Dipl.-Ing. Hannes Plank, BSc

Location-aware Optical Communication with Time-of-Flight Sensors

DISSERTATION

zur Erlangung des akademischen Grades

Doktor der technischen Wissenschaften

eingereicht an der

Technischen Universität Graz

Betreuer

Univ.-Prof. Dipl.-Inform. Dr.sc.ETH Kay Uwe Römer

Institut für Technische Informatik

Graz, Februar 2019

EIDESSTATTLICHE ERKLÄRUNG

Ich erkläre an Eides statt, dass ich die vorliegende Arbeit selbstständig verfasst, andere als die angegebenen Quellen/Hilfsmittel nicht benutzt, und die den benutzten Quellen wörtlich und inhaltlich entnommenen Stellen als solche kenntlich gemacht habe. Das in TUGRAZonline hochgeladene Textdokument ist mit der vorliegenden Dissertation identisch.

Datum

Unterschrift

Kurzfassung

Diese Disseration präsentiert die grundlegende Forschungsarbeit um das erste optische Kommunikationsystem mit Time-of-Flight Sensoren zu ermöglichen, welches die präzise 3D Lokalisierung der Kommunikationpartner ermöglicht. Elektronische Geräte interagieren und kooperieren zusehens miteinander und eine Lokalisierung des Kommunikantionspartners kann diese Interaktion und Kooperation signifikant verbessern. Diese Lokalisierung ist bedeutend für technische Gebiete wie Robotik, sichere Kommunikation, Augmented und Virtual Reality und automatisierte Autos. Existierende Innenraumlokalisierungsmethoden sind oft komplex und benötigen zusätzliche Komponenten und eigene Infrastrukturen. Diese Arbeit zeigt, dass Time-of-Flight Sensoren den Ansprüchen dieser Anwendungen gerecht werden können, indem 3D Lokalisierung mit optischer Kommunikation verbunden wird. Time-of-Flight Sensoren erzeugen Tiefenbilder, indem die Phasenverschiebung zwischen ausgestrahltem und reflektiertem Licht gemessen wird. Dieses Messprinzip wird in dieser Arbeit dazu wiederverwendet um Daten mittels Phasenumtastung zu verschicken und zu empfangen.

Am Anfang dieser Disseration werden die potentiellen Anwendungen und Anforderungen an ein location-aware optisches Kommunikationssystem untersucht. Die identifizierten Probleme und Herausforderungen werden dann mit dem Design von zwei Systemen und einigen Datenverarbeitungsmethoden behandelt. Die daraus resultierenden Implementierungen ermöglichen die Charakterisierung von den entwickelten Methoden und zeigen die Durchführbarkeit. Die Neuheit des verfolgtem Ansatz wird durch die Entwicklung eines Demonstrators bewiesen, welcher den Anwendungsfall von 3D Tags für Augmented Reality Systeme demonstriert.

Die Ergebnisse dieser Disseration zeigen, dass Time-of-Flight Sensoren gut geeignet sind um das vorgeschlagene location-aware Kommunikationsparadigma zu realisiern. Die vorgestellten Systeme und Methoden lösen die grundlegenden Probleme und Herausforderungen und ermöglichen eine weite künftige Verbreitung.

Abstract

This doctoral thesis presents the fundamental research to establish the first location-aware optical communication system with Time-of-Flight sensors. The ever increasing interaction and cooperation between electronic devices can vastly benefit from communication partner localization. Determining the relative position of a communication partner is important for fields like robotics, secured communication, augmented and virtual reality and automated driving. Existing indoor localization solutions are often complex and require dedicated infrastructures. This work shows that using Time-of-Flight sensors can fulfill the demanding requirements by combining precise 3D localization with optical communication. Time-of-Flight sensors capture depth images by measuring the phase difference between emitted and reflected light. This depth sensing mechanism is repurposed in this work to transmit and receive data via phase-shift-keying.

The thesis first investigates the potential applications with a focus on the requirements on a location-aware communication system. The identified problems and challenges are then addressed by two system designs and several proposed processing methods. The resulting implementations enable the characterization of the proposed methods and proof the feasibility. The novelty of Time-of-Flight based location-aware communication is demonstrated by implementing and evaluating the promising use case of 3D tags for augmented reality systems.

The results of this doctoral thesis show that Time-of-Flight sensors are well suited to implement the proposed location-aware communication paradigm. The proposed system design and methods solve the fundamental issues and challenges, enabling location-aware communication on a wide range of systems.

Acknowledgements

First and foremost, I like to thank Prof. Kay Uwe Römer for supervising this thesis. I also want to thank Christian Steger for his advice and guidance to conduct this doctoral thesis. Special thanks go to Norbert Druml, who initiated this project and competently guided me throughout this thesis. I am grateful for my line manager Gerald Holweg who handled the organizational aspects of this work. I am also thankful to the core team of the Time-of-Flight sensor at Infineon Technologies Austria AG, especially Markus Dielacher, Robert Lobnik and Martin Flatscher. Their support enabled me to effectively implement and evaluate my proposed methods. Valuable advice and support was also provided by Thomas Herndl. This work was supported by two excellent master theses by Armin Schönlieb and Theresa Egger. Matthias Almer and Stefan Kamper also made valuable contributions with their bachelor theses. I would also like to thank my co-workers for their support, the great work environment and the productive atmosphere.

I want to thank the Austrian Federal Ministry for Transport, Innovation and Technology as well as the Austrian Research Promotion Agency (FFG) for funding the research activities under the grant agreement number 852328.

Finally, I want to express my deepest gratitude towards my friends and family, who supported me throughout my life.

Graz, Februar 2019 Hannes Plank

Extended Abstract

In recent years, spatial computing applications such as augmented and virtual reality, robotics, indoor navigation and automated driving increased the demand for precise localization systems. Since these applications also require interaction between localized devices, it is advantageous to combine communication and localization into a single system. Conventional wireless communication methods are however unable to fulfill the demanding localization requirements. This doctoral thesis fills this gap by performing the fundamental research on location-aware optical communication with Time-of-Flight (ToF) sensors. Combining Time-of-Flight depth sensing with optical communication enables to precisely localize communication partners at high update rates. Unlike alternative localization methods, the presented approach does not require any infrastructure and integrates localization and communication on a single miniaturized device. Depending on the application, the proposed methods enable data exchange and localization between ToF cameras, or between ToF cameras and multiple miniaturized optical transceivers.
Figure 1 illustrates the structure of the conducted project. The absence of directly related work caused this doctoral thesis to take a pioneering approach to enable the first ToF-based location-aware communication system. Potential applications are therefore first investigated with regards to their requirements. An analysis and characterization of the ToF sensor allows the identification of the fundamental problems which need to be solved to fulfill these requirements. Several system designs and methods are proposed to fulfill the identified requirements, while keeping the limitations and challenges of the ToF system in mind. An experimental evaluation produces results which prove the feasibility of the approach.

The unique combination of line-of-sight communication, precise localization, fast update rate and low-cost implementation enables novel applications. Secured communication between two ToF cameras allows to exchange the mutual distance measurements[1]. Mutually verifying the consistency of these measurements has the potential to effectively mitigate relay attacks[2]. This avoids distance-bounding protocols, and enables secured and authenticated wireless connections over several meters. Resilience against relay attacks can also enable a new generation of vehicle-to-vehicle communication systems. Time-critical traffic warnings could be relayed between subsequent cars[3]. Indoor localization

[1] *Plank et al., Optisec3D - A new Paradigm in Secure Communication and Authentication featuring Time-of-Flight*, Proceedings of the 2016 International Conference on Embedded Wireless Systems and Networks, Graz, Austria, 2016

[2] *Plank et al., Survey on Camera based Communication for Location-aware Secure Authentication and Communication*, EMC2 Summit at CPS Week, Vienna, Austria, 2016

[3] *Plank et al., Time-of-Flight based Optical Communication for Safety-Critical Applications in Autonomous Driving*, DECSoS workshop at Computer Safety, Reliability, and Security (SAFECOMP), Springer Publishing, Trondheim, Norway, 2016

and 6-degrees-of-freedom (6-dof) motion tracking can be unified into a single system with a ToF-based beaconing system. Sensing the beacon positions in 3D reduces the required amount of beacons compared to 2D camera systems. The pose of the ToF sensor can be determined by aligning the beacon positions. If the position and orientation of a ToF sensor are derived by an alternative indoor localization system, it is possible to localize embedded devices such as sensor-network nodes[4]. A promising application is a 3D tag system for augmented reality (AR) systems. Any electronic device can be equipped with a ToF tag, which enables the device to be identified and localized by a ToF sensor. If ToF sensors are integrated into AR devices, the localized position of the ToF tag can be mapped into the view of the user. This enables to overlay information and user interfaces directly over the physical location of any electronic device. Each of the introduced applications has a certain set of requirements. This thesis develops a system design and several processing methods to fulfill the combined requirements of the potential applications. Future ToF-based location-aware communication systems can either extend the system design to combine all proposed methods into a versatile system or tailor subsets to specific applications.

Time-of-Flight sensors work by emitting pulsed infrared light with an active illumination unit. This light signal is reflected by the scene and projected by the lens onto the pixel

[4] *Plank et al., Localization and Context Determination for Cyber-physical Systems based on 3D Imaging*, Solutions for Cyber-Physical Systems Ubiquity, IGI Global, 2017

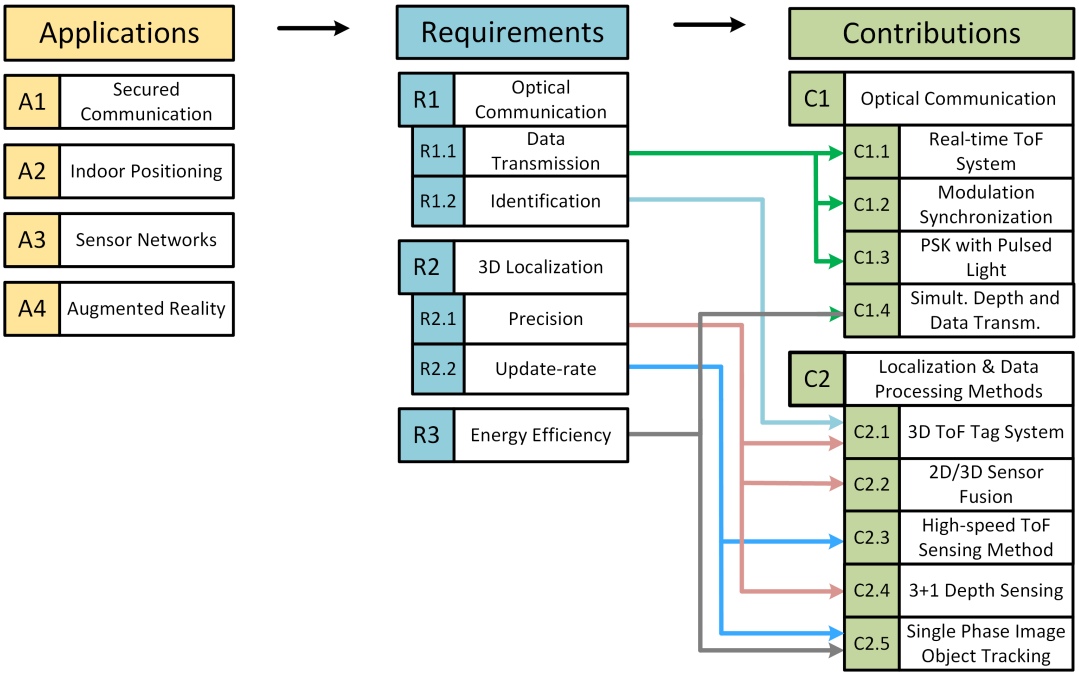


Figure 1: As a first step, potential applications of location-aware optical communication with Time-of-Flight sensors are investigated. The identified requirements are then addressed by the contributions of this thesis, which include system design concepts and processing methods. An experimental evaluation shows that the proposed methods enable ToF-based location-aware communication systems to fulfill the requirements.

array. Continuous-wave ToF sensors continuously emit the lightpulses during the exposure time. A photonic mixer device (PMD) on each pixel correlates the received light with a pixel reference signal. This correlation value depends on the phase shift and the amount of reflected light. Capturing multiple images with different equidistant phase offsets between the emitted and reflected light enables the removal of the influence of the light intensity and the calculation of the distance. This distance-sensing principle can be extended to also transmit and receive data. The active illumination unit is able to transmit data while the pixelarray can be used as receiver. An evaluation approach is established to investigate and characterize this communication principle. A ToF sensor is integrated into a real-time FPGA system, which enables to control the sensor behavior with specifiable time limits. The developed modulation method uses groups of phase-shifted lightpulses to transmit and receive data with phase-shift keying modulation. Unlike conventional image sensors, it is possible to transmit multiple bits per symbol at constant signal strength. This modulation method however requires synchronized modulation signals. The introduced synchronization method samples the aliasing signal and iteratively adapts the modulation frequency of the Time-of-Flight sensor[5]. The results show that the synchronization method can achieve a 100% success rate over a distance of up to 9 m. The data transmission rate depends on the distance and reaches up to 6 kbit/s with a range of 4.5 m.

The ToF sensor needs to alternate between optical communication and capturing depth images to localize the communication partners. A location-aware optical communication system can adapt the proposed system design to subsequently execute such different sensor operations. The sensor operations combine sensor parameter adjustments and processing steps that include localization, synchronization, receiving and transmitting data. The implemented operations can be used as building blocks for future ToF-based location-aware communication protocols.

The sensor operations depend on real-time execution, which is not available on all hardware platforms. An alternative system design is therefore presented which does not require a real-time system to control the ToF sensor. Real-time sensor configuration changes can be avoided by reducing the optical communication functionality to suit specific applications. The simplified design enables to use unmodified Time-of-Flight sensors for the promising application of 3D ToF tags. These tags enable to equip any electronic device with a localization and communication interface for ToF cameras. During operation, the Time-of-Flight sensor continuously captures a distinct pattern of depth images. This involves a 30 Hz stream of short-range depth images with an interleaved 5 Hz stream of long-range ToF images. The ToF tags can detect this pattern with a photo diode and a receiver circuit. This enables to synchronize their data transmission events to the ToF camera. The tags can then reliably transmit data via on/off keying, when the ToF sensor is sensitive to incoming light. When the ToF sensor captures long-range ToF images, the tags remain silent. This enables the ToF sensor to localize the tags without signal distortions. The alternative system is then extended to implement the promising use case of 3D tags for augmented reality smartphones. The basic functionality of a ToF tag is to present relevant information at its location. In the implemented use case, a temperature sensor streams its readings to a Time-of-Flight camera. The 3D tag position is mapped to the image stream of a color camera. The displayed temperature is overlaid at the

[5] *Plank et al., Synchronization of Time-of-Flight 3D Sensors for Optical Communication*, IEEE International Conference on Communications (ICC), Paris, France, 2017

mapped position. A ToF tag can also provide a user interface, enabling to directly control any electronic device with an AR device. The ToF tag system is also used to determine the localization accuracy of location-aware optical communication. The evaluated system achieves an average accuracy of 5 mm within a range of 3.5 m with an update-rate of 5 Hz. The evaluation shows that implementing this use case with alternative technologies is significantly more complex and often restricts the flexibility of the system.

The experimental evaluation of the proposed system designs proofs the functionality and feasibility of ToF-based location-aware optical communication. Potential applications might however additionally require a certain localization performance and energy efficiency. Several processing methods are therefore developed to tackle these requirements. Combining ToF depth data with color images can significantly improve the depth resolution while reducing noise. This increases the localization precision and helps to associate the correct depth pixels with the communication partner. A computationally efficient sensor fusion algorithm is developed, which fuses depth and color data on the graphics processing unit (GPU) of smartphones[6]. Relative motion between the ToF camera and its communication partners can cause motion artifacts. This effect is most severe for small objects and can impair the 3D localization. A method is therefore developed, which enables a ToF system to compensate motion artifacts for localization[7]. Low power consumption of the location-aware communication system is important in most applications. The active illumination unit of the ToF sensor consumes a significant energy portion of the system. Reducing the number of light pulses for communication partner localization can drastically reduce the required energy. A method is therefore proposed, which enables to reduce the number of phase images for localization. Instead of capturing four or eight different images for a single localization, the proposed method is able to determine the distance to a communication partner from a single phase image[8]. The method exploits the fact that the communication partners are continuously localized at varying distances. This allows to create a phase-depth lookup table during a brief tracking initialization phase. The evaluation show that the method can also be applied to general 3D object tracking, and reduces the power consumption by up to 68%.

The results of this doctoral thesis show that Time-of-Flight sensors are well suited for the proposed location-aware optical communication paradigm. The developed systems and processing methods provide solutions to identified problems and challenges. Future systems are therefore able to use our research and results to adapt the proposed methods to the specific needs of the intended applications. Location-aware optical communication can further be compatible to traditional depth sensing. This enables the ability to provide valuable additional location-aware communication functionality to a wide range of existing and future devices.

---

[6] *Plank et al., High performance Time-of-Flight and color sensor fusion with image-guided depth super resolution*, Design, Automation Test in Europe Conference Exhibition (DATE), Dresden, Germany, March 2016

[7] *Plank et al., High-performance Indoor Positioning and Pose Estimation with Time-of-Flight 3D Imaging*, International Conference on Indoor Positioning and Indoor Navigation (IPIN), location, September 2017

[8] *Plank et al., Energy-efficient Time-of-Flight distance sensing method for 3D object tracking*, International Conference on Digital Image Computing: Techniques and Applications (DICTA), Canberra, Australia, December 2018

Contents

1 Introduction **1**
- 1.1 Motivation . . . 1
- 1.2 Location-aware Optical Communication with Time-of-Flight Sensors . . . 4
  - 1.2.1 Introduction . . . 4
  - 1.2.2 Problem Statement . . . 6
  - 1.2.3 Goals and Research Questions . . . 6
  - 1.2.4 Contributions . . . 7
  - 1.2.5 Structure of this Thesis . . . 8

2 Related Work **9**
- 2.1 Optical Communication with Image Sensors . . . 9
- 2.2 Single-Anchor Localization . . . 11
- 2.3 Location-aware Communication Methods . . . 12
- 2.4 Comparison to the State of the Art . . . 15

3 Application Requirements and Sensor Characterization **17**
- 3.1 Applications and Requirements . . . 18
  - 3.1.1 A1: Secured Communication . . . 18
  - 3.1.2 A2: Indoor Positioning . . . 19
  - 3.1.3 A3: Sensor Networks . . . 19
  - 3.1.4 A4: Augmented-Reality . . . 20
- 3.2 Time-of-Flight Sensor Characterization . . . 21
  - 3.2.1 Distance Sensing Principle . . . 22
  - 3.2.2 Suitability for Optical Communication . . . 24
- 3.3 System Design Goals . . . 25
  - 3.3.1 R1: Optical Communication . . . 25
  - 3.3.2 R2: 3D Localization . . . 26
  - 3.3.3 R3: Energy Efficiency . . . 26

4 Methods and System Design **28**
- 4.1 Optical Communication with Time-of-Flight Sensors . . . 29
  - 4.1.1 Concept and Methods . . . 29
  - 4.1.2 System Design for ToF-based Optical Communication . . . 33
- 4.2 Localization and Time-of-Flight Data Processing . . . 39
  - 4.2.1 Concepts and Methods . . . 39
  - 4.2.2 Implementation of a Time-of-Flight Tag Localization System . . . 46

5 Results and Use Case Evaluation **50**
5.1 Optical Communication . . . 50
5.1.1 Evaluation Setup . . . 50
5.1.2 Modulation Frequency Synchronization . . . 51
5.1.3 Data Transmission Characteristics . . . 53
5.1.4 Discussion . . . 55
5.2 Localization . . . 55
5.2.1 Evaluation Setup . . . 56
5.2.2 3D Localization Error . . . 56
5.2.3 Discussion . . . 58
5.3 Processing Methods . . . 59
5.3.1 Time-of-Flight and Color Sensor Data Fusion . . . 59
5.3.2 High-speed Time-of-Flight Sensing . . . 60
5.3.3 3D Object Tracking with Single Phase Images . . . 61
5.4 Limitations . . . 62
5.5 Use Case Evaluation of Time-of-Flight 3D Tags . . . 63
5.5.1 ToF Tags for Augmented Reality . . . 63
5.5.2 Alternative Use Case Implementations . . . 65

6 Conclusion and Future Work **68**
6.1 Conclusion . . . 68
6.2 Directions for Future Work . . . 69
6.2.1 ToF-based Location-aware Communication Systems and Methods . . . 69
6.2.2 Applications . . . 70

7 Publications **72**
7.1 OptiSec3D - a new paradigm in secure communication and authentication featuring Time-of-Flight . . . 76
7.2 Survey on camera based communication for location-aware secure authentication and communication . . . 82
7.3 High performance time-of-flight and color sensor fusion with image-guided depth super resolution . . . 88
7.4 Time-of-Flight based optical Communication for Safety-Critical Applications in Autonomous Driving . . . 94
7.5 Localization and context determination for cyber-physical systems based on 3D imaging . . . 106
7.6 Synchronization of time-of-flight 3D sensors for optical communication . . . 128
7.7 High-performance indoor positioning and pose estimation with time-of-flight 3D imaging . . . 134
7.8 Fast and energy-efficient time-of-flight distance sensing method for 3D object tracking 142

References **150**

List of Figures

1 As a first step, potential applications of location-aware optical communication with Time-of-Flight sensors are investigated. The identified requirements are then addressed by the contributions of this thesis, which include system design concepts and processing methods. An experimental evaluation shows that the proposed methods enable ToF-based location-aware communication systems to fulfill the requirements. viii

1.1 The opportunities of location-aware optical communication with Time-of-Flight sensors. A) ToF sensors are able to use depth sensing to localize their communication partners. B) The localized communication partners can act as beacons for fast and precise 6-degrees-of-freedom localization. C) Two ToF sensors mutually localize themselves and exchange their measurements for verification. D) The 3D localization enables to display information at the position of the communication partners on augmented-reality devices. 2
1.2 Location-aware optical communication with ToF sensors can provide any electronic device with a direct connection to augmented-reality systems. Users are able to interact with these devices by means of an overlaid user interface. 3
1.3 The principle of distance sensing with Time-of-Flight sensors. Reproduced with changes from [1]. 5

3.1 Potential applications are evaluated to obtain the requirements, which are then addressed by several system designs and methods. 17
3.2 A relay attack on a secured optical ToF connection is infeasible, because the distance measurements d can be exchanged and mutually verified. Obtained with changes from [2]. 18
3.3 Location-aware optical communication can be combined with existing indoor positioning systems to localize sensor network nodes. 20
3.4 3D localization enables to determine the pixel position of the communication partner on a color camera stream. This enables to overlay information in augmented-reality devices. 21
3.5 The photonic mixer device is located at each pixel. Incoming photons generate charges in the silicon, which are redirected by the PMD into either capacitor A or B. Reproduced with changes from [1]. 22
3.6 The charge content of the pixel capacitors A and B during exposure. The ratio between A and B depends on the phase difference φ_0. Reproduced with changes from [1]. 22
3.7 The phase value of an object depends on the object reflectivity and distance. Reproduced with changes from [3]. 23
3.8 Calculating the distance requires four different samples of the autocorrelation function. Reproduced with changes from [1]. 24

4.1 The contributions of this thesis enables future location-aware optical communication systems to meet their localization and communication requirements. 28
4.2 The algorithm for modulation frequency synchronization. Reproduced with changes from [1]. 29
4.3 Example of the modulation signal synchronization procedure. The initial aliasing frequency is estimated in the first step. The second step reveals that the arbitrarily assigned sign of the estimated frequency F_{est} was incorrect, and therefore a PLL adjustment of two times the negative F_{est} leads to synchronization. Reproduced with changes from [1]. 30
4.4 Taking two subsequent samples with 90°phase offset enables to measure the transmitted phase φ_1. 32
4.5 It is possible to transmit PSK modulated data during depth sensing. Pulse groups in the pixel reference signal and the illumination signal are simultaneously phase shifted. 33
4.6 The optical communication system is organized in a layer architecture and implemented on an FPGA with integrated CPU. 34
4.7 The ToF-FPGA hardware system consists of a ToF sensor, which is connected to a Zynq 7000 FPGA via I2C and a parallel data interface. Reproduced with changes from [4]. 35
4.8 UML class diagram of the ToF configuration layer. 36
4.9 ToF operations are the building blocks of a location-aware communication protocol. Each operation captures a series of phase images, followed by a processing step and optional sensor configuration adaption. 37
4.10 The core of the sensor fusion algorithm. Sparsely mapped depth values distribute an influence value within a certain area. Edges in the color image exponentially decrease the influence. 40
4.11 Motion artifacts are most severe when sensing small object such as markers. Motion causes the object to be projected to a different set of pixels in the subsequent phase images t_1 and t_2. 41
4.12 Pixel binning reduces the noise of a depth image at the cost of reduced resolution. By capturing a high-resolution phase image, it is possible to use the edge information as guidance to produce an improved high-resolution depth image. 42
4.13 If phase values of an object are associated with the distance during a training phase, it is possible to estimate the distance to the object with a single phase image. . . . 43
4.14 A) At least three light sources are mounted in close proximity in the proposed 6-dof ToF tag. B) A barrier between the light source can increase the directional light emission characteristics. 44
4.15 The geometic relation between a ToF sensor and 6-dof ToF tag. At least 3 points need to be determined in each coordinate system to obtain the rotation matrix R. 45
4.16 The architecture of the ToF tag system is organized in a pipeline architecture. . . 47
4.17 A ToF and color camera are rigidly mounted and facing the same direction. The 3D position of the ToF tag P_{ToF} can be transformed and projected into the stream of the color camera. 48
4.18 A ToF sensor captures long and short-range ToF images. The ToF tag is time-synchronized with the ToF sensor and responds when the ToF sensor captures short-range depth images 49

5.1 An evaluation transmitter enables to evaluate the proposed synchronization and modulation method with different parameters. 51
5.2 Evaluation of the modulation frequency synchronization method with different initial frequency differences. Reproduced with changes from [1]. 52

5.3 A) Evaluation of the modulation frequency synchronization method at different distances. B) The effect of the increased exposure time on the sampling rate. Reproduced with changes from [1]. 53
5.4 The signal strength of the received signal at different distances and exposure times. 54
5.5 The received signal strength of a 6.5 MHz signal at different pixel reference signal frequencies. This demonstrates the signal suppression effect of the photonic mixer device. 54
5.6 Three tags are mounted in a equilateral triangle configuration to evaluate the localization accuracy at different orientations and distances. 56
5.7 The average localization error at different distances. 57
5.8 The CDF plot of the localization error between 1 and 4.5 m (A) and between 1 and 3.5 m (B). 58
5.9 The additional localization error caused by the tag orientation, measured at 2 m distance. 58
5.10 The error distribution of the proposed motion artifact resilient distance estimation method. Reproduced with changes from [5]. 61
5.11 Results of the proposed energy efficient 3D object tracking method. A) The introduced distance estimation error is below 1 cm. B) The number of training measurements influences the distance estimation error. Reproduced with changes from [3]. 61
5.12 Location-aware optical communication with ToF sensors enables the use case of 3D tags for augmented reality smartphones. 63
5.13 The implemented use case supports two different ToF tag types. A) Tags are able to transmit location-specific information. B) An electronic device can use a ToF tag to present a user interface. 64

7.1 The publications of this thesis cover the investigation of potential applications and the development of system designs and methods for location-aware optical communication with Time-of-Flight sensors. 72

List of Tables

2.1 Comparison of related localization and location-aware communication approaches. 15

3.1 The minimum requirements for each application A1-A4 lead to the system design goals. These goals are met by the contribution C1 and C2. 25

5.1 The bit error ratio at different distance and data transfer rates. 53
5.2 Distance and tag orientation requirements. The localization method need to conduct valid ToF distance measurements and to receive data from the tags. 57
5.3 Comparison of upscaling algorithms on the Middlebury 2006 dataset [6], using the root mean square error (RMSE) and the percentage of bad matching pixel (PbmP) as metrics. Reproduced with changes from [7]. 60

List of Abbreviations

ADC	Analog to Digital Converter
AR	Augmented-Reality
BER	Bit Error Ratio
CDF	Cumulative Distribution Function
DoA	Direction of Arrival
DoF	Degrees of Freedom
FPGA	Field Programmable Gate Array
FPS	Frames per Second
GPU	Graphics Processing Unit
IMU	Inertial Measurement Unit
IR	Infrared
MIMO	Multiple-Input Multiple-Output
MEMS	Microelectromechanical System
OFDM	Ortogonal Frequency Division Multiplexing
PbmP	Percentage of bad matching Pixels
PLL	Phase Locked Loop
PMD	Photonic Mixer Device
PSK	Phase Shift Keying
RF	Radio Frequency
RFID	Radio Frequency Identification
RMSE	Root Mean Square Error
RSS	Received Signal Strength
SLAM	Simultaneous Mapping and Localization
SoC	System on Chip
ToF	Time-of-Flight
V2I	Vehicle to Infrastructure
V2V	Vehicle to Vehicle
VDMA	Video Direct Memory Access

Glossary

Location-aware Communication
Communication method that features communication partner localization.

Time-of-Flight
The time it takes light to travel over a certain distance. This principle is used for depth sensing with Time-of-Flight sensors.

Photonic Mixer Device
An electronic circuit on each pixel of continuous-wave Time-of-Flight sensors. It enables the measurement of a correlation value between the pixel reference signal and the received light.

Time-of-Flight Tag
An electronic device providing information to a Time-of-Flight system and can be localized in 3D.

Augmented-Reality
Information is overlaid at a certain location within the field of view of the user.

Relay Attack
An attack scenario on secured communication between at least two devices. The connection is relayed to an unsuspecting third party at a different location.

Phase-Shift-Keying
Modulation method to transmit digital data. Information is modulated onto a carrier frequency by shifting the phase of the signal.

FPGA
Field Programmable Gate Array, an integrated circuit with customizable connections between logic blocks. Often used for high performance data processing and hardware prototyping.

Chapter 1

Introduction

This Chapter provides an introduction to this doctoral thesis. Section 1.1 motivates the topic by showing the opportunities of the developed location-aware optical communication methods. Section 1.2 provides the background, an overview of the identified problems and the contributions of this thesis.

1.1 Motivation

Electronic devices are becoming increasingly connected and cooperative. A manifold of different wireless communication methods and protocols have emerged in the last decades to meet this trend. Localization is often an important aspect of device to device interaction and can define the nature and context of the exchange. Localization and communication have however traditionally been treated as separate disciplines and conventional wireless communication methods do not offer precise localization. This doctoral thesis aims to close this gap by combining 3D localization and optical communication in a unified system based on Time-of-Flight sensors.

Location-aware communication in this doctoral thesis is defined as a bidirectional communication method, where at least one involved party is able to localize the others in 3D. The communication link becomes location-aware, since the received and transmitted signals are simulteneously used for localization. Therefore, no external infrastructure is necessary and all involved systems are miniaturizable. This thesis investigates Time-of-Flight sensors for location-aware optical communication, which is illustrated in Fig. 1.1 A). ToF sensors feature an active illumination unit which emits pulsed infrared light for depth sensing. This lightsource can be repurposed to transmit data. During depth sensing, the pixel array receives the reflected light. In optical communication, the pixel data enables to obtain the 3D location and the received data. With the proposed methods, ToF sensors are enabled to communicate with other ToF sensors or miniaturized optical transceivers. Such tags are equipped with a photo diode to receive data and synchronize to a ToF sensor. An LED emits pulsed infrared light which can be modulated with phase-shift keying.

The illustration in Fig. 1.1 B) shows that the ToF sensor is able to localize and communicate with multiple communication partners simultaneously. The source separation is caused by the lens which projects the lightsources to different locations on the pixel array.

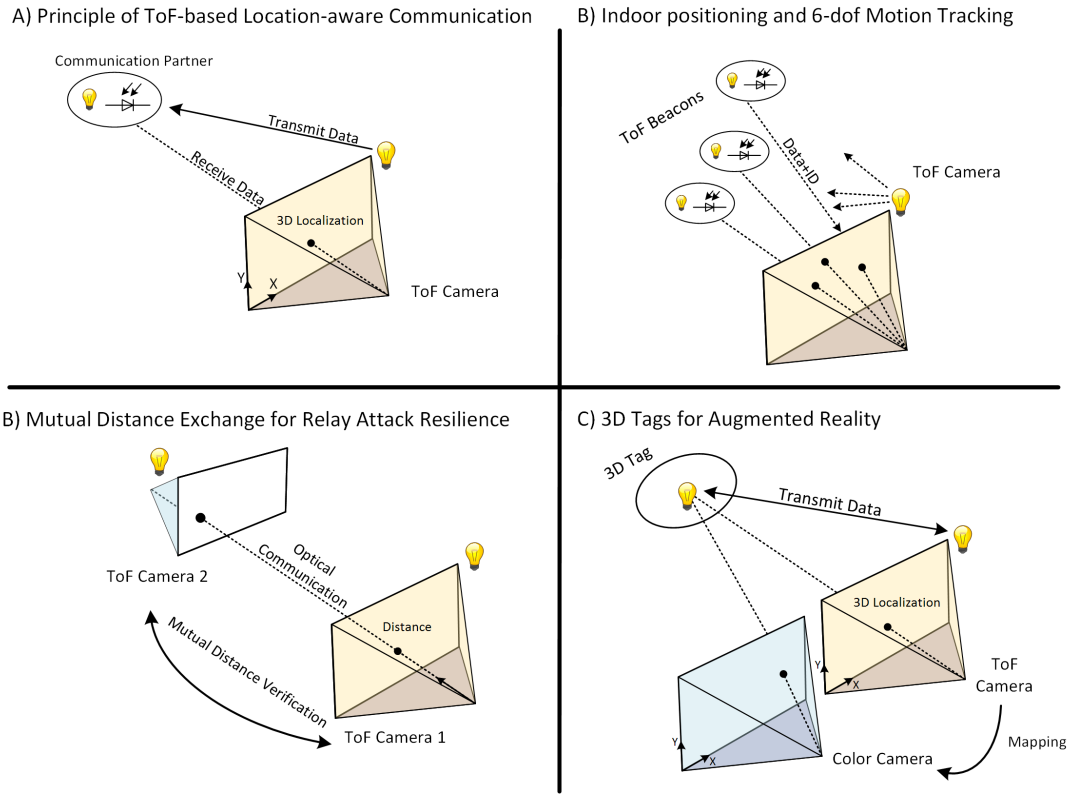


Figure 1.1: The opportunities of location-aware optical communication with Time-of-Flight sensors. A) ToF sensors are able to use depth sensing to localize their communication partners. B) The localized communication partners can act as beacons for fast and precise 6-degrees-of-freedom localization. C) Two ToF sensors mutually localize themselves and exchange their measurements for verification. D) The 3D localization enables to display information at the position of the communication partners on augmented-reality devices.

The ability to simultaneously measure the 3D position of multiple partners at a high update rate can lead to a new class of indoor-positioning systems. This beaconing approach can unify indoor localization and 6 degrees-of-freedom (6-dof) motion tracking into a single system. Compared to existing 2D active beacon systems, 3D beacon localization increases the robustness by reducing the number of required visible beacons. 6-dof motion tracking has long been a problem in the robotics domain, but virtual and augmented-reality systems introduced additional applications in consumer electronics. As the evaluation of the developed methods shows, ToF-based location-aware optical communication can fulfill the demanding requirements of these fields.

Fig. 1.1 C) shows a mutual location-aware connection between two Time-of-Flight sensors. If the measured 3D position is mutually exchanged between these systems, both systems can verify that the other system is authentic. This helps to mitigate relay attacks, in which the unmodified communication between two parties is relayed to an unsuspecting third party. Common contactless communication systems such as NFC, smartcards or entry tokens for vehicles need to apply distance bounding methods to ensure that their communication partner is in close proximity. Between ToF sensors, it is possible to es-

tablish secured and authenticated connections over several meters. Augmented-reality (AR) systems in smartphones and smartglasses have seen promising developments in the recent years. These devices however require sophisticated motion tracking or simultaneous mapping and localization (SLAM) systems to perceive their geometric relation to the environment. As Fig. 1.1 D) shows, location-aware communication can directly connect AR systems with other electronic devices. The relative 3D position of a localized communication partner can be mapped into the field of view of any AR display. By introducing 3D Time-of-Flight tags for augmented-reality, it is possible to give any electronic device an augmented-reality interface. Fig. 1.2 presents this use case for head-mounted smartglasses. The 3D position of the ToF tag can be mapped stereoscopically into the view of the user. Electronic devices provide information and user interfaces, which are overlaid at the object positions. Like normal depth sensors, a ToF-based location-aware communication system also delivers a stream of depth images. This enables the system to co-exist with different applications, such as 3D scanning and gesture control. The latter can be used to interact with Time-of-Flight tags. In the example in Fig. 1.2, the user can activate a light by touching a virtual overlaid button.

These novel applications are compatible with the existing use cases for ToF sensors. A sizeable portion of the devices that will benefit from location-aware communication are already equipped with Time-of-Flight sensors. This work therefore does not only establish a new optical communication paradigm, but it also offers promising additional functionality to existing technology.

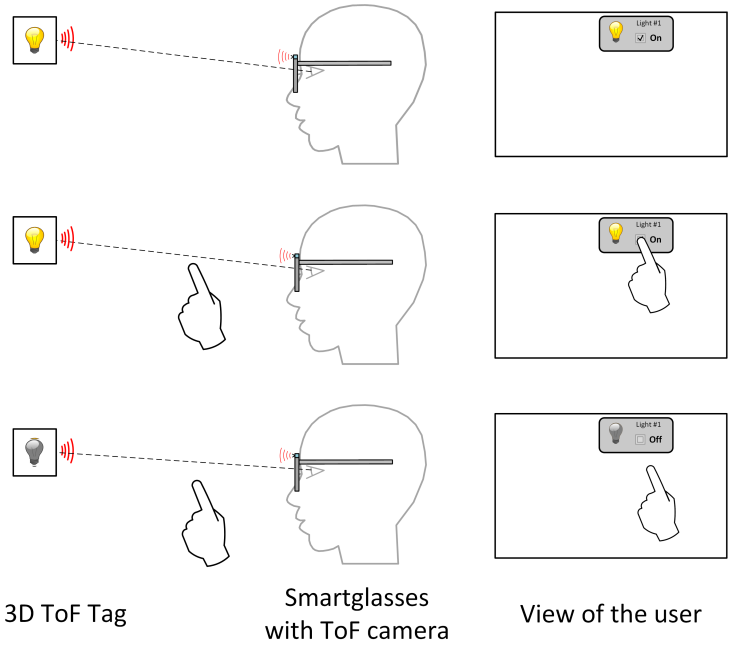


Figure 1.2: Location-aware optical communication with ToF sensors can provide any electronic device with a direct connection to augmented-reality systems. Users are able to interact with these devices by means of an overlaid user interface.

1.2 Location-aware Optical Communication with Time-of-Flight Sensors

1.2.1 Introduction

Continuous-wave Time-of-Flight sensors produce depth images by continuously emitting pulsed infrared light with an active illumination unit. The light is reflected by the scene and projected by a lens to the pixel array. On each pixel, there is a photonic mixer device (PMD), which produces a correlation value between a pixel reference signal and the received light pulses. This pixel output depends on the distance between the sensor and scene but also on the amount of received light. The state-of-the-art method to determine the distance is to capture multiple subsequent phase-images. Each of these images is taken with a different phase offset between the emitted light and the pixel reference signal. Selecting four equidistant phase offsets enables to eliminate the influence of the reflected light and error sources. Therefore it is possible to calculate the phase difference between the pixel reference signal and the received light. The distance is proportional and can be calculated based on the frequency of the light pulses and the speed of light. A thorough analysis of this measurement principle is provided in Section 3.2.

A challenge of this work is to repurpose such ToF sensors for optical communication. As the review of the related work in Chapter 2 shows, ToF sensors have barely been considered for optical communication. This thesis, however, shows that the depth-sensing principle of common ToF sensors makes them well-suited for all identified applications. With the active illumination unit, a powerful optical transmitter is already part of the system. ToF illumination units are designed to emit light pulses at frequencies of up to 100 MHz, since the distance precision partly depends on the light pulse frequency. Most continuous-wave ToF sensors are also capable of shifting the phase of these light pulses to capture multiple phase images per depth image. For data transmission, the phase-shifting system can use phase-shift keying (PSK) to modulate the outgoing light pulses. A ToF pixel is able to receive and demodulate such signal with its photonic mixer device. The ToF pixels need to be digitized at a certain sampling rate to reach the required data-transfer rate. The proposed system design restricts the number of digitized pixels to a certain region around the received light signal. The system dynamically adapts the region to accommodate movements by tracking the projected lightsource position.

Using the PMD to demodulate PSK data comes with certain advantages compared to conventional image sensors. As there is no amplitude modulation involved, each symbol can be transmitted at the same signal strength. The proposed system is able to decode a symbol by capturing two phase images. Depending on the number of different phases, this allows to transmit multiple bits per symbol. The most distinctive advantage is that the phase images do not contain any background light. In the PMD, the generated charges of the received light pulses are distributed into either of two capacitors, depending on the logic state of the pixel reference signal. The pixel output is the difference between the digitized voltages of the capacitors. Background light is not pulsed and thus evenly distributed into the capacitors during an exposure. This can lead to pixel saturation and increased noise, but does not produce a signature in the image. This property enables the direct detection of the modulated lightsources because the sensor is only sensitive to pulsed light.

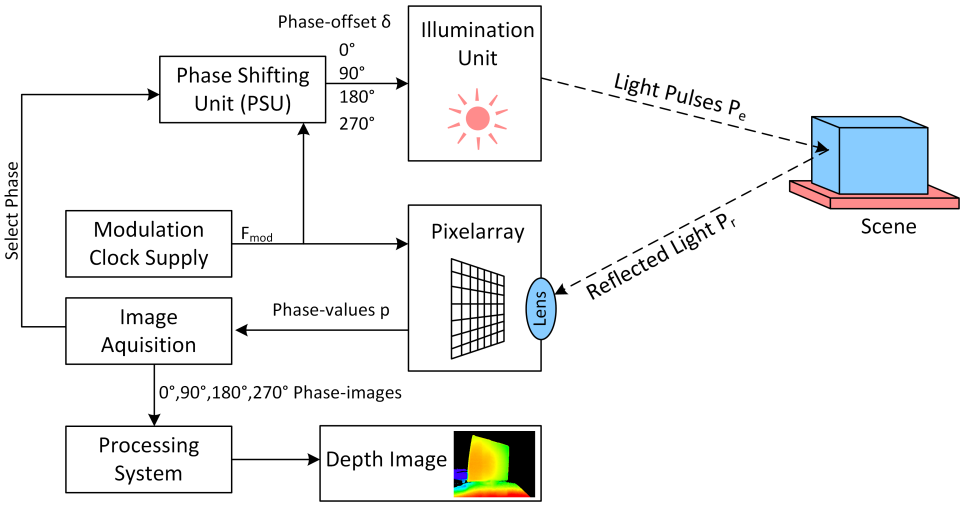


Figure 1.3: The principle of distance sensing with Time-of-Flight sensors. Reproduced with changes from [1].

Communication partner localization adds the location-awareness to the system. The pixel location of the received signal can be used to determine the direction if certain camera parameters are known. Unlike 2D camera systems, a ToF sensor has the ability to directly measure the distance to the communication partner. The direction and distance allow to calculate the precise relative position to the ToF sensor. As shown in Section 3.1, this 3D localization approach improves and establishes a number of different applications. Depending on the final application, the ToF sensor executes alternating phases of optical communication and 3D localization.

Stereo camera systems might be the most viable alternative to ToF sensors for location-aware optical communication. A certain base line between the individual cameras is however necessary to triangulate the distance to the communication partner. This conflicts with the everlasting trend of miniaturization. Depth sensing with structured light also requires a certain base line between image sensor and projector. The distance estimation works by detecting a projected pattern on the sensed object and using triangulation to determine the distance. This might not work for small objects, as the projected pattern needs to be detectable. With 2D image sensors, the 3D position of an object can also be determined by triangulation. This requires the object to be equipped with multiple lightsources or visual landmarks. It also increases the complexity and does not allow for miniaturized devices. As shown in Section 2.4, radio-frequency-based (RF) systems are unable to fulfill the localization requirements of the intended applications.

A major benefit of ToF-based location-aware optical communication is that ToF sensors are already available in different systems. A good portion of these systems can benefit from an additional location-aware communication system. The approach of this thesis aims to remain compatible with regular depth sensing. The communication partner localization requires depth images, which can be provided to the rest of the system. This enables location-aware optical communication as additional feature, which can coexist with the original sensor operation.

1.2.2 Problem Statement

Location-aware optical communication is a novel use case for ToF sensors. Due to the lack of comparable systems, the necessary methods and potential applications have not yet been investigated. This doctoral thesis thus needs to first research and define the enabled applications and their requirements. These requirements are then analyzed with regards to the characteristics of ToF sensors. The proposed methods and system designs need to address the arising problems. The experimental evaluation of the developed designs needs to show that location-aware communication with ToF sensors can meet the requirements of the intended applications. The key problems and challenges can be summarized as:

- Optical communication for ToF sensors has been barely considered in previous work. The ToF sensor needs to be capable of sampling a modulated lightsource at high rates. An efficient modulation method needs to exploit the properties of the ToF distance sensing mechanism.
- The modulation method requires that the ToF sensor synchronizes its modulation signal source to the frequency of the communication partners.
- Location-aware sensor operation involves alternating periods of localization and optical communication. A system design thus needs to propose a method to quickly switch the ToF sensor behavior.
- Motion artifacts can be caused by the distance sensing principle. This affects small and fast moving objects and could restrict the miniaturization of the communication partners. A processing method needs to compensate these artifacts.
- Localization precision and range are important for the intended applications. Methods are needed to reduce the distance sensing error and improve the localization performance.
- Energy efficiency is crucial for all battery-powered devices. It is important to limit the processing effort and the amount of emitted light.
- The novelty of the proposed system needs to be proven with a use case evaluation. Implementing the use case with ToF-based location-aware optical communication should have significant advantages to alternative methods.

1.2.3 Goals and Research Questions

The main goal of this work is to provide fundamental research on location-aware optical communication with Time-of-Flight sensors. As the first step, potential applications need to be identified and analyzed. The requirements of these applications define the problems which need to be addressed in this work. A system design needs to provide insight on how Time-of-Flight sensors can be repurposed as efficient optical transceivers. This involves the development of an efficient data transmission method. A real-time system enables to rapidly switch the sensor operation between localization and communication. A potential application for location-aware optical communication should be implemented and evaluated. The novelty of the proposed system should be analyzed by comparing

the implementation with alternative methods. These goals are motivated by the following research questions:

- Is it possible to turn Time-of-Flight sensors into optical transceivers, while simultaneously measuring the precise relative position of the communication partners?
- How can such location-aware optical communication method shape the future of embedded systems?

1.2.4 Contributions

This thesis was initiated by the idea to create a ToF-based optical communication system with distance checking to prevent relay-attacks. This use-case is introduced in Section 3.1.1. The remaining introduced concepts, methods and ideas were introduced by the author. This doctoral thesis thus makes the following contributions:

- Providing an efficient modulation method for optical communication with Time-of-Flight sensors. The distance sensing mechanism of ToF sensors is converted to transmit and receive signals with phase-shift keying of pulsed infrared light. The mechanism is implemented to proof the feasibility and to conduct an experimental evaluation.
- A synchronization method is designed and implemented, enabling ToF sensors to synchronize their pixel reference to external signals. Synchronization is crucial for efficient optical communication, but also enables additional applications such as external illumination units.
- A system design is provided that implements the communication method to proof the feasiblity of location-aware sensor operation. The design offers an abstraction of the sensor operation and processing methods, enabling to create a customized ToF-based location-aware communication system for any potential application.
- A system design is provided that is optimized for the application of 3D Time-of-Flight tags for augmented-reality. By using an alternative communication method, the system complexity is drastically reduced. The system design works with unmodified ToF sensors and does not require a real-time system. A promising use case is implemented and evaluated with this system.
- A distance tracking method is provided that significantly reduces the required energy for communication partner localization. The method reduces the energy by up to 68%, while only introducing a minor distance estimation error. The method can also be applied to general object tracking applications such as 3D face tracking on smartphones.
- A ToF sensor calibration method is provided, which enables to localize small objects without motion artifacts. This enables fast 3D localization of miniaturized optical communication partners. The method also enables to attach retroreflectors to the communication partners, significantly improving the localization precision and range.

1.2.5 Structure of this Thesis

Chapter 2 introduces and discusses the related work of this thesis. Potential applications and their requirements are presented in Chapter 3. Chapter 4 contains the developed system designs and methods. This includes an optical communication method, the design of the ToF tag system, and several processing methods which enable to meet the requirements. Chapter 5 provides the results and an evaluation of the use case of 3D ToF tags for augmented-reality smartphones. Chapter 6 concludes the thesis and provides directions for future work. This work resulted in 8 academic publications which are included in the final Chapter 7.

Chapter 2

Related Work

This chapter introduces and discusses the related work of this doctoral thesis. Section 2.1 is dedicated to optical communication methods for 2D image sensors. These methods are able to provide reliable and fast line-of-sight connections, but only offer directional localization. The vast majority of existing 3D localization methods in this field are based on an infrastructure of multiple devices. For true point-to-point location-awareness, as proposed in this thesis, single-anchor localization methods are required. These methods are investigated in Section 2.3. Section 2.3 discusses 3D location-aware communication methods, which combine 3D localization with communication. In the final Section 2.4, the metrics of the related work are compared to the approach of this doctoral thesis.

2.1 Optical Communication with Image Sensors

Due to the wide availability of digital cameras, there is considerable research effort on image sensor-based optical communications. Optical communication between image sensors and beacons enables precise indoor-positioning and motion tracking [8]. In the automotive domain, image sensor communication is considered a promising method for vehicle to vehicle (V2V) and vehicle to infrastructure (V2I) communication. The challenge of receiving data with an image sensor is that the vast majority of existing cameras are not designed for optical communication. The introduced work in this section exploits the properties of existing sensors, proposes adaptions or even creates dedicated communication image sensors.

Conventional cameras usually expose their image sensor to the incoming light for a certain time period and then digitize the charge content of the pixels. A simple way to use this mechanism for receiving data from a modulated light source is to obtain a sample per captured image. This enables modulation schemes such as on/off keying [9] or color-shift keying [10]. The data transfer rate of these methods is however restricted by the camera frame rate. High-speed cameras can tackle this problem [11], but increase the required processing power, energy consumption and might lead to a reduced resolution. In certain applications, it is possible to use an array of light sources in a multiple-input multiple-output (MIMO) approach [12]. In order to sample information from these sources in parallel with multiple pixels, the sensor needs to resolve the light sources on the pixel array. This means that the projected light source positions require a certain pixel distance.

An interesting approach is to exploit the rolling shutter mechanism of a camera to increase the sampling rate [13, 14]. Image sensors in consumer electronics almost exclusively use a rolling shutter configuration. The pixels of most of these sensors are not able to reject incoming light signals. Therefore a pixel is light-sensitive until it is digitized in the readout process. Since pixels are usually digitized row by row, the exposure period of each row is slightly delayed by the row digitization time. Since a light source is usually projected on multiple pixels, it is possible to use the readout delay to take multiple samples of the light source from a single image. An established application for rolling shutter communication are lighting installations which also serve as beacons for indoor positioning [8]. Philips Lighting offers a commercial system, enabling to provide indoor navigation and location-based services [15].

Dynamic vision sensors are image sensors with a different pixel readout paradigm. In contrast to conventional imaging devices, these sensors do not produce a series of discrete images. The content of a pixel is only digitized and sent to a processing system when the received light changes. This leads to systems which can react to motion without considerable delay. Chau et al. propose a dynamic vision sensor with active pixels for visible light communication [16, 17]. Their proposed image sensor uses a token passing system during the readout phase. This enables fast sampling rates, since just the pixels which receive a signal are digitized. In this sensor, the first pixel in a row receives a token. When the pixels are selected for readout, the token is passed to the next active pixel in the column. A pixel is only able to receive a token when the optical signal changed since the last readout. Beside the proposed pixel architecture, Chau et al. investigate characteristics such as frequency response and exposure time variation by simulation. The main drawback of dynamic vision sensors is the additional circuitry on each pixel which reduces the resolution and in some cases light sensitivity. The proposed ToF system in this thesis also avoids to digitize all pixels per frame when sampling a modulated light source. If a light source is detected, only a rectangular pixel region around the light source is digitized. This is a trade-off compared to dynamic vision sensors, as not all digitized pixels contain a light source.

The use case of optical communication in the automotive domain in V2V and V2I triggered the development of dedicated communication image sensors. Takai et al. [18] introduce an image sensor which includes dedicated communication pixels along with conventional pixels. These communication pixels can be digitized like a conventional image, producing a binary image. This binary image contains the visible light sources and enables to localize the pixel position. When a communication partner is located, the corresponding communication pixels can be selected and sampled at a high frequency. With Manchester coding, it is possible to achieve a data transfer rate of 10 Mbps. The process of signal localization and sampling is continuously repeated in order to accommodate the movements of the communication partners. Goto et al. extend this sensor with optical orthogonal frequency division multiplexing (OFDM) [19], leading to a data transfer rate of 55 Mbps. Such dedicated communication pixels are a promising method for image sensor communication. A trade-off is that the communication pixels require space on the sensor array, which reduces the resolution of the conventional image and also might cause aliasing artifacts. A minor drawback is that it might be difficult to apply sub-pixel resolution methods on the binary image to increase the precision of the directional localization. As this thesis shows, Time-of-Flight sensors feature unique properties making them well-suited for op-

tical communication. Dargie and David were the first to recognize the potential of using the photonic mixer device as optical demodulator [20]. They correctly assessed, that the pixel modulation suppresses the background light, which enables robust signal reception. Unfortunately there has been no further investigation or implementation. Over a decade later, Yuan et al. created the first communication system with a Time-of-Flight sensor [21]. The decision of using a ToF sensor was motivated by providing smartphones with ToF sensors a method to receive data from imperceptible infrared light sources. In the vision of the authors, these light sources can be mounted at signs and posters to provide further information to the user. The basic communication principle is to use a photo diode to reconstruct the modulation signal of ToF light pulses. The reconstructed signal is then amplified and phase-modulated with a phase shifter circuit. The signal is then immediately sent back to the ToF sensor via LEDs. In that approach, the ToF sensor is not modified and captures a constant stream of depth images. The received information is decoded by analyzing depth variations, which appear at the position of the LEDs. Due to the low frame rate of ToF sensors, the authors propose to use an array of LEDs in a MIMO arrangement. With a 3x3 LED array and a ToF framerate of 60 Hz, the system has a data transfer rate of 960 bps with 95% bit accuracy. The proposed method of Yuan et al. avoids some challenging problems, such as modulation signal synchronization. There is also potential to extend the approach by customizing the ToF sensor operation and using single phase images for signal demodulation. As experiments in this doctoral project however showed, the reconstruction of a ToF modulation signal with photo diodes is a major challenge. A state-of-the-art ToF sensor uses a modulation frequency between 60 and 100 MHz. Even with fast photo diodes, such high frequencies require a certain signal strength. This works in free-space optical communication, where photo detectors directly receive a bundled light signal. However, a ToF sensor distributes its illumination light evenly on the scene. This also adds multi-path effects, since a diode receives its light over a wide angle. This severely distorts the reconstructed signal due to the different light-path distances. Another issue are multiple ToF sensors, which will interfere with each other. Due to these reasons, the approach presented in this thesis avoids reconstructing the ToF signal but proposes a method for ToF sensors to synchronize to the light sources.

2.2 Single-Anchor Localization

Single-anchor localization is related to the localization principle of this work. The challenge is that just two parties are involved and hence the relative position cannot be be obtained by triangulation or trilateration. An anchor has to emit an optical or RF signal in order to localize other devices. These devices either just receive the signal or transmit a response. In the context of this work, the anchor is a Time-of-Flight sensor, and the other devices are either other ToF sensors or ToF tags. Both anchor and tags may change their position, so prior knowledge about the surroundings cannot be exploited. The main goal of state-of-the-art single-anchor localization approaches is indoor localization. If the position of the anchor in a world coordinate system (e.g. floorplan) is available, it is also possible to calculate the world position of each localized tag. Single-anchor approaches often combine directional localization with distance sensing. Cidronali et al. [22] use a switched beam array antenna to measure the signal direction of arrival (DoA). A tag is equipped with

an omnidirectional antenna and a base-station can estimate the DoA of the tag with its own directional antenna. The distance is measured by the received signal strength (RSS) leading to an average localization error of 1.7 m. Other systems with directional antennas achieve a comparable precision [23].

Since RF signals are reflected by most surfaces, the signal takes multiple paths before it arrives at the receiver. Due to the time of flight, each signal path arrives at a different time. This effect is exploited in multipath-assisted approaches to improve localization. A recent system is SALMA by Grosswindhager et al. [24], which is capable of localizing multiple tags in a crowded environment. Unlike optical systems, a line-of-sight connection is not required and results in a 20 cm positioning error. If there is a line-of-sight connection, the system can achieve a 8 cm positioning error. Due to the intended application as an indoor navigation system, the tags are localized in a 2D coordinate system. According to the authors, the approach could be extended to 3D localization. This would however require additional research effort and increase the computational complexity.

In the optical domain, calibrated cameras offer precise directional localization due to the signal projection on the pixel array. In order to calculate the relative position, the distance between the camera and the tag is required. The distance can be coarsely estimated by the visual appearance of the object if geometry of the tagged object is known. However, smaller objects cannot be resolved by the pixel array beyond a certain distance. Using the received optical signal strength or the pixel size of the light source offers only coarse distance estimates [25]. Moreover these parameters also depend on the orientation between the light source and the camera due to the radiation characteristics.

A promising optical single-anchor system is SmartLight by Liu et al. [26]. The authors propose a de-focused projector, which emits light at a distinctive frequency at each pixel. Since modern LED lighting installations sometimes consist of LED arrays behind a lens, this projector can potentially be implemented as a ceiling-mounted light. The device to be localized receives the light of multiple defocused pixels with a light sensor. The included signals can be separated by a Fourier transformation. Due to the defocused projection, the received frequencies depend on the 3D position of the lightsensor. The detected frequencies are then marked on a bitmap. It is possible to derive the directional localization via the center pixel position of the light source on the bitmap. The distance is estimated by the size of the detected frequency cluster. In an indoor positioning scenario, SmartLight achieves a 10 cm positioning error with a commercial projector instead of a lighting installation. One could argue that this system inverts the problem of 2D camera-based light source position estimation by the use of a projector. A projector however removes intrusive imaging hardware, preserving the privacy of the user. Another advantage is that use of high pixel frequencies can decrease the influence of ambient light. This is also a characteristic of ToF sensors, which feature modulated pixels. The SmartLight system also moves the sensor complexity to the base-station and allows the lightsensor system to locate itself with low cost and complexity.

2.3 Location-aware Communication Methods

This thesis defines the term location-aware communication as a communication link where at least one participant is able to measure the relative 3D position of the other participants.

This section reviews contributions which follow this paradigm by combining localization and communication. As the previous section showed, location-aware optical communication between just two entities is hard to implement with RF signals. A communication link can also be location-aware when both communication partners are localized within the same coordinate system. The global positioning system (GPS) and its derivatives offer a localization system, which provides location-awareness in a manifold of applications. With upcoming 5G network, RF-based localization is not just a feature of mobile communication, but becomes an important requirement due to beamforming [27]. These multi-anchor approaches however require an infrastructure, which is not omnipresent in indoor environments.

A majority of point-to-point location-aware communication approaches are motivated by radio frequency identification (RFID) technology. RFID enables to deploy active or passive tags, which can be identified by an active reader. Knowledge of the relative position of RFID tags can be valuable in applications like logistics, theft prevention and people tracking. Yang et al. present Tagoram, which is a method to locate a moving RFID tag with a stationary off-the-shelf RFID reader [28]. By conducting multiple measurements of the tag at different positions, a virtual antenna array is created. Multiple antennas enable the challenging angular localization and result in 6.35 cm median accuracy in a real-world application.

Goller et al. propose a method to visually locate RFID tags without active illumination [29]. Assuming a camera is mounted at a fixed position, the authors use blob detection and tracking to visually observe the movement of tagged objects in the scene. At the same time, the positions of the connected tags are coarsely tracked via the received signal strength (RSS) of the RF signal. If an RFID tag is moved, the visual movement of the blob is associated with the RF-based movement observation. A potential application field are theft prevention systems, where the identity of an item is associated with an observed person. In such scenario the detection probability exceeds 90% with four moving tags. Duan et al. also fuse computer vision with RFID in their work [30]. Along with a blob-detector, the authors use the RFID phase data for distance estimation. By matching the phase to visual blobs, it is possible to associate the optical data with the RFID phase data. In the evaluation, four different moving tags could be tracked at 10.33 mm accuracy with 97% matching accuracy. This approach also requires a camera at a fixed position and moving Tags.

Sample et al. propose to use integrated LEDs to localize stationary RFID tags with a moving RFID reader [31]. In that approach, passive UHF tags are equipped with LEDs which are activated during RFID read operations. A synchronized color camera is mounted next to the reader. The camera subsequently takes two images, one of them with the activated LED. The pixel position with the maximum brightness change between the images is used to calculate the direction of the tag. To improve robustness, the tag is coarsely located with RF methods, before optical localization. The accuracy of this directional localization depends on the resolution of the color camera and can be further enhanced with super-resolution methods. According to the authors, the final accuracy depends on the quality of the camera calibration. For 3D localization, the authors propose to either sense at least two tags with a known distance between them, or to use a stereo camera system. Both approaches allow determining the relative 3D position by triangulation. When the stereo cameras are mounted at a distance of 200 mm, the RFID tags could

be localized with an error of 10-20 mm at 2 m distance. Stereo systems are a viable optical alternative to the ToF approach in this work. Unlike Time-of-Flight or structured light, stereo can measure depth without active illumination. With stereo cameras however comes the drawback that two sensors need to be mounted at a certain distance which prohibits miniaturization. Stereo cameras also require relatively high resolution for precise triangulation, which reduces the frame rate and consequently the optical communication ability.

Ashok et al. have created a low-power RFID tag localization system for smartglasses [32]. This system enables smart-glasses to detect which electronic device or RFID tag the user is looking on. This enables the display of the devices information on a screen in the users field of view. The reader uses multiple photo diodes to determine the direction of the tag, which is equipped with an infrared LED. The RF reader can contact the tags and trigger a short light pulse. Since each photo diode is mounted at a different location and angle, it is possible to use the received signal strength to determine the direction of the tag. Using multiple photo diodes instead of a camera to estimate the angle of arrival decreases the system complexity and enables the use of short light pulses without any synchronization. In the case of the work of Ashok et al., this leads to a directional error between 1°and 2°. Avoiding the use of a camera not only decreases the system complexity, but also avoids potential privacy violations. A disadvange of photo sensor arrays is that there is no source separation from lens projection. Therefore only a single tag can be sensed at the same time. The approach of Ashok et al. avoids this by synchronizing the light pulses the RF communication link. Such single light pulse localization might not be sufficient for directional localization in an environment with other infrared light sources. The distance is calculated from the received light intensity, which leads to a distance uncertainty of 40 cm. Besides the LED radiation pattern, another reason for this significant distance error is that light takes multiple paths from the tag to the photo diode.

The recent work of Xie et al. presents a method to localize RFID tags with sufficient precision for augmented reality [33]. Instead of active LEDs, they use a combination of passive RFID tags and a Kinect depth camera. The goal of the system is to associate objects in the depth image with RFID tags. Therefore the tags are first coarsely localized using the RFID phase information and two antennas. Additional information is necessary to associate this coarse location with a 3D object in the depth image. The complete setup, which includes the RFID reader and Kinect camera, rotates around a vertical rotation axis during operation from -40°to +40°. During this rotate-scanning procedure, the system measures the distance to each tag by the RFID phase value. When the platform directly faces the tag, the measured distance reaches a minimum. The vertical angle is determined by sensing the angle-of-arrival between both antennas. This coarse localization is then matched with the 3D pointcloud and is associated with a 3D object. Since this approach matches RFID tags to visible objects, the work is evaluated with a focus on the matching performance. If a tag is matched, the precision of the object localization depends on the depth sensing system. The matching performance depends on the proximity of the RFID tags and reaches over 90% if the tags are separated by at least 50 cm. Sensing information about tags or electronic devices with a depth camera enables to map the relative 3D position into any other coordinate system if the transformation parameters are known. In the case of augmented reality, it is possible to use a see-through display to directly overlay the information. A major difference between the approach of Xie et al. and this thesis

is that optical communication does not require object recognition. If the location-aware optical communication approach is used to localize tags, the tags can be mounted at any arbitrary point and not on distinct objects.

A constraint of optical communication with small embedded devices is the power consumption of the modulated light source. The use of modulated retroreflectors can potentially reduce the power consumption of such devices. Kahn et al. propose using retroreflectors for their smart dust system [34]. One side of a corner-cube retroreflector is slightly rotated by a microelectromechanical system (MEMS). This enables to transmit information to a camera system with an illumination unit. In the smart dust system, a number of sensor nodes are deployed in an area and are later localized by an aircraft. With a Time-of-Flight system, such modulated reflector could reflect the illumination signal and lead to both 3D localization and data exchange. Since significantly more light is reflected into the direction of the ToF sensor, the signal-to-noise ratio and range are improved as well. Ollikkala et al. arrive at this conclusion in their investigation of retroreflectors for Time-of-Flight sensors [35].

2.4 Comparison to the State of the Art

This section provides a comparison between the related work and the approach of this thesis. Table 2.1 compares the localization accuracy, update rate, evaluated range, and the constraints of the related approaches. The metric for accuracy is the average of the combined localization error of all dimensions. The approaches were evaluated either evaluated in a 2D area or within a 3D volume. If not specified by the authors, the diagonal of the evaluated area is assumed to be the the maximum range. The listed limitations of the presented approaches are system characteristics which prevent miniaturization or add significant complexity or cost. All of the presented approaches use the same localization setup as the proposed system in this thesis: Each approach involves two or more devices, where at least one device is able to identify and determine the relative 3D position of the other devices. This single-anchor localization principle enables or improves the applications which are presented in this work. The communication method of all introduced approaches is based on RF communication and thus superior to ToF-based optical communication.

