Institute for Visualization and Interactive Systems

University of Stuttgart
Universitätsstraße 38
D–70569 Stuttgart

Diplomarbeit Nr. 3737

Evaluation of depth-camera-systems for usage in semi-controlled assembly environments

Michael Matheis

Course of Study:	Informatik
Examiner:	Prof. Dr. Albrecht Schmidt
Supervisor:	Yomna Abdelrahman, M.Sc.
Commenced:	December 1, 2015
Completed:	June 1, 2016
CR-Classification:	I.4.1, I.2.10, H.5.m

Abstract

With the availability of affordable depth-camera-systems like the Microsoft Kinect, Depth Imaging has seen a fast-growing number of applications in many different fields over the last years. Such systems can however be based on different measurement principles with widely differing parameters and hence are difficult to evaluate against a single benchmark. While accuracy and precision of depth-camera-systems inherently vary significantly with measuring distance and changing environments, and therefore impose heavy constraints on real world applications, they even allow for automated quality assurance in controlled environments. Context aware assistive systems in manual assembly environments push these boundaries by employing quality assurance in more open environments, where distracting influences by the worker or the work-space environment cannot be ruled out. The thesis concerns itself with the exploration and evaluation of different depth measuring approaches (e.g. Time of Flight, Structured Light, Stereo Vision) for usage in semi-controlled assembly environments. The still underexplored effects of material properties on measurements are experimentally evaluated and the resulting limitations of each approach for usage in assembly environments are discussed.

Contents

List of Figures

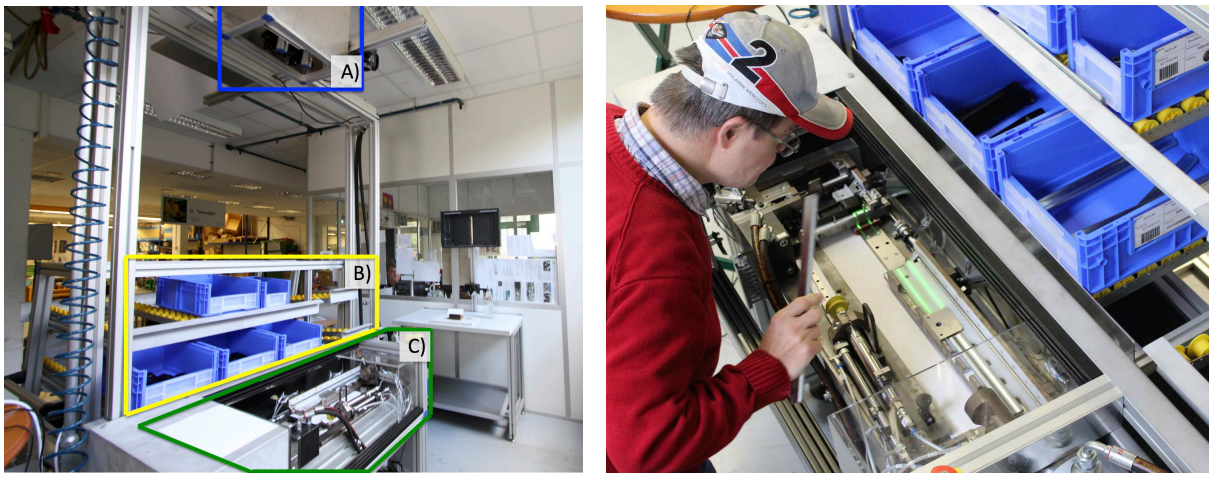
1 Introduction

Obtaining depth information for points in a 2D Image, also known as Depth Imaging, is an increasingly important branch of Computer Vision, finding application in a vast range of fields as various as geology, automation, interaction, robotics and microbiology. Depth-camera-systems are able to capture depth images in a single shot, allowing to measure whole objects and scenes in real-time.

(a) **(b)**

Figure 1.1: (a) Color image of a scene and (b) its corresponding depth image [Sze10]

First used mostly in scientific contexts and highly specialized applications, Depth Imaging has become almost ubiquitous in recent years through the availability of compact and affordable depth-camera-systems. Such systems can however be based on different measurement principles with widely differing parameters and hence are difficult to evaluate against a single benchmark. Due to the wide range of applications, requirements on depth-camera-systems can also be hugely different, depending on usage and environment.



(a) (b)

Figure 1.2: (a) Workplace with assistive system (b) providing instructions to the worker with in-situ projection [FBB+15]

1.1 Motivation

Accuracy and precision of depth-camera-systems inherently vary significantly with measuring distance and changing environments, and therefore impose heavy constraints on real world applications. On one hand Depth Imaging is used for quality assurance and inspection in industrial manufacturing, where accurate measurements are very important, but the measurement environment is usually strictly controlled to ensure optimal conditions and avoid any disrupting factors. On the other hand Depth Imaging is used in Robot Navigation and Interactive systems where accuracy is not the primary issue, but instead largely different and uncontrolled environments have to be considered.

Recently Funk et al. introduced a context aware assistive system for manual assembly environments [FS15]. The system consists of a projector and depth-camera-system mounted above the workplace as shown in Figure 1.2(a). By comparing the actual depth data to target states it can provide the worker automatically with the instructions for the current working-step and give immediate feedback.

In order to reliably detect smaller parts or slight mistakes in the assembly, high accuracy and precision is required. While typical working environments however do not necessarily provide optimal measurement conditions, they are still clearly constraint in many aspects. Assistive systems in assembly environments can therefore provide a good vantage point for pushing the boundaries of accurate depth measurement in open and less controlled environments.

1.2 Scenario

To establish a working definition, the assembly scenario and its requirements on depth-camera-systems can be narrowed down along five primary dimensions: measurement area/volume, scene dynamic, environmental factors, material properties and system placement.

Measurment volume – The area or volume to be covered can vary from whole outdoor locations down to samples of microscopic size. While assembly workspaces can be of larger roomsize, where the workers have to move around, here only workplaces for assembling smaller parts are considered. So the volume is naturally restricted by the workers' arm lengths.

Scene dynamic – A scene can be completely static or it can change over time. Change can again be differentiated into continuous change (motion) and spontaneous change, that leaves time frames in which the scene can be considered static for measurements. Assuming the workpieces can be placed within a working area during assembly, motion can be excluded as a factor and the measurement time needs only be restricted in order to keep the system's response time acceptable.

Environmental factors – For outdoor applications systems have to cope with direct sunlight, bigger temperature changes, rain and possibly other weather effects. Indoors these factors are either not present or can be avoided. (By not puting a workspace at a window with direct sunlight, using climate control, etc.) Still there are other factors to consider, like indirect sunlight, artificial light sources, reflective surfaces, other active sensors in proximity as well as room climate. It is assumed that avoidable environmental factors are keept at a minimum for assembly workspaces and other factors are at least ensured to lie within a specified range (as also required by law in some cases), so that extreme cases do not have to be considered. Additionally assisitive systems might themselves employ light sources (projectors) in order to provide in-situ instructions and feedback to the worker.

Material properties – Another crucial factor are the material properties of the scenes or objects to be measured, as some surface material can be more or less challenging to measure depending on the used method. This is tightly related to the accuracy that is aimed for. Individual and detailed material composition of objects becomes more important, the higher the requirements for accuracy are. In the assembly scenario the work-pieces and their material properties are usually well specified, but also very diverse. The capability of measuring a wide range of surface material is therefore a key requirement, especially when requiring millimeter accuracy as not unlikely in manual assembly.

System placement – Last but not least are requirements concerning the possible placement of depth-camera-systems, which also includes the question whether the sensor is fixed to certain position or attached to a moving platform, which yields a whole different class of applications. For the assembly scenario the placement should ideally not restrict the worker in any way. In practice this results in a placement of at least a meter above the workplace.

According to this categorization the target scenario can be roughly summed up as following: Fixed Sensor placement above workplace with at least one meter distance from assembly area. Accurate measurement of components made from many different materials with about 1 millimeter tolerance. Measurement of scene possible in less than a second in bright indoor environment without direct sunlight.

1.3 Objective / aims

One of the main challenges of the assembly scenario lies in the combination of greater variety of possible materials with high accuracy requirements while providing less optimal measurement conditions. Accuracy and precision of depth-camera-systems is known to depend on material properties. Currently it is however still difficult to make more concrete statements about limitations and capabilities of different systems.

This thesis concerns itself with the exploration and evaluation of different depth measuring approaches (e.g. Time of Flight, Structured Light, Stereoscopy) for usage in semi-controlled assembly environments. It therefore also aims to clarify prerequisites and limitations of using depth-camera-systems for quality assurance in manual assembly environments. The focus lies on the underexplored effects of material properties on measurements, which are experimentally evaluated.

1.4 Overview

The thesis is structured into seven chapters as follows:

Chapter 1 – Introduction: introduces Depth Imaging and describes the assembly scenario providing the context for evaluation.

Chapter 2 – Background: provides a general overview of depth measuring methods and theoretical background to current state of the art optical depth measurement technologies.

Chapter 3 – Related work: investigates previous evaluations aiming at assessing the capabilities and limitations of depth-camera-systems.

Chapter 4 – Current depth-camera-systems: describes the three depth-camera-systems used for experimental evaluation in more detail and puts them into context with other available systems on the market.

Chapter 5 – Experiments: reports setup and results of the conducted experimental evaluations.

Chapter 6 – Discussion: discusses limitations of each depth measurement method on the basis of the experimental results.

Chapter 7 – Conclusion: summarizes findings and presents further research opportunities.

2 Background

Given the requirements of the assembly scenario described in Section 1.2, it is clear that only some depth measurement methods come into consideration. This chapter gives a brief overview on depth measurement in general and then goes on to provide the theoretical background for optical methods, which are investigated in this thesis.

2.1 Depth measurement

There are many different methods to determine the depth of objects. They can be divided in contact based methods (which might even involve destroying the object in question) and non-contact methods. Non-contact methods are either based on looking at what is reflected from an object's surface (reflective methods) or by looking at what passes through it (transmissive methods). Though the later one requires using x-rays and computer tomography, which is neither practical nor advisable in a working environment with human interaction. This leaves methods based on the reflectance of light, microwaves or ultrasound. While microwaves are suitable for measuring depth of some types of objects, they easily pass through many different materials and are therefore not suitable for measuring objects of a wide range of materials. Ultrasonic

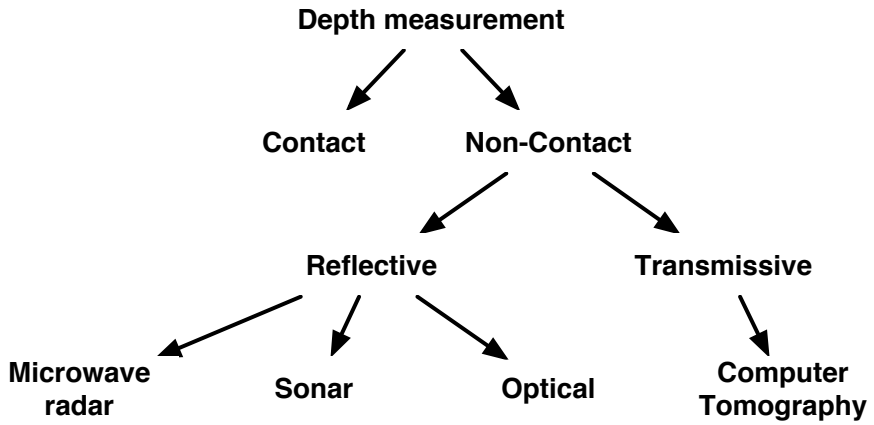
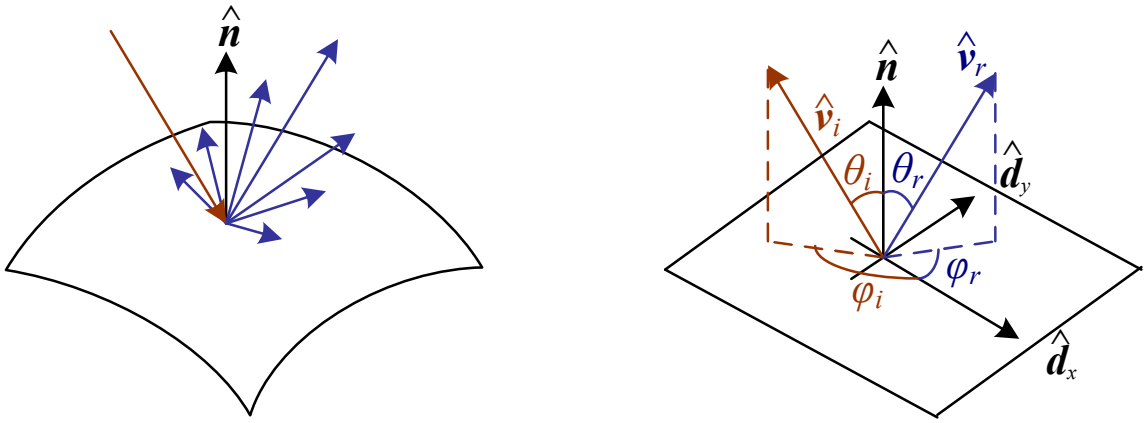


Figure 2.1: classification of depth measurement methods

waves on the other hand get reflected by most surfaces but are quite different from electromagnetic waves and hard to use for accurate and precise measurements. They are however interesting for helping in cases where optical measurements fail. This finally leaves only optical methods, working either within the visible spectrum of light or close to the lower end in the near infrared spectrum. The underlying basis for measurements is therefore given by an object's surface reflectance.

2.2 Surface reflectance

When light hits a surface it can be absorbed, transmitted or reflected. Reflective measurement methods can only work with the portion of light that is reflected by the surface. While absorbed light can be ignored, light that is transmitted may be reflected back to the sensor from a different point causing disruption or false readings. When regarding an opaque surface however only reflectance has to be considered. Generally the reflected light can be scattered in all possible directions as shown in Figure 2.2(a). How exactly the light is reflected depends on the surface's microstructure and the wavelength of the light. [NIK91]



(a) **(b)**

Figure 2.2: (a) Light Scattering on a surface and (b) its parametrization via the bidirectional reflectance function (BRDF) [Sze10]

In theory a surface's reflectance for a specific wavelength λ can be fully described by its bidirectional reflectance distribution function $BRDF_\lambda(\hat{v}_i, \hat{v}_r)$, which defines the ratio of light getting reflected into direction $\hat{v}_r$ when light is hitting the surface from direction $\hat{v}_i$. The directions can be specified by two angles θ and φ relative to the surface's normal vector $\hat{n}$ as shown in Figure 2.2(b): $BRDF_\lambda(\theta_i, \varphi_i, \theta_r, \varphi_r)$ If the surface has isotropic

reflectance i.e. if rotating the surface along the normal does not change the reflection as it is the case with many surfaces, only the differences between the horizontal angles φ_i and φ_r matters and its BRDF can thus be defined as follows:

$$\boldsymbol{BRDF}_{isotropic,\lambda}(\theta_i, \varphi_i, \theta_r, \varphi_r) = \boldsymbol{BRDF}_{\lambda}(\theta_i, 0, \theta_r, \varphi_r - \varphi_i) \tag{2.1}$$

This simplest version of the BRDF which assumes an opaque, isotropic and uniform surface therefore still possesses three degrees of freedom. Physically the BRDF is only restricted by *conservation of energy* (2.2), i.e. in sum there cannot be more light reflected than incoming, and *Helmholtz reciprocity* (2.3), i.e. switching incoming and outgoing light direction does not change the reflectance-ratio.

$$\forall \hat{\boldsymbol{v}}_i \int_{\hat{\boldsymbol{v}}_r \in \Omega} \boldsymbol{BRDF}_{\lambda}(\hat{\boldsymbol{v}}_i, \hat{\boldsymbol{v}}_r) \boldsymbol{cos}\theta_r d\hat{\boldsymbol{v}}_r \leq 1 \tag{2.2}$$

$$\boldsymbol{BRDF}_{\lambda}(\hat{\boldsymbol{v}}_i, \hat{\boldsymbol{v}}_r) = \boldsymbol{BRDF}_{\lambda}(\hat{\boldsymbol{v}}_r, \hat{\boldsymbol{v}}_i) \tag{2.3}$$

While the BRDF can be almost arbitrarily complex, which enables applications as holography, most ordinary surfaces exhibit more regular reflectance. Thus there have been a lot of efforts of modeling the BRDF of certain types of surfaces through simpler formulas, especially for usage in computer graphics. There are empirical models mainly just trying to fit obtained reflectance measurements as close as possible and theoretical models that are solely based on physical parameters that (at least in theory) can be measured independently. [War92]

From the perspective of reflective measurement methods the ideal case would be a perfect matte surface (2.4), i.e. a surface which equally diffuses incoming light into all directions (also known as *lambertian surface*), so that a sensor is able to pick up the reflections from any angle.

$$\boldsymbol{BRDF}_{\lambda}(\hat{\boldsymbol{v}}_i, \hat{\boldsymbol{v}}_r) = \boldsymbol{c} > 0 \tag{2.4}$$

The worst case would be of course a surface that does not reflect any light, either by absorbing all light or by being translucent and letting all light pass through. The contrast of surfaces which diffuse light into all directions are those that basically reflect incoming light only in one direction like a perfect mirror, whose BRDF would be 1 for all cases where $\boldsymbol{v}_i$ is $\boldsymbol{v}_r$ rotated by 180° around the surface normal $\hat{\boldsymbol{n}}$ and 0 for all other cases. In that case the sensor would only be able to pick up reflections coming from one specific direction and therefore render all methods discussed in this thesis inapplicable.

