration. From this data we build a lower resolution dataset that emulates the measurement that would be performed by the SPAD array. We apply our forward model operator, A_r , following Eq. 4.2 with a downsampling ratio of 4×4 . This results in a $64 \times 64 \times 75$ temporally resolved dataset that forms the low-resolution input to our algorithm, d . For the high spatial resolution dataset, c , we take the total time-integrated photon counts from the full 256×256 pixel array to form an intensity image. The lifetimes are estimated by fitting a single exponential decay model to both d and i_{HR} , bounded between 1 ns and 7 ns using prior knowledge of the lifetime distribution. The algorithm input images, along with the resulting lifetime image, are depicted in Fig. 4.3. We test the validity of our approach by

¹ Black lines visible in the images are due to temporal quantisation and rendition of the data (the temporal axis is shown with a factor 3 in order to keep aspect ratios consistent).

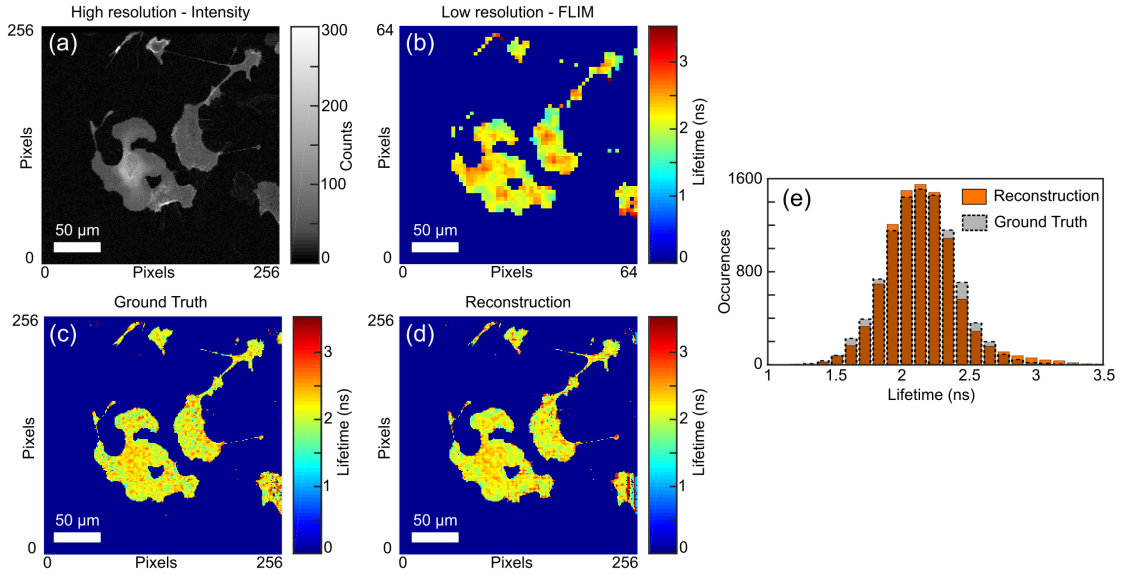


Figure 4.3: Spatial upsampling of the fluorescence lifetime of cancer cells. (a) Full 256×256 resolution intensity image. (b) Down-sampled low resolution fluorescence lifetime image. (c) Ground truth image at the full 256×256 resolution. (d) Result of our reconstruction algorithm with a 4×4 pixel upsampling. (e) Distribution of the reconstructed lifetimes.

comparing the distribution of the measured lifetimes with those obtained with the algorithm, shown in Fig. 4.3e. There is a high level of fidelity to the ground truth data with the overall shape of the ground truth distribution. The algorithm mean lifetime of the reconstructed image was 2.18 ns with a standard deviation 0.41 ns, in close agreement with the ground truth lifetimes of (2.14 ± 0.25) ns. The same approach could also be used to improve the acquisition speed of point scanning imaging systems such as the one used to acquire the data in Fig. 4.3. Our results show a reduction in the number of time-resolved measurements (spatial points) by at least a factor of 16 can be achieved with the amount of time needed to acquire the high resolution image being small in comparison. We note, however, that substantial time is still required for the reconstruction post-measurement.

4.4 CONCLUSION

Our method shows that with a simple optical set-up and a conventional camera, the spatial resolution of a SPAD array sensor can be increased significantly. In simulations, a factor of 12 could be achieved on each spatial dimension, corres-

ponding to a factor of 144 in pixel count, even in the presence of noise. Low fill factor limitations could be overcome by moving the SPAD sensor slightly out of focus. Holes in the SPAD measurement due to dead pixels are filled in by the reconstruction. This has been demonstrated on measurement data with an upsampling factor of 3×3 on LIDAR data and a factor of 4×4 on FLIM data due to hardware limitations. On additional datasets that have not been captured with our hardware set-up we demonstrated upsampling of 8×8 and 16×16 after blurring and downsampling the original 256×256 pixels SPAD data. These results, as well as those from simulated measurements, suggest that using a CCD sensor with higher pixel density, our method would allow higher upsampling factors also with our original hardware set-up. Additionally, our method is not limited by a low signal-to-noise ratio or the presence of ambient light as evidenced by the reconstructions in the Supplementary Information (“Upsampling Results on Other Data Sets” Section).

The main limitation of our method is the long run time of the reconstructions, which scales with the size of the reconstruction as well as the original SPAD measurement. A full calibration of the light transport matrix, which would include all optical effects for the specific hardware set-up accurately, might yield even better results on experimental data. On the other hand, it would supposedly also make the model more bound to a specific set-up, and less flexible in the application to new unknown hardware systems. However, it could be a worthwhile enhancement for a fixed (commercial) system. Considering the availability of small form factor SPAD and CCD sensors, both could be combined into a single, convenient device.

4.5 METHODS

4.5.1 EXPERIMENTAL SETUP: LIDAR

For the high temporal resolution dataset, we use a 32×32 SPAD array with in-pixel Time Correlated Single Photon Counting (TCSPC) capabilities of 55 ps bin width [65, 76]. The sensor layout consists of $7 \mu\text{m}$ diameter sensors with a $50 \mu\text{m}$ pitch and is of the same basic design now commercialised by Photon Force Ltd. Exposure times of up to 13 s are used. The high spatial resolution is obtained using an Andor iXon emCCD with a 512×512 pixel array cropped to 96×96 pixels to match the field of view of the SPAD array. The emCCD is used without gain such that it operates as a conventional CCD. Exposure times of the order of 100 ms are used. The same camera objective (12 cm fisheye) is used for

both sensors in parallel, separated with a beamsplitter. The illumination source is Ti:Sapph oscillator of 130 fs pulse duration at a repetition rate of 80 MHz and a centre wavelength of 800 nm which flood illuminates the scene. The SPAD camera acquisition is synchronised with the laser using an Optical Constant Fraction Discriminator to minimise electronic jitter.

4.5.2 MULTIPHOTON TIME-DOMAIN FLUORESCENCE LIFETIME IMAGING (FLIM)

Cells were left to equilibrate on a heated microscope insert at 37 °C, perfused with 5 % CO₂ prior to imaging. Images were acquired in the dark using a multiphoton LaVision TRIM scan head mounted on a Nikon Eclipse inverted microscope with a 20X water objective. Illumination is provided by a Ti:Sapphire femtosecond laser used at 920 nm (12 % power). Clover signal was passed through band pass filters 525/50 nm emission and acquired using a PicoHarp 300 TCSPC FLIM system (Picoquant). A 254 μm² field of view correlating to 256 pixel² was imaged at 600 Hz with a 10 line average.

4.5.3 MAMMALIAN CELL LINES, CULTURING CONDITIONS AND TRANSFECTIONS

SKOV3 cells were maintained in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10 % FBS, 2 mM L-Glutamine and 1X PenStrep. Cell lines were maintained in 10 cm dishes at 37 °C and 5 % CO₂. SKOV3 cells were transfected in the morning using Amaxa Nucleofector (Lonza) kit V, program V-001 with 5 μg Raichu-Rac1_Clover-mCherry DNA (adapted from [87]) following manufacturers guidelines and replated on 6 cm TC-treated dishes at 37 °C and 5 % CO₂. Cells were collected and replated onto 35 mm glass bottom MatTek dishes that were previously coated overnight with laminin (10 μg ml⁻¹) diluted in PBS. These were left overnight at 37 °C, 5 % CO₂. The next morning prior to imaging, the dishes were washed twice with pre-warmed PBS and replaced with pre-warmed FluoroBrite DMEM supplemented with 10 % FBS, 2 mM L-Glutamine and 1X PenStrep.

4.5.4 RECONSTRUCTION PARAMETERS

See Table 4.1.

Table 4.1: Parameters used in the reconstruction of all images in the main text and Supplementary Information.

	α	β	γ	δ
LIDAR (Fig. 4.2)	10^{-4}	10^{-3}	10^{-7}	0
FLIM (Fig. 4.3)	1	10^{-3}	10^{-7}	10^{-5}
Sim. monkey (Fig. 4.9)	10^{-7}	10^{-3}	10^{-7}	0
Sim. table (Fig. 4.13)	10^{-7}	10^{-3}	10^{-7}	10^{-5}

4.6 SUPPLEMENTARY INFORMATION

4.6.1 LIGHT TRANSPORT AND IMAGE FORMATION

A light-in-flight image is a three-dimensional data cube with two spatial and one temporal dimension. It can be understood as an image where each spatial pixel consists not of a single value like in a conventional intensity image, but of a temporal histogram which contains information about how much light the pixel receives at a given time t after the emission from a light source. Following the notation of O'Toole et al. [81], the flux of photons Φ for each pixel at time t is then

$$\Phi(t) = (s * l)(t) + a(t) \quad (4.4)$$

where s is the scene response defined by the geometry and reflectance of the scene, l is the temporal distribution of the illumination pulse, and a is the ambient light present in the scene. This distribution is sampled using the SPAD sensor which produces a signal for each detected photon that is then time-stamped by the respective timing electronics. All detected photon events are then sorted into a temporal histogram. For each SPAD pixel, a histogram is measured:

$$n(t) = \eta_{\text{SPAD}} (\Phi * j)(t) + \gamma_{\text{SPAD}} \quad (4.5)$$

where η_{SPAD} is the photon detection probability of the SPAD, j is the jitter accounting for uncertainties in the time-stamping, and γ_{SPAD} denotes dark counts of the sensor. Temporal discretisation is determined by the width of each histogram bin. The whole light-in-flight image is a spatial grid of these temporal histograms and therefore has the form of a three-dimensional data cube $i(x, y, t) = n_{x,y}(t)$.

In addition to the SPAD sensor, we utilize a CCD sensor that sees the same image as the SPAD sensor, but naturally has no time resolution. It can thus only measure the integrated signal

$$p = \eta_{\text{CCD}} \int_{\Delta T} \Phi(t) dt + \gamma_{\text{CCD}} \quad (4.6)$$

that sums up all light intensity measured during an exposure time ΔT by the sensor with quantum efficiency η_{CCD} , including dark counts γ_{CCD} . The whole two-dimensional CCD image is a grid of pixel values $c(x, y) = p_{x,y}$.

In the following mathematical considerations, all images are vectorized, e.g. the vector c contains all pixels of a CCD image in a linear sequence, a vector d analogously contains all entries of the light-in-flight data cube.

In our set-up, both sensors share the same objective lens via a beam splitter. Due to this fixed imaging system, only one single initial alignment calibration is necessary to ensure that the time-integrated SPAD-image matches the CCD-image. It can easily be performed using a calibration target like a printed pattern and then choosing the appropriate crop of the CCD sensor that matches the integrated SPAD image.

4.6.2 FORWARD MODEL AND RECONSTRUCTION OF HIGH-RESOLUTION LIGHT-IN-FLIGHT IMAGES

Our goal is to fuse a low spatial resolution SPAD image of size $m \times n \times \tau$ (a data cube with $m \times n$ spatial pixels and τ time bins) with an intensity image of dimension $M \times N$ in order to recover a high resolution data cube of the size $M \times N \times \tau$. To achieve this, the light transport from the scene to the SPAD sensor is described by a matrix A of size $m \cdot n \times M \cdot N$, which models all effects that the signal undergoes in the spatial domain on its way from the high resolution state as it is measured by the CCD camera, to the low resolution state as measured by the SPAD array. In principle, this transport matrix could be gauged experimentally by illuminating the scene with suitable patterns and probing the mapping from CCD pixels to SPAD pixels. This process is difficult to conduct in practice, as the illumination would have to be modified in such a way that it is capable of illuminating only certain exact pixels in the CCD image. To circumvent this tedious calibration step, we instead develop the matrix as a forward model based on the known (linear) effects on the signal, which can then also be used to produce simulated SPAD

measurements from a high-resolution ground truth. The matrix A in our model is given as the product of three matrices:

$$A = P \cdot S \cdot B \quad (4.7)$$

each representing a distinct step in the transport.

As described in the main paper, the SPAD sensor is moved slightly out of focus in order to acquire information from all scene points despite the low fill factor of the sensor. Instead of calibrating for the point spread function of the system, the blur that is accounted for in the matrix B is approximated as a 2D Gaussian. In synthetic experiments with different blur kernels, a Gaussian kernel yielded the best reconstruction results, even when the simulated measurements had been produced using other possible shapes like a disk kernel (which would correspond to the isolated defocus blur as it constitutes a convolution with the aperture shape) instead. B is thus the convolution matrix that, when multiplied with a vectorized image, yields the convolution of said image with a Gaussian kernel.

The matrix S acts as a mask, corresponding to the distribution of the active pixel area on the sensor, selecting the part of the incident light that is actually measured by the SPAD pixels. In practice, S is a diagonal matrix containing only zeros and ones - depending on whether a high resolution image pixel falls onto passive or active SPAD sensor area.

The matrix P performs the downsizing of the resulting image from $M \times N$ to $m \times n$ pixels by summing up corresponding patches of pixels (for $k \times k$ downscaling, P would add up patches of size $k \times k$ to obtain a SPAD pixel value).

The matrix A acts only on the spatial dimensions of the light-in-flight image and is thus applied on each temporal bin of the light-in-flight data cube:

$$r = A_\tau \cdot i_{\text{HR}}, \quad (4.8)$$

where r is the vectorized raw SPAD measurement, i_{HR} is the vectorized high-resolution transient image and $A_\tau \in \mathbb{R}^{mn\tau \times MN\tau}$ applies A to all time bins of i_{HR} . It is obtained as the Kronecker product of an identity matrix of size $\tau \times \tau$ and the matrix A :

$$A_\tau = \mathbb{1}_\tau \otimes A \quad (4.9)$$

More details on the matrices used in the model can be found in the section below.

In the case of noisy SPAD data (real measurement or simulations including noise), the data cube is first denoised using total variation in all three dimensions. The contrast of the CCD image is adjusted such that a potential offset of the pixel values is removed in order to eliminate sensor specific noise. Its intensity is

furthermore scaled to match the integrated intensity of the SPAD measurement, so that the total intensity integrated over all pixels and time bins is the same for the CCD and the SPAD measurement. This is necessary due to different quantum efficiencies and exposure times, as well as different measurement units of the sensors and corresponds to a scaling of factor

$$f = \frac{\eta_{\text{SPAD}} \cdot \Delta T_{\text{SPAD}}^{\text{eff}}}{e \cdot \eta_{\text{CCD}} \cdot \Delta T_{\text{CCD}}} \quad (4.10)$$

under the assumption of temporally constant ambient light, with an effective exposure time $\Delta T_{\text{SPAD}}^{\text{eff}}$ of the SPAD sensor and a conversion factor e between photoelectrons and digital pixel intensity values. These individual quantities constituting f need not be known; instead we use

$$f = \frac{\sum_{x,y,t}^{m,n,\tau} d(x, y, t)}{\sum_{x,y}^{M,N} c(x, y)}, \quad (4.11)$$

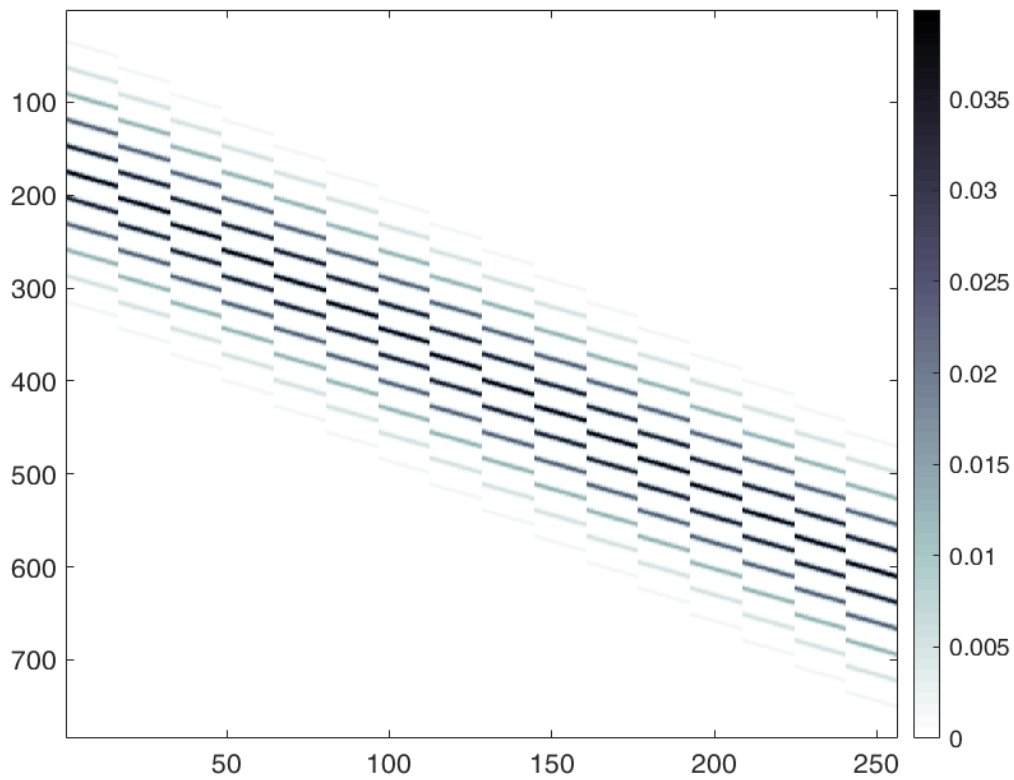
treating intensity values in the SPAD and CCD images as effectively dimensionless quantities and units as implicit.

4.6.3 FORWARD MODEL DETAILS

In order to fuse a low spatial resolution SPAD measurement of size $m \times n \times \tau$ (a data cube with $m \times n$ spatial pixels and τ time bins) with an intensity image of dimension $M \times N$, we first model the light transport from the scene to the SPAD sensor. It is described by a matrix A of size $m \cdot n \times M \cdot N$ that is the product of three matrices, as given in Eq. 4.7, where P , S and B each represent a distinct step in the transport and are described in more detail in the following. All matrices are given for a toy example of a 4×4 SPAD array with $\tau = 16$ time bins and a CCD image of resolution 16×16 .

Since our SPAD sensor is moved out of focus, we model the resulting blur by a 2D Gaussian distribution. The matrix B then looks as shown in Fig. 4.4. When multiplied with a vectorized image of resolution $M \times N$ (represented as a column vector of length $M \cdot N$), it yields a blurred image of the same resolution (plus additional padding due to the blur).

The SPAD sensor has a low fill factor. We therefore include in our model the fact that only 2% of the light actually reaches the active pixel area and neglect the rest of the light. A mask as shown on the left side of Fig. 4.5 is used to model

Figure 4.4: Matrix B .

the distribution of active area on the sensor area. It is reshaped into a matrix that can be multiplied with the blurred vectorized image, as can be seen on the right side of Fig. 4.5.

Lastly, the matrix P performs the downscaling of the vectorized $M \times N$ image (plus padding from the blur) to the resolution $m \times n$ by summing up respective image areas (in this case patches of 4×4 pixels - see Fig. 4.6).

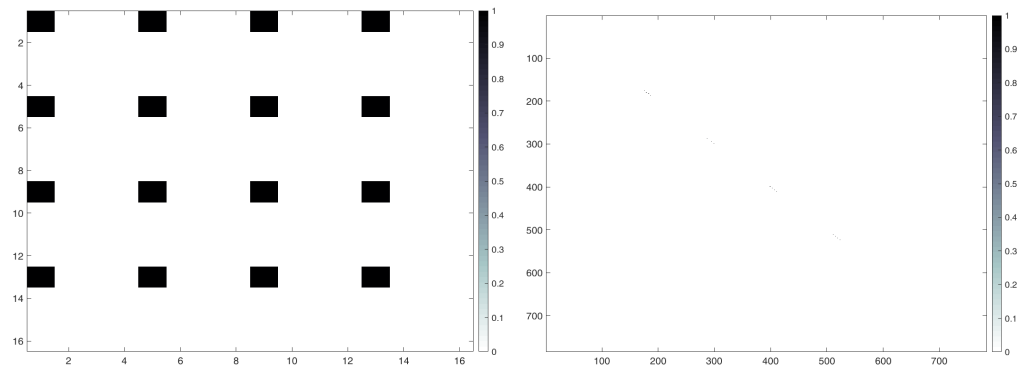


Figure 4.5: Left: Mask of the active pixel areas on the whole sensor area. Right: Matrix S - due to the dimensions of S , the active areas are only barely visible as black dots on the diagonal.

