2 Depth Camera Assessment

The driving question of this chapter is how competitive cheap consumer depth cameras, namely the Microsoft Kinect and the SoftKinetic DepthSense, are compared to state-of-the-art Time-of-Flight depth cameras. Several depth camera models from different manufactures are put to the test on a variety of tasks in order to judge their respective performance and to reveal their weaknesses. The evaluation will concentrate on the area of application for which all cameras are specified, i.e. near field indoor scenes. The characteristics and limitations of the different technologies as well as the available devices are discussed and evaluated with a set of experimental setups. In particular, the noise level and the axial and angular resolutions are compared. Additionally, refined formulae to generate depth values based on the raw measurements of the Kinect are presented.

2.1 Depth Camera Overview

Depth or range cameras have been developed for several years and are available to researchers as well as commercially for certain applications for about a decade. PMD Technologies (PMDTec), Mesa Imaging, 3DV Systems and Canesta were the companies driving the development of Time-of-Flight (ToF) depth cameras. In recent years additional competitors like Panasonic, SoftKinetic or Fotonic announced or released new models.

The cameras produced by all these manufacturers have in common that they illuminate the scene with infrared light and measure the time until the light is received. There are two main principles of operation: pulsed light and continuous wave amplitude modulation. The former is limited by having to measure very short time intervals in order to achieve a distance resolution which corresponds to a few centimeters in depth (e.g. ZCam of 3DV Systems). The continuous wave modulation approach avoids this by measuring the phase shift between emitted and received modulated light, which corresponds directly to the time of flight and in turn to the depth. However, ambiguities in form of multiples of the modulation wavelength may occur here.

B. Langmann, *Wide Area 2D/3D Imaging,*
DOI 10.1007/978-3-658-06457-0_2, © Springer Fachmedien Wiesbaden 2014

In the past the ToF imaging sensors suffered from two major problems: a low resolution and a low sensitivity resulting in high noise levels. Additionally, background light caused problems when used outdoors. Currently, ToF imaging chips reaching resolutions of up to 200×200 pixels are on the market and chips with 352×288 pixels are in development. Moreover, for a few years some ToF chips have featured methods to suppress ambient light (e.g. Suppression of Background Illumination - SBI).

Other depth cameras or measuring devices, such as laser scanners or structured light approaches, were not able to provide (affordably) high frame rates for full images with a reasonable resolution. This was true until in 2010 Microsoft (PrimeSense) released the Kinect. Instead of relying on a pattern varying in time as widely applied previously, it works with a fixed irregular pattern consisting of a large number of dots produced by an infrared laser LED and a diffractive optical element. The Kinect determines the disparities between the emitted light beam and the observed position of the light dot with a two megapixel grayscale imaging chip. The identity of a dot is determined by utilizing the irregular pattern. It is assumed that the depth of a local group of dots is calculated simultaneously, but the actual method remains a secret up until today. Once the identity of a dot is known the distance to the reflecting object can be easily triangulated. In addition to the depth measurements, the Kinect includes a color imaging chip as well as microphones.

Given the low cost of the Kinect as a consumer product and the novelty as well as the non-disclosure of its functional principle, the reliability and accuracy of the camera should be evaluated. Instead of an investigation of a specific application, the approach taken to judge the performance of the Kinect is to develop a set of experimental setups and to compare the results of the Kinect to state-of-the-art ToF depth cameras.

The performance of ToF cameras using the Photonic Mixer Divece (PMD) was widely studied in the past. Noteworthy are for example [55, 11]. The measurement quality at different distances and using different exposure times is evaluated. Lottner et al. discuss the influence and the operation of unusual lighting geometries in [45], i.e. lighting devices not positioned symmetrically around the camera in close distance. In [66] depth cameras from several manufactures are compared, which are PMDTec, Mesa Imaging and Canesta. The application considered is 3D reconstruction for mobile robots. And in [2] PMD cameras are compared to a stereo setup. They use the task of scene reconstruction to judge the performance of both alternatives. The most closely related paper is [61], in which two ToF cameras are compared to the Kinect and to a laser scanner. The application in mind is navigation

