# Dissertation

submitted to the

Combined Faculties for the

Natural Sciences and for Mathematics

of the Ruperto-Carola University of Heidelberg, Germany

for the degree of

Doctor of Natural Sciences

put forward by
Diplom-Physiker Mirko Schmidt
born in Berlin

Oral examination: July 20th, 2011

# Analysis, Modeling and Dynamic Optimization of 3D Time-of-Flight Imaging Systems

Referees: Prof. Dr. Bernd Jähne

Prof. Dr. Karl-Heinz Brenner

# Zusammenfassung

Die vorliegende Arbeit befasst sich mit der Optimierung von 3D-Laufzeitkamerasystemen. Diese neuartigen Kameras erfassen Entfernungsbilder, indem sie die beobachtete Szene aktiv beleuchten und die Laufzeit (Time-of-Flight, ToF) des rückgestreuten Lichtes bestimmen. Dabei werden Tiefenbilder aus mehreren Rohbildern konstruiert, wobei typischerweise zwei dieser Bilder simultan mit Hilfe spezieller korrelierender Sensoren aufgenommen werden.

Der wissenschaftliche Beitrag dieser Arbeit setzt sich aus vier Entwicklungen zusammen: Präsentiert wird ein *physikalisches Sensor-Modell*, welches eine Analyse und Optimierung des Prozesses der Rohbildaufnahme ermöglicht. Hierauf gestützt wird ein auf einer logarithmischen Kennlinie beruhendes *ToF Sensor-Design* vorgeschlagen.

Aufgrund von Asymmetrien der beiden parallelen Auslesestufen des Sensors ist gegenwärtig eine mehrfache Akquisition der Rohbilder notwendig. Dies ermöglicht eine Korrektur systematischer Fehler. Die vorliegende Arbeit präsentiert eine Methode zur dynamischen *Kalibrierung und Kompensation* dieser Asymmetrien. Sie erlaubt die Erzeugung von zwei Tiefenkarten aus den ursprünglichen Rohdaten (eines Tiefenbildes), und bewirkt so eine Verdopplung der Bildwiederholrate.

Da mehrere zu unterschiedlichen Zeiten aufgenommene Rohbilder zu einem einzigen Tiefenbild kombiniert werden, treten bei der Abbildung dynamischer Szenerien Bewegungsartefakte auf. Diese Arbeit stellt eine neue, einfache und robuste Methode zur *Detektion und Korrektur* solcher Artefakte vor.

Die in dieser Arbeit präsentierten Algorithmen besitzen eine Berechnungskomplexität, die auch auf Systemen mit limitierten Ressourcen (z.B. eingebetteten Systemen) eine Ausführung in Echtzeit erlaubt. Die Algorithmen werden unter Nutzung eines kommerziellen ToF Systems demonstriert.

# Abstract

The present thesis is concerned with the optimization of 3D Time-of-Flight (ToF) imaging systems. These novel cameras determine range images by actively illuminating a scene and measuring the time until the backscattered light is detected. Depth maps are constructed from multiple raw images. Usually two of such raw images are acquired simultaneously using special correlating sensors.

This thesis covers four main contributions: A *physical sensor model* is presented which enables the analysis and optimization of the process of raw image acquisition. This model supports the proposal of a new *ToF sensor design* which employs a logarithmic photo response.

Due to asymmetries of the two read-out paths current systems need to acquire the raw images in multiple instances. This allows the correction of systematic errors. The present thesis proposes a method for *dynamic calibration and compensation* of these asymmetries. It facilitates the computation of two depth maps from a single set of raw images and thus increases the frame rate by a factor of two.

Since not all required raw images are captured simultaneously motion artifacts can occur. The present thesis proposes a robust method for *detection and correction* of such artifacts.

All proposed algorithms have a computational complexity which allows real-time execution even on systems with limited resources (e.g. embedded systems). The algorithms are demonstrated by use of a commercial ToF camera.

# Acknowledgments

I would like to thank all those who contributed to the success of this work. First and foremost, I want to express my special gratefulness to my thesis supervisor Prof. Dr. Bernd Jähne for offering the possibility to work on such an interesting and up-to-date topic. He gave me the freedom to explore my own ideas and was always open for discussion and supervision of my work.

I thank Prof. Dr. Karl-Heinz Brenner for agreeing to act as the second referee.

I would like to thank my colleagues from Sony Deutschland GmbH. Especially I am indebted to my former supervisor Klaus Zimmermann for the good cooperation, and for moving the project forward so actively. In addition, I thank him for giving me the opportunity to meet so many experts from the companies developing the technology I was working with. I would like to thank Yanagihara-san (Technology Development Group) and Dr. Dietmar Schill for their trust in the project and in my work, and for letting the flow of money not dry out. I gratefully acknowledged the financial support of Sony Deutschland GmbH within the Time-of-Flight project.

I am thankful to the people from Sony EuTEC, especially Markus Kamm, Christian "Bob" Unruh, Zoltan Facius, Marco Hering, Muhammad Atif and Dennis Harres for the lively exchange of ideas and inspiration, and for giving me a fantastic time at the Sony lab.

I enjoyed working at the Heidelberg Collaboratory for Image Processing (HCI) very much. I would like to thank Barbara Werner, Karin Kruljac and Evelyn Wilhelm for the unbureaucratical handling of all administrative matters, and for contributing substantially to the good atmosphere in our institute.

I thank all my colleagues from the HCI lab. Especially I am grateful to Dr. Martin Schmidt for making my start into the topic of Time-of-Flight (ToF) imaging easy, and for helping me uncountable times with very specific programming issues. I thank Dr. Michael Erz, Dr. Daniel Kondermann and Rahul Nair for the good cooperation with countless of fruitful discussions, and for giving me detailed and extremely valuable feedback on a manuscript of this thesis.

I thank Dr. Lena Maier-Hein and Sven Mersmann from the Deutsches Krebsforschungszentrum (DKFZ) for the excellent collaboration, for sharing useful information, and for convincing me that ToF imaging can help save lives.

I would like to thank my friends Anika Paape, Anika Brüning, and my brother Florian Schmidt for proof-reading this work.

I thank all my friend for their words of encouragement and emotional support. Especially I am obliged to Ferdinand Schweser and Sandro Köhler for diverting me from my work, when it was necessary. They helped me not getting lost in my research, and not forgetting what is really important in life.

I am deeply grateful to my family. My parents Kerstin and Michael Schmidt always gave me the freedom to follow my inspiration. Their dedication gave me the opportunities I have today. I am very thankful to my grandparents Elke and Bernhard Lange, and Hanna and Dr. Kurt Schmidt. Their support during the many years of my studies provided the foundation of this work. I also would like to thank my cousin Dr. Steffen Schmidt who vigorously encouraged me to do a doctorate – when I was about 15 years old. My family always believed in me, which I appreciate very much.

Last, but not least, I would like to thank my fiancee Enkhjargal Algaa for her understanding and love. She sacrificed a lot to simplify things for me. Her patience enabled me to complete this work, for which I am very grateful.

Heidelberg, 9. Mai 2011 Mirko Schmidt

# Contents

**1. Introduction** 1

1.1. Motivation 1

1.2. Overview: Depth Imaging Techniques 2

1.2.1. Triangulation 3

1.2.1.1. Stereoscopy (Passive) 3

Minimal Base Stereoscopy 5

Depth-from-X 5

Methods from the Field of Computational Photography 5

1.2.1.2. Active Stereoscopy 6

1.2.2. Interferometry 7

1.2.3. Time-of-Flight Imaging 7

1.2.4. Summary 8

1.3. Outline 8

1.4. Contribution 9

1.5. Notation 10

**2. Time-of-Flight Depth Imaging** 11

2.1. General Principle 11

2.1.1. Temporal Order & Parallel Acquisition 13

2.2. Technical Implementation 14

2.2.1. Pulse-Based Time-of-Flight Systems 14

2.2.2. Continuous-Wave Time-of-Flight Systems 15

2.2.2.1. Correlating Sensors 16

2.2.2.2. Multi-Tap ToF Sensors 18

2.2.3. Comparison and Unification 19

2.3. Difficulties and Shortcomings of Current ToF Systems 20

2.3.1. Basic Difficulties 20

2.3.1.1. Statistical Uncertainty 20

2.3.1.2. Lateral Resolution 21

2.3.1.3. Interfering Ambient Light 21

2.3.1.4. Dynamic Range . . . . . . . . . . . . . . . . . . . . . . 22
2.3.1.5. Systematic Errors . . . . . . . . . . . . . . . . . . . . . 22
2.3.1.6. Non-Ambiguity Range . . . . . . . . . . . . . . . . . . . 23
2.3.2. Errors Caused by Insufficient Sampling . . . . . . . . . . . . 24
2.3.2.1. Flying Pixel . . . . . . . . . . . . . . . . . . . . . . . . 24
2.3.2.2. Motion Artifacts . . . . . . . . . . . . . . . . . . . . . . 24
2.3.2.3. Frame Rate . . . . . . . . . . . . . . . . . . . . . . . . 24
2.3.3. Deviations Related to an Imperfect Propagation of Light . . . 25
2.3.3.1. Scene-induced Interferences . . . . . . . . . . . . . . . . 25
2.3.3.2. Deviations Caused by the Optics . . . . . . . . . . . . 25
2.3.4. Further Deviations . . . . . . . . . . . . . . . . . . . . . . . . 25
2.3.4.1. Interference of Multiple ToF Systems . . . . . . . . . 25
2.4. Summary . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 26

**3. A Physical Model of a ToF Sensor** **27**
3.1. Physical Model . . . . . . . . . . . . . . . . . . . . . . . . . . 27
3.1.1. Motivation and Related Work . . . . . . . . . . . . . . . . . . 29
3.1.2. Assumptions . . . . . . . . . . . . . . . . . . . . . . . . . . . 30
3.1.3. Structure of the Model . . . . . . . . . . . . . . . . . . . . . . 30
3.1.4. Optimizations . . . . . . . . . . . . . . . . . . . . . . . . . . . 32
3.1.5. Suppression of Ambient Light . . . . . . . . . . . . . . . . . . 34
3.1.5.1. Observations . . . . . . . . . . . . . . . . . . . . . . . . 34
3.1.5.2. Modeling the SBI System . . . . . . . . . . . . . . . . . 35
3.2. Parameterization and Experimental Verification of the Model . . . . . 37
3.2.1. Noise Behavior . . . . . . . . . . . . . . . . . . . . . . . . . . 37
3.2.1.1. Method . . . . . . . . . . . . . . . . . . . . . . . . . . . 37
3.2.1.2. Results of Noise Investigation . . . . . . . . . . . . . . 38
3.2.2. Systematic Depth Deviation . . . . . . . . . . . . . . . . . . . 39
3.2.2.1. Setup . . . . . . . . . . . . . . . . . . . . . . . . . . . . 39
3.2.2.2. Results of Investigation of Systematic Deviations . . . 40
3.2.3. Summary . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 41
3.3. Utilization of the Model . . . . . . . . . . . . . . . . . . . . . . 43
3.3.1. Virtual Prototype: Sony Total System Simulator . . . . . . . . 43
3.3.2. Evaluation of Algorithms: Generation of Realistic Data for given Ground Truth Information . . . . . . . . . . . . . . . . . 44
3.3.3. Standardization of ToF Systems: Extension of EMVA 1288 Standard . . . . . . . . . . . . . . . . . . . . . . . . . . . . 47

**4. Investigation of a ToF Sensor with a Nonlinear Photo Response 49**

4.1. Motivation . . . . . 49

4.1.1. Related Work . . . . . 49

4.2. Methodology for Theoretical Investigation of the Phase Estimation Error . . . . . 50

4.3. Impact of a Nonlinear Photo Response . . . . . 51

4.4. Acceptable Limits of Linearity . . . . . 54

4.4.1. EMVA 1288 Linearity Measure . . . . . 54

4.4.2. Evaluation . . . . . 56

4.4.3. Discussion & Conclusion . . . . . 57

4.5. Exploiting a Nonlinear Photo Response: A Logarithmic ToF Sensor . 60

4.5.1. Theoretical Investigation . . . . . 61

4.5.2. Realistic Simulations . . . . . 61

4.5.2.1. Response and Noise of the Logarithmic Sensor . . . . 62

4.5.2.2. Implications on the Dynamic Range and Comparison with Standard Sensors . . . . . 63

4.5.3. Conclusion: Investigation of a Logarithmic ToF Sensor . . . . . 66

4.6. Summary . . . . . 67

**5. Dynamic Sensor Calibration 69**

5.1. Motivation . . . . . 69

5.1.1. Related Work . . . . . 70

5.2. Problem Definition . . . . . 71

5.2.1. Erroneous Measurement Process . . . . . 72

5.2.2. State-of-the-Art: Averaging Technique . . . . . 73

5.3. Calibration and Rectification . . . . . 73

5.3.1. Implicit Calibration . . . . . 74

5.3.2. Dynamic Sensor Calibration . . . . . 75

5.3.3. Raw Data Rectification . . . . . 76

5.3.3.1. Extension: Using Interleaved Datasets . . . . . 77

5.3.3.2. Frame Rate Increase . . . . . 77

5.4. Experimental Verification . . . . . 78

5.4.1. Evaluation . . . . . 82

5.4.1.1. Consistency . . . . . 83

5.4.1.2. Temporal Noise . . . . . 85

5.4.2. Cross Temperature Test . . . . . 86

5.4.2.1. Experimental Setup . . . . . 86

5.4.2.2. Execution . . . . . 87

5.4.2.3. Results . . . . . . . . . . . . . . . . . . . . . . . . . 88
5.4.3. Computational Performance . . . . . . . . . . . . . . . . . . . . 91
5.5. Conclusion and Outlook . . . . . . . . . . . . . . . . . . . . . . . 92

**6. Reduction of Motion Artifacts 95**
6.1. Related Work and Motivation . . . . . . . . . . . . . . . . . . . . 95
6.2. Robust Correction of Motion Artifacts . . . . . . . . . . . . . . . . 97
6.2.1. Origins of Motion Artifacts . . . . . . . . . . . . . . . . . . 97
6.2.2. Detection . . . . . . . . . . . . . . . . . . . . . . . . . . . . 100
6.2.3. Correction . . . . . . . . . . . . . . . . . . . . . . . . . . . 101
6.2.4. A Variant: Burst Internal Detection . . . . . . . . . . . . . . 102
6.2.5. Performance . . . . . . . . . . . . . . . . . . . . . . . . . . 104
6.2.5.1. Geometric Properties . . . . . . . . . . . . . . . . . . 104
6.2.5.2. Computational Performance . . . . . . . . . . . . . . 105
6.2.5.3. Limitations . . . . . . . . . . . . . . . . . . . . . . . 106
6.3. Experimental Results . . . . . . . . . . . . . . . . . . . . . . . . . 106
6.3.1. Controlled Scenario: Rotating Target . . . . . . . . . . . . . 106
6.3.1.1. Quantitative Evaluation . . . . . . . . . . . . . . . . . 108
A Measure for Quantification of Motion Artifacts . . . . 111
6.3.1.2. Implementation of the Measure . . . . . . . . . . . . 111
6.3.1.3. Results and Discussion . . . . . . . . . . . . . . . . . 112
6.3.2. Natural Scenes . . . . . . . . . . . . . . . . . . . . . . . . . 115
6.3.2.1. Scenario 1: Rapid Gestures . . . . . . . . . . . . . . . 117
6.3.2.2. Scenario 2: Juggling Performance . . . . . . . . . . . 119
6.3.2.3. Summary . . . . . . . . . . . . . . . . . . . . . . . . 121
6.4. Conclusion and Outlook . . . . . . . . . . . . . . . . . . . . . . . 121

**7. Conclusion and Outlook 123**
7.1. Summary and Conclusion . . . . . . . . . . . . . . . . . . . . . . 123
7.2. Inference . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 125
7.3. Outlook . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 126

**A. Quadratic Photo Response – Details 129**

**Acronyms and Notation 134**

**List of Publications 135**

**Bibliography 137**

# Chapter 1.

# Introduction

## 1.1. Motivation

A vast number of applications rely on depth maps. Many tasks in the areas of gaming, robotics, automotive, home security etc. are based on range images. 3D Time-of-Flight (ToF) cameras have the potential of efficiently generating such depth images. These cameras employ a modulated light source to actively illuminate an observed scene. The distance is determined by measuring the time it takes for the emitted radiation to travel to an object and back to the camera. Time-of-Flight systems utilize special sensors or shutters in order to perform this measurement simultaneously in each pixel; enabling the generation of dense depth maps.

Within the last years ToF range imaging has become a considerable alternative to traditional techniques: Unlike laser scanners Time-of-Flight systems generate dense depth maps without requiring any moving parts. In contrast to stereoscopic techniques, ToF cameras determine distances by use of very simple computations, requiring only little computational power. Therefore Time-of-Flight depth imaging is the potential to become a cheap yet robust and reliable technique.

Further advantages as its good scaling properties and low price at mass production have drawn the interest of big companies. For instance the car manufacturer AUDI decided to use ToF cameras in series production of its Q7 model. The provided depth information is the essential input of a collision avoidance system.

Another example for the successful usage of range imaging system in a widely distributed product is provided by MICROSOFT: The KINECT device enables to steer a game console by gestures. It facilitates a more direct man-machine interaction and thus an immersive playing experience. Although this system does not utilize the Time-of-Flight principle (yet), ToF could help to decrease the costs of such devices. Since MICROSOFT recently purchased two Time-of-Flight manufacturers, 3DVSYSTEMS and CANESTA, this seems to be an attractive option.

While the Time-of-Flight hardware is getting cheaper and finds its way into first mass market applications, the ToF data processing pipeline is still very simple. Enhancing the processing, however, enables to significantly rise the quality of generated depth data with only little effect on the devices' costs.

The goal of the present thesis is an improvement of this data processing pipeline. A thorough analysis of the complete chain will be performed. It will include all the steps from the acquisition of the individual raw values by a special sensor to the computation of depth information from these data. At various points in this pipeline improvements will be suggested.

The work presented in this thesis was funded by Sony Deutschland GmbH, Stuttgart.

## 1.2. Overview: Depth Imaging Techniques

This work is concerned with Time-of-Flight depth imaging systems. These novel devices provide a possibility to acquire dense depth maps of dynamic scenes.

Such depth images are an essential input for many applications, for instance in the areas of gaming, robotics, automotive, home security, machine vision, biometrics, etc. A depth map is a matrix which comprises a depth information in each entry. In Fig. 1.1 the acquisition of a depth map is visualized schematically. Here, each gray value corresponds to a specific distance.

The acquisition system is capable of determining a depth information ($d$) for objects located on the projection beams within its field of view (FOV). The scene is imaged by an optical system into an image plane and sampled at discrete points given by the pixel positions. If a depth estimation is performed for each pixel the generated depth map is called to be dense (otherwise sparse). By use of two spatial coordinates ($x_1$ and $x_2$) these sampled data may be addressed.

The depth estimation is characterized by a specific lateral resolution ($\Delta x_1$ and $\Delta x_2$) and depth resolution ($\Delta d$). Furthermore, depending on the system implementation, the imaged objects might not be located in arbitrary distances to the acquisition system. Instead, a minimum ($d_{min}$) and maximum ($d_{max}$) distance exist which define the depth dynamics given as $d_{max}/d_{min}$.

To allow a classification of the ToF method this section will briefly introduce some typical optical depth imaging techniques. It will handle stereoscopic and interferometric methods, followed by a very short introduction into Time-of-Flight imaging. A summary and comparative overview on all techniques will be given in Sect. 1.2.4.

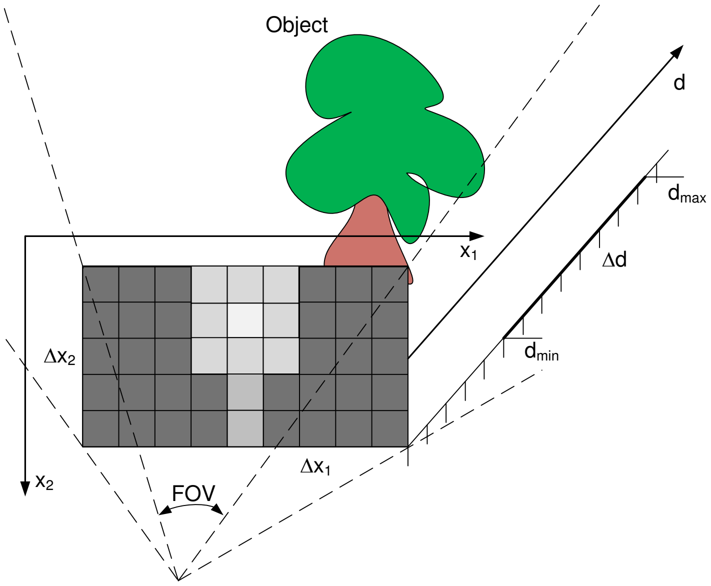


**Figure 1.1.:** Schematic visualization of a depth image and the parameters describing it.

## 1.2.1. Triangulation

Triangulation methods determine the depth of single points of the scene by individually analyzing the relation between two projection rays of two optical systems imaging these specific points. The depth information is extracted from the angles of the triangle formed by the two rays and the baseline connecting both and going through a common image plane. Triangulation methods (also called stereoscopic methods) may be distinguished into passive and active techniques which both will be explained in the following.

### 1.2.1.1. Stereoscopy (Passive)

Passive stereoscopic methods make use of two cameras. A simplified two-dimensional case is visualized in Fig. 1.2. Both cameras are observing the scene and generate images. In the chosen example, these two images are projected into a common plane. A specific point in the scene is identified in the images and its position $x_l$ in the left image and $x_r$ in the right image is used to compute the disparity

$$x_p = x_l + x_r \,. \tag{1.1}$$

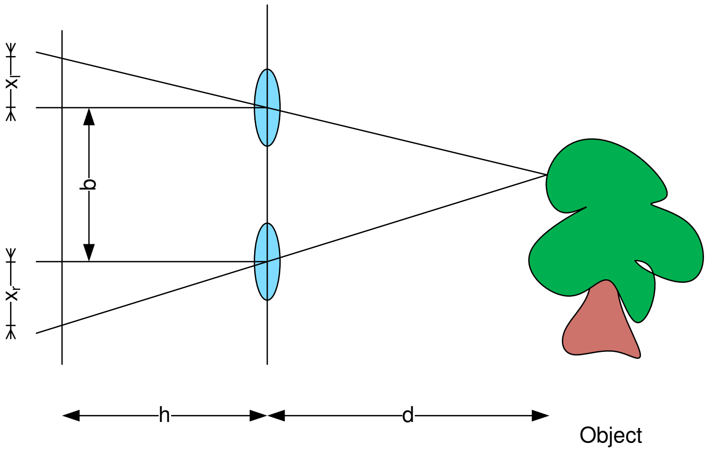


**Figure 1.2.:** Depth estimation using two cameras (passive stereoscopy).

The depth is then determined as

$$d = \frac{b \cdot h}{x_p}\,, \tag{1.2}$$

where $b$ is the length of the baseline and $h$ is the distance between the image plane and the plane through the centers of the central projection of the two cameras.

By using Gaussian error propagation the statistical uncertainty of the generated depth information may be estimated as

$$\Delta d = \frac{d^2}{b \cdot h} \Delta x_p\,. \tag{1.3}$$

Passive stereoscopic methods rely on the identification of corresponding regions in the acquired images. Solving this so called *correspondence problem*, however, is computationally demanding. Therefore today's real-time implementations of passive stereo systems require powerful computers or dedicated hardware [SS02].

Furthermore the identification of corresponding regions requires features to be found in both images which are not distributed densely in natural scenes. Thus, the generation of dense depth maps requires a propagation of the depth information into areas of low confidence. Therefore, depending on the distribution of features the validity of the depth estimation of single pixels may vary significantly, or even get lost completely in some situations [HS09].

A multitude of variants based on the passive stereoscopic approach have been arising, for example the following:

**Minimal Base Stereoscopy** A passive stereoscopic method circumventing the correspondence problem called *Minimal Base Stereoscopy* was proposed by the author in [Sch08b] and filed for patent applications in [Sch08c; Sch09]. This method uses an extremely small stereoscopic base such that corresponding regions of the two images are always imaged by the same pixels.

By pursuing this strategy the problem of finding corresponding regions is transformed into a problem of finding corresponding intensities of single points. This can be computed by evaluating a single expression which significantly reduces the computational complexity compared to the standard approach. However, employing minimal base stereoscopy the depth resolution is determined by the intensity resolution of the used image device. Thus, in order to acquire high quality depth maps cameras with a high resolution of intensity are required.

Furthermore, the method performs best for features with a high contrast which are not guaranteed to be available numerously in natural scenes. Hence, the approach can be said to achieve a simplified and therefore faster processing pipeline by trading precision of the depth estimation.

**Depth-from-X** Many further variants based on or related to passive stereoscopy exist. So called depth-from-motion techniques use video sequences and process the individual images acquired at different times similar to the described classical approach processing images acquired simultaneously at different positions. Examples can be found in publications by Kirchgeßner et al. [KSS00] or Knorr et al. [KKS08].

Especially popular in the field of microscopy is the generation of depth maps by analyzing stacks of images taken while applying different focus settings of the optical system (so called focal series). For instance methods called depth-from-focus or depth-from-defocus are using this principle [DW88; XS93]. Their similarity with stereoscopic methods was described by Schechner et al. [SK98].

The latter methods can be regarded as variants of stereoscopy using a very small baseline (determined by the diameter of the optics). Since the depth precision drops fast for distances which are big compared to the baseline (i.e. $d \gg b$, see (1.3)) these methods deliver high resolution depth maps only for microscopic or macroscopic objects. This also applies to the following methods using Computational Photography.

**Methods from the Field of Computational Photography** Within the last decade a new field of imaging, the so called *Computational Photography* was established. Computational photography combines an adaption of the physical image acquisition process with a sophisticated processing of the digital image, aiming at overcoming the limitations of traditional film cameras and enabling novel imaging applications (c.f. [Ras+06]).

These modifications include an adaption of the optics which brought up depth imaging techniques as for example the plenoptic camera suggested by Ng et al. [Ng+05] and going back to the idea of integral imaging by Lippmann in 1908 [Lip08]. Other methods are using wavefront coding proposed by Dowski and Cathey [DC95] or a special optic's aperture implemented using coded masks [Vee+07].

A unified description of these approaches based on an analysis of the performed light field projections was proposed by Levin et al. [LFD08].

#### 1.2.1.2. Active Stereoscopy

Classical passive stereoscopic approaches are challenged by finding corresponding projection beams by analyzing features in the acquired images. In contrast, active methods (also called *structured light* techniques) determine these correspondences by employing a steerable projection unit. The projector generates a pattern on the imaged surface. Usually a single pattern is not sufficient for an unambiguous determination of the correspondence. By use of time- or color-multiplexing this ambiguity may be eliminated.

By employing many different patterns it is possible to acquire high resolution depth information. So for instance Wiegmann and Kowarschik [WWK06] presented a system for scanning human faces with micrometer resolution.

Another possibility to avoid ambiguities is to incorporate additional spatial information, for example by using pseudo-random patterns and analyzing the local distortion in the detected image. This approach was pursued by the company PRIMESENSE who developed a fast 3D input device which is today very famous as MICROSOFT KINECT. Not many details about the technology were disclosed, but it is known that the KINECT device uses an infrared light source to project a dot pattern onto the scene. A camera positioned in some distance to the projector images the pattern and computes a displacement map using a single dedicated digital signal processor (DSP). The device is able to produce depth maps with 640 × 480 pixels at relatively high frame rates of about 30Hz. However, it has to be mentioned that because of the spatial operations not each pixel comprises an independent depth measurement. Therefore the device's “true spatial resolution” can be assumed to be far below the stated 640 × 480 pixels.

As it can be seen from these examples a multitude of depth imaging techniques based on stereoscopy have been realized. Further implementations make use of multiple cameras and/or multiple projection units, color channels etc. A more detailed description of the methods mentioned here may be found in the books by Jähne et al. [JGH99] or Hartley and Zisserman [HZ00].

### 1.2.2. Interferometry

Interferometric methods are based on indirectly measuring the phase difference of light reflected by an object and light of a reference beam. The depth information can be computed very fast and very precisely. Usually the depth resolution is in the range of the wavelength of the used light, which is about several hundred nanometers. However, the high precision results in the fact that also disturbing influences of the same magnitude have effects on the measurements. This means that for interferometric measurements very much effort has to be put into stabilizing the setup including mechanical and thermal isolation.

Furthermore interferometric methods require the surface of the measured objects to be smooth. In particular the roughness of the objects must be smaller than the wavelength of the utilized light.

For these reasons typical applications of interferometric methods are high precision measurements of small objects in a scientific or industrial environment. For depth map acquisition in a consumer or automotive environment other techniques are better suited because of their greater robustness.

More information about interferometric methods may be found for instance in [Har03].

### 1.2.3. Time-of-Flight Imaging

Time-of-Flight depth imaging techniques use an active light source illuminating the scene discontinuously and measure the time until the backscattered light is detected. Such methods have been used since decades for one-dimensional range measurements (Light Detection And Ranging, LIDAR, see [Sha09]). By varying the direction of the scanning beam the construction of point clouds is possible which enables the generation of depth maps. However, such laser scanners employ moving parts which makes them bulky and mechanically vulnerable.

In the recent years various systems using solid-state matrix detectors have been realized. These cameras enable the instantaneous determination of the time-of-flight for many pixels, and thus facilitate the acquisition of dense depth maps at high frame rates. Therefore they are suitable for imaging dynamic scenes.

Compared to the stereoscopic approach these ToF cameras have the advantage that a simple processing of the raw data delivers a ready to use depth map. ToF imaging generates dense depth maps without requiring object features. Furthermore it does not require a minimum distance between the light source and the camera (like a stereoscopic base) which means ToF systems can be drastically miniaturized. Therefore ToF imaging is a robust and (in mass production) cheap technique which makes it very promising for many applications.

The present thesis will focus on these 3D ToF cameras, so detailed explanations of the technique will be given in the following chapters.

### 1.2.4. Summary

Three principles are mainly used for the optical acquisition of depth images: triangulation, interferometry and Time-of-Flight. Interferometry is unsuitable for acquisitions in an uncontrolled environment because of its sensitivity. Triangulation approaches require high computational power and/or a sophisticated illumination of the scene. In contrast, ToF employs simple computations and easy to implement light sources. Compared to the stereoscopic approach or laser scanners, ToF imaging can be extremely miniaturized and thus has the potential to cheaply generate depth maps for a growing number of applications.

For these reasons the present thesis will focus on Time-of-Flight imaging.

## 1.3. Outline

The goal of this work is to provide a thorough analysis of the state-of-the-art Time-of-Flight depth imaging technology. It will describe various shortcomings and difficulties of current implementations, and suggest possibilities to overcome some of these challenges by optimization of the sensor design or the data processing pipeline. The content of this work is the following

**Chapter 2:** This chapter will explain the principle of ToF depth imaging in detail. It will introduce an abstract formalism for the general description of Time-of-Flight systems. Based on this formalism the pulse-based and continuous-wave ToF approach will be discussed. Furthermore an overview on the difficulties and shortcomings of today's ToF implementations will be given.

**Chapter 3:** A physical model of a ToF system is presented which aims at providing a better understanding of the Time-of-Flight technology. By use of measurements of a real camera this model is parameterized with values providing a physical meaning. Utilizing this parameterization a simulation tool will be derived which reproduces the sensor behavior and generates realistic data.

**Chapter 4:** In this chapter the effect of a nonlinear photo response of ToF sensors will be investigated. An intentionally nonlinear sensor employing a logarithmic characteristic curve will be proposed. A thorough analysis based on the developed physical model will reveal the great potential of this novel type of ToF sensor.

**Chapter 5:** A dynamic calibration method for compensation of the inequalities in the photo response of multi-tap sensors will be suggested. By optimizing the processing pipeline this method allows a drastic increase of the frame rate of today's systems. Using a commercial two-tap system a doubling of the camera's frame rate will be demonstrated, leading to a performance of 60Hz − 80Hz.

**Chapter 6:** This chapter will propose a method for the robust detection and correction of motion artifacts. These artifacts occur if the required raw data are not acquired simultaneously, which is the case in all of today's implementations. The developed method is based on an analysis of the temporal signals of the raw channels of single pixels. It is very simple and hence may be implemented in an extremely efficient manner. By use of a commercial ToF system the applicability of the method even in highly dynamic scenes will be demonstrated.

**Chapter 7:** A detailed summary and conclusion will be given in this chapter. Furthermore an outlook will be provided.

## 1.4. Contribution

The following is a list of novel contributions of this thesis:

- introducing an abstract formalism for description of Time-of-Flight depth imaging systems, enabling an unified explanation of all ToF systems as well as its errors and difficulties, showing that these effects exist in all system implementations

- development of a model of a ToF camera, focusing on the description of a two-tap sensor including a system for suppression of ambient light, whereas the assigned parameters provide a physical meaning, published in [SJ09]

- a careful parameterized simulation tool derived from this model being able to reproduce the sensor behavior and generate realistic data, thus facilitating the evaluation of algorithms working with ToF data, and the development of virtual prototypes, published in [Sch+10] and used in cooperation projects presented in [MH+10b; MH+11]
- thorough investigation of the influence of a nonlinear photo response on the accuracy and precision of the generated depth data
- proposal of a logarithmic ToF sensor which facilitates a high depth dynamic while having systematic and statistical errors comparable to linear sensor implementations
- proposal of a dynamic calibration method, correcting the inequalities of the different taps in multi-tap sensors, enabling a drastically increased frame rate, leading (to the authors knowledge) to the best performance of a today's commercially available ToF system without any need for hardware adaptions, applied for a patent in [SZ10a] and published in [SZJ11]
- proposal of a method for detection and correction of motion artifacts based on an analysis of temporal raw data signals, leading to a simple to implement, computationally cheap, efficient and robust solution, applied for a patent in [SZ10b]

## 1.5. Notation

This thesis explains a lot of effects and algorithms occurring in or working with data of matrix sensors. Therefore many quantities and parameters are described as maps. These maps are represented by matrices, symbolized by italic, bolt letters, for example $\boldsymbol{M}$. Single elements of such a matrix are represented by a lowercase letter, for instance $m$.

The descriptions in this work make use of several indices symbolized by lowercase letters (e.g. $i$). The maximum value of such an index is denominated by the corresponding capitalized letter (for instance $I$).

A detailed list of the used nomenclature is given in the Appendix on page 134.

# Chapter 2.

# Time-of-Flight Depth Imaging

This chapter describes the principle of Time-of-Flight imaging in detail. Two approaches for ToF depth imaging have been implemented: Pulse-based and continuous-wave (phase-based) systems. These two approaches are usually distinguished in the literature and discussed separately. However, the underlying ideas, principles as well as their challenges and shortcomings are very similar, if not identical. Therefore this work seeks to describe and discuss both approaches on an abstract level. Also new proposals for overcoming these limitations will be explained using an abstract description.

For this it is important to introduce some terms and definitions, which will form the basis of the following chapters. Starting with a general description of the principle underlying all ToF systems, new denominations will be explained in Sect. 2.1. This theoretical consideration will be followed by an explanation of technical implementations in Sect. 2.2. It will include a description of the pulse-based and continuous-wave ToF approach, and propose an unification of both. Subsequently a detailed explanation of the shortcomings of current ToF systems, focusing on the statistical errors and artifacts of the generated depth maps will be given in Sect. 2.3. A summary will be provided in Sect. 2.4.

## 2.1. General Principle

3D Time-of-Flight (ToF) cameras acquire depth images by determining the time it takes for emitted light to travel the distance from a source to an object and back to the camera (see Fig. 2.1). The time delay measured $\tau$ is proportional to this distance. Assuming the light source to be located near the camera the object's distance $d$ may be computed as

$$d = \frac{\tau \cdot c_0}{2}\,, \tag{2.1}$$

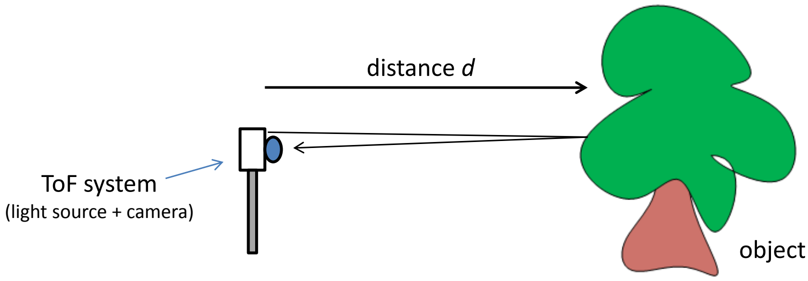


**Figure 2.1.:** Time-of-Flight depth imaging: The ToF system measures the time delay it takes for the light from the source to the object and back. From this time delay the depth $d$ is computed.

with $c_0$ being the speed of light. Time-of-Flight cameras are capable to measure this time delay $\tau$ simultaneously in each pixel which enables the fast generation of dense depth maps $\boldsymbol{D}$.

All ToF systems determine the time delay $\tau$ by measuring the incident irradiance during a single or multiple given time windows. This process is similar to the image capturing process of conventional cameras, which also detect electromagnetic radiation during a specific time window. However, ToF systems have to use windows which are several magnitudes shorter to ensure a sufficient temporal resolution, resulting in an acceptable depth resolution.

The acquired quantities $y$ are called *raw values* (or *samples*). ToF systems are able to measure these raw values in parallel in all pixels, giving a *raw image* $\boldsymbol{Y}$.

These raw data do not directly correspond to depth values, but have to be processed. Therefore, also ToF imaging can be regarded as a form of Computational Photography (c.f. definition given on page 5).

In a typical depth imaging scenario besides the distance $\boldsymbol{D}$ of the object also other quantities are unknown. Especially its reflectivity and the intensity of present non-modulated light may vary. In this work these general unknown quantities will be called *scene unknowns*. All the scene unknowns influence the determined raw values. Therefore the reconstruction of the depth $\boldsymbol{D}$ from a single measurement $\boldsymbol{Y}$ is an underdetermined problem. Thus, multiple raw images acquired under different conditions are required to determine all scene unknowns, including the depth. For an unambiguous determination, the *number of acquired raw images* $R$ must be greater or equal the number of scene unknowns. So, in typical applications $R = 3$ raw images would suffice, but most of today's ToF systems use at least $R = 4$ images (see section Sect. 2.2.2.1).

Regarding single pixels this means generally $R$ *raw channels* are used to compute the values of the *processed channels* which contain information about the scene unknowns.

### 2.1.1. Temporal Order & Parallel Acquisition

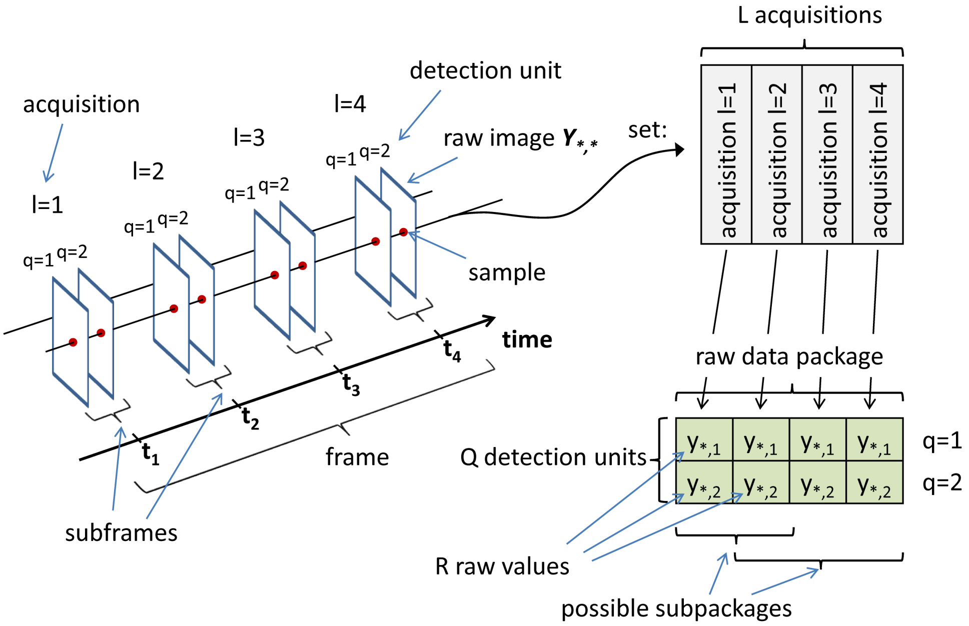


**Figure 2.2.:** Visualization of the given definitions, here depicted for $Q = 2$ detection units performing $L = 4$ acquisitions to gather $R = 8$ raw images. Please note that all raw images of a subframe are acquired simultaneously.

ToF systems need $R$ raw images for computing a set of processed channels. Ideally, these raw images would be taken simultaneously but unfortunately todays ToF systems have not implemented such a fully parallel acquisition due to technical difficulties. Systems do exist, however, which are able to acquire a subset of the required raw images in parallel.

This is done by employing multiple *detection units* per pixel. Each detection unit $q$ is able to measure the incident light in a special measurement mode[1] $n$, giving a

[1] The term "measurement mode" abstractly describes the fact that the incident light is sampled using different states of the ToF system (including the sensor). Depending on the specific implementation, this measurement mode corresponds for instance to determining a certain sample of the correlation function (continuous-wave systems, c.f. Sect. 2.2.2). In another implementation

sample $y_{n,q}$ of the raw image $\boldsymbol{Y}_{n,q}(= \boldsymbol{Y}_r)$. The number of detection units per pixel will be denoted as $Q$, and the total number of measurement modes as $N$. For a visualization of these and the following definitions, please see Fig. 2.2.

As mentioned, all systems currently available use fewer detection units than raw values needed ($Q < R$). So in order to acquire all required raw images, $L$ acquisitions are necessary ($L = R/Q$). Each acquisition $l$ corresponds to an integration of the light incident at the sensor over a certain timespan.

Each pixel collects data building a *set* of $L$ acquisitions. These data are processed in order to estimate the scene unknowns.

Each acquisition $l$ will produce $Q$ raw images which will be denoted to belong to the same *subframe*. The entirety of all subframes (and consequently of all raw images) will be called *raw data frame* (or *frame*).

The samples acquired by a single pixel in one frame constitute a *raw data package*. This raw data package comprises the data of all detection units $Q$ and all acquisitions $L$.