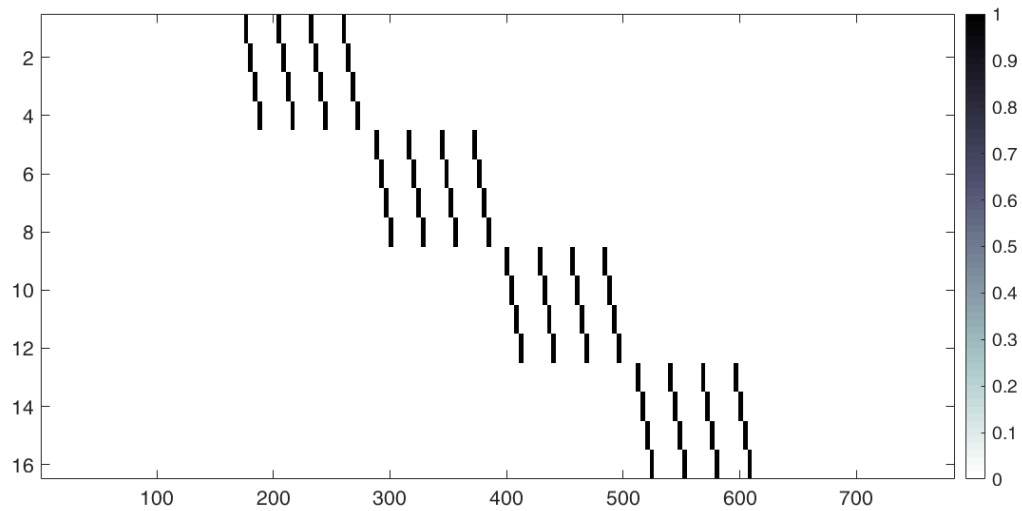
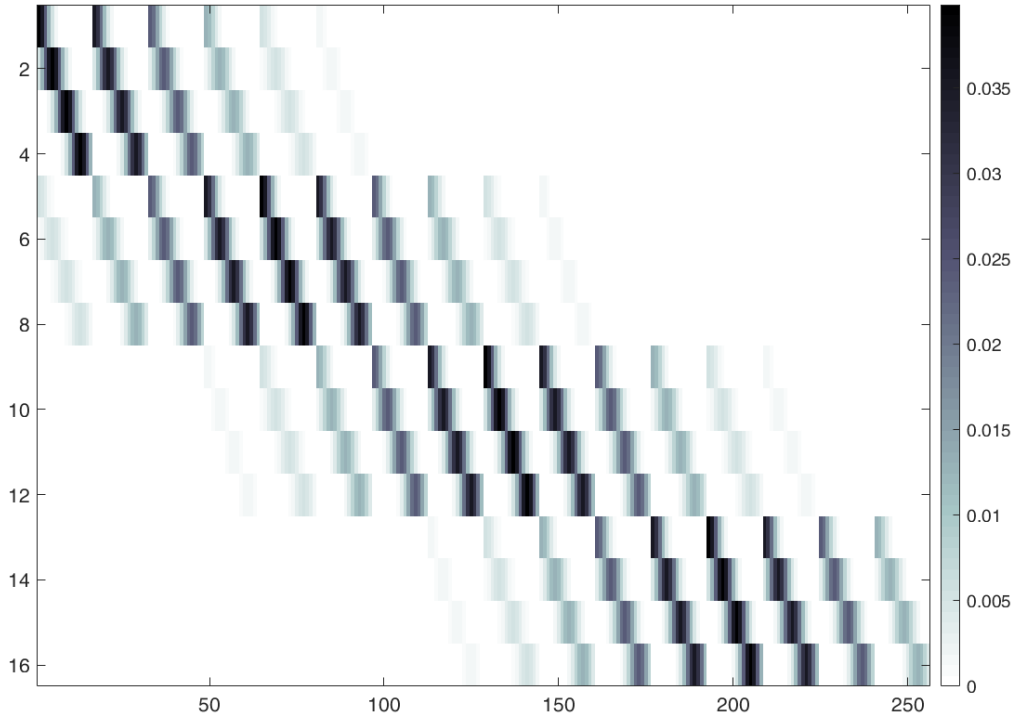


Figure 4.6: Matrix P .

The matrix A which is the product of the matrices P , S , and B , then looks as shown in Fig. 4.7.

Since this matrix acts only on a single image and not on a full SPAD data cube, we construct a matrix A_τ as the Kronecker product of an identity matrix of size $\tau \times \tau$ and the matrix A :

$$A_\tau = \mathbb{I}_\tau \otimes A$$

Figure 4.7: Matrix A .

When multiplied with a vectorized high resolution data cube x_{HR} (size $M \cdot N \cdot \tau \times 1$), it yields a low resolution SPAD measurement r (size $m \cdot n \cdot \tau \times 1$), including blur and mask, by applying A to each time frame of the data cube:

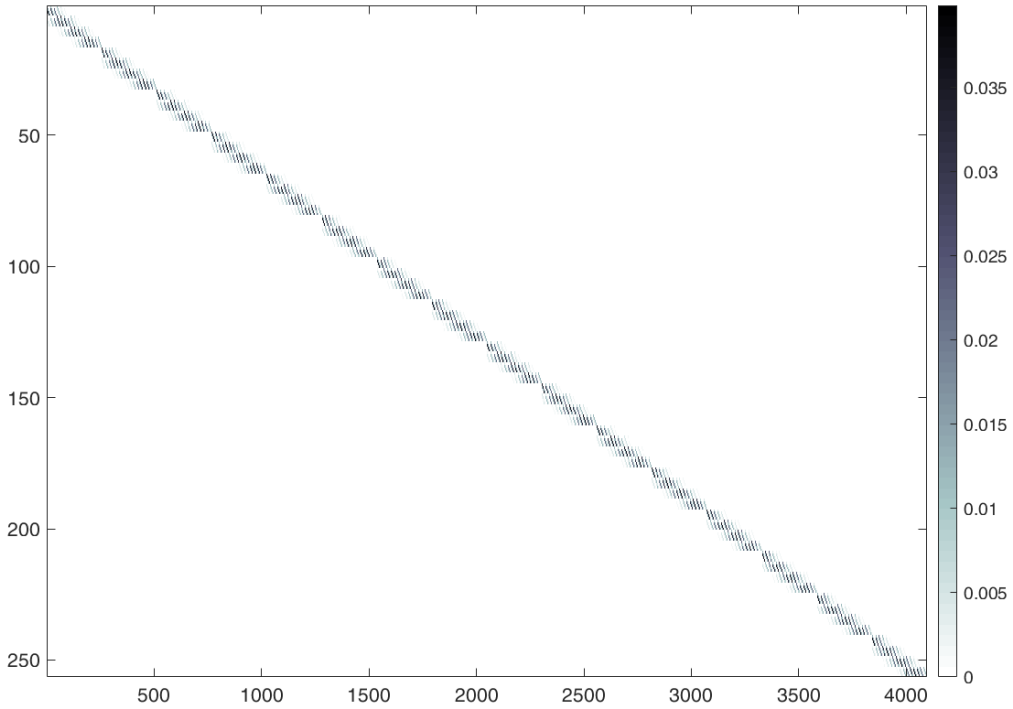
$$d = A_{\tau} \cdot i_{\text{HR}}.$$

A_{τ} is of dimension $m \cdot n \cdot \tau \times M \cdot N \cdot \tau$ as depicted in Fig. 4.8.

Using this forward model, a high resolution data cube can be reconstructed from a low resolution SPAD measurement and a high resolution CCD image as described in the paper.

4.6.4 NUMERICAL SIMULATIONS

In order to simulate the fusion of a CCD sensor image and a corresponding SPAD measurement, we created artificial light-in-flight images of three-dimensional

Figure 4.8: Matrix A_τ .

scenes by using a time-of-flight renderer² that raytraces the scene, stores the time each ray has travelled from the light source to the camera and sorts them into a histogram for each camera pixel. The result is a three-dimensional data cube of the light-in-flight image. An integration over the temporal dimension of the high-resolution ground-truth rendering serves as the simulated CCD image (see e.g. Fig. 4.10). Simulated SPAD measurements are created from this high-resolution data cube using the forward model described earlier. Figure 4.9 shows the light-in-flight data cube as volume renderings seen from different angles and, for additional visualisation, as depth maps. These depth maps are created in a naive way, using the time bin with the highest photon count as depth information, and are meant to provide additional visualisation for scenes with negligible amounts of multiply scattered light. They should not be considered comprehensive visualisations of the reconstruction results as they are not created using state-of-the-art methods and always constitute a reduction of the data to two dimensions. This holds for all depth maps depicted in this paper. In Fig. 4.9,

² Further information on the used renderer can be found in [89].

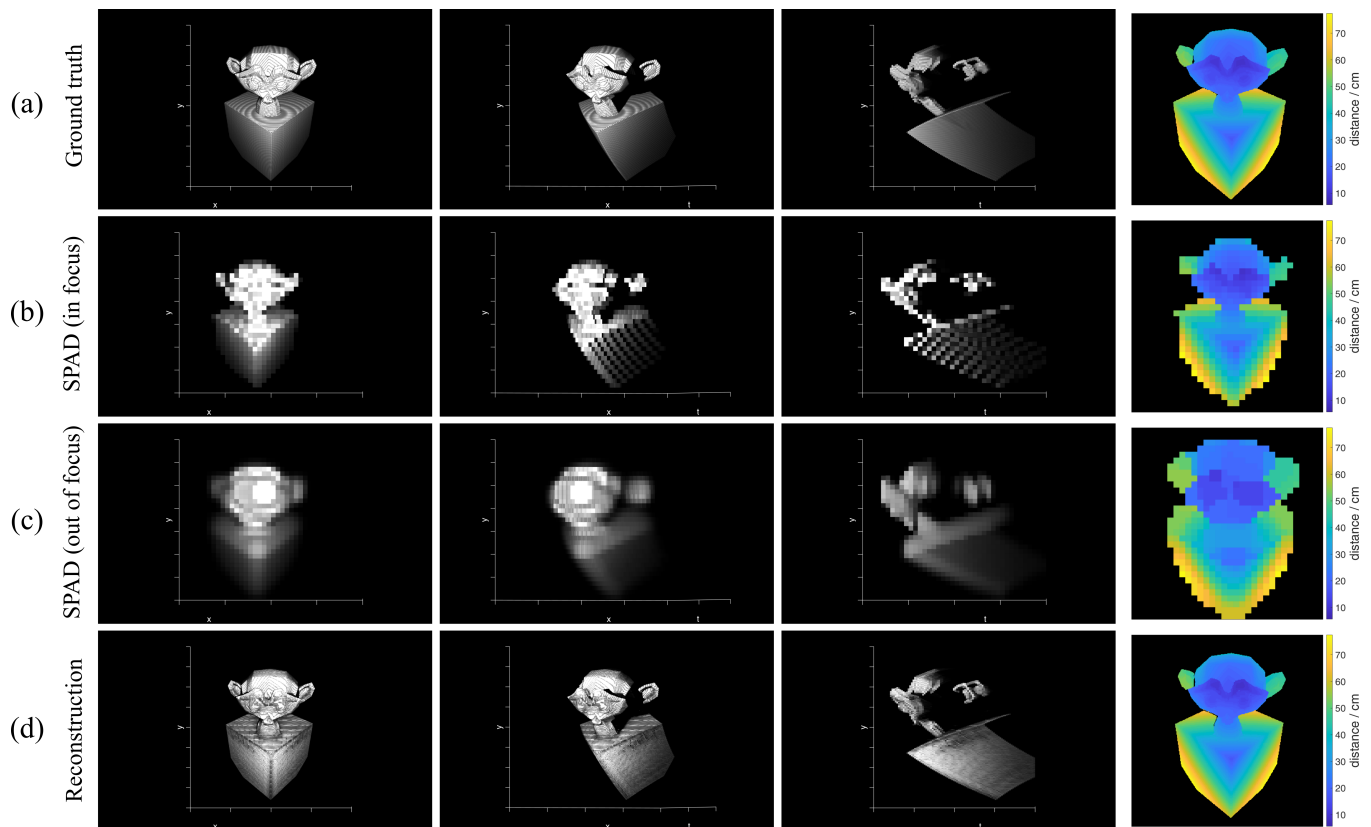


Figure 4.9: (a) High resolution ground truth simulated light-in-flight image. (b) The same scene in simulated SPAD sensor view when the image is in focus. (c) Simulated SPAD view when the image is out of focus. (d) Reconstruction from the out-of-focus measurements and a high-resolution time-integrated image of the scene. Columns 1-3 show a volume rendering of the scene from different perspectives, column 4 shows a depth image of the scene obtained by using the time bin with the highest photon count as the depth information per pixel.

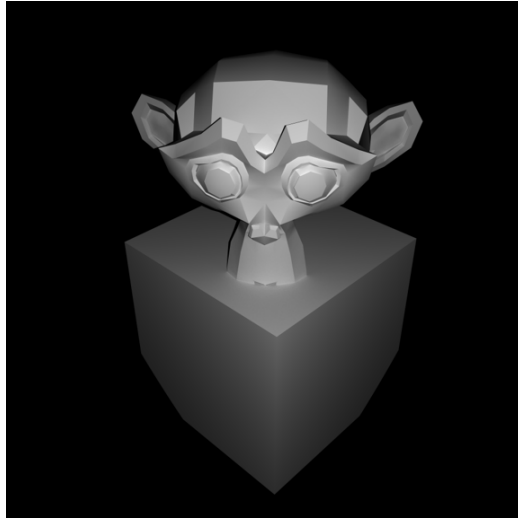


Figure 4.10: Simulated CCD image of the artificial three-dimensional scene.

the ground truth data cube is shown in row (a) and the corresponding simulated SPAD measurements created using the forward model in row (c). The scene is visibly blurred in the spatial dimension – a standard variation of 6 high-resolution pixel widths was used for the Gaussian kernel. Row (b) shows what the SPAD measurement would look like without any blur: due to the low fill factor of the SPAD sensor, information is lost and holes appear in the image in the temporal dimension due to spatially sparse sampling.

Using only the simulated measurement from (c) and the simulated high-resolution CCD image, the scene can be reconstructed as seen in row (d) according to Eq. 4.3 of the main text. Suitable regularization parameters α , β and γ were found experimentally by performing a parameter sweep over several orders of magnitude for each parameter. The total variation prior was found to be unnecessary for this scene, as it contains mostly direct reflections, thus δ was set to zero. $\alpha = 1$, $\beta = 10^{-3}$ and $\gamma = 10^{-7}$ were found to yield the best and most stable reconstruction results and were used for all reconstructions of this simulated scene throughout this paper.

Figure 4.11 shows reconstructions of the same scene in different spatial resolutions with and without noise. The width of the Gaussian blur used in the simulation was adjusted for each resolution, from $\sigma = 1.5$ CCD pixel widths in the 96×96 case to $\sigma = 6$ CCD pixel widths in the 384×384 reconstruction. The noise is modeled as a combination of Poissonian and Gaussian noise in order to

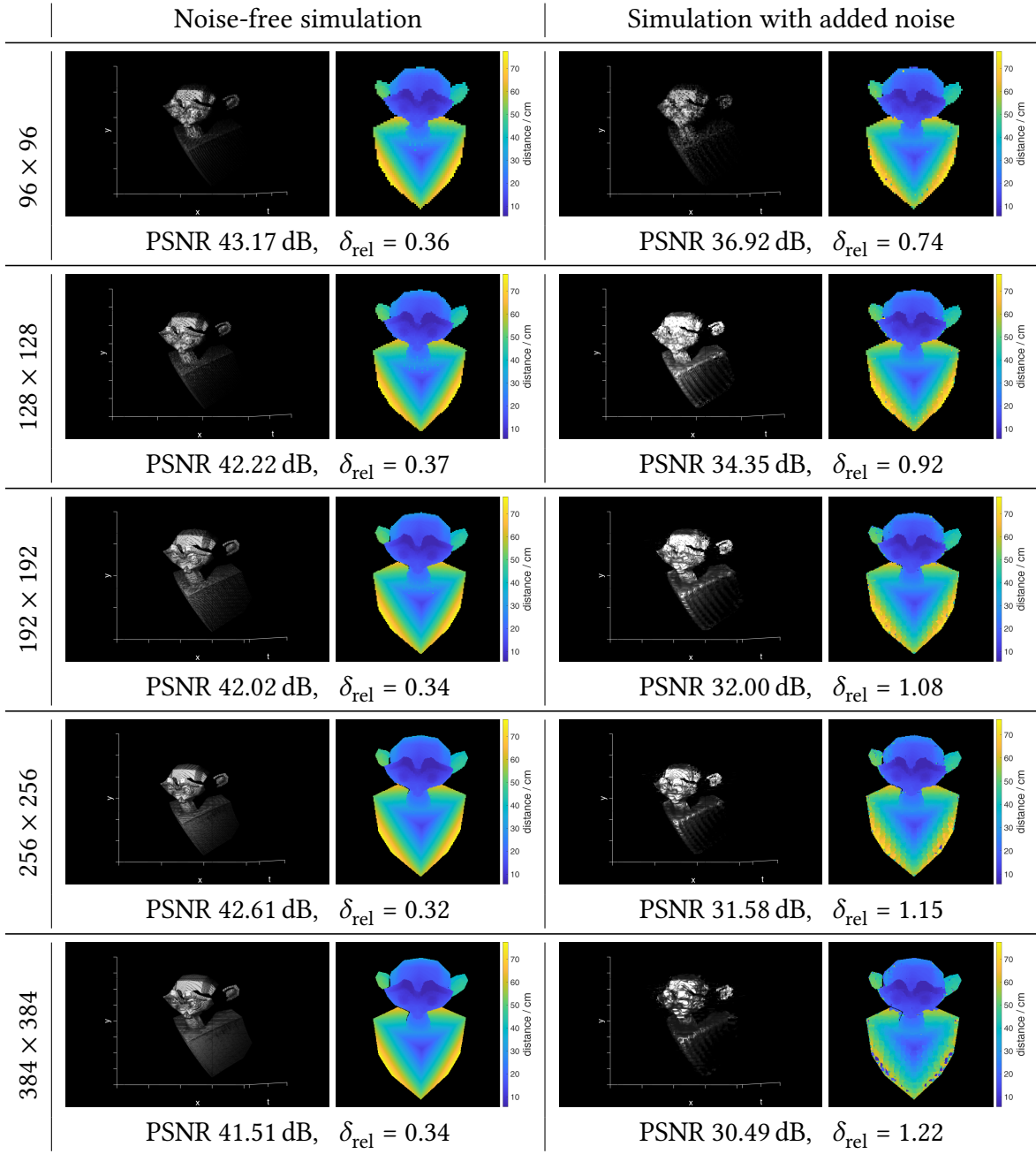


Figure 4.11: Reconstructed simulated scene in different spatial resolutions (96×96 to 384×384) with 95 time bins without (left column) and with (right column) added noise. For each reconstruction, the deviation from the ground truth is given as the peak signal-to-noise ratio (PSNR) and the relative error $\delta_{\text{rel}} = \|i_{\text{reco}} - i_{\text{truth}}\|_2 / \|i_{\text{truth}}\|_2$. Both refer to the whole three-dimensional light-in-flight image, not the depth image.

account for shot noise as well as other effects such as thermal or readout noise. In order to model the simulated measurements as closely to the real measurements as possible, the pixel count range of the artificial data was adjusted to typical values measured with the set-up described in the main paper before applying the Poisson noise (roughly $5 \cdot 10^6$ total photon counts, which corresponds to ~ 100 photon counts per pixel and time bin - meaning per 'voxel' in the data cube - on average). Additionally, areas in the measured data cube (after background subtraction) that do not contain any signal (only noise) were analysed and the pixel values were found to follow a Gaussian distribution (due to background subtraction having already been applied here, noise that is actually of Poissonian nature with high mean values is treated as a Gaussian distribution with lower mean). The parameters of this distribution were fitted to the data and used for the artificial Gaussian noise that was added to the simulated data. As a result, signal-independent Gaussian noise with $\mu = 5.6$ and $\sigma = 6.1$ was added to the simulated data, in addition to the Poisson noise.

As measures of the reconstruction performance, the peak signal-to-noise ratio (PSNR) as well as the relative error $\delta_{\text{rel}} = \|i_{\text{reco}} - i_{\text{truth}}\|_2 / \|i_{\text{truth}}\|_2$ are stated for each resolution. While in the noise-free case the quality of the reconstruction is approximately constant, the addition of noise decreases the reconstruction performance both numerically and perceptually. The degradation also increases with higher spatial upsampling factors. However, even for a spatial resolution of $M \times N = 384 \times 384$, which corresponds to a factor of 12 in both spatial dimensions of the SPAD measurement, a meaningful reconstruction could be achieved even in the presence of noise.

An additional simulated scene featuring a diffuse table, three diffuse walls forming a corner, and a specular mirror located behind the table, is shown on the left side of Fig. 4.12. Due to inter-reflections between the objects, a light-in-flight image of this scene has a complex temporal distribution that can not be reduced to single light bounces on each surface. On the right side of Fig. 4.12, the temporal intensity profile of a spatial pixel of the light-in-flight image (marked by a red square in the left image) is shown: After an initial peak of light that undergoes a single reflection on its path from the light source to the camera, additional, less intense light is detected by the same pixel, due to interreflections of the surrounding walls.

Figure 4.13 shows five frames from the simulated light-in-flight image at different points in time, the first row depicting the ground truth, the second the 32×32 pixels measurement that has been created from it. The third row shows the reconstruction of a 256×256 light in flight image from the 32×32 SPAD measure-

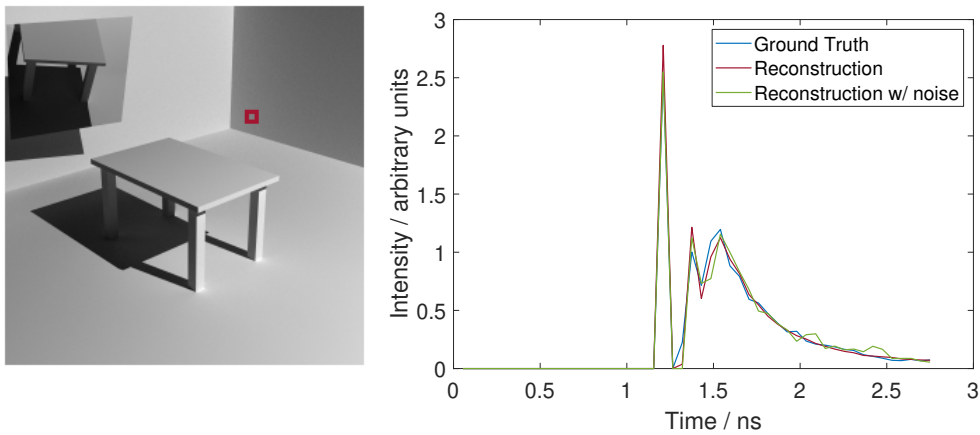


Figure 4.12: Simulated scene featuring a diffuse table and walls, as well as a specular mirror behind the table. The red square marks the spatial pixel whose temporal light intensity profile is plotted on the right. After a primary (direct) reflection, additional light that has scattered off the surrounding walls is reflected towards the camera from this same scene point.

ment and the simulated CCD image (as shown in Fig. 4.12, left). It uses the same parameters as above, except $\delta = 10^{-5}$, therefore employing a total variation prior. The last row shows how the reconstruction quality degrades when Gaussian and Poisson noise is added (using the same scaling and parameters as above). This and Fig. 4.12 (right) show that the reconstruction matches the ground truth well, with some artefacts in the late time-frames that contain only multiply reflected light, especially with added noise. The time frame at $t = 1.430$ ns (column four) shows how spatial detail is resolved in the reconstruction that is not visible in the simulated measurement due to the low resolution.