The listed approaches in Table 2.1 are introduced and discussed in Section 2.2 and Section 2.3. Duan et al. manage to achieve the best accuracy of 10 mm within a 4x8 m area [30]. Their approach uses a combination of a 2D 720p camera with RF-based RFID tag localization. The drawback of the method is that the object can only be localized in 2D on a planar surface. If the object is moving, the visual motion can be matched to the

	Accuracy	Evaluated range	Update rate	Localization method	Limitations
Ashok et al. [32]	400 mm (max)	9 m	10 Hz	Optical and RF	
Duan et al. [30]	10 mm (avg)	9 m	60 Hz	Photogrammetry	Localization on 2D plane
Liu et al. [26]	100 mm (avg)	3 m	60 Hz	Optical signal strength	
Sample et al. [31]	20 mm (max)	2 m	not specified	Stereo vision	Visible light
Grosswindhager et al. [24]	80 mm (med)	4 m	4.5 Hz	RF	Localization on 2D plane
Xie et al. [33]	not specified	5 m	not specified	Structured light	Rotating antenna
This work	20 mm (max)	4.5 m	30/5 Hz	Time-of-Flight	

Table 2.1: Comparison of related localization and location-aware communication approaches.

trajectory of the corresponding RF signal. These restrictions inhibit this approach for the ToF-based location-aware optical communication applications.

The approach of Ashok et al. also combines optical localization with RF communication [32]. Their work aims to provide smart glasses with the capability to detect objects within the field of view. This use case is related to the ToF tag system for augmented reality devices in this thesis. However instead of directly overlaying information at the tag position, the approach of Ashok et al. enables to display information about present devices on the side of the field of view. The relatively large ranging error originates from the RF distance sensing method which would impair a mapping to another viewport. The system, however, requires significantly less processing power and energy than a ToF-based location-aware communication system.

Sample et al. [31] propose to localize RFID tags with an LED and a stereo camera. Due to the stereo triangulation principle, the localization accuracy is relatively accurate with an average error of 2 cm. Visible light flashes might, however, distract the environment and sensing them might be impaired by other light sources. To best knowledge of the author, an infrared stereo system for location-aware optical communication has not yet been proposed. Such a system could lead to comparable performance as the ToF-based system and is further discussed in Section 5.5.2.

The single-anchor approach of Grosswindhager et al. is a promising RF-based localization method. Since the intended application is indoor positioning and navigation, the implementation is limited to 2D floorplan localization. According to the authors, it is possible to adapt the approach for 3D localization with additional effort. If the current accuracy of 80 mm can be maintained with 3D localization, it might be possible to meet some of the application requirements listed in Section 3.1.

The recent approach of Xie et al. [33] aims for a use case which is similar to ToF tags for augmented reality. The focus on their work is on matching pointclouds from structured light sensors with RFID tags. If an object is, however, correctly matched with a tag, the localization performance is identical to the 3D object sensing capabilities of the structured light sensor. Matching RFID tags with pointclouds or other visual data might restrict the potential applications. This is further discussed in Section 2.3.

This doctoral thesis establishes the first location-aware communication system with Time-of-Flight sensors. Such ToF-based system offers the unique combination of precise 3D localization, long range, high update rate and optical communication functionality. The major drawback of the related approaches (except stereo sytems) are the imprecise distance sensing capability and the localization error. Time-of-Flight distance sensing enables to precisely localize any miniaturized communication partner with a single ToF sensor. The optical communication capabilities of ToF sensors can be considered superior to conventional 2D cameras and enable to implement all intended applications without an additional communication system. Repurposing ToF sensors as location-aware optical transceivers is not trivial and requires extensive system design considerations. This thesis therefore provides two different system designs and several signal processing methods. The fundamental research results thus greatly reduce the future effort for location-aware optical communication systems, possibly leading to wide adoption.

Chapter 3

Application Requirements and Sensor Characterization

This chapter discusses the background and requirements of the proposed location-aware communication approach. These requirements are found by evaluating potential applications while considering the characteristics of Time-of-Flight sensors. As illustrated in Fig. 3.1, the intended applications produce a set of requirements, which are then considered in the developed system designs and methods. This chapter therefore first introduces the intended applications and their requirements in Section 3.1. In Section 3.2, the characteristics of the ToF sensor are analyzed in detail. The application requirements and the ToF system characteristics lead to the contributions of this thesis, which are introduced in Section 3.3.

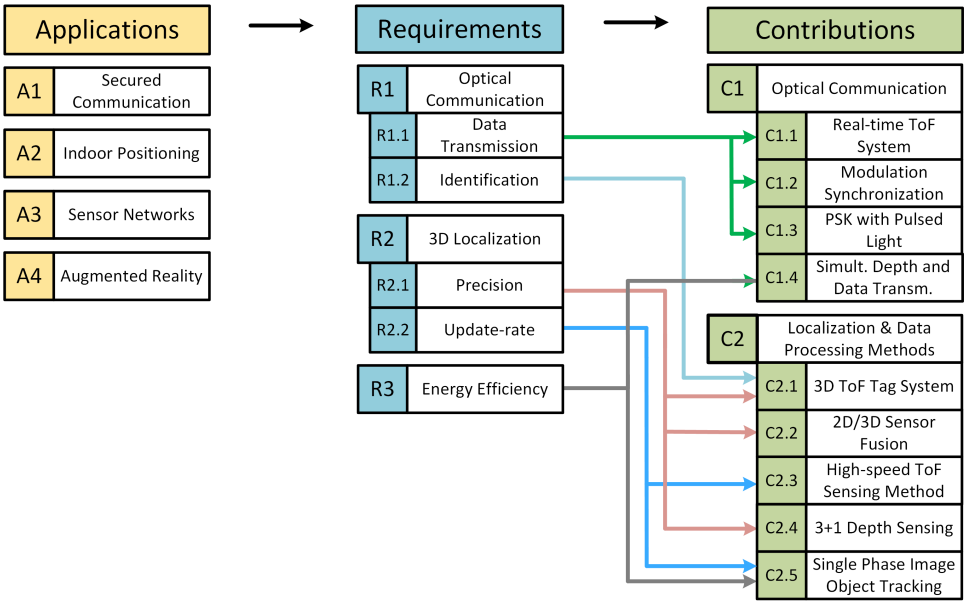


Figure 3.1: Potential applications are evaluated to obtain the requirements, which are then addressed by several system designs and methods.

3.1 Applications and Requirements

This section introduces potential applications of the proposed location-aware optical communication approach. These applications are either enabled by this work or improved by adding functionality or reducing complexity and costs. As seen in Fig. 3.1, the requirements of these applications are matched with the ToF sensor characteristics. This evaluation defines the system design goals, which specify how the applications requirements can be fulfilled. The developed methods and system implementations in this work satisfy these goals, providing the foundation for any future implementation.

3.1.1 A1: Secured Communication

A potential application for location-aware communication is secured communication. Certain wireless communication systems are vulnerable to relay attacks. This attack scenario involves an attacker, who forwards communication signals from one communication partner to another one without any modifications. This attack is severe for connections which are established from proximity or context. A prominent example are passive wireless entry key tokens for cars. Francillion et al. [36] demonstrated the severity of this attack by successfully gaining access to 10 different car models. Forwarding the communication to an unsuspecting third party also poses a risk for contactless payment systems. In that case, the signal of a payment device could be transmitted to a terminal at a different location, paying for the purchase of an attacker. Research on relay attack prevention involved the development of numerous distance bounding protocols [37]. These protocols restrict the distance by evaluating the round-trip time of such communication links. Due to the challenging implementation of such measurement, the vulnerability for relay attacks still persists.

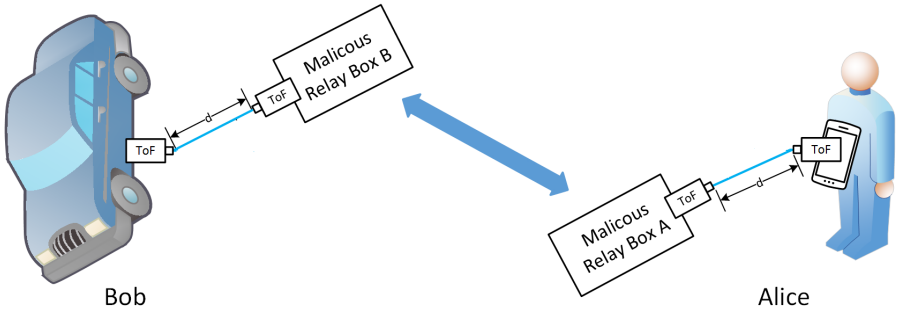


Figure 3.2: A relay attack on a secured optical ToF connection is infeasible, because the distance measurements d can be exchanged and mutually verified. Obtained with changes from [2].

Location-aware communication can considerably mitigate the risk of relay attacks. With the results of this work it is possible to establish a communication protocol with mutual distance checks. Two Time-of-Flight sensors can exchange information via optical communication and use conventional Time-of-Flight imaging to measure the distance to each other. Using conventional cryptography, the devices establish a secured and authenticated connection, and exchange their distance measurements. As seen in Fig. 3.2, the distance measurement d of a system needs to be identical with the other's. Different distance

measurements are an indication for a relay attack. An attacker on such a system would need to manipulate the distance measurement of both involved Time-of-Flight cameras. Such attack can be considered unfeasible and could be detected by the ToF cameras. In order to implement a secured communication protocol, a certain data transfer rate is required to handle the protocol overhead. The mutual distance checks do not have to happen frequently, so a low communication partner localization update rate is sufficient. The precision of the localization is supposed to be within the margin of a few centimeters, which Time-of-Flight sensors can provide. Since this secured communication method is not continuously operating, the energy efficiency requirements can be given a low priority.

3.1.2 A2: Indoor Positioning

Indoor positioning is still a challenging problem with no widely adopted solution. Emerging spatial computing applications such as virtual and augmented-reality require a combination of indoor positioning and motion tracking in a single system. This work therefore proposes a high-performance indoor positioning method, fast and accurate enough to be used as a six degrees-of-freedom (6-DoF) motion-tracking system [5]. In this approach a number of beacons are distributed in the environment. Each beacon consists of a modulated light source and a retroreflector. The ToF system locates and identifies these beacons. Each beacon has an associated 3D position in the target world coordinate system. If at least three beacons are visible in the field of view, the system can calculate its position and orientation within the target coordinate system by estimating the rigid transformation.

The optical communication link between beacon and ToF system only needs transmit an identification number to the ToF sensor. The localization update rate should be as high as possible to guarantee robust tracking. This update rate can be later improved by an inertial measurement unit (IMU). The precision of the beacon localization is crucial and can be greatly enhanced with retro-reflectors. An energy efficient implementation of the communication system is crucial for these applications, due to continuous operation on battery powered devices.

3.1.3 A3: Sensor Networks

A challenging problem in sensor networks is accurate and reliable node localization. This is especially difficult in large sensor networks and when sensor data needs to be associated with a accurate location. As shown in Chapter 7, this work proposes a localization approach using location-aware optical communication. In this approach, the ToF sensor is already part of an existing 6-dof indoor positioning system. As seen in Fig. 3.2, the ToF system uses this indoor positioning system to determine its own position and orientation in the world coordinate system. The ToF system determines the relative position P_R with regards to each visible node. From the position and orientation of the ToF sensor, it is possible to calculate the node position P_W in the world coordinate system. This procedure also associates the node position with its identity via optical communication. The 3D location P_W can then be forwarded to the node, either via optical communication or a secondary RF-based link.

For this approach, the optical communication link can either just feature a low data-transfer rate or be limited to simple identification broadcast. Data transfer from the ToF

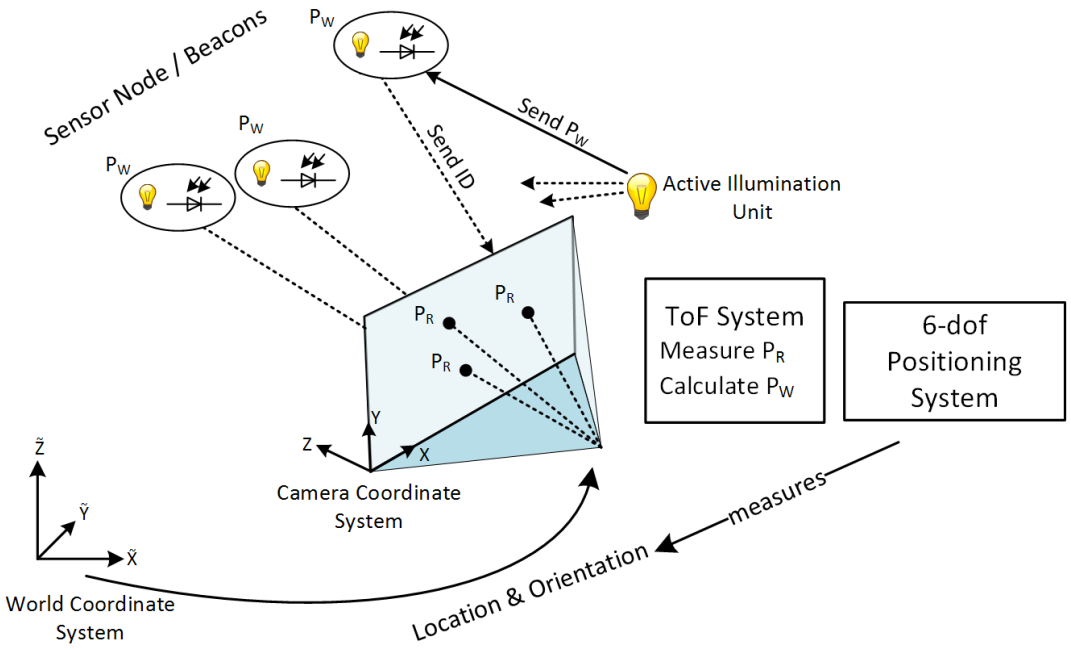


Figure 3.3: Location-aware optical communication can be combined with existing indoor positioning systems to localize sensor network nodes.

sensor to the node is optional, because the sensor data can be externally associated with the location. The localization should be accurate and feature a medium update rate, in order mitigate the latency between the indoor positioning system and the ToF system. Since the ToF sensor is part of the indoor positioning system, the optical communication must not impair the depth sensing functionality. The power consumption of the system is not crucial, since static nodes do not need to be frequently localized.

3.1.4 A4: Augmented-Reality

Most augmented-reality (AR) systems require complex simultaneous localization and mapping (SLAM) systems in order to overlay virtual information in the field of view of the user. This work provides a way for direct AR, which avoids complex systems and additional infrastructures. If an object features a small optical transceiver (ToF tag), it is able to be localized and identified by a ToF sensor. This allows to directly overlay information about the object into the AR view. With a bidirectional communication link, it is possible to interact with this device. The use case of ToF tags for augmented-reality smartphones is addressed in this work with a system design and an implemented use case. Section 5.5 provides an evaluation of this use case.

In order to map the information at the location of the ToF tag into the field of view, the relative position and identity of the tag need to be known. The AR view can either be a video stream of a color camera, as used in smartphones, or an overlay directly mapped into the field of view of the user. The relative position is provided by the location-aware communication. The 3D position in the AR view can be calculated by a geometric

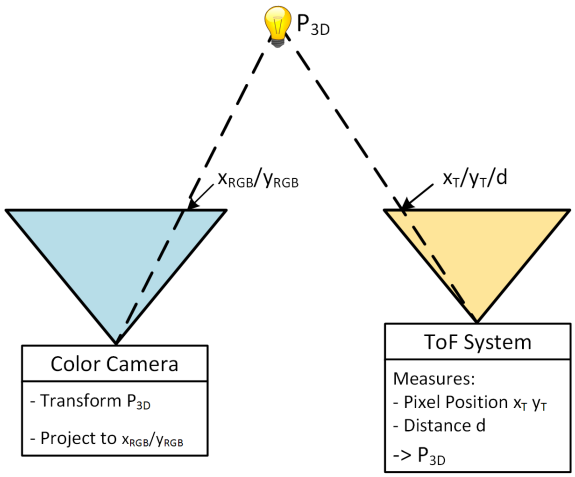


Figure 3.4: 3D localization enables to determine the pixel position of the communication partner on a color camera stream. This enables to overlay information in augmented-reality devices.

transformation between the coordinate system of the ToF sensor and the viewport. For this approach, the optical communication can be limited to tag identification. If an additional information source or conventional communication link is used, the optical communication can be reduced to identification. As for application A2, the localization needs to be as frequent as possible. For stationary tags, the localization update rate can be improved with IMU sensor fusion. The final position of the displayed information about the tag depends on the measured 3D position, which requires accurate localization.

Time-of-Flight sensors are already used in head-mounted AR devices for gesture recognition. As seen in Fig. 3.4, combining optical communication with gesture detection enables a novel method to interact with electronic devices. This requires the communication system to also provide a constant stream of depth images to the gesture recognition system. The power consumption is crucial for this application, since future AR devices are expected to be worn continuously. Therefore this work proposes methods to reduce the power consumption of communication partner localization [3]. A combination of depth sensing and data transmission, as proposed in Section 4.1.1, can further decrease the required power.

3.2 Time-of-Flight Sensor Characterization

At the time of writing this doctoral thesis, location-aware optical communication has not been considered in any existing ToF sensor design. The characteristics and functionality of ToF sensors therefore needs to be analyzed before defining the system design goals for the proposed optical communication system. This section characterizes the Infineon REAL3™ ToF sensor based on Time-of-Flight technology of pmdtechnologies. The developed concepts and methods in this doctoral thesis are, however, transferable to the majority of continuous-wave Time-of-Flight sensors. The section further provides insight into the operating principle of continuous-wave ToF sensors. The sensor characteristics are then discussed with respect to the application requirements.

3.2.1 Distance Sensing Principle

As introduced in Chapter 1, continuous-wave ToF sensing is based on measuring the phase difference of reflected light pulses. This is accomplished by photonic mixer devices (PMD), located at each pixel. When photons arrive at the pixel array, they generate charge-hole pairs in the silicon.

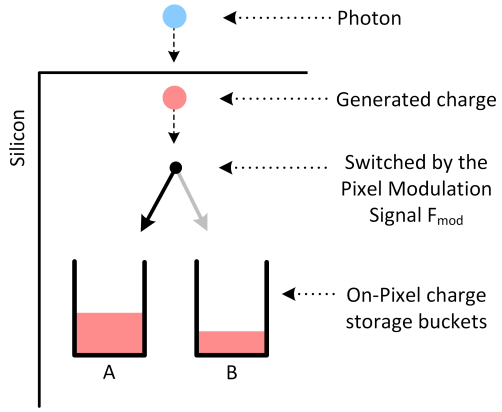


Figure 3.5: The photonic mixer device is located at each pixel. Incoming photons generate charges in the silicon, which are redirected by the PMD into either capacitor A or B. Reproduced with changes from [1].

As illustrated in Fig. 3.5, the PMD acts as a switch, redirecting the charges into either of two capacitors A or B. This switch is controlled by the pixel reference signal P_{Ref}, which has the same frequency as the emitted light. Due to these capacitors, the ToF sensor can integrate the received signal for a certain exposure time. This is illustrated in Fig. 3.6. After the exposure, the sensor readout phase digitizes the contents of A and B for each pixel.

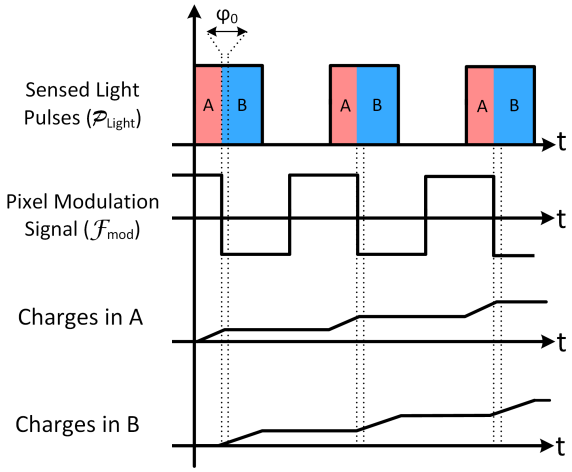


Figure 3.6: The charge content of the pixel capacitors A and B during exposure. The ratio between A and B depends on the phase difference φ_0. Reproduced with changes from [1].

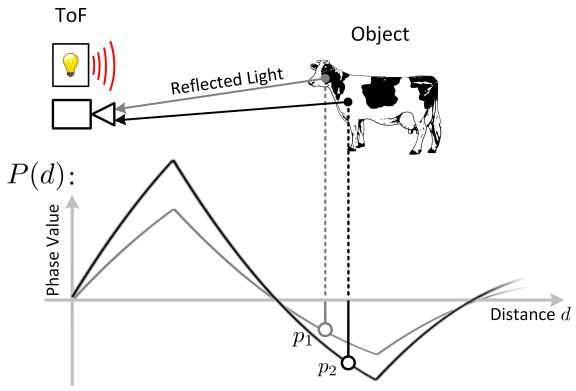


Figure 3.7: The phase value of an object depends on the object reflectivity and distance. Reproduced with changes from [3].

The raw sensor output, referred to as phase value, is the difference between the digitized values of A and B. Besides the exposure time and the distance, this value also depends on the intensity of the received light. As seen in Fig. 3.7, the received light intensity depends on the reflectivity of the sensed object and the distance due to the light intensity degradation by the inverse square law. For depth sensing, the influence of the light intensity needs to be removed to calculate the distance. The state-of-the-art approach is to capture at least four different phase images of the same scene. Each image is captured with a different phase offset between the emitted signal and the pixel reference signal. This is accomplished by a phase shifting unit which shifts the phase of the emitted light. This work proposes to use this phase shifter to modulate the emitted light pulses with phase-shift keying for communication. The relation between phase offset δ and phase value L at a constant distance is commonly referred to as autocorrelation function $c(\delta)$. As seen in Fig. 3.8, the phase-shift between outgoing and received light φ_d corresponds to the phase-shift of the autocorrelation function. Capturing four different phase images with equidistant phase offsets produces four different samples L_i of the autocorrelation function. With these four samples, it is possible to calculate φ_d with the following relation:

$$\varphi_d = atan\left(\frac{L_{90^\circ} - L_{270^\circ}}{L_{0^\circ} - L_{180^\circ}}\right) \tag{3.1}$$

This operation assumes a sinoidal autocorrelation function. The shape of this function, however, also depends on the shape and distribution of the emitted light pulses. Therefore this introduced wiggling error is compensated by calibration. With the speed of light c and the modulation frequency f_p, it is possible to calculate the distance for each pixel.

$$d_{obj} = \frac{c}{2} \cdot \frac{\varphi_{obj}}{2\pi f_P} \tag{3.2}$$

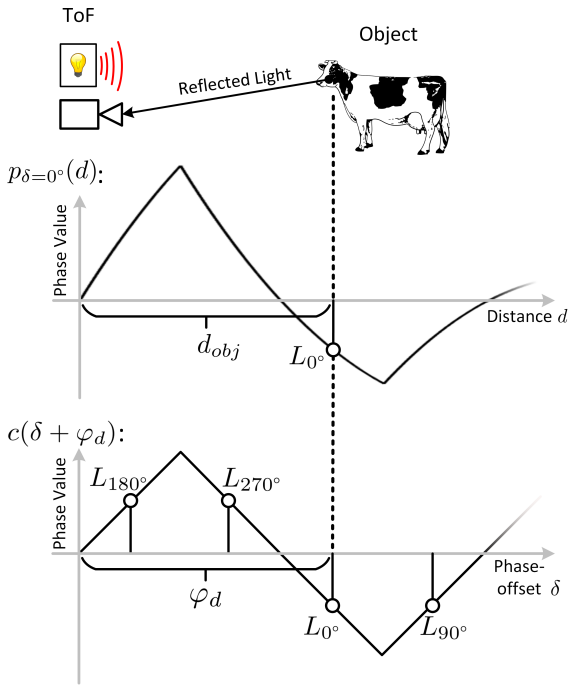


Figure 3.8: Calculating the distance requires four different samples of the autocorrelation function. Reproduced with changes from [1].

3.2.2 Suitability for Optical Communication

The established system implementation shows that the ToF sensor is capable of at least 500 full sensor readouts per second. It is possible to define a pixel region of interest, which restricts the digitized pixels to a sub-image. This enables to only digitize the image region which contains the modulated light source. As discussed in Section 5.1.2, this leads to a sampling rate of up to 7300 Hz with 0.3 μs exposure time. Due to the active illumination unit, ToF systems are equipped with all necessary features to also transmit data. The active illumination unit can be configured to emit a modulated light signal, rather than just distance sensing pulses. Furthermore, the illumination unit is suited for pulse frequencies of over 80 MHz. Since the outgoing signal can be phase-shifted with an integrated phase-shifter, it is possible to introduce phase-shift keying of pulsed light as an efficient modulation method. This thesis focuses on this modulation method as it can also be received with ToF sensors. Conventional free-space optical communication methods like on/off keying or pulse position keying can however also be implemented on ToF sensors to transmit data [38].

As the results of this work indicate, ToF sensors are capable optical receivers. Each pixel features a PMD, which demodulates groups of PSK-modulated light pulses. This enables to receive multiple bits per frame, while the modulated light source is visible in each image. The latter enables rapid light source tracking and consequently fast directional communication-partner localization. Another advantage of the PMD as demodulator is that the pixel output is only sensitive to pulsed light. Other infrared light sources such as sunlight can cause noise and pixel saturation. They, however, do not directly influence

the sensed phase values, because the photonic-mixer device integrates the received signal alternatingly into two capacitors. Continuous light signals are distributed evenly into these capacitors. The final phase value is the result of a subtraction of the digitized charges in capacitor A and B, which removes the evenly distributed light signal. Phase-shift keying of pulsed light also requires modulation-signal synchronization between the ToF sensor and its communication partner. The frequency of the received light pulses and the pixel reference frequency are not identical. A fast and robust synchronization method is therefore a design goal which is introduced in Section 4.1.1.

A system design that is capable of meeting the necessary application requirements needs to be able to change the ToF sensor behavior frequently with minimal delay. Different operation modes such as signal sampling, data transmission and depth imaging are required to alternate the sensor operation multiple times per second. As the system implementation of this work shows, such configuration changes are possible with a real-time sensor configuration system.

3.3 System Design Goals

This section specifies the final system design goals, which are met by the contributions of this thesis. The goals are derived by considering the application requirements and the ToF sensor characteristics. This is summarized in table 3.1. The most demanding requirement in each category is set as a goal for the corresponding contribution. This way, the contributions enable to create systems which are able to meet the requirements of all applications. Contribution C1 introduces the optical communication method, while C2 enables precise localization. The system implementations allow to experimentally verify that the goals can be fulfilled. Contribution C2 also entails several processing methods, which enable a future location-aware communication system to reach application specific requirements. The most important methods are evaluated with additional prototype implementations. The remainder of this section discusses the requirement groups (R1-R3) with regards to the ToF sensor characteristics.

	R1 Optical Communication	R2 Localization	R3 Energy efficiency
A1 - Secured Communication	5 kbit/s	5 cm / 1 Hz	low priority
A2 - Indoor Positioning	0.05 kbit/s	2 cm / 30 Hz	important
A3 - Sensor Networks	0.5 kbit/s	2 cm / 10 Hz	low priority
A4 - Augmented-Reality	0.05 kbit/s	2 cm / 30 Hz	crucial
System Design Goal	5 kbit/s	2 cm / 30 Hz	crucial
Contributions	C1	C2	C2

Table 3.1: The minimum requirements for each application A1-A4 lead to the system design goals. These goals are met by the contribution C1 and C2.

3.3.1 R1: Optical Communication

A major goal of this work is to provide methods to establish an optical communication link with a data transfer rate of at least 5 kbit/s. The quality of the received signal depends

on the sampling rate, received optical power and modulation method. The sampling rate is proportional to the data transfer rate and depends on the speed of the analog-to-digital converters (ADCs) of the ToF sensor, the exposure time and the amount of digitized pixels per sample. In the proposed approach, the number of digitized pixels is reduced to a sub-image, containing the light source. The received signal strength depends on the exposure time of the ToF sensor. A prolonged exposure time reduces the sampling rate, creating a trade-off between signal strength and data transfer rate. The modulation method has the most influence on the data transfer rate and communication range. This work proposes a modulation scheme in Section 4.1.1 which exploits the photonic mixer device for demodulation.

Another approach to optical communication with Time-of-Flight sensors is to choose a slower data transmission method in favor for less system complexity. This approach still allows for communication partner identification, which is a requirement for all potential applications. If a low bitrate is sufficient, synchronized on/off keying can be used as alternative modulation approach. This reduces the system complexity and does not require a real-time system. The reduced data transfer rate is still enough for identification broadcast, which can be sufficient for applications A2 and A4. A dedicated system design is proposed in this work, enabling to implement the use case of 3D tags for augmented-reality.

3.3.2 R2: 3D Localization

The performance of the 3D localization can be rated by the precision and update rate. Except for retroreflective communication partners, there is a trade-off between and update rate and precision. A prolonged illumination phase increases precision, but lowers the update rate.

The communication partner is localized by the ToF sensor by measuring its direction and distance. A fast and more efficient method than traditional depth sensing is introduced in Section 4.2.1. The direction can be calculated by the position of the received communication signal on the sensor array. Therefore, the signal direction can be derived during optical communication and does not require a separate measurement. Directional changes are thus updated faster than the complete 3D localization. This is useful for application A2 and A4, where changes of the direction are more frequent than distance changes. The localization precision can be further enhanced by using sub-pixel precision or sensor fusion methods [7]. As presented in Section 4.2.1, reducing the number of raw images required for a single depth image can dramatically improve the 3D localization update rates.

3.3.3 R3: Energy Efficiency

The energy efficiency is crucial in all applications with continuous sensor operation and limited energy reserves. The intended class of devices for A2 and A4 are smartphones and head-mounted AR devices. The power consumption of ToF systems largely depends on the amount of emitted light, and the required processing operations. Section 4.2.1 presents an approach to reduce the required computational complexity during depth image processing, while simultaneously increasing the signal-to-noise ratio. As shown in Publication 8, all further data processing effort is insignificant due to the proposed methods in this work.

The amount of emitted light during localization can be reduced by restricting the distance update rate in favor of a higher directional update rate. This work also proposes to combine depth sensing and data transfer in Section 4.1.1. The light pulses used for distance localization can be reused for optical communication. This benefits the applications A2 and A4 with low data transfer rate requirements.

Chapter 4

Methods and System Design

This chapter presents the contributions of this thesis. Several methods are proposed, enabling location-aware optical communication which can meet the requirements, specified in Chapter 3. As seen in Fig. 4.1, all key requirements groups R1, R2 and R3 are addressed by the contributions C1 and C2. The key goal of this thesis is to research designs and methods to create the foundation of ToF-based location-aware communication. Rather than providing a single system implementation with limited capabilities, the fundamental problems of a wide variety of future ToF-based location-aware communication systems are addressed. This removes the obstacles for future research and opens the way for future adoption.
Contribution C1 contains the methods to effectively receive and transmit data with Time-of-Flight sensors. The second major contribution C2 comprises several Time-of-Flight data processing methods, enabling robust and efficient communication partner localization. Communication and localization are combined to location-aware optical communica-

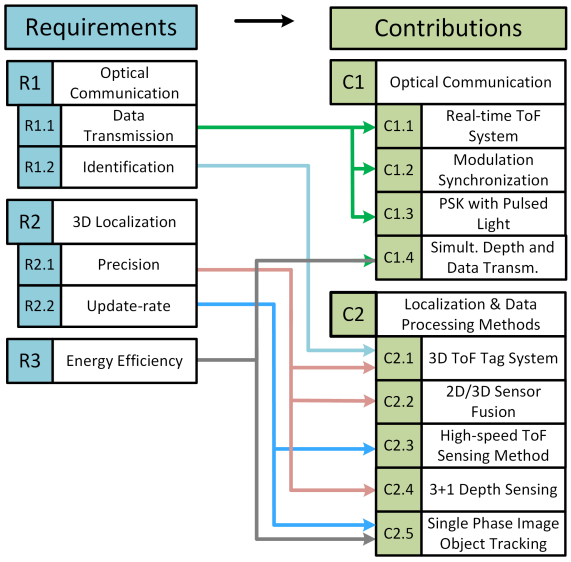


Figure 4.1: The contributions of this thesis enables future location-aware optical communication systems to meet their localization and communication requirements.

tion by several design proposals and also a prototype implementation. The implemented systems furthermore allow to conduct an experimental characterization, which is presented in Chapter 5. An FPGA-based real-time system includes the methods to receive data with a ToF sensor from modulated light sources. The system proofs the feasiblity of the proposed solutions for important data transmission challenges (R1.1). The implemented ToF Tag localization system provides a unique method to fulfill the requirements R1.2 and R2.1., and enables to provide an implementation of the application of 3D tags for ToF sensors. This system is later used in chapter 5 to evaluate a potential use case for this location-aware communication link.

4.1 Optical Communication with Time-of-Flight Sensors

The first contribution C1 of this thesis is a novel optical communication approach with Time-of-Flight sensors. The proposed methods enable ToF sensors to emit and receive data amd meet the data transmission requirement R1.1. A prerequisite is modulation frequency synchronization, which is introduced in Section 4.1.1. Section 4.1.1 discusses the modulation method of phase-shift keying of pulsed light, which repurposes the photonic mixer device for demodulation. Since location-aware optical communication also require ToF sensors to capture depth images, a method for simultaneous data transmission and depth sensing in presented in Section 4.1.1. Section 4.1.2 finally presents a system design, which shows how location-aware optical communication with Time-of-Flight sensors can be implemented with a real-time system.

4.1.1 Concept and Methods

Modulation Frequency Synchronization

Modulation frequency synchronization is a necessary step for the developed optical communication method. In the proposed synchronization method, the pixel reference signal F_{ref} of the ToF sensor is synchronized to the signal of the incoming light pulses F_p. In

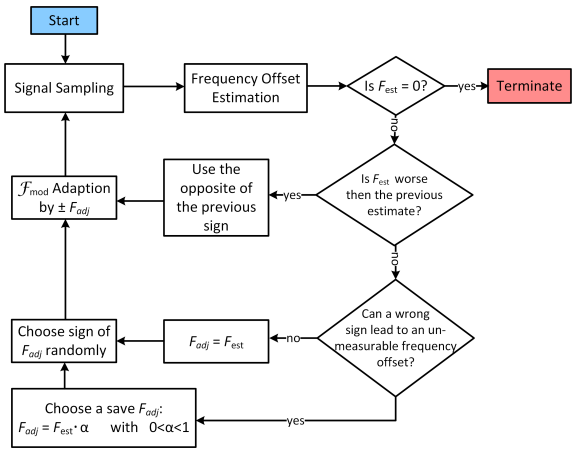


Figure 4.2: The algorithm for modulation frequency synchronization. Reproduced with changes from [1].

the system design, which is proposed in Section 4.1.2, the synchronization mechanism is implemented in the ToF operations layer as one of the fundamental ToF operations. During the synchronization phase, the ToF sensor is configured to sample an incoming signal at a constant pixel reference frequency F_{ref}. A certain number of samples are collected by averaging the pixel values of the received signal in each phase image. The sampled signal oscillates with the aliasing frequency F_a, as seen at the top of Fig. 4.3. F_a is the frequency difference between the reference frequency F_{ref} and the received light pulse frequency F_p. This frequency difference is minimized during the synchronization by changing the PLL parameters of the ToF sensor. Since only the absolute value of the frequency difference is available, the system can arbitrarily decide wether frequency difference is positive or negative. This procedure illustrated in Fig. 4.3. Fig. 4.2 shows the developed iterative algorithm, where the sign of F_a is arbitrarily assumed in the first step. If the guessed sign was incorrect (see step 2), the control step went in the wrong direction and the frequency

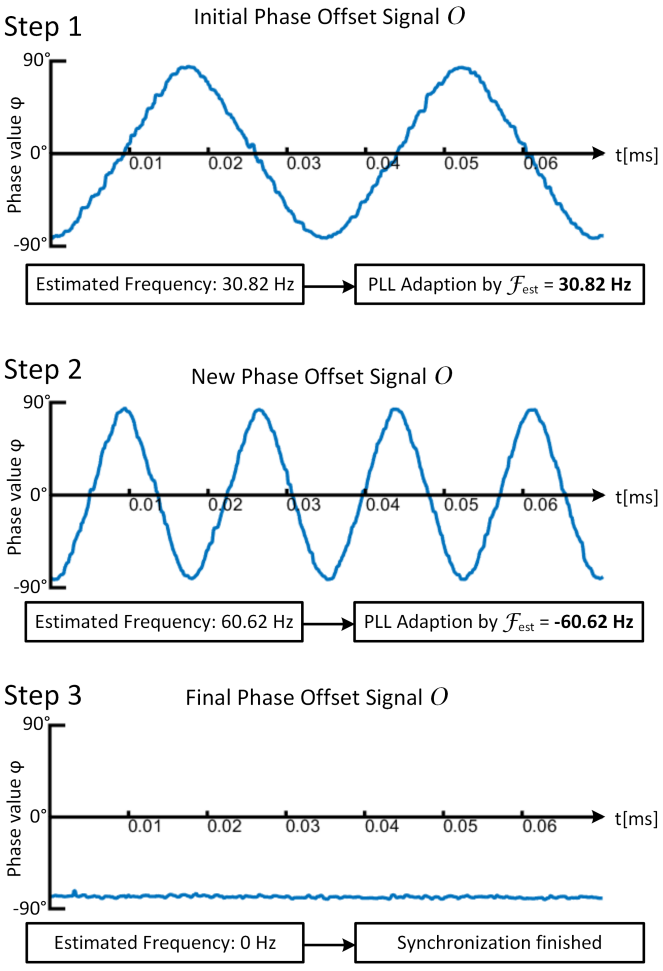


Figure 4.3: Example of the modulation signal synchronization procedure. The initial aliasing frequency is estimated in the first step. The second step reveals that the arbitrarily assigned sign of the estimated frequency F_{est} was incorrect, and therefore a PLL adjustment of two times the negative F_{est} leads to synchronization. Reproduced with changes from [1].

of F_a increases. The algorithm can now adjust F_{ref} accordingly and measures the remaining F_a, until the remaining frequency difference F_{error} is below a tolerable threshold. The results of this method are presented in Chapter 5.

Phase-Shift Keying with the Photonic Mixer Device

This doctoral thesis and the supervised master thesis of Theresa Egger [38] propose phase-shift keying of pulsed infrared light as modulation method for ToF sensors. The received signal can be demodulated with Time-of-Flight sensorsby exploiting the properties of the photonic-mixer device. This enables to receive and transmit multiple bits per symbol, enabling ToF sensors to fulfill the data transmission requirement R1.1. As introduced in Section 3.2, the functionality of the photonic mixer device can be interpreted as correlation operation. The received light signal F_{Light} is correlated on each pixel with the pixel reference signal F_{ref} of the ToF Sensor. The result of the correlation is continuously integrated during each exposure. This integrated correlation result is the phase-value p and is obtained by the charge difference between capacitor A and B. Due to the correlation, the phase value depends on the phase difference between F_{ref} and F_{Light}.

After synchronization, the communication partner emits groups of light pulses which follow a reference frequency. When data is modulated onto this signal, a group of pulses is phase-shifted in relation to the reference frequency. The ToF sensor is able to use the PMD to measure the correlation between these light pulse groups and its own pixel reference signal.

Because the pixel reference signal is synchronized to the reference frequency, it is possible to demodulate the contained information with additional steps. This work chooses to capture two phase images to obtain the phase-difference of a transmitted symbol from the reference frequency. This is illustrated in Fig. 4.4. The goal is to measure the phase angle φ_1 of phasor P_0, which can be seen in the phasor diagram. The orientation of the phasor φ_1 contains the transmitted data. Determining the angle of P_0 can be accomplished by capturing two phase values p_0 and p_{90}. Due to the correlation operation, the measured phase value is the projection of the phasor P_0 on the imaginary axis. The second phase value $p_{90°}$ is captured at a 90° phase offset between F_{ref} and F_{Light}. This enables to determine the real and imaginary part of the phasor P_0. The necessary phase offset can be created either by the ToF sensor by shifting F_{ref} or when the communication partner shifts its transmitted signal F_{Light}. In the implemented system, the evaluation transmitter transmits a symbol in two groups of light pulses, where the second group is shifted by 90°. The resulting phase values p_0 and p_{90} are used to calculate the phase angle φ_1.

$$\varphi_1 = atan\left(\frac{p_{0°}}{p_{90°}}\right) \tag{4.1}$$

Relative phase-shift keying is used to encode the transmitted information. The transmitted phase angle φ is thus encoded in the difference of the subsequently received phase angles φ_1 and φ_2. Receiving data with this modulation method is part of the ToF operations layer in the software design. Most ToF sensors are able to transmit data with this method. This enables efficient communication between ToF sensors without hardware modifications.

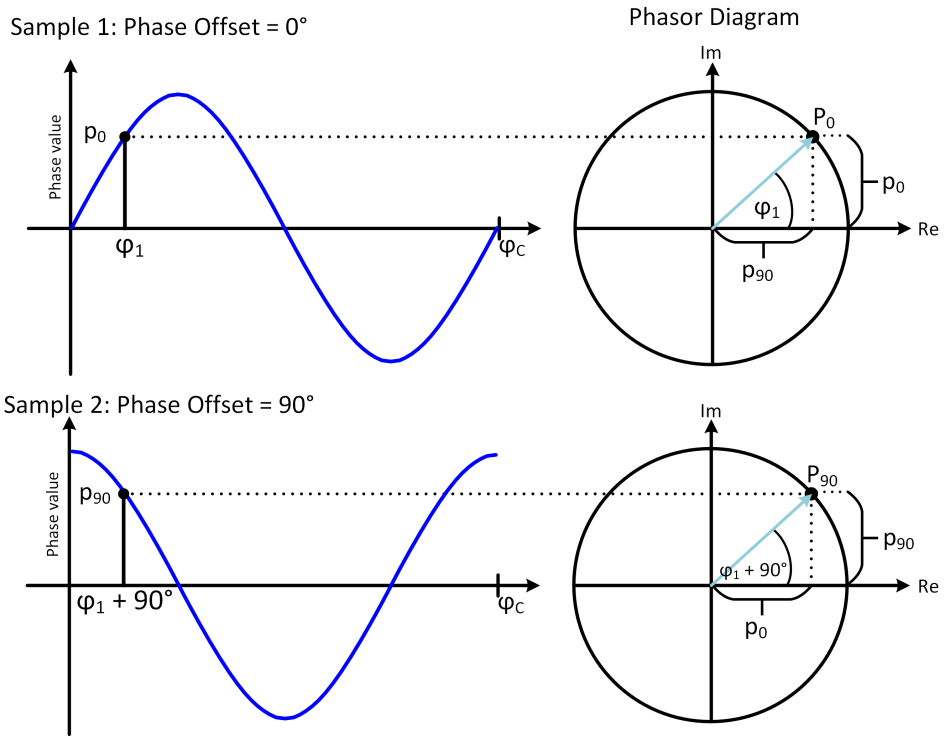


Figure 4.4: Taking two subsequent samples with 90°phase offset enables to measure the transmitted phase φ_1.

Simultaneous Depth Sensing and Data Transmission Method

This section introduces a design suggestion which can be implemented with future ToF sensors. As the previous section showed, data transmission and depth sensing reduce the received data rate. Furthermore, dedicated data transmission light pulses consume energy and are restricted by eye-safety regulations. It is, however, possible to combine PSK data transmission and depth sensing into a single sensor operation. This increases the energy efficiency, which is an important requirement (R3) in certain applications. The basic principle is that during the exposure of a continuous-wave ToF phase image, the phase of the light pulses and also the pixel reference signal are both PSK modulated. This principle is illustrated in Fig. 4.5. In the proposed phase-shift keying method, each individual symbol consists of a group of light pulses. The light pulses are shifted relative to their reference signal. These phase-shifted pulse groups can also be transmitted during depth sensing. If the pixel reference signal is also shifted, the ToF sensor receives the reflected light pulses with the same phase shift as in normal depth sensing. One could argue that this method prolongs the pixel exposure time, since the introduced phase shift must not interfere with the reflected light pulses from the previous pulse. The sum of the delays t_a is however minuscule and can be calculated using the symbol rate R_S, the exposure time t and the modulation frequency f_M.

$$t_a = \frac{R_S \cdot t}{f_M} \tag{4.2}$$

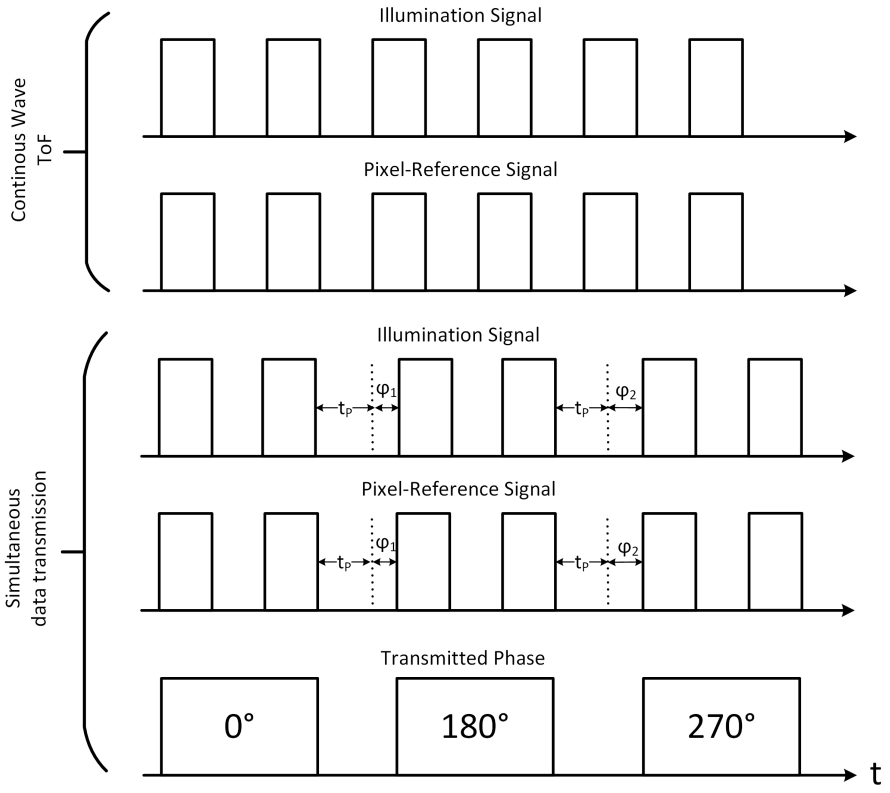


Figure 4.5: It is possible to transmit PSK modulated data during depth sensing. Pulse groups in the pixel reference signal and the illumination signal are simultaneously phase shifted.

In a realistic example, 7300 symbols are transmitted per second, the exposure time for a phase image is 2 ms and the frequency of the modulation signal is 80 MHz. In this example, a 8-phase long-range ToF depth image has an increased exposure time of just 0.18 μs which is just 0.0091% longer than the original exposure time.

4.1.2 System Design for ToF-based Optical Communication

The purpose of this system design is to provide the necessary mechanism and sensor operations to establish a location-aware communication link with a ToF sensor [4]. Major challenges are to establish an efficient modulation method, conduct modulation signal synchronization and to combine depth sensing with optical communication. The developed methods need to be able to support a second ToF sensor or multiple ToF tags as potential communication partners. As seen in Figure 4.6, the system is organized in a layered architecture, starting with the ToF and FPGA hardware as the physical layer. With this structure, the complexity of the system can be divided between the layers. Each layer features well-defined software or hardware interfaces. Data exchange is restricted to neighboring layers and only higher layers can control the behavior of the lower layer.

The hardware layer consists of a Time-of-Flight sensor which is connected to an FPGA system. The FPGA also features a dedicated CPU, directly connected to the programmable logic. The hardware layer provides drivers for the involved IP blocks as

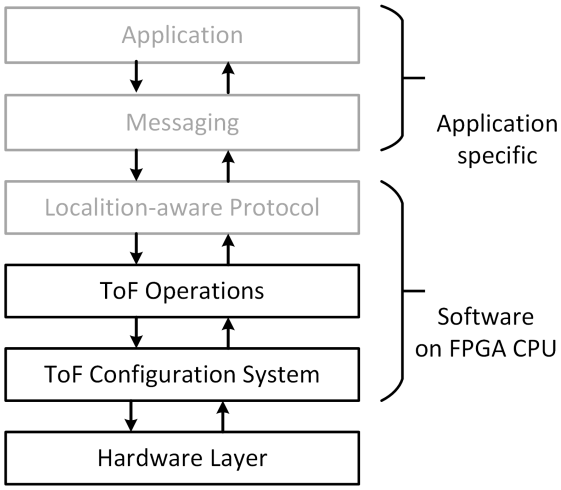


Figure 4.6: The optical communication system is organized in a layer architecture and implemented on an FPGA with integrated CPU.

interface to the upper layer. These drivers allow to configure and control the ToF sensor, and provide the captured ToF data. The ToF configuration system layer is the lowest software layer and is executed on a CPU which is integrated in the FPGA Package. This layer ensures that all ToF operations and configuration changes happen with the correct timing. With these interleaved configuration changes, it is possible to repurpose ToF sensors from depth sensors to location-aware optical transceivers. During communication and tracking, it is regularly necessary that the ToF sensor behaviour changes rapidly during operation. The ToF operation layer therefore can execute ToF operations, which are a series of ToF measurements with subsequent processing operations. This layer can adjust the sensor configuration depending on the image data. The processing operations in this layer involve sensor synchronization and signal decoding. These major contributions are further discussed in Sections 4.1.1 and 4.1.1.

A location-aware protocol completes the functionality of a location-aware communication system. The design of such a protocol depends on the intended applications and their requirements. General design guidelines to implement such protocols are presented in Section 4.1.2.

ToF-FPGA Hardware System

In order to explore and implement optical communication methods on ToF sensors, the system needs to control the behavior of the sensor with real-time guarantees. This requires a direct connection to the sensor to receive data directly from the ToF sensor without buffering delay. With these requirements, the Xilinx Zynq 7000 system was selected for this work. This system-on-chip (SoC) consists of an FPGA, 512 MB RAM and two Cortex A9 ARM CPU cores. As Time-of-Flight sensor, an evaluation board was selected, which features an Infineon REAL3$^{\text{TM}}$ sensor based on Time-of-Flight technology of pmdtechnologies. The sensor features a 352x288 resolution and two OSRAM SFH4715s LEDs as illumination source. As shown in Fig. 4.7, the ToF sensor is connected to the Zynq 7000 via a parallel interface and an I2C connection. The image data from the Tof sensor is

transferred to the programmable logic of the Zynq 7000 via a 12bit parallel data interface. A video direct memory access (VDMA) IP block stores this data on the main memory. The software on the ARM processor implements the proposed software design. The main purpose of ToF sensors is to provide a continuous stream of phase images. In this system, the operating mode of the sensor needs to be frequently changed during runtime. This requires to change certain sensor registers with precise timing. A GPIO line of the ToF sensor is therefore connected to the FPGA platform. The GPIO is routed to an IP block in the FPGA which can trigger an interrupt on the ARM processor. This interrupt synchronizes the ToF sensor operation with the FPGA system. The optical communication system therefore can rapidly change the ToF sensor behavior.

An evaluation transmitter serves as an evaluation communication partner for the system. A microcontroller controls a custom circuit, which emits PSK-modulated light with an SFH4715 LED. The device can either be configured to emit synchronization pulses or a predefined data pattern for evaluation.

Real-Time Time-of-Flight Sensor Configuration Layer

The main purpose of the real-time ToF configuration layer is to adjust the ToF sensor behavior during operation. It is the base layer of the FPGA CPU software, and abstracts the sensor configuration details and timings to the ToF operation layer. The structure is illustrated in Fig. 4.8 as an UML class diagram. All sensor control operations are executed in the main loop in the ConfigurationSystemLayer class. The class interfaces with the FPGA via the drivers of the IP blocks. If requested by an upper layer, the class writes the sensor configuration changes via I2C. When the corresponding image data is received, a callback function in the ToF operation layer is invoked. As shown in Fig. 4.6, the complete communication system consists of a layered architecture. Besides the lower hardware layer, the real-time ToF configuration layer only interfaces with the upper ToF operations layer. This interface consists of a shared data structure (ConfigurationSystemData) and a control interface class (ConfigurationSystemInterface). All data is exchanged between the layers by mutually accessing the shared data structure. The upper layer in the

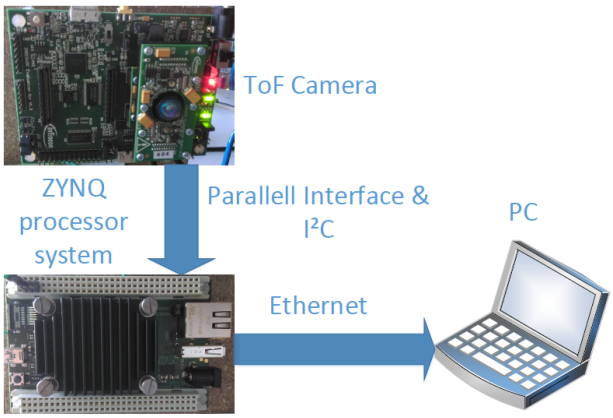


Figure 4.7: The ToF-FPGA hardware system consists of a ToF sensor, which is connected to a Zynq 7000 FPGA via I2C and a parallel data interface. Reproduced with changes from [4].

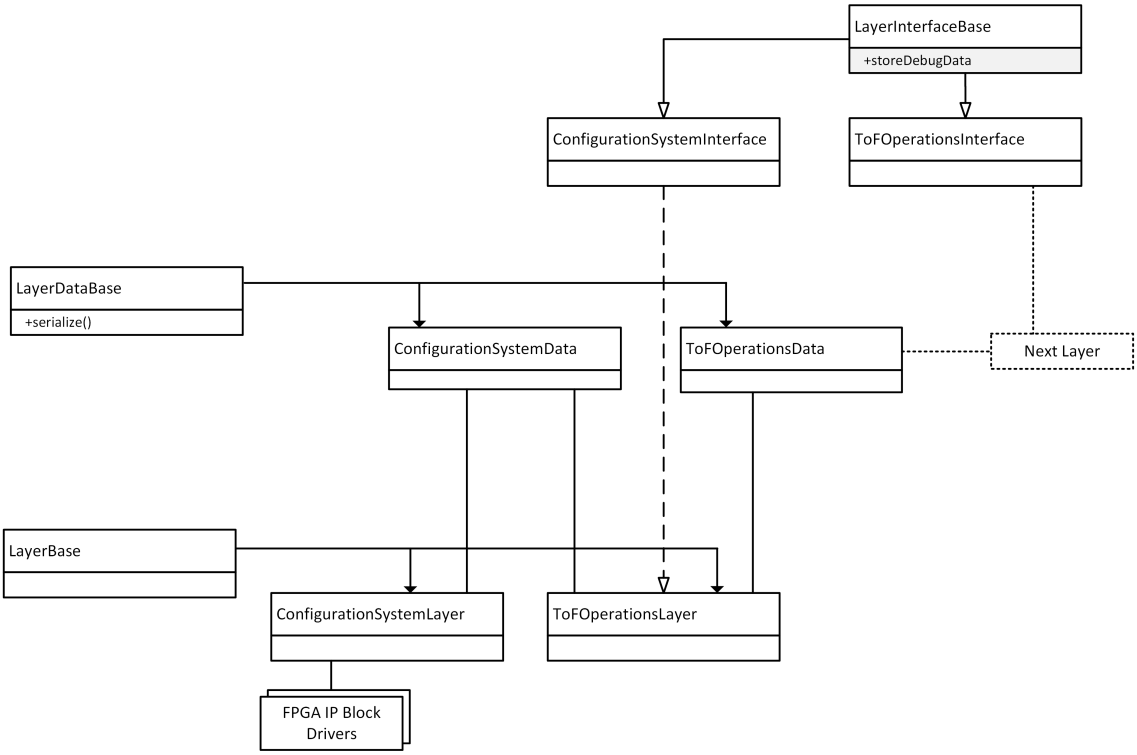


Figure 4.8: UML class diagram of the ToF configuration layer.

communication system can change the behavior of the lower layer via the control interface class. This encapsulation enables to store the exchange between the layers at dedicated timestamps. Step-by-step debugging is not possible is such a real-time system, since halting the CPU does not halt the ToF sensor nor its communication partners. Therefore these interface classes inherit debugging and data dumping functionality (LayerDataBase, LayerInterfaceBase). This way it is possible to record the data exchange between all layers enabling to recreate potential problems offline in a testbed.

During operation, the real-time ToF configuration layer is regularly ordered by the upper layer to change the ToF sensor configuration. These configuration changes involve the exposure time, phase offset, modulation frequency and the pixel region-of-interest. In order to configure the sensor behavior during operation, a GPIO of the ToF sensor is connected to the FPGA. The GPIO is configured to indicate when it is possible to change the ToF sensor configuration. The system transfers the new configuration in this precise moment. When the ToF sensor captures the new sets of images, the resulting data is transferred to the RAM via an FPGA IP block. The ToF configuration system layer then invokes the callback method of the upper layer when the data is available.

Location-aware Time-of-Flight Sensor Operations

In this system design, the location-aware ToF operations are the basic building blocks for any future ToF-based location-aware communication system. At the start of a ToF operation, the system captures either a single or a series of phase images with a certain

sensor configuration. The operation then processes the data and immediately reacts to the extracted information. These ToF operations are triggered by a control logic in the protocol layer. The order and frequency of the executed ToF operations are defined by the protocol and depend on the demands of the application. When a ToF operation is triggered, the ToF operation layer configures the ToF sensor, the sensor performs its operations, and a processing step is executed. In this processing step, the system decides if it needs to react to the received image data or repeat the operation. In certain operations, it is necessary to change the configuration of the ToF sensor and to immediately repeat the operation.

As seen in Fig. 4.9, the system design includes four different ToF operations. A basic ToF operation is to capture a single phase image with a specified exposure time, phase offset and modulation frequency. Multiple images can be used to create a depth image. A single phase image can also be used to locate the signals of all visible communication partners. The pixel positions of the light sources are the input for the receive data operation. In this operation, the ToF sensor continuously digitizes a sub-image containing the light source for a certain amount of samples. Reducing the region of interest enables a high sampling rate, compared to a complete sensor readout process. The processing part of the ToF operation also decodes the received signal by using the method presented in Section 4.1.1. The pixel position also needs to be tracked to update the region-of-interest on the sensor. As shown in Section 4.1.1, modulation signal synchronization is crucial for ToF PSK communication. In the synchronization operation, the ToF sensor samples

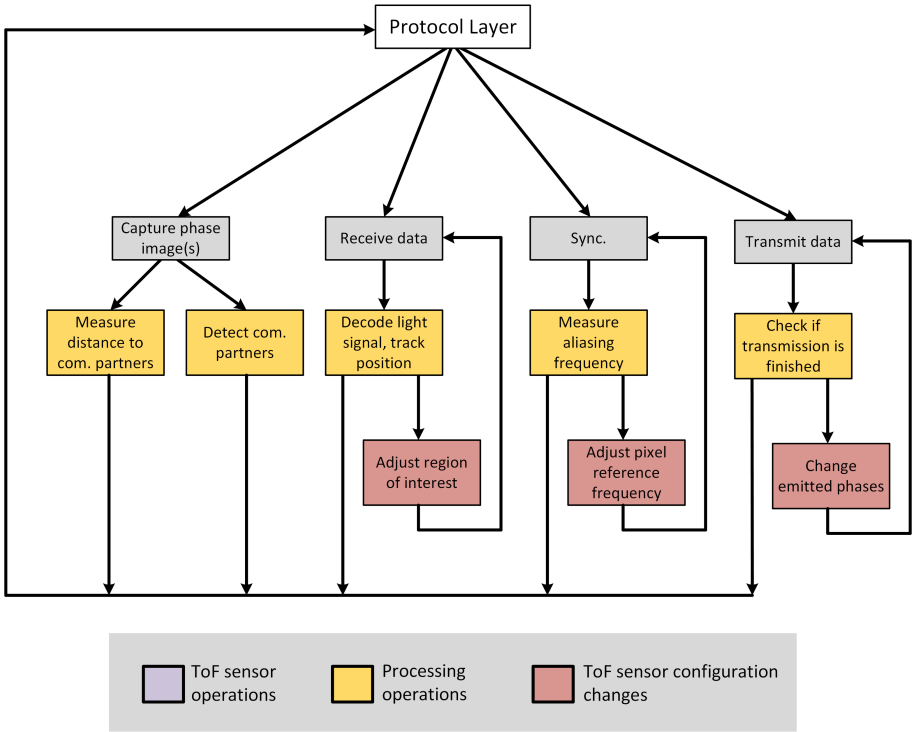


Figure 4.9: ToF operations are the building blocks of a location-aware communication protocol. Each operation captures a series of phase images, followed by a processing step and optional sensor configuration adaption.

the light source similar to the receive data operation. The processing logic measures the aliasing frequency and adjusts the modulation frequency of the ToF sensor according to the synchronization algorithm. This process is repeated until the frequency difference is tolerable. The transmit data operation captures pseudo images in which the phase of the outgoing light pulses in encoded. This operation is not limited to transmitting data, it can also be used to transmit a synchronization signal to the communication partner. Besides modulation signal synchronization, the communicating systems might also require time synchronization. The received light signal can be used by a system to synchronize its response. In Chapter 4, a system design is introduced which uses this mechanism to synchronize the communication partners.

Design Suggestions for Location-aware Optical Communication Systems

The functionality of a location-aware optical communication system with ToF sensors is defined by its application. Since localization and communication are conducted with a single sensor, this causes a trade-off between the update rate, localization precision and the data transfer rate. Parameters such as localization range and update rate also have vastly different characteristics in different applications. Therefore a location-aware communication protocol needs to be adapted to the intended application. The protocol defines the ToF sensor operations, and specifies the behavior of all involved participants. On a ToF system, such protocol can be implemented as software layer between the data exchange layer and the ToF operations layer. Depending on the specified protocol characteristics, the ToF operations are executed in a certain pattern, as shown in Fig. 4.9. The frequency and duration of each ToF operation can then be adapted to the demands of the application.

A basic implementation of the protocol would synchronize and then regularly localize, receive and send data to multiple communication partners. The protocol would start by invoking the send data operation to broadcast a request to all communication partners to continuously emit pulsed light. The next ToF operation captures a phase image which enables to directionally localize the signal source of each partner. A synchronization operation is then executed for each communication partner to synchronize the modulation frequencies. After the synchronization, the system enters the send data operation to transmit or to broadcast data. The data might contain a request to receive certain information about the communication partners like their identity. The communication partners are constantly listening and respond to this request with a specified delay. The ToF sensor on the other hand immediately enters the receive operation, which enables to sample the received signal. To determine the distance, the ToF sensor takes a depth image by executing the phase image operation multiple times to capture the required phase images. With the measured distances to the communication partners, it is possible to determine their 3D position. The complete process is then continuously repeated, enabling location-aware ToF communication with multiple communication partners.

High data rate applications such as secured communication might reduce the frequency of 3D localization operations, and increase the receive and transmit data operations. Localization systems might only use the transmit data operation to trigger the communication partners to transmit their IDs. The capture phase image operation is more frequently executed in a localization system to track the 3D positions. 3D tags for augmented reality

systems is an application in which the direction to the communication partners requires more frequent updates than the distance. Therefore more single phase image operations are executed and less complete depth images are captured. The capture phase image operation can also be used to remain compatible with normal depth sensing systems, by providing a depth image stream to the rest of the system. Section 1.2 introduces methods to improve the proposed design, so that it can fulfill the demanding requirements of all applications. These methods can be implemented in form of processing methods in the ToF operations layer, or can be utilized by the protocol to reduce the amount of certain ToF operations.

4.2 Localization and Time-of-Flight Data Processing

The second contribution C2 of this thesis consists of methods to improve communication partner localization. Precise, robust and energy-efficient localization requires additional data processing methods. Section 4.2.1 describes a sensor fusion method, fusing color images with Time-of-Flight depth data. This decreases noise and improves the resolution of the depth images. The localization of moving communication partners might be impaired by motion artifacts. An alternative ToF sensor calibration method is presented in Section 4.2.1, which avoids the effect of motion artifacts. Section 4.2.1 introduces the idea to combine pixel binning with image-guided upscaling. This can reduce the SNR of the depth image, while preserving the resolution. A method for 3D object and communication partner tracking with single phase-images is introduced in Section 4.2.1. Using just a single phase-image for distance estimation reduces the energy consumption significantly. Section 4.2.1 introduces a concept to determine the orientation of the communication partner, which leads to 6-dof localization. In the final Section 4.2.2, a localization system is presented, which is able to localize and identify 3D tags with Time-of-Flight sensors.

4.2.1 Concepts and Methods

In ToF-based location-aware optical communication, the ToF sensor localizes the communication partner by combining a separate direction and distance measurement. The direction can be determined by the pixel position of the light source p_{ToF}. The distance d is obtained by conventional depth imaging, or by using the methods provided in this section. The direction and distance to the communication partner enable to calculate the relative 3D position P_{ToF}. With a calibrated ToF sensor, it is possible to calculate P_{ToF} with the pinhole camera model. The pseudo-inverse $\tilde{C}_{ToF}$ of the ToF camera matrix enables to calculate the direction vector, which is then multiplied by the distance d.

$$P_{ToF} = d \cdot \frac{\tilde{C}_{ToF} \cdot p_{ToF}}{\|\tilde{C}_{ToF} \cdot p_{ToF}\|} \tag{4.3}$$

The presented methods in this section focus on the energy efficiency (R3), precision (R2.1) and the update-rate (R2.2) of the localization. The feasibility of integrating these methods in a ToF-based location-aware communication system depends on the requirements of the intended application. Some of the proposed methods have additional applications which are further investigated in the corresponding publications in Chapter 7.

Time-of-Flight and Color Sensor Data Fusion

Electronic devices with ToF sensors are often also equipped with color cameras. The edge information in color images can be used to decrease the noise and to increase the resolution of the depth image. Edges in high-resolution color images often correspond with edges in depth images. In certain applications for location-aware communication, the localized position is mapped to a color images. If the surrounding depth pixel values are also mapped to the color image, it is possible to improve the depth quality with color information. This improves the distance estimation and consequently the accuracy of the communication partner localization. Besides the improved noise, the increased depth resolution can help to distinguish the communication partner from the background. These improvements contribute to the requirement of precise communication partner localization (R2.1). Other applications such as the artificial bokeh effect in computational photography or 3D scanning can also benefit from a depth image which is mapped 1-1 to a high-resolution color image.