A *subpackage* is each possible subset of the raw data package, comprising data of all detection units $Q$ and consecutive acquisitions $l$.

## 2.2. Technical Implementation

ToF depth sensing is based on measuring the time of flight of light emitted onto a scene and detected back at the camera. Continuous-wave as well as pulse-based ToF systems have been put into practice. Pulse-based ToF systems employ discrete pulses of light and measure its time of flight. Continuous-wave ToF systems use periodically modulated light sources, and determine the phase shift between the incident optical signal (backscattered from the scene) and a reference signal. Both approaches will be discussed and compared in the following sections.

### 2.2.1. Pulse-Based Time-of-Flight Systems

Pulse-based ToF systems use a light source which emits discrete pulses of light. These pulses are backscattered by objects of the scene and detected by the system.

this mode could correspond to measuring the incident light while the active light source is deactivated (typical for pulse-based systems, c.f. Sect. 2.2.1). For the general description given here it is sufficient to understand the different raw images to be sampled in a different manner.

Due to the time of flight the backscattered pulse is delayed. The image sensor is integrating the intensity of the incident light over a certain exposure time. By using an extremely fast shutter it is possible to determine the mismatch between the progress of the incident pulse and the integration window. This facilitates the estimation of the pulse delay from the amount of light detected by the sensor.

Such system can be implemented using a physical shutter in combination with a conventional 2D image sensor. For example 3DVSYSTEMS is using this technique [YIM06]: Their ZCAM system employs a thin GaAs plate attached to a conventional image sensor. The plate can be modulated in transmissivity with frequencies up to 1GHz [3DV09], enabling a very fast shuttering.

Another possibility is to implement fast electronic shutters directly on the imaging device. This approach is taken for example by the company TRIDICAM [Elk+06].

Pulse-based ToF systems usually use a sequence of some tens up to several thousand pulses to acquire the raw images of a frame. This is simply done for increasing the precision of the computed depth map.

For more technical details about pulse-based ToF imaging please refer to the dissertation of Erz [Erz11].

### 2.2.2. Continuous-Wave Time-of-Flight Systems

Utilizing a *continuous-wave*, amplitude-modulated light source the depth can be determined by measuring the phase shift between the emitted and the received optical signal. This *phase-based* approach exploits the fact that the backscattered light is delayed by a time $\tau$ relative to the emitted signal, which results in a phase shift $\varphi$:

$$\varphi = 2\pi \cdot \nu_0 \cdot \tau \,, \tag{2.2}$$

with $\nu_0$ being the modulation frequency of the light source.

Continuous-wave ToF cameras measure this phase shift in each pixel, i.e. these systems are able to acquire phase maps $\boldsymbol{\Phi}$. From this phase map a depth map $\boldsymbol{D}$ can be computed comprising depth estimations

$$d \;=\; \frac{\varphi \cdot c_0}{4 \cdot \pi \cdot \nu_0} \,. \tag{2.3}$$

Continuous-wave ToF systems determine this phase shift by use of correlating sensors.

#### 2.2.2.1. Correlating Sensors

The working principle of digital image sensors could shortly be described as follows: Incident photons generate charge carriers due to the inner photoelectric effect. The generated electrons are accumulated during the exposure time. In a read-out cycle, these electrons are converted into a voltage, amplified, digitized and output as digital values.

To measure the phase shift between the incident optical signal and the electronic reference signal special sensors have been developed. These sensors employ pixels which contain (one or multiple) lock-in detection units. The lock-in mechanism provides to vary the sensitivity of the process of detecting photons over time. This variation of the sensitivity is steered using a reference signal.

Normally, the emitted light signal and the reference signal are periodical, but also systems using non-periodical signals have been demonstrated (see for instance [Büt+07]). The base frequencies of reference signal and light signal are usually set to identical values (homodyne ToF systems), but it should be mentioned that also systems using different frequencies (heterodyne ToF systems, see [Con+06]) have been realized.

By synchronizing the reference signal and the light source signal, the value determined by a single detection unit corresponds to a sample of the cross-correlation function of reference and light source signal. By introducing an additional, controllable phase shift $\theta$ between both signals, it is possible to sample the cross-correlation function at various angles $\theta$.

As an example, the correlation function $c(\theta)$ of a sinusoidal electro-optical signal $S(t)$ with an electronic reference signal $R(t)$, delayed by a phase angle $\theta$, is given by:

$$S(t) = b_{ls} + a_{ls} \sin(2\pi \cdot \nu_0 t - \varphi)\,, \tag{2.4}$$

$$R(t) = H(\sin(2\pi \cdot \nu_0 t + \theta))\,, \tag{2.5}$$

$$\begin{aligned} c(\theta) &= \int\limits_0^{mT_0} S(t)\, R(t)\, dt = \int\limits_0^{mT_0} S(t)\, H\left(\sin(2\pi \cdot \nu_0 t + \theta)\right) \\ &= mT_0 \left( \frac{a_{ls}}{\pi} \cos(\varphi + \theta) + \frac{b_{ls}}{2} \right)\,. \end{aligned} \tag{2.6}$$

Here, $\nu_0$ is the modulation frequency, $T_0$ is the oscillating period and $m$ is the number of integrated oscillating periods (correlation range). $\varphi$ is the phase shift to be estimated, introduced by the delay of the incident light (see (2.2)). $H$ is the Heaviside step function, meaning that $R(t)$ is assumed to be rectangular. The constant $b_{ls}$

describes the offset of the light source, and $a_{ls}$ is its amplitude[2]. The full derivation is available in [Sch08a].

The ToF sensor is able to sample the correlation function by applying various delay angles $\theta$ to the electronic reference signal. Usually, $\tilde{N}$ equidistant sampling points located at the phase angles $\theta_{\tilde{n}} = \tilde{n} \cdot 2\pi/\tilde{N}$ are used to reconstruct the offset $a_0$, amplitude $a_1$ and phase shift $\varphi$ of the electro-optical input[3]. As shown by Xu [Xu99], Plaue [Pla06] and Frank et al. [Fra+09] the optimal solution in a least square sense is given by

$$\begin{aligned} a_0 &= \frac{2}{\tilde{N}} \sum_{\tilde{n}=0}^{\tilde{N}-1} c_{\tilde{n}} \,, && (2.7) \\ a_1 &= \frac{2\pi}{\tilde{N}} \left| \sum_{\tilde{n}=0}^{\tilde{N}-1} c_{\tilde{n}} e^{-\mathrm{i}\theta_{\tilde{n}}} \right| , && (2.8) \\ \varphi &= \arg\left( \sum_{\tilde{n}=0}^{\tilde{N}-1} c_{\tilde{n}} e^{-\mathrm{i}\theta_{\tilde{n}}} \right) , && (2.9) \\ & \text{with } c_{\tilde{n}} = \frac{c(\theta_{\tilde{n}})}{mT_0} \,, \end{aligned}$$

with $\arg(\tilde{z})$ being the argument of the complex expression $\tilde{z}$.

As outlined in Section 2.1, at least $\tilde{N} = 3$ sampling points are required for an unambiguous estimation of these scene unknowns. Most available ToF systems use $\tilde{N} = 4$ samples, because of a better noise performance (see Philip and Carlsson [PC03]) and simpler reconstruction formulas, which are then given as

[2] Please note that $a_{ls}$ and $b_{ls}$ significantly influence the determined values of $a_1$ and $a_0$. However, $a_{ls}$ and $b_{ls}$ are parameters describing the modulation of the light source while $a_1$ and $a_0$ are describing the sampled correlation function. Therefore these parameters are not equal, i.e. $b_{ls} \neq a_0$ and $a_{ls} \neq a_1$.

[3] To clarify the difference between $\tilde{N}$ and $N$ (c.f. Sect. 2.1): $N$ is the number of available measurement modes per detection unit, while $\tilde{N}$ is the number of acquired samples of the correlation function. Each map of samples of the correlation function $c_{\tilde{n}}$ is acquired utilizing a dedicated measurement mode $n$, meaning a mapping between $n$ and $\tilde{n}$ exists which is defined by the chosen indexing. In this work the indexing is chosen to result in a mapping corresponding to the identity transformation, i.e. $n = \tilde{n}$.

$$
\begin{aligned}
a_0 &= \frac{c_0 + c_1 + c_2 + c_3}{2}\,, && (2.10)\\
a_1 &= \frac{\pi}{2}\sqrt{(c_2 - c_0)^2 + (c_3 - c_1)^2}\,, && (2.11)\\
\varphi &= \operatorname{atan}\left(\frac{c_3 - c_1}{c_2 - c_0}\right). && (2.12)
\end{aligned}
$$

#### 2.2.2.2. Multi-Tap ToF Sensors

Firstly proposed in 1995 by Schwarte et al. [Sch+95] and Spirig, Seitz, and Heitger [SSH95] multi-tap ToF sensors have been developed. These sensors use multiple detection units (also called *taps*) per pixel, and thus are able to perform multiple measurements of the correlation function in parallel. Today a two-tap approach is used by many manufacturers.

One sensor developed and used by the company PMD Technologies is the so called *Photonic Mixing Device* (PMD) (see Fig. 2.3). This sensor uses two quantum wells (i.e. two taps) to store the electrons generated by incident photons. The key element is an electronic switch, implemented as a variable electrical field. Incident photons generate electrons which are sorted by this switch into the one or the other quantum well. By synchronizing the switch with the modulated light source, the number of accumulated electrons in each tap corresponds to a sample of the correlation function. Therefore, the sensor is capable of acquiring two samples $c(\theta)$ and $c(\theta + 180°)$ in parallel.

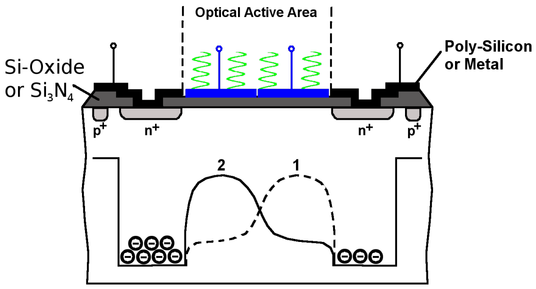


**Figure 2.3.:** Schematic representation of the PMD two-tap ToF sensor. Incident photons generate electrons which are sorted by an electric field into two quantum wells. The switch is synchronized with the modulated light source, thus the number of electrons in each tap corresponds to a sample of the correlation function.

The discrete switching signal is well approximated by a rectangular shaped reference signal. Thus, in combination with a sinusoidal modulation of the light source, the mathematical model given by (2.4) and (2.5) is justified. Hence, the derived correlation function (2.6) and the used reconstruction formulas (2.7)–(2.9) can be assumed to be correct.

Many investigations in this work will focus on this sensor and the ToF camera system CamCube made by PMDTechnologies, because of its relatively open design and processing pipeline. However, it should be noted that currently also other ToF camera manufacturers are using very similar approaches, for example Mesa Imaging [Ogg+04] and Canesta [GYB04]. Therefore, the results derived here for the PMD sensor should be regarded to be valid in a more general sense.

| Manufacturer | $L$ | $Q$ | $R$ | Approach (Official) | Implementation | Source |
|---|---|---|---|---|---|---|
| PMDTechnologies | 4 | 2 | 8 | phase-based | correlating sensor | [RH07] |
| Mesa Imaging (SR3000) | 4 | 1 | 4 | phase-based | correlating sensor | [MES06] |
| Mesa Imaging (SR4000) | 4? | 2 | 8? | phase-based | correlating sensor | [MES11] |
| Optrima / SoftKinetic | ? | ? | ? | phase-based | correlating sensor | [KN05] |
| Canesta / Microsoft | ? | 2 | ? | pulse-based | correlating sensor | [GYB04] |
| TriDiCam | 2 | 2 | 4 | pulse-based | electronic shutter | [Elk+06] |
| 3DVSystems / Microsoft | 3 | 1 | 3 | pulse-based | physical shutter | [YIM06] |

**Table 2.1.:** Overview of today's commercial ToF systems: The table shows how the measurement process of today's commercial ToF systems can be described using the introduced abstract formalism. The number of acquisitions per frame $L$, detection units per pixel $Q$ and raw images per frame $R$ are given. Furthermore it is shown what approach is used by each manufacturer (official term used by manufacturer) and which physical implementation was chosen. Question marks indicate that no solid information has been disclosed.

### 2.2.3. Comparison and Unification

Independent of the usage of discrete pulses or continuously modulated light all ToF systems are based on a convolution of the incident optical signal with a temporal window. Both approaches aim to determine phase maps of scenes, of which besides the depth also other parameters are unknown. Therefore multiple samples have to be taken which is not done simultaneously by today's systems. However, ToF cameras have been realized which perform multiple measurements in parallel. Table 2.1 gives an overview of how the measurement process of today's commercial ToF systems can be described using the abstract formalism introduced in Sect. 2.1, and what implementation they are using. Since not all manufacturers are disclosing the full information some of the table cells do not contain values. Instead, they are labeled with a question mark.

## 2.3. Difficulties and Shortcomings of Current ToF Systems

Depth maps generated by current ToF systems suffer from a variety of difficulties and shortcomings compared to other depth acquisition methods. Simplified overviews on these errors can be found for example in [KBK08; FAT11].

The known limitations may be divided into *basic difficulties* (relating to the ToF technology) (Sect. 2.3.1), errors caused by an *insufficient sampling* (Sect. 2.3.2), deviations caused by a *suboptimal propagation of light* (Sect. 2.3.3), and *further deviations* (Sect. 2.3.4). Each of the following sections will shortly discuss one of these groups of difficulties. This will serve as an overview in order to provide a better understanding and allow a grading of the work presented in the subsequent chapters.

### 2.3.1. Basic Difficulties

Compared to conventional 2D imaging and other depth imaging techniques ToF imaging shows a lot of fundamental difficulties which will be discussed in the following.

#### 2.3.1.1. Statistical Uncertainty

ToF depth imaging is based on the integration of light intensities. The detection of light involves quantum mechanical processes. Especially the generation of charge carriers by incident photons in the image sensor is a Poisson process which always introduces Poisson noise (c.f. Sect. 3.1.4, and [Sei08]). Therefore statistical errors of the depth measurement are inevitable.

Other noise sources like timing inaccuracies of the exposure window, dark currents etc. result in an additional uncertainty of the depth estimation. Therefore the typical depth resolution of today's ToF systems is, compared to other depth imaging techniques (see Section 1.2), relatively low.

#### 2.3.1.2. Lateral Resolution

Most of today's ToF systems use special sensors. Currently the pixels of these sensors are relatively large for two reasons: Firstly, the depth precision increases with an increasing amount of detected light. Therefore each pixel is equipped with a big sensitive area and charge storage unit to ensure that much light is collected during the exposure time. Another reason is the complex pixel electronics: ToF sensors employ pixels which provide additional functionality compared to pixels of conventional image sensor. Therefore ToF cameras use large pixels which results in a low lateral resolution compared to conventional sensors.

However, many depth imaging applications do not necessitate the same lateral resolution as known from 2D imaging. This is because natural scenes do often show only little "depth texture", meaning on object surfaces depth values are normally similar instead of showing large variations. So typical depth images are fairly smooth and thus can be captured well even with low resolution depth cameras.

#### 2.3.1.3. Interfering Ambient Light

Interfering ambient light leads to an earlier saturation of the quantum wells, so less of the backscattered light emitted by the active light source can be detected. This results in a decreased signal-to-noise ratio (SNR) of the raw images and thereby in a worse depth estimation. Manufacturers try to decrease the influence of non-modulated light by various techniques:

A simple method is to use a *burst mode* for driving the active illumination: Instead of operating the (modulated) light source at a constant power level it is run discontinuously. Accordingly, the detector is switched to an operation mode sensing the returning light only when a signal is expected. The idea behind this strategy is that the average power of the light source is kept constant while its peak output may be highly increased. So, restrictions limiting the mean light output (e.g. for keeping the temperature of the light source under a given threshold, or limits originating from eye safety restrictions) can be fulfilled while the ratio of active light to ambient light is extended.

Please note that the "burst mode" is a driving scheme which may be applied independently of the modulation of the light emitter and demodulation of the detection units. It is an operating scheme for activation of the light source and sensor; using time scales which are several magnitudes slower than the modulation of the active light.

Another possibility to reduce the influence of non-modulated light are active circuits which are implemented on the image sensor. These systems seek to separate the modulated light from non-modulated light while or shortly after integrating photons.

Many manufacturers use such systems. For example PMDTechnologies implements a system called Suppression of Background Illumination (SBI) [Möl+05], and Canesta uses a system called Sunshield [BS05; Can08]. Also Mesa Imaging [MES08], SoftKinetic/Optrima [Nie+08] and TriDiCam [Elk+06] are employing compensation techniques. This work will denote such on-sensor implementations as systems for active *Suppression of Ambient Light*.

#### 2.3.1.4. Dynamic Range

ToF imaging is based on analyzing active light backscattered by a scene, where the active light source is normally located near the camera. By doubling the distance to the scene only a quarter of the emitted light is reaching the target because of the distance square law. Additionally, the detector should not saturate in presence of interfering light sources (see Sect. 2.3.1.3). For these reasons ToF cameras need an enormous dynamic range to measure the incident light with sufficient precision while avoiding under- and over-exposures.

#### 2.3.1.5. Systematic Errors

Today's ToF systems suffer from a multitude of systematic errors. Such errors arise whenever the physical implementation does not correspond to the theoretical model assumed to describe the system. Critical components are for instance the modulation of the emitted light and the temporal variation of the sensitivity of the detection units. If these mechanism are not implemented perfectly the derived reconstruction formulas are not correct. This results in an erroneous estimation of depth data from acquired raw images.

For example in case of a continuous-wave ToF system designed to use a sinusoidal modulation $S(t)$ of the light source and rectangle switching function $R(t)$ the reconstruction of the scene unknowns is done using (2.7)–(2.9). If the assumptions are violated, i.e. if the modulation of the light source is not (perfectly) sinusoidal or the switching function is not rectangular, the reconstruction formulas (2.7)–(2.9) are not valid anymore. If these equations are used without additional correction systematic errors occur. A prominent example of an error originating from this mismatch is the so called *wiggling error* (see for instance Rapp [Rap07]). This periodical depth error is caused by higher harmonics in the light source modulation (meaning $S(t)$ is not purely sinusoidal). It will be investigated in Section 3.2.2.

Similarly, pulse-based ToF systems will deliver suboptimal depth estimates, if the *temporal* shape of the light pulse (and/or shutter window) is not implemented optimally.

Also a non-linear photo response of the image sensor might cause deviations of the estimated range information. More details about this effect will be given in Chapter 4.

Furthermore several components of ToF systems are sensitive to temperature variations, such that a drift of the estimated depth with temperature is observed. Usually, also the light source is affected. This is problematic as the light source heats up itself due to the high power. The result is a temporal drift of the estimated depth.

Moreover the optics introduces systematic errors: Besides the geometric distortion which is known from 2D cameras and can be calibrated using the same methods, as presented by Lindner and Kolb [LK07], the optics alter the optical path length of light rays. This is because its average refractive index differs from that of air, resulting in an offset in the estimated depth information.

#### 2.3.1.6. Non-Ambiguity Range

ToF systems using periodically modulated light sources have only a limited non-ambiguity range. This is due to the fact that the depth estimation is based on an estimation of the phase shift between the incident optical signal and a given reference signal. Because of the periodicity of the signals the system is able to determine only the real phase modulo $2\pi$. Thus, for an object causing a phase shift of

$$\varphi_{object} = 2\pi \cdot k + \varphi\,, \text{ with } k \in \mathbb{N}\,,$$

$k$ remains unknown. This means objects being located at distances greater than the non-ambiguity range $d_{amb} = c_0/(2 \cdot \nu_0)$ appear to be in the foreground.

For a typical modulation frequency of $\nu_0 = 20\text{MHz}$ this non-ambiguity range is $d_{amb} = 7.5\text{m}$. However, there are techniques to extend that range, e.g. by combining measurements made with multiple modulation frequencies (e.g. discussed by Gokturk et al. [GYB04; Can08]), or by applying phase unwrapping techniques (investigated by Choi et al. [Cho+10], Droeschel et al. [DHB10], or McClure et al. [McC+10]).

Pulse-based ToF systems show similar effects, if driven at very high repetition rates. (If low repetition rates are used, ambiguity effects can be avoided. However, the system then requires longer exposure times to collect the same amount of light.)

### 2.3.2. Errors Caused by Insufficient Sampling

In contrast to conventional 2D imaging, a spatially or temporally insufficient sampling can cause significant artifacts in ToF imaging systems. Furthermore the required combination of a multitude of samples results in a low frame rate of current implementations. In the following sections these artifacts and effects will be discussed.

#### 2.3.2.1. Flying Pixel

If a depth boundary, for example an edge between a foreground and a background object, is imaged by a single ToF pixel, artifacts occur. The computed depth value then does not belong to one of the imaged surface elements in the fore- or background. It is also not restricted to a value between these distances due to the non-linearity of the reconstruction formulas (see (2.8)–(2.9)). This means the effect does not correspond to the blurring of edges in 2D imaging. Instead, any depth value in the available depth range is possible, depending on all the scene unknowns describing the region imaged by the regarded pixel (e.g. the reflectivity of the involved objects, among others). This effect is known as “flying pixels”.

#### 2.3.2.2. Motion Artifacts

Today’s ToF systems are not able to acquire all raw images simultaneously. If one or multiple of the scene unknowns change during the acquisition of raw images used for computation of one depth map the reconstructed values are incorrect. Incorrect in this sense means that a computed value does not correspond to the state of the scene before nor after the event. Furthermore it is normally not between these values, but lays somewhere in the available range. Therefore this error can significantly decrease the quality of the depth map.

These distortions are called motion artifacts. They are mostly observed on the edges of objects and in fine structures. Chapter 6 will give a very detailed discussion of this effect and propose methods to robustly detect and correct them.

#### 2.3.2.3. Frame Rate

All Time-of-Flight systems have to perform multiple measurements in order to generate a single depth value. A ratio of raw data to processed data of eight to one is not unusual (see Sect. 5.2.2). Compared to a system in which the raw data can be used directly (e.g. in 2D imaging) this results in a reduction of the effective frame rate.

For this reason all ToF systems currently available have frame rates of $30 - 40$Hz at maximum. Many applications, however, require higher frame rates. For example gesture recognition necessitates at least 60Hz.

This work will propose a method which drastically improves the frame rate of current systems by use of a dynamic calibration approach (see Chap. 5).

### 2.3.3. Deviations Related to an Imperfect Propagation of Light

The optimal propagation of light plays an important role for the quality of the constructed depth map, as described in the following.

#### 2.3.3.1. Scene-induced Interferences

ToF depth imaging is based on the assumption that the emitted light follows a straight line to the target and back to the camera. If the light is not going directly but being multiply reflected (or deflected) by any other object, the depth estimation is not correct. These multi reflections lead to various deviations. For example right angular corners imaged from inside appear to be “round”.

#### 2.3.3.2. Deviations Caused by the Optics

Reflections and scatter inside the optics cause a mixing of light backscattered by targets in different distances, which leads to deviations in the depth estimation.

Since in ToF imaging light intensities of very high dynamic range are used (see Sect. 2.3.1.4) even little scatter occurring in optics designed for 2D cameras can cause significant errors. Better results are obtained by utilizing special high dynamic range (HDR) optics, or of course optics optimized for ToF imaging.

### 2.3.4. Further Deviations

#### 2.3.4.1. Interference of Multiple ToF Systems

ToF imaging is an active technique. If multiple identical ToF systems are used to image a single scene these systems may disturb each other. Such errors can be reduced by utilizing different system parameters, for instance different optical wavelengths, slightly different modulation frequencies or pulse repetition rates, respectively. A very robust method is to use different modulation schemes for each participating camera system, as investigated by Büttgen [Büt+07].

## 2.4. Summary

This chapter gave an introduction into the matter of Time-of-Flight depth imaging. After explaining its principle on an abstract level, typical implementations were discussed including the pulse-based and phase-based approach. Furthermore an overview on the difficulties and shortcomings of today's ToF systems was given which are listed here again:

- basic difficulties
    - statistical uncertainty
    - lateral resolution
    - interfering ambient light
    - dynamic range
    - systematic errors
    - non-ambiguity range
- errors caused by insufficient sampling
    - flying pixel
    - motion artifacts
    - frame rate
- deviations related to an imperfect propagation of light
    - scene-induced interferences
    - deviations caused by the optics
- further deviations
    - interference of multiple ToF systems

The next chapter will investigate some of these effects and properties by presenting a physical model of a ToF imaging system.

# Chapter 3.

# A Physical Model of a ToF Sensor

It is a goal of this thesis to provide a better understanding of current ToF systems in order to improve the quality of its data. As outlined in Chapter 2 (in particular in Sections 2.2.2.2 and 2.2.3) special correlating image sensors are widely used in today's implementations. A detailed physical model of such a ToF sensor was developed for the purpose of comprehension of the process of generating depth data with these devices. This model will be presented in Sect. 3.1. Its parameterization enables it to derive a powerful simulation tool which was carefully verified using different experimental scenarios. These aspects will be outlined in Sect. 3.2. Some examples for a successful utilization of the model will be given in Section 3.3.

Parts of the work presented in this chapter were published in [SJ09; Sch+10] and used in cooperation projects presented in [MH+10b; MH+11].

## 3.1. Physical Model

The model aims to build a general framework for description of ToF sensors. Although the goal was to develop a generic model, it was found to be useful to orient its design on a specific ToF system. This helps to provide a clear and comprehensible structure, and a tangible physical meaning of the model parameters.

The specific system chosen here is a PMD CamCube ToF camera by PMDTechnologies. It uses a sensor based on the Photonic Mixing Device which was described in Section 2.2.2.2. This sensor employs two detection units in each pixel and is combined with a continuous-wave light source modulated at $\nu_0 = 20\text{MHz}$.

The sampling of the PMD sensor is well described by the correlation function (2.6) given in Section 2.2.2.2. Using the samples acquired by such a camera system the optimal estimation of the scene unknowns is performed by applying Equations (2.7)–(2.9), which are shown here again to provide a better overview:

$$a_0 = \frac{2}{\tilde{N}} \sum_{\tilde{n}=0}^{\tilde{N}-1} c_{\tilde{n}} , \tag{3.1}$$

$$a_1 = \frac{2\pi}{\tilde{N}} \left| \sum_{\tilde{n}=0}^{\tilde{N}-1} c_{\tilde{n}} e^{-\mathrm{i}\theta_{\tilde{n}}} \right| , \tag{3.2}$$

$$\varphi = \arg \left( \sum_{\tilde{n}=0}^{\tilde{N}-1} c_{\tilde{n}} e^{-\mathrm{i}\theta_{\tilde{n}}} \right) , \tag{3.3}$$

$$\text{with } c_{\tilde{n}} = \frac{c(\theta_{\tilde{n}})}{mT_0} \text{ and } \theta_{\tilde{n}} = \tilde{n} \cdot 2\pi/\tilde{N} .$$

As shown by Lange et al. [Lan00] the variance of these values may be estimated for the special case of $\tilde{N} = 4$ as:

$$\sigma_{a_0}^2 = \frac{\sigma^2}{4} , \tag{3.4}$$

$$\sigma_{a_1}^2 = \frac{\sigma^2}{2} , \tag{3.5}$$

$$\sigma_{\varphi}^2 = \frac{\sigma^2}{2a_1^2} . \tag{3.6}$$

This is derived by assuming an equal variance $\sigma^2$ of all acquired raw values $c_{\tilde{n}}$, and applying Gaussian error propagation.

However, in practice this simplified assumption does not hold since the variance of an acquired sample $c_{\tilde{n}}$ depends on its value. Additionally systematic deviations occur which are caused by a variety of factors, e.g.

- a non sinusoidal light modulation $S(t)$,
- a non rectangle switching function $R(t)$,
- a non-linear photo-response, and
- the influence of on-sensor systems for suppression of ambient light.

Furthermore spatial variations from pixel to pixel like *photo response non-uniformity* (PRNU), *dark signal non-uniformity* (DSNU), and *dark current non-uniformity* (DCNU) (see [EMV10]) must be considered.

### 3.1.1. Motivation and Related Work

To describe these effects a detailed physical model of a ToF sensor is required. Similar models were proposed for 2D sensors, for example by the "Standard 1288 for Measurement and Presentation of Specifications for Machine Vision Sensors and Cameras" [EMV10] (abbreviated as *EMVA 1288 Standard*) presented by the European Machine Vision Association (EMVA).

The goal of the developed model is to simulate the data produced by a ToF camera as realistically as possible. Hence, its thorough parameterization will allow the optimization of present ToF systems as well as the prediction of the characteristics of cameras not existing yet.

Prior models do not include all the effects discussed in the previous section. They rather focused on the simulation of whole 3D scenes. Hasouneh et al. [Has+06] took a MATLAB-based approach where the resulting point cloud of a 3D scene is represented as superposition of single point responses. The influence of an area light source and inhomogeneous illumination of the scene was simulated by Peters et al. [Pet+07]. Keller et al. [Kel+07; KK09] presented a real-time simulation tool for synthetic ToF data. It uses the GPU to generate data of whole 3D scenes, which can be static or moving.

All these approaches focus on the simulation of ToF data for a given 3D scene. This includes issues of rendering, an adequate camera model, reflectance characteristics of the imaged objects and the position and size of the light sources. From the given ideal depth image the simulated samples are generated using a measured correlation function of a real ToF camera. Then, a very simple noise model is employed to simulate the temporal fluctuations of the acquired raw data. From these noisy samples a depth image is computed. Unfortunately the employed noise model is not able to represent the statistical uncertainty of the image formation process in an adequate manner. This makes it hard to use these simulations for verification of ToF algorithms under realistic conditions.

In contrast the work presented here focuses on the effects influencing the quality of the generated depth image and its origins. So, it concentrates on modeling the sensor and its noise sources very carefully. As the focus of this work was put on the internal effects, ideal depth and reflectivity maps were assumed as input, and any issues with real world imaging were neglected.

### 3.1.2. Assumptions

Investigating the errors occurring in a ToF system is not possible by regarding the isolated sensor. Instead, a complete ToF system including a light source, the target response, the image acquisition and analysis (i.e. raw data processing) has to be modeled. The focus of this development lies on the sensor and its noise sources, so questions about the appropriate camera model, the shape and position of the light source and scene-induced interferences like multi-reflections of the active illumination were neglected.

The model does not simulate an area light source but employs a point light source instead (which was shown by Keller [KK09] to be a good approximation). The light source was assumed to be located at exactly the same position as the sensor. The model uses maps containing information about the ideal scene depth, the target's reflectivity and the distribution of interfering ambient light to cover the scene unknowns in a simple way.

### 3.1.3. Structure of the Model

A phase-based ToF measuring setup is a system consisting of a modulated light source, a target which has some effect on the light and a ToF camera which generates data from the detected optical signal.

The model is separated into modules to ensure a high flexibility. Fig. 3.1 depicts the structure and the information flow between the different modules. Each box represents a processing unit. These units have different complexity and may consist of sub units as it is shown for the `target response` and `sampling` module in the figure.

**Excitation.** The excitation module computes the function which represents the optical excitation. Furthermore a synchronization signal is generated which will be used in the sampling module.

**Light source.** The excitation function is converted into a light signal within the light source module. The appropriate physical unit of this signal is “mean number of detectable photons during one time step”, so it corresponds to a temporal density of photons.

**Target response.** The target response module simulates the response of the probe. Parameters like the target's reflectivity are used here and the influence of additional (non-modulated) ambient light is taken into account. Because of the

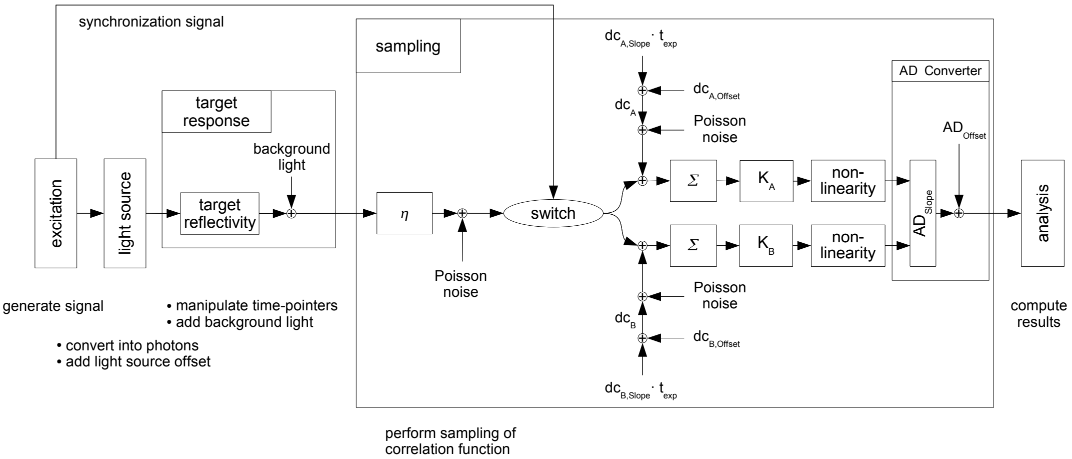


**Figure 3.1.:** Schematic representation of the model: the main modules `excitation`, `light source`, `target response`, `sampling` and `analysis` can be seen.

target's distance from the light source and ToF camera the light signal is being shifted here against the synchronization signal.

**Sampling.** The sampling of the correlation function at different phase angles is performed in the sampling module. Incident photons generate electrons with a given probability $\eta$. This generation is a binomial selection process which follows a Poisson distribution (see Seitz [Sei08]). So, Poisson noise is added here by this quantum mechanical process.

A switch sorts the generated electrons into the two quantum wells `A` and `B`. Then, dark current electrons are added which are also affected by Poisson noise[1]. The sum $\Sigma$ of all collected electrons of the two taps is converted into a voltage by two distinct amplification factors $K_A$ and $K_B$. In each path this voltage is transformed by a nonlinear function which simulates the effect of a nonlinear photo response and pixel saturation. Both resulting voltages are digitized and output as digital numbers which represent the sensor raw data.

**Analysis.** From the determined samples of the correlation function the estimated scene unknowns are computed here, i.e. a phase map $\mathbf{\Phi}$, a derived depth map

[1] Please note that for the desired description of the ToF system as a black box model the exact knowledge about the origin of the noise is not required. Using the black box approach makes it also impossible to determine the contribution of each source unambiguously. Therefore these sources are combined here as dark current noise. For a detailed investigation of all sources contributing to the noise of a ToF sensor please refer to Lange and Seitz [LS01].

$\boldsymbol{D}$, and two maps describing the offset $\boldsymbol{A_0}$ and amplitude $\boldsymbol{A_1}$ of the electro-optical signal.

All modules work on a single vector which contains the signal over time. This signal can be described as a temporal density of detectable photons, but its specific physical meaning slightly changes between the modules.

Since the aim is to model phenomena which are faster than one oscillating period of the light source the temporal sampling density has to be set to a value which is at least one hundred times higher. So, for typical integration times of $10^6$ oscillating periods or more for a single depth image, a vector containing at least $10^8$ entries is generated.

This might be no problem for simulating a single pixel, but the goal here is to model a whole ToF sensor comprising up to millions of pixels with acceptable consumption of computing time and memory. For this reason further optimizations are required.

### 3.1.4. Optimizations

In order to simulate a large number of ToF pixels simultaneously, it is an interesting question which of the discussed operations are pixel-dependent and which are identical for all pixels. Because of its size, the processing of the time-dependent signal vector consumes a lot of computation time and memory. Thus, it is desirable to separate it into a part which is equal for all pixels and a difference term. Since the time-dependent signal is affected by noise and therefore differs randomly from pixel to pixel, this is not trivial. Fortunately, it can be shown that it is possible to separate the noise in an easy way:

The process of adding Poisson noise is a function which generates random numbers which are distributed according to the Poisson distribution with a parameter $\lambda$. The Poisson distribution is given by

$$P_\lambda = \frac{\lambda^k}{k!} e^{-\lambda} . \tag{3.7}$$

The parameter $\lambda$ describes the mean of the values which corresponds to the number of generated electrons here. $P_\lambda$ is the probability of detecting $k$ electrons for a given $\lambda$.

Since the Poisson distribution is reproductive, which means that

$$\begin{aligned} X_1 &\sim P_{\lambda_1} , \\ X_2 &\sim P_{\lambda_2} , \\ \Rightarrow X_1 + X_2 &\sim P_{\lambda_1+\lambda_2} , \end{aligned} \tag{3.8}$$

multiple time steps of collecting electrons for the same tap can be grouped, and the accounting for Poisson noise may be performed only once per group. This “grouping” is exactly what the sorting module does – so it is possible to perform the switching first, and to add the Poisson noise afterwards.

In each tap two Poisson processes take place: the generation of electrons from incident photons as well as the creation of a specific number of dark current electrons. Both quantities are affected by Poisson noise and may be combined in order to speed up the simulation further.

Separating the time-dependent signal from its noise is the key step for optimizing the speed of the simulation: All other pixel-dependent operations can simply be rearranged. This includes for instance the application of multiplicative factors which describe the reflectivity of the target and the quantum efficiency of each pixel. Further examples are additive factors like the amount of incident background light and interfering dark current electrons.

Since it is confirmed that separating the time-dependent signal from its noise is allowed, it is possible to compute the switching function only once and to use these values to simulate all pixels. If the excitation function is periodical and the integration time of a subframe is an integer multiple of an oscillating period, a further increase of speed is achieved by computing the switching function for a single oscillating period only, and multiplying the result by the number of oscillating periods per subframe.

After rearranging the model according to this explanation it looks as shown in Fig. 3.2.

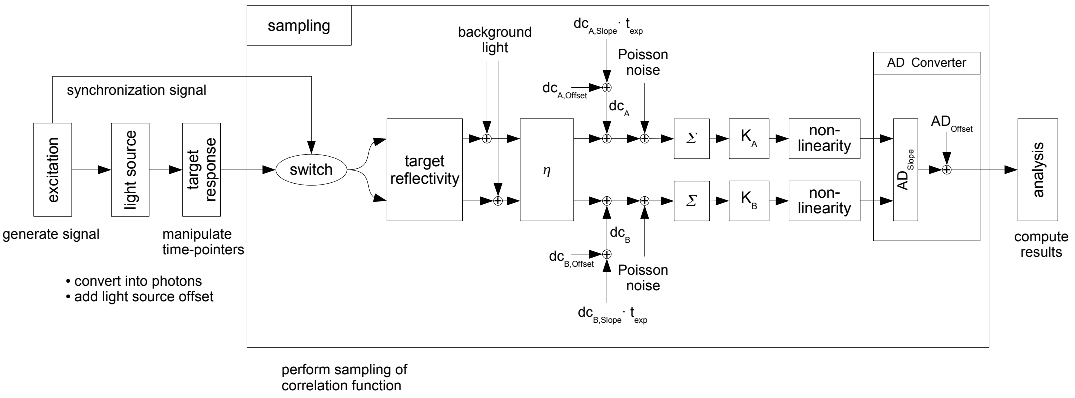


**Figure 3.2.:** Schematic representation of the model after combining Poisson processes.

By performing the sampling operation at four phase angles $\theta = \{0, 90, 180, 270\}^\circ$ of the input signal the acquisitions of four subframes are simulated, with each subframe consisting of two raw images. So, eight raw images are generated which corresponds to the output of a real PMD ToF camera: All phase angles were sampled using both raw channels `A` and `B`. Each pair of raw images corresponding to the same phase angle is summed and divided by two in order to decrease the influence of spatial inhomogeneities of the sensor. This averaging technique will be discussed in detail in Sect. 5.2.2. Equation (2.9) is used to reconstruct a phase image from which the depth image is being computed by using Equation (2.3).

### 3.1.5. Suppression of Ambient Light

Due to its modular structure the model can easily be extended to describe even more complex systems. It is a very interesting question for developers and users of ToF systems how robust the system reacts to interfering non-modulated light. This interfering illumination causes an earlier saturation of the quantum wells, so less of the backscattered active light containing depth information is being detected. This results a decreased precision of the depth estimation (see also Section 2.3.1.3).

Therefore, an interesting task for ToF manufacturers is to design systems which actively decrease the influence of non-modulated light. One system developed by PMDTechnologies is called Suppression of Background Illumination (SBI), which is implemented, inter alia, in the CamCube ToF camera.

The manufacturer did not publish detailed descriptions about the SBI, but it is possible to gather some information by analyzing the data produced by the camera.

#### 3.1.5.1. Observations

When irradiating the camera sensor with increasing intensities and analyzing the acquired intensity values of both channels `A` and `B` of a subframe of a specific angle $\theta$, the following behavior can be observed: For low intensities there is a linear relation between the intensity of the light source and the sensor output. At a particular point one of the channels shows a behavior similar to saturation, i.e. there is almost no variation of the raw data while further increasing the intensity of the light. At the same point the output of the other raw channel starts decreasing while still increasing the irradiation level.

This behavior can be explained as follows: The charge stored in the two quantum wells $\Sigma$ is continuously compared with a reference value $n_{SBI,start}$. As soon as the amount of stored charges of one quantum well exceeds this value, i.e. the difference $n_\Delta$ of both gets positive, a compensation process is triggered. During this process two compensation currents are injected into the quantum wells which contain roughly the same charge as the difference $n_\Delta$. By doing that the quantum well which contained more electrons at the beginning of the process is reset to $n_{SBI,start}$. The other quantum well is set to a value which is below its original value.

The loss of information due to this process is not critical: The most interesting quantity which is reconstructible from the raw data is the phase shift $\varphi$ which gives the depth information $d$. To estimate $\varphi$ only the difference of the two channels `A` and `B` is of importance, not their absolute value (see Equation (2.9)).

#### 3.1.5.2. Modeling the SBI System

These observations were modeled in a separate module and included into the system (see Fig. 3.3): The amount of charge carriers of the two quantum wells $\Sigma$ is continuously read into the SBI circuit. It computes the maximum of both and subtracts a reference value $n_{SBI,start}$. This difference is, if positive, multiplied by a factor $S_{DK}$ and an offset $S_{D0}$ is added. These parameters were introduced to model possible deviations from an ideal system.

The computed and transformed difference value is affected by Poisson noise. It is fed into two paths which generate the compensation currents for the two quantum wells by multiplying with a factor $S_{AK}$ (or $S_{BK}$) and adding an offset $S_{A0}$ ($S_{B0}$, respectively). The generated compensation currents are also affected by Poisson noise, which is considered by the model.