Since the intensity level of the multiply reflected light is very low (only about 5% of the intensity of the direct reflections, as can be seen from the color legends in Fig. 4.13), its reconstruction is affected strongly by additional noise. If necessary, lower noise levels can be achieved in experiment by averaging multiple measurements of the same scene, which would extend the acquisition time accordingly.

The supplemental material of the paper contains video renderings of the light-in-flight image and its reconstructions.

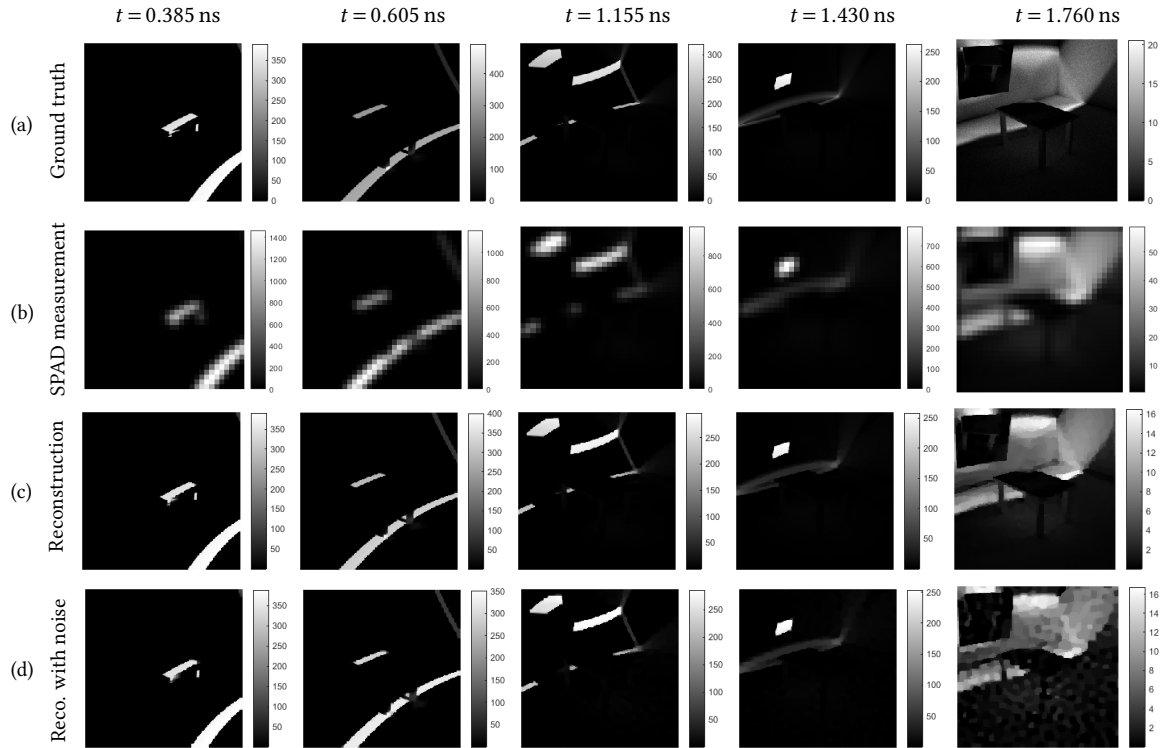


Figure 4.13: Five frames from the light-in-flight image of the scene shown on the left side of Fig. 4.12 at different points in time (see column headings) illustrating multiple light bounces in the scene. Due to severe differences in brightness between direct and higher order reflections, each frame's brightness has been adjusted to the respective depicted intensity range (see color legends). (a) Ground truth simulation. (b) Simulated SPAD measurement of size 32×32 (without added noise). (c) Reconstruction to 256×256 without added noise. (d) Reconstruction to 256×256 with added noise.

4.6.5 IMAGE RETRIEVAL RUNTIMES

Data set	time bins τ	Run time / mins
Golfball	25	11.6
Waterglass	44	78.3
Basketball	34	37.8
Letters	51	126.7
Steps	32	31.3
Simulation	38	19.0

Table 4.2: Reconstruction runtimes for data sets depicted in Fig. 4.2 of the main paper, as well as the simulated data. All reconstructions have a spatial resolution of 96×96 pixels.

4.6.6 UPSAMPLING RESULTS ON OTHER DATA SETS

In order to evaluate and compare our method further, we used it to reconstruct high resolution light-in-flight images from measured and simulated data provided by Lindell et al. [83, 90]. Each real measurement data set consists of a SPAD measurement in 256×256 pixels spatial resolution and a time bin width of 26 ps, as well as a 1024×1024 pixel CCD image of the same scene. The data was captured with the SPAD sensor in focus, so in order to simulate experimental conditions as in our set-up and thus make the data compatible to our reconstruction, we blurred the SPAD measurements spatially with a two-dimensional Gaussian and then downsampled it to a spatial resolution of 64×64 pixels. We then used our method to upsample this 'simulated out-of-focus measurement' data to 512×512 and 1024×1024 for the full time resolution and a temporally rebinned version of the data cube³, respectively. Figure 4.14 shows these results for three different datasets. Especially the results depicted in the first row demonstrate how high resolution detail, including intensity information, from the CCD image has been transferred into the reconstructed light-in-flight image (see detail in Fig. 4.15). For additional illustration and to provide a certain degree of comparability to the upsampling results in [83], (naive) depth images are shown in the last column.

³ According to [83] the FWHM of their acquisition system is ~ 440 ps, which corresponds to approximately 17 time bins. We therefore perform a temporal rebinning of factor 6 for faster reconstruction times and smoother results.

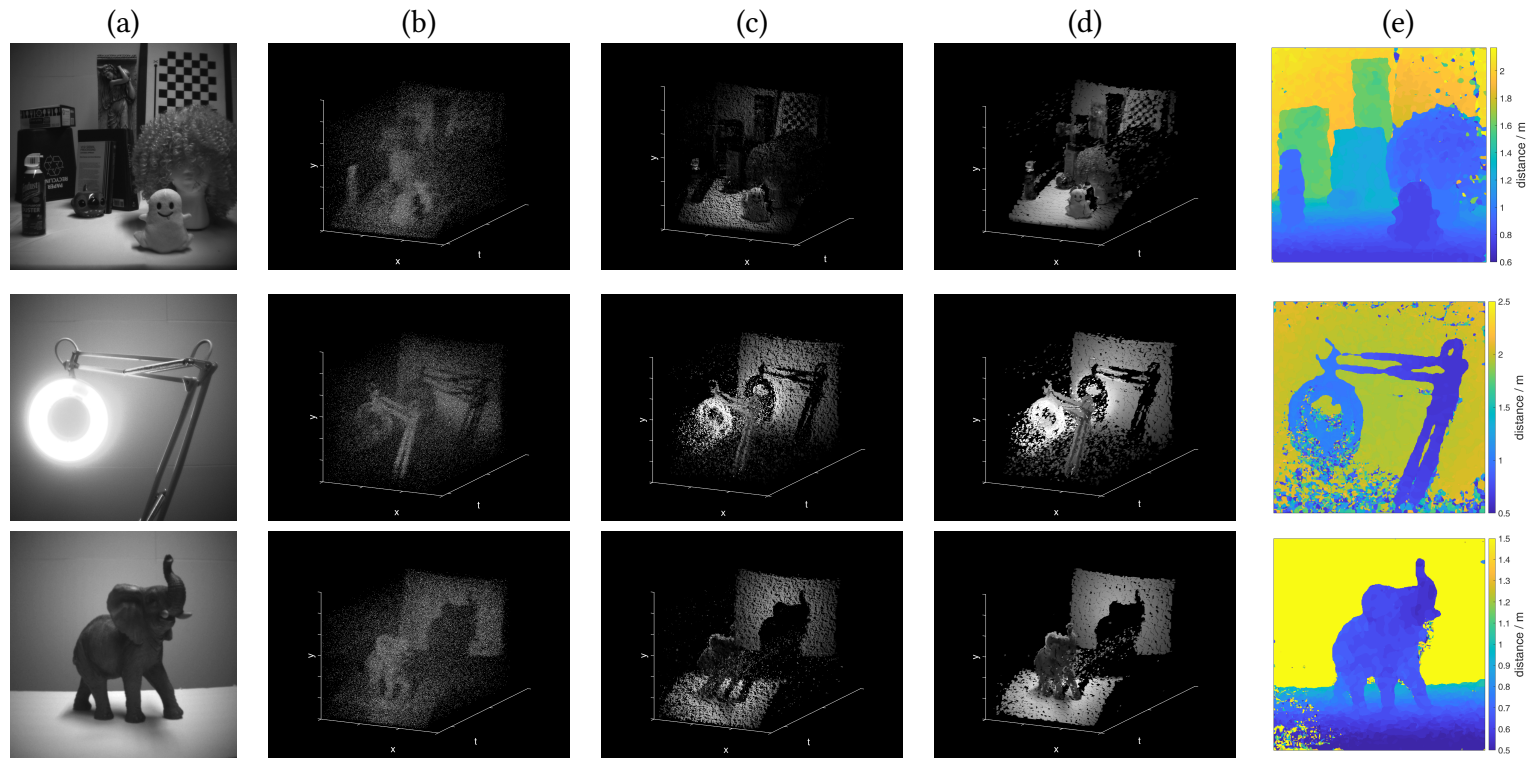


Figure 4.14: (a) CCD image and (b) SPAD image from [83] [90]. (c) Reconstruction using our method in full temporal resolution. (d) Reconstruction using our method after temporally rebinning the SPAD data sixfold. (e) Simple depth images created from the rebinned reconstruction.

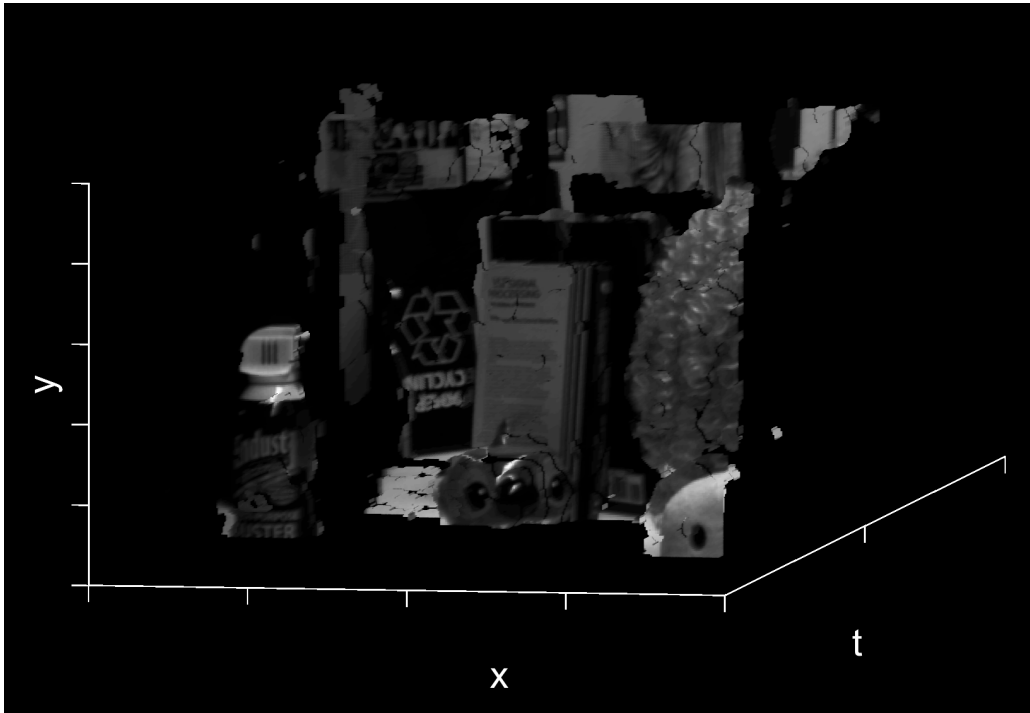


Figure 4.15: Detail of the first scene shown in Fig. 4.14 from a slightly different angle. It shows how high resolution detail from the CCD camera has been fused into the reconstructed data-cube, including fine intensity details like letters and symbols on objects in the scene.

Figure 4.16 shows details of reconstructions, depicted as (naive) depth maps, from a simulated dataset [90] with very low signal-to-noise ratio (on average 2 signal photons and 50 background photons per pixel). As with the real measurements, to make the data compatible to our method, we first blurred the (now synthetic) SPAD image spatially and downsampled it by a factor of 3. We then upsampled it back to original size using our reconstruction method. For comparison, results from Lindell et al. for the same details can be found in [83] in Figure 4. Despite being noisier than theirs, our depth maps demonstrate that our model is capable of dealing with very low signal levels and produce meaningful depth maps even though no sophisticated method to extract the depth map from the reconstructed light-in-flight image is used.

Reconstructions of both captured and simulated data from [90] show a certain level of "patchiness" in the depth maps, which can also be observed in a video

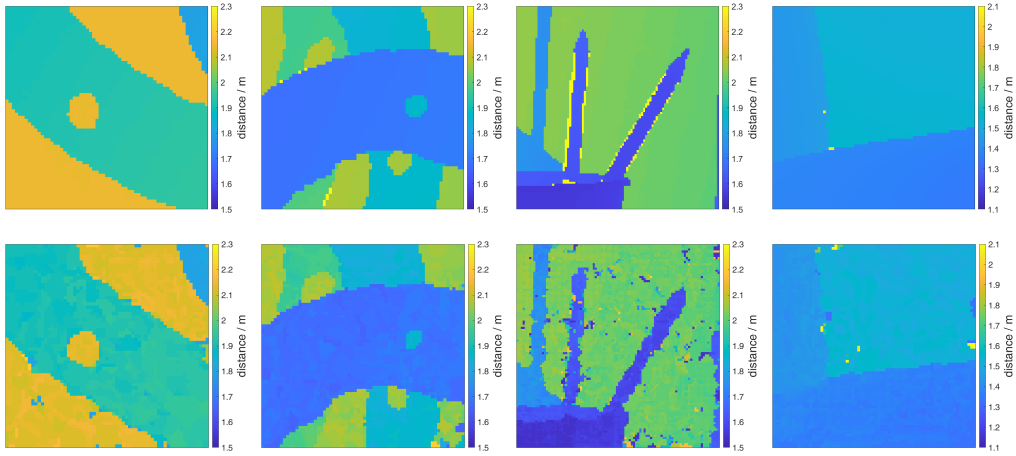


Figure 4.16: Depth maps of details from a scene simulated with low signal-to-noise ratio of only 2 signal photons and 50 background photons per pixel. Top: ground truth. Bottom: Results using our reconstruction.

view of the reconstructed data cubes (see supplemental material). Since this behaviour is also visible in reconstructions on simulations with high signal-to-noise ratio, and not present in our own captured data, we suspect that it is caused by the artificial blur being applied to the SPAD measurement in a resolution of only 256×256 pixels (as this is the maximum resolution available). The original scene signal at this point has already undergone sampling by a SPAD device, including all uncertainties and losses that come with it (low fill factor, spatial quantization). Using this data to create the 64×64 measurement that serves as input for our method, and then upsampling it to 1024×1024 , is basically an attempt to reconstruct more details than the unfocused SPAD data contains, justifying a lack in quality (i.e. in the form of 'patchiness') in the results. With the optical blur during the acquisition, before the sampling step, we specifically address this problem by spreading scene information into multiple pixels. Better results would therefore be expected from data that had been acquired with a set-up like ours, as it would include more scene information in the SPAD measurement than we can account for with the given measurement data sets.

In comparison to the work by Lindell et al. [83], our method is capable of upsampling whole light-in-flight images including multiply reflected light. Instead of a neural network, it employs convex optimization and an appropriate image formation model. It is still possible to reconstruct depth maps from the reconstructed scenes in cases where this mapping makes sense (mostly direct reflections). Our method is not limited to, but allows usage of a low resolution

two-dimensional SPAD array that does not require scanning of the scene, like the set-up used by [83]. As we downsample their 256×256 SPAD data by a factor of 4×4 after blurring it, we demonstrate upsampling by a factor of up to 16×16 on these external data sets (from 64 to 1024 pixels edge length).

The content of the following chapter (Chapter) has been published as:

Clara Callenberg, Zheng Shi, Felix Heide, and Matthias B. Hullin. 2021. *Low-Cost SPAD Sensing for Non-Line-of-Sight Tracking, Material Classification and Depth Imaging*. ACM Trans. Graph. 40, 4, Article 61 (August 2021), 12 pages. DOI:<https://doi.org/10.1145/3450626.3459824>.

LOW-COST SPAD SENSING FOR NON-LINE-OF-SIGHT TRACKING, MATERIAL CLASSIFICATION AND DEPTH IMAGING

ABSTRACT Time-correlated imaging is an emerging sensing modality that has been shown to enable promising application scenarios, including lidar ranging, fluorescence lifetime imaging, and even non-line-of-sight sensing. A leading technology for obtaining time-correlated light measurements are single-photon avalanche diodes (SPADs), which are extremely sensitive and capable of temporal resolution on the order of tens of picoseconds. However, the rare and expensive optical setups used by researchers have so far prohibited these novel sensing techniques from entering the mass market. Fortunately, SPADs also exist in a radically cheaper and more power-efficient version that has been widely deployed as proximity sensors in mobile devices for almost a decade. These commodity SPAD sensors can be obtained at a mere few cents per detector pixel. However, their inferior data quality and severe technical drawbacks compared to their high-end counterparts necessitate the use of additional optics and suitable processing algorithms. In this paper, we adopt an existing evaluation platform for commodity SPAD sensors, and modify it to unlock time-of-flight (ToF) histogramming and hence computational imaging. Based on this platform, we develop and demonstrate a family of hardware/software systems that, for the first time, implement applications that had so far been limited to significantly more advanced, higher-priced setups: direct ToF depth imaging, non-line-of-sight object tracking, and material classification.

5.1 INTRODUCTION

Time-correlated imaging, or the recording of the optical response of a scene to transient illumination, allows to analyze the temporal dimension of light transport, a feature that is not accessible in pure intensity imaging. Time-correlated

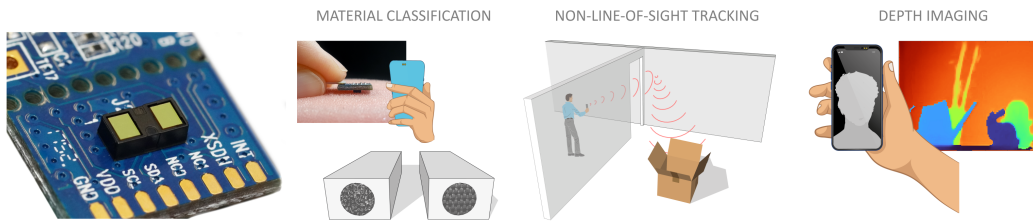


Figure 5.1: We propose the use of cheap, small off-the-shelf distance sensors (far left) for a variety of computational imaging and vision tasks, and demonstrate and evaluate their capabilities in emerging sensing applications like (from left to right) material classification, non-line-of-sight tracking, and depth imaging.

optical measurements have established themselves as a valuable source of information for scene understanding [91] and analysis in computer graphics, computer vision, scientific imaging, healthcare and life sciences (e.g., fluorescence lifetime imaging), consumer electronics, defense, robotics and autonomous driving, and even locating hidden objects outside the direct line of sight, or “looking around a corner” [66].

The approaches available for recording time-correlated measurements are rich and varied, but most require bulky and expensive hardware and are too fragile to be used outside of lab settings. A notable exception is the emerging technology of single-photon avalanche diodes (SPADs). SPADs are single-photon sensitive devices [92] that can be fabricated in CMOS technology [93] and, when combined with precise time-tagging, provide time-resolved images at picosecond resolution [65]. As direct time-of-flight sensors, SPADs have become the workhorse for a wide range of emerging fields [76, 94], such as pulsed light detection and ranging (lidar) in autonomous vehicles [95], non-line-of-sight (NLOS) sensing [96–99], as well as fluorescence lifetime imaging microscopy (FLIM) [77] and extremely high dynamic range imaging [100]. The sensors employed in all these works are custom-made research-grade devices that need to be combined with an ultrafast laser source, which is bulky and expensive (at least tens of thousands of US Dollars in total) and therefore out of reach for most real-life applications.

In parallel to the research on implementing novel applications using SPADs, however, the technology has already been fully democratized: small form-factor SPAD-based ranging systems are available at a price of USD 3, that integrate a pulsed laser source and a time-correlated single photon counter (TCSPC) along with an I²C interface. Billions of consumer-grade mobile devices use them as low-cost proximity sensors [101–103], for instance to turn off the display when the phone is placed on the ear [104]. In this work, we investigate the suitability

of such extremely cheap SPAD sensors for implementing computational imaging applications such as the ones discussed above. This is not a given since, despite the shared core technology, consumer-grade SPADs differ significantly from their high-end counterparts in terms of their feature set and performance. In lab settings without mass-market economies in mind, or in high-end industrial prototyping (robotics, autonomous mobility), cost is a far lesser concern and each of the components can be selected for optimum performance. Where research systems offer thousands of histogram bins with low temporal jitter on the order of a few tens of picoseconds, and sometimes detector arrays with relatively high spatial resolutions [77, 98], consumer SPADs offer single-point or low resolution measurements with coarse temporal binning. Equally severe constraints are imposed by operation safety (consumer devices must be safe under all imaginable circumstances, even when pointed directly at the user’s eye as is often the case with cell phones), and the tight power budget of mobile devices. Consumer SPAD systems therefore operate their lasers at the very minimum of what is required for close-range sensing, which is not enough for less light-efficient scenarios like non-line-of-sight sensing. On the API side, consumer SPADs are highly integrated devices that cannot be synchronized to external devices, and by default output heavily digested range data instead of raw timestamped photon events. This renders them unsuitable for most computational imaging applications.

The goal of this study is to implement typical application scenarios like range imaging, material classification, and object tracking around a corner, on a popular type of consumer SPAD (STMicroelectronics VL53L1X), and to identify possible avenues to improve the performance of such systems. In particular, these are our contributions:

- We propose to use an off-the-shelf sensor evaluation kit as a low-cost alternative to high-end SPAD sensors, and equip the board with a custom firmware to output raw photon count histograms.
- We introduce hardware add-ons such as collimating optics and galvanometer scanners to meet the needs of a selection of key applications for time-resolved imaging. We further propose reconstruction pipelines based on inverse filtering, deep learning, and other computational sensing paradigms that are capable of handling the low-resolution time-tagged measurements produced by our system.
- We validate the proposed platform for some of the most iconic application modes of time-resolved imaging, namely non-line-of-sight object tracking, material classification, and depth imaging.

- We propose cost-neutral feature additions to the sensor hardware that would greatly improve their interfacing to external hardware, and their suitability as a general-purpose sensing platform for time-resolved light transport.