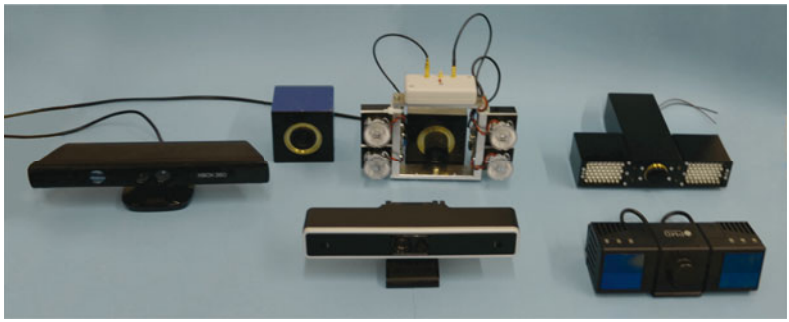


Figure 2.1: The depth cameras involved in the comparison. In the top row left to right are the MicroSoft Kinect, two ZESS MultiCams, the PMDTec 3k-S displayed and in the bottom row the SoftKinetic DepthSense 311 and finally the PMDTec CamCube 41k.

for mobile robots and the methodology is the reconstruction of a 3D scene with known ground truth.

The following comparison was previously published in [L8] and involves different commercially available depth cameras, some of which are shown in Fig. 2.1 as well as several versions of our MultiCam. The Microsoft Kinect and the SoftKinetic DepthSense as recent consumer depth cameras compete with two established Time-of-Flight cameras based on the Photonic Mixer Device (PMD) by PMDTec.

2.1.1 Microsoft Kinect

The Microsoft Kinect camera generates an irregular pattern of dots (actually, a sub-pattern is repeated 3×3 times) with the help of a diffractive optical element and an infrared laser diode. It incorporates a color and a two megapixel grayscale chip with an IR filter, which is used to determine the disparities between the emitted light dots and their observed position. In order to triangulate the depth of an object in the scene, the identity of an observed dot on the object must be determined. This can be performed with much more certainty with the irregular pattern than with an regular pattern. The camera is built with a 6 mm lens for the color chip and an astigmatic lens for the grayscale chip, which skews the infrared dots to ellipsoids. These deformations provide a depth estimate and together with the triangulation a depth map is calculated. In the standard mode the depth map contains 640×480 pixels and each pixel is a raw 11-bit integer value. The depth

values describe the distance to the imaginary image plane and not to the focal point. There are currently two formulae to calculate the depth in meters publicly known, cf. [49]. An integer raw depth value d is mapped to a depth value in meters with a simple formula by

$$\delta_{simple}(d) = \frac{1}{-0.00307d + 3.33} \ . \tag{2.1}$$

A more precise method based on a higher order function is be given by

$$\delta_{tan}(d) = 0.1236 \cdot \tan\left(\frac{d}{2842.5} + 1.186\right) \ . \tag{2.2}$$

Since the depth map has about $300k$ pixels, calculating the latter formula 30 times per second can be challenging or even impossible, especially for embedded systems.

Using the translation unit described in Section 2.2.1 refined formulas have been determined:

$$\delta^{refined}_{simple}(d) = \frac{1}{-0.8965 \cdot d + 3.123} \tag{2.3}$$

$$\delta^{refined}_{tan}(d) = 0.181 \cdot \tan\left(0.161 \cdot d + 4.15\right) \ . \tag{2.4}$$

See Section 2.2.1 for a comparison of these formulae.

2.1.2 PMDTec CamCube 41k

The CamCube 41k by PMDTec, cf. [51], contains a 200×200 pixel PMD chip and includes two lighting units. Modulated infrared light with frequencies up to 21 MHz is emitted and the phase shift between the emitted and received light is calculated. The phase corresponds to the distance of the reflecting object and it is determined using the so-called four phase algorithm. For this algorithm four phase images P_1 to P_4 are recorded at different phase offsets and with the arc tangent relationship the phase difference can be retrieved as

$$\Delta\varphi = \operatorname{atan2}\left(P_2 - P_4, P_1 - P_3\right) \ . \tag{2.5}$$

The distance can be derived from the phase difference with

$$\delta(\Delta\varphi) = \Delta\varphi \cdot \frac{c}{4\pi \cdot \nu} \ , \tag{2.6}$$

where c is the speed of light and ν is the modulation frequency. More details will be discussed in Section 3.1.