By employing the property of the Poisson distribution of being reproductive (c.f. (3.8)) the model can be optimized regarding the speed and memory consumption of a numerical implementation. This leads to the scheme shown in Fig. 3.4. This model can be simulated much faster because the SBI compensation currents are computed only once per quantum well, just before the read out cycle starts. In a continuous system the quantum well containing more electrons is kept on a constant level, so the additional noise caused by the SBI is canceled out by the controlling loop. This was implemented by setting the quantum well which contained the higher number of electrons at the time of activation of the SBI to $n_{SBI,start}$ at the end of the compensation process.

The model was implemented in HEURISKO, an image processing script language. The simulation of a ToF camera system acquiring one $1000 \times 1000\text{pixel}^2$ depth image using four subframes takes about 10s on a Windows XP Pentium 4 2.80GHz machine. The

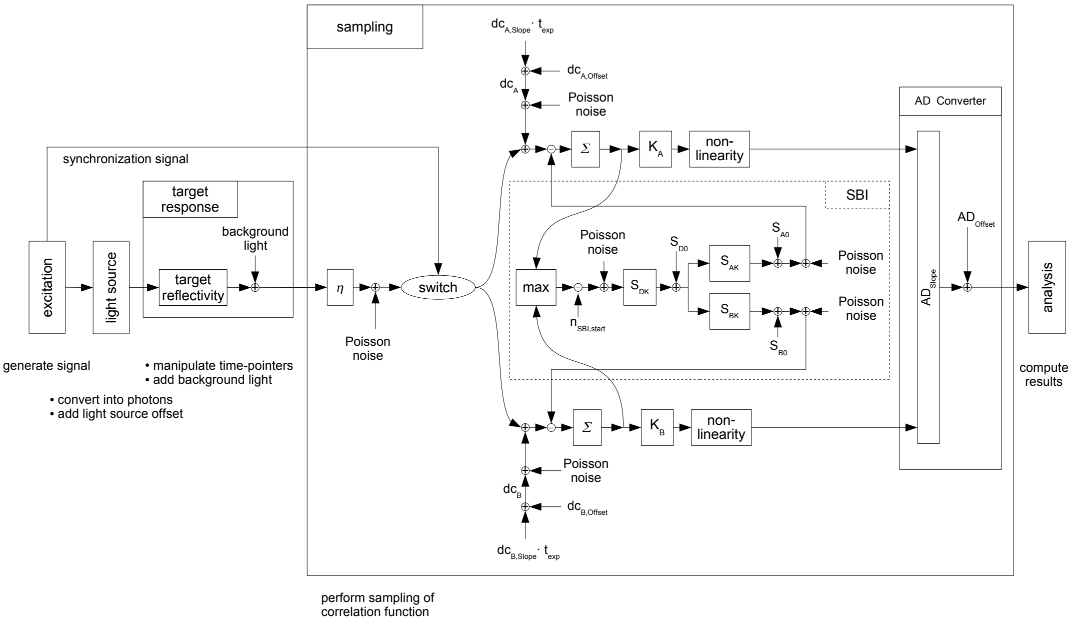


**Figure 3.3.:** Schematic representation of the model, including the SBI circuit.

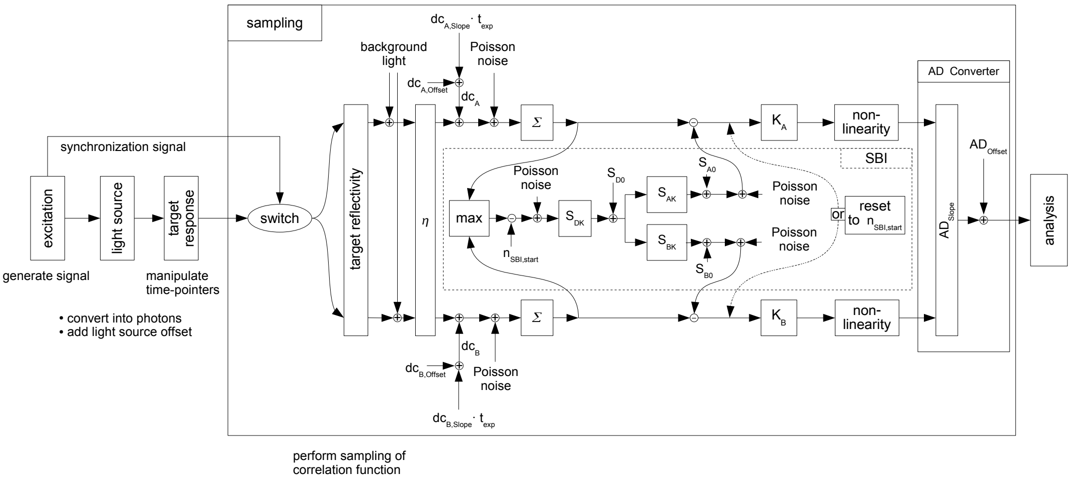


**Figure 3.4.:** Schematic representation of the model, including a SBI circuit, after combining Poisson processes.

source code was not optimized for high speed computation but rather to serve as a flexible framework, enabling a simple implementation of new modules.

## 3.2. Parameterization and Experimental Verification of the Model

The model represents the theoretical framework to describe and understand a given ToF system. To derive a simulation tool which is able to generate realistic data, i.e. to reproduce real camera data, an appropriate parameterization has to be found.

To verify the different aspects of such a parameterization two scenarios were analyzed: In a first investigation the internal sensor parameters were determined and the mean as well as the statistical uncertainty of simulated sensor raw data were compared to that of real sensor data. A second scenario concentrated on the systematic error of the estimated depth, caused by a suboptimal modulation of the light source. Both scenarios will be discussed in detail in the following sections.

### 3.2.1. Noise Behavior

#### 3.2.1.1. Method

To determine a parameterization of the sensor model a setup similar to a radiometric calibration setup for conventional 2D cameras was utilized: A PMD CamCube ToF camera was mounted on an integrating sphere attached to a calibrated light source, allowing to illuminate the sensor homogeneously with variable intensities. The irradiation $H$ was varied while the mean value $\mu_y$ and the variance $\sigma_y^2$ of the raw values of one raw image[2] were observed.

For low intensities the camera behaves like a conventional linear camera, because the SBI is not active. By applying the *photon transfer method* [Jan07] it was possible to determine the quantities $K_A$, $K_B$, and $\eta$. The idea of this technique is to exploit the fact that the number of detected electrons is affected by Poisson noise, which has the property of $\mu = \sigma^2$, meaning the statistical mean of the signal is equal to its variance. So, by analyzing the relation of the known number of incident photons, the generated raw values, and its variance, it was possible to estimate the searched parameters. See [EMV10; EJ09] for further details.

[2] For this experiment the raw channel `A` of the subframe acquired with $\theta = 0°$ was analyzed.

The highest observed mean raw value divided by $K$ gave the parameter $n_{SBI,Start}$. The dark currents $dc_A$ and $dc_B$ and their distribution were estimated from the variance of the dark signal $\sigma^2_{y,0}$ [3]. All other non-uniformities were neglected in this simulation; especially the SBI module was set to ideal parameters.

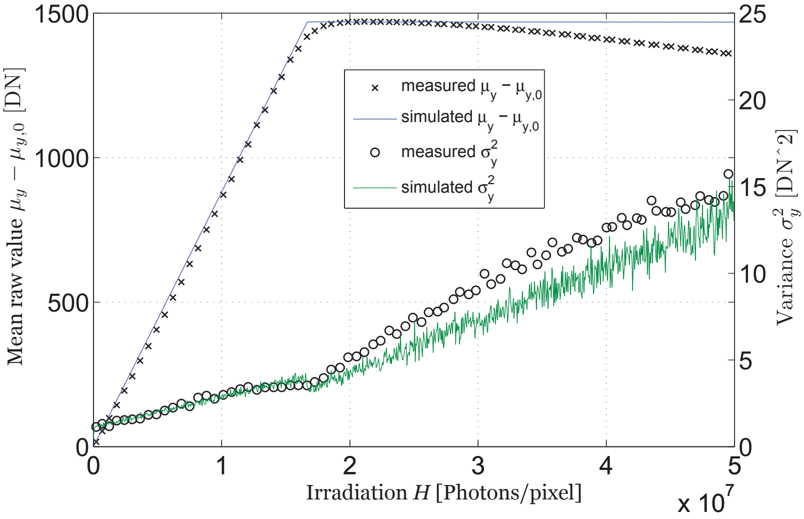


**Figure 3.5.:** Difference of mean raw value and mean dark raw value $\mu_y - \mu_{y,0}$, and variance $\sigma^2_y$ plotted over irradiation $H$. At $H = 1.7 \times 10^7$photons/pixel the SBI is activated.

### 3.2.1.2. Results of Noise Investigation

In Fig. 3.5 the measured difference of the mean raw value and the mean dark raw value $\mu_y - \mu_{y,0}$ was plotted over the irradiation $H$. The measured variance of the raw value $\sigma^2_y$ was plotted as well. Also the computed corresponding quantities as a result of the simulation were plotted in the same figure. It can be seen that the model provides a good reproduction of the observed data.

The results of the simulation and the measured quantities are very similar in the linear range up to an irradiation of $H = 1.7 \times 10^7$photons/pixel. At this point the SBI is activated which causes the sharp bend in the observed and simulated data. With increasing irradiations the model still gives a good approximation of the

[3] The dark currents $dc_A$ and $dc_B$ were set directly in the model; varying the exposure time $t_{exp}$ would have allowed to determine the parameters $dc_{A,\text{offset}}$ and $dc_{A,\text{Slope}}$ (and similarly for raw channel B, see Fig. 3.4) and thus to model the DCNU correctly. However, this was not of interest for this experiment.

real ToF camera, but starts to show slight deviations. The observed variance $\sigma_y^2$ is above the simulated quantity, which was expected because the SBI module was using an ideal parameterization. Please note that even this ideal SBI module introduces additional noise compared to a ToF system without SBI.

## 3.2.2. Systematic Depth Deviation

### 3.2.2.1. Setup

In a second scenario the systematic error of depth data generated by the simulation was investigated and compared to data measured with the real camera. The expected observation was a periodical deviation between the depth estimated by the system and the real depth. This “wiggling” called error is caused by higher harmonics of the optical signal. A theoretical discussion of this phenomenon was given by Rapp [Rap07].

To determine the phase deviation of the real ToF system the camera and a plane target were mounted on movable positioning tables. The light source was detached from the camera and mounted at a fixed position to the target. So, the target’s surface was irradiated from a constant distance while the backscattered light was detected by the ToF camera (see Fig. 3.6). This directly illuminated target acts like a plane emitter which has a constant irradiance independent of its distance. Thus, the acquired depth data does not contain deviations caused by near-field effects of the optical systems (especially of the light source) nor effects caused by a varying amplitude of the optical signal.

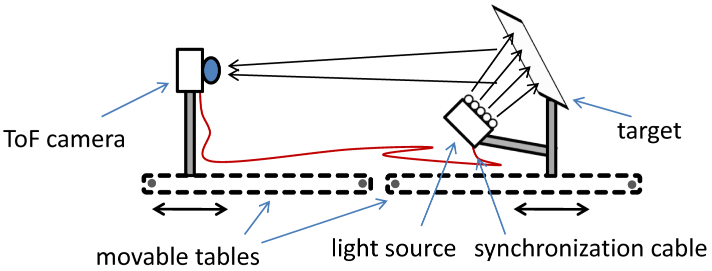


**Figure 3.6.:** Setup for measuring the (isolated) depth dependent error of the depth estimation. The light source is mounted at the target which thus acts as a plane emitter. So, the irradiance is kept on a constant level, which prevents intensity-related errors as well as near-field effects.

The lengthened cable from the camera to the light source introduces an additional but constant offset of the measured phase which can easily be corrected.

A telephoto lens was used to image only a small, homogeneously irradiated area in the middle of the target. The tables were moved to specific positions in order to vary the distance between the active target and the camera, and to analyze the depth estimated using data of some center pixels.

To model the depth error the temporal modulation of the optical signal was measured using a fast photo diode (Femto Photoreceiver HCA-S-400M-SI-FS). The acquired signal was averaged over 16 oscillating periods in order to decrease the noise. In Fig. 3.7 the measured modulation of the light source is plotted. This real shape was integrated into the model and the simulation was run using a varying distance between target and camera, i.e. varying phase shifts.

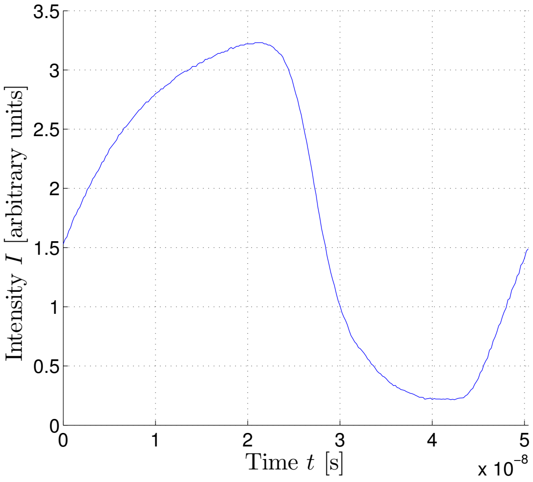


**Figure 3.7.:** Modulation of the PMD light source: The intensity $I$ is plotted over the time $t$ for one oscillating period.

#### 3.2.2.2. Results of Investigation of Systematic Deviations

Figure 3.8 shows the measured and simulated depth deviations over the real depth $d_{real}$. The relation between the real depth and the chosen distance $d_{distance}$ between camera and target is given as

$$d_{real} = 2 \cdot d_{distance} + d_{real,0} \, . \tag{3.9}$$

Because of the detached light source, the light has to travel the distance between target and camera only once, which explains the factor 2 in (3.9). The distance offset $d_{real,0}$, which results from the lengthened cable and some camera internal delays of the signal, is unknown and unimportant for this investigation. The depth data delivered by the camera in an area of 10 × 10pixel$^2$ near the optical axis was averaged and used as “measured depth data”.

The measured depth deviation has a periodical structure with a wavelength of a quarter of the non-ambiguity range, i.e. $c_0/(8 \cdot \nu_0) \approx 1.87$m. Since $d_{real,0}$ is unknown it was set to a value which fits best to the simulated data. From Fig. 3.8 it can be seen that the model generated a very well reproduction of the measured deviation: The wavelength and amplitude of measured and simulated depth deviation are in very good agreement.

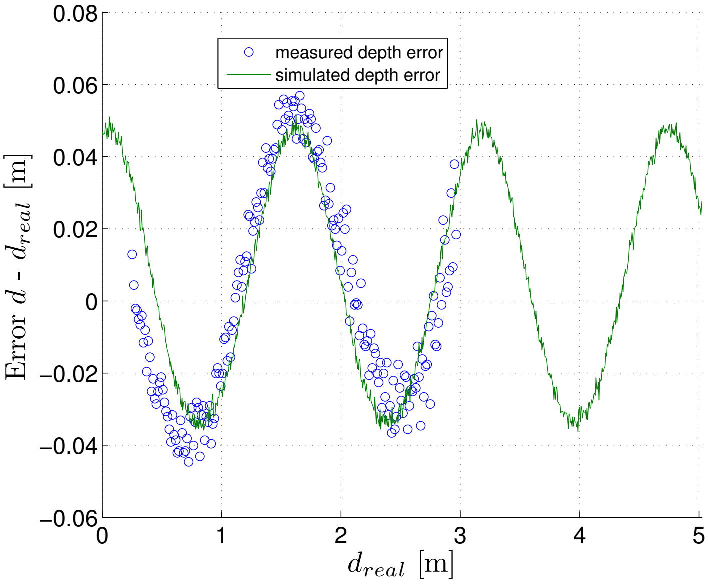


**Figure 3.8.:** Mean depth deviation of the simulated and measured distance from the real distance, plotted over the real distance $d_{real}$.

### 3.2.3. Summary

The developed system is a physical model of ToF cameras with a clear focus on the sensor. It offers a very high flexibility due to its modular structure.

An arbitrary optical excitation may be used to simulate the sampling of a target response by a ToF sensor. The system is able to simulate two detection units per pixel, which can use any function as switching function. All spatial parameters like the reflectivity of the target seen by a single pixel, the local amount of background light, or the quantum efficiency $\eta$ are treated as maps and may be specified for each pixel individually. Additionally, a special module was integrated which simulates an on-sensor circuit for suppressing ambient light. The simulation of sensor data runs at low computational effort.

The derived simulation was parameterized using measurements of a PMD CamCube ToF camera. This camera implements a continuous-wave approach by using a two-tap correlating sensor. As a verification two scenarios were analyzed: The camera's response to an increasing, homogeneous irradiation as well as the systematic phase deviation caused by higher harmonics of the optical excitation. In both scenarios the model gave a precise reproduction of the observed data.

To summarize, the model is currently able to reproduce all properties and shortcomings directly related to the technology or implementation of today's ToF systems. These were discussed as "basic difficulties" in Sect. 2.3.1 and consist of:

- statistical uncertainty,
- limited lateral resolution,
- influence of interfering ambient light (and on-sensor systems to compensate for it),
- a need for a high dynamic range,
- systematic errors, and
- a limited non-ambiguity range.

The well parameterized simulation hence enables the generation of realistic ToF data, as acquired with a real PMD sensor. Therefore, it is a powerful tool to evaluate the performance of algorithms working with ToF data and to estimate the limits of current ToF systems. Examples for the utilization of the model will be given in Section 3.3 and Chapter 4.

The design of the model was focusing on two-tap continuous-wave cameras, but in principle every ToF system using no more than two detection units per pixel may be simulated. So for instance, also the behavior of pulse-based systems using only one tap (e.g. the ZCam by 3DVSystems) may be reproduced by adapting the switching function and modifying the analysis module. However, in this case some of the model parameters might change or even loose their physical meaning.

## 3.3. Utilization of the Model

The developed system is a generic model of ToF cameras which is useful for different kinds of investigations. The following sections will show some possible applications of the system by shortly introducing projects which are utilizing the ToF model and/or the simulation developed here. Each of the following projects was joint work in collaboration with other research groups. Therefore, they will not be described in detail; the description will rather focus on the ToF model and explain how it contributed to the project.

### 3.3.1. Virtual Prototype: Sony Total System Simulator

The modern development process of a camera for use in a consumer systems is an extremely complex procedure. Especially the design of the optics and image sensor are crucial. Since technology evolves rapidly and production cycles are shortening, a parallel design of all system components is desired. This requires to test how different components interact, even if they do not exist yet. Therefore it is a current trend to use more and more simulations to predict the properties of each component, and so of the whole future system. The goal of this trend is to develop a complete system as a *virtual prototype* which is solely based on simulations.

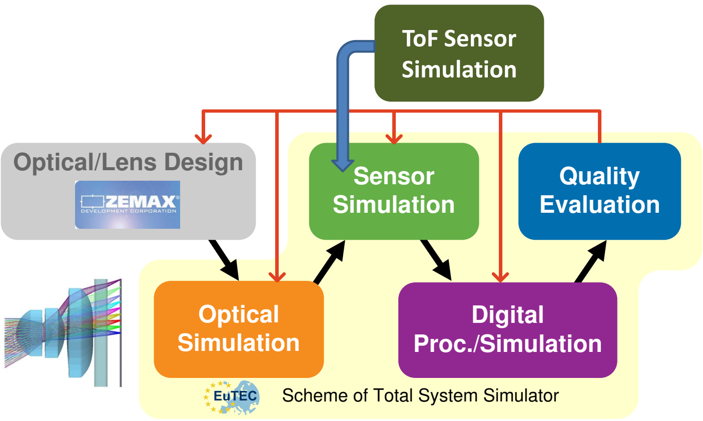


**Figure 3.9.:** Structure of the SONY TOTAL SYSTEM SIMULATOR. The developed ToF sensor simulation was adapted to replace the current sensor simulation module, hence the development of a virtual prototype of a new ToF system gets feasible.

A framework for the development of such virtual prototypes of camera systems is the TOTAL SYSTEM SIMULATOR (TSS) developed at SONY DEUTSCHLAND. The framework is implemented in MATLAB. It is designed as a chain of modules which enable a realistic simulation of a complete camera (cf. Fig. 3.9). Based on an interface to optical design tools (e.g. Zemax) it provides an efficient simulation of the camera optics. A further module emulates the sampling process by the image sensor, followed by a module which simulates the digital processing of the data. A final module is used to evaluate the quality of the simulation result, and to optimize the parameters of each of the preceding modules.

This system thus enables the optimization of hundreds of design parameters. Among other things it allows a co-design of the camera optics and processing, so it facilitates feasibility studies of completely new camera concepts.

The ToF sensor model presented here was fused into the TSS. For this its source code was ported from heurisko to MATLAB and integrated into the TSS architecture as a replacement of the current sensor module. Currently, a module for realistic simulation of the active illumination is being devised. When finished, this extended TSS will provide the development of new ToF camera systems by enabling virtual ToF prototypes.

### 3.3.2. Evaluation of Algorithms: Generation of Realistic Data for given Ground Truth Information

The presented ToF simulation tool is able to reproduce the behavior of a real ToF sensor. Hence, it enables the generation of realistic data for given ground truth information. The possibility of using realistic data and corresponding ground truth data is a key feature for the objective evaluation of algorithms.

In collaboration with the German Cancer Research Center (Deutsches Krebsforschungszentrum (DKFZ), Heidelberg, Germany) a framework for evaluation of algorithms working with ToF data was developed. It was used to verify a new algorithm for fine registration of ToF range data with high-resolution surface data in a medical context.

The motivation of this work is the idea that pre-interventionally acquired volume data of the patient could be used to support a physician during a surgery. This pre-acquired data could be registered intra-operatively with “live” range data of the patients organs and used by a computer to provide for example aids for exact positioning of medical instruments. Such a registration requires to match the noisy surfaces generated from ToF range data onto pre-interventionally acquired high-resolution surfaces.

A widely used method for geometric alignment of 3D models is the Iterative Closest Point (ICP) algorithm [BM92]. This algorithm assumes that the input points are measured with zero-mean, identical and isotropic Gaussian noise. However, the process of generating 3D points from ToF range data leads to highly anisotropic noise.

An adapted version of the ICP algorithm better suited for coping with anisotropic noise than the original ICP algorithm was proposed by Maier-Hein et al. [MH+10a]. The goal of the collaboration project presented here was to evaluate this algorithm in a medical context.

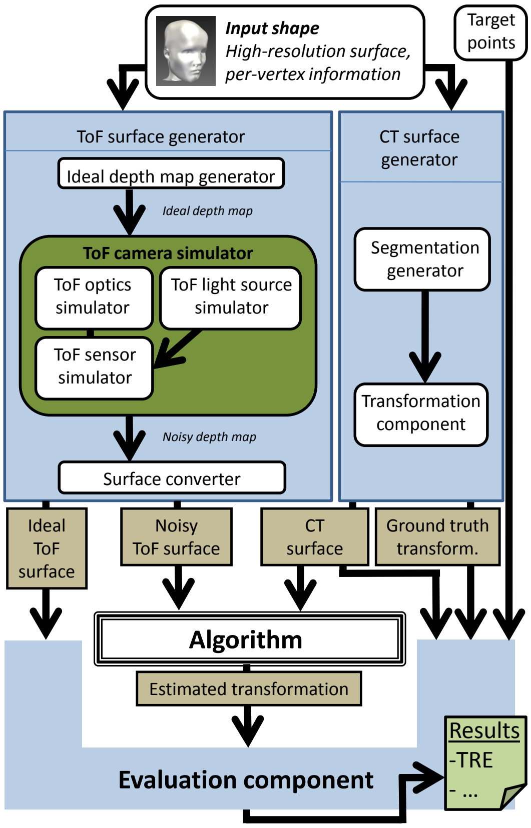


**Figure 3.10.:** Developed framework for evaluation of algorithms. (Modified according to [MH+10b].)

For this study, an evaluation framework was developed (c.f. Fig. 3.10). The framework uses real volumetric medical data as input which were acquired with a computed tomography (CT) scan. From the volumetric data simulated ToF data and corresponding CT surface data are generated by two modules: "ToF surface generator" and “CT surface generator”. The output of both modules is used as input for the algorithm under evaluation, which is analyzed by a following “Evaluation component”. Please note that the description given here focuses on the ToF simulation. A detailed discussion of the complete framework was published in [MH+10b].

The author’s contribution to the algorithm evaluation framework is a “ToF camera simulator” module. This module is an extended version of the developed ToF sensor simulation, which was parameterized to simulate a PMD CamCube ToF camera. Compared to the physical model presented in Section 3.1 two major extensions were made:

To account for a realistic depth-dependent attenuation of the active light the simulated illumination was modified by a function. This function implements the distance square law, meaning that for a doubled distance only a quarter of the active light reaches the target. It corresponds to a point light source located at the camera, which is a good approximation of the light source used by the real camera, if the target’s distance is $d_{target} > 30$cm (avoidance of near-field effects).

As a second extension a simple simulation of the camera optics was implemented to account for a finite lateral resolution. This module focuses on the optics blur which was simulated by convoluting the input of the `sampling` module (cf. Section 3.1.4, Fig. 3.2) with a specific kernel. This blur kernel represents the point spread function (PSF) of the simulated optics. It was approximated by a (space-invariant) Gaussian with a full width at half maximum (FWHM) of 20$\mu$m. This was found to be a reasonable approximation of the PSF of the original optics near the center of its field of view[4].

The evaluation framework was used for comparison of two algorithms which were given the task of matching of the noisy surfaces generated from simulated ToF range data onto high-resolution surfaces. As a virtual test object a human liver was used. The performance of the standard ICP as well of the new anisotropic ICP algorithm were investigated. Using various simulation setups and starting conditions it was shown that the anisotropic ICP outperforms the standard ICP. The total registration error, a quantity for measuring the misalignment of the two input meshes after convergence, was reduced by up to 70%.

[4] This approximation is based on measurements of the optical transfer function (OTF) of the PMD CamCube optics. The measurements were performed in collaboration with René Reichele (Institut für Technische Optik, Stuttgart University), Michael Erz, and Roland Rocholz (both: Heidelberg Collaboratory for Image Processing, Heidelberg University) within the Lynkeus project. The results were, however, not published yet.

A visualization of the result of the anisotropic ICP algorithm is given in Figure 3.11: It shows the noisy submesh generated from the simulated ToF data, registered to a reference mesh of a human liver.

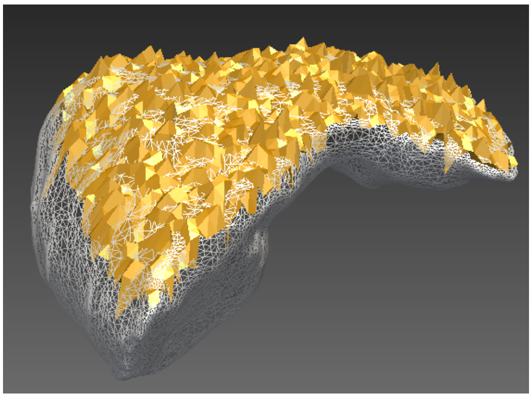


**Figure 3.11.:** Noise submesh generated from simulated ToF depth data, registered to a reference mesh of a human liver.

The study showed the advantage of the new algorithm in a medical context by using a human organ as demonstration object. Real ToF data of such an organ would have been hard to acquire because of strict directives (for example prohibiting the use of a (yet medically uncertified) ToF camera in an operating room). Therefore, an unique opportunity for testing the new ICP algorithm was opened up by the presented evaluation framework, of which the developed ToF model was an integral component.

The same framework was applied in a further cooperation project published in [MH+11], in which a ToF based augmented reality device for medical applications was proposed.

### 3.3.3. Standardization of ToF Systems: Extension of EMVA 1288 Standard

The developed ToF model is able to physically describe any given ToF camera. Although designed with focus on a specific system it is a generic model which may be used to emulate many different ToF camera implementations. The determined parameters describing a given system represent specific properties. Therefore, the author sees the model as an important element for the development of standards to characterize and compare ToF systems.

A detailed measurement and comparison of ToF systems from different manufacturers was performed by Erz et al. [EJ09]. A further generalization of this approach, involving aspects introduced by the presented ToF model may be found in [Erz11]. These efforts are contributing to work which seeks to develop a standard for characterization of ToF systems, which will be realized as an extension of EMVA 1288 Standard [EMV10].

This standard will certainly include measures to describe the accuracy of a depth measurement (i.e. covering systematic errors), the statistical depth error, and the distance non-uniformity (DNU, describing the error from pixel to pixel). It will enable a comparison and objective characterization of ToF cameras from different manufacturers.

# Chapter 4.

# Investigation of a ToF Sensor with a Nonlinear Photo Response

In the previous chapter a ToF camera was modeled and parameterized assuming a linear system. However, in reality such a perfectly linear imaging device does not exist. Every implementation of a ToF system shows nonlinear effects which influence the quality of the measured quantities.

## 4.1. Motivation

Especially the question of how a nonlinear photo response alters the determined depth information is of very much interest. Focusing on continuous-wave ToF systems this chapter will answer that question by considering three types of non-linearity and performing theoretical investigations. The methodology of these investigations will be introduced first in Sect. 4.2 and then applied to all investigated types.

In Sect. 4.3 the photo response will be modeled as a power function, which will help to understand the nature of the depth deviation. Section 4.4 will investigate the acceptable limits of the sensor's linearity in order to achieve a specific accuracy of the depth estimation. For this, a more natural shape of the photo response will be assumed and characterized using a measure defined by the EMVA 1288 Standard.

The characteristics of a possible logarithmic ToF sensor are analyzed in Section 4.5. Besides theoretical considerations realistic simulations utilizing the ToF model from Chapter 3 will be employed to describe this system.

### 4.1.1. Related Work

In 2D imaging the effect of a nonlinear photo response is quite obvious and was, to the authors knowledge, not investigated thoroughly. Methods aiming at a characterization of the non-linearity of 2D imaging sensors are provided e.g. by the EMVA 1288 Standard [EMV10]. An analysis of the nonlinear characteristics of ToF sensors was given by Erz and Jähne [EJ09].

To the authors knowledge an investigation of the influence of a nonlinear photo response on the accuracy or statistical uncertainty of the generated depth data was not published yet. Such analysis will be provided in this chapter. Furthermore a logarithmic ToF sensor will be proposed in Sect. 4.5. Similar intentionally nonlinear sensors are known from 2D imaging; for example logarithmic or semi-logarithmic imagers were presented in [Kav+00; Sch+00; THI98; Sto+04; Har+05].

## 4.2. Methodology for Theoretical Investigation of the Phase Estimation Error

Continuous-wave ToF systems are able to sample the correlation function of an incident electro-optical signal with an electronic reference signal. A nonlinear photo response causes the acquired samples to be altered which results in a distorted depth information. Here, this theoretical explanation will be outlined to serve as a basis for numerical investigations in the following sections.

For this, a sinusoidally modulated light source signal (4.1) and a rectangular reference signal (4.2) are assumed, which correspond exactly to the assumptions made in Section 2.2.2.1 (c.f. (2.4) and (2.5), page 16).

Because of the nonlinear photo response, the electro-optical signal sampled by the sensor is distorted. This distortion is modeled by a mapping function $\gamma$. The sensor samples the correlation function of this distorted signal $\gamma(S(t))$ and the reference signal $R(t)$, and generates distorted samples $c^{(\gamma)}(\theta)$ as (4.3).

$$S(t) = b_{ls} + a_{ls} \sin(2\pi \cdot \nu_0 t - \varphi) \tag{4.1}$$

$$R(t) = H(\sin(2\pi \cdot \nu_0 t + \theta)) \tag{4.2}$$

$$c^{(\gamma)}(\theta) = \int_0^{mT_0} \gamma(S(t))\, R(t)\, dt = \int_0^{mT_0} \gamma(S(t))\, H\left(\sin(2\pi \cdot \nu_0 t + \theta)\right)\, dt \tag{4.3}$$

Applying the reconstruction formulas (2.7)–(2.9) on samples acquired by such a system will cause systematic errors. Especially the phase shift (2.9) computed from the distorted samples will deviate from a phase shift computed from undistorted samples as

$$\begin{aligned}\Delta\varphi &= \varphi^{(\gamma)} - \varphi \\ &= \arg\left(\sum_{\tilde{n}=0}^{\tilde{N}-1} c_{\tilde{n}}^{(\gamma)} e^{-\mathrm{i}2\pi(\tilde{n}/\tilde{N})}\right) - \arg\left(\sum_{\tilde{n}=0}^{\tilde{N}-1} c_{\tilde{n}} e^{-\mathrm{i}2\pi(\tilde{n}/\tilde{N})}\right), \end{aligned} \tag{4.4}$$

with:

$$\theta_{\tilde{n}} = n \cdot 2\pi/\tilde{N}\,, \quad c_{\tilde{n}}^{(\gamma)} = \frac{c^{(\gamma)}(\theta_{\tilde{n}})}{mT_0}\,, \text{ and } \; c_{\tilde{n}} = \frac{c(\theta_{\tilde{n}})}{mT_0}\,.$$

The analyses given in the following sections will focus on a system using $\tilde{N} = 4$ equidistant samples and assume a bounded signal $0 < S(t) < 1$. Modeling the photo response by employing individual mapping functions $\gamma$ will enable to simulate different sensor characteristics. These simulations will be carried out by numerically evaluating Equation (4.4).

## 4.3. Impact of a Nonlinear Photo Response

This section aims at giving an impression about the nature of the expected depth error caused by a nonlinear photo response. Therefore the non-linearity $\gamma$ of the photo response is modeled as a power function with exponent $\alpha$,

$$\gamma : S(t) \to S(t)^{\alpha}\,, \tag{4.5}$$

and the phase difference $\Delta\varphi$ (4.4) is evaluated (numerically). For this, a fully modulated light source ($b_{ls} = a_{ls} = 0.5$, c.f. (4.1)) is assumed.

In Fig. 4.1 $\Delta\varphi$ is plotted in dependence of $\alpha$ and $\varphi$. The figure shows that $\Delta\varphi$ varies periodically with $\varphi$. A better visualization of this property is given in Fig. 4.2, where the error of the estimated phase $\Delta\varphi$ was evaluated for a fixed exponent $\alpha = 3$ and plotted over the phase $\varphi$. The wavelength of the variation is a quarter of the non-ambiguity range, so it is identical to the wavelength of the wiggling error caused by higher harmonics of the light source modulation (c.f. Sections 2.3.1.5 and 3.2.2).

The amplitude of the phase error is approximately 0.023rad, which corresponds for a typical modulation frequency of the active light source of $\nu_0 = 20$MHz to a depth error of $\Delta d = 2.7$cm.

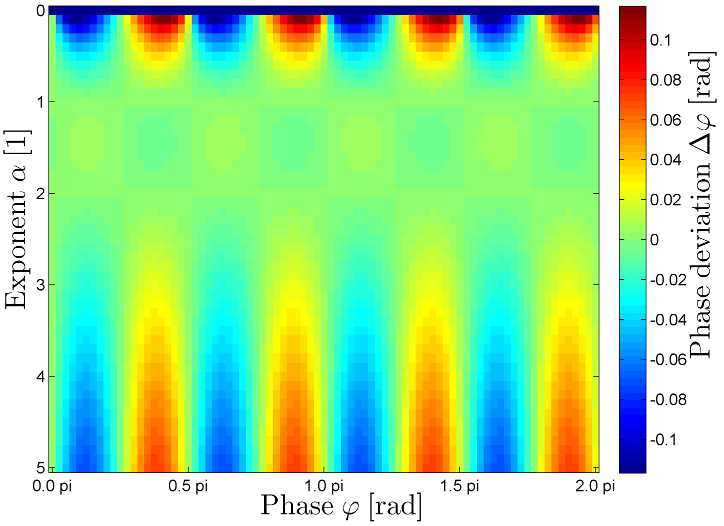


**Figure 4.1.:** Theoretical error of estimated phase, $\Delta\varphi$, in dependence of the exponent $\alpha$ and phase shift $\varphi$.

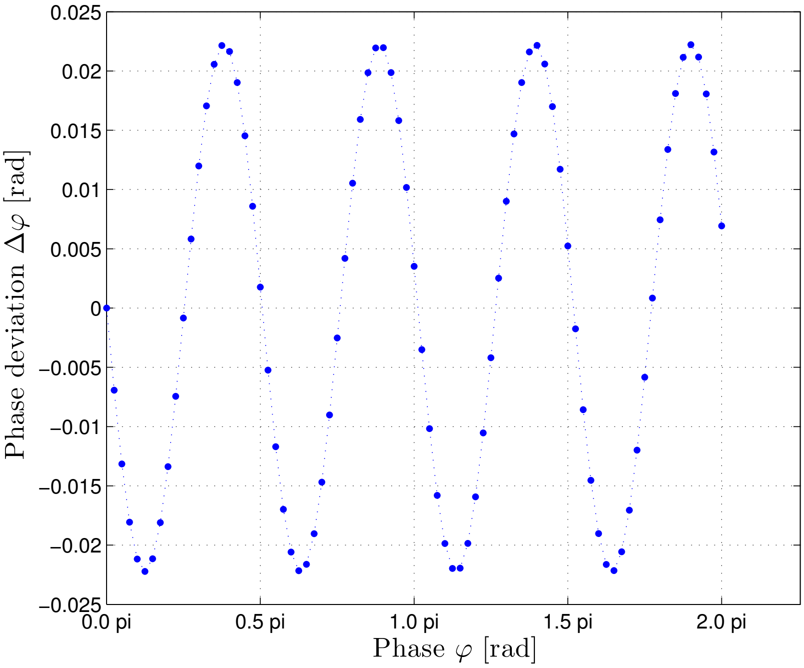


**Figure 4.2.:** Theoretical error of estimated phase, $\Delta\varphi$, in dependence of $\varphi$, for a fixed exponent $\alpha = 3$.

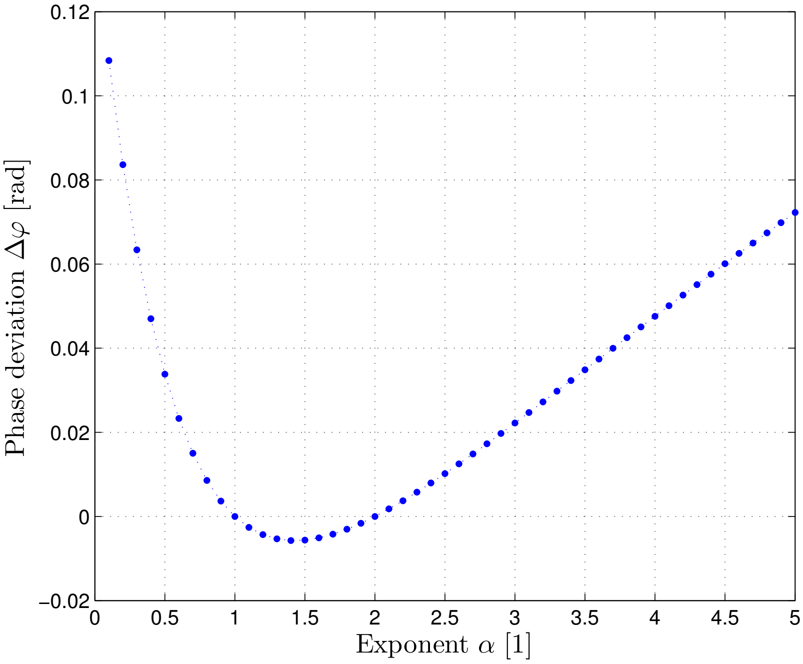


**Figure 4.3.:** Theoretical error of the estimated phase, $\Delta\varphi$, in dependence of the exponent $\alpha$, for a fixed phase shift $\varphi = 3/8 \cdot \pi$.

Returning to Fig. 4.1, a variation of $\Delta\varphi$ with $\alpha$ is visible which becomes more clearly in the following plot: Figure 4.3 shows a profile which is $\Delta\varphi$ over $\alpha$ for a fixed phase shift $\varphi = 3/8 \cdot \pi$ (corresponding to an extremum). It can be seen that for exponents $\alpha = 1$ and $\alpha = 2$ no phase deviation occurs. This is the case independently of $\varphi$ and – as evaluated using further simulations – independent of the modulation amplitude and offset of the light signal.

This absence of an error is obvious for $\alpha = 1$, since this case represents a perfectly linear photo response of the sensor, so no systematic deviation was expected. Interestingly, also exponent $\alpha = 2$ leads to a system behavior which does not introduce systematic deviations. This means that no additional systematic errors of the depth estimation are introduced by a quadratic photo response, if the system is using a (perfectly) sinusoidal modulation of the light source, rectangular shaped reference signal and $\tilde{N} = 4$ equidistant sampling points. An analytical proof for this property will be given in Appendix A.

The approximation of the nonlinear distortion by use of a power function with a single exponent is sufficient for non-linearities of a small extent. Real systems, however, have characteristic curves which are better described by a mixture of nonlinear terms of different orders. To investigate this point, the following section will perform further studies by assuming a more realistic shape of the non-linearity (containing many higher order terms).

## 4.4. Acceptable Limits of Linearity

The previous section outlined what kind of non-linearity of a ToF system's photo response causes errors of the depth estimation. This section will illuminate the topic from a different point of view and assume a more realistic shape of the non-linearity. It will answer the question of the requirements on the linearity of a ToF system, in order to reach a specific accuracy of the depth estimation. The analysis is done by incorporating a measure defined by the EMVA 1288 Standard for characterization of the non-linearity of a system's photo response.

The goal of this section is to give an estimation of the systematic error of the depth measurement $\Delta\varphi$ in dependence of a realistic non-linearity of the sensor's photo response. For this, the sensor's characteristic curve is modeled as a circular arc which has a homogeneous curvature over the whole range. This is a reasonable approximation of the shape of the characteristic curve of a real imaging sensor driven below its saturation (see for instance example section in [EMV10]).

Figure 4.4 visualizes the mapping function $\gamma$ with a parameter $\lambda$. This non-linearity parameter $\lambda$ describes the deviation of the investigated characteristic curve from a perfectly linear curve. Mathematically, $\gamma$ is defined as

$$\gamma(S) = P_y - \sqrt{r^2 - (S - P_x)^2}\,, \tag{4.6}$$

where $(P_x, P_y)$ are the coordinates of the center and $r$ is the radius of the arc:

$$\begin{aligned} r &= \frac{1}{2}\left(\frac{1/4 \cdot \rho^2 + \lambda^2}{\lambda}\right), && (4.7)\\ P_x &= 1 - 1/\sqrt{2} \cdot (\rho/2 - \lambda + r)\ , && (4.8)\\ P_y &= 1/\sqrt{2} \cdot (\rho/2 - \lambda + r)\ , && (4.9) \end{aligned}$$

with $\rho = \sqrt{2}$ being the length of the chord.