At a total system cost of USD 150 including all parts, our prototype system in the most expensive configuration is two to three orders of magnitude lower in cost than existing time-tagged research instrumentation. Software and data are provided as supplementary material. We hope that our work will help unlock a wide range of fascinating sensing applications on hardware that millions of users are already carrying in their pockets.

5.2 RELATED WORK

This paper builds upon a substantial body of prior art, both from a technological and an application-centered point of view.

SINGLE-PHOTON AVALANCHE DIODES SPADs are emerging as a promising mass-market sensor technology capable of detecting electron-hole pairs generated by single photons incident on the sensor. When combined with time-correlation electronics, these sensors allow for accurate time-tagging of such individual photon detection events, or time-correlated single-photon counting (TCSPC). SPADs are reverse-biased photodiodes that are operated well above their breakdown voltage [105]. Every photon incident on a SPAD has a probability of triggering an electron avalanche, the so-called photon detection efficiency (PDE). The resulting avalanche event is time-stamped, providing a temporal resolution of tens to hundreds of picoseconds. Compared to other single-photon photodetectors such as photomultiplier tubes and multichannel plates, SPADs are small and versatile, and are able to work at fast rates under ambient lighting conditions without requiring a high bias voltage. SPADs and avalanche photodiodes (APDs) have been successfully employed for a wide range of TCSPC applications [106] in optical telecommunication, fluorescence lifetime imaging, and remote sensing systems (e.g., LIDAR). While SPAD sensors can be fabricated in CMOS technology [105], research instrumentation has been prohibitively costly at tens of thousands of USD, outside the range of most practitioners and in particular computer vision and graphics researchers without optical laboratories. In this work, we introduce a scalable research platform for time-tagged photon counting that is affordable and easy to use, since it is based on smartphone proximity sensors.

TIME-OF-FLIGHT (ToF), TRANSIENT AND DEPTH IMAGING Transient imaging captures the impulse response of light transport in a scene, thereby completely characterizing light transport as a linear time-invariant system. The idea of transient imaging was originally proposed by Abramson [73] as “light-in-flight recording”, using a holographic technique to reconstruct the propagation of a picosecond light pulse over time. With recent developments in ultrafast sensing technology, there exists now a variety of hardware options for transient imaging, turning it into an emerging imaging modality with manifold applications in computer graphics and computer vision [107]. Transient images and related time-of-flight techniques have been used for fast and robust depth sensing and foreground-background segmentation [83, 108, 109] where they often outperform passive methods on scenes with complex geometry and untextured regions [110, 111], while not requiring a wide stereo baseline. In addition, the availability of time-of-flight histogram data allows for insightful visualization of light transport [74], material estimation [112, 113], fluorescence lifetime microscopy [114], and even the reconstruction of objects beyond the direct line of sight.

Technologies used for time-resolved imaging differ widely in cost as well as individual advantages and drawbacks. At the high end of the price spectrum, streak cameras offer very high temporal resolution, but require additional instrumentation and computation for imaging two-dimensional scenes [21, 22, 74]. At much lower prices, amplitude modulated continuous wave (AMCW) ToF sensors, specifically intended for depth imaging at relatively high spatial resolutions, have been shown to be suitable for time-resolved imaging [17, 36]. For a more detailed comparison of this work with amplitude modulated ToF sensors, see Section 5.7.

NON-LINE-OF-SIGHT (NLOS) TRACKING Conventional cameras capture scenes that are in their direct line of sight, but computational sensing techniques have been recently proposed to reconstruct objects that are obscured from direct view, using secondary effects like indirect reflections [66] or partial shadows [115] as an information source. Such ability to see occluded parts of the scene would have numerous obvious benefits in traffic safety, search and rescue, healthcare (endoscopy), and defense, but has yet to find its way into practical applications. While a large body of work has been dedicated to the challenge of reconstructing detailed scene geometry [66, 67, 98, 99, 116–120], some applications do not require a full 3D reconstruction. Often, it could be sufficient to be able to detect objects and track their motion. Thanks to a greatly reduced number of degrees of freedom, this problem can be addressed with less detailed input data and even steady-state (intensity, no time of flight) input images under passive [115] or active [121, 122]

illumination, and it has led to the first industry demonstrators to integrate robust non-line-of-sight sensing technology [123]. Nonetheless, with an expensive bill of components, these demonstrators are unlikely to converge to mass-market products. To our knowledge, this work marks the first reported instance of using low-budget SPAD sensors for non-line-of-sight tracking.

MATERIAL CLASSIFICATION Indoor and outdoor scenes for robotic or scene understanding tasks almost always contain a diverse set of materials. Being able to robustly identify materials can be beneficial in many computer vision tasks such as acquisition, object recognition and segmentation. Classifying materials based on optical measurements is still a largely unsolved problem in computer vision, as a result of the ambiguity in appearance measurements. For example, polystyrene foam and white paper can appear very similar in conventional intensity RGB images, which makes material classification challenging based on their visual appearances. However, the interaction of light with many materials gives rise to a unique temporal point spread function (TPSF), which can be resolved with time-correlated detectors. A recent line of work [113, 124] proposes classification methods that use temporally resolved measurements to identify the materials via subsurface scattering. Comparing to reflectance-based methods that rely on single-view RGB images [125–127], temporally-resolved approaches are more robust to changes in illumination, and they are not as easily fooled by replicas such as printed pictures of the target materials. In this work, we demonstrate that it is, in fact, possible to achieve extremely reliable distinction between five whitish materials using a cellphone-class SPAD sensor.

5.3 SYSTEM DESCRIPTION

The centerpiece of our system is the VL53L1X time-of-flight sensor module by STMicroelectronics. The 12-pin package, priced around USD 3 for large volumes, has a footprint of 15 mm^2 and integrates a 940 nm light source and a 16×16 SPAD array sensor with a field of view of 27° imaged by a miniature lens. Ambient light influx is reduced by an appropriate filter. From this device, we obtain 24-bin histograms of time-of-flight data with a bin width of 1.3 ns, thus spanning a range of ca. 4.7 meters from the sensor in direct view. The data can be read as one combined measurement from the sensor as a whole, or from a region of interest (ROI) that can be any rectangle of sensor pixels with a minimum size of 4×4 . Read-out of individual pixels is not possible.

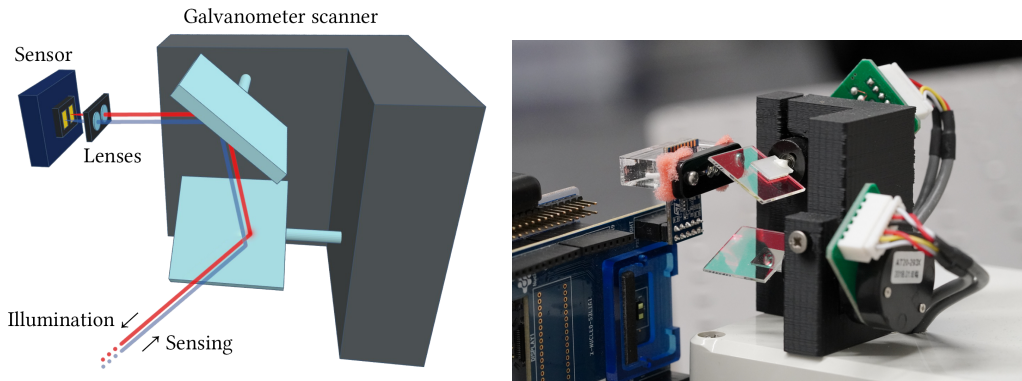


Figure 5.2: Left: Setup schematic for depth and NLOS imaging: Two plano-convex lenses collimate the light emitted from the VL53L1X’s light source and the SPAD sensor’s field of view. Two rotatable mirrors allow deflection of the beam. Right: Lenses and galvanometer scanner pair in front of the sensor and light source in the lab.

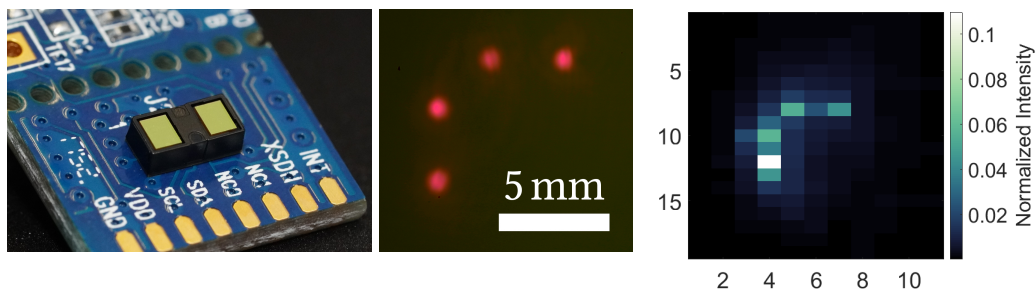


Figure 5.3: Left: The VL53L1X time-of-flight sensor module by STMicroelectronics on a commercially available breakout board. Center: Profile of collimated laser beam at 0.4 m from the lens. Right: Point spread function of the full system shown in Figure 5.2, obtained using a small retroreflective target.

For our prototype, we rely on the sensor evaluation kit P-NUCLEO-53L1A1 which adds a 32-bit microcontroller and USB interface. The system hosts up to three sensor units, one premounted and two more on satellite boards that connect to pin headers to either side of the main sensor. The stock firmware reads preprocessed range data from the sensor, and forwards it to the host via a virtual serial port. We replace it with a custom firmware that reads and forwards raw photon count histograms.

For some of the applications demonstrated in this paper, we use additional optical equipment for increased flexibility of the system.

ADDITION 1: GLASSES The sensor by default is configured for weakly directional forward sensing. For some purposes, it is more desirable to work with narrower illumination and viewing beams. We equip the sensor with a pair of small plano-convex lenses ($f = 6$ mm, $\varnothing = 3$ mm) mounted at an adjustable distance from the exit/entrance apertures to fine-tune collimation (Figure 5.2). Using a bare image sensor placed at roughly 40 cm from the emitter, we characterize the profile of the laser beam as shown in Figure 5.2 (top right). It is clearly visible that the emitter employs four laser diodes arranged in an arc shape.

ADDITION 2: GALVO SCANNERS With the illumination and sensing beams collimated through the lens pair, we mounted a pair of galvanometer scanners with hot mirrors to scan the sensing path in two angular dimensions (Figure 5.2). We used the cheapest unbranded product “20kpps Laser Galvanometer Set” that is available through the most common global sales platforms, and equipped it with larger infrared mirrors and a custom 3D-printed mount. An Arduino microcontroller board provides the analog input signals for these scanners.

Being able to control the sensing beam like this, we characterized the system’s point spread function by using a small retroreflective target (Figure 5.3, right).

5.4 MATERIAL CLASSIFICATION

We use the VL53L1X to classify different materials based on their temporal and spatial response to the illumination emitted by the device. When placing the sensor right onto the surface of a material, the infrared light from the VL53L1X light source penetrates the material, is scattered inside, and part of it is reflected back to the SPAD sensor as illustrated in Figure 5.4. Depending on the structure of the material, the signal measured by the sensor can vary temporally and spatially. By training a neural network, characteristics of different materials can be learned and they can later be distinguished by holding the sensor to an object.

For the material classification, we use contact measurements without any additional equipment. This makes this application particularly suitable for usage in small confined spaces, as well as for scenarios where the sensor is integrated into (consumer) devices like smartphones or cameras. A possible use case could be the distinction between a real finger and a dummy to improve the security of fingerprint sensors.

For the measurement, we read out the whole sensor area in ROIs of 4×4 pixels, which yields 16 independent ROI measurements arranged on the sensor in a 4 by 4 grid. This configuration allows the maximum number of independent ROIs

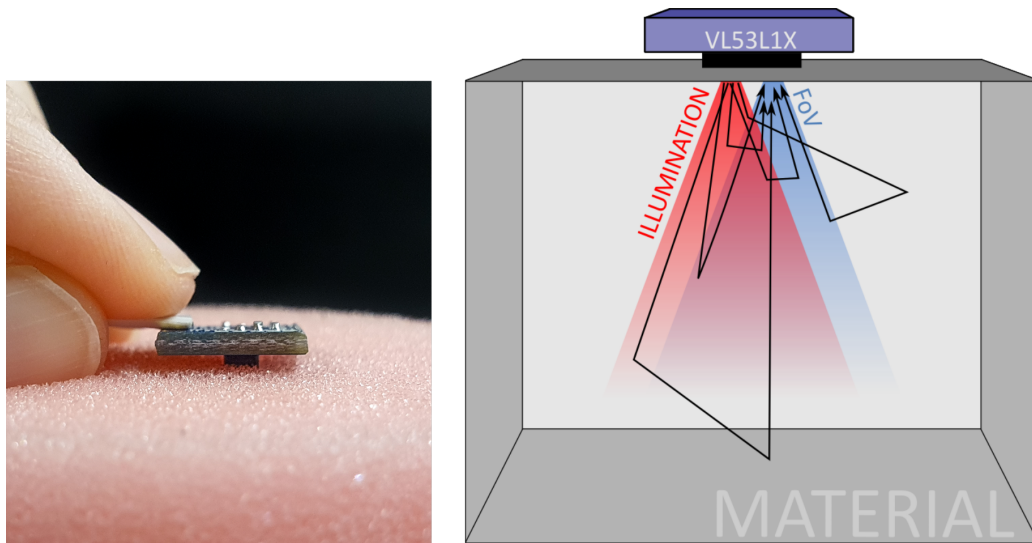


Figure 5.4: For the material measurement, the VL53L1X is placed in direct contact of the sample. The active illumination penetrates the material and is scattered inside as illustrated on the right. Depending on the material's structure, the signal measured by the SPAD sensor varies spatially and temporally.

on the sensor and constitutes a good compromise between captured information and acquisition time.

We record data for five different materials — foam, paper, skin, towel, and wax — by holding the sensor to the material 40 times in different positions and orientations and recording 25 histograms on all 16 ROIs. This procedure takes about 10 minutes per material. For evaluation purposes, we repeated the measurement for each material in the presence of ambient illumination in the form of a 100 W incandescent lightbulb that was placed in a distance of $\sim 30\text{-}40$ cm from the sample, as well as for different color variants of each material (lightgreen foam, colored paper, another person's skin, darkblue towel and red wax). Due to the low temporal resolution of the SPAD sensor and the very short range of a few millimeters, most of the information is contained in the first few time bins of each histogram. We truncate the measurements to 16 time bins from the original 24, then reshape the data to matrices of size 16×16 with one spatial and one temporal dimension. Renditions of this measurement data for the five different materials are shown in Figure 5.5.

A two-dimensional principal component analysis of the data shows a certain clustering of the materials - almost independent of the presence of ambient

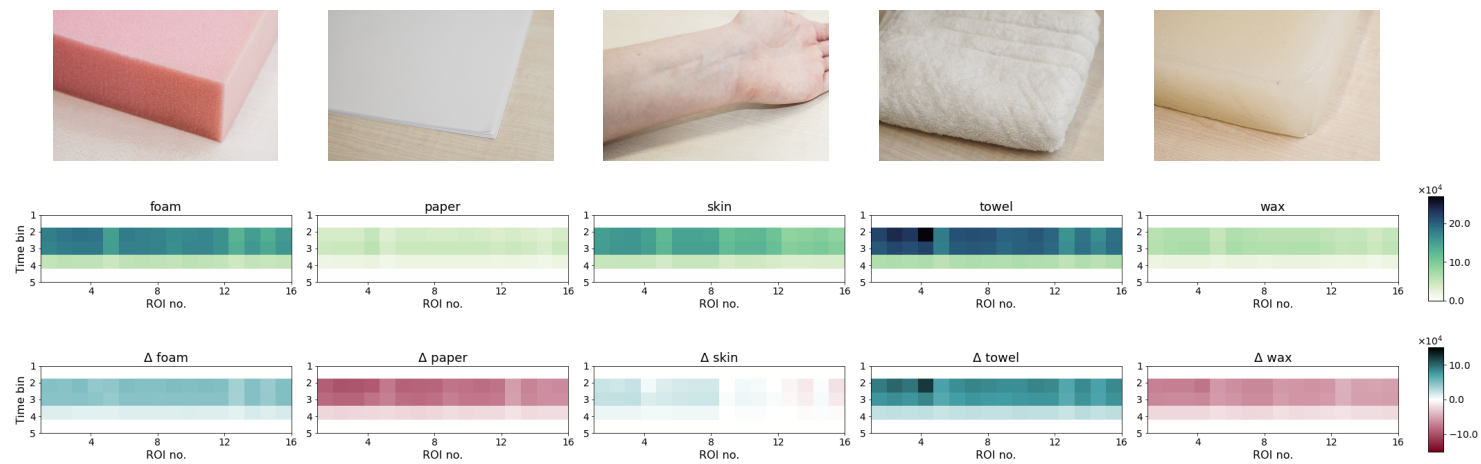


Figure 5.5: Measurements for five different materials. Top: Photograph of material. Center: SPAD histogram data averaged over 1000 measurements. Bottom: Deviation of each material from the mean of all materials. *Center and bottom: Since no visible information is contained in later time bins, these plots only show bins 1 to 5.*

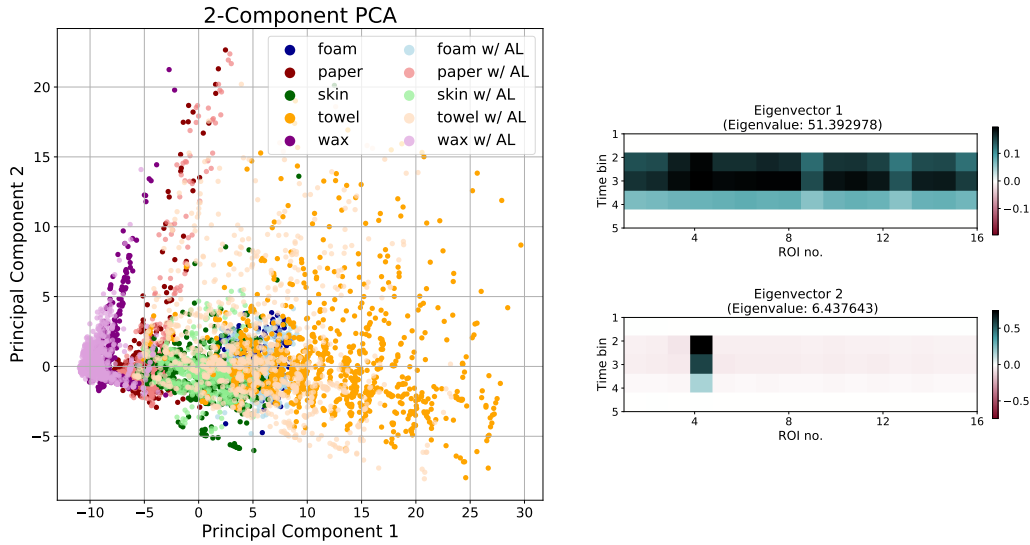


Figure 5.6: Left: Two-dimensional PCA of the material data with and without ambient light (AL). Right: The first two eigenvectors with their corresponding eigenvalues, reshaped to the input data format. As in Figure 5.5, only the first five time bins are shown.

light -, but poor separation between the categories. The PCA and the first two eigenvectors are shown in Figure 5.6.

In order to classify the different materials, we train a convolutional neural network (CNN) with two 3×3 convolution layers and two fully connected layers on data from 35 of the 40 collected positions while 5 serve as test data. We perform and evaluate the training for different constellations of data for the five materials: without ambient light, with added ambient light, and with added color variants for each material. To ensure comparability in the results, the total number of measurements in the training was kept consistent - results are shown in Table 5.1 and Figure 5.7(a)-(c). Additionally, we evaluated the performance for all material variants separately, as shown in Figure 5.7(d).

While the addition of ambient light has almost no effect on the performance, adding materials with different reflection properties (and slightly different structure) decreases the performance quality which can, however, be almost completely avoided by adding more training data (see Table 5.1). As can be seen in the confusion matrix in Figure 5.7(d), especially paper, skin and towel are mostly confused among the variants of one type of material.

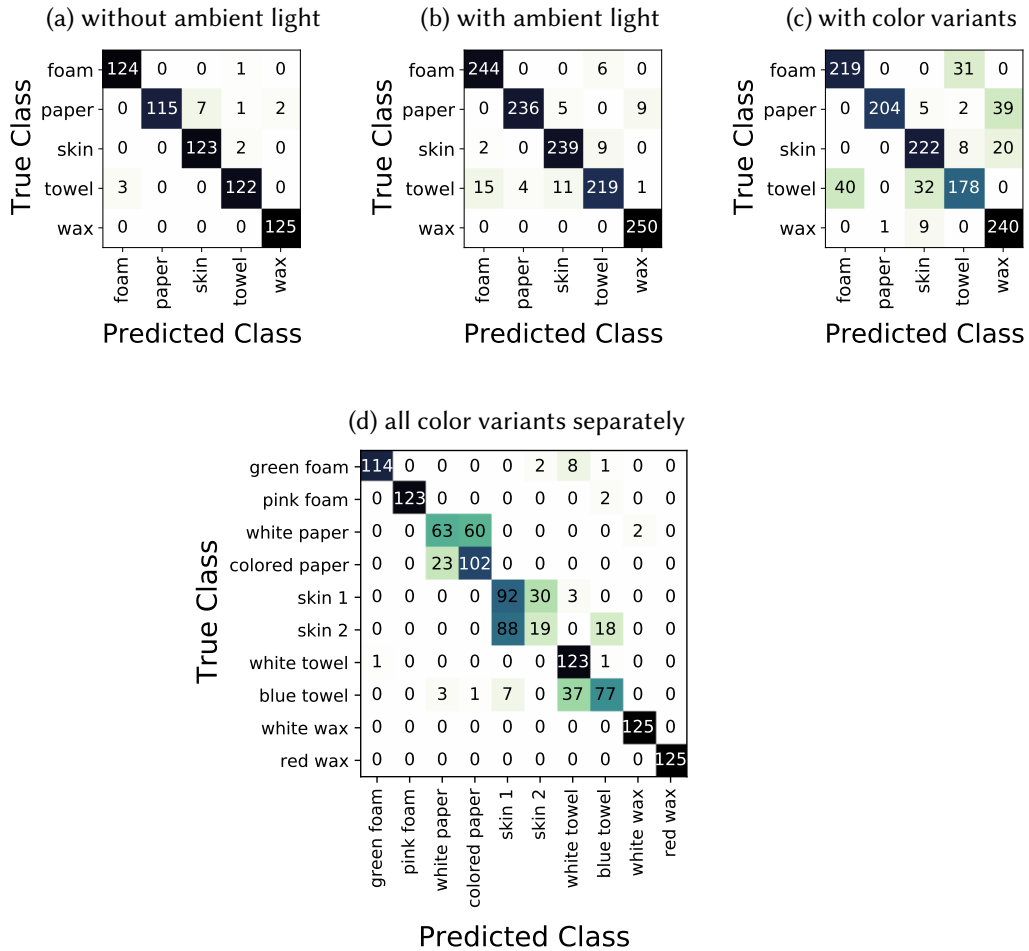


Figure 5.7: Confusion matrices for the classification of five material classes on a test dataset. (a) Measurements without ambient light, (b) measurements with and without ambient light in the training and test datasets, (c) data of two color variants for each material in the training and test datasets, (d) evaluation of classification of all color variants treated as separate classes. Numbers are the absolute count of data points in the test sets. The number of training data samples was kept consistent across (a)-(c) for better comparability.