The CamCube features a method to suppress background illumination called SBI to allow for outdoor imaging. It provides the possibility to synchronize the acquisition of images with the means of a hardware trigger. A wide variety of different lenses can be mounted on the camera due to the standard CS-mount adapter. The camera is connected to a computer via USB.

2.1.3 PMDTec 3k-S

The 3k-S PMD camera is a development and research version from PMDTec and it employs an older PMD chip with only 64×48 pixels. It features a SBI system and contains a C-mount lens adapter and uses firewire (IEEE-1394) to communicate with the computer. This PMD camera is known to be significantly less sensitive than cameras with newer PMD chips even though the pixel pitch is $100\,\mu$m compared to $45\,\mu$m of the 41k PMD chip.

2.1.4 PMDTec 100k

PMDTec is developing depth imaging chips of higher resolution and a version of a 100k PMD chip with 352×288 pixels and a pixel pitch of $17.5\,\mu$m is tested as well. The chip features also a SBI system and is electronically similar to earlier versions. However, it contains inhomogeneities and certain deficiencies which will be discussed later on.

2.1.5 Softkinetic DepthSense 311

The newly established company SoftKinetic released a depth imaging camera named DepthSense 311 in 2012, which includes an additional color camera in a binocular setup and microphones. The camera is also based on the ToF principle and modulates infrared light with frequencies between 14 and 16 MHz. The modulation frequency is continuously changed and 500 phase images are acquired per second resulting in up to 60 fps. The lateral resolution of the depth map is 160×120 pixels and color images of 640×480 pixels are delivered. The camera is connected via USB and contains a fixed lens with an opening angle of 57.3 degrees in width for the depth chip.

2.2 Experimental Evaluation

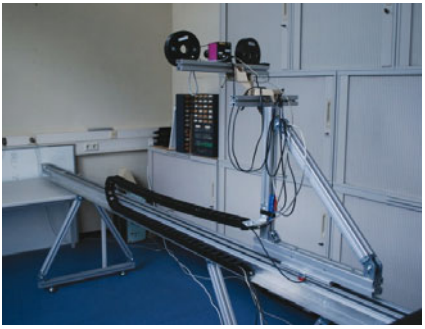
In this section the evaluation methods and the most notable results of the comparison will be discussed. In the first part in Section 2.2.1, the radial

resolution in terms of precision and accuracy of all cameras will be compared. For the second set of experiments only the CamCube will serve as a reference for the Kinect, since only the different technologies ToF and triangulation are compared. In order to make the results comparable an appropriate lens has to be chosen for the CamCube. Since the Kinect uses a fixed 6 mm lens and the grayscale chip has a resolution of 1280×1024 (only 640×480 depth pixels are transmitted) with a pixel pitch of $5.2\,\mu$m, this results in a chip size of $6.66\,\text{mm} \times 5.33\,\text{mm}$. The CamCube has a resolution of 200×200 pixels with a pixel pitch of $45\,\mu$m resulting in a chip size of $9\,\text{mm} \times 9\,\text{mm}$. Therefore, the corresponding lens for the CamCube would have a focal length of 8.11 mm for the width and about 10.13 mm for the height. As a compromise a lens with a focal length of 8.5 mm was chosen.