The proposed approach introduces an image-guided upscaling method, that can be executed on mobile devices [39]. In contrast to previous work, the proposed algorithm is optimized to be efficiently executed on GPUs. In the first step, the depth values are mapped into the color camera image space, using the same method as introduced in Section 4.2.2. The mapped values are then sparsely associated with the closest element of a 2D array datastructure. The array has the same resolution as the color image. For each of these mapped depth values, there is a processing program executed in parallel on the GPU. The idea of the algorithm is that each depth value spreads its influence among the surrounding pixels on the array. As seen on Fig. 4.10, the program parses therefore all radial pixel paths on the color image from the depth value position. For each of the color image pixels, an influence value is placed on the array. To incorporate information from the color image, the influence value decreases exponentially each time the parsed radial path crosses a color image edge. After the influence of each depth value is distributed among the buffer, the depth values are interpolated. Depending on the adjustable path length, multiple depth values placed their influence for each pixel in the array. The influence values serve as interpolation weight. This results in a depth image with increased resolution, because hardly any depth influence is propagated over color image edges. The depth noise is also reduced because the interpolation averages depth values on the same surface.

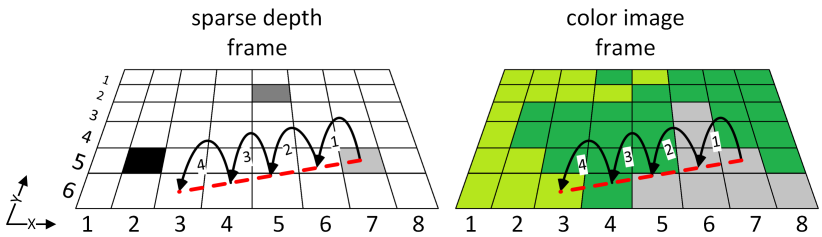


Figure 4.10: The core of the sensor fusion algorithm. Sparsely mapped depth values distribute an influence value within a certain area. Edges in the color image exponentially decrease the influence.

High-speed Time-of-Flight Sensing

Conventional ToF sensors capture at least four different phase images in order to create a single depth image. As illustrated in Fig. 4.11, object movements relative to the camera can cause motion artifacts. The reflection of an object arrives at different sets of pixels in each phase image. When calculating depth images, the same pixel in each phase image is used to calculate the depth image. The motion artifacts occur, when these phase values do not stem from the same object. These artifacts can completely invalidate distance measurements of small objects, which can be a problem when localizing communication partners. The proposed method helps to fulfill the requirement of precise localization R2.1 by compensating for such motion artifacts.

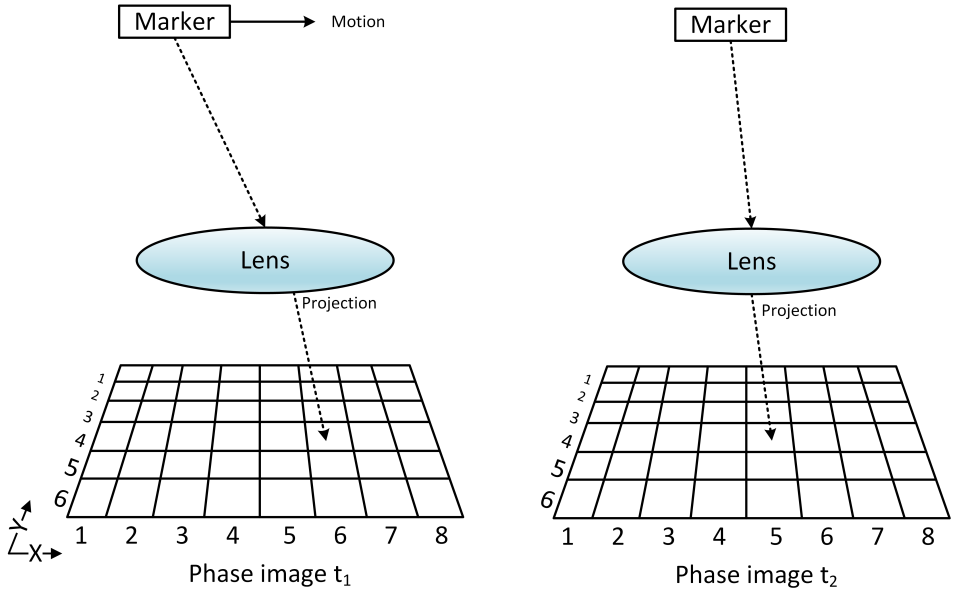


Figure 4.11: Motion artifacts are most severe when sensing small object such as markers. Motion causes the object to be projected to a different set of pixels in the subsequent phase images t_1 and t_2.

A calibration method is therefore proposed in Publication 7, which enables to measure the 3D position of small and fast moving objects. Conventional ToF sensing can lead to motion artifacts, because the object projected to different pixel positions in the captured phase-images. Tracking the object does not help with conventional calibration, because a depth value needs to be calculated from the same pixel in the phase-images. The reason is that there are pixel and ADC specific error sources which influence the output of each pixel. A major culprit is also the uneven light distribution of the illumination unit and the lens.

The introduced calibration method allows to track small objects, and use different phase-values from different pixels to calculate the distance. The first step is to provide a phase value correction function. The goal is that the sensor produces the same phase-value, regardless of the pixel position of the object. This function has certain parameters, which are obtained by a novel calibration procedure. A ToF sensor is therefore mounted on a rotation platform. Several reference images of a 10x10 mm retroreflector are captured

at different sensor orientations. These reference measurements are then interpolated to derive a calibration parameters for each individual pixel. In the second calibration step, the parameters are derived to remove the remaining depth estimation error. The ToF system is mounted on a linear stage, facing a planar target. Several phase images are captured at different distances. A distance correction factor is derived for each adjusted distance. The sensed depth value is then corrected by an interpolation of this factor. This calibration step removes the ToF wiggling error, and the distance distortion, which was introduced by the first calibration step. The method is best suited for fast indoor localization or motion tracking systems. As the evaluation showed, the calibration approach leads to an additional distance error of up to 3 cm. This enables to preserve the localization during fast motions albeit with decreased positioning accuracy.

Binned Phase Image Upscaling

Precise communication partner localization is crucial for applications like indoor positioning and augmented reality. Since ToF sensors actively emit light, the depth sensing precision decreases over distance. Pixel binning is a common method for image sensors to increase their sensitivity. By combining a number of neighboring pixels, it is possible to increase the SNR at the cost of decreasing the image resolution. The proposed method can combine the advantages of pixel binning with a minor resolution trade-off. Depth images with lower SNR increase the localization precision (R2.1).

In conventional ToF imaging, at least four phase images are required to calculate a single depth image. Fig. 4.12 illustrates the proposed method. Three of four phase images are captured with pixel binning, and a forth is captured conventionally at the original sensor resolution. A binned version of the forth image is then created digitally by

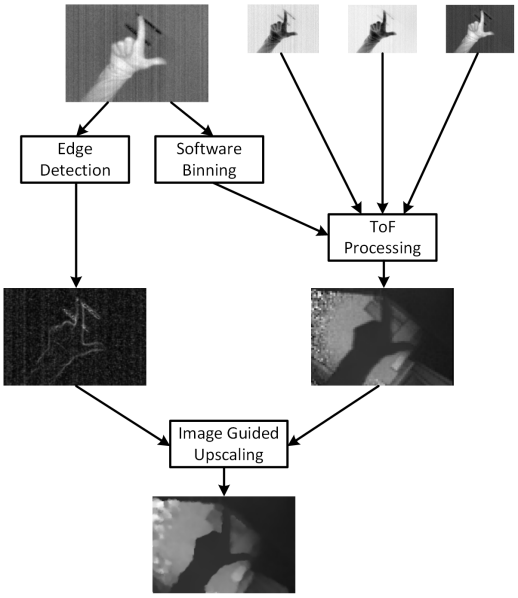


Figure 4.12: Pixel binning reduces the noise of a depth image at the cost of reduced resolution. By capturing a high-resolution phase image, it is possible to use the edge information as guidance to produce an improved high-resolution depth image.

the processing system. These four low resolution binned phase images are used to create a low resolution depth image. The original high-resolution phase image can be converted to a high-resolution edge image, using an edge detection filter. This edge image is then used as a guidance to upscale the low resolution depth image. The color and ToF sensor fusion algorithm, which was introduced in Section 4.2.1, can be applied here. The resulting image features a comparable resolution as a conventional ToF image, but the binned pixels and the upscaling method effectively de-noise the image.

3D Object Tracking with Single Phase Images

Location-aware ToF communication systems continuously capture depth images of the communication partners for localization. The proposed method in this section removes redundancy by enabling to calculate the distance to the communication partners with just a single phase image. Reducing the number of captured phase images significantly decreases the energy consumption (R3) of the system. Alternatively, the localization update-rate (R2.2) can be increased, since less light is emitted per distance measurement. The proposed method tracks the 2D location of an object or communication partner and obtains an average phase value of the object. During an initialization phase, the ToF sensor uses conventional ToF sensing to locate the object. The corresponding phase images are processed to obtain the average phase value. The individual phase values are combined with the measured distance and are stored in a lookup table. This later allows to obtain the distance to the object with just a phase-value. If sufficient phase-distance pairs are captured, the system selects an appropriate phase offset and starts to localize the communication partner with single phase images. As seen in Fig. 4.13, the discrete values in the lookup table can be interpolated by fitting a polynomial. This allows to input a measured phase value and to retrieve a distance. During tracking, the object must never cross the peak of this phase value function to avoid ambiguity. An algorithm is therefore proposed in Publication 8, which derives the best suitable phase offset from several different lookup tables.

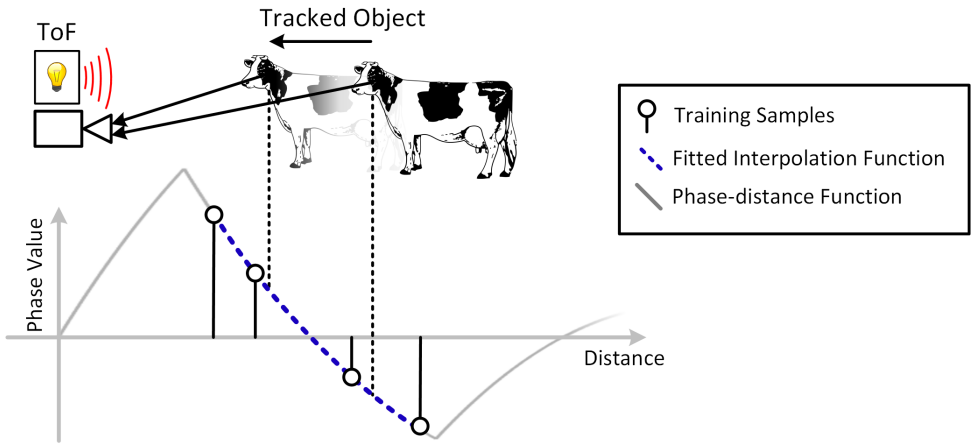


Figure 4.13: If phase values of an object are associated with the distance during a training phase, it is possible to estimate the distance to the object with a single phase image.

Tracking the 3D position of an object with significant energy savings is also valuable for general 3D object tracking. With the increasing availability of ToF sensors in smartphones, this method can potentially enable ubiquitous 3D face tracking at low energy consumption. As the experimental evaluation shows, it is possible to also use this method for 3D face tracking with an additional distance error of below 1 cm.

6-Degrees-of-Freedom ToF Tag Localization

This section proposes a concept to extend ta ToF tag system to also derive the rotation parameters of a ToF tag. This enables 6-degrees-of-freedom (6-dof) localization and can lead to a number of new applications for ToF tags. In augmented reality use cases, it enables to display a 3D object at the tag position. In indoor positioning, it is possible to derive the direction and location of a ToF system by just sensing a single available beacon. This can dramatically reduce the number of required beacons and consequently the complexity and implementation effort of such system.

The output of the proposed method is a geometric transformation, which can be applied to any point in the tag coordinate system C_{Tag}, transforming it to the Time-of-Flight camera coordinate system C_{ToF}. Such transformation is usually described with translation vector $\vec{t}$ and rotation matrix R. This rigid transformation covers all 6 degrees of freedom (6-dof). The translation vector is the vector between the 3D position of the ToF tag and the ToF sensor and can be determined with normal ToF tag localization. In order to obtain the rotation matrix R, it is first necessary to determine the direction vector from the tag to the ToF sensor. This can be accomplished by mounting at least three different light sources in close proximity on the tag, as seen in Fig. 4.14. Any suitable light source (e.g. radial symmetric VCSEL or LED) shows such a directional light emission characteristic. This characteristic can be enhanced by mounting a barrier between the light sources. The ToF sensor now needs to measure the light intensities L_1, L_2, L_3 of each different light source by taking multiple phase images. Therefore the tag activates one LED at a time, so that a single LED is visible in each captured image. The differences of these intensities are used to calculate the direction vector d. This vector originates from the center of the light sources P_C and points to the optical center U_C of the ToF sensor. The ToF sensor

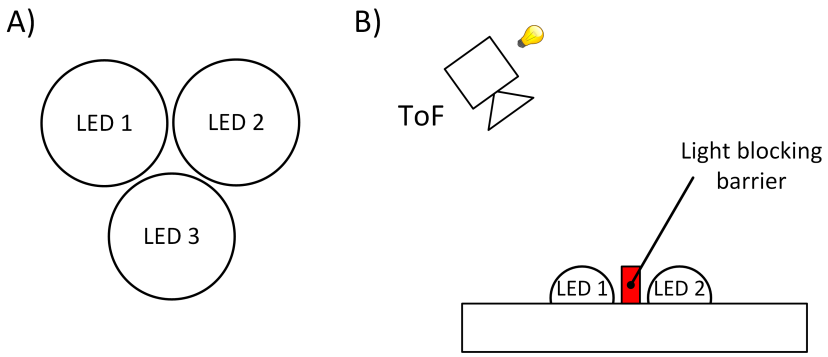


Figure 4.14: A) At least three light sources are mounted in close proximity in the proposed 6-dof ToF tag. B) A barrier between the light source can increase the directional light emission characteristics.

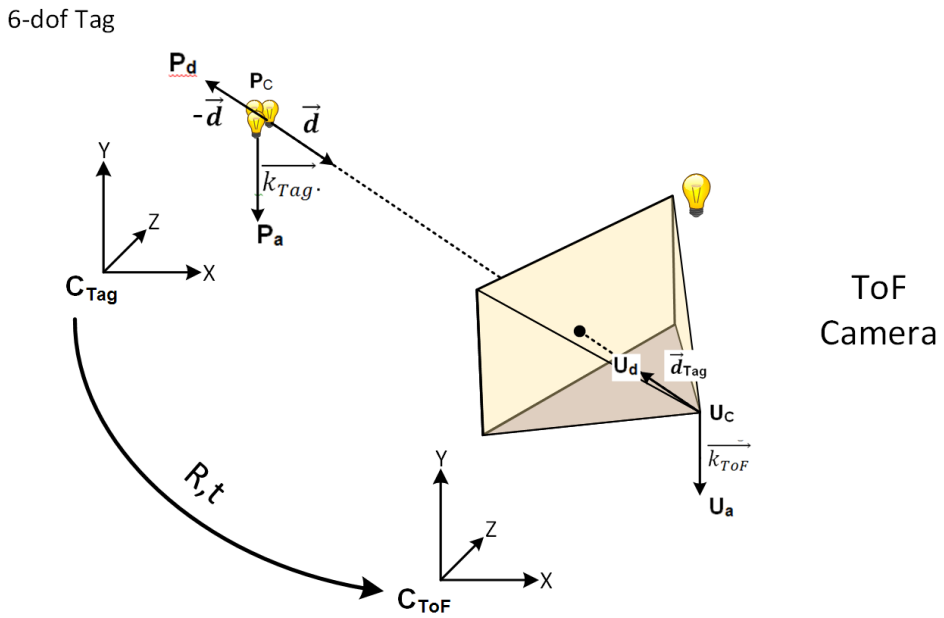


Figure 4.15: The geometic relation between a ToF sensor and 6-dof ToF tag. At least 3 points need to be determined in each coordinate system to obtain the rotation matrix R.

now needs to measure direction vector $\vec{d}_{Tag}$ from its origin to the tag. This directional localization is accomplished by using the pixel position of the projected light sources and calibrated camera parameters.

Fig. 4.15 shows the geometric relations of the tag localization. Calculating the rotation matrix R between the two coordinate systems requires a set of at least 3 points in each coordinate system. The relative geometric relation between the points in each coordinate system needs to be identical. The first pair of points is given by P_C and U_C. The normalized direction vectors can be used to obtain the second pair of points P_d, U_d:

$$P_d = P_C - \vec{d} \tag{4.4}$$

$$U_d = U_C + \vec{d}_{Tag} \tag{4.5}$$

Additional information about the orientation of the tag and the ToF sensor is required to obtain the last set of points. The vectors $\vec{k}_{ToF}$ and $\vec{k}_{Tag}$ define a common direction for both devices. A common measurable vector is the direction to the center of the earth. On the ToF system, $\vec{k}_{ToF}$ can be sensed with an accelerometer. Any state-of-the-art smartphone or head-mounted AR device features such a sensor. The accelerometer provides $\vec{k}_{Tag}$ at high update rates. On the ToF tag, $\vec{k}_{Tag}$ can be set to a constant vector if the tag is rigidly mounted. If the tag is mounted on a moving object, the tag also requires an accelerometer. With these vectors, it is possible to calculate the last set of points P_a, U_a:

$$P_a = P_C + \vec{k}_{Tag} \tag{4.6}$$

$$U_a = U_C + \vec{k}_{ToF} \tag{4.7}$$

It is now possible to calculate the rotation matrix R from these sets of points using a singular value decomposition [40]. With the rotation matrix R and the translation vector $\vec{t}$ obtained, all necessary parameters for 6-dof localization are determined.

4.2.2 Implementation of a Time-of-Flight Tag Localization System

This section presents a distinct variant of a ToF-based location-aware communication systems which is tailored to a specific application. The emphasis in this system is on communication partner localization, allowing to obtain the experimental results, which are presented in Chapter 5. The communication system which was introduced in Section 4.1, provides the foundation for a universal location-aware communication system that can be adapted to all intended applications. However, that approach requires a real-time system, which cannot be implemented in many of platforms without hardware modifications. The ToF tag system presented in this section takes a different approach [41]. The system is only capable of serving a single application and therefore it is possible to use unmodified ToF sensors. The simplifications further allow to implement and to evaluate the important use case of 3D tags for augmented reality. Implementing this use case enables to conduct experiments to characterize its performance and to detect potential problems.

As seen in Fig. 4.1, this system design provides a method meets the requirement of communication partner identification R1.2. Communication partner identification can be interpreted as a simplified communication link, where the communication partner identity is frequently broadcasted. Such identification in combination with 3D localization is the unique core of a Time-of-Flight location-aware communication system. Therefore this system design focuses on reliably identifying, localizing and tracking the communication partners with a ToF sensor. Using on/off keying as modulation method avoids modulation frequency synchronization and consequently a real-time system. Using complete phase images instead of a reduced region of interest reduces the data transfer rate, but removes the real-time region-of-interest tracking.

With these simplifications, the system can be implemented in software which is executed on an operating system. The complete ToF tag system consist of a ToF sensor with a processing system and multiple tags. The tags include a microcontroller, a photo diode as receiving unit and an LED to transmit data. The ToF sensor continuously captures depth images in a certain pattern. A combination of long and short range depth image capture produces such pattern, while also providing two separate streams of depth images at different range and frame-rate. Such pattern is well-suited for the use-case of ToF tags, and also used in head-mounted devices to scan the environment while simultaneously conducting gesture detection. The ToF tag uses the light from the ToF sensor to detect the pattern and to synchronize to the captured phase images. The tag can then activate its LED to transmit data when the ToF sensor is capturing certain phase images. The software on the processing system uses image processing to localize and track the visible tags. The datastream of the tracked tags is decoded and allows to identify the tracked elements. In order to evaluate the potential application of 3D tags for augmented reality, the 3D tag position is mapped into the image space of a color camera stream. The localization performance and the usability are evaluated in Chapter 5.

Software Design

The purpose of the ToF system is to process the captured ToF images, localize the visible communication partners and decode their data. The ToF system uses an unmodified Camboard pico flexx as ToF sensor, which is connected to a PC via USB 3.0. The software identifies and localizes the tags and maps their position to a color image stream. As seen in Fig. 4.16, the core of the C++ software design is organized in a pipeline architecture. Each step in the pipeline is represented by a class which executes processing operations on the received data. The processed data originates from the ToF sensor, a color camera and an optional inertial measurement unit. At each processing step, additional data is added and automatically forwarded to the next processing step.

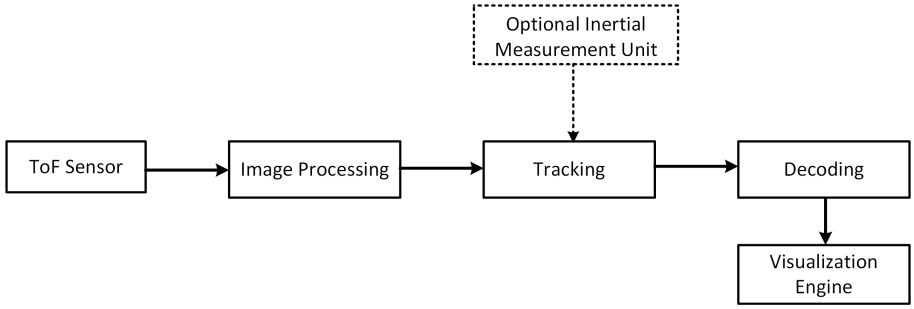


Figure 4.16: The architecture of the ToF tag system is organized in a pipeline architecture.

The image processing step uses the middleware software library of the ToF sensor to interface with the operating system driver to obtain depth and phase images. The 2D position p_{ToF} of the projected light sources is detected with sub-pixel accuracy. After compensating the lens distortion, the direction vector to the light source can be calculated from this 2D position. When the ToF sensor captures depth images, the communication partners do not emit light. This way, the distance to the communication partners d can be measured without distortion. The next step in the pipeline is to track the 2D position of the communication partners. With each new phase image, new light source data is associated with the closest tracked element. The positions of the tracked elements are then updated with the recent light source position. The associated data is converted to a datastream for each of the tracked elements. The tracking process also calculates the 3D position of the tag P_{ToF} in the ToF sensor coordinate system. The data stream of an inertial measurement unit (IMU) can optionally be integrated in this step. It is possible to aid the data association and increase the directional localization rate by fusing the IMU data with the detected light source positions using a Kalman filter. The message decoding process then uses the obtained phase values of the image processing process to reconstruct the transmitted messages. The visualization system finally transforms the 3D position of the tag P_{Tag} to the 3D coordinate space of a color camera, which is illustrated in Fig. 4.17.

$$P_{Color} = t + R \cdot P_{Tag} \tag{4.8}$$

For this transformation, a rotation matrix R and a translation vector t are required. The ToF and color camera are therefore rigidly mounted next to each other facing the same direction. This enables them to be calibrated like a stereo camera system, using the

method of Zhang et al. [42]. To display information at the 2D position of the tag in the color stream p_{Color}, the 3D position P_{Color} needs to be projected to the color image plane. Again using the pinhole camera model, the projection is a multiplication of P_{Color} with the color camera matrix C_{Color}.

$$\tilde{p}_{Color} = C_{Color} \cdot P_{Color} \tag{4.9}$$

After converting from homogenous coordinates, the visualization system can render any information at the 2D position p_{Color}. This enables to connect virtually any electronic device to any augmented reality system. A further investigation of this use case in provided in Chapter 5.

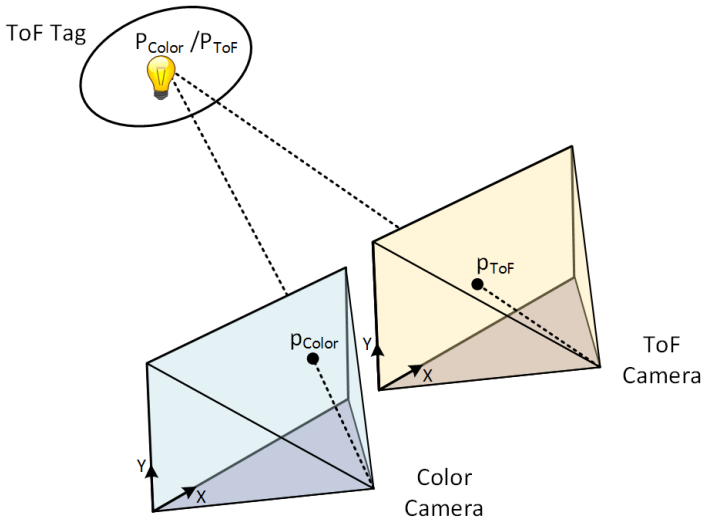


Figure 4.17: A ToF and color camera are rigidly mounted and facing the same direction. The 3D position of the ToF tag P_{ToF} can be transformed and projected into the stream of the color camera.

Synchronized ToF Tag Communication

The goal of the ToF tag system is to localize the ToF tags in 3D and to simultaneously receive data from all visible tags. A time synchronization mechanism is required, because the ToF tags need to know when the ToF sensor is able to receive their data. This means that tags need to transmit the data in the exact moment when the ToF sensor exposes its sensor to capture certain phase images. When the ToF sensor, however, takes depth images to measure the distance to the tags, the tags must not emit any light. In the ToF tag system, such distinct pattern occurs when a ToF sensor is configured to produce two streams of depth images simultaneously. This enables to combine short and long-range depth sensing. As illustrated in Fig. 4.18, a long-range depth image is captured at 5 Hz. To remove ambiguous distances, 8 different phase images are captured per depth image. Between each of these long-range depth images, 6 short-range depth images are captured, leading to a frame rate of 30 Hz. This pattern allows the tags to synchronize their time reference to the emitted pattern. By using a photo diode with a filtering circuit, it is

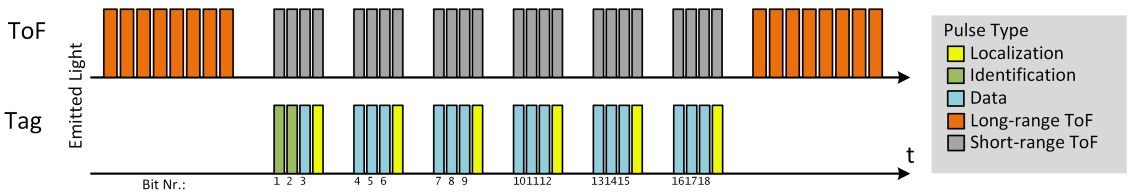


Figure 4.18: A ToF sensor captures long and short-range ToF images. The ToF tag is time-synchronized with the ToF sensor and responds when the ToF sensor captures short-range depth images

possible for the tag to detect the capture time of the phase images. Forwarding this signal to a microcontroller enables to match the timestamps of the pulses to a reference model. The required shift of the reference model is the time offset between the tag and the ToF sensor.

When synchronized, the firmware of the ToF tags knows when and how to transmit data. During the long-range depth image, the tag remains silent. This allows the ToF sensor to measure the distance without any signal distortion by the LED. During the 6 short-range depth images, the ToF tag transmits its identity and data via on/off keying. The ToF system uses this depth sensing pattern to conduct directional localization at 30 Hz and distance estimation at 5 Hz, while receiving data from each communication partner at 80 bit/s.

Chapter 5

Results and Use Case Evaluation

This chapter presents the experimental evaluation of this doctoral thesis. With the developed prototype systems and the implemented processing methods, the chapter characterizes the key functionality of ToF-based location-aware communication systems. Section 5.1 presents characteristics of the introduced modulation method which includes the synchronization mechanism. Section 5.2 characterizes the accuracy of the localization method, which is crucial for many applications. Section 5.3 contains the results of the proposed ToF data processing methods. Section 5.5 finally presents a use case evaluation. In the evaluated use case, the ToF tag system connects an augmented reality system with embedded devices. The use case evaluation investigates the suitability of the presented location-aware communication method and discusses alternative implementations.

5.1 Optical Communication

The goal of this section is to evaluate the key data transmission metrics. This includes the performance of the synchronization method, the data transfer rate and bit error ratio (BER). Several experiments are conducted to observe these characteristics with different parameters.

5.1.1 Evaluation Setup

The experiments are conducted with the FPGA-based optical communication system, which was introduced in Chapter 4. The evaluation setup is illustrated in Fig. 5.1. An evaluation transmitter emulates the signal of a potential communication partner. This allows to test the synchronization and communication performance in a controlled setup with different values. A clock-generation circuit on the evaluation transmitter generates a rectangular signal with 26 MHz frequency. This signal is fed to a frequency dividing circuit which provides four different signals at 6.5 MHz with equidistant phase offsets. A multiplexer then controls which of these signals is fed to the illumination unit. Alternating the phase-shifted signals creates a PSK-modulated illumination signal. For synchronization, the 26 MHz signal is directly connected to the illumination unit. The illumination unit consists of an SFH4715s LED, which is activated by a MOSFET transistor when the logic state of the illumination signal is high. A second MOSFET connects the anode and cathode of the LED via a resistor when the illumination signal is in low state. This helps

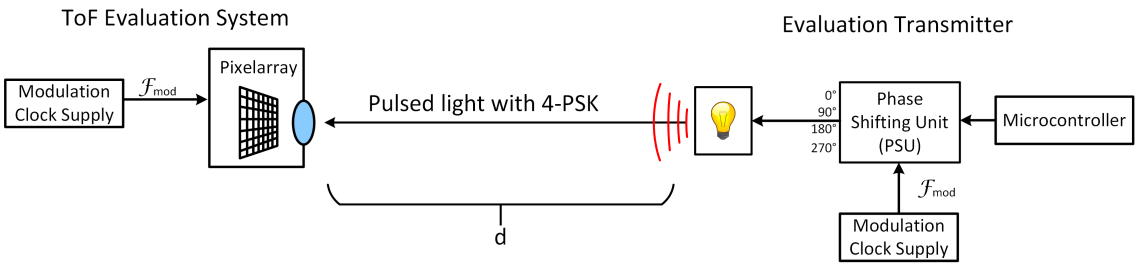


Figure 5.1: An evaluation transmitter enables to evaluate the proposed synchronization and modulation method with different parameters.

to discharge the LED after each light pulse, which increases the fall time of the optical signal. The multiplexer is controlled by a microcontroller which can select the appropriate illumination signal with precise timing. This enables to transmit test data to measure the performance of data transmission and synchronization. The distance between the communication partners has a significant impact on the received signal strength. The evaluation transmitter is therefore mounted on a linear rail, which enables to vary the distance in predefined steps.

5.1.2 Modulation Frequency Synchronization

Modulation frequency synchronization is an important requirement for efficient optical communication with ToF sensors. Synchronization enables to use the photonic mixer device for efficient PSK demodulation. A synchronization method was presented in Section 4.1.1, in Publication 6. The proposed synchronization method minimizes the frequency difference between the modulation frequency of the ToF sensor and its communication partner. This frequency difference is caused by manufacturing tolerances and temperature differences of the clock supply circuit. The proposed demodulation method tolerates a certain remaining synchronization error. The dynamic frequency drift therefore does not need to be considered.

The evaluation transmitter continuously emits light pulses at 26 MHz during the synchronization procedure. The main goal of the first experiment is to determine the maximum initial frequency difference dF. This frequency difference is measured by obtaining discrete samples of the aliasing signal. A too large dF can hence not be measured and the synchronization is impossible. In the first experiment, a series of synchronization procedures are executed at different initial frequency differences dF. The first experiment is conducted at ideal conditions with the shortest possible exposure time. The distance is 30 cm, with an exposure time of 0.3 μs, leading to a sampling rate of 7300 samples. The initial dF is manually adjusted in the experiments and the system then synchronizes to the evaluation transmitter 100 times. If the iterative algorithm reaches more than 20 iterations, the synchronization attempt is classified as failed. The tolerable remaining synchronization error is defined to be 10 Hz, which sufficient for optical communication. Fig. 5.2 A) shows the average number of iterations of the successful synchronization attempts. The larger dF is set, the greater the initial frequency estimation error is. More iterations are consequently required for a larger dF. Fig. 5.2 B) shows the remaining synchronization error for all successful synchronizations. Since a minor dF leads to an adequate

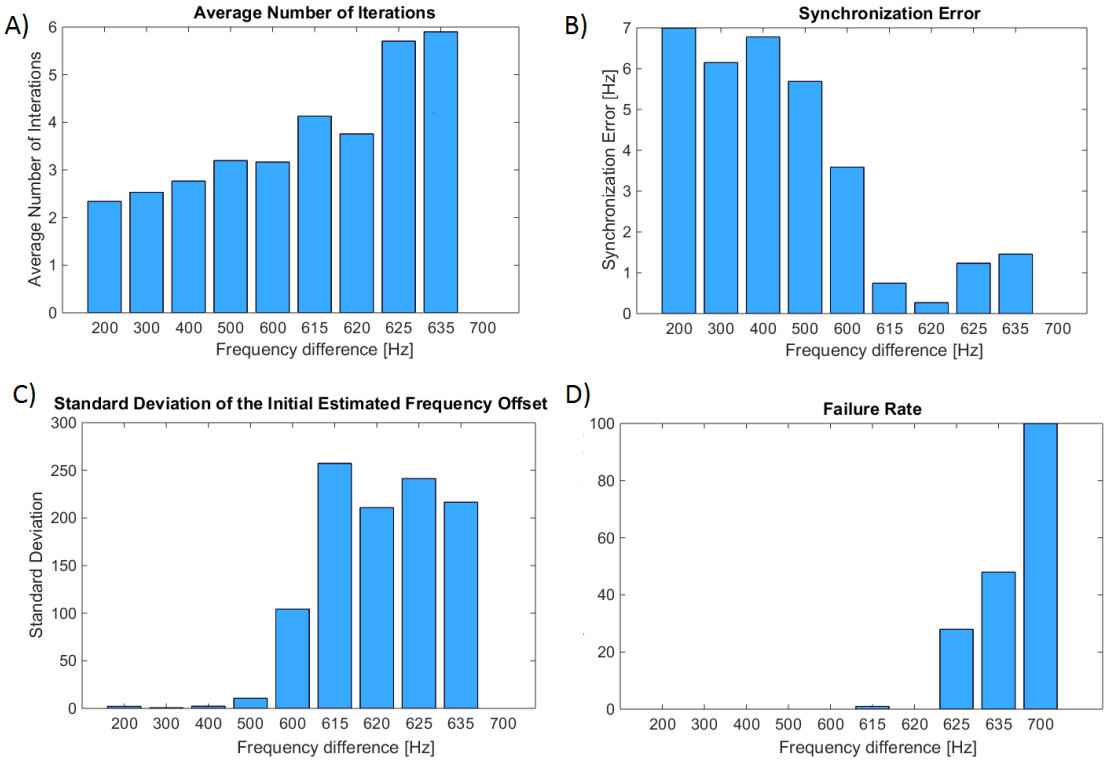


Figure 5.2: Evaluation of the modulation frequency synchronization method with different initial frequency differences. Reproduced with changes from [1].

initial frequency estimation, the iterative algorithm stops just after a few iterations. More iterations are caused by erroneous frequency estimations, and the algorithm finally terminates with a tolerable synchronization error. If a system wants to further minimize this synchronization error, reducing the maximum tolerable synchronization error is a valid option, but might lead to more iterations. Fig. 5.2 C) shows the standard deviation of the initial frequency difference estimation. This confirms that a larger dF is hard to measure, but the synchronization procedure can still be successful due to its iterative nature. The failure rate in percent is finally presented in Fig. 5.2 D). The system is capable of synchronizing to a signal with up to 600 Hz frequency difference. This limit depends on the frequency estimation method, the theoretical limit is half of the sampling rate. This result shows that the proposed modulation frequency synchronization procedure is a valid approach to synchronize to communication partners.

The second experiment observes the synchronization performance at different distances. Sampling the modulated light signal over longer ranges requires to prolong the exposure time to increase the signal strength. As in the first experiment, 100 synchronization attempts are executed by the system for each different distance. The frequency difference dF is set to 400 Hz. Fig. 5.3 A) shows the failure rate of the synchronization approach for different distances and exposure times. A short exposure of 0.3 μs is only sufficient for short-range synchronization. Prolonging the exposure to 30.7 μs gives a range of 4 m with 100% success rate. The furthest distance with 100% success rate was achieved at 9 m with 124.9 μs exposure time. As Fig. 5.3 B) shows, prolonging the exposure time

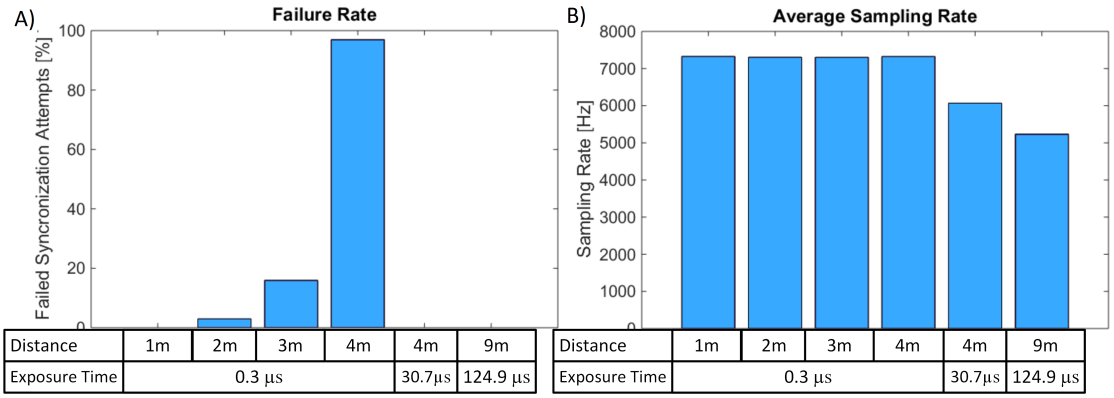


Figure 5.3: A) Evaluation of the modulation frequency synchronization method at different distances. B) The effect of the increased exposure time on the sampling rate. Reproduced with changes from [1].

has a minor effect on the sampling rate. At 9 m distance, the system is still capable of sampling the lightsource signal at over 5000 Hz. The reason is that the ToF sensor spends most of its operation time with pixel digitization. Prolonging the exposure just prolongs the shorter exposure periods.

5.1.3 Data Transmission Characteristics

In the first experiment, the bit error ratio (BER) is measured at different distances and data transfer rates. The BER is the number of bit errors divided by the number of transmitted bits. Time-of-Flight sensors are integrating receivers, which allows the adjustment of the received signal strength by varying the exposure time. A prolonged exposure however reduces the sampling rate and consequently the data transfer rate. For each different measurement, 4096 bits of test data are transferred from the evaluation light transmitter to the ToF sensor. The measurements are conducted at different distances and different data transfer rates. A typical ToF sensor is able shift its emitted light pulses by four equidistant phase offsets. Consequently, 4-PSK is used in this evaluation as modulation method.
Table 5.1 presents the obtained bit error ratios in percent at different distances and data transfer rates. If the transmitted data did not include any bit errors, the BER is stated

Distance \ Data Transfer Rate	6376 bps	5939 bps	4671 bps
1 m	$2.9 \cdot 10^{-3}$	$< 2.4 \cdot 10^{-4}$	$< 2.4 \cdot 10^{-4}$
2 m		$< 2.4 \cdot 10^{-4}$	$< 2.4 \cdot 10^{-4}$
3 m		$< 2.4 \cdot 10^{-4}$	$< 2.4 \cdot 10^{-4}$
4 m		$< 2.4 \cdot 10^{-4}$	$< 2.4 \cdot 10^{-4}$
5 m		$4 \cdot 10^{-4}$	$2 \cdot 10^{-3}$
6 m			$3.7 \cdot 10^{-2}$

Table 5.1: The bit error ratio at different distance and data transfer rates.

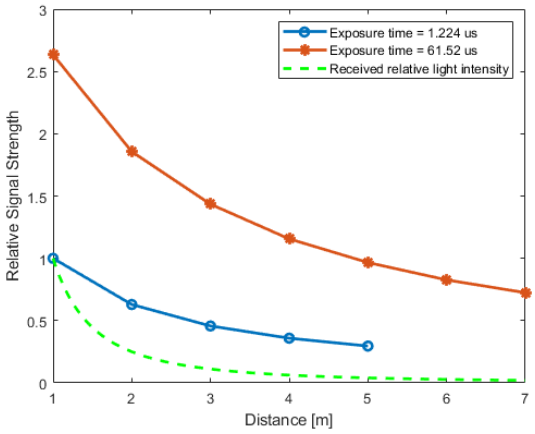


Figure 5.4: The signal strength of the received signal at different distances and exposure times.

as to be lower than the measurable ratio of $2.4 \cdot 10^{-4}$. The table shows that it is possible to obtain a range of at least four meters at a 5936 bps data rate. Extending the exposure of the ToF sensor time extends the range, but significantly reduces the data transfer rate.

Figure 5.4 shows the relation between the exposure time and the signal strength at different distances. The signal strength is the amplitude of the received phase signal. The phase signal is obtained by averaging the pixels that contain the modulated lightsource. The signal strength plots in the figure are normalized to the signal strength of the minimum exposure time of 1.224 μs at 1 m distance. The relation shows, that increasing the exposure time by a factor of 70 (from 1.224 μs to 61.52 μs) increases on the received signal strength only by an average factor of 3.0. The reason is that pixels are saturated at lower distances. Stray light causes a corona around the saturated pixels, which still contains the transmitted information. The result shows that an overexposed ToF sensor is still capable of receiving data. Figure 5.4 also includes the theoretical relative light intensity at the pixel array, which follows the inverse square law. The curve of the signal strength does not match

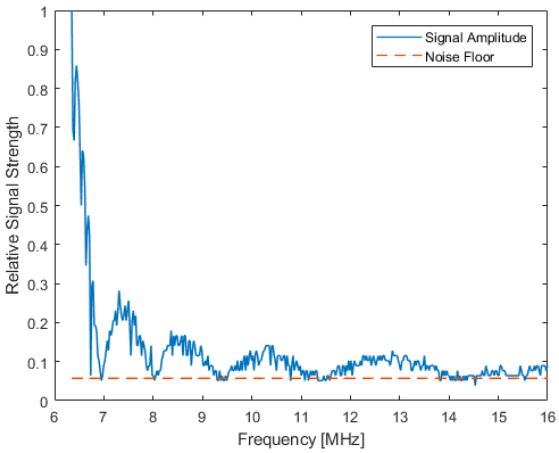


Figure 5.5: The received signal strength of a 6.5 MHz signal at different pixel reference signal frequencies. This demonstrates the signal suppression effect of the photonic mixer device.

the light intensity, due to the pixel saturation. The final evaluation examines the receives signal strength over different pixel reference frequencies. Communication partners could potentially exploit the frequency suppression to avoid interference. For this evaluation, the ToF receiver still receives the test pattern, but varies the pixel reference frequency. The signal strength of the asynchronous phase signal is plotted in Fig. 5.5. The plotted signal strength is relative to the synchronous frequency of 6.5 MHz. The measurement is conducted at 1 m distance with 1.224 μs exposure time. It can be observed, that there are certain frequency ranges, where the received light signal is suppressed beyond dedectability.

5.1.4 Discussion

The conducted experiments for sensor synchronization show, that the performance of the synchronization method sufficient to synchronize ToF sensors for optical communication. The initial frequency offset needs to be under 600 Hz, and the exposure time needs to be adjusted to the distance or signal strength of the light signal. Due to the iterative algorithm, the success rate is consistently at 100% for a range up to 9 m. This method is therefore a valuable foundation for any future ToF PSK communication link.

The data transmission characteristics show that it is possible to achieve a 4 m range with a data transmission rate of 5.9 kbps. This meets the requirements of the intended applications. The number of different transmitted phases N influences the number of transmitted bits per symbol and consequently the data transfer rate by $ld(N)$. The evaluation shows that it is possible to adjust the exposure time of the ToF sensor to increase the signal strength and consequently the range. The improvement comes however with a significantly reduced data transfer rate. Increasing the signal strength without reducing the data transfer rate requires modification of the ToF sensor system or brighter lightsources. It is possible to use LASER lightsources such as VCSELs, at an eye-safe duty cycle in future work. At a data rate of 5.9 kbps and exposure time of 61.52 μs, the ToF sensor is only sensitive to incoming light at 28% of the operation time. If lightsource follows this pattern, it is possible to significantly increase the signal strength without additional power consumption.

The frequency sweep of the pixel reference signal shows that the received light signal is effectively suppressed at certain frequencies. This enables to create communication systems using multiple frequencies in order to avoid interference between communication partners. It could also be possible, that this is used in scenarios, where multiple ToF sensors operate in the same area with different communication partners.

5.2 Localization

In this section, the accuracy and range of the 3D communication partner localization method are evaluated. The combination of communication and accurate localization is a distinctive property of ToF-based location-aware communication. The performance of potential applications such as indoor positioning, sensor node localization and 3D tags directly relays on accurate 3D localization.

5.2.1 Evaluation Setup

The proposed localization method combines photogrammetric angular localization with Time-of-Flight depth sensing to obtain the 3D position relative to the ToF sensor. Therefore it is important to evaluate the localization accuracy in each dimension. The error of the X and Y dimension is mainly caused by the angular localization, while the error in the Z axis largely depends on the Time-of-Flight depth sensing method. A major challenge of evaluating the performance of a camera-based localization system is to obtain ground-truth data. To characterize the Z localization error, a linear rail is used, which enables to adjust the distance to the communication partners at sub-millimeter accuracy. Setting a ground-truth position to obtain the X and Y error, would however require a complex setup which might not be accurate enough. The relative positions between multiple communication partners can, however, be accurately configured. As shown in Fig 5.6, the evaluation rig consists of three ToF tags mounted at 30 cm distance. Such equilateral triangle configuration allows for accurate ground truth position adjustment and guarantees that all communication partners are situated on a plane. The angle between the plane normal and the Z axis can be calculated, which enables to accurately adjust the orientation of the communication partners. In order to obtain multiple measurements at different distances with accurate ground truth in the Z dimension, the communication partners are mounted on a linear rail.

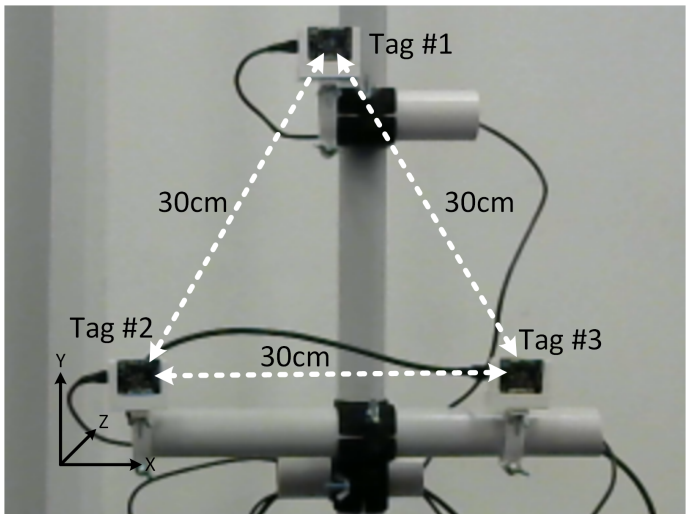


Figure 5.6: Three tags are mounted in a equilateral triangle configuration to evaluate the localization accuracy at different orientations and distances.

5.2.2 3D Localization Error

The goal of the conducted experiments is to provide insight on the localization performance of ToF-based communication partner localization. For a successful localization, the ToF sensor needs to conduct at least one valid distance measurement and needs to receive identification data from the tag. Both signals are affected by the orientation and the distance of the tags. Table 5.2 shows distances and orientations in which the ToF tag system is able to localize the tags.

Distance	Angle 0°	22.5°	45°
4.5 m	✓	✗	✗
4.0 m	✓	✓	✗
3.0 m	✓	✓	✓

Table 5.2: Distance and tag orientation requirements. The localization method need to conduct valid ToF distance measurements and to receive data from the tags.

In the first experiment, the localization error is measured over a range of different distances. With an adjustable linear rail, the three ToF tags are localized 50 times at each distance which ranges from 1.0 m to 4.5 m with a spacing of 50 cm. The measured positions of the tags are compared with the ground-truth for each dimension to obtain the localization error. The error for each dimension is then averaged for each distance, and plotted in Fig. 5.7. The combined localization error E_C is the average distance between the measured 3D position and the ground-truth. It is calculated from the individual axis errors E_X, E_Y, E_Z.

$$E_C = \sqrt{E_X^2 + E_Y^2 + E_Z^2} \tag{5.1}$$

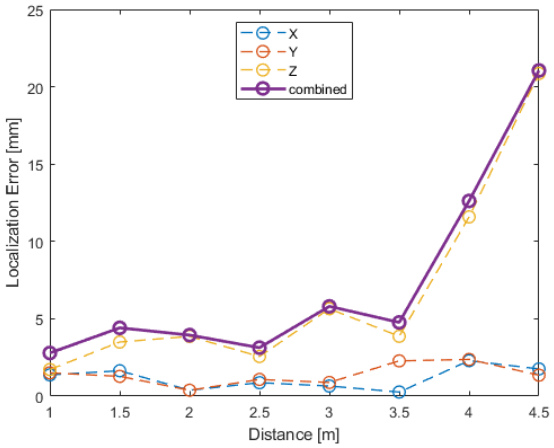


Figure 5.7: The average localization error at different distances.

Figure 5.8 shows the cumulative distribution function CDF plot of the complete measurement series of experiment 1. Unlike the plot in the previous figure, the localization error is not averaged. Since the measurements are equally distributed over the distance range, the plot provides estimates for the maximum localization error and the corresponding probability. Since the localization accuracy severely degrades for distances over 3.5 m, the CDF plot is also provided for a range of 1 to 3.5 m in Fig. 5.8 B.
The second experiment explores the localization performance for different tag orientations. If the tag faces the ToF sensor at an angle, the received signal strength decreases depending on the radiation characteristics of the tag LED. Furthermore the distance measurement might be impaired, since less light is reflected from the surface of the tag. In the experiment, the triangular tag rig is placed in front of the ToF sensor at a distance of

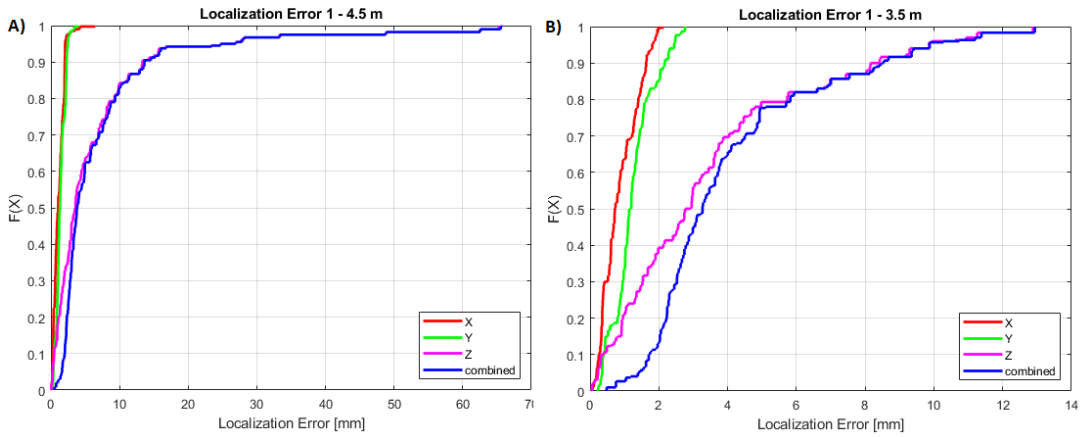


Figure 5.8: The CDF plot of the localization error between 1 and 4.5 m (A) and between 1 and 3.5 m (B).

2 m. Localizations are conducted with different rotations around the Y axis between 11° to 45°. Fig. 5.9 shows the additional average localization error for different orientations of the tag rig.

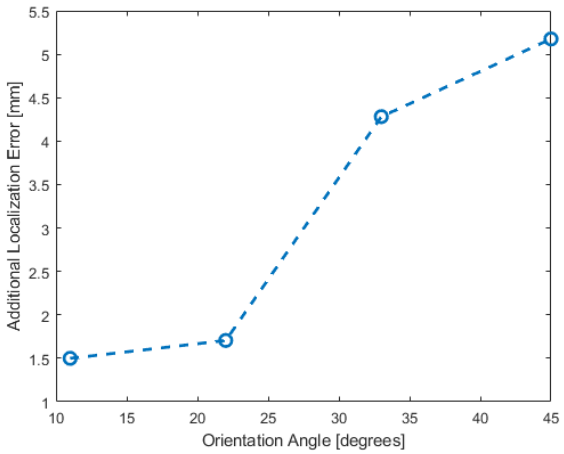


Figure 5.9: The additional localization error caused by the tag orientation, measured at 2 m distance.

5.2.3 Discussion

As discussed in Chapter 3, the potential applications require a typical accuracy of 2 cm. This requirement is met for all conducted measurements for a range of up to 3.5 m. At the maximum evaluated range of 4.5 m, the localization error is below 2 cm with a probability of 94%. The experiment also shows that the ToF distance sensing accuracy has a significant influence on the localization accuracy. It can be observed, that the accuracy severely degrades at a distance longer than 3.5 m. The reason is that the signal strength of the reflected lightpulses in each pixel is less and less sufficient for a proper depth measurement. The system takes the average distance measurement from multiple

pixels to estimate the distance to the communication partner. This accuracy decreases dramatically after 3.5 m, as fewer pixels contribute to the distance estimation.

The ToF sensor used in the evaluation is designed for use in mobile devices and is therefore miniaturized. Applications with less demanding requirements on power consumption and sensor size can potentially improve the localization accuracy. Another way to improve the depth measurements is to increase the exposure time of the depth images. For depth sensing in the ToF tag system, 8 phase images are captured with 1.2 ms exposure time. Reducing the number of synchroniziation light pulses allows to further increase the exposure time for depth sensing.

The second experiment shows the additional localization error caused by the orientation of the communication partner. At 2 m distance, an additional localization error of 5.2 mm is introduced at a 45° angle. This value depends on the geometry of the communication partner and the radiation characteristics of the LED. Even if the data signal can be received by the ToF sensor, the size of the projected lightsource on the pixel array decreases with the distance. This impairs the super-resolution method which is used to estimate the position of the lightsources on the pixel array. The localization system also uses distance measurements in proximity of the lightsource to estimate the distance. Rotating the lightsource and its frontal surface further impairs the distance estimation. Model-based approaches such as fitting the point cloud of the object to a reference model have the potential to compensate this error source.

5.3 Processing Methods

Several optional processing methods were introduced in Section 1.2. These methods aim to help ToF-based location-aware communication systems to meet the localization requirements. The most important methods were implemented as prototypes and evaluated in this chapter. The evaluation of the methods shows that they can provide future performance enhancements. This potentially enables a wide adoption of ToF-based location-aware optical communication. Besides location-aware communication, the developed methods are helpfull for various additional applications. Due to the lack of comparable work in the field of location-aware communication, the experiments of the evaluation concentrate on these existing applications.

5.3.1 Time-of-Flight and Color Sensor Data Fusion

This method combines depth and color data in order to obtain a high-resolution depth image. The edge information in the color image guides a custom depth upscaling algorithm. Unlike the numerous previous 2D and 3D sensor fusion approaches, the developed algorithm is designed for efficient computation on a graphics processing unit (GPU). The algorithm is capable of upscaling a 288x256 depth image to a resolution of 640x480 at over 10 frames per second (FPS) on a state-of-the-art smartphone GPU. If this method is used to refine the depth image during communication partner localization, a smaller sub-image can be used for faster upscaling.
In order to compare the quality of the upscaled depth image, the algorithm is tested with the Middlebury 2006 dataset [6]. This dataset contains color images and high-resolution ground-truth depth images. These depth images are downscaled by a certain factor and

	Aloe				Art				Moebius			
	8x		16x		8x		16x		8x		16x	
	RMSE	PbmP	RMSE	PbmP	RMSE	PbmP	RMSE	PbmP	RMSE	PbmP	RMSE	PbmP
Nearest Neighbor	4.89	77.76	5.96	77.14	5.73	78.49	7.50	80.47	4.91	77.83	5.62	78.42
Bilinear	4.21	72.17	5.32	73.27	4.98	74.50	6.79	77.24	4.18	73.05	4.92	74.26
Kopf et al. [43]	3.45	57.13	5.47	59.35	4.20	59.41	3.00	54.75	2.89	54.37	4.05	56.65
Dai et al. [44]	3.81	59.72	5.03	69.30	4.06	61.87	4.11	58.58	4.25	64.79	4.03	68.44
Ferstl et al. [45]	2.28	31.69	3.50	49.94	2.83	38.98	1.92	27.89	2.66	52.45	2.33	41.53
Ours	3.36	54.31	4.99	65.44	3.59	53.28	2.59	47.82	2.59	47.82	3.94	63.70

Table 5.3: Comparison of upscaling algorithms on the Middlebury 2006 dataset [6], using the root mean square error (RMSE) and the percentage of bad matching pixel (PbmP) as metrics. Reproduced with changes from [7].

noise is added. The images are then upscaled by the proposed algorithm. Two metrics are available to compare the approach with state-of-the-art upscaling methods. The root-mean-square-error (RMSE) compares the overall pixel-per-pixel difference. As shown in [7], the percentage of bad matching pixels (PbmP) is the percentage of pixels in the upscaled image d, which differ by a certain value of δ from the ground-truth image g. While the RMSE is a metric for the similarity of the images, the PbmP is sensible to artifacts.

$$PbmP = \frac{1}{N} \sum_{(x,y)} (|d(x,y) - g(x,y)| > \delta) \tag{5.2}$$

Table 5.3 compares the RMSE and PbmP with several state-of-the-art methods. It is possible to observe, that the proposed method is superior to all evaluated methods, except for the approach of Ferstl. et al. [45] Their approach is based on optimization, which is not suitable for high-performance applications. The results show that the proposed method is computationally efficient while still providing decent upscaling quality.

5.3.2 High-speed Time-of-Flight Sensing

This method proposes a calibration method for ToF sensors, enabling to measure the distance to fast moving objects. The ToF motion artifact problem is most severe in applications where the ToF sensor is mounted to a moving and rotating object, like robotics, smartphones or head-mounted AR/VR devices. In conventional ToF imaging, multiple phase images are required to create a depth image or to localize a communication partner. Due to the movement, each phase image captures the object with different sets of pixels. The calibration method needs to enable to combine phase values from different pixels to create a depth image.

The calibrated ToF sensor is mounted on a rotation platform for the experimental evaluation. A set of phase images is captured of the same object at a constant 1 m distance at different angles. The object is a 1x1 cm retroreflective marker, which can be identified in each captured phase image. The calibration operation is applied to the average phase value of the marker. These corrected phase values are then arbitrarily combined to calculate the distance. This simulates the distance calculation of a fast moving object. The random combination causes different sets of pixels to be used for distance calculation. The histogram of the distance estimation error is shown in Fig. 5.10. The standard deviation is 1.25 cm, and the error shows a bias of about 0.48 cm, which originates from the distance calibration procedure. These results show that the proposed calibration

method introduces a certain error when estimating the distance of small objects. This error will however only occur during fast movements, in which traditional ToF sensing would not allow any distance measurements. The calibration procedure can consequently be useful to increase the robustness of communication partner localization.

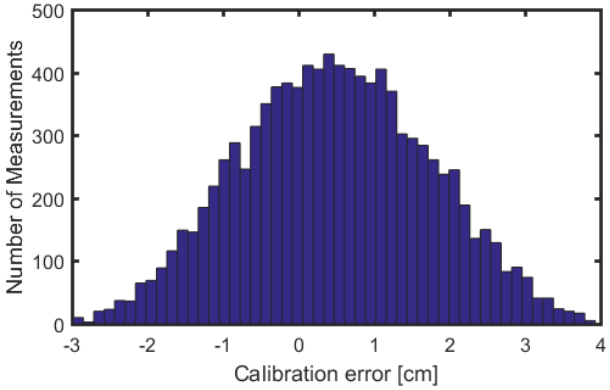


Figure 5.10: The error distribution of the proposed motion artifact resilient distance estimation method. Reproduced with changes from [5].

5.3.3 3D Object Tracking with Single Phase Images

An important requirement for many ToF systems is low power consumption. Localizing the communication partner in 3D requires a sizeable amount of energy due to the active illumination unit. The proposed method attempts to reduce this energy by just using a single phase image to determine the distance to a communication partner. Since only reflected light is sensed and processed in this method, it can also be applied to any 3D object tracking application. With the increasing amount of front-facing ToF sensors in

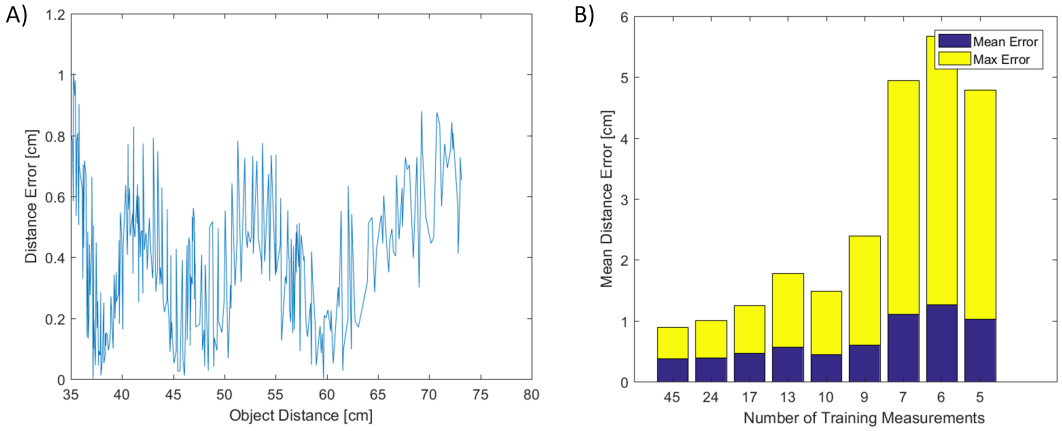


Figure 5.11: Results of the proposed energy efficient 3D object tracking method. A) The introduced distance estimation error is below 1 cm. B) The number of training measurements influences the distance estimation error. Reproduced with changes from [3].

smartphones, this method has the potential to allow efficient 3D face-tracking on smartphones. Publication 8 evaluates this use case by continuously tracking a human head over a distance between 35 to 75 cm. The result indicates that this method can be applied to communication partner localization as well. For the experiment, 300 depth images of a human head are captured at different distances and stored with their corresponding phase images. From these measurements, a random subset is selected as training measurements for the distance estimation method. The most suitable phase offset is selected, and the distance is estimated in each of the remaining phase images. In Fig. 5.11 A), the distance estimation error is plotted over the range of the recorded dataset. The ground truth for the distance is obtained by using the recorded depth image. With just 24 training measurements, it is possible to obtain a depth estimation error below 1 cm. The noticeable periodicity can be explained by the fitting error of the interpolation function. Fig. 5.11 B) shows the mean and maximum distance estimation error in the dataset, with different numbers of training measurements. It is possible to just obtain 9 training measurements to achieve an estimation error below 1 cm. This estimation error is acceptable for many object tracking applications. Power measurements on the evaluated ToF sensor show a power consumption saving of 68% compared to four-phase ToF sensing. These results serve as proof of concept for communication partner localization and show that energy savings are possible in future ToF-based location-aware communication systems.

5.4 Limitations

This section discusses the limitations of the location-aware optical communication approach. The presented work proposes optical line-of-sight connections between a ToF sensor and multiple communication partners. The ToF sensor can receive data from multiple sources in parallel, because the light sources are projected to different pixel regions. The presented optical communication approach and its results are only capable to sample a single light source, due to the restricted pixel area. Digitizing all pixels and receiving data from multiple sources in parallel is possible, but drastically reduces the sampling rate to 500 Hz. A ToF sensor can only broadcast data to all present communication partners. Connections to individual devices therefore need to share this channel. The potential applications, however, do not require significant data transfer rates to multiple communication partners. For the application of secured communication, significant data might be exchanged, but only between two entities. The developed systems and methods assume that just a single ToF sensor operates in an environment. Multiple ToF sensors require additional considerations and might face performance limitations. In the presented work, the communication partners detect the illumination signal from a ToF sensor to synchronize their data transmission timings. When multiple ToF sensors simultaneously communicate with the same device, it needs to synchronize the response to multiple ToF sensors. This also requires synchronization between the ToF sensors to avoid simultaneous transmission. Multiple infrared light sources in the same environment might also impair the depth sensing performance of the ToF sensors. While ToF sensors are able to actively suppress background light, the measurement might be affected by the saturation of the capacitors on the photonic mixer device. These issues can be addressed by synchronizing the active ToF sensors in the same area. For such time synchronization, a ToF sensor

could detect the signals of other ToF sensors in an initialization phase.

The results show, that the communication link is limited to a range of 4 m at a data transfer rate of 6 kbps. As stated by the inverse square law, the received intensity exponentially decreases over the distance. The range and data transfer rate could be improved by increasing the light intensity. However eye-safety requirements constrain the emitted intensity in most applications. Recent developments such as VCSEL arrays can however lead to future improvements.

The localization performance is restricted by the resolution of the pixel array and the depth sensing performance. Sub-pixel accuracy methods can be applied to improve the accuracy of the angular localization. The depth sensing range is however restricted by the amount of reflected light and the ToF sensor characteristics. In the evaluated system, the 3D localization range is therefore limited to 4.5 m. Retroreflective materials and active backscattering can improve the localization range in future work.

5.5 Use Case Evaluation of Time-of-Flight 3D Tags

This section provides a use case evaluation of a distinct application for ToF-based location-aware communication. The ToF tags are small optical transmitters, which are able to transmit data to a ToF sensor. The ToF system is able to receive data and localize these tags. This system enables to implement the use case of 3D tags for augmented reality. The 3D position can be used to map the tag position into the field of view of an augmented reality device. This allows unprecedented user interaction with embedded electronics. For this evaluation, the ToF tag system was used to implement two use cases for augmented reality smartphones. This section further features performance metrics of the use case and a comprehensive discussion on how alternative technologies could replicate the use case.