### 4.4.1. EMVA 1288 Linearity Measure

The EMVA 1288 Standard defines a measure for the non-linearity of the photo response of an imaging systems [EMV10, section 6.7]. Although developed for the

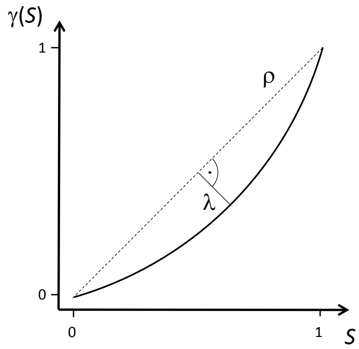


**Figure 4.4.:** Modeling the distortion $\gamma$ of the sensor's photo response as circular arc with a parameter $\lambda$ defining its curvature.

description of 2D imaging sensors this measure provides also a good tool for the characterization of the non-linearity of ToF sensors.

The EMVA Standard evaluates imaging sensors or cameras as a black box system, i.e. the complete characterization is based on analyzing the system's response to a well defined input. The linearity of the system is determined by illuminating the sensor homogeneously at varying irradiation levels while collecting the raw data $y$ output by the system. Varying the illumination level results in a different irradiation $H$. The raw values of each illumination level are averaged (giving $\mu_y$) and the average of the dark value (raw value acquired without any light) is subtracted. A straight line is fit to these values $\mu_y - \mu_{y,dark}$ over $H$.

Then, for each value the relative deviation $\delta_y$ from the regression is estimated. The mean of the maximal and minimal deviation gives the so called linearity error $LE_{\mathrm{EMVA}}$:

$$LE_{\mathrm{EMVA}} = \frac{\max(\delta_y) - \min(\delta_y)}{2}\,. \tag{4.10}$$

This linearity error is the central quantity used by the EMVA 1288 Standard to describe the non-linearity of an imaging system. For further details, please refer to [EMV10].

In the following it will be explained by use of Fig. 4.5 how the concept of the linearity measure $LE_{\mathrm{EMVA}}$ is related to the parameter $\lambda$. The points of $\{\gamma(S), S \in [0,1]\}$ located on the arc are interpreted as data samples. The EMVA Standard uses a linear fit of these data samples in order to estimate $LE_{\mathrm{EMVA}}$. Here, such a fit will not be (exactly) computed because it would require additional assumptions, e.g. a model of how the data are sampled. Instead, this linear fit is approximated as a constructed line which is parallel to the chord connecting the terminal points of the arc. The constructed line is located so that the maximal orthogonal distance of samples on both sides of the line is equal. This constructed line is visualized in Fig. 4.5 as solid line.

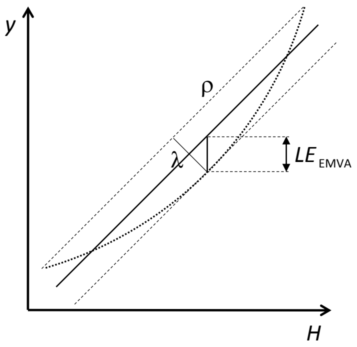


**Figure 4.5.:** Relation between curvature parameter $\lambda$ and the linearity measure $LE_{\text{EMVA}}$ defined by the EMVA 1288 Standard.

Using basic geometry the relation between $\lambda$ and $LE_{\text{EMVA}}$ reveals as

$$\lambda = 2\,\rho \cdot LE_{\text{EMVA}} = 2\,\sqrt{2} \cdot LE_{\text{EMVA}}\,. \tag{4.11}$$

This estimation is only an approximation, neglecting for example the fact that a linear fit would not be located exactly in the center of the arc. However, the focus of the following analysis is put on small non-linearities. Furthermore also the complete EMVA standard is only valid for systems showing small deviations from a linear system, so the measure $LE_{\text{EMVA}}$ is only valid for rather linear sensors. Thus, the relation (4.11) is assumed to be sufficiently accurate for the following investigation.

### 4.4.2. Evaluation

The phase error (4.4) was evaluated numerically by use of (4.6). The error $\Delta\varphi$ depends on multiple factors: It varies with the the extent of the non-linearity, which is here expressed using the non-linearity error $LE_{\text{EMVA}}$. Furthermore it varies with the phase $\varphi$, which was expected from the results of Sect. 4.3. But the phase error $\Delta\varphi$ also depends on the offset $b_{ls}$ and amplitude $a_{ls}$ of the optical signal $S(t)$. To understand these dependencies some figures will be shown in the following, where $\Delta\varphi$ is plotted over each one of these variables, while the others remain fixed.

Figure 4.6 shows the phase error $\Delta\varphi$ for varying phase $\varphi$ and non-linearity parameter $\lambda$. For each $\lambda$ the corresponding linearity error $LE_{\mathrm{EMVA}}$ was determined using (4.11).

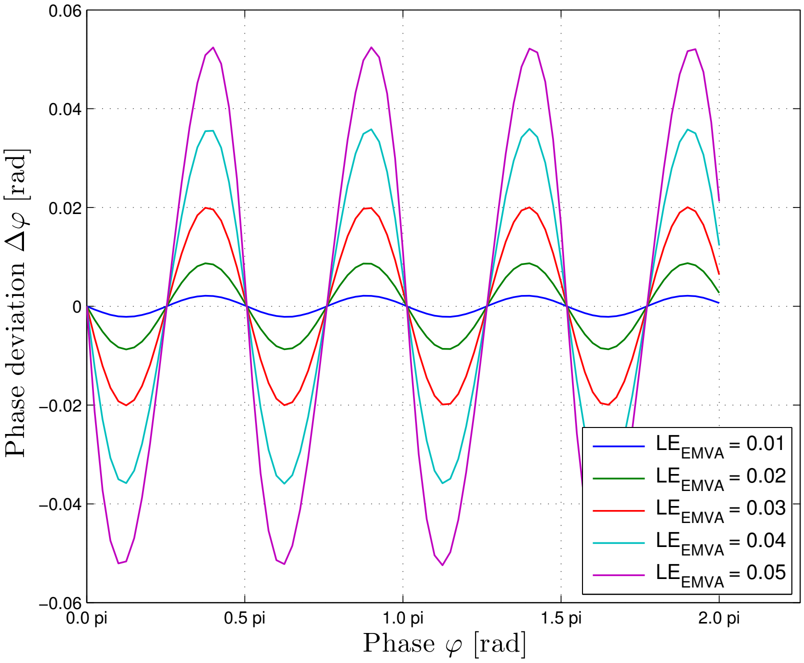


**Figure 4.6.:** Theoretical error of estimated phase, $\Delta\varphi$, in dependence of the phase $\varphi$, for various non-linearity errors $LE_{\mathrm{EMVA}}$.

The phase error shows similar characteristics as discovered in Sect. 4.3: It varies periodically with $\varphi$, having the same wavelength and its extrema at the same positions. Now the phase was set to $\varphi = 3/8 \cdot \pi$, which represents an extremum. For this fixed $\varphi$ the phase error $\Delta\varphi$ was evaluated using varying linearity errors $LE_{\mathrm{EMVA}}$ of the distorted photo response. Figure 4.7 shows $\Delta\varphi$ over $LE_{\mathrm{EMVA}}$ for varying modulation amplitudes $a_{ls}$ of the light source signal $S$, while its offset was set to a constant value $b_{ls} = 0.5$.

Figure 4.8 shows the corresponding plot of $\Delta\varphi$ over $LE_{\mathrm{EMVA}}$ for a varying offsets $b_{ls}$ and a fixed amplitude $a_{ls} = 0.1$.

### 4.4.3. Discussion & Conclusion

This section aimed to analyze the effect of a nonlinear photo response by assuming a characteristic curve corresponding to a circular arc. Such circular arc was assumed

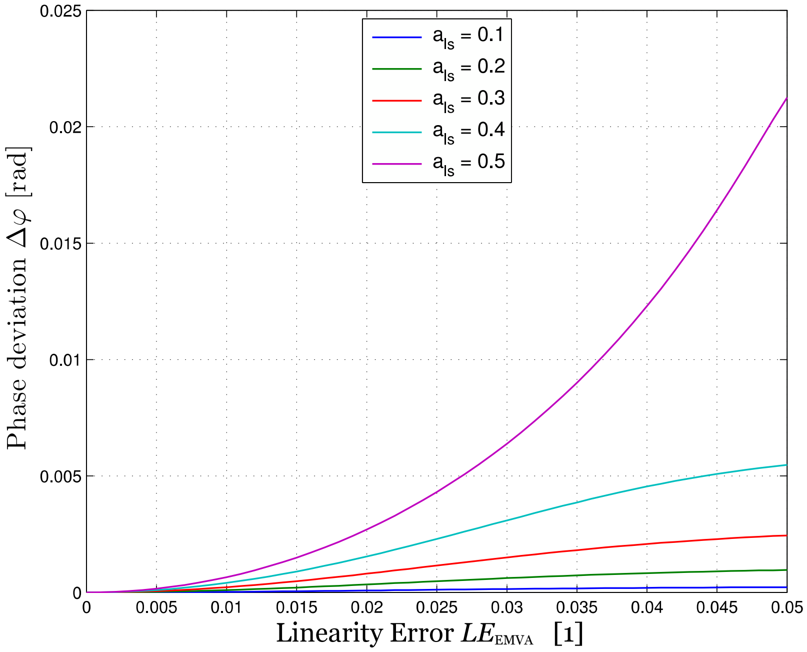


**Figure 4.7.:** Theoretical error of estimated phase, $\Delta\varphi$, in dependence of non-linearity error $LE_{\mathrm{EMVA}}$ for varying amplitudes $a_{ls}$ and a fixed offset $b_{ls} = 0.5$ of the light source signal. The phase was set to $\varphi = 3/8 \cdot \pi$.

to be a reasonable model of a real sensor driven below its saturation. The extent of the non-linearity of this arc was measured using the linearity error defined by the EMVA 1288 Standard. This measure seeks to express the non-linearity of an imaging system in a single number. Thus is is a good tool to give an estimation of the non-linearity, but it is not suitable to provide details about the characteristics of the photo response.

For this reason, also the analysis given here, linking the linearity error $LE_{\mathrm{EMVA}}$ with the error of the phase estimation $\Delta\varphi$, should be understood as a rough estimation rather than an exact relation. A real imaging system could have for example a characteristic curve with a shape differing from a that of the assumed circular arc. The non-linearity of such a system could be described by a specific linearity error $LE_{\mathrm{EMVA}}$, but cause a phase error $\Delta\varphi$ which differs very much from the estimates given here.

The obtained results show that the error of the phase estimation $\Delta\varphi$ depends on a variety of factors. It is influenced by the extent of the non-linearity, which was

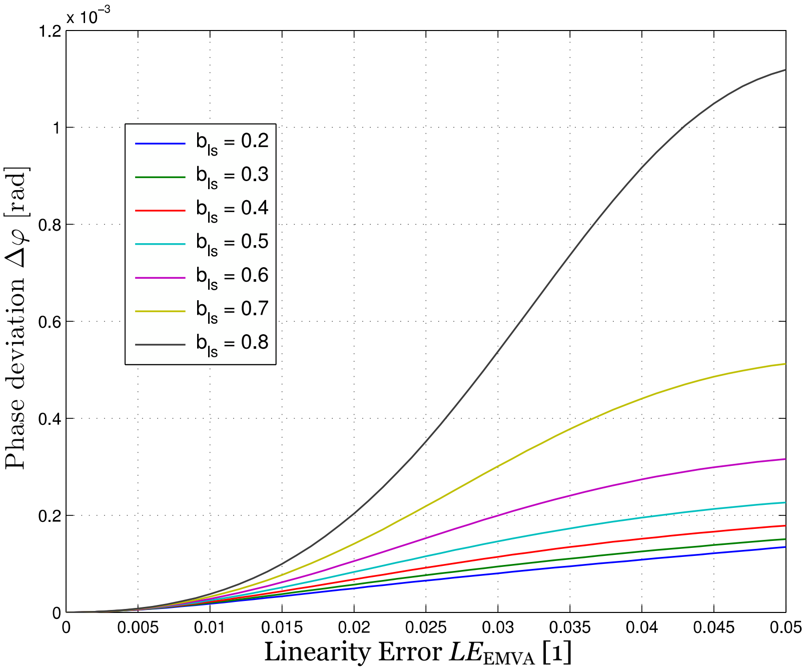


**Figure 4.8.:** Theoretical error of estimated phase, $\Delta\varphi$, in dependence of non-linearity error $LE_{\text{EMVA}}$ for varying offsets $b_{ls}$ and a fixed amplitude $a_{ls} = 0.1$ of the light source signal. The phase was set to $\varphi = 3/8 \cdot \pi$.

characterized by $LE_{\text{EMVA}}$ here. Furthermore $\Delta\varphi$ depends on all parameters of the electro-optical signal $S(t)$, namely its offset $b_{ls}$, amplitude $a_{ls}$ and phase $\varphi$.

Figure 4.6 suggests that the average error of the phase estimation increases with $LE_{\text{EMVA}}$, corresponding to an increased extent of the non-linearity. Furthermore it varies periodically with $\varphi$. Figure 4.7 indicates that the error increases with increasing the modulation amplitude $a_{ls}$. It has to be noted, however, that the simulated maximum value $a_{ls} = 0.5$ is unrealistically high and has only been simulated for completeness. In practice modulation amplitudes exceeding $a_{ls} = 0.3$ are very unlikely. The maximal phase error for $a_{ls} = 0.3$ and a linearity error $LE_{\text{EMVA}} = 0.05$ is $\Delta\varphi = 2.5 \times 10^{-3}$rad. Assuming modulation frequency of $\nu_0 = 20$MHz this corresponds to a depth error of $\Delta d = 3$mm which is negligible in many applications.

Figure 4.8 was simulated assuming a fixed modulation amplitude of $a_{ls} = 0.1$ and a varying offset of the electro-optical signal. Such modification could result for instance from a variation of the non-modulated ambient light. The figure suggests that the error $\Delta\varphi$ increases with increasing offset $b_{ls}$. Here, the maximal phase error for $b_{ls} = 0.8$ and a linearity error $LE_{\text{EMVA}} = 0.05$ is $\Delta\varphi = 1.12 \times 10^{-3}$rad. Assuming a modulation frequency of $\nu_0 = 20\text{MHz}$ this corresponds to a depth error of $\Delta d = 1.3\text{mm}$. Hence, it is also negligible in most applications.

These results reveal an interesting property of ToF systems: Even sensors having a rather nonlinear photo response generate data which leads to very small errors of the depth estimation. The following section will analyze this feature in more detail and suggest a ToF sensor which uses a consciously distorted characteristic curve in order to facilitate a higher dynamic range.

## 4.5. Exploiting a Nonlinear Photo Response: A Logarithmic ToF Sensor

As mentioned in Section 2.3.1.4 ToF imaging requires sensors with an enormous dynamic range. Such a large dynamic range enables the system to cope with strong sources of interfering ambient light. But also applications in a controlled environment and without any interfering light benefit from an increased dynamic range:

The active light source has a spatial extent which is normally much smaller than the distances of the imaged objects[1]. Hence, it can be approximated as a point light source. Therefore the intensity of the active illumination reaching the target drops with approximately $1/d^2$ (distance square law). For this reason the intensity of the light backscattered by objects near the camera is much higher than this of the light backscattered by far objects. Thus, the intensity of the detected backscattered light easily varies over several magnitudes, even for scenes with moderate depth dynamic[2].

A typical method to reach a high dynamic range in ToF imaging is to use pixels with immense fullwell capacities, which requires big pixel areas and therefore limits the lateral resolution of the system (c.f. shortcoming in Sect. 2.3.1.2).

Here, another possibility will be investigated which is based on the idea of using an intentionally nonlinear characteristic curve. In recent years so called *High Dynamic Range Cameras* (HDRC) utilizing pixels with a logarithmic or semi-logarithmic photo response were presented (see e.g. [Kav+00; Sch+00; THI98; Sto+04; Har+05]). However, no ToF system using a logarithmic photo response has been realized yet. Therefore it is an interesting question of how a logarithmic photo response would influence the characteristics of a ToF camera system.

[1] Typical dimensions: light source: 10cm, object's distance: $> 1\text{m}$

[2] The depth dynamic of a scene is the ratio of the distances of the most remote to the most closest object (see Sect. 1.2).

This logarithmic photo response is modeled as a mapping function:

$$\gamma(S) = \frac{1}{\log(g+1)} \cdot \log(S+1)\,. \tag{4.12}$$

The parameter $g$ defines the gain of the input dynamic range, in other words this factor describes the multiple of the light being detectable before the sensor gets saturated. The following investigations will focus on system behavior for a parameter $g = 10$. This specific setting was chosen because the parameter is high enough to lead to system characteristics which are very different from that of the linear system. On the other hand the parameter is small enough to allow comparability of the derived logarithmic system with the original linear camera.

First, a theoretical consideration will be given, followed by more realistic simulations based on the physical model presented in Chapter 3.

### 4.5.1. Theoretical Investigation

The results of Sects. 4.3 and 4.4 suggest that also a logarithmic distortion of the photo response introduces systematic errors of the depth deviation. Employing the same methodology (c.f. Sect. 4.2) as in these sections the error of the phase estimation $\Delta\varphi$ was evaluated by use of simulations. As a result, the error varies periodically with $\varphi$, and depends on the amplitude $a_{ls}$ and offset $b_{ls}$ of the light source signal. The maximum error (computed for $\varphi = 3/8 \cdot \pi$) is visualized as a surface plot in Fig. 4.9. The triangle structure of the diagram results from the fact that only parameter combinations giving $S \in [0, 1]$ were simulated.

It can be seen that the error depends only very slightly on the offset $b_{ls}$. It increases with higher amplitudes $a_{ls}$. The typical error is very small and has a maximum value of $5 \cdot 10^{-5}$rad.

### 4.5.2. Realistic Simulations

The theoretical estimation given in Sect. 4.5.1 does not regard sensor noise nor is able to take the real modulation of the light source signal into account. Therefore the physical model presented in Chapter 3 was adapted to perform a more realistic study of the influence of the logarithmic photo response on the accuracy and statistical error of the depth measurement.

As a basis for simulation of such a logarithmic sensor the model and parameterization of the PMDTechnologies CamCube 2.0 camera (see Sect. 3.2) was used. The

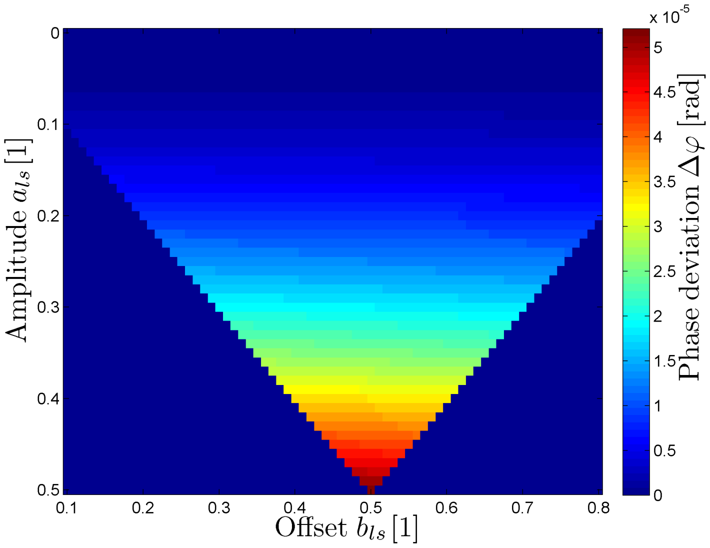


**Figure 4.9.:** Theoretical error of estimated phase, $\Delta\varphi$, in dependence of the amplitude $a_{ls}$ and offset $b_{ls}$ of the light source signal, simulated for a fixed phase $\varphi = 3/8 \cdot \pi$ and gain $g = 10$.

logarithmic photo response (4.12) is modeled as a separate non-linearity module. Since the simulated camera is a two-tap sensor system, two copies of this module are integrated into the model. Each logarithmic module is located between the simulation of the overall system gain $K$ and the AD converter (see Fig. 3.4, page 36).

For this virtual nonlinear camera the special photo response characteristic prevents a saturation of the quantum wells, therefore the SBI circuit was no longer required and thus was deactivated.

#### 4.5.2.1. Response and Noise of the Logarithmic Sensor

The logarithmic sensor was characterized using a setup according to the EMVA 1288 Standard: The virtual sensor was irradiated homogeneously with light of varying intensities. While the irradiation $H$ was increased the mean and variance of the raw data were analyzed. The result is given in Fig. 4.10, showing the mean raw value minus the mean raw value of the dark image $\mu_y - \mu_{y,0}$ and the variance of the raw values $\sigma_y^2$ over the irradiation $H$.

As expected, the curve representing the mean raw value $\mu_y - \mu_{y,0}$ has a logarithmic shape. The curve's slope is high at low irradiations $H$ and decreases for higher values of $H$. This means that the sensor's light sensitivity decreases for increasing $H$ which results in an increased dynamic range.

The variance $\sigma_y^2$ increases with $H$, reaches a maximum at $H \approx 2 \cdot 10^7$ photons/pixel and slightly decreases again. Over the full range of $H$, it seems to be relatively constant, which is a typical property of cameras having a logarithmic photo response. Please note that the relatively low variance of the raw data $\sigma_y^2$ follows from a highly decreased sensitivity (relative to the standard linear sensor).

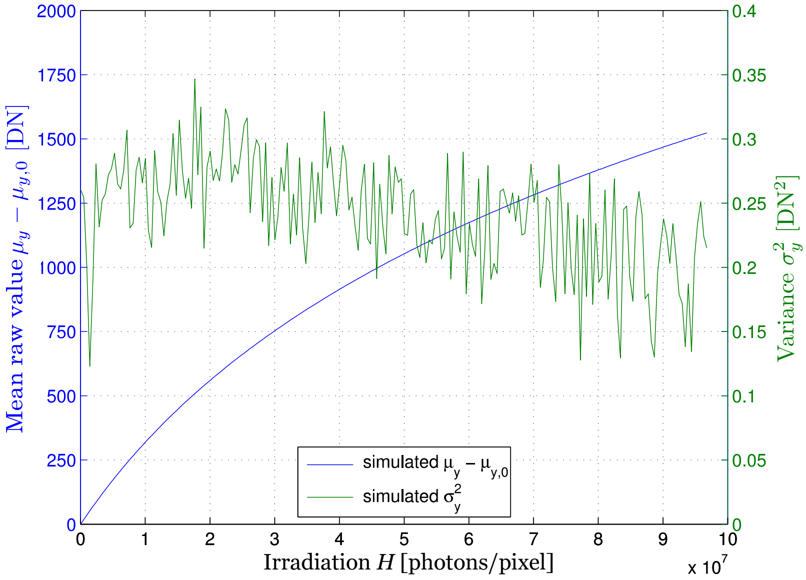


**Figure 4.10.:** Simulated response of the proposed logarithmic ToF sensor to a homogeneous illumination at varying intensities.

### 4.5.2.2. Implications on the Dynamic Range and Comparison with Standard Sensors

The main advantage of a logarithmic ToF sensor compared to a linear sensor is a high dynamic range which facilitates the imaging of scenes with an increased depth dynamic. Thus, a good scenario for evaluating the behavior and understanding the practical benefit of a logarithmic ToF system is imaging a simple target at various distances and regarding systematic and statistical errors of the determined depth data. The following analysis utilizes the physical sensor model (from the prior Section 4.5.2.1) for simulation of a plane target at different distances while incorporating a realistic active illumination. Especially the real temporal modulation of the light

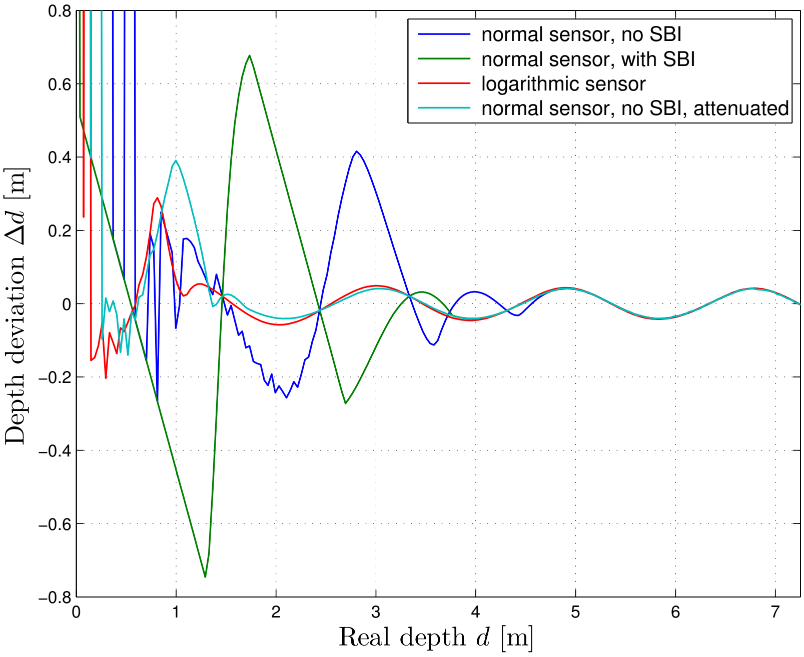


**Figure 4.11.:** Comparison of the systematic error $\Delta d$ of the depth estimation for different sensor types.

source as measured for the CamCube camera and its spatial attenuation as $1/d^2$ were implemented.

To assign the peak intensity of the light source the virtual target was placed in a distance of $d = 1$m. The peak intensity was now set to a value corresponding to a photon flux of the backscattered light of $2 \cdot 10^{10}$ photons hitting one sensor pixel per second. Since the light source is being modulated, the mean photon flux is about the half of this quantity. The simulation was run without any additional non-modulated light, i.e. assuming a fully modulated light source and no ambient light. The integration time was set to $t_{exp} = 0.025$s per subframe.

Besides the system using a logarithmic sensor three other settings were simulated using the same setup and system properties: A ToF system using the standard linear sensor (1) and parameterization (as derived in Sect. 3.2), which is equipped with a SBI circuit for compensation of non-modulated light. This SBI circuit was deactivated (2) for a second evaluation. (Please note that a deactivation of the SBI system is not possible using the real camera.) A third simulation was employing the same linear sensor with deactivated SBI, but assuming an (3) attenuated intensity of the detected light by factor 8 (corresponding to an optics aperture narrowed by 3 f-stops).

For all these four virtual ToF systems the imaging of a plane target in various distances was simulated. At each distance a phase map $\boldsymbol{\Phi}$ and depth map $\boldsymbol{D}$ were computed from sensor raw data. By averaging the values over all pixels the deviation $\Delta d$ of the estimated depth from the real depth and variance of the depth data $\sigma_d^2$ were computed.

Fig. 4.11 shows the depth deviation $\Delta d$ plotted over the real depth. For big distances ($d \approx$ 5m…7m) the four systems generate very similar depth data. The typical periodical depth variation caused by the imperfect modulation of the light source (wiggling error, see Sects. 2.3.1.5 and 3.2.2) can be seen. Fig. 4.12 shows a magnification of the same depth range, from which the expected slightly increased systematic error of the logarithmic system compared to the other systems is visible (c.f. Sect. 4.5.1). The maximum difference of the depth deviations between these systems is about 3mm.

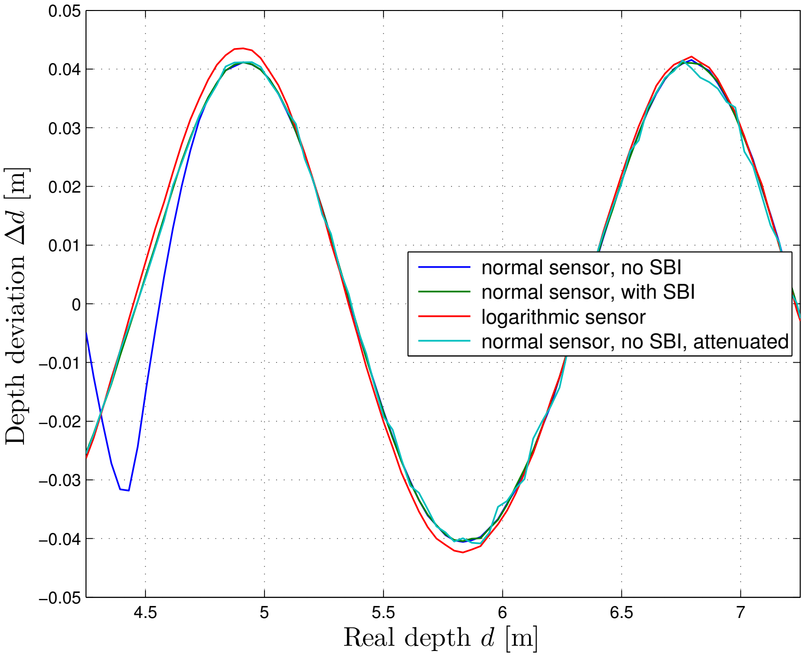


**Figure 4.12.:** Magnified area of Fig. 4.11: Comparison of the systematic error $\Delta d$ of the depth estimation for different sensor types.

In the performed simulation a smaller distance of the target corresponds to an increased intensity of the detected light. Returning to Fig. 4.11 it can be seen that the systems start to show severe systematic errors of the estimated depth at different distances, which are caused by successive saturation of the raw channels. Following the curves in the figure from right to left it is visible that the standard sensor with deactivated SBI is the first system showing significant deviations ($d$ = 4.6m, dark blue curve). Next, the standard sensor with activated SBI circuit ($d$ = 3.8m, green curve) and the attenuated system ($d$ = 1.6m, light blue curve) show deviations. The system using the logarithmic sensor is able to cope with the light reflected by the nearest target ($d$ = 1.3m, red curve) corresponding to the highest light intensity.

Furthermore, for each system the statistical error of the depth estimation was investigated in dependence of the target's distance. For this, the statistical variance of the depth values determined by all pixels was computed at each distance. Fig. 4.13 shows a plot of the variance $\sigma_y^2$ over the depth $d$. It is visible that the uncertainty of the depth data generated by each system is approximately proportional to the distance of the target. Please note that the computed values are only valid in the range of unsaturated raw data which correspond to the distances in which the depth estimation is correct (c.f. Fig. 4.11).

For big distances ($d \approx 5 \ldots 7$m) the systems using the standard sensor (with and without SBI) produce data with a similar uncertainty (dark blue and green curve). Compared to these two systems the error of the logarithmic system is slightly increased (red curve). The statistical error of the system using an attenuated sensor (light blue curve) is much higher.

### 4.5.3. Conclusion: Investigation of a Logarithmic ToF Sensor

In the previous sections an investigation of a logarithmic ToF imaging system was performed. Since such system has not been realized yet, these investigations were based on a theoretical analysis (Sect. 4.5.1) and realistic simulations (Sect. 4.5.2).

The theoretical consideration has shown that a logarithmic characteristic curve introduces systematic errors in the depth estimation which are, however, extremely small and hence negligible in most applications. The physical model from Chapter 3 was adapted in order to simulate the logarithmic sensor and to investigate its response. This virtual sensor was used for a comparison study with modified versions of the original linear sensor. The study showed that the logarithmic sensor has an increased dynamic range which facilitates the recording of sceneries with high depth dynamic. Among the compared systems, the logarithmic system was able to cope with the highest light intensity. However, it should be mentioned again that the simulation was run assuming no additional non-modulated light (i.e. also no back-

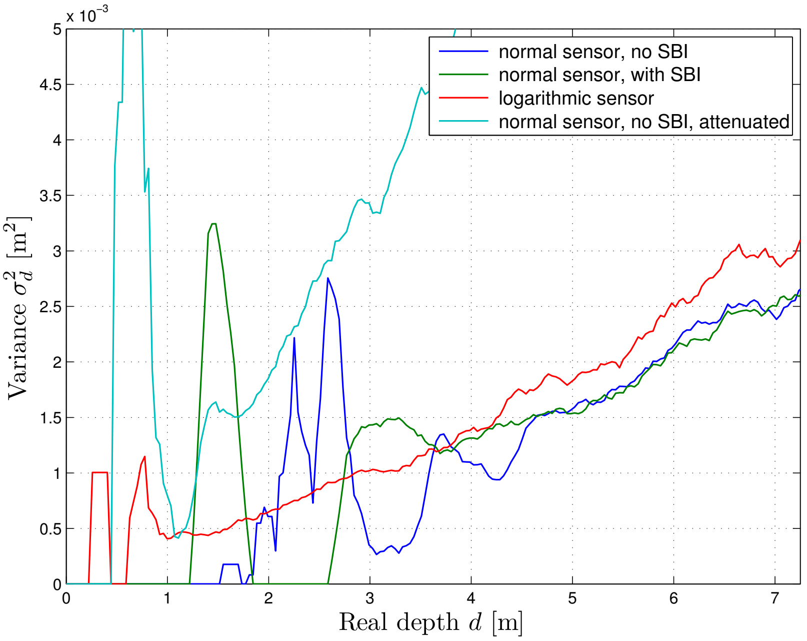


**Figure 4.13.:** Comparison of the statistical error of the depth estimation $\sigma_d^2$ for different sensor types.

ground light). So for example the SBI system – which is optimized for neutralization of non-modulated light – was not simulated to work under optimal conditions.

The statistical error of the depth data generated by the logarithmic system was only slightly above the error of the original system and much smaller than the error of the attenuated system, which was the only system being able to image a near target. Concluding these facts, the approach of a logarithmic ToF sensor seems to be a very promising concept, enabling a highly increased dynamic range while showing systematic and statistical errors which are only slightly increased relative to a comparable linear sensor.

## 4.6. Summary

The subject of this chapter was the investigation of the impact of a nonlinear photo response on the accuracy of the depth estimation. By evaluating different kinds of distortion the phase error was characterized. The acceptable limits of the sensor's linearity in order to reach a specific accuracy of the depth information were explored. This was done by utilizing the EMVA 1288 linearity measure and determining the error of the phase estimation theoretically. The results of these investigations suggest that even large deviations from a linear photo response cause systematic errors which are still manageable.

This inspired the investigation of a ToF sensor using a logarithmic photo response. Since such sensor has not been implemented yet its analysis was based on theoretical investigations and realistic simulations using an extended version of the physical ToF model from Chapter 3. According to the results of these considerations a logarithmic ToF sensor seems to be a very promising concept.

The analyses performed here were focusing on a two-tap ToF system following the continuous-wave approach. However, because of the similarity of all Time-of-Flight implementations (c.f. Sect. 2.2.3) these results should be regarded to be valid in a more general sense. So for example systems using more than two taps or driven in a pulsed mode will lead to comparable results.

All investigations performed in this chapter assumed a perfect non-linearity module which is identical for both taps. However, in practical implementations the differences of the nonlinear behavior of the taps might be a critical issue. The next chapter will focus exactly on this question. It will turn out that the different characteristic curves of the taps are actually contributing to shortcomings even of today's ToF systems, resulting in a limited frame rate and reduced quality of the generated depth maps. A method resolving these issues based on a dynamic calibration will be presented.

# Chapter 5.

# Dynamic Sensor Calibration

The subject of the previous chapter was the investigation of the effect of a (global) nonlinear photo response on the data produced by a ToF sensor. However, the model of an equal and homogeneous non-linearity is not an optimal description of real Time-of-Flight imagers: Multi-tap sensors employ several detection units and each of them has its own amplification path with a specific characteristic curve. Differences of the photo response of these amplification paths can lead to large distortions in the reconstructed depth image. Therefore, using today's sensors it is not possible to acquire the required raw images for reconstruction of a depth map using different taps. Instead, each tap acquires a raw image on its own and the systematic errors are canceled out by averaging these raw images.

This chapter presents a method to implicitly calibrate the photo response characteristic of multi-tap 3D Time-of-Flight sensors. The calibration data are gathered from arbitrary live acquisitions. The proposed correction of raw data supersedes the commonly used averaging technique. Thus it is possible to compute multiple depth maps from a single set of raw images. This enables an increase in frame rate of factor two or more depending on the sensor design. Furthermore motion artifacts are significantly reduced.

The method presented in this chapter was applied for a patent in [SZ10a]. Furthermore, parts of the work presented here were published in [SZJ11].

## 5.1. Motivation

In this chapter a method is proposed which performs an implicit scene-based calibration of multi-tap correlating ToF sensors. The resulting calibration routine allows the computation of additional independent depth images, so the effective frame rate can be increased (from currently 30Hz on average to 60Hz or, using an extension, even 120Hz).

The goal of this investigation is to provide a novel technique to increase the frame rate of ToF systems based on today's hardware. Please note that this method is not intended to replace the initial depth calibration routines, inherent to every ToF system to achieve absolute accuracy.

Using a specific camera system and a simple implementation of the proposed technique, it will be shown that doubling the frame rate of an available ToF system is possible. Thus the feasibility of this approach is shown for the entire class of ToF cameras employing correlating multi-tap sensors (c.f. Sect. 2.2.3). Since this work is intended to stimulate the design of new ToF systems, limitations given by the provided proof of concept implementation (for example necessity of a deactivated SBI) do not restrict the applicability of the concept itself.

Starting with a definition of the problem caused by an unequal photo response of the taps in Sect. 5.2, the approach of an implicit dynamic calibration and raw data rectification will be outlined in Sect. 5.3. Experimental results and a detailed evaluation will be given in Sect. 5.4. Conclusion and outlook are provided in Sect. 5.5.

### 5.1.1. Related Work

Much work performed in the field of calibration of Time-of-Flight cameras relates to the compensation of the deviations of distance or intensity measurements. For instance Kahlmann, Remondino, and Ingensand [KRI06], Lindner and Kolb [LK06; LK07], Rapp [Rap07], and Stürmer, Penne and Hornegger [SPH08] presented methods to decrease systematic deviations of the estimated scene unknowns.

An investigation dealing with the raw data of ToF systems, aiming to understand errors of the estimated scene unknowns is given in Chapter 3 and was published by the author and Jähne [SJ09]. A work focusing on the radiometric characteristics of ToF sensors was published by Erz and Jähne [EJ09].

The approach of computing depth maps by use of fewer acquisitions was mentioned by Lottner et al. [Lot+07] in a work aiming at a reduction of the motion artifacts. However, it could not be put into practice because of observed considerable deviations of the generated depth data from the expected depth.

A similar idea was presented by Hussmann and Edeler [HE09]. Their method suffers from large distortions which they noticed as an increase of the (spatial) standard deviation of the determined depth values.

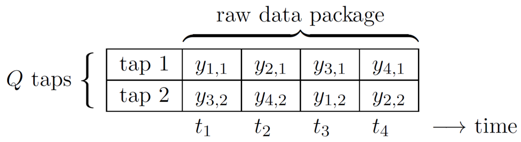
As it will be shown here, these deviations result from the substantial inequality of the different taps. Correcting these inequalities is crucial in order to generate high quality depth maps, facilitating the application of the method.

## 5.2. Problem Definition

As outlined in Sect. 2.2.2 the majority of the available ToF sensors is capable of acquiring multiple samples of the correlation function simultaneously. Today's sensors usually make use of two taps.

In a general formulation, *each pixel* of the sensor has $Q$ detection units (taps) which parallelly acquire measurement values. Each of these detection units may be driven in $N$ different measurement modes and each one of these modes aims to measure one specific sample out of a set of $\tilde{N}$ required samples of the correlation function. In this work the indexing is chosen in such a way that the sampling mode $n$ measures the sample with index $\tilde{n}$, so $\tilde{n} \equiv n$ (c.f. footnote on page 17). The method being presented here is not restricted to multi-tap sensors sampling the correlation function as discussed in Sect. 2.2.2. Therefore the following reasoning will use the index of the sampling mode $n$ rather than $\tilde{n}$. Please note that the line of argument and results may be applied directly on sensors working as discussed in Sect. 2.2.2 by setting $\tilde{n} = n$.

The theoretical value to be measured by a particular detection unit $(q, \text{with } q \in \{1, \ldots, Q\})$ in a specific measurement mode $(n, \text{with } n \in \{1, \ldots, N\})$ will be denoted as $u_{n,q}$. The result of this measurement is a digital value which will be denoted as $y_{n,q}$. Usually $N > Q$ is valid, thus to acquire the required $N$ samples, multiple $(L)$ acquisitions are necessary. A typical raw data package is depicted in Fig. 5.1.



| | | | | |
|---|---|---|---|---|
| tap 1 | $y_{1,1}$ | $y_{2,1}$ | $y_{3,1}$ | $y_{4,1}$ |
| tap 2 | $y_{3,2}$ | $y_{4,2}$ | $y_{1,2}$ | $y_{2,2}$ |
| | $t_1$ | $t_2$ | $t_3$ | $t_4$ |

**Figure 5.1.:** A typical raw data package for $Q = 2$ taps and $N = 4$ measurements of the correlation function.

In the case of using a sensor as discussed in Sect. 2.2.2 the values $u_{n,q}$ to be measured correspond to samples of the correlation function (2.6) $c(\theta_n) = c_n$ [1]:

$$\begin{aligned} u_{1,1} = u_{1,2} &= c_0 \quad (5.1) \\ u_{2,1} = u_{2,2} &= c_1 \quad (5.2) \\ u_{3,1} = u_{3,2} &= c_2 \quad (5.3) \\ u_{4,1} = u_{4,2} &= c_3 \quad (5.4) \end{aligned}$$

Ideally, the acquired values $y_{n,q}$ would be identical to $u_n$, and hence in the case of (5.1)–(5.4) they would be equal to the samples $c_n$ used in (2.7)–(2.9) for reconstructing the scene unknowns.

$$y_{n,q} \quad \sim \quad u_{n,q}\,, \text{with} \;\; n \in \{1,\ldots,N\}\,, q \in \{1,\ldots,Q\} \qquad (5.5)$$

Unfortunately it is not possible to use these values $y_{n,q}$ directly, because the measurement process introduces errors which have to be compensated by an adequate processing.