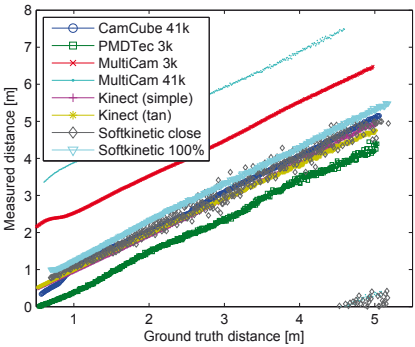
2.2.1 Depth Accuracy Evaluation

All cameras were mounted on a translation unit, which is able to position the cameras at distances between 50 cm and 5 meters from a wall with a positioning accuracy better than 1 mm. The cameras were leveled and were pointing orthogonally at the wall. 100 images were taken per position with a step size of 1 cm, which resulted in 45000 images per camera. The same lens, the same modulation frequency of 20 MHz as well as the same exposure times (5 ms for distances below 2.5 meters and 10 ms for higher distances) were used for all PMD based cameras. In Fig. 2.2 single depth measurements, the estimated standard deviation (SD) as well as the average distance error of the measurements to the ground truth after a linear correction are shown for a single pixel of each evaluated camera. The average error is computed by performing a linear regression for a subset of depth measurements (2 to 4 meters) and correction all depth measurements afterwards. This removes constant and linear errors in the measurements, which are caused by systematic errors or inaccuracies in the setup. Similar plots were made for different pixels to ensure that the results are representative for the cameras.

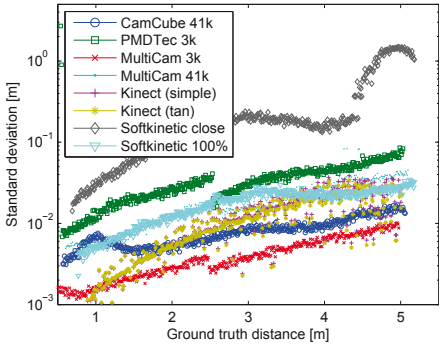
Here the CamCube shows measurement errors for small distances due to overexposure and both PMD based ToF cameras display the wobbling behavior as previously discussed, e.g. in [34]. The distance error for all cameras is comparable in the optimal region of application $(2 - 4m)$ with a slight advantage for the Kinect. More complex calibration methods exist for PMD based cameras, see [41] or [58], which are able to reduce the distance error further. The estimated standard deviation of the distance



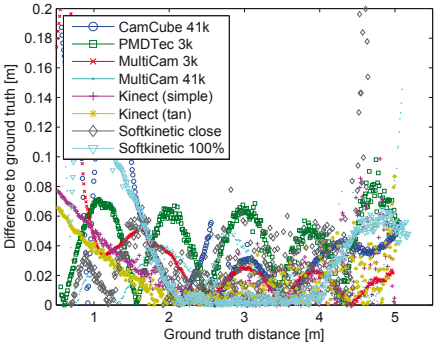
(a) Translation unit



(b) Raw depth measurements



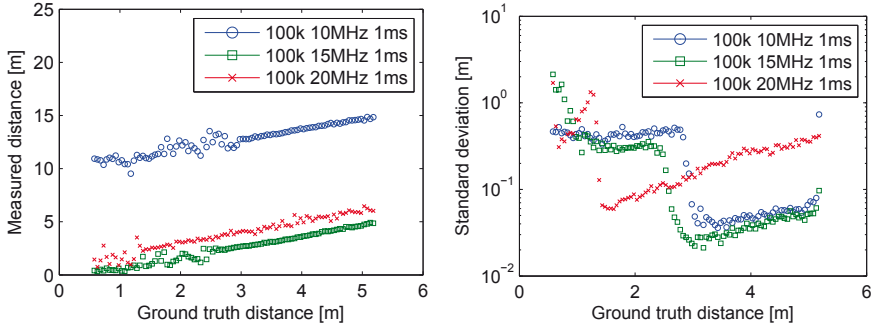
(c) Standard deviation ($N = 100$)



(d) Average measurement error ($N = 100$)

Figure 2.2: Measurement results and analysis for the different depth cameras performed with the translation unit.

measurements based on 100 frames shows significant differences. The Kinect displays a low variance for short distances but higher noise levels than e.g. the CamCube for distances larger than two meters. The variance of the PMDTec 3k camera is higher due to its limited lighting system and its low sensitivity. The SoftKinetic DepthSense is in this respect slightly inferior to other state-of-the-art depth cameras. The experimental 100k PMD chip is evaluated in Fig. 2.3. The higher measurement noise is a consequence of the much smaller pixel size. The pixels are also very easily overexposed. Consequently, a small exposure time of 1 ms was applied. Nevertheless, an overexposure is observed for distances closer than 1.5 meters. Shorter



(a) Depth measurements **(b)** Standard deviation ($N = 100$)

Figure 2.3: Measurement results and analysis for the experimental 100k PMD chip performed with a translation unit.

exposure times will reduce this minimal distance but will limit the range of the camera. Therefore, the 100k PMD chip is limited to scenes with small distances and reflectivity differences.