5.5.1 ToF Tags for Augmented Reality

The field of augmented reality has reached several milestones in the previous years. Head-mounted devices such as the Microsoft HoloLens, the Meta 2 or the Magic Leap One are leading the way for a wide adaption in the future. Middleware for smartphones such

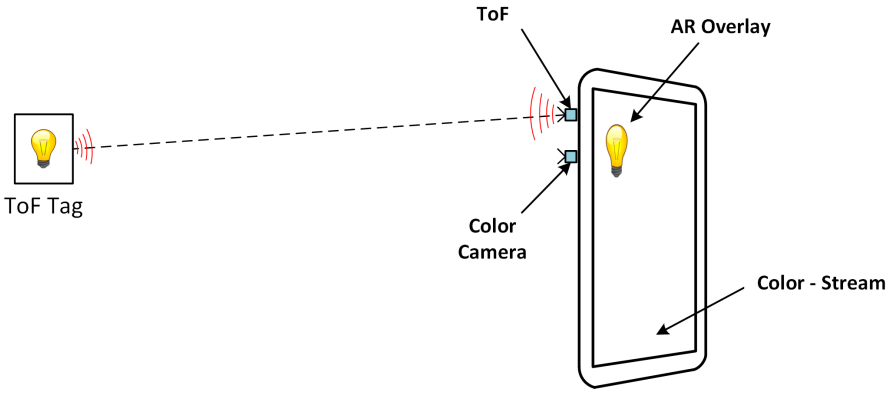


Figure 5.12: Location-aware optical communication with ToF sensors enables the use case of 3D tags for augmented reality smartphones.

as the Apple ARKit and the ARCore for Android are already providing widely adopted augmented reality middleware. These solutions are able to augment information with simultaneous mapping and localization (SLAM) and combined IMU and 2D camera-based motion tracking techniques. A shortcoming of the existing approaches is that there is a lack of interaction methods between the AR devices and embedded electronics.
A major purpose of ToF-based location-aware optical communication is to close this gap by introducing ToF tags for augmented reality. Numerous AR devices are already equipped with ToF sensors to provide point clouds to SLAM systems and to capture close-range depth images for gesture recognition. The ToF tag system is compatible with these use cases, since the system captures both long and short-range depth images. As discussed in Section 5.5.2, the ToF tag approach is simple and cost-efficient compared to its alternatives. While ToF tags can serve any kind of augmented reality device, AR smartphones are the most widely adapted system. The evaluated use case introduces two different ToF tags for a smartphone/tablet-based AR system. As shown in Fig. 5.13 A), the first tag type enables to attach information to a specific location. In the implemented example, a temperature and a humidity sensor are connected to the tag. The sensor data is continuously streamed to the ToF sensor via on/off keying. The demo tablet displays a 30 Hz color image stream. The information about the tag is overlaid in this stream at the position of the tag. This type of ToF tag can serve as transmitter unit for any electronic device, enabling to bring any information to augmented reality systems.

As seen in Fig. 5.13 B), ToF tags also enable remote control of electronic devices. The ToF system recognizes the tag and overlays a user interface at the position of the tag in the color image stream. In case of the implemented example, it is possible to toggle a light switch by touching the user interface of the tag. The ToF system transmits the command to the tag which returns the status of the switch. The command can be either transmitted via the Internet or via optical communication. In case of the implemented use case, the ToF sensor transmits the command by altering the capture times of the depth images. The ToF sensor sends out a specific pattern of light pulses, and the tag calculates the correlation to a set of reference patterns. Future work can use the active illumination unit

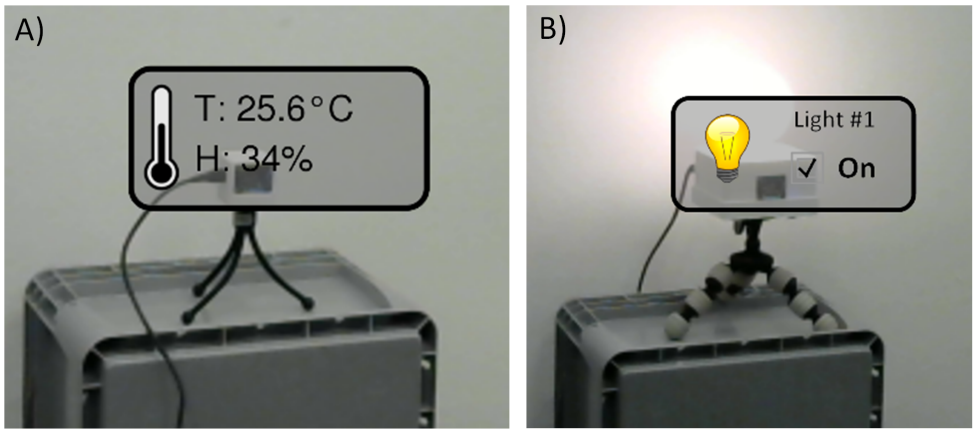


Figure 5.13: The implemented use case supports two different ToF tag types. A) Tags are able to transmit location-specific information. B) An electronic device can use a ToF tag to present a user interface.

for advanced modulation methods, such as pulse-position keying or phase-shift keying. The implemented use case shows that ToF-based location-aware optical communication can give any electronic device an augmented reality user interface. Without buttons or screens, users can get the interface directly presented into their field of view.

5.5.2 Alternative Use Case Implementations

This section reviews alternative approaches to implement the use case with different methods. These methods are discussed with focus on functionality, complexity and comparison to ToF-based location-aware communication. Alternative location-aware communication systems need sufficient localization accuracy and update rate to support the use case of 3D tags for augmented reality systems. As the review of the related work in Chapter 2 shows, most introduced localization approaches do not fulfill these requirements. The approach of Xie et al. [33] however targets a similar use case. Their approach associates 3D point clouds with RFID tags, which enables to attach information in an augmented reality system. A requirement of the system, however, is that the tagged objects are sizeable and not obstructed to be associated with their RFID tags.

The essential information for the use case is the relative position of the tag to the AR system. This position can be acquired, if the AR system and the tag are localized in a common coordinate system. The AR system needs to be localized with 6-dof (3D position and orientation), while the tags only require 3D localization. These parameters can be obtained by an external system. Sweeping laser beams [46, 47] and camera systems [48] have seen usage in commercial systems. These methods, however, require an external infrastructure, restricting the operation range and adding complexity.

If the position of a static tag within a coordinate system is known or manually configured, only the AR device requires localization. The most common method for self-localization in AR devices is SLAM in which the geometry of the scene is continuously gathered by the device. The data structure of the mapped geometry might also feature salient landmark points, which are matched by the device for localization. This enables the device to not only track its 6-dof pose, but to re-locate itself in its own mapped data. To obtain the pose in the coordinate system of the tag, the AR device needs to calculate the transformation between its own mapped data and the tag coordinate system. A proven method to match such coordinate system is to first conduct a 3D scan of the surroundings [49]. The scanned data is manually aligned with the target coordinate system. The AR device can then match the 3D scan with its own mapped data, enabling 6-dof indoor positioning. In a tag system, the tag position could be placed into the target coordinate system. This method has the advantage, that the tagged device does not need to be present and no line-of-sight connection is required. This approach however is complex and requires a certain infrastructure to become widely adopted. Mapping an environment with a visual method is impaired by dynamic objects and changes in scene geometry and illumination. The tagged information also needs to be stationary, while ToF tags can also be applied to moving objects.

Sensing at least three identifiable light sources with a 2D camera enables 6-dof localization. It is possible to implement the use case with such a system, if a tag is part of such a visual beaconing system. Despite the lack of a 3D sensor, it is possible to derive the 6-dof pose of the AR system by solving the perspective-n-point problem. Such beacon-based

positioning approach however requires an extensive infrastructure. The accurate location of each beacon needs to be derived and transmitted to the AR system. The accuracy of the localization vastly depends on at which positions the beacons are visible in the 2D image. Such a beacon-based positioning system requires a 2D infrared camera, which is not part of common AR systems.

An established method to directly transmit information to augmented reality systems are visual markers [50]. Markers such as QR codes [51] encode information, but can also be localized with 6-dof by any 2D camera system. Such visual markers could also be used to transmit information when being displayed at a screen [52]. The markers, however, need to be sensed at a certain size on the pixel array, which severely limits the range and applicability of the approach. Increasing the camera resolution improves the range, but leads to a lower localization rate. The performance of the system is also restricted by the amount of background light. Low-light conditions require a longer exposure time, which increases motion blur. An additional communication link to the tagged device is necessary to implement the use case of remote control of electronic devices.

Stereo camera systems are an alternative to Time-of-Flight sensors for the 3D tag use case. Sample et al. propose to use stereo cameras to localize RFID tags which are coupled with an LED [31]. Such visible light localization might however disturb the surroundings with their flashing light sources. An infrared stereo camera system with optical communication capabilities could be able to implement the use case. The advantage to ToF is that the system does not require an active illumination source to determine the 3D position of the communication partner. Due to the higher sensor resolution of stereo cameras, the direction of the communication partner can be more accurately determined than with ToF systems. This higher resolution, however, also reduces the frame rate. Such system would also require image processing to detect the signal of the communication partners. While ToF sensors are only sensitive to modulated light, conventional image-sensors are sensitive to all infrared light sources. A major disadvantage are the two cameras, which need to be mounted at a certain distance. This is in conflict to the miniaturization of smartphones and other AR devices. While a ToF sensor can be used for additional use cases, the stereo infrared system would mostly be restricted to location-aware communication. Using such infrared stereo system instead of ToF sensors would therefore increase the system complexity and size while reducing the functionality of the device. An alternative depth sensing method for AR devices is structured light. A pattern projector projects a pattern on the scene and an IR camera obtains the depth by triangulation. Using such a system for the 3D tag use case is challenging, because it is difficult to obtain the distance to small objects. While a ToF sensor only needs a single pixel to derive the distance, a structured light system need to identify its projected pattern. Albeit the resolution of the imager in a structured light sensor is superior, a certain part of the projected pattern needs to be visible on the object.

In case of augmented reality for smartphones, the 3D location is required to map the tag position into the image space of the color camera. If the color camera itself however is able to directly sense the tag, no mapping is required. Color cameras in smartphones usually block infrared light because it affects photography. If, however, a certain percentage of pixels is sensitive to IR light, it is possible to determine the direction of the tag. As introduced in Section 2.1, Takai et al. [18] created an image sensor with communication pixels for automotive purposes. Such sensor might however not be suitable

for smartphones. Dedicating a percentage of pixels for different purposes removes these pixels from the color image, potentially causing under-sampling artifacts. Selectively placing IR-blocking filters over the color pixels is a compromise, which might not lead to the same results as a dedicated IR blocking filter in the lens. While the mechanism of the use case could be covered by such a camera, the missing 3D localization does not allow to scale the displayed information. The tag information of a close tag would be displayed at the same size as a tag which is further away. This also affects the rendering order when tag contents overlap. This problem might be insignificant for smartphone AR, but the next step in augmented reality are smartglasses and other headmounted devices. In these devices, the overlaid information is rendered in stereoscopic 3D, which strictly requires 3D tag localization.

In conclusion, the use case of 3D tags for augmented reality devices is not straightforward to implement with alternative methods as they require complex systems and infrastructures. Using depth sensing for location-aware communication removes the complexity and trade-offs of the alternatives. Infrared stereo cameras are the most viable alternative, however, Time-of-Flight sensors are already part of several AR systems, and are most miniaturized depth sensing solution. The proposed ToF tag method is compatible with these applications and is able to provide its functionality as additional feature to these systems. Location-aware optical communication with Time-of-Flight sensors might therefore change the future of embedded systems by enabling direct spatial interaction to a large number of devices.

Chapter 6

Conclusion and Future Work

6.1 Conclusion

This doctoral thesis presents fundamental research to establish a location-aware communication system with Time-of-Flight sensors. The potential of ToF sensors for optical communication has barely been touched by industry or academia. Due to this lack of previous work, this doctoral thesis concentrates on researching the requirements of future systems and solving the emerging problems by proposing and implementing several system designs and methods. The experimental results indicate that the requirements can be fulfilled with the developed methods. Time-of-Flight sensors are therefore well suited for location-aware communication and are able to fulfill the requirements of promising future applications. The thesis thus provides the fundamental building blocks for a broad variety of future ToF-based location-aware communication solutions.

In the first part of the thesis, potential applications are investigated to define their requirements. By characterizing the functionality of the ToF sensor, the challenges and problems can be identified and translated to design goals. These goals are then addressed by two system designs and several processing methods. The first system design uses a real-time system to evaluate the optical communication capabilities and location-aware sensor operation. An experimental evaluation shows that the Time-of-Flight pixel is suitable for fast and efficient PSK demodulation of pulsed infrared light. A requirement of the introduced modulation method is to synchronize the modulation frequency of the ToF sensor. The developed synchronization mechanism can iteratively adjust the frequency of a ToF sensor to any modulated light source. The system design further demonstrates that ToF sensors can combine localization and communication to location-aware communication. A set of ToF sensor operations and processing methods is introduced, which can serve as building blocks for a wide variety of location-aware communication protocol implementations. The second system is designed to fulfill the requirements for a specific application without the requirement of a real-time system. Aimed for wide and versatile use, the system design works with established ToF sensing systems because no sensor modifications are necessary. The novel use case of 3D tags for augmented reality smartphones is implemented with this system. This promising application is evaluated with regard to usability, performance and alternative technologies. As the evaluation shows, ToF-based 3D tags perform well in the intended applications and are less complex than alternative

methods. Several processing methods are finally introduced to further improve location-aware communication, by increasing precision, update-rate and energy-efficiency. The most promising methods are implemented and evaluated with a focus on their potential applications.

In conclusion, Time-of-Flight sensors can be used as capable optical transceivers. Optical communication and precise 3D localization can be combined to the first 3D location-aware communication system based on Time-of-Flight sensors. The evaluations of the system designs, use case and processing methods show promising results.

6.2 Directions for Future Work

The doctoral thesis provides methods and system designs which can serve as foundation for future ToF-based location-aware communication system. Future work involves further investigation of the proposed methods, improving the proposed designs and engineering protocols and systems. With the provided research and the increasing availability of ToF sensors, it is also possible to find further potential applications for location-aware communication systems.

6.2.1 ToF-based Location-aware Communication Systems and Methods

Application-specific location-aware communication systems are able to tailor the proposed methods to suit intended applications. Another option are parametric location-aware communication systems, which can be adapted to specific applications. The proposed system designs in this thesis were established for characterization and use case evaluation. Future systems can be structured more efficiently and be implemented on different hardware platforms. Instead of using a real-time system to configure the ToF sensor during operation, a custom sensor firmware can implement this functionality. Without real-time configuration changes, the location-aware communication protocol could be implemented as an operating system driver. Such architecture could enable to implement location-aware communication on a vast number of devices such as smartphones without hardware modifications.

The next level for location-aware communication might be 6-dof localization. Several additional applications arise if the relative orientation of a communication partner can also be determined. A concept for 6-dof localization is provided in Section 4.2.1. Future work involves to proof the feasiblity with an implementation, integrate the 6-dof localization into a protocol and to research and explore the arising applications. With such localization, it is also possible to create an unprecedented single-anchor 6-dof indoor localization system, if the communication partner acts as beacon. The simultaneous data transmission and depth sensing method which was introduced in Section 4.1.1 has potential for future work. The method enables ToF sensors to modulate information in both the pixel reference signal and the emitted light pulses. This enables ToF sensors to transmit data during depth sensing without additional energy consumption. When omnipresent ToF sensors can transmit data and receive data without additional cost, vast networking opportunities arise.

6.2.2 Applications

With the results of this doctoral thesis, ToF-based location-aware optical communication becomes significantly less complex to implement. Besides the already proposed potential applications, there are further application fields which might benefit of a ToF-based location-aware optical communication system. In automated and autonomous driving, the surroundings need to be sensed by multiple and partly redundant methods. Establishing location-aware optical connections between subsequent cars improves safety by redundant localization at high update rates. The optical connection allows to transmit data such as traffic, speed and brake warnings with short latency. Applying secured distance exchanges allows to verify that the received signal is not relayed by an attacker. As this work shows, it is also possible to create high performing beacon-based indoor positioning systems with ToF sensors. Due to the high update rate, these systems can also be used as motion tracking systems for virtual reality devices and robotics. Using passive camera systems for position tracking is a complex challenge and currently impairs the use of smartphones as VR devices. By providing just three ToF tags as beacons, any smartphone with a rear-facing ToF sensor can use efficient 6-dof motion tracking at high update rates.

The evaluated use case of ToF tags for smartphone augmented reality can be transferred to smartglasses and other head-mounted devices. Instead of mapping the 3D position to a color camera stream, the 3D position is stereoscopically mapped into each eye of the user. This allows to overlay the tag information at the exact 3D location of the tag. Since the proposed system design still provides a stream of depth images, gesture control enables the user to touch and interact with the virtual interfaces. Depending on the future development of such AR devices, ToF-based location-aware optical communication has the potential to extend the internet-of-things with ubiquitous spatial user interaction.

Chapter 7

Publications

This chapter presents the publications which were established in the context of this doctoral thesis. This dissertation project led to the publication of eight academic papers and the filing of eleven patent applications. In the first phase, potential applications for ToF-based location-ware optical communication were investigated. The application requirements were then evaluated with regards to the ToF sensor performance. Several system designs and processing methods were developed to enable future systems to fulfill the application requirements. The context of the contributions are shown in Figure 7.1. The application of secured communication with Time-of-Flight sensors initiated this thesis. The application was therefore thoroughly investigated in publications 1, 2, and 4. A summary of this application and its requirements is provided in Section 3.1. Publication

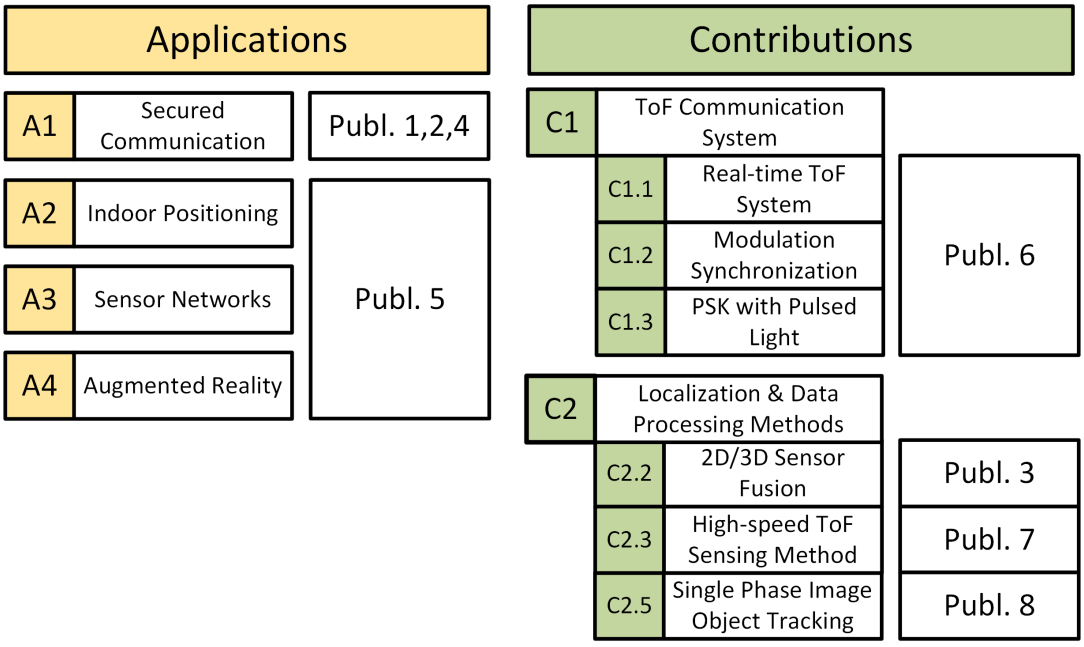


Figure 7.1: The publications of this thesis cover the investigation of potential applications and the development of system designs and methods for location-aware optical communication with Time-of-Flight sensors.

5 reviews the application of indoor positioning, sensor node localization and augmented reality. These applications and their requirements are also presented in Section 3.1.

The second phase of this thesis lead to the development of the evaluation system for optical communication, the ToF tag system and several processing methods. Publication 6 presents the modulation frequency synchronization approach, which is also summarized in Section 4.1.1. The developed data processing methods are introduced in Section 4.2. The color and ToF sensor fusion system is presented in Publication 3. The camera calibration method for sensing fast moving objects at high update rates was introduced in Publication 7. Publication 8 presents a method to estimate the distance to an object from single phase images.

Publication 1: Plank et al., *Optisec3D - a new paradigm in secure communication and authentication featuring Time-of-Flight*, Proceedings of the 2016 International Conference on Embedded Wireless Systems and Networks, 2016.

Publication 2: Plank et al., *Survey on camera based communication for location-aware secure authentication and communication*, EMC2 Summit, CPS Week, 2016.

Publication 3: Plank et al., *High-performance Time-of-Flight and color sensor fusion with image-guided depth super resolution*, Design, Automation and Test in Europe Conference, 2016.

Publication 4: Plank et al., *Time-of-Flight based optical communication for safety-critical applications in autonomous driving*, SAFECOMP, 2016.

Publication 5: Plank et al., *Localization and context determination for cyber-physical systems based on 3D imaging*, Solutions for Cyber-Physical Systems Ubiquity, 2017.

Publication 6: Plank et al., *Synchronization of Time-of-Flight 3D sensors for optical communication*, International Conference on Communication, 2017.

Publication 7: Plank et al., *High-performance indoor positioning and pose estimation with Time-of-Flight 3D imaging*, International Conference on Indoor Positioning and Indoor Navigation, 2017.

Publication 8: Plank et al., *Fast and energy-efficient Time-of-Flight distance sensing method for 3D object tracking*, Digital Image Computing: Techniques and Applications, 2018.

OptiSec3D - A new Paradigm in Secure Communication and Authentication featuring Time-of-Flight

Hannes Plank* Matthias Almer† Robert Lobnik* Christian Steger†
Thomas Ruprechter* Holger Bock* Josef Haid* Gerald Holweg* Norbert Druml*
*Infineon Technologies Austria AG, Design Center Graz
†Graz University of Technology, Institute for Technical Informatics
{hannes.plank, robert.lobnik, thomas.ruprechter, holger.bock, josef.haid,
gerald.holweg, norbert.druml}@infineon.com
matthias.almer@student.tugraz.at steger@tugraz.at

Abstract

Information security and trust represent fundamental requirements for today's information and communication systems as well as interconnected embedded systems. If these requirements are not tackled properly, security attacks, such as relay attacks, may compromise these systems severely.

Here we introduce OptiSec3D, a new paradigm in secure communication systems. The presented approach for secure communication and authentication uniquely enhances the key enabling Time-of-Flight 3D localization technology with optical communication abilities. By extending these features with state-of-the-art security anchors, new levels of trust and security can be reached for information and communication systems.

This work not only provides the background and concept of this secure communication and authentication approach, but also demonstrates its feasible implementation by means of a very first prototype. Furthermore, we outline how security attacks, such as relay attacks, can be counteracted and how future OptiSec3D-based security-critical systems will look like.

Keywords

Information Security, Optical Communication, Contactless Authentication, Time-of-Flight, 3D Localization

1 Introduction

Today, trust and information security is one of the key requirements for embedded systems that are interconnected through wired or wireless communication technologies. If such embedded systems are not designed properly, security breaches may have far-reaching consequences. For example, passive keyless entry and start systems, which are widely used in the automotive domain and deemed secure in the past, have been compromised because of missing distance checks (i.e., evaluation of whether the transponder is in the range of the reader or not). This example, cf. [5], demonstrates that in order to provide secure embedded communication systems, further parameters (such as distance and localization information of communication partners) are essential for proper trust establishment.

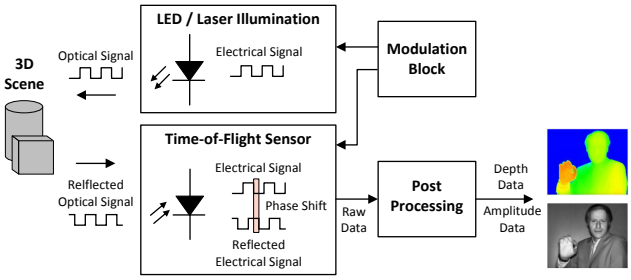


Figure 1. Working principle of PMD-based Time-of-Flight 3D sensing. Obtained with changes from [3].

This paper tackles this gap and proposes an innovative solution for secure communication systems, which is based on the key enabling Time-of-Flight technology and state-of-the-art security anchors (such as Infineon's security controllers). Time-of-Flight is a depth sensing method that provides distance information by measuring the travel time of emitted light. There are direct and indirect Time-of-Flight measurement approaches, cf. [14]. This work focuses on the indirect method, which calculates depth distance information by evaluating the phase shift of an emitted infrared light with the help of photonic mixing devices (PMD), cf. [10]. Figure 1 depicts the basic working principle of a PMD-based camera system. In theory, since the emitted infrared light is modulated, data can also be exchanged between two Time-of-Flight imaging systems.

We propose a new secure communication interface through uniquely extending the Time-of-Flight technology's 3D environment sensing with optical communication capa-

EWSN'16, February 15–17, 2016, Graz, Austria.

bilities and state-of-the-art security anchors. The 3D information gathered by the camera can then be used to verify, e.g., localization information (e.g., both communication partners check if they both measure the same distance) or the communication partner's 3D geometry.

Summarizing, this paper makes the following contributions:

- It introduces OptiSec3D, a new approach for secure communication systems featuring Time-of-Flight depth sensing.
- It outlines how today's information security challenges can be tackled with help of OptiSec3D.
- It demonstrates that the presented concept can be feasibly implemented by means of a very first prototype.

This paper is structured as follows. Section II gives a short introduction into the related work covering the topics of related communication techniques. Furthermore, it discusses today's security challenges in contactless communication. In Section III, our novel security concept based on Time-of-Flight is presented. Followed by Section IV which demonstrates our very first OptiSec3D system prototype. Finally, our results are concluded and some details about our future work are given in Section V.

2 Related Work

2.1 Communication Technologies

Given the fact that Time-of-Flight 3D sensing systems use an active and modulated illumination source (such as an LED or a laser operating in the infrared spectrum, which is depicted in Figure 1), data can be exchanged between two systems. Yuan et al. [16] demonstrated with an early prototype that data transfer between a modulated light source and a Time-of-Flight sensor is possible. The authors developed an LED array, which is able to transfer information to a Time-of-Flight camera. The communication in this system is however one-way only. The LED array system senses the modulation frequency and repeats the signal with information modulated into the phase shift. There is no information sent from the Time-of-Flight camera and furthermore this approach does not deal with security aspects at all.

Another well-known interface technology that relates to the introduced secure communication concept is Near Field Communication (NFC). Today, NFC is used in our everyday life, e.g., in the fields of transportation, payment, loyalty and coupons, logistics, healthcare, and access control (cf. [4]). Given these application fields, the security of an NFC-based reader / transponder system is of high importance; as summarized by the authors in [8] and [7]. In order to make security attacks more difficult, standardized security methods are employed, such as elliptic-curve cryptography. Furthermore, the distance between reader and transponder is limited to a maximum of 10cm: the reader emits an alternating magnetic field, which is used to power the transponder and to exchange data with it. Therefore, by limiting the reader's output power, the range of operation is limited as well. However, a maliciously modified reader, which emits a very strong magnetic field, enables communication over a distance of meters.

In the field of optical communication systems, famous

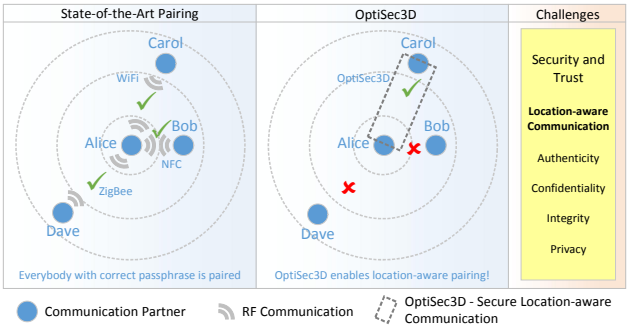


Figure 2. Comparison between state-of-the-art pairing, the OptiSec3D approach, and the challenges ICT systems are confronted with.

and simple solutions are the Consumer IR and IrDA technologies. While Consumer IR is used to control consumer electronics over distances of some meters, IrDA was designed for short range only. A novel technology in the field of optical data communication is Li-Fi, which was used in 2014 in the first commercial product (cf. [13]). Li-Fi works in the visible light spectrum, uses LEDs, and achieves data rates that can compete with RF-based communication, such as WiFi. Harald Haas, the inventor of Li-Fi, predicts that in 25 years Li-Fi technology will be present in every light bulb and thus will lead to the Internet-of-Things, which connects every electronic device to the Internet.

2.2 Challenges in Today's Contactless Authentication Solutions

Providing security and trust represents a fundamental requirement for today's information and communication systems. Ravi et al. outline in [12] the challenges that are faced when designing and developing secure embedded systems. As illustrated in Figure 2, the missing awareness for location and distance is one of the key challenges for today's information and communication technology systems (such as NFC, WiFi, ZigBee). For instance, WiFi- or Bluetooth-based pairing procedures accept all communication partners that provide the correct pass-phrases, regardless of distance or location verifications. This missing verification enables a set of malicious attack vectors.

Several authentication and identification systems, although deemed secure in the past were compromised and spoofed successfully. As an example for a compromised type of authentication system, Francillon et al. showed in [5] that passive keyless entry and start systems, which are widely used in modern cars, can be hacked easily and with only little effort. The authors performed relay attacks on ten cars of eight manufacturers and were able to maliciously unlock, start, and drive these cars in each case. This incident reveals the need for further security measures, in particular location awareness, at system-level. If these keyless entry and start systems implemented proper verification of distance between car and key, such relay attacks would not be successful.

In passive and contactlessly powered RFID and Near Field Communication-based systems, the communication

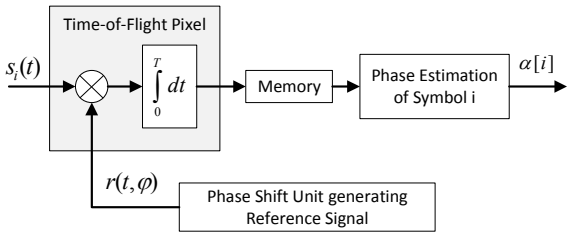


Figure 3. Physical principles of Time-of-Flight-based optical data transmission.

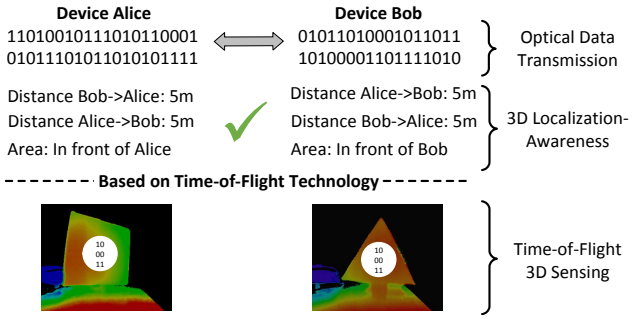


Figure 4. OptiSec3D exploits optical data transfer and 3D localization based on Time-of-Flight.

distance is restricted (typically to a maximum of 10 cm) by the technology. Thus, it is restricted by the strength of the magnetic field, which is emitted by the reader. Again, this poses an issue for the passive transponder, because it is hardly capable to verify the distance to the reader: for example, a maliciously modified reader emitting a very strong magnetic field can enable communication over a distance of meters, which makes a whole set of attacks possible.

In order to make such contactless and RF-based communication more secure, for example, the authors of [1] proposed a distance bounding protocol. This type of protocol estimates the round trip time of radio signals and thus can estimate the distance of the communication partners and can identify whether, e.g., a relay attack takes place. Hancke et al. showed in [6] that distance bounding protocols can also be feasibly implemented for RFID- and NFC-based communication systems. However, in [2], the authors demonstrated that even such sophisticated protocols are not 100% secure against cleverly designed attacks.

As another example, in [11], Prabhakar et al. presented a survey summarizing security issues in today's biometric-based authentication systems. The usage of cameras that are only capable to gather two-dimensional biometric information is one of the identified root causes for these so far unsolved security challenges. For instance, a photo or video can be employed for spoofing face recognition systems. Recently, Krissler demonstrated in [9] the vulnerability of today's fingerprint-based authentication systems: with the help of a fingerprint, which was recreated from high resolution pictures of Germany's Defense Minster taken during a public event, he was able to successfully spoof the biometric system. If biometric-based authentication systems employed 3D camera systems and communication between authentication partners, such spoofing attacks would not be successful at all. These examples support the view that fundamental challenges exist in several information and communication technology systems used today which makes them vulnerable to attacks:

- Communication-based systems usually do not verify distance or area information of their communication partners.
- Vision-based authentication systems usually do not support communication between authentication partners.

Summarized, in order to provide secure authentication and communication for critical fields of applications, it is essential to verify additional parameters (such as distance and localization information of communication partners). Although there is research in this field, there is a major gap in literature concerning the outlined challenges, which is addressed in our work.

3 OptiSec3D

3.1 Physical Principles

The principle of Time-of-Flight based communication works by phase modulation of the emitted infrared light. Figure 3 shows the principle of a pixel capturing the transferred symbol i. Time-of-Flight cameras are designed to measure phase differences between a reference signal $r(t,\phi)$ and incoming light $s_i(t)$. The incoming signal pulses are integrated for the integration time t. A shorter integration time means a higher frame-rate. The phase difference is converted to a voltage by a photonic mixture device, cf. [15], and is then digitized. The difference to the normal depth sensing purpose of Time-of-Flight systems is that the incoming light is sent directly from the communication partner, and the signal amplitude is relatively high. This allows to dramatically shortening the integration time and thus increasing the communication bandwidth. Since lots of pixels sense the optical signal at the same time, phase noise can be limited by averaging the measured phase. Using longer integration times for sensing the communication partner, and using short integration times for communication can be alternated multiple times per second. Furthermore, Time-of-Flight communication is based on modulated infrared light, which ensures high robustness against environmental influences, such as other light-sources.

3.2 Concept and Key Features

As highlighted in Figure 4, OptiSec3D will introduce a radically new communication technology. This is achieved by uniquely extending the key enabling Time-of-Flight 3D sensing technology with optical data communication (while using the same Time-of-Flight hardware) and with state-of-the-art security anchors. Today, there is no solution similar to OptiSec3D. The following key features will be enabled:

- Optical communication in conjunction with 3D location-awareness (e.g., 3D distance and area checks) will make communication-based and authentication-based applications (e.g., secure pairing of devices, payment) more secure and will reduce the probability for malicious attacks.

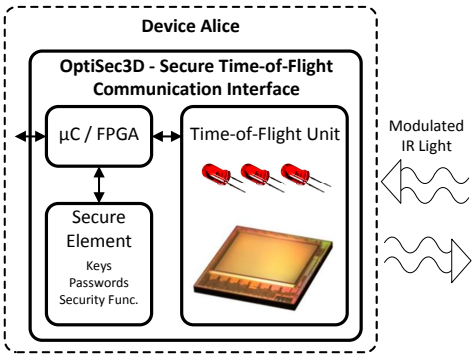


Figure 5. Proposed OptiSec3D communication interface.

- Optical communication in conjunction with 3D shape checks (e.g., biometrics such as face recognition and fingerprints) will make authentication-based applications more secure against spoofing attacks.
- OptiSec3D's unique union of optical communication and 3D sensing will represent a key enabling technology and will open whole new fields of applications (such as long-distance secure pairing of devices by pointing towards the communication partner).

Figure 5 depicts a proposed OptiSec3D communication node from a design perspective. It consists of the Time-of-Flight unit, which is made of the illumination and the sensor units, a micro-controller (or FPGA), and a security anchor given as a secure element. On the one hand, the micro-controller implements Time-of-Flight specific algorithms (such as processing of exchanged data, calculating 3D depth data from raw sensor data, distance and area verifications) and the protocols to the attached embedded system. On the other, the secure element implements security related functionality (such as encryption, decryption, authentication protocols, key handling). Thus, a modular and secure communication technology will be given that can be easily integrated or plugged into embedded systems.

3.3 OptiSec3D-based Secure Authentication

OptiSec3D supports several possibilities to implement secure authentication methods. As a fundamental feature, two Time-of-Flight communication partners easily detect each other through very distinctive amplitude as well as depth values (exemplified by white circuits in Figure 6 and demonstrated in Figure 9) within the sensed 3D-scenery. These values are highly local and are caused by the Time-of-Flight illumination unit. Thus, 3D localization (relative position, distance, etc.) of the communication partner is achieved with little effort. After the communication partner was detected, the authentication procedure can start. To outline some examples, following approaches are feasible.

First, the Time-of-Flight optical communication channel can be used to implement typical asymmetric authentication methods, such as elliptic-curve based cryptography.

Second, the communication partners use the optical Time-of-Flight channel to transmit not only data but also the measured distances. This distance information can then be used by the communication partners to decide, e.g., whether

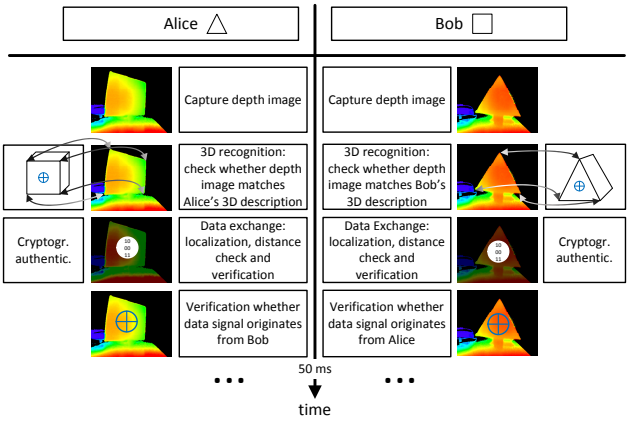


Figure 6. Example of an OptiSec3D secure communication procedure during a certain timeslot.

general Time-of-Flight optical communication between both partners is continued, or the authentication procedure is started, etc.

Third, a Time-of-Flight system can use 3D information stored in a database to compare it with the sensed 3D data. Thus, it is able to recognize the communication partner (e.g., a car, a payment terminal, or a human's biometric information). It is possible to finish the recognition in a very small time interval, which makes it possible to re-authenticate the communication partner multiple times per second. Figure 6 shows a basic example for a secure communication flow employing the Time-of-Flight technology. The procedure in the figure is performed multiple times per second. For example, the majority of secure authentication use-cases features at least one communication partner with static appearance. An example would be a reader at a payment terminal. This known and unchangeable geometry can be stored and provided as a library of 3D datasets. The necessary algorithms to match a depth image with an existing dataset are well established and can be executed in the required timespan.

3.4 Counter Measure against Relay Attacks

Relay attacks are very severe in the RF domain. In the Time-of-Flight domain, distance, 3D localization, shape of the communication partners, and 3D-origin of the communication are known. Furthermore, the Time-of-Flight communication partners use an optical line-of-sight connection. These key-features make security attacks difficult to implement. If, for example, a relay-attacker managed to get into the field-of-view of a Time-of-Flight communication system, he would have to reproduce the exact distances (an authentication only succeeds if both communication partners measure the same distance) between the communication partners. This is outlined in Figure 7. In combination with state-of-the-art security features (message authentication, encryption, authentication protocols, etc.) and the Time-of-Flight's location awareness (communication partners sense depth images of each other and have exact knowledge from which direction and distance the signal of the communication partner originates), further crucial security barriers are given. These security features make it very difficult for attackers, even

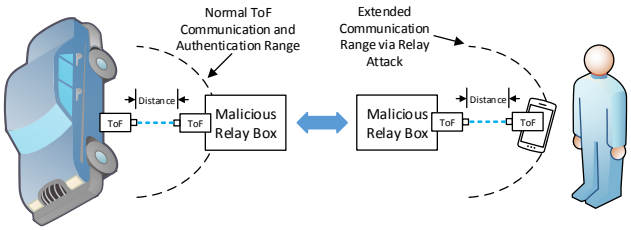


Figure 7. Theoretical concept for a Time-of-Flight based relay attack.

practically impossible, to intercept communication, to spoof distance information, and to maliciously relay the communication. The outlined security measures can be carried out multiple times per second and are thus robust against communication partners in motion. Furthermore, the computation and verification tasks can be done in parallel and do not halt or delay communication.

4 System Prototype and Results

In order to proof the feasibility of the OptiSec3D concept, a very first prototype system was developed. Figure 8 shows the basic approach. Two Infineon REAL3™ sensor evaluation boards, which are based on Time-of-Flight technology of pmdtechnologies, are facing each other. A common clock source is provided to the sensors, in order to avoid carrier detection and clock synchronization issues. Both evaluation boards are connected to a PC through USB 3.0. Matlab is used to configure the cameras, to process the data, and to evaluate the results.

4.1 Depth Sensing and Detection of Devices

A very important requirement is to ensure an easy detectability of transmitting communication partners within the 3D environment. For this detectability test, the two evaluation boards of the prototype system were operated in depth-sensing mode employing the same modulation frequency (but without synchronization). Figure 9 shows Time-of-Flight amplitude and depth data that displays the Time-of-Flight communication partner (implementing two illumination LEDs) mounted on a wooden fixing. Within the left images, the communication partner has its illumination LEDs

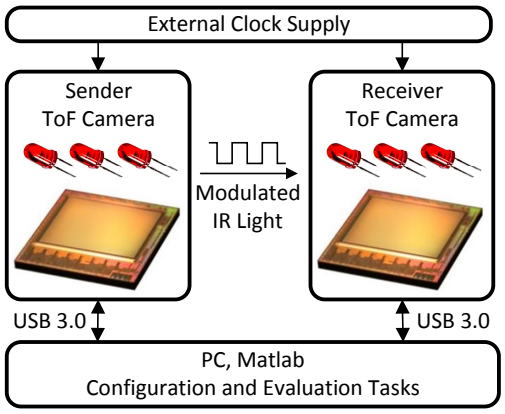


Figure 8. Proposed OptiSec3D communication interface.

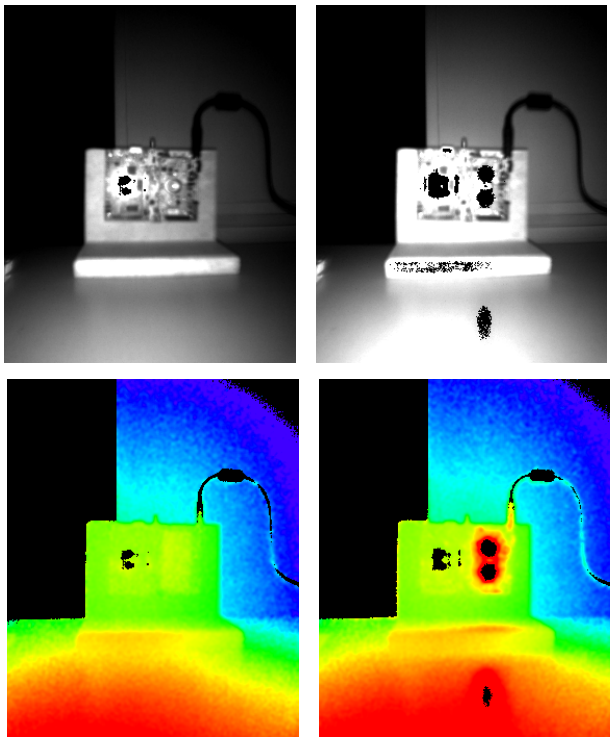


Figure 9. Time-of-Flight amplitude and depth data showing another camera with deactivated illumination (left image) and activated illumination (right image).

deactivated. Whereas, within the right images, the communication partner has its illumination LEDs activated. Thanks to this local and very distinctive amplitude and depth inconsistencies, an OptiSec3D device is able to detect other transmitting communication partners in the 3D environment with only little effort.

4.2 Optical Data Transfer

Successful data transfer between Infineon's REAL3™ evaluation boards can be practically achieved through various approaches. In the following, two implementations are exemplarily presented.

In the first approach, the sender camera emits modulated light with a phase of, e.g., 0° and 180°. The receiving camera samples with a higher frequency and evaluates the Time-of-Flight pixels' raw data. No depth or amplitude images are calculated (cf. Figure 1). The visualization of the sensed raw data is given by Figure 10. One can easily detect the two raw data peaks, which prominently stand out from the noise-level, and which are caused by the two illumination LEDs of the sender. Given this behavior and by considering a Time-of-Flight system's characteristics, one can implement a basic Time-of-Flight-based communication system by evaluating the sensed raw data only.

In the second approach, a higher order modulation scheme is demonstrated, such as an 8-PSK. For this purpose, the receiving device now implements a proper sampling and phase calculation method. The transmitted phase and amplitude information is sampled four times and calcu-

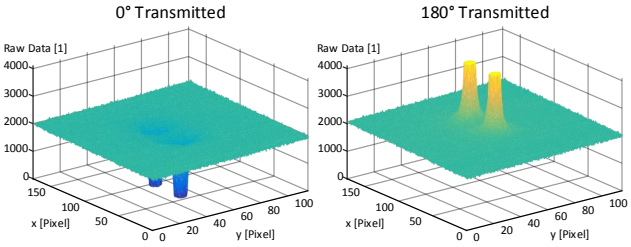


Figure 10. Sensed raw data of the sender who transmits light with 0°(left image) and 180°(right image) phase shift.

lated according to Equation (1) and (2) by the receiver. Thus, a more robust communication can be achieved that also provides higher data rates. For demonstration purposes, Figure 11 depicts the sensed and calculated phase images on receiver side.

$$\varphi = \arctan\left(\frac{A_{90^\circ} - A_{270^\circ}}{A_{0^\circ} - A_{180^\circ}}\right) \quad (1)$$

$$A = \frac{\sqrt{(A_{90^\circ} - A_{270^\circ})^2 + (A_{0^\circ} - A_{180^\circ})^2}}{2} \quad (2)$$

5 Conclusion

Information security is a major challenge in today's interconnected world. Recent concepts, such as the Internet-of-Things, further intensify the need for secure information and communication technologies as well as dedicated hardware security which is hardened against all kinds of security attacks.

This paper introduces OptiSec3D, a new approach for secure information and communication systems. OptiSec3D uniquely enhances the key enabling Time-of-Flight 3D localization technology with optical communication abilities and state-of-the-art security anchors. Thus, new levels of trust and security can be reached for information and communication systems. This work demonstrates OptiSec3D's feasible implementation with the help of a very first prototype. Furthermore, it shows how the OptiSec3D technology can counteract sever security attacks, such as relay attacks.

Our future work concerns the development of an OptiSec3D communication node that also employs dedicated hardware security (i.e. security controllers) for security-critical applications.

6 Acknowledgments

The authors would like to thank the Austrian Federal Ministry for Transport, Innovation and Technology as well as the ARTEMIS Joint Undertaking, which funded the research activities under the grant agreement n° 852328 and n° 621429 respectively.

7 References

[1] S. Brands and D. Chaum. Distance-bounding Protocols. In *Workshop on the Theory and Application of Cryptographic Techniques on Advances in Cryptology (EUROCRYPT)*, pages 344–359, 1994.

[2] C. Cremers, K. Rasmussen, B. Schmidt, and S. Capkun. Distance Hijacking Attacks on Distance Bounding Protocols. In *IEEE Symposium on Security and Privacy (SP)*, pages 113–127, May 2012.

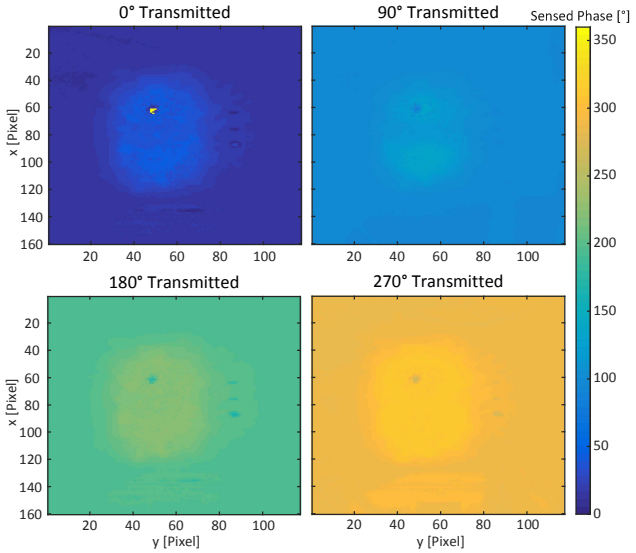


Figure 11. Sensed and calculated phase data of the sender who transmits data symbols of, e.g., 0°, 90°, 180°, and 270° phase information.

[3] N. Druml, G. Fleischmann, C. Heidenreich, A. Leitner, H. Martin, T. Herndl, and G. Holweg. Time-of-Flight 3D Imaging for Mixed-Critical Systems. In *13th International Conference on Industrial Informatics (INDIN)*, pages 1432–1437, July 2015.

[4] K. Finkenzeller. *RFID Handbook: Fundamentals and Applications in Contactless Smart Cards and Identification*. John Wiley & Sons, Inc., New York, NY, USA, 2 edition, 2003.

[5] A. Francillon, B. Danev, and S. Capkun. Relay Attacks on Passive Keyless Entry and Start Systems in Modern Cars. In *Network and Distributed System Security Symposium (NDSS)*, February 2011.

[6] G. Hancke and M. Kuhn. An RFID Distance Bounding Protocol. In *First International Conference on Security and Privacy for Emerging Areas in Communications Networks (SecureComm)*, pages 67–73, September 2005.

[7] E. Haselsteiner and K. Breitfu. Security in Near Field Communication (NFC). Strengths and weaknesses. In *Workshop on RFID Security*, 2006.

[8] M. Hutter, J.-M. Schmidt, and T. Plos. RFID and Its Vulnerability to Faults. In *Proceedings of the 10th International Workshop on Cryptographic Hardware and Embedded Systems (CHES)*, 2008.

[9] J. Krissler. Ich sehe, also bin ich ... Du. Gefahren von Kameras für (biometrische) Authentifizierungsverfahren, 2014.

[10] R. Lange and P. Seitz. Solid-state time-of-flight range camera. *IEEE Journal of Quantum Electronics*, 37(3):390–397, March 2001.

[11] S. Prabhakar, S. Pankanti, and A. Jain. Biometric recognition: security and privacy concerns. *Security Privacy, IEEE*, 1(2):33–42, March 2003.

[12] S. Ravi, A. Raghunathan, P. Kocher, and S. Hattangady. Security in Embedded Systems: Design Challenges. *ACM Transactions on Embedded Computing Systems (TECS)*, 3(3):461–491, August 2004.

[13] N. Savage. Li-Fi gets ready to compete with Wi-Fi. *Spectrum, IEEE*, 51(12):13–16, December 2014.

[14] J. Y. Wang. Imaging laser radar - An overview. In R. W. McMillan, editor, *Lasers '86; Proceedings of the Ninth International Conference on Lasers and Applications*, pages 19–29, 1987.

[15] Z. Xu, T. Möller, H. Kraft, J. Frey, and M. Albrecht. Photonic mixer device, April 2008. US Patent 7,361,883.

[16] W. Yuan, R. Howard, K. Dana, R. Raskar, A. Ashok, M. Gruteser, and N. Mandayam. Phase messaging method for time-of-flight cameras. In *IEEE International Conference on Computational Photography (ICCP)*, May 2014.

Survey on Camera based Communication for Location-Aware Secure Authentication and Communication

Hannes Plank*, Christian Steger†, Thomas Ruprechter*, Gerald Holweg*, Norbert Druml*
*Infineon Technologies Austria AG, Graz, Austria
†Institute for Technical Informatics, Graz University of Technology, Austria
{hannes.plank, thomas.ruprechter, gerald.holweg, norbert.druml}@infineon.com, steger@tugraz.at

***Abstract*—Contactless communication and authentication in the RF-domain is still impaired from relay attacks. Countermeasures like distance bounding protocols are difficult to implement properly and are still continuously broken by sophisticated attacks. This work reviews a variety of approaches to counteract relay attacks by adding location-awareness to the security concept. Camera based optical line of sight communication is an alternative to RF communication and provides location-awareness by the ability of sensing communication partners. Concepts and methods of this field are introduced and their contribution to location-aware security analyzed.**

Finally, we present OptiSec3D, which is our novel approach to use Time-of-Flight 3D cameras for secure communication. Our proposed system combines state-of-the-art cryptographic security with 3D depth recognition and camera based optical communication. Various attack scenarios are identified and discussed. We show how our system can improve the security of previous camera-based communication systems and how it can drastically reduce the risk of relay attacks.

***Keywords*—Time-of-Flight, contactless authentication, camera based communication, security, 3D sensing**

I. Introduction

The demand for secure contactless authentication and communication methods is growing steadily. Applications range from mobile payment, access control, device pairing, electronic passports to keyless car entry and start systems. In these applications, the devices face the challenge to establish a secure connection over an insecure channel. While cryptographic methods, like Diffie-Hellman key exchange, digital signatures and message authentication can be combined to a secure protocol, the risk of relay attacks persists.

An example of a relay attack on a keyless car entry system is illustrated in Fig. 1. An attacker relays the communication between an entry token to the car. Access can be gained without the owner's presence. Francillon et al. [1] successfully executed this attack on 10 different car models and were able to gain entry and to start the engine. Distance bounding protocols can limit the maximum allowed distance between two communication partners, but have been successfully attacked. If both communication partners can determine their relative position in a reliable way, the vulnerability against relay-attacks might improve [2].

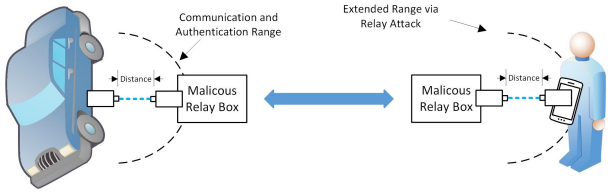


Fig. 1. Concept of a relay attack on passive keyless car entry systems

This work discusses security aspects of location-awareness with a focus on camera based communication systems. Camera based communication systems can have the ability to sense their communication partner, and visually detect attackers. However, most approaches have a limited range and just rely on the optical channel for security. We introduce our OptiSec3D approach, which overcomes the boundaries of other camera based communication systems by using Time-of-Flight 3D cameras. Our security concept relies on state-of-the-art hardware-accelerated cryptography in combination with Time-of-Flight 3D sensing and its optical communication ability.

II. Time-of-Flight 3D Sensing

Time-of-Flight is a sensing technology that provides distance information by measuring the travel time of emitted light. There are two fundamental Time-of-Flight principles, the direct and the indirect measurement approaches. The indirect method, which is illustrated in Fig. 2, measures distance information by analyzing the phase shift between the emitted modulated infrared light and the reflected light with the help of

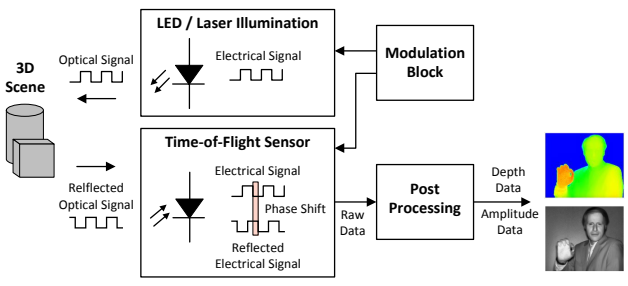


Fig. 2. The concept of PMD-based Time-of-Flight 3D sensing, obtained with changes from [3].

photonic mixing devices (PMD), as explained by the authors in [4].

Thanks to this concept (monocular, no baseline required, little power consumption), PMD-based Time-of-Flight cameras can be easily miniaturized and integrated into small embedded devices in a very robust way. The raw data, which is gathered by the sensor, is then post processed by a processing pipeline. The outcome of this pipeline is 3D depth information of the scenery and amplitude information, which can be used as a robust infrared image. Given these features, future Time-of-Flight camera systems will enable the concept of *ubiquitous real-time geometry devices* that provide 3D data for our everyday applications.

III. Location-Awareness in the RF-domain

This section reviews approaches aiming to tackle the remaining problems of secure communication over an insecure channel by introducing distance or location-awareness. Relay attacks, also named mafia fraud attacks, prove to be sever and difficult to avoid. As listed by Fischlin et al. [5], relay attacks have been successfully implemented on applications such as Bluetooth [6], [7], smartcards [8], [9], [10], [11], electronic passports [12], voting systems [13] and passive keyless car entry systems [1].

Distance bounding protocols [14] are the most common protection against relay attacks. The core concept is to introduce an upper bound on the distance between communication partners. This bound is usually enforced by timing the delay between a sent message the reply of the partner. This limited reply time hampers an attacker to rely the message to a third party which is not within this boundary.

Brelurut et al. [15] provide a recent survey about distance bounding protocols. According to Clulow et al. [16], some hardware is not fit to handle the tight timing constraints of distance bounding. Sometimes, the clocks are not precise enough to reliably employ distance bounding. The protocols are often just implemented in software layers, which also doesn't allow strict distance bounding. This is due the fact that the time a message travels through the air is very low compared to the response time of an electronic device. An attacker can be faster at relaying messages, spoofing the original distance. There are approaches, like Clulow et al. outlined in [16], to avoid this by using probabilistic models about timing behavior to detect attackers. Despite the past research effort, which resulted in over 40 distance bounding protocols, they still show vulnerability against attacks [16], [17]. Cremers et al. [17] proposed distance hijacking attacks, where the attacker uses unsuspecting third parties to verify the distance measurement.

If two devices are not just aware of their distance, but can securely determine their relative location, their vulnerability for relay attacks might be reduced [2]. Using the characteristics of wireless networks for indoor localization has been well researched and was surveyed by Zhu et al. [18]. A lot of work has been accomplished on the localization in Wifi networks. The challenge of Wifi positioning is that Wifi networks were not designed for localization. The main motivation for Wifi localization is not security, but indoor navigation and location based services.

Wifi fingerprinting is a well-established technique, offering coarse localization. It requires a time-consuming calibration of the signal strength indicator (RSSI) in the network [19]. Because of the Wifi signal characteristics, the RSSI depends on the location. The problem with Wifi fingerprinting is that it requires an RSSI map of the Wifi access point and is thus volatile to environmental changes. SecureAngle [20] is a system to overcome fingerprinting based techniques by using multi-antenna access points to create unique angle signatures of the clients. There are approaches to enrich Wifi-based localization with additional signals, cf. Niu et al. with Zigbee [21]. Tippenhauer et al. [22] show that Wifi-based localization can be spoofed by replaying localization signals. There are countermeasures possible but according to Tippenhauer, Wifi-based localization cannot be reliably used in security applications.

A different paradigm for localization is sound positioning [23]. According to Clulow [16], sound positioning offers precise localization due to the slower propagation speed of sound. Sound-based security mechanisms are however vulnerable to relay attacks because the communication can be relayed over a faster medium.

Visible light communication (VLC) has recently gained momentum through the rise of the Internet of Things (IoT). The security aspects of visible light communications are discussed by Rohner et al. [24] in their survey. Physical security features of the visible light channel were introduced by Mostafa et al. in [25] and [26]. They use null-steering, artificial noise and beamforming to mitigate eavesdropping.

IV. Camera-Based Communication

There are several approaches to use cameras for optical communication. The obvious benefit for secure communication is that due to the optical line-of-sight communication principle, an attacker needs to be within the field of view of the cameras. This section provides an overlook on camera communication principles. Security aspects of camera communication are reviewed in Section V.

Analyzing image data can lead to recognition of the communication partner and validation of its location. The limited frame-rate of common image sensors is compensated by the high numbers of pixels. This guarantees enough bandwidth for a lot of applications. The most common way to use 2D cameras for communication is using displays to stream 2D barcodes. This principle is known as pixelated multiple input, multiple output (MIMO) and was first introduced by Hranilovic and Kschischang [27]. A good overview on barcode generation and streaming is provided by Kokan and Gupta [28]. Hao et al. [29] propose a system to use colored barcodes to reach a good performance on small displays like smartphones.

It is also possible to embed imperceptible information into the display output. Li et al. [30] demonstrate a method of en-

coding information with pixel translucency changes. Humans are unable to see the modulation but cameras can use it for communication.

The previously introduced concepts were based on sending information over a display device. With the Bokode method [31], it is possible to transmit information to cameras just with a single point light source. When cameras are out-of-focus, light sources arrive as focus spots on the sensor. In the Bokode system, the light rays of the point source are arranged in a coded pattern. When the rays arrive at the out-of-focus camera, this pattern is projected onto the sensor. This enables a two way optical communication link without bulky displays. The distance of Bokode is restricted similarly as conventional barcode displaying methods, however the angle is also restricted to under 20 degrees.

While optical data streaming with 2D-cameras works with sufficient reliability and bandwidth, these systems are either bulky or limited in range and angle. The next section introduces the communication ability of Time-of-Flight 3D cameras and how it can tackle the flaws of 2D camera based communication.

A. Communication with Time-of-Flight 3D Cameras

In the past years, depth sensing cameras became increasingly popular and sophisticated. The smallest system are Time-of-Flight cameras, which are close to becoming largely available in consumer devices [32]. To best of our knowledge, the only attempt to use Time-of-Flight cameras for communication has been accomplished by Yuan et al. [33]. Yuan proposes a system, which enables a one-way communication link from an array of LEDs to a ToF camera. The signal from the Time-of-Flight illumination unit is sensed by IR-photodiodes, amplified and sent back phase-modulated via the LED array.

Our communication method [34] uses Time-of-Flight cameras for communication and mutual depth sensing. Unlike the previously introduced MIMO approaches, the signal originates from the Time-of-Flight illumination unit, which is a single laser or LED. The information is transmitted by using the camera's phase-shifting feature. Hence just two Time-of-Flight cameras without any additional communication hardware are required for our approach. The necessary bandwidth is reached by employing high frame-rates and adapting the Time-of-Flight post-processing pipeline. We developed a proof-of-concept prototype, as shown in Fig. 3, which demonstrates the communication ability of Time-of-Flight cameras. This optical communication principle is used by our proposed OptiSec3D security concept, which is explained further in Section VI.

V. Cameras in Secure Communication

While imaging devices are frequently used in biometric access control systems (face-, iris- hand-, and veinscanning), the use for secure device authentication is not so well explored. The Seeing-is-Believing (SiB) system of McCune et al. [35] is the first use of cameras for secure device authentication to our knowledge. It is based on generating a barcode, containing a hash of a device's public key. The visual channel is assumed

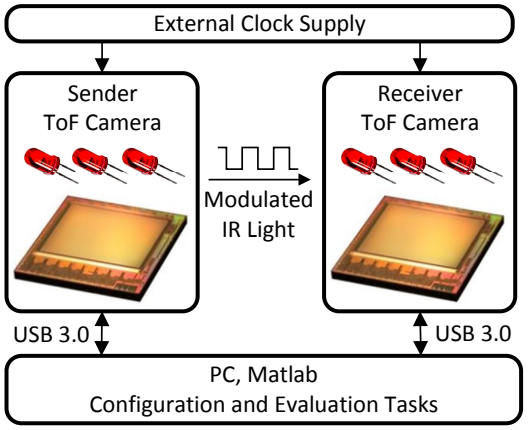


Fig. 3. The design principle of our proof-of-concept prototype.

to be secure in this work, and the rest of the authentication runs over an insecure RF-channel. Saxena et al. improve the SiB principle to devices with limited displaying capabilities like LEDs.

The SBVLC system proposed from Zhan et al. [36] aims to be an alternative to NFC by securely transmitting barcodes using smartphones. The authors create and test a 2D/3D geometric security model, to analyze the eavesdropping possibility. A similar concept is CamTalk [37], which like the other approaches, relies on the presumption that the visual channel is relatively immune to attacks.

These camera based secure communication approaches presume that two good devices face each other. It is however a possible attack scenario that a user unknowingly points a device at a malicious relay box. Camera to display communication suffers practical restrictions, such as range and system size. We thus developed the OptiSec3D security concept, which is based on Time-of-Flight 3D sensing. It is introduced in the next section and aims to defeat the identified relay attack scenarios, while being more convenient than 2D-camera based systems.

VI. The OptiSec3D Security Concept

The OptiSec3D security concept enables secure communication and unified cryptographic and 3D location-aware authentication between two devices over the optical channel. Infineon's REAL3™ camera system, which is based on Time-of-Flight technology of PMDTechnologies, is used for secure communication and 3D-sensing. The proposed system combines several security features to a novel concept: **State-of-the-art cryptography** can counteract most of the known attack scenarios but is vulnerable against relay attacks. Relay attack attempts on the proposed system are very easy to detect due to the **nature of optical line of sight data transmission** alone. With the ability of the communication partners to sense each other by depth imaging, this adds the security feature of **machine recognition**. 3D authentication, integrity tracking and continuous **secure mutual position checking** can effectively defeat relay attacks.

A. State-of-the-Art Cryptographic Authentication and Communication

The common way to establish a secure connection over an insecure channel is to agree to a common secret. This common secret is used to derive a common key which is used to efficiently encrypt messages with a symmetric encryption method such as AES. The limited bandwidth of a Time-of-Flight camera based communication system calls for a cryptographic key exchange and authentication protocol requiring as little data exchange as possible. Elliptic curve based cryptography (ECC) allows to securely use short keys and is thus suited well for our purpose. The Elliptic Curve based Diffie-Hellman (ECDH) key exchange protocol enables the communication partners Alice and Bob to establish their common secret, while communicating over the insecure optical channel. The common key can't be derived by a third party in an easy way as it requires solving the Diffie-Hellman problem, which is mathematically complex. If however an attacker manages to listen and to alter the messages between Alice and Bob, this exchange is compromised. To counter these man-in-the-middle attacks, the messages need to be authenticated by using a digital signature. Digital signatures can prove that a message originates from a certain communication partner.

When a common secret is securely established, it is used to derive a common key for fast symmetric encryption. Message authentication codes (MAC) ensure the integrity of encrypted messages. A checksum of the message and a shared secret are combined by Alice and attached to the message. Bob uses the shared secret to compute the MAC himself and compares it with Alice's version. A nonce is a number which changes during every transaction. Adding this number to the unencrypted message prevents the attacker from recording and replaying messages.

These cryptographic principles are combined to a protocol. Numerous established light-weight cryptographic protocols are already available. The protocol is embedded into the security concept of OptiSec3D and provides cryptographic trust on the message-transfer level. A hardware security controller can be used as security anchor, e.g., for secure key-storage and numerous hardware accelerated cryptographic operations.