### 5.2.1. Erroneous Measurement Process

As investigated by Erz et al. [EJ09; Erz11] each tap of today's ToF sensors has an individual characteristic curve. Following the notation from Chap. 4 this characteristic curve will be modeled as transformation $\gamma$ (5.6). Ideally, $\gamma$ is a linear function and identical for all taps $q$ and sampling modes $n$. However, due to imperfect fabrication processes, $\gamma_{n,q}$ differs for each sampling mode $n$ ($n \in \{1,\ldots,N\}$) and tap $q$ ($q \in \{1,\ldots,Q\}$).

$$y_{n,q} = \gamma_{n,q}(u_{n,q}), n \in \{1,\ldots,N\}, q \in \{1,\ldots,Q\} \qquad (5.6)$$

Please note that the characteristic curves $\gamma_{n,q}$ are also different for each pixel. Thus, (5.6) extents to

$$\boldsymbol{Y}_{n,q} = \boldsymbol{\Gamma}_{n,q}(\boldsymbol{U}_{n,q})\,. \qquad (5.7)$$

For simplicity, the following reasoning will focus on a single pixel. In an implementation, the method derived here is applied to all pixels of the sensor in the same way.

[1] If, however, a multi-tap sensor utilizing signals different from those assumed in Sect. 2.2.2 (e.g. a non-rectangular reference signal) is used, the values $u_{n,q}$ might correspond to different quantities. But still, the dynamic calibration and rectification method proposed here could be employed. This means the presented algorithm is not restricted to sensors working exactly as modeled in Sect. 2.2.2.

### 5.2.2. State-of-the-Art: Averaging Technique

A possible strategy to compensate errors introduced by the different characteristic curves is to perform an averaging over all taps, i.e. each sample of the correlation function is measured by each of the $Q$ detection units individually, and all these values are averaged arithmetically [2]. Thus, capturing $N$ samples for correction using the averaging technique requires $L = N$ acquisitions.

For example, the CamCube 2.0 ToF System by PMDTechnologies uses a sensor with $Q = 2$ taps and acquires $N = 4$ samples of the correlation function. Therefore a raw data package consists of $R = 8$ values, of which half is acquired with tap 1, and the other half with tap 2 (c.f. Fig. 5.1).

The acquired values $y_{n,q}$ are used to compute the samples $c_n$ by (5.8)–(5.11), which are utilized in (2.7)–(2.9) for reconstructing the scene unknowns.

$$c_0 = (y_{1,1} + y_{1,2})/2 \tag{5.8}$$

$$c_1 = (y_{2,1} + y_{2,2})/2 \tag{5.9}$$

$$c_2 = (y_{3,1} + y_{3,2})/2 \tag{5.10}$$

$$c_3 = (y_{4,1} + y_{4,2})/2 \tag{5.11}$$

This strategy has the effect that differences of the various characteristic curves $\gamma$ cancel out. However, this is only valid for differences of linear order, i.e. higher order deviations of the different characteristic curves $\gamma$ are not compensated. Furthermore, any implementation of this strategy will be slow since each sample of the correlation function $c_n$ is measured multiple times (namely by each tap $q$) to generate a single set of scene unknowns.

## 5.3. Calibration and Rectification

A possibility to supersede this averaging technique is to determine $\gamma_{n,q}$ of the ToF system by performing a photometric calibration (see e.g. Erz and Jähne [EJ09]). Such approach explicitly determines each $\gamma_{n,q}$ by illuminating the sensor with a well defined input and by analyzing the (raw data) output of the ToF system. However, such an explicit calibration requires a tunable and preferably homogeneous light source, e.g. an integrating sphere. Furthermore, this explicit calibration is slow and

[2] This particular averaging strategy is used by `PMDTechnologies` to the author's knowledge. The raw data processing methods of other ToF manufacturers are not disclosed and thus not known to the author.

thus expensive in a production line. The most critical issue is, however, that $\gamma_{n,q}$ are usually not stable over time, because they depend on a variety of factors especially on the temperature. (This temperature dependence of the estimated correction parameters will be investigated in Sect. 5.4.2.)

Therefore instead of such an explicit calibration, the method proposed here aims at performing an implicit calibration where the differences between two read-out paths $\gamma$ are estimated and compensated. The following section (Sect. 5.3.1) will introduce a *rectification operator*. The goal of this operator is to correct the raw data such as they were measured using a single tap. By defining the requirements on this operator it will be described in an abstract manner.

The section subsequent to the following one (Sect. 5.3.2) will then specify this operator and explain how exactly it is constructed.

### 5.3.1. Implicit Calibration

The approach chosen here performs an implicit calibration of the sensor inhomogeneities from arbitrary raw data acquired from a scene. It uses a rectification operator $r_{n,q}$ which is applied to correct the sensor raw data $\{y_{n,q}\}$ (5.12).

$$\hat{y}_{n,q} = r_{n,q}(y_{n,q}) = r_{n,q}(\gamma_{n,q}(u_{n,q}))\,, \tag{5.12}$$

with $\hat{y}_{n,q}$ being the rectified data of sample $y_{n,q}$. Note that $\gamma_{n,q}$ and $u_{n,q}$ are unknowns, which are not determined by the calibration process.

The goal of the rectification process is to generate a set of corrected raw data $\{\hat{y}_{n,q}\}$ such that each corrected output value $\hat{y}_{n,q}$ only depends on the theoretical input value $u_{n,q}$, and is no longer depending on the detection unit $q$ or sampling mode $n$ used for the measurement. Thus the requirement for $r_{n,q}$ is:

$$\begin{aligned} u_{n_1,q_1} = u_{n_2,q_2} \Rightarrow \hat{y}_{n_1,q_1} = \hat{y}_{n_2,q_2}\,, &\text{for all } n_1, n_2 \in \{1, \ldots, N\}\,, \\ &\text{and } q_1, q_2 \in \{1, \ldots, Q\}\,. \end{aligned} \tag{5.13}$$

Since a relative calibration is desired the data of only $Q-1$ taps have to be rectified. W.l.o.g. we choose $q = 1$ as the tap of which the data are trivially corrected, i.e. remain uncorrected. The raw data of all other taps are corrected for each possible sampling mode $n$, see (5.14).

$$r_{n,q}(y_{n,q}) = \begin{cases} y_{n,q} & \text{, if } q = 1 \\ r_{n,q}(\gamma_{n,q}(u_{n,q})) = r_{n,1}(\gamma_{n,1}(u_{n,1})) = \hat{y}_{n,1} & \text{, if } q \neq 1\,\text{,} \\ & \text{for each possible } u_{n,1} \\ & \text{and } u_{n,q} = u_{n,1}\,\text{,} \\ & n \in \{1, \dots, N\} \end{cases} \tag{5.14}$$

This means that there are $(Q-1) \cdot N$ independent nontrivial and $N$ trivial rectification operators $r_{n,q}$ for each pixel. The rectification operators are used to compensate deviations caused by the different detection units $q$ individually for each sampling mode $n$. Please note that this is only an implicit definition of $r_{n,q}$. It will be shown in the next section how $r_{n,q}$ is constructed.

### 5.3.2. Dynamic Sensor Calibration

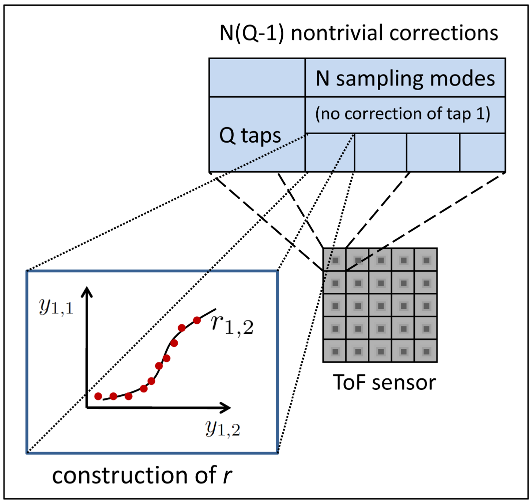


**Figure 5.2.:** Overview calibration: The rectification operator $r_{n,q}$ is a polynomial fit of $\{y_{n,1}\}$ over $\{y_{n,q}\}$ depicted here for $q = 2$, $n = 1$. For every pixel, $N(Q-1)$ nontrivial rectification operators have to be computed.

The rectification operators $r_{n,q}$ can be constructed by analyzing raw data delivered by a ToF system. Under the assumption that the observed scene is (temporarily) not changing, each tap (of a pixel) measures the same theoretical input, hence:

$$u_{n,q} \quad = \quad u_{n,1}, \; n \in \{1, \dots, N\} \tag{5.15}$$

Due to aforementioned different characteristic curves, the sensor output measured by different taps is usually not identical: $y_{n,q} \neq y_{n,1}$. The rectification operator $r_{n,q}$ is generated in such a way that (5.16) is valid for each pair $(y_{n,q}, y_{n,1})$.

$$r_{n,q}(y_{n,q}) = y_{n,1} \tag{5.16}$$

The rectification operator $r_{n,q}$ expresses the correlation of actually measured data ($y_{n,q}$) and the data which would have been measured with tap $q = 1$ ($y_{n,1}$). For an ideal sensor, $r_{n,q}$ would be the identity function.

The generation of $r_{n,q}$ can be done by collecting multiple pairs $\{(y_{n,q}, y_{n,1})_i\}$ and fitting a polynomial function to this data set. The rectification operator $r_{n,q}$ is then the polynomial function. It has to be computed individually for all taps $q \neq 1$, all sampling modes $n$, and all pixels. Please see Fig. 5.2 for a visualization of proposed calibration technique.

The assumption that the scene is static does not need to be fulfilled for all pixels simultaneously. Instead, static subsequences of the raw data signal can be found for every pixel individually and might be used for generation of $r_{n,q}$. Such static subsequences are usually present in all kinds of natural sequences. They can be identified by comparing the absolute temporal gradient of the raw data signal with a predefined threshold: If the absolute gradient of the raw data signal of a particular pixel is below this threshold, the pixel images a static object, so pairs of $\{(y_{n,q}, y_{n,1})_i\}$ can be extracted from the acquired raw data package.

### 5.3.3. Raw Data Rectification

The rectification operators $r_{n,q}$ may be used to compensate the effect of the different characteristic curves of the different taps. So, the averaging technique described in Sect. 5.2.2 is not needed anymore. Thus, each raw data package can be split into separate packages, which may be used to compute individual sets of scene unknowns. A package consisting of $L$ acquisitions of $Q$ taps can be split into $(L \cdot Q/N)$ subpackages of length $N/Q$. For instance each raw data package of a two-tap camera using $N = 4$ samples may be split into two *subpackages* carrying the full information necessary for reconstruction of the scene unknowns (see Fig. 5.3).

By pursuing this strategy multiple sets of scene unknowns can be computed from a single raw data package, and hence the frame rate is increased. Please note that since each subpackage carries the full information to compute the scene unknowns, these computed quantities are *independent.* For the given example the frame rate of the depth maps and all other computed scene unknowns is doubled.

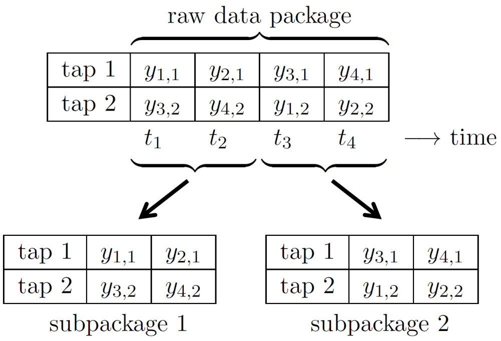

raw data package
tap 1 | $y_{1,1}$ | $y_{2,1}$ | $y_{3,1}$ | $y_{4,1}$
tap 2 | $y_{3,2}$ | $y_{4,2}$ | $y_{1,2}$ | $y_{2,2}$
$t_1$ $t_2$ $t_3$ $t_4$ ⟶ time
tap 1 | $y_{1,1}$ | $y_{2,1}$
tap 2 | $y_{3,2}$ | $y_{4,2}$
subpackage 1
tap 1 | $y_{3,1}$ | $y_{4,1}$
tap 2 | $y_{1,2}$ | $y_{2,2}$
subpackage 2

**Figure 5.3.:** Splitting a raw data package into two independent subpackages.

### 5.3.3.1. Extension: Using Interleaved Datasets

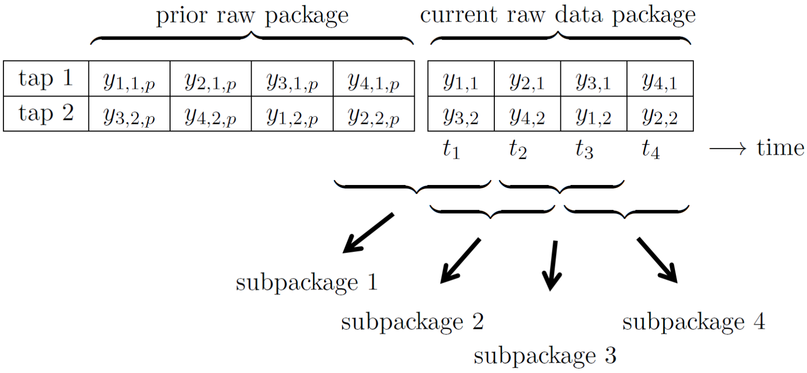
A further increase of the frame rate is possible by using interleaved subpackages. This requires data of the raw data package acquired prior to the considered one. These data will be denoted with an additional index $p$ here. A raw data package consisting of $L$ acquisitions can be split into $L$ interleaved subpackages. This enables the computation of $L$ sets of scene unknowns for each raw data package, corresponding to an increase of the frame rate by a factor of $L$. For instance the raw data package of the example discussed may be split into four interleaved subpackages (see Fig. 5.4).

Please note that using interleaved subpackages does not produce the same values as interpolating the scene unknowns generated from independent (i.e. not interleaved) subpackages would do. In other words, using interleaved subpackages does not correspond to applying a simple interpolation. The reason is that the reconstruction of scene unknowns from raw data is performed by nonlinear operations (see (2.8), (2.9)).

Here, the computed quantities are not independent since each subpackage has an overlap with its consecutive subpackage (see next section).

### 5.3.3.2. Frame Rate Increase

Using the proposed raw data rectification enables splitting the raw data packages into subpackages, which enables a significant frame rate increase. The averaging technique described in Sect. 5.2.2 is capable to compute one set of scene unknowns for each set of $N$ measurements (length of a raw data package). By applying proposed raw data rectification, each raw data package can be split into subpackages of length $N/Q$, of which each can be used to compute an independent set of scene unknowns.



|  |  |  |  |  |  |  |  |  |
|---|---|---|---|---|---|---|---|---|
| tap 1 | $y_{1,1,p}$ | $y_{2,1,p}$ | $y_{3,1,p}$ | $y_{4,1,p}$ | $y_{1,1}$ | $y_{2,1}$ | $y_{3,1}$ | $y_{4,1}$ |
| tap 2 | $y_{3,2,p}$ | $y_{4,2,p}$ | $y_{1,2,p}$ | $y_{2,2,p}$ | $y_{3,2}$ | $y_{4,2}$ | $y_{1,2}$ | $y_{2,2}$ |

**Figure 5.4.:** Splitting a raw data package into four interleaved subpackages.

Thus, compared to the averaging technique using $L = N$ acquisitions the frame rate increase is $N/(N/Q) = Q$, which corresponds to the number of taps used.

By using interleaved subpackages, a set of scene unknowns can be computed every time a new measurement is done (length of new data: 1). Consequently, from $N$ subpackages of the sequence, $N$ sets of scene unknowns may be computed, giving a frame rate increase of $N/1 = N$, which is the number of samples. It has to be noted that the same speedup would be feasible using an adapted averaging technique with a "sliding window". However, using data rectified by the proposed method significantly decreases the overlap of the used subpackages and therefore decreases the dependency of the generated sets of scene unknowns. Subpackages constructed from rectified data have an overlap of $(N/Q)/N = 1/Q$, compared to an overlap of $(N-1)/N$ when using the averaging technique. For example, rectified interleaved subpackages of a two-tap sensor employing $N = 4$ samples would have 50% overlap, compared to 75% overlap when using interleaved data in combination with an averaging technique.

## 5.4. Experimental Verification

For the experimental verification using real data a PMD CamCube 2.0 camera (PMDTechnologies, Siegen, Germany) was employed. This ToF camera utilizes a correlating sensor with two taps and thus represents a considerable class of commercially available 3D ToF systems[3].

[3] Among others, this class includes also cameras from Canesta and Mesa Imaging (c.f. Sect. 2.2.3).

A sequence of 250 raw data packages (per pixel) was acquired, which included four static subsequences. For the gathering of calibration data covering a big fraction of the available raw data range, nearly homogeneous “targets” of various reflectivities were presented at various distances to the camera. The objects serving as targets were casually chosen and positioned, since the quality of the targets was not important, but rather the fact that the input was (temporarily) static and of various intensities. In particular a cardboard (distance $d = 1\text{m}$), the lab’s carpet ($d = 2\text{m}$), the wall ($d = 4\text{m}$) and a piece of paper ($d = 0.5\text{m}$) were used. At the end of the sequence a rotating target was imaged which will serve to evaluate the success of the frame rate increase. This target consists of two opposing quadrants rotating around a common axis. A schematic representation is given in Fig. 5.5.

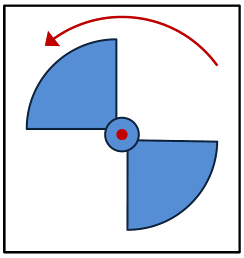


**Figure 5.5.:** Schematic representation of the rotating target used in the performed experiments.

As discussed in Sect. 3.1.5 the PMD camera has a system for active compensation of background light (called SBI) built in, which introduces a highly non-linear feature to the characteristic curve $\gamma$. For the proof of concept, correctly dealing with this highly specific system property does not provide any benefits. Therefore, the author decided to keep the algorithms simple and to acquire data without activation of the SBI. Since the SBI is activated automatically at high intensities, the absence of strong light sources ensured that the SBI was deactivated.

The sequence was processed offline using MATLAB scripts. Static subsequences were searched individually for each pixel. This was done by accepting all samples whose squared temporal gradient was below a threshold $\xi$:

$$\text{accept } y_{n,q}[t_1], \text{ if } \left(y_{n,q}[t_1] - y_{n,q}[t_0]\right)^2 < \xi \tag{5.17}$$

With $y_{n,q}[t_0]$ and $y_{n,q}[t_1]$ being two consecutive values acquired at time steps $t_0$ and $t_1$ ($t_0 < t_1$) of a specific raw channel and pixel. For the performed experiments

$\xi = 4000\text{DN}^2$ [4] was chosen. From these static subsequences, on average 191.3 pairs of $(y_{n,q}, y_{n,1})_i$ per pixel and raw channel were collected.

From previous investigations (see Sect. 3.2 and [EJ09]) it was known that a typical characteristic curve of the camera system at hand is well approximated by a linear function. Therefore a linear function (polynomial of degree 1) is well suited to model also the difference of two different characteristic curves.

Thus, for each pixel, each sampling mode $n$, and $q = 2$, a linear function (5.18) was fit to data points $\{(y_{n,q}, y_{n,1})_i\}$ using a least square fit, giving $r_{n,q}$ (5.19).

$$y_{n,1} = \beta_{n,q} + \alpha_{n,q} \cdot y_{n,q} \ , q = 2, n \in \{1, \ldots, N\} \tag{5.18}$$

$$r_{n,2}(y_{n,q}) = \beta_{n,q} + \alpha_{n,q} \cdot y_{n,q} \tag{5.19}$$

Here, $\beta_{n,q}$ is the offset and $\alpha_{n,q}$ the slope of the rectification operator $r_{n,q}$.

The process of generating $r_{n,q}$ is visualized in Fig. 5.6 for $q = 2$, $n = 1$ and a single representative pixel with coordinates $x_1 = 100$, $x_2 = 80$. The blue crosses represent all pairs $\{(y_{1,2}, y_{1,1})_i\}$ present in the input sequence. All the pairs belonging to static subsequences (identified by applying (5.17)), were used to compute $r_{1,2}$ and are labeled with red circles in the figure. The computed correction operator $r_{1,2}$ is visualized as green solid line.

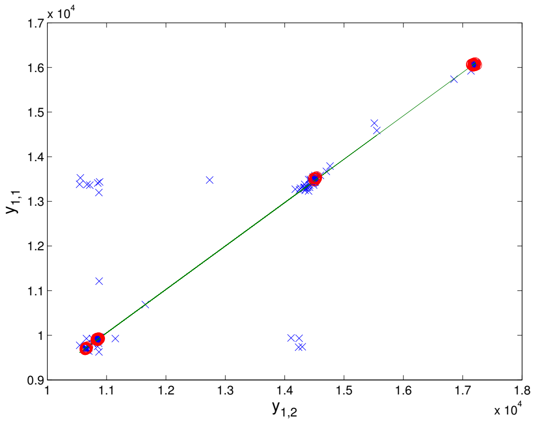


**Figure 5.6.:** Generation of the rectification operator $r_{1,2}$ for a single pixel with spatial coordinates $x_1 = 100$, $x_2 = 80$.

Fig. 5.6 suggests that these samples are clustered and not evenly distributed over the input range, which might result in a bad numerical fit. However, these clusters

[4] [DN] = Digital Number (physical unit of the sensor raw data)

correspond to the static subsequences, each representing a single target of the test sequence. Therefore, the clustered characteristic of the data is a result of the limited extent of the acquired sequence and does not allow any conclusions about the presented method.

The derived correction parameters for $n = 1$ and $q = 2$ are visualized in Fig. 5.7 for all pixels of the sensor. The structure in Fig. 5.7.a (horizontal stripes) probably contains clues about the internal hardware layout of the sensor (different amplification paths etc.). The artifacts visible in Fig. 5.7.b result from an imperfect distribution of the used data points and correspond to the shapes of the targets imaged in the input sequence.

Also parameters for $n \in \{2, 3, 4\}$ were computed, which are not visualized here[5].

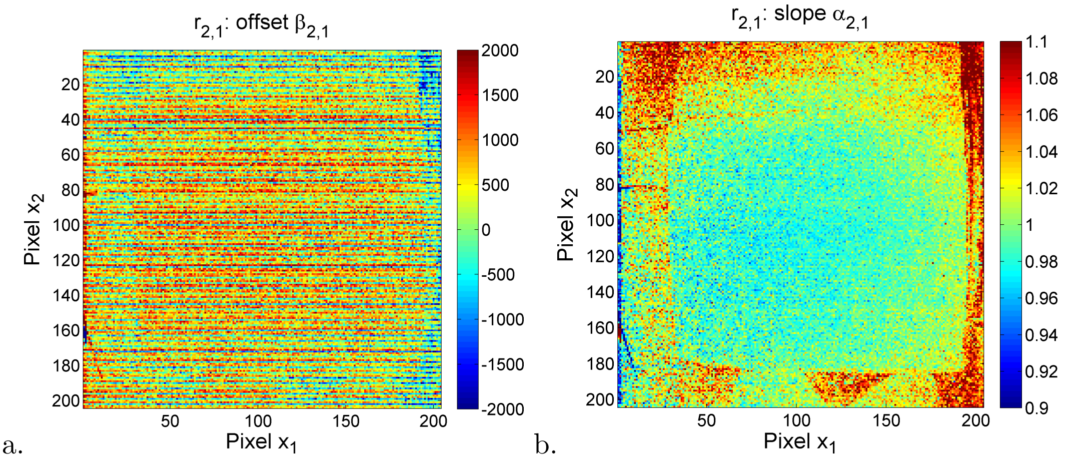


**Figure 5.7.:** Computed parameters of the rectification operators $\{r_{1,2}\}$, plotted for each pixel: **a** offset $\beta_{1,1}$, **b** slope $\alpha_{1,1}$.

The correction was applied to a single frame showing the mentioned rotating depth target: For each pixel, the raw data package was split into two subpackages (c.f.

[5] A computation of further parameters varying $q$ was not necessary, because the sensor uses only $Q = 2$ taps, and a correction of the first tap $q = 1$ is not required (see Sect. 5.3.1).

Fig. 5.3). All raw data measured with tap $q = 2$ were being corrected, while all data acquired with the first tap were trivially corrected:

$$\begin{aligned}
\hat{y}_{1,1} &= r_{1,1}(y_{1,1}) = y_{1,1} && (5.20)\\
&\vdots \\
\hat{y}_{4,1} &= r_{4,1}(y_{4,1}) = y_{4,1} && (5.21)\\
\hat{y}_{1,2} &= r_{1,2}(y_{1,2}) = \beta_{1,2} + \alpha_{1,2} \cdot y_{1,2} && (5.22)\\
&\vdots \\
\hat{y}_{4,2} &= r_{4,2}(y_{4,2}) = \beta_{4,2} + \alpha_{4,2} \cdot y_{4,2} && (5.23)
\end{aligned}$$

This operation was performed for each pixel individually, giving 8 corrected raw images: $\hat{\boldsymbol{Y}}_{1,1}, \hat{\boldsymbol{Y}}_{2,1}, \hat{\boldsymbol{Y}}_{3,1}, \hat{\boldsymbol{Y}}_{4,1}, \hat{\boldsymbol{Y}}_{1,2}, \hat{\boldsymbol{Y}}_{2,2}, \hat{\boldsymbol{Y}}_{3,2}, \hat{\boldsymbol{Y}}_{4,2}$.

From these corrected data, two *single* phase maps $\hat{\boldsymbol{\Phi}}_1$ and $\hat{\boldsymbol{\Phi}}_2$ were computed by using the assignments (5.5) and (5.1)–(5.4), and applying (2.9) on the data of each subpackage:

$$\begin{aligned}
\hat{\boldsymbol{\Phi}}_1 &= \arctan[(\hat{\boldsymbol{Y}}_{4,2} - \hat{\boldsymbol{Y}}_{2,1})/(\hat{\boldsymbol{Y}}_{3,2} - \hat{\boldsymbol{Y}}_{1,1})] && (5.24)\\
\hat{\boldsymbol{\Phi}}_2 &= \arctan[(\hat{\boldsymbol{Y}}_{4,1} - \hat{\boldsymbol{Y}}_{2,2})/(\hat{\boldsymbol{Y}}_{3,1} - \hat{\boldsymbol{Y}}_{1,2})] && (5.25)
\end{aligned}$$

For comparison, also two *uncorrected single* phase maps $\boldsymbol{\Phi}_1$, $\boldsymbol{\Phi}_2$ using uncorrected data of the subframes were computed as (5.26) and (5.27). Furthermore, an *averaged* phase map using the averaging technique described in Sect. 5.2.2 was generated by applying (5.8)–(5.11) and (5.28).

$$\begin{aligned}
\boldsymbol{\Phi}_1 &= \arctan[(\boldsymbol{Y}_{4,2} - \boldsymbol{Y}_{2,1})/(\boldsymbol{Y}_{3,2} - \boldsymbol{Y}_{1,1})] && (5.26)\\
\boldsymbol{\Phi}_2 &= \arctan[(\boldsymbol{Y}_{4,1} - \boldsymbol{Y}_{2,2})/(\boldsymbol{Y}_{3,1} - \boldsymbol{Y}_{1,2})] && (5.27)\\
\boldsymbol{\Phi}_{avg} &= \arctan[(\boldsymbol{C}_3 - \boldsymbol{C}_1)/(\boldsymbol{C}_2 - \boldsymbol{C}_0)] && (5.28)
\end{aligned}$$

From these phase maps depth maps were computed using (2.3), which are shown in Fig. 5.8.

### 5.4.1. Evaluation

The objective of this chapter is to show that a dynamic sensor calibration can be used to compensate for the inhomogeneities of the different taps in multi-tap ToF sensors, enabling an increased frame rate.

From Fig. 5.8 it can be seen that depth maps generated from uncorrected subpackages are heavily distorted (b, c). In contrast, the two depth maps generated from rectified subpackages (d, e) look very similar. By comparing the two separate depth maps, the motion of the rotating target can be recognized (counter clockwise). The comparison of (b, c) and (d, e) indicates that computing two independent depth maps from split raw data packages as proposed gives much better results, if the presented raw data rectification is used.

A quantitative analysis of these results is challenging. Please note that the presented method is working with camera raw data, delivered by an *uncalibrated* ToF camera. It is not meaningful to evaluate the absolute accuracy of the computed single depth maps, because also the absolute accuracy of the averaged depth map is unknown. Figure 5.8 shows slight deviations in the averaged depth map (a) compared to the single depth maps (d, e). However, without the ground truth of the dynamic scene and without an absolute raw data calibration of the ToF system (including temporal sensor effects), an evaluation of the absolute accuracy is not possible.

#### 5.4.1.1. Consistency

The goal of the method proposed in this chapter is increasing the frame rate, i.e. producing multiple consistent depth maps per frame. To measure this consistency, a *consistency measure* $\sigma_d$ of corresponding regions of two computed single depth maps was analyzed, which is defined here as follows:

$$\Delta d(x_1, x_2) \;=\; d_1(x_1, x_2) - d_2(x_1, x_2)\,, \tag{5.29}$$

$$\sigma_d \;=\; \sqrt{\frac{1}{K} \sum_{(x_1,x_2)\in \mathrm{A}} (\Delta d(x_1, x_2) - \mu_{\Delta d})^2}\,, \tag{5.30}$$

where $d_1(x_1, x_2)$ and $d_2(x_1, x_2)$ are the depth values of the two analyzed depth maps at position $(x_1, x_2)$, $\mu_{\Delta d}$ is the arithmetic mean of $\Delta d$ over the regarded area $A$, and $K$ is the number of pixels inside this area. Since the consistency measure $\sigma_d$ incorporates also statistical temporal fluctuations of the depth values, its theoretical limit is given by the temporal noise $\hat{\sigma}_t$ of the quantities $d_1$ and $d_2$. Assuming a perfectly consistent correction (i.e. $d_1 = d_2$), this optimal value reveals as

$$\sigma_{d,\mathrm{ideal}} = \sqrt{2}\,\hat{\sigma}_t\,. \tag{5.31}$$

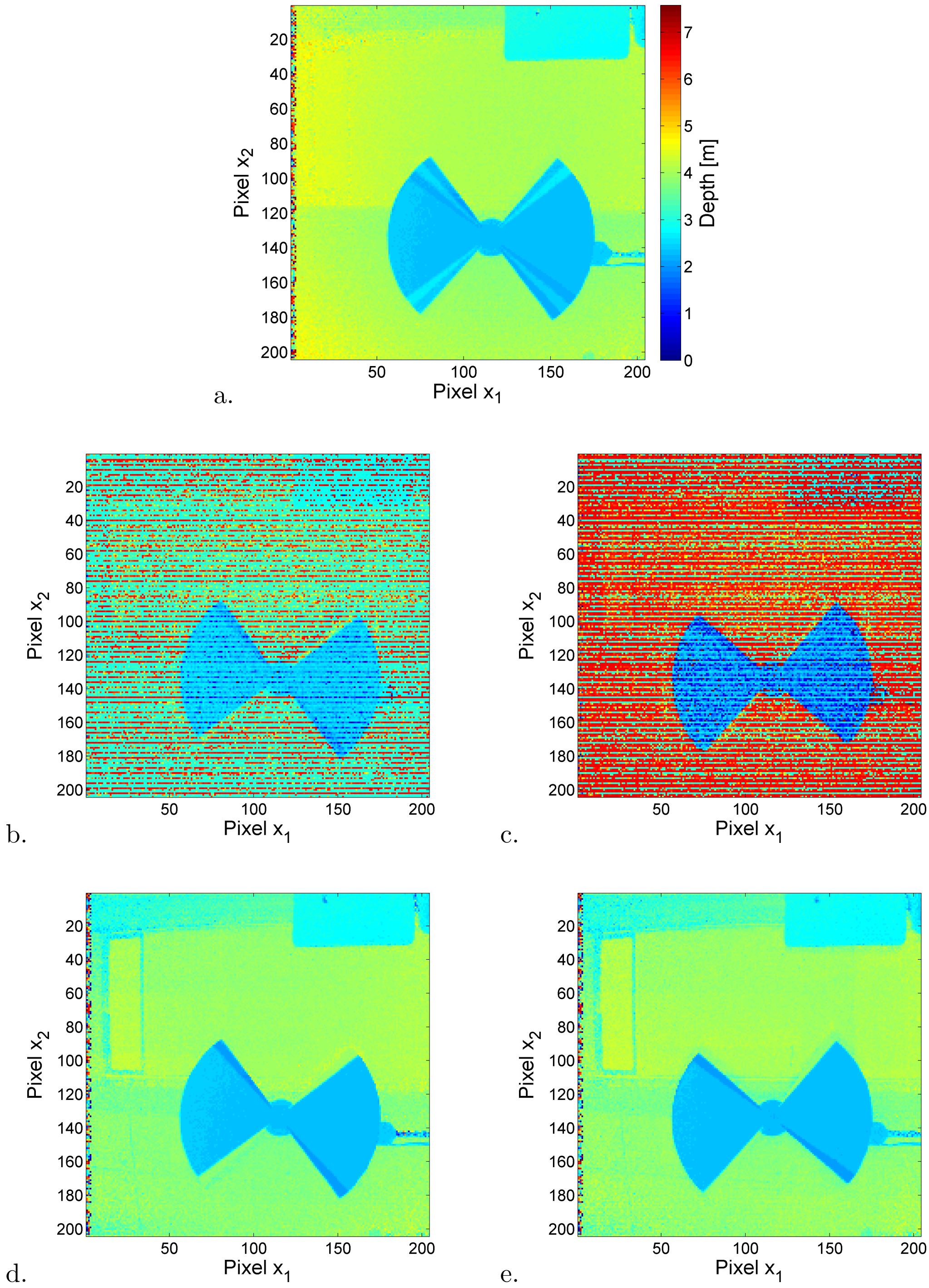


**Figure 5.8.:** Depth maps of rotating target. **a** Depth map using averaged raw data, computed from $\mathbf{\Phi}_{avg}$ (state-of-the-art). **b, c** Two depth maps generated from subpackages without correction (from $\mathbf{\Phi}_1$, $\mathbf{\Phi}_2$). **d, e** Two depth maps generated from subpackages corrected using proposed method (from $\hat{\mathbf{\Phi}}_1$, $\hat{\mathbf{\Phi}}_2$).

Here, as region $A$ for the analysis a small static area of the scene ($x_1 = 40 \dots 60$, $x_2 = 40 \dots 60$) was chosen. By computing the measure $\sigma_d$ on 25 consecutive depth maps the evaluation of its statistical properties was possible[6].

So, the consistency measure of the single depth maps generated from the corrected raw data was computed as $\hat{\sigma}_d = (0.1031 \pm 0.0065)$m. The corresponding quantity computed for the uncorrected single depth maps was determined as $\sigma_d = (2.8057 \pm 0.0031)$m. The value of the ideal measure was computed[7] as $\sigma_{d,\text{ideal}} = (0.0988 \pm 0.0173)$m.

Thus, the single depth maps generated from split raw data packages show a significantly higher consistency, if proposed raw data rectification is applied. Furthermore it can be seen that the presented method performs very close to the theoretical optimum.

Please note that motion artifacts (visible at the borders of the rotating target) are significantly removed in the single depth maps compared to the averaged depth map, since the data for computing each depth map were gathered in less time. For a detailed analysis of this property please refer to Chapter 6.

#### 5.4.1.2. Temporal Noise

In a second evaluation the temporal noise was analyzed. For this, the (temporal) standard deviation of the depth values of all pixels of the same area $A$ was computed over 25 consecutive frames. By averaging these values over the whole patch $A$ the mean temporal noise was determined. The (spatial) standard deviation of these values was taken to describe the statistical uncertainty of the mean temporal noise over the whole patch.

The resulting mean statistical depth error of the depth map computed using the averaging technique is $\sigma_{t,avg} = (0.0487 \pm 0.0080)$m. Since the single depth maps are generated using roughly half of the available light an increase of the noise by factor $\sqrt{2}$ leading to $\sigma_{t,exp} = (0.0689 \pm 0.0113)$m was expected.

The averaged statistical depth error for a depth value from one[8] of the two single depth maps computed from corrected data is determined as $\hat{\sigma}_t = (0.0699 \pm 0.0122)$m. This coincides nicely with the expected increase by a factor of $\sqrt{2}$.

[6] The consistency measure (5.30) was computed individually for each pair of depth maps. This was done for all of the 25 consecutive frames. The statistical mean and standard deviation of the computed measures were then determined and are given as value $\sigma_d$ and its error here.

[7] This computation was using the temporal noise and its error. The determination of these quantities is described in the following Section 5.4.1.2.

[8] For this evaluation the depth map computed from $\hat{\mathbf{\Phi}}_1$ was used.

### 5.4.2. Cross Temperature Test

An important reason for using a dynamic calibration method is the fact that the calibration parameters are not stable over time. In particular they depend on the temperature of the camera system. This section will investigate the influence of different temperatures on the quality of the generated single depth maps. It is based on experiments involving an active manipulation of the temperature of the camera housing. The quality of the generated depth maps will be evaluated using the consistency measure defined above.

#### 5.4.2.1. Experimental Setup

The basis of the following experiment is the same PMD CamCube 2.0 ToF camera used for the investigations in Sect. 5.4.1. Some hardware modifications were necessary in order to enable control of the temperature of the camera housing:

**Separation of light sources and camera body.** The original PMD CamCube 2.0 ToF camera system consists of three cubes: The camera body is placed in the center and enclosed by two light source elements which are attached directly at the camera.
For the experiments the distance between the light sources and the camera body was increased to about 1cm. This improves the thermal separation and enhances the control over the camera housing temperature.

**Active temperature control.** In order to actively manipulate the camera's temperature two Peltier elements (Quick-Cool QC-127-1.4-8.5M)were mounted on its top and bottom side. They were equipped with heatsinks and fans to improve the heat transportation and thus to increase the accessible temperature range. The Peltier elements were steered by a controller (Cooltronic TC 3224-RS232) enabling an active cooling or heating of the ToF camera.

**Temperature sensor.** A precise temperature probe was attached at one side of the camera body, approximately in the middle between the Peltier elements.

A photography of the modified ToF system is given in Fig. 5.9.

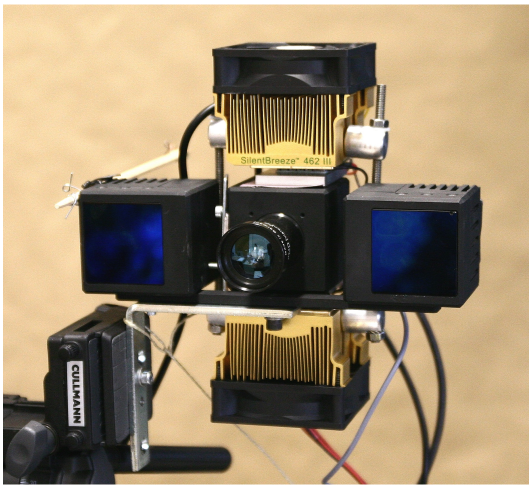


**Figure 5.9.:** Front view of the adapted ToF camera (middle) with attached cooling/heating elements (top, bottom) used for the performed experiments. The cubes on the left and right side of the camera constitute the ToF light source.

#### 5.4.2.2. Execution

For the experiment a number of 40 raw data sequences was acquired with each sequence comprising 250 raw frames. By using a sight on the camera and small targets in the lab it was possible to capture similar content in all sequences: In each sequence the camera was imaging the lab's carpet, a homogeneous surface covered with wrapping paper, and the lab's wall with the rotating target in front of it.

Every ten sequences the temperature was varied. Each time after selecting a new temperature an idle time of at least 15min was taken to allow the system to reach thermal equilibrium. During these idle times the camera was not switched off in order to minimize errors possibly caused later by the heating of the camera during acquisition.

The measured housing temperatures detected by the attached sensor present during acquisition of each sequence are given in Table 5.1. In this table also the mean temperature $\mu_T$ of each of the four groups is given.

On each sequence the dynamic estimation of calibration parameters was applied (using the same settings as described above). These calibration parameters then were used to rectify a specific frame of *all* acquired sequences. So, the calibration parameters were not only applied to the sequence used for their determination, but also to each other sequence.

| $j$ | $T \pm 0.02$ [°C] | $j$ | $T \pm 0.02$ [°C] | $j$ | $T \pm 0.02$ [°C] | $j$ | $T \pm 0.02$ [°C] |
|---|---|---|---|---|---|---|---|
| 1 | 18.25 | 11 | 27.85 | 21 | 36.32 | 31 | 43.56 |
| 2 | 18.24 | 12 | 27.85 | 22 | 36.29 | 32 | 43.50 |
| 3 | 18.23 | 13 | 27.86 | 23 | 36.39 | 33 | 43.44 |
| 4 | 18.23 | 14 | 27.88 | 24 | 36.12 | 34 | 43.48 |
| 5 | 18.22 | 15 | 27.89 | 25 | 36.01 | 35 | 43.48 |
| 6 | 18.21 | 16 | 27.89 | 26 | 35.99 | 36 | 43.51 |
| 7 | 18.19 | 17 | 27.93 | 27 | 36.00 | 37 | 43.47 |
| 8 | 18.18 | 18 | 27.92 | 28 | 35.99 | 38 | 43.51 |
| 9 | 18.12 | 19 | 27.94 | 29 | 36.00 | 39 | 43.58 |
| 10 | 18.10 | 20 | 27.92 | 30 | 35.97 | 40 | 43.57 |
| $\mu_T$ | $18.20 \pm 0.05$ | $\mu_T$ | $27.89 \pm 0.03$ | $\mu_T$ | $36.11 \pm 0.16$ | $\mu_T$ | $43.51 \pm 0.05$ |

**Table 5.1.:** Temperature of the camera housing during acquisition of each sequence (with index $j$) and mean temperature $\mu_T$ of each group.

The frame chosen for rectification was taken from the end of each sequence (frame number 218). By using exactly the same method as described above from each set of rectified raw images two single depth maps were generated. This was done for each combination of calibration and rectification sequences.

The consistency of each pair of single depth maps was evaluated by analyzing a homogeneous region in the background ($x_1 = 62 \dots 84$, $x_2 = 35 \dots 58$) and using the consistency measure $\sigma_d$ from Sect. 5.4.1.1.

#### 5.4.2.3. Results

The determined consistency error $\sigma_d$ for each pair of single depth maps, generated using each combination of calibration and rectification sequences is visualized in Fig. 5.10. Patches of size 10 × 10 can be seen, corresponding to regions of homogeneous temperatures.

