Matti Laukkanen

Performance Evaluation of Time-of-Flight Depth Cameras

School of Electrical Engineering

Thesis submitted for examination for the degree of Master of Science in Technology.

Espoo 23.11.2015

Thesis supervisor:

Prof. Ville Kyrki

Thesis advisor:

M.Sc. (Tech.) Paul Kemppi

Aalto University
School of Electrical
Engineering

AALTO UNIVERSITY
SCHOOL OF ELECTRICAL ENGINEERING

ABSTRACT OF THE
MASTER'S THESIS

Author: Matti Laukkanen

Title: Performance Evaluation of Time-of-Flight Depth Cameras

Date: 23.11.2015 Language: English Number of pages: 7+55

Department of Electrical Engineering and Automation

Professorship: Automation technology Code: AS-84

Supervisor: Prof. Ville Kyrki

Advisor: M.Sc. (Tech.) Paul Kemppi

In this thesis, accuracy of Time-of-Flight (ToF) depth cameras was studied. ToF camera is relatively new technology, which is a potential challenger for traditional 3D imaging methods. However, the accuracy of ToF measurements is limited by several factors, such as the properties of the target or the intensity of background illumination.

In the experimental part of this thesis, three ToF cameras are compared. The cameras are Mesa SR4000, Panasonic D-IMager EKL3106 and Microsoft Kinect v2. For the evaluation, five different test scenarios were carried out in indoor and outdoor environments. The goal of the measurements was to study how the temperature of the camera, background illumination and distance and reflectivity of the imaged objects affect the depth accuracy.

The results showed that although all tested cameras had difficulties in some conditions, ToF camera technology has developed rapidly. Current cameras perform more robustly in challenging conditions, which enables a whole new range of applications.

Keywords: Time-of-Flight, Depth camera, SR4000, D-IMager, Kinect v2

AALTO-YLIOPISTO
SÄHKÖTEKNIIKAN KORKEAKOULU

DIPLOMITYÖN
TIIVISTELMÄ

Tekijä: Matti Laukkanen

Työn nimi: Time-of-Flight -syvyyskameroiden suorituskyvyn arviointi

Päivämäärä: 23.11.2015 Kieli: Englanti Sivumäärä: 7+55

Sähkötekniikan ja automaation laitos

Professuuri: Automaatiotekniikka Koodi: AS-84

Valvoja: Prof. Ville Kyrki

Ohjaaja: DI Paul Kemppi

Tässä työssä tutkittiin Time-of-Flight (ToF) syvyyskameroiden tarkkuutta. ToF kamera on suhteellisen uusi teknologia ja se tarjoaa varteenotettavan vaihtoehdon perinteisemmille 3D-kuvantamistekniikoille. Mittausten tarkkuutta kuitenkin rajoittavat jotkin ympäristötekijät, kuten taustavalaistuksen voimakkuus tai kuvattavan kohteen ominaisuudet.

Työn kokeellisessa osiossa vertailtiin kolmea kameraa, jotka olivat Mesa SR4000, Panasonic D-IMager EKL3106 ja Microsoft Kinect v2. Vertailua varten tehtiin viisi erilaista mittausta sisä- ja ulkotiloissa. Mittauksissa tutkittiin kameran sisäisen lämpötilan, kuvattavan kohteen etäisyyden ja heijastavuuden sekä taustavalaistuksen vaikutusta mittaustarkkuuteen.

Tulokset osoittivat, että vaikka kaikilla kameroilla oli tietyissä olosuhteissa ongelmia, on ToF-kamerateknologia kehittynyt nopeasti. Nykyiset kamerat suoriutuvat hyvin yhä useammissa olosuhteissa, mikä mahdollistaa kasvavan joukon uusia sovelluksia.

Avainsanat: Time-of-Flight, syvyyskamera, SR4000, D-IMager, Kinect v2

Preface

This research was conducted at VTT Technical Research Centre of Finland as a part of Reconfigurable ROS-based Resilient Reasoning Robotic Cooperating Systems (R5-COP) project.

First of all I would like to thank my instructor Paul Kemppi from VTT for sharing his expertise and lending a helping hand whenever needed. I especially appreciate his assistance on PCL, ToF camera drivers and other software issues. I would also like to thank my thesis supervisor Prof. Ville Kyrki from Aalto University for valuable feedback and great last-minute improvement suggestions.

Several VTTers also deserve thanks. First, thanks to our team leader Juha Zidbeck and project manager Jukka Koskinen, who made this work possible. I would also like to thank Seppo Horsmanheimo for useful pieces of advice regarding writing, and Joonas Elo for peer pressure, as he was working on his master's thesis at the same time.

Finally I would like to thank my family and all friends who gave their support during this thesis work. Very special thanks go to a very special person - my precious and beloved Erika, who was always supporting and encouraging me during this writing process. You are the sunshine of my life.

Otaniemi, 23.11.2015

Matti Laukkanen

Contents

Abbreviations

API	Application Programming Interface
CWM	Continuous Wave Modulated
CV	Coefficient of Variation
FOV	Field of View
IT	Integration Time
LED	Light Emitting Diode
LRF	Laser Rage Finder
PCL	PointCloud Library
PMD	Photonic Mixer Device
ROS	Robot Operating System
SD	Standard deviation
SDK	Software Development Kit
SL	Structured Light
SNR	Signal to Noise Ratio
SPAD	Single Photon Avalanche Diode
SR4000	MESA SwissRanger 4000
ToF	Time of Flight

1 Introduction

Depth imaging means sensing the three-dimensional geometrical structures of our surrounding world. The third dimension adds lots of valuable information about the scene if compared to traditional 2D images. That is why depth cameras have been used for several decades in various applications, such as remote sensing of Earth and mapping the environment [1]. In last few years, as low-cost depth sensors have appeared on the markets, a whole new range of applications has emerged. A good example is Kinect [2], a cheap structured light sensor released in 2010. It quickly became popular and ever since it has been studied extensively and used in numerous applications, including but not limited to a vision sensor for low-cost mobile robot [3], human machine interface and healthcare [4].

Several distinct technologies for 3D imaging exist, some of which are based on triangulation and others on propagation time of electromagnetic radiation. The most recent depth imaging technology is Time-of-Flight (ToF) camera. Compared to other systems, ToF cameras have some promising features: high frame rate, compact size and very low computational requirements. Additionally, performance in different lighting conditions is robust, which makes ToF camera a great challenger against other technologies.

On the other hand, Time-of-Flight cameras have very limited operating range and they have traditionally been suffering from rather low resolutions and high prices. However, rapid development of semiconductor technology has led to huge improvements recently. At the end of 2013 Microsoft announced the new version of Kinect [5], based on Time-of-Flight. It features an image sensor of 512 x 424 pixels, which is many times more than in previous ToF cameras. Moreover, due to broad consumer markets and mass production, the price is one order of magnitude lower than before.

Since the first ToF cameras appeared about two decades ago, they have been studied actively [6]. There have been several general surveys of the technology with the error sources addressed [7, 8, 9], quantitative performance evaluations in different models [10, 11, 12], and comparisons with other depth imaging technologies, such as structured light [9, 13, 14, 15] and stereo cameras [16, 17].

Majority of the studies have been conducted in indoor environments, but also outdoor performance has been studied. Dynamic weather conditions, such as rain and bright sunlight pose many challenges to ToF cameras [18, 19, 20, 21, 22], but recent advances in sensor technology yielded more promising results. Even underwater performance has been studied with Kinect v2 [23].

The goal of this thesis is to provide a comprehensive overview of the Time-of-Flight camera technology. To support the theory part, performance of three recent ToF cameras is evaluated in the experimental part. Based on earlier work, a set of test scenarios is selected so that the devices can be evaluated and compared, especially in situations that are problematic for the ToF technology. Since the devices

have slightly different characteristics, also the impact of various sensor designs on performance is studied.

The evaluated cameras are MESA SR4000 (released in 2010), Panasonic D-IMager EKL-3106 (2012) and Microsoft Kinect v2 (2014). Whereas SR4000 and Kinect v2 have been studied rather extensively, D-IMager has only few references in the literature. It has been used in a few applications such as gesture recognition [24], people tracking [25] and robotic vision [26], but apparently no comprehensive performance analysis has been done with it. Therefore, as the same tests are carried out with all three cameras, the following contributions are provided:

1. Repeat the tests carried out by the other authors with SR4000 and Kinect v2 and verify their results,

2. Provide performance evaluation data for the Panasonic D-IMager,

3. Compare the three cameras with each other.

The goal of the comparison is also to reflect the evolution of ToF cameras over the last few years, and to find out how well the consumer-grade Kinect v2 can challenge the more expensive sensors.

Performance has to be defined, as it is very application-specific term and has alternative meanings in different contexts. In this thesis, "performance" is conceived as depth accuracy in varying conditions. For example, depth error relative to distance could be measured. To evaluate and compare the ToF cameras with each other, a reference point needs to be chosen. In the above example this could be achieved e.g. by measuring the ground truth using an additional sensor, after which the data from each ToF camera could be compared to a common reference point. To evaluate the cameras, several test scenarios are carried out.

The thesis is structured as follows: In Chapter 2, the working principles of ToF camera are explained and its advantages and disadvantages against other prominent depth imaging technologies are briefly discussed. In addition, error sources that affect the depth measurements are introduced. In Chapter 3, the experimental setup is introduced. The specifications of the evaluated cameras are presented and their differences are discussed. Moreover, software and methods used in the experiments are outlined. In Chapter 4, more detailed descriptions of the test scenarios are given and the results are analysed. Finally in Chapter 5, the contents of the thesis are summarized and conclusions of the impact of the results are made.

2 Background

In this chapter, the Time-of-Flight camera technology is discussed. First, the structure of a ToF camera and its working principles are presented. Then, the most prominent other depth imaging technologies are briefly reviewed, and their advantages and disadvantages are compared. Finally, the most important error sources affecting depth measurements in ToF cameras are described.

2.1 Time-of-Flight Camera

Time-of-Flight camera physically resembles a traditional digital camera. It consists of a lens collecting light to a image sensor, which is a two-dimensional array of photosensing elements, pixels. Additionally, ToF camera has an active light source that is used to illuminate the scene. The camera registers the light reflected from the scene and each pixel in parallel calculates depth information, forming a depth image from the whole scene. All of these components are packed into a compact package with no moving parts. Two examples of state-of-the-art Time-of-Flight cameras are shown in Figure 1.

Figure 1: Two Time-of-Flight cameras, MESA SR4000 [29] and Microsoft Kinect v2 [5].

Light emitting diodes (LEDs) are usually used as a light source, because of their fast response time [27]. Most commercial sensors use near infrared (NIR) wavelengths not visible to human eye, usually around 850 nm [6, 28]. NIR region of the spectrum used is because many materials have relatively high reflectance on that band [29] and it does not interfere with human vision.

Pixels of the image sensor can be implemented in several ways, depending on operation principle discussed more in next section. Currently most systems are analogue, using photodetectors and capacitors to collect and store electricity from light pulses, after which analogue signals are converted to digital [7]. Also fully digital ToF cameras have been researched, based on single photon avalanche diodes (SPAD) [30] which can register even single photons. A fully digital system helps reducing noise related to analogue signals and digital conversion process [7, 28].

2.2 Theory of Operation

The basic working principle of Time-of-Flight cameras is that they illuminate the scene with modulated light source and then measure the returning light that reflects

back to the sensor. Because the speed of light is constant, the distance of the object where the light was reflected from can be calculated from the time difference of the emitted and returning light signals. Two different techniques of illumination are used in ToF cameras, either pulsed light or continuous wave (CW) modulated light. [7].

2.2.1 Pulsed modulation

In the pulsed method the depth measurement is straightforward. The illumination unit is switched on and off rapidly, thus generating short light pulses. A timer starts when the light pulse is sent and stops when the reflected light is detected by the sensor. The distance to the object can then be calculated by [27]

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

where Δt is the round-trip time of the light pulse and c is the speed of light. However, the ambient illumination usually contains same wavelengths as the light source of the ToF camera. Therefore the light received by camera is a sum of the emitted light and ambient light. This would lead to distance calculation errors, so the light is recorded also when the illumination unit switched off so that the background can be subtracted from signal. This can be done by using outgoing light signal as a control signal for the detector sensor. Moreover, one short light pulse contains rather low amount of energy [30], and due to unideal components in the system [7], received signal is noisy. In order to increase the Signal-to-noise ratio (SNR), multiple cycles of pulses, typically millions, are recorded for a specific period of time and the depth information is computed from the average. This time interval is called integration time (IT) [31]. The concept of this pulsed modulation method is illustrated in Figure 2. By using the integration time of Δt and two out-of-phase sampling windows C_1 and C_2, the averaged distance is computed by [31]:

$$d = \frac{1}{2} c \ \Delta t \frac{Q_2}{Q_1 + Q_2}, \tag{2}$$

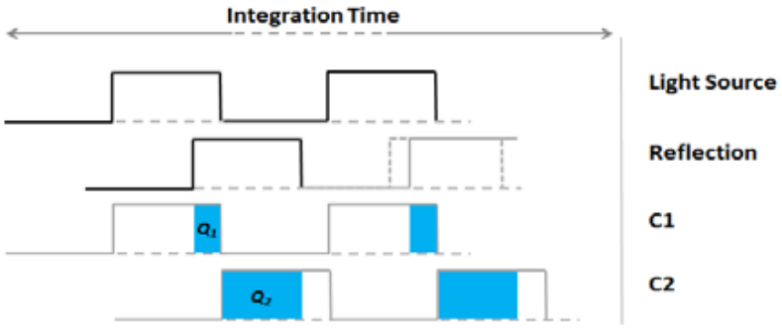

Figure 2: Pulsed Time-of-Flight method. Received signal is sampled in two out-of-phase windows in parallel. [31]

where Q_1 and Q_2 are the accumulated electrical charges received over the integration time period.

Depth resolution of the pulsed method is limited by the speed of the electronics in the camera. From (1) it can be calculated that to achieve depth resolution of 1 mm, one would need a light pulse of approximately 6.6 picoseconds. However, current LEDs/laser diodes limit the rising and falling times and repetition times of the pulses [27]. In addition, achieving such speeds in the receiver circuit is challenging with current silicon-based technology in room temperatures. [31]

2.2.2 Continuous Wave Modulation

Instead of direct measurement of the round-trip time of a light pulse, the CW modulation method is based on measuring phase difference of sent and received signals. The light is modulated by altering the input current to the light source, generating waveform signal [27]. Different shapes of modulation signal can be used, but usually either square or sinusoidal waves are used [32]. CW modulation technique lowers the requirements for the light source, so better depth resolution can be achieved compared to pulsed light.

There are various styles to demodulate the received signal and extract amplitude and phase information from it. A traditional way is to calculate cross correlation function of the original modulation signal and returned signal [32]. Cross correlation can be calculated by measuring the returned signal at selected phases, which can be implemented using mixers and low-pass filters in the detector. This, however requires very complicated circuitry [8].

Another, more efficient approach is to sample the modulated returned light synchronously using special pixel structure. Received modulated light is simultaneously mixed with reference signal and sampled at four different phases (0°, 90°, 180°, 270°), as illustrated in Figure 3 [32]. Advantage of this synchronous sampling technique is simpler circuit design and smaller pixel sizes, which allows putting more pixels in to a sensor, resulting in higher resolution. In the literature, this kind of pixel architecture is referred as photonic mixer device (PMD) [9, 27, 33] or lock-in-pixel [7, 8].

As with the pulsed method, multiple samples are measured and averaged to increase the SNR. Using four equally spaced sampling windows Q_1 to Q_4 timed by the reference signal (see Figure 3), received signal is sampled at different phases for a integration time period. Assuming that the modulation signal is sinusoidal wave without harmonic frequencies, discrete Fourier transform (DFT) equations can be used to calculate phase ϕ, amplitude A and offset B as follows [32, 31]:

$$\phi = arctan\left(\frac{Q_3 - Q_4}{Q_1 - Q_2}\right) \tag{3}$$

$$A = \frac{\sqrt{(Q_1 - Q_2)^2 + (Q_3 - Q_4)^2}}{2} \tag{4}$$

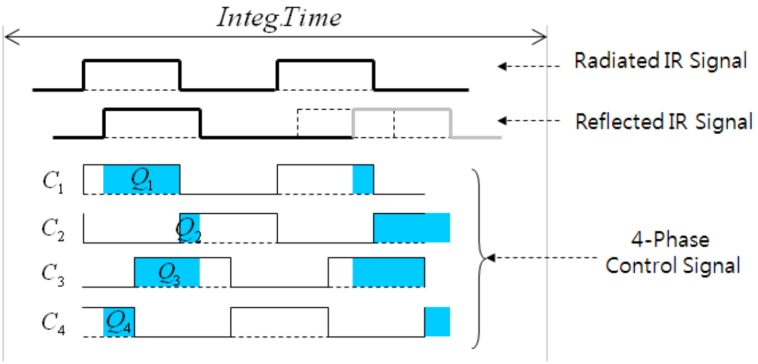

Figure 3: Demodulation principle of CW method. $C1$ to $C4$ are control signals with 90° phase delay from each other. [34, p. 3]

$$B = \frac{Q_1 + Q_2 + Q_3 + Q_4}{4}. \tag{5}$$

From the phase ϕ, the distance can be finally calculated as [31]:

$$d = \frac{c}{4\pi f}\phi. \tag{6}$$

The intensity, i.e. amplitude A of the light decreases proportionally to the travelled distance in a known way. Hence, the received amplitude value from (4) can be used as a confidence measure for the distance measurements [29]. Additionally, the reflected signal is often superimposed to background illumination, which causes error to the measurement. Thus, the offset (5) is used to distinguish modulated light component from the background light [32].