Table 5.1: Classification accuracy on the test dataset with and without ambient light, as well as with color variants of each material.

	foam	paper	skin	towel	wax
w/o ambient light	99.2 %	92.0 %	98.4 %	97.6 %	100.00 %
w/ ambient light	97.6 %	94.4 %	95.6 %	87.6 %	100.00 %
w/ color variants	87.6 %	81.6 %	88.8 %	71.2 %	96.00 %
w/ color variants (2× training size)	99.6 %	97.6 %	97.2 %	95.2 %	100.00 %

While the results for training and testing on a single sensor are very good and the classification performs very well even in a live application where the sensor is held to different materials and the classification runs at interactive rates (as demonstrated in the supplemental video), it has to be considered that there is a certain hardware variation between individual copies of the sensor and that classification accuracy decreases significantly when performed on a sensor instance that has not been used for the acquisition of the training data. However, due to the relatively short time needed for the acquisition of training data and the training itself, as well as the individuality of potential use cases that probably require tailoring to the particular situation and used materials, we do not consider this to be a substantial drawback. Future work could focus on a generalization of the method that allows for the calibration of a particular sensor instance in order to make the data consistent across different devices.

5.5 TRACKING OBJECTS “AROUND THE CORNER”

Observing objects hidden from the direct line of sight is a common application of time-resolved imagers. We show that the VL53L1X can be used to track an object “around the corner” by illuminating a wall facing the hidden area and recording the echoing light signal that is reflected from the target object. To this end, we train a neural network to recognize the target position from the SPAD data of four measurements on the wall, as described in detail in the following subsections.

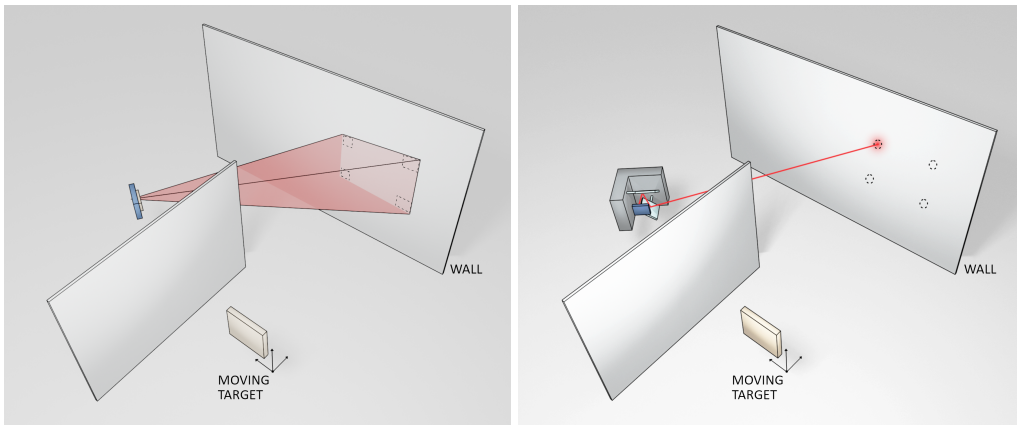


Figure 5.8: Two variants of NLOS tracking setting. Left: The SPAD package’s integrated light source directly illuminates the wall in front of the occluded target. The histograms from the four corners of the SPAD image are read out. Right: The illumination and SPAD view is collimated using the *glasses* and deflected to four points on the wall using a mirror galvanometer.

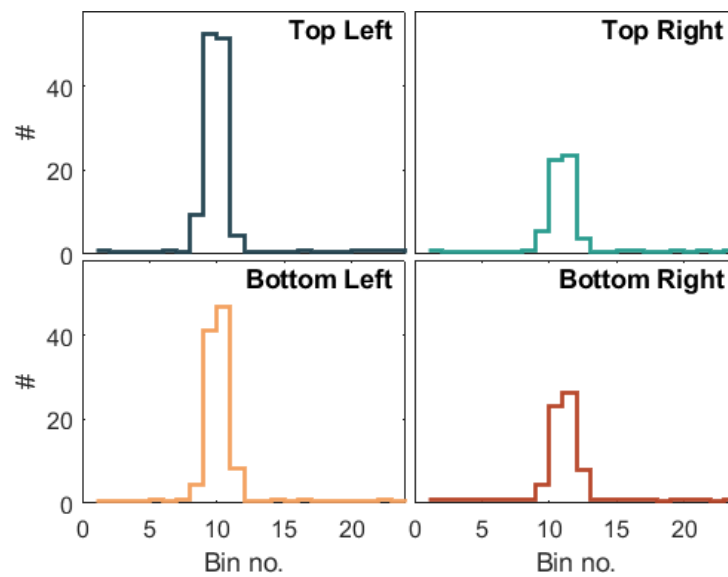


Figure 5.9: Histograms for all four corners measured with the setting without mirrors for one target position, averaged over 30 measurements. The peaks from the reflection on the wall are clearly visible.

5.5.1 DATA ACQUISITION AND PROCESSING

We propose two possible setups for the non-line-of-sight tracking task. In the first configuration, the bare light source and sensor are pointed at the relay wall and the four corners of the SPAD sensor are read out as ROIs of size 5×5 pixels (see Figure 5.8, left). The second configuration employs the “glasses” and galvo-mirror system described in Section 5.3 to focus the illumination and SPAD view to four points on the wall (Figure 5.8, right). This way, better spatial separation between the probe positions can be achieved, and the luminous efficacy – and therefore the signal-to-noise ratio in the measurements – is increased.

A square-shaped piece of cardboard laminated with retroreflective film serves as the target. It is placed facing the relay wall at different positions in the hidden volume by a Universal Robots UR10 robotic arm for optimal precision. We perform all experiments with a big ($50 \text{ cm} \times 50 \text{ cm}$) and a small ($30 \text{ cm} \times 30 \text{ cm}$) target.

We record the SPAD signal for 800 random target positions with 10 histograms each while averaging two consecutive measurements to mitigate intensity fluctuation in the data (training data acquisition takes about 5 hours). In the configuration using the galvo-mirrors, we additionally normalize the histogram by the total intensity collected in the first five bins, which only contain signal caused by light reflected back directly from the mirror system. This way, fluctuations and deviations caused by internal factors of the sensor and illumination system can be accurately compensated for.

Assuming a completely diffuse reflection of the illumination at the relay wall, the light intensity falls off quadratically from thereon. We correct for this effect by multiplying each bin content with the square of the corresponding bin number in order to keep the signal intensity of light reflected off the target consistent across the whole measurement volume. This has shown to greatly improve the reconstruction accuracy.

5.5.2 POSITION RECONSTRUCTION

Figure 5.9 shows an example measurement of four corners in the “no mirror” setup for one target position, averaged over 30 measurements to reduce noise. The direct peak from the wall is clearly visible in all four histograms, while the indirect peak position is not obvious to the naked eye. Due to the indirect peak being this low, we propose to use a neural network to determine the position of the target in the hidden space. Below, we outline four different approaches to reconstruct the target position, three of which rely on training a neural network,

while the last is a ‘classical’ approach that does not depend on previously recorded data.

1. *Direct position prediction:* We train a neural network (a multilayer perceptron with five hidden layers of size 50) to directly predict the target position from the four histograms. It takes the four histograms as input and reconstructs the coordinates of the target from these.
2. *Distance prediction and multilateration:* This approach consists of two stages. First, for each of the four probe points, its distance to the target is estimated from the measured histogram. To do this, an MLP with five hidden layers of size 16 and a final dense output layer of size 1 is trained to yield the distance from the wall point to the target from the histogram it takes as input. In this manner, the distances of all four points on the wall to the target can be predicted and used to multilaterate the target position from the four wall points by solving a simple optimization problem using the L-BFGS-B algorithm [128].
3. *Distance prediction and multilateration with histogram shift:* Before processing the histogram, we find the ‘direct peak’ that corresponds to the reflection on the relay wall by calculating the weighted mean (center of mass) of the histogram. We then shift the histogram such that it starts with the center of the direct peak. For sub-bin precision of this method, we upsample the histogram by a factor of ten and resubsample it after the shift. Then we proceed as described in approach 2. By making the histogram independent of the distance between the SPAD sensor and the wall and thus training the neural net to predict distances to the target from any point on the wall, the tracking method becomes invariant to the positioning of the SPAD relative to the relay wall: For multilateration, the position of the probe points can be calculated from the positions of the four direct peaks in the histograms (yielding the distance between the SPAD and the wall) and the angles between the four beam directions, which are a known system property.
4. *Peak finding and multilateration:* For a non-learning-based comparison, we determine the position of the indirect peak from the latter part of the histogram (behind the direct peak) as the weighted mean (center of mass) and use the distance between this and the direct peak – converted from bin widths to spatial distances – for multilateration of the target position. This approach does not require any training and thus generalizes across target sizes and SPAD/wall locations.

Table 5.2: RMSE (root mean square error) of the reconstructed target position in all spatial dimensions for all methods and target configurations in meters. We also evaluate the reconstruction methods’ ability to generalize across target sizes, where the model is trained on data recorded with the big target (if training is required) but tested on small target data, listed as “zero-shot” performance

Reconstruction Approach	w/ Mirror Performance (RMSE)			w/o Mirror Performance (RMSE)		
	Big target	Small target	Zero-shot	Big target	Small target	Zero-shot
Direct Position	0.160	0.255	0.288	0.269	0.316	0.319
Distance-Multilateration	0.299	0.304	0.307	0.397	0.421	0.423
HistShift-Distance-Multilateration	0.304	0.316	0.312	0.413	0.453	0.445
Peak-Finding-Multilateration	0.498	0.516	0.516	0.508	0.535	0.535

We evaluate all methods quantitatively and demonstrate their “zero-shot” ability across the different target sizes. All configurations and the corresponding reconstruction accuracies are shown in Table 5.2.

As is expected due to the higher light efficiency, the setup using lenses and mirrors consistently yields better results for all evaluated methods. Since the estimation of the indirect peak with classical methods is almost impossible, the ‘peak finding and multilateration’ approach, however, performs poorly in all configurations. The best accuracy is achieved by training a neural net to directly predict the target coordinates from the four given histograms, which we therefore consider our proposed method. It performs particularly well with the big target and mirrors, but also yields fair results with the small target and even without the additional mirrors, but just the bare VL53L1X device.

Predicting the distance between the probe points on the wall and the target results in lower accuracy, but the performance drops only slightly when additionally shifting the histogram to make the model independent of the SPAD and wall position. All methods generalize reasonably well across target sizes.

Figure 5.11 shows the full distribution of distances in all spatial dimensions between the true and the predicted target position in the test dataset for the proposed method, where x denotes the dimension to the left and right (parallel to the wall), y is the distance from the wall, and z is the height from the floor. As visible in these plots, the method performs especially well at predicting the distance from the wall. The height from the floor apparently poses the greatest challenge, which could be attributed to the spacing of the probe points on the wall – due to the oblique projection of the sensor’s field of view onto the wall, the left and right corner points are set wider apart than the top from the bottom, providing a larger baseline in x -direction. The height reconstruction accuracy

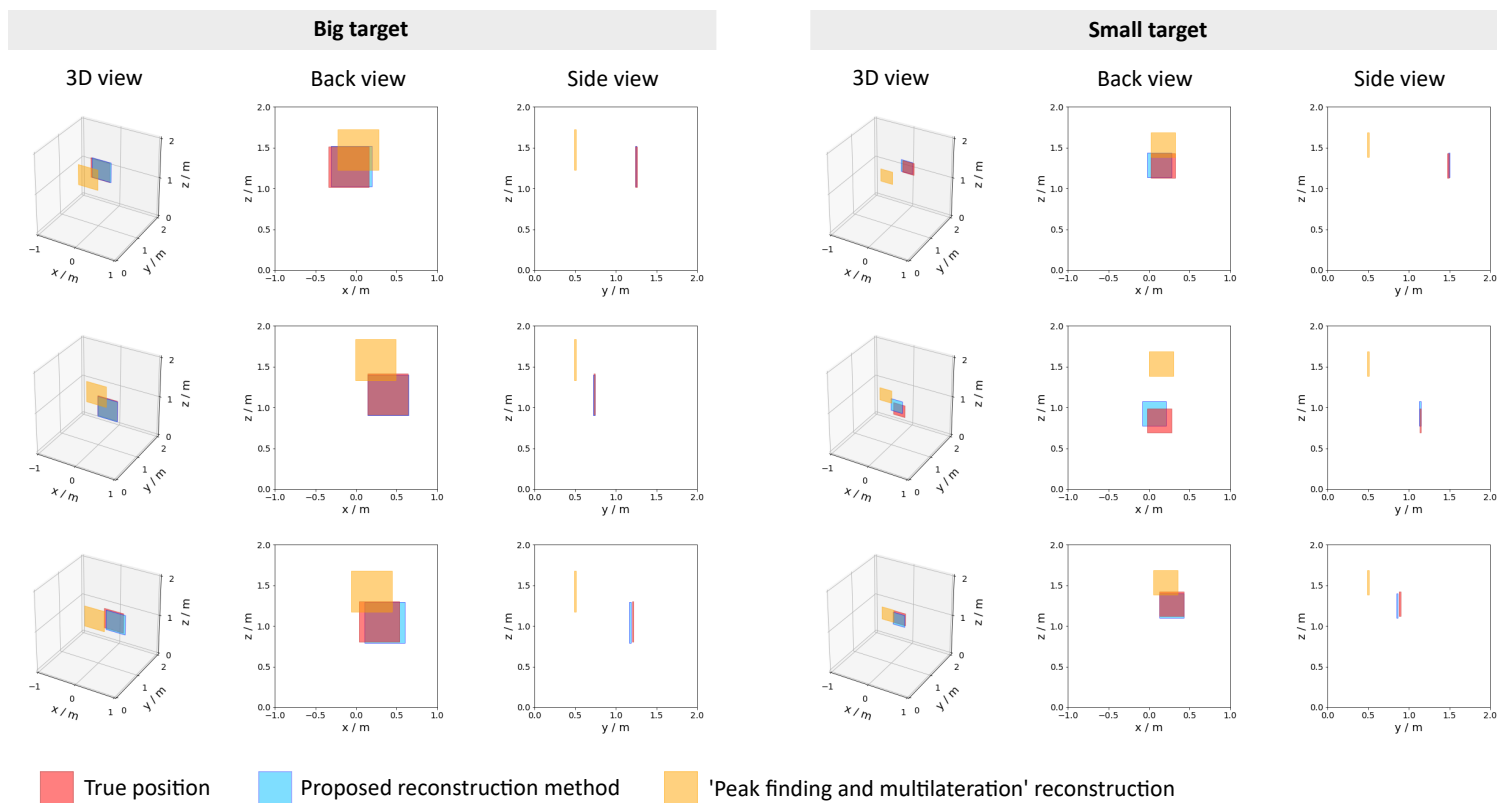


Figure 5.10: Exemplary reconstructions for both target sizes. Shown are the true target position, the reconstruction using the proposed method (direct prediction of position), and the peak finding and multilateration method for comparison. For better visibility, three different projections are shown for each target position: a three-dimensional view, a view from behind the target in viewing direction to the wall, and a side view. The relay wall is located in the x - z -plane at $y = 0$. Note that the red and the blue squares overlap almost completely in some plots which is why they appear as one purple square.

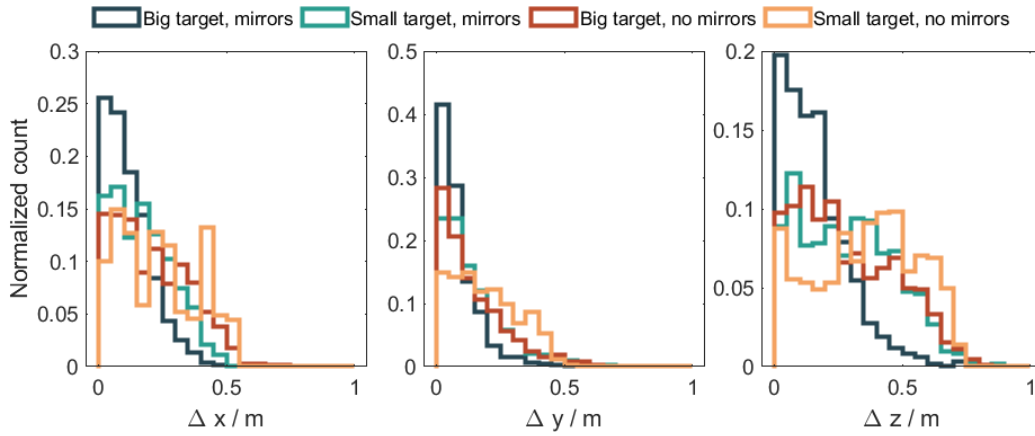


Figure 5.11: Distances in all spatial dimensions between the true and predicted target position in the test dataset for the 'direct position prediction' method.

also benefits the most from the better light efficiency through the bigger target and the collimating lenses.

A set of exemplary reconstructions with the proposed method and the 'peak finding and multilateration' approach for comparison is shown in Figure 5.10.

5.6 DEPTH IMAGING

The VL53L1X can yield a spatially resolved transient image by scanning all possible 4×4 ROIs on the 16×16 pixel sensor, which yields a 13×13 pixel measurement. This measurement, however, features a substantial blur due to the overlapping ROIs and the poor optical quality of the imaging lens. We therefore instead use the imager setup shown in Figure 5.2. We rotate the mirrors to 128×128 positions and capture a single all-sensor measurement per mirror position to acquire the final transient data cube (acquisition time ~ 30 minutes). In order to correct for internal automatic intensity corrections of the sensor, we normalize each pixel's histogram by the amount of light that is backscattered from the lens and its mounting which is collected in the first five temporal bins of each measurement.

Additionally, using a small retroreflector we measure the point spread function of the system as shown in Figure 5.3 (right) and employ the fast deconvolution method by Krishnan and Fergus [129] to deconvolve each temporal slice of the data cube with the measured PSF.

To keep intensity values relatively consistent to each other along the full depth of the measurement volume, we compensate the intensity falloff of the

light travelling from the illuminated point in the scene back to the sensor by multiplying each temporal bin n_i with i^2 .

Figure 5.12 (a)–(c) shows three different scenes scanned with our setup, where column (b) shows the raw data (each pixel’s intensity summed over all time bins) and column (c) shows the data after applying the above-mentioned corrections.

From the $128 \times 128 \times 19$ ‘corrected’ data cube, we calculate detailed depth maps in two different ways. In the first approach, we calculate the weighted mean of the histogram to use as the given pixel’s depth value d as

$$d = \frac{\sum_i i \cdot n_i}{\sum_i n_i}$$

where n_i is the intensity of bin i . This way we achieve sub-bin accuracy in the depth estimation, allowing even smooth depth gradients to be faithfully reproduced despite the large bin width that corresponds to ~ 40 cm (or ~ 20 cm in depth due to forth and back travel of the light). Results are shown in Figure 5.12(d), where the second scene has been recorded in the presence of ambient light (fluorescent ceiling lights). While this method produces detailed depth images, it lacks the ability to distinguish between fore- and background contributions and, instead of separating them, yields a mixture of both. This is especially relevant for highly specular surfaces that contribute to the measurement with direct and indirect reflections, where the latter take a longer time to arrive back at the sensor, as well as depth edges of objects where pixels contain contributions from the foreground object and the background. Additionally, this method suffers from a bias in very near and very far distances due to noise and secondary reflections dragging the center of mass of the histogram to the center of the distribution. This effect is furthermore dependent on the albedo of the imaged surface, as low peaks (in comparison to background noise) in the histogram will have less impact on the resulting mean than high peaks.

To mitigate these effects, we use a second approach where we fit Gaussian functions to the histogram of each pixel. While this method takes longer to compute, it yields sharper and more reliable results as shown in Figure 5.12(e). Especially the highly specular surface of the bucket in the third scene is much more accurately reconstructed than with the first method. Note that both methods produce false results for objects that are beyond the measurement range, as visible in the second scene where the background behind the shelves is too far away to be measured correctly.