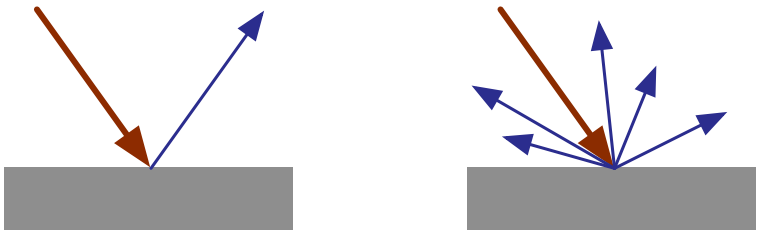


Figure 2.3: Specular and diffuse reflection

Most surfaces however exhibit both diffuse and (mirror-like) specular reflection in various degrees and for some types of surfaces the BRDF can be well approximated[1] for example by using the Ward-model [War92] given two constants ρ_d and ρ_s, corresponding with general diffuse respectively specular reflectance of the surface.

2.3 Passive Stereo Vision

A popular and well researched way of Depth Imaging is basically imitating human binocular vision. Our two eyes give us two slightly different views of a scene, which our brain can combine to create depth perception. Similarly in Computer Vision two camera images can be combined to calculate the distance from the camera-system. All this requires is picking up already present light reflected from external light sources. Hence this method can be classified as passive.

In more general terms we are trying to determine the position of a point $\boldsymbol{p}$ in 3D space given corresponding images from two (or more) cameras at known location. This is known as the problem of triangulation, which is illustrated in Figure 2.4. There we have two camera images with a correspondence in the points $\boldsymbol{x_0}$ and $\boldsymbol{x_1}$. Given the position in the images we can obtain two lines along the direction from which the light entered the cameras. The position of $\boldsymbol{p}$ can then be approximated by finding the point that minimizes the distance to each of the lines.

The main challenge of Stereo Vision hence reduces to finding the correspondences in the two given images. For this however it is very important that the two cameras have the same response. The ideal case would be having two perfect identical cameras aligned on a horizontal axis facing in the exact same direction. The depth value of a point identified in both images is then directly correlated and inversely proportional to the horizontal difference in the image coordinates.

[1] How well certain BRDF models fit measured data has been analyzed in [NDM05]

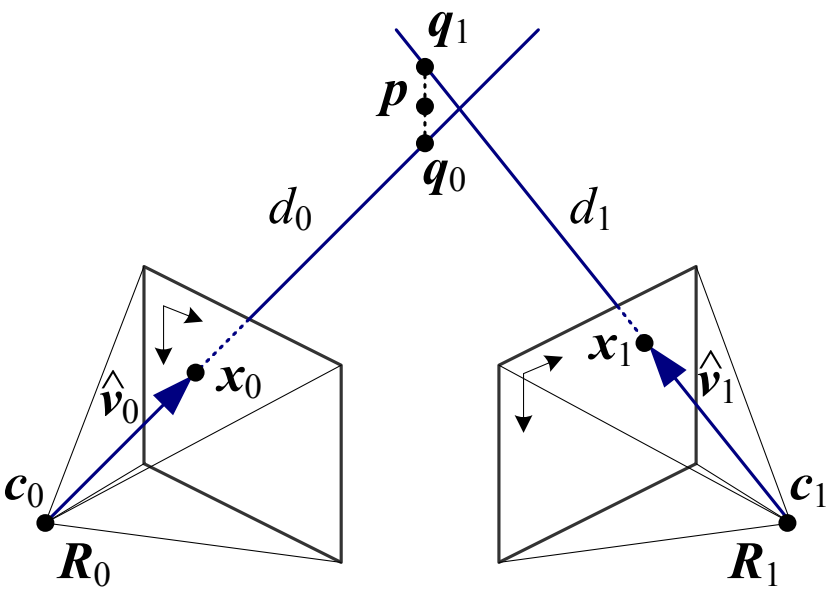
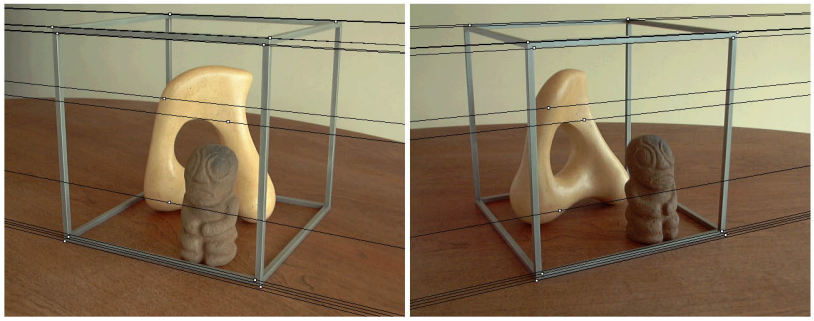


Figure 2.4: Stereo Triangulation [Sze10]

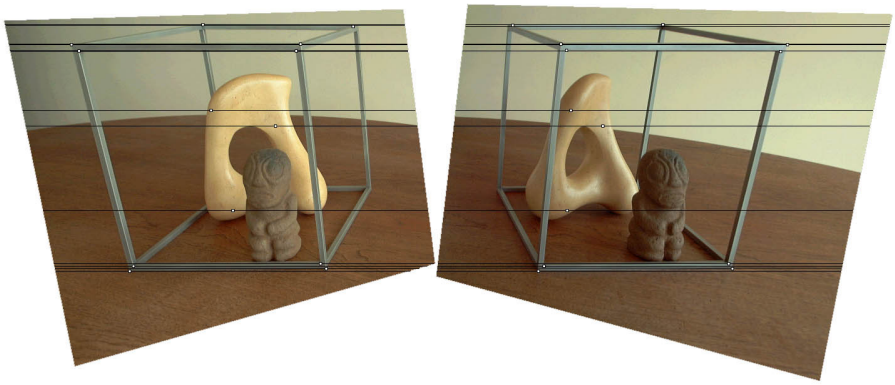
In reality optical lenses usually cause distortions in the image and no pair of cameras is exactly identical. To compensate for this the cameras have to be calibrated. For example by taking several images of a known checkerboard pattern from different distances and angles. This also allows to determine the relative position and orientation of the cameras to each other. The calibration information can then be used to rectify the camera images so that both images are displaced only on the horizontal axis, even if the cameras are not aligned horizontally and orientated differently. [LZ99] This limits the search for correspondences only to one dimension of the image instead of searching in two dimensions.

Finding correspondences in stereo images is however not a trivial task. Due to the different view angles, the pixel values of a point on a surface can be quite different between the images, depending on the scene lighting and surface reflectance. (As can also be seen in Figure 2.5) It might therefore be necessary to compare larger regions of the images in order to find correct matches.

Accordingly, there are plenty of different stereo correspondence algorithms with widely varying computational complexity. They can be divided into *local* (window based) and *global* algorithms. [SS01] While *local* algorithms only regard a limited region for the calculation of each pixel's disparity value, *global* algorithms determine a single disparity value in dependence of (potentially) all other values. The best choice of algorithm depends on scene assumptions and computational time constraints. This gives room for



(a)



(b)

Figure 2.5: A pair of stereo images (a) before and (b) after rectification [LZ99]

optimization, but if the system should be able to process several frames within a second, the possibilities are ultimately limited.

Though even with an optimal correspondence algorithm, there are principle limitations of Passive Stereo Vision. First the scene has to be sufficiently lit for the camera to pick up surface texture. If a surface however has no visible texture there is no way to identify a certain point on that surface and find its location in the second camera image. The same problem arises if too much light is reflected from glossy surfaces into the camera causing over-saturation of the sensor resulting in the loss of texture information. Additionally a point has to be visible to both cameras to determine its depth, which may be impossible to achieve for all relevant points due to occlusion of one scene object through another. Though this can in principle be overcome by adding additional cameras where necessary.

2.4 Active Stereo Vision

A simple solution to include surfaces with little or no visible texture into the measurement consists in actively projecting a light pattern onto these surfaces as shown in Figure 2.6. Since the pattern only serves to adding texture it can be optimized just for that purpose. (Fernandez et al. propose a random greyscale pattern [FFS12]) The resolution of the pattern should match the resolution of the stereo cameras. (A lower resolution pattern reduces accuracy, a higher resolution pattern might be completely useless) Since the projected pattern has to be clearly visible at different ranges, a projection technique with high depth of field like laser projection is preferred.

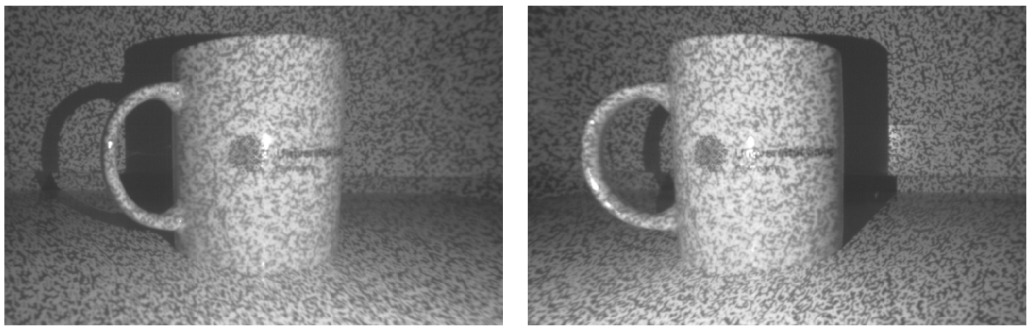


Figure 2.6: Stereo images of a cup with added texture by a projected pattern

Instead of visible light, invisible near infrared (IR) light (typically at around 800nm wavelength) can be used to employ Active Stereo Vision in unobtrusive interactive systems. This has an additional advantage: indoor environments with artificial lighting usually have little or no active IR light sources, which could disturb the pattern projection or cause additional unintended specular reflections.

Additionally the active projection opens up the possibility of increasing accuracy by taking multiple measurements while moving or changing the pattern. [USMF93] (At the expense of a lower frame-rate and increased motion blur)

Adding a projector to the stereo cameras does however not only increase complexity and costs but also require that points to be measured are both visible to the cameras as well as illuminated by the projector. Using multiple systems is usually not simply possible with active measurements methods, but since the projected pattern can be random, an overlapping of patterns from multiple projectors is unlikely to cause disruption. Therefore Active Stereo Vision setups can easily be scaled up by adding additional systems, which is a noteworthy trait when considering active measurement methods.

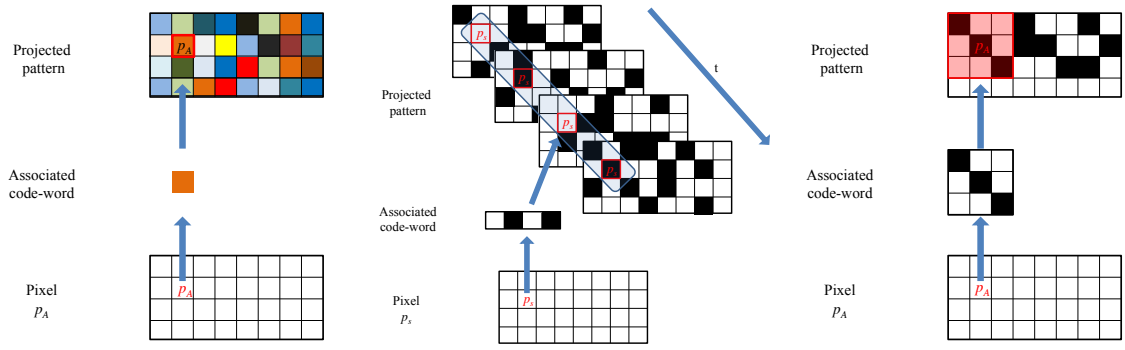
2.5 Structured Light

Instead of using a setup with two cameras (and possibly an additional projector) it is also possible to measure depth with just one camera and a projector. Since in optical systems the path of the light is always the same in both directions (*helmholtz reciprocity*), the principle of triangulation as shown in Figure 2.4 still works as long as both light paths $\boldsymbol{d_0}$ and $\boldsymbol{d_1}$ can be determined.

The paths can be calculated given the relative position and orientation of the camera and the projector and their optical parameters, which can be obtained by a calibration procedure similar to the calibration of stereo cameras [Tro95] If the light path for each point on the projected image and on the recorded camera image is known, the only problem remaining is to find out the corresponding projector image coordinate for a point on the camera image.

The most simple way of doing this is by illuminating only a single pixel or line of the projected image, which eliminates the correspondence problem by giving it only one possible solution. But this requires taking a lot of images in succession if the depth of the whole scene should be measured. To decrease the number of necessary images to be taken, the coordinates can be encoded into the projected images.

There are many different encoding strategies, they can however be categorized by a three important properties as illustrated in Figure 2.7.



(a) color encoding **(b)** temporal encoding **(c)** spatial encoding

Figure 2.7: Different encoding possibilities [DZC12]

The first is the number of different color or intensity levels used. The higher the number the more information can be encoded into a single image, but the harder it is to correctly extract the coding from the camera image. Especially when measuring objects with widely varying surface reflectance, different color or intensity levels can be hard to

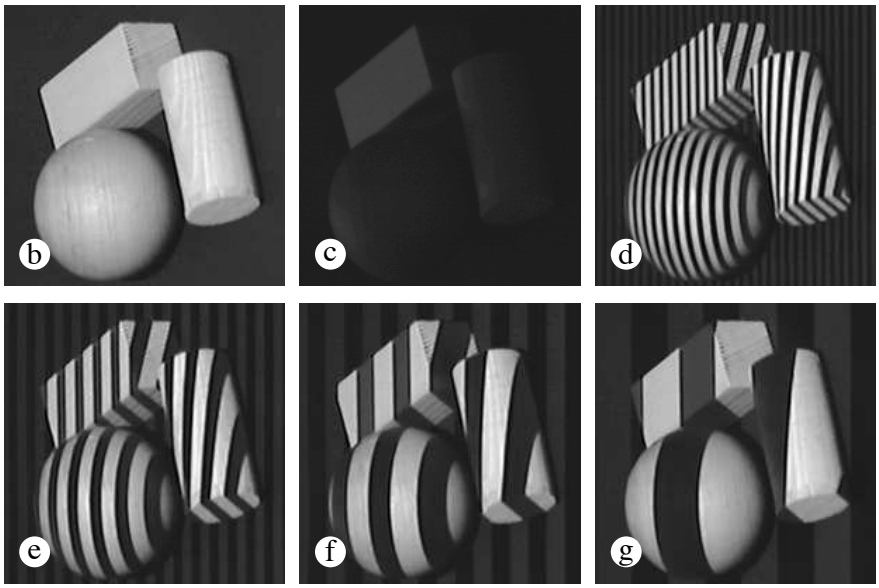


Figure 2.8: Binary coded structured light patterns [Tro95]

distinguish. In such cases binary encoding may be the only option to obtain robust results.

Second is the number of successive images that have to be taken for one complete measurement. In order to use multiple images camera and projector have to be synchronized, which requires expensive hardware to work at high frame-rates. If a static pattern is used, the frame-rate is only limited by the camera alone.