When calculating distances from phase difference as in (6), one important thing has to be considered. Since the modulation signal is periodical, its phase wraps around every 2π. This means that also distances can be measured unambiguously only in a certain range:

$$L = \frac{c}{2f}, \tag{7}$$

which depends only on the modulation frequency f. For example, using typical 20 MHz modulation frequency, unambiguous range is 7.5 meters. Therefore, reflected signal coming from distance of 8.6 meters would be mapped to 1.1 meters, since they are in same phase. Operating range can be extended by using lower modulation frequency, but it results to decreasing of resolution [16]. Another way is to take multiple measurements with different modulation frequencies. First, coarse distance is calculated using lower frequency, and after that higher frequency is used to increase accuracy of the measurement [31]. However, using multiple frequencies shortens the integration time that can be used, thus decreasing the SNR. [34, p. 37].

To summarize, pulsed method gives better operating range (up to 1500 m [35]), but depth resolution is worse than in CW modulated method. On the other hand, CW method has an ambiguity problem, which limits the operating range. Most commercial ToF cameras are based on continuous wave modulation [7, 35].

2.3 Data Formats

Three-dimensional data can be stored and presented in various ways. The data types differ in memory usage, computational effectiveness and amount of information they can preserve. *Depth image* and *point cloud*, both commonly used with ToF cameras, are presented and their common uses and limitations are considered. Examples of these data types are shown in Figure 4. Other representations exist as well, but they are out of scope of this work. For further information, please see [1, Ch. 4].

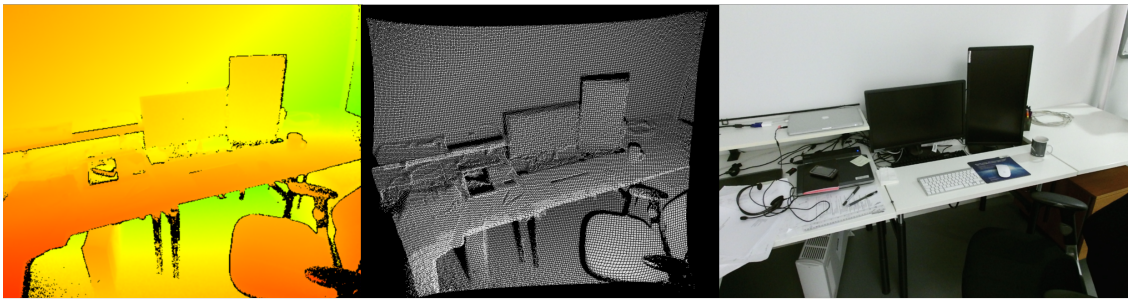

Figure 4: Depth image, point cloud and RGB image taken from same scene.

Depth image, *range image* or sometimes called *depth map* is a data type that ToF cameras usually output directly. Similarly to traditional intensity image, depth image is a two-dimensional array with the width of x and the height of y pixels, but each pixel containing a depth value instead of intensity. Depending on the device, pixels of the output image can also contain intensity and amplitude information, which can be used as a confidence measure [29].

Because a depth image is acquired from a single viewpoint, it can only have one depth value for each x, y position. If there are occluding, i.e. overlapping objects in the scene, only the nearest point from the camera can be stored in a depth image. Therefore, depth image is considered being actually 2.5D instead of 3D [1, p. 72].

A point cloud is an unstructured data format, where the 3D data is stored as a set of points, each having x, y, z coordinates. Points can also include additional data, such as RGB values. The cloud can hold more complex geometric shapes than range image, allowing to combine data from multiple viewpoints to a single structure. Since each point needs to hold at least three floating point values, the memory consumption of point cloud is higher than in a depth image.

A point cloud can be converted to a depth image by projecting the data points from single perspective to a 2D plane, but all occluding points are lost. On the other hand, a depth image can also be converted to a point cloud if the intrinsic

parameters, such as focal length and lens distortion of the camera are known. Most commercial Time-of-Flight cameras provide both depth images and point clouds as output.

2.4 Comparison to Other Technologies

In addition to Time-of-Flight cameras, there are several other technologies traditionally used for acquiring depth information from the scene. These systems can be divided into two categories, passive and active. Passive systems, such as stereo cameras, do not use any source of illumination and are only relying on ambient lighting. On the other hand, active systems (where ToF camera also belongs to) illuminate the scene with a light source, and the depth information is computed in various ways. [1, p. 11]

Each system has its limitations, and best choice always depends on the application. Properties that need to be considered include operating range, resolution, physical size and performance in different environmental conditions. In this chapter, some most popular depth imaging techniques are discussed. Different approaches for both active and passive systems are briefly presented, and their advantages and disadvantages compared to ToF cameras are considered.

2.4.1 Passive 3D Imaging systems

Passive 3D imaging systems, generally based on traditional 2D gray-level or RGB cameras, are the oldest and the most studied depth imaging technology. Several different approaches fall into the passive category, some employing only a single camera while others use multiple cameras. They all have in common the fact that they use only ambient lighting, so the depth information is extracted from the details in 2D images. Single-camera methods, such as shape-from-focus or structure-from-motion use sequences of images, and geometry of the scene is computed from variations between the frames. The most well-known multi-camera method is stereo vision, where two or more fixed cameras image the same scene from slightly different perspectives. Stereo camera system is illustrated in Figure 5. Corresponding features (e.g. edges, corners) are searched from both images and the depth information can then be determined using *triangulation*. For more details on passive methods, see [1, ch. 2].

Stereo vision systems have been researched for several decades, so they are rather well-known. With modern technology, stereo vision system can be implemented by using even cheap consumer cameras and personal computers, so the costs can be very low [31]. The depth resolution and the operating range of a Stereo vision system can also be rather easily changed by selecting suitable camera equipment and adjusting the baseline of the cameras [7].

Compared to ToF cameras, stereo vision as well as other passive systems have a few major drawbacks. First, they are highly sensitive to changes in lighting and

cannot operate at all in darkness, since they do not employ a light source like ToF cameras. Secondly, the computation of depth values is based on details and features in the image. This means that passive methods perform poorly in non-textured and uniformly coloured surfaces that do not have many details. Third, the correspondence problem, i.e. finding corresponding features in different images is very complex and computationally intensive process. Last, setting up and calibrating a stereo camera system requires great amount of work and knowledge. Several commercial implementations exist, but they usually are fixed and not easily configurable [1, ch. 2.9.1].

In contrast, Time-of-flight cameras perform well in low-detailed scenes and in different lighting conditions. Depth computation is done directly on the chip, so they can provide depth data at high frame rate without need for heavy computing. Operating range can be changed easily by adjusting integration time, and no calibration is needed [16].

2.4.2 Structured Light

Structured light (SL) approach is also based on triangulation principle, but it uses active illumination unit that projects specific pattern of light on to a scene. Camera sees the distorted pattern from a different viewpoint, and the geometry of the surface can then be calculated from deformations in the observed pattern. Many alternative light patterns can be used, such as lines or speckles. [7]

The most successful structured light device has been Kinect [2] by Microsoft. Since its launch in 2010, numerous applications have been developed and its performance has been extensively studied. Several comparisons to Time-of-Flight cameras have also been done, with varying conclusions [7, 9, 13]. A few years ago ToF cameras still had such a low resolution that Kinect outperformed state-of-the-art ToF camera models in accuracy, when operated in short ranges and in a controlled environment. However, the operating range of the Kinect is very limited (3.5 meters) [13], so it is suitable only for limited applications.

Structured light systems generally, not limited only to Kinect, have also some other disadvantages compared to ToF cameras. First, they suffer from partial occlusion because the light projector and the camera are in different viewpoints. This

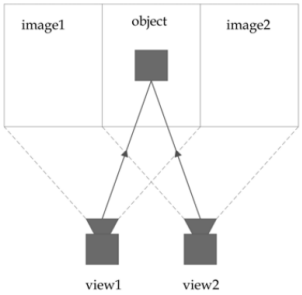

Figure 5: A stereo camera system. [16]

leads to missing depth information in parts of the image [7]. Secondly, SL systems are very sensitive to lighting conditions, so they generally do not work well in bright sunlight. Also ToF camera is sensitive to background light, but the received signal is easier to differentiate from the background component [15].

Last, because of the triangulation method, the computation of depth values is significantly more demanding than in ToF cameras.

2.4.3 Laser rangefinders

A *laser rangefinder* (LRF), also called *Lidar* or *laser scanner*, is based on the same principle as ToF camera, measuring the time-of-flight of a reflected light. Instead of illuminating the whole scene, it uses a laser beam to illuminate and measure only a single point at a time. Typically a laser scanner device also has a mirror rotating at high speed, so it can measure distances in a 2D plane. To acquire complete 3D information, the laser scanner needs to be tilted up and down to sequentially scan the scene row by row to form a point cloud. Thus, this kind of actuated system clearly differs from cameras, as it *scans* the scene instead of simultaneously capturing the whole view.

Compared to other 3D measurement technologies, laser scanners generally have high resolution and long operating range up to hundreds of meters. Laser scanners are widely used in many high accuracy demanding applications, such as remote sensing [1, ch. 9.4] or in mobile robotics for localization and mapping purposes [6] [7]. Stoyanov et al. experimented by using actuated laser scanner to define ground truth data for comparison of other depth sensors [13]. They indicated that although being most accurate of tested sensors, actuated laser scan data still had rather large variance, mainly because of inconsistencies in timing of the samples and actuating the sensor.

However, actuated laser scanner systems have some significant drawbacks. First of all, moving parts cause many problems. Mobile parts are vulnerable to mechanical stress and they add weight and complexity to the system as mentioned above. Row by row scanning also increases the time needed to capture the whole 3D scene, making actuated laser scanners unsuitable for high-speed dynamic applications [7]. Recently, manufacturers such as Velodyne [36] has brought to markets more advanced multi-channel LRF sensors, which can scan up to 64 rows simultaneously and are more compact in size. These multi-channel sensors still suffer from poor vertical field of view, and although prices have been dropped lately, they are still rather expensive compared to recent ToF camera models.

2.5 Error Sources

The accuracy of depth measurement is limited by the errors added to the signal. Many different errors originating from various sources and ways to compensate them have been identified by various researchers [6, 7, 32]. Errors can be classified in to

two categories: systematic errors, which can generally be calibrated out, and random errors, whose presence cannot be predicted but which can filtered out.

Some most important error sources affecting ToF depth measurement are presented in this chapter. The focus will be particularly in the CW modulated technology, although most error sources concern both types of ToF cameras. For more detailed overview of errors, please see [7].

2.5.1 Systematic Errors

Circular error, also referred as *wiggling error* [7] originates from unideal modulation process. In calculations of CW method, the modulation signal is assumed sinusoidal but in reality it has irregularities, resulting in an offset that depends on the distance measured. Effect of wiggling error is shown in Figure 6. Offset follows a sinusoidal shape, thus the name circular error.

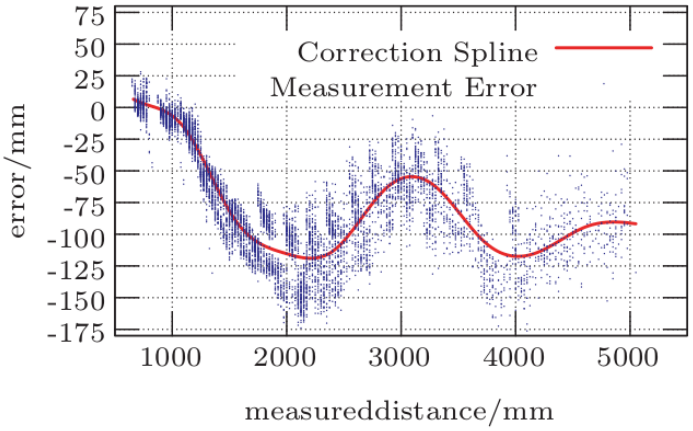

Figure 6: Circular error, measured with SR3000 camera. Blue dots are measurements, and red line is a polynomial model fitted to dataset for correction. [6]

Circular error can be compensated either by comparing a measurement to a reference ground truth, or by modeling the error using multiple measurements [7]. The first mentioned requires an additional sensor, and the one is more time-consuming as it requires large number of measurements.

Amplitude related errors result either from low amplitudes or overexposed pixels. Amplitude of the received light is closely related to the depth, so generally higher amplitude means better depth accuracy. There are a few main reasons to amplitude related errors. First, inhomogeneous light source causes different level of illumination in different parts of a scene. Usually the light power decreases when moving from center towards the edges, as seen in Figure 7. Lower amplitude values result to more distant measurements in the edges, although the object is flat wall. On the other hand, the center of the image seems to be closer because the pixels are

overexposed. Objects being near to the camera or too long integration times can cause overexposure, which in worst case can lead pixels to be saturated [7].

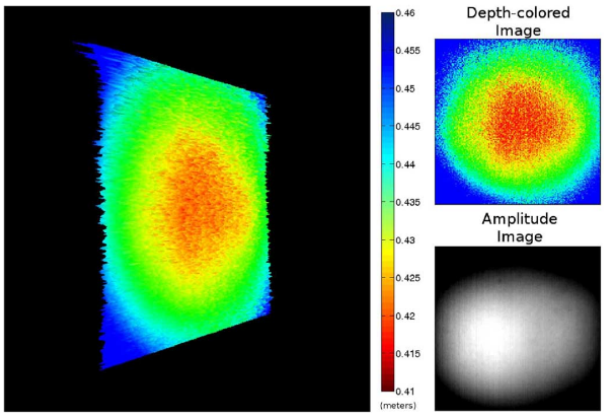

Figure 7: Amplitude related error due to inhomogeneous light. Object is a flat wall at 0.43 meters. [7]

Another cause for amplitude errors is different reflectivity of materials [6]. For example, while white paper reflects approximately 80% of the light at 850nm wavelength, black rubber tire reflects only 2% [29]. Low-reflective materials usually appear to be closer to the camera, and very high-reflective, mirror-like materials appear to be farther than it should be. If the scene includes multiple materials with very different reflection factors, depth measurements are invalid and distinguishing the shapes becomes difficult [34].

Temperature related errors exist because semiconductors technology, especially photodetectors used in ToF cameras, are highly sensitive to temperature changes [37]. When turned on, temperature starts gradually rising, causing drift to the whole depth image. Reports suggest that cameras should be let warm up to measurements to stabilize [7, 35]. For example, SR4000 camera measurements include a drift of a couple centimeters in the beginning, but error stabilizes to a few millimeters after waiting approximately 40 minutes. Warm up time depends on used integration time and distance of measured objects [35].

Also external temperature affects to the measurement. Kahlmann et al. measured object at different temperatures between -10°C and 40°C taking only single frames, so that camera did not warm up [37]. Their results showed that the temperature of the object clearly affects to measured distance, drift being around 8mm/°C.

Fixed-pattern noise (FPN) is related to imperfect manufacturing process of the image sensor. Each pixel has different characteristics of sensing light, causing constant offset [6]. Also the placement of the pixel in the chip has effect. Rows or columns of pixels are connected in series, which results to small signal delays in consecutive pixels when reading out data. FPN is always constant for each pixel, thus it can be calibrated out rather reliably.

2.5.2 Random Errors

Noise consists of several different sources, the most substantial being *photon shot noise* [32]. Shot noise limits theoretically reachable SNR and also the depth resolution. Knowing the statistical properties of shot noise, the absolute limit for *depth resolution* of CW ToF measurement can be derived as [32]:

$$\sigma_d = \frac{L}{\sqrt{8}} \frac{\sqrt{B}}{2A}, \tag{8}$$

where A, B and L are amplitude, offset and unambiguous range, as defined in (4), (5) and (7). From the equation it can be seen that a large offset, i.e. bright background illumination increases the noise. The only way to increase this accuracy is averaging [32]. However, using more frames decreases the frame rate, and if the object is moving between frames, averaging does not work. Also some filtering methods, such as median filtering, can be used to improve data quality, but it does not work in every case, e.g. in a very detailed scene [8].

