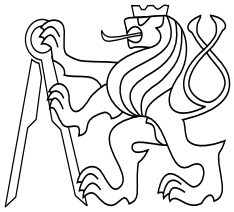
CENTER FOR
MACHINE PERCEPTION



MASTER'S THESIS

ISSN 1213-2365

3D Camera Calibration

Bc. Jan Smíšek

smiseja1@fel.cvut.cz

CTU–CMP–2011–06

June 2, 2011

Available at
http://cmp.felk.cvut.cz/~smiseja1/

Thesis Advisor: Ing. Tomáš Pajdla, Ph.D.

Research Reports of CMP, Czech Technical University in Prague, No. 6, 2011

Published by

Center for Machine Perception, Department of Cybernetics
Faculty of Electrical Engineering, Czech Technical University
Technická 2, 166 27 Prague 6, Czech Republic
fax +420 2 2435 7385, phone +420 2 2435 7637, www: http://cmp.felk.cvut.cz

3D Camera Calibration

Bc. Jan Smíšek

June 2, 2011

Declaration

I hereby declare that following thesis is my own work and I used only sources (literature, projects, SW etc.) quoted in enclosed reference list.

In Prague, June 2, 2011

Bc. Jan Smíšek

Acknowledgement

I would like to express gratitude to my diploma thesis supervisor Tomáš Pajdla for providing me with much needed aid and explanations of the topic. I would like to thank Gerhard Paar and Ben Huber from Joanneum Research for helping me with their Swissranger camera. I would also like to thank Michal Jančošek from CMP for his help with multi-view reconstruction.

I very much appreciated the support and patience of my girlfriend Klára and my parents during my work on the thesis.

Abstract

We studied the topic of depth sensing camera calibration. Two devices Microsoft Kinect and Swissranger SR-4000, that work on different physical principles, were investigated. Both 3D cameras were described and subjected to experiments in order to evaluate their performance. Several systematic error sources were identified and we proposed methods to compensate for them. A comparison of reconstruction performance of both 3D cameras and a stereo-pair of conventional cameras was presented. Finally, we showed an application of the depth sensing camera together with conventional color camera in area of complex scene reconstruction.

Abstrakt

Tato diplomová práce se zabývala problematikou kalibrace dálkoměrných fotoaparátů. Byla prostudována dvě zařízení (Microsoft Kinect a Swissranger SR-4000) pracující na odlišných fyzikálních principech. Oba 3D fotoaparáty byly popsány a podrobeny experimentům za účelem zhodnocení jejich přesnosti. Podařilo se identifikovat některé zdroje systematických chyb a byly představeny postupy, které tyto vlivy kompenzují. Oba dálkoměrné fotoaparáty byly porovnány z hlediska jejich přesnosti oproti klasické stereovizní metodě. Na závěr je ukázáno možné použití výsledků práce na rekonstrukci 3D modelu složité scény.

Czech Technical University in Prague
Faculty of Electrical Engineering

Department of Control Engineering

DIPLOMA THESIS ASSIGNMENT

Student: **Bc. Jan Smíšek**

Study programme: Cybernetics and Robotics
Specialisation: Systems and Control

Title of Diploma Thesis: **3D Camera Calibration**

Guidelines:

1. Review the literature about standard camera and 3D camera calibration [1-9] and its implementation [2].
2. Propose, design and implement a camera and 3D camera calibration technique for a triangulation-based 3D camera (e.g. Kinect) and a 3D-TOF camera (e.g. Swiss Ranger).
3. Validate the calibration technique on simulated data for both cameras and on real data from a 3D Kinect camera and (if data will be available) also for the 3D TOF camera.