Third is the number of pixels used to encode one coordinate. Instead of using different color levels, the range of encoded values can as well be increased by spatially combining multiple neighboring pixels to encode a single value. This comes at the expense of reduced accuracy, but allows for fast and robust measurements.

While all these properties can be arbitrarily combined in one encoding strategy, most strategies use either temporal or spatial encoding but not both at once. A very common approach is to use multiple binary coded stripe patterns like shown in Figure 2.8. The different sized stripe patterns are slightly displaced so that the coordinate of a pixel is encoded by its unique sequence of either being illuminated or dark in the series of projected patterns. A more detailed investigation of different codification strategies can be found in [SPB04].

2.6 Time of Flight

In theory measuring distance by the time of flight principle is very simple: A light pulse is emitted on a surface and gets reflected back to the sensor. The time between emitting the pulse and receiving its reflection is measured and, since the speed of light is known, directly yields the distance covered by the light.

In practice however it is both difficult to emit a very brief and bright enough light pulse and to measure the exact time of arrival of its reflection on the sensor. Especially when aiming to get measurements of high accuracy at close distances, it would require expensive highly accurate equipment. The commonly used alternative is to illuminate the scene with a continuously modulated light source and determine the time of flight from the phase differences. Typically the intensity of the light source is modulated by a sinusoidal wave function, this approach is hence often called continuous wave intensity modulation (CWIM) in literature.

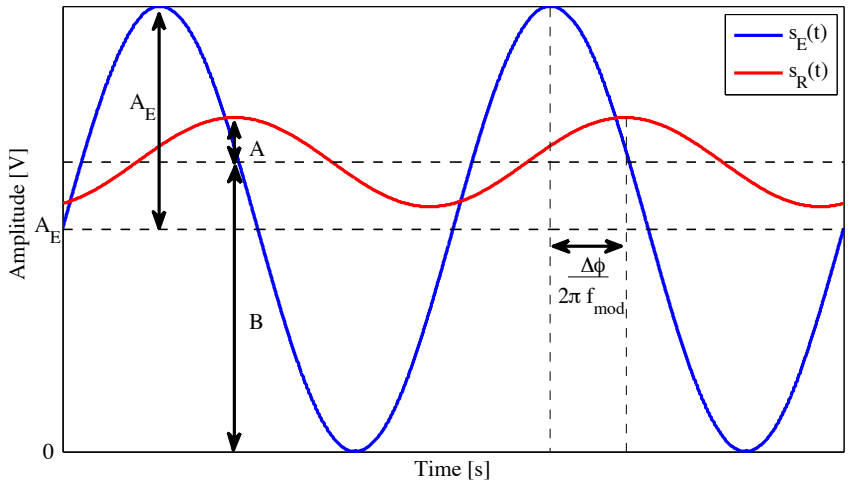


Figure 2.9: Continuously modulated signal emitted (blue) and received by the sensor (red) [DZC12]

Correlating the received light on the sensor with the emitted signal (s_R and s_E in Figure 2.9) and hence determining the phase difference has some practical problems. Since there is always some background illumination adding an offset (B) that is not known, one measurement does not solve the equation. Also the reflected (maximal) amplitude (A) can be any portion of the emitted amplitude (A_E) depending on the surface reflectivity. Therefore there are three unknowns, requiring at least three measurements. Usually there are four measurements taken at fixed points along the modulation period. Though in reality intensity cannot be measured at a single point of time, but only by

integrating arriving photons over a period of time, which is called integration time and is an important parameter for time of flight measurements.

Another important parameter is the modulation frequency, which directly determines the maximum range. Since at a certain distance the time it takes for emitted light to get back to the sensor is greater than the modulation period, light reflected beyond that distance cannot be distinguished from light emitted one period later.

2.7 Light field

A light field describes the flow of light through space and thus contains information for both direction and intensity of individual light rays. While a conventional camera only captures the intensity of the incident light, a plenoptic camera as introduced by Adelson and Wang [AW92] can also capture its direction. Through an additional array of microlenses an incoming light ray gets refracted differently depending on its direction and hence hits a different portion of the imaging sensor. Given that each microlense covers multiple sensor pixels, the direction of the light ray, and hence the distance of the point it was reflected from can be determined more or less accurately.

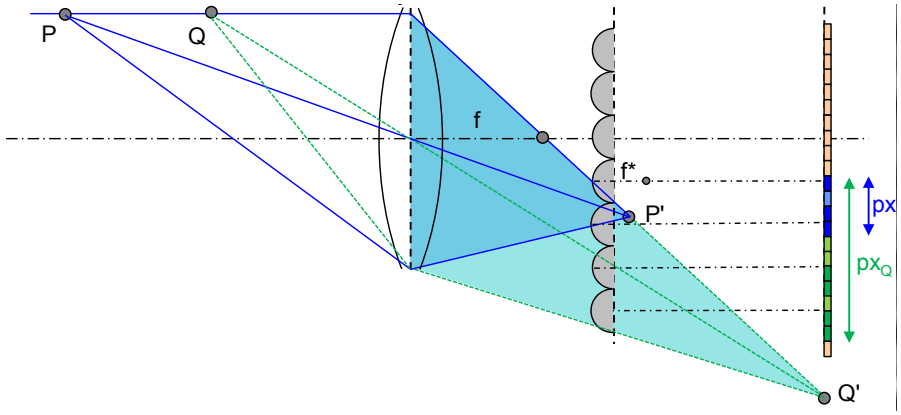
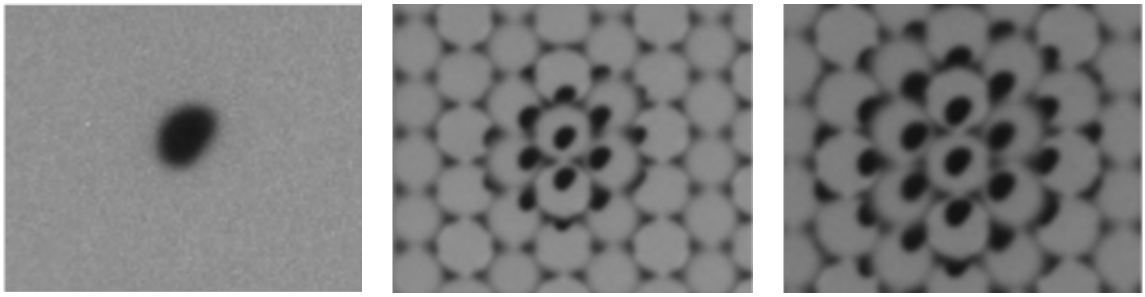


Figure 2.10: Functional principle of a plenoptic camera [LJH14]

Since this basic design limits the effective image resolution to the size of the microlense array and also requires much higher sensor resolution, a different design was proposed by Lumsdaine and Georgiev [LG09]. By focusing the microlense array differently, a higher (lateral) image resolution can be achieved with the same number of microlenses at the cost of reduced directional resolution. Figure 2.10 shows the general configuration of a plenoptic camera with two points at different distances given as an example. Light from closer points is directed to more microlenses and therefore spread out over a larger

portion of the sensor, as can also be seen in Figure 2.11 showing the images of a point seen through a microlense array from two different distances.



(a) **(b)** **(c)**

Figure 2.11: A point (a) seen through a microlense array from far (b) and closer (c) distances [LJH14]

As with Stereo Vision, it has first to be determined which points on the sensor image correspond with each other, in order to calculate a depth value. Due to the much higher sensor resolution and added complexity, the necessary correspondence algorithms are however even more demanding. This is one reason why commercial plenoptic cameras only showed up recently. Also like Stereo Vision, a depth value can only be determined, if there are sufficient visible differences in the image. Untextured surfaces can therefore not be measured, unless an active pattern projection is used.

Since only a single camera is needed the light field technology allows for much more compact design and can easily combined with additional optical systems. Given the potential for usage in Computational Photography and other areas, it can be expected that this technology will receive much more attention in the near future. Since the essential principle of determining depth is very similar to Stereo Vision, many aspects regarding limitations can however be assumed to be comparable.

2.8 Combination of multiple approaches (Fusion)

A combination of multiple measurements into a single result, also called *Fusion*, can help to improve accuracy, robustness or measuring range.

The term "*Fusion*" is used to describe either the combination of multiple samples of a single camera system from different perspectives or the combination of measurements from multiple camera systems using different measurement methods. While the former is an important issue especially for mobile setups, its application is limited for static setups, where measurements from multiple perspectives can only be achieved by adding

additional camera systems at fixed positions. In such cases a combination of different measurement approaches is therefore much more promising.

Combining passive stereo with an active method operating at infrared light is an obvious candidate. Passive stereo usually can provide higher resolution while it does not interfere with the active measurement, which can be used to overcome its shortcomings. *Fusion* of passive Stereo and Time-of-flight measurement has first been proposed in 2006 [KS06] and is since then thoroughly investigated [NRL+13] As the ToF-data can be used to reduce the search space and eliminate ambiguities in stereo-matching, the *Fusion*-approach asks for a whole new class of correspondence algorithms, which allow for more accurate results at less execution time. [EHH15]

A combination of to active methods operating within the same frequency range is rather difficult due to interference, it is however possible to use structured light patterns simultaneously for active stereo vision. [JJSL13]

Fusion also allows for including measurement methods, which are unsuitable to determine depth when used alone, but can help reconstructing surface details. An enhancement of ToF-data by measuring the polarization of the light reflected from surface, has been recently demonstrated. [KTS+15] Assuming that in most environments the ambient light is not polarized in any specific way, the polarization can be captured passively by taking multiple photos with a standard camera using a polarization filter. Projectors used for providing in-situ feedback however usually emit polarized light which may cause interference or even render this approach inapplicable.

In order to deal with the obvious inability of the discussed methods to measure transparent surfaces, adding another type of sensor might even be the only solution in such cases. The inclusion of ultrasonic sensors can help to overcame this limitation. [YZYM15] Due to the lack of higher resolution arrays of ultrasonic sensors, this approach is however of limited use in static setups.

3 Related work

Due to the long history of 3D vision and the recent boom through the availability of cheap depth-sensors[1], there is an enormous amount of work investigating the usage of depth-sensors in all kinds of different scenarios. (From robotics [SLAL11] over archaeology [MGH09] to Plant Phenotyping [PBM+14]) There are however only relatively few works aimed at evaluating depth-sensors in a more systematic manner and even less that include a side-by-side comparison of different measurement methods. This chapter investigates a range of evaluations aiming at assessing the capabilities and limitations of depth-camera-systems.

Most of the works evaluating depth sensors try to obtain general values for effective accuracy and resolution of a sensor. This is usually either done by measuring a simple target with known geometry like a planar surface from a fixed distance [BBK07] or by comparing the sensors measurements of a scene with one or more arbitrary objects to a very accurate measurement (ground truth) of that same scene, usually obtained by using a laser scanner. [GRV+13] Due to its high availability there are several evaluations of the first Microsoft Kinect sensor ($\text{Kinect}_{\text{SL}}$), which uses a Structured Light (SL) approach and its successor ($\text{Kinect}_{\text{ToF}}$), which employs the Time of Flight (ToF) principle. Both sensors are also described more detailed in chapter 4. Other ToF-Sensors often used in evaluations are the SwissRanger SR4000 from Mesa Imaging[2] and the 3k-S or CamCube from PMDtec[3].

Since accuracy and precision of depth measurements are highly dependent on the measured scene as well as the environment, a simple quantitative comparison can however only be given in respect to a specified scene and environment or regarding "ideal" measurement conditions. In order to get a more complete picture of the limitations, the effects of different sources for errors should be investigated as independently as possible. In the remainder of this chapter each of the relevant error sources is discussed along with the related work investigating their effects on the measurement. The categorization of error sources is partially adopted from [SLK15]

[1] The release of the Microsoft Kinect alone has resulted in over 3000 related publications just within 3 years [BMNK13]

[2] Recently acquired by Heptagon: www.hptg.com/industrial/

[3] www.pmdtec.com (Both mentioned products have been discontinued)

3.1 Systematic (non-environmental) errors

Even in the most simple cases with optimal conditions (high-reflective diffuse planar surface in a disturbance-free environment) there can be non-linear errors in the depth measurement, due to the imperfection of the sensor components. Distortion of optical lenses for example may cause varying depth offsets at different measured points. Errors of this kind which are persistent when the measurement is repeated are called systematic errors.

Many works make a distinction between systematic (predictable) and random (less predictable) errors. There are however broader [FA11] and narrower [SLK15] definitions of systematic error. Here a narrow definition is used by only regarding errors, which are independent of scene and environmental factors and therefore intrinsically systematic.

For triangulation-based methods like Structured Light (SL) and Stereo Vision (Stereo) systematic errors are mainly due to limited pixel resolution and non-linear response of the camera sensors as well as optical distortions. While pixel resolution is an inevitable limitation, the other two can be compensated by photometric calibration used commonly in Computer Vision [Sze10].

Systematic errors can be modeled as a function of the hardware specifications (see [BH87] for Stereo and [Tro95] for SL) in order to optimize sensor design and calibration methods.

Sensors using the Time of Flight (ToF) principle exhibit additional systematic errors due to the more complex measurement procedure necessary when using continuous wave intensity modulation (CWIM) [LNL+13]. Since the light source cannot generate the intended modulation function precisely, a systematic offset changing over distance ("wiggling") can be observed. Foix and Aleny also identified systematic errors caused by different integration times and irregularities in the image sensor [FA11].

3.2 Temperature Drift

Due to inevitable deformations of hardware components by temperature change measurements are also drifting with temperature change. Temperature drift is still relevant in environments with constant temperature, since the light sources in active sensors emit significant amounts of heat as well and therefore heat up the sensor components

when turned on.[4] Most sensors reach a stable state after some period (warm-up time). The $\text{Kinect}_{\text{SL}}$ has been shown to stablize after 60 minutes [CALT12]. As Sarbolandi et al. point out, ToF sensors generally need higher illumination-power to cover same area and range, which results in greater heat dissipation and therefore requires better cooling. [SLK15] Which is why the $\text{Kinect}_{\text{ToF}}$ (having a compact housing) uses active cooling which helps to achieve stable measurements after 60 minutes as well.

3.3 External Light Sources

Ambient background light can cause interference by over-saturating the sensor or outshining the signal or pattern projected by the sensor.

Indoor illumination does usually only emit light in the visible spectrum and therefore causes little or no interference for sensors operating in the invisible NIR-range as also reported by [CALT12]

Daylight or other strong sources of IR-light however are problematic. ToF sensors have been shown to be still operating with strong ambient IR-light, while comercial SL-Sensors like the $\text{Kinect}_{\text{SL}}$ already show errors with considerably lower amounts of ambient light [LHL12] and cease to work where ToF-sensors still yields measurements [SLK15] (Though drastic errors appear at higher levels of ambient light)

Another type of interference may caused by light sources emitting pulsating light projecting patterns onto the measured scene. This especially an issue when using multiple active sensors in parallel. Since these cases not only add additional ambient light put may also distort the projected signal, inteference can be much more intense.

The parallel usage of multiple $\text{Kinect}_{\text{SL}}$ or $\text{Kinect}_{\text{ToF}}$ has been shown to be rather error-prone [SLK15] There are however methods to use multiple ToF sensors together, but they require a special setup and deeper access to the sensor hardware. [HLC+12] In contrast AS and SL can be easily used together when only the SL pattern is projected. [JJSL13]

[4] The drift caused by self-heating could also be considered as a systematic error. Temperature is however an independent environmental factor in any case.

3.4 Depth Inhomogeneity

Since a single pixel of sensor data does not correspond to a single point on a surface but a small area, its associated depth values can be inhomogenous at object edges (depth discontinuities). This may lead to noisy edges or even completey different depth values.