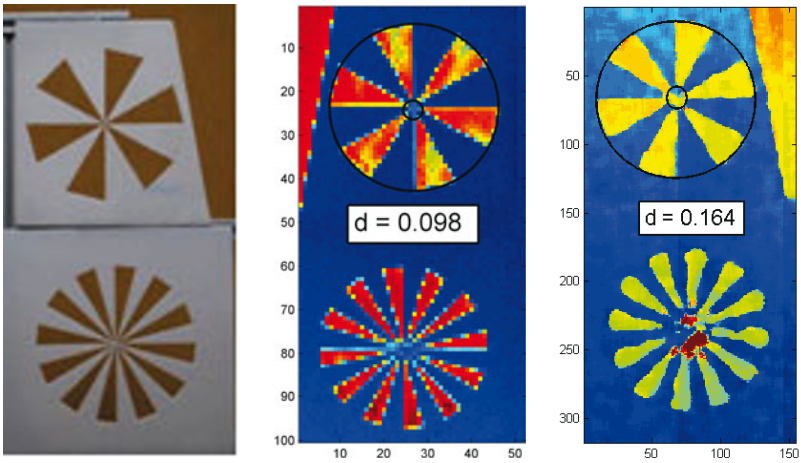
2.2.2 Estimation of the Lateral Resolution

A 3D Siemens star, see Fig. 2.4, is a tool to determine the angular or lateral resolution of depth measurement devices. In [8] it was used to compare laser scanners. In the context of depth cameras it promises insights, in particular for the Kinect, for which the effective resolution is not known. The lateral resolution r of an imaging device can be calculated as

$$r = \frac{\pi d M}{n} \tag{2.7}$$

with n being the number of fields of the star (here 12 and 24 respectively), d being the ratio of the diameter of an imaginary circle in the middle of the star containing incorrect measurements to the diameter M of the star.

For the 3D Siemens stars frames were taken at different distances and in Fig. 2.4 the respective parts of one set of the resulting images are shown. In theory, the CamCube has an opening angle of 55.8 degrees with a 8.5 mm lens, which leads to an angular resolution of 0.28 degrees. Using the 3D Siemens stars in one meter distance an estimate for the angular physical resolution of the CamCube is 0.51 cm, which corresponds to 0.29 degrees and confirms the theoretical value.



(a) 3D Siemens stars **(b)** CamCube **(c)** Kinect

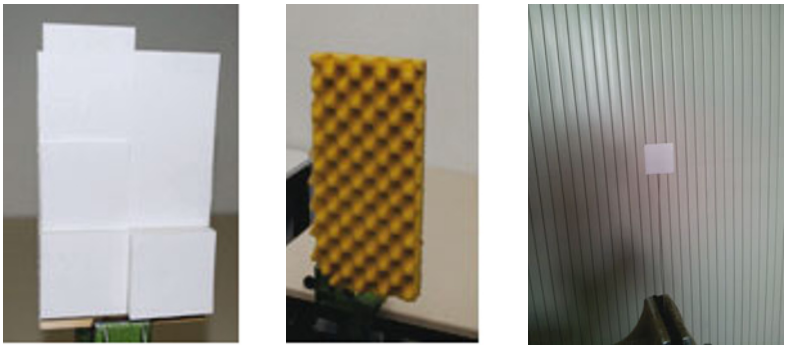
Figure 2.4: 3D Siemens stars with 20 cm diameter and measurement results in 1 meter distance.

The Kinect has an opening angle of 58.1 degrees and with 640 pixels in width it has a theoretical angular resolution of 0.09 degrees (0.12° in height). In practice an angular resolution of 0.85 cm and 0.49 degrees was determined. This corresponds to a resolution of 118 pixels in width. The significant difference is due to the fact that multiple pixels are needed in order to generate one depth value (by locating the infrared dot). Even though the Kinect contains a two megapixel grayscale chip and transfers only a VGA depth map, this still does not compensate the need of multiple pixels in order to locate the dots. Additionally, the Kinect performs to our knowledge either a post-processing or utilizes regions of pixels in the triangulation, which may lead to errors at boundaries of objects.