Bibliography/Sources:

[1] R. Hartley and A. Zisserman. Multiple View Geometry in Computer Vision.
[2] J-Y. Bouguet. Camera Calibration Toolbox.
[3] S. May, D. Droeschel, D, Holz . Robust 3D-Mapping with Time-of-Flight Cameras.
[4] C. Schuon, C. Theobalt, J. Davis, S. Thrun, LidarBoost: Depth Superresolution for TOF 3D Shape Scanning.
[5] F. Chiabrando, R. Chiabrando, D. Piatti, F. Rinaudo. Sensors for 3D Imaging: Metric Evaluation and Calibration of a CCD/CMOS Time-of-Flight Camera.
[6] S. Fuchs and G. Hirzinger. Extrinsic and Depth Calibration of ToF-cameras.
[7] A.Kolb, E. Barth, R. Koch, R. Larsen. Time-of-Flight Sensors in Computer Graphics.
[8] N. Burrus. Kinect Calibration.
[9] ROS Kinect.

Diploma Thesis Supervisor: Ing. Tomáš Pajdla, Ph.D.

Valid until summer semester 2011/2012

prof. Ing. Michael Šebek, DrSc.
Head of Department

ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE

prof. Ing. Boris Šimák, CSc.
Dean

Prague, February 23, 2011

List of Figures

List of Tables

List of Procedures

Contents

1 Introduction

Precise cameras are instruments of great importance for the planetary rover missions . Common configuration is a high-resolution color stereo-pair of CCD cameras mounted on an extensible bar. New design conception is to augment the stereo- pair of cameras by a 3D-TOF camera (which can also deliver a depth map of captured scene utilizing time-of-flight principle). The thesis focuses on calibration of depth sensing cameras in order to improve their performance. A method, that allows using depth sensing camera together with a stereo-pair of cameras is also described.

Such procedures can be possibly used in future planetary rover missions. Work was conducted as a part of ProViSco project in cooperation with Joanneum Research in Graz, Austria.

1.1 Background

A contemporary way how the planetary rovers operate (e.g. Mars rovers Spirit and Opportunity [53, 54]) is that they perform a predefined task, capture the scientific data and transmit them back to Earth for further processing. Because of the wast distance between Mars and Earth the data bandwidth is so narrow that it usually takes several days to transmit the images acquired in one day. Until then the rover can be working on other predefined tasks - risking that the analysis of previously transmitted will discover some object of scientific interest and require the rover to return. The other method would be to wait until the data are analyzed. Both approaches are extremely time consuming and significantly reduce the scientific output of the rover mission (given that the rover lifetime is limited) [11].

The project PRoViScout aims to overcome this obstacle - it will transfer the mission-planning intelligence directly to the rover. The main idea is to allow the robot to operate relatively autonomously both in navigation and also in selection of targets with scientific interest. This method could be for example applied in the future ESA ExoMars rover mission (see image 1.1).

1.2 Motivation for using a 3D Camera

To allow precise terrain reconstruction for navigation and for scientific targets identification purposes a pair of cameras is nowadays used (see Fig. 1.2a). The advantage of stereo vision over other range measuring devices is, that it acquires high resolution color images of the scene together with the depth information. The main idea is to find corresponding points in images from both cameras and calculate the distance using triangulation. The major problem is to identify the correspondences precisely - which could be difficult for scenes without many visual details and it also has high computational demands.

Figure 1.2b shows the operational principle of Time-of-Flight camera vision systems. The range information is measured by emitting a modulated near-infrared light signal and computing the phase difference of the received reflected light signal. Using the