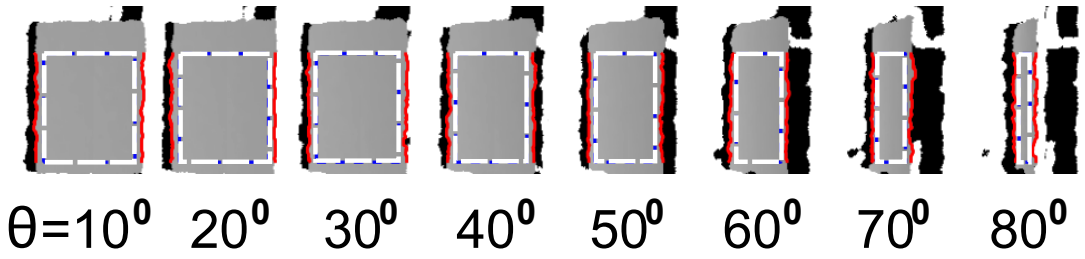


Figure 3.1: lateral noise at edges (in red) of a rectangular shape measured from different angles with the $\text{Kinect}_{\text{SL}}$ [NIL12]

Nguyen et al. measured axial and lateral noise on edges for the $\text{Kinect}_{\text{SL}}$ from different angles. [NIL12] They used the data to derive a noise model that can be used to improve *Fusion*-algorithms combining multiple measurements from different perspectives.

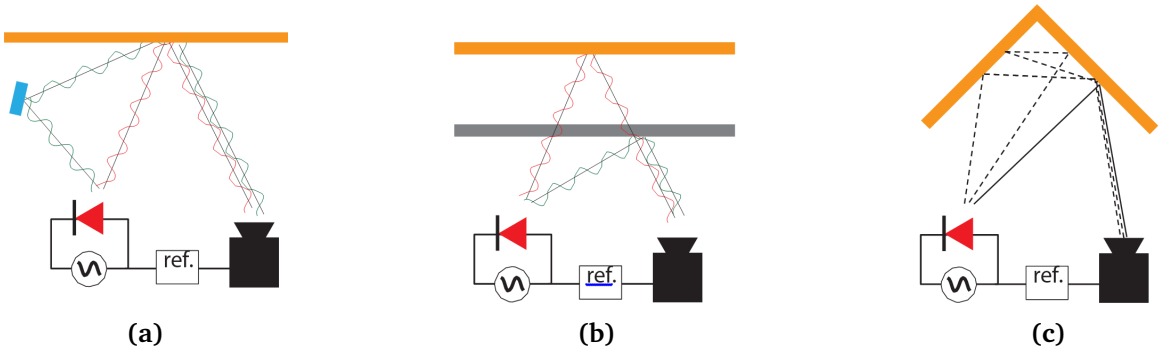
For ToF-sensors errors caused by depth inhomogeneity can take any value within the sensor's measuring range leading to enormous outliers [LNL+13] and are therefore also called "flying pixels". Reynolds et al. investigated these errors in more detail and introduced a method for filtering them. [RDP+11]

A comparison of the $\text{Kinect}_{\text{SL}}$ and $\text{Kinect}_{\text{ToF}}$ however shows that the amount of errors related to depth inhomogenity is similar for both methods. [SLK15]

3.5 Multipath interference

The ToF principle assumes that light emitted from the sensor is reflected back exclusively on the most direct way. The emitted light can however also be reflected from another surface on the scene or in the environment before it gets reflected off the surface point to be measured and therefore taking multiple paths as visualized in Figure 3.2.

Multipath interference is major issue for ToF-sensors and is extensively researched. [God12] For SL multipath reflections are only a problem, if parts of the coded light pattern get directly reflected on other surfaces. Accordingly the $\text{Kinect}_{\text{SL}}$ shows very



(a) (b) (c)

Figure 3.2: Multipath interference (a) from external reflection (b) from semitransparent surface (c) from internal reflections [BFI+14]

little errors in a multi-path test, where the $\text{Kinect}_{\text{ToF}}$ has problems getting any valid measurements [SLK15].

3.6 Dynamic Scenery

If an object moves during measurement it might cause motion blur. How much a measurement is affected depends on the exposure time respectively the duration of the whole measurement cycle for ToF. The effects of motion blur for depth images do however not just result in blurred depth images, as with normal color images, but instead may cause significant distortions. The effect also depends on the used method. For ToF motion blur can result in depth values much higher or lower than the actual values, but may also be detected and marked as invalid. This is investigated in more detail in [HLC+12].

For the $\text{Kinect}_{\text{SL}}$ it has been found that motion blur causes a clear bias towards closer values, while the $\text{Kinect}_{\text{ToF}}$ delivers more reliable values and marks invalid measurements. [SLK15]

3.7 Surface Properties

Due to their great variety the effects of surface properties on depth measurements are difficult to evaluate. The most complete survey of different surface properties has been done by Hansard et al. by measuring arbitrary objects with various surface properties with the $\text{Kinect}_{\text{SL}}$ and the SR4000 ToF-Sensor. [HLC+12] They coarsely classified the objects based on surface roughness and specularity and presented the root mean square

error (RMSE) for each complete measurement, resulting in much higher RMSE-values for specular surfaces due to many invalid pixels from specular highlights.

Additionally they measured semitransparent surfaces for which they took a more systematic sample set by applying a matte spray to a translucent cylinder in order to obtain a whole range of different degrees of translucency. A more refined version of this experiment has been done by Sarbolandi et al. showing similar results. [SLK15]

Another important property is surface albedo, though it has not been in the focus of any of the discussed evaluations. It has been included in an evaluation of the $\text{Kinect}_{\text{SL}}$ by Chow et al., but they only compared measurements of two (unspecified) black and white surfaces and found no significant differences. [CALT12] A later evaluation of the $\text{Kinect}_{\text{ToF}}$ shows significant differences when comparing measurements of a white and black surfaces with 99% and 5% albedo. [SLC+15]

3.8 Incident Angle

Last but not least an object's geometry influences measurement accuracy as well. According to Hansard et al. however "accuracy is not so much dependent on the geometric complexity", but rather on the deviation of the surface normal to the optical axis of the sensor [HLC+12] unless the geometric structures are smaller than the sensor resolution or having concave shapes causing multipath interferences.

Effects of measuring a surface from different angles have been investigated in many evaluations, usually by rotating a white planar target in front of the sensor [CALT12] or moving the whole camera system [LHL12]. The $\text{Kinect}_{\text{SL}}$ has been shown to obtain measurements at angles of 75° and below while the $\text{Kinect}_{\text{ToF}}$ also yields depth values at over 80° but only with larger errors.

It remains however an open question how other error sources and especially different material properties effect the performance of measurements from different angles.

4 Current depth-camera-systems

While the discussed Active Stereo (AS), Structured Light (SL) and Time of Flight (ToF) principles have been used for depth measurement for a few decades already [Bla04], affordable mass-produced depth-camera-systems have only shown up recently with the release of the first Kinect (using SL) by Microsoft in 2010. Since then several consumer-grade systems appeared, either using ToF or a SL-approach similar to the Kinect. With the second Kinect Microsoft however pushed the market again in 2013 by providing a ToF-system with much higher resolution than other ToF-systems at that time.

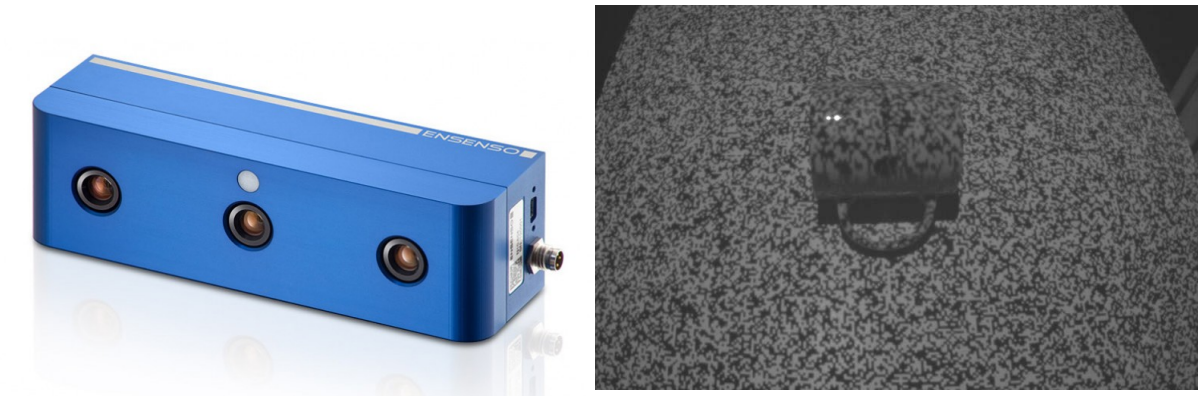
In this thesis three different depth-camera-systems are used for experimental evaluation of each measurement principle: The AS-based Ensenso N10 from IDS Imaging as well as the first ($\text{Kinect}_{\text{SL}}$) and the second Kinect ($\text{Kinect}_{\text{ToF}}$) from Microsoft. This chapter describes the three systems in more detail and puts them into context with other available systems on the market.

Since there are neither active systems using light field technology, nor suitable fully integrated Fusion systems available as products, only systems using exclusively either AS, SL or ToF are considered.

4.1 Active Stereo

Passive stereo systems are readily available, given that they only consist of standard cameras and software. Active stereo systems however are rather rare and not to be found in consumer products. Their main advantage over other active systems, the simultaneous usage of multiple systems, is usually not a requirement or option for simpler and consumer-oriented applications.

In theory a passive system could be used for AS by simply adding a pattern projector, but making the system to operate exclusively in the invisible near-infrared-range (NIR) for unobtrusive measurements is more difficult. Also because the projected pattern has to fit the stereo cameras, a fully integrated solution is preferred.



(a) (b)

Figure 4.1: (a) Ensenso N10 Stereo camera from IDS Imaging [IDS] (b) Pattern projected on a flat surface with a cup

IDS Imaging provides such a solution with their Ensenso-series, which is meant for industrial usage in rough environments. The models differ in sensor resolution and focal length resulting in various effective accuracy.

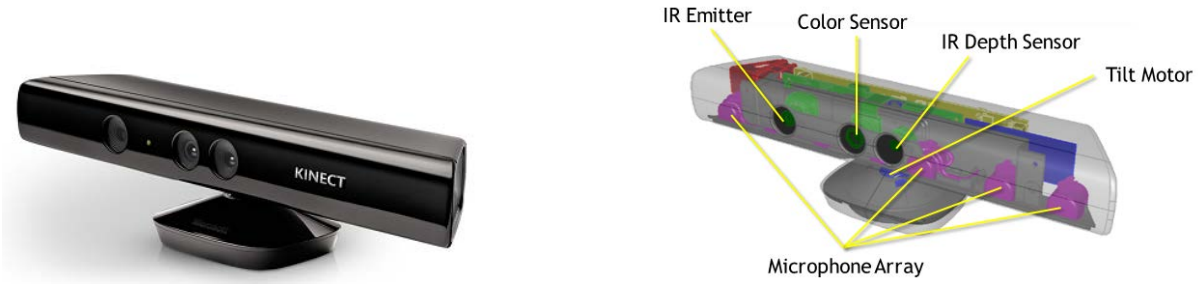
For this thesis the Ensenso N10-802-18 (Shown in Figure 4.1) was used, which has a specified operation range from 0.65m to 2m. The N10-Series uses stereo cameras with a resolution of 720 x 480 pixels. With a focal length of 8mm however the used model offers a rather small Field of View, providing similar effective accuracy as models with higher resolution and bigger Field of View. At a distance of 0.75m the pixel size is about half a millimeter, allowing for a theoretical depth-accuracy of one millimeter. While the cameras are able to deliver 30 frames per second, the effective performance for depth measurement also depends on the available processing power, as the used correspondence algorithm is computationally expensive.

All models feature a LED-based pattern projector, which is synchronized to the camera shutters and only illuminates the scene with a binary pattern as shown in Figure 4.1(b) during exposure time. (Therefore having a lower power consumption, generating less heat and possibly causing less interference)

The Usage of standard optical components theoretically allows for using imaging sensors with much higher resolution. Compact stereo cameras with high resolution are already available (For example the ZED from Stereolabs [Ste]) But stereo matching algorithms are computationally expensive and therefore limit framerate at higher resolutions or require specialized hardware solutions. This is especially an issue when multiple systems are used to capture depth images from different perspectives.

4.2 Structured Light

Structured Light technology is already used in professional products for a longer time, the Microsoft Kinect ($Kinect_{SL}$) developed in cooperation with PrimeSense however was the first affordable consumer-grade SL-system. It was originally intended to be a gaming interface for the Microsoft Xbox 360, but later also released independently along with a software development kit as “Kinect for Windows”. In addition to the IR camera and pattern projector used to measure depth, the system features a RGB camera to also provide a color picture as as shown in Figure 4.2. Such systems are therefore also called RGB-D cameras in literature.



(a) **(b)**

Figure 4.2: (a) The $Kinect_{SL}$ and (b) a schematics of its components [Mica]

The exact specifications and operation details are not made available by Microsoft and the firmware does not allow full access to the system. According to Koshelham et al. the resolution of both cameras is 1280x1024 pixels though only 640x480 pixels are streamed due to bandwith limitations. [KE12] The $Kinect_{SL}$ uses a static spatially coded pattern (shown in Figure 4.3), which allows to deliver depth images of 640x480 pixels at 30fps.

The operation range is specified from 0.8m to 4m distance, although the depth errors are well above 1cm for all but the closest distances. Additionally the $Kinect_{SL}$ uses quantization of the depth values with a step size of already 1.8mm at 0.8m distance [SJP11]. More details about the operation of the $Kinect_{SL}$ can also be found in [DZC12].

While more products using the same technique as the $Kinect_{SL}$ have since been released, advancements were made mainly in miniaturization for better usage in mobile applications. (See Structure Sensor from Occipital[1])

[1] www.structure.io

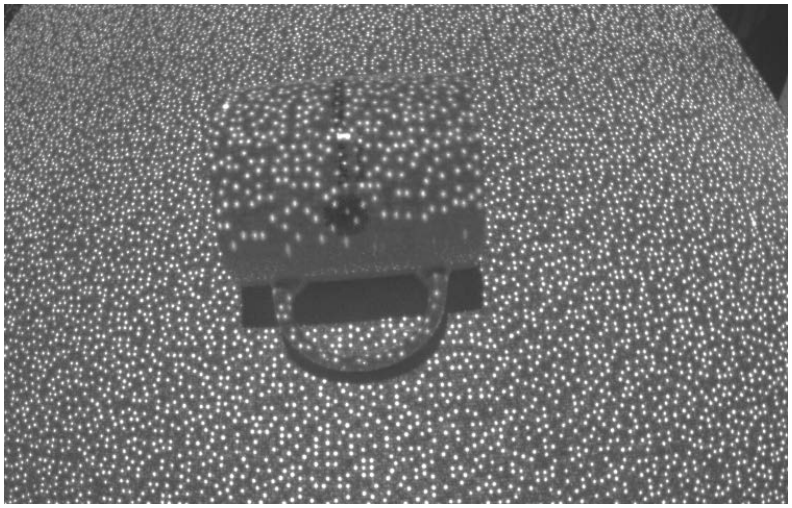


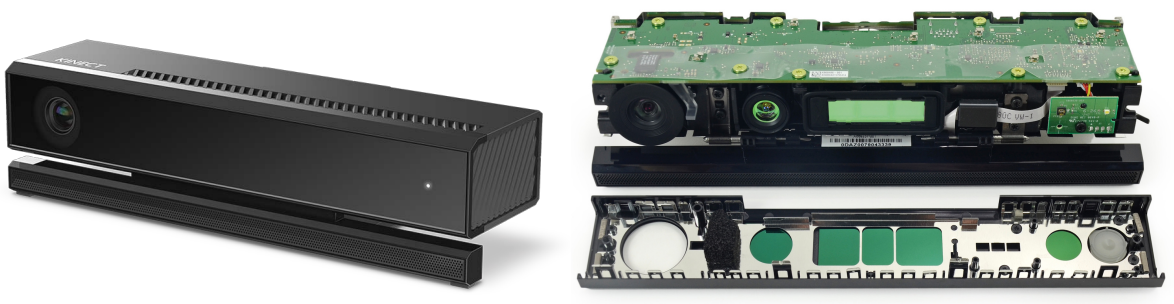
Figure 4.3: Pattern projected by the $\text{Kinect}_{\text{SL}}$ on a flat surface with a cup

The Structured Light method in general however can be easily scaled to allow for measuring small structures at high resolution. Usually though commercial solutions are not aimed to be used in interactive environments and therefore use visible light and temporal coded patterns, either requiring expensive high-speed components or resulting in longer measuring times.