Within the patches on the main diagonal the consistency error is minimal ($\sigma_d \in [0.05\text{m}, 0.08\text{m}]$), whereas in other patches the error increases up to $\sigma_d \approx 0.60\text{m}$. This means that combinations of calibration and rectification sequences matching the same temperature range lead to a minimal consistency error, indicating that the presented rectification works best if (as proposed) the required calibration parameters are gathered from the same sequence.

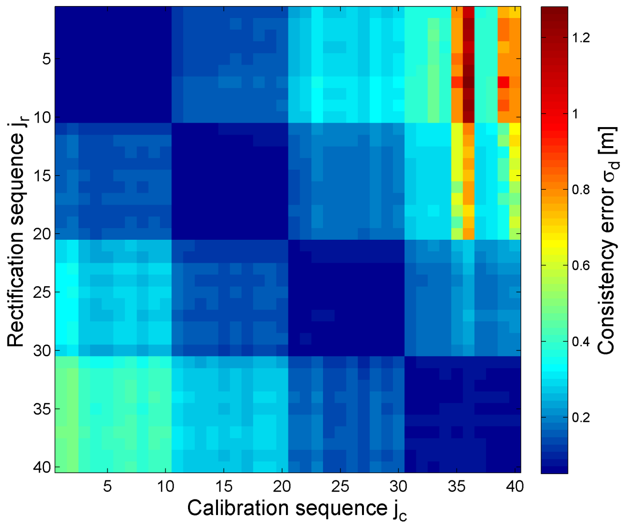


**Figure 5.10.:** Consistency error $\sigma_d$ for each possible combination of sequences used for calibration and being rectified. Patches of size 10 × 10 can be seen, corresponding to a similar temperature. The consistency error is minimized for patches on the main diagonal, corresponding to combinations matching the same temperature range.

The stripes in the upper right corner of Fig. 5.10 (e.g. with the coordinates: Calibration sequence $j_c = 36$, rectification sequence $j_r = 1 \ldots 20$) are outliers caused by an improper estimation of the calibration parameters. This probably originates from erroneous data acquisition, since it coincides with the observation of a very unstable system behavior during acquisition at these high temperatures (including lost camera connections, software driver crashes, etc.).

For a more detailed analysis the averaged consistency error and its uncertainty was determined for each patch by computing the statistical mean and standard deviation of the consistency measure over all pairs belonging to one temperature range (i.e. over the whole patch). These quantities are given in Table 5.2. The temperature differences between the four investigated temperature ranges are $(9.69 \pm 0.04)$°K, $(8.22 \pm 0.04)$°K, and $(7.40 \pm 0.04)$°K.

These data were used to compute the consistency error in dependence of the difference of the temperature present during acquisition of the calibration and rectification sequences. This relation is plotted in Fig. 5.11. For instance a sequence acquired at 27.89°C (green curve) rectified with calibration data gathered from a sequence

| | | | calibration $T$ [°C] | | | |
|---|---|---|---|---|---|---|
| | | $\sigma_d$[m] | 18.20 ± 0.05 | 27.89 ± 0.03 | 36.11 ± 0.16 | 43.51 ± 0.05 |
| rectification | $T$ [°C] | 18.20 ± 0.05 | 0.058 ± 0.003 | 0.148 ± 0.012 | 0.293 ± 0.021 | 0.597 ± 0.273 |
| | | 27.89 ± 0.03 | 0.148 ± 0.015 | 0.064 ± 0.004 | 0.181 ± 0.016 | 0.411 ± 0.166 |
| | | 36.11 ± 0.16 | 0.279 ± 0.028 | 0.143 ± 0.014 | 0.068 ± 0.007 | 0.195 ± 0.027 |
| | | 43.51 ± 0.05 | 0.414 ± 0.034 | 0.273 ± 0.015 | 0.155 ± 0.017 | 0.073 ± 0.007 |

**Table 5.2.:** Averaged consistency error $\sigma_d$[m] computed for each patch from Fig. 5.10, corresponding to homogeneous temperature ranges.

acquired at a temperature of 9.69°K below (18.20°C) results in a consistency error of $\sigma_d = 0.148$m.

For comparison, also the theoretical optimum $\sigma_{d,\text{ideal}}$ is given for each temperature range. It was computed by use of the temporal noise $\hat{\sigma}_t$. This noise was measured as described in Sect. 5.4.1.2 for each of the 40 sequences rectified with calibration parameters determined for the very same sequence. Within each temperature range these noise values were averaged, giving a mean noise value. From these quantities $\sigma_{d,\text{ideal}}$ was computed for each range as follows:

| $T$ [$°C$] | $\sigma_{d,\text{ideal}}$ [$m$] |
|---|---|
| 18.20 ± 0.05 | 0.0530 ± 0.0083 |
| 27.89 ± 0.03 | 0.0561 ± 0.0088 |
| 36.11 ± 0.16 | 0.0559 ± 0.0088 |
| 43.51 ± 0.05 | 0.0544 ± 0.0083 |

The figure (Fig. 5.11) indicates that the consistency error is rather independent from the absolute temperature, but determined by the difference of temperatures present during acquisition of the calibration and rectification sequence. For all investigated temperature ranges the consistency error is minimal, if this difference is zero. Deviations of this difference from zero led to an error which was increased. This result is statisticly significant for all analyzed temperature ranges.

For this reason it can be concluded that the proposed raw data rectification aiming at increasing the frame rate performs best if the calibration parameters are gathered from data acquired in the same temperature range as the data being rectified. Thus, the dynamic calibration outperforms static calibration approaches (using a fixed temperature), even for temperature differences as little as (7.40 ± 0.04)°K. Furthermore the comparison with the theoretical limit of the consistency error $\sigma_{d,\text{ideal}}$ shows that the presented approach performs very close at the theoretical optimum.

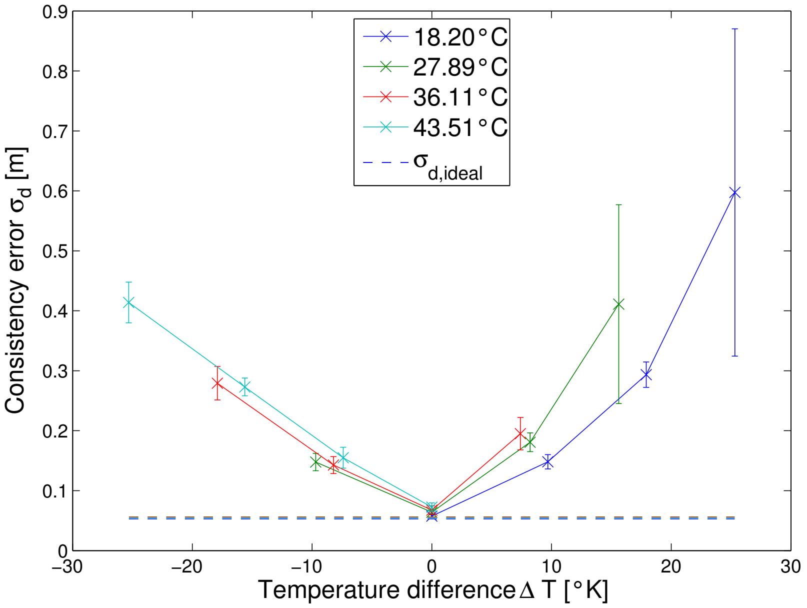


**Figure 5.11.:** Consistency error $\sigma_d$ in dependence of the temperature difference present during acquisition of the sequences used for calibration and rectification, plotted for different acquisition temperatures of the rectified sequences. For comparison, also the theoretical optimum $\sigma_{d,\text{ideal}}$ is given for each temperature range (see dashed lines in corresponding colors).

### 5.4.3. Computational Performance

All presented experiments were performed using a MATLAB implementation. The program runtime for loading the complete raw data sequence, generating the rectification operators $r_{n,q}$ and rectifying the data of a frame was about 50 seconds on a standard notebook PC [9]. This performance was obtained without computational optimizations. Since the generation of the rectification operators can be implemented recursively and applying them is very simple, the complete algorithm may be implemented computationally very efficiently. A realtime implementation of the proposed method is hence feasible, even on systems with limited hardware resources.

[9] Intel Core 2 Duo CPU P8600, 2.40GHz, 3GB RAM

## 5.5. Conclusion and Outlook

This chapter provided a proof of concept for performing an implicit dynamic calibration of the characteristic curves of the different taps of a ToF sensor. It has been demonstrated that the derived raw data rectification can be used for boosting the frame rate of ToF systems. The experimental results show that doubling the frame rate of a commercial two-tap ToF system is definitely feasible. The generated single depth maps are consistent and their statistical uncertainty increases as expected. By utilizing interleaved subpackages, a frame rate increase by a factor of four for the same system is possible.

The state-of-the-art averaging technique described in Sect. 5.2.2 makes the differences of the various characteristic curves $\gamma$ cancel out. However this is only valid, if these differences are described by a linear function. The approach presented here is able to handle higher order deviations by employing a higher order polynomial as rectification operator. Thus it is suitable to deliver data of higher accuracy compared to state-of-the-art solutions.

In opposite to the averaging technique the developed approach uses the acquisition of raw data performed in less time. As a result motion artifacts are significantly reduced.

The demonstrated method makes use of the static subsequences of a given raw data sequence. For applications in which such static subsequences do not occur (e.g. automotive), the generation of the rectification operators could be handled by temporarily interpolating the sensor raw data.

Using an extended hardware setup the influence of temperature variations on the consistency of the generated single depth maps was investigated. According to the results, differences of the camera temperature present during the acquisition of data used for calibration and data being rectified cause a decreased consistency of generated single depth maps. Thus, it was shown that the dynamic calibration method proposed here outperforms static calibration approaches. Furthermore, a comparison with the theoretical limit of the consistency error $\sigma_{d,\text{ideal}}$ revealed that the presented approach performs very close at the theoretical optimum.

In an application temperature changes could occur gradually or suddenly. It is therefore beneficial to implement routines allowing a temporal adaptation of the generated rectification operators.

Current ToF systems acquire raw data packages in a burst mode fashion. For optimally exploiting the proposed technique of enhancing the frame rate, it is advisable to adjust the temporal sampling of these acquisitions, such that the generated single depth maps correspond to an equitemporal sampling of the scene.

# Chapter 6.

# Reduction of Motion Artifacts

Time-of-Flight systems are capable of determining depth information by performing multiple measurements using different measurement modes. Each of these modes is used for capturing a particular raw image. The combination of raw images facilitates the estimation of all scene unknowns including the depth. Multi-tap sensors as discussed in the previous chapter enable to acquire some of the required raw images in parallel.

However, today's ToF systems are not able to perform all necessary measurements simultaneously but have to acquire them consecutively. Thus multiple (generally $L$) acquisitions are needed.

If the observed scene changes between these acquisitions, motion artifacts occur which significantly decrease the quality of the reconstructed scene unknowns. As an example please see the reconstructed depth map of a person performing a fast hand gesture in Fig. 6.1.

The goal of this chapter is to detail the origin of motion artifacts and to show possibilities to detect and significantly reduce them.

Beginning with a description of related work and motivating the following investigations in Sect. 6.1 the developed method for correction of motion artifacts will be explained in Sect. 6.2. Experimental results and a detailed evaluation will be given in Sect. 6.3. A conclusion and outlook are provided in Sect. 6.4.

The method presented in this chapter was applied for a patent in [SZ10b].

## 6.1. Related Work and Motivation

A method for reduction of such motion artifacts was proposed by Lindner and Kolb [LK09]. Their approach is based on an optical flow analysis of the individual raw images. The detected flow field is used to warp the raw images, such that corresponding

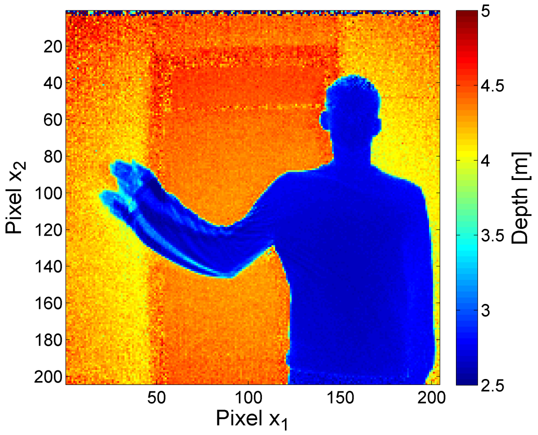


**Figure 6.1.:** Depth map of a person performing a fast gesture with his hand. Motion artifacts are visible on the edges perpendicular to the motion.

regions align. The method is able to reduce motion artifacts but it is computationally very demanding. Furthermore it has to deal with challenges like an ambiguous or incomplete estimation of the optical flow field and an appropriate normalization of the individual raw images.

A morphological method aiming to reduce motion artifacts was proposed by Gokturk, Yalcin and Bamji [GYB04]. Starting from a depth segmentation into foreground and background objects, pixels located near depth edges are found. These pixels are replaced by synthetic values using a spatial filtering process. This method reduces motion artifacts caused by depth edges between two different layers (foreground – background).

Lottner et al. [Lot+07] proposed to employ data of an additional high resolution 2D sensor being monocularly combined with the 3D sensor. Edges detected in the 2D image were used to identify critical areas in the raw images of the ToF system. By use of neighboring raw values and incorporating information from the 2D image these critical samples were replaced. Unfortunately, this approach requires additional hardware (2D sensor) which has to be spatially aligned and temporally synchronized with the ToF system.

For the special case of continuous-wave ToF sensors using $\tilde{N} = 4$ samples of the correlation function, Schmidt [Sch08a, page 88-92] derived how the estimated values of phase shift $\varphi$ and amplitude $a_1$ are altered due to motion artifacts. He also presented a method for detecting motion artifacts based on the symmetry of the correlation function (also valid only for $\tilde{N} = 4$).

A very similar idea was used by Hussmann et al. [HHE11] who evaluated the sum of two raw images acquired simultaneously. The difference of two of these sums (corresponding to two different subframes) indicates regions disturbed by motion artifacts. Unfortunately, the method is vulnerable against variations of the intensity measured by different pixels. Therefore it requires a photometric calibration of the camera and is restricted to a very limited depth range (they report $90\text{cm} - 100\text{cm}$). Furthermore, only lateral movements along one dimension may be corrected. Hence, the method is limited to very specific applications; they propose its usage for observing objects on a conveyor belt.

In contrast, the approach presented here tackles the problem of motion artifacts on a more abstract level. The occurrence of these artifacts is interpreted as a consequence of disturbances of the raw data acquired by an arbitrary ToF camera. The presented approach does not employ any spatial information or relations between different pixels but solely temporal information of single pixel signals. By detecting temporal discontinuities of the raw signals the events causing motion artifacts can be identified. By replacing raw values inducing artifacts with undisturbed values of prior acquisitions it is possible to prevent distortions.

The method is not limited to artifacts caused by depth edges, nor to artifacts occurring in a specific depth layer. It is also not restricted to object movements along a specific direction and requires no calibration. Instead, the approach presented here detects any disturbing influences and is able to correct most of them. Furthermore, the method is not limited to a specific ToF implementation but is generally valid for all ToF systems. Because of the method's simplicity its implementations are computationally very efficient.

## 6.2. Robust Correction of Motion Artifacts

### 6.2.1. Origins of Motion Artifacts

In typical depth imaging applications, three quantities are unknown and have to be determined for each pixel individually: the object's distance, its reflectivity and the intensity of non-modulated light (comprising ambient light and non-modulated light emitted by the light source of the ToF system and backscattered by the scene). To determine these three scene unknowns at least three measurements have to be performed. In a general formulation $N$ raw images have to be acquired by the ToF system. This is done using $L$ acquisitions, where $L \neq N$ is allowed if multi-tap systems are used. The unknowns are computed for each pixel individually using a *set* of such acquisitions.

There are several reasons, why current ToF systems do not deliver optimal depth maps for moving sceneries. One reason is the motion blur affecting each raw image. Since the ToF camera is integrating the incident signal over a certain time window, edges and fine details of moving objects are blurred.

A further and usually more serious reason is the temporal delay between the raw data acquisitions. If one or multiple of the scene unknowns change during the process of acquisition of a raw data set, the computed depth of affected pixels is incorrect. More precisely, if at least one of the three unknowns (depth, background light, reflectivity) changes, the reconstruction of all scene unknowns generates incorrect results.

So an obvious but technically hard to implement option is to reduce the number of required acquisitions. This approach was investigated and put into practice by the dynamic calibration method presented in Chap. 5. Its results in terms of a reduction of motion artifacts will be discussed in Sect. 6.3.1 and compared to results of the method investigated here.

Motion artifacts are caused by changes in the scene during acquisition of the required raw images. Thus, in a narrow sense, dynamic scenes imaged with today's ToF systems permanently determine scene data which is slightly altered due to motion effects. However, significant artifacts of the estimated depth map occur only for rapid changes in the scene. It should be noted that also other data channels generated by the ToF system, e.g. describing the measured intensity of background light or the modulation amplitude of the detected signal (typical for continuous-wave ToF systems), will contain corrupt data in this case. The following work will focus on the computed depth, but the reasoning and derived algorithm is also valid for all other scene unknowns.

Motion artifacts are caused by moving or rapidly changing features, for instance moving depth- or reflectivity edges. If the movement is parallel to the projection beam of a specific pixel, the signal deviations affecting this pixel are small due to the usually low speed of the objects (Fig. 6.2, movement 1). In contrast, laterally moving features effect fast changes of raw values because the edge is entering or leaving the area imaged by a specific pixel (Fig. 6.2, movement 2). These rapid changes result in large errors in the reconstructed scene unknowns. For this reason, motion artifacts are usually visible at the edges between objects of foreground and background.

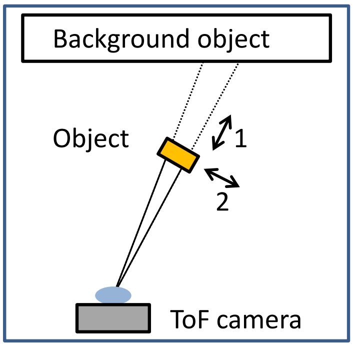


**Figure 6.2.:** A single pixel imaging a moving object. Movement 1 will produce only little motion artifacts. In contrast, movement 2 causes large discontinuities in the raw data signals, and thus will generate significant motion artifacts.

Regarding the temporal signal of one raw channel of one pixel, a discontinuity occurs at the instant of time the edge hits the pixel. This occurrence of a temporal discontinuity will be denoted as *event* in the following.

As an example, Fig. 6.3 depicts a ToF system using sets comprising samples of $L = 4$ acquisitions. For simplicity only the first raw channel of each acquisition is visualized[1]. For each of these raw channels a possible temporal progress of the raw signals is shown. These (unknown) signals are sampled at discrete points in time (red dots). Please note that the four raw values acquired for each set are sampled at different times. The occurrence of events causes discontinuities of the raw signals. These discontinuities of each raw value (relative to the corresponding value of the prior set) are also shown in the figure (in square brackets, 0: continuous samples, d: discontinuity).

If all raw values constructing the depth value are acquired before an event occurs, the computed depth value is correct (Fig. 6.3, set 2). In case of all raw values constructing the depth value being acquired after the event, the depth value is also correct (but represents another state of the scene) (Fig. 6.3, set 4). However, if the depth value is constructed by combining raw values acquired before with raw values acquired after the event, the computed depth information is incorrect (Fig. 6.3, set 3). Incorrect in this sense means that the value does not represent the state of the scene neither before nor after the event. Generally, it is also not between these values, meaning it is not comparable to an averaged measurement, but lays somewhere in the

[1] A multi-tap system using two taps would generate two raw values per acquisition. This means each acquisition would measure samples of two raw channels.

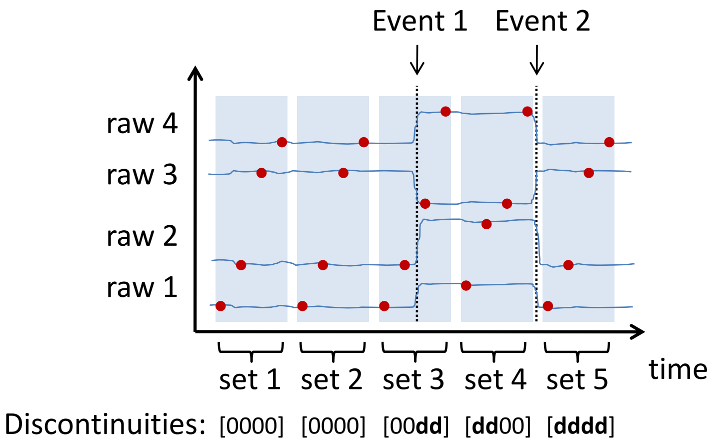

Event 1
Event 2
raw 4
raw 3
raw 2
raw 1
time
set 1
set 2
set 3
set 4
set 5
Discontinuities: [0000] [0000] [00**dd**] [**dd**00] [**dddd**]

**Figure 6.3.:** Temporal progress of four raw channels of which samples (red dots) are acquired at different time steps. Discontinuities are caused by events. These discontinuities are detected for each set by comparing its samples with the corresponding values of the previous set.

available depth range[2]. Therefore, such motion artifacts highly degrade the quality of generated depth maps.

The solution proposed here focuses on the detection and correction of this kind of motion artifacts caused by temporal discontinuities of the raw signals. They originate for example from lateral movement of scene features. This lateral movement is the most critical contribution to motion artifacts in practical systems. Its correction will lead to depth maps, of which the effect of residual motion artifacts is negligible for most applications.

### 6.2.2. Detection

The basic assumption here is that raw values acquired by each pixel of a ToF system vary smoothly over time. Significant artifacts are caused by laterally moving depth- and/or reflectivity-edges. By analyzing the temporal signal of a single raw channel of a single pixel, such event can be identified as a discontinuity. Thus, a powerful tool for identifying motion artifacts is to evaluate the temporal gradient of each raw channel. If the absolute temporal gradient exceeds a predefined threshold, the regarded raw channel is labeled as discontinuous for the current time step (see Fig. 6.3, bottom).

[2] For an analytical derivation describing a special case (continuous-wave ToF system, $\tilde{N} = 4$), see [Sch08a].

Optionally, if a ToF system capable of performing multiple measurements simultaneously is used, all simultaneously acquired raw values can be labeled as discontinuous if one of them was detected to be discontinuous. This analysis is performed for each raw channel of each pixel individually.

An event may occur between the acquisition of two subsequent samples or during the acquisition of one sample. Current systems normally have very short integration times (for single samples) compared to the delay between the samples. Therefore the possibility of an event occurring during the acquisition of two samples is much more likely. However, if it occurs during the acquisition of one sample, it will (depending on its exact temporal occurrence and the setting of the threshold) lead to a detected discontinuity of the current sample or of the subsequent sample. Both cases will be handled correctly by the proposed method.

Significant motion artifacts correlate with such discontinuities but not each detected discontinuity of a raw data signal indicates a motion artifact. Motion artifacts occur only if the event causing the discontinuity happens within a set of raw data processed to compute a depth value, i.e. between the individual acquisitions forming a set. Therefore, to determine if for a given set a motion artifact is generated, it has to be evaluated (using the discontinuity information) if the event occurred inside or outside the set.

Under the assumption that only one event occurs within two consecutive sets there exist only the cases shown in Fig. 6.4. From these possible cases only case 5, case 6 and cases potentially laying between both cause motion artifacts. From these critical cases a rule can be derived which indicates that the event occurred inside the set:

Rule:

```
IF (the first sample is not discontinuous)
    AND (at least one of all following samples is discontinuous),
THEN (a motion artifact will occur).
```

With this knowledge of what exactly on the level of raw data causes motion artifacts it is not only possible to detect but also to correct them.

### 6.2.3. Correction

Figure 6.5 gives a visualization of the correction algorithm proposed here. The rule derived in Sect. 6.2.2 is evaluated for each pixel. If a raw data set is found to generate

| set | | previous set | | | | | | | current set (to be analyzed) | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| acquisition $l$ | | 1 | | 2 | ... | | $L$ | | 1 | | 2 | ... | | $L$ | comment |
| case 1 | E | d | | d | ... | | d | | 0 | | 0 | ... | | 0 | no event in set |
| case 2 | | 0 | E | d | ... | | d | | d | | 0 | ... | | 0 | no event in set |
| | | | | | $\ddots$ | | $\vdots$ | | $\vdots$ | | | $\ddots$ | | $\vdots$ | $\vdots$ |
| case 3 | | 0 | | 0 | ... | E | d | | d | | d | ... | | 0 | no event in set |
| case 4 | | 0 | | 0 | ... | | 0 | E | d | | d | ... | | d | no event in set |
| case 5 | | 0 | | 0 | ... | | 0 | | 0 | E | d | ... | | d | **event in set** |
| | | | | | $\ddots$ | | $\vdots$ | | | | | $\ddots$ | | $\vdots$ | $\vdots$ |
| case 6 | | 0 | | 0 | ... | | 0 | | 0 | | 0 | ... | E | d | **event in set** |
| case 7 | | 0 | | 0 | ... | | 0 | | 0 | | 0 | ... | | 0 | no event in set |

**Figure 6.4.:** Possible occurrence of an event (indicated by the letter E) and detected discontinuities. The right set is to be analyzed. Only case 5, case 6, and cases potentially laying between both cause motion artifacts.

a motion artifact (c.f. Fig. 6.5, set 3), the proposed correction method simply is to overwrite raw values of channels experiencing a discontinuity with the corresponding values of the prior acquisition. If also the corresponding values of the prior acquisition were corrected, the procedure should use the original (uncorrected) values[3]. In case of a correction, the altered pixels are labeled as corrected which might provide useful information for further processing steps.

In such case of a correction, the computed information value (of this single pixel) does not represent the current state of the scene but corresponds to a prior state. However, this temporal misestimation of the scene state is less than the temporal distance of two depth maps and thus is negligible in most applications.

### 6.2.4. A Variant: Burst Internal Detection

Today's ToF systems usually acquire the required raw images in a burst mode fashion in order to capture a similar state of the scene and thus to minimize motion artifacts. This means the temporal distance between consecutive acquisitions of a single frame is normally much smaller than the delay between the last acquisition of one frame and first of the following one. Since the proposed method for correction of motion artifacts requires that at most one event occurs during two consecutive sets, this delay between two bursts is the principal factor limiting the effectiveness of the algorithm.

A very promising possibility to overcome this limitation is opened up by the dynamic calibration method presented in Chap. 5. By using a ToF system with a multi-tap

[3] However, this case should never arise since it is only possible if multiple events occur in two consecutive frames, which violates the assumption made in Sect. 6.2.2.

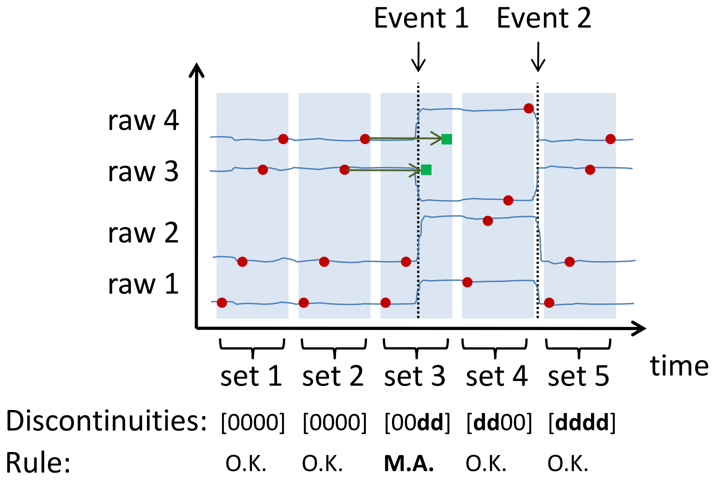

Event 1
Event 2
raw 4
raw 3
raw 2
raw 1
set 1
set 2
set 3
set 4
set 5
time
Discontinuities: [0000] [0000] [00dd] [dd00] [dddd]
Rule: O.K. O.K. M.A. O.K. O.K.

**Figure 6.5.:** Proposed correction algorithm. Discontinuous values of sets found to generate motion artifacts are overwritten using corresponding values of the prior set (green squares).

sensor this method enables to split each set of acquisitions into subsets carrying the full information to reconstruct the scene unknowns. So for example, the data of a PMD CamCube 2.0 camera from PMD Technologies was demonstrated to be split successfully into two independent subsets per frame, enabling a doubling of the frame rate.

Exactly this feature facilitates here to perform the comparison between two different raw values for detection of discontinuities within one burst, or – using the terminology from Chap. 5 – within one raw data package. This enables a more robust correction of all subsets of a given set except from the first one. So for instance, in the case of using a PMD CamCube 2.0 camera the second set (giving $\mathbf{\Phi}_2$)[4] can be corrected for motion artifacts using this burst internal detection. Please see Fig. 6.6 for a visualization. Using split sets enables the correction of Event 3 which follows shortly after Event 2. Without splitting of the sets the correction would have led to an artifact[5].

[4] Please note that in this chapter $\mathbf{\Phi}_2$ always represents the phase map generated from rectified raw data. This was called $\hat{\mathbf{\Phi}}_2$ in Chap. 5 and means that the notation was adapted in the present chapter in order to improve the document's readability.

[5] Using the full set of four acquisitions the two Events 2 and 3 generate a discontinuity pattern of [dddd] in set 5. Since all values are detected as discontinuous, a correction is not performed according to the rule from Sect. 6.2.2, which leads to an artifact in this case.

The reason of the failure of the standard method here is the violation of the assumption of only one event within two compared sets.

In the following this burst internal detection will be referred to as *BID* method, whereas the discussed comparison using two (normal) sets will be called *standard* mode.

Please note that also an uncorrected depth map computed from a subset shows less distortion caused by motion artifacts. This is the case because the data required for computing the map is acquired in less time, causing that a moving object is able to affect fewer pixels (compared to a depth map generated using the averaging technique, c.f. Sect. 5.4.1.1 and Fig. 5.8).

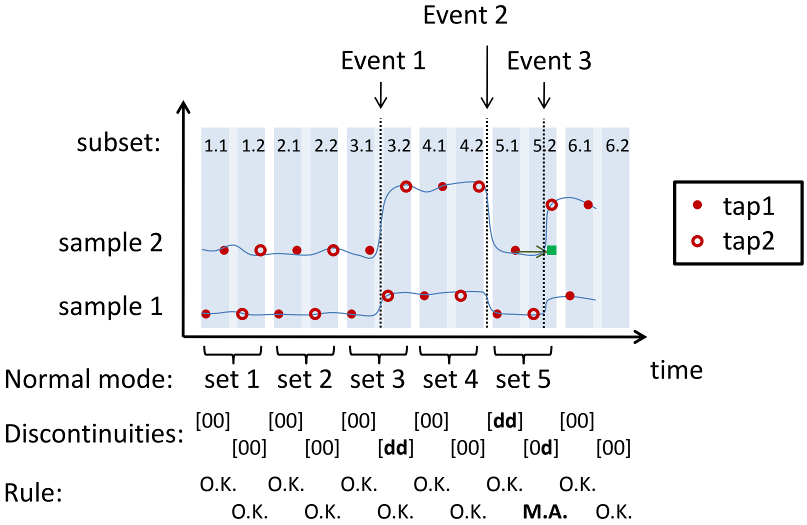

Event 2
Event 1
Event 3
subset:
1.1 1.2 2.1 2.2 3.1 3.2 4.1 4.2 5.1 5.2 6.1 6.2
sample 2
sample 1
● tap1
○ tap2
Normal mode:
set 1 set 2 set 3 set 4 set 5
time
Discontinuities:
[00] [00] [00] [00] [dd] [00]
[00] [00] [dd] [00] [0d] [00]
Rule:
O.K. O.K. O.K. O.K. O.K. O.K.
O.K. O.K. O.K. O.K. M.A. O.K.

**Figure 6.6.:** The dynamic calibration method proposed in Chap. 5 facilitates to split each set of acquisitions into two subsets. Using this technique each sample of the correlation function is determined using single measurements of different taps. This enables to correct artifacts caused by events with a short temporal distance (Event 3). Furthermore fewer of the occurring events require a correction because of the higher sampling density (Event 1, c.f. Fig. 6.5).

## 6.2.5. Performance

### 6.2.5.1. Geometric Properties

The proposed procedure preserves the geometry of a moving object for two reasons: First, the correction affects only a small fraction of the object (its edges). Second, opposite edges of the object are affected in an opposite manner, so the average spatial effect of the applied correction is zero. For example, an in the image horizontally moving object generates motion artifacts on its left and right edge (see Fig. 6.7). The correction algorithm overwrites the raw values of affected pixels with the corresponding values of the prior acquisitions. This results in a shift of these edges in the opposite direction of the current movement (in Fig. 6.7 to the left). Since both edges are moved by the same distance, the area of the object does not change. Solely its position is adjusted and corresponds to the object's location during the acquisition of the first raw channel.

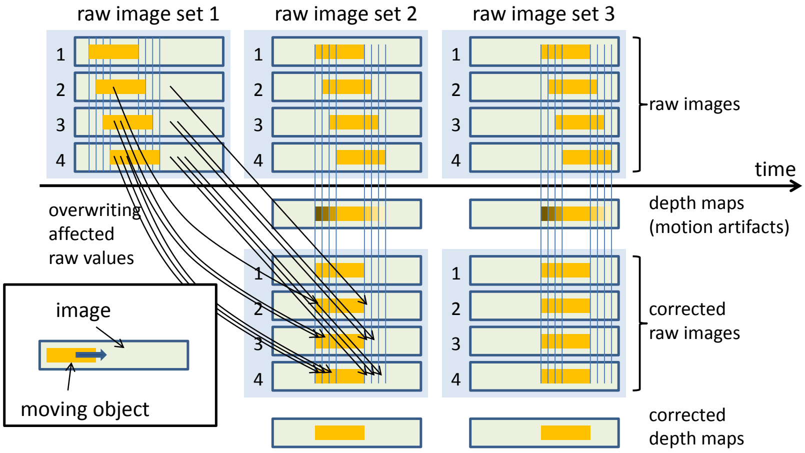


**Figure 6.7.:** The proposed raw data correction preserves the geometry: A horizontally moving object causes motion artifacts on its left and right edge due to the temporal delay between the raw image acquisitions. The algorithm overwrites raw values of pixels which would cause motion artifacts with corresponding values of the prior acquisitions. Thus, the corrected depth maps are free of motion artifacts and represent the state of the scene during acquisition of the first raw image.

#### 6.2.5.2. Computational Performance

The computational complexity of the proposed algorithms for detecting and correcting motion artifacts is $O(k)$, with $k$ being the number of pixels. Both methods are based on only a few operations per pixel and raw channel. Therefore, these algorithms may be implemented in a computationally extremely efficient manner. Implementations for real-time applications are thus definitely feasible, even on systems with limited hardware resources (e.g. embedded systems running inside a ToF camera).

#### 6.2.5.3. Limitations

The algorithm is able to handle a single discontinuity event occurring in two consecutive raw data sets correctly. I.e. more than one depth- or reflectivity edge imaged by one pixel while acquiring two sets of raw data used for the detection will lead to an erroneous reconstruction. The use of a multi-tap system in combination with the dynamic calibration method from Chap. 5 allows to employ multiple sets out of a burst (BID method). This significantly relaxes the requirement on the temporal distance of two events.

The assumption of one event within two sets could be violated for example by a small fast moving object. The proposed correction algorithm would handle the situation by replacing the raw values of the object by raw values describing the background. So, the fast object would disappear in the generated depth map. However, by analyzing the regions marked as corrected the system would still be able to detect the presence of such an object.

## 6.3. Experimental Results

The presented method is simple to implement, but an appropriate evaluation is challenging. A meaningful quantitative evaluation requires some knowledge about the imaged scene (ground truth information), which is only providable for simple settings. On the other hand the practicability of the method is demonstrated best using natural scenes with a high complexity. Therefore, the experimental verification of the algorithm will be twofold:

Section 6.3.1 investigates a scenario employing the algorithm on data acquired using a controlled setup. It proposes a measure to describe the distortion introduced by motion artifacts and will compare different variants of reducing these errors.

In Sect. 6.3.2 the algorithm's performance will be investigated using two natural scenes. Since an acquisition of ground truth data for such complex scenes is extremely difficult the analysis of the results will focus on a discussion of the produced visual output.

### 6.3.1. Controlled Scenario: Rotating Target

The proposed method for detection and correction of motion artifacts was investigated in a scenario using the rotating target also employed in Chap. 5. This depth target was imaged by a CamCube 2.0 ToF camera (PMD Technologies, Siegen, Germany). The camera acquires $R = 8$ raw images using $L = 4$ acquisitions. From these raw images $N = 4$ samples of the correlation function are computed by use of the averaging technique (c.f. Sect. 5.2.2), which are employed to reconstruct the scene unknowns.

A sequence of 250 frames was acquired and processed using MATLAB scripts. From this sequence a single frame was chosen, and the required discontinuity information was determined as follows: A raw value was labeled as discontinuous if its squared temporal gradient was above a threshold $\zeta$:

$$\text{label } y_{n,q}[t_1], \text{ if } \left(y_{n,q}[t_1] - y_{n,q}[t_0]\right)^2 > \zeta \tag{6.1}$$

With $y_{n,q}[t_0]$ and $y_{n,q}[t_1]$ being two consecutive values acquired at time steps $t_0$ and $t_1$ ($t_0 < t_1$) by a specific raw channel of a certain pixel. Here, $\zeta = 5 \cdot 10^5 \mathrm{DN}^2$ was chosen.

Using these discontinuity information the algorithms for detection and correction of motion artifacts were run. From the raw data corrected for motion artifacts a depth map was generated utilizing the (standard) averaging technique. For comparison also a depth map without using any correction was computed.

The results are shown in Fig. 6.8. In Fig. 6.8.a the uncorrected depth map of the rotating target is visualized. Please note the considerable motion artifacts at the target's laterally moving edges. The employed camera uses $L = 4$ acquisitions. Between these acquisitions $L-1 = 3$ pauses occur which lead to three distinguishable distorted regions at each laterally moving edge of the target. These artifacts are clearly visible at the upper left and bottom right edge. The upper right and bottom left edges, however, seem to show only two distorted regions each. This is a visual illusion: In fact, also here three differently distorted regions are generated, but one of them coincidentally produces depth values which match the depth values of the background.

As aforementioned the erroneous combination of raw samples causing motion artifacts results in depth values which can lay anywhere in the available depth range. This means that the distortion depends on the properties of the imaged moving objects (e.g. their depth and reflectivity). The visibility of only two distorted regions for two of the four edges is a good example for the difficulty of identifying regions distorted by this kind of artifacts. In Sect. 6.3.1.1 this point will be detailed.

Applying the correction strategy proposed in Sect. 6.2.3 gave a corrected depth map (see Fig. 6.8.b). It can be seen that the corrected depth map reproduces the plane surface of the depth target better than the original depth map. The artifacts at the edges were successfully removed, and the target's geometry was preserved.

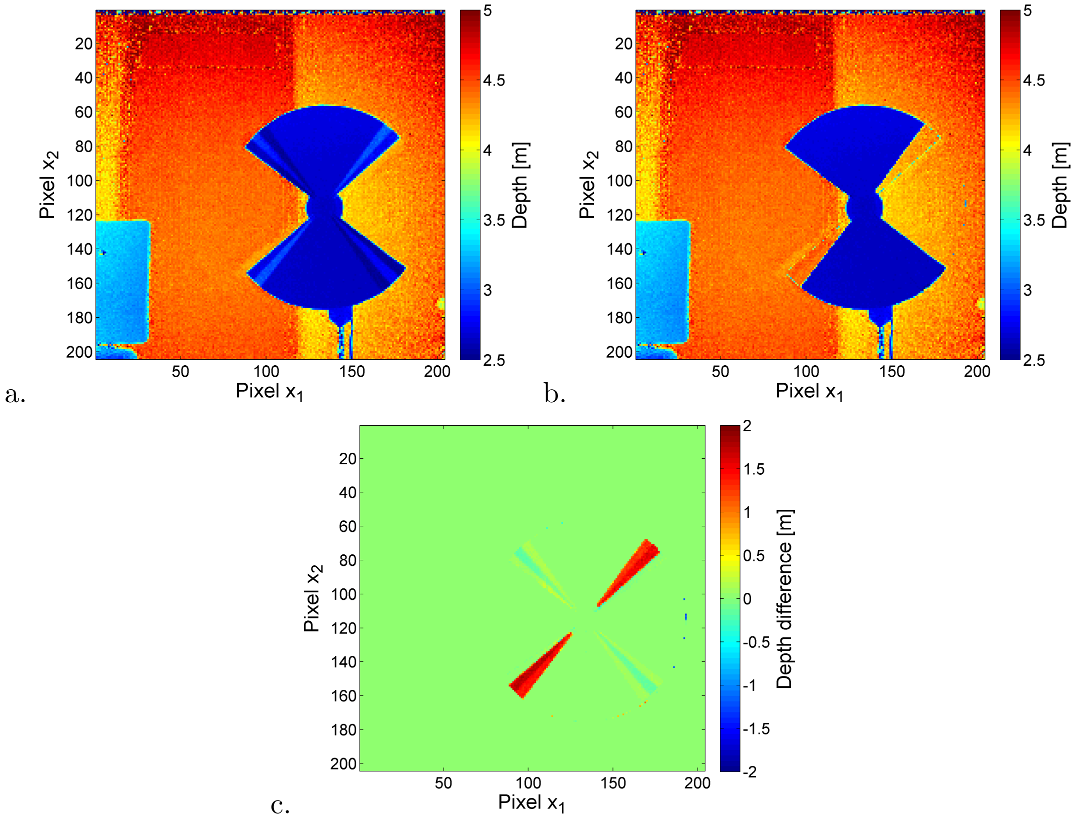


**Figure 6.8.:** Correction of motion artifacts for depth map of a rotating depth target. **a** Motion artifacts are visible on the edges perpendicular to the motion. **b** Corrected depth map without motion artifacts. **c** Difference map of a and b.

This result was obtained by evaluating the rule for detection of motion artifacts from Sect. 6.2.2. The derived information of erroneous acquisitions requiring to be corrected is visualized in Fig. 6.9. It is shown for each of the acquisitions $l \in \{2, 3, 4\}$ (Fig. 6.9.a - c). Please note that according to the correction algorithm an adjustment of the samples of the first acquisition $l = 1$ is never required. Fig. 6.9.d shows the sum of the three computed masks, representing the number of corrected acquisitions.