The Gaussian fit method also allows a variety of active tweaking of the produced depth images. For example, varying sensitivity to back- or foreground can easily be implemented by changing the threshold for which of two detected peaks to be

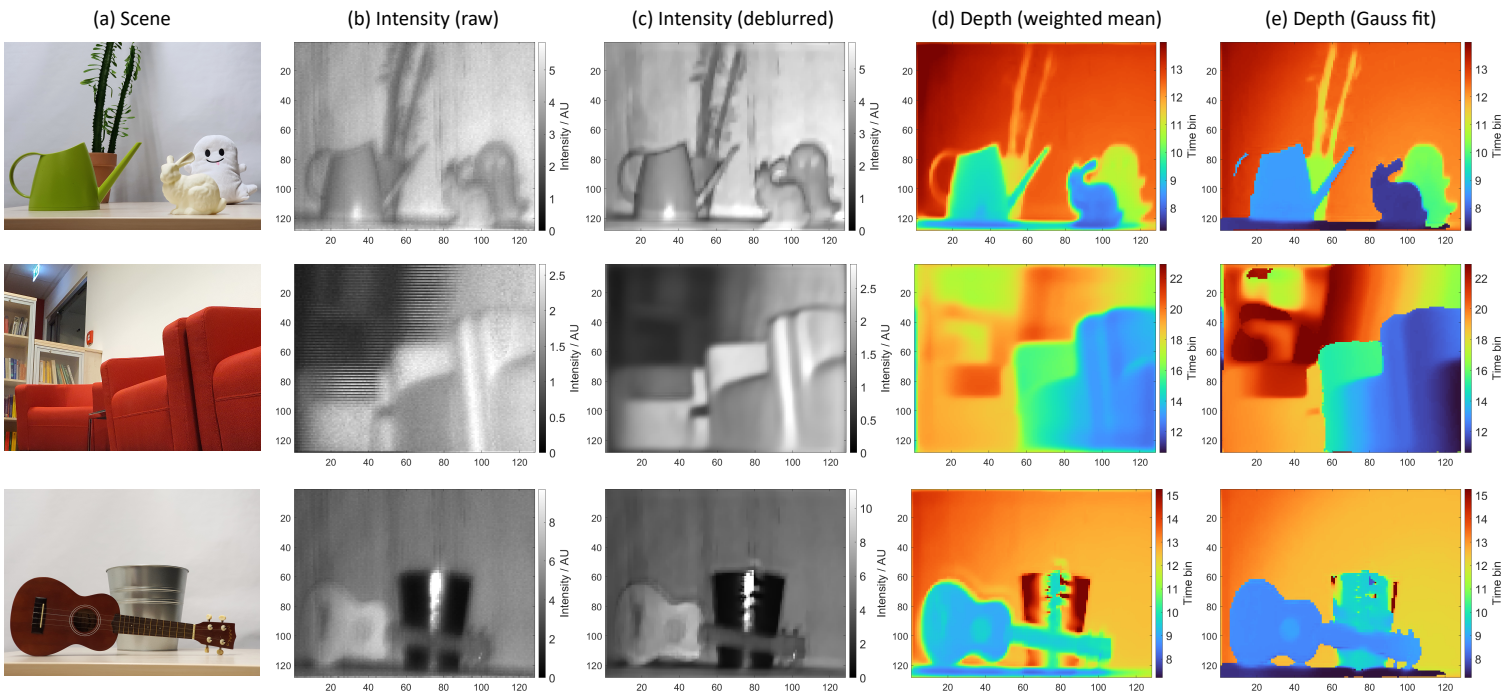


Figure 5.12: Measurements taken with our scanning setup of three different real scenes shown in column (a). Column (b) shows the measured intensity of each pixel as raw data. After applying several corrections to the data, we obtain clearer and less noisy data as shown in (c). From this data, depth maps are calculated in two different ways, once as a weighted mean as shown in column (d), and once by fitting Gaussian distributions to the measured histograms (column (e)).

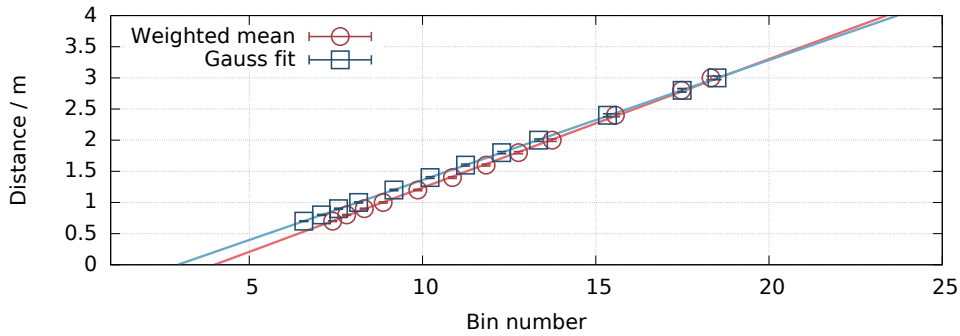


Figure 5.13: Calibration of the relation between bin number and true distance for both methods used to determine the position of the depth peak.

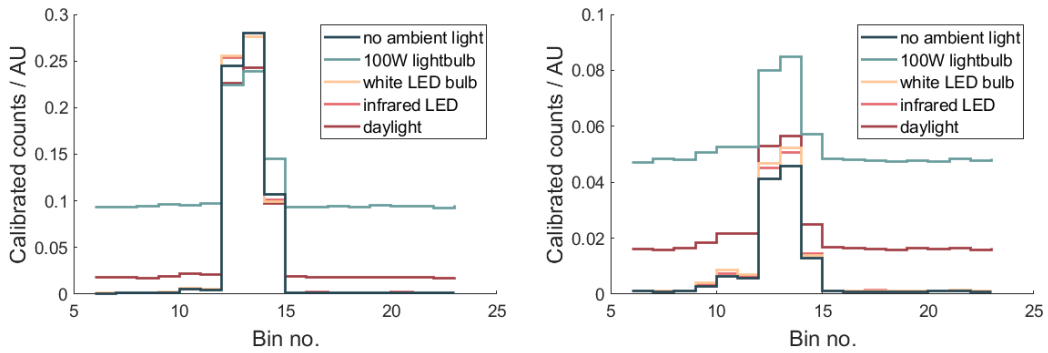


Figure 5.14: Exemplary histograms for a bright (left) and a dark (right) scene pixel (after intensity calibration using the first five bins of the raw histogram) for different ambient light scenarios (no ambient light, a 100 W incandescent lightbulb, a 230 lm white LED lightbulb, a 940 nm infrared LED, and indirect daylight through an open window).

used in the depth map. For semitransparent objects, contributions from the object itself and the background can be easily separated and selected independently of each other.

A calibration measurement of the relation between the obtained distance in units of bin numbers and the true distance is shown in Figure 5.13 for both methods. From the fitted linear relation, distances can be easily and accurately converted from bin numbers to meters for each method. This, however, does not account for the above-mentioned bias of the weighted mean-method towards differences in surface albedo.

In order to analyze the VL53L1X's response and robustness to different parameters like ambient light, object reflectivity, object distance and integration time, we imaged a scene with three flat targets arranged at different depths from the

sensor under different ambient lighting conditions. Each target consists of five patches that appear at different brightness at the 940 nm illumination of the light source. Ambient light sources (a 940 nm LED and a 100 W incandescent light bulb) have been placed next to the VL53L1X, in a distance of ca. 90 cm from the closest target, directly illuminating the scene. Ambient daylight (third row) was indirectly illuminating the target through an open window in a few meters distance from the scene. An exemplary histogram for one pixel, averaged over 10 measurements each, with different sources of ambient light is given in Figure 5.14: only the 100 W incandescent light bulb and, to some extent, daylight produce levels of ambient light that noticeably impact the measured histograms. Especially the former results in a decreased signal-to-noise ratio, which we observed to be more severe with higher total influx of light on the sensor, indicating that there is a general saturation threshold which was, however, not reached in any of our experiments. The ambient light background level is flat with no pile-up effects and can therefore be easily subtracted from the signal, however leaving an increased level of noise in the histograms. This has almost no effect on depth maps created using the Gauss fit method, but does induce a bias towards larger distances in the weighted mean method. Results for the whole scene are shown in Figure 5.15: Column (a) shows the level of total brightness for each measured pixel (integrated over all time bins) - note that the scale in the last row is adjusted to the significantly higher light intensity. Columns (b) and (c) show the depth maps of the scene acquired with the aforementioned methods. As illustrated in columns (d) and (e), the additional ambient light almost exclusively affects the weighted mean method (and more, the brighter the target), while in the Gauss fit depth maps, only slight deviations around depth edges are visible.

Since the VL53L1X does not allow adjustment of the integration time for each histogram, we acquired ten measurements for each ambient light situation and analyzed the variance of the depth values acquired from them as plotted in columns (f) and (g). Again, the weighted mean method shows larger variance as it is more susceptible to the increased noise level. In the Gauss fit depth maps, there is almost no variance except at depth edges where peak fitting is less precise due to contributions from back- and foreground. In conclusion, there is no significant benefit from averaging multiple histograms, and thus extended acquisition times, for this method. Especially given these limitations, we consider the weighted mean method a “quick-and-dirty” approach for a fast but noisy depth image while the Gauss fit method provides the accurate results.

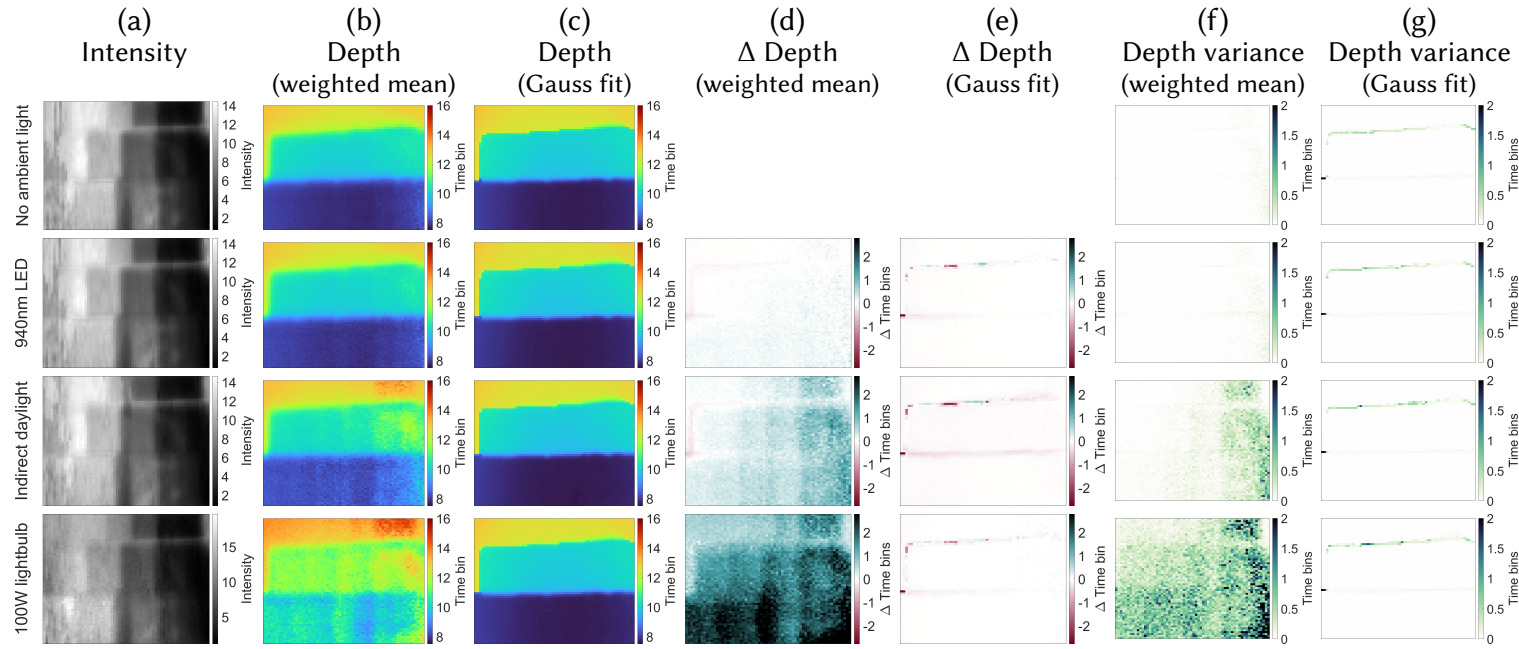


Figure 5.15: Measurements of three targets in different depths, each with five patches of different reflectivity at the illumination wavelength, for different levels / sources of ambient light (no ambient light, a 940 nm infrared LED, a 100 W incandescent lightbulb, and indirect daylight through an open window). (a) Total intensity integrated over all time bins - please note the adjusted colorscale in the last row. (b) Depth map acquired with the 'weighted mean' method. (c) Depth map acquired with the 'Gauss fit' method. (d)/(e) Difference of the depth maps at different ambient light levels to the depth maps acquired without any ambient light (first row). (f)/(g) Variance of the calculated depth over 10 measurements.

5.7 COMPARISON TO AMCW TOF

Due to their relatively low cost in comparison to high-end SPAD setups, amplitude modulated continuous wave (AMCW) ToF devices can be considered the closest alternative technology for the tasks shown in this paper. We therefore provide a short overview of the similarities and differences of the two technologies. In general, cheap SPAD sensors like the VL53L1X are still orders of magnitude cheaper than correlation-based ToF devices, making them especially suitable for applications where multiple sensors are involved, as well as more widespread in existing hardware where they are currently used, for instance, as proximity sensors, while correlation ToF sensors are still a rather specialized feature of a small number of device models. The designated application of correlation ToF sensors is depth imaging, which they are therefore best suited for, without requiring scanning of the scene. While ToF sensors directly measure a single depth value per pixel, SPAD sensors yield a response histogram for each pixel, providing complex light transport information. Heide et al. [17] and Kadambi et al. [36] have shown that such temporally resolved information can be recorded using photonic mixer device (PMD) measurements, which however requires hours of post processing time. Even non-line-of-sight imaging with PMD sensors has been shown by Heide et al. [130]; their setup uses six 250 mW laser diodes as illumination sources, while the VL53L1X features an eye-safe class 1 laser source.

Su et al. [113] have demonstrated material classification using raw PMD measurements on similar material samples as used in our paper, but from larger distances. Since in our approach the whole device, including the sensor and light source, has such a small form factor and can be placed directly on the material surface, spatial and angular characteristics of the scattered light can be exploited. We achieve much better classification accuracies than Su et al. and do not require a combination of multiple samples or manual segmentation of the recorded data.

5.8 DISCUSSION AND FUTURE WORK

Although cheap SPADs deliver low-quality data and are narrowly optimized for short-distance, single-point ranging applications, we have been able to demonstrate that, by configuring these sensors to output raw photon counts, they can be opened to a wider range of computational sensing tasks. They are readily available in existing consumer devices. With nonstandard programming, they could instantly enable new features for a wide audience without the need for additional hardware. Furthermore, the affordability of these chips even qualifies

them for use in larger quantities (arrays), which would be prohibitively expensive for most research-grade SPAD systems.

The limitations of our prototype system (an off-the-shelf sensor evaluation kit with custom firmware) are numerous. Although the sensor can capture at 60 histograms per second, overhead from API and serial communication reduces the effective capture rate to about 9 histograms per second. Despite the relatively fine grid of individual detector pixels, they cannot be read out individually but only in blocks. We assume that the manufacturers want to ensure a minimum photon count and hence signal-to-noise ratio. For our purposes, however, the main effect is a reduced spatial resolution. While the spatial resolution is potentially improvable with additional hardware [8], the temporal resolution of 24 bins per histogram, too, is barely enough for most applications. Additional temporal blur due to laser pulse length and detection jitter further reduces the capability to discriminate components with similar time of flight. Finally, the light output of our system is limited by the strict requirements of low power consumption and eye safety in mobile applications. Most experimental work on non-line-of-sight sensing, on the other hand, uses optical powers that are at least 3–4 orders of magnitude higher.

To facilitate the broader adoption of the proposed sensor platform, we hope that the sensor manufacturer alleviates these limitations by making minor technical modifications to the sensor design and its API. An API change could be official support of histogram readout, at the same level of configurability that is offered for the ranging mode. An increased number of histogram bins, even if temporal blur is the limiting factor and reduced photon count per bin leads to a higher shot noise level, would enable computational recovery of a higher-resolved signal. With the option of using region-of-interest sizes down to a single pixel, it would be possible to obtain noisy photon counts at full sensor resolution, which can be more useful for many applications than pre-binned counts. In the future, we hope that an even wider range of applications could be enabled by deeper hardware modifications to the sensor. A major difference between existing scientific time-tagged SPAD systems and the mass-market proximity sensor used in this work is the ability of synchronized triggered acquisition. The proposed system is fully free-running in the sense that it does not facilitate external triggered acquisition but only delivers repeated measurements using the internal source. As such, the user can not “start” or “stop” measurements but only read out continuous measurement streams. In the future, a trigger interface could enable the triggering of the sensor by external light sources, or vice versa. With changes only to the logic portion of the chip (introducing a time-to-digital converter per pixel), full

image data could be captured. Finally, we envision a bare-sensor version (without lens) that would allow for the use of higher-quality optics for better image quality.

With the ongoing establishment of SPAD sensors as a technology for time-resolved imaging, which is accompanied by the in-depth analysis and modelling of these sensor devices [131], we believe that the data quality even from cheap and simple sensors will improve rapidly and that they will open up affordable research using SPAD sensors across disciplines.

CONCLUSION AND OUTLOOK

The work presented in this thesis is based on and part of a rapidly evolving body of related scientific work and technological developments. Time-of-Flight camera systems are a relatively new technology and the field of time-resolved imaging advances rapidly. Automation, smart consumer devices, and emerging media like virtual and augmented reality drive the fast progress in this field, both hardware and software wise, as the need for a thorough automated capture and comprehension of physical surroundings and conditions grows. At the same time, the different fields of applications pose a variety of demands and challenges on the required technology like price, size, eye-safety, portability, etc.

Similarly, imaging and photography has generally become an increasingly important topic over the last years, as smartphones and online services have made the production and sharing of images easy and accessible. The demand for small, portable cameras has therefore greatly increased and computational photography as a means of enhancing image quality and features while exhausting the technological and optical limits of the imaging hardware has become more and more important. In recent years, even the combination of ToF and intensity cameras became a relevant mode in smartphone cameras, mostly to achieve out-of-focus blur and bokeh effects as an aesthetically desired natural optical feature of bigger cameras with larger apertures. This fixed and well-calibratable setup provides a convenient basis for all sorts of algorithms that combine intensity and ToF data, like the one presented in Chapter 4. The sensor type that is used in Chapter 5 is already used in many smartphones as proximity sensors. With the trend of using much higher quality SPAD sensors in recent devices like Apple iPhones and iPads [132], the results that have been shown for this cheap sensor type will likely scale to even better quality and enable many new sensing modes for these devices.

In applications like robotics and autonomous driving, depth sensing technologies are essential as machines need to be able to capture and process their environment in order to interact with and move inside it. SPAD sensors are becoming an increasingly important technology in this context, as they are able to capture the full light response of their surroundings, making them usable in different conditions and allowing them to draw inferences even about objects that may not be in the direct line of sight. SPAD sensor hardware is progressing rapidly, with the first megapixel SPAD sensor having been presented recently [133] and a 3.2 megapixel SPAD sensor being announced by Canon [134], and with new architectures like back-illuminated sensors emerging [135, 136]. Metrics like resolution, fill factor, quantum efficiency, and temporal resolution of SPAD sensors are therefore expected to improve even more in the future. Similarly, however, CCD and CMOS sensors are still improving as well, keeping a method for combining SPAD and ToF sensors like the one presented in Chapter 4 to improve SPAD data resolution a relevant option. This is especially true wherever cost, size, and power are limiting factors, like in consumer devices.

The progress in ToF hardware is closely connected to the advances in computational imaging algorithms, as is also reflected in the work presented in this thesis. The snapshot difference imaging approach presented in Chapter 3 and the low-cost SPAD sensing applications in Chapter 5 are in line with a general trend of making imaging sensors more versatile, and using programmable hardware to encode more image information than one would obtain through simple integration of light. Multiple approaches to general two-bucket sensors that allow for a flexible switching between buckets have recently been proposed [137–139], allowing to encode the measured image spatially and temporally, enabling for example compressive sensing techniques. SPAD sensors have been shown to be useful to greatly increase the dynamic range in imaging applications [100], and methods to fuse intensity and SPAD sensor data in order to obtain very high dynamic range photographs have recently been proposed [140].

ToF hardware is likely going to improve further in the future, as the above-mentioned developments suggest. All methods presented in this thesis generally scale to improved sensor characteristics, although processing times may become an issue, especially with the sensor fusion approach presented in Chapter 4. Reducing the runtime of the high-resolution data reconstruction, for example by finding more efficient ways to solve the proposed optimization problem, would be a sensible leverage point for a future improvement of this method. A deep learning approach is also conceivable, as fusion of intensity and SPAD image

data has been demonstrated for robust and noise-efficient depth estimation [83, 141] – although not yet for reconstruction of full transient images as are required for applications like FLIM or NLOS imaging.

Even with the general improvement of high-end ToF technology, though, cost-efficient solutions and hardware that is able to perform under a variety of other external restrictions will most probably remain a relevant challenge. The extension of working modes for existing hardware may also enable new applications for already existing and widespread, as well as future consumer devices. A specific issue that could be investigated based on this thesis is whether the material classification demonstrated in Section 5.4 can be extended to larger distances between the SPAD device and the material. Especially in setups where a SPAD camera is aligned with an intensity or RGB camera, material classification could help in the recognition and identification of objects. The usage of low-cost SPAD sensors for this purpose could provide an easy and cheap way to improve the understanding that robots or autonomous cars can obtain about their environment or objects they might interact with. Also conceivable are consumer applications like the automatic estimation of caloric content or nutrients of a meal by scanning it with a smartphone camera system that includes ToF sensors.

Generally, ToF imaging and ToF-based sensor systems will likely open up many new applications in the future in a variety of different fields like medicine and life science, robotics, autonomous driving, AR/VR, and consumer photography. Virtual, augmented, and mixed reality might be on track to becoming comparably influential and widespread as smartphones today, for example through smart glasses or contact lenses [142]. In that case, ToF will most probably become an essential and ubiquitous sensing mode, whose capabilities and prospects are therefore worth exploring extensively, even today.

BIBLIOGRAPHY

- [1] Wolfgang Baier, *Quellendarstellungen zur Geschichte der Fotografie*, Schirmer-Mosel, 1977, ISBN: 9783921375150 (cit. on p. 1).
- [2] J. Mauritsson, P. Johnsson, E. Mansten, M. Swoboda, T. Ruchon, A. L'Huillier and K. J. Schafer, *Coherent Electron Scattering Captured by an Attosecond Quantum Stroboscope*, *Physical Review Letters* **100** (2008), URL: <https://doi.org/10.1103/physrevlett.100.073003> (cit. on p. 1).
- [3] Kazunori Akiyama et al., *First M87 Event Horizon Telescope Results. IV. Imaging the Central Supermassive Black Hole*, *The Astrophysical Journal. Letters (Online)* **875** (2019) (cit. on p. 1).
- [4] Wilhelm Conrad Röntgen, *On a New Kind of Rays*, *Nature* **53** (1896) 274, URL: <https://doi.org/10.1038/053274b0> (cit. on p. 1).
- [5] Ahmed Kirmani, Tyler Hutchison, James Davis and Ramesh Raskar, *Looking around the corner using transient imaging*, *2009 IEEE 12th International Conference on Computer Vision*, IEEE, 2009 159 (cit. on p. 1).
- [6] John G. Skellam, *The Frequency Distribution of the Difference Between Two Poisson Variates Belonging to Different Populations*, *Journal of the Royal Statistical Society* **109** (1946) 296, ISSN: 09528385, URL: <http://www.jstor.org/stable/2981372> (cit. on pp. 5, 36).
- [7] Clara Callenberg, Felix Heide, Gordon Wetzstein and Matthias B. Hullin, *Snapshot Difference Imaging using Time-of-Flight Sensors*, *ACM Trans. Graph. (Proc. SIGGRAPH Asia)* **36** (2017) 220:1 (cit. on p. 5).
- [8] C. Callenberg, A. Lyons, D. den Brok, A. Fatima, A. Turpin, V. Zickus, L. Machesky, J. Whitelaw, D. Faccio and M. B. Hullin, *Super-resolution time-resolved imaging using computational sensor fusion*, *Scientific Reports* **11** (2021) 1689, ISSN: 2045-2322, URL: <https://doi.org/10.1038/s41598-021-81159-x> (cit. on pp. 7, 112).