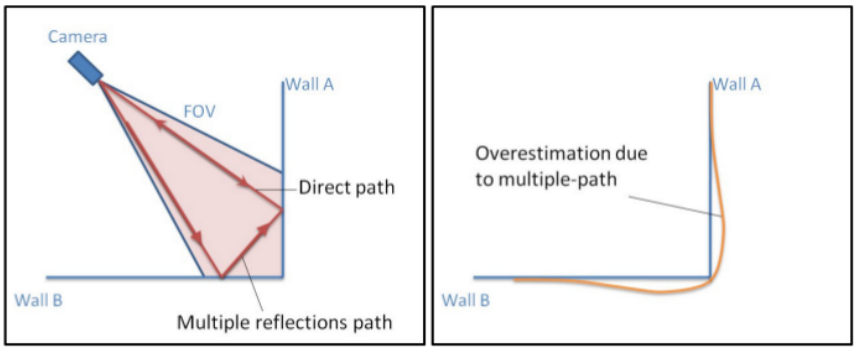

Figure 8: Multipath reflections on a corner. Due to interfering light waves, corner appears rounded. [29]

Multipath reflection errors appear because light reflects from multiple surfaces before returning to receiver. In corners, multiple light rays interfere and cause wrong depth measurement in the pixel, resulting in rounded-off corners as seen in Figure 8. Concave objects may reflect light in such way that it never reaches back to sensor, causing discontinuities called *jump edges* or *flying pixels* in the depth image. Several approaches have been proposed to filter out jump edges, such as locally comparing neighbouring pixels [6, 39].

Dealing with multipath reflections originated from corners seems to be still an open question. Several approaches have been proposed, but they either require modifications to hardware or involve additional equipment, such as an external light projector. Moreover, most implementations require so many samples that they cannot be used real-time [38].

Light scattering errors come from multiple light reflections between lens and image sensor inside the camera. As illustrated in Figure 9, some part of light scatters

to neighbouring pixels, resulting to depth errors. Scattering error occur only when very near objects are present in the scene [7].

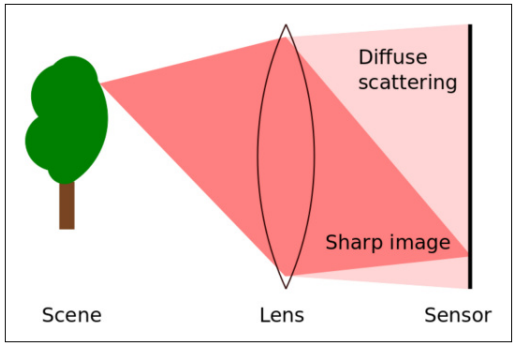

Figure 9: Scattering effect. Light scatters inside lens and spreads all over imaging sensor, affecting neighbouring pixels. [40]

Scattering error can be reduced by filtering with amplitude and phase data. Some results show that up to 90% of scattering error can be eliminated [40]. However, this approach requires access to raw data, which some camera models does not provide.

3 Experimental Setup

In this chapter, the three Time-of-Flight camera models selected for evaluation are briefly presented. After that, by combining theoretical aspects, earlier studies and the characteristics of the selected cameras, a set of test scenarios is defined to evaluate the performance of the cameras. The software used for the measurements is presented, and some prerequisites regarding the tests are discussed.

3.1 Evaluated Camera models

The goal of this thesis is to evaluate and compare the performance of three different Time-of-Flight cameras, which are Mesa SwissRanger 4000, Panasonic D-IMager EKL3106 and Microsoft Kinect v2. The cameras are shown in Figure 10. The reason behind that specifically these three cameras were chosen is not only that they happened to be available, but they also represent state of the art technology from a few last years, and therefore they can give good overview of how ToF cameras have evolved. Moreover, the properties of the cameras slightly differ from each other, and the impact of these differences in measurement accuracy is examined.

Figure 10: The three Time-of-Flight cameras evaluated in this thesis. From left: Mesa SR4000, Panasonic D-IMager EKL3106 and Microsoft Kinect for Windows v2.

SwissRanger 4000 (or simply SR4000) is a Time-of-Flight camera manufactured by company Mesa Imaging, and it was released in 2010. It has 176 x 144 pixel image sensor and operating range of either 5.0 or 10.0 meters, depending on model. Although there are separate models offered for different ranges, also different modulation frequencies appear to work with the model intended for 5 meter range, which is evaluated in this thesis. SR4000 offers adjacent modulation frequencies for multi-camera systems, which are 29/30/31 MHz for a 5 m range and 14.5/15.0/15.5 MHz for a 10 m range. There are also two undocumented frequencies of 10 and 60 MHz, which correspond 15 and 2.5 meter ranges respectively. Each model is calibrated for a specific range and measurements outside that range is not as good quality as specified, and therefore some options are not mentioned in data sheet.

SR4000 is widely supported, as official drivers and software are provided for Windows and Linux systems. It also has a couple of features that the other two cameras do not. Firstly, the integration time is configurable; it can be either set to a fixed value between 0.3 and 25.8 ms with 0.1 ms steps, or an automatic mode, where integration time is automatically adjusted depending on the amount of light reflected from the scene. Secondly, SR4000 provides additional confidence data for each pixel, which can be used for filtering out unreliable depth values.

In 2013, Mesa released their next model, SR4500. It is very similar to its predecessor, but features dust and water-resistant housing and improved background illumination suppression. Operating ranges and the size of the image sensor appear to be same in both models. At the moment of writing this thesis in mid-2015, both SR4000 and SR4500 apparently are still on the markets.

Panasonic D-IMager EKL-3106 was released in early 2012 and later discontinued three years after, during spring 2015. The product family has also two other models, EKL-3104 and EKL-3105, which differ by operating range and accuracy. The model evaluated in this thesis, EKL-3106, has more powerful LEDs for illumination, which allows it to work better in a bright background light. According to manufacturer, it can operate in background illumination up to 100,000 lux, which roughly corresponds to direct sunlight from clear sky.

EKL-3106 has an image sensor of 160 x 120 pixels, and the data sheet specifies the operating range to be from 1.2 to 5.0 meters when no background light is present, and from 1.5 to 5.0 meters when there is 100.000 lux background light. However, the camera outputs range data up to 15 meters, which was also confirmed to be actual operating range when testing the camera. The modulation frequency can be chosen from three options, allowing to use multiple cameras simultaneously without interference. However, there are neither mentions in the data sheet nor in the literature of what these frequencies are exactly.

A major limitation is that the drivers for the D-IMager cameras are provided only for Windows, and apparently no unofficial libraries for other operating systems, such as Linux, exist. Moreover, during the experiments it turned out that D-IMager works reliably only on a Windows 7 machine, as the camera driver did not work in Windows 8 for an unknown reason.

Microsoft Kinect v2 is the newest one of the evaluated cameras, originally released with the Xbox One video game console in the end of 2013, and later Windows-compatible version of the sensor was released for developers in July 2014. Unlike its predecessor, the first version of Kinect, the new version is based on CW modulated Time-of-Flight technology. It features an image sensor with resolution of 512 x 424 pixels, which is significantly larger than any earlier ToF camera model from competitors. Moreover, the price of Kinect v2 is one order of magnitude lower than previous camera models in the markets. The retail price of Kinect v2 is only $199, while the prices of SR4000 and D-IMager were several thousand dollars at the time of release.

Kinect v2 also has a built in RGB camera, microphones and cooling fan system. Color camera enables one to easily combine RGB images and depth images into colored point clouds, which might have advantages in some specific applications. On the other hand, presence of additional components makes the Kinect v2 more bulky and require more processing due to multiple data streams. The device requires an USB3.0 port from the host computer, which clearly limits its usage especially with older computers. There has been some reports that only some USB3.0 chipsets are supported, and there are also problems with USB3.0 extension cables, as not a single one of several tested cables did work with Kinect v2 [23].

According to Microsoft, Kinect v2 works in range between 0.5 and 4.5 m, but actually it is capable of measuring distances significantly longer than that. According to Breuer et al. Kinect v2 uses three samples with different modulation frequencies of approximately 16, 80 and 120 MHz, which results in 9.37 m non-ambiguity range [11].

The drivers, SDK and software for Kinect v2 are officially provided only for Windows 8, with no support for other operating systems. However, team of OpenKinect has developed libfreenect2, open source drivers for Linux systems [41]. These drivers gives access to functionalities that are not included in the official SDK, such as configuring the operating range [11].

Comparison of specifications

The evaluated cameras are all based on CW modulated ToF technology and their illumination units work in same wavelength region. However, they exhibit more or less notable differences in their specifications, which are listed in Table 1.

As it can be seen from the table, SR4000 and D-IMager are very similar in terms of image sensor size and field of view. Kinect v2 stands out from the three, as it has almost ten times more pixels and a FOV of nearly twice as large as the other two cameras. The differences in resolution and FOV are demonstrated in Figure 11, where three point clouds from the same scene acquired by different cameras are visualized. Kinect v2 clearly outperforms the lower resolution sensors in accuracy mostly due to better pixel to FOV ratio, as smaller pixels can capture smaller details.

Another factor affecting the accuracy is the data precision that sensors output. If looking at Figure 11 and depth data formats in Table 1, it can be seen that D-IMager clearly has the lowest data precision, which results in decreased depth accuracy of the measurements. Depth values of D-IMager are limited to range from 0 to 1500, and if considering the non-ambiguity range of 15 meters, one step increase in depth value corresponds to 1 cm in distance. For comparison, Kinect v2 can theoretically achieve approximately 1 mm accuracy and SR4000 even better than that. However, in reality, accuracies this high are generally not achievable, because of several error sources that cause noise to the data.

The difference in data formats is even more dramatic in intensity images. D-IMager provides only 8-bit precision, whereas SR4000 and Kinect v2 provide 16 bit

Table 1: Specifications of the evaluated cameras. Composed from user manuals and data sheets. [5][29][42]

	SR4000	**D-IMager**	**Kinect v2**
Released	2010	2012	2014
Image sensor size	176 x 144	160 x 120	512 x 424
FOV	43.6 x 34.6	60 x 44	70 x 60
Range (specified)	0.5 - 5.0 m	1.2 - 5.0 m	0.5 - 4.5 m
Max range (actual)	up to 15 m	15 m	~9m
Depth Resolution	± 10mm at 5m, ± 15mm at 10m	3cm at 0 lx, 14cm at 100 klx	N/A
Depth data	16bit (0-65535)	11bit (0-1500)	13bit (0-8096)
Amplitude data	16bit (0-65535)	8bit (0-255)	16bit (0-65535)
Dimensions (mm)	65 x 65 x 68	170 x 54 x 50	250 x 66 x 67
Weight	470 g	550 g	970 g
Connection type	USB2.0	USB2.0	USB3.0
Cooling system	passive	passive	Active
Power rating	12V, 0.8A (1.0A peak)	24V, 0.35A (2.5A peak)	12V, 2.67A

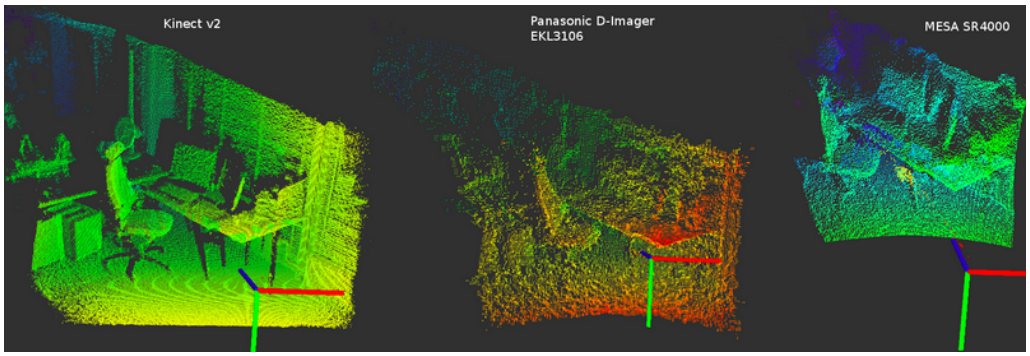

Figure 11: Point clouds from an office scene imaged with three different ToF cameras. The clouds are slightly rotated for better visualization, blue axis pointing to the direction away from the camera. The clouds are colored by the value of z-axis, colors scaled to the range between minimum and maximum value.

data. It is unknown how the precision of data affects depth measurements, as the sampling process and depth calculations are done directly on the chip of the camera.

The physical dimensions and weights of the cameras differ a little. Kinect v2 has an RGB camera and active cooling system, which adds weight and requires more space. SR4000 and D-IMager weigh almost the same, but the latter is significantly larger in width. This is because D-IMager has more powerful illumination system, as it is intended for high background light environment. As shown in Figure 12, the

LED alignments of each camera are very different. SR4000 has total of 24 LEDs uniformly aligned around the lens, whereas D-IMager has a total of 18 LEDs, placed in 3x3 grids to both sides of the lens. Kinect v2 has only three high power LEDs, and they are all placed to one side of the lens. If looking the power requirements of the cameras from the table, D-IMager draws significantly more power at peak than SR4000, and Kinect v2 lies between them. Considering the power requirements and number of LEDs, a wild guess would be that D-IMager has the most powerful illumination unit, Kinect v2 being second and SR4000 having the least powerful LEDs. The impact of different LED configurations is studied in the experiments.

Figure 12: The alignments of LEDs in the ToF cameras.

The operating ranges reported by the manufacturers are nearly identical for all cameras. However, all models support also ranges longer than specified. D-IMager operates in fixed 15 m range, but SR4000 and Kinect v2 can be set to work in different ranges by the driver software. The lower end of the operating range, i.e. minimum depth values that can be detected, depends highly on near-infrared reflectivity of the imaged material [21]. If the object reflects large amounts of emitted light, depth measurements tend to saturate and become invalid in near ranges. On the other hand, materials with low reflectivity can be perceived at lower distances, as less photons are returning to the sensor. The shortest measurable distances for each camera will be studied in the experiments.

3.2 Software

In the experiments the cameras were used with two different laptops, one with Windows 7 and the other with Ubuntu Linux 14.04, because the D-IMager camera had only Windows drivers. Operating all three cameras in the Windows environment was not an option, because the existing software applications for SR4000 and Kinect v2 are very limited compared to their open source counterparts on Linux. Therefore, D-IMager data was recorded on a Windows machine, and the other two cameras on a Linux machine.

For the D-IMager, a simple program for saving the image data from the camera was implemented using the API provided by Panasonic. It periodically reads the depth and intensity values from the camera and saves them to the disk as images

and point clouds. The interval of how often images are saved can be configured by the user.

In the Linux environment, Robot Operating System (ROS) [43] was used. It is a flexible framework that provides publisher-subscriber type messaging system as well as numerous libraries and tools for many purposes. The software modules in ROS are implemented as nodes, which communicate with each other through the master node that keeps track of nodes and message topics. There are existing ROS packages for SR4000 and Kinect v2 that read the image data from the camera driver and publish it to the ROS network as image messages [44, 45]. On top of existing software, a ROS node that subscribes to the image topics and periodically saves images and point clouds to the disk, was implemented. Additionally, the package called rviz, a 3D visualization tool for ROS was used for inspecting and visualizing images and point clouds.

All of the data processing was done in Linux environment. Figures from the data sets were generated using Matplotlib, an open source plotting library for Python [46]. It is very similar than Matlab, containing functions for reading and processing image files and plotting graphs. Since each of the cameras provide data in different format, range and intensity values were scaled to the common range so that they can be directly compared. Point clouds were read and handled with PCL, an open source point cloud library [47]. Recorded clouds were read from files and passed to a ROS node, where the data processing was done. The data extracted from the clouds was then plotted with Matplotlib, and the clouds itself were visualized in rviz.

3.3 Test Arrangements

3.3.1 Modifications to power cords

Power cords of the cameras were modified so that they were cut in half and RCA connectors were added to the both ends. Therefore the cameras can be connected either to an AC adapter or to a battery pack. This allows using cameras in places where there is no power outlet nearby, such as outdoors. The battery pack consists of two 12 V lead-acid batteries connected in series with a switch, so it is capable to provide either 24 V DC output for D-IMager or 12 V DC for SR4000/Kinect v2.

3.3.2 Interference

Because the evaluated Time-of-Flight cameras use active, continuously modulated illumination source and they work in the same wavelength region, interference needs to be taken into account. In preliminary tests with SR4000 and Kinect, a noticeable interference was observed when the cameras were facing to the same direction. Changing the modulation frequency of SR4000 showed no effect, as measurement errors were present with each frequency. Therefore, to be on the safe side, all measurements were done with a single camera at a time to eliminate errors related to interference.

3.3.3 Reference Data

To reliably compare the depth accuracy of ToF cameras, ground truth of the measured objects is needed. This is challenging, as no perfectly accurate method to measure objects and their relative distances exists.

However, considering the comparative nature of the experiments and the magnitude of errors in the ToF range measurements, resolution of a few millimeters is sufficient. This was obtained by using objects of known dimensions, measured with a tape measure or a ruler. Distances from the camera to the object were measured with a Bosch DLE 50 handheld laser rangefinder, which has accuracy of ± 1.5 mm according to the manufacturer. For each distance, several measurements were taken and their mean was used as a reference value.

The origin of the camera's coordinate system is in the middle of the image plane, which lies inside the camera. This is also the point that the distance measurements are related to. Therefore, when measuring the object distance from the front panel of the camera, a distance offset has to be applied. It was assumed that the origin of each camera is at the position of the hole for camera mount. The distance from front panel to the center of the hole was measured to be approximately 28 mm for SR4000, 26 mm for D-IMager and 31 mm for Kinect v2. However, Fankhauser et al. reported that based on their tests, true offset for the Kinect would be 28 mm [21]. This was confirmed in our initial tests. Therefore, an offset of 28 mm was used for SR4000 and Kinect v2, and 26 mm for D-IMager. These values were added to the measurement obtained by the laser rangefinder.