4.3 Time of Flight

The Time of Flight principle is commonly used for optical distance measuring. Depth-camera-systems using ToF able to capture whole depth images in one shot are however a more recent development. Such systems cannot be build completely from standard components and their development is thus tied to the availability of the necessary specialized imaging sensors. Since the development of new imaging sensors is only economically viable when aimed for mass production, ToF-systems have long been restricted to lower resolutions. Due to the high production volume it was therefore possible to equip the second Kinect ($\text{Kinect}_{\text{ToF}}$) with a sensor of much higher resolution than other ToF-systems at that time.

Like previous ToF-Systems the $\text{Kinect}_{\text{ToF}}$ shown in Figure 4.4 uses continuous wave intensity modulation (CWIM), but is able to deliver a depth image with 512x424 pixels at 30 fps. And like its predecessor it also features a RGB camera with 1920x1080 pixels. According to Pagliari and Pinto [PP15] the $\text{Kinect}_{\text{ToF}}$ uses multiple modulation frequencies (120Mhz, 80Mhz and 16Mhz) in order to achieve high depth accuracy over a longer range, which is specified from 0.5m to 4.5m. In contrast to triangulation based



(a) (b)

Figure 4.4: (a) Kinect for Windows v2 Sensor ($\text{Kinect}_{\text{ToF}}$) [Micb] and (b) opened casing revealing the IR-camera and light source sitting next to each other in the center of the device [iFi]

methods, depth accuracy is not inherently tied to measurement distance for ToF-systems. While the $\text{Kinect}_{\text{ToF}}$ has been shown to have similar depth accuracy to the $\text{Kinect}_{\text{SL}}$ at close distance, it still keeps that level of accuracy even at higher distances where the $\text{Kinect}_{\text{SL}}$ shows errors of more than 1cm. [PP15]

ToF-systems are also better to miniaturize as no base line for triangulation is needed. The Light source and Image sensor could theoretically be in the same spot. This allows for very compact design and integration into mobile devices. Such small ToF-systems are for example available from PMD[2].

Meanwhile professional ToF-systems aimed for industrial use with similar capabilities as the $\text{Kinect}_{\text{ToF}}$ are also available from other manufactures like Basler.[3]

[2] www.pmdtec.com
[3] www.baslerweb.com/en/produkte/kameras/3d-kameras/time-of-flight-kamera

5 Experiments

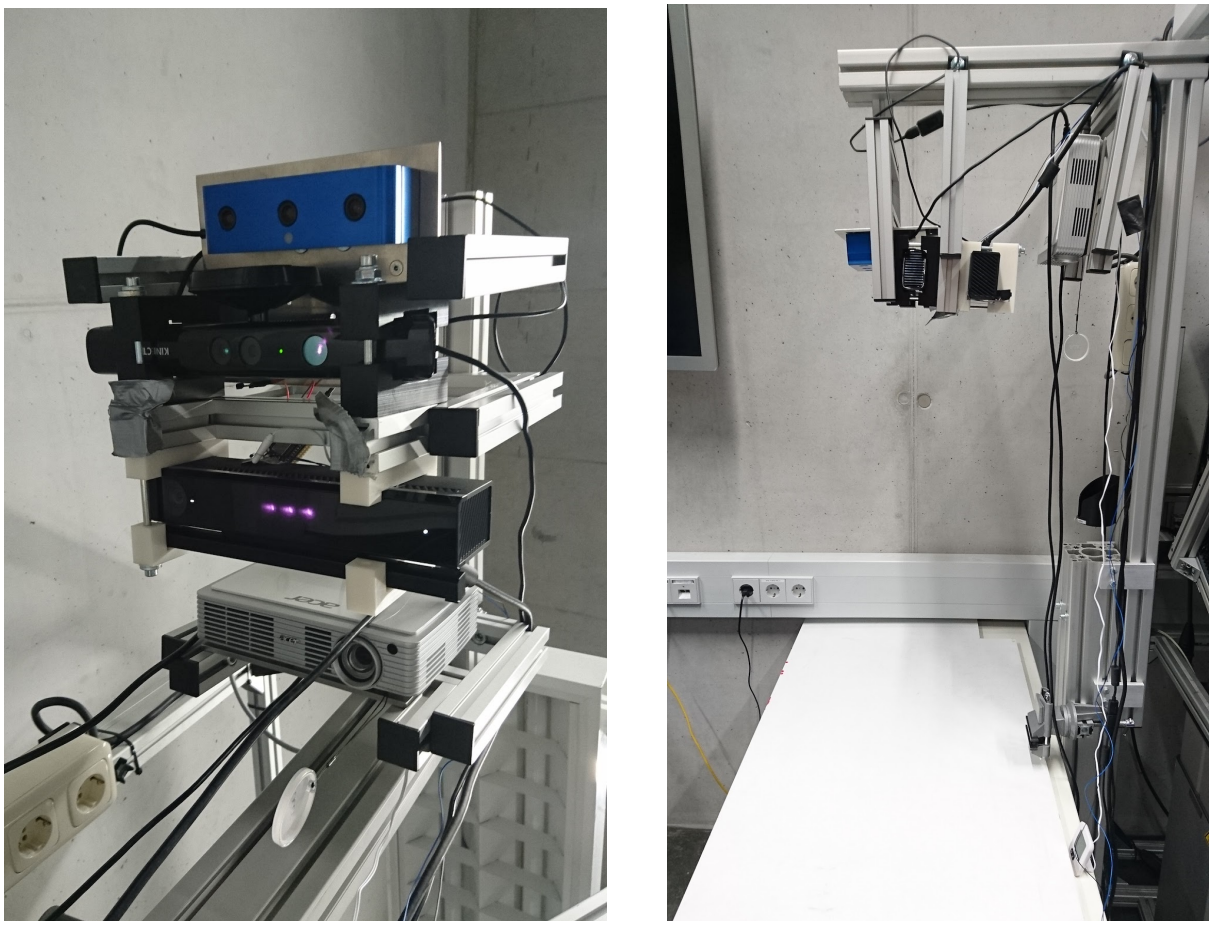
In order to assess the effects of different surface material on depth-measurements in practice, an experimental evaluation is needed. Due to the huge variety of surface properties, an extensive evaluation would require hundreds of different samples. To limit the scope of the evaluation only smooth surfaces with varying albedo and gloss were considered. This chapter describes the experiments, that have been conducted in the context of this thesis.

The experiments are not aimed at obtaining absolute accuracy values allowing for direct comparison, as many other evaluations do. Instead the focus is on revealing how different surface material relatively impacts measurements. The different systems can then be compared in respect to the magnitude of relative differences. Therefore it is not necessary to perform an intrinsic or extrinsic calibration of the systems. Furthermore comparing based on relative differences essentially rules out systematic errors including calibration errors.

5.1 General Setup

The experimental setup is based on the assistive system prototype from Funk et al. [FS15] in order to provide conditions similar to those in actual application. For comparison the Ensenso N10, the Microsoft $\text{Kinect}_{\text{SL}}$ and $\text{Kinect}_{\text{ToF}}$ (described in Chapter 4) have been mounted on a carrier at about 1.6m above the floor as shown in Figure 5.1. The setup has been placed in the lab next to other workspaces, like in a typical working environment.

All measurements were taken indoors with no daylight at constant room temperature of about 22°C. Also interferences by infrared light from any other source have been ruled out by checking the raw output of the IR-Cameras with the pattern projector turned off. Additionally, effects of illumination from the room lighting and other non-IR light sources on the depth-measurements have been tested and found to be essentially non-existent. Some light could be picked up when aiming a non-IR light source directly or via a mirroring surface at the IR-Cameras. Mirroring surfaces are however much more



(a) View from below **(b)** Side view

Figure 5.1: Experimental setup with three different depth-camera-systems mounted side by side

problematic for other reasons (see chapter 2) and having to measure parts containing active light sources seems unlikely or at least avoidable.

To capture depth-data all three sensors were connected to a Windows-PC using the official drivers and SDKs provided by the manufactures. (Ensenso SDK 1.3.167[1], Kinect for Windows SDK v1.8[2] and Kinect for Windows SDK 2.0[3]) To facilitate recording depth-data and quickly switching between the systems, a software tool was written in C#.[4]

[1] www.ensenso.com/support/sdk-download/
[2] www.microsoft.com/en-us/download/details.aspx?id=40278
[3] developer.microsoft.com/en-us/windows/kinect/tools
[4] Since the two different versions of the Kinect for Windows SDK (1.8 for the $\text{Kinect}_{\text{SL}}$ and 2.0 for the $\text{Kinect}_{\text{ToF}}$) cannot be used in a single application, the depth-data recording tool has been conceived

Raw depth images, as provided by the SDKs, were saved directly as a bit-stream with 16bits per pixel. The depth-data was then evaluated without any further pre-processing using MATLAB.

To exclude differences caused by sensor noise, all depth-values are obtained by temporal averaging over 20 consecutive frames. Comparing averaging of up to 200 frames showed, that averaging more than 20 frames has only little effect on measurements. But also are 20 frames a sensible limit to stay well within one second of response time, if temporal filtering is to be applied in assistive systems.

5.2 Warm-up Test

Even at constant room temperature all of the systems are affected by temperature drift (see section 3.2) due to internal heat-up caused by the active components. To rule out temperature drift as a factor for differences in measurements it has to be known how much the measurements fluctuate during the warm-up phase and after what time the system becomes sufficiently stabilized.

5.2.1 Setup

To asses temperature drift during the warm-up phase a planar surface on a table at a distance of about 0.85m (as shown in Figure 5.1(b)) was measured for two hours in intervals of 10 seconds. All the systems were already connected and powered long before the measurement started. This test therefore specifically assesses drift caused by warm-up of the active light emitting components, which are only turned on during measurements.

5.2.2 Results

For each point of time the depth values were averaged both temporally (20 frames) and spatially. For comparison the average value during the second hour was taken as a stabilized reference value. Figure 5.2 shows deviation from this average over time. While the Ensenso N10 is only slightly affected by temperature drift and the $\text{Kinect}_{\text{SL}}$ stabilizes quickly, the $\text{Kinect}_{\text{ToF}}$ exhibits a significant drift before the value stabilizes.

as a set of client-server applications with a central control (server) for the individual clients, each controlling one of the systems.

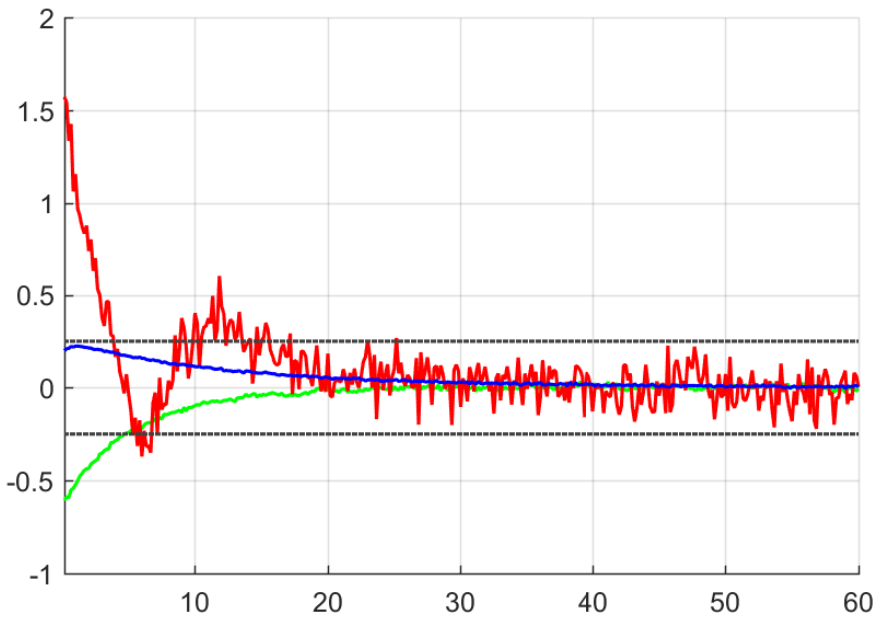


Figure 5.2: Deviation of average depth value [mm] during warm-up over 60 minutes Blue: Ensenso N10, Green: $\text{Kinect}_{\text{SL}}$, Red: $\text{Kinect}_{\text{ToF}}$

After less than 10 minutes the value for the $\text{Kinect}_{\text{SL}}$ changes only by less than 0.25mm. Due to active cooling the $\text{Kinect}_{\text{ToF}}$ exhibits a non-monotonic curve. The cooling shows effect after 6 minutes and allows the depth value to stay within a range of 0.5mm after 30 minutes.

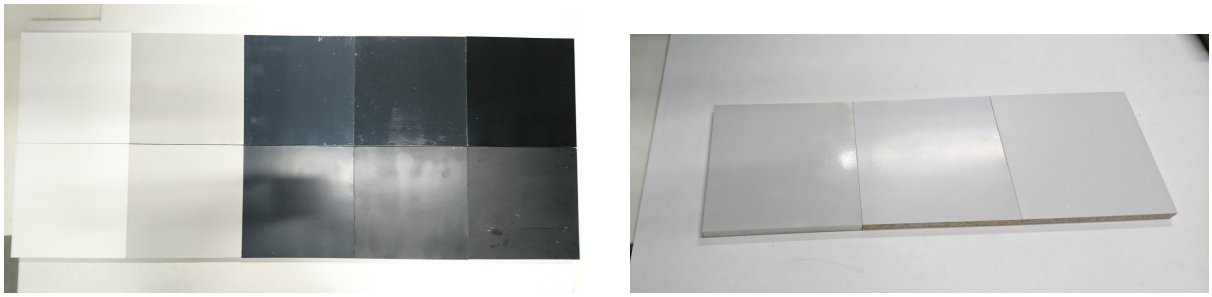
Consequently for the following experiments the depth-camera-systems were turned on at least one hour prior to taking measurements.

5.3 Samples

Given the experimental setup it was very important to have samples of same shape and size. For this purpose different spray paints were applied to 20 x 20cm wooden boards with 1.5cm thickness. (Normal ink printed on paper has shown not to be opaque enough to completely cover the material beneath.) Several gray colors were taken from the industrial RAL color standard. Since impact on measurements was expected to be most noticeable with low-reflective shades, multiple darker colors were chosen along with high reflective white and lighter gray for reference. (See Table 5.1) Additionally, some of the colors were applied in different degrees of gloss. The glossiness is specified by the manufacturer as 80 gloss units for the glossy paint, 30-35 gloss units for the satin

Number	CIELAB Lightness	Color name	Used gloss levels
RAL 9016	~95	Traffic white	glossy, matte
RAL 7035	~81	Light grey	glossy, satin matte, matte
RAL 7043	~40	Traffic grey B	glossy
RAL 7016	~34	Anthracite grey	glossy, matte
RAL 7021	~31	Black grey	glossy, matte
RAL 9005	~25	Jet black	glossy, satin matte, matte

Table 5.1: Colors from the RAL color standard used for creating the samples



(a) **(b)**

Figure 5.3: Some of the samples used: (a) matte and glossy samples sorted by lightness (b) light grey samples in different degrees of gloss: glossy, satin matte and matte

matte paint and 5-10 gloss units for the matte paint. (measured at 60° according to DIN 67530) [MOT]

One unpainted wooden board was also included into the measurement for comparison. The unpainted surface is white matte and should therefore show similar measurements as the sample with the white matte spray paint applied.

Each color can be correlated with a certain lightness value in the CIELAB color space. This specification is however of limited use since these values are given in respect to human color perception and have no validity regarding reflectivity in the invisble NIR-range. But at least for grayscale colors the lightness value seems to be roughly transferable, as can be confirmed by comparing the brightness of the samples in a image taken by one of the IR-Cameras. Though the colors are standardized, commercially available paint has been shown to deviate from the standard [HM09]. Unfortunately it was not possible to have the samples measured directly within the given time. Nevertheless the sample set can still provide evidence for wider ranges of darker material, even if the bounds of these ranges are not known precisely.

5.4 Distance test

It is evident that surfaces with low albedo and higher portions of specular reflections pose a problem for reflective depth measurement, but to what extend does this restrict the measurement of parts with such properties? Since this can be expected to be also depending on the measurement conditions, a first step is to investigate the restrictions in best possible conditions in respect to possible setups in assembly environments.