This observation agrees with estimates that the dot pattern consists of about 220×170 dots which can be interpreted as the theoretical limit of the lateral resolution.

2.2.3 Depth Resolution Test Objects

Fig. 2.5 shows three objects to visualize the angular and axial resolution of the depth cameras. The first object consists of a ground plane and three



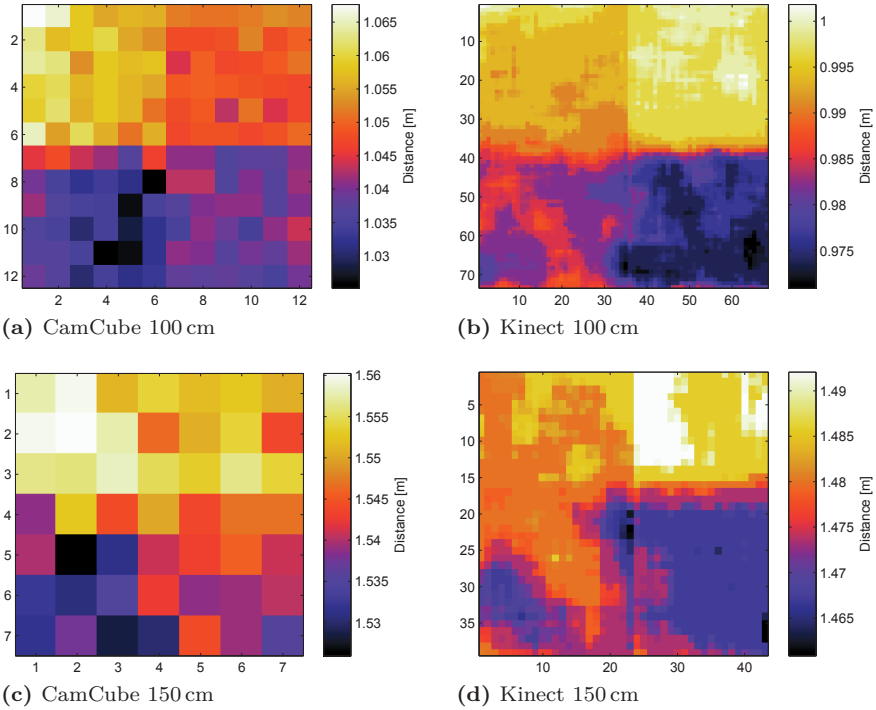
(a) Stairs **(b)** Sinusoidal pattern **(c)** White 2 cm × 2 cm plane

Figure 2.5: Resolution test objects to evaluate and visualize the angular and axial resolution of depth cameras.

6 cm × 6 cm cuboids of different heights of 3, 9 and 1 mm. The second object has a surface close to a sinusoidal formed plane with an amplitude of 1.75 cm and a wave length of 3.5*cm*. Moreover, a 2 cm × 2 cm white plane mounted on a 0.5 mm steel wire was placed in some distance to a wall. Then the depth cameras were positioned at different distances to the plane and it was checked whether they were able to distinguish between the plane and the wall.

In Fig. 2.6 some results for the cuboids are shown, for which 10 depth maps were averaged. Both cameras are able to measure the different distances with high accuracy in one meter distance. At 1.5 meters distance the precision decreases and at 2 meters both cameras cannot resolve the pattern reliably. In Fig. 2.7 a rendering of the sinusoidal structure is given. Again both cameras are able to capture the pattern correctly, but the detail of preservation is higher for the Kinect.