B. 3D Recognition

Using the depth (and maybe amplitude) images gathered by the Time-of-Flight cameras, it is possible for the communication partners to recognize each other. An initial 3D geometry recognition ensures a higher level of security. Not only the geometry of the communication partner is detected, but also the expected source of modulated light. This alone makes relay attacks very hard, because an attacker would need to be able to emit light from the same location as the communication partner.

An example is the contactless payment use-case, where a payment device detects the 3D shape of a payment terminal. Since the shape and source of light are stored in a database, the device detects a relay attack if the information originates from a second malicious device.

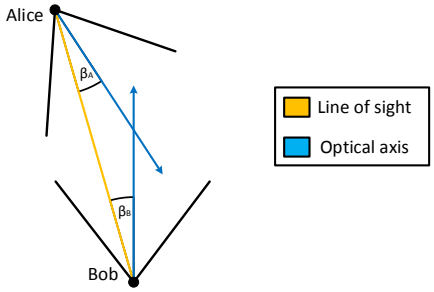


Fig. 4. Alice and Bob both measure their distance (length of the line of sight) and the angles β_A and β_B.

It is sufficient for just one device to sense a malicious device to prevent relay attacks. In the case of a payment terminal, it does not make sense to track the customer in 3D because an attacker would look the same. Whether 3D recognition is performed continuously or just at the beginning of the communication phase is subject to an evaluation.

C. 3D Integrity Checking

Integrity tracking of the communication partner ensures that no third party can secretly alter or relay the transmission after the initial 3D-recognition.

Assuming that geometrical appearance does not change during the communication, time-consuming continuous 3D recognition can be replaced by tracking geometrical changes. As depth image based simultaneous localization and mapping (SLAM) approaches like Kinect Fusion by Izadi et al. [38] proofed to be reliable, it is possible to estimate the relative camera pose between two captured depth frames. This allows communication partners to be in motion while the system is still capable to detect an interfering malicious device.

D. 3D Position Verification

While 3D recognition can prevent a relay attack originating from a different location, it does not cover an imposing device with the same geometric properties. The countermeasure against this attack scenario is to track the relative position of the other device during the communication. Due to the strong signal of the modulated light source, each system can detect the distance and angle (β) between the line of sight (LOS) and the optical axis, as shown in Fig. 4. Each device regularly forwards the other's measured distance and angle β as encrypted message. These measured parameters are supposed to be similar for each device and dissimilar parameters indicate relay attacks. Because it is very hard to spoof Time-of-Flight depth measurements, this is the most important security feature of OptiSec3D and can be enough to avert relay attacks.

E. 3D Secure Communication Protocol

The introduced security concepts are combined to a secure communication protocol. Unlike traditional security protocols, the proposed protocol includes the Time-of-Flight technology

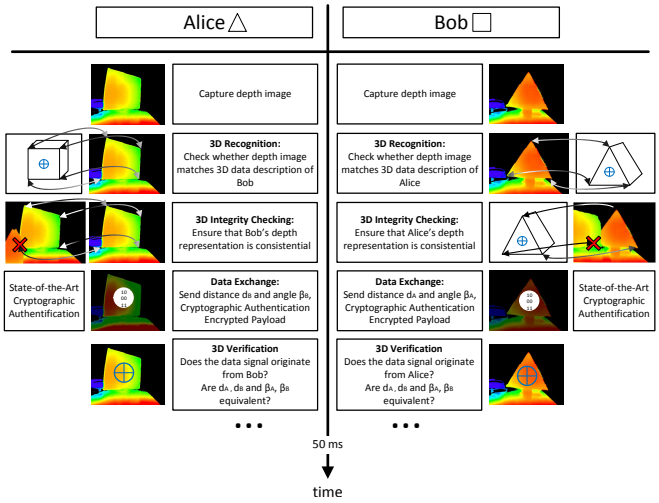


Fig. 5. Activities during a time-frame to implement the 3D security protocol

by introducing phases of communication and depth measurements and evaluation. The secure communication is managed by a state-of-the-art cryptographic protocol, as described in Section VI. Fig. 5 illustrates the idea of a future time-frame in the protocol. The Time-of-Flight cameras alternate between communication and depth sensing. It employs the previously discussed 3D security features. It is not possible to transmit information and sense depth at the same time.

An important part of the security protocol is to exchange encrypted messages containing the measured distance and angle β of the communication partner. Relay attacks are detected, if the measured positions are inconsistent.

VII. Evaluation of Potential Threats

As mentioned in Section III, relay attacks are one of the biggest remaining risks for secure communication and authentication applications. The OptiSec3D concept aims to make these attacks unfeasible, and hence we identify and analyze relay attack scenarios in this section.

A. General relay attack

Due to the nature of optical communication, an attacker would need to get into the view of the camera systems. This is so unlikely that previous camera based security approaches, which were introduced in Section V, assume this channel to be sufficiently secure [35]. The longer reachable distance and small lightsource (compared to a display) of Time-of-Flight cameras, can make this kind of attack feasible if there are no countermeasures. The OptiSec3D approach hence does not rely on the optical channel for security.

The 3D recognition can estimate the expected position of the light source. Signals originating from any different location are not accepted. This restricts the communication coarsely to the line of sight between the illumination unit and image sensor. As the other attack scenarios show, 3D recognition is only an additional feature and not mandatory to prevent relay attacks.

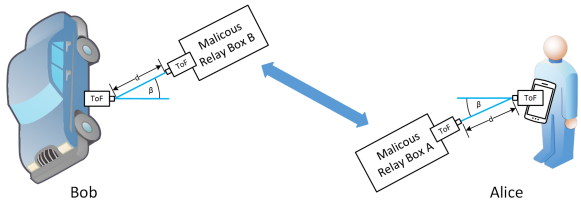


Fig. 6. Distance and angle checking: A relay attack on OptiSec3D would need to recreate the exact angle β and distance d with relay box B.

B. Relay attack with 3D spoofing

In this scenario, an attacker tries to influence the camera system to take a wrong depth measurement. There have been attacks to depth sensing systems in the field of biometric face-scanning, using 3D-masks [39]. However spoofing depth measurements without physically changing the environment are nearly impossible due to the nature of the Time-of-Flight depth sensing principle. The next scenario shows, why OptiSec3D is not just immune to spoofing, but also secure against relay attacks even if a relay box geometrically impersonates a valid communication partner.

C. Relay attack with a malicious duplicate device

This scenario describes an attack, where the communication partner has been replaced with a malicious relay box. This attack is prevented by mutual distance and angle measurement and verification, as described in Section VI-D. Fig. 6 illustrates the situation for an attacker. Relay box B would need to have the exact relative position as Alice and relay box A. Even then, the attack could be detected by 3D recognition.

D. Relay attack by laser reflection

We consider this an unlikely attack scenario. It involves an attacker pointing a modulated IR laser beam onto a communication partner to override its light source. Despite the fact that a IR-laser could create a signature similar to the illumination unit of a Time-of-Flight system, this attack faces the same issues as the previous duplicate device attack scenario.

VIII. Conclusion

In this work, we addressed relay attacks, which still pose a threat to RF-based secure communication. Concepts to enhance security in this field through location-awareness were reviewed. We showed camera based communication efforts, and how they could benefit security applications.

Finally, we introduced our OptiSec3D security concept. It is a work-in-progress concept, supported by a proof-of-concept prototype. Attack scenarios were identified and analyzed in this work. We conclude, that our system can immensely reduce the risk of relay attacks.

Future work involves the implementation of the OptiSec3D concept in form of a secure communication system, which can be easily integrated into embedded systems.

ACKNOWLEDGMENT

The authors would like to thank the Austrian Federal Ministry for Transport, Innovation and Technology as well as the ARTEMIS Joint Undertaking, which funded the research activities under the grant agreement n° 852328 and n° 621429 respectively.

REFERENCES

[1] A. Francillon, B. Danev, and S. Capkun, "Relay Attacks on Passive Keyless Entry and Start Systems in Modern Cars," in *Network and Distributed System Security Symposium (NDSS)*, February 2011.

[2] N. Chandran, V. Goyal, R. Moriarty, and R. Ostrovsky, *Advances in Cryptology - CRYPTO 2009: 29th Annual International Cryptology Conference*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, ch. Position Based Cryptography, pp. 391–407.

[3] N. Druml, G. Fleischmann, C. Heidenreich, A. Leitner, H. Martin, T. Herndl, and G. Holweg, "Time-of-Flight 3D Imaging for Mixed-Critical Systems," in *13th International Conference on Industrial Informatics (INDIN)*, July 2015, pp. 1432–1437.

[4] R. Lange and P. Seitz, "Solid-state time-of-flight range camera," *IEEE Journal of Quantum Electronics*, vol. 37, no. 3, pp. 390–397, March 2001.

[5] M. Fischlin and C. Onete, *Applied Cryptography and Network Security: 11th International Conference (ACNS)*. Springer Berlin Heidelberg, 2013, ch. Terrorism in Distance Bounding: Modeling Terrorist-Fraud Resistance, pp. 414–431.

[6] A. Levi, E. Çetintaş, M. Aydos, Ç. K. Koç, and M. U. Çağlayan, *Computer and Information Sciences - ISCIS 2004*, ch. Relay Attacks on Bluetooth Authentication and Solutions.

[7] K. Haataja and P. Toivanen, "Two practical man-in-the-middle attacks on bluetooth secure simple pairing and countermeasures," *IEEE Transactions on Wireless Communications*, vol. 9, no. 1, pp. 384–392, January 2010.

[8] S. Drimer and S. J. Murdoch, "Keep your enemies close: Distance bounding against smartcard relay attacks," in *USENIX Security Symposium on USENIX Security Symposium*, ser. SS'07, 2007. [Online]. Available: http://dl.acm.org/citation.cfm?id=1362903.1362910

[9] G. Hancke, "A practical relay attack on iso 14443 proximity cards," Tech. Rep., 2005.

[10] Z. Kfir and A. Wool, "Picking Virtual Pockets using Relay Attacks on Contactless Smartcard," in *Security and Privacy for Emerging Areas in Communications Networks*, September 2005, pp. 47–58.

[11] M. Roland, J. Langer, and J. Scharinger, "Applying relay attacks to Google Wallet," in *5th International Workshop on Near Field Communication (NFC)*, February 2013, pp. 1–6.

[12] T. R. Martin Hlavac, "A Note on the Relay Attacks on e-passports: The Case of Czech e-passports," in *IACR Eprint archive*, 2007. [Online]. Available: http://eprint.iacr.org/2007/244.pdf

[13] Y. Oren and A. Wool, "Rfid-based electronic voting: What could possibly go wrong?" in *RFID, IEEE International Conference*, April 2010.

[14] S. Brands and D. Chaum, "Distance-bounding Protocols," in *Workshop on the Theory and Application of Cryptographic Techniques on Advances in Cryptology (EUROCRYPT)*, 1994, pp. 344–359.

[15] A. Brelurut, D. Gerault, and P. Lafourcade, "Survey of Distance Bounding Protocols and Threats," in *International Symposium on Foundations & Practice of Security (FPS)*, 2015.

[16] J. Clulow, G. P. Hancke, M. G. Kuhn, and T. Moore, *Security and Privacy in Ad-Hoc and Sensor Networks*, 2006, ch. So Near and Yet So Far: Distance-Bounding Attacks in Wireless Networks.

[17] C. Cremers, K. Rasmussen, B. Schmidt, and S. Capkun, "Distance hijacking attacks on distance bounding protocols," in *IEEE Symposium on Security and Privacy (SP)*, May 2012, pp. 113–127.

[18] L. Zhu, A. Yang, D. Wu, and L. Liu, *International Conference on Life System Modeling and Simulation*. Springer Berlin Heidelberg, 2014, ch. Survey of Indoor Positioning Technologies and Systems.

[19] L. Schauer, F. Dorfmeister, and M. Maier, "Potentials and limitations of WIFI-positioning using Time-of-Flight," in *International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, Oct 2013.

[20] J. Xiong and K. Jamieson, "SecureAngle: Improving Wireless Security Using Angle-of-arrival Information," in *ACM SIGCOMM Workshop on Hot Topics in Networks*, ser. Hotnets-IX, New York, NY, USA, 2010, pp. 11:1–11:6.

[21] J. Niu, B. Wang, L. Shu, T. Duong, and Y. Chen, "ZIL: An Energy-Efficient Indoor Localization System Using ZigBee Radio to Detect WiFi Fingerprints," *IEEE Journal on Selected Areas in Communications*, 2015.

[22] N. O. Tippenhauer, K. B. Rasmussen, C. Pöpper, and S. Čapkun, "Attacks on public wlan-based positioning systems," in *International Conference on Mobile Systems, Applications, and Services*, ser. MobiSys '09. New York, NY, USA: ACM, 2009, pp. 29–40.

[23] C. De Marziani, J. Urefia, A. Hernandez, M. Mazo, F. Alvarez, J. Garcia, J. Villadangos, and A. Jimenez, "Simultaneous measurement of times-of-flight and communications in acoustic sensor networks," in *IEEE International Workshop on Intelligent Signal Processing*, September 2005, pp. 122–127.

[24] C. Rohner, S. Raza, D. Puccinelli, and T. Voigt, "Security in visible light communication: Novel challenges and opportunities," *Sensors & Transducers Journal*, 2015.

[25] A. Mostafa and L. Lampe, "Physical-layer security for indoor visible light communications," in *IEEE International Conference on Communications (ICC)*, June 2014, pp. 3342–3347.

[26] A. Mostafa and L. Lampe, "Physical-Layer Security for MISO Visible Light Communication Channels," *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 9, pp. 1806–1818, September 2015.

[27] S. Hranilovic and F. Kschischang, "A pixelated mimo wireless optical communication system," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 12, no. 4, pp. 859–874, July 2006.

[28] S. G. Vidyashree Kokane, "A Survey on Barcode Generation and Modulation Techniques," *IEEE Journal of Selected Topics in Quantum Electronics*, 2013.

[29] T. Hao, R. Zhou, and G. Xing, "Cobra: Color barcode streaming for smartphone systems," in *Proceedings of the 10th International Conference on Mobile Systems, Applications, and Services*, ser. MobiSys '12. New York, NY, USA: ACM, 2012, pp. 85–98.

[30] T. Li, C. An, X. Xiao, A. T. Campbell, and X. Zhou, "Real-Time Screen-Camera Communication Behind Any Scene," in *13th Annual International Conference on Mobile Systems, Applications, and Services*, 2015.

[31] A. Mohan, G. Woo, S. Hiura, Q. Smithwick, and R. Raskar, "Bokode: imperceptible visual tags for camera based interaction from a distance," *ACM Transactions on Graphics (TOG)*, vol. 28, no. 3, p. 98, 2009.

[32] Google, "ATAP Project Tango – Google," 2 2014. [Online]. Available: http://www.google.com/atap/projecttango/

[33] W. Yuan, R. Howard, K. Dana, R. Raskar, A. Ashok, M. Gruteser, and N. Mandayam, "Phase messaging method for time-of-flight cameras," in *IEEE International Conference on Computational Photography (ICCP)*, May 2014.

[34] H. Plank, M. Almer, R. Lobnik, C. Steger, T. Ruprechter, H. Bock, J. Haid, G. Holweg, and N. Druml, "OptiSec3D - A new Paradigm in Secure Communication and Authentication featuring Time-of-Flight," in *International Conference on Embedded Wireless Systems and Networks (EWSN)*, February 2016.

[35] J. McCune, A. Perrig, and M. Reiter, "Seeing-is-believing: using camera phones for human-verifiable authentication," in *IEEE Symposium on Security and Privacy*, May 2005, pp. 110–124.

[36] B. Zhang, K. Ren, G. Xing, X. Fu, and C. Wang, "SBVLC: Secure barcode-based visible light communication for smartphones," in *INFOCOM, 2014 Proceedings IEEE*, April 2014, pp. 2661–2669.

[37] M. Xie, L. Hao, K. Yoshigoe, and J. Bian, *Security and Privacy in Communication Networks: 9th International ICST Conference, SecureComm 2013, Sydney, NSW, Australia, September 25-28, 2013, Revised Selected Papers*. Springer International Publishing, 2013, ch. CamTalk: A Bidirectional Light Communications Framework for Secure Communications on Smartphones, pp. 35–52.

[38] S. Izadi, D. Kim, O. Hilliges, D. Molyneaux, R. Newcombe, P. Kohli, J. Shotton, S. Hodges, D. Freeman, A. Davison, and A. Fitzgibbon, "KinectFusion: Real-time 3D Reconstruction and Interaction Using a Moving Depth Camera," in *ACM Symposium on User Interface Software and Technology*, 2011, pp. 559–568.

[39] N. Erdogmus and S. Marcel, "Spoofing Face Recognition With 3D Masks," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 7, pp. 1084–1097, July 2014.

High Performance Time-of-Flight and Color Sensor Fusion with Image-Guided Depth Super Resolution

Hannes Plank*, Gerald Holweg*, Thomas Herndl*, Norbert Druml*
*Infineon Technologies Austria AG, Graz, Austria
{hannes.plank, gerald.holweg, thomas.herndl, norbert.druml}@infineon.com

***Abstract*—In recent years, depth sensing systems have gained popularity and have begun to appear on the consumer market. Of these systems, PMD-based Time-of-Flight cameras are the smallest available and will soon be integrated into mobile devices such as smart phones and tablets. Like all other available depth sensing systems, PMD-based Time-of-Flight cameras do not produce perfect depth data. Because of the sensor's characteristics, the data is noisy and the resolution is limited. Fast movements cause motion artifacts, which are undefined depth values due to corrupted measurements. Combining the data of a Time-of-Flight and a color camera can compensate these flaws and vastly improve depth image quality.**

This work uses color edge information as a guide so the depth image is upscaled with resolution gain and lossless noise reduction. A novel depth upscaling method is introduced, combining the creation of high quality depth data with fast execution. A high end smart phone development board, a color, and a Time-of-Flight camera are used to create a sensor fusion prototype. The complete processing pipeline is efficiently implemented on the graphics processing unit in order to maximize performance. The prototype proves the feasibility of our proposed fusion method on mobile devices.

The result is a system capable of fusing color and depth data at interactive frame rates. When there is depth information available for every color pixel, new possibilities in computer vision, augmented reality and computational photography arise. The evaluation shows, our sensor fusion solution provides depth images with upscaled resolution, increased sharpness, less noise, less motion artifacts, and achieves high frame rates at the same time; thus significantly outperforms state-of-the-art solutions.

***Index Terms*—Time-of-Flight, 3D sensing, sensor fusion, GPGPU, image processing**

I. Introduction

Indirect depth sensing based on Time-of-Flight works by emitting a modulated light signal. As shown on Fig. 1, the light signal is reflected by the scene and captured by the Time-of-Flight sensor. The Time-of-Flight sensor measures the phase difference of the incoming to the outgoing light for each pixel. A Photonic Mixer Device (PMD) on each pixel enables the conversion of the phase to a voltage, which is proportional to the depth. These devices occupy a lot of space on the sensor and limit the image resolution. In addition, various systematic errors are introduced during the measurement process, causing not all pixels to contain valid depth values. The most common is the low-signal error, when not sufficient light is reflected. This is also the main noise source. Another inherent weakness of the PMD-based Time-of-Flight technology is motion artifacts. These artifacts are caused by a fast moving scenery and due to the fact that several raw images need to be taken

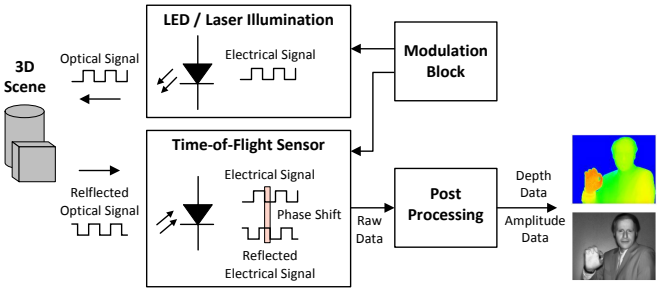


Fig. 1. The working principle of an indirect PMD-based Time-of-Flight depth sensing system. Obtained with changes from [2].

consecutively (each with a certain exposure time) in order to compute one depth image. There are numerous countermeasures for these errors, but using additional information is the best way to cope with these weaknesses. While a combination of several different depth sensing systems makes sense and has been thoroughly researched, practical applications are limited due to the increased complexity. Fusing depth and color can enhance depth resolution and reduce noise. High-resolution color cameras are ubiquitous on mobile devices like smart phones and tablets. Research has shown that depth image quality can be improved immensely by using color information, as shown by the authors in [1].

Color images offer additional information about the depth data, because there is a correlation between depth and color discontinuities. When upscaling a depth image, depth discontinuities can adapt to color discontinuities, improving the resolution and reducing noise. When every color image pixel is associated with a depth value, new applications for depth sensors arise. Fields like augmented reality, computational photography, and computer vision in general can benefit of a unified color and depth sensor.

However, there is major gap in literature concerning efficient sensor fusion solutions targeting mobile devices. This paper tackles this gap and proposes a novel solution which not only improves the resolution of Time-of-Flight depth images, but also performs loss-less noise reduction, reduces motion artifacts, and maintains interactive frame-rates by exploiting massive parallel processing. Summarizing, this paper makes the following contributions:

- It introduces a novel depth upscaling approach, designed to unify depth image refinement with high computation

performance.

- It demonstrates a feasible prototype implementation on a mobile platform using a combined Time-of-Flight and color camera system.
- It presents an evaluation which not only provides a comparison to other upscaling methods but also shows the feasibility of depth and color sensor fusion on mobile platforms.

II. Related Work

Fusing depth information with other camera data is well-explored. A lot of contributions focus on fusing different depth sensing systems, cf. [3]. By fusing Time-of-Flight cameras with stereo systems it is possible to accelerate the stereo matching computation by limiting the search space with Time-of-Flight depth measurements. In contrast to passive stereo systems, Time-of-Flight depth sensing can measure depth on homogenous surfaces and during bad lighting conditions.

Fusing Time-of-Flight depth data with any camera system requires calibration. In [4], Herrera et al. propose a method for calibrating a depth and color camera system. It is however possible to reduce the calibration procedure to a 2D stereo calibration, cf. [5], because recent Time-of-Flight cameras are also able to capture a 2D gray-scale image. This enables the usage of the well-established Bouguet Matlab calibration toolbox.

Since depth cameras usually offer lower image resolutions than RGB cameras, improving the resolution of the depth sensor is a well researched topic. It is possible to increase the depth resolution by regarding color discontinuities. In [1], Langmann et al. review a number of approaches. The joint bilateral filter by Kopf et al. [6] is a popular upsampling method. It is a modification of the original bilateral filter by Tomasi [7] and works by weighting the depth image with the color distance of the color image. Sung et al. [8] extend the joint bilateral filter with color segmentation based boundary refinement. Yang et al. [9] construct a cost volume and apply the bilateral filter. The best cost is then selected and the process is applied iteratively.

A different way to fuse depth and color data is global optimization. These global approaches formulate a cost function, which is numerically minimized, cf. [5] and [10]. This leads to the best results among depth super resolution methods but the optimization is computationally complex. In [11], Dai et al. promise a high performing global approach. The algorithm interprets the pixel-grid as graph with intensity differences as edge weights. The minimum spanning tree (MST) is calculated and the sparsely mapped depth values are interpolated along the tree. The drawback of this method is that the MST computation is hard to implement efficiently on a GPU, as described by the authors in [12].

The demand for an efficient color and depth fusion method, capable of running on limited hardware formed just recently, when depth sensors became available on mobile platforms. In [13], Chuchvara et al. propose a GPU based sensor fusion method, which creates a surface mesh from the mapped depth data. This mesh is rendered to a frame-buffer with OpenGL, and a non-local means based noise reduction method is executed. This method uses the GPU hardware with OpenGL efficiently for interpolation, but does not increase the depth resolution.

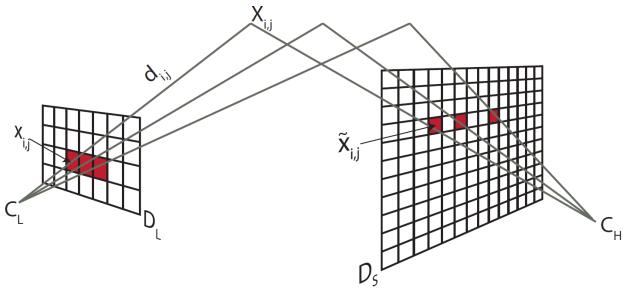


Fig. 2. Depth data mapping from depth D_L to color camera D_S image space. Obtained with changes from [5].

Despite the vast amount of research, there is major gap in literature concerning efficient sensor fusion solutions for mobile devices, which is addressed in our work.

III. Depth Image Mapping

The desired sensor fusion result is a color image with depth associated to each pixel. As a first step, it is necessary to warp the depth data to the color image. Since warping introduces undefined areas, the cameras are mounted as close as possible. Several images of a checkerboard target in various orientations and distances are captured for camera calibration. The checkerboard pattern enables the calibration software to derive edge reference points with sub-pixel accuracy. The Time-of-Flight camera is able to capture infrared-based gray-scale images, enabling to calibrate the cameras like a stereo system. The Bouguet Matlab Calibration toolbox was used to compute the intrinsic and extrinsic camera parameters.

After distortion correction, the depth values with pixel position $x_{i,j}$ and depth $d_{i,j}$ are transformed to 3D points $X_{i,j}$ in color camera space, as defined in (1) and as illustrated in Fig. 2. $\tilde{P}_D$ denotes the pseudoinverse of the depth camera calibration matrix. The extrinsic camera parameters in form of rotation matrix R and translation vector T directly transform the depth measurements to the 3D space of the color camera with center C_H.

$$X_{i,j} = T + Rd_{i,j}\frac{\tilde{P}_D x_{i,j}}{\|\tilde{P}_D x_{i,j}\|} \tag{1}$$

The homogenous 3D depth coordinates $X_{i,j}$ are projected to the 2D image space D_S of the color camera by a multiplication with the color camera projection matrix P_C, cf. (2). This leads to warped depth measurements with the coordinates $\tilde{x}_{i,j}$ in color image space.

$$\tilde{x}_{i,j} = P_C \cdot X_{i,j} \tag{2}$$

After applying the lens distortion of the color camera, the depth coordinates are associated with the nearest color pixel.

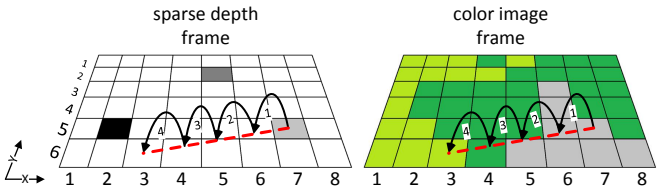


Fig. 3. Pixel parsing along circular paths for each depth value.

Reducing the noise of the depth image with an edge-preserving 3x3 median filter before warping benefits the final depth image greatly while not causing significant computation overhead. Alternatively, other well-established depth image denoising techniques can be applied. The next step after the depth image mapping is to interpolate this sparsely mapped depth frame by using color information.

IV. Image-Guided Depth Upsampling

The depth upscaling is the most computational complex step in the sensor fusion procedure. With a well-designed algorithm, the sparse depth data on the color image is not just interpolated, but also denoised and the resolution refined. This work targets to unify the creation of a high quality 1-1 mapped color and depth image, while maintaining high computation performance. Data parallelism with as less mutual data dependence as possible is desired for an efficient GPU implementation. It is also important to consider memory usage and addressing, as GPU memory operations are slow.

A preparation step before the upscaling is extracting edge information from the color image. While any edge detector can be used, the Sobel operator proved to be simple and effective. If the Sobel operator is not just applied to the intensity of the image but also to the saturation, then the edges between surfaces with equal brightness but different colors are detected as well.

The edge image and the sparsely mapped depth frame form the input data for the upscaling algorithm which is executed in parallel for each mapped depth value. The algorithm can be interpreted as the propagation of depth value influence over a local circular area on the color image. A radial path from the mapped depth value position to each influenced pixel is parsed.

Fig. 3 illustrates an example of a depth influence propagation path. The influence of a depth value on the output depends on the sum of edge values S_E between the mapped depth value (at 7/5) and each pixel. The color edges crossed at step 3 and 4 decrease the depth influence. Each pixel of the resulting high-resolution depth image is influenced by several depth values which are then interpolated by their propagated influence. The interpolation weight w quantifies the influence of a depth value. The requirement of a function to convert S_E to w, is that small changes of S_E lead to large changes of w. Depth values need to be restricted from propagating over edges if there are valid depth measurements on the other side of the edge. The natural exponential function e^x suits this purpose well, cf. (3).

$$w = e^{\frac{-S_E}{\sigma}} \tag{3}$$

The parameter σ scales the argument of the exponential function, in order to restrict the range of w to the boundaries of the 32bit floating point data type. The k depth values d_n which influence pixel i are interpolated with the interpolation formula by using the ratios of the weights w_n, resulting in the final depth d_i, cf. (4).

$$d_i = \frac{\sum_{n=1}^{k} d_n \cdot w_n}{\sum_{n=1}^{k} w_n} \tag{4}$$

Depth values propagate their influence to the influence sphere of other depth values to reduce noise. Since color edges greatly reduce the influence, the noise reduction is edge preserving.

A. Efficient Processing

The introduced algorithm can be efficiently implemented with some modifications. The interpolation formula in the last section can be reformulated as update formula. d_{sum} and w_{sum} are stored for each pixel and replace the upper and lower fraction terms in the interpolation formula. They are updated every time a depth value operates on a pixel, cf. (5) and (6).

$$d_{sum} = d_{sum} + d_n \cdot w_{new} \tag{5}$$

$$w_{sum} = w_{sum} + w_{new} \tag{6}$$

The final depth d_i is calculated by dividing d_{sum} and w_{sum} on each pixel i, as given by (7).

$$d_i = \frac{d_{sum}}{w_{sum}} \tag{7}$$

This operation requires at least 3 GPU memory access operations per pixel, which can be reduced. If d_i is calculated and updated during the depth influence propagation, d_{sum} can be extracted by multiplying d_i with w_{sum}. Without storing d_{sum} for each pixel, the depth d_i and w_{sum} are updated at each pixel operation according to (8)-(10).

$$w = e^{\frac{-S_E}{\sigma}} \tag{8}$$

$$w_{sum} = w_{sum} + w_n \tag{9}$$

$$d_i = \frac{d_i \cdot (w_{sum} - w_n) + d_n \cdot w_n}{w_{sum}} \tag{10}$$

To satisfy the requirement that S_E is the sum of all edges along the pixel path between pixel i and the mapped depth value d_n, the surrounding pixels need to be parsed in a specific order. Fig. 4 illustrates the dependence hierarchy, caused by the radial parsing principle. Each newly processed pixel depends on its predecessor along the pixel path. Calculating the pixel paths during computation is avoided by a lookup table. The table stores the relative pixel positions on the paths in the correct order.

A small array is used to buffer the sum of edge weights on the path S_E for each pixel, so it can be looked up during the

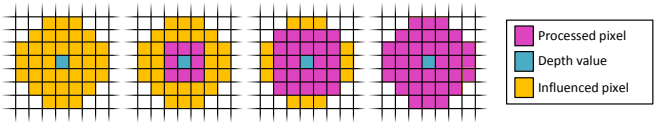


Fig. 4. The pixel parsing hierarchy of each mapped depth value.

next pixel operation on the path. This averts the prerequisite to parse the pixels along the radial pixel paths, as long the hierarchy shown in Fig. 4 is satisfied. As a consequence, the lookup table can be cleaned of all redundant operations, so just n pixel operations for n influenced pixels are required.

A further optimization is to selectively abort calculations on a pixel path. When a pixel path is classified as irrelevant, the latest buffered value of S_E is marked invalid, and all further pixel processing operations are canceled. The abortion criteria is satisfied, when a depth value does not have a significant influence on the final depth. This can happen when a pixel path crosses too many edges and its influence is minuscule compared to the other depth values in the area. Irrelevant paths are detected, by comparing the already existing weights w_n of a pixel to the current weight. If the difference is below a certain value γ, the remaining operations on the current pixel path are canceled. The evaluation showed that stopping the processing on irrelevant influence propagation paths also improves the quality of the upscaled depth image as the irrelevant pixel paths introduce an error to the upscaled depth image.

B. GPU Implementation

In our prototype, the complete sensor fusion procedure is implemented on the graphics processing unit, using OpenCL. The system features a color and Time-of-Flight camera system as well as a high-end smart phone development kit for computation. The camera system is shown in Fig. 6 and consists of a Logitec B910 color camera and an Infineon REAL3™ sensor based on Time-of-Flight technology of pmdtechnologies. With this prototype, the ability of fusing depth and color while maintaining interactive frame-rates at same time is demonstrated. An Android application implements the framework for handling the cameras and Android NDK is employed to handle the GPU computation.

The processing pipeline consists of 3 GPU programs, for mapping the depth data, calculating the edge image, and upscaling the depth data with the proposed method. The upscaling computation is by far the most computational intensive operation, requiring 95% of the execution time.

Because the parameter space of the local OpenCL workgroup size is restricted to a few dozen configurations, the implementation can optimize itself on a large variety of GPU devices. Local memory buffering was avoided in favor for optimized cache usage. This avoids any synchronization mechanisms and allows to use all computation units simultaneously without data corruption.

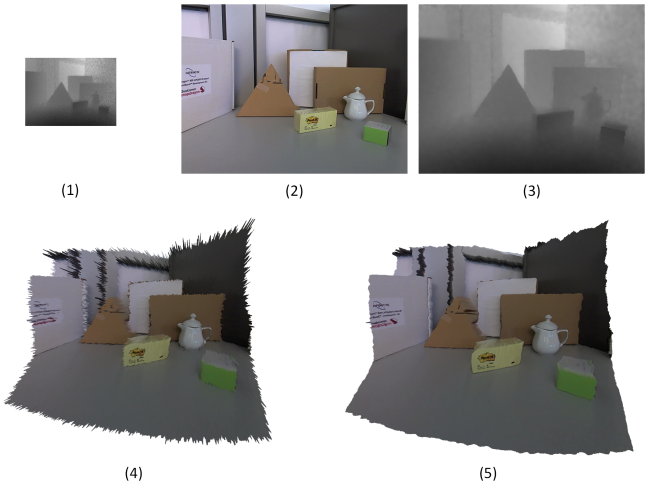


Fig. 5. 1: Original depth image, captured with the Time-of-Flight camera; 2: Color image; 3: The resulting mapped and upscaled depth image; 4 and 5: 3D visualization of raw depth data compared to our upsampling method

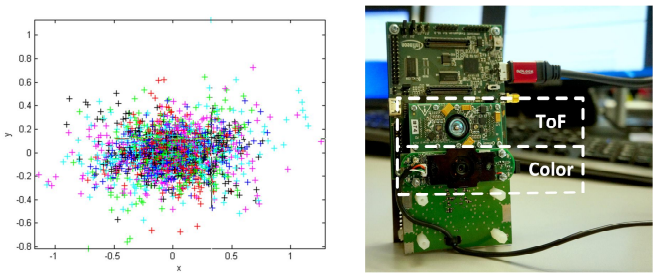


Fig. 6. 1: The reprojection error of the calibration target; 2: The color and depth camera system.

V. Results and Evaluation

A. The Prototype System

One of the goals of this work was to proof that state-of-the art mobile devices are capable of running the proposed sensor fusion method at interactive frame rates. The prototype is capable of fusing 288x256 depth images with 640x480 color images at a rate of 10 frames per second (fps). Fig. 5 shows a scene captured and processed by the prototype. The 3D visualization clearly shows the significant quality impact of our proposed depth interpolation method. The evaluation in Section V-D reveals, the upscaling algorithm can also handle large resolutions, up to several megapixel.

As mentioned in Section III, the camera system is calibrated like a stereo camera system, using the 2D amplitude image of the Time-of-Flight camera. Despite the low depth resolution, it is possible to calibrate the system properly. The reprojection error of the interest points on the 46 calibration images is visualized in Fig. 6. It can be observed that the vast majority of reprojected calibration interest points deviate by less than one pixel. A reprojection error below 0.5 pixel is insignificant,

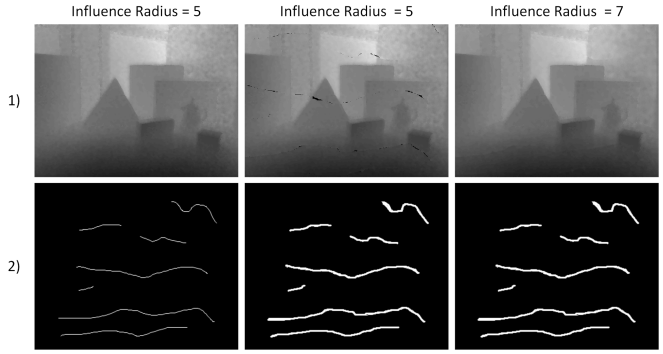


Fig. 7. Interpolation of invalid depth values. 1: The upscaled depth image; 2: The masks which were used to simulate invalid measurements due to, e.g., motion artifacts.

as the mapped depth values are associated with the closest color pixel.

B. Motion Artifacts Compensation

Another important aspect of depth image refinement is the removal of motion artifacts. Advanced Time-of-Flight cameras emit light pulses with different phases and modulation frequencies per depth measurement. Fast movements during this procedure can cause invalid depth measurements. While cameras are able to detect them, it is up to the postprocessing pipeline to interpolate the missing information. The proposed depth upscaling algorithm interpolates these invalid measurements without modifications. As Fig. 7 shows, even larger motion artifacts are compensated if the influence radius of the depth values is increased. Since the cameras are not synchronized, the motion artifacts in Fig. 7 are simulated by creating a mask.

C. Performance

Our prototype system is capable of fusing the sensor data at a rate of 10 fps, when executed on the graphics processor on the mobile platform. In comparison, the sensor fusion procedure runs at a rate 126 fps on an Nvidia Geforce GTX 550 TI desktop GPU.

The depth upscaling stage is the most computational intensive stage, taking 96% of the computation time. The first stage, the depth value mapping takes less than 1%, and the edge detection takes approximately 3% of the computation time. Since the parameter space for the OpenCL local work-group size parameter is limited due to the image dimensions, a brute-force approach quickly determines the optimal configuration. This ensures optimal performance among a wide range of devices.

The complexity class of the upscaling algorithm is $O(k \cdot n_{color})$ where n_{color} is the number of color pixels and k the average number of depth values influencing a color pixel. k depends directly on the influence radius and number of mapped depth pixels. The higher the value of k, the better the noise reduction.

D. Comparison with other depth upscaling methods

With available ground-truth depth data, it is possible to quantify the depth image quality refinement. By calculating the difference between the upscaled depth image and a ground-truth reference image, depth upsampling algorithms can be compared. The popular Middlebury 2006 dataset [14] consists of color and high-quality depth images. By downscaling and adding noise, realistic depth input data is created. Beside the root mean square error (RMSE), the percentage of bad matching pixels is commonly used to evaluate depth upscaling and stereo matching algorithms [15]. The percentage of bad matching pixel (PbmP), is a measure for the number of pixels differing more than a defined tolerance δ from the ground-truth, as given by (11).

$$PbmP = \frac{1}{N} \sum_{(x,y)} (|d(x,y) - g(x,y)| > \delta) \qquad (11)$$

Our evaluation uses the common value 1 for δ, assuming the depth values are mapped to a 0-255 value range.

The proposed solution is compared to simple nearest neighbor and bilinear interpolation, joint bilateral filtering [6] (Kopf et al.), minimax path based interpolation [11] (Dai et al.), and anisotropic total variation upscaling [5] (Ferstl et al.). The joint bilateral filter was selected for comparison because it is a popular and fast upscaling method. It however needs a lot more operations per pixel compared to our proposal and works best if applied iteratively. As our evaluation shows, the joint bilateral filter is significantly outperformed, especially when the upscaling factor is large. The minimax path based upscaling algorithm showed promising results. One of the conditions for its proclaimed fast performance is that every color pixel is just influenced by two depth values. However, due to this fact, it is not capable to reduce depth noise. The anisotropic total variation upscaling method is based on energy minimization. It leads to better results, but is not suitable for real-time applications due to its global approach. We compare with this method to demonstrate what is possible with global optimization, and how the other faster methods perform.

Three scenes of the Middlebury 2006 dataset were selected, and the ground-truth depth images were downscaled by the factor of 8 and 16. Gaussian noise was added and the depth images were upscaled with the selected algorithms. The Aloe dataset was specifically selected, because it features a salient texture on flat surfaces, which misguides a lot of depth upscaling algorithms. Table I shows the root mean square error and the percentage of bad matching pixel for three scenes of the Middlebury 2006 dataset.

Besides the quantitative comparison, it is also interesting compare the visual results. Fig. 8 shows the resulting depth image of the previously compared algorithms.

VI. Conclusion

Sensor fusion and Time-of-Flight 3D imaging in general is a fast-paced field of research. Fusion of depth and color data on mobile platforms will enable whole new fields of applications.

	Aloe				Art				Moebius			
	8x		16x		8x		16x		8x		16x	
	RMSE	PbmP	RMSE	PbmP	RMSE	PbmP	RMSE	PbmP	RMSE	PbmP	RMSE	PbmP
Nearest Neighbor	4.89	77.76	5.96	77.14	5.73	78.49	7.50	80.47	4.91	77.83	5.62	78.42
Bilinear	4.21	72.17	5.32	73.27	4.98	74.50	6.79	77.24	4.18	73.05	4.92	74.26
Kopf et al. [6]	3.45	57.13	5.47	59.35	4.20	59.41	3.00	54.75	2.89	54.37	4.05	56.65
Dai et al. [11]	3.81	59.72	5.03	69.30	4.06	61.87	4.11	58.58	4.25	64.79	4.03	68.44
Ferstl et al. [5]	2.28	31.69	3.50	49.94	2.83	38.98	1.92	27.89	2.66	52.45	2.33	41.53
Ours	3.36	54.31	4.99	65.44	3.59	53.28	2.59	47.82	2.59	47.82	3.94	63.70

TABLE I
COMPARISON OF UPSCALING ALGORITHMS ON THE MIDDLEBURY 2006 DATASET, USING THE ROOT MEAN SQUARE ERROR (RMSE) AND THE PERCENTAGE OF BAD MATCHING PIXEL (PBMP).

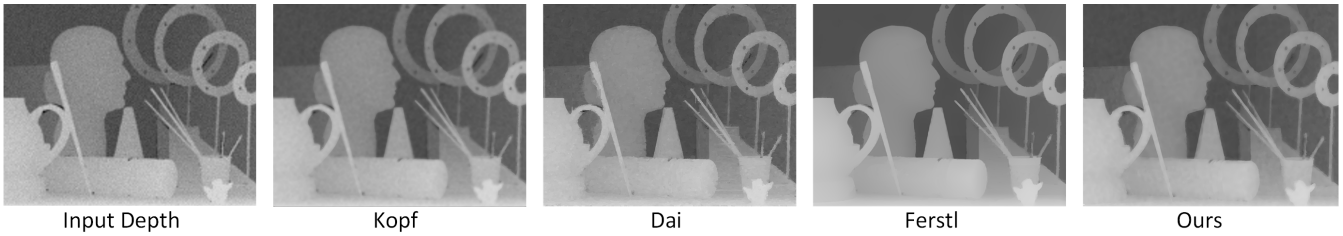


Fig. 8. Visual comparison of depth upscaling methods on the Middlebury 2006 dataset Moebius.

Currently, there are several sensor fusion approaches which can increase depth resolution and reduce noise, but the design goal of reaching high performance on mobile devices was neglected so far.

This work demonstrates the feasibility of high-performance and high-quality depth and color sensor fusion targeting mobile devices by employing a state-of-the-art smart phone development platform. It uses an Infineon REAL3™ image sensor based on Time-of-Flight technology of pmdtechnologies, which is on the brink of being integrated into new generations of smart phones and tablets (cf. Google Tango project). The output of this system is a color and depth image, where each color pixel is directly associated with a depth pixel.

The presented innovative method works non-iteratively and has an $O(n \cdot k)$ complexity. Our sensor fusion solution provides depth images with upscaled resolution, increased sharpness, less noise, less motion artifacts, and achieves high frame rates at the same time. Given these facts, this work significantly outperforms state-of-the-art and thus sets new benchmarks in the field of mobile computing and will enable a set of new types of mobile applications.

Our future work concerns the development of an interface abstracting the color and Time-of-Flight camera into a unified RGB-D sensor. This involves further optimizations, resolution scalability, camera synchronization, and auto-calibration.

ACKNOWLEDGMENTS

The authors would like to thank the Austrian Federal Ministry for Transport, Innovation and Technology as well as the ARTEMIS Joint Undertaking, which funded the EMC^2 project under the grant agreement n° 621429.

REFERENCES

[1] B. Langmann, K. Hartmann, and O. Loffeld, "Comparison of Depth Super-Resolution Methods for 2D/3D Images," *International Journal of Computer Information Systems and Industrial Management Applications*, vol. 3, pp. 635 –645, 2011.

[2] N. Druml, G. Fleischmann, C. Heidenreich, A. Leitner, H. Martin, T. Herndl, and G. Holweg, "Time-of-Flight 3D Imaging for Mixed-Critical Systems," in *13th International Conference on Industrial Informatics (INDIN)*, July 2015, pp. 1432–1437.

[3] R. Nair, K. Ruhl, F. Lenzen, S. Meister, H. Schfer, C. Garbe, M. Eisemann, M. Magnor, and D. Kondermann, "A Survey on Time-of-Flight Stereo Fusion," *Lecture Notes in Computer Science*, vol. 8200, pp. 105–127, 2013.

[4] D. Herrera C., J. Kannala, and J. Heikkil, "Accurate and Practical Calibration of a Depth and Color Camera Pair," in *Computer Analysis of Images and Patterns*, ser. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2011, vol. 6855, pp. 437–445.

[5] D. Ferstl, C. Reinbacher, R. Ranftl, M. Ruether, and H. Bischof, "Image Guided Depth Upsampling Using Anisotropic Total Generalized Variation," in *IEEE International Conference on Computer Vision (ICCV)*, December 2013, pp. 993–1000.

[6] J. Kopf, M. F. Cohen, D. Lischinski, and M. Uyttendaele, "Joint Bilateral Upsampling," *ACM Transactions on Graphics (TOG)*, vol. 26, no. 3, July 2007.

[7] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Sixth International Conference on Computer Vision*, January 1998, pp. 839–846.

[8] S.-Y. Kim, J.-H. Cho, A. Koschan, and M. Abidi, "Spatial and Temporal Enhancement of Depth Images Captured by a Time-of-Flight Depth Sensor," in *20th International Conference on Pattern Recognition (ICPR)*, August 2010, pp. 2358–2361.

[9] Q. Yang, R. Yang, J. Davis, and D. Nister, "Spatial-Depth Super Resolution for Range Images," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2007, pp. 1–8.

[10] J. Park, H. Kim, Y.-W. Tai, M. Brown, and I. Kweon, "High quality depth map upsampling for 3D-TOF cameras," in *IEEE International Conference on Computer Vision (ICCV)*, November 2011, pp. 1623–1630.

[11] L. Dai, F. Zhang, X. Mei, and X. Zhang, "Fast Minimax Path-Based Joint Depth Interpolation," *IEEE Signal Processing Letters*, vol. 22, no. 5, pp. 623–627, May 2015.

[12] V. Vineet, P. Harish, and P. J. Narayanan, "Large Graph Algorithms for Massively Multithreaded Architectures," 2009.

[13] A. Chuchvara, M. Georgiev, and A. Gotchev, "A speed-optimized RGB-Z capture system with improved denoising capabilities," *SPIE Proceedings*, vol. 9019, March 2014.

[14] H. Hirschmuller and D. Scharstein, "Evaluation of Cost Functions for Stereo Matching," in *Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on*, June 2007, pp. 1–8.

[15] D. Scharstein and R. Szeliski, "A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms," *International Journal on Computer Vision*, vol. 47, no. 1-3, pp. 7–42, April 2002.

Time-of-Flight based Optical Communication for Safety-Critical Applications in Autonomous Driving

Hannes Plank[1], Gerald Holweg[1] Christian Steger[2], and Norbert Druml[1]

[1] Infineon Technologies Austria AG, Graz, Austria,
Hannes.Plank@infineon.com,
WWW home page: www.infineon.com/
[2] Graz University of Technology, Institute for Technical Informatics, Austria

Abstract. The automotive field has seen tremendous research effort in the past years to increase safety by using vehicle to vehicle (V2V) communication. There are however several issues in RF based V2V communication impairing the safety applications requiring hard-reality communication. Channel congestion, latency and the vulnerability to denial of service attacks demand for an alternative solution to transmit time-critical safety messages.
In this work, we discuss the security and safety aspects of free-space optical communication solutions for line-of-sight V2V communication. We show scenarios to demonstrate the high demand of a reliable low-latency communication link between cars and evaluate the requirements in communication security.
Image sensor based communication is widely accepted as the most promising optical communication link for V2V. This work offers a detailed discussion on how it is possible to increase speed and robustness by using Time-of-Flight 3D image sensors for communication. Finally, we show how communication partners can be localized and how location-aware communication can greatly benefit secured communication by mitigating relay and denial-of-service attacks.

1 Introduction

As long lasting field tests like Google's self-driving car project [9] prove, autonomous cars have a significant lower accident rate, compared to human drivers. In fact, cars operated by humans and pedestrians are the main threat for autonomous cars, due to their sometimes irrational or unpredictable behavior. In certain scenarios, autonomous cars are unable to prevent accidents, when the front view is occluded by other cars. An example is shown in Figure 1, where a pedestrian steps on the road between two subsequent cars. The accident might be preventable by a reliable communication link between the cars. The car in the front can sense the pedestrian and can detect the fatal movement in the moment

2

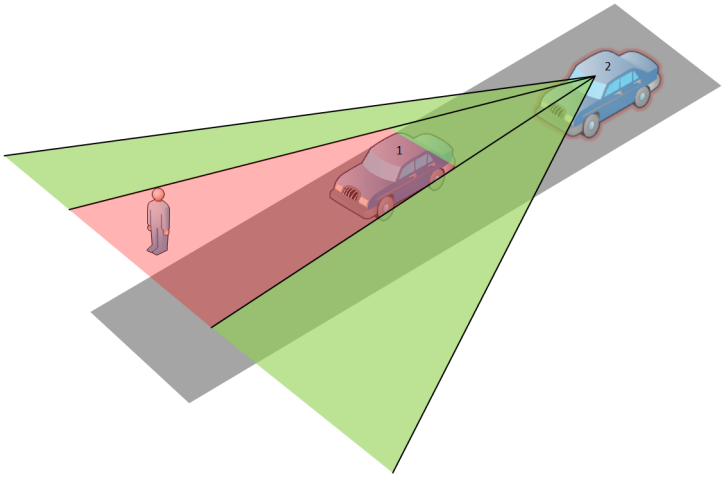


Fig. 1. Example for the need of low delay V2V communication: If vehicle 1 detects a movement of the pedestrian towards the street, it can warn vehicle 2.

of passing by. The front car can send a warning message to the back car, which can take immediate action.

Such scenarios require reliable real-time and low latency V2V communication. In this work, we show how recent RF based V2V communication is unable meet these requirements and how optical communication can tackle this problem.

In V2V communication, vehicles in a certain area form an RF ad hoc network. Since they share the same channel, a delay is introduced due to collision avoidance protocols. While RF based V2V communication is successfully used to transmit accident warnings over a longer range, problems appear in close domain, where time requirements are more critical.

Optical line-of-sight communication between subsequent cars has the huge advantage that the communication is limited between two partners and no collisions can appear. Messages can be transmitted directly with very low delay and without the risk of channel congestion. Such free-space optical communication has been traditionally conducted by narrow field of view photo detectors. In a dynamic environment like between moving cars, image sensor based detectors are better suited. Image sensor based communication systems offer a wide field of view and the ability to locate the communication partner, separating the source of information from other interfering light sources creating a robuster communication link. A rather unexplored aspect of image sensor based communication is location-awareness, where the information source during communication can be assigned to a position.

Time-of-Flight sensors are image sensors, originally designed for capturing depth images. However as we show in this work, they are a promising technology enabling robust image sensor based communication with a novel, more efficient

modulation method. While traditional image sensors measure an induces charge, proportional on the amount of incoming photons, Time-of-Flight cameras apply a modulation to each pixel and are able to measure phase shift differences of pulsed incoming light. The signal of non-pulsed light is drained from the pixel, eliminating all background interference.

Security is a crucial aspect in V2V communication and certain attack scenarios exist, which can impact the safety of vehicular networks. The nature of image sensor based optical communication offers the significant advantage of localizing the communication partner. Even if the distance cannot be determined by the image sensor itself, localization data is available from other sensors, such as RADAR, LIDAR, or other depth sensing systems. We propose a V2V communication concept, enriching state-of-the-art cryptographic security protocols with localization information, which mitigates relay attacks.

This work is structured the following way: Section 2 discusses V2V communication with focus on its limitations in safety and security in low-delay applications. Recent image sensor based free-space optical communication approaches in the automotive domain are then introduced in Section 3. In Section 4, we provide insight about the workings of Time-of-Flight sensors, and how it can be used for robust image sensor based communication links. Finally we introduce our own concept for a secure and location-aware optical communication link.

2 Problems of RF based V2V communication

Vehicular ad hoc networks (VANET) are going to make huge contributions to the safety and convenience of assisted and autonomous driving. Using the RF channel for communication however is not the ideal solution for applications which require very low latency and high reliability. In the RF domain, the communication medium is shared among multiple entities and thus requires collision avoidance protocols. According to Cailean et al. [5], channel congestion is the main reason for unreliability of otherwise properly configured VANETs. Even if a channel is reserved for high priority emergency warnings, a certain delay is always introduced by collision avoidance mechanisms. As shown by Bilstrup et al. [3], current V2V can experience large channel access delays and can not be considered a real-time communication protocol. Mass events like large car crashes can event trigger a broadcast storm [21], where many nodes in a VANET broadcast messages at high frequencies. Agarwal et al. therefore emphasize the need of directional communication for safe and reliable message exchange in dense environments [1].

Crash prediction and avoidance mechanisms are the most common applications, requiring low latency communication. Even if a crash is unavoidable, crash prediction can improve passenger safety by airbag pre-triggering [10]. Crash scenarios as presented in Section 1 exemplify the requirement of low latency in V2V communication.

Optical communication between subsequent cars is a promising technology which is congestion free and can offer a very low communication delay. Cailean

et al. [5] propose a combination of visible light communication and RF based solutions, being the key to meet modern demands of safety critical V2V communications systems.

2.1 Security

The dynamic environment of VANETs introduces certain security threats, which are not easily countered by cryptographic methods. VANETs based on RF are vulnerable to denial of service (DoS) attacks, which are simple to execute by RF jamming. Different countermeasures are discussed by Hasbullah et al. [11], however they are based on switching to different channels, RF transmission technologies or the utilization of multiple transceivers. While these countermeasures can mitigate the severity of a DoS attack, a sophisticated jamming attack can also block evasion channels.

False accident warnings [22] can impair the traffic flow, as they cause a slow down of the surrounding vehicles. In a network where all cars are authenticated and communication packets are signed, the warnings could be matched with the driver. This however compromises privacy, since vehicles could be tracked by an attacker or authority.

3 Image Sensor Based V2V Communication

The problems in the RF domain of V2V communication can partially be solved by additionally connecting vehicles with an optical line-of-sight based communication system. A line-of-sight connection will limit the communication to two cars and thus prevent channel congestion issues and therefore also reduce latency.

The dynamic environment of V2V communication is a special application of visible light communication, which requires a wide field of view. If a traditional optical communication system with a single photo detector is used, the wide field of view would bundle all lightsources onto the sensor, which would result in a comparatively low SNR. It is therefore necessary to separate lightsources, which is conveniently archived with cameras. The incoming light is projected by the lens onto the sensor, separating the lightsources. When the pixels on the image sensor are used to sample the communication partner's light source, it is possible to only select suitable pixels.

In the past years, various image sensor based communication approaches for V2V communication appeared [5]. The first advances were driven by the fact that cars usually already frequently encounter lightsources [24]. Road, tunnel, traffic or brake lights are increasingly driven by LEDs, which can be modulated. Modulation schemes have been proposed, enabling to modulate visible lightsources without inducing perceivable flickering [4].

Originally most image sensors are designed to produce video streams with typically 30 frames per second and thus lack of sampling speed for communication. Since most sensors require a full readout of all pixels, the frame rate is

severely limited, and dedicated high speed hardware is required to reach feasible speeds for V2V communication [15].

An approach to boost the bandwidth of mostly conventional image sensors are multiple input, multiple output (MIMO) approaches where information is emitted by multiple LEDs in parallel [18]. MIMO approaches rely on the separability of lightsources which however limits the distance. Increasing the resolution of image sensors could counteract the distance problem, but either requires a larger camera or smaller pixels reducing the light sensitive surface area.

A general problem of free-space visible light communication are atmospheric turbulences, which are discussed further in Section 3.1. These turbulences are a severe problem for MIMO communication approaches. When light is scattered by small participles like raindrops or fog, the different light sources in MIMO approaches are mixed and can no longer be separated by the receiver.

The demand for fast image sensor based communication lead to the development of dedicated optical communication image sensors (OCI). The sensor of Takai et al. [19] features dedicated communication pixels which are interleaved with imaging pixels. The imaging pixels directly provide a flag image, using comparator circuits. This flag image is used to detect the position of high intensity objects such as LEDs. The communication pixels are designed to be very sensitive to illumination changes, and can be read out separately using an address generator.

3.1 The Influence of Atmospheric Turbulences

A challenge for free-space optical communication are atmospheric turbulences which involve all kinds of obstructions such as fog, rain, snow or dust. Kim et al. [12] simulate rainy conditions on an optical communication channel. They use on/off keying for transmission and measure a SNR loss of 60%. They show that using modified fixed decision thresholding (MFDT) reduces the bit error rate dramatically.

Fog is the most challenging obstruction, since it absorbs, scatters and reflects light. This means that the sender might also sense a reflection of his own signal. Kim et al. also investigate the influence of fog on visible light V2V communication [13]. Experiments with a fog chamber show massively decreased SNR, which could be compensated to a certain degree by focusing more light on the sensor with an additional lens. Interestingly, their experiments also show that red light exhibits the lowest attenuation coefficient. This confirms the use of red taillights or (invisible) near infrared light (NIR) for optical communication under foggy conditions.

4 Time-of-Flight Sensors

Time-of-Flight cameras are designed to measure the distance on each pixel between the camera and the scene. The basic principle, as illustrated in Fig. 2, is

6

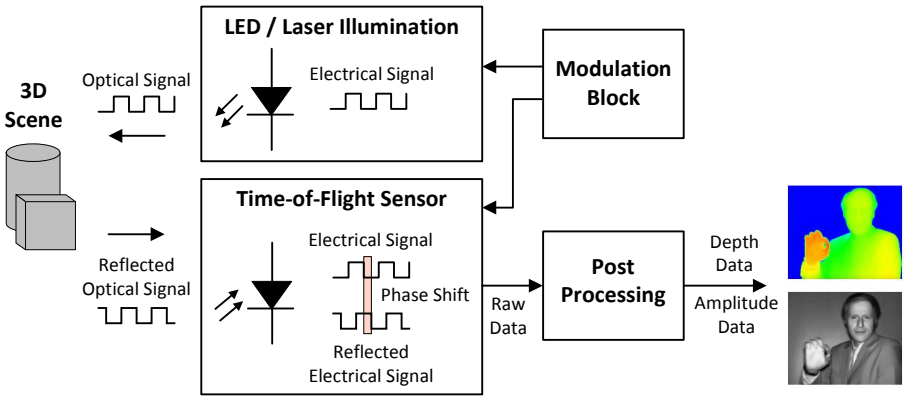


Fig. 2. The principle of Time-of-Flight depth sensing, obtained with changes from [6].

.

to measure the time it takes emitted light to travel from the camera to the scene and back to the sensor.

During a measurement, an active illumination unit next to the sensor emits pulsed infrared light. The ToF image sensor registers the phase difference between emitted and reflected light. Each pixel is equipped with a photonic mixture device (PMD) [14], capable of converting a phase shift to a measurable voltage. The basic principle of a PMD is illustrated in Fig. 3. When photons arrive at the sensor, they induce charges in the sensor. Each PMD is equipped with two buckets to store the charges. A reference signal decides into which bucket the charges are transferred. The reference signal is a pulsed signal with the same frequency as the signal emitted by the illumination unit. After a certain integration time, the contents of the buckets are read-out and digitized. The charge difference between the buckets is proportional to the phase-shift of these signals and thus proportional with distance between camera and scene. Because ToF sensors are designed to measure the phase shift of pulsed light signals, ToF sensors can be used as novel visible light receivers for optical communication.

Captured raw images are usually severely biased by different amount of light being reflected by different surface materials. It is common to capture multiple raw images with different phases shifts of the illumination signal and use post-processing methods to reconstruct depth images. Most ToF cameras feature therefore an integrated phase shifting unit (PSU) for emitting phase shifted signals. This enables phase shift modulation of the emitted signal without additional hardware.

Conventional free-space optical communication suffers from background light. Direct lightsources like the sun and a bright environment will reduce the SNR of direct detection based systems. A significant advantage of Time-of-Flight based optical communication is that the pixels are exclusively sensitive to pulsed light. The reason is that PMD pixels of a ToF sensor measure charge differences be-

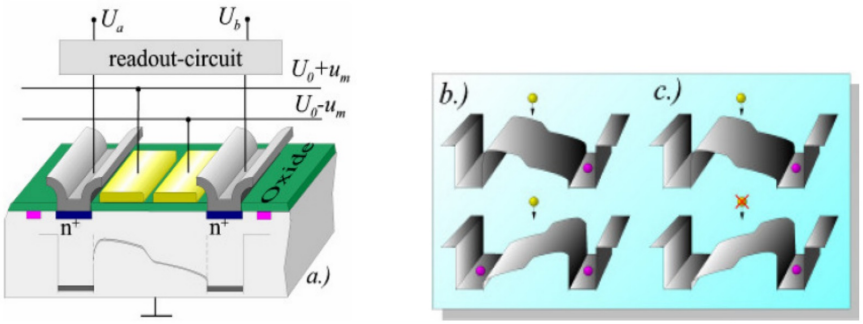


Fig. 3. The working principle of a ToF pixel. A: A pixel is supplied with a modulation signal u_m. B,C: Charges are sorted into bucket A or B, depending on the modulation signal. Obtained with changes from [20].

tween two buckets. If non-pulsed light arrives at the pixels, the photons are evenly distributed into both buckets with no change of the charge difference. The pulse frequency of ToF cameras depends on the intended application and is typically in the range between 1 and 100 MHz.

The sender and the receiver must use the same frequency, otherwise the signal is suppressed. This enables to establish multiple non interfering VLC connections in a close environment by using different frequencies.

Each pixel has a limited capacity to store charges during exposure. The ToF sensor we use in our work features a suppression of background illumination (SBI) circuit on each pixel [20]. The SBI circuit actively drains the buckets of the PMD element during exposure. This prevents the pixels from saturating from background illumination, and enables to use ToF sensors for robust depth sensing and communication in bright environments.

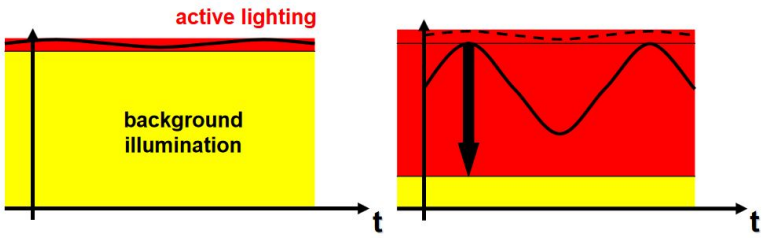


Fig. 4. Pulsed signal detection characteristics of a PMD pixel with (right) and without (left) suppression of background illumination. Obtained from [8].

8

4.1 Optical Communication with Time-of-Flight Sensors

Phase shift keying (PSK) of pulsed light has not been widely considered as viable modulation method in visible light communications [16]. The first and to our knowledge only approach to use Time-of-Flight cameras for optical communication has been accomplished by Yuan et al. [25]. They established a one-way MIMO communication link, using a ToF camera in depth sensing mode as receiver. Photodiodes are used to sense the pulsed illumination signal of a ToF camera. This signal is amplified and phase shifters are used to modulate information using PSK. The modulated signal is then emitted back to the ToF camera with an LED array. The signal is received by the ToF camera and the encoded information is extracted by analyzing the depth image.

In our approach, we use two ToF cameras for mutual communication, serving as receiver and sender. Capturing depth data is optional, and raw images are used for communication. Limiting the exposure time to a minimum enables high frames. The integrated phase shifting unit is used for PSK modulation and the illumination unit is used as emitter.

Compared to on-off keying, pulse position or pulse shift keying based modulation schemes, phase shift keying offers certain advantages. At first, phase shift keying is more robust to signal intensity variations. In a dynamic environment, modulated light does not arrive homogeneously on a set of pixels. During the exposure time, a moving lightsource illuminates an undefined number of pixels with varying intensity. This motion blur might prevent to recover the intensity and consequently causes an information loss. A phase shifted pulsed signal however will have a consistent phase shift, independent of the number of photons arriving at the sensor. The SNR suffers from less sensed photons, but can be compensated by averaging all influenced pixels.

Image sensors produce a relative large amount of data, which takes significant time to read and digitize, limiting the frame rate. Using a high speed camera for optical communication counters this problem, but requires powerful hardware. Phase shift keying however increases the symbol size per transmitted frame. On-off keying is able to transmit one bit per frame on a synchronized system, while it is possible to transmit multiple bits per frame using PSK.

While lightsources are separated by projection onto image sensors, there might be the possibility of interference from other optical communication systems. Usually, the signal strength from a direct connection is so significantly stronger than light from indirect paths, that interference is not problematic. Fog might however reflect sufficient light from other ToF lightsources. Optical interference can be avoided with ToF sensors, by using different pulse frequencies for communication. Unlike RF, pulsed optical communication does not need channel permissions and can be used without limitations over a wide spectrum. Time-of-Flight cameras are also designed to use a wide range of modulation frequencies. Due to the requirement of unambiguous range extension in depth sensing, ToF cameras are able to alternate between multiple modulation frequencies. Our experiments and simulations show that a frequency difference of hundred Hertz is sufficient to effectively suppress a signal.