In some experiments, where other factors than absolute distance accuracy are studied, exact dimensions of the objects are not necessarily needed. In these cases, multiple data sets were recorded and one could be used as a reference. For example, when studying how bright sunlight influences to the distance measurements, a data set recorded during overcast weather was used as a reference.

3.3.4 Sample Averaging

Due to random errors, ToF depth images include noise. Averaging over multiple consecutive frames is effective way to improve data quality, as the impact of random errors, such as shot noise, decreases.

Lachat et al. investigated the effect of averaging over consecutive frames with Kinect v2 [22]. They analysed pixel-wise mean values and standard deviations for several different sample sizes to determine which is optimal for improving depth accuracy. They found out that averaging over 50 frames gives the best outcome; The result is smoother than with smaller sample sizes, and precision does not significantly change any more if the number of frames is increased. Therefore, a sample size of 50 frames was used in the experiments of this thesis.

4 Experiments

4.1 Warm Up Test

As mentioned before, Time-of-Flight camera starts to heat up right after it is powered up. The temperature gradually increases, until it stabilizes after a brief period of time. The depth measurements are unstable during this warming up period, error being up to several centimetres. The purpose of this test is to determine how the increasing temperature affects the measurements, and how long it takes for each ToF camera to warm up and measurements to stabilize.

4.1.1 Setup

In this test, a stationary object with a fixed distance is imaged for a specific amount of time. Only the temporal change of distance measurements was observed, so all other parameters had to remain constant during the process. This was done by placing a ToF camera on a tripod, pointing to a planar target. A white 120 x 90 cm paperboard was used as a target. The test was repeated several times and recording was done with only a single camera at a time. Ambient temperature in the room during the measurements was measured to be approximately 23-24 °C.

The distance to the target was chosen to be 1.2 meters, because it is the shortest specified distance for D-IMager. However, at this distance, some pixels in the center of D-IMager appeared to saturate and therefore give erroneous depth values. Therefore, test was also repeated with a longer distance of 3 meters. Exact reference distances measured by laser rangefinder were 1.19 and 3.0 meters from the front panel of each camera.

Duration of the measurement and frequency of capturing images were also considered. In previous studies, nearly all authors have used a 60 minute recording time with image acquisition interval varying between 1 to 20 seconds [10][11][20][22][35]. Their results suggest that SR4000 needs 30 - 40 minutes to stabilize, while Kinect v2 needs approximately 30 minutes. For D-IMager, there is no data available. To see if depth measurements remain stable for a longer period of time, the time of measurement was decided to be extended to two hours (120 minutes). One frame was saved every 10 seconds, totaling 720 frames per data set.

Before measurements, cameras were kept unplugged from power source for at least several hours to completely cool down. When turned on, recording of the data was immediately started. To ensure consistent heating, all cameras were operated in continuous mode so that they captured 30 frames per second during the whole measurement time. One sample was saved every ten seconds and other frames were discarded. This way, the camera would not cool down between frames, as it would do if the capturing was paused after each measurement [10].

Some of the data sets were recorded during evening hours when it was dark, and office lights with motion detector sensor switched on and off as someone moved nearby. These changes in the background lightning potentially caused error to

the measurements. However, when inspecting preliminary results, changes in background lightning did not seem to have any impact as the fluorescent tubes do not emit significant amounts of near infrared light that ToF cameras could detect.

4.1.2 Results

The results of the first warm up tests at 1.2 m distance are shown in Figure 13, where both the center pixel values and the mean values of 10 x 10 pixels ROI are plotted. Averaging over multiple pixels on a planar target is beneficial, as the variations of the mean distances over time are significantly smaller than single pixel values.

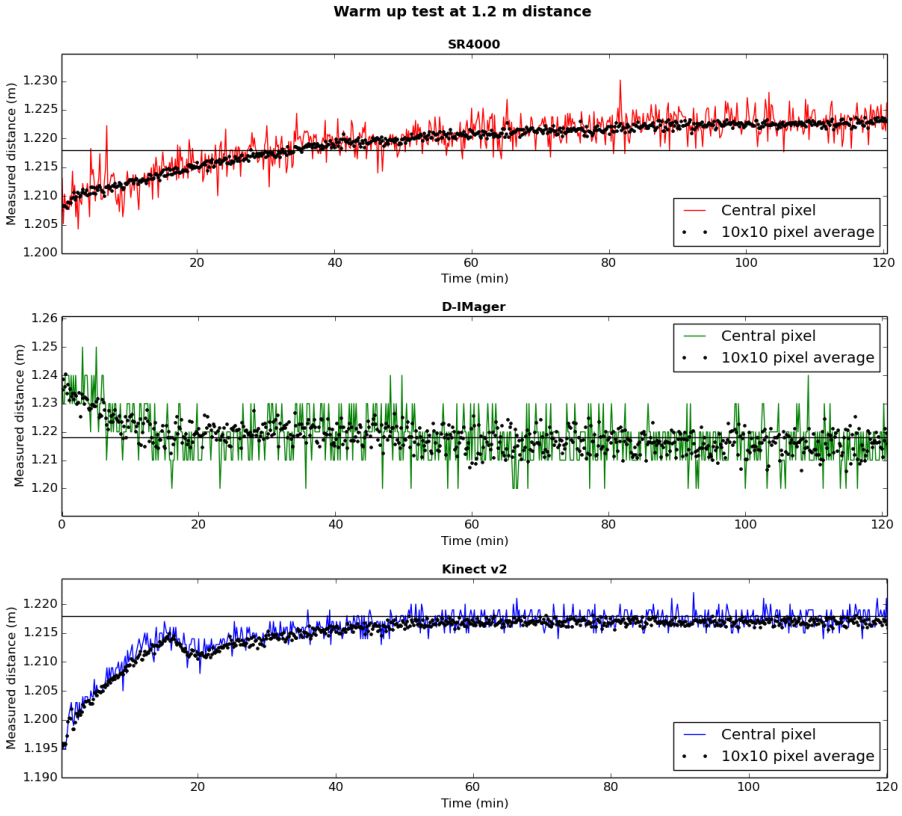

Figure 13: Warm up test results at 1.2 m distance.

SR4000's curve is different than expected, because the warm up time is significantly longer than 30-40 minutes that other authors have reported [10, 35]. In this case the depth values keep increasing even after 60 minutes. This might have something to do with automatic integration time, which will be discussed later.

D-IMager behaves contrarily to the other ToF cameras, as it outputs too high depth values when it is cold. Stabilization time is rather quick, as the values settle down in approximately 15 minutes and remain at nearly constant range of $\pm 1cm$. The poor precision of the D-IMager's depth data can be clearly seen, as the center

pixel values vary in discrete 1 cm intervals. The initial drift is approximately 2 - 3 cm, which corresponds to roughly 2% of the measured distance.

The Kinect's curve is as expected, nearly identical to earlier results [11, 22]. First, the depth values start to rise immediately after powering up. At approximately time T=16 min, the curve suddenly starts to descend for a while, after which it starts to gradually ascend again until it reaches the final value at approximately T=50 min. The spike is caused by the cooling fan, which appears to start always at a same time. After 50 minutes the center pixel values fluctuate within $\pm 3mm$ and the 10x10 pixel averages within ± 1 mm.

Next, the results from 3 m distance are considered. The results are expected to differ slightly from the 1.2 m distances, because the scene is different as only a part of the center region of the image is covered by the object. As the warm up rate of the sensor depends on the amount of received photons, less reflected light from more distant objects should extend the warm up time. The results from the 3 metres distance are shown in Figure 14.

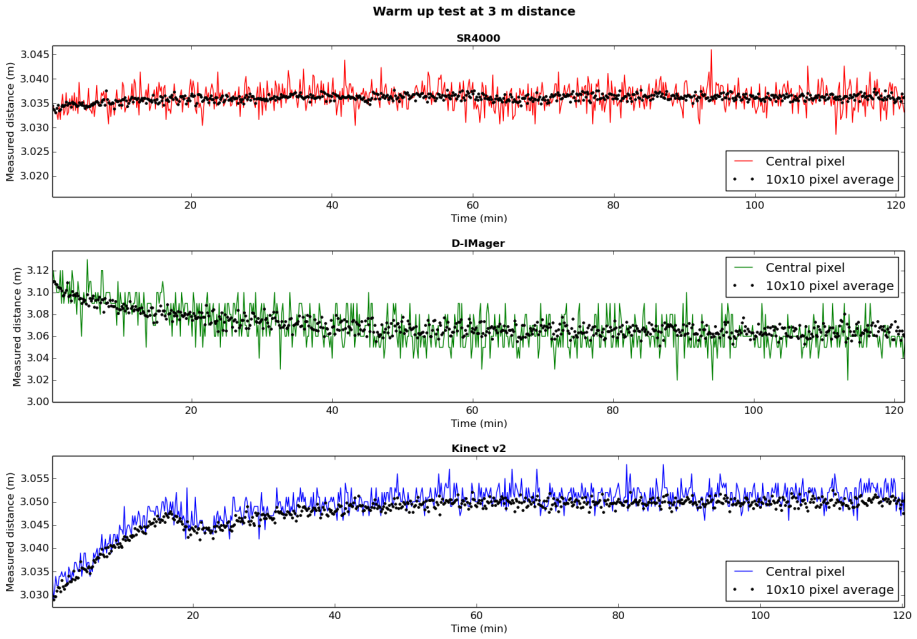

Figure 14: Warm up test results at 3 m distance.

Now the behavior of SR4000 differs even more from the results reported in the literature. The distance values seem to be rather stable from very beginning of the measurement. The initial depth drift is only 3 mm which is only a fraction of the corresponding at the 1.2 m distance. Moreover, the drift diminishes after 5-10 minutes, after which the values fluctuate within $\pm 5mm$ (center pixel) and $\pm 1mm$ (100 pixel average).

As expected, the D-IMager takes significantly longer time to stabilize than at shorter distance. The warm up time is approximately three times longer, as it takes

40-50 minutes to stabilize. Initial depth error is approximately 6 cm, which is 2% of the distance. After stabilising the center pixel values stay within $\pm 2cm$ and averages within $\pm 1cm$, which is half less accurate than before. This is expected result, since the accuracy decreases at longer distances.

Kinect behaves almost in the same way than at 1.2 m: during the first 16 minutes period the depth values increase by 2 cm at rather constant rate, after which the cooling fan starts and values decrease slightly. Then the values again start to increase slowly until at approximately 50 minutes they settle. The final values vary within ± 5 mm (center pixel) and ± 2 mm (100 pixel average). The active cooling system can probably regulate the internal temperature better than the passive systems, and the warm up time is consistent in different measurements.

Because of SR4000's strange behavior with the automatic integration time, the warm up test was repeated at 1.2 m distance also with two constant integration times. The values for the IT were chosen to be 30 and 100 from the range 0 - 255, which correspond 3.3 ms and 10.3 ms times respectively. The results with the fixed integration times in contrast to automatic setting are shown in Figure 15.

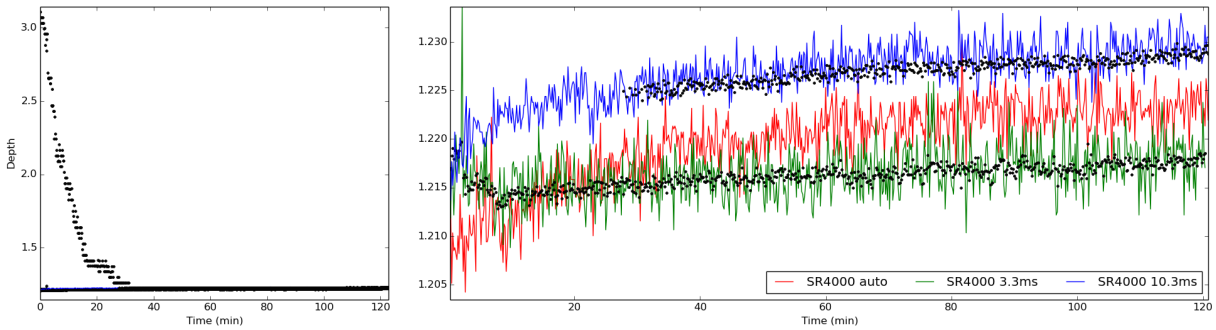

Figure 15: Warm up test result with different different integration times. Graph on the right is a zoomed in version of the left one.

The results are interesting. With an integration time of 3.3 ms the process of warming up appears to slow down even more. Actually, the depth values even decrease during the first few minutes, after which they start to gradually increase by 2 mm over the 120 minutes time period, which is considerably slower rate than with the other settings. The depth values of the center pixel fluctuate within a range of $\pm$ 4 - 5 mm, which slightly more than with automatic IT. This indicates that the integration time might be too short, because the sensor does not warm up properly and its accuracy decreases.

Also 10.3 ms integration time causes strange behaviour. If looking at the left part in Figure 15, it can be seen that the 10x10 pixel averaged values are very unstable during the first 30 minutes, after which they settle down. The root of the problem was found to be saturated pixels in the amplitude images. In Figure 16 three raw amplitude images of SR4000 from different moments of time are shown. At first

there are quite many saturated pixels in the center region of the image, but they start to disappear as the camera gets warmer. After approximately 30-40 minutes the saturated pixels are all gone.

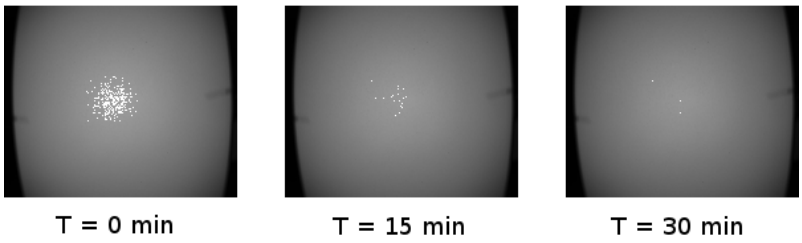

Figure 16: Saturated pixels in the amplitude images of SR4000. The amount of the saturated pixels decrease over time, as the sensor warms up.

Lastly, behaviour of the amplitude values during the warm up test is examined. Figure 17 shows temporal changes of amplitudes with different ToF cameras and settings. All values are scaled to a common range, between 0 and 1.0. In the leftmost graph, the effect of automatic integration time is clearly visible, as the amplitude constantly rises and falls between certain values. The changes however are not very big, only 2 - 3%.

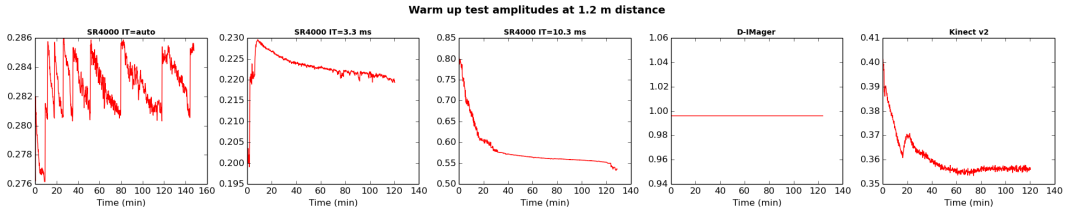

Figure 17: Temporal change of amplitudes at 1.2 m distance.

The second and third amplitude graphs in turn explain the behaviour of SR4000 with constant integration times. With the 3.3 ms IT, amplitude quickly increases by 15% during the first several minutes, which was seen in Figure 15 as unstable depth values. The 10.3 ms curve on the other hand experiences rather steep decrease, as the amplitude drops by 40% within 30 minutes. A closer look reveals that the initial value, 80%, corresponds to approximately 52400 arbitrary units used by SR4000. According to user manual [29], amplitude values above 32767 (50%) indicate saturation, which in part explains the invalid depth values. However, on this basis the pixels should have been saturated during the entire two hour measurement period, but reasonable values were still obtained after 30 minutes.

The D-IMager's amplitude remains constantly saturated for the entire measurement. This is reasonable, as the 1.2 m distance that was used in the experiment is below its specified operating range. Nevertheless, D-IMager is still able to provide stable depth measurements. The Kinect's amplitude curve is nearly identical to its corresponding depth curve, but only a mirror image relative to x-axis.

To summarize, the warm up period is important for the ToF cameras, as all tested devices experience a significant drift in measurements immediately after powering up. When the sensor is cold, pixels even tend to saturate and provide inconsistent values, as it was in SR4000's case. The time needed for the ToF camera to warm up and stabilize appears to depend on the amount of received light. Thus, the details in the imaged scene, such as the distance and NIR reflectivity of the object, most likely have an impact. Therefore, in subsequent experiments the cameras are let to warm up at least 60 minutes before measurements to ensure stability.

As these tests were carried out in at constant temperature, the results are rather clear. An interesting follow-up for this experiment would be to study how the measurement drifts change, if the ambient temperature was changing during the measurement. Moreover, in outdoor environments also other temperature affecting factors are present, as the wind potentially cools the sensor, whereas the heat of the sun causes excessive warming. Therefore, the warm up test should be repeated in dynamic conditions to study how well the ToF cameras can adapt to changing temperature.

Additionally, even longer measurement times than two hours should be considered because the amplitude and depth values do not completely settle down in that time. Although the fluctuation is rather small after some time, it would be interesting to see how the values behave in a long term.