The experiment with the small plane yields surprising results. For the CamCube the 2 cm × 2 cm plane stays visible with correct depth value even in 4.5 m distance. The plane has a size of only 0.7 pixels when the camera is placed at this distance, but this is still enough to gain a correct measurement. The pixel will observe a mixture of signals with different phases, but the one coming from the small plane is the strongest and therefore the measurement still yields sufficiently reliable values. The Kinect displays a completely different behavior. Here a second camera with an optical IR filter was



(a) CamCube 100 cm **(b)** Kinect 100 cm

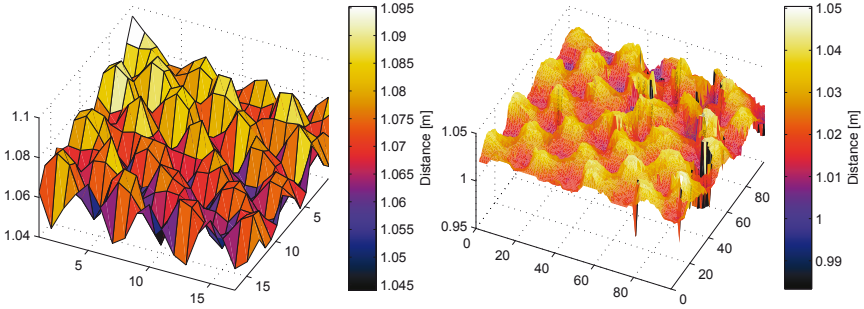
(c) CamCube 150 cm **(d)** Kinect 150 cm

Figure 2.6: Cuboids of different heights recorded using the Kinect and the Camcube. 10 frames were averaged for each distance.

employed to observe the infrared dots on the plane. In 1.75 meters distance the small plane is invisible to the Kinect, as the number of dots on the plane is less than five. In 1.5 meter distance the plane is visible in about 50% of the cases depending on the lateral position of the plane, for an example see Fig. 2.8. In one meter distance the plane is visible and correctly measured all the time with about $10 - 20$ dots on the plane. The explanation for this behavior is the same as for the 3D Siemens stars.

2.2.4 Angular Dependency of Measurements

Since the measurements of the Kinect are based on triangulation, it is doubtful that objects can be measured accurately at all angles. To evaluate



(a) CamCube 100 cm **(b)** Kinect 100 cm

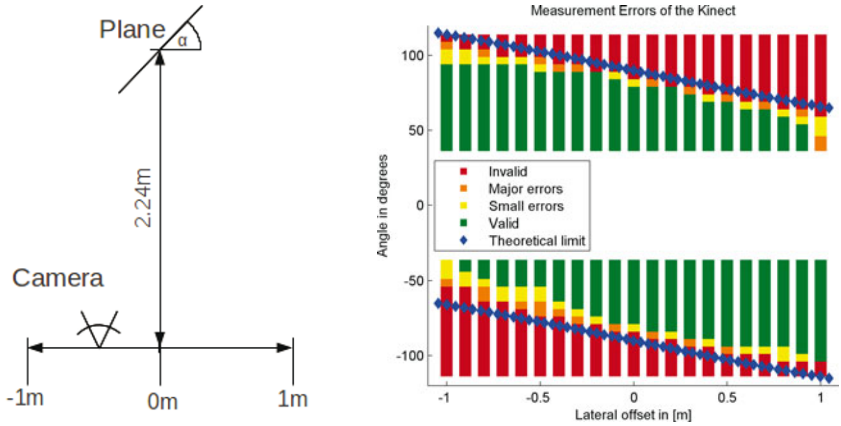
Figure 2.7: Sinusoidal structure measured with both cameras in 1 m distance.

(a) Color image **(b)** Depth map **(c)** IR image from different camera

Figure 2.8: Measurement result of the Kinect for the small plane in 1.5 m distance.

the range of angles resulting in accurate measurements the camera is moved horizontally and a plane is installed in a fixed distance to the camera path. Angles from -40 to -110 degrees and from 40 to 110 degrees with a step size of 5 degrees are applied and the camera is positioned with offsets from -1 to 1 meter using a step size of 10 cm. High accuracy positioning and rotation devices are used for this purpose. This leads to a total number of 30×21 images. For each image the measurement quality is evaluated and grades are assigned: All pixels valid, more than 80% valid, more than 20% valid and less than 20% valid.