5 Secured and Robust Communication with Time-of-Flight Sensors

While a lot of attack risks are mitigated by the nature of optical line-of-sight communication, location and context-awareness can further enhance security. In this Section we analyze how the position of a communication partner can be determined and how this localization benefits secure communication.

5.1 Context and Location-Aware Communication

Despite the separation of lightsources, camera based optical communication has the advantage that the sensor can help to provide valuable information about the communication partner. When a line-of-sight connection is established between two subsequent cars for crash prevention, it is important that the cars are located in the same lane.

When image sensors are used for communication, it is not always the case that the produced images are recorded. High sampling speeds, optical filters or active suppression of background light can be the reason that 2D images cannot be used to classify communication partners directly. It is however feasible to fuse data from the communication sensor with other sensors such as LIDAR, 2D or 3D stereo cameras, which are increasingly available in modern cars. The position of the projected lightsource on the image sensor can be used to determine the 3D direction of the communication partner. To determine the direction vector $\overrightarrow{d}$, the camera is approximated with the common pinhole model. The pixel position of the projected light source $x_{i,j}$ is multiplied with the pseudo inverse of the camera calibration matrix $\tilde{P}_D$.

$$\overrightarrow{d} = \tilde{P}_D x_{i,j} \tag{1}$$

The relative direction between camera and communication partner can be transformed to the global coordinate system of the car, when the position of the camera to the origin of the global coordinate system is known. The result is a line-of-sight vector, starting at the center of the camera. This vector can be used to localize the partner's illumination source by combining it with data from other sensors. It is possible to calculate the intersection between the optical line-of-sight and the sensed 3D pointcloud from on-board depth sensors. The intersection point is the 3D position of the communication partner's illumination source.

The line-of-sight can also be directly projected onto 2D images. The projected line can be used as search space to determine the exact source of information using 2D information. Given the fact that car models can be robustly determined by 2D cameras [2], it is possible to use a database to determine and verify the expected lightsource position. Then it is also possible to use triangulation between the communication image sensor and the 2D camera to determine the distance.

5.2 Benefits for Secured Communication

When two subsequent cars communicate over a line-of-sight connection, the transmitted data is already hard to manipulate. Denial of service attacks take a lot more effort, as the line-of-sight connection needs to be either blocked, or overridden by a stronger signal. When using Time-of-Flight sensors, blocking a line-of-sight connection is even harder, as overriding a lightsource need to be done with modulated light at the same pulse frequency. While message manipulation and eavesdropping is hard, state-of-the-art encryption is highly recommended. Wasef et al. [23] propose a public key infrastructure for ad hoc vehicular networks, which can be extended to line-of-sight connections. The security of state-of-the-art encryption, authentication and key exchange protocols can be further be enhanced by location-awareness.

A problem that still impairs secured connections of all kinds are relay attacks. These attacks forward unaltered encrypted connections between two unsuspecting parties. Francillion et al. demonstrate the severeness of these attacks in passive keyless car entry systems [7]. In the application of line-of-sight communication between cars, an adversary could for instance relay the signal of the front car to the car in the back, allowing a potential appearing road hazard to be undetected.

As we previously showed [17], location-aware communication can mitigate relay attacks by the exchange of localization information. Each communication partner forwards the sensed relative position of the other partner over a secured connection. The exchanged localization information has to correspond to the measured position of the communication partner. During a relay attack, the origin of information is not identical with the communication partner. Using GPS position data as localization information is simple, but vulnerable to spoofing attacks. A more reliable way is to use localization information determined by using an additional depth sensing system as described in Section 5.1. The mutual relative distance is sufficient for a successful localization verification.

6 Conclusion and Future Work

In this work, we discussed the issues of RF based V2V communication with applications demanding low latency. We showed why camera based optical communication is a viable solution for location-aware communication. Our own approach is based on Time-of-Flight depth sensing technology, which has the novel ability to use phase shifted pulsed light signals for communication. Along with suppression of background illumination and robustness against interference, this makes it a promising technology for optical communication in the V2V domain. Location-aware communication benefits communication security as it can mitigate relay attacks.

For our future work, we will implement a reliable communication solution, using Time-of-Flight cameras and an FPGA based processing system. We will implement a secured communication protocol, using state-of-the-art hardware

security anchors. Interfaces are going to be added, so our solution can be integrated into a diversity of embedded systems and can serve as a reliable secured communication solution.

References

1. A. Agarwal and T. D. C. Little. Role of directional wireless communication in vehicular networks. In *Intelligent Vehicles Symposium (IV), 2010 IEEE*, pages 688–693, June 2010.
2. M. M. Arzani and M. Jamzad. Car type recognition in highways based on wavelet and contourlet feature extraction. In *Signal and Image Processing (ICSIP), 2010 International Conference on*, pages 353–356, Dec 2010.
3. K. Bilstrup, E. Uhlemann, E. G. Strm, and U. Bilstrup. On the ability of the 802.11p MAC method and STDMA to support real-time vehicle-to-vehicle communication. *Eurasip Journal on Wireless Communications and Networking*, 2009, 2009.
4. A. M. Cailean, B. Cagneau, L. Chassagne, M. Dimian, and V. Popa. Miller code usage in visible light communications under the phy i layer of the ieee 802.15.7 standard. In *Communications (COMM), 2014 10th International Conference on*, pages 1–4, May 2014.
5. A. M. Cailean, B. Cagneau, L. Chassagne, V. Popa, and M. Dimian. A survey on the usage of DSRC and VLC in communication-based vehicle safety applications. *Proceedings of the 2014 IEEE 21st Symposium on Communications and Vehicular Technology in the BeNeLux, IEEE SCVT 2014*, pages 69–74, 2014.
6. N. Druml, G. Fleischmann, C. Heidenreich, A. Leitner, H. Martin, T. Herndl, and G. Holweg. Time-of-Flight 3D Imaging for Mixed-Critical Systems. In *13th International Conference on Industrial Informatics (INDIN)*, pages 1432–1437, July 2015.
7. A. Francillon, B. Danev, and S. Capkun. Relay Attacks on Passive Keyless Entry and Start Systems in Modern Cars. In *Network and Distributed System Security Symposium (NDSS)*, February 2011.
8. V. Frey. Pmd cameras for automotive and outdoor applications, 2010.
9. Google. Google self-driving car, 2016.
10. G. Gstrein and W. Sinz. Improvement of airbag performance through pretriggering.
11. H. Hasbullah, I. A. Soomro, and J.-l. A. Manan. Denial of Service (DOS) Attack and Its Possible Solutions in VANET. *World Academy of Science, Engineering and Technology*, 4(5):411–415, 2010.
12. Y.-h. Kim, W. A. Cahyadi, and Y. H. Chung. Experimental Demonstration of LED-based Vehicle to Vehicle Communication under Atmospheric Turbulence, 2015.
13. Y. H. Kim, W. A. Cahyadi, and Y. H. Chung. Experimental Demonstration of VLC-Based Vehicle-to-Vehicle Communications Under Fog Conditions. *IEEE Photonics Journal*, 7(6):1–9, 2015.
14. R. Lange and P. Seitz. Solid-state time-of-flight range camera. *IEEE Journal of Quantum Electronics*, 37(3):390–397, March 2001.
15. S. Nishimoto, T. Nagura, T. Yamazato, and T. Yendo. Overlay Coding for Road-to-Vehicle Visible Light Communication using LED Array and High-Speed Camera, 2011.
16. P. Pathak, X. Feng, P. Hu, and P. Mohapatra. Visible Light Communication, Networking and Sensing: A Survey, Potential and Challenges. *IEEE Communications Surveys & Tutorials*, 17(c):1–1, 2015.

17. H. Plank, C. Steger, T. Ruprechter, G. Holweg, and N. Druml. Survey on Camera based Communication for Location-Aware Secure Authentication and Communication, 2016.
18. R. D. Roberts. A MIMO protocol for camera communications (CamCom) using undersampled frequency shift ON-OFF keying (UFSOOK). *2013 IEEE Globecom Workshops, GC Wkshps 2013*, pages 1052–1057, 2013.
19. I. Takai, S. Ito, K. Yasutomi, K. Kagawa, M. Andoh, and S. Kawahito. LED and CMOS image sensor based optical wireless communication system for automotive applications. *IEEE Photonics Journal*, 5(5), 2013.
20. M. Tobias, K. Holger, F. Jochen, A. Martin, and L. Robert. Robust 3D Measurement with PMD Sensors. *Range Imaging Day, Zürich*, 7(Section 5):8, 2005.
21. O. K. Tonguz, N. Wisitpongphan, J. S. Parikh, F. Bai, P. Mudalige, and V. K. Sadekar. On the broadcast storm problem in ad hoc wireless networks. In *2006 3rd International Conference on Broadband Communications, Networks and Systems*, pages 1–11, Oct 2006.
22. P. Tyagi and D. Dembla. Investigating the security threats in Vehicular ad hoc Networks (VANETs): Towards security engineering for safer on-road transportation. *Proceedings of the 2014 International Conference on Advances in Computing, Communications and Informatics, ICACCI 2014*, pages 2084–2090, 2014.
23. A. Wasef, R. Lu, X. Lin, and X. Shen. Complementing public key infrastructure to secure vehicular ad hoc networks [security and privacy in emerging wireless networks]. *IEEE Wireless Communications*, 17(5):22–28, October 2010.
24. H. B. C. Wook, T. Komine, S. Haruyama, and M. Nakagawa. Visible light communication with led-based traffic lights using 2-dimensional image sensor. In *CCNC 2006. 2006 3rd IEEE Consumer Communications and Networking Conference, 2006.*, volume 1, pages 243–247, Jan 2006.
25. W. Yuan, R. Howard, K. Dana, R. Raskar, A. Ashok, M. Gruteser, and N. Mandayam. Phase messaging method for time-of-flight cameras. In *IEEE International Conference on Computational Photography (ICCP)*, May 2014.

Chapter 1
Localization and Context Determination for Cyber-Physical Systems Based on 3D Imaging

Hannes Plank
Infineon Technologies Austria AG, Austria

Josef Steinbaeck
Infineon Technologies Austria AG, Austria

Norbert Druml
Independent Researcher, Austria

Christian Steger
Graz University of Technology, Austria

Gerald Holweg
Infineon Technologies Austria AG, Austria

ABSTRACT

In recent years, consumer electronics became increasingly location and context-aware. Novel applications, such as augmented and virtual reality have high demands in precision, latency and update rate in their tracking solutions. 3D imaging systems have seen a rapid development in the past years. By enabling a manifold of systems to become location and context-aware, 3D imaging has the potential to become a part of everyone's daily life. In this chapter, we discuss 3D imaging technologies and their applications in localization, tracking and 3D context determination. Current technologies and key concepts are depicted and open issues are investigated. The novel concept of location-aware optical communication based on Time-of-Flight depth sensors is introduced. This communication method might close the gap between high performance tracking and localization. The chapter finally provides an outlook on future concepts and work-in progress technologies, which might introduce a new set of paradigms for location-aware cyber-physical systems in the Internet of Things.

DOI: 10.4018/978-1-5225-2845-6.ch001

INTRODUCTION

3D imaging technologies have seen rapid developments in the past years. The introduction of the Microsoft Kinect depth sensor to the consumer market in 2010 triggered a massive research interest and effort. In 2016, the first mass-produced smartphone appeared, featuring depth sensing based on Time-of-Flight. The availability of such ubiquitous and miniaturized depth sensing solutions can tremendously help any kind of electronic device to sense and understand its environment.

A crucial part of operation for certain devices is localization. While depth sensors provide geometric information about the immediate surrounding, determining location and orientation within a certain coordinate system is a challenge of its own. This chapter explores the opportunities depth sensing systems provide to localization. A focus is set on applications in fields such as consumer electronics, internet of things and autonomous robots. Localization and tracking of electronic devices has a long history and has seen the use of a variety of different principles. This work focuses on the fields of high performance localization based on optical and sensor fusion solutions. Localization principles in general can be categorized into passive, guided and cooperative solutions.

A passive system is able to determine its position in a local or global coordinate system without external help. An increasing number of applications also require information about the orientation of the device. A combination of position and rotation sensing is often referred to as pose determination. A pose has of six degrees of freedom (dof) and completely describes the static position and orientation of an entity in 3D space. Each axle in 3D space presents one degree of freedom for the position and one degree for rotation around the axle. Passive 6-dof localization is often used in computer vision based positioning systems, where features are matched with prerecorded databases. Early examples are cruise missiles using terrain contours for navigation.

A well-known example for guided localization is GPS, where devices use the aid of satellites to determine their position. Cooperative localization solutions use a communication channel, which is often used for active identification and position forwarding. Optical tracking, using image sensors and active markers is an example for cooperative tracking. In such system, an external base-station with an image sensor can sense a tracked device equipped with active LED markers, and has the ability to toggle the LEDs for identification. Another example are beacon based systems, where active beacons forward information about their location.

When classifying the location-awareness of cyber-physical systems, it is important to distinguish between localization and tracking. While these terms are sometimes used ambiguously, tracking is commonly used in a relative context, where the registration of movements is important. Tracking the pose of a device does not always lead to the determination of a position within a meaningful coordinate system. However relative position and rotation changes can be detected. For certain applications, this is sufficient and no broader localization is required. Examples for such systems are instruments measuring rotations, such as gyroscopes or compasses, some 3D scanning solutions or human interface devices.

Localization is often associated with position determination without focus on detecting relative pose changes. A combination of tracking and localization is used in a lot of location-aware systems and leads to localization at a high update rate. Tracking and localization are often performed by different sensors, because localization solutions often lack of the desired accuracy to track relative pose changes. While localization can provide the position and orientation within a greater context, tracking sensors provide the accuracy and update-rate required for the application.

A great example of sensor fusion for localization and tracking is Wikitude (2016). This smartphone application provides augmented reality on smartphones. It annotates the video stream of the internal camera with meaningful information about the environment and displays it on the screen. GPS or WIFI is used for positioning. The absolute orientation is determined by the gravity and compass sensors. The gyro sensors are used to track movements to enable a high update rate of the rotation. This enables to robustly attach information to certain landmarks and directions in the smartphone video stream.

Arising technologies such as virtual and augmented reality, autonomous driving and indoor localization demand precise pose determination at very high update rates. These demands are tackled in state-of-the-art systems with a sensor fusion approach, combining optical and inertial sensors. Optical sensors are used in the form of 2D and 3D cameras, LiDAR (light detection and ranging) and optical triangulation.

Data fusion with inertial sensors can compensate the flaws of optical sensors. Most optical positioning sensors require a line of sight connection and sometimes feature a slow update rate and too much latency as well as visual artifacts such as motion blur. Inertial sensors are commonly miniaturized, using MEMS technology and feature high update rates. These sensors are already well-established in mobile devices. Inertial sensors however base their measurements on differential movements and rotations. In order to measure absolute movements and rotation, the measurements need to be integrated. This introduces an integration error which introduces drift (Woodman, 2007). This drift can be compensated by fusing these measurements with non-drifting data, as optical systems can produce.

Using cameras for localization is traditionally accomplished by sophisticated computer vision methods, which are often solely based on 2D images. A common approach is simultaneous localization and mapping (SLAM), where a 3D representation of the environment is created during localization. Depth sensors are capable to improve the performance of such monocular SLAM systems (Steinbrücker, Sturm, & Cremers, 2011). A prominent example is Google Tango, which uses a number of sensors, including a Time-of-Flight depth camera for SLAM based localization on mobile devices.

Depth sensing systems are devices capable to directly gather geometric information about the immediate environment. A measurement typically consists of a coordinate in three dimensions, usually relative to the depth sensor itself. In a depth camera, every pixel produces such measurement. If the shutter of the camera triggers all pixels at the same time, every measurement is captured simultaneously. Since all available depth imaging systems are limited in range and show systematic measurement errors, some systems also attach a confidence measure to each measurement.

One of the reasons that the vast research effort on optical localization systems are based on 2D cameras, is that the field of depth sensors is much younger than 2D sensors and they are not yet part of common vision systems. This might change in the smartphone and tablet domain, since as of 2016 the first smartphone featuring a Time-of-Flight depth camera appeared on the consumer market. Measuring depth based on the Time-of-Flight principle is the most miniaturized solution available and has the flexibility to be used in a manifold of applications.

DEPTH IMAGING FOR 6-DOF LOCALIZATION

This section introduces current depth imaging solutions with focus on Time-of-Flight technology. Depth sensors alone are usually not directly associated with localization, although they provide 3D-context awareness of the immediate surrounding. A sensor fusion approach, incorporating depth sensors however offers advantages in SLAM based systems (Steinbrücker, Sturm, & Cremers, 2011), and can improve

tracking precision in general. In this work, we focus on depth sensing based on Time-of-Flight, since it is the most miniaturized solution, and is the only depth sensor, which can be found in mass produced smartphones. We also present a concept in this chapter, where Time-of-Flight sensors are directly used for location-aware optical communication, closing the gap between depth sensing and localization.

THE PRINCIPLE OF TIME-OF-FLIGHT DEPTH SENSING

Time-of-Flight imaging is based on measuring how long light takes to travel from the sensing system to the scene and back to a photosensitive sensor. These systems can be distinguished by their operating principle, as illustrated in Figure 1. Direct Time-of-Flight systems emit short light pulses and measure the time between emission and reception. Each pulse corresponds with one measurement. A prominent example is LiDAR, where bundled light pulses are emitted by a laser and detected by photodiodes. The angle of these pulses is modulated to receive a spatial image of the surroundings. Indirect ToF imaging sensors relay on a continuous wave approach.

The operating principle of this indirect approach is illustrated in Figure 2 and works by emitting pulsed infrared light at a wide radiation angle. An image sensor receives the reflected light and is able to measure the phase-shift between incoming and outgoing signal, which is proportional to the distance.

The emitted light pulses usually have a frequency of up to 100 MHz. They originate from an active illumination unit, which typically consists of an infrared LED or vertical cavity surface emitting laser (VCSEL). The pulses travel to the sensed object and back to the image sensor. The lens projects the light onto a ToF image sensor.

Each pixel of the ToF sensor contains a photonic mixer device, which produces a signal corresponding to the phase-shift between the outgoing and incoming light (Tobias, 2005). The incoming photons generate charge-hole pairs in the silicon, which are moved into either of two charge buckets A and B. This is decided by the logical state of the modulation signal. This signal is generated on the ToF sensor and is also used to emit light pulses. In order to control the PMD devices, the signal is directly supplied to each pixel. The charge difference between bucket A and B is the output of the ToF sensor and is related to the phase-shift of the incoming and outgoing signals. This phase output value however also depends on the amount of incoming light. Reflection properties of the sensed material and as well the

Figure 1. The different principles of Time-of-Flight based depth measuring

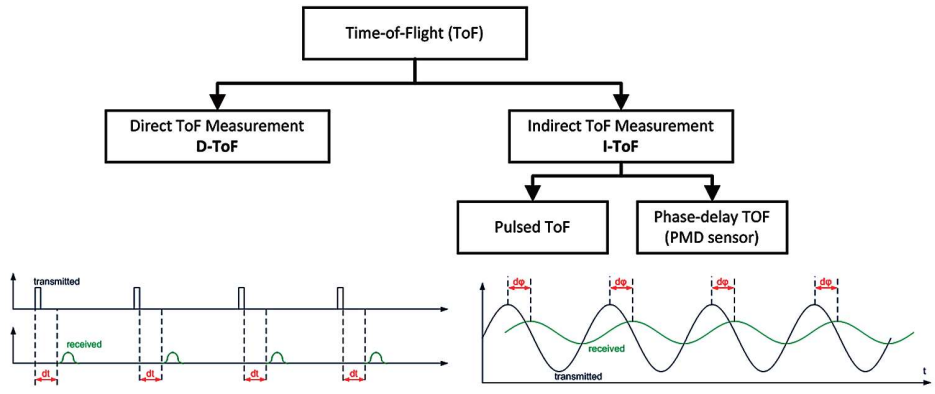
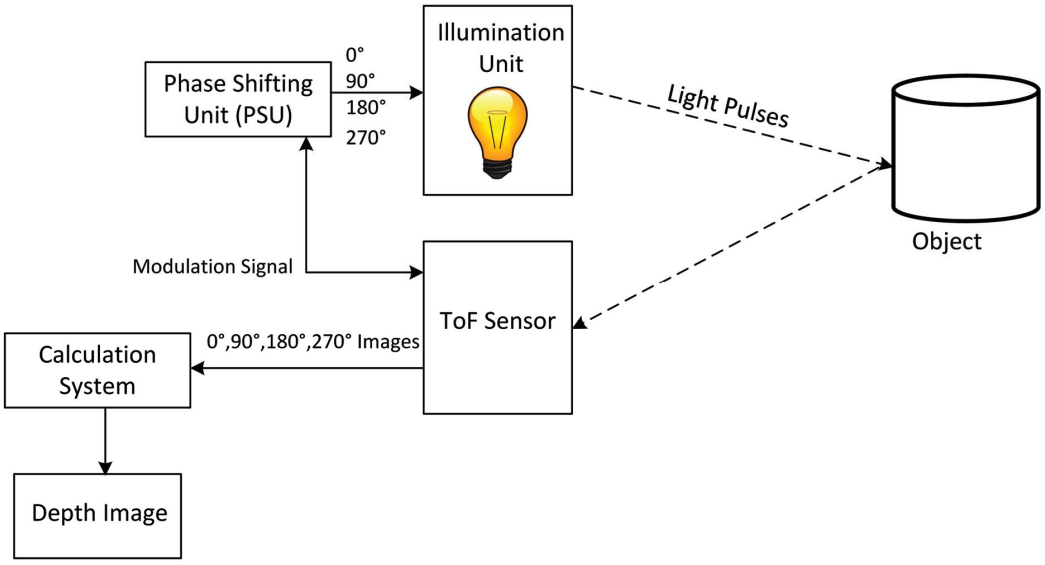


Figure 2. The working principle of continuous wave indirect Time-of-Flight 3D imaging



distance influence this value. In order to determine the exact phase-shift, the most common way is to conduct four or more different measurements. In these measurements, the phase-shift offset between the outgoing light and the modulation signal is changed, producing four or more different phase images.

A well-established procedure is to take four measurements I with four equidistant phase offsets (e.g. 0°, 90°, 180°, 270°) and calculate the phase-offset by the following relation:

$$p = \operatorname{atan}\left(\frac{I_0 - I_{180}}{I_{90} - I_{270}}\right)$$

Since p is proportional to the phase-shift of a pulsed light signal, the phase values wrap and start again at zero, if the distance is too long. Time-of-Flight sensors are capable to change their modulation frequency in order to produce another different set of four phase images to unwrap these ambiguous phases. The final measurement for a depth image consists then of eight phase-images, which results in a larger range, while maintaining precision. The tradeoff is a decreased frame-rate and potential motion artifacts. The eight-phase mode is commonly used in SLAM applications, which prefer depth quality over frame-rate.

The drawback of Time-of-Flight sensors is limited resolution of the depth image, since each pixel contains the circuit of the photonic mixer device. The limited photo-sensitive area on the silicon is compensated partly by using micro-lenses directly applied above the silicon to focus light to the photo-sensitive areas of a pixel. Since continuous wave Time-of-Flight sensing is the most miniaturized depth sensing system available, increasing the sensor size to enhance resolution or photosensitivity is often not feasible. Another drawback is the sensitivity to incoming background light. Despite countermeasures such as background illumination suppression (BSI) and infrared filters, very bright light sources such as reflected sunlight can reduce the signal to noise ratio, but do not directly influence the measurement.

ALTERNATIVE DEPTH IMAGING APPROACHES

Each available depth imaging system has its own trade-offs and no system is yet predominant. Compared to Time-of-Flight, all image-sensor based approaches need a certain size to conduct measurements based on triangulation. Stereo depth sensors usually consist of two cameras which are mounted in parallel with a certain distance. This baseline between these cameras is necessary, because depth is measured by the pixel offset between features in two images. A short baseline impairs depth quality and range. However, progression in research and increased sensor resolution made rather small baselines feasible, like they are found in form of dual camera systems in smartphones. Stereo cameras usually produce sparse depth measurements, since it is not possible to measure feature disparities of homogenous surfaces. The depth image's x/y resolution however is superior to most other depth sensing principles, because high resolution image sensors can be utilized for stereo. A sensor fusion approach, combining ToF and stereo depth sensors (Zhu, Wang, Yang, Davis, & Pan, 2011), is a promising solution, when high quality depth images are required.

A variation of stereo sensors is structured light (Scharstein, 2003). Structured light utilizes a projector in order to project a pattern onto the scene. An infrared camera senses the projected pattern. Since the projected pattern is predefined, the operating principle is similar to a stereo camera setup, as both systems are based on extracting the feature disparity caused by the distance to the camera. The famous Microsoft Kinect sensor is based on this technology. Unlike stereo, dense depth images cannot be gathered, as the projected pattern is also reflected on homogenous surfaces. Due to the active illumination principle, structured light based systems can also operate in low light conditions. The drawback of this system is that the active illumination requires more energy, like ToF, and the system is also impaired by ambient light. Unlike ToF, such systems also require a certain distance between projector and camera.

Depth imaging based on LiDAR is most commonly used in automotive applications. Unlike image sensor based approaches, the angular variation of LiDAR measurements does not stem from optical projection onto an imaging sensor, but from mechanical scanning devices. This mitigates multi-path interference, and allows the usage of highly sensitive photo elements, such as single photon avalanche diodes. This increases the effective range at the cost of more complex sensing systems.

DIRECT LOCALIZATION WITH DEPTH SENSORS

Depth imaging sensors currently are not commonly used in the field of localization of electronic devices. One of the reasons is that most of the research on image based localization is focused on the far more established and ubiquitous color cameras. 3D data does not have as much variation and distinctive features as 2D images. Geometry is more generic and repetitive than reflected light. Due to the measurement principles, 3D imaging also suffers from resolution and limited range. Geometry however is more consistent, since it is not influenced by different illumination. With depth sensors, it is possible build dense 3D models of the environment, as demonstrated by Newcombe et al. (2011) in their Kinect Fusion approach. Such 3D models can be used for re-localization and mapping the environment for systems such as autonomous robots. Another application for such high-quality depth maps is augmented reality. Due to the integration of depth data to a dense model, it is possible to embed virtual objects, using the high-quality depth data.

Biswas and Veloso (2012) present an indoor navigation system based on depth information using a plane filtered pointcloud and Monte Carlo localization. Since solely depth sensors are used, the robots

can also dynamically avoid obstacles, which are not in the reference map. While these approaches might work in a smaller context, with distinctive geometry, sensor fusion approaches are more favorable for localization. Later in this chapter, we explore the concept of using Time-of-Flight depth sensors in a novel way to establish location-aware optical communication links to embedded devices. If these devices forward positional information, highly accurate cooperative localization is possible by solely using miniaturized depth sensors.

DEPTH AND COLOR SENSOR FUSION

Depth and color cameras can be combined to create a unified RGB-D sensor, capable to capture images containing both depth and color information. When a SLAM or visual localization system is provided with depth information, the system can either build better 3D maps, but also benefit in robustness from the immediately available distance data (Steinbrücker et al., 2011). The process of 3D and 2D data fusion can also improve depth image resolution, if the required processing power is available. This is due to the principle that depth edges often correlate with edges in color and intensity images. Color edges without any depth discontinuity can be simply ignored, when the lower resolution depth image does not show any variation. Since depth imaging systems usually lack high resolution, color edge information can be used to interpolate and upscale depth images in a meaningful way.

Research has produced a large number of image-guided depth upscaling algorithms (Chetverikov, Eichhardt, & Jankó, 2015). The input for most methods is a high resolution color image and sparse depth measurements which are mapped into the color image space. This sparse RGB-D image is produced by 3D measurements which are mapped to the 2D image. The requirement for such mapping is knowledge about the intrinsic camera parameters of both cameras, which include distortion coefficients, focal length and pixel offsets of the optical centers. These parameters are usually derived by checkerboard calibration, which is also possible for Time-of-Flight depth cameras, since they are able to produce a grey-scale intensity image. The extrinsic parameters of a dual camera system are in this case a translation vector T and a rotation matrix R, which describe the transformation from the depth camera's coordinate system to the color camera. These parameters can also be gathered by capturing images of a checkerboard pattern, and using a toolchain, such as Camera Calibration toolbox (Bouguet, 2016). If intrinsic and extrinsic camera parameters are known, depth measurements with depth $d_{i,j}$ and pixel position $x_{i,j}$ can be mapped to 3D space of the color camera, using the pseudoinverse P of the intrinsic depth camera matrix:

$$X_{i,j} = T + Rd_{i,j} \frac{Px_{i,j}}{Px_{i,j}}$$

The 3D measurements $X_{i,j}$ can be projected to 2D image space coordinates $v_{i,j}$ by multiplication with the intrinsic matrix I of the color camera.

$$v_{i,j} = I\, X_{i,j}$$

Methods to interpolate these projected depth images involve various principles, such as energy minimization (Ferstl, Reinbacher, Ranftl, Ruether, & Bischof, 2013), graph based methods (Dai, Zhang, Mei, & Zhang, 2015) or edge aware interpolation (Plank, Holweg, Herndl, & Druml, 2016). Most methods are designed with focus on depth image quality and not for efficiency. In location-aware cyber-physical systems, interactive framerates are desired. When the depth images are used to build 3D maps in a SLAM system, several depth images per second are desired. Methods to create high resolution depth images on restricted hardware are relatively rare. While there are approaches, which emphasize on low computational complexity (Dai et al., 2015), there are not many implementations which can be executed on parallel processing systems such as GPUs. The joint bilateral filter, developed by Kopf et al. (Kopf, Cohen, & Lischinski, 2007) can be executed in parallel and works by weighting image filter kernels on color similarity. This works well with relatively small upscaling factors. If, however, more than just a few pixels between depth values need to be interpolated, depth values influence pixels despite edges between the interpolated pixel and the original depth value. We therefore developed an edge-aware interpolation approach which is optimized for GPUs on mobile devices (Plank, Holweg, Herndl, & Druml, 2016). In this approach, sparse depth values are mapped to high-resolution color images. Each depth value propagates its influence among circular paths to the surrounding pixels. If edges are crossed, the influence drastically decays. If no edges are crossed, the influence is evenly distributed, suppressing sensor noise. Our prototype implementation is capable of producing 13 frames per second, when executed on GPUs on mobile devices. Beside the capability of SLAM systems to create better geometric models of the environment, the availability of high resolution 1-1 mapped depth and color images, enables better context awareness, since combined depth and color data benefits 2D object detection algorithms (Yu & Zhao, 2012).

A rather unexplored issue is synchronization among camera systems. Most academic work assumes that color and depth sensors operate in synchronous mode, since the technical solution to a synchronous camera system seems trivial. In practical applications, however, synchronization is often not feasible as it usually requires tight cooperation across multiple hardware vendors. The vast majority of image sensors are developed to be integrated into monocular systems, not offering any option for hardware synchronization. For RGB-D SLAM systems however, synchronization can be avoided by using additional sensors. Such system is usually in motion while the environment remains static. If color and depth cameras gather information at different times, this can be corrected by using motion tracking data from an inertial sensor. Inertial sensors can operate at a high update-rate. If all measurements, including the depth and color images are accurately timestamped, the actual pose difference between depth image and color image can be calculated by transforming the depth data by the relative pose derived by the motion tracking system. By using physical models of the tracking system, the relative pose can be even more refined.

Ovren et al. (2013) introduce this approach in their attempt to correct the effect of rolling shutter image sensors. A rolling shutter is caused by the pixel readout process, and means, that not all color pixels are captured at the same time. Inertial based depth image alignment is only possible with static scenes, because they are only able to compensate the motion of the camera system.

LOCALIZATION WITH 2D IMAGING SENSORS

Feature Based Localization

Color information alone can be directly used for localization by matching visual input data against databases (Vedaldi & Soatto, 2008). This works by finding features, which are regions or points in input images which are significant and distinctive. Feature descriptors are an abstract representation of these regions with the goal of being able comparable to other features, while being resilient against pose and intensity variations. These feature descriptions can be stored in a visual dictionary, associating these features with localization information. These databases are either generated systematically via 3D scanning systems, or using topologic information. If these features are recognized by a vision system, the pose of the system can be calculated by triangulation. This kind of visual localization can also be used to initialize a visual SLAM system (Lepetit, Arth, Pirchheim, Ventura, & Schmalstieg, 2015). SLAM can provide more accurate localization, by creating a 3D data representation of the sensed environment.

Visible Light Localization

The motivation of visible light localization (VLC) is caused by the increasingly ubiquitous LED lights. Due to the fast switching speed, it is possible to transmit information from repurposed existing illumination systems. With appropriate modulation, it is possible to use lights for communication without perceivable flickering. Visible light based localization has different applications with different demands in precision. Current products, such as Qualcomm Lumicast (Jovicic, 2016), repurpose image sensors for visible light localization. The distinction between vision based methods with active targets and visible light localization is that base-stations transmit information to help with localization. This can be either IDs or positional information of the base station.

Do and Yoo (2016) provide an extensive survey on methods and implementations of visible light based positioning systems. Such systems usually consist of base-stations, which emit encoded light to electronic devices, equipped with photo-detection sensors. The base stations are either re-used light sources, such as traffic lights or lamps, or dedicated infrared beacons. The simplest solution is based on proximity detection. Such systems can be implemented with just a photodiode as receiver. It is however only possible to detect the base-station itself, so only a very coarse localization with an uncertainty of several meters (Campo-Jimenez, Martin Perandones, & Lopez-Hernandez, 2013) is possible.

Finger printing based methods can achieve more precise positioning, but require pre-recorded maps for localization. Time difference of arrival (TDOA) is another method, which works by receiving light signals from multiple base-stations. For 3D localization, at least three base-stations need to be in the direct field of view. The tracked device directly measures the distance to the base-station in this method. This is accomplished by measuring the time it takes the light pulses of each base station to travel to the device. The position is then determined by trilateration. Such localization systems require a good synchronization of the base stations and the localized device. A single photodiode can be used to receive the signals, and a method to separate the received signals of the base-stations has to be employed. This is possible by using time or frequency division multiplexing (Liu, et al., 2014). TDOA localization is not limited to the optical domain. It is possible to also use radio or sound waves; however, multi-path effects need to be considered. Image sensors might be capable to measure time-differences, and also offer light-source separation due to the projection via lenses. The position on the pixel can be used to detect the angles

between the base-stations. Due to the pixel readout process, image sensors usually cannot be sampled at the required rates. It either requires dedicated image sensors, featuring customized electronic shutter units or direct pixel readout. Dynamic vision sensors (Censi, Strubel, Brandli, Delbruck, & Scaramuzza, 2013) are a promising development, and might be able to conduct such measurements. An active vision sensor does not produce images, but events which describe local pixel value changes. The difficult synchronization between device and base-stations might deem such image sensor based approaches unfeasible, because it is also possible to determine the position on triangulation alone. Time-of-Flight 3D imaging sensors might be capable to support TDOA based localization, but to our knowledge, this has not yet been investigated and is subject for future research.

Another method to determine the distance between optical receivers and LEDs is to measure the received signal strength (RSS). Calculating the distance to light sources is based on modelling the light source and its propagation path to photo sensors. After calibration, distances can be associated with the output of optical receivers. When receiving the signal from multiple base-stations, the signal differences between the base-stations can be used to calculate the distance. This mitigates the influence of background illumination. RSS based positioning has the potential to be simply implemented and widely adopted, since no synchronization is necessary. The problem however is that the received signal strength depends on the orientation of the photo detector relative to the LEDs as well. The light strength also depends on the orientation of the LEDs, since light is radiated inhomogeneously. It is however possible to combine RSS with angle based localization methods (Mauro Biagi, 2015).

Localization based on triangulation requires systems, which are able to measure the angle of arrival (AOA). This can be either accomplished by an array of photodiodes (Lee & Jung, 2012) or by using image sensors. AOA systems are in general more complicated, but do not require synchronization. With the help of a 3D camera, it is possible to localize the relative positions of the base stations, and combine trilateration and triangulation in order to improve the localization accuracy. It is also possible to avoid determining the positions of the base-stations beforehand. If only relative movements need to be detected, the base stations can be supplied with their relative locations from the 3D camera system via optical communication. In the next section, we present our OptiSec3D approach, which enables these concepts by combining Time-of-Flight depth sensing with optical communication.

LOCATION-AWARE OPTICAL COMMUNICATION BASED ON TIME-OF-FLIGHT SENSORS

The operating principle of Time-of-Flight depth sensors requires an image sensor, capable of demodulating phase differences of pulsed light. In this section, we present our effort to create a novel location-aware optical communication system. We further go into detail, how it might benefit future localization and tracking systems in the fields of IoT and cyber physical systems.

The most significant feature of image sensor based optical communication is the directional awareness of the communication partner. If depth imaging sensors are used for optical communication, it is even possible to track communication partners in 3D. While there exists a manifold of image sensor based optical communication systems, Time-of-Flight sensors have not yet been widely explored for optical communication.

A first attempt was made by Yuan et al. (2014), who establish a one-way communication link between a Time-of-Flight camera and an array of modulated LED lights. The sending device avoids the

required synchronization by recovering the Time-of-Flight sensor's modulation signal with a photodiode. The emitting LEDs are supplied with a phase-modulated modulation signal and manipulate depth measurements of the ToF sensor. These depth measurements are analyzed and the received information is extracted. Since the Time-of-Flight sensor is operated in normal depth sensing mode at relatively low frame-rates in this approach, multiple LEDs are used to transmit information in parallel. Such multiple input approaches are limited in range, since the pixel array cannot resolve individual LEDs when a certain distance is exceeded.

If a system however is capable of configuring and controlling Time-of-Flight 3D imaging systems with a direct connection and a real-time system with low level configuration access, optical communication parameters can be changed to increase readout speeds and it is also possible to use just single modulated LEDs to send information. In our OptiSec3D approach, we utilize Time-of-Flight sensors as optical transceivers, which are also capable to incorporate depth measurements into the communication protocol. Our approach has the potential to reach a throughput of several kilobits per second.

Operation Principle of the OptiSec3D Approach

Indirect Time-of-Flight sensing works by emitting pulsed infrared light. The active illumination unit of such system can be used as transmitter, since it is designed to emit pulsed infrared light at different phase-shifts. The pixels of the receiving Time-of-Flight image sensor are capable to demodulate the phase-shifted signal. This allows optical communication based on pulsed light phase-shift keying (PLPSK). The vast advantage of PLPSK is that multiple bits can be encoded in one image. In most image sensor based approaches, simple binary modulation schemes, such as on/off keying (Roberts, 2013) or pulse position keying are used. They support the transmission of one bit per frame at best. PLPSK takes advantage of the photonic mixer device (PMD), located on each pixel of a Time-of-Flight sensor. The PMD decodes phase differences of incoming light pulses, by sorting the incoming charges into charge storage buckets A and B on the pixels. After the readout process, the voltage difference of these buckets is proportional to the phase difference between the own modulation signal and the incoming light pulses. If these light pulses are phase-modulated, the output of the ToF sensor contains the decoded phase offset signal. Since at least four equidistant phases are used during communication, it is possible to decode phase differences by just using one frame, instead of at least four frames used during depth measurement.

Channel Characteristics

Due to the measurement principle of the PMD on each pixel, the sensor is sensitive to pulsed light within a certain frequency range. Non-pulsed background light does not have a direct influence on the measurement. The photons of continuous light arrive during both switching states of the PMD with near equal intensity. This fills both charge buckets equally, leading to increased noise but no measurement bias. Due to this principle, extensive image processing is unnecessary, as it can be assumed, that all detected signals originates from potential communication partners.

Another side-effect of the PMD pixels is that sensors can choose modulation frequencies from a large spectrum to communicate. Stray light from different connections from different systems do not directly influence the measurement, as long the light pulse frequency is just several thousand Hertz apart. There exist no experimental evaluations so far, but a viable spectrum of 16 to 26 MHz can potentially yield to

1000 different channels, with a rather large distance of 10000 Hz. Using fast-switching VCELs, instead of LEDs, the upper boundary of the spectrum can be extended to over 100 MHz.

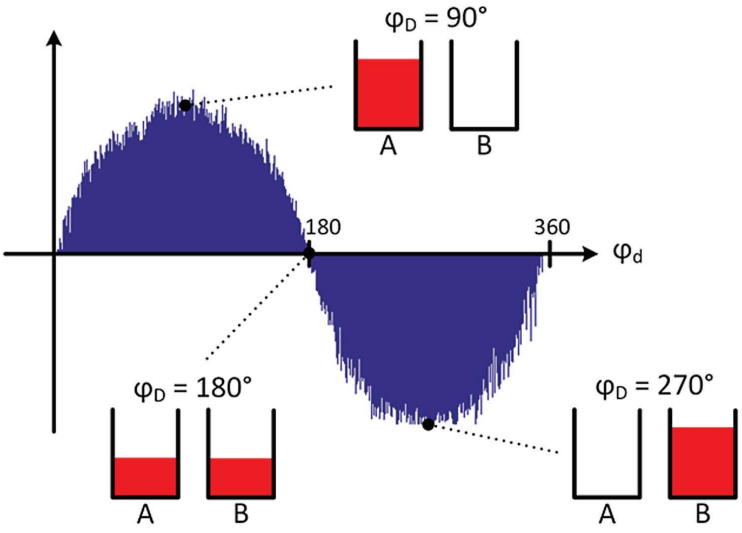
Image sensor based optical communication usually suffers from the low frame-rates of the sensors. While there exist experimental dedicated communication image sensors (Takai, 2013), the necessary digitization of complete frames has been holding back high speed image sensor based communication links so far.

Time-of-Flight sensors need to capture up to eight phase images in order to create one depth image. Therefore, the readout and analog-digital conversion circuitry is often optimized for fast readout and digitization. In order to demodulate a line-of-sight communication signal, just a very short exposure time is required. These characteristics enable high frame-rates during optical communication. We manage to operate our Time-of-Flight sensor at 500 frames per second with full sensor readout. This however can be massively increased, if the sensor is only read-out partially. By configuring the readout region to a minimum of 16x32 pixels, we are able to reach 7300 frames per second. This can lead to a transmission throughput of 14600 bits per second, when using 4-PLPSK. This high framerate leads to fast light source tracking capabilities, supported by fast adoption of the readout window, to accommodate moving targets.

Sensor Synchronization

A technical challenge throughout many communication systems is synchronization. In Time-of-Flight based communication, it is important that the frequency of the modulation signal of a ToF camera matches the frequency of its communication partner. If the modulation frequency of a ToF sensor is different from the frequency of the incoming light pulses, the measured phase values start to drift. Figure 3 shows sampled phase values, when the sender continuously emits pulsed infrared light without phase-shift. If this signal is measured and digitized, the frequency of this signal is the absolute modulation frequency difference between both communication partners.

Figure 3. The observed phase with corresponding charge bucket contents, if sender and receiver are not synchronized



Synchronization can be accomplished by sampling the incoming light pulses, and calculating the frequency difference of the communication partner. The modulation frequency of Time-of-Flight cameras is usually configurable to accommodate different use-cases. In the case of OptiSec3D, the frequency is adapted by configuring a phase locked loop (PLL) on the sensor. Synchronization can be reached, if the PLL is adapted by the measured frequency difference.

Implementation of OptiSec3D

The core of each OptiSec3D communication partner is an Infineon Real3™ 3D imaging sensor, based on Time-of-Flight technology of pmd technologies. The Xilinx Zynq 7000 platform is used in our platform to operate the sensor with software executed on its integrated ARM processors, while the FPGA is used as glue logic and for imaging data transmission. The software uses an I2C bus to configure the ToF imaging sensor. This live-configuration of the sensor allows changing the internal workings of a normal depth sensor in such way, that it is possible to transmit and decode data. This works by limiting the number of digitized pixels per frame to a small area around modulated lightsources. With this configuration, the sensor is able to sample the image of the communication partner at over 7300 frames per second. The received signal directly contains the transmitted decoded information, since ToF pixels are sensitive to phase-shift differences.

Communication Modes

Time-of-Flight cameras could be either used to communicate with each other, or with different electronic devices. While it is not difficult to implement a PLPSK transmitter, receiving PLPSK is not trivial, since a photonic mixer device is required. ToF cameras however are able employ alternative modulation schemes, such as pulse position modulation. A receiver would just need to be able to detect the presence of light within certain time-slots. Time-shifts could be implemented either directly by accessing the illumination unit, by varying the frame-rate, or the number of read-out pixels.

Optical line-of-sight communication is not limited to two communication partners. The aforementioned concepts can be employed to multiplex communication between multiple partners. For synchronization, the adapted frequency for each communication partner can be stored, and the PLL adapted each time, when switching the focus to a different partner. If multiple devices want to communicate with a single node, all other devices could alternatively adjust their frequency to the node. If the focus of an application is on low latency rather than throughput, the sensor can be read-out completely, instead of adapting the readout region to the location of each communication partner.

Localization Principles with Location-Aware Optical Communication

Location-awareness and re-usability are the main motivations for optical communication based on image sensors. While 2D image sensors are only capable to determine the incident angles of a line-of-sight connection, Time-of-Flight sensors can locate communication partners in 3D.

Since Time-of-Flight imaging systems can be used as optical transmitter and receiver, the simplest use case is optical communication between two Time-of-Flight cameras. When both communication partners measure their mutual distance, they can both locate each other in 3D. In the application of e.g. encrypted device authentication, both partners can forward their mutual distance measurement and check

for consistency. This effectively defeats relay attacks, where an attacker relays communication without alteration, using two relay boxes. In such relay attack, the distance between sender and relay box A is not consistent to the distance of receiver and relay box B.

Communication between Time-of-Flight cameras could also be employed in localization solutions. In that case, a stationary camera with a light emitter serves as beacon. Electronic devices, equipped with a Time-of-Flight camera can contact one beacon to determine its position and orientation. An example could be an autonomous robot, desiring to navigate around a building.

The beacon's purpose is to forward its own position and the angles relative to the device. When simple LED beacons are used, at least three of them are necessary to determine the camera's position, even when the camera can determine the distance to each beacon. If the beacon however features an image sensor, as depicted in Figure 4, the incidence angle of the line-of-sight can be determined by the beacon and forwarded to the device. This enables 5-dof localization of embedded devices, using a single beacon. The only unknown degree of freedom is the roll angle, since the beacons optical signature is invariant to rotations around this axis. Sensor fusion with a gravity sensor or computer vision methods can effectively help reach full 6-dof localization. In the case of autonomous robots, this angle might already be locked due to the camera mounting method.

While beaconing with Time-of-Flight sensors enables a miniaturized localization system, simpler and cheaper beacons might be desirable. If three or more beacons are visible at the same time, a device using an imaging sensor can reach 6-dof localization by solving the perspective-n-point problem. When using depth sensors, the relative 3D position between camera and each beacon is directly available. This enhances the positioning robustness, since both trilateration and triangulation can be used to determine the position.

When no absolute localization within a predetermined coordinate system is desired, it is sufficient to use beacons which initially do not hold information about their location. A device can determine the relative positions of the beacons by using a 3D camera. With optical communication, it is possible to assign IDs to each of them and re-localize them in the local coordinate system. It is also possible to forward localization information to these beacons, so that simpler devices with 2D cameras can later on use them for navigation. In the last section of this chapter, the idea of forwarding positional information to small IoT devices is further discussed.

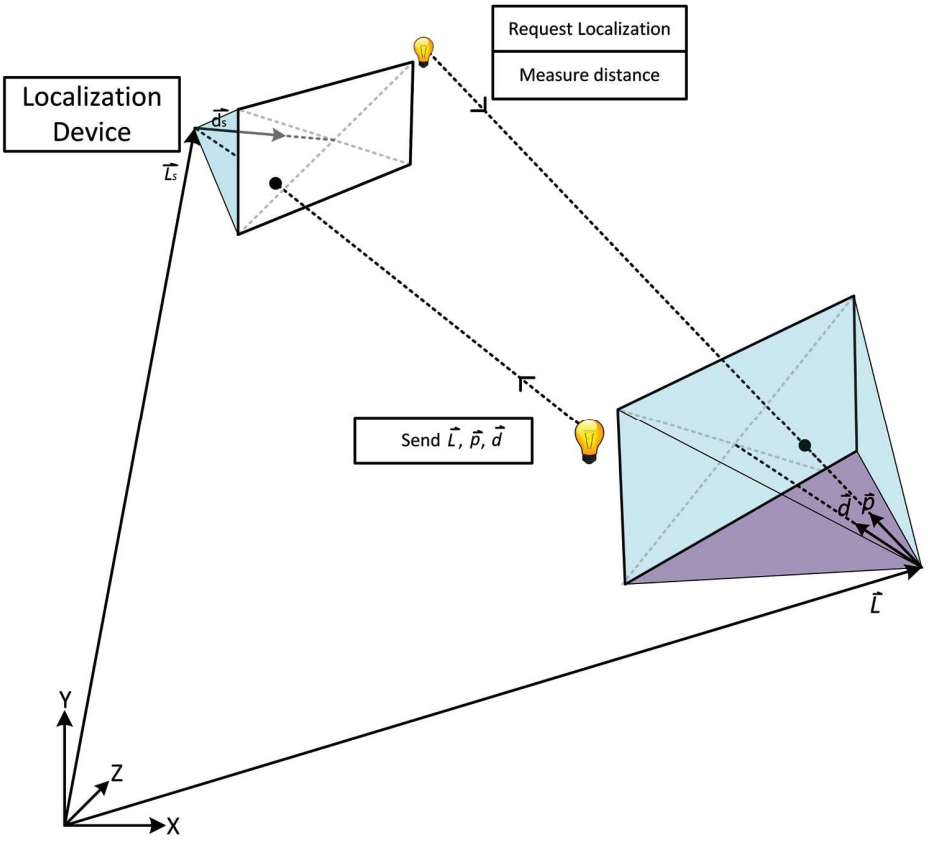
Augmented Internet of Things

So far, the main focus of augmented reality (AR) is fusing virtual and real worlds in order to receive an augmented world. The main mechanisms so far are putting virtual objects into a live camera stream (e.g. Google Tango, Qualcomm Vuforia), or embedding them into the viewport of a user (e.g. Microsoft HoloLens). We propose a concept with our OpticSec3D approach, to use 3D location-aware optical communication in order to enable embedded devices to interact with the augmented world. This would enable electronic devices within the viewport of an AR, to transfer information and interaction possibilities to the augmented world. These devices could be any kind of system, requiring human interaction, such as light switches, payment terminals, heating and climate control. They could also serve as virtual signs, or display promotions in supermarkets, or be used for pairing with local WIFI or Bluetooth connections.

A Time-of-Flight depth sensor on such AR system can receive optical signals from such devices and can use distance measurements to determine the 3D position. When the 3D position relative to the AR device is known, it can be embedded into the augmented reality. The concept, of mapping a 3D posi-

Localization and Context Determination for Cyber-Physical Systems Based on 3D Imaging

Figure 4. Depth sensing, combined with optical communication leads to 5-dof localization with just one base-station. The base-station is equipped with an image sensor, and transmits its position $\vec{L}$ *and the direction* $\vec{p}$ *vector of the localization device. The localization device can calculate its own position* $\overrightarrow{L_s}$ *and direction* $\overrightarrow{d_s}$ *by combining the received information with a distance measurement*

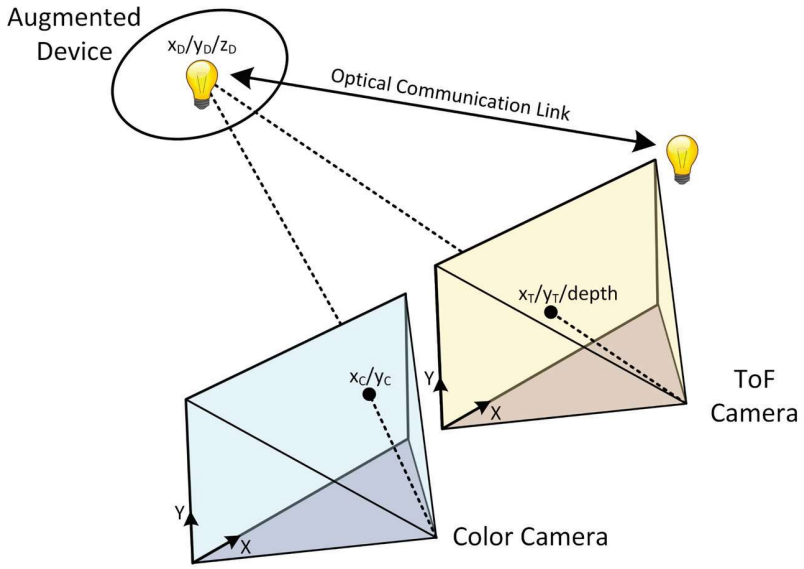


tion into AR space, is shown in Figure 5. Since the pixel position of the depth camera is different to the augmented color image stream, the 3D position needs to be transformed to color camera image space and projected to the 2D image space.

When the 3D position of a stationary device is determined, there are two possibilities. On the one hand, AR device can be equipped with a variety of pose tracking sensors. In this case, the 3D position can be placed into the augmented coordinate system and the pose tracking sensor would keep track of the position at this point. If the AR system does not feature such additional tracking system, the position solely relies on tracking the modulated light source. This requires continuous tracking, which means, that depth measurements need to be incorporated into the communication protocol.

It is however sufficient to conduct these measurements infrequently, since only larger changes in distance create different pixel mappings to the augmented reality stream. For rotations, it is possible to track the incident angle of the modulated light sources during communication with low latency.

Figure 5. With 3D-location-aware optical communication, electronic devices can be embedded into augmented reality. A Time-of-Flight sensor is able to combine depth measurements with pixel coordinates (x_T, y_T) to calculate the 3D coordinates (x_D, y_D, z_D) of the augmented device. Using calibrated camera parameters, it is possible to transform and project the 3D position to the image coordinates (x_C y_C) of an augmented image stream



CASE STUDIES ON OPTICAL LOCALIZATION AND TRACKING SYSTEMS

As of 2016, the upcoming applications for high-performance localization and tracking solutions are found in the automotive field in automated driving systems, and in consumer electronics in augmented and virtual reality devices. In this section, we introduce such high-performance localization and tracking mechanisms and how they already found their way into products.

Google Tango

Google started Tango with the goal to provide smartphones and tablets a human-like understanding of the environment. The central ambition is to produce a reference design and software to enable smartphone and tablet vendors to implement Tango's functionality into their devices. A large part of Tango is to support application developers by providing support, tools and development kits. A Tango enabled mobile device is able to record and locate itself in a 3D map of its environment and track its pose at a very high update rate. This enables indoor localization and augmented reality. Google categorizes the aspects of its platform into motion tracking, area learning and depth perception. These main features are mutually dependent from each other and rely on a sophisticated software and hardware implementation.

Motion Tracking

Project Tango devices feature 6-dof motion tracking. While current mobile devices are often capable of orientation tracking, by using gyroscopic magnetic and gravity sensors, Tango adds sophisticated 3D pose tracking. This is realized with a sensor fusion approach, including inertial sensors and a motion tracking camera. The motion tracking camera has a very wide field of view and records black and white images. Tango detects features and uses frame-to-frame feature correspondences to detect camera movements. The data from the inertial sensors complements the data of the motion tracking camera to receive a higher update rate and to increase robustness during strong motions and featureless images. The outcome from motion tracking is continuously available pose information, containing the position and orientation of the device, but neither provides a reference coordinate system or localization in a larger context.

Area Learning

The goal of area learning is to build a 3D model of the environment. This is accomplished by a simultaneous mapping and localization approach. Data from the color, depth and motion tracking sensors are combined to create a 3D presentation of the environment. Image features which are especially unique are saved as landmarks. According to Google, these landmarks are saved about every 50 cm of tracked camera displacement. Landmarks along with the available 3D data enable fast localization within previously recorded datasets. The captured area learning data can be exported and stored into area description files (ADF).

These files enable Tango devices to remember the environment without re-scanning. It is also possible to use externally provided area description files for localization. An example would be indoor navigation system, where the building was initially recorded with Tango. Waypoints for navigation could be placed into such an area description file to directly support navigation.

The area description file along with navigation information is then provided to Google Tango enabled devices, and they are able to precisely locate themselves in the coordinate system of the provided file.

Navigation and location based services are possible, if the ADF is associated to mapping data.

Jeon et al. (2016) enhance this concept, by scanning the indoor environment with a 3D laserscanner, and use their own method to build a database. Google Tango enabled devices can use the database for localization, but also to push new data to the database. This enables to update the database automatically with change to the environments.

Depth Understanding

The latest iteration of Google Tango incorporates a depth sensor based on Time-of-Flight technology. This sensor is important for area learning, but also gives the Tango platform an immediate understanding of the geometric context. Since Time-of-Flight is a dense depth sensing method, it is possible to gather 3D information on texture-less flat surfaces. This helps to create a denser map during area learning, but also enables the augmented reality applications which are one of the most significant selling points of the platform. With depth understanding, it is possible to integrate virtual objects into the augmented world. Besides gaming and other entertainment purposes, depth sensing enables to measure distances without measuring tape, and even preview furniture in its determined surrounding before buying.

Head-Mounted Virtual Reality Devices

Virtual reality is currently the most demanding application for indoor positioning and tracking systems. The demand for operating range of consumer-grade head mounted devices (HMD) is limited, but the precision, update rate and latency are critical. User experience motion sickness, if there is a discrepancy between their motion and visual stimuli. A common term is the motion to photon delay, which is the time from a user's movement to the point, when the displays in the HMD update their images to accommodate the movement.

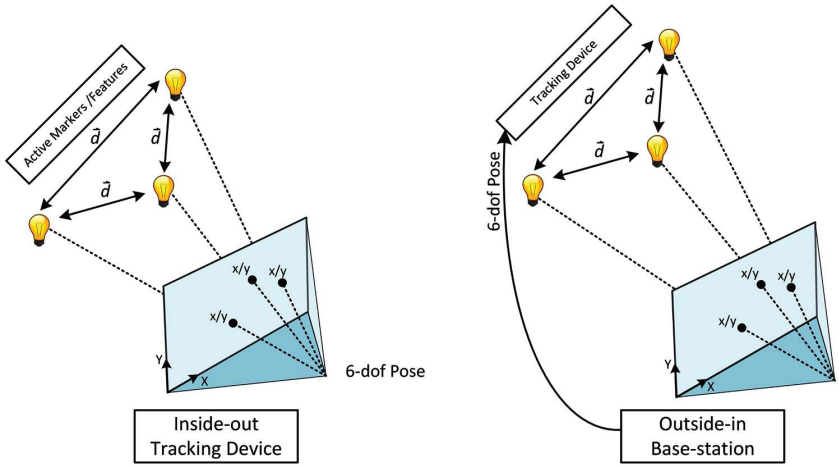
The main measure to reduce this delay is a high display frame-rate. This is currently 90 frames per second in the current top products such as the Oculus Rift and the HTC Vive. On the other hand, the tracking sensors require an even higher update rate. Since visual output and tracking are usually not synchronized, a higher tracking update rate enables to associate frames with better tracking results.

The common denominator of current VR tracking systems is that they are based on sensor fusion approaches, using MEMS-based inertial sensors for a high update rates in combination with optical sensors for positioning and error correction.

For the optical tracking part, there exist two major paradigms, as illustrated in Figure 6. While the exact definition is debated, the main difference is the position of the optical localization device (Foxlin, 2002). An inside-out tracking system has the localization device mounted on the device, while outside-in systems have a stationary tracking device, facing the tracked object. Fixed reference points are used to localize the device within a given coordinate system. The fixed locations can be either active beacons, or passive visual markers. Even normal visual features might be sufficient in future iterations of HMD tracking systems. In 2016, Oculus invited journalists to demonstrate a prototype, based on Inside-out tracking and just using visual features of a typical living room environment (Orland, 2016).

The advantage of inside-out tracking is that the system is not limited to certain boundaries, as long as enough reference points are within the field of view of the tracking sensors. A problem of practical

Figure 6. The difference between outside-in and inside-out tracking. 2D-projections (x, y) of features or active markers with known relations $\vec{d}$, are used to determine the 6-dof pose of the tracking device. In inside-out tracking, the tracking device observes the static environment. In outside-in tracking systems a stationary base-station observes the device and forwards the pose over a communication link



implementations however is that usually with 2D image sensors, only incident angles of the line-of-sight to the reference points can be measured. Since it is not feasible in consumer electronic, to let the user calibrate the positions of the reference points, the points need to be located. This can be achieved by image to image correspondences, but requires view disparity and is subject to inaccuracy. Another reason, why as of 2016, inside-out tracking is not used in virtual reality is that the pose tracking precision of active visual markers depends on the distribution of the markers. At least three markers are required to be in sight at the same time, and they must not be arranged in a straight line.

An outside-in tracking system has one or more localization sensors mounted on a fixed point, while the tracked objects are observed. This currently dominant tracking paradigm can be categorized into the following principles:

Tracking With LED Markers and Image Sensors

This technology is currently used by Oculus Rift, OSVR and PlayStation VR. These outside-in tracking systems are based on 2D infrared cameras, which sense active markers on the HMD. It case of the Oculus Rift, these markers consist of LED lights, which are toggled by the tracking system for identification. Since the positions of these markers on the HMD are fixed and known, finding the pose of the head-mounted device can be mapped to the perspective-n-point problem. This problem describes finding the 6-dof pose of a set of points in 3D space, by analyzing 2D projections on images, gathered by a calibrated camera. While the pose can be determined by sensing three or more points, the robustness and precision are enhanced by additional points. The problem with outside-in tracking is the stationary position of the camera. This limits the usage to viewport of the tracking camera. Due to limited sensor resolution, the precision decreases with distance limiting the tracking solution to medium sized rooms. This can however be counteracted by using multiple tracking cameras.

Tracking With Laser Beacons and Photodiodes

This tracking solution is used by HTC Vive HMD in its Lighthouse tracking system. It is based on photosensitive element instead of image sensors. These elements are distributed among the surface of the device. Two base stations are positioned in front of the user. Each feature two rotating lasers, which scan the X and Y axis of the room at precisely 60 Hz. A synchronization pulse sequence is broadcasted to all sensors. Then each element counts the time until the laser beam reaches its position. With the timing information, it is possible to reconstruct the angle of the laser, when it reached the element. By using two lasers, it is possible to triangulate the position of the element. Using the 3D positions of all elements, it is possible to determine the 6-dof pose of the HMD. The advantages of this system, are that it has an increased range, minimum computation overhead, potentially lower latency, and the possibility to track multiple devices without additional effort in the base stations (Deyle, 2016). The update-rate however is too low for optical tracking alone, and thus sensor fusion with inertial sensors is necessary. Another advantage is that the base-station is independent of the tracked device. The positional information could be calculated on the tracked device itself.

The scanning principle of Lighthouse also permits 3rd parties to use the base stations to develop their own tracking solutions. Since inexpensive microcontrollers and photodiodes can be used, this tracking system has the potential to be widely used for localization and positioning in areas like internet of things or smart-homes. State-of-the-art HMDs are dedicated to electronic entertainment. The devices

are localized by the tracking system into a coordinate system, defined by the base stations. This pose is then transformed into the coordinate system of the virtual world. If the pose of the base station within a mapping system of the real world is known, virtual reality tracking systems can provide high fidelity localization.

Another aspect is awareness of the immediate surroundings. Virtual worlds usually do not have boundaries, so the users need to be prevented from stumbling over objects or colliding with walls. An elegant solution would be to detect these obstacles by using a depth camera. The first generation of consumer grade HMDs does not feature such a camera, but let the user manually define virtual boundaries before usage.

VEHICLE LOCALIZATION

This section briefly introduces some special use-cases where different forms of localization and tracking are employed in the very important vehicular context.

Fleet Management

A company owing a number of cars (for example a taxi company or a truck company) can strongly benefit from knowing the current position of all their cars in order to improve the offered service. Since such systems do not require a very precisely measured location they get along with a GPS device. Those systems are inexpensive and are currently widely used for fleet management, since they can be built up using a mobile phone with GPS capability. A famous example is the taxi company Uber, which displays the current position of the closest available vehicles via their smartphone application. Additionally, the GPS in the car replaces the taximeter, since it tracks the actually driven distance and time.

Assisted Driving Functions

Some assisted driving functions require environment data of the direct surroundings, for example the detection of other cars and objects for parking assistance or collision avoidance (Winner, 2015). For that purpose, sensors for object detection are mounted on modern vehicles. Those include radar sensors, LiDAR sensors, ultrasonic sensors and cameras. Since each of those sensors has shortcomings in certain environments, typically more than one sensor technology is implemented and considered by the assisted driving system.

Other assisted driving functions require the vehicle to not only know its direct surroundings, but also know the current position on the map. It can be very beneficial for a collision avoidance system to have information about the current context (for example an urban area or a motorway) (Levinson, 2008). This can easily be acquired using GPS localization.

Then there are also systems which communicate with infrastructure objects (Vehicle-to-infrastructure, V2I) or other cars (Vehicle-to-vehicle, V2V) in order to gain additional knowledge of the surrounding environment. This can be information about dangers on the upcoming road, the timing of the next traffic lights or other relevant information for the vehicle. Today, the information is communicated using the mobile network or a Wi-Fi-like short range network. Yet, the presented novel optical line-of-sight communication techniques may be used in future.

Automated Driving

Although fully autonomous driving is not reality yet, automakers are putting a lot of effort in that topic in order to preserve their position in the market. Localization of a vehicle using only the GPS is considered to be insufficient for the use in (partwise) automated driving systems (Bar, 2014). The current approaches depend on up-to-date HD maps and the exact position of the vehicle within that map. The exact position in addition with the data from multiple environmental perception sensors enables the potential for fully autonomous driving.

The solely use of current GPS systems for localization is not precise and reliable enough for direct mapping into HD maps. Therefore, the issue can be resolved by using the environmental perception sensors (radar, LiDAR, cameras) of the vehicle. For instance, Time-of-Flight based sensors can measure the distance to certain points of interest in the environment and align them with the corresponding points within the HD maps. Using the inertial sensors and odometric data, it is possible to keep tracking the exact position on the map even if the GPS is in a non-functional state (for example in a tunnel).