4.2 Distance Test

The distance to the target affects to the measurement in two ways. First, because the intensity of the light is inversely proportional to the squared distance, fewer photons are returning to the image sensor and the SNR of the distance measurements decreases at longer distances. Secondly, due to differences between actual modulation signal and idealized model used in calculations, a systematic distance offset called "wiggling error" (see Chapter 2.5) occurs. To study how these two error sources influence to depth accuracy, the distance test was carried out.

4.2.1 Setup

In this test, all parameters except distance should remain constant. The same white paperboard that was used in the previous test was used as a target also in this test. It was taped to an easily movable wheeled cabinet, whereas the cameras were in fixed position. The distance range was chosen to be from 0.5 to 5 meters, because it covers the specified ranges for all three cameras. The measurements were done in 10 cm intervals, and 50 frames were recorded at each distance.

Since the cameras should stay in a constant temperature, they were switched on one hour beforehand, and the power was kept on during the whole measurement process. Because the cameras would interfere with each other, only one camera could be used at a time and the two others were turned to face away from the target.

Because the cameras were constantly swapped, camera placements and reference measurements had to be done with special care. The measurement scheme is illustrated in Figure 18. A drawer with a width of 60 cm was used as a platform and cameras were placed so that the front panel of the camera and the edge of the drawer were parallel. Each time when the target was moved to next point, the reference distance was determined with a hand-held laser rangefinder. To ensure the target being perpendicular to the optical axis of the camera, measurements d_1 and d_2 were taken from two edges of the drawer. The position of the target was fine tuned until the measurements from both sides were within $\pm$ 1 mm from the desired distance (e.g. 1199 mm and 1201 mm were accepted for 1200 mm). Using trigonometry, it can be calculated that the angle between the optical axis of the camera and the normal of the target was no more than 0.2 degrees. Finally, offset values for each camera were added to the reference distances, as described in Chapter 3.

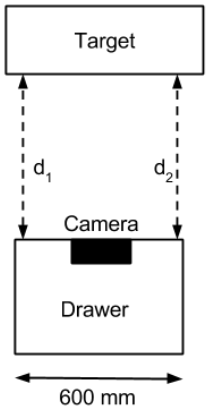

Figure 18: Illustration of the measurement setup in the distance test. Reference distances d_1 and d_2 were measured using a laser rangefinder.

Now that the temperature, the angle of the target and the background lighting are considered, only remaining uncontrollable variable is reflections from other objects. Particularly with the longer distances when the target does not cover the entire image, unwanted reflections are possible. A plain wall would have been better for this test, but because no such premises with enough free space were available, a smaller target had to be used. Figure 19 shows an example frame from Kinect at 2 meters distance, where the target is clearly reflected from the floor. Additionally, there are cabinets on the both sides of the target and a glass wall behind the target at approximately 8 meters distance, which could cause errors. After some manual inspection of the frames, it was concluded that the impact of the possible multipath reflections for the center pixel was small.

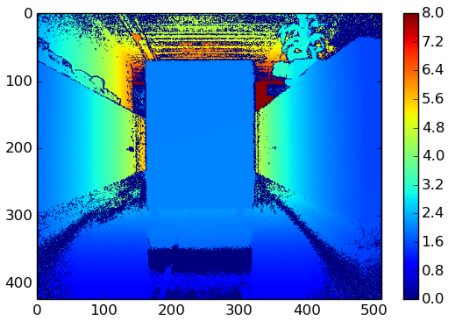

Figure 19: Example depth image captured using Kinect v2 at 2.000 m distance

4.2.2 Results

Initial results of the distance test are shown in Figure 20, where the measured distance values of each ToF camera's center pixel are plotted against reference distances. Some interesting features marked with circles: a) and b) indicate erroneous distance values caused by overexposure, and c) depicts measurements mapped to wrong distance due to ambiguous phase.

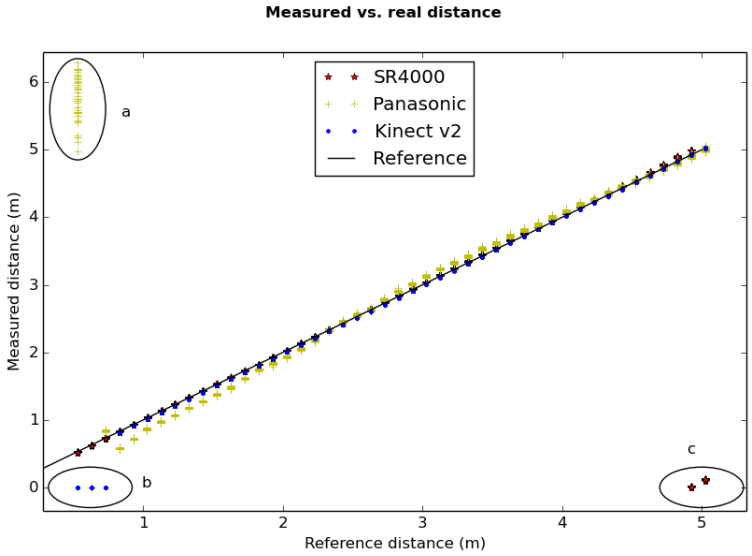

Figure 20: Results of the distance test. Highlights a) and b) indicate incorrect depth values due to pixel saturation, and c) indicates ambiguous values mapped to wrong distance.

The measurable operating ranges differ slightly from specified. SR4000 works as expected, since it gives approximately correct values through the entire measured range. At 4.93m some of the values are incorrectly mapped to close to zero, and at 5.03 all of the values are wrapped back to the lower end of the range. This is reasonable, because used modulation frequency 30 MHz limits measurable values to

5 meters.

D-IMager works somehow even at lower distances than specified 1.2 m; the measurements between 0.7 and 1.1 meters are fairly correct, although offset errors are rather large. Distances below 0.7 m are problematic, as the output values are very unstable. At reference distance 0.53 m the measurements are scattered between 5.0 and 6.5 meters, which is one order of magnitude greater. On the other hand, at 0.63 m reference distance, all values are suddenly zeros. Whereas zero distances can be safely classified as invalid, non-zeros might be challenging to distinguish between real and incorrect.

Kinect v2 outputs only zero values below 0.83 m, although it should work down to 0.5 m according to Microsoft. The used target has fairly high reflectivity, which might explain the difference in results. As the Kinect v2 is mainly targeted to gaming applications, values reported by the Microsoft probably apply for average home environment, where there are no highly reflective materials present. Nevertheless, bad pixels can be filtered out reliably, because saturated distance values are consistently zeros.

To take a closer look at the results, reference distance values were subtracted from the measurements. Resulting distance errors are plotted in Figure 21. The wiggling shape of the systematic depth error is clearly visible in all graphs. However, in the Kinect v2 graph at distances below 2.5 m there are some anomalies, which might be caused by errors in the reference distance measurements. The measurement process is prone to errors, as it involved plenty of manual work and moving the cameras. It is also worth noting that the wiggling error of D-IMager is one order of magnitude higher than the other two cameras: Its depth offset at different distances varies between -18 and 14 cm, whereas the measurements of SR4000 and Kinect v2 remain within a few centimetres from the reference distance. This significant difference could be explained by D-IMager's lower modulation frequency, which also decreases the range accuracy.

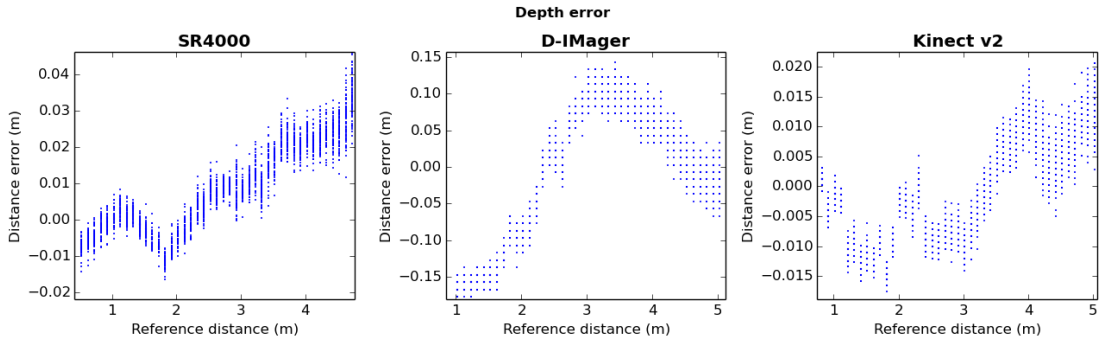

Figure 21: Depth measurement errors of the central pixel.

Due to the systematic nature of the wiggling error, it can be calibrated out as described in Section 2. However, the correction spline should be calculated separately for each pixel, because they have individual responses. In Figure 22 are illustrated

least squares fitted error functions for three different pixels of SR4000. The offsets between nearby pixels vary by a few millimeters, but the shapes of the splines are very similar. Therefore, relatively good results could still be achieved by using a single calibration spline for the whole image. Similar behaviour was observed also with the other two sensors; there were minor differences between individual pixels, but the overall shape of the error curve was same.

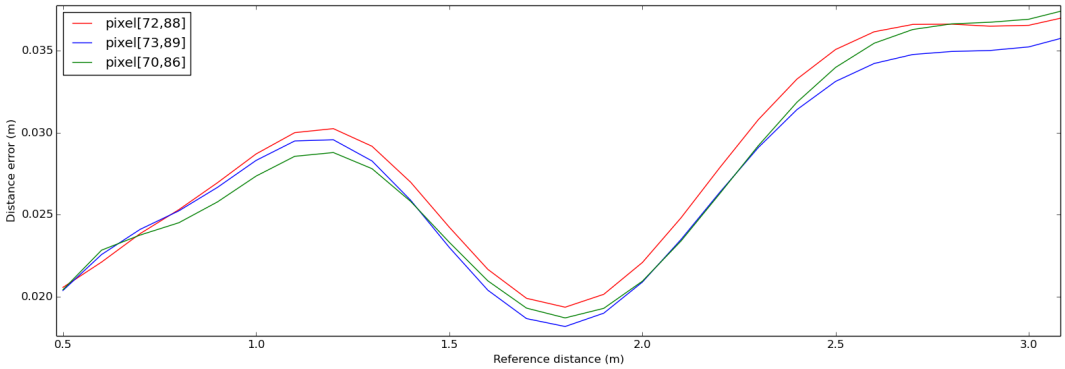

Figure 22: Least square fitted correction splines for three different pixels of SR4000.

In addition to the systematic distance error, also random errors depend on distance. One measure for the accuracy of ToF camera is the deviation of data points at each distance. In Figure 23 the calculated standard deviations of distance errors for 50 frames at each distance are presented. As it can be seen from the graph, the range accuracy generally decreases as the distance increases, which is explained by the decreasing SNR, as less photons are returning back to the sensor from distant objects. Kinect performs best in the entire range with the standard deviation varying between 1.0 and 3.9 mm. SR4000 is a runner-up with accuracy varying between 1.5 and 4.5 mm, except the distances greater than 4.5 m, where the measurements become unstable, as the standard deviation quickly rises because of phase wrapping. D-IMager differs significantly, as the deviations of measurements vary between 6 and 22 mm, which is nearly one order of magnitude greater than the other two cameras.

When comparing to the values listed in Table 1, the results meet the requirements rather well. The accuracy of SR4000 is specified as tolerance of ± 10 mm at 5 m distance for an object with 99% NIR reflectivity. When examining closer the raw measurements, there are a few individual data points that do not fit into specified tolerance range, but it can be explained by reflectivity of the object, which is lower than 99%. The accuracy of D-IMager is specified as $\sigma = 3$ cm when there is no interfering background light present. This condition is satisfied in a whole operating range, as the maximum deviation is 2.2 cm at 4.9 m distance. There are no such accuracy values specified for the Kinect v2, so no comparison can be done.

Next, the amplitudes relative to the distance are examined. Amplitudes relative to the distance are presented in Figure 24. The blue curves in the subplots repre-

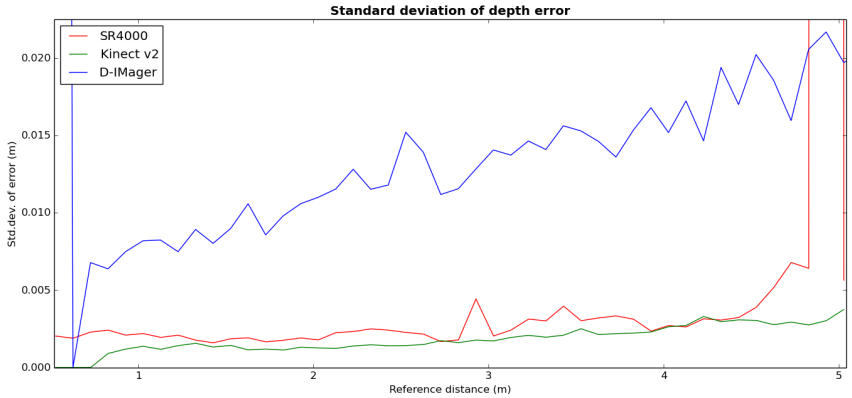

Figure 23: Standard deviations of the distance errors. Note that the peaks at unstable ranges (0.6 m for D-IMager and 4.8 - 5.0 m for SR4000) are omitted.

sent the mean amplitude values of the central pixel, averaged over 50 frames. All amplitude values were scaled to between 0 and 100 so that different values from different cameras could be compared directly. Because the light intensity decreases with the distance, the amplitude deviations at different distances are not comparable. Therefore, the coefficient of variation (CV) was calculated for each distance. It is defined as the ratio of the standard deviation to the mean:

$$CV = \frac{\sigma}{\mu}, \tag{9}$$

where σ is the standard deviation and μ is the mean value. CV values are also plotted in Figure 24.

As it can been seen from Figure 24, SR4000 behaves completely differently from the other two cameras. This can be explained by the fact that automatic integration time setting was used. At longer distances, when the target does not completely cover the view, reflections come also from other objects. SR4000 then tries to adjust integration time based on average objects in the scene. As the target is moved further away from the camera, less reflected light returns to the sensor and the integration time is increased, which results to higher amplitude value. Unfortunately the camera did not provide integration times used in each frame, so this phenomenon cannot be further examined. On the other hand, D-IMager and Kinect appears to obey the inverse-square law, i.e. for a point light source, the light intensity is inversely proportional to the square of the distance. Hence, it can be concluded that both cameras use constant integration time.

An interesting observation in the D-IMager's amplitude curve is that even though the pixel is saturated between 0.7 and 1.5 meters, the distance values in the respective range are reasonable. Same does not apply to Kinect, as it outputs only invalid distance values when the amplitude of the pixel is saturated between 0.5 and 0.7 m.

The CV roughly indicates the quality of amplitude measurements, the lower

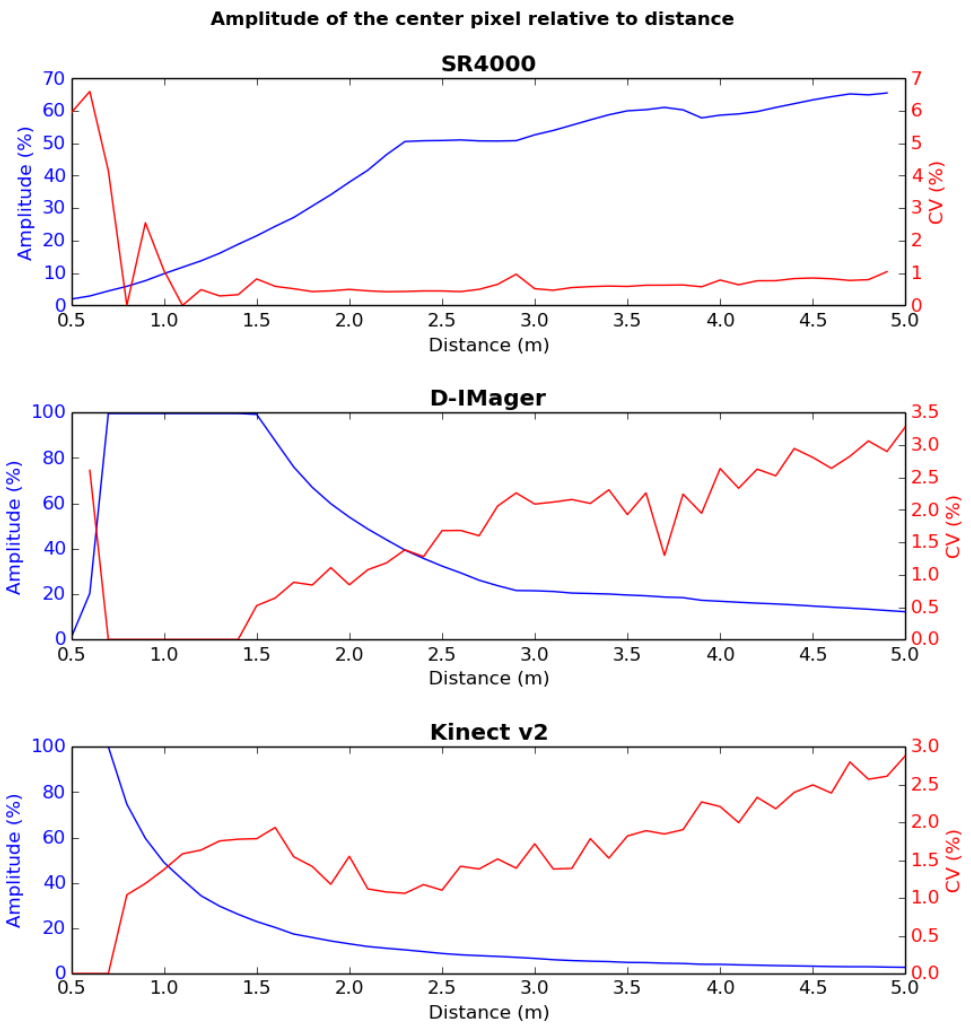

Figure 24: Amplitudes averaged over 50 frames (blue curves) and corresponding coefficients of variation (red curves).

value being better. When the amplitude decreases, SNR also decreases accordingly, which results to higher CV. Because of the automatic integration time, SR4000's CV remains rather stable between 1.2 and 4.8 meters. However, below 1.2 meters the behavior is very unstable due to very low amplitude values, as the lower mean values have more impact on the result. Moreover, it seems that SR4000 has problems with choosing a right integration time. On the contrary, with the constant integration times of D-IMager and Kinect, CVs appear to be increasing roughly in the same proportion as the amplitude drops.