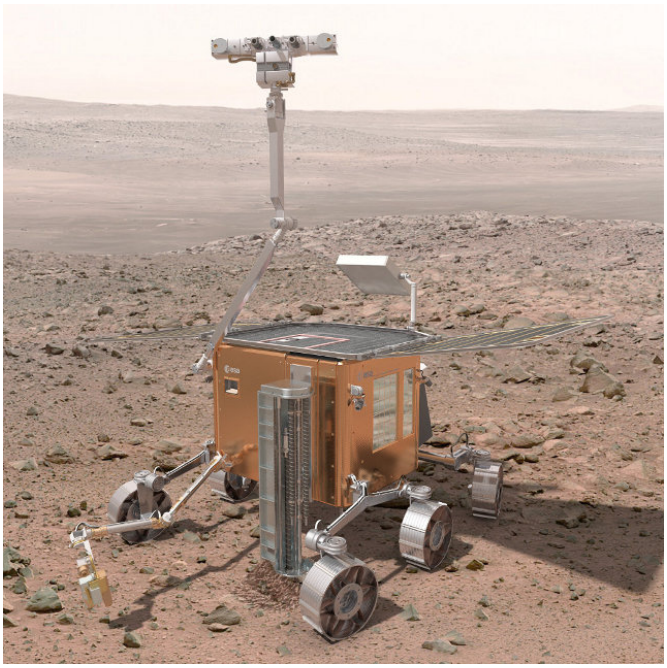
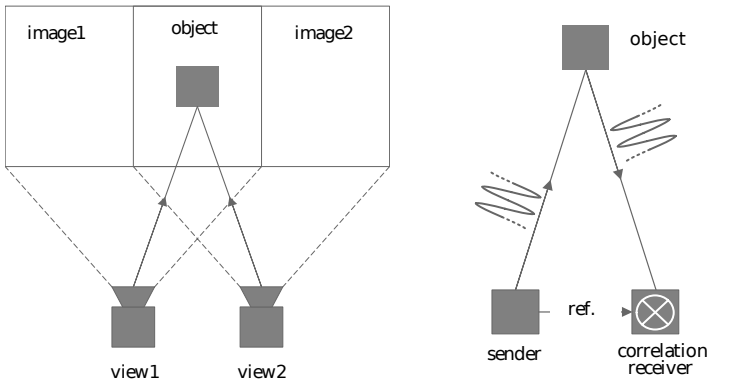


Figure 1.1 ExoMars rover - phase B1 concept (Courtesy of ESA).



a) Stereo vision based system **b)** Time of Flight camera

Figure 1.2 Comparison of working principles of Stereo vision and TOF depth sensing systems (Courtesy of S. Hussmann).

TOF camera the distance calculation is done in each individual pixel of the sensor. For further discussion and comparison of those principles please refer to [34].

1.3 ProVisG / PRoViScout Projects

Both ProVisG and PRoViScout projects aim to increase the scientific outcome of the future rover missions ([9] and [10]).

PRoVisG will build an unified European framework for Robotic Vision Ground Processing. State-of-art computer vision technology will be collected inside and outside Europe to exploit better the image data gathered during future planetary missions. The idea is to provide the operator on Earth with all visual data, as if he would be standing on the surface of other planet.

The ProViScout target is to equip the rover with enough *intelligence*, so that it would be able to handle local navigation and hazard avoidance on its own. Moreover, the rover will autonomously decide about scientific importance of the surrounding environment and automatically collect relevant data.

2 State of the Art

2.1 Calibration of Microsoft Kinect

Although the Kinect was introduced as a gaming controller for Xbox 360 console from Microsoft, a way to use it with personal computer was quickly discovered by the open-source community [5]. Many different Kinect related computer visions projects are nowadays represented by Open Kinect initiative [7] - above others the *libfreenect* drivers necessary for using the device with PC. Currently new drivers *OpenNI* from manufacturer of Kinect's chipset Prime Sense have been released [8]. These drivers allow more options to control the hardware e.g. to use full resolution of the RGB-camera. Microsoft is planning to release official Kinect SDK for PC till spring 2011 [2].

Question of device calibration is threated by Nicolas Burrus [23] from University of Madrid. He maintains a software package *Kinect RGB Demo* currently in version 0.5. His procedure consists of photogrammetric calibration, correction for image distortions, stereo calibration between IR and RGB-camera and reconstruction of 3D scene as a colored point cloud. The program also includes some methods to reconstruct complex 3D environments. The depth calibration model seems to be disparity based even though it is not stated explicitly.

A group around Robot Operating System [1] published details on their calibration procedure that contains photogrammetric calibration, correction for image distortions, stereo calibration between IR and RGB-camera and reconstruction of 3D scene as a colored point cloud. They present extensive analysis of how Kinect probably operates. They use a disparity based depth measurement model that is similar to ours.