FUTURE CONCEPTS

Inside-Out Tracking Without Active Markers

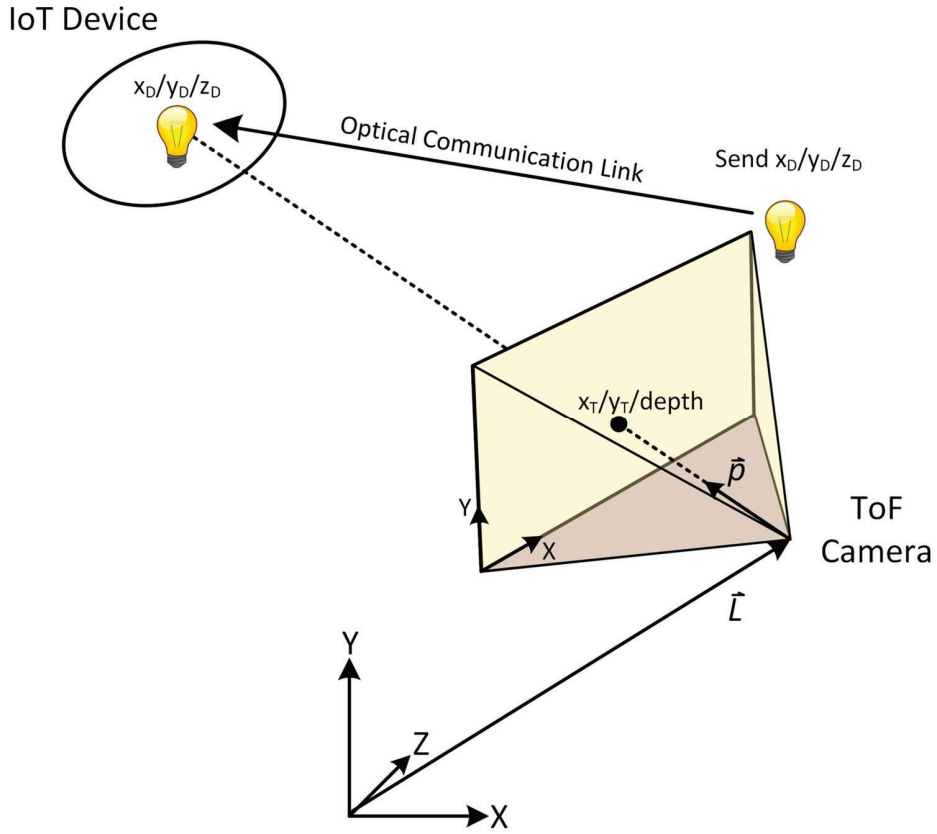
Vision based inside-out tracking has is already extensively been used in augmented reality. Google Tango, Microsoft HoloLens and Qualcomm Vuforia are well known examples. In these applications, the required latency and update rate allows feature based tracking and localization. In augmented reality, latency is more tolerable, than in virtual reality, because it does not cause motion sickness. The 2D imaging pipeline of an inside-out tracking system introduces such latency which creates a massive technical hurdle in high fidelity tracking and localization applications such as virtual reality. Despite these circumstances, Oculus presented a working prototype in 2016, using four motion capture cameras in order to provide 6-dof tracking without requiring any peripherals.

Position Forwarding to IoT devices

With an increasing number of 6-dof location-aware devices on the consumer market, it is possible to use image sensors along with optical communication to detect and localize embedded devices of all kind. An example could be a smartphone equipped with a location-aware AR platform such as Google Tango and an optical communication solution. By sensing the presence of embedded IoT devices, it would be possible to localize them and use a communication channel to directly forward them their position. Potential applications involve secure device pairing, location-aware temperature control, geometry-aware audio systems or all kinds of intelligent sensors. If desired, surveillance cameras could be provided their pose in order to be able to reliably track persons throughout buildings.

In professional settings such as fabrication, workshops or warehouses, the location of tools and parts can be determined by using optical communication along with positioning.

Figure 7. A device is aware of its own location $\vec{L}$ and orientation $\vec{p}$ is able to localize embedded IoT devices in 3D. It can forward the localization information via optical communication.



CONCLUSION

This chapter discusses concepts, methods and opportunities for optical localization of cyber-physical systems and future IoT devices. A focus is placed on Time-of-Flight depth imaging systems, and how they can benefit existing and future localization systems. Examples of high-performance localization and tracking systems in existing products are introduced in form of case-studies.

As the field of visible light communication shows, optical communication is a crucial aspect of modern optical localization systems. We introduce our own location-aware optical communication approach, based on Time-of-Flight depth sensors. With this approach, it is going to be possible to localize communication partners in 3D with high accuracy and very low latency. We are also confident that 3D-location aware optical communication will close the gap between augmented reality and IoT, by enabling embedded devices to participate in an augmented world.

Synchronization of Time-of-Flight 3D Sensors for Optical Communication

Hannes Plank*, Armin Schoenlieb, *, Christoph Ehrenhoefer, * Christian Steger, †, Gerald Holweg*, Norbert Druml*
*Infineon Technologies Austria AG, Graz, Austria
{hannes.plank, arminjosef.schoenlieb-ee, christoph.ehrenhoefer-ee, gerald.holweg, norbert.druml}@infineon.com
†Graz University of Technology, Austria, steger@tugraz.at

***Abstract*—Time-of-Flight 3D imaging systems are promising transceivers for image sensor based optical communication. 3D sensing based on Time-of-Flight is the most miniaturized depth imaging technology available and is currently being integrated into consumer electronics such as smart phones. Optical line-of-sight communication with a depth imaging system offers complete location-awareness of the communication partner. This enables new opportunities in fields like secure authentication, augmented reality or vehicle to vehicle communication. We show how Time-of-Flight systems are capable of using phase shift keying of pulsed light (PLPSK) to transmit data over a line-of-sight connection. PLPSK is still unexplored in the domain of image sensors and enables Time-of-Flight sensors to transmit multiple bits per frame at rates of over 7 kHz. A serious problem however are asynchronous modulation signals, causing frequent transmission errors and impairing proper communication. In this work, we formulate and discuss the severity of the problem and propose a synchronization procedure. We evaluate our solution with a prototype system, and show that it is possible to reach a synchronization success rate of nearly 100% over a distance of 9 meter.**

I. Introduction

Optical line-of-sight communication based on image sensors is a viable alternative to radio based communication in certain domains. Image sensors proof to be capable optical receivers [1] and offer source separation due to projection of incoming light onto the sensor. The most significant distinctions however is the ability to capture images of the communication partner, and to determine the precise direction of the signal. This directional awareness opens a manifold of opportunities for certain applications.

In our work, we want to go further and provide complete location-awareness thus including the distance between communication partners. We therefore propose Time-of-Flight (ToF) depth sensors for optical communication. Depth imaging based on ToF is the most miniaturized depth imaging system available and is becoming available integrated in consumer devices such as smart-phones or virtual reality head-mounted-devices. ToF cameras are capable of capturing depth images for a range up to 15 meter, and as we show in this work are capable of transmitting information to each other. Such complete location-aware communication is highly desired in domains such as relay attack resilient device authentication, vehicle to vehicle communication or augmented reality.

In this work, we first introduce the concept of ToF 3D sensors and discuss their suitability for optical communication. We then stress the necessity of modulation frequency synchronization, and provide a solution which is supported by a prototype system and evaluation.

A. Time-of-Flight Depth Sensors

ToF cameras produce depth images, where each pixel encodes the distance between the camera and the sensed scene. This works as shown in Fig. 1 by emitting infrared light and measuring the time it takes to travel to the scene and back to the sensor. In the continuous wave approach [2], pulsed light $\mathcal{P}_{Light}$ is emitted by the illumination unit during the capture process. The pixels of the sensor are supplied with a square wave signal $\mathcal{F}_{mod}$, which has the same frequency as the emitted light pulsed. The travel time of the light pulses cause a phase-shift φ between the received and emitted signals. A Photonic Mixer Device (PMD) [3] is located at each pixel and produces a value proportional to the phase shift φ. The phase of the pulses is shifted for each image to compensate for light intensity variations. Multiple images are then processed to one depth image. Due to this measurement principle, ToF sensors are capable of emitting phase shifted pulsed light, and the pixels are capable to precisely measure these phase differences.

II. Time-of-flight Depth Sensing Systems as Optical Communication Transceivers

Optical communication with Time-of-Flight imaging sensors is rather unexplored. A first attempt was accomplished by Yuan et al., who establish a one-way communication link between a ToF camera and an array of LEDs [4]. In their

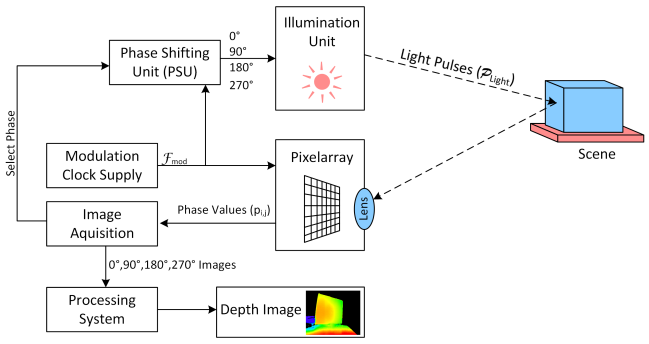


Fig. 1. The principle of Time-of-Flight depth sensing. Light pulses are sent out and reflected by the scene. The pixel measure the phase differences of which is proportional to the depth.

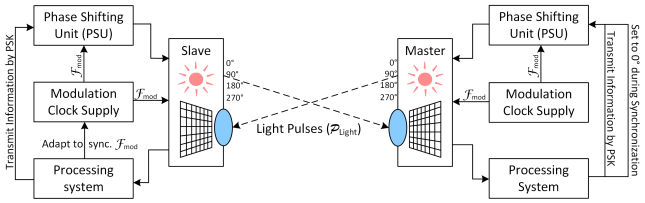


Fig. 2. The principle of optical communication based on Time-of-Flight. Two ToF cameras communicate by phase shift keying. They using their built-in phase shifting unit for modulation. The phase shift is demodulated at each pixel. The slave system synchronizes its modulation signal F_{mod} to the master's.

approach, LEDs are modulated via PSK to manipulate the depth measurements of a Time-of-Flight camera in normal depth sensing operation. They use a photodiode to recover the modulation signal of a Time-of-Flight camera and thus avoiding the issue of synchronization.

In our approach, we want establish a communication link between two Time-of-Flight cameras without requiring additional sensors or multiple transmitters to boost bandwidth. As shown in Fig. 2, it is possible to use the phase shift unit to modulate the outgoing pulsed light with phase shift keying (PLPSK) to transmit information. The phase shifts are directly demodulated on each pixel by a Photonic Mixer Device (PMD). As illustrated in Fig. 3, when light arrives onto the sensor, it creates charges in the silicon. There are two areas per pixel where these charges are stored, until they are read and digitized. The PMD acts as a switch, deciding into which of these buckets the charges are transferred.

A hardware unit inside the Time-of-Flight sensor generates a square wave signal $\mathcal{F}_{mod}$. This signal is supplied to the phase-shifting-unit (PSU), which shifts and forwards the signal to the illumination unit emitting the pulsed light signal $\mathcal{P}_{Light}$. The original modulation signal $\mathcal{F}_{mod}$ is also supplied to each pixel where it triggers the PMD. If the logic level of $\mathcal{F}_{mod}$ is low, incoming charges are collected in bucket A and otherwise in bucket B. After a certain exposure time t_{expo}, the sensor reads and digitizes the contents of all buckets. The final product of each pixel are the phase values $p_{i,j}$, which are the differences between the voltages of $A_{i,j}$ and $B_{i,j}$. The phase values $p_{i,j}$ are proportional to the phase shift between the modulation signal $\mathcal{F}_{mod}$ and the received light pulses $\mathcal{P}_{Light}$. Continuous light sources do not influence $p_{i,j}$, because their generated charges are distributed evenly among A and B. Consequently no complex image processing is necessary to determine which pixel receive the modulated light signal.

The receiver can directly sample these unprocessed phase values $p_{i,j}$ and directly extract the transmitted information. Our evaluation shows, that we can reach a sampling rate of over 7300 frames per second with our prototype system.

III. The Problem with Asynchronous Modulation Signals

Synchronization is generally an important aspect in free-space optical communication and the procedure depends on the modulation method. Approaches like Schmidt [5] or Hu et al. [6] use software for synchronization. In this Section, we

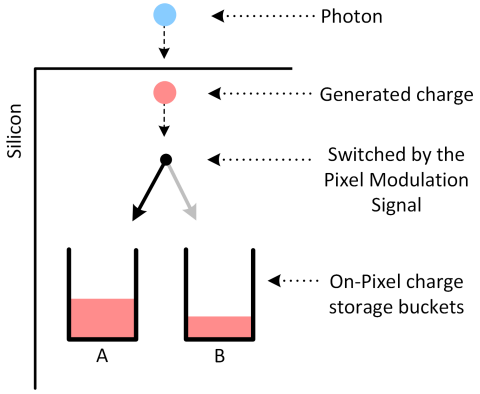


Fig. 3. Illustration of the operating principle of the Photonic Mixer Device which is located on each pixel of a ToF sensor. Charges are transferred into either charge bucket A or B, depending on the modulation signal $\mathcal{F}_{mod}$.

discuss how asynchronous modulation signals affect the phase value measurement, and why it is necessary to physically adapt the modulation signal $\mathcal{F}_{mod}$.

In depth sensing, the reflected light pulses P_{Light} and $\mathcal{F}_{mod}$ at pixels have the same frequency as they originate from the same signal source. When Time-of-Flight systems however communicate, they are supplied with signals from different sources. There is usually a frequency offset ΔF between these modulation signals due to frequency imprecision of the oscillators. Fig. 5 shows the operating principle of a PMD pixel with no frequency offset. When light pulses arrive at the pixels, the generated charges are sorted into either of the charge storage buckets A and B, depending on the modulation signal $\mathcal{F}_{mod}$. In this paper, we define the phase offset φ between P_{Light} and $\mathcal{F}_{mod}$ to be zero, when the pixel modulation signal $\mathcal{F}_{mod}$ raises from high to low in the middle of a light pulse, evenly distributing the charges to A and B and producing a phase value $p_{i,j} = 0$. The ratio of charges which are stored in A and B does thus not change, if the frequency offset ΔF is zero.

If there is however a frequency offset ΔF between the pixel modulation signal $\mathcal{F}_{mod}$ and the received light pulses $\mathcal{P}_{Light}$, the phase offset φ between these signals changes over time. As shown in Fig. 4, after a certain time, the initial phase offset φ_0 has changed to $\varphi_0 + \varphi_d$. If we interpret the phase offset as angle between 0° and 360°, the time dependence can be

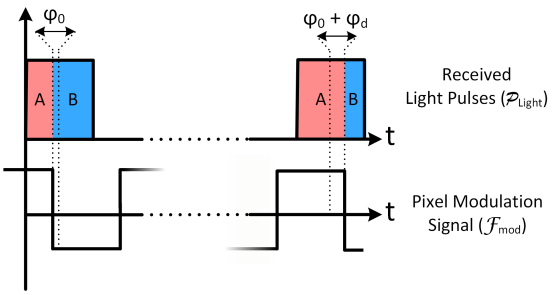


Fig. 4. A Photonic Mixter Device receives a pulsed signal. The frequency of the received pulses differs from the pixel modulation signal, causing the sorting ratio between A and B to change.

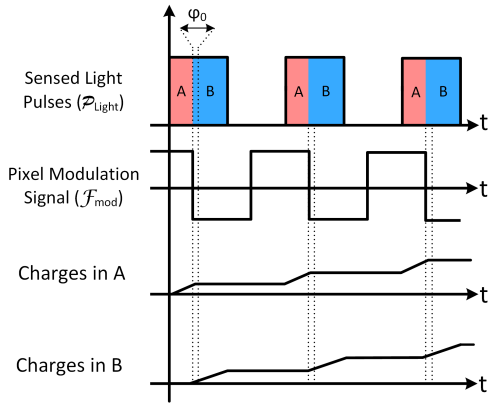


Fig. 5. Synchronous operation of a Photonic Mixer Device. Charges from light pulses are stored into either bucket A and B. The sorting ratio of A and B does not change over time.

expressed as:

$$\varphi(t) = [(\varphi_0 + t \cdot \Delta F) \mod 1] \cdot 360° \quad (1)$$

This phase drift has a serious impact on the phase shift sensing capabilities of the sensor. If these signals were synchronous, a ToF sensor would measure constant phase values $p_{i,j}$ for light signals $\mathcal{P}_{Light}$, assuming the intensity and exposure time do not change. Phase shifting the light pulses for communication would always yield to the same changes of $p_{i,j}$ and proper decoding.

With a frequency offset $|\Delta F| > 0$ however, the measured phase values change over time. Fig. 6 visualizes this time-dependency by showing a sampled phase value $p_{i,j}$ over time. The incoming pulsed light signal $\mathcal{P}_{Light}$ and the pixel modulation signal $\mathcal{F}_{mod}$ do not change over time. Their frequencies differ however by the offset ΔF. Within the time period T_D, the phase offset $\varphi(t)$ changes from 0° to 360°, creating a periodic triangular signal. This phase offset signal $\mathcal{O}$, is created by the changing content of charge bucket A and B.

The frequency of the phase offset signal $\mathcal{O}$ is identical with the absolute value of the frequency offset $|\Delta F|$. We exploit this property in our approach to extract the frequency offset ΔF and to physically adapt the source of the modulation signal $\mathcal{F}_{mod}$. The signal can be reproduced experimentally, when a receiver samples the average of phase values $p_{i,j}$ of

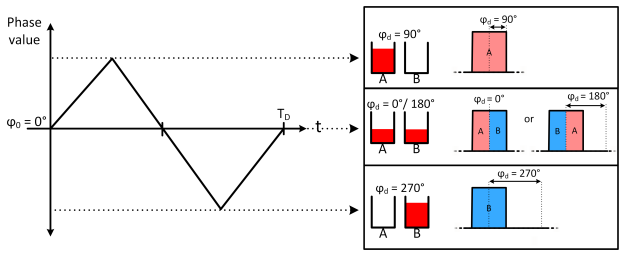


Fig. 6. The difference between the content of A and B determines the measured phase value. If the modulation clock of sender and receiver are not synchronized, the measured phase values change over time. We use the frequency of this signal to synchronize our system.

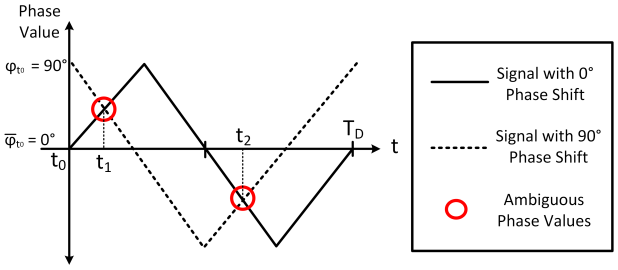


Fig. 7. The sender sends phase shifted signals to transmit information. At t_1 and t_2, the receiver can not differentiate between the signals, due to asynchronous modulation clocks.

the received illumination signal $\mathcal{P}_{Light}$. The sender camera in this setup continuously emits light pulses without phase shift keying. As shown in Section V, we use such setup in our prototype to estimate and compensate the frequency offset ΔF.

A. The Effect of Asynchronous Modulation Signals on Communication

This Section discloses why a frequency offset ΔF between the modulation signals of the sender and receiver poses such a serious problem for communication. The problem is not the impact on the measured phase values $p_{i,j}$. This could be compensated by predicting phase values via software or by employing differential coding. The real issue are ambiguous phase measurements, where the receiver can not distinguish between differently phase shifted light pulses anymore.

Fig. 7 illustrates this problem by using 2-PLPSK as example. The sender in this example encodes information by either transmitting pulsed light with 0° or 90° phase shift. In the Figure, the received values for both signal are shown. Due to $|\Delta F| > 0$, the received phase values change over time. At t_0 the receiver can clearly distinguish between 0° and 90°. However at point t_1 and t_2, the receiver is not be able to measure distinct phase values.

This ambiguity of the received phase values severely impairs efficient communication. The rate of ambiguous phase occurrence R_a depends on the frequency offset ΔF and on how many different phases N_P are used for modulation. For 2-PLPSK modulation, as illustrated in Fig 7, there are two ambiguous phases per period. Increasing the number of modulation phases N_P introduces two new ambiguous phases for each existing phase within T_D, leading to the following generalization:

$$R_a = N_P^2 \cdot (1 - \frac{1}{N_P}) \cdot \Delta F \quad (2)$$

This relation however does not take into account, that ambiguous phases often occur at the same time. For example in 4-PLPSK, there are time points, where the sender is unable to distinguish phase values of 0° and 180° but also 90°and 270°. This leads to the ambiguous symbol rate R_s, which can be expressed as:

$$R_s = N_P \cdot 2 \cdot \Delta F \quad (3)$$

Our synchronization approach aims to limit R_a to an acceptable minimum by reducing the frequency offset ΔF.

When the modulation phases are equidistant, the time intervals T_s between ambiguous symbols only depend on ΔF and N_P:

$$T_s = \frac{1}{N_P \cdot 2 \cdot \Delta F} \tag{4}$$

This enables uninterrupted communication for save time periods T_s. The length of these time periods stays constant, provided ΔF does not change. As we show in Section IV-B, we can adjust our synchronization procedure to reach a desired T_s with the tradeoff of prolonged synchronization time.

IV. Our Synchronization Approach

We propose a method to synchronization Time-of-Flight modulation signals which is based on determining the frequency offset ΔF, and adapting the modulation signal source for compensation. Time-of-Flight cameras encounter a variety of different application scenarios and therefore the modulation frequency of the sensor is usually configurable. In our approach, the frequency adaption is carried out by a slave device, which adapts its modulation frequency to its communication partner (master).

A. Capturing the Phase Offset Signal $\mathcal{O}$

The first step is to reliably sense the phase offset signal $\mathcal{O}$. This works by the master device emitting light pulses $\mathcal{P}_{Light}$, and the slave sampling the phase values of the receiving pixels. In most image sensors, all pixels are read and digitized per frame. This is however time-consuming and limits the bandwidth of optical communication.

We therefore restrict the readout process to a small subimage of n times m pixel, containing the received signal $\mathcal{P}_{Light}$. Each pixel of this subimage produces a phase value $p_{i,j}$, which is the difference of the digitized voltages of charge bucket A and B. The location of the subimage (u/v) is frequently adjusted to the position of the lightsource, actively tracking the communication partner.

A sample p of the phase offset signal $\mathcal{O}$, is produced by averaging N phase values $p_{i,j}$ containing a signal. A threshold Z is experimentally determined and used to classify these pixels from the ones containing just background noise.

$$p = \frac{1}{N} \cdot \sum_{i=u}^{u+n} \sum_{j=v}^{v+m} \begin{cases} p_{i,j}, & \text{if } |p_{i,j}| \geq Z \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

As the evaluation shows, averaging uncalibrated phase values is sufficient to sample $\mathcal{O}$ for frequency estimation and consequently synchronization.

B. Frequency Estimation of the Phase Offset Signal $\mathcal{O}$

The modulation signals are synchronized by estimating the frequency offset ΔF, and adjusting the frequency generation circuitry of F_{mod} of the Time-of-Flight sensor. As mentioned in Section III, the frequency of the phase offset signal $\mathcal{O}$ is identical to the absolute value of the frequency offset $|\Delta F|$. Therefore ΔF is estimated by capturing and estimating the frequency of $\mathcal{O}$.

This frequency estimation needs to be carried out several times during the synchronization process and thus has a huge influence on the optical communication overhead. In our prototype, we chose a fast but reliable way for frequency estimation.

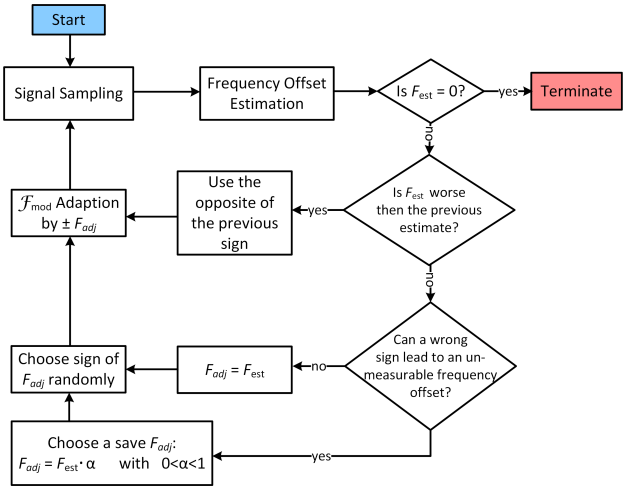


Fig. 8. Flow chart of our proposed synchronization algorithm. Signal $\mathcal{O}$ is sampled, its frequency estimated and the modulation signal $\mathcal{F}_{mod}$ is adapted by F_{est}.

This estimation algorithm is based on parsing the periodic signal $\mathcal{O}$ for zero-crossings and peaks. Samples are classified as peaks if they are the maximum value within a certain neighborhood. Noise can lead to the detection of multiple peaks and zero-crossings per actual feature. A median filter is applied to clusters of peaks and zero-crossings to determine the actual location of the feature. It is then validated, if the detected zero-crossings and peaks form a valid periodic signal. This is accomplished by checking if subsequent zero-crossings have one peak in the middle, producing a set of detected half periods. Finally, the estimated frequency of $\mathcal{O}$ is derived from the median of detected period lengths of the validated half periods.

C. Synchronization Algorithm

The estimated frequency offset F_{est} can be used to adapt the modulation PLL of the slave device, synchronizing both systems. The frequency estimation however may not be accurate enough to compensate the frequency offset ΔF in single step. We propose a multi-step process, where F_{mod} is physically adapted by F_{adj} in each step. Since more samples per period lead to a preciser estimate, ΔF converges to the final remaining synchronization error ΔF_{sync}.

Our proposed synchronization algorithm also needs to handle the sign uncertainty of F_{est}. Just by estimating the frequency of the phase offset signal $\mathcal{O}$, it is impossible to determine, if the estimated frequency offset F_{est} is positive or negative. We propose to randomly adapt the modulation frequency by either $F_{adj} = +F_{est}$ or $F_{adj} = -F_{est}$ and observe the impact. The new resulting frequency offset ΔF_{new} is either mitigated or amplified. In the later case, the randomly chosen sign of F_{est} was wrong and the system needs to compensate ΔF by using the opposite sign.

The multi-step approach of the frequency offset ΔF mitigation and the sign uncertainty lead to the development of

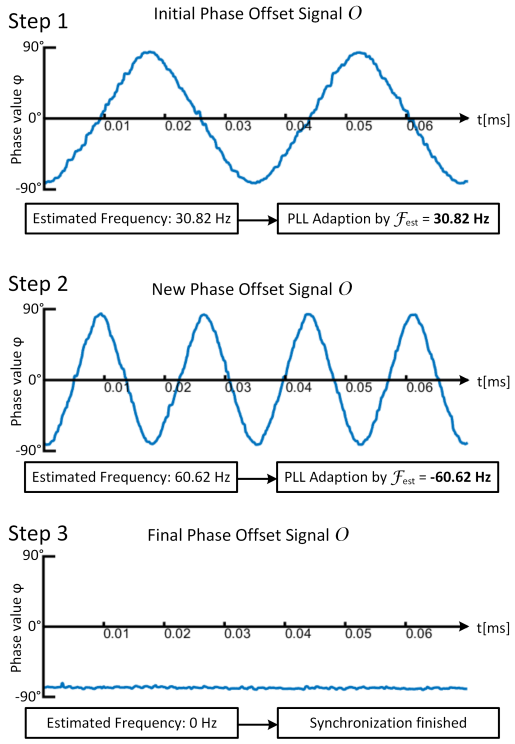


Fig. 9. Example for our synchronization procedure with data from our prototype. The frequency of the phase offset signal $\mathcal{O}$ is estimated and adapted in each step.

the proposed synchronization algorithm as shown in Fig. 8. Additionally, there is a robustness measure in place, preventing the system to adapt F_{mod} in a way so that the frequency offset ΔF gets too large to be measured. The algorithm calculates the maximum adaptable value F_{max} and reduces the adjustment by a factor $F_{adj} = F_{est} \cdot \alpha$ where $\alpha = \frac{F_{est}}{F_{max}}$.

Fig 9 showns an example of a synchronization procedure. It shows the recorded dataset of a successful calibration conducted by our prototype system. The system samples the phase offset signal $\mathcal{O}$ and estimates the frequency. It is uncertain if F_{est} is positive or negative and thus the system randomly chooses to decrease F_{mod} by $F_{adj} = F_{est}$. In our prototype this is accomplished, by sending a configuration command to the ToF sensor which adjusts the phase-locked-loop (PLL) of the modulation signal.

When $\mathcal{O}$ is sampled again in step 2, the system recognizes that the adjustment was wrong, since F_{est} is now larger than in the initial estimate. Because, modulation signal $\mathcal{F}_{mod}$ was wrongfully increased in the last step, it is now decreased by F_{est}. In step 3, the $\mathcal{F}_{mod}$ is already considered synchronized, as no periodic signal can be detected.

The final synchronization error ΔF_{final} depends on the last frequency estimation. The lowest frequency our estimation algorithm can detect has a period length equal to the sampling period. The following relation gives an upper boundary of the synchronization error ΔF_{final} depending on the sampling rate S and number of samples N_S.

$$\Delta F_{final} \leq \frac{S}{N_S} \tag{6}$$

Save time periods T_s without any ambiguous phases are desired for communication. Increasing the number of samples N_S reduces ΔF_{final} to an arbitrary level at the cost of synchronization time. This also allows to adjust T_s to a certain value. In Section IV-A the relation between the frequency offset ΔF_{final} and the save communication time period T_s is shown in Eq. 4. Using the relation in Eq. 6, it is possible to rearrange Eq. 4 to receive Eq. 7 which enables to determine the required amount of samples reach a certain save communication time period T_s.

$$N_S = 2 \cdot N_P \cdot T_s \cdot S \tag{7}$$

V. Evaluation and Discussion

In order to systematically evaluate the requirements and performance of our synchronization procedure, we implement a prototype system. We use two Infineon REAL3™ sensors based on Time-of-Flight technology of pmdtechnologies. Each ToF camera is connected to a processing system based on the Xilinx Zynq 7000 platform. Our system features an FPGA, two ARM Cortex A9 cores and 256 MB RAM. We receive the data from the Time-of-Flight camera via parallel interface, and use I^2C to configure the camera's parameters. Our synchronization algorithm is implemented as part of the processing software. The hardware implementation on the FPGA is responsible for the data transfer from the camera to the RAM.

In the following experiments, we want to investigate, how our synchronization approach performs under difficult conditions. We quantify the performance of our system by the synchronization failure rate. If ΔF is be reduced so that no frequency offset can be detected anymore, a synchronization is successful. A synchronization attempt fails, if the procedure aborts after 20 iterations with a measurable offset ΔF.

A. Initial Frequency Offset

The initial frequency offset ΔF, is a direct consequence of the frequency inaccuracies of the modulation signals between master and slave. In our system, the maximum tolerable value of ΔF is restricted by the quality of the frequency estimation and the sampling rate. Due to restricted sensor readout, we are able to sample the phase offset signal $\mathcal{O}$ at 7300 Hz. This restricts the theoretical maximum measurable frequency offset ΔF_{max} to 3650 Hz. For reference, our prototype system uses a HC-49S 9C Quarz oscillator, creating a 26 MHz modulation signal $\mathcal{F}_{mod}$. The frequency inaccuracy is rated at 10 ppm, which is 260 Hz for our 26 MHz modulation frequency. If both modulation signals show the worst specified frequency offset, our synchronization method needs to compensate a frequency offset ΔF of 520 Hz.

We test how our system handles different frequency offsets by setting it up with initial phase offsets ranging from 200 to 700 Hz. For each of these phase offsets, we attempt the synchronization procedure 100 times to determine the failure rate. This results are presented in Fig. 10.

As Fig. 10 shows, our synchronization procedure works up to a frequency offset of 600 Hz without any failure. At 700 Hz, our system did not accomplish any successful synchronization attempts. A larger frequency offset, requires the synchronization procedure to take more iterations to terminate. This

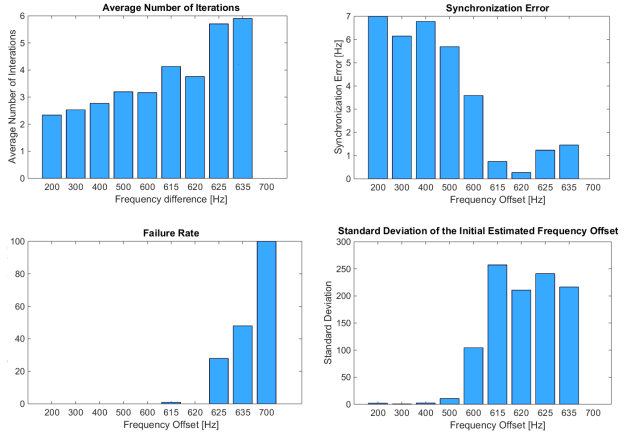


Fig. 10. Evaluation of the initial frequency offset ΔF. Our system handles frequency offsets of up to 600 Hz without failure. The results also show that our synchronization algorithm can compensate wrongful initial frequency estimates.

is caused by incorrect frequency estimations due to limited sampling points per period of phase offset signal $\mathcal{O}$.

Another aspect we want to investigate is the final synchronization error ΔF_{sync}. As discussed in Section IV-C, a remaining frequency offset ΔF_{sync} is tolerable, but reduces the communication bandwidth. In Fig. 10 we observe an inverse relation between the average number of iterations and the synchronization error.

The reason is the nature of our synchronization approach. We aim to synchronize the master and slave system as fast as possible. If feasible, the algorithm attempts to compensate the complete frequency offset ΔF at every step. The frequency estimation error however decreases with the estimated frequency, because more samples are available per period. The initial frequency offset estimation is often sufficient enough to terminate the frequency adaption algorithm, yet a certain tolerable offset ΔF_{sync} remains. When the synchronization is on the other hand approximated in multiple steps causing ΔF to incrementally decrease, the final synchronization error might be significantly lower. In conclusion we recommend a slower approximation approach, if a small final synchronization error is favored to a short synchronization period.

B. Distance

In this experiment we evaluate the range of our prototype system. We chose a realistic initial frequency offset of 400 Hz, and at first kept the smallest possible exposure time of 0.3 μs yielding to a sampling rate of around 7300 Hz.

As Fig. 11 reveals, with this setup it is possible to reach a range of 3 meter with a failure rate of less than 20%. We consider this failure rate tolerable, because it is possible to just repeat the synchronization in event of failure.

By prolonging the exposure time to 30.7 μs, it is possible to synchronize without failures at 4 meter distance. Even though the exposure time was increased 100 times, the average sampling rate decreased just by 17.2 %. The reason is that most of the frame-time, the sensor is occupied with pixel readout. This enables us to reach a range of 9 meter without any failed

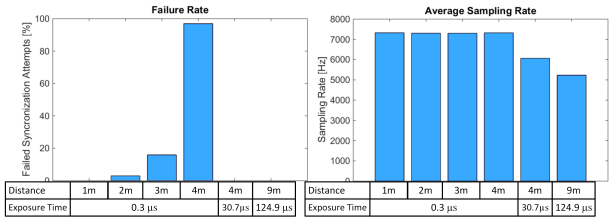


Fig. 11. Evaluation of synchronization range with an initial frequency offset of $\Delta F = 400Hz$. The first experiments for 1-4 m distance were conducted by using the minimum exposure time of 0.3 μs. Increasing the exposure to 30.7μs enables synchronization over 4 m. A further increase to 124.9 μs exposure time enables reliable synchronization over a distance of 9 meter.

synchronization attempts while maintaining a sample rate of over 5000 Hz with an exposure time of 124.9 μs.

VI. Conclusion

This work introduces Time-of-Flight 3D imaging sensors for free-space optical communication, using PLPSK. We emphasize the problem of asynchronous modulation signals and show that synchronization is vitally important for a successful communication link.

Our proposed solution is based on estimating the frequency offset ΔF between the modulation signals of two communicating systems. This is accomplished by evaluating the phase offset signal $\mathcal{O}$ and dynamically adapting the modulation signal frequency. We prove the feasibility of our method with a prototype system, executing our proposed synchronization procedure. We evaluate the performance of our solution under difficult conditions and offer an analysis on the requirements and boundaries of our approach.

Future work involves the development of a low-level communication protocol, incorporating the synchronization approach and offering a reliable location-aware optical communication link.

Acknowledgment

The authors would like to thank the Austrian Federal Ministry for Transport, Innovation and Technology as well as the Austrian Research Promotion Agency (FFG) for funding the research activities under the grant agreement number 852328.

References

[1] T. Yamazato, I. Takai, and H. Okada, "Image-sensor-based visible light communication for automotive applications," *Communications ...*, no. July, pp. 88–97, 2014.

[2] N. Druml, G. Fleischmann, C. Heidenreich, A. Leitner, H. Martin, T. Herndl, and G. Holweg, "Time-of-Flight 3D Imaging for Mixed-Critical Systems," in *13th International Conference on Industrial Informatics (INDIN)*, July 2015, pp. 1432–1437.

[3] M. Tobias, K. Holger, F. Jochen, A. Martin, and L. Robert, "Robust 3D Measurement with PMD Sensors," *Range Imaging Day, Zürich*, vol. 7, no. Section 5, p. 8, 2005. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.132.5821{\&}rep=rep1{\&}type=pdf

[4] W. Yuan, R. Howard, K. Dana, R. Raskar, A. Ashok, M. Gruteser, and N. Mandayam, "Phase messaging method for time-of-flight cameras," in *IEEE International Conference on Computational Photography (ICCP)*, May 2014.

[5] S. Schmid, G. Corbellini, S. Mangold, and T. R. Gross, "An LED-to-LED Visible Light Communication System with Software-Based Synchronization," pp. 1264–1268, 2012.

[6] W. Hu, "LightSync : Unsynchronized Visual Communication over Screen-Camera Links," pp. 15–26.

High-performance Indoor Positioning and Pose Estimation with Time-of-Flight 3D Imaging

Hannes Plank*, Theresa Egger*, Christoph Steffan*, Christian Steger†, Gerald Holweg*, Norbert Druml*
*Infineon Technologies Austria AG, Graz, Austria
{hannes.plank, theresa.egger, christoph.steffan, gerald.holweg, norbert.druml}@infineon.com
†Graz University of Technology, Austria, steger@tugraz.at

***Abstract*—In recent years, fields such as industrial automation, virtual and augmented reality and autonomous robotics increased the demand for location-awareness of electronic devices. Image sensor based inside-out localization and tracking systems are sufficiently accurate to determine the position and orientation of electronic devices. Without additional sensors however, these systems are impaired in reaching high update-rates, handling fast motions, and tend to be unable to provide localization with low latency.**

We present a new localization approach in our work, using Time-of-Flight 3D sensors in combination with small reflective markers. This allows to establish high-performance optical localization systems, delivering the position and orientation of a device at a rate of several hundred Hz. A novel Time-of-Flight 3D sensing procedure is introduced, enabling to measure the 3D positions of fast moving targets at unprecedented frame-rates. With this work, we aim to close the gap between indoor positioning and motion tracking, enabling a new class of location-aware devices.

I. Introduction

In the past, indoor positioning and pose tracking have been widely regarded as separate disciplines. If electronic devices are however capable to determine their pose (position and orientation) at a high update-rate, the demands for indoor positioning and motion tracking applications can be satisfied in a unified system. Consumer electronics in the fields of augmented and virtual reality are among the most demanding applications for such pose estimation systems. Since pose-depended information is embedded into the user's viewpoint, low latency and high pose estimation update-rates are crucial for the users experience.

Positioning and tracking systems based on 2D cameras are well established. Since conventional image sensors need to capture light for a certain time and then digitize every pixel on each frame, the frame-rate is limited. State-of-the-art solutions hence require additional sensors to satisfy the demand for high update-rates and low latencies.

This work proposes a system consisting solely on ToF sensors. It can be included on a device and is capable of determining its pose at high update-rates. The basic concept is illustrated in Fig. 1. Time-of-Flight camera emits light and captures 3D images of reflective markers. In contrast to active beacon based systems, small randomly distributed reflectors drastically reduce the infrastructure effort of a positioning system. A device equipped with a ToF imaging sensor, can use distance measurements to determine the relative 3D position

978-1-5090-6299-7/17/$31.00 © 2017 European Union

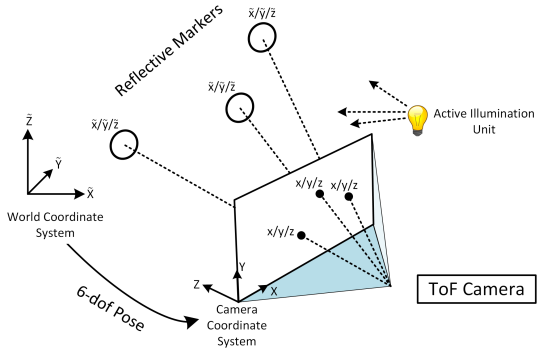


Fig. 1. The principle of our proposed positioning system. A Time-of-Flight 3D imaging system determines its pose by sensing reflective markers.

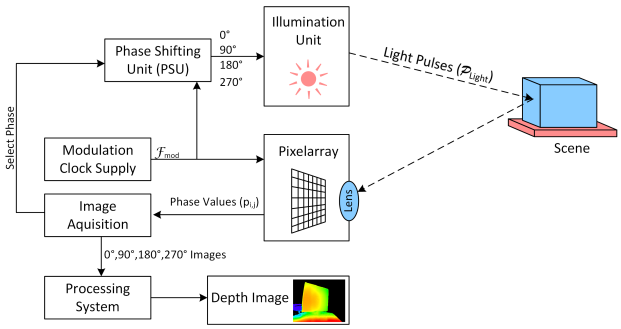


Fig. 2. The principle of Time-of-Flight 3D imaging, obtained with changes from [4].

of each visible marker. Since ToF sensors are capable of measuring distances with the correct scale, the pose to a real world coordinate system can be determined by applying 3D pointcloud matching methods. Since the signal quality is vastly improved from reflecting targets [1], short exposure times in the range of 30 to 200 microseconds are possible.

A. Time-of-Flight 3D Imaging

Time-of-Flight sensors [2] are miniaturized depth sensing systems, which are increasingly included into augmented reality smartphones and embedded systems [3]. ToF sensors work by measuring the time it takes for light to travel from the camera to the scene and back to the sensor. This produces images, where each pixel encodes the distance between the camera and corresponding objects.

As illustrated in Fig 2, an active illumination unit is located close to the imaging sensor and emits pulsed infrared light. The light pulses are reflected by the scene and projected by the lens onto the sensor. Due to the travel time, the phase of the incoming pulsed light signal $\mathcal{P}_{Light}$ is shifted. This phase-offset is measured by a photonic mixer device (PMD) located on each pixel i, j [5]. When photons arrive at the pixel, the PMD transfers the generated charges into either of two capacitors A or B on each pixel. Whether the charges are stored in A or B is controlled by the signal $\mathcal{F}_{mod}$. This modulation signal determines the frequency of the emitted pulsed light signal, and is routed to each pixel. Subtracting the digitized voltages of A and B leads to the phase value p. While p is related to the distance of the reflected object, it also depends on the amount of reflected light. Since each material reflects light differently, it is necessary to normalize these effects. An established way to calculate the actual phase φ is the four-phase algorithm. It works by capturing four different images of the same scene. In each image, the phase of the emitted light pulses $\mathcal{P}_{Light}$ is shifted by equidistant offsets of 0°, 90°, 180°, 270°. With the resulting four phase values p, the actual phase offset φ can be calculated.

$$\varphi = \arctan \frac{p_{90^\circ} - p_{270^\circ}}{p_{0^\circ} - p_{180^\circ}} \tag{1}$$

The phase offset φ is proportional to the Time-of-Flight of the light and thus the distance d and can be determined using the modulation frequency f and the speed of light c.

$$d = \frac{c}{2} \cdot \frac{\varphi}{2\pi f} \tag{2}$$

Due to the measurement principle, the phase values are only directly influenced by pulsed infrared light within a certain frequency interval. Continuous lightsources such as lamps or sunlight do not produce a signature in either the phase or the depth images. This reduces the image processing effort, since reflective markers can be extracted intuitively due to their high signal to noise ratio.

If either the camera or the sensed objects move, traditional Time-of-Flight imaging produces artifacts. We thus propose a novel Time-of-Flight distance measuring procedure in Section III. In combination with a new sensor calibration procedure, it is possible to measure 3D positions of reflective markers during fast camera movements at several hundred Hz. The resulting measurement principle is not sensitive to the amount of reflected light, allowing to use differently sized reflectors.

This leads to a simple and compact inside-out positioning system, where the camera is mounted on the tracked object, and senses the distances to fixed infrastructural markers for localization. Such system is capable of tolerating fast movements and thus able to simultaneously fulfill the requirements of indoor positioning and motion tracking applications.

B. Marker-based Optical Positioning with 3D Cameras

Camera-based positioning systems [6] need to identify reference points in their field of view in order to calculate their location and orientation. With 2D cameras, only projections of reference points can be measured. While the pose can be calculated by solving the perspective-n-point problem, it is advisable to identify the reference points. Marker-based 2D systems thus either employ large markers visually encoding information, multiple view geometry methods or optical communication. This impairs the deployment of 2D based positioning systems, since the infrastructure effort might be too large for many applications.

When reflective markers are sensed with depth cameras [7], it is possible to use the distance measurements and the incident angel to determine the 3D positions of the visible markers. The identity of the markers can be determined by their geometric relation to other visible markers. Therefore it is possible to match the visible pointcloud reference model, providing the orientation and position of the camera system.

This allows to use small, hardly noticeable reflectors as markers, in our case with a surface of 1 cm^2. Since reflectivity increases the signal strength, very robust 3D ToF measurements with comparatively short exposure times are possible. In our approach, the markers themselves are passive reflectors, and hence do not encode identification information. Due to ToF distance measurements, the markers and can be identified due to the relative positions of the surrounding markers. Since Time-of-Flight sensors are increasingly miniaturized and mass-produced, potentially allowing a manifold of future electronic devices to become location-aware.

C. Our Contribution

In this work, we introduce a novel optical indoor-positioning and motion tracking method based on Time-of-Flight sensors. By sensing reflective markers, such system has the capability to determine its pose with all 6-dof with low latency and at a high update-rate. This is enabled by a new depth sensing procedure, capable of tracking markers in 3D without motion artifacts.

II. Background

The field of 2D camera-based positioning is well-established however hybrid systems, offering both positioning and sufficient update-rates for motion tracking are rare. In recent years, virtual and augmented reality and as well the expanding industrial automation increased the demand for such unified systems. The current generation of consumer products, such as the Oculus Rift or the HTC Vive currently use outside-in tracking solutions, providing a pose relative to the base-stations. With the Google Tango platform, robust inside-out motion tracking and indoor localization became available on mobile devices. The common denominator of these systems are high-performance tracking capabilities enabled by sensor fusion approaches which might use extended Kalman filtering [8]. While imaging sensors provide stable low speed tracking, MEMS-based inertial sensors provide high update-rates. The combination of multiple sensors however increases the cost, size and complexity of such systems, limiting the tracking and positioning systems to expensive devices. In this work, we aim to provide an optical system, capable of satisfying the demands of motion tracking and positioning applications with a single sensor.

Using 3D cameras for marker-based positioning has been primarily conducted with stereo cameras so far. Maeda et

al. propose a hybrid system, using a stereo camera with active and passive markers [9]. Optical communication is used to transmit reference coordinates to the camera system. A visible light communication-based positioning system is also introduced by Nakazawa et al. [10]. Optical communication for positioning is a promising research field [11] due to the large range of applications. In Section V-C, we show that we are considering to combine the proposed positioning system with optical communication.

The feasibility of using passive retro-reflective markers in combination with a 3D stereo system was demonstrated by Mehling [7]. We follow a similar approach, in our work using the relative 3D positions of the detected markers for identification. To best of our knowledge, the potential of Time-of-Flight 3D imaging sensors for marker-based positioning has not yet been explored. In contrast to stereo, ToF cameras directly measure the distance and do not need to solve the correspondence problem and can lead to miniaturized systems. When sensing reflective markers, the positioning precision remains nearly constant over a range of 10 meters, while allowing potential update-rates in the range of several hundred Hz.

A. Error Analysis for Single View Positioning

In this Section, we investigate the potential positioning accuracy of 3D ToF sensors, compared to 2D based methods. Another goal is to determine, how camera parameters such as field of view and sensor resolution influence the accuracy.

We define the positioning error E_{pos} as the worst possible distance deviation, which is introduced due to measurement imprecision. This error however depends on the number of visible reference points and their positions relative to the camera. We therefore create a basic scenario to investigate the influence of camera parameters on E_{pos}. As illustrated in Fig. 3 in this scenario, a 2D or 3D system wants to determine the distance $d_{C,P}$ from itself to the center of two reference points P_1 and P_2. With a 2D system, the distance $d_{C,P}$ is calculated by triangulation. This requires precise registration of the 2D position of the projections of P_1 and P_2. The resolution and field of view (FoV) are thus the camera parameters influencing E_{pos}. Assuming square pixels, the horizontal FoV angle Θ and number of pixel per row N_{row} can be used to determine the angular resolution ψ (Eq. 3) of a camera, which is the minimum angle at which a sensor can distinguish positions of tracked markers. Since most marker-based positioning systems apply sub-pixel accuracy methods, the factor f_{sub} represents the number of different positions which can be distinguished between two pixels.

$$\psi = \frac{N_{row}}{\Theta} \cdot f_{sub} \tag{3}$$

In this scenario the distance d_{P_1,P_2} between P_1 and P_2 are previously known by the positioning system. The distance $d_{C,P}$ can be derived by triangulation, since calibrated cameras are able to measure the angle α between its position and P_1 and P_2. As shown in Fig. 3, the camera directly points to the center of the points, and the distances between the camera

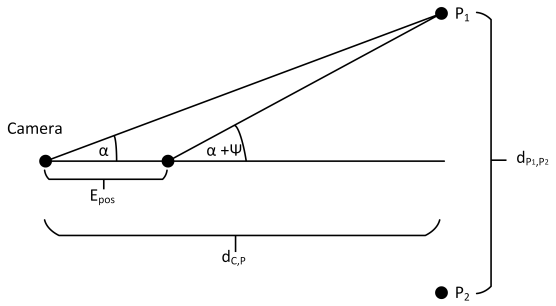


Fig. 3. Triangulation-based distance measurement

and the points are equal. This leads to the relation between the desired distance d_{P_1,P_2} and $d_{C,P}$.

$$d_{C,P} = \frac{\frac{d_{P_1,P_2}}{2}}{\tan(\alpha)} \tag{4}$$

If the angle α is measured by an image sensor, an uncertainty of $\pm\frac{\psi}{2}$ is introduced. In the worst case, the measured angle $\tilde{\alpha}$ thus deviates by ψ. The distance $\tilde{d}_{C,P}$ for this worst deviation can therefore be expressed as:

$$\tilde{d}_{C,P} = \frac{\frac{d_{P_1,P_2}}{2}}{\tan(\alpha + \psi)} \tag{5}$$

The positioning error in this analysis is the difference between the ground-truth distance $d_{C,P}$ and worst deviating measurement $\tilde{d}_{C,P}$:

$$E_{pos} = \|d_{C,P} - \tilde{d}_{C,P}\| \tag{6}$$

Inserting Eq. 5 in Eq. 6 and substituting α with the flawed angle measurement $\arctan(\frac{d_{P_1,P_2}}{d_{C,P}})$, the term for the positioning error for 2D cameras is:

$$E_{pos} = \|d_{C,P} - \frac{\frac{d_{P_1,P_2}}{2}}{\tan[\arctan(d_{P_1,P_2}) + \psi]}\| \tag{7}$$

In order to visualize the influence of sensor resolution, we analyze the error for different common camera resolutions. In Fig. 4, the error E_{pos} is plotted for distances between 1 and 10 meters, with a point to point distance d_{P_1,P_2} of one meter. We choose a common horizontal field of view of 55°and assume a conservative estimate for the sub-pixel accuracy factor f_{sub} of 2. For comparison with direct depth measurements, we also plot the error curve for depth measurements with our CamBoard pico flexx Time-of-Flight camera, which has a rated measurement error of 1% of the distance between 0.5 - 4 m.

The curves for the 2D distance estimation in Fig. 4 show, that the relation between the distance target points and distance error increases in a non-linear fashion. It shows the importance of sensor resolution for 2D image based positioning. Unfortunately, a higher sensor resolution also decreases the frame-rate. The Time-of-Flight sensor we use in our work features a resolution of 224x171 pixel. As the comparison with the 2D cameras shows, the resolution limit can be tackled by depth measurements and sub-pixel accuracy methods. The comparatively low resolution however enables fast sensor-

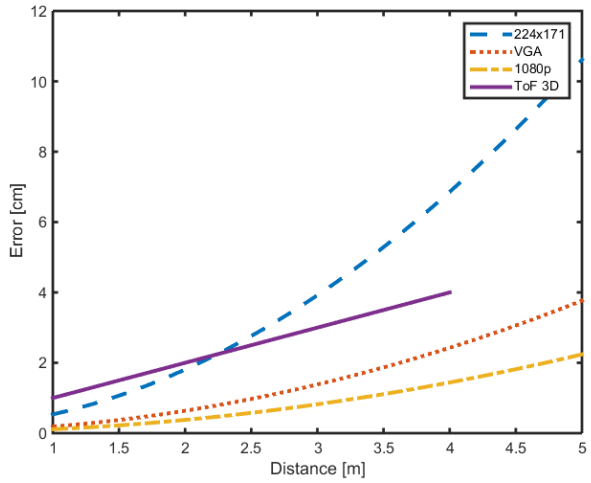


Fig. 4. Comparison of 2D distance estimation between triangulation and Time-of-Flight Imaging

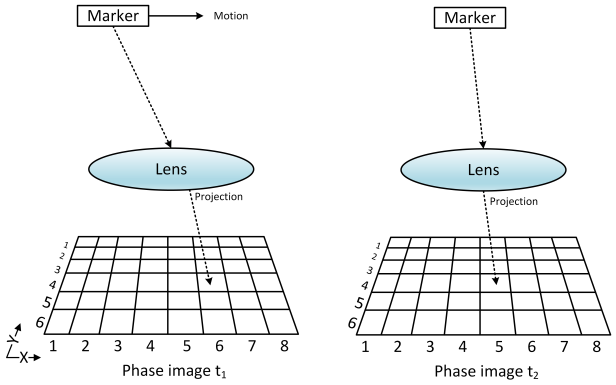


Fig. 5. Relative motion can change the pixel position of a projected marker between subsequent phase images.

readout and consequently a high positioning update-rate.

III. Motion Resilient Distance Measurements with Time-of-Flight Sensors

Our proposed positioning system aims to enable high update-rates to satisfy the demands of positioning and motion tracking applications. Strong unpredictable positional and rotational accelerations should therefore not impair the system. As explained in Section IV-C, unlike 2D camera systems, 3D measurements allow to identify markers by their geometric relation. Furthermore, the analysis in Section II-A shows that 3D measurements can increase the positioning precision. A disadvantage however is that conventional Time-of-Flight depth imaging faces problems when sensing fast moving objects. We therefore propose a ToF distance capture and calibration solution in this Section, enabling 3D position measurements of fast moving reflective markers.

A. The Motion Problem

A big challenge for tracking fast movements with optical systems is motion blur. With reflective markers, the comparatively large amount of reflected light allows for short sensor exposure times, effectively mitigating motion blur. However a severe and commonly uncompensated problem of ToF sensors are motion artifacts. These erroneous depth measurements are caused by inconsistencies among subsequent phase images. As shown in Section I-A, ToF sensors need to capture typically four different phase images in order to produce a single depth image. After each exposure, the charge contents of all pixels needs to be digitized and transferred to a processing system. This introduces a time delay between subsequent images. As illustrated in Fig. 5, when moving objects are projected onto the sensor, the projected position changes with each captured phase image.

The problem is that different sets of phase image pixels cannot be used to calculate distances due to a lack of calibration. In contrast to most conventional 2D image sensors, the output signal of a ToF pixel does not have a linear relation to the incoming light. This introduces a non-linear relation between fabrication tolerances and the pixel output. Each pixel measures slightly different phase values for the same incoming light signal. The normal ToF sensor calibration process assumes that the same pixel is used to measure all four phase values. We therefore introduce a new calibration procedure, enabling to gather reliable distance measurements from such distributed signal positions.

B. Sensor Calibration Method

In this Section, we show a new calibration approach enabling ToF sensors to measure the distance to fast moving point lightsources, such as reflective markers. As shown in Fig. 5, when subsequent images of moving objects are captured, the light arrives at different sets of pixels.

Instead of capturing four phase images and then calculating the distance for each pixel, we gather the phase value p of the marker directly in each phase image. The distance to the marker is then calculated from these four phase measurements, utilizing a new sensor calibration approach. As discussed in Section III-A, a new calibration procedure and model are necessary since different pixels are used to measure phase values.

Generally, a calibration model should map raw measurements to results, eliminating device-depended errors. During the calibration process, reference measurements are conducted to derive the parameters of the calibration model. The proposed calibration process in this approach is designed to be simple and capable of being performed automatically during operation. We therefore use conventional ToF distance sensing to measure reference distances during the calibration process.

A requirement for our calibration model is a phase value p of a marker. Our previous experiments lead to the conclusion that the average pixel values of a projected point light source can be used for distance calculation [**?**]. Averaging the marker signal contained by a $n \times m$ pixel patch also removes certain pixel-depended errors. Potential background noise is removed by restricting the averaging of the signal of the pixels $p_{i,j}$ with a signal above a certain threshold Z.

$$p = \frac{1}{N} \cdot \sum_{i=u}^{u+n} \sum_{j=v}^{v+m} \begin{cases} p_{i,j}, & \text{if } |p_{i,j}| \geq Z \\ 0, & \text{otherwise} \end{cases} \tag{8}$$

The next step is to calibrate the phase value p so the produced value only depends on the distance or the amount of reflected light. This means, that p should remain constant for all possible projected 2D marker positions $M_{x,y}$. Unlike conventional image sensor calibrations, we do not use discrete pixel positions, but use continuous coordinates and interpolation. The measured phase value p is corrected by a calibration factor $\eta_{(x,y)}$, which ideally eliminates the influence of the projected marker position $M_{x,y}$ on p. Therefore $\eta_{(x,y)}$ is a function of the projected marker position. In our prototype, we implement $\eta_{(x,y)}$ as a linear interpolation between support values $\tilde{\eta}_{x,y}$. These support values are measurements, gathered during the calibration procedure and are conducted for a certain reference distance d_{ref}. A major cause for phase value variation is inhomogeneous light signal distribution on the sensor. This is caused by the active illumination unit, the lens and sensor manufacturing tolerances.

In conventional ToF depth sensing, the amount of light does not have a direct influence on the depth measurement since linear shifts do not influence the phase calculation due to normalization during the arcustangens operation. When the markers however are projected to different positions $M_{x,y}$ on the sensor, each phase value is influenced separately. Therefore, we chose to equalize the influence of incoming light for all possible marker positions $M_{x,y}$. Since the four equidistant phase measurements sample a periodic signal, it is also possible to calculate the amplitude of this signal. This amplitude only depends on the amount of incoming light. Therefore it is possible to take conventional four phase measurements p_i of a stationary marker, and calculate the support value $\tilde{\eta}_{x,y}$.

$$\tilde{\eta}_{x,y} = \frac{1}{\sqrt{(p_{0^\circ} - p_{180^\circ})^2 + (p_{90^\circ} - p_{270^\circ})^2}} \tag{9}$$

$\eta_{(x,y)}$ is a continous function, supported by the actual calibration values $\eta_{x,y}$. In our approach, we use a linear interpolation of the three closest $\eta_{x,y}$ values to derive η at position (x/y). Dividing the phase value of a marker p at position (x, y) by $\eta_{(x,y)}$ produces the calibrated phase value $\overline{p}$.

$$\overline{p} = \frac{p}{\eta_{(x,y)}} \tag{10}$$

This process is conduced on a dense set of 2D positions with constant marker distance. This correction works for different marker parameters such as reflectivity and size. It also allows to change the exposure time of the sensor without influences on the calibration parameters. The final phase $\overline{\varphi}$ of the incoming light pulses is calculated analog to Eq. 1.

The last stage of the calibration model is to map the corrected phase measurement $\overline{\varphi}$ to a distance $\overline{d}$. We introduce parameter $\gamma_{(\varphi)}$, a scalar factor depending on the measured phase $\overline{\varphi}$. $\gamma_{(\varphi)}$ is also a linear interpolation of support values $\tilde{\gamma}_{(\varphi)}$. These support values are determined during the calibration process by conducting reference distance measurements d_r and calibrated reference phase measurements $\overline{\varphi}$.

$$\tilde{\gamma}_{(\varphi)} = \frac{d_r}{\overline{\varphi}} \tag{11}$$

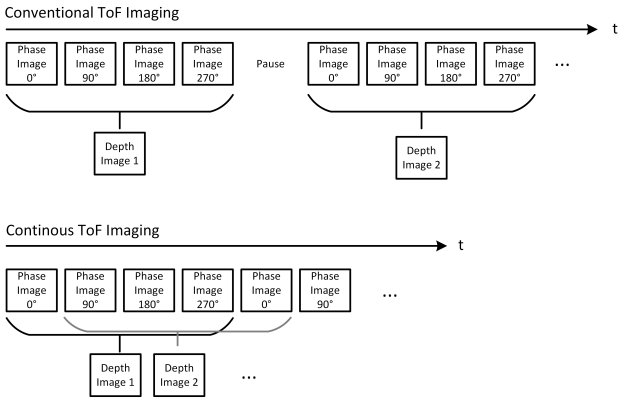


Fig. 6. Continous ToF imaging leads to a higher frame-rate compared to conventional ToF imaging.

This scaling factor depends on the phase-offset due to the wiggling error. As stated in Eq. 1, converting the phase-measurements to the actual phase shift uses the $\arctan$ operation, requiring a sinusoidal signal to accurately determine the phase-shift. Due to the non-sinusoidal shape of the pulsed light, the wiggling error is introduced. Because this error varies with phase, $\tilde{\gamma}_{(\varphi)}$ depends on the phase offset φ. During the calibration process, multiple values of $\tilde{\gamma}_{(\varphi)}$ are captured for different distances, avoiding the necessary offset error calibration, conducted during normal ToF depth sensing. Since the distance is proportional to the phase-offset φ, a multiplication of $\gamma_{(\varphi)}$ directly leads to the final distance $\overline{d}$.

$$\overline{d} = \gamma_{(\varphi)} * \varphi \tag{12}$$

This distance is updated from with each captured phase image, using the continuous ToF capture mode, depicted in Fig. 6. In Section V, we evaluate the calibration procedure and measure the precision when sensing fast moving objects.

C. High Frame-rate Time-of-Flight

Conventional Time-of-Flight systems usually capture four or more phase images, and then produce a single depth image. High-speed distance measurements with Time-of-Flight cameras have been explored so far by Stühmer et al. [12]. This approach is however aimed to track larger objects using a model-based approach. In our system, we want to produce as many distance measurements per second as possible. If there is no pause or delay between phase images, ToF systems can have problems with overheating, excessive power consumption or eye-safety violations. By using retroreflective markers however, these problems are insignificant, because the illumination time per frame can be reduced. This allows to configure the ToF sensor to capture phase images at a constant rate, which allows continuous ToF sensing. As shown in Fig. 6, continous ToF sensing works by instead of taking four different images to capture a depth image, the last four phase images are used for each depth image. This approach allows to produce a new distance measurement after each gathered phase image, quadrupling the depth frame-rate.

IV. Our Proposed Indoor Positioning System

So far we showed how Time-of-Flight sensors can measure the 3D positions of reflective markers at high frame-rates without motion artifacts. The final step is to determine the position and orientation of the ToF camera with respect to a world coordinate system (e.g. floorplan of a building). Since our proposed system directly measures 3D marker positions, a single distance measurement of three visible markers is sufficient. As we show in Section IV-C, the relative position of the markers can be matched with a reference geometry model.

While post processing, motion models or sensor fusion approaches can further increase the robustness and accuracy, we aim to provide the framework of a base-system in this work, capable of delivering accurate positioning information solely on optical information. Further processing methods can then be added and might be chosen and adapted to fit the application.

A. 2D Marker Registration

The first step in a marker-based optical positioning system is to determine the projected 2D position of each visible marker on the sensor. When Time-of-Flight sensors capture phase images, the signal is only influenced by pulsed infrared light. Since the active illumination unit next to the sensor illuminates the scene, only the reflected light is captured. The reflectivity of the markers allows to use comparatively short exposure times, while maintaining an appropriate signal to noise ratio. Since no other significant signals are captured, it is possible to detect the sub-pixel accurate marker positions fitting a 2D Gaussian distribution 13 with no further processing.

$$f(x,y) = I \cdot exp\left(-\frac{(x-x_0)^2}{2\cdot\sigma_x^2} + \frac{(y-y_0)^2}{2\cdot\sigma_y^2}\right) \tag{13}$$

The coarse position of the marker is detected by thresholding and used to initialize the minimization procedure [13]. If we constrain the amplitude parameter, to $\pm 20\%$ of the highest pixel value, we achieve a 100% convergence rate. We also add image rotation as parameter to the minimization process to accommodate potential motion blur and non square markers. An example for a fitted Gaussian distribution is shown in Fig 7. The grid shows the phase signal of a 1 cm marker captured with 30 μs exposure time. The parameters x_0 and y_0 correspond to the sub-pixel estimation of the marker position. In future work, it could be possible to analyze the rotation angle and the ratio between the variances σ_x^2 and σ_y^2, to detect motion blur. This could provide the motion direction and be used to predict positions.

B. 3D Marker Position Calculation

The measured distance d and the estimated sub-pixel position are used to calculate the 3D position M_{3D} of a marker, assuming a calibrated camera. The intrinsic camera parameters of a Time-of-Flight sensor can be obtained the same way as 2D cameras, since ToF cameras are able to produce 2D images as well. A common intrinsic camera calibration procedure is to capture multiple checkerboard pattern from different distances and orientations and determine the parameters by iterative minimization methods [14]. Assuming the lens distortion is compensated, the 3D position of a point relative to the camera can be calculated by multiplying the normalized 2D marker position M_{2D} with the pseudo inverse P_{Cam} of the camera matrix.