The correlation between the amplitude CV values and the depth accuracies is illustrated in the scatter plots in Figure 25. Distances with unstable behaviour (null values from saturated pixels and phase-wrapping region of SR4000) were left out. As it can be seen from the plots, there is rather strong correlation with the

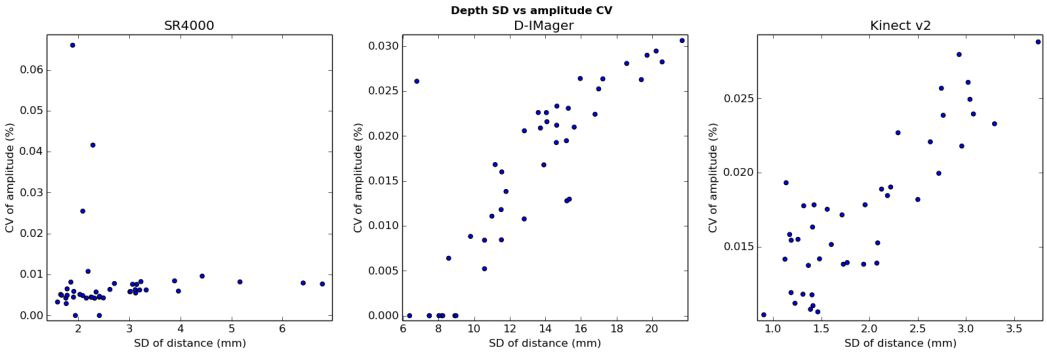

Figure 25: Standard deviations of depth measurements versus coefficients of variation of amplitude.

variables when constant integration time is used. Calculated correlation coefficients for the scatter plots are R=0.86 for Kinect and R=0.85 for D-IMager. Instead, the SR4000 appears to have almost no correlation, as the correlation coefficient is as low as R=0.05. However, if the outlier points at top-left and bottom-right regions of the scatter plot (corresponding distances lower than 1.2 m and greater than 4.5 m respectively) are left out, the correlation coefficient increases to 0.68, which means moderately high correlation.

This information about the correlation between the amplitude and depth accuracy might be used to predict the reliability of the ToF measurement. However, the usage of CV values is impractical in real-time applications, as it requires averaging over multiple samples. But, as mentioned before, the CV curve roughly obeys the complement of amplitude, which means that also the plain amplitude values could be used for analysis. This obviously does not apply for automatic integration time, as the amplitude values change with integration time.

Suggestions for further work include repeating the test with different modulation frequencies (cameras that allow it to be configured, e.g. SR4000 and D-IMager) to see how the calibration spline changes. Also longer distances should be considered, since all three cameras used in this thesis work also in longer distances, up to three times longer than specified by the manufacturer. Interesting aspects would be to study how the calibration spline progresses, or how the depth accuracy changes with the longer distances.

4.3 Illumination Unit Test

The quality of the ToF distance measurements highly depend on the amplitude. In the previous test, amplitudes of the center pixels were measured relative to the distance. This time the cameras are kept stationary, and the focus is in the variation of the light intensity in the different parts of image area. It is known that the light power, thus also the received amplitude decreases when moving towards the edges from the center of the image, causing distance errors.

The three evaluated cameras have very different LED alignments and powers. The purpose of this test is to examine, do the different LED configurations have effect to the distance measurements, and if so, how much do they affect? Additionally, advantages and disadvantages of different lighting unit designs are considered.

4.3.1 Setup

Following procedure was done separately with each camera: first, the camera was placed on a tripod, facing towards a white wall. To minimize the effect of unwanted reflections, the tripod was placed on a table so that the camera was equally distant from floor and ceiling. The position of the camera was then adjusted manually, so that it would be as perpendicular as possible. This was done with the help of the visualization tool by looking at the amplitude-colored point cloud and simultaneously fine tuning the angle carefully. The angle was considered sufficient, when the inhomogeneities of the light pattern appeared symmetric. After the adjustment was done, the camera was kept acquiring images for at least one hour, so it would properly warm up. Finally, 50 frames were recorded.

With the results of the previous test and some experimenting, the distance was chosen to be approximately 1.50 meters. This is because the pattern from the LEDs was best visible, and none of the cameras was saturated. Longer distances were excluded because of problems with multipath reflections that arise from other surfaces. The FOV of the Kinect v2 is so large, that reflections from the ceiling were observed even at distances below 2 meters.

4.3.2 Results

First, to see how the amplitude changes in different parts of the image, averaged amplitude images were created from the data sets of the each camera. The 50-frame average images are shown in Figure 26. The images are colored with a heat map, so that the variations in the values are clearly visible. For the sake of clarity, the limits of the color maps were slightly adjusted so that the extreme values in both ends of the scales were smoothened.

Both SR4000 and D-IMager have very similar round shaped light patterns, whereas Kinect's pattern is more of a rectangular shaped. This is somewhat surprising, because it was presumed that the LEDs arranged on both sides of the D-IMager would form a more elliptical pattern. This indicates that the vignetting in the corners of the image is not necessarily caused by illumination, but other factors instead. Kinect's light pattern is also interesting, as it is created with only three diffused laser diodes, whereas the other two cameras use LEDs.

The differences in the amplitude values are fairly high. Each camera has its highest values in the center region, the maximums being 56%, 89% and 23% for SR4000, D-IMager and Kinect respectively. When compared to the results of the wiggling error test at 1.5m distance (see Figure 24), corresponding values are very similar: 22%, 22% and 99%. For the cameras with constant integration time the

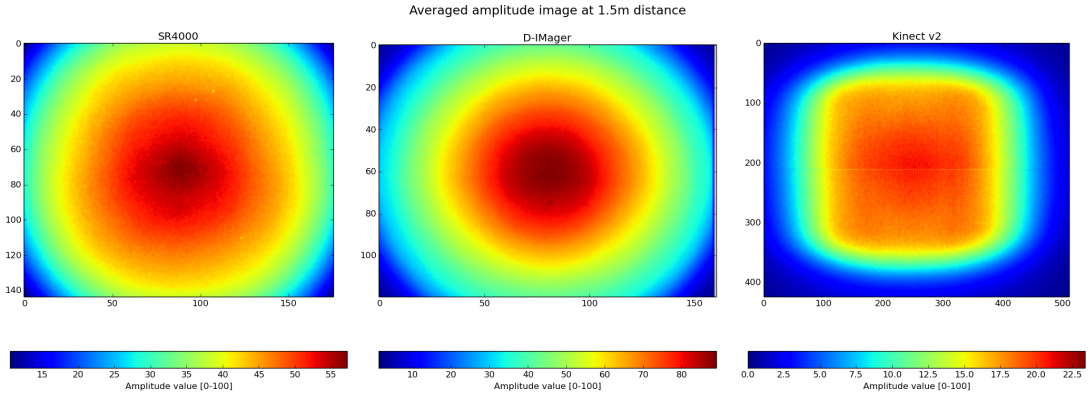

Figure 26: Averaged amplitude images.

values are consistent, but in the case of SR4000, the difference is major. It can be explained by the differences in the scene, as now there was only a flat wall covering the entire image area, whereas in the earlier test there were also distant objects present. Since the integration time is determined according to averaged returning light in the image, integration times might differ.

In all images, the amplitude decreases significantly towards the edges of the image. The drop is most significant with D-IMager, as the amplitude values vary from 89% in the center pixel to as low as 2% in the corners, which means a decrease of 87 percentage points (pp). The equivalent difference between the maximum and minimum amplitudes for SR4000 is 47%, and for Kinect it is only 22%.

Whereas the amplitudes of SR4000 and D-IMager decrease rather uniformly towards the edges, Kinect has an approximately 300 x 300 pixel area at the center region where the amplitude remains relatively high, until it quickly decreases near the border region. Considering the FOVs and the sizes of the image sensors, the bright area in the center of Kinect's image covers approximately same sized area from the scene as the entire image planes of the two other cameras. Therefore, Kinect clearly outperforms its competitors in the uniformity of illumination.

Next, depth accuracy in different parts of the sensor is considered. The averaged depth images from 50 frames are presented in Figure 27. As it can be seen from the images, there are significant depth distortions in the edge regions of SR4000 and D-IMager, but Kinect's errors are small. This is most likely because the open source driver for Kinect uses calibration parameters by default, whereas the other cameras output raw data from the sensor, which does not include correction for lens distortions. In addition, mean depth values do not depict the accuracy of the sensors very well because of systematic errors. So that the accuracy of the cameras could be analysed better, the standard deviations of depth values in each pixel are examined instead. They are illustrated in Figure 28. Because the deviations increase rapidly in the corner areas, the colormaps of the images were adjusted for clarity so that

the values above 99th percentile were truncated.

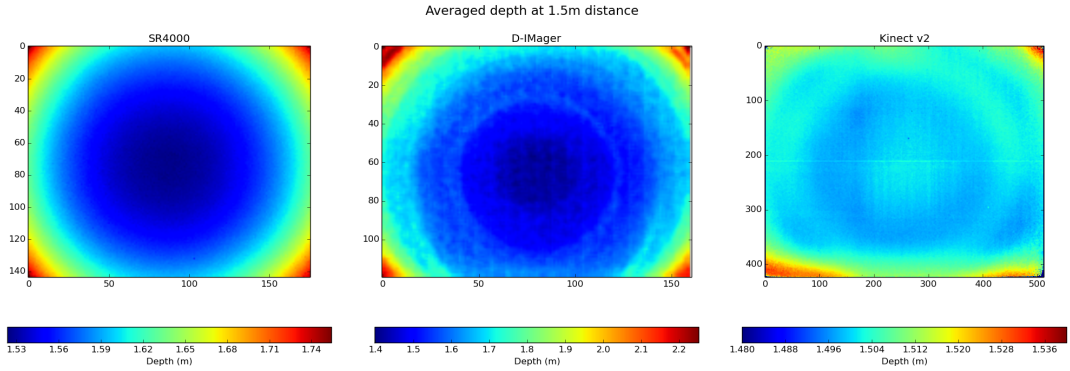

Figure 27: Averaged depth images.

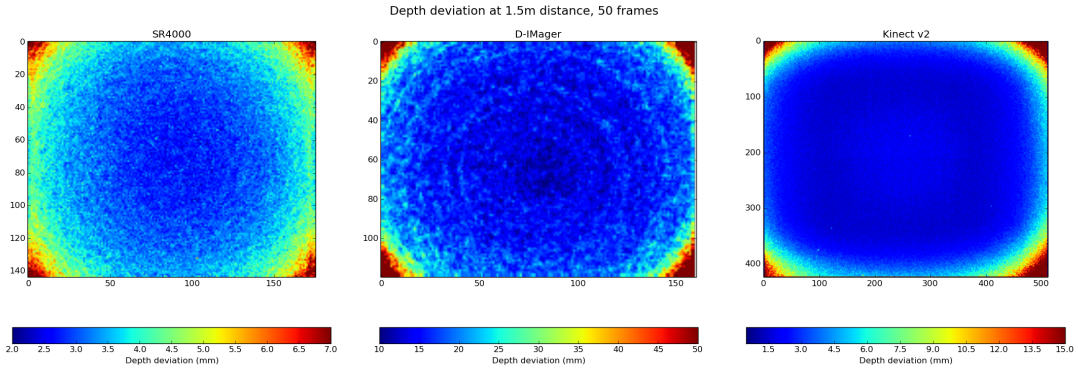

Figure 28: Standard deviations of depth values per pixel.

When comparing Figures 26 and 28, it can be seen that depth accuracies clearly decrease at a same rate with amplitudes. For further analysis, the percentile distributions of the depth deviations are listed in Table 2.

Table 2: Percentile distribution of pixel depth deviations (mm).

Percentile	p_0	p_{25}	p_{50}	p_{75}	p_{90}	p_{99}	p_{100}
SR4000	2.0	3.0	3.3	3.7	4.3	6.4	10
D-IMager	8.5	14.1	16.3	18.9	23.1	57.4	115
Kinect v2	1.0	1.6	1.8	3.0	5.3	15.0	44

Kinect v2 is the most accurate, as 50% of its pixels have a depth SD of 1.8 mm, which is lower than the single best pixel in SR4000. D-IMager is once again the worst, as the deviations are one order of magnitude higher. On the other hand,

SR4000 does not experience as extreme vignetting in corners as the other two and its accuracy remain more uniform throughout the image area. However, considering the field of views, Kinect v2 can see a significantly larger area. Therefore the furthermost corner pixels can be discarded and still achieve good results. For example Lachat et al. [22] suggested using only the 360 x 300 pixel center region of Kinect v2 in applications that demand high accuracy.

To summarize, all tested cameras suffer from the decreasing of the amplitude in the border regions of the image. The number and the placement of the LEDs does not seem to have significant effect, as the light patterns and vignetting of SR4000 and D-IMager are very similar. Kinect's laser-based illumination instead differs slightly, as the center region of the image is uniform and the amplitude and depth accuracy quickly drops in the edges. For high-accuracy applications, using only the center region of the image is therefore recommended. It should also be noted that this test focused only on one distance and material. For better assessment, also different distances and materials with reflectivity characteristics should be tested, as the behaviour of the amplitude might differ considerably.

4.4 Material Reflectivity Test

Now that the depth accuracy in relation to the distance and amplitude variation has been analysed, it is time to investigate the measurement errors originating from the different NIR reflectivity characteristics of the materials. Since the ToF distance measurement is calculated from the amount of received photons, result highly depends on how well the object reflects light. Signal returned from a low-reflective material is weaker, which causes the decrease of SNR and therefore distance errors. On the other hand, very high-reflective materials, such as mirrors, reflect light beams away so that they necessarily do not return to the camera [48]. The purpose of this test is to study how different materials are seen by ToF cameras, and are there any major differences between evaluated cameras.

4.4.1 Setup

For studying the effects of different materials on distance measurements, a test object with different materials was built (shown in Figure 29). Inspired by Lachat et al. [22] and Langmann et al. [9], materials with varying reflectivities and thicknesses were attached to the same paperboard that was used in previous experiments. The material samples are listed in Table 3, starting from the top-left:

The cameras were first placed perpendicular to the target, at a distance of approximately 1 meter, whereof 50 frames were acquired with each device. To better reflect the characteristics of the materials, several different distances and angles were also used.

Table 3: Materials used in the reflectivity test.

Material	**Thickness**
Rubber mat	3 mm
Compact Disk	1 mm
Printed checkerboard	<1 mm
Cardboard box	10 mm
Aluminum plate	2 mm
Blue, green and red papers	<1 mm
Foam profile	10 - 33 mm

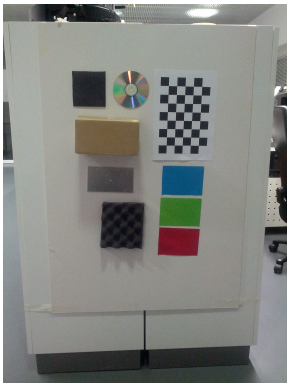

Figure 29: Test object of different materials.

4.4.2 Results

First, the amplitude values are examined. For a preliminary visual inspection, averaged amplitude image from the Kinect was created. It is shown in Figure 30. In the top-left corner is the test object imaged as a whole, and the other images are zooms to the different materials with appropriate colormaps applied.

There are some significant differences in the material reflectivities. For example, in the printed checkerboard pattern, the white squares (plain copy paper) return approximately 50% of the maximum value, whereas black squares return only roughly 10%. Also colored papers have a notable difference; whereas the red and green papers have nearly same amplitude values 40%, the blue paper reflects significantly less NIR light as it has only approximately 30% amplitude.