#### 6.3.1.1. Quantitative Evaluation

The quantitative evaluation of the degradation of depth maps caused by motion artifacts, and thus also the evaluation of the performance of algorithms reducing

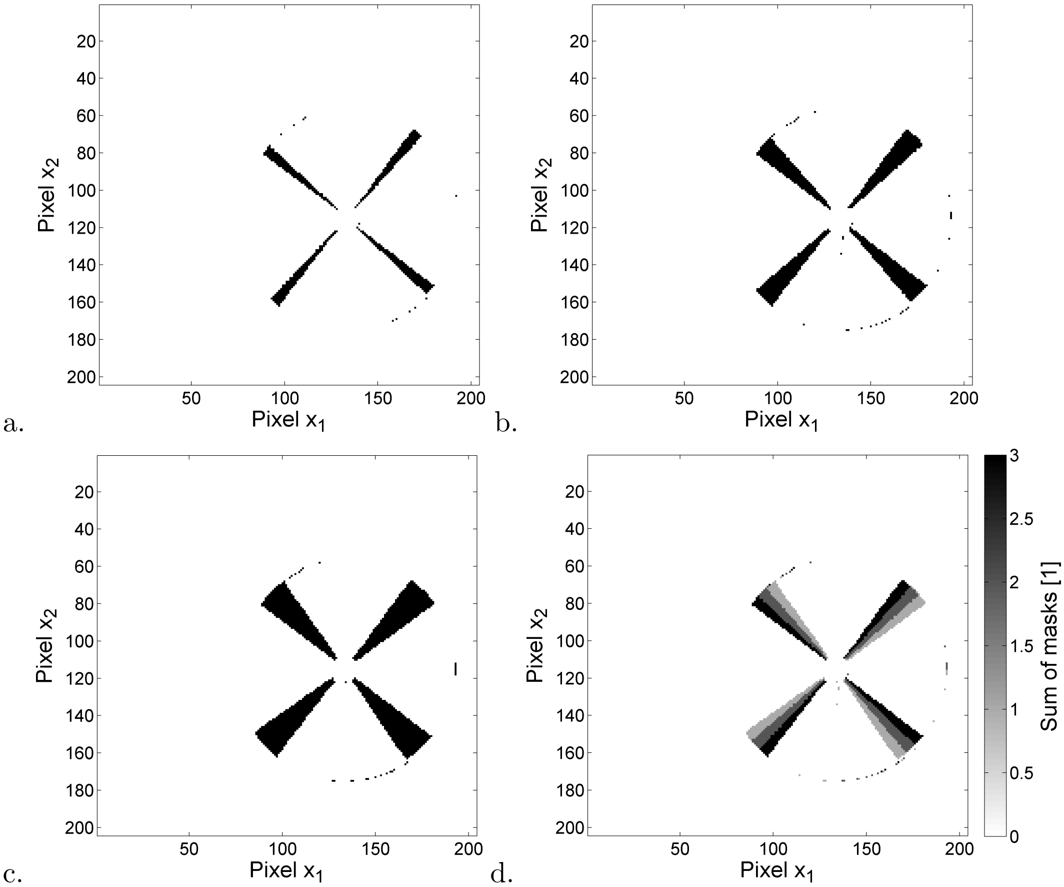


**Figure 6.9.:** Masks labeling erroneous acquisitions requiring correction: Black pixels indicate that the corresponding samples of the acquisitions (**a**) $l = 2$, (**b**) $l = 3$ and (**c**) $l = 4$ were found to be invalid. **d** The sum of the three masks, representing the number of acquisitions to be corrected.

these artifacts is demanding. Essentially a measure is required which scores the image distortion caused by motion artifacts. This is difficult especially for the fact that the erroneously reconstructed scene unknowns may be located anywhere in the available range. To the authors knowledge such measure does not exist so far. Therefore, in the following a list of requirements on an appropriate measure as well as a tangible proposal will be given.

Since the characteristics of motion artifacts of an arbitrary ToF camera are not generally predictable the author suggests that not the extent of the artifact (i.e. the value of the erroneous estimation), but solely the number of affected pixels is evaluated. An ideal measure for characterization of motion artifacts should furthermore fulfill the following requirements:

**Black box** The measure should be applicable on black box systems. Hence, no knowledge about camera parameters should be required.

**Generality** Optimally, motion artifacts produced by all possible ToF systems should be able to be characterized by the measure. Thus, it should be independent on product- and implementation-specific properties, especially it should not require explicit information about the camera optics, image size, frame rate, etc.

**Content independence** An ideal measure does not require a specific scene to be imaged.

Although the last item implies that not a specific scene has to be imaged, the desired evaluation of “wrong” scene estimations requires some ground truth information about the scene content. This information is hard to acquire for arbitrary scenes.

A further requirement is the following: For a given velocity of an imaged object the area affected by motion artifacts will decrease if the frame rate of the ToF system increases. For this reason, a measure should evaluate the speed of moving test objects relative to the system’s frame rate. This is not in conflict with the second item (generality), since the velocity of objects relative to the frame rate can be measured without explicitly knowing the frame rate, for instance by applying optical flow techniques.

In the following a measure will be proposed which fulfills these requirements.

**A Measure for Quantification of Motion Artifacts** The measure proposed here characterizes the *relative distorted area* in the maps of the reconstructed scene unknowns for a particular frame. It is defined as

$$\rho = \frac{A_{art}}{A_{max}} , \tag{6.2}$$

with $A_{art}$ being the area distorted by motion artifacts and $A_{max}$ being the (theoretically) *maximal area* distorted:

The area $A_{max}$ of a given frame is defined as the sum of the length of all line segments perpendicular to the direction of motion, multiplied with its velocity (measured in pixel/frame). Thus $A_{max}$ corresponds to the area of the image changing between two consecutive frames. This area represents the maximum number of pixel values affected by motion artifacts. It, therefore, embodies the worst case scenario.

The area $A_{art}$ is the area of pixel values *actually distorted* by motion artifacts. The quotient of both, $\rho$, expresses the distorted area in the maps of reconstructed scene unknowns relative to the area distorted in the worst case. A value of $\rho = 1$ stands for the worst case. Techniques preventing or correcting motion artifacts result in a decreased measure $\rho$. Hence, $\rho$ can be understood as the sensitivity of the ToF system to motion artifacts.

The determination of $A_{art}$ is not trivial since affected pixel values may lay anywhere in the available range. One possibility is to use a simple scene with a well-defined foreground and background, and to label all pixel values deviating from the typical values of these objects by more than a predefined threshold as artifacts. As shown in the example using the rotating target (Sect. 6.3.1) artifacts may coincidentally match the values of the foreground or background object. Therefore a robust determination of $A_{art}$ is a segmentation problem which should be solved by incorporating all available channels, i.e. all raw channels and all estimations of scene unknowns. At this point a general rule for the determination of affected pixels valid for a general ToF system cannot be given. In the following a procedure will be explained which provides a good labeling for the sequence of the rotating target at hand.

#### 6.3.1.2. Implementation of the Measure

The quantitative evaluation will focus on the four laterally moving edges of the rotating target. A circular region at the center (radius: $\upsilon_1 = 14$pixel) was excluded from the analysis. Furthermore, only pixels within a $\upsilon_2 = 60$pixel radius circular region around the rotation axis of the target were analyzed; so basically everything apart from the target was excluded from the analysis as well.

By regarding 25 frames preceding to the analyzed one the target's speed of rotation was determined as $\omega = (0.20 \pm 0.01)$rounds/frame. The maximum area affected by motion artifacts $A_{max}$ generated by the four edges is thus given as

$$A_{max} = 4 \cdot (\upsilon_2^2 - \upsilon_1^2) \cdot \omega \cdot \pi = (8555 \pm 430)\text{pixel}^2 \tag{6.3}$$

The area actually distorted by artifacts $A_{art}$ was determined for each of the analyzed cases (which will be discussed in the next section) individually. It was done by applying a thresholding scheme on the the maps of the estimated scene unknowns.

For example the map $A_{art}$ describing the artifacts of the depth map generated using the standard averaging technique was prepared using the maps of the depth and amplitude channel. Both were generated from raw data using the averaging technique (see Sect. 5.2.2) and by applying (5.8)–(5.11), (2.9) and (2.3), and (2.8), respectively. A visualization of these maps is given in Fig. 6.10.a and b.

In order to isolate the artifacts a threshold scheme was applied to these maps: All pixels from the depth maps with a distance of 2.578m $< d <$ 2.771m or $d >$ 3.856m were excluded from the map labeling the artifacts. All other pixels were marked as candidates for pixels affected by motion artifacts, resulting in the map visualized in Fig. 6.10.c.

Additionally, a map of candidates was generated using the amplitude map. Here, all pixels with an amplitude value of 475DN $< a_1 <$ 2000DN were labeled as candidates for artifacts resulting in a map given in Fig. 6.10.d.

Both maps of candidates were combined by an OR-operation. The resulting map is given in Fig. 6.10.e.

This combined map was further processed in order to "clean up" the result: Image regions apart from the target were removed, and the circular area in the center was cut out[6]. Furthermore the holes in artifact regions were filled and some erroneous detections from the background were removed. In Fig. 6.10.f the final result of these operations is visualized.

The sum of labeled pixels in this final map was used as area $A_{art}$. In this example it was computed as $A_{art} = 2092\text{pixel}^2$.

#### 6.3.1.3. Results and Discussion

For the further analysis four different depth maps were computed which are visualized in Fig. 6.11. First, a depth map was generated by employing the state-of-the-art

[6] corresponding to the radii $\upsilon_1$, $\upsilon_2$ used in the estimation of $A_{max}$, see (6.3)

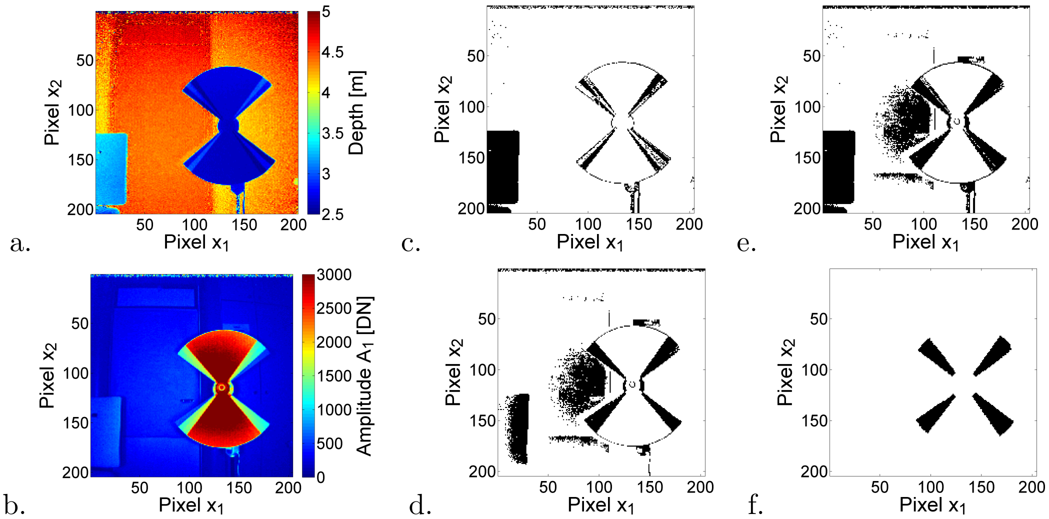


**Figure 6.10.:** Generation of the mask $A_{art}$ labeling the artifacts of the depth map which was generated using the standard averaging technique: **a** Depth map and the derived intermediate mask **c**. **b** Amplitude map and the intermediate mask derived from it **d**. **e** Combined mask and final result **f**.

averaging technique (i.e. using the phase map $\mathbf{\Phi}_{avg}$, see Fig. 6.11.a). This depth map was corrected for motion artifacts using the proposed algorithm (via a phase map $\mathbf{\Phi}_{avg,mc}$, see Fig. 6.11.b). Furthermore, using the dynamic calibration method from Chap. 5 a depth map using only the second subset of acquisitions was generated (i.e. using $\mathbf{\Phi}_2$, Fig. 6.11.c). This map was corrected using the variant of the burst internal detection ($\mathbf{\Phi}_{2,BID}$, Fig. 6.11.d).

The defined measure for characterization of motion artifacts was applied to these depth maps giving the results visualized in Fig. 6.12.

This required the determination of a mask labeling the artifacts in the processed maps. Analog to the procedure described in Sect. 6.3.1.2 these masks were generated using a combination of thresholds applied to the depth and amplitude maps and refined by a following "clean up" step. The raw combined masks and the resulting final maps are visualized in Fig. 6.13. The relative error of each area affected by artifacts $A_{art}$ was assumed to be $\sigma_{A_{art}}/A_{art} = 10\%$.

The measure $\rho$ serves for characterization of the distortion of maps of estimated scene unknowns due to motion artifacts. It expresses the system's sensitivity relative to a worst case scenario (represented by the case of $\rho = 1$). The depth map computed using the averaging technique is evaluated with $\rho_{\mathbf{\Phi}_{avg}} = 0.2445 \pm 0.0367$ which is

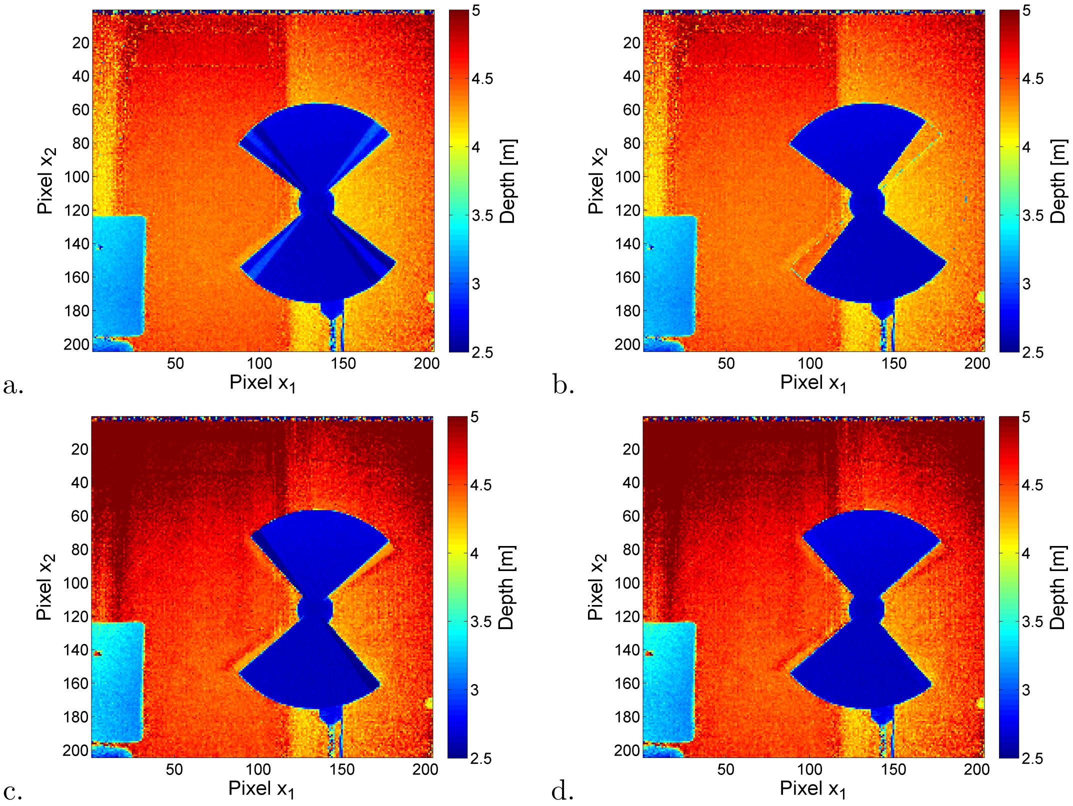


**Figure 6.11.:** Computed depth maps and results of the proposed algorithm for correction of motion artifacts: **a** Depth map computed using the (state-of-the-art) averaging technique (via $\mathbf{\Phi}_{avg}$) and the derived map corrected for motion artifacts **b** (using $\mathbf{\Phi}_{avg,mc}$). **c** Depth map using the second subset of the given set of acquisitions (using $\mathbf{\Phi}_2$). **d** Result of the correction for motion artifacts using the variant of burst internal detection (i.e. via $\mathbf{\Phi}_{2,BID}$).

about four times better than the worst case performance. This is due to the fact that the camera uses a burst mode for acquisition of its raw images. The length of a burst is about one quarter of the temporal distance between two frames, resulting in a value of the sensitivity measure $\rho$ of about one quarter.

After applying the proposed method for compensation of motion artifacts the sensitivity measure drops to $\rho_{\mathbf{\Phi}_{avg,mc}} = 0.0178 \pm 0.0027$, which corresponds to an increase of the performance by factor 13.7. Thus, the depth map generated by the correction algorithm is significantly less sensitive to motion artifacts.

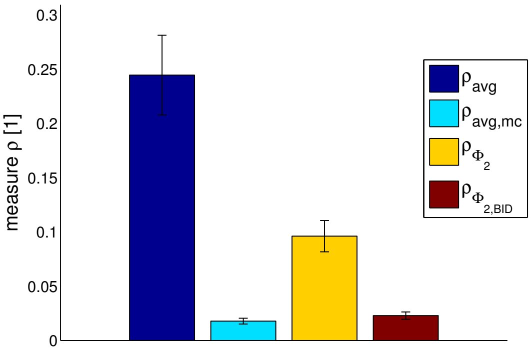


**Figure 6.12.:** Determined values of the sensitivity measures $\rho$ for description of motion artifacts. Lower values indicate a better performance. $\rho_{avg}$ represents the state-of-the-art.

The depth map computed from the second half of the set of acquisitions (corresponding to a computation from the single phase map $\mathbf{\Phi}_2$ from Chap. 5) gives a measure of $\rho_{\mathbf{\Phi}_2} = 0.0961 \pm 0.0144$. This is about a third of the value determined for the map computed using the averaging technique (i.e. via $\mathbf{\Phi}_{avg}$). The reason for this is that $\mathbf{\Phi}_{avg}$ is constructed using four acquisitions with three pauses between of them. In contrast, $\mathbf{\Phi}_2$ employs only two acquisitions captured with one pause. During three pauses the target's edges take a distance which is about three times more than the distance taken in one pause. Thus, employing the camera at hand and using a depth map computed from a single phase map as proposed in the previous chapter results in a reduction of motion artifacts by a factor of three.

By applying the burst internal detection variant described in Sect. 6.2.4 on this single depth map, its motion artifacts were corrected ($\mathbf{\Phi}_{2,BID}$). For this depth map a sensitivity measure of $\rho_{\mathbf{\Phi}_{2,BID}} = 0.0229 \pm 0.0034$ was determined. This is slightly above the value computed for the corrected depth map generated using the averaging technique ($\mathbf{\Phi}_{avg,mc}$). The strength of the BID approach is, however, that its higher temporal sampling density facilitates to cope with scenes showing a much higher dynamic. This will be demonstrated in the following section by use of natural test scenes.

### 6.3.2. Natural Scenes

In this section the discussed methods will be applied to sequences of more natural scenes. It will be shown that these highly dynamic scenes are challenging for state-of-the-art ToF camera systems.

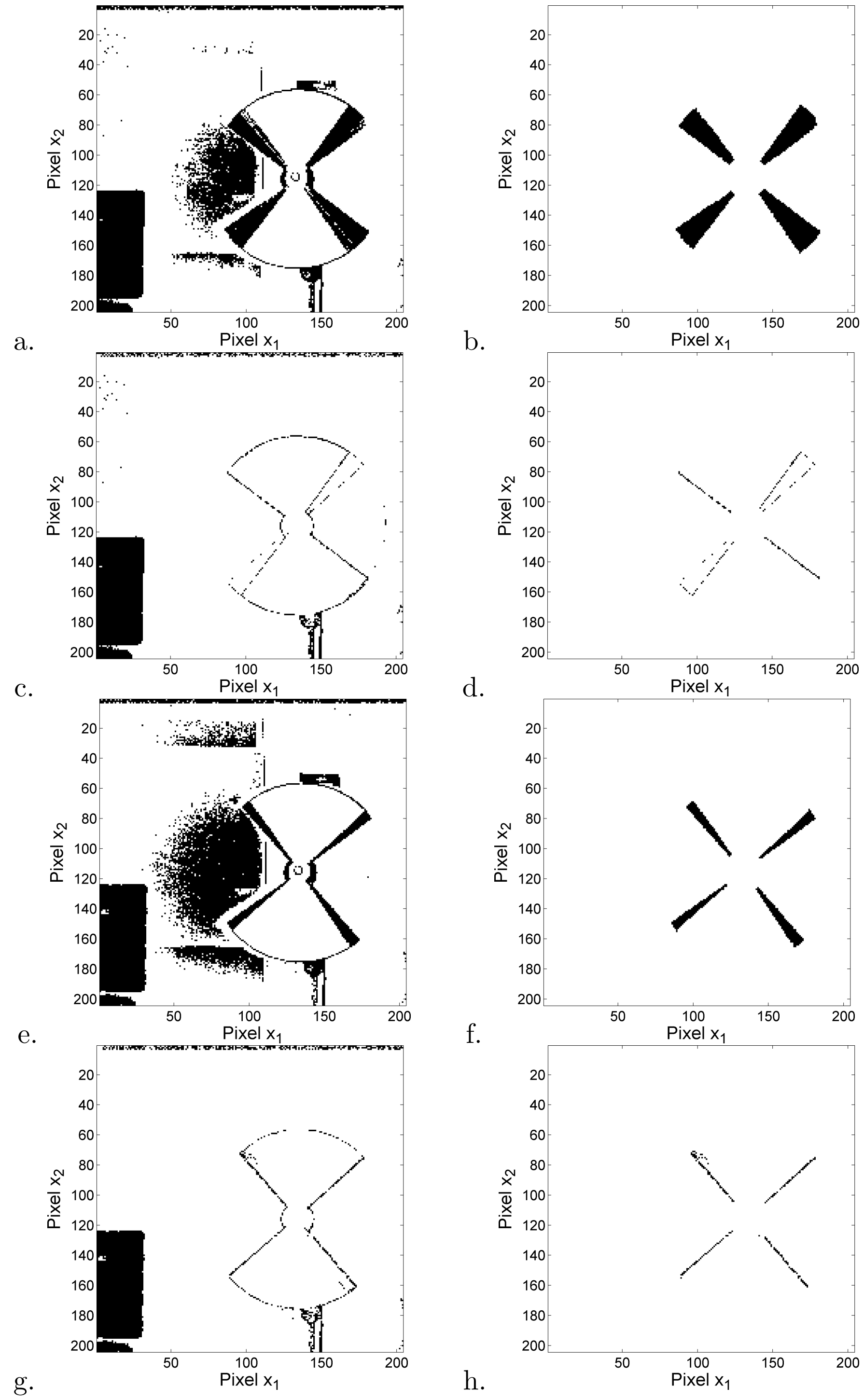


**Figure 6.13.:** Masks labeling the artifacts used for determination of $A_{art}$: **a** Raw combined mask and **b** result of the cleaning step for the depth map generated using the averaging technique (using $\mathbf{\Phi}_{avg}$). **c**, **d** Corresponding masks for the depth map corrected for motion artifacts ($\mathbf{\Phi}_{avg,mc}$). **e** Depth map generated using the second half of the split frame ($\mathbf{\Phi}_2$) and result of the clean up step **f**. **g**, **h** Corresponding masks for the depth map generated from $\mathbf{\Phi}_2$ using the burst internal detection ($\mathbf{\Phi}_{2,BID}$).

Also these sequences were acquired using a PMD CamCube 2.0 camera. The complete processing is performed using exactly the same procedures as described above.

#### 6.3.2.1. Scenario 1: Rapid Gestures

The first scenario reproduces the acquisition of 3D input for gesture recognition systems, which is an interesting application of ToF cameras. A sequence was acquired showing a person performing a rapid gesture with his hand: In the analyzed frame the right arm is moving fast from an upper to a lower position.

In Fig. 6.14 the generated depth maps are shown. Fig. 6.14.a was reconstructed using the averaging technique. It shows a significant artifact at the hand and parallel to the lower arm. Applying the (standard) correction algorithm on this image using four acquisitions leads to the result shown in Fig. 6.14.b. Here, the artifacts next to the lower arm were successfully corrected, but areas around the finger tips remained distorted. Additionally the algorithm introduced new artifacts visible above of the arm. This artifact is caused by the rapid succession of edges (background–arm and arm–background) which violates the assumption of at most one event within two compared sets of acquisitions.

Fig. 6.14.c shows the depth computed from two acquisitions (via $\mathbf{\Phi}_2$). It contains less artifacts than the map generated using the averaging technique. The remaining artifacts can hardly be recognized by the eye because they match the distance of the foreground and background objects. By employing the burst internal detection a depth map corrected for motion artifacts was generated. This map is visualized in Fig. 6.14.d. It appears sharper especially around of the ball of the hand and the upper part of the lower arm.

For a better comparison a cross section at $x_1 = 48$ going through the imaged arm is given in Fig. 6.15. This plot shows the depth values determined using the averaging technique, the values generated using two acquisitions, and the values corrected by the BID method. All values are shown over the pixel coordinate $x_2$. Although no ground truth information is available it can be seen that the depth values computed using the averaging technique (blue curve) are unsteady and obviously distorted by artifacts. Also the reconstruction using $\mathbf{\Phi}_2$ (green curve) introduces artifacts by assigning values describing a step of 30cm height to the flat surface of the imaged arm. The BID method (red curve) corrects these artifacts. Only one pixel with a distorted depth value remains at the edge.

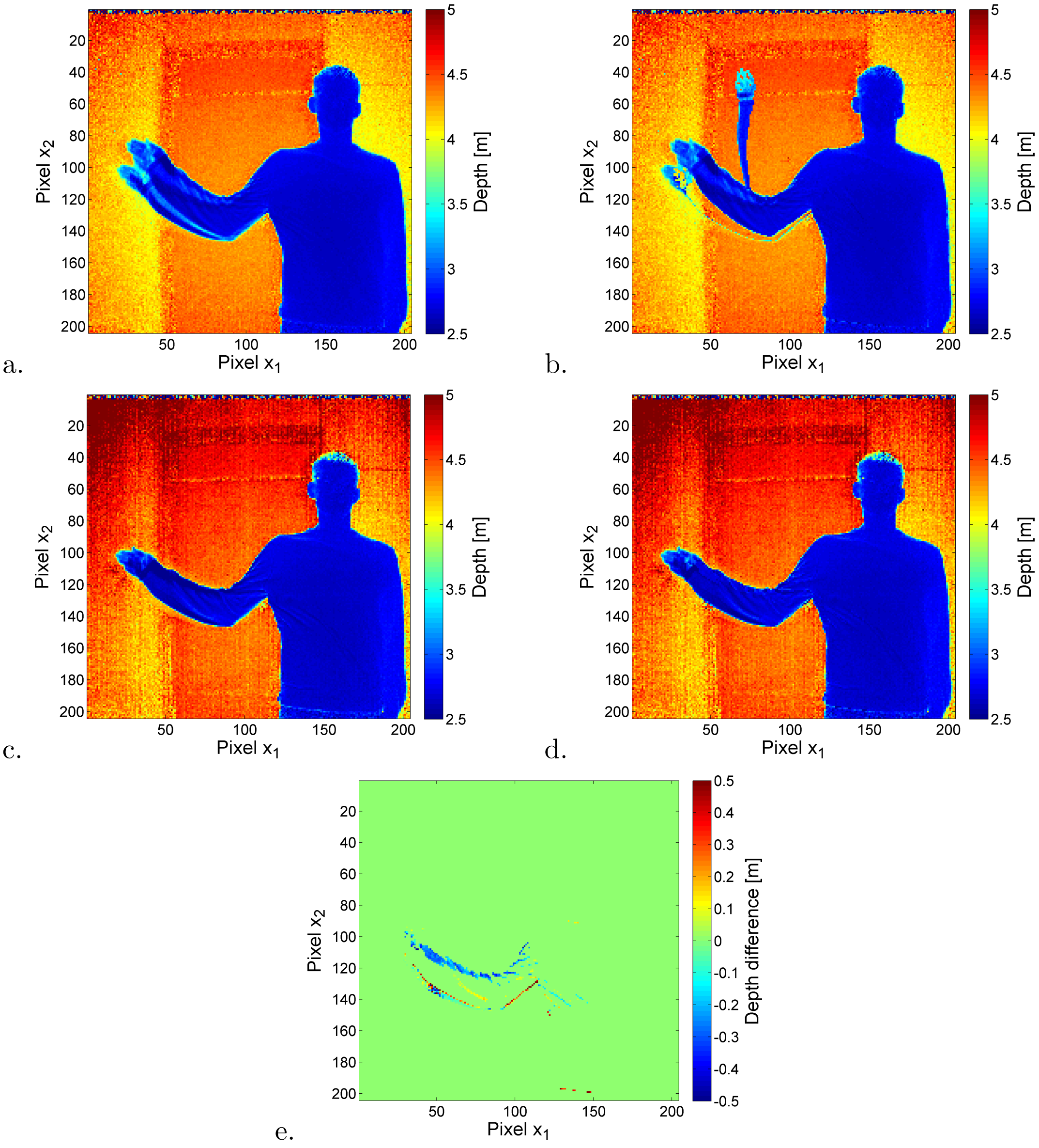


**Figure 6.14.:** Depth maps of a dynamic scene showing a person rapidly moving his arm from an upper to a lower position. The state-of-the-art is given in a, while b-d represent results of this work: **a** Map generated using the averaging technique with considerable artifacts parallel to the lower arm. **b** Result of the developed algorithm (standard version). Some artifacts were corrected, but additional errors were introduced. **c** Depth map generated by use of two acquisitions (via $\mathbf{\Phi}_2$). **d** Result of the correction using the burst internal detection (BID) variant. **e** Difference image of the depth maps computed using $\mathbf{\Phi}_2$ and the BID method ($\mathbf{\Phi}_{2,BID}$).

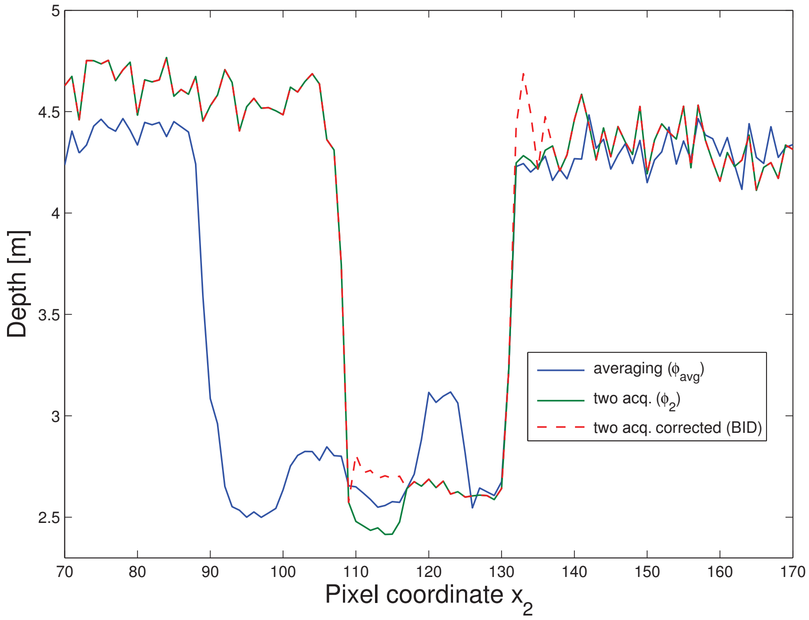


**Figure 6.15.:** Cross section at $x_1 = 48$ through the depth maps generated using the averaging technique (blue), using 2 acquisitions (green) and the derived map corrected for motion artifacts by employing the BID method (red).

A difference image of the both depth maps generated via $\mathbf{\Phi}_2$ and using the BID method is given in Fig. 6.14.e. It reveals that depth artifact of up to 40cm distortion were corrected. Please note that noticeable differences of the depth map using the averaging technique and the one using only two acquisitions, for instance expressed by the different hue of the background, can be explained as a consequence of working with an uncalibrated camera. This property does not allow any conclusions about a difference in quality of both approaches (c.f. discussion in Sect. 5.4.1).

Since no ground truth information is given for the observed scene a detailed quantitative analysis is not possible at this point.

#### 6.3.2.2. Scenario 2: Juggling Performance

In a second scenario a sequence showing a juggling artist[7] was acquired. For one specific frame the different depth maps were computed as explained in the prior section. They are visualized in Fig. 6.16.

[7] At this point the author would like to thank Dr. Christoph Sommer for his performance.

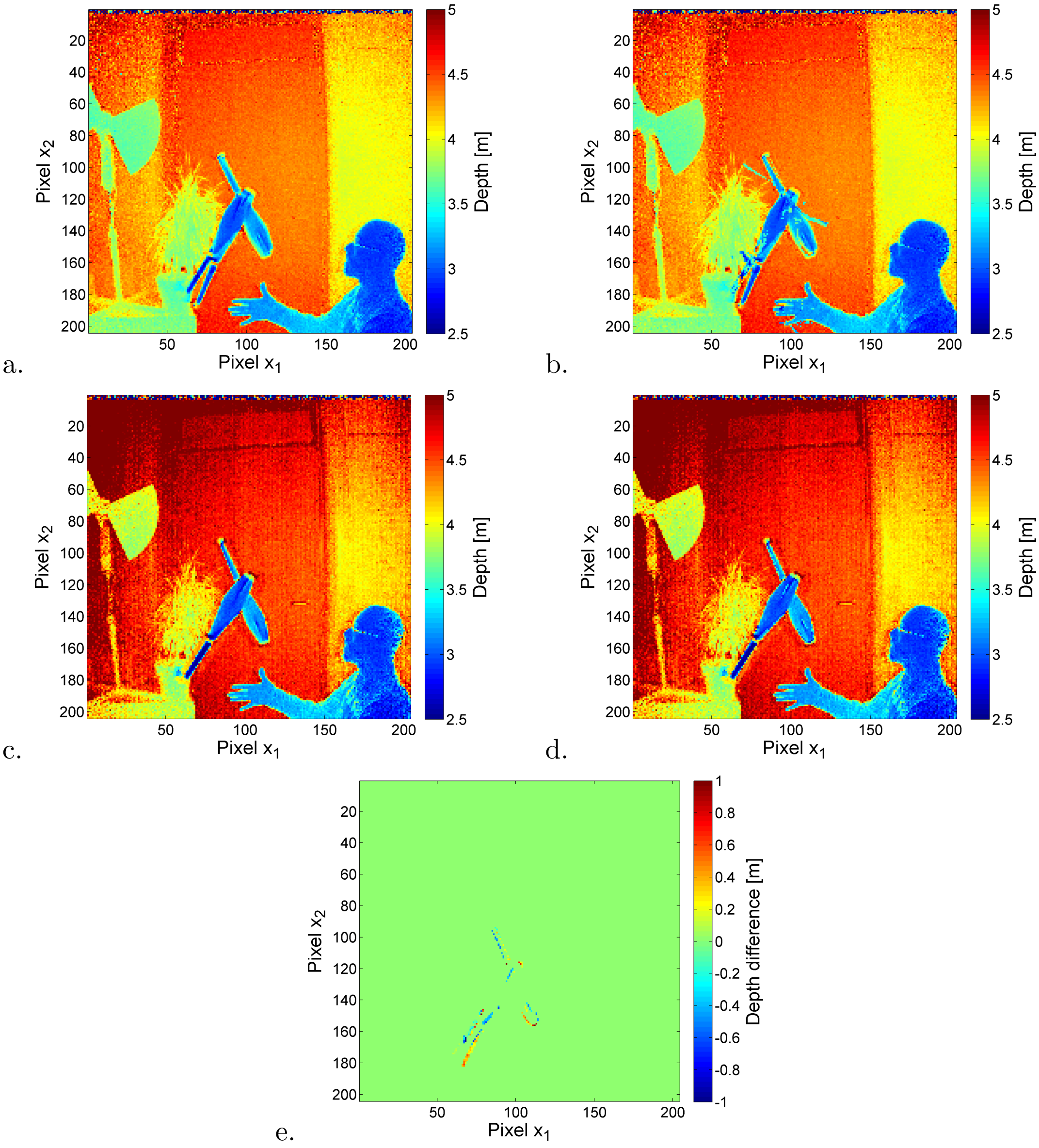


**Figure 6.16.:** Depth maps of a dynamic scene showing a juggling artist. The state-of-the-art is given in a, while b-d represent results of this work: **a** Map generated by use of the averaging technique. **b** Result of the presented algorithm (standard version). **c** Depth map generated by use of two acquisitions (via $\mathbf{\Phi}_2$). **d** Result of the burst internal detection (BID) variant. **e** Difference image of the depth maps computed using $\mathbf{\Phi}_2$ and the BID method ($\mathbf{\Phi}_{2,BID}$).

The depth map using the averaging technique and the map corrected for motion artifacts are given in Fig. 6.16.a and Fig. 6.16.b. Both maps show considerable artifacts around the rapidly spinning clubs.

The depth map computed using data of two acquisitions and the derived map corrected for motion artifacts using the method of burst internal detection are shown in Fig. 6.16.c and Fig. 6.16.d. A difference between both is hardly visible. However, the computed difference image given in Fig. 6.16.e shows that some pixel show deviations in the depth estimation of up to 80cm.

#### 6.3.2.3. Summary

In these scenarios of imaging rapid movements the correction of motion artifacts based on the method of burst internal detection surpasses all other evaluated methods. Compared to depth maps generated using the averaging technique this advantage of being able to cope with highly dynamic scenes is achieved by accepting an increased statistical uncertainty of the depth estimation. This increase was determined as a factor of $\sqrt{2}$ (see Sect. 5.4.1.2). Compared to a depth map generated using two acquisitions but no correction of motion artifacts the BID approach allows to generate only one depth map per frame[8]. It thus trades the doubling of the frame rate achieved by the dynamic calibration approach from Chap. 5 against the ability to image highly dynamic scenes with a significant reduction of motion artifacts.

## 6.4. Conclusion and Outlook

This chapter investigated the origin of motion artifacts which are a well known issue of today's ToF systems. Analyzing the temporal raw data signal was found to be a powerful tool for identification and correction of these artifacts. The derived algorithms for detection and correction of motion artifacts may be applied to data acquired by all kinds of Time-of-Flight cameras, including pulse-based and continuous-wave systems.

The proposed detection of artifacts is based on a comparison of raw values with corresponding values of the prior acquisition. Rapid changes of scene unknowns cause events which result in discontinuous raw channel signals. The algorithm is able to

[8] To be more accurate: The BID method can be applied to all independent subsets of acquisitions except from the first one. If a camera performing more acquisitions per burst or employing more detection units per pixel is used, the application of the BID method on several sets per burst becomes feasible.

handle one of such events within two consecutive sets of acquisitions correctly. Since many of today's ToF systems acquire its raw images in a burst mode fashion the delay between two consecutive bursts is the limiting factor of the algorithm's performance.

As a variant, the burst internal detection (BID) was proposed which may be applied in multi-tap ToF systems. In combination with the dynamic calibration presented in Chap. 5 this method is able to perform the comparison of raw channels and thus the detection of discontinuities within each burst. Therefore the algorithm is able to cope with one event within each set of acquisitions, which allows the correction of very dynamic scenes.

The method was verified experimentally by use of a commercial ToF system acquiring sequences showing different scenes. The usage of a well defined environment with a special target enabled to demonstrate the functionality of the algorithm. By introducing a measure for evaluation of motion artifacts a quantitative analysis was given. Both variants (standard and BID) of the proposed method for correction of motion artifacts decrease the system's sensitivity for these artifacts by almost a factor of 14.

The applicability of the developed algorithms was demonstrated using natural scenes. One of these scenes showed a person performing a fast gesture and a second scene showed a juggling artist. For these highly dynamic scenes the standard version of the proposed algorithm introduced additional artifacts due to the fact that the assumption of only one event occurring within two sets was violated. The BID version, however, was able to eliminate all artifacts and deliver an undistorted depth map. This was achieved by accepting an increase of depth noise by a factor $\sqrt{2}$ compared to the state-of-the-art averaging technique.

The proposed method performs a local analysis using only raw values and temporal relations of the pixels being corrected. This feature results in a low complexity of the algorithm and thus enables real-time implementations, even on systems with limited hardware resources.

# Chapter 7.

# Conclusion and Outlook

This work provided a thorough analysis of the state-of-the-art Time-of-Flight depth imaging technology and suggested possibilities to overcome some particular difficulties of today's implementations.

## 7.1. Summary and Conclusion

Chapter 2 introduced an abstract formalism for the general description of Time-of-Flight camera systems. Based on this formalism the two implementations of ToF imaging – the pulse-based as well as the continuous-wave approach – were discussed. Following, both variants were compared and an unification was proposed. Furthermore, an overview on the difficulties and shortcomings of today's ToF systems was given which were classified into (1) *basic difficulties*, (2) *errors caused by an insufficient sampling*, (3) *deviations related to an imperfect propagation of the light* and (4) *further deviations*. These errors were described using the general formalism, showing that they occur in all of today's systems independently of the chosen implementation.

In Chapter 3 a physical model of a ToF system was presented. This model aims at providing a better understanding of the Time-of-Flight technology and the discussed difficulties. The developed system models the sensor as a black box, including a system for suppression of ambient light (SBI). Although its design was focusing on a specific camera the model can be employed to describe any of today's available systems. By use of measurements of a real camera (PMD CamCube 2.0) this model was parameterized with values providing a physical meaning. Utilizing this parameterization it was possible to derive a simulation tool which reproduces the sensor behavior and generates realistic data. This tool reproduces all effects and deviations from the class of *basic difficulties*. It was employed in various projects, of which three were presented:

The simulation was used to evaluate an algorithm for a medical ToF application in which the use of a real ToF systems is currently not allowed. Furthermore the ToF sensor model was fused into the Sony Total System Simulator (TSS), a framework for exact simulation of a complete camera. This extended TSS hence will allow the development of virtual prototypes of novel ToF cameras. A third application is the usage of the model for the characterization of ToF systems. This supports the development of a standard for description of ToF cameras which will become an extension to the EMVA 1288 Standard.