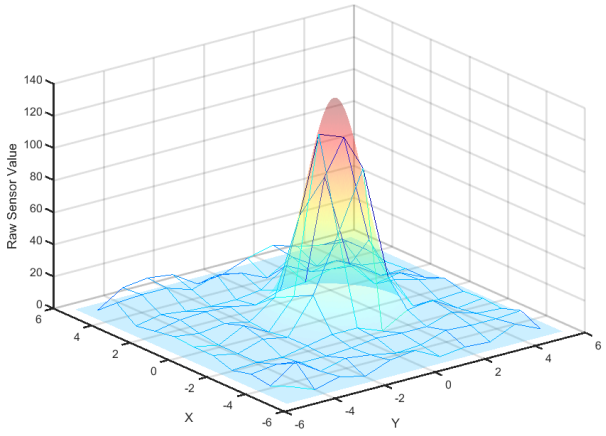


Fig. 7. A 2D Gaussian distribution is fitted to a phase image to determine the marker position with sub-pixel accuracy

$$M_{3D} = d \cdot P_{Cam} \cdot M_{2D} \tag{14}$$

C. Marker Identification and Pose Calculation

Finding the position and orientation of a camera, when sensing markers in 3D is a pointcloud matching problem, where the visible set of points P_M is matched with reference points P_{Ref}. As shown by Mehling [7], it is possible to use the geometry of the visible set of markers for identification. In our evaluation we find the point correspondences by matching the distances between visible points to the distances between the markers.

When the identity and 3D position of the markers are known, it is possible to find a least-square solution for the pose parameters [15]. The pose contains the transformation vector t and rotation matrix R defining the transformation from world to camera coordinates. This allows to determine the orientation and location of the camera system with all 6-dof, when the world coordinates of the available markers are available in a reference model.

The pose parameter is the rigid transformation between the pointcloud P_M of the i visible markers and their corresponding reference pointcloud P_{Ref}. With the centroids C_M and C_{Ref} of P_M and P_{Ref} it is possible to assemble matrix H, as mentioned by Ho [16].

$$H = \sum (P_M^i - C_M) \cdot (P_{Ref}^i - C_{Ref}) \tag{15}$$

A singular value decomposition (SVD) of H leads to the rotation matrix R (Eq. 16,17).

$$[U,S,V] = SVD(H) \tag{16}$$

$$R = V \cdot U^T \tag{17}$$

The transformation vector t can then be calculated by:

$$t = -R \times C_M + C_{Ref} \tag{18}$$

D. Limitations

Reflective markers are a simple and effective way to increase the signal-to-noise ratio significantly. This enables to extend the distance measuring range beyond the Time-of-Flight sensor's specifications. After a certain distance however, a ToF distance measurement becomes ambiguous. This happens because the phase of the incoming light pulses is shifted by a complete period. For a typical modulation frequency of 80 MHz, this distance is 1.8 m. A potential countermeasure for this problem is to measure the amount of light which is reflected by the marker and classify the phase interval.

Another problem might be the sensor dynamic. Each pixel has two capacitors where charges are stored until pixel readout. If too much light arrives, these capacitors saturate, and may cause an invalid measurement. This might be a problem if markers of close and far distances are sensed simultaneously in a single image. If there are enough markers in the background, close markers can be disregarded. It might also be feasible, to choose smaller or less reflective markers for positions which are potentially closer to the camera, or dynamically adapt the sensor exposure time.

In this work, we mention potential update rates of several hundred Hz. We previously conducted experiments [4] with ToF sensors connected to an FPGA platform, reaching 500 frames per second with full sensor readout. These high framerates however are not yet reachable with existing ToF cameras, due the amount of produced data. It is therefore subject to future work to create a capable processing system, providing devices such as PCs or smartphones with positioning information.

V. Results and Evaluation

In this Section, we use a prototype implementation to conduct experiments to receive insight on the potential performance of a Time-of-Flight based positioning system. We use the CamBoard pico flexx Time-of-Flight camera, jointly developed by Infineon and pmdtechnologies for this evaluation. This miniaturized Time-of-Flight sensor features a resolution of 224x171 pixel, and has a specified distance error of 1% of the measured distance.

The markers consist of reflective fabric, which is commonly used in high visibility clothing. We use 1x1cm planar patches in our test setup. As seen in Fig. 8, the ToF camera is mounted on a linear slide and rotation platform, to provide accurate ground-truth data for the position analysis.

A. Sensor Calibration

In this experiment, we investigate the calibration procedure, introduced in Section V-A. The goal of the calibration is to enable the ToF camera to measure distances to moving markers. As we showed in Section III-A, when the four phase images are gathered, the light arrives at different pixels on each image. We therefore want to test, if the distance to a marker can be derived from phase values gathered by different pixels.

In our evaluation setup, the calibrated ToF camera is mounted on a rotation disc (Fig. 8) and different phase images of a single reflector are captured with a constant ground-truth

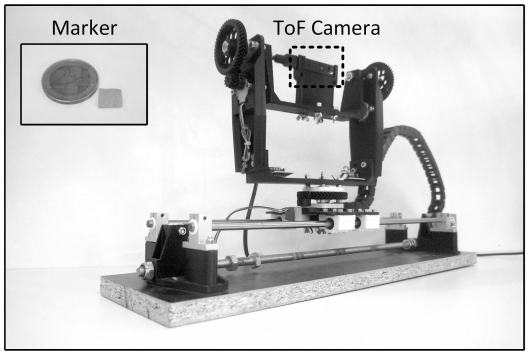


Fig. 8. Setup for calibration and evaluation. The Camboard pico flexx Time-of-Flight camera is mounted on a linear stage and rotation platform.

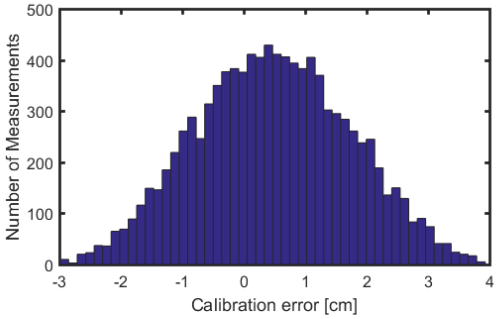


Fig. 9. The distribution of the remaining distance error after calibration.

distance d_{Gt} of 1m. The phases are randomly combined to calculate the distance d_{Cal} with different sets of pixels. This mimics the data of fast motions. The distance error is the difference between the ground truth distance d_{Gt} and d_{Cal}.

$$E_d = d_{Gt} - d_{Cal} \tag{19}$$

Fig. 9 shows the distribution of the distance error E_d introduced by using calibration method.

The standard deviation of the error is 1.25 cm and the error shows an average bias of 0.48cm. The distance variations are caused by the inaccuracy of calibration parameter $\eta_{(x,y)}$, while the distance bias is caused by $\gamma_{(\varphi)}$. Compared to the rated Time-Flight error, we are satisfied with the result, but believe that a denser set of calibration values could potentially improve the calibration error. It should be noted, that we tested the calibration with constant modulation frequency under stable temperature conditions. In future work, we will evaluate, if the temperature can be compensated equally to normal ToF depth sensing, where a fitted polynom or lookup table is sufficient.

B. Positioning Analysis

In this experiment, the positioning precision of the prototype implementation is tested with randomly distributed markers. A single distance image is captured and used to locate the markers and create a reference world geometry. The ToF camera is moved and rotated to 10 different positions and orientations in a 3x4 m room. Using a linear stage and rotation platform allows to capture the ground-truth for each

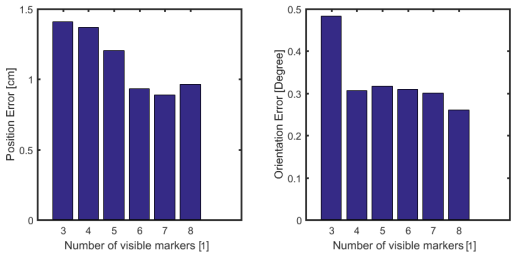


Fig. 10. Left: The average positioning error E_P in cm; Right: The average orientation error E_α in degree

measurement. The average distance between the markers and the camera system is 190 cm.

This evaluation uses the average positioning error E_P as metric for positioning precision, which is distance between the measured position P_C and the ground truth position P_{Gt}. The orientation error is quantified as the average angle deviation E_α between the measured and ground truth Euler angles of the rotation. In Fig. 10 we plot the positioning and orientation errors with different number of visible markers.

While final precision of the proposed positioning system depends on the geometry of the captured scene, this evaluation shows that it is possible to use Time-of-Flight 3D imaging systems for centimeter-accurate indoor positioning. Along with the high achievable update-rates, and the simple implementation with a single sensor, we believe that these results prove the feasibility of this positioning solution for a wide range of applications.

C. Outlook and Future Work

While we are satisfied with the robustness of the current approach, we identify a potential weakness in matching the reference geometry model with the measured marker positions. The problem is that the geometric relation between each set of observed markers needs to be unique. This can be counteracted by using active beacons in conjunction with passive markers. Such optical communication would also allow to transfer the geometric reference model to the ToF system without an additional communication system. In the future we therefore plan to combine our work on ToF camera optical communication [**?**] with this indoor positioning approach. Due to the depth measurement principle of Time-of-Flight cameras, it is possible to decode phase-shifted pulsed infrared light. By limiting the sensor readout to the region of the communication partner, it is possible to achieve a symbol-rate of 7300 baud. Already deployed Time-of-Flight systems are able to send information by using their active illumination units. A feasible transmission method is relative pulse-position keying, where a photodiode can be used to receive data from a ToF camera without synchronization. Since depth measurements can be included in a communication protocol, the positions of the base stations can serve as unique reference points, while simultaneously providing the reference geometry model.

VI. Conclusion

We present a framework for a novel indoor position system, based on Time-of-Flight 3D sensors. A major contribution is a new distance measuring method, enabling Time-of-Flight sensors to sense the 3D positions of fast moving reflective markers. Consequently it is possible to determine the position and orientation within a reference coordinate system at a high update-rate with low latency. The characteristics of our proposed system are experimentally evaluated with a prototype setup.

We believe that 6-dof positioning with Time-of-Flight distance measurements is a promising method for indoor localization and pose estimation, satisfying the demands for consumer and industrial applications.

Acknowledgment

The authors would like to thank the Austrian Federal Ministry for Transport, Innovation and Technology as well as the Austrian Research Promotion Agency (FFG) for funding the research activities under the grant agreement number 852328.

References

[1] A. V. H. Ollikkala and A. J. Mäkynen, "Range imaging using a time-of-flight 3D camera and a cooperative object," *2009 IEEE Intrumentation and Measurement Technology Conference, I2MTC 2009*, no. May, pp. 817–821, 2009.

[2] R. Schwarte, Z. Xu, H.-G. Heinol, J. Olk, R. Klein, B. Buxbaum, H. Fischer, and J. Schulte, "New electro-optical mixing and correlating sensor: facilities and applications of the photonic mixer device (pmd)," pp. 245–253, 1997. [Online]. Available: http://dx.doi.org/10.1117/12.287751

[3] N. Druml, G. Fleischmann, C. Heidenreich, A. Leitner, H. Martin, T. Herndl, and G. Holweg, "Time-of-Flight 3D Imaging for Mixed-Critical Systems," in *13th International Conference on Industrial Informatics (INDIN)*, July 2015, pp. 1432–1437.

[4] H. Plank, A. Schoenlieb, C. Ehrenhoefer, C. Steger, G. Holweg, and N. Druml, "Synchronization of time-of-flight 3d sensors for optical communication," May 2017.

[5] B. Büttgen and P. Seitz, "Robust Optical Time-of-Flight Range Imaging Based on Smart Pixel Structures," vol. 55, no. 6, pp. 1512–1525, 2008.

[6] R. Mautz and S. Tilch, "Survey of Optical Indoor Positioning Systems," 2011.

[7] M. Mehling, "Implementation of a Low Cost Marker Based Infrared Optical Tracking System," *Master Thesis*, p. 101, 2006.

[8] A. T. Erdem and A. Ozer, "Fusing Inertial Sensor Data in an Extended Kalman Filter for 3D Camera Tracking," vol. 24, no. 2, pp. 538–548, 2015.

[9] M. Maeda, T. Ogawa, K. Kiyokawa, and H. Takemura, "Tracking of user position and orientation by stereo measurement of infrared markers and orientation sensing," in *Eighth International Symposium on Wearable Computers*, vol. 1, Oct 2004, pp. 77–84.

[10] Y. Nakazawa, H. Makino, and K. Nishimori, "LED-Tracking and ID-Estimation for Indoor Positioning using Visible Light Communication," pp. 87–94, 2014.

[11] T.-h. Do and M. Yoo, "An in-Depth Survey of Visible Light Communication," 2016.

[12] J. Stuemer, S. Nowozin, A. Fitzgibbon, R. Szeliski, T. Perry, S. Acharya, D. Cremers, and J. Shotton, "Model-Based Tracking at 300Hz using Raw Time-of-Flight Observations," 2015.

[13] T. F. Coleman and Y. Li, "On the convergence of interior-reflective newton methods for nonlinear minimization subject to bounds," *Math. Program.*, vol. 67, no. 2, pp. 189–224, Nov. 1994.

[14] Z. Zhang, "Flexible camera calibration by viewing a plane from unknown orientations," in *Proceedings of the Seventh IEEE International Conference on Computer Vision*, vol. 1, 1999, pp. 666–673 vol.1.

[15] P. Besl and N. McKay, "A Method for Registration of 3-D Shapes," pp. 239–256, 1992.

[16] N. Ho, "Finding the optimal rotation and translation between corresponding 3d points," May 2013.

Fast and Energy-efficient Time-of-Flight Distance Sensing Method for 3D Object Tracking

1st Hannes Plank
CRE
Infineon Technologies Austria AG
Graz, Austria
hannes.plank@infineon.com

2nd Gerald Holweg
CRE
Infineon Technologies Austria AG
Graz, Austria
gerald.holweg@infineon.com

3rd Christian Steger
Institute for Technical Informatics
Graz University of Technology
Graz, Austria
steger@tugraz.at

4th Norbert Druml
ATV Sense and Control
Infineon Technologies Austria AG
Graz, Austria
norbert.druml@infineon.com

***Abstract*—We present a new energy-efficient distance sensing method for 3D object tracking with Time-of-Flight sensors. The field of 3D object tracking with 3D cameras recently gained momentum due to the advent of front-facing depth cameras in smartphones. Tracking the user's head with 3D cameras will enable novel user experiences, but can lead to power consumption issues due to the active illumination.**

State-of-the-art continuous-wave Time-of-Flight imaging requires at least four different phase-images, while our approach can produce 3D measurements from single phase-images. This reduces the amount of emitted light to a minimum, improves latency and enables higher framerates. As our evaluation shows, after a brief initialization phase, our method can reduce the power consumption of a Time-of-Flight system by up to 68%.

***Index Terms*—Time-of-Flight, face tracking, smartphones, depth sensing, phase-images, 3d object tracking**

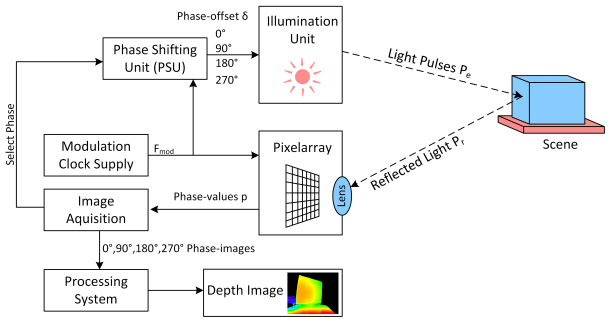


Fig. 1. The operation principle of a continuous-wave Time-of-Flight sensor. The distance for each pixel is measured via the phase difference between emitted and received light.

I. Introduction

With the introduction of front-facing depth sensors on smartphones, face tracking is a promising human computer interface opportunity. A manifold of potential applications exist, such as dynamic render-resolution, adaptive content and intelligent autofocus. The most demanding application is pseudo-holographic 3D content which can be found in games, 3D scanned scenes and innovative user interfaces. These novel interaction methods however require 3D tracking systems with high framerates, low latency and limited power consumption.

Time-of-Flight (ToF) sensors are well suited for this domain of 3D tracking. The measurement principle allows for a miniaturized depth sensing solution in combination with high framerates. In contrast to structured light and stereo-based systems, the depth is not derived by triangulation but via phase measurements of modulated light. The basic operation principle is depicted in Fig. 1. An active illumination unit emits pulsed infrared light, which is reflected and projected on the Time-of-Flight pixel array. At each pixel, a photonic mixer device (PMD) measures the correlation between incoming and outgoing light pulses. After exposure, the content of the pixelarray is digitized, producing a phase-image. Since the output of the PMD is also influenced by the amount of reflected light, ToF sensors usually capture at least four different phase-images to calculate a depth image.

The illumination unit has a major share of a ToF sensor's power consumption and might restrict the use of ubiquitous face tracking for mobile devices. In this work, we propose a distance sensing mode, enabling ToF sensors to track objects in 3D by just using single phase-images. Our approach exploits the fact that in 3D object tracking, the same object is sensed repeatedly. This enables our algorithm to create a mapping between phase-values and distances during a brief initialization phase using conventional ToF sensing.

As our evaluation shows, our method can estimate the distance to an object with an error below 1 cm in a range between 35 to 75 cm. Capturing just a single phase-image either reduces the power consumption of the illumination unit by up to 68% or alternatively can to quadruple the framerate. In both cases, the acquisition time and consequently the latency is reduced by typically more than 2 ms, further contributing to the tracking system's performance.

II. Related Work

Before ToF cameras found their way into mobile devices, reducing the number of required phase-images was motivated by obtaining higher framerates and less motion artifacts. Grünwald [1] introduced a 3-phase method, which can perform as well as 4-phase ToF imaging. In this method, the phase-images are obtained by shifting the emitted light signal by 0°, 120° and 240°. These phase-offsets are not yet established, a ToF sensors would need to support them, which might be the reason why we do not find any further investigation into this method yet.

Hussmann et al. proposed a method to obtain a depth image from two phase-images in 2009 [4] and later extended their approach for single phase-images [5]. In this approach, the sensor outputs the voltages of the on-pixel charge storage buckets A and B separately, instead of subtracting them. These values also depend on the phase and are used to calculate the distance. Besides inferior precision, the major problem with their single phase-image approach is that according to the authors, the method only works when ambient light is suppressed. Non-pulsed ambient light will fill both buckets A and B equality, which has a direct impact in their measured distance. In their work, the authors propose intermediate ambient light measurements as compensation. This will lead to a performance similar to their two-phase method but does still not reach the precision of a four phase approach.

In the domain of 3D tracking, it is not necessary to provide a complete depth image, but rather a distance measurement to an object. Felix Heide et al. [3], built a sophisticated ToF tracking system. It uses the Doppler effect of pulsed infrared light to track objects in 3D while simultaneously measuring the speed. This method can produce a 3D velocity field, where each pixel is associated with a motion vector. This system uses orthogonal modulation frequencies to directly measure velocity in phase-images. The project also uses a color camera in order to measure the optical flow of the scene, and requires at least two phase-images, captured at the same time to measure velocity.

Jan Stühmer et al. [7] propose an object tracking system based on single phase-images from ToF sensors. A model is used to calculate the signature of the object in phase-images, depending on the distance. The calculated phase-response is then matched with the actual phase-images to calculate the tracking position in 3D. Model-based approaches might be difficult to use in practice, since they require prior knowledge about the tracked object. In the application of facetracking, the reflectivity of human skin varies per person and might be influenced by sweat, skin color and makeup.

Our work provides a way to track any object in 3D by just using single phase-images. No prior knowledge about the object is required and ambient light does not influence the measurement. Compared to conventional ToF sensing, our method introduces just a minor distance estimation error.

III. Time-of-Flight 3D Imaging

Continuous-wave Time-of-Flight imaging is the most common ToF-based depth sensing approach. It works by measuring the phase-difference between continuously emitted light pulses. On each pixel, a photonic mixer devices (PMD) demodulates the incoming signal. The output of the PMD for each pixel is a phase-value p, which corresponds to the correlation between emitted P_e and received light P_r at the distance d_{obj} of the sensed object.

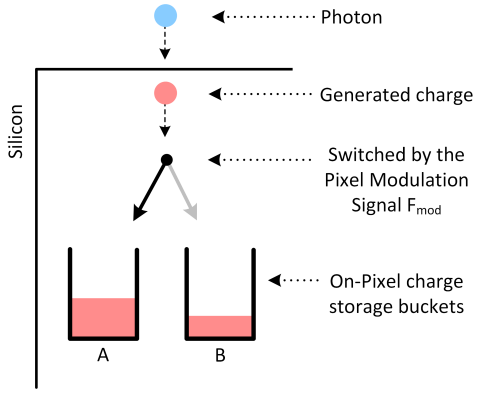


Fig. 2. Operation principle of a photonic mixer device. Signal F_{mod} controls whether charges are stored in A or B.

As shown in Fig. 2 the PMD works by storing charges which are generated by incoming photons into either of two capacitors A or B on each pixel. Each pixel is supplied with the binary reference signal F_{mod}, which triggers whether the charges are transferred to A or B. F_{mod} has the same frequency as the emitted light pulses P_e, but features a configurable phase-offset δ. After the exposure, the charges in A and B on each pixel are digitized in the readout process. The phase-value p results from the subtraction $A - B$.

Besides measurement biases from the sensor, the phase-value p depends on the distance d_{obj} and the amount of received light per pixel. We define the relation between p and d_{obj} as the distance-correlation function $p(d)$, which is sketched in Fig 4. There are several approaches to describe this relationship [2], [6], in practice however, a stochastic simulation is required to model the behavior of the sensor and illumination unit [1]. The figure shows that sensing materials with different reflectivity has a significant impact on the magnitude of the distance-correlation function. This is the major reason why it is difficult to create a depth image from a single phase-image. However if the same material is sensed in different phase-images from the same angle, the phase-values solely depend on the distance. This principle is used in our method to enable 3D object tracking with single phase-images.

In conventional distance sensing, the reflectivity and sensor-specific influences need to be compensated. A common approach is to capture four different phase-images, each with different phase-offset δ between P_r and F_{mod}. It is common to shift the emitted signal P_{Light} by four equidistant phase-offsets (e.g. 0°, 90°, 180°, 270°). If the distance to the object d_{obj} is constant, we can express the relation between measured phase-value p and δ as autocorrelation function $c(\delta)$. $c(\delta)$ is a periodic function, with minima and maxima located between

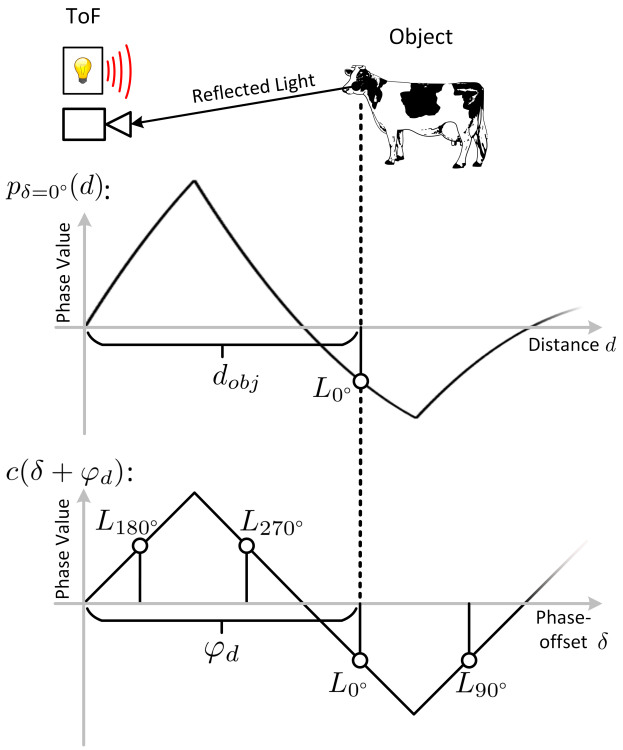


Fig. 3. The established four-phase distance measuring method. The distance d_{obj} is proportional with phase φ_d), which is calculated by four samples L of the autocorrelation function $c(\delta + \varphi_d)$.

the zero-crossings. The shape of $c(\delta)$ depends on the sensor's characteristics and the waveform of the emitted lightpulses.

Fig 3 shows the autocorrelation function in the four phase measurement principle. With $\delta = 0°$, and distance d_{obj}, a phase-value can be seen as a sample $L_{0°}$ of $c(\delta + \varphi_d)$. The goal is to measure the phase of the autocorrelation function φ_d, since it is proportional to the distance d_{obj}. The aforementioned phase-offsets can be used to sample the correlation function, resulting in four samples $L_{0°}$, $L_{90°}$, $L_{180°}$, $L_{270°}$. The common approach is to assume a sinusoidal correlation function, and later correct the deviation by calibration [6]. Four equidistant samples of a sinus function enable to calculate the actual phase-shift φ_{obj} by:

$$\varphi_d = atan2\left(\frac{L_{90°} - L_{270°}}{L_{0°} - L_{180°}}\right) \quad (1)$$

Since φ_d is proportional to the flight time of light, it is possible to calculate the distance to an object d_{obj} by using the speed of light c and lightpulse modulation frequency f_P.

$$d_{obj} = \frac{c}{2} \cdot \frac{\varphi_{obj}}{2\pi f_P} \quad (2)$$

IV. 3D Object Tracking with Single Time-of-Flight Phase-Images

We propose a method to derive the distance between a ToF sensor and an object from a single phase-image in 3D

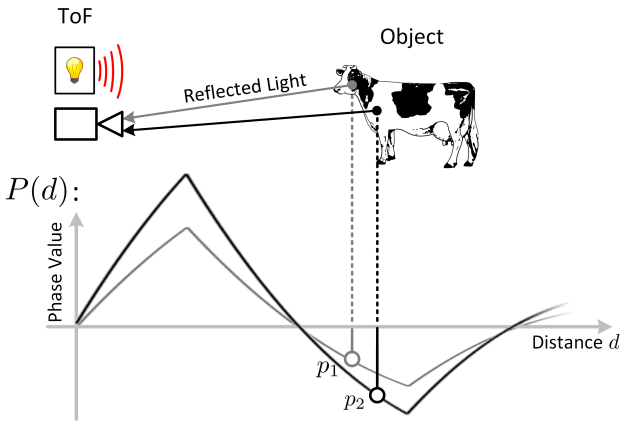


Fig. 4. A ToF sensor captures a phase-value p for each pixel. The value of p depends on the distance, but is also proportional to the amount of received light. Different material reflectivity therefore has an impact on the phase-distance function $P(d)$

object tracking. We obtain phase-images from the Time-of-Flight sensor by subtracting the raw image of the sensor with a dark image. The dark image is a raw image, captured without exposure before the phase-image. Subtracting the dark image compensates certain sensor errors and helps on removing the analog-digital-converter (ADC) offset.

An established approach for 3D tracking is to use a combination of 2D and 3D images to determine the 3D object position. When the 2D object position on the pixel plane is known, it is possible to identify regions with similar light reflectivity on each of the captured phase-images. Our approach is to create a mapping between the average region phase-values and conventionally measured distance in an initialization phase. After the initialization, the distance can be estimated by a lookup operation. By choosing the ideal phase-offset, it is possible to maintain the best accuracy while avoiding distance ambiguity.

A. Combined 2D-3D tracking

Tracking is a correspondence problem and works best with as much available data as possible. Since Time-of-Flight cameras can provide an 2D images along with depth images, a hybrid tracking approach might be the best option. These images are commonly referred to as amplitude images,

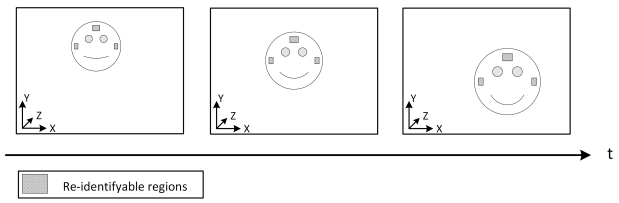


Fig. 5. If certain regions of an object can be identified among subsequent phase-images, the average distance can be estimated from a single phase-image with our approach.

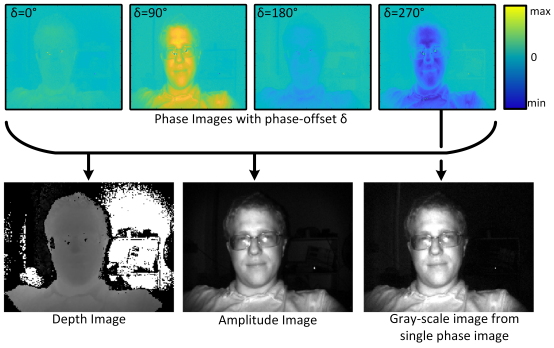


Fig. 6. State-of-the-art close range depth sensing uses at last four phase-images with phase-offset δ to acquire a depth and amplitude image. Phase-images can alternatively be mapped to gray-scale images.

since the amplitude of the autocorrelation function $c(\delta)$ is proportional to the amount of received light [6]. Assuming again a sinusoidal autocorrelation function, the amplitude a can be calculated for each pixel:

$$a = \frac{\sqrt{(L_{270^\circ} - L_{90^\circ})^2 + (L_{180^\circ} - L_{0^\circ})^2}}{2} \tag{3}$$

In our proposed distance estimation method, we are not able to calculate a, because we just use a single phase-image. As illustrated in Fig. 4, phase-values also depend on the amount of received light. Therefore it is possible to map the phase-vales to gray-scale values g for each pixel.

$$g = |p| \cdot scale \tag{4}$$

Fig 6 shows four different phase-images and their resulting amplitude and depth images. It also contains a mapped gray-scale image for comparison. Upon closer inspection, the gray-scale image has a slightly lower dynamic than the phase-image. By looking at phase-image δ=0° and δ=180° in the figure, it becomes apparent that not every phase-image leads to a proper gray-scale image. The reason is that the distance of the face in these images is located at a peak of the phase-distance function. The ideal phase-offset to capture the best gray-scale image also enables the best distance estimation accuracy. In Section IV-C, we propose a method on how to chose the best suitable phase-offset for any distance d.

When an object is located at pixel position $\vec{x}$, the 3D position $\vec{X}$ can be calculated by using the pseudo inverse of camera matrix $\tilde{P}_D$ and distance d_{obj}. In the next Section, we introduce our approach and show how a single phase-image can be used to estimate the distance d_{obj}.

$$\vec{X} = d_{obj} \cdot \frac{\tilde{P} \cdot \vec{x}}{\|\tilde{P} \cdot \vec{x}\|} \tag{5}$$

B. Distance estimation from Single Phase-images

We propose to use 2D tracking methods on gray-scale images to get the 2D object position. This enables to identify regions in the object with a similar phase-value. As seen in Fig. 5, these regions represent the same area on object and

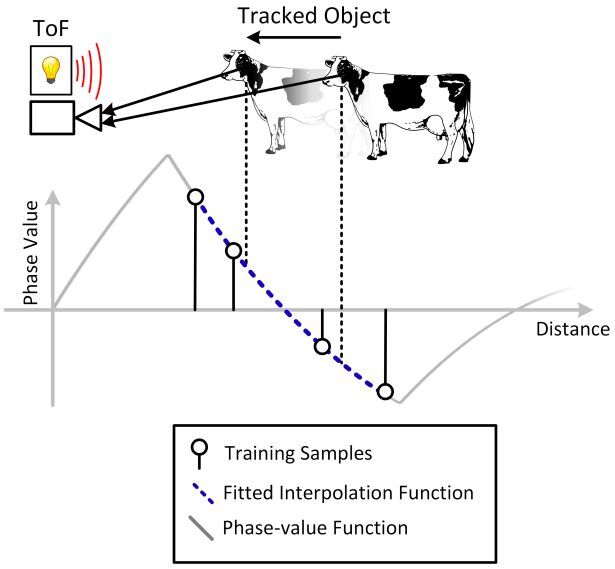


Fig. 7. Phase-images can be used to apply 2D tracking and identify certain regions in each frame.

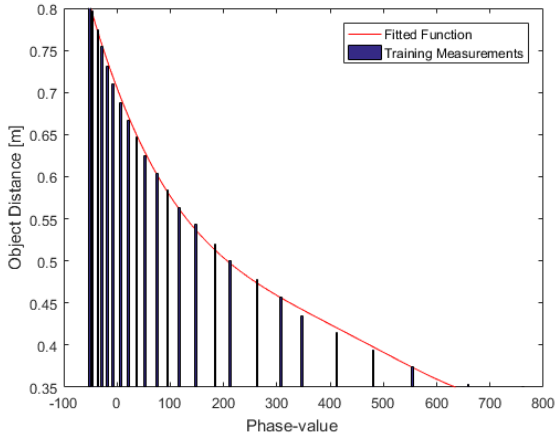


Fig. 8. Training measurements are used to fit a polynomial function, in order to map phase to depth values. This enables to estimate the distance with a single phase-image.

are identified in every phase-image. Since the regions always reflect the same amount of light, the phase-distance function can be used for distance estimation. The first step is to derive a relationship between the region's phase-values and distances by an initialization procedure. Later this mapping can be used to estimate the distance from a single phase-value. Fig. 7 illustrates this principle. Since regions of the object can be re-identified with 2D tracking methods, their average phase-value follows the phase-distance function.

The aim of the initialization procedure is to create a mapping between region phase-values p_r and depth measurements d_r. This can be accomplished by letting the ToF camera operate in normal depth sensing mode for a certain number of images, and using the phase-images to extract phase-values of

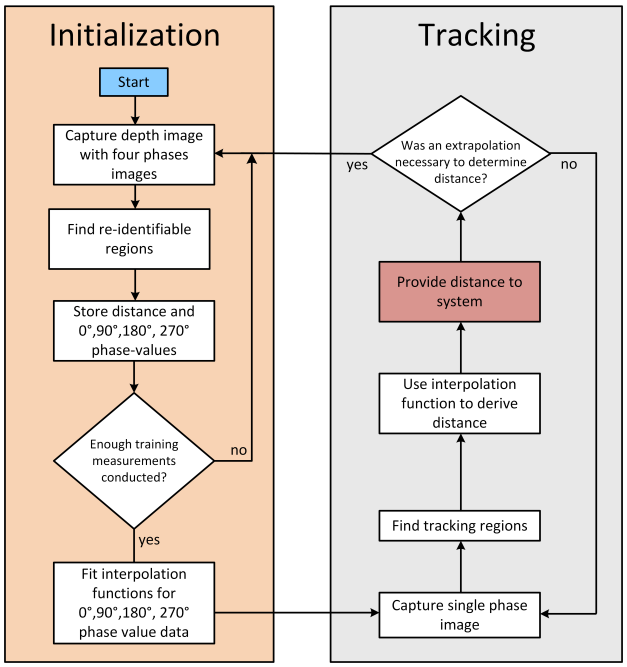


Fig. 9. The algorithm of our proposed Time-of-Flight sensing mode. After an initial training phase, it is possible to estimate the distance to a tracked object solely by capturing single phase-images.

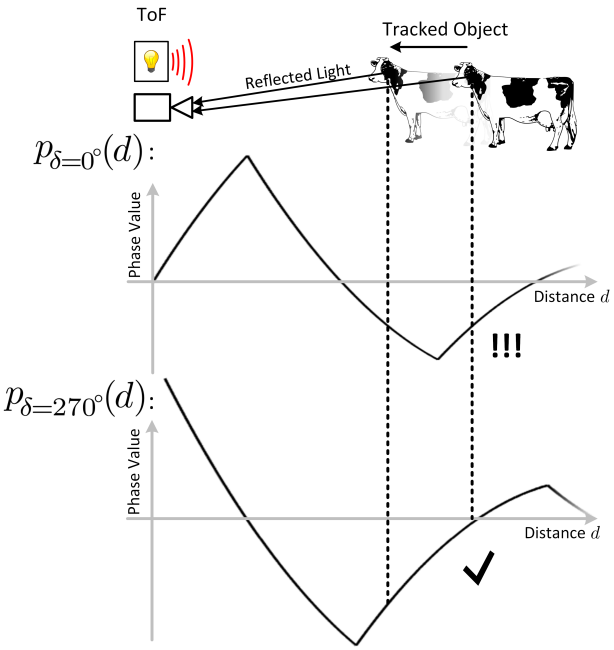


Fig. 10. Example for phase-value ambiguity during tracking. A: When the tracked object moves over the peak between two phase-images, the phase-values are ambiguous. B: This can be avoided by selecting a different phase-offset.

the region for each phase-offset δ. After a certain amount of training data is gathered, a fourth-order polynomial $d(p_r)$ is fitted to the data for each phase offset δ.

$$d(p_r)_\delta = a \cdot p_r^4 + b \cdot p_r^3 + c \cdot p_r^2 + d \cdot p_r + e \qquad (6)$$

These interpolation functions are used estimate the object distance d_{obj} with a single phase-value p with phase-offset δ, using $d(p_r)_\delta$. In Section IV-C, we introduce a method to calculate the ideal δ. Fig. 8 shows an example for $d(p_r)$, fitted with 24 training measurements. Besides simpler interpolation methods, fitting a function has the advantage of mitigating the influence of training data noise. A fitted fourth-order polynomial worked well in our evaluation dataset.

The number of required training measurements depends on the application. It is possible to fit the mapping function with a few training measurements and then conduct new training measurements during tracking. Fig. 9 shows the algorithm of our proposed method. When the object moves out of the trained range, the distance is extrapolated, and new training measurements can be conducted. Adding sporadic training measurements does not decrease tracking framerate if the ToF camera is able to maintain the same framerate with conventional ToF sensing.

C. Phase-offset Adjustment

An important aspect of our tracking method is to select a suitable phase-offset between the emitted pulsed light signal P_e and the reference signal F_{mod}. It is possible for a ToF system to observe the distance of the tracked object and configure the sensor during operation to switch the phase-offset δ. We propose to evaluate after each captured phase-image, whether the current phase-offset is ideal. Due to the widely used four-phase depth calculation algorithm, most ToF sensors offer four equidistant phase-offsets, i.e. 0°, 90°, 180°and 270°. This work therefore uses these particular phase-offsets as example, while different phase-offsets can work as well.

As seen on Fig. 10 A, it is possible that different distances produce the same phase-value p. Between two captured phase-images, the object in the figure moves over the peak of the phase-distance function $p_{\delta=0°}$, leading to ambiguous phase-values. This can be avoided by adjusting the phase-offset δ between emitted light and pixel reference signal. Fig. 10 B shows the adjusted phase-distance function $p_{\delta=270°}$, which does not have a peak in this period.

Another criteria for the ideal phase-offset is the distance precision. Due to the quantization of charge storage A and B, a ToF sensor is most sensitive when producing a phase value at the maximum gradient of the autocorrelation function $c(\varphi)$. As introduced in Section IV, imperfections in the emitted light pulses give the autocorrelation function a sinusoidal shape. As a consequence, the ideal phase-offset is located at a zero-crossings of $c(\varphi)$, featuring the steepest gradient $\frac{dc}{d\varphi}$. The same phase-offset is also ideal to avoid phase-value ambiguity, since it produces phase-values close to the zero-crossings. The offset location between the peaks is also ideal for obtaining the best gray-scale image from the phase-image. Fig. 11 illustrates

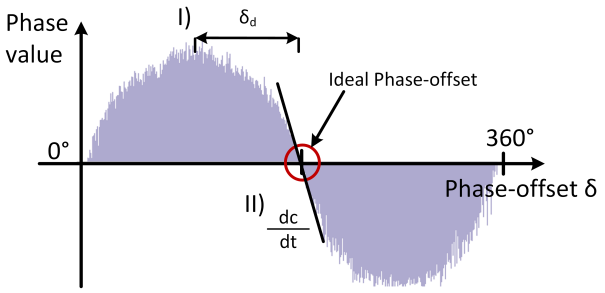


Fig. 11. The criteria for selecting the ideal phase-offset, illustrated with a sampled autocorrelation function. Both criteria point to the same ideal phase-offset. I: The ideal phase-value is in the middle of two maxima. II: The phase-value is located at the largest gradient.

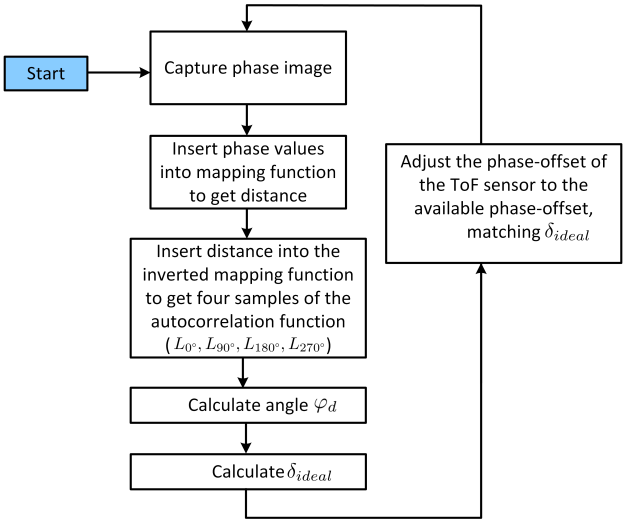


Fig. 12. Algorithm to adjust phase-offset δ during tracking.

these criteria on a measured autocorrelation function. Since each criteria leads to the same ideal phase-offset, there is no tradeoff involved.

The ideal phase-offset δ_{ideal} can be obtained for any distance d_{obj}. If applied to the emitted light signal, it shifts the autocorrelation function $c(\varphi+\delta)$ to a zero-crossing at $\varphi = \varphi_d$, which leads to δ_{ideal}:

$$\delta_{ideal} = |\varphi_d - \pi| \tag{7}$$

In Fig. 12, we provide an algorithm to adjust the phase-offset to δ_{ideal} during tracking. The distance d_{obj} is calculated from a single phase-image with our method. The distance can then be used to obtain the four samples of the correlation function $L_{0°}, L_{90°}, L_{180°}, L_{270°}$, using the training dataset. As stated in Eq. 1, it is possible to calculate φ_d with these four samples. Eq. 7 leads to the ideal phase-offset δ_{ideal}. The ToF sensor can then be configured to a phase-offset matching δ_{ideal}.

D. Discussion and Limitations

In our method, we obtain 2D images from phase-images, since the phase-values are heavily influenced by the amount of received light. These images are suitable for 2D tracking methods and are used in our approach to detect re-identifiable areas with similar phase-offset on the object. We extract the average phase-values from these areas for distance estimation. The quality of our distance estimation method therefore relies on the 2D area-identification method. As Section V shows, basic 2D face-detection is sufficient to reliably detect regions for our distance estimation method. An advantage of using gray-scale images from ToF cameras is that due to the modulated illumination unit, the influence of other lightsources is suppressed. A phase-value p represent the charge difference in bucket A and B. Lightsources such as sunlight, which continuously emit light do therefore not have a direct influence on p, as they fill A and B at the same rate. This means that objects have a distinctive 2D signature in the phase-image which is not influenced by natural or artificial lighting. This can compensate for the rather limited 2D resolution and provides quality input data for machine learning based tracking approaches.

A limitation of this work is that it just conducts positional tracking so far. Object rotations can influence our distance estimation method, or require additional training measurements. In the intended application in smartphones, this problem is however mitigated, since the user is always facing the device and consequently the camera.

V. Results and Evaluation

A. Experiment Setup

We consider facetracking the most significant application of our work. In this Section, we evaluate our method with respect to this domain. This involves tracking a face in the typical smartphone operation range, and using a well-established 2D method in order to extract our re-identifiable regions. To show the feasibility of our approach, we apply our method on a pre-recorded dataset. This enables us to compare the results of our proposed tracking method with different parameters on the same data.

We use the CamBoard pico flexx ToF camera, which is jointly developed by Infineon and pmdtechnologies. The sensor features a resolution of 224x171 pixel and has a specified accuracy of 1% of the measured distance. For each depth image, the camera also provides four phase-images. After the initialization, we use single phase-images for depth estimation, and the associated depth as ground truth.

B. Distance Estimation

The goal of this evaluation is to investigate the precision compared to normal ToF sensing. We record a dataset of 300 measurements, each of them containing four phase-images and a conventional depth image. The exposure time of the ToF camera is set to 0.5 ms and modulation frequency to 80 MHz. The ToF sensor features a framerate of up to 45 fps at this setup. We consider these parameters typical for close range ToF sensing. The measured depth is used for training measurements, but also as ground truth for the distance estimation evaluation. In addition, we capture a dark

image without exposure for each measurement in order to compensate the ADC offset of the correlation function in the raw phase-images.

After subtracting the dark image, we coarsely extract the head position in the gray-scale images using the Viola-Jones detection algorithm [8]. We select the human forehead as tracking region, since the diffuse light reflection of human skin increases the robustness against head rotations. Fig. 13

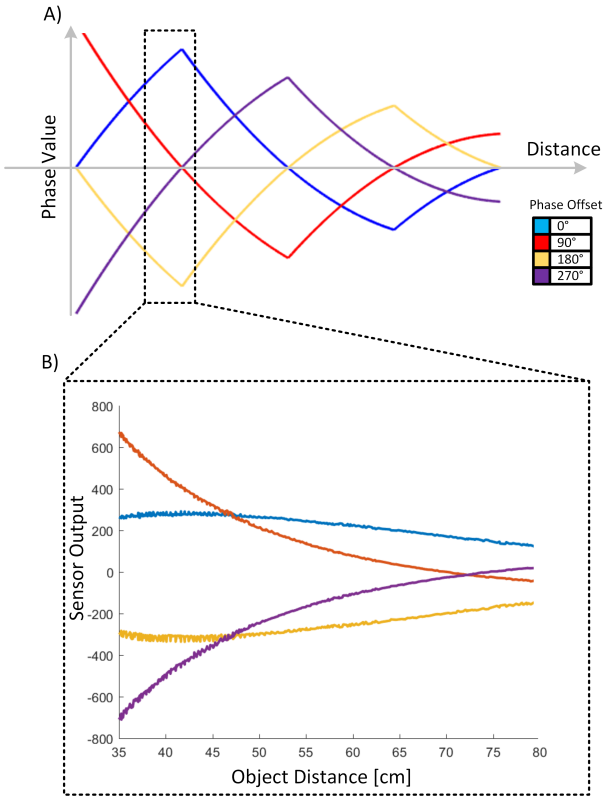


Fig. 13. A: The correlation function between the received and emitted ToF signal for different phase-offsets. B: The phase-values, captured during evaluation. The actual measurement is influenced by the sensor characteristics and intensity of reflected light.

shows the relation between the phase-distance functions for different phase-offsets and the measured phase-values, sorted by associated distance. In the complete tracking range, the ideal phase-offset is 90° or 270°, since our ToF sensor features an additional 90° phase-offset.

In order to test our proposed tracking method, we use a subset of the dataset as initialization measurements in order to fit phase-to-depth mapping functions. This is visualized in Fig. 8 with 24 training samples and a fourth-order polynomial fitted to the phase-values. Analog to our proposed algorithm, these training measurements were taken at approximately equidistant distance intervals.

We use conventional ToF depth images as ground-truth to calculate the distance estimation error. Figure 14 shows the distance estimation error of our method with 24 training

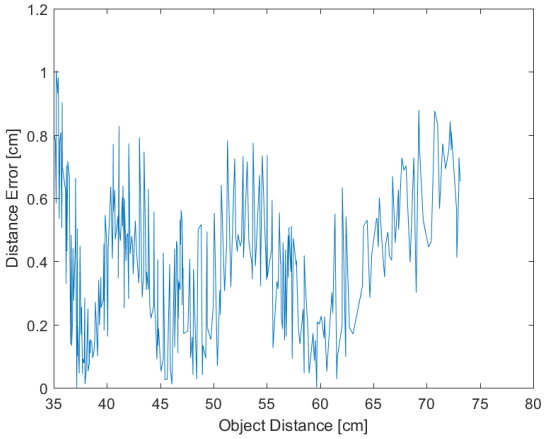


Fig. 14. The absolute distance estimation error of our method with 24 training measurements in relation to object distance. Conventional ToF sensing is used as ground truth.

measurements, sorted by object distance. For the evaluated range of 35 to 75 cm, the maximum error is below 1cm.

Since training measurements during the initialization phase require more energy and time than our method, it is beneficial to reduce them. In Fig. 16, we show the mean and maximum distance estimation error of our method with a varying number of training measurements. The maximum estimation error significantly increases when too few samples are used for fitting. Fig. 15 shows the histogram of the estimation error for 311 estimated distances in the range of 35 to 75 cm. The histogram visualizes that the vast majority of distances can be estimated at an accuracy of below 1 cm with our method.

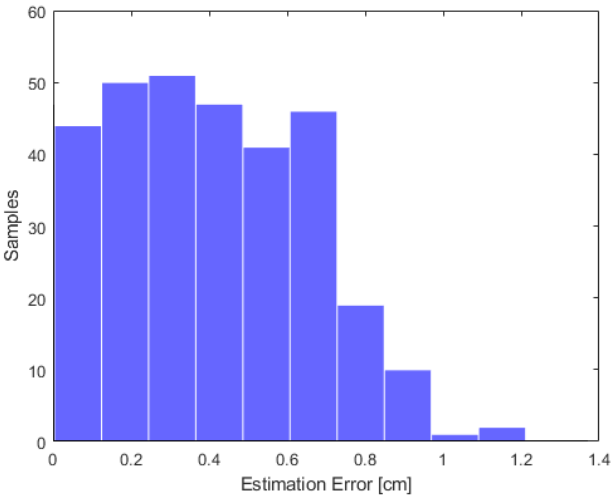


Fig. 15. Histogram of the distance estimation error for 311 samples in the range of 35 to 75cm.

We are satisfied with the results, since our method can enable fast face tracking with up to 68% energy savings with the expense of an estimation error of below 1 cm. We are aware

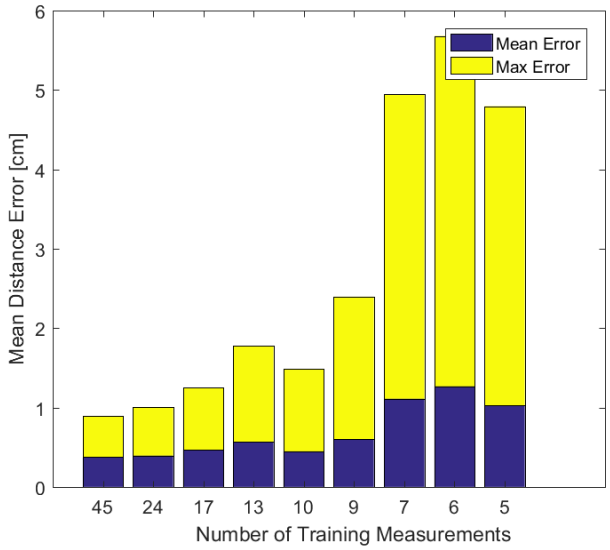


Fig. 16. The absolute distance estimation error with different number of training measurements. More training measurements decrease the error at the expense of framerate and power consumption.

that our method requires a capable 2D tracking subsystem in order to extract re-identifyable regions for depth estimation. However with the ongoing advances in machine-learning based 2D object recognition and segmentation, we are positive that our method will be able to supplement these methods with fast and energy-efficient 3D localization, enabling ubiquitous face and object tracking in mobile devices.

C. Power Consumption Analysis

We evaluate the power consumption by comparing our method after the initialization phase with conventional Time-of-Flight sensing. For a typical close-range object tracking scenario, we analyze an exposure time of 0.5 ms and a framerate of 30 fps. By using a shunt resistor, we analyze the power consumption via oscilloscope, after subtracting idle consumption. This allows us to identify three sensor operation phases with different consumption level. In conventional close-range ToF imaging, the sensor illuminates the scene four times at 1.77 W. For each image, the readout and data transfer process takes 0.68 W and lasts 0.89 ms. For each depth image, a dark-image without illumination is captured, consuming as much power as a single readout process. Our approach also requires this dark-image, but just a single illumination and readout phase. The results are summarized in Table I. For this scenario, our method consumes 68% less power than conventional ToF sensing. In the range of common ToF illumination times from 0.3 ms to 2 ms, our approach saves from 66% up to to 72% power.

	Conventional 4-phase ToF		Our method after initialization	
	Power [mW]	Distribution	Power [mW]	Power Saving
Dark Image	18.1	9.2%	18.1	0.0%
Readout	72.6	36.8%	18.1	75.0%
Illumination	106.5	54.0%	26.6	75.0%
Total	**197.2 mW**		**62.9 mW**	**68.1%**

TABLE I
POWER CONSUMPTION OF 4-PHASE TOF IN COMPARISON TO OUR METHOD. OUR METHOD CONSUMES 68% LESS POWER AFTER THE INITIALIZATION PHASE.

VI. CONCLUSION

In this work, we have introduced a novel approach of conducting distance measurements in 3D object tracking. Our work exploits the fact that phase-images already contain distance information. Other influences are compensated by using training data gathered during an initialization phase. This work enables fast and power-efficient 3D tracking systems, only requiring single phase-images instead of four or more images as in state-of-the-art ToF sensing. This helps important applications especially in mobile devices, because the power consumption is reduced up to 68%. As our evaluation shows, our proposed solution only introduces an additional distance error of less than 1 cm.

In future work, we aim to extend our system to 6-dof tracking, and create a ToF-based 3D tracking system, implementing the proposed depth estimation method.

ACKNOWLEDGMENTS

The authors would like to thank the Austrian Federal Ministry for Transport, Innovation and Technology as well as the Austrian Research Promotion Agency (FFG) for funding the research activities under the grant agreement number 852328.

REFERENCES

[1] J. Gruenwald. Investigation of systematic errors in time-of-flight imaging. Master's thesis, Graz University of Technology, 2013.
[2] M. Gupta, S. K. Nayar, M. B. Hullin, and J. Martin. Phasor imaging: A generalization of correlation-based time-of-flight imaging. *ACM Trans. Graph.*, 34(5):156:1–156:18, Nov. 2015.
[3] F. Heide, W. Heidrich, M. Hullin, and G. Wetzstein. Doppler time-of-flight imaging. *ACM Transactions on Graphics*, 34(4):36:1–36:11, 2015.
[4] S. Hussmann and T. Edeler. Performance improvement of a 3d-tof pmd camera using a pseudo 4-phase shift algorithm. *2009 IEEE Intrumentation and Measurement Technology Conference, I2MTC 2009*, (May):542–546, 2009.
[5] S. Hussmann and A. Hermanski. One-phase algorithm for continuous wave tof machine vision applications. *IEEE Transactions on Instrumentation and Measurement*, 62(5):991–998, May 2013.
[6] R. Lange and P. Seitz. Solid-state time-of-flight range camera. *IEEE Journal of Quantum Electronics*, 37(3):390–397, Mar 2001.
[7] J. Stuehmer, S. Nowozin, A. Fitzgibbon, R. Szeliski, T. Perry, S. Acharya, D. Cremers, and J. Shotton. Model-based tracking at 300hz using raw time-of-flight observations. In *ICCV*, pages 3577–3585, Dec 2015.
[8] P. Viola and M. Jones. Robust real-time object detection. In *International Journal of Computer Vision*, 2001.

Bibliography

[1] H. Plank, A. Schoenlieb, C. Ehrenhoefer, C. Steger, G. Holweg, and N. Druml, "Synchronization of Time-of-Flight 3D sensors for optical communication," in *2017 IEEE International Conference on Communications (ICC)*, pp. 1–6, May 2017.

[2] H. Plank, C. Steger, T. Ruprechter, G. Holweg, and N. Druml, "Survey on camera based communication for location-aware secure authentication and communication," in *EMC2 Summit, CPS Week*, 2016.

[3] H. Plank, C. Steger, G. Holweg, and N. Druml, "Fast and energy-efficient Time-of-Flight distance sensing method for 3D object tracking," in *International Conference on Digital Image Computing: Techniques and Applications (DICTA)*, pp. 210–218, December 2018.

[4] A. Schoenlieb, "Design and implementation of a ToF-based data transmission system," Master's thesis, FH Joanneum Kapfenberg, 2016.

[5] H. Plank, T. Egger, C. Steffan, C. Steger, G. Holweg, and N. Druml, "High-performance indoor positioning and pose estimation with Time-of-Flight 3D imaging," in *2017 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, pp. 1–8, Sep. 2017.

[6] H. Hirschmuller and D. Scharstein, "Evaluation of cost functions for stereo matching," in *Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on*, pp. 1–8, June 2007.

[7] H. Plank, G. Holweg, T. Herndl, and N. Druml, "High performance Time-of-Flight and color sensor fusion with image-guided depth super resolution," in *2016 Design, Automation Test in Europe Conference Exhibition (DATE)*, pp. 1213–1218, March 2016.

[8] T.-H. Do and M. Yoo, "An in-depth survey of visible light communication based positioning systems," *Sensors*, vol. 16, no. 5, 2016.

[9] P. Luo, Z. Ghassemlooy, H. L. Minh, X. Tang, and H. Tsai, "Undersampled phase shift ON-OFF keying for camera communication," in *2014 Sixth International Conference on Wireless Communications and Signal Processing (WCSP)*, pp. 1–6, Oct 2014.

[10] S. Chen and C. Chow, "Color-shift keying and code-division multiple-access transmission for RGB-LED visible light communications using mobile phone camera," *IEEE Photonics Journal*, vol. 6, pp. 1–6, Dec 2014.

[11] S. Nishimoto, T. Nagura, T. Yamazato, T. Yendo, T. Fujii, H. Okada, and S. Arai, "Overlay coding for road-to-vehicle visible light communication using led array and high-speed camera," in *2011 14th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, pp. 1704–1709, Oct 2011.

[12] R. D. Roberts, "A MIMO protocol for camera communications (camcom) using undersampled frequency shift on-off keying (UFSOOK)," in *2013 IEEE Globecom Workshops (GC Wkshps)*, pp. 1052–1057, Dec 2013.

[13] G. Woo, A. Lippman, and R. Raskar, "VRCodes: Unobtrusive and active visual codes for interaction by exploiting rolling shutter," *ISMAR 2012 - 11th IEEE International Symposium on Mixed and Augmented Reality 2012, Science and Technology Papers*, no. November, pp. 59–64, 2012.

[14] T. Nguyen, C. H. Hong, N. T. Le, and Y. M. Jang, "High-speed asynchronous optical camera communication using LED and rolling shutter camera," *International Conference on Ubiquitous and Future Networks, ICUFN*, vol. 2015-Augus, pp. 214–219, 2015.

[15] "Philips Lighting: Indoor positioning." http://www.lighting.philips.com/, 2018.

[16] J. C. Chau and T. D. C. Little, "Analysis of CMOS active pixel sensors as linear shift-invariant receivers," in *2015 IEEE International Conference on Communication Workshop (ICCW)*, pp. 1398–1403, June 2015.

[17] J. C. Chau and T. D. Little, "Scalable imaging VLC receivers with token-based pixel selection for spatial multiplexing," in *Proceedings of the 1st ACM MobiCom Workshop on Visible Light Communication Systems*, VLCS '14, (New York, NY, USA), pp. 21–26, ACM, 2014.

[18] I. Takai, S. Ito, K. Yasutomi, K. Kagawa, M. Andoh, and S. Kawahito, "LED and CMOS image sensor based optical wireless communication system for automotive applications," *IEEE Photonics Journal*, vol. 5, pp. 6801418–6801418, Oct 2013.

[19] Y. Goto, I. Takai, T. Yamazato, H. Okada, T. Fujii, S. Kawahito, S. Arai, T. Yendo, and K. Kamakura, "A new automotive VLC system using optical communication image sensor," *IEEE Photonics Journal*, vol. 8, pp. 1–17, June 2016.

[20] W. Dargie and K. David, "An alternative approach using a photonic mixer device (PMD) as an optical detection for a high speed infrared link," in *International Workshop IOG 2002, Mannheim, Germany*, Sept 2002.

[21] W. Yuan, R. E. Howard, K. J. Dana, R. Raskar, A. Ashok, M. Gruteser, and N. Mandayam, "Phase messaging method for Time-of-Flight cameras," in *2014 IEEE International Conference on Computational Photography (ICCP)*, pp. 1–8, May 2014.

[22] A. Cidronali, S. Maddio, G. Giorgetti, and G. Manes, "Analysis and performance of a smart antenna for 2.45-GHz single-anchor indoor positioning," *IEEE Transactions on Microwave Theory and Techniques*, vol. 58, pp. 21–31, Jan 2010.

[23] M. Rzymowski, P. Woznica, and L. Kulas, "Single-anchor indoor localization using ESPAR antenna," *IEEE Antennas and Wireless Propagation Letters*, vol. 15, pp. 1183–1186, 2016.

[24] B. Grosswindhager, M. Rath, J. Kulmer, S. Hinteregger, M. Bakr, C. A. Boano, K. Witrisal, and K. Römer, "UWB-based single-anchor low-cost indoor localization system," in *Proceedings of the 15th ACM Conference on Embedded Network Sensor Systems*, SenSys '17, (New York, NY, USA), pp. 34:1–34:2, ACM, 2017.

[25] H. Huang, A. Yang, L. Feng, G. Ni, and P. Guo, "Indoor positioning method based on metameric white light sources and subpixels on a color image sensor," *IEEE Photonics Journal*, vol. 8, pp. 1–10, Dec 2016.

[26] S. Liu and T. He, "SmartLight: Light-weight 3D indoor localization using a single LED lamp," in *Proceedings of the 15th ACM Conference on Embedded Network Sensor Systems*, SenSys '17, (New York, NY, USA), pp. 11:1–11:14, ACM, 2017.

[27] W. Roh, J. Seol, J. Park, B. Lee, J. Lee, Y. Kim, J. Cho, K. Cheun, and F. Aryanfar, "Millimeter-wave beamforming as an enabling technology for 5G cellular communications: theoretical feasibility and prototype results," *IEEE Communications Magazine*, vol. 52, pp. 106–113, February 2014.

[28] L. Yang, Y. Chen, X.-Y. Li, C. Xiao, M. Li, and Y. Liu, "Tagoram: Real-time tracking of mobile RFID tags to high precision using COTS devices," in *Proceedings of the 20th Annual International Conference on Mobile Computing and Networking*, MobiCom '14, (New York, NY, USA), pp. 237–248, ACM, 2014.

[29] M. Goller, C. Feichtenhofer, and A. Pinz, "Fusing RFID and computer vision for probabilistic tag localization," in *2014 IEEE International Conference on RFID (IEEE RFID)*, pp. 89–96, April 2014.

[30] C. Duan, X. Rao, L. Yang, and Y. Liu, "Fusing RFID and computer vision for fine-grained object tracking," in *IEEE INFOCOM 2017 - IEEE Conference on Computer Communications*, pp. 1–9, May 2017.

[31] A. P. Sample, C. Macomber, L. Jiang, and J. R. Smith, "Optical localization of passive UHF RFID tags with integrated LEDs," in *2012 IEEE International Conference on RFID (RFID)*, pp. 116–123, April 2012.

[32] A. Ashok, C. Xu, T. Vu, M. Gruteser, R. Howard, Y. Zhang, N. Mandayam, W. Yuan, and K. Dana, "What am i looking at? Low-power radio-optical beacons for in-view recognition on smart-glass," *IEEE Transactions on Mobile Computing*, vol. 15, pp. 3185–3199, Dec 2016.

[33] L. Xie, C. Wang, Y. Bu, J. Sun, Q. Cai, J. Wu, and S. Lu, "TaggedAR: An RFID-based approach for recognition of multiple tagged objects in augmented reality systems," *IEEE Transactions on Mobile Computing*, pp. 1–1, 2018.

[34] J. M. Kahn, R. H. Katz, and K. S. J. Pister, "Next century challenges: Mobile networking for smart dust," in *Proceedings of the 5th Annual ACM/IEEE International*

Conference on Mobile Computing and Networking, MobiCom '99, (New York, NY, USA), pp. 271–278, ACM, 1999.

[35] A. V. H. Ollikkala and A. J. Makynen, "Range imaging using a Time-of-Flight 3D camera and a cooperative object," in *2009 IEEE Instrumentation and Measurement Technology Conference*, pp. 817–821, May 2009.

[36] A. Francillon, B. Danev, and S. Capkun, "Relay attacks on passive keyless entry and start systems in modern cars," in *Network and Distributed System Security Symposium (NDSS)*, February 2011.

[37] S. Drimer and S. J. Murdoch, "Keep your enemies close: Distance bounding against smartcard relay attacks," in *USENIX Security Symposium on USENIX Security Symposium*, SS'07, 2007.

[38] T. Egger, "Design and implementation of an optical communication system between Time-of-Flight cameras and embedded devices," Master's thesis, Graz University of Technology, 2017.

[39] H. Plank, "Design and implementation of a sensor fusion system featuring a Time-of-Flight and color camera," Master's thesis, TU Graz, 2015.

[40] N. Ho, "Finding the optimal rotation and translation between corresponding 3d points," 2013.

[41] S. Kamper, "Optical communication between a Time-of-Flight camera and smart tags." Bachelor's Thesis, 2018.

[42] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, pp. 1330–1334, Nov 2000.

[43] J. Kopf, M. F. Cohen, D. Lischinski, and M. Uyttendaele, "Joint bilateral upsampling," *ACM Transactions on Graphics (TOG)*, vol. 26, July 2007.

[44] L. Dai, F. Zhang, X. Mei, and X. Zhang, "Fast minimax path-based joint depth interpolation," *IEEE Signal Processing Letters*, vol. 22, pp. 623–627, May 2015.

[45] D. Ferstl, C. Reinbacher, R. Ranftl, M. Ruether, and H. Bischof, "Image guided depth upsampling using anisotropic total generalized variation," in *IEEE International Conference on Computer Vision (ICCV)*, pp. 993–1000, December 2013.

[46] "HTC Corporation: VIVE VR-SYSTEM." https://www.vive.com, 2 2019.

[47] K. Römer, "The lighthouse location system for smart dust," in *Proceedings of the 1st International Conference on Mobile Systems, Applications and Services*, MobiSys '03, (New York, NY, USA), pp. 15–30, ACM, 2003.

[48] "Facebook Technologies LLC: Oculus Rift." https://www.oculus.com/, 2 2019.

[49] "Google: ATAP Project Tango." http://www.google.com/atap/projecttango/, 2 2014.

[50] X. Zhang, S. Fronz, and N. Navab, "Visual marker detection and decoding in AR systems: a comparative study," in *Proceedings. International Symposium on Mixed and Augmented Reality*, pp. 97–106, Oct 2002.

[51] S. Tiwari, "An introduction to QR code technology," in *2016 International Conference on Information Technology (ICIT)*, pp. 39–44, Dec 2016.

[52] M. Stafford, A. Rogers, S. Wu, C. Carver, N. S. Artan, and Z. Dong, "TETRIS: Smartphone-to-smartphone screen-based visible light communication," in *2017 IEEE 14th International Conference on Mobile Ad Hoc and Sensor Systems (MASS)*, pp. 570–574, Oct 2017.