Highly reflective mirror-like materials (CD and aluminium plate) are problematic, because depending on the viewing angle, they can either reflect all light back to the sensor and cause saturation, or alternatively the light is reflected away from the sensor and no return signal is received at all. This effect can be seen best in the aluminum plate, as one of its edge has nearly zero amplitude value, but simultaneously the other edge is saturated. On the other hand, low reflective materials also return very low amplitude values, but for other reason. This is because the light is

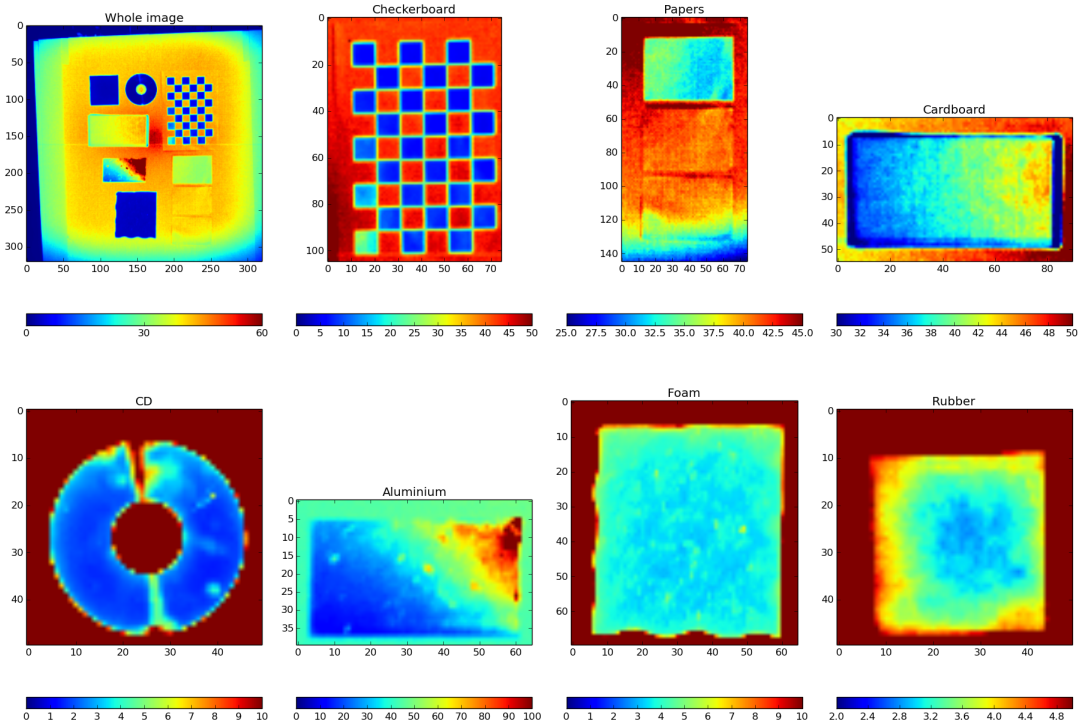

Figure 30: Amplitudes of different materials seen by Kinect.

absorbed rather than reflected.

Next, the distance errors and deviations are analysed. As the test object was imaged from different angles, calculating the distance error was not straightforward. Therefore, the parameters of the reference plane were obtained from the point clouds with least square fitting, and then each point's distance to the plane was determined. Since the thickness of each material sample was known, distance error could be determined. To see whether the distance error was positive or negative, point clouds were visually examined. An example of such distance error image is illustrated in Figure 31. The data sets of each camera were analysed in multiple ways using raw images, averaging, corrected images and point clouds. The resulting amplitudes and depth deviations are collected in Table 4.

The values listed are indicative only, as the both amplitude and depth accuracy highly depend on the angle from which the object is seen and in which part of the image the object lies, since the amplitude decreases towards the edges. Fields marked with an asterisk (*) denote unstable behavior with high variation and occurrence of invalid zero values in certain angles. Despite the values not being fully comparable, the materials can be classified into three categories: high-reflective (CD, aluminum), low-reflective (rubber, foam, black ink), and medium-reflective (papers, cardboard). The low-reflective materials return very low amplitude values

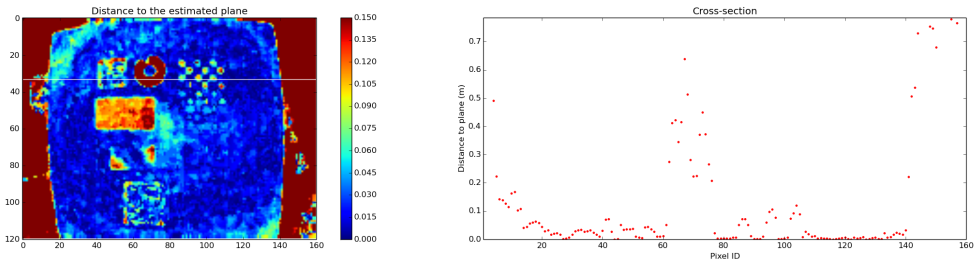

Figure 31: Amplitudes of different materials seen by Kinect.

Table 4: Measured amplitudes and depth errors of different materials. Amplitude values are scaled to a range of 0-100, and depth values are in millimetres.

	SR4000		**D-IMager**		**Kinect v2**	
Material	μ_A	σ_d	μ_A	σ_d	μ_A	σ_d
Rubber	0.7	20-30	9	30-35	3-4	2-3
Foam	1.5	10-20	12	30-40	3-4	2-4
Aluminium	15-100*	2-4*	50-100*	8-20*	10-100*	1-2*
CD	0.1*	43-120*	5*	50-150*	1-3*	3-5*
Cardboard	20	2-3	100	6-12	30-40	1.5
Red paper	17	4	100	6-12	40	1
Green paper	18	4	100	6-12	40	1
Blue paper	15	4	100	6-12	30	2
White checkerboard	20	3-4	100	6-12	42-47	1
Black checkerboard	2-5	7-11	22-45	15-25	5-9	2

and have approximately two times higher depth deviation than middle-reflective materials, wherehas the high-reflective materials have large variations in the amplitudes and the depth deviations of approximately one order of magnitude higher than the medium-reflective materials.

There are slight differences between the cameras. Whereas the behavior with different material categories is rather similar between SR4000 and D-IMager, Kinect v2 seems to be more robust with all kinds of materials. Especially with the high-reflective materials it does not overestimate distances that much, though invalid values are still a problem.

Next, the short distance performance of the cameras was tested with the rubber mat. As the reflectivity of the rubber is very low, pixels do not saturate that easily and also lower distances should be detectable. 50 frames were recorded with each camera at several different distances. The reference distances and corresponding depth deviations of the center pixels are listed in Table 5. As it can be seen from the results, all three cameras work rather well also in shorter distances, since the

rubber mat can be seen even at 15 cm distance. The depth deviations of Kinect and D-IMager are similar than with the higher reflectivity object (e.g. paper) at 1 meter distance, but SR4000 performs significantly worse. This is most likely caused by the automatic integration time, as the paperboard around the rubber mat reflects significantly more light and integration time is hence adjusted lower.

Table 5: Depth standard deviations of the center pixel over short distances (mm).

Reference distance	**15cm**	**20cm**	**25cm**	**30cm**
SR4000	8.1	7.3	9.0	11.2
D-IMager	6.8	5.3	6.6	7.5
Kinect v2	1.5	1.5	1.5	1.3

To summarize, different reflectance characteristics of the materials are challenging for Time-of-Flight cameras. For most materials the depth deviation was only up to a few centimetres at 1 m distance, which is relatively good result for many applications. In controlled environments where the reflectivity characteristics of the materials are roughly known, invalid depth values can be filtered by the amplitude values. However, in unknown environments the amplitude-based filtering becomes more complicated, because the returned amplitude values vary significantly between different materials.

4.5 Outdoor Tests

As mentioned before, bright background illumination poses challenge for the depth measurements of Time-of-Flight camera. When the camera measures the reflected light pulses, it also senses ambient lighting. If the intensity of background light is high, distinguishing the light pulses from the background component becomes more difficult. This is true especially for the sunlight, because its spectrum contains same near infrared wavelengths that most ToF cameras use.

There have been several studies on the subject. Whereas the SR4000 has been reported to perform poorly in direct sunlight [17], Kinect v2 has recently been studied with more promising results. Although the depth accuracy and available range significantly decreases in bright daylight, it still has potential as a low-cost sensor in some outdoor applications [21, 23].

The goal of this test is to examine how bright sunlight affects the depth measurements, and how much the data quality decreases. The D-IMager is particularly interesting because of its high-powered illumination unit and background light suppression capabilities, but also because it has not been evaluated before.

4.5.1 Setup

This experiment consists of two parts, the first of which is to assess the effects of bright sunlight to the depth measurements. It was done simply by imaging a brick wall that was half sunlit and half shady. In the same manner than in previous tests, each camera at a time was placed on a tripod pointing towards the wall at a 2.13 meters distance, whereof 50 frames were recorded. The measurement setup is illustrated in Figure 32. The measurements were carried out on a sunny afternoon in July, when the sun was at its highest. Because a light intensity meter was not available, the intensity of the sun could not be measured accurately. However, the illuminance of direct sunlight at afternoon should be something between 85 - 100 klx [21], which is considered to be enough to severely interfere with ToF measurement.

Figure 32: The outdoor measurement scenarios. On left: half sunny and half shaded brick wall and on right: allotment plot in Otaniemi.

Originally, the purpose of the above-mentioned test was to see how the automatic integration time of SR4000 reacts to two totally differently illuminated regions in a same image. However, it turned out to be practical for the other cameras also, because the effect of sunlight can be observed directly with the same setup. The shaded area is used as a reference, and the risk that additional error sources occurs, e.g. from moving the camera, is reduced.

The second scenario was carried out in an allotment plot, which was intended to simulate a potato field for a certain application, which however is not in the scope of this thesis. Anyhow, the dataset itself was decided to be included, because it contains some interesting aspects. The objective was to assess which of the three ToF cameras is best suited for outdoor use, more particularly detecting the crop rows in the field. Therefore, the allotment plot was imaged with each camera three times: during overcast (cloudy day), on a sunny day so that the sun shines directly to the sensor, and so that the sun illuminates behind the sensor. Cameras were held in hand while walking between the crop rows, and different angles and distances to the ground level were used. Because the data was not necessarily recorded from the same place, no ground truth was available. Instead, the sunny day samples are compared to the overcast data only rather superficially.

4.5.2 Results

First, the results from the brick wall are presented. Mean amplitudes, amplitude deviations and depth deviations are presented in Table 6, and for clarification the mean amplitudes are shown also in Figure 33. Again, as the values vary in different parts of the image, ranges are given instead of single values.

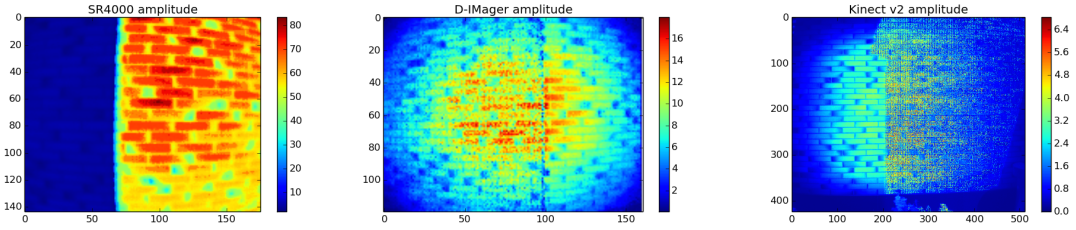

Figure 33: Averaged amplitude images from the brick wall.

Table 6: Amplitudes (%) and depth deviations (m) in different background lighning.

	Amplitude Mean	**Amplitude SD**	**Depth SD**
SR4000 (shade)	2 - 5	2 - 3	0.02 - 0.04
SR4000 (sun)	65 - 85	35 - 50	1.4 - 1.9
D-IMager (shade)	7 - 15	0.4 - 0.6	0.03 - 0.04
D-IMager (sun)	7 - 15	0.8 - 1.2	0.06 - 0.11
Kinect v2 (shade)	2 - 4	0.1 - 0.15	0.004 - 0.007
Kinect v2 (sun)	2 - 7	0.5 - 1.5	0.7 - 1.1

As it can seen from the figure and the table, cameras have significant differences. Amplitude values of SR4000 are nearly twenty times higher in the sunny region of the image compared to shaded part. Also depth deviation is two order of magnitudes higher, which means that the errors are even higher than the reference distance itself.

D-IMager performs significantly better. Its mean amplitudes are nearly identical in both regions of the image, and both amplitude and depth deviations are only two times higher in the sunny region than in shaded part.

Kinect is slightly more sensitive to background lightning. Its mean amplitudes in the sunlight are approximately twice as high than without background light, but the deviation of amplitudes per pixel is one order of magnitude higher. The impact of background light is even more dramatic in the depth values, as the deviation increases approximately hundredfold.

As it appears that pixels easily saturate and output erroneous values in sunlight, errors accumulate when using consecutive frames. Therefore, single pointclouds from each camera are examined instead. They are shown in Figure 34.

As it can be seen from the figure, results are very similar with the averaged images. SR4000 performs poorly, as it gets no signal at all from the sunny part. Actually, when inspecting consecutive clouds from SR4000, it either got no signal or then the whole sunny part of the image was full of random values from the whole operating range, which explains the high depth deviations.

D-IMager is interesting, because there is no significant difference between the two regions. The points in the sunny region are slightly further from the plane and only several outliers are visible. The high-powered illumination unit clearly is advantageous in the presence of sunlight. Then again, poor data precision of D-IMager clearly limits the performance. It would be interesting to see how D-IMager would perform if it had same LED system, but better resolution and data precision.

With the Kinect v2, however, the difference is considerable as there are significantly more outlier points visible in the bright light. On the other hand, due to large number of pixels in the sensor, there are still many pixels that give correct values and the plane is clearly visible. Outliers points can be filtered out rather easily e.g. by statistical outlier filtering, so the cameras can be used in bright ambient illumination but with limited accuracy.

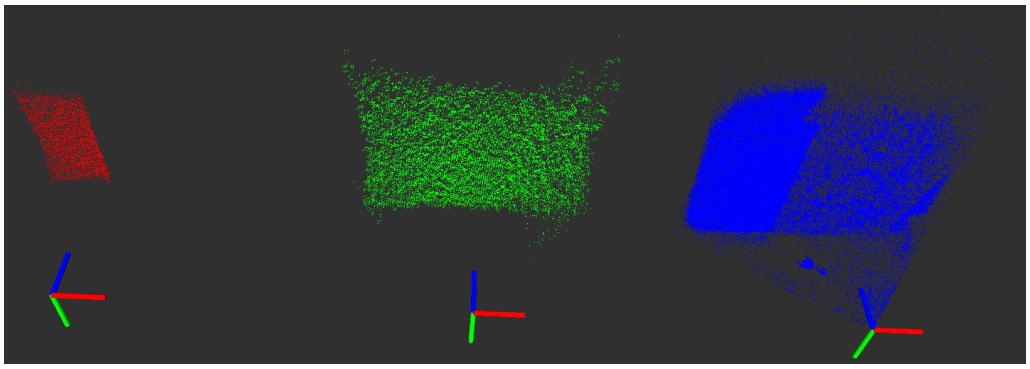

Figure 34: Point clouds from the brick wall. SR4000 in the left, D-IMager in the middle and Kinect v2 in the right.

Next, the allotment field scenario is discussed. Because the data sets were recorded by hand, comparison turned out to be difficult. As there were necessarily no similar frames (viewing angle, shapes of the scene) available from different sun directions, the analysis could not be made. Therefore, only a visual inspection was performed.

As previously, SR4000 was badly saturated in both sunny data sets, and thus it was left out from the comparison. Point clouds from similar scenes captured by D-IMager and Kinect v2 are illustrated in Figure 35. Similarly than with the brick wall, Kinect's cloud includes significantly more outliers in the low-range region than D-IMager does. These outliers might either originate from saturated pixels as discussed before, or they might be incorrectly registered values due to phase wrapping. Either way the D-IMager is better, as it seems to have more powerful LEDs, but it also has an unambiguous range twice as long as Kinect, so it is not that vulnerable to

phase wrapping. However, again due to the poor data precision and low resolution, accuracy quickly decreases on longer distances.

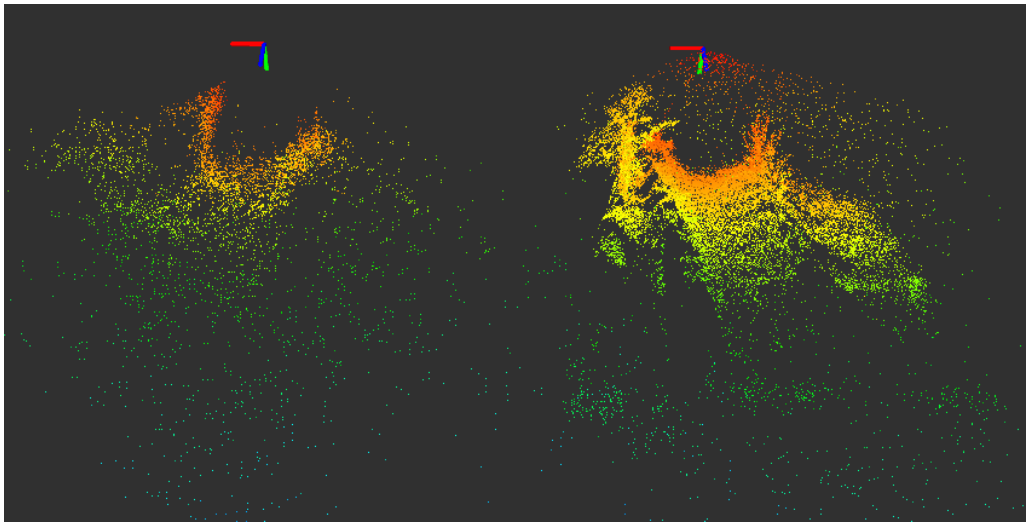

Figure 35: Point clouds from the allotment plot. D-IMager on the left and Kinect v2 on the right. For clarity, the viewing angle has been rotated to be in front of the camera.