In Chapter 4 an investigation of a nonlinear photo response on the accuracy and precision of the generated depth data was performed. It was found that the nonlinearity introduces a periodical error which is similar to the wiggling error known from previous investigations. Using realistic simulations based on the tool from Chap. 3 this new error was estimated to be about a magnitude smaller than the known wiggling phenomenon. These positive results inspired the proposal of a logarithmic ToF sensor. This new kind of sensor facilitates a highly increased depth dynamic while generating data with systematic and statistical errors comparable to linear implementations. Thus, the logarithmic Time-of-Flight sensor is a promising concept for future depth imaging systems.

In Chapter 5 a dynamic calibration method for compensation of the inequalities in the photo response of multi-tap sensors was proposed. This method supersedes the commonly used technique of computing intermediate samples by use of multiple acquisitions (averaging technique). The presented dynamic calibration and data rectification method enables to compute multiple depth maps per raw frame. This results in an increase of the frame rate. Since the proposed method works on the level of camera raw data it is not restricted to the depth maps, but can be applied on all processed channels.

Using a commercial two-tap system a doubling of the camera's frame rate was demonstrated. In a detailed quantitative evaluation the consistency of the generated single depth maps was shown. Furthermore, using many sequences acquired at different temperatures it was proven that the dynamic calibration surpasses static approaches.

The enhanced frame rate let to a performance of $60\text{Hz} - 80\text{Hz}$ at a lateral resolution of $200 \times 200$ pixel. To the authors knowledge this is the highest frame rate of a ToF imaging system currently available. It was reached without any adaption of the hardware, but only by an optimized processing of the raw data. This enhanced processing pipeline may be implemented in a computationally very efficient manner. Hence, real-time implementations of the proposed methods are definitely feasible. Since the speed-up is equal to the number of detection units ($Q$) used by the sensor, future sensors employing more taps will profit even more from this algorithm. A patent application on the method was filed in [SZ10a].

In Chapter 6 a method for the robust detection and correction of motion artifacts was proposed. This method is based on an analysis of the temporal signals of the raw channels. It was explained that and how motion artifacts can be predicted from special constellations of temporal discontinuities in the individual raw channels. These constellations can be formulated in a single rule. The evaluation of this rule enables the robust detection of raw values leading to artifacts. A simple strategy of overwriting these values by values of the prior acquisition prevents the artifacts. The method can be combined with the dynamic calibration from Chap. 5 leading to a very efficient variant called Burst Internal Detection (BID). Both variants – the standard as well as the BID algorithm – were verified experimentally with great success. Using sequences acquired in a controlled scenario and in two highly dynamic natural scenes the method was demonstrated to significantly correct motion artifacts. These results were evaluated also quantitatively by proposing a novel measure for characterization of motion artifacts.

The presented method is very simple and hence may be implemented in an extremely efficient manner. It is applicable to all ToF systems combining multiple consecutive acquisitions of raw data into one depth map (i.e. all systems for which motion artifacts are an issue). The method neutralizes a big fraction of the distortions caused by motion artifacts present in all of today's ToF systems. Thus the author expects it to become an integral part of future implementations. A patent application on this method was filed in [SZ10b].

## 7.2. Inference

This work sought to describe the working principle of Time-of-Flight systems as well as their difficulties using a general formalism. A list of shortcomings of current implementations was given, and some of them were studied using a developed physical model. This model is able to reproduce realistic sensor data and thus is a powerful instrument for the development of algorithms working with ToF data.

From the list of difficulties of current ToF systems (see page 26), three issues were tackled in the subsequent chapters:

The influence of a nonlinear photo response was analyzed and inspired the investigation of a logarithmic ToF sensor. This sensor turned out to be a promising concept. It is a possible solution for the challenges related to the limited *dynamic range*.

A dynamic calibration method was proposed which facilitates to compute multiple depth maps from the data of a single frame. The algorithm is applicable on data acquired by multi-tap ToF devices which constitute the biggest class of today's available systems. It thus drastically relaxes the problems for some applications resulting from the limited *frame rate* of current implementations.

Furthermore the problem of erroneous combinations of raw data caused by movements of the imaged objects was addressed. A very simple solution was proposed which efficiently detects and corrects these errors. Thus it can be seen to mostly solve the problem of *motion artifacts*.

## 7.3. Outlook

Possible extensions of the work presented here include investigations from the following areas:

The developed physical model is currently very much focused on the ToF sensor and uses ideal maps as input. In order of being able to simulate the complete acquisition of a 3D scene this sensor model should be extended by a renderer module. Such module would generates the required maps from a given 3D representation and hence enable to simulate complex artificial scenes.

The model currently insufficiently supports temporally changing content. A corresponding expansion would allow the simulation of much more complicated scenarios, and include the reproduction of errors like motion artifacts. Hence, algorithms could be tested under much more realistic conditions. The same applies to the following extension: The current module uses input maps which have the same lateral resolution as the simulated sensor. An adaption overcoming this restriction and facilitating a spatial supersampling would be beneficial. This would allow the simulation of artifacts at object boundaries, for example flying pixels (c.f. Sect. 2.3.2.1).

In Chapter 4 a logarithmic Time-of-Flight sensor was proposed. Before putting this sensor into practice additional simulations are advisable. An incorporation of measurement results and observations of non-linearity modules of real sensor implementations would improve the validity of these simulations.

The dynamic calibration for increasing the camera's frame rate was demonstrated to work with a commercial ToF system. Since this system acquires raw frames in a burst mode fashion the algorithm's output of single depth maps does not correspond to an equitemporal sampling of the scene. In order to exploit the proposed technique optimally, these temporal acquisitions of the individual raw frames should be adjusted.

The idea of a dynamic self-correction of errors could be carried over to other challenges of current Time-of-Flight systems. So for example the geometric distortions introduced by the optics could be detected and compensated using techniques known from 2D imaging (e.g. a combination of feature tracking and bundle adjustment, see [HZ00]). In conjunction with the available distance data this could enable to correct the systematic depth errors known from today's systems (wiggling).

Also the proposed method for detection and correction of motion artifacts suffers from the burst mode acquisition of raw images. A temporal adjustment of these samples will lead to a more robust correction of artifacts. It will, hence, enable the algorithm to cope with scenes showing more dynamics.

Finally it should be emphasized again that all presented solutions, especially the the dynamic calibration algorithm for an increased frame rate and the method to compensate motion artifacts are not restricted to depth maps. In fact, they may be applied to improve the quality of all processed channels of the ToF system. Furthermore, they are not limited to Time-of-Flight *depth* imaging but might be used in any application employing multi-tap ToF sensors. So for instance Fluorescence Lifetime Imaging (FLIm, see [Erz11; Fra11; Lin11]) could profit from the developed methods.

# Appendix A.
# Quadratic Photo Response – Details

The goal of this section is to show that under specific conditions the depth estimation performed with a ToF sensor using a quadratic photo response does not suffer from systematic errors caused by the non-linearity of its characteristic curve. This is the analytical proof to the observation made in Sect. 4.3, where a non-linearity parameter $\alpha = 2$ led to a vanishing systematic error of the phase estimation $\Delta\varphi$, and thus to a vanishing error of the depth estimation.

As in Sect. 4.3 a continuous-wave, correlating ToF system using a sinusoidal modulation of the light source signal and a rectangular reference signal is assumed:

$$\begin{aligned} S(t) &= b_{ls} + a_{ls}\sin(\nu_0 t - \varphi) && \text{(A.1)} \\ R(t) &= H(\sin(\nu_0 t + \theta))\,, && \text{(A.2)} \end{aligned}$$

Furthermore it is assumed, that $\tilde{N} = 4$ equidistant sampling points located at the phase angles $\theta = \tilde{n} \cdot 2\pi/\tilde{N}$ are used to reconstruct the phase shift $\varphi$ of the electro-optical input. This simplifies (2.9) (given here again as (A.3)) to (A.4):

$$\begin{aligned} \varphi &= \arg\left(\sum_{\tilde{n}=0}^{\tilde{N}-1} c_{\tilde{n}} e^{-\mathrm{i}\theta_{\tilde{n}}}\right) && \text{(A.3)} \\ \varphi &= \operatorname{atan}\left(\frac{c_3 - c_1}{c_2 - c_0}\right) && \text{(A.4)} \end{aligned}$$

The ideal (perfectly linear) correlating sensor generates these samples $c$ as

$$c = \int S(t) \cdot R(t)\, dt = \int\limits_{\xi}^{\xi + 1/2\cdot\pi} S(t)\, dt\,, \tag{A.5}$$

with $\xi$ corresponding to a certain sampling mode, giving a specific sample $c_{\tilde{n}}$.

Accordingly, a correlating sensor with a quadratic characteristic curve acquires samples $c^{(2)}$ as

$$c^{(2)} = \int S(t)^2 \cdot R(t)\, dt = \int\limits_{\xi}^{\xi+1/2\cdot\pi} S(t)^2\, dt\,. \tag{A.6}$$

The phase shift is reconstructed from such samples using (A.4). In this function, the numerator and denominator of the argument of the arcus tangent (atan) function are each a difference $\delta$ of integrals of type (A.5) and (A.6), respectively. The difference of two samples $i$ (acquired using an ideal characteristic curve) is $\delta_i$:

$$\begin{aligned}
i_1 &= \int\limits_{\xi+\pi}^{\xi+3/2\cdot\pi} S(t)\, dt \\
&= \frac{b\nu_0\,\pi + 2\,a\cos(\varphi+\nu_0\,\xi+\nu_0\,\pi) - 2\,a\cos(\varphi+\nu_0\,\xi+3/2\,\nu_0\,\pi)}{2\,\nu_0} && \text{(A.7)}\\
i_2 &= \int\limits_{\xi}^{\xi+1/2\cdot\pi} S(t)\, dt \\
&= \frac{2\,a\cos(\nu_0\,\xi+\varphi) - 2\,a\cos(\varphi+\nu_0\,\xi+1/2\,\nu_0\,\pi) + b\nu_0\,\pi}{2\,\nu_0} && \text{(A.8)}\\
\delta_i &= i_1 - i_2 \\
&= \frac{-1}{\nu_0}[a\cos(\nu_0\,\xi+\varphi) - a\cos(\varphi+\nu_0\,\xi+1/2\,\nu_0\,\pi)\ldots \\
&\quad -a\cos(\varphi+\nu_0\,\xi+\nu_0\,\pi) + a\cos(\varphi+\nu_0\,\xi+3/2\,\nu_0\,\pi)] && \text{(A.9)}
\end{aligned}$$

$\delta_i$ corresponds to the numerator in (A.4) for $\xi = \pi/2$, and to the denominator for $\xi = 0$. Thus, the difference of two samples acquired using a square photo response is $\delta_s$:

$$\begin{aligned}
\delta_s &= \int\limits_{\xi+\pi}^{\xi+3/2\cdot\pi} S(t)^2\, dt - \int\limits_{\xi}^{\xi+1/2\cdot\pi} S(t)^2\, dt \\
&= \frac{-1}{4\nu_0}\,[8\,ab\cos(\nu_0\,\xi+\varphi) + a^2\sin(2\,\nu_0\,\xi+2\,\varphi)\ldots \\
&\quad -a^2\sin(2\,\varphi+2\,\nu_0\,\xi+\nu_0\,\pi) - 8\,ab\cos(\varphi+\nu_0\,\xi+1/2\,\nu_0\,\pi)\ldots \\
&\quad -a^2\sin(2\,\nu_0\,\pi+2\,\varphi+2\,\nu_0\,\xi) - 8\,ab\cos(\varphi+\nu_0\,\xi+\nu_0\,\pi)\ldots \\
&\quad +a^2\sin(2\,\varphi+2\,\nu_0\,\xi+3\,\nu_0\,\pi) + 8\,ab\cos(\varphi+\nu_0\,\xi+3/2\,\nu_0\,\pi)] && \text{(A.10)}
\end{aligned}$$

Since the difference $\delta_{is}$ of both

$$\begin{aligned}
\delta_{is} &= \delta_i - \delta_s \\
&= \frac{1}{4\nu_0}\,[4\,a\cos(\nu_0\,\xi+\varphi) - 4\,a\cos(\varphi+\nu_0\,\xi+1/2\,\nu_0\,\pi)\ldots \\
&\quad -4\,a\cos(\varphi+\nu_0\,\xi+\nu_0\,\pi) + 4\,a\cos(\varphi+\nu_0\,\xi+3/2\,\nu_0\,\pi)\ldots \\
&\quad -8\,ab\cos(\nu_0\,\xi+\varphi) - a^2\sin(2\,\nu_0\,\xi+2\,\varphi)\ldots \\
&\quad +a^2\sin(2\,\varphi+2\,\nu_0\,\xi+\nu_0\,\pi) + 8\,ab\cos(\varphi+\nu_0\,\xi+1/2\,\nu_0\,\pi)\ldots \\
&\quad +a^2\sin(2\,\nu_0\,\pi+2\,\varphi+2\,\nu_0\,\xi) + 8\,ab\cos(\varphi+\nu_0\,\xi+\nu_0\,\pi)\ldots \\
&\quad -a^2\sin(2\,\varphi+2\,\nu_0\,\xi+3\,\nu_0\,\pi) - 8\,ab\cos(\varphi+\nu_0\,\xi+3/2\,\nu_0\,\pi)] \\
&= 0 \qquad \text{(A.11)}
\end{aligned}$$

is zero, also difference of two phase estimations performed with a ToF system using an ideal characteristic curve and one using a square characteristic curve is zero:

$$\begin{aligned}
\Delta\varphi &= \varphi^{(2)} - \varphi \\
&= \operatorname{atan}\left(\frac{c_3^{(2)} - c_1^{(2)}}{c_2^{(2)} - c_0^{(2)}}\right) - \operatorname{atan}\left(\frac{c_3 - c_1}{c_2 - c_0}\right) \\
&= 0 \qquad \text{(A.12)}
\end{aligned}$$

Thus a system using a quadratic photo response does not show systematic deviations of the depth estimation compared to a system using an ideal photo response. □

# Acronyms and Notation

## Acronyms and Abbreviations

| | |
|---|---|
| CCD | Charge Coupled Device |
| CMOS | Complementary Metal–Oxide–Semiconductor |
| DCNU | Dark Current Non-Uniformity |
| DSNU | Dark Signal Non-Uniformity |
| DSP | Digital Signal Processor |
| FPGA | Field-Programmable Gate Array |
| GPU | Graphics Processing Unit |
| HDR | High Dynamic Range |
| PMD | Photonic Mixing Device |
| PRNU | Photo Response Non-Uniformity |
| SNR | Signal-to-Noise Ratio |
| ToF | Time-of-Flight |

## General notation

| | |
|---|---|
| $\boldsymbol{M}$ | Matrix, for instance an image or a map of values |
| $m$ | Element of Matrix $\boldsymbol{M}$ |
| $m[t_0]$ | Element of Matrix $\boldsymbol{M}$ at time step $t_0$ |
| $m^{x_1,x_2}$ | Element of Matrix $\boldsymbol{M}$ with spatial coordinates $(x_1, x_2)$ |
| $\mathrm{i}$ | Imaginary unit $\sqrt{-1}$ |

## Greek Symbols

| | |
|---|---|
| $\varphi, \theta$ | Phase of a periodic signal |
| $\sigma$ | Standard deviation of a normal distribution |
| $\mu$ | Statistical mean |

## ToF related Symbols

| | |
|---|---|
| $N$ | Number of measurement modes |
| $\tilde{N}$ | Number of samples of the correlation function |

| | |
|---|---|
| $Q$ | Number of taps per pixel |
| $L$ | Number of acquisitions per frame |
| $R$ | Number of raw channels |
| $\boldsymbol{C}_{\tilde{n}}$ | Map of sample $\tilde{n}$ of the correlation function |
| $\boldsymbol{U}_{\tilde{n},q}$ | Map of theoretical (unknown) values of sample $\tilde{n}$, to be measured with detection unit $q$ |
| $\boldsymbol{Y}_r$ | Measured raw image of raw channel $r$, $(r \in [1,\dots,R])$ |
| $\boldsymbol{Y}_r[t_0]$ | Raw image of raw channel $r$, $(r \in [1,\dots,R])$ at time step $t_0$ |
| $\boldsymbol{Y}_{n,q}[t_0]$ | Raw image acquired in sampling mode $n$ $(n \in [1,\dots,N])$ using detection unit $q$ $(q \in [1,\dots,Q])$ at time step $t_0$ |
| $y_{n,q}^{x_1,x_2}[t_0]$ | Raw value of sampling process $n$ with detection unit $q$ of the pixel with coordinates $(x_1, x_2)$ at time step $t_0$ |
| $\hat{y}_{n,q}^{x_1,x_2}[t_0]$ | Corrected raw value of sampling process $n$ with tap $q$ of the pixel with coordinates $(x_1, x_2)$ at time step $t_0$ |
| $\boldsymbol{\Gamma}_{n,q}$ | Nonlinear transformation, modeling process of measuring theoretical (unknown) value of sample $n$ using detection unit $q$: $\boldsymbol{Y}_{n,q} = \boldsymbol{\Gamma}_{n,q}(\boldsymbol{U}_{n,q})$ |
| $\boldsymbol{R}_{n,q}$ | Rectification operator (map), correcting the raw values: $\hat{\boldsymbol{Y}}_{n,q} = \boldsymbol{R}_{n,q}(\boldsymbol{Y}_{n,q})$ |
| $\boldsymbol{\Phi}$ | Phase map |
| $\boldsymbol{D}$ | Depth map |
| $\boldsymbol{A_0}$ | Non-modulated light (map) |
| $\boldsymbol{A_1}$ | Modulation amplitude (map) |
| $\nu_0$ | Modulation frequency |
| $T_0$ | Oscillating period of a periodical signal |
| $c_0$ | Speed of light |
| $\tau$ | Time of flight |

# List of Publications

Parts of this work were already published as

**[SJ09]** **M. Schmidt** and B. Jähne. "A physical model of Time-of-Flight 3D imaging systems, including suppression of ambient light". In: *3rd Workshop on Dynamic 3-D Imaging.* Ed. by R. Koch and A. Kolb. Vol. 5742. Lecture Notes in Computer Science. Springer, 2009, pp. 1–15.

**[Sch+10]** **M. Schmidt**, M. Erz, K. Zimmermann, and B. Jähne. "Exact modelling of Time-of-Flight Cameras for Optimal Depth Maps". In: *International Conference on Computational Photography (ICCP).* Poster. 2010.

**[SZ10b]** **M. Schmidt** and K. Zimmermann. "3D Time-of-Flight Camera and Method (Reduction of Motion Artifacts of Depth Maps Acquired by 3D Time-of-Flight Cameras Utilizing a Real-Time Algorithm)". European patent application EP10188353. 2010.

**[SZ10a]** **M. Schmidt** and K. Zimmermann. "3D Time-of-Flight Camera and Method (Enabling Increased Frame Rate of Multi-Tap 3D Time-of-Flight Cameras Utilizing a Dynamic Calibration Algorithm)". European patent application EP111507008. 2010.

**[MH+10b]** L. Maier-Hein, **M. Schmidt**, A. M. Franz, T. dos Santos, A. Seitel, B. Jähne, J. M. Fitzpatrick, and H.-P. Meinzer. "Accounting for Anisotropic Noise in Fine Registration of Time-of-Flight Range Data with High-Resolution Surface Data". In: *Medical Image Computing and Computer-Assisted Intervention (MICCAI) 2010.* Vol. 6361. Lecture Notes in Computer Science. Springer, 2010, pp. 251–258.

**[SZJ11]** **M. Schmidt**, K. Zimmermann, and B. Jähne. "High Frame Rate for 3D Time-of-Flight Cameras by Dynamic Sensor Calibration". In: *Proceedings IEEE International Conference on Computational Photography (ICCP).* 2011, pp. 1–8.

**[MH+11]** L. Maier-Hein, A. Franz, M. Fangerau, **M. Schmidt**, A. Seitel, S. Mersmann, T. Kilgus, A. Groch, K. Yung, T. dos Santos, and H.-P. Meinzer. “Towards Mobile Augmented Reality for On-Patient Visualization of Medical Images”. In: *Bildverarbeitung für die Medizin (2011)*. Ed. by H. Handels, J. Ehrhardt, T. Deserno, H.-P. Meinzer, and T. Tolxdor. Springer, 2011, pp. 389–393.

# Bibliography

[3DV09] 3DVSystems. *ZCam Documentation (Camera Driver CD ROM)*. 2009.

[BM92] P. J. Besl and N. D. McKay. “A Method for Registration of 3-D Shapes”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 14 (2 Feb. 1992), pp. 239–256. ISSN: 0162-8828. DOI: 10.1109/34.121791. URL: http://portal.acm.org/citation.cfm?id=132013.132022.

[BS05] C. Bamji and K. Salama. “Method and system to differentially enhance sensor dynamic range using enhanced common mode reset”. U.S. patent application 20,080,048,100. 2005.

[Büt+07] B. Büttgen, M.-A. E. Mechat, F. Lustenberger, and P. Seitz. “Pseudonoise optical modulation for real-time 3-D imaging with minimum interference”. In: *IEEE Transactions on Circuits and Systems I: Regular papers* 54.10 (Oct. 2007), pp. 2109–2119. ISSN: 1549-8328. DOI: 10.1109/TCSI.2007.904598.

[Can08] Canesta Inc. *Introduction to 3D Vision in CMOS (White paper)*. 2008.

[Cho+10] O. Choi, H. Lim, B. Kang, Y. S. Kim, K. Lee, J. Kim, and C.-Y. Kim. “Range unfolding for Time-of-Flight depth cameras”. In: *17th IEEE International Conference on Image Processing (ICIP) 2010*. 2010, pp. 4189–4192. DOI: 10.1109/ICIP.2010.5651383.

[Con+06] R. Conroy, A. Dorrington, M. Cree, R. Künnemeyer, and B. Gabbitas. “Shape and Deformation Measurement using Heterodyne Range Imaging Technology”. In: *12th Asia-Pacific Conference on Non-Destructive Testing*. A-PCNDT, 5.-10. November 2006. 2006.

[DC95] E. Dowski, Jr. and W. T. Cathey. “Extended depth of field through wave-front coding”. In: *Applied Optics* 34.11 (Apr. 1995), pp. 1859–1866.

[DHB10] D. Droeschel, D. Holz, and S. Behnke. “Probabilistic Phase Unwrapping for Time-of-Flight Cameras”. In: *ISR/ROBOTIK 2010 - (41st International Symposium on Robotics) and ROBOTIK 2010 (6th German Conference on Robotics).* 2010.

[DW88] T. Darrell and K. Wohn. “Pyramid based depth from focus”. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR).* Ann Arbor, MI, 1988, pp. 504–509.

[EJ09] M. Erz and B. Jähne. “Radiometric and Spectrometric Calibrations, and Distance Noise Measurement of TOF Cameras”. In: *3rd Workshop on Dynamic 3-D Imaging.* Ed. by R. Koch and A. Kolb. Vol. 5742. Lecture Notes in Computer Science. Springer, 2009, pp. 28–41. DOI: 10.1007/978-3-642-03778-8_3.

[Elk+06] O. Elkhalili, O. Schrey, W. Ulfig, W. Brockherde, B. Hosticka, P. Mengel, and L. Listl. “A 64x8 pixel 3-D CMOS time-of flight image sensor for car safety applications”. In: *IEEE 32nd European Solid-State Circuits Conference (ESSCIRC). Proceedings.* 2006, pp. 568–571. URL: http://www.scientificcommons.org/20268102.

[EMV10] EMVA Standard 1288. *EMVA Standard 1288 - Standard for Characterization of Image Sensors and Cameras.* Release 3.0. European Machine Vision Association, Dec. 2010. URL: www.standard1288.org.

[Erz11] M. Erz. “Charakterisierung von Laufzeitkamerasystemen für Lumineszenzlebensdauermessungen”. Dissertation. IWR, Fakultät für Physik und Astronomie, University of Heidelberg, 2011. URL: http://www.ub.uni-heidelberg.de/archiv/11598.

[FAT11] S. Foix, G. Alenya, and C. Torras. “Lock-in Time-of-Flight (ToF) Cameras: A Survey”. In: *Sensors Journal, IEEE* 99 (Jan. 2011), pp. 1–1. DOI: 10.1109/JSEN.2010.2101060.

[Fra+09] M. Frank, M. Plaue, H. Rapp, U. Köthe, B. Jähne, and F. A. Hamprecht. “Theoretical and experimental error analysis of continuous-wave time-of-flight range cameras”. In: *Optical Engineering* 48 (2009), p. 013602. DOI: 10.1117/1.3070634.

[Fra11] R. Franke. to appear. Dissertation. University of Heidelberg, 2011.

[GYB04] S. B. Gokturk, H. Yalcin, and C. Bamji. “A Time-Of-Flight Depth Sensor - System Description, Issues and Solutions”. In: *Conference on Computer Vision and Pattern Recognition (CVPR) Workshop.* Vol. 3. 2004. DOI: `http://doi.ieeecomputersociety.org/10.1109/CVPR.2004.291`.

[Har03] P. Hariharan. *Optical Interferometry.* San Diego, CA, USA, 2003.

[Har+05] K. Hara, H. Kubo, M. Kimura, F. Murao, and S. Komori. “A linear-logarithmic CMOS sensor with offset calibration using an injected charge signal”. In: *Solid-State Circuits Conference.* San Francisco, CA, Feb. 2005, pp. 354–603. DOI: `10.1109/ISSCC.2005.1494015`.

[Has+06] F. Hasouneh, S. Knedlik, V. Peters, and O. Loffeld. “PMD Based Mobile Node Position Monitoring”. In: *Position, Location, And Navigation Symposium* (Apr. 2006), pp. 569–573.

[HE09] S. Hussmann and T. Edeler. “Performance improvement of a 3D-TOF PMD camera using a pseudo 4-phase shift algorithm”. In: *Instrumentation and Measurement Technology Conference, 2009. I2MTC ’09. IEEE.* 2009, pp. 542–546. DOI: `10.1109/IMTC.2009.5168509`.

[HHE11] S. Hussmann, A. Hermanski, and T. Edeler. “Real-Time Motion Artifact Suppression in TOF Camera Systems”. English. In: *IEEE Transactions on Instrumentation and Measurement* 60.5 (May 2011), pp. 1682–1690.

[HS09] H. Hirschmüller and D. Scharstein. “Evaluation of Stereo Matching Costs on Images with Radiometric Differences”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 31.9 (2009), pp. 1582–1599.

[HZ00] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision.* Cambridge, UK: Cambridge University Press, 2000.

[Jan07] J. R. Janesick. *Photon Transfer.* Bellingham, Washington, USA: SPIE Press, 2007.

[JGH99] B. Jähne, P. Geißler, and H. Haußecker, eds. *Handbook of Computer Vision and Applications.* San Diego: Academic Press, 1999.

[Kav+00] S. Kavadias, B. Dierickx, D. Scheffer, A. Alaerts, D. Uwaerts, and J. Bogaerts. “A logarithmic response CMOS image sensor with on-chip calibration”. In: *IEEE Journal of Solid-State Circuits* 35, issue 8 (Aug. 2000), pp. 1146–1152. DOI: `10.1109/4.859503`.

[KBK08] A. Kolb, E. Barth, and R. Koch. "ToF Sensors: New Dimensions for Realism and Interactivity". In: *CVPR 2008 Workshop on Time-of-Flight-based Computer Vision (TOF-CV)*. 2008. URL: `http://www.inb.uni-luebeck.de/publications/pdfs/KoBaKo08.pdf`.

[Kel+07] M. Keller, J. Orthmann, A. Kolb, and V. Peters. "A Simulation Framework for Time-Of-Flight Sensors". In: *Proceedings of the International IEEE Symposium on Signals, Circuits & Systems (ISSCS)*. Vol. 1. 2007, pp. 125 –128.

[KK09] M. Keller and A. Kolb. "Real-time Simulation of Time-Of-Flight Sensors". In: *Simulation Practice and Theory* 17 (2009), pp. 967–978.

[KKS08] S. Knorr, M. Kunter, and T. Sikora. "Stereoscopic 3D from 2D video with super-resolution capability". In: *Image Communication* 23.9 (2008), pp. 665–676. ISSN: 0923-5965. DOI: `10.1016/j.image.2008.07.004`.

[KN05] M. Kuijk and D. V. Nieuwenhove. "TOF rangefinding with large dynamic range and enhanced background radiation suppression". U.S. patent 7,268,858. 2005.

[KRI06] T. Kahlmann, F. Remondino, and H. Ingensand. "Calibration for increased accuracy of the range imaging camera SwissRanger". In: *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* XXXVI.5 (2006), pp. 136–141.

[KSS00] N. Kirchgeßner, H. Scharr, and U. Schurr. "3D-Modellierung von Pflanzenblättern mittels eines Depth-from-Motion Verfahrens". In: *Proceedings of the 22th DAGM Symposium on Pattern Recognition.* Kiel, Germany, 2000, pp. 381–388.

[Lan00] R. Lange. "3D Time-of-Flight Distance Measurement with Custom Solid-State Image Sensors in CMOS/CCD-Technology". PhD thesis. Department of Electrical Engineering and Computer Science at University of Siegen, 2000.

[LFD08] A. Levin, W. T. Freeman, and F. Durand. "Understanding camera tradeoffs through a Bayesian analysis of light field projections". In: *Proceedings of the European Conference on Computer Vision (ECCV)* (Oct. 2008).

[Lin11] Z. Lin. to appear. Dissertation. IWR, Fakultät für Mathematik und Informatik, University of Heidelberg, 2011.

[Lip08] *PHOTOGRAPHY. - Reversible Prints. Integral Photographs.* Note by M. G. Lippmann. (translation by Frédo Durand, MIT CSAIL). Mar. 1908. URL: `http://people.csail.mit.edu/fredo/PUBLI/Lippmann.pdf`.

[LK06] M. Lindner and A. Kolb. "Lateral and Depth Calibration of PMD-Distance Sensors". In: *International Symposium on Visual Computing (ISVC06).* Vol. 2. Lake Tahoe, Nevada: Springer, 2006, pp. 524–533. ISBN: 978-3-540-48626-8.

[LK07] M. Lindner and A. Kolb. "Calibration of the intensity-related distance error of the PMD TOF-Camera". In: *SPIE: Intelligent Robots and Computer Vision XXV* 6764 (2007), pp. 6764–35.

[LK09] M. Lindner and A. Kolb. "Compensation of Motion Artifacts for Time-of-Flight Cameras". In: *Dyn3D '09: Proceedings of the DAGM 2009 Workshop on Dynamic 3D Imaging.* Jena, Germany: Springer, 2009, pp. 16–27. ISBN: 978-3-642-03777-1. DOI: `10.1007/978-3-642-03778-8_2`.

[Lot+07] O. Lottner, A. Sluiter, K. Hartmann, and W. Weihs. "Movement Artefacts in Range Images of Time-of-Flight Cameras". In: *International Symposium on Signals, Circuits & Systems - ISSCS 2007.* Vol. 1. 2. Iasi, Romania, 2007, pp. 117–120. URL: `http://scs.etc.tuiasi.ro/isscs2007/`.

[LS01] R. Lange and P. Seitz. "Solid-state time-of-flight range camera". In: *IEEE Journal of Quantum Electronics* 37.3 (2001), pp. 390–397.

[McC+10] S. H. McClure, M. J. Cree, A. A. Dorrington, and A. D. Payne. "Resolving depth measurement ambiguity with commercially available range imaging cameras". In: *Proceedings of SPIE: the International Society for Optical Engineering* 7538 (2010).

[MES06] MESA Imaging AG. *SwissRanger SR-3000 Manual.* Version 1.02. Oct. 2006.

[MES08] MESA Imaging AG. *SR-4000 Data Sheet.* Version: August 2008. 2008. URL: `http://www.mesa-imaging.ch/pdf/SR4000_Data_Sheet_rev1.0.pdf`.

[MES11] MESA Imaging AG. *SR4000 User Manual.* Version 2.0. version 2.0, consultation date 04/15/2011. Apr. 2011. URL: http://www.mesa-imaging.ch/dlm.php?fname=customer/Customer_CD/SR4000_Manual.pdf.

[MH+10a] L. Maier-Hein, T. R. dos Santos, A. M. Franz, and H.-P. Meinzer. "Iterative closest point algorithm in the presence of anisotropic noise". In: *Bildverarbeitung für die Medizin 2010 (BVM)* (2010). Ed. by T. M. Deserno, H. Handels, H.-P. Meinzer, and T. Tolxdorff, pp. 231–235.

[MH+10b] L. Maier-Hein, M. Schmidt, A. M. Franz, T. dos Santos, A. Seitel, B. Jähne, J. M. Fitzpatrick, and H.-P. Meinzer. "Accounting for Anisotropic Noise in Fine Registration of Time-of-Flight Range Data with High-Resolution Surface Data". In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2010.* Vol. 6361. Lecture Notes in Computer Science. Springer, 2010, pp. 251–258. DOI: 10.1007/978-3-642-15705-9_31.

[MH+11] L. Maier-Hein, A. Franz, M. Fangerau, M. Schmidt, A. Seitel, S. Mersmann, T. Kilgus, A. Groch, K. Yung, T. dos Santos, and H.-P. Meinzer. "Towards Mobile Augmented Reality for On-Patient Visualization of Medical Images". In: *Bildverarbeitung für die Medizin (2011).* Ed. by H. Handels, J. Ehrhardt, T. Deserno, H.-P. Meinzer, and T. Tolxdorff. Springer, 2011, pp. 389–393.

[Möl+05] T. Möller, H. Kraft, J. Frey, M. Albrecht, and R. Lange. "Robust 3D Measurement with PMD Sensors". In: *Range Imaging Day Zurich.* Vol. Section 5. 2005.

[Ng+05] R. Ng, M. Levoy, M. Brédif, G. Duval, M. Horowitz, and P. Hanrahan. *Light Field Photography with a Hand-held Plenoptic Camera.* Tech. rep. Stanford, Feb. 2005.

[Nie+08] D. V. Nieuwenhove, W. V. D. Tempel, R. Grootjans, and M. Kuijk. "Time-Of-Flight distance sensor with enhanced dynamic range". In: *International Journal of Intelligent Systems Technologies and Applications* 5 (2008), pp. 246–254.

[Ogg+04] T. Oggier, M. Lehmann, R. Kaufmann, M. Schweizer, M. Richter, P. Metzler, G. Lang, F. Lustenberger, and N. Blanc. "An all-solid-state optical range camera for 3D real-time imaging with sub-centimeter depth resolution". In: *Society of Photo-Optical Instrumentation Engineers (SPIE)*

*Conference Series*. Ed. by L. Mazuray, P. Rogers, and R. Wartmann. Vol. 5249. Feb. 2004, pp. 534–545. DOI: `10.1117/12.513307`.

[PC03] J. Philip and K. Carlsson. “Theoretical investigation of the signal-to-noise ratio in fluorescence lifetime imaging”. In: *Journal of the Optical Society of America A, optics, image science and vision.* 20 (Feb. 2003), pp. 368–379.

[Pet+07] V. Peters, O. Loffeld, K. Hartmann, and S. Knedlik. “Modeling and Bistatic Simulation of a High Resolution 3D PMD-Camera”. In: *EUROSIM 2007 (6th EUROSIM Congress on Modelling and Simulation), Ljubljana, Slovenia.* 2007.

[Pla06] M. Plaue. *Analysis of the PMD Imaging System.* Tech. rep. Interdisciplinary Center for Scientific Computing, University of Heidelberg, 2006.

[Rap07] H. Rapp. “Experimental and Theoretical Investigation of Correlating TOF-Camera Systems”. Diplomarbeit. IWR, Fakultät für Physik und Astronomie, Universität Heidelberg, 2007.

[Ras+06] *Computational Photography.* Eurographics 2006, 2006. URL: `http://www.cs.northwestern.edu/~jet/docs/EG2006STAR_CompPhotog.pdf`.

[RH07] T. Ringbeck and B. Hagebeuker. “A 3D Time of Flight camera for object detection”. In: *Measurement* 9 (2007). URL: `http://www.ifm.com/obj/O1D_Paper-PMD.pdf`.

[Sch+00] M. Schanz, C. Nitta, A. Bussmann, B. Hosticka, and R. Wertheimer. “A high-dynamic-range CMOS image sensor for automotive applications”. In: *IEEE Journal of Solid-State Circuits* 35, issue 7 (Aug. 2000), pp. 932–938. DOI: `10.1109/4.848200`.

[Sch08a] M. Schmidt. “Spatiotemporal Analysis of Range Imagery”. Dissertation. IWR, Fakultät für Physik und Astronomie, University of Heidelberg, 2008. URL: `http://www.ub.uni-heidelberg.de/archiv/8879/`.

[Sch08b] M. Schmidt. “Optische Methoden zur Form- und Positionserkennung von Körpern in Werkzeugmaschinen”. MA thesis. Friedrich-Schiller-Universität Jena, 2008.

[Sch08c] M. Schmidt. “Verfahren zur dreidimensionalen Bilddatenerfassung (A Method for 3D Image Acquisition)”. German patent application DE102007007775A1. 2008.

[Sch09] M. Schmidt. “Three-dimensional Image Acquisition Method”. PCT patent application PCT/EP2008/001121. 2009.

[Sch+10] M. Schmidt, M. Erz, K. Zimmermann, and B. Jähne. “Exact modelling of time-of-flight cameras for optimal depth maps”. In: *International Conference on Computational Photography (ICCP) 2010.* Poster. 2010.

[Sch+95] R. Schwarte, H. G. Heinol, Z. Xu, and K. Hartmann. “New active 3D vision system based on rf-modulation interferometry of incoherent light”. In: *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*. Ed. by D. P. Casasent. Vol. 2588. Oct. 1995, pp. 126–134. URL: `http://adsabs.harvard.edu/abs/1995SPIE.2588..126S`.

[Sei08] P. Seitz. “Quantum-noise limited distance resolution of optical range imaging techniques”. English. In: *IEEE Transactions on Circuits and Systems I: Regular papers* 55.8 (Sept. 2008), pp. 2368–2377. ISSN: 1549-8328. DOI: `10.1109/TCSI.2008.918231`.

[Sha09] J. Shan. *Topographic Laser Ranging and Scanning: Principles and Processing.* Ed. by C. K. Toth. Boca Raton, FL: CRC Press, 2009.

[SJ09] M. Schmidt and B. Jähne. “A physical model of Time-of-Flight 3D imaging systems, including suppression of ambient light”. In: *3rd Workshop on Dynamic 3-D Imaging.* Ed. by R. Koch and A. Kolb. Vol. 5742. Lecture Notes in Computer Science. Springer, 2009, pp. 1–15. DOI: `10.1007/978-3-642-03778-8_1`.

[SK98] Y. Y. Schechner and N. Kiryati. “Depth from Defocus vs. Stereo: How Different Really are They?” In: *International Conference on Pattern Recognition (ICPR).* 1998, pp. 1784–1786.

[SPH08] M. Stürmer, J. Penne, and J. Hornegger. “Standardization of intensity-values acquired by Time-of-Flight-cameras”. In: *Computer Vision and Pattern Recognition Workshops.* IEEE Computer Society Conference on CVPR. Anchorage, AK, 2008, pp. 1–6.

[SS02] D. Scharstein and R. Szeliski. “A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms”. In: *Intational Journal on Computer Vision* 47 (1–3 Apr. 2002), pp. 7–42. ISSN: 0920-5691. DOI: `10.1023/A:1014573219977`.

[SSH95] T. Spirig, P. Seitz, and F. Heitger. “The lock-in CCD. Two-dimensional synchronous detection of light”. In: *IEEE J.Quantum Electronics* 31 (1995), pp. 1705–1708.

[Sto+04] G. Storm, J. Hurwitz, D. Renshaw, K. Findlater, R. Henderson, and M. Purcell. “Combined linear-logarithmic CMOS image sensor”. In: *Solid-State Circuits Conference.* Feb. 2004, pp. 116–517. DOI: `10.1109/ISSCC.2004.1332621`.

[SZ10a] M. Schmidt and K. Zimmermann. “3D Time-of-Flight Camera and Method (Enabling Increased Frame Rate of Multi-Tap 3D Time-of-Flight Cameras Utilizing a Dynamic Calibration Algorithm)”. European patent application EP111507008. 2010.

[SZ10b] M. Schmidt and K. Zimmermann. “3D Time-of-Flight Camera and Method (Reduction of Motion Artifacts of Depth Maps Acquired by 3D Time-of-Flight Cameras Utilizing a Real-Time Algorithm)”. European patent application EP10188353. 2010.

[SZJ11] M. Schmidt, K. Zimmermann, and B. Jähne. “High Frame Rate for 3D Time-of-Flight Cameras by Dynamic Sensor Calibration”. In: *Proceedings IEEE International Conference on Computational Photography (ICCP).* 2011, pp. 1–8. DOI: `10.1109/ICCPHOT.2011.5753121`.

[THI98] N. Tu, R. Hornsey, and S. Ingram. “CMOS active pixel image sensor with combined linear and logarithmic mode operation”. In: *IEEE Canadian Conference on Electrical and Computer Engineering.* Waterloo, Ont., Canada, May 1998, pp. 754–757. DOI: `10.1109/CCECE.1998.685607`.

[Vee+07] A. Veeraraghavan, R. Raskar, A. Agrawal, A. Mohan, and J. Tumblin. “Dappled Photography: Mask Enhanced Cameras for Heterodyned Light Fields and Coded Aperture Refocusing”. In: *ACM SIGGRAPH 2007.* 2007.

[WWK06] A. Wiegmann, H. Wagner, and R. Kowarschik. “Human face measurement by projecting bandlimited random patterns”. In: *Optics Express* 14, Issue 17 (2006), pp. 7692–7698.

[XS93] Y. Xiong and S. A. Shafer. “Depth from focusing and defocusing”. In: *Proceedings CVPR’93, New York City, NY.* Ed. by Y. Xiong and S. A. Shafer. Washington, DC, 1993, pp. 68–73.

[Xu99] Z. Xu. “Investigation of 3D-Imaging Systems Based on Modulated Light and Optical RF-Interferometry (ORFI)”. In: *ZESS Forschungsberichte* 14 (1999).

[YIM06] G. Yahav, G. J. Iddan, and D. Mandelboum. “3D Imaging Camera for Gaming Application”. In: *International Conference on Consumer Electronics (ICCE) 2007* (2006). DOI: 10.1109/ICCE.2007.341537. URL: http://www.3dvsystems.com/technology/3D\%20Camera\%20for\%20Gaming-1.pdf.