Fully optimal measuring conditions would for example also include taking the measurements in an isolated environment completely free of any other reflective surfaces. Getting anywhere close to such conditions is not only completely impractical in any working environment, but would also violate regulations in many countries. While it has been made sure that there are no larger highly reflective or mirroring surfaces close by, the measurement environment has knowingly be chosen to be less optimal in that respect. Exploring these environmental factors is subject of further research.

In this case optimal conditions are assumed to be given by excluding external IR-light-sources, measuring only a flat surface from a straight angle avoiding highlights caused by specular reflections and reducing the measuring distance to a minimum. But even though the irradiance on the measured surface is greater at shorter distances, the closest possible distance might not be the most optimal depending on the system. Therefore multiple distances should be tested.

5.4.1 Setup

The complete set of samples was measured from three different distances at about 0.85m (close), 1.1m (mid) and 1.5m (far). Closer distances were not considered, since the minimum measurement distance of the $\text{Kinect}_{\text{SL}}$ is at 0.8m. But also because any considerable closer distance would restrict the available workspace and therefore be impractical in many cases. In order to avoid specular highlights, the samples were placed slightly off the optical axis of the systems. For each distance all samples were measured individually by successively placing them at the exact same location. This guarantees exact same environmental conditions for every sample and also allows for a direct comparison of the depth-values.

5.4.2 Results

For each sample the depth values were averaged both temporally (20 frames) and spatially (partial area of the sample surface). The difference of this averaged value to

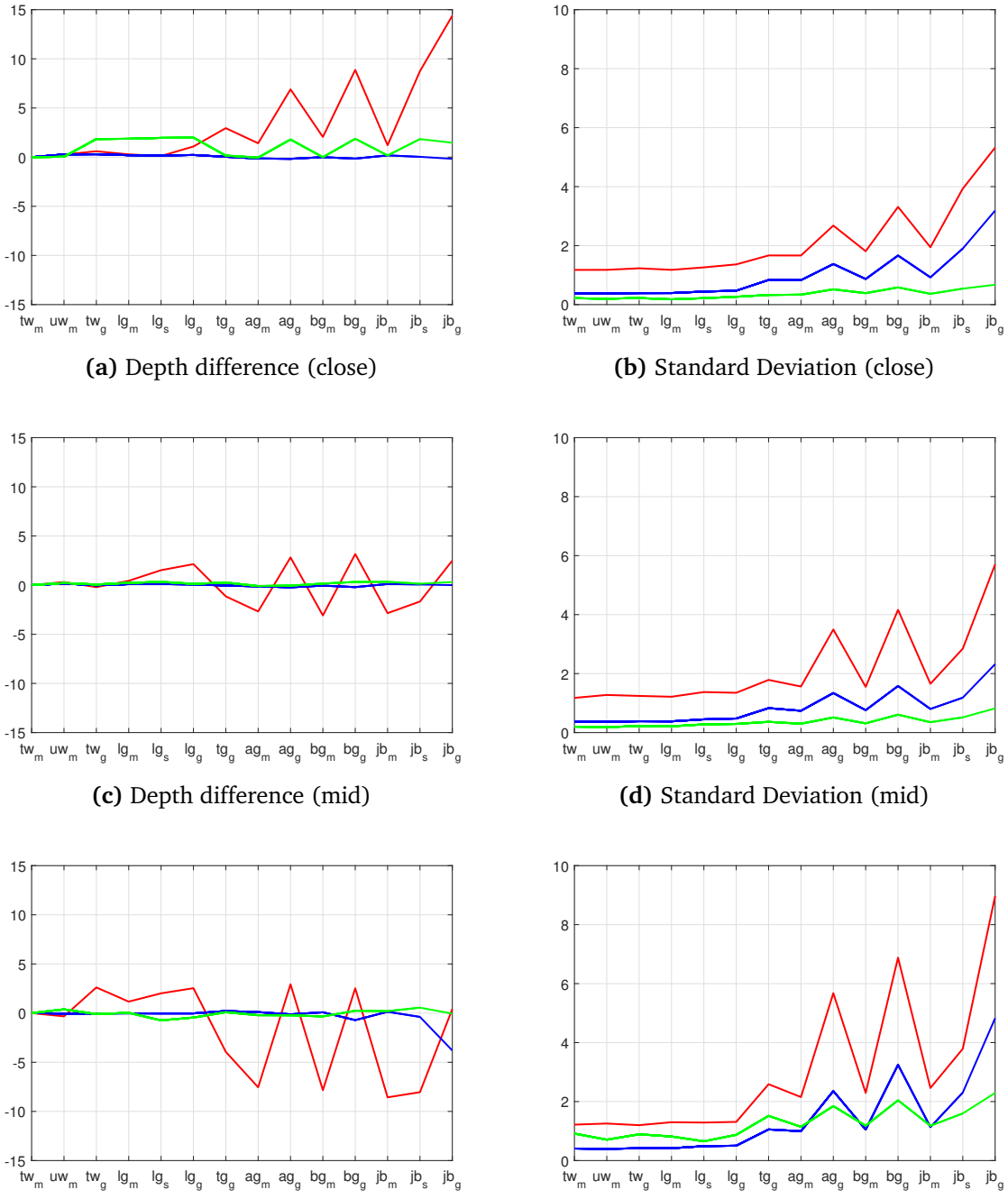
that of the reference measurement (white matte sample) can be seen as the relative offset or additional error caused by the difference in surface reflectance. While larger offsets clearly indicate a problem, smaller offsets are not necessarily without problems, as they could be hidden by the averaging.

Another important factor is the amount of noise. Temporal averaging can deal with lower levels of noise, but it becomes ineffective when noise levels are too high. As a second measure the standard deviation for each individual pixel of the depth image over 20 frames was determined. The average standard deviation can be seen as an overall measurement of noise in a measurement.

Figure 5.4 shows the average depth differences to the reference measurement and the average standard deviations for each sample by system. While the Ensenso N10 and the $\text{Kinect}_{\text{SL}}$ show little differences in the average depth value, the $\text{Kinect}_{\text{ToF}}$ exhibits considerable offsets for all the darker samples. Measuring close to the minimum depth range seems to be a problem for both the $\text{Kinect}_{\text{SL}}$ and $\text{Kinect}_{\text{ToF}}$ as the differences are higher at close distance than at mid distance. Surprisingly the $\text{Kinect}_{\text{ToF}}$ already shows clear differences for the bright glossy samples.

A closer look at the depth difference for each individual pixel as shown in Figure 5.6(a) also reveals a much bigger local discrepancy for the glossy samples. The regions further away from the optical axis of the system are showing bigger differences than the closer regions, which are receiving more light in a more direct angle from the IR-emitter of the $\text{Kinect}_{\text{ToF}}$.

Though the Ensenso N10 does show little differences for all samples at close and mid distance, a look on the standard deviations clearly shows a strong effect for the dark samples. The increase in noise at far distance is quite drastic and already seems to mark a maximum distance at which darker surfaces can still be measured. Though the $\text{Kinect}_{\text{SL}}$ seems to perform best, its low standard deviation is not the result of actual low noise levels at the sensor, but rather due to quantization of the depth values. The quantization step size is about 3mm at 1m distance [SJP11]. Therefore only differences of at least 3mm can be considered. This is visible in Figure 5.5. Larger regions with a single value are the result of this quantization, which is not visible in the difference images of the other systems. It also explains why the $\text{Kinect}_{\text{SL}}$ shows so similar offsets of about 2mm at close distance (Figure 5.4(a)), since 2mm is the quantization step size at that distance.



(a) Depth difference (close)

(b) Standard Deviation (close)

(c) Depth difference (mid)

(d) Standard Deviation (mid)

(e) Depth difference (far)

(f) Standard Deviation (far)

Figure 5.4: Difference to averaged depth value [mm] and average standard deviation [mm] of pixels between frames measured from different distances. Blue: Ensenso N10, Green: $\text{Kinect}_{\text{SL}}$, Red: $\text{Kinect}_{\text{ToF}}$

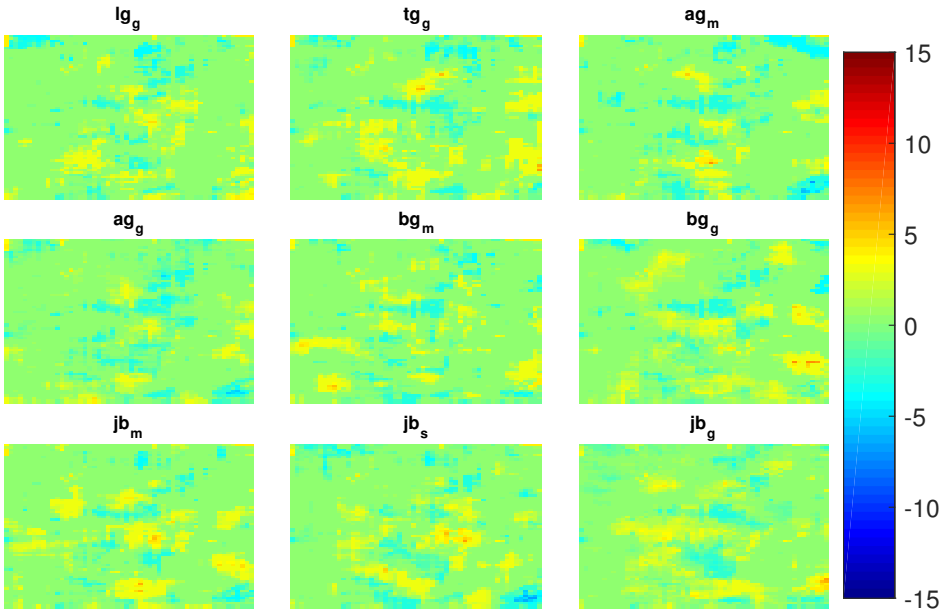
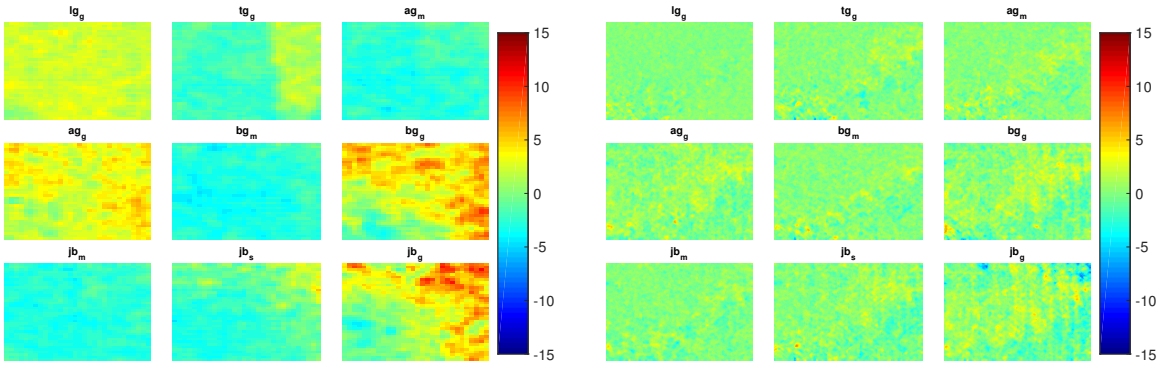


Figure 5.5: Difference images for $\text{Kinect}_{\text{SL}}$ at mid distance showing effects of quantization.



(a) **(b)**

Figure 5.6: Difference images for $\text{Kinect}_{\text{ToF}}$ (a) and Ensenso N10 (b) at mid distance (Bottom left is closer to depth image center)

5.5 Angle test

One of the main problems when measuring objects with arbitrary geometry is the fact, that surfaces have to be measured from sharper angles. This is a challenge for reflective depth measurement as projected patterns get distorted and the light received per area decreases. For matte surfaces with high albedo this is usually only a problem at very sharp angles higher than 70°. Given the results of the previous test, this can be expected to be a very different case for darker and glossier surface material.

5.5.1 Setup

In order to conduct the angle test in the same manner as the distance test, a contraption was used to hold the samples in place at different angles. First the angle was adjusted then each sample was placed on the holder at the exact same position for measurement.

(a) **(b)**

Figure 5.7: Angle test setup. (a) contraption to hold the samples in place at different angles. (b) Side view of test setup

The holder was placed in 1m distance to the camera-systems and again slightly off the optical axis to avoid specular highlights. For this reason the actual measurement angle differs from the stated angle, which was determined by using a spirit level. Also since not a single point but an area is measured, there are actually multiple measurement angles involved at the same time. Generally the results should be considered as concerning an angle slightly less sharp than the stated angle.

Measurements were taken at 0°, 15°, 30°, 45°, 60°, 75°, 80° and 85°. The actual surface area considered for evaluation was kept roughly the same at all angles, resulting in less depth pixels for the sharper angles. This mainly has an effect on measurements with many invalid pixels where a bigger evaluation area may still yield some more depth values. The alternative, using the same section of the depth image at all angles, has been tested to show little difference in the results otherwise.

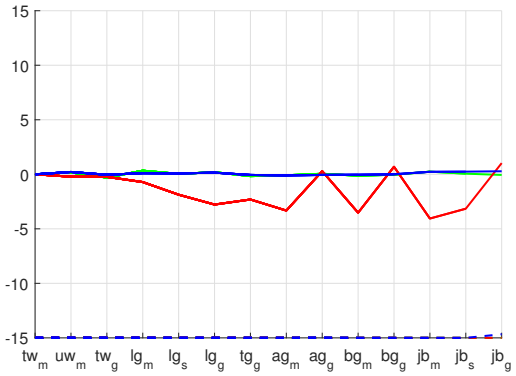
5.5.2 Results

As for the distance test the depth values were averaged and compared to the reference measurement of the white matte sample. (As described in Section 5.4.2) Likewise the standard deviation was determined.

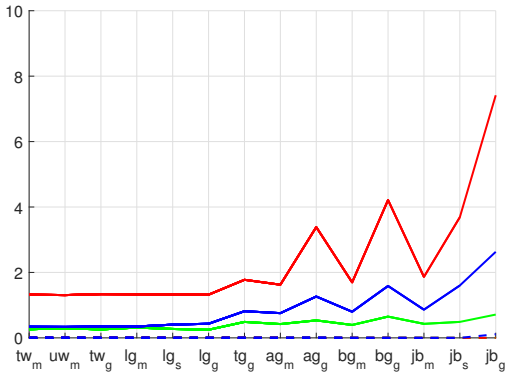
The resulting values in both measures are shown in Figures 5.8, 5.9 and 5.10. Since especially at sharper angles it was not always possible to obtain a depth value for each pixel, the ratio of invalid pixels at each measurement is shown in dashed lines. The results at 0° were included for reference, and can also be seen as another result for the distance test. (Showing how much the offsets of the $\text{Kinect}_{\text{ToF}}$ can already change with a slight change in distance/position)

The axis limits have been kept the same as for the distance test. All values have been shown to either lie within the bounds or being completely off. (At higher angles the $\text{Kinect}_{\text{ToF}}$ shows offsets up to 800mm.)

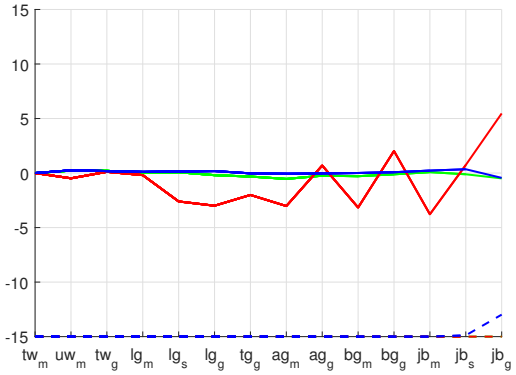
The $\text{Kinect}_{\text{ToF}}$ shows clear offsets already at a straight angle, but for non-straight angles the effects are even more striking. Already at 45° the offsets for all the darker samples are at 10mm or more, which is way too much to expect any usable measurement. At rather straight angles the effects of gloss are clearly visible in the considerable higher noise levels, which shows that glossy surfaces are problematic not only because of their lower diffuse reflectance and specular highlights. With the exception of an outlier at 60° for the brightest glossy sample, it can also be observed that at less straight angles starting from 30°, the effects of gloss generally are reduced and the measurements are similar to the next darker matte sample.