During the measurements, the heat of the sun was considered as a potential source of problems. When the dark surfaces of the cameras and the laptops were continuously exposed to sunlight and absorbed heat, they felt significantly warmer than in indoors. No problems were faced though, because the measurement times were relatively short. However, Fankhauser et al. reported the cooling fan of Kinect v2 having problems adapting to changing temperature in the outdoors [21]. Therefore, solutions to reduce the heat from the sun, e.g. shades or reflective coatings, should be considered when operating in direct sunlight.

5 Conclusion

The goal of this thesis was to introduce the Time-of-Flight camera technology, emphasizing its advantages and limitations compared to other depth imaging technologies. Two different ToF approaches were presented: direct measurement based on roundtrip time of a light pulse and indirect measurement based on phase difference of modulated light source, the latter of which was in focus of this work. A time-of-Flight camera provides high frame rate with a minimal need of data post-processing and is robust in different lightning conditions. On the other hand, there are also some fundamental limitations, such as a limited operating range. Additionally, several error sources of ToF cameras were identified and briefly discussed.

In the experimental part, three state-of-the-art ToF cameras, Mesa SR4000, Panasonic D-IMager EKL3106 and Microsoft Kinect v2, were compared. For evaluating the cameras, five tests were carried out. The tests scenarios were chosen based on previous work by other authors, with the focus in demonstrating the circumstances that are problematic for ToF cameras. As the same tests were repeated with all cameras, the devices could be compared directly and study how different sensor attributes affect to the measurements.

The first test was to study how the measurements change when the camera warms up after powered on. Result was that each camera needs a certain amount of time to heat up before measurements stabilize, but the length of the time period is highly dependent on several factors. With lower integration times and more distant objects the warm up process is slower, as less light enters the sensor. Automatic integration time of SR4000 was also found to be potentially problematic, as the constant adjustment of the IT affects to the amplitude values. As a rule of thumb, one hour warm up time is suggested just in case.

In the next two tests, depth accuracy was analysed with respect to the received amplitude. First, depth accuracy relative to distance was studied, while the second case focused on the drop of amplitude inside the image area at fixed distance. As expected, depth accuracy decreases with longer distances. Whereas the relationship between the amplitude and distance is rather linear with constant integration time (D-IMager and Kinect), automatic mode in SR4000 again caused unpredictable behaviour as the integration time and thus also amplitude changed non-linearly with the distance. Moreover, distance-related systematic wiggling error originating from un-ideal modulation of the light was observed with all cameras.

The next objective was to examine how different lightning solutions affect to the accuracy. All of the three cameras have different alignments: SR4000 has 24 LEDs aligned around the lens, whereas D-IMager has total of 18 LEDs, 9 of which on the each side of the lens. Kinect v2 on the other hand uses different approach, as the light is produced with three laser diodes. The alignment of LEDs appeared to have no significant effect, as the light pattern is very similarly circular shaped in both cameras, with the amplitude linearly decreasing towards the edges. Kinect v2 differs significantly, as its light pattern is more rectangular with a bright area

in the center and amplitude quickly decreasing in the edge region. Depth accuracy was found to be inversely proportional to the amplitude. In all cameras, the depth deviation was one order of magnitude higher in the corners than in the center of the image. For best performance, ignoring the values of a few furthermost corner pixels is suggested.

Another important factor affecting to the performance is the choice of materials. As different materials have very different albedo, they are also perceived differently by the ToF camera. Very high reflective mirror-like materials, such as a CD used in the experiment, were found to be especially problematic, because the emitted light does not get back to the sensor in certain angles. On the other hand, very low-reflective materials are also problematic, because the proportion of noise increases as the intensity drops. Both depth overestimation and underestimation of depth were observed, as well as increasing deviation of the values.

Finally, cameras were tested also in outdoors to study how much bright background light decreases the data quality. SR4000 was found to be virtually unusable, as sunlight saturated the whole sensor. Kinect v2 was performing moderately, as there were some saturated pixels, but most of the pixels were still working sufficiently. D-IMager was the most robust in the direct sunlight, as there were no saturated pixels. The depth deviation in the sunlight was approximately one order of magnitude higher than with no background light present. Another problem in outdoors was found to be open space, as invalid depth values are returned due to reflections from distant objects and phase wrapping. As a partial solution, ToF camera is suggested to be tilted slightly downwards so that the optical axis of the camera reaches the ground level within the unambiguous range. This however does not solve the problem completely, as the upper part of the image sensor can still receive distant values that map to invalid range.

Overall, Kinect v2 can be considered as the winner of the trio. Compared to the other two devices, it offers approximately ten times more pixels and a significantly larger FOV at a fraction of the price. It is also the most accurate, as it can achieve up to millimetre precision. SR4000 is at its best almost as accurate, but on the other hand it is highly sensitive to background light. D-IMager differs significantly from the other two, as its depth accuracy is nearly tenfold lower. The big difference is probably mostly caused by the poor data precision, which allows only depth values on 1 cm intervals. It would be interesting to see how much the performance of the D-IMager would improve, if the data precision was better.

Compromises must be made when seeking the optimal performance of a Time-of-Flight camera. For example, higher modulation frequency increases the depth resolution, but decreases the unambiguous operating range. Increasing the lighting power of the camera on the other hand improves the performance in bright background light and in long distances, but as a trade-off, low range performance decreases as the pixels saturate due to the large amount of light. The best choice of the parameters depends highly on the application.

Despite its problems, Time-of-Flight camera is promising technology for depth

imaging. Especially lately, as the sensors have been developing rapidly and mass-produced devices entered markets, low resolution and high price are no longer a problem as before. With low-cost ToF cameras, a whole new application areas can be expected to emerge.

References

[1] Nick Pears, Yonghuai Liu, and Peter Bunting. *3D Imaging, Analysis and Applications*. Springer, 2012. ISBN: 9978-1-4471-4062-7.

[2] Microsoft. *Kinect*, 2010. http://www.xbox.com/en-US/xbox-360/accessories/kinect. Accessed on 15.1.2015.

[3] Willow Garage. *TurtleBot*, 2011. http://www.willowgarage.com/turtlebot. Accessed 15.1.2015.

[4] Erik E. Stone and Marjorie Skubic. Evaluation of an inexpensive depth camera for passive in-home fall risk assessment. In *5th International Conference on Pervasive Computing Technologies for Healthcare*, pages 71-77. 2011.

[5] Microsoft. *Kinect for Windows v2*, 2014. http://www.microsoft.com/en-us/kinectforwindows/meetkinect/features.aspx. Accessed on 15.1.2015.

[6] Stefan May, David Droeschel, Dirk Holz, Stefan Fuchs, Ezlo Malis, Andreas Nüchter and Joachim Hertzberg. Three-dimensional mapping with time-of-flight cameras. In *Journal of Field Robotics*, Vol 26: pages 934–965. 2009.

[7] Sergi Foix, Guillem Alenyà and Carme Torras. Lock-in Time-of-Flight (ToF) Cameras: A Survey. In *IEEE Sensors Journal* 11(9). pages 1917-1926. 2011.

[8] Burak S. Gokturk, Hakan Yalcin and Cyrus Bamji. A Time-Of-Flight Depth Sensor - System Description, Issues and Solutions. In *IEEE Conference on Computer Vision and Pattern Recognition Workshop, 2004 (CVPRW'04)*. 2004.

[9] Benjamin Langmann, Klaus Hartmann and Otmar Loffeld. Depth Camera Technology Comparison And Performance Evaluation. In *Proceedings of the 1st International Conference on Pattern Recognition Applications and Methods*. pages 438-444. 2012.

[10] Alexandra von Beringe. Performance evaluation of a Range Camera SR4000. *Master's Thesis. Institute of Photogrammetry and Remote Sensing, Vienna University of Technology*. 2012.

[11] Timo Breuer, Christoph Bodensteiner and Michael Arens. Low-cost commodity depth sensor comparison and accuracy analysis. In: *SPIE Security+ Defence. International Society for Optics and Photonics*. pages. 92500G-92500G-10. 2014

[12] Lin Yang, Longyu Zhang, Haiwei Dong, Abdulhameed Alelaiwi and Abdulmotaleb El Saddik. Evaluating and Improving the Depth Accuracy of Kinect for Windows v2. In *IEEE Sensors Journal vol. 15 no. 8*. pp. 4275-4285. 2015.

[13] Todor Stoyanov, Athanasia Louloudi, Henrik Andreasson and Achim J. Lilienthal. Comparative Evaluation of Range Sensor Accuracy in Indoor Environments. In: *European Conference on Mobile Robotics (ECMR)*. pages 19-24. 2011.

[14] Andry M. Pinto, Paulo Costa, A. Paulo Moreira, Lufs F. Rocha, Eduardo Moreira and Germano Veiga. Evaluation of Depth Sensors for Robotic Applications. In *International Conference on Autonomous Robot Systems and Competitions (ICARSC)* pp. 139-143. 2015.

[15] Hamed Sarbolandi, Damien Lefloch and Andreas Kolb. Kinect Range Sensing: Structured-Light versus Time-of-Flight Kinect. In: *Computer Vision and Image Understanding*. 2015.

[16] Stephan Hussmann, Thorsten Ringbeck, Bianca Hagebeuker. A Performance Review of 3D TOF Vision Systems in Comparison to Stereo Vision Systems. In *Stereo Vision,* Asim Bhatti(Ed.). InTech Open Access Publisher. 2008.

[17] Wajahat Kazmi, Sergi Foix, Guillem Alenya and Hans J. Andersen. Indoor and outdoor depth imaging of leaves with time-of-flight and stereo vision sensors: Analysis and comparison. In:*ISPRS Journal of Photogrammetry and Remote Sensing, 88* pages 128-146. 2014.

[18] D. Falie and V. Buzuloiu. Wide range time of flight camera for outdoor surveillance. In *Microwaves, Radar and Remote Sensing Symposium, MRRS 2008. IEEE.* 2008.

[19] Haris Balta, Geert De Cubber, Daniela Doroftei, Yvan Baudoin and Hichem Sahli. Terrain Traversability Analysis for off-road robots using Time-Of-Flight 3D Sensing. In *7th IARP International Workshop on Robotics for Risky Environment-Extreme Robotics, Saint-Petersburg, Russia*. 2013.

[20] A.V.H. Ollikkala and A.J. Mäkynen. Range Imaging Using a Time-of-Flight 3D Camera and a Cooperative Object. In *Instrumentation and Measurement Technology Conference, I2MTC'09. IEEE.* 2009.

[21] Péter Fankhauser, Michael Bloesch, Diego Rodriguez, Ralf Kaestner, Marco Hutter and Roland Siegwart. Kinect v2 for Mobile Robot Navigation: Evaluation and Modeling. In *IEEE International Conference on Advanced Robotics (ICAR)* (submitted). 2015.

[22] E. Lachat, H. Macher, M.A. Mittet, T. Landes and P. Grussenmeyer. First Experiences with Kinect v2 Sensor for Close Range 3D Modelling. In *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences (1)*. pages 93-100. 2015.

[23] Thomas Butkiewicz. Low-cost coastal mapping using Kinect v2 time-of-flight cameras. In *Oceans - St. John's*. IEEE, 2014.

[24] Masaki Takahashi, Mahito Fujii, Masahide Naemura and Shin'ichi Satoh. Human gesture recognition using 3.5-dimensional trajectory features for hands-free user interface. In *Proceedings of the first ACM international workshop on Analysis and retrieval of tracked events and motion in imagery streams* pp. 3-8. 2010.

[25] Drazen Brscic, Takayuki Kanda, Tetsushi Ikeda and Takahiro Miyashita. Person tracking in large public spaces using 3-D range sensors. In *IEEE Transactions on Human-Machine Systems 43(6)*, pages 522-534. 2013.

[26] Xuan Wang. 2D Mapping Solutions for Low Cost Mobile Robot. Master's Thesis. Royal Institute of Technology, School of Computer Science and Communication, Sweden. 2013.

[27] Xuming Luan. Experimental investigation of photonic mixer device and development of TOF 3D ranging systems based on PMD technology. *Diss., Department of Electrical Engineering and Computer Science, University of Siegen* 2001.

[28] Cyrus S. Bamji, et al. A 0.13 μm CMOS System-on-Chip for a 512x424 Time-of-Flight Image Sensor With Multi-Frequency Photo-Demodulation up to 130 MHz and 2 GS/s ADC. In *IEEE Journal of Solid-State Circuits 5(50(1)*. 2015.

[29] Mesa Imaging AG. *SR4000 User Manual*. 2011.

[30] Cristiano Niclas, Claudio Favi, Theo Kluter, Frédéric Monnier and Edoardo Charbon. Single-Photon Synchronous Detection. In *IEEE Journal of Solid-State Circuits* 44(7). pages 1977-1989. 2009.

[31] Larry Li. Time-of-Flight Camera - An Introduction. *Texas Instruments - Technical White Paper*, 2014.

[32] Robert Lange and Peter Seitz. Solid-State Time-of-Flight Range Camera. In *IEEE Journal of Quantum Electronics* 37(3). pages 390-397. 2001.

[33] Tobias Möller, Holger Kraft, Jochen Frey, Martin Albrecht and Robert Lange. Robust 3D Measurement with PMD Sensors. *Range Imaging Day, Zürich*. 2005.

[34] Miles Hansard, Seungkyu Lee, Ouk Choi and Radu Horaud. *Time of Flight Cameras: Principles, Methods, and Applications*. Springer, 2012. ISBN: 978-1-4471-4658-2.

[35] Dario Piatti and Fulvio Rinaudo. SR-4000 and CamCube3. 0 time of flight (ToF) cameras: Tests and comparison. In *Remote Sensing 4(4)*. pages 1069-1089. 2012.

[36] Velodyne. *Velodyne VLP-16*, 2014. http://velodynelidar.com/lidar/hdlproducts/vlp16.aspx. Accessed on 19.1.2015.

[37] Timo Kahlmann, Fabio Remondino and H. Ingensand. Calibration for Increased Accuracy of the Range Imaging Camera Swissranger. In *Image Engineering and Vision Metrology (IEVM) (36:3)*. pages 136-141. 2006.

[38] Nikhil Naik, Achuta Kadambi, Christoph Rhemann, Shahram Izadi, Ramesh Raskar and Sing Bing Kang. A Light Transport Model for Mitigating Multipath Interference in Time-of-flight Sensors. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* pp. 73-81. 2015

[39] A.A. Dorrington, J.P. Godbaz, M.J. Cree, A.D. Payne and L.V. Streeter. Separating true range measurements from multi-path and scattering interference in commercial range cameras. In *IS&T/SPIE Electronic Imaging* pp. 786404-786404. International Society for Optics and Photonics. 2011

[40] Henrik Schäfer, Frank Lenzen and Christoph S. Garbe. Model Based Scattering Correction in Time-Of-Flight Cameras. In *Optics express 22(24)*. pages 29835-29846. 2014.

[41] OpenKinect. libfreenect2. https://github.com/OpenKinect/libfreenect2. 2015.

[42] Panasonic. *D-IMager EK3106 Sensor Specifications*. 2012. http://www2.panasonic.biz/es/densetsu/device/3DImageSensor/en/product.html Accessed on 8.4.2015.

[43] Morgan Quigley, Ken Conley, Brian Gerkey, Josh Faust, Tully Foote, Jeremy Leibs, Rob Wheeler and Andrew Ng. ROS: an open-source Robot Operating System. *ICRA workshop on open source software. Vol. 3. No. 3.2.* 2009.

[44] Thiemo Wiedemeyer. IAI Kinect2. Institute for Artificial Intelligence, University Bremen, 2014–2015. https://github.com/code-iai/iai_kinect2. Accessed on 20.7.2015.

[45] Swissranger_camera ROS package. http://wiki.ros.org/swissranger_camera. Accessed on 11.2.2015.

[46] John D. Hunter. Matplotlib: A 2D graphics environment. In: *Computing in science and engineering 9(3)* pages 90-95. 2007.

[47] Radu Bogdan Rusu and Steve Cousins. 3D is here: Point Cloud Library (PCL). In *IEEE International Conference on Robotics and Automation (ICRA)*, May 9-13, 2011.

[48] Damien Lefloch, Rahul Nair, Frank Lenzen, Henrik Schäfer, Lee Streeter, Michael J. Cree and Andreas Kolb. Technical foundation and calibration methods for time-of-flight cameras. In *Time-of-Flight and Depth Imaging. Sensors, Algorithms, and Applications* pp. 3-24. Springer Berlin Heidelberg. 2013.