- [9] Clara Callenberg, Zheng Shi, Felix Heide and Matthias B. Hullin, *Low-Cost SPAD Sensing for Non-Line-Of-Sight Tracking, Material Classification and Depth Imaging*, ACM Trans. Graph. (SIGGRAPH) **40** (2021) (cit. on p. 8).
- [10] Wolfgang Demtröder, *Experimentalphysik 2: Elektrizität und Optik; 3rd ed.* Springer, 2004 (cit. on pp. 12, 14, 15, 17).
- [11] B. Thide, *Electromagnetic Field Theory*, Dover Publications, Incorporated, 2011, ISBN: 9780486477732 (cit. on p. 13).
- [12] Eugene Hecht, *Optics*, Pearson, 2012, ISBN: 9788131718070 (cit. on p. 17).
- [13] Wolfgang Demtröder, *Experimentalphysik 3: Atome, Moleküle und Festkörper; 3rd ed.* Springer-Lehrbuch, Springer, 2005 (cit. on pp. 18, 20).
- [14] Bahaa E. A. Saleh and Malvin Carl Teich, *Fundamentals of Photonics (Wiley Series in Pure and Applied Optics)*, Third Edition, John Wiley & Sons, 2019, ISBN: 978-1-119-70211-5 (cit. on pp. 19, 21).
- [15] William R Leo, *Techniques for nuclear and particle physics experiments: a how-to approach; 2nd ed.* Springer, 1994, URL: <https://cds.cern.ch/record/302344> (cit. on p. 19).
- [16] Ulrich Teubner and Hans Josef Brückner, *Optical Imaging and Photography: Introduction to Science and Technology of Optics, Sensors and Systems*, De Gruyter, 2019, ISBN: 9783110472943, URL: <https://doi.org/10.1515/9783110472943> (cit. on pp. 21, 22).
- [17] Felix Heide, Matthias B. Hullin, James Gregson and Wolfgang Heidrich, *Low-Budget Transient Imaging using Photonic Mixer Devices*, ACM Trans. Graph. **32** (2013) 45:1 (cit. on pp. 23, 32, 36, 58, 91, 111).
- [18] Cristiano Niclass, Claudio Favi, Theo Kluter, FrÉdÉric Monnier and Edoardo Charbon, *Single-Photon Synchronous Detection*, IEEE Journal of Solid-State Circuits **44** (2009) 1977, URL: <https://doi.org/10.1109/jssc.2009.2021920> (cit. on p. 23).
- [19] Edoardo Charbon, Matt Fishburn, Richard Walker, Robert K. Henderson and Cristiano Niclass, "SPAD-Based Sensors", *TOF Range-Imaging Cameras*, Springer Berlin Heidelberg, 2013 11, URL: https://doi.org/10.1007/978-3-642-27523-4_2 (cit. on pp. 23, 27).

-
- [20] Akira Takahashi, Mitsunori Nishizawa, Yoshinori Inagaki, Musubu Koishi and Katsuyuki Kinoshita, *New femtosecond streak camera with temporal resolution of 180 fs, Generation, Amplification, and Measurement of Ultrashort Laser Pulses*, ed. by Rick P. Trebino and Ian A. Walmsley, vol. 2116, International Society for Optics and Photonics, SPIE, 1994 275, URL: <https://doi.org/10.1117/12.175863> (cit. on p. 23).
- [21] Liang Gao, Jinyang Liang, Chiye Li and Lihong V Wang, *Single-shot compressed ultrafast photography at one hundred billion frames per second*, *Nature* **516** (2014) 74 (cit. on pp. 23, 58, 91).
- [22] Jinyang Liang, Cheng Ma, Liren Zhu, Yujia Chen, Liang Gao and Lihong V Wang, *Single-shot real-time video recording of a photonic Mach cone induced by a scattered light pulse*, *Science advances* **3** (2017) e1601814 (cit. on pp. 23, 58, 91).
- [23] Rudolf Schwarte, Zhanping Xu, Horst-Guenther Heinol, Joachim Olk, Ruediger Klein, Bernd Buxbaum, Helmut Fischer and Juergen Schulte, *New electro-optical mixing and correlating sensor: facilities and applications of the photonic mixer device (PMD)*, *Sensors, Sensor Systems, and Sensor Data Processing*, ed. by Otmar Loffeld, vol. 3100, International Society for Optics and Photonics, SPIE, 1997 245, URL: <https://doi.org/10.1117/12.287751> (cit. on pp. 24, 25).
- [24] T. Spirig, P. Seitz, O. Vietze and F. Heitger, *The lock-in CCD-two-dimensional synchronous detection of light*, *IEEE Journal of Quantum Electronics* **31** (1995) 1705, URL: <https://doi.org/10.1109/3.406386> (cit. on pp. 25, 26).
- [25] Sergi Foix, G. Alenyà and Carme Torras, *Lock-in Time-of-Flight (ToF) Cameras: A Survey*, *IEEE Sensors Journal* **11** (2011) 1917 (cit. on p. 25).
- [26] *Introduction to the Time-of-Flight (ToF) System Design – User’s Guide*, tech. rep., Texas Instruments, 2014 (cit. on p. 26).
- [27] Robert Lange, *3D time-of-flight distance measurement with custom solid-state image sensors in CMOS/CCD-technology*, PhD thesis: Universität Siegen, 2000, URL: <https://dspace.ub.uni-siegen.de/handle/ubsi/178> (cit. on p. 26).
- [28] S. Cova, M. Ghioni, A. Lacaita, C. Samori and F. Zappa, *Avalanche photodiodes and quenching circuits for single-photon detection*, *Appl. Opt.* **35** (1996) 1956, URL: <http://www.osapublishing.org/ao/abstract.cfm?URI=ao-35-12-1956> (cit. on p. 27).

- [29] Shree K Nayar, Gurunandan Krishnan, Michael D Grossberg and Ramesh Raskar, *Fast separation of direct and global components of a scene using high frequency illumination*, ACM Trans. Graph. (Proc. SIGGRAPH) **25** (2006) 935 (cit. on pp. 29, 38).
- [30] Jack Tumblin, Amit Agrawal and Ramesh Raskar, *Why I want a gradient camera*, Proc. IEEE CVPR, vol. 1, 2005 103 (cit. on p. 29).
- [31] Patrick Lichtsteiner, Christoph Posch and Tobi Delbruck, *A 128×128 120 dB 15 s latency asynchronous temporal contrast vision sensor*, Solid-State Circuits, IEEE Journal of **43** (2008) 566 (cit. on pp. 29, 33, 43).
- [32] Ramesh Raskar, Kar-Han Tan, Rogerio Feris, Jingyi Yu and Matthew Turk, *Non-photorealistic Camera: Depth Edge Detection and Stylized Rendering Using Multi-flash Imaging*, ACM Trans. Graph. (Proc. SIGGRAPH) **23** (2004) 679 (cit. on pp. 29, 41).
- [33] Cyrus Bamji et al., *A 0.13 um CMOS System-on-Chip for a 512×424 Time-of-Flight Image Sensor With Multi-Frequency Photo-Demodulation up to 130 MHz and 2 GS/s ADC*, IEEE Journal of Solid-State Circuits **50** (2015) 303 (cit. on p. 31).
- [34] Robert Lange, Peter Seitz, Alice Biber and Rudolf Schwarte, *Time-of-flight range imaging with a custom solid state image sensor*, Industrial Lasers and Inspection (EUROPTO Series), International Society for Optics and Photonics, 1999 180 (cit. on pp. 32, 34).
- [35] Miles Hansard, Seungkyu Lee, Ouk Choi and Radu Horaud, *Time of Flight Cameras: Principles, Methods, and Applications*, Springer, 2012 (cit. on p. 32).
- [36] A. Kadambi, R. Whyte, A. Bhandari, L. Streeter, C. Barsi, A. Dorrington and R. Raskar, *Coded time of flight cameras: sparse deconvolution to address multipath interference and recover time profiles*, ACM Trans. Graph. **32** (2013) (cit. on pp. 32, 91, 111).
- [37] Felix Heide, Lei Xiao, Andreas Kolb, Matthias B. Hullin and Wolfgang Heidrich, *Imaging in scattering media using correlation image sensors and sparse convolutional coding*, Opt. Express **22** (2014) 26338, URL: <http://www.opticsexpress.org/abstract.cfm?URI=oe-22-21-26338> (cit. on p. 32).
- [38] Ryuichi Tadano, Adithya Kumar Pediredla and Ashok Veeraraghavan, *Depth Selective Camera: A Direct, On-Chip, Programmable Technique for Depth Selectivity in Photography*, 2015 IEEE International Conference on Computer Vision (ICCV), IEEE, 2015, URL: [https://doi.org/10.1109%2Ficcv.2015.410](https://doi.org/10.1109/Ficcv.2015.410) (cit. on p. 32).

-
- [39] Albert Wang, Sriram Sivaramakrishnan and Alyosha Molnar, *A 180nm CMOS image sensor with on-chip optoelectronic image compression*, *Custom Integrated Circuits Conference (CICC), 2012 IEEE*, IEEE, 2012 1 (cit. on p. 32).
- [40] Albert Wang and Alyosha Molnar, *A Light-Field Image Sensor in 180 nm CMOS*, *IEEE Journal of Solid-State Circuits* **47** (2012) 257, URL: <https://doi.org/10.1109/JSSC.2011.2164669> (cit. on p. 33).
- [41] Assaf Zomet and Shree K Nayar, *Lensless imaging with a controllable aperture*, *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, vol. 1, IEEE, 2006 339 (cit. on p. 33).
- [42] Sanjeev J. Koppal, Ioannis Gkioulekas, Travis Young, Hyunsung Park, Kenneth B. Crozier, Geoffrey L. Barrows and Todd E. Zickler, *Toward Wide-Angle Microvision Sensors*, *IEEE Trans. Pattern Anal. Mach. Intell.* **35** (2013) 2982 (cit. on p. 33).
- [43] Massimo Gottardi, Nicola Massari and Syed Arsalan Jawed, *A 100 μ W 128 \times 64 Pixels Contrast-Based Asynchronous Binary Vision Sensor for Sensor Networks Applications*, *IEEE Journal of Solid-State Circuits* **44** (2009) 1582 (cit. on p. 33).
- [44] Hanme Kim, Stefan Leutenegger and Andrew J Davison, *Real-time 3D reconstruction and 6-DoF tracking with an event camera*, *European Conference on Computer Vision*, Springer, 2016 349 (cit. on pp. 33, 34, 55).
- [45] David Weikersdorfer, David B Adrian, Daniel Cremers and Jörg Conradt, *Event-based 3D SLAM with a depth-augmented dynamic vision sensor*, *Robotics and Automation (ICRA), 2014 IEEE International Conference on*, IEEE, 2014 359 (cit. on p. 33).
- [46] Trygve Willassen et al., *A 1280 \times 1080 4.2 μ m split-diode pixel HDR sensor in 110nm BSI CMOS process*, *Int. Image Sensor Workshop*, 2015 (cit. on p. 34).
- [47] Johannes Solhusvik, Jiangtao Kuang, Zhiqiang Lin, Sohei Manabe, J Lyu, Howard Rhodes et al., *A comparison of high dynamic range CIS technologies for automotive applications*, *Proc. 2013 Int. Image Sensor Workshop (IISW)*, 2013 (cit. on p. 34).
- [48] Arnaud Darmont, *High Dynamic Range Imaging: Sensors and Architectures*, SPIE, 2012 (cit. on p. 34).
- [49] Shree K Nayar and Vlad Branzoi, *Adaptive Dynamic Range Imaging: Optical Control of Pixel Exposures Over Space and Time.*, *ICCV*, 2003 1168 (cit. on p. 34).

- [50] Gordon Wan, Xiangli Li, Gennadiy Agranov, Marc Levoy and Mark Horowitz, *CMOS image sensors with multi-bucket pixels for computational photography*, Solid-State Circuits, IEEE Journal of **47** (2012) 1031 (cit. on p. 34).
- [51] Mirko Schmidt, *Analysis, modeling and dynamic optimization of 3D time-of-flight imaging systems*, PhD thesis: University of Heidelberg, 2011 (cit. on pp. 35, 36).
- [52] Youngbae Hwang, Jun-Sik Kim and In So Kweon, *Difference-based image noise modeling using Skellam distribution*, IEEE Trans. PAMI **34** (2012) 1329 (cit. on p. 36).
- [53] Shikhar Shrestha, Felix Heide, Wolfgang Heidrich and Gordon Wetzstein, *Computational Imaging with Multi-Camera Time-of-Flight Systems*, ACM Trans. Graph. (Proc. SIGGRAPH) (2016) (cit. on p. 36).
- [54] Wan-Chun Ma, Tim Hawkins, Pieter Peers, Charles-Felix Chabert, Malte Weiss and Paul Debevec, *Rapid acquisition of specular and diffuse normal maps from polarized spherical gradient illumination*, Proc. EGSR, Eurographics Association, 2007 183 (cit. on p. 38).
- [55] Tongbo Chen, Hendrik PA Lensch, Christian Fuchs and Hans-Peter Seidel, *Polarization and phase-shifting for 3D scanning of translucent objects*, Proc. IEEE CVPR, IEEE, 2007 1 (cit. on p. 38).
- [56] Matthew O'Toole, Supreeth Achar, Srinivasa G Narasimhan and Kiriakos N Kutulakos, *Homogeneous codes for energy-efficient illumination and imaging*, ACM Trans. Graph. (Proc. SIGGRAPH) **34** (2015) 35 (cit. on pp. 38, 55).
- [57] Matthew O'Toole, Ramesh Raskar and Kiriakos N Kutulakos, *Primal-dual coding to probe light transport*, ACM Trans. Graph. (Proc. SIGGRAPH) **31** (2012) 39 (cit. on p. 38).
- [58] Lawrence B. Wolff, *Polarization-based material classification from specular reflection*, IEEE Transactions on Pattern Analysis and Machine Intelligence **12** (1990) 1059 (cit. on p. 38).
- [59] Chao Liu and Jinwei Gu, *Discriminative illumination: Per-pixel classification of raw materials based on optimal projections of spectral BRDF*, IEEE Trans. PAMI **36** (2014) 86 (cit. on p. 40).
- [60] Robert J Woodham, *Photometric method for determining surface orientation from multiple images*, Optical engineering **19** (1980) 191139 (cit. on p. 42).

-
- [61] Dabi Wei, Paul Masurel, Toru Kurihara and Shigeru Ando, *Optical flow determination with complex-sinusoidally modulated imaging*, *Signal Processing, 2006 8th International Conference on*, vol. 2, IEEE, 2006 (cit. on p. 44).
- [62] Hidekata Hontani, Go Oishi and Tomohiro Kitagawa, *Local estimation of high velocity optical flow with correlation image sensor*, *European Conference on Computer Vision*, Springer, 2014 235 (cit. on p. 44).
- [63] Carlo Tomasi and Roberto Manduchi, *Bilateral filtering for gray and color images.*, *Iccv*, vol. 98, 1, 1998 2 (cit. on p. 51).
- [64] N. Matsuda, O. Cossairt and M. Gupta, *MC3D: Motion Contrast 3D Scanning*, *Proc. ICCP*, 2015 (cit. on p. 55).
- [65] Genevieve Gariepy, Nikola Krstajić, Robert Henderson, Chunyong Li, Robert R Thomson, Gerald S Buller, Barmak Heshmat, Ramesh Raskar, Jonathan Leach and Daniele Faccio, *Single-photon sensitive light-in-flight imaging*, *Nature communications* 6 (2015) 6021 (cit. on pp. 58, 65, 88).
- [66] Andreas Velten, Thomas Willwacher, Otkrist Gupta, Ashok Veeraraghavan, Mounqi G. Bawendi and Ramesh Raskar, *Recovering three-dimensional shape around a corner using ultrafast time-of-flight imaging*, *Nature Communications* 3 (2012) 745, ISSN: 2041-1723 (cit. on pp. 58, 88, 91).
- [67] Mauro Buttafava, Jessica Zeman, Alberto Tosi, Kevin Eliceiri and Andreas Velten, *Non-line-of-sight imaging using a time-gated single photon avalanche diode*, *Optics express* 23 (2015) 20997 (cit. on pp. 58, 91).
- [68] Daniele Faccio and Andreas Velten, *A trillion frames per second: The techniques and applications of light-in-flight photography*, *Reports on Progress in Physics* 81 (2018), ISSN: 00344885, arXiv: 1808.00428 (cit. on p. 58).
- [69] T. Durduran, R. Choe, W. B. Baker and A. G. Yodh, *Diffuse optics for tissue monitoring and tomography*, *Reports on Progress in Physics* 73 (2010), ISSN: 00344885, arXiv: 15334406 (cit. on p. 58).
- [70] Guy Satat, Barmak Heshmat, Dan Raviv and Ramesh Raskar, *All Photons Imaging Through Volumetric Scattering*, *Scientific Reports* 6 (2016) 1, ISSN: 20452322, URL: <http://dx.doi.org/10.1038/srep33946> (cit. on p. 58).

- [71] Ashley Lyons, Francesco Tonolini, Alessandro Boccolini, Audrey Repetti, Robert Henderson, Yves Wiaux and Daniele Faccio, *Computational time-of-flight diffuse optical tomography*, *Nature Photonics* **13** (2019) 575, ISSN: 1749-4885, URL: <http://www.nature.com/articles/s41566-019-0439-x> (cit. on p. 58).
- [72] K Dowling, M J Dayel, S C W Hyde, P M W French, M J Lever, J D Hares and A K L Dymoke-Bradshaw, *High resolution time-domain fluorescence lifetime imaging for biomedical applications*, *Journal of Modern Optics* **46** (1999) 199, URL: <https://www.tandfonline.com/doi/abs/10.1080/09500349908231265> (cit. on p. 58).
- [73] Nils Abramson, *Light-in-flight recording by holography*, *Optics letters* **3** (1978) 121 (cit. on pp. 58, 91).
- [74] Andreas Velten, Di Wu, Adrian Jarabo, Belen Masia, Christopher Barsi, Chinmaya Joshi, Everett Lawson, Mounji Bawendi, Diego Gutierrez and Ramesh Raskar, *Femto-photography: capturing and visualizing the propagation of light*, *ACM Trans. Graph.* **32** (2013) 44 (cit. on pp. 58, 91).
- [75] Lucrezia Cester, Ashley Lyons, Maria Braidotti and Daniele Faccio, *Time-of-Flight Imaging at 10 ps Resolution with an ICCD Camera*, *Sensors* **19** (2019) 180 (cit. on p. 58).
- [76] Justin A. Richardson, Lindsay A. Grant and Robert K. Henderson, *Low dark count single-photon avalanche diode structure compatible with standard nanometer scale CMOS technology*, *IEEE Photonics Technology Letters* **21** (2009) 1020, ISSN: 10411135 (cit. on pp. 58, 65, 88).
- [77] Robert K. Henderson, Nick Johnston, Haochang Chen, David Day Uei Li, Graham Hungerford, Richard Hirsch, David McLoskey, Philip Yip and David J.S. Birch, *A 192x128 Time Correlated Single Photon Counting Imager in 40nm CMOS Technology*, *ESSCIRC 2018 - IEEE 44th European Solid State Circuits Conference* (2018) 54 (cit. on pp. 58, 88, 89).
- [78] Mushfiqur Rouf, Rafal Mantiuk, Wolfgang Heidrich, Matthew Trentacoste and Cheryl Lau, *Glare encoding of high dynamic range images*, *CVPR 2011*, IEEE, 2011 289, ISBN: 978-1-4577-0394-2, URL: <http://ieeexplore.ieee.org/document/5995335/> (cit. on p. 59).
- [79] P Favaro and S Soatto, *3-D Shape Estimation and Image Restoration*, Springer London, 2007, ISBN: 978-1-84628-176-1, URL: https://books.google.co.uk/books?id=fro%7B%5C_%7DStKBG6sC%20http://link.springer.com/10.1007/978-1-84628-688-9 (cit. on p. 59).

-
- [80] Qilin Sun, Jian Zhang, Xiong Dun, Bernard Ghanem, Yifan Peng and Wolfgang Heidrich, *End-to-end Learned, Optically Coded Super-resolution SPAD Camera*, *ACM Transactions on Graphics* **39** (2020) 1, ISSN: 0730-0301, URL: <https://dl.acm.org/doi/10.1145/3372261> (cit. on p. 59).
- [81] Matthew O’Toole, Felix Heide, David B Lindell, Kai Zang, Steven Diamond and Gordon Wetzstein, *Reconstructing transient images from single-photon sensors*, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017 1539 (cit. on pp. 60, 67).
- [82] David B Lindell, Matthew O’Toole and Gordon Wetzstein, *Towards transient imaging at interactive rates with single-photon detectors*, *2018 IEEE International Conference on Computational Photography (ICCP)*, IEEE, 2018 1 (cit. on p. 60).
- [83] David B Lindell, Matthew O’Toole and Gordon Wetzstein, *Single-photon 3D imaging with deep sensor fusion*, *ACM Trans. Graph.* **37** (2018) 113 (cit. on pp. 60, 81–85, 91, 117).
- [84] Michael Grant and Stephen Boyd, *CVX: Matlab Software for Disciplined Convex Programming, version 2.1*, <http://cvxr.com/cvx>, 2014 (cit. on p. 61).
- [85] Michael Grant and Stephen Boyd, “Graph implementations for nonsmooth convex programs”, *Recent Advances in Learning and Control*, ed. by V. Blondel, S. Boyd and H. Kimura, Lecture Notes in Control and Information Sciences, Springer-Verlag Limited, 2008 95 (cit. on p. 61).
- [86] Gurobi Optimization, LLC, *Gurobi Optimizer Reference Manual*, 2018, URL: <http://www.gurobi.com> (cit. on p. 61).
- [87] Reina E. Itoh, Kazuo Kurokawa, Yusuke Ohba, Hisayoshi Yoshizaki, Naoki Mochizuki and Michiyuki Matsuda, *Activation of Rac and Cdc42 Video Imaged by Fluorescent Resonance Energy Transfer-Based Single-Molecule Probes in the Membrane of Living Cells*, *Molecular and Cellular Biology* **22** (2002) 6582, ISSN: 0270-7306, URL: <https://ayudaenaccion.org/ong/blog/solidaridad/historia-de-las-ong/%20https://mcb.asm.org/content/22/18/6582> (cit. on pp. 63, 66).
- [88] Kirsty J. Martin, Ewan J. McGhee, Juliana P. Schwarz, Martin Drysdale, Saskia M. Brachmann, Volker Stucke, Owen J. Sansom and Kurt I. Anderson, *Accepting from the best donor; analysis of long-lifetime donor fluorescent protein pairings to optimise dynamic FLIM-based FRET experiments*, *PLoS ONE* **13** (2018) 1, ISSN: 19326203 (cit. on p. 63).