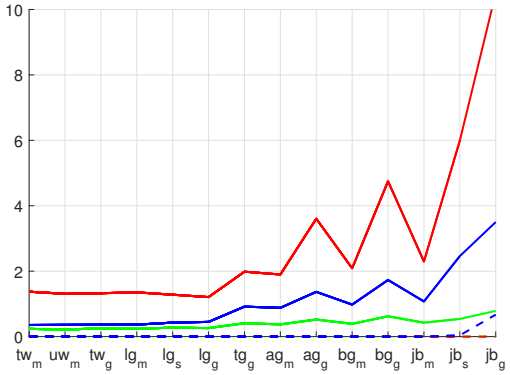
(a) Depth difference (0°)



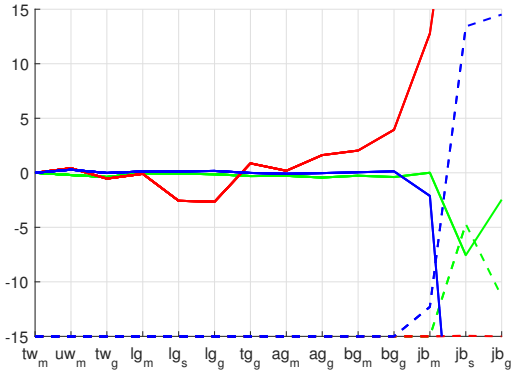
(b) Standard Deviation (0°)



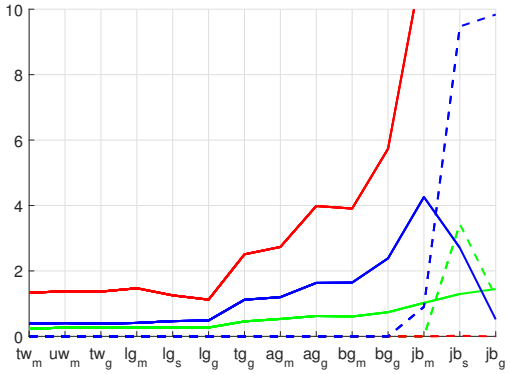
(c) Depth difference (15°)



(d) Standard Deviation (15°)

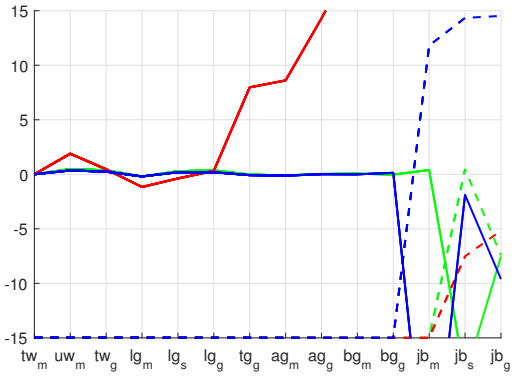


(e) Depth difference (30°)

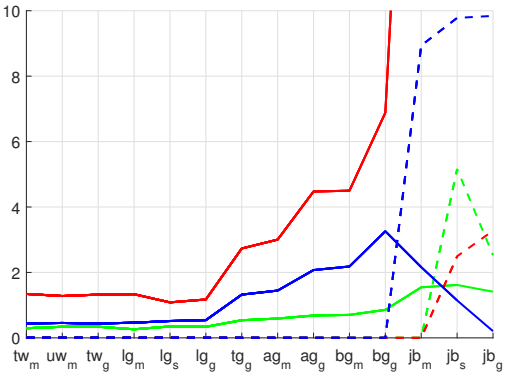


(f) Standard Deviation (30°)

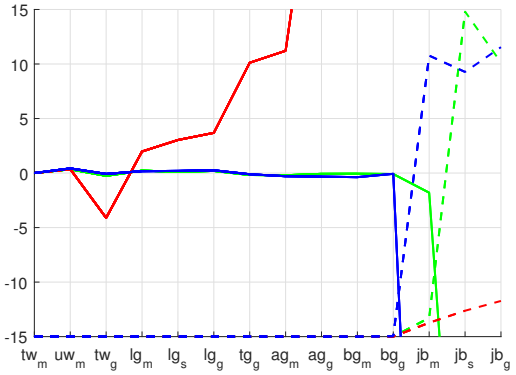
Figure 5.8: Difference to averaged depth value [mm] and average standard deviation [mm] of pixels between frames measured from 1m distance at 0°, 15° and 30°. The ratio of invalid pixels is shown by the dashed lines. Blue: Ensenso N10, Green: $\text{Kinect}_{\text{SL}}$, Red: $\text{Kinect}_{\text{ToF}}$



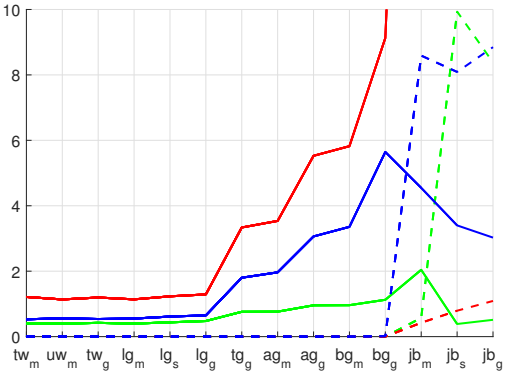
(a) Depth difference (45°)



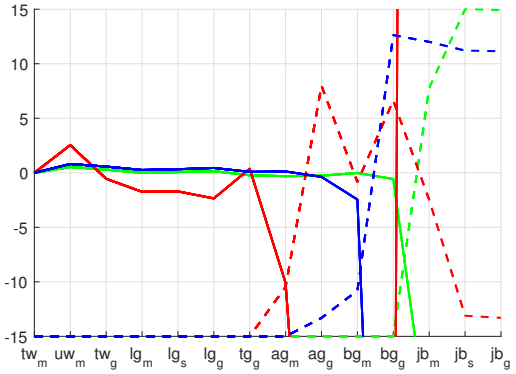
(b) Standard Deviation (45°)



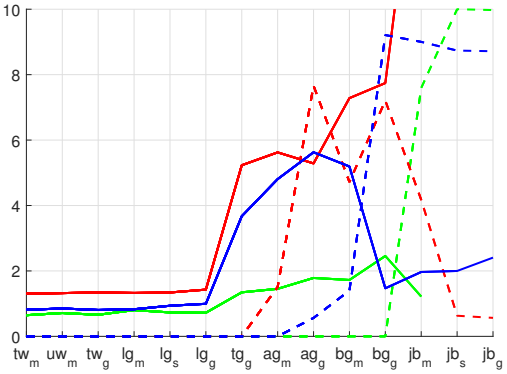
(c) Depth difference (60°)



(d) Standard Deviation (60°)

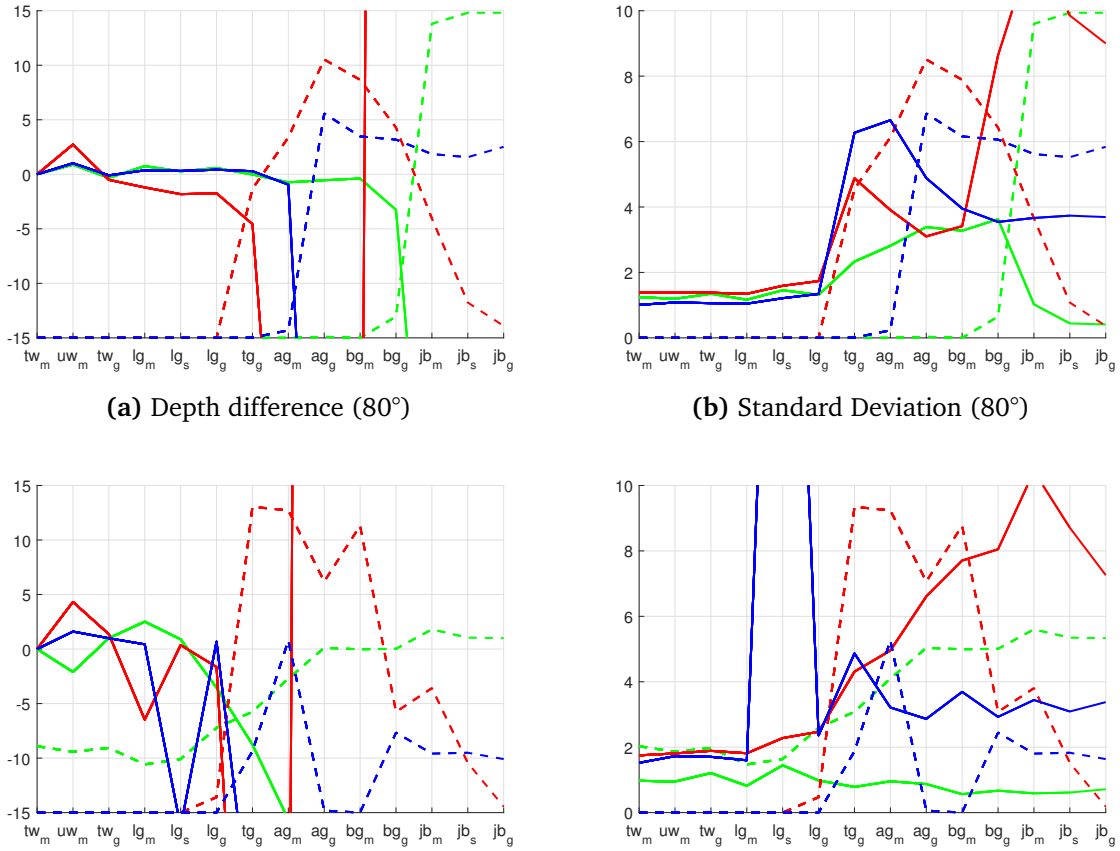


(e) Depth difference (75°)



(f) Standard Deviation (75°)

Figure 5.9: Difference to averaged depth value [mm] and average standard deviation [mm] of pixels between frames measured from 1m distance at 45°, 60° and 75°. The ratio of invalid pixels is shown by the dashed lines. Blue: Ensenso N10, Green: $\text{Kinect}_{\text{SL}}$, Red: $\text{Kinect}_{\text{ToF}}$



(a) Depth difference (80°)

(b) Standard Deviation (80°)

(c) Depth difference (85°)

(d) Standard Deviation (85°)

Figure 5.10: Difference to averaged depth value [mm] and average standard deviation [mm] of pixels between frames measured from 1m distance at 80° and 85°. The ratio of invalid pixels is shown by the dashed lines. Blue: Ensenso N10, Green: $\text{Kinect}_{\text{SL}}$, Red: $\text{Kinect}_{\text{ToF}}$

The $\text{Kinect}_{\text{SL}}$ and Ensenso N10 show rather robust measurements up to 60°, though the Ensenso N10 shows considerable more noise for the darker samples. In contrast to the $\text{Kinect}_{\text{ToF}}$ both systems seem to have a clear threshold at which samples are too dark or glossy to be measured. This threshold is shifted from the black satin matte sample at 30° to the glossy anthracite grey sample at 80° for the Ensenso N10. Finally, at 85° measurements from all systems are becoming equally unusable for all of the darker samples. Although the enormous outlier for the glossy light grey sample by the Ensenso N10 at 85° is caused by an error in measurement, likely to be caused by the interference from one of the other systems that have not been turned off in time to take the measurement.

It is worth noting that the $\text{Kinect}_{\text{SL}}$ and the Ensenso N10 much rather indicate invalid measurements, where the $\text{Kinect}_{\text{ToF}}$ instead delivers completely wrong values. This is an important quality, especially when the data may be combined with measurements from another system.

6 Discussion

Though the experiments described in the previous chapter are showing rather clear results, it is important to keep in mind that all the results are relative to a system's general measurement qualities and cannot simply be used for direct comparison. Both the distance and angle test provide no information on how good the measurement of a system at a certain distance or angle are, but only how much they are affected by different surface material. Therefore conclusions have to be drawn carefully. It is only because the results are showing substantial differences, that they allow to draw more far-reaching conclusions.

This chapter discusses limitations of each of the depth measurement methods on basis of the new experimental results and previous findings with the aim of clarifying implications for applications in assembly environments. Table 6.1 at the end of this chapter provides a brief summary on how the different methods compare in respect to different environmental factors.

6.1 Active Stereo

The test conditions completely ignore one advantage of the Stereo Vision approach by using surfaces with no texture at all. So it might be seen as an edge case, although untextured surfaces are not that uncommon in assembly. But moreover the critical low-albedo surfaces are unlikely to have a clearly visible texture.

The evaluated Ensenso N10 shows considerable effects only for very dark surfaces or at sharper angles. While showing little depth-offset in average, the measurements are noticeable noisier for darker surfaces. A more luminous pattern projector may be able to lessen these effects and achieve results more similar to those of the $\text{Kinect}_{\text{SL}}$. The generally much stronger increase in noise however, suggests a more general weakness of the stereo vision approach. Since the correspondence algorithms are looking for any kind of match, a possible explanation for this is the higher chance of false correspondences caused by random noise.

It might seem that limitations for measuring low-albedo surfaces could be overcome by simply increasing output power of the projector. The Distance test however suggests, that projectors would have to be much stronger to measure darker surfaces from longer distances. Also the the output power can only be increased up to a certain point, since imaging sensors have a limited bandwidth. Too strong projection would easily oversaturate the sensor and therefore prevent the pattern to be picked up from higher-albedo surfaces. This could be mitigated by taking multiple measurements with different light powers, but comes at the cost of increased complexity and measurement times.

A factor that cannot be changed however, is the inherent susceptibility of Active Stereo for occlusion and shadowing by an object's geometry, which is especially a problem in the assembly scenario, where smaller details of an object can be relevant. Though through the possibility of easily using multiple systems, Active Stereo still has an considerable advantage.

6.2 Structured Light

The quantization of depth values used by the $\text{Kinect}_{\text{SL}}$ may cover smaller differences and let noise levels appear even lower than they are. But since the fluctuations in the depth values are well beyond the rather small quantization step size of about 3mm at 1m, the results are still useful for comparison, especially for darker surfaces where the other systems show considerable noisier measurements.

Overall the $\text{Kinect}_{\text{SL}}$ appears to be least affected when measuring darker and glossy surfaces. Effects are only noticable for very dark surfaces or at sharp angles higher than 60°. Even though the projector of the $\text{Kinect}_{\text{SL}}$ is stronger than that of the Ensenso N10, the Structured light approach may also have an systematic advantage. It is less likely to produce false values at higher noise, as it looks for a known pattern instead of possibly corresponding noisy values.

Occlusion and shadowing is also a problem for the Structured Light approach, but in contrast to Active Stereo it is not possible to simply use multiple systems at same time without special measures.

6.3 Time of Flight

It is likely that the $\text{Kinect}_{\text{ToF}}$ would show substantially different results, if tested isolated with less reflective surfaces in the environment. It is however hardly practical to avoid reflective surfaces in working environments, which are usually also required to have

brighter surfaces for ergonomic reasons. Even if the test environment can be considered less optimal, it hardly poses an unlikely case for many working environments. Moreover the large depth offsets for darker surfaces in almost all tested conditions, do not suggest that a slight change in the environmental conditions could mitigate the problem.

What exactly causes the offsets is not clear. A possible explanation is to see it as an extreme case of multipath interference. Since little light is reflected off the direct path, it might be hard to distinguish this signal from that, which comes combined from many other paths and the general noise of the imaging sensor. If multipath effects are involved, then it presumably is an intrinsic problem of the ToF method and cannot be fixed without changing the whole measurement process.

In any case a more powerful light source would clearly not be a solution. On the one hand the light source of the $\text{Kinect}_{\text{ToF}}$ is already strong and needs cooling to keep temperature. And on the other hand already measurements of less dark surfaces at close distance are affected, suggesting that a change of higher magnitude is needed. In the case of involved multipath effects it would also be clear, that increasing light output would likewise increase these multipath effects.

Main advantages of ToF like higher ranges, compactness and less interference from other IR-sources are of less importance in the assembly scenario. Since light source and IR-camera can be put close together, occlusion and shadowing are a much smaller problem for ToF. This is however is a comparatively slight advantage.