In Fig. 2.9 the results for the test setup to identify difficulties in measuring sharp angles are shown. Measuring errors for angles up to 20 degrees less than the theoretical limit, i.e. the angle in which the front side of the plane



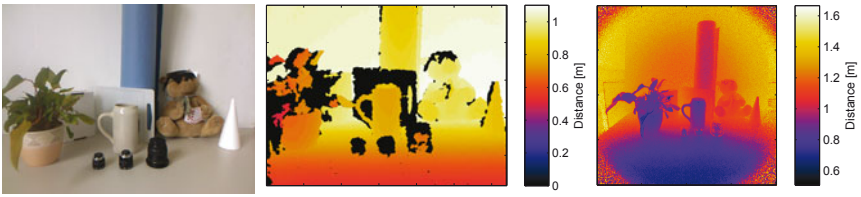
(a) Test setup **(b)** Results for the Kinect

Figure 2.9: Setup to test the ability of the Kinect to measure planar objects with different angles and positions relative to a camera path and results.

is invisible, are encountered. It is noteworthy that the left side of the depth map is affected significant higher than the right side. This is where the grayscale camera is located and therefore, the angle under which the incident light strikes the plane is here smaller than on the right side.

2.2.5 Limitations

In this evaluation and in previous experiments the different types of cameras displayed different weaknesses. The Kinect showed problems with dull (LCD monitors) or shiny surfaces or surfaces under a sharp viewing angle. Obviously, mounting the Kinect is relatively difficult and the lens is not exchangeable, which limits its application. Different lenses in combination with different diffractive optical elements might for example allow for larger distances. These drawbacks might be solved in different hardware implementations, but the largest problems are caused by systematic limitations. A significant part of a typical depth map contains no measurements due to shading: certain regions of the objects seen by the grayscale camera are not illuminated by the IR light beam. Depending on the application these areas can cause huge problems. In Fig. 2.10 a challenging test scene is shown. Here black indicates invalid measurements in the depth map for the Kinect.



(a) Color image **(b)** Kinect (black: invalid) **(c)** PMDTec CamCube

Figure 2.10: Resulting depth maps of a difficult test scene using the Kinect and the PMDTec CamCube.

Daylight is another source of problems. Since the grayscale chip of the Kinect uses an optical filter only infrared light disturbs the measurements. Therefore, a high power infrared LED with a peak wavelength at 850 nm and an infrared diode with corresponding characteristics have been tested to give an impression at which levels of infrared ambient light the Kinect can be used. It has been determined that measuring errors occur for an irradiance of $6 - 7\,\mathrm{W/m^2}$ depending on the distance. For comparison: sunlight at sea level has an irradiance of about $75\,\mathrm{W/m^2}$ for wavelengths between 800 and 900 nm.

The limitations of PMD based ToF cameras are mainly motion artifacts, which occur when objects move significantly during the acquisition of the four phase images. Another problem are mixed phases, which are produced when a pixel observes modulated light with different phase shifts due to reflections or borders of objects inside a pixel. Additionally, the low resolution and the higher power requirements limit the application of ToF cameras to some degree.

2.3 Summary

In this chapter a consumer depth camera, the Microsoft Kinect working with a novel depth imaging technique, is compared to state-of-the-art continuous wave amplitude modulation Time-of-Flight cameras. A set of experimental setups was devised to evaluate the respective strengths and weaknesses of the cameras as well as the underlying technology.

It was found that the new technique as well as the available device poses a strong competition in the area of indoor depth imaging with small distances. Only the problems caused by the triangulation, namely shading due to

different viewpoints, measuring difficulties of sharp angles and measuring of small structures are major weaknesses of the Kinect.

The Kinect as well as the DepthSense are not able to measure distances under high illumination, especially in sunlight. The PMD Time-of-Flight imaging chips with SBI are able to measure distances in sunlight, but with significantly lower exposure times, i.e. one millisecond, leading to a reduced quality. Therefore, acquiring full frame depth measurements at high frame rates in an outdoor environment or for longer distances is the domain of ToF chips with SBI mechanism up until today. For indoor scenes higher resolutions like the currently developed 100k PMD chip by PMDTec may level the playing field again.

In the following chapters PMD chips will be used as a basis to develop depth imaging techniques for medium and long distances due to the severe limitations of the other devices.

http://www.springer.com/978-3-658-06456-3

Wide Area 2D/3D Imaging
Development, Analysis and Applications
Langmann, B.
2014, XIV, 136 p. 71 illus., Softcover
ISBN: 978-3-658-06456-3