- [89] Jonathan Klein, Martin Laurenzis, Dominik L. Michels and Matthias B. Hullin, *A Quantitative Platform for Non-Line-of-Sight Imaging Problems*, *British Machine Vision Conference 2018, BMVC 2018, Northumbria University, Newcastle, UK, September 3-6, 2018*, 2018 104 (cit. on p. 74).
- [90] Stanford Computational Imaging Lab, *Single-Photon 3D Imaging With Deep Sensor Fusion*, URL: <http://www.computationalimaging.org/publications/single-photon-3d-imaging-with-deep-sensor-fusion/> (visited on 15/05/2019) (cit. on pp. 81–83).
- [91] Corneliu Rablau, *LIDAR—A new (self-driving) vehicle for introducing optics to broader engineering and non-engineering audiences*, *Education and Training in Optics and Photonics*, Optical Society of America, 2019 11143_138 (cit. on p. 88).
- [92] Franco Zappa, Simone Tisa, Alberto Tosi and Sergio Cova, *Principles and features of single-photon avalanche diode arrays*, *Sensors and Actuators A: Physical* **140** (2007) 103 (cit. on p. 88).
- [93] Edoardo Charbon, *Towards large scale CMOS single-photon detector arrays for lab-on-chip applications*, *Journal of Physics D: Applied Physics* **41** (2008) 094010 (cit. on p. 88).
- [94] Yoann Altmann, Stephen McLaughlin, Miles J Padgett, Vivek K Goyal, Alfred O Hero and Daniele Faccio, *Quantum-inspired computational imaging*, *Science* **361** (2018) (cit. on p. 88).
- [95] Brent Schwarz, *LIDAR: Mapping the world in 3D*, *Nat. Photonics* **4** (2010) 1749 (cit. on p. 88).
- [96] David B. Lindell, Gordon Wetzstein and Matthew O’Toole, *Wave-based non-line-of-sight imaging using fast f - k migration*, *ACM Trans. Graph.* **38** (2019) 116 (cit. on p. 88).
- [97] Wenzheng Chen, Fangyin Wei, Kiriakos N Kutulakos, Szymon Rusinkiewicz and Felix Heide, *Learned feature embeddings for non-line-of-sight imaging and recognition*, *ACM Trans. Graph.* **39** (2020) 1 (cit. on p. 88).
- [98] Xiaochun Liu, Ibón Guillén, Marco La Manna, Ji Hyun Nam, Syed Azer Reza, Toan Huu Le, Adrian Jarabo, Diego Gutierrez and Andreas Velten, *Non-line-of-sight imaging using phasor-field virtual wave optics*, *Nature* (2019) 1 (cit. on pp. 88, 89, 91).

-
- [99] Felix Heide, Matthew O’Toole, Kai Zang, David B Lindell, Steven Diamond and Gordon Wetzstein,
Non-line-of-sight imaging with partial occluders and surface normals,
ACM Trans. Graph. **38** (2019) 1 (cit. on pp. 88, 91).
- [100] Atul Ingle, Andreas Velten and Mohit Gupta,
High Flux Passive Imaging With Single-Photon Sensors, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019
(cit. on pp. 88, 116).
- [101] STMicroelectronics, *STMicroelectronics Ships 1 Billionth Time-of-Flight Module*,
Press release, 2019,
URL: <https://www.st.com/content/st.com/en/about/media-center/press-item.html/t4210.html> (cit. on p. 88).
- [102] Sabbir Rangwala, *The iPhone 12 – LiDAR At Your Fingertips*, Forbes (2020),
URL: <https://www.forbes.com/sites/sabbirrangwala/2020/11/12/the-iphone-12lidar-at-your-fingertips/> (cit. on p. 88).
- [103] Junko Yoshida, *ST & Apple, through Thick and Thin*, EETimes (2018),
URL: <https://www.eetimes.com/st-apple-through-thick-and-thin/>
(cit. on p. 88).
- [104] David Patrick Baxter, *Application using a single photon avalanche diode (SPAD)*,
US Patent 9,058,081, 2015 (cit. on p. 88).
- [105] Samuel Burri, Harald Homulle, Claudio Bruschini and Edoardo Charbon,
LinoSPAD: a time-resolved 256x1 CMOS SPAD line sensor system featuring 64 FPGA-based TDC channels running at up to 8.5 giga-events per second,
Proc. SPIE **9899** (2016) 98990D (cit. on p. 90).
- [106] Desmond O’Connor, *Time-correlated single photon counting*,
Academic Press, 2012 (cit. on p. 90).
- [107] Adrian Jarabo, Belen Masia, Julio Marco and Diego Gutierrez,
Recent advances in transient imaging: A computer graphics and vision perspective,
Visual Informatics **1** (2017) 65 (cit. on p. 91).
- [108] Matthew O’Toole, Felix Heide, Lei Xiao, Matthias B. Hullin, Wolfgang Heidrich and Kiriakos N. Kutulakos,
Temporal Frequency Probing for 5D Transient Analysis of Global Light Transport,
ACM Trans. Graph. **33** (2014) (cit. on p. 91).
- [109] Christoph Peters, Jonathan Klein, Matthias B. Hullin and Reinhard Klein,
Solving Trigonometric Moment Problems for Fast Transient Imaging,
ACM Trans. Graph. **34** (2015) (cit. on p. 91).

- [110] Nikolai Smolyanskiy, Alexey Kamenev and Stan Birchfield, *On the importance of stereo for accurate depth estimation: An efficient semi-supervised deep neural network approach*, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2018 1007 (cit. on p. 91).
- [111] Andreas Meuleman, Seung-Hwan Baek, Felix Heide and Min H Kim, *Single-Shot Monocular RGB-D Imaging Using Uneven Double Refraction*, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020 2465 (cit. on p. 91).
- [112] Nikhil Naik, Shuang Zhao, Andreas Velten, Ramesh Raskar and Kavita Bala, *Single View Reflectance Capture Using Multiplexed Scattering and Time-of-flight Imaging*, *ACM Trans. Graph.* **30** (2011) 171:1 (cit. on p. 91).
- [113] Shuochen Su, Felix Heide, Robin Swanson, Jonathan Klein, Clara Callenberg, Matthias Hullin and Wolfgang Heidrich, *Material Classification Using Raw Time-Of-Flight Measurements*, *Proc. IEEE CVPR*, 2016 (cit. on pp. 91, 92, 111).
- [114] Vytautas Zickus, Ming-Lo Wu, Kazuhiro Morimoto, Valentin Kapitanov, Areeba Fatima, Alex Turpin, Robert Insall, Jamie Whitelaw, Laura Machesky, Claudio Bruschini et al., *Fluorescence lifetime imaging with a megapixel SPAD camera and neural network lifetime estimation*, *Scientific Reports* **10** (2020) 1 (cit. on p. 91).
- [115] Katherine L Bouman, Vickie Ye, Adam B Yedidia, Frédo Durand, Gregory W Wornell, Antonio Torralba and William T Freeman, *Turning corners into cameras: Principles and methods*, *IEEE International Conference on Computer Vision (ICCV)*, 2017 2289 (cit. on p. 91).
- [116] Victor Arellano, Diego Gutierrez and Adrian Jarabo, *Fast back-projection for non-line of sight reconstruction*, *Opt. Express* **25** (2017) 11574 (cit. on p. 91).
- [117] Matthew O’Toole, David B. Lindell and Gordon Wetzstein, *Confocal Non-line-of-sight imaging based on the light cone transform*, *Nature* (555 2018) 338 (cit. on p. 91).
- [118] Chia-Yin Tsai, Aswin C Sankaranarayanan and Ioannis Gkioulekas, *Beyond Volumetric Albedo—A Surface Optimization Framework for Non-Line-Of-Sight Imaging*, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019 1545 (cit. on p. 91).

-
- [119] Julian Iseringhausen and Matthias Hullin,
Non-line-of-sight reconstruction using efficient transient rendering,
ACM Trans. Graph. **39** (2020) 1 (cit. on p. 91).
- [120] Javier Grau Chopite, Matthias B. Hullin, Michael Wand and
Julian Iseringhausen, *Deep Non-Line-of-Sight Reconstruction*,
IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020
(cit. on p. 91).
- [121] Jonathan Klein, Christoph Peters, Jaime Martín, Martin Laurenzis and
Matthias B Hullin,
Tracking objects outside the line of sight using 2D intensity images,
Scientific reports **6** (2016) 32491 (cit. on p. 91).
- [122] Wenzheng Chen, Simon Daneau, Fahim Mannan and Felix Heide,
Steady-state non-line-of-sight imaging,
Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,
2019 6790 (cit. on p. 91).
- [123] N. Scheiner, F. Kraus, F. Wei, B. Phan, F. Mannan, N. Appenrodt, W. Ritter,
J. Dickmann, K. Dietmayer, B. Sick et al., *Seeing Around Street Corners:
Non-Line-of-Sight Detection and Tracking In-the-Wild Using Doppler Radar*,
*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
Recognition*, 2020 2068 (cit. on p. 92).
- [124] Kenichiro Tanaka, Yasuhiro Mukaigawa, Takuya Funatomi, Hiroyuki Kubo,
Yasuyuki Matsushita and Yasushi Yagi, *Material classification using
frequency-and depth-dependent time-of-flight distortion*,
Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,
2017 79 (cit. on p. 92).
- [125] Barbara Caputo, Eric Hayman and P Mallikarjuna,
Class-specific material categorisation,
Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1,
vol. 2, IEEE, 2005 1597 (cit. on p. 92).
- [126] Ce Liu, Lavanya Sharan, Edward H Adelson and Ruth Rosenholtz,
Exploring features in a bayesian framework for material recognition,
2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition,
IEEE, 2010 239 (cit. on p. 92).
- [127] Manik Varma and Andrew Zisserman,
A statistical approach to material classification using image patch exemplars,
IEEE transactions on pattern analysis and machine intelligence **31** (2008) 2032
(cit. on p. 92).

- [128] Richard H Byrd, Peihuang Lu, Jorge Nocedal and Ciyou Zhu, *A limited memory algorithm for bound constrained optimization*, *SIAM Journal on scientific computing* **16** (1995) 1190 (cit. on p. 102).
- [129] Dilip Krishnan and Rob Fergus, “Fast Image Deconvolution using Hyper-Laplacian Priors”, *Advances in Neural Information Processing Systems 22*, ed. by Y. Bengio, D. Schuurmans, J. D. Lafferty, C. K. I. Williams and A. Culotta, Curran Associates, Inc., 2009 1033 (cit. on p. 105).
- [130] Felix Heide, Lei Xiao, Wolfgang Heidrich and Matthias B Hullin, *Diffuse mirrors: 3D reconstruction from diffuse indirect illumination using inexpensive time-of-flight sensors*, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014 3222 (cit. on p. 111).
- [131] Quercus Hernandez, Diego Gutierrez and Adrian Jarabo, *A Computational Model of a Single-Photon Avalanche Diode Sensor for Transient Imaging*, 2017, arXiv: 1703.02635 [physics.ins-det] (cit. on p. 113).
- [132] sensing-aiot, *Apple LIDAR Demystified: SPAD, VCSEL, and Fusion...* 2021, URL: <https://web.archive.org/web/20220107210846/http://4da.tech/?p=582> (cit. on p. 115).
- [133] Kazuhiro Morimoto, Andrei Ardelean, Ming-Lo Wu, Arin Can Ulku, Ivan Michel Antolovic, Claudio Bruschini and Edoardo Charbon, *Megapixel time-gated SPAD image sensor for 2D and 3D imaging applications*, *Optica* **7** (2020) 346, URL: <http://www.osapublishing.org/optica/abstract.cfm?URI=optica-7-4-346> (cit. on p. 116).
- [134] Canon Global, *Canon develops SPAD sensor with world-highest 3.2-megapixel count, innovates with low-light imaging camera that realizes high color reproduction even in dark environments*, 2022, URL: <https://web.archive.org/web/20220107210846/https://global.canon/en/news/2021/20211215.html> (cit. on p. 116).
- [135] T. Al Abbas, N. A. W. Dutton, O. Almer, S. Pellegrini, Y. Henrion and R. K. Henderson, *Backside illuminated SPAD image sensor with 7.83 μ m pitch in 3D-stacked CMOS technology*, *2016 IEEE International Electron Devices Meeting (IEDM)*, 2016 8.1.1 (cit. on p. 116).
- [136] Oichi Kumagai et al., *7.3 A 189 \times 600 Back-Illuminated Stacked SPAD Direct Time-of-Flight Depth Sensor for Automotive LiDAR Systems*, *2021 IEEE International Solid- State Circuits Conference (ISSCC)*, vol. 64, 2021 110 (cit. on p. 116).

-
- [137] Navid Sarhangnejad et al., *Dual-Tap Computational Photography Image Sensor With Per-Pixel Pipelined Digital Memory for Intra-Frame Coded Multi-Exposure*, *IEEE Journal of Solid-State Circuits* 54 (2019) 3191, URL: <https://doi.org/10.1109/jssc.2019.2932623> (cit. on p. 116).
- [138] Yi Luo, Jacky Jiang, Mengye Cai and Shahriar Mirabbasi, *CMOS computational camera with a two-tap coded exposure image sensor for single-shot spatial-temporal compressive sensing*, *Optics Express* 27 (2019) 31475, URL: <https://doi.org/10.1364/oe.27.031475> (cit. on p. 116).
- [139] Mian Wei, Navid Sarhangnejad, Zhengfan Xia, Nikita Gusev, Nikola Katic, Roman Genov and Kiriakos N Kutulakos, *Coded two-bucket cameras for computer vision*, *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018 54 (cit. on p. 116).
- [140] Yuhao Liu, Felipe Gutierrez-Barragan, Atul Ingle, Mohit Gupta and Andreas Velten, *Single-Photon Camera Guided Extreme Dynamic Range Imaging*, *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2022 1575 (cit. on p. 116).
- [141] Zhanghao Sun, David B. Lindell, Olav Solgaard and Gordon Wetzstein, *SPADnet: deep RGB-SPAD sensor fusion assisted by monocular depth estimation*, *Opt. Express* 28 (2020) 14948, URL: <http://www.osapublishing.org/oe/abstract.cfm?URI=oe-28-10-14948> (cit. on p. 117).
- [142] H. Ho, E. Saeedi, S. S. Kim, T. T. Shen and B. A. Parviz, *Contact lens with integrated inorganic semiconductor devices*, *2008 IEEE 21st International Conference on Micro Electro Mechanical Systems*, 2008 403 (cit. on p. 117).

List of Figures

1.1	Overview of the technologies used in the three main chapters	3
2.1	Reflection and refraction of light at the surface between two materials with different refractive indices.	14
2.2	Illustration of the angle γ_i between the electric field vector and the plane of incidence.	15
2.3	Band model for insulators, semiconductors and conductors.	19
2.4	Doping of a semiconductor by introducing impurity atoms into the lattice.	19
2.5	Schematic of a p-n junction.	20
2.6	Schematic sketch of the structure and photo-electron collection mechanism of a PMD pixel.	24
2.7	Measurement principle of a correlation ToF system.	25
2.8	Example of a light response distribution as it could be measured with a SPAD setup.	27
3.1	Snapshot Difference Imaging – Teaser Figure	30
3.2	Principle of operation of a time-of-flight (ToF) pixel.	32
3.3	Components of our difference imaging system based on the Texas Instruments OPT8241-CDK module.	37
3.4	Polarization difference imaging principle.	38
3.5	Polarization difference image results for two different scenes.	39
3.6	Difference image of a color calibration chart.	41
3.7	Depth edge difference images with different alignments and distances of the light sources.	42
3.8	Directional difference image acquired with two light sources in wide vertical spacing.	43
3.9	Spatial gradient imaging principle.	44
3.10	Spatial gradient image of an aluminum resolution chart.	45
3.11	Example filters (modulation patterns) for use with our system.	45
3.12	Temporal gradient and RGB image of a rotating fan.	46

3.13	Histograms of the absolute pixel variance in the analog and digital difference images of the scene depicted in the bottom row of Fig. 3.5.	47
3.14	Variance of each pixel, averaged over the whole difference image, for the PMD sensor and the TI sensor.	47
3.15	Pixel variance in difference images of a color chart recorded with a TI ToF camera and a Point Grey Flea3 camera.	49
3.16	Difference images without and with ambient illumination, acquired with a PMD ToF camera and a conventional PointGrey Flea3 camera with larger and smaller aperture setting.	50
3.17	Reconstruction of two color channels from a single exposure by exploiting photon statistics.	52
3.18	Source images of the polarization difference image shown in Fig. 3.5a (bottom), reconstructed from the statistics of 400 difference images.	53
3.19	Modified ribbon cable with (top to bottom) connectors for sensor board, LS1 and LS2.	55
4.1	Schematic sketch of the imaging set-up for the LIDAR experiment and effect of the optical blur.	59
4.2	CCD image, SPAD measurement, reconstructed light-in-flight image and depth images extracted from the SPAD measurement and the reconstruction for different scenes.	62
4.3	Spatial upsampling of the fluorescence lifetime of cancer cells.	64
4.4	Matrix B	71
4.5	Mask of the active pixel areas on the whole sensor area.	72
4.6	Matrix P	72
4.7	Matrix A	73
4.8	Matrix A_T	74
4.9	High resolution ground truth simulated light-in-flight image and simulated SPAD sensor measurements in and out of focus.	75
4.10	Simulated CCD image of the artificial three-dimensional scene.	76
4.11	Reconstructed simulated scene in different spatial resolutions without and with added noise.	77
4.12	Simulated scene featuring a diffuse table and walls and temporal light intensity profile for one pixel.	79
4.13	Five frames from the light-in-flight image of the scene shown on the left side of Fig. 4.12 at different points in time illustrating multiple light bounces in the scene.	80
4.14	CCD images and SPAD images from Lindell et al., reconstructions using our method.	82

4.15	Detail of the first scene shown in Fig. 4.14 from a slightly different angle.	83
4.16	Depth maps of details from a scene simulated with low signal-to-noise ratio.	84
5.1	Low-Cost SPAD Sensing teaser figure.	88
5.2	Setup schematic for depth and NLOS imaging; Lenses and galvanometer scanner pair in front of the sensor and light source in the lab.	93
5.3	The VL53L1X time-of-flight sensor module by STMicroelectronics on a commercially available breakout board; Collimated laser beam profile; Point spread function of the full system.	93
5.4	Material measurement principle.	95
5.5	Measurements for five different materials.	96
5.6	Two-dimensional PCA of the material data with and without ambient light and first two eigenvectors.	97
5.7	Confusion matrices for the classification of five material classes on a test dataset.	98
5.8	Two variants of NLOS tracking setting	100
5.9	Histograms for all four corners measured with the setting without mirrors for one target position.	100
5.10	Exemplary reconstructions for both target sizes.	104
5.11	Distances in all spatial dimensions between the true and predicted target position in the test dataset for the 'direct position prediction' method.	105
5.12	Measurements taken with our scanning setup of three different real scenes.	107
5.13	Calibration of the relation between bin number and true distance for both methods used to determine the position of the depth peak.	108
5.14	Exemplary histograms for a bright and a dark scene pixel for different ambient light scenarios.	108
5.15	Measurements of three targets in different depths, each with five patches of different reflectivity at the illumination wavelength, for different levels / sources of ambient light.	110

List of Tables

4.1	Parameters used in the reconstruction of all images in the main text and Supplementary Information.	67
4.2	Reconstruction runtimes for data sets depicted in Fig. 4.2 of the main paper, as well as the simulated data.	81
5.1	Classification accuracy on the test dataset with and without ambient light, as well as with color variants of each material.	99
5.2	RMSE (root mean square error) of the reconstructed target position in all spatial dimensions for all methods and target configurations in meters.	103

ACKNOWLEDGEMENTS

Completing this thesis would not have been possible without the help and support of many people that I shared parts or all of this ‘PhD journey’ with. I’m grateful to all of them and would like to especially thank...

MY SUPERVISORS...

MATTHIAS HULLIN for introducing me to the world of computational imaging, sharing his ideas, visions, and enthusiasm about this research area. For his extensive help, support, and encouragement throughout the whole time, and especially during stressful paper deadline periods. For always being available for discussions, often even at quite unreasonable times. For giving me the chance to visit and present my work at several international conferences and workshops and to collaborate with inspiring people all over the world.

REINHARD KLEIN for co-supervising my PhD studies and co-examining my thesis, and for helpful discussions and meetings.

MY CO-AUTHORS – ALL OF THEM, BUT ESPECIALLY...

ASHLEY LYONS AND DANIELE FACCIO for a great collaboration on the super-resolution project, for always being supportive and encouraging and helping a lot with the paper and publication process. Also for giving me the opportunity to present my and my colleagues’ work at a very interesting and enjoyable workshop in Como.

FELIX HEIDE for collaborating on two of my projects, always sharing great ideas and discussions, and giving valuable advice and support.

DENNIS DEN BROK for all the things I have learned from him, his ideas and inspirations, his extensive and continuous support and help. For regularly making

me see things in new ways and discover new perspectives, for always being open to ideas and discussions. For his jokes and all the fun and good times.

MY COLLEAGUES...

SEBASTIAN THIEL, JAVIER GRAU CHOPITE, AND JONATHAN KLEIN for being great office mates that were always open to discussing research problems and providing new ideas when I got stuck. For shared fun, frustration, joy, hard work, silliness, and achievements.

RALF SARLETTE for always being helpful with absolutely everything, for finding perfect solutions to any problem using his technical expertise and sometimes probably magic.

ALL THE OTHER CURRENT AND FORMER COLLEAGUES who have made my time working on my thesis projects a pleasure by sharing lunch times, coffee breaks, corridor chats, Band Hero nights, barbecues, pub quizzes, and more.

MY FRIENDS AND FAMILY...

BIRGIT STAPF for always listening and giving the wisest of advice, for all the emotional support, helpful discussions, and feedback.

JONAS MÜLLER for his endless support and lots of help. For enduring me when I was stressed out and making sure I wasn't drowning in chaos before deadlines. For always being there and encouraging and motivating me when I needed it.

MY PARENTS for their unconditional support, for always having an open ear, and for believing in me at all times.