	AS	SL	ToF
Long Range	–	–	++
High Ambient IR-light / Sunlight	–	– –	++
Depth Inhomogenity	○	–	–
Multipath interference	++	+	– –
Occlusion / Shadowing	– –	–	+
Dynamic Scenery	+	–	○
Multi-system interference	++	–	– –
Low-albedo Surfaces	+	++	– –

Table 6.1: Overview of relative strengths and weaknesses of Active Stereo (AS), Structured Light (SL) and Time of Flight (ToF)

7 Conclusion

This thesis explored and evaluated different Depth Imaging technology for suitability and limitations of usage in semi-controlled assembly environments. The Active Stereo (AS), Structured Light (SL) and Time of Flight (ToF) methods have been further investigated, by experimentally evaluating three current depth-camera-systems: The Ensenso N10 (AS) from IDS Imaging, the first Kinect (SL) and second Kinect (ToF) from Microsoft. The effects of surface albedo and gloss on measurements from different angles were investigated with a systematic sample set.

7.1 Implications for usage in Assembly Scenario

While the ToF method has several advantages over the triangulation based methods, it appears obviously unsuitable for the measurement of darker surface material (CIELAB Lightness $\leq \sim 40$) in assembly environments. This might already be a criterion for exclusion in many application scenarios, unless the requirements clearly exclude the need for measuring darker surfaces. For glossy surfaces the effects have been shown to be weaker or stronger depending on the measurement angle, though the noise levels were found to consistently equal or higher than those of matte surfaces. Furthermore due to the difficulty of recognizing invalid measurements in such cases, a combination of ToF with another method cannot simply be used for compensation.

The restrictions regarding darker surface material are much less pronounced for both AS and SL. It should be possible to further reduce these restrictions by using more luminous pattern projectors and if necessary combine multiple measurements with different luminosity values. Though SL seems to have an advantage, AS may be preferred for its easy use of multiple systems at the same time.

7.2 Further Research

This thesis presented experiments to asses effects of two different material properties (albedo and gloss) and demonstrated that there are considerable limitations especially

for ToF-systems. In order to identify the boundaries of these limitation more precisely, the sample set could be extended to be more fine grained. Given exact reflectance and gloss levels of the samples for the NIR-range, the results could then be used to model a system's relationship between the material properties and error magnitude respectively noise level.

The sample set could also be extended to other material properties by including less smooth surfaces with different micro-texture. Tough it is more difficult to produce systematic samples in this case, a wider range of commonly used material could be tested to provide a more realistic cross section than only smooth samples do.

Another direction to turn into is to investigate whether and how different material properties influence effects from other error sources like ambient IR light, depth disparities and multipath interference.

An evaluation of different environments could help to shed more light on the huge variations found in the ToF-measurements for low-albedo surfaces and clarify the role of multipath interference. This could be done systematically by measuring in a low-reflection and adding increasingly more reflective surfaces at varying distances.

Additonally the effects of multipath interference from internal reflections could be investigated by using concave shaped samples.

Bibliography

[AW92] E. H. Adelson, J. Y. A. Wang. “Single Lens Stereo with a Plenoptic Camera.” In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 14.2 (1992), pp. 99–106 (cit. on p. 25).

[BBK07] C. Beder, B. Bartczak, R. Koch. “A Comparison of PMD cameras and Stereo Vision for the Task of Surface Reconstruction Using Patchlets.” In: *Computer Vision and Pattern Recognition* (2007), pp. 1–8 (cit. on p. 29).

[BFI+14] A. Bhandari, M. Feigin, S. Izadi, C. Rhemann, M. Schmidt, R. Raskar. “Resolving multipath interference in Kinect: An inverse problem approach.” In: *IEEE SENSORS 2014 Proceedings* (2014), pp. 614–617. URL: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6985073 (cit. on p. 33).

[BH87] S. D. Blostein, T. S. Huang. “Error analysis in stereo determination of 3-d point positions.” In: *IEEE transactions on pattern analysis and machine intelligence* 9.6 (1987), pp. 752–765 (cit. on p. 30).

[Bla04] F. Blais. “Review of 20 years of range sensor development.” In: *Journal of Electronic Imaging* 13.1 (2004), p. 231 (cit. on p. 35).

[BMNK13] K. Berger, S. Meister, R. Nair, D. Kondermann. “A State Of the Art Report on Research in Multiple RGB-D sensor Setups.” In: *Time-of-Flight and Depth Imaging. Sensors, Algorithms, and Applications* 8200 (2013), pp. 257–272 (cit. on p. 29).

[CALT12] J. C. . Chow, K. D. Ang, D. D. Lichti, W. F. Teskey. “Performance Analysis of a Low-Cost Triangulation-Based 3D Camera: Microsoft Kinect System.” In: *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XXXIX-B5.July (2012), pp. 175–180 (cit. on pp. 31, 34).

[DZC12] C. Dal Mutto, P. Zanuttigh, G. M. Cortelazzo. *Time-of-Flight Cameras and Microsoft Kinect™*. 2012, pp. 107–108. URL: http://link.springer.com/10.1007/978-1-4614-3807-6 (cit. on pp. 22, 24, 37).

[EHH15] G. Evangelidis, M. Hansard, R. Horaud. “Fusion of Range and Stereo Data for High-Resolution Scene-Modeling.” In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2015), pp. 1–1. URL: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=7031946 (cit. on p. 27).

[FA11] S. Foix, G. Aleny. “Lock-in Time-of-Flight (ToF) Cameras : A Survey.” In: *Sensors Journal, IEEE* 11.3 (2011), pp. 1–11 (cit. on p. 30).

[FBB+15] M. Funk, A. Bächler, L. Bächler, O. Korn, C. Krieger, T. Heidenreich, A. Schmidt. “Comparing Projected In-Situ Feedback at the Manual Assembly Workplace with Impaired Workers.” In: *Petra´15* (2015), pp. 1–8 (cit. on p. 10).

[FFS12] S. Fernandez, J. Forest Collado, J. Salvi. “Active stereo-matching for one-shot dense reconstruction.” In: *Pattern Recognition* (2012) (cit. on p. 21).

[FS15] M. Funk, A. Schmidt. “Cognitive Assistance in the Workplace.” In: *Pervasive Computing, IEEE* 14.3 (2015), pp. 53–55 (cit. on pp. 10, 41).

[God12] J. P. Godbaz. “Ameliorating systematic errors in full-field AMCW lidar.” PhD thesis. 2012 (cit. on p. 32).

[GRV+13] H. Gonzalez-Jorge, B. Riveiro, E. Vazquez-Fernandez, J. Martínez-Sánchez, P. Arias. “Metrological evaluation of Microsoft Kinect and Asus Xtion sensors.” In: *Measurement: Journal of the International Measurement Confederation* 46.6 (2013), pp. 1800–1806. URL: http://dx.doi.org/10.1016/j.measurement.2013.01.011 (cit. on p. 29).

[HLC+12] M. Hansard, S. Lee, O. Choi, R. Horaud, M. Hansard, S. Lee, O. Choi, R. Horaud, F. Cameras. *Time of Flight Cameras : Principles , Methods , and Applications*. 2012 (cit. on pp. 31, 33, 34).

[HM09] R. Hiesgen, G. Meichsner. “Farbtonübereinstimmung bei Lacken aus dem Dekor-und Industrielackbereich.” In: *Farbe & Lack* 115 (2009), pp. 132–135 (cit. on p. 45).

[IDS] IDS Imaging Development Systems GmbH. *Ensenso N10 Stereo 3D camera*. URL: https://en.ids-imaging.com/store/ensenso-n10.html (cit. on p. 36).

[iFi] iFixit. *Xbox One Kinect Teardown*. URL: https://www.ifixit.com/Teardown/Xbox+One+Kinect+Teardown/19725 (cit. on p. 39).

[JJSL13] W. Jang, C. Je, Y. Seo, S. W. Lee. “Structured-light stereo: Comparative analysis and integration of structured-light and active stereo for measuring dynamic shape.” In: *Optics and Lasers in Engineering* 51.11 (2013), pp. 1255–1264 (cit. on pp. 27, 31).

[KE12] K. Khoshelham, S. O. Elberink. “Accuracy and Resolution of Kinect Depth Data for Indoor Mapping Applications.” In: *Sensors* 12.12 (2012), pp. 1437–1454. URL: http://www.mdpi.com/1424-8220/12/2/1437/ (cit. on p. 37).

[KS06] K. Kuhnert, M. Stommel. “Fusion of Stereo-Camera and PMD-Camera Data for Real-Time Suited Precise 3D Environment Reconstruction.” In: *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*. Oct. 2006, pp. 4780–4785 (cit. on p. 27).

[KTS+15] A. Kadambi, V. Taamazyan, B. Shi, R. Raskar, A. Iccv, P. Id. “Polarized 3D : high-quality depth sensing with polarization cues.” In: *Proceedings of the IEEE International Conference on Computer Vision*. 2015 (cit. on p. 27).

[LG09] A. Lumsdaine, T. Georgiev. “The Focused Plenoptic Camera.” In: *Computational Photography (ICCP), 2009 IEEE International Conference on. IEEE*. 2009 (cit. on p. 25).

[LHL12] B. Langmann, K. Hartmann, O. Loffeld. “Depth Camera Technology Comparison and Performance Evaluation.” In: *Proceedings of the 1st International Conference on Pattern Recognition Applications and Methods* (2012), pp. 438–444. URL: http://www.scitepress.org/DigitalLibrary/Link.aspx?doi=10.5220/0003778304380444 (cit. on pp. 31, 34).

[LJH14] T. Luhmann, C. Jepping, B. Herd. “Untersuchung zum messtechnischen Genauigkeitspotenzial einer Lichtfeldkamera.” In: *DGPF Tagungsband 23 / 2014*. 2014 (cit. on pp. 25, 26).

[LNL+13] D. Lefloch, R. Nair, F. Lenzen, H. Schäfer, L. Streeter, M. J. Cree, R. Koch, A. Kolb. “Technical foundation and calibration methods for time-of-flight cameras.” In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 8200 LNCS (2013), pp. 3–24 (cit. on pp. 30, 32).

[LZ99] C. Loop, Z. Z. Z. Zhang. “Computing rectifying homographies for stereo vision.” In: *Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149)* 1 (1999), pp. 125–131 (cit. on pp. 19, 20).

[MGH09] S. P. McPherron, T. Gernat, J. J. Hublin. “Structured light scanning for high-resolution documentation of in situ archaeological finds.” In: *Journal of Archaeological Science* 36.1 (2009), pp. 19–24. URL: http://dx.doi.org/10.1016/j.jas.2008.06.028 (cit. on p. 29).

[Mica] Microsoft Corporation. *Kinect for Windows specifications*. URL: https://msdn.microsoft.com/en-us/library/jj131033.aspx (cit. on p. 37).

[Micb] Microsoft Corporation. *Kinect for Windows v2 Sensor*. URL: https://news.microsoft.com/kinect-for-windows-v2-sensor/ (cit. on p. 39).

[MOT] MOTIP DUPLI GmbH. *TECHNICAL INFORMATION AEROSOL-ART*. URL: http://www.motipdupli.com/en/INT/products/dupli-color/decoration/color/aerosol-art/ipg-1050/tm-1050.html (cit. on p. 45).

[NDM05] A. Ngan, F. Durand, W. Matusik. "Experimental analysis of BRDF models." In: *Proceedings of the Eurographics Symposium on Rendering* (2005), pp. 117–126. URL: http://dl.acm.org/citation.cfm?id=2383654.2383671 (cit. on p. 18).

[NIK91] S. K. Nayar, K. Ikeuchi, T. Kanade. "Surface reflection: physical and geometrical perspectives." In: *IEEE Transactions on Pattern Analysis & Machine Intelligence* 7 (1991), pp. 611–634 (cit. on p. 16).

[NIL12] C. V. Nguyen, S. Izadi, D. Lovell. "Modeling kinect sensor noise for improved 3D reconstruction and tracking." In: *Proceedings - 2nd Joint 3DIM/3DPVT Conference: 3D Imaging, Modeling, Processing, Visualization and Transmission, 3DIMPVT 2012* July 2015 (2012), pp. 524–530 (cit. on p. 32).

[NRL+13] R. Nair, K. Ruhl, F. Lenzen, S. Meister, H. Schäfer, C. S. Garbe, M. Eisemann, M. Magnor, D. Kondermann. "A survey on Time-of-Flight stereo fusion." In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 8200 LNCS (2013), pp. 105–127 (cit. on p. 27).

[PBM+14] S. Paulus, J. Behmann, A.-K. Mahlein, L. Plümer, H. Kuhlmann. "Low-Cost 3D Systems: Suitable Tools for Plant Phenotyping." In: *Sensors* 14.2 (2014), pp. 3001–3018. URL: http://www.mdpi.com/1424-8220/14/2/3001/ (cit. on p. 29).

[PP15] D. Pagliari, L. Pinto. "Calibration of Kinect for Xbox One and comparison between the two generations of microsoft sensors." In: *Sensors* 15.11 (2015) (cit. on pp. 38, 39).

[RDP+11] M. Reynolds, J. Doboš, L. Peel, T. Weyrich, G. J. Brostow. "Capturing Time-of-Flight data with confidence." In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (2011), pp. 945–952 (cit. on p. 32).

[SJP11] J. Smisek, M. Jancosek, T. Pajdla. "3D with Kinect." In: *Image (Rochester, N.Y.)* (2011) (cit. on pp. 37, 47).

[SLAL11] T. Stoyanov, A. Louloudi, H. Andreasson, A. J. Lilienthal. "Comparative evaluation of range sensor accuracy in indoor environments." In: *Proceedings of the 5th European Conference on Mobile Robots, ECMR 2011* (2011), pp. 19–24 (cit. on p. 29).

[SLC+15] J. Steward, D. Lichti, J. Chow, R. Ferber, S. Osis. “Performance Assessment and Calibration of the Kinect 2.0 Time–Of–Flight Range Camera for Use in Motion Capture Applications.” In: *FIG Working week 2015*. May. 2015, pp. 17–21 (cit. on p. 34).

[SLK15] H. Sarbolandi, D. Lefloch, A. Kolb. “Kinect range sensing: Structured-light versus Time-of-Flight Kinect.” In: *Computer Vision and Image Understanding* 000 (2015), pp. 1–20. URL: http://linkinghub.elsevier.com/retrieve/pii/S1077314215001071 (cit. on pp. 29–34).

[SPB04] J. Salvi, J. Pagâ, J. Batlle. “Pattern codification strategies in structured light systems.” In: *Pattern Recognition* 37 (2004), pp. 827–849 (cit. on p. 23).

[SS01] D. Scharstein, R. Szeliski. “A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms.” In: *International journal of computer vision* 47.1-3 (2001), pp. 7–42 (cit. on p. 19).

[Ste] Stereolabs Inc. *ZED - 3D Camera for AR/VR and Autonomous Navigation*. URL: https://www.stereolabs.com/zed/specs/ (cit. on p. 36).

[Sze10] R. Szeliski. *Computer Vision: Algorithms and Applications*. Springer Science & Business Media, 2010 (cit. on pp. 9, 16, 19, 30).

[Tro95] M. Trobina. “Error Model of a Coded-Light Range Sensor Error Model of a Coded-Light Range Sensor.” In: *Technique Report, Communication Technology Laboratory* (1995) (cit. on pp. 22, 23, 30).

[USMF93] C. W. Urquhart, J. P. Siebert, J. P. McDonald, R. J. Fryer. “Active Animate Stereo Vision.” In: *Procedings of the British Machine Vision Conference 1993*. British Machine Vision Association, 1993, pp. 1–8. URL: http://www.bmva.org/bmvc/1993/bmvc-93-008.html (cit. on p. 21).

[War92] G. J. Ward. “Measuring and modeling anisotropic reflection.” In: *ACM SIGGRAPH Computer Graphics* 26.2 (1992), pp. 265–272 (cit. on pp. 17, 18).

[YZYM15] M. Ye, Y. Zhang, R. Yang, D. Manocha. “3D Reconstruction in the Presence of Glasses by Acoustic and Stereo Fusion.” In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2015 (cit. on p. 27).

All links were last followed on May 30, 2016.

Declaration

I hereby declare that the work presented in this thesis is entirely my own and that I did not use any other sources and references than the listed ones. I have marked all direct or indirect statements from other sources contained therein as quotations. Neither this work nor significant parts of it were part of another examination procedure. I have not published this work in whole or in part before. The electronic copy is consistent with all submitted copies.

---

place, date, signature