A research group from Intel Labs Seattle [12] is using the Kinect in indoor enviroment reconstruction, interactive projection systems and object recognition for robotic manipulation.

2.2 Calibration of Time of Flight Cameras

The topic of TOF camera calibration is already threated extensively in the literature. Several systematic error sources were identified and authors propose different methods to correct for them. A complete overview of TOF calibration methods and related computer vision techniques is given in [39] and in [45]. Rapp in [33] provides a comparison of TOF camera devices from different manufacturers.

Since capturing a lot of refference datais often necessary, new ways how to do it effectively are proposed. Kahlmann *et al.* uses a automatic optical bench with moving target. Other approach is to attach the camera on robotic manipulator, which was done by Fuchs *et al.* in [27]. A procedure which is now widely use is to augment the TOF camera with other conventional calibrated camera and use it to measure the pose of the target [30, 25, 18, 42]. This also allow to combine both principles to achieve higher precision.

Attempts to increase low resolution of TOF cameras were made mostly by *super-resolution* technique - where multiple point clouds of captured object from slightly different position are registered together and used to interpolate more dense depth image

raster (as in [50]). The topic of registration of multiple TOF and RGB cameras using ICP[1] is treated in [38, 44]. The question of TOF camera self-calibration is discussed by Becerro in [17].

[1] Iterative Closest Point is an algorithm, that matches togther clouds of points order to minimizethe difference between them.

3 Sensors

3.1 Perspective Camera

Perspective or pinhole camera is a common geometric model of an ideal camera. As is shown in Fig. 3.1 the model consists of *image plane* π and *camera center* C. The distance between π and C is called *focal length* f. *Optical axis* is a line going through C and perpendicular to π and its intersection is denoted as *principal point*. For further explanation the reader can refer to [51].

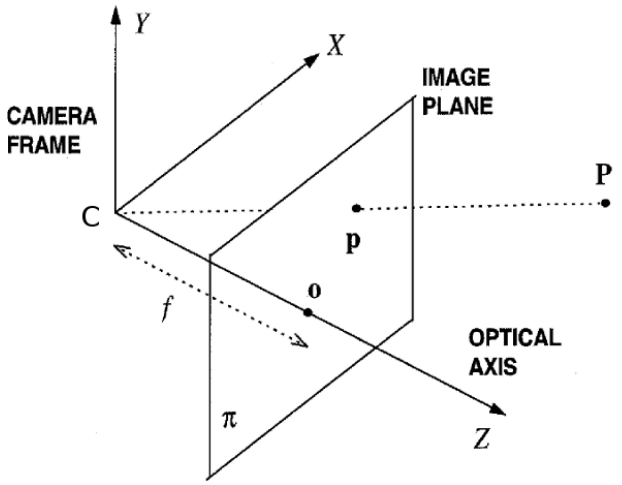


Figure 3.1 The perspective *pinhole* camera model (Figure taken from [51]).

3.1.1 Intrinsic Camera Parameters

Intrinsic parameters of the camera are kept in form of matrix

$$K = \begin{bmatrix} f/k_x & s & x_0 \\ 0 & f/k_y & y_0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{3.1}$$

where
- f is the focal length in [mm],
- $k_x,\ k_y$ are pixel size in respective directions in [mm/px],
- s is the skew factor (usually $s = 0$),
- $(x_0,\ y_0)^{\mathrm{T}}$ are the coordinates of the principal point in image plane [px].

3.1.2 Extrinsic Camera Parameters

Extrinsic parameters represent position and orientation of the camera in the world reference coordinate system. To be specific, such transformation can be realized using:
- $\mathbf{T}$ is the the translation 3x1 vector defining the position of camera center,

- R is an orthogonal 3x3 matrix representing rotation between coordinate systems.

Let $X_w(x_w,\ y_w,\ z_w)$ be a point in world reference frame and let $X_c(x_c,\ y_c,\ z_c)$ represent the same point in camera coordinate system (see Figure 3.2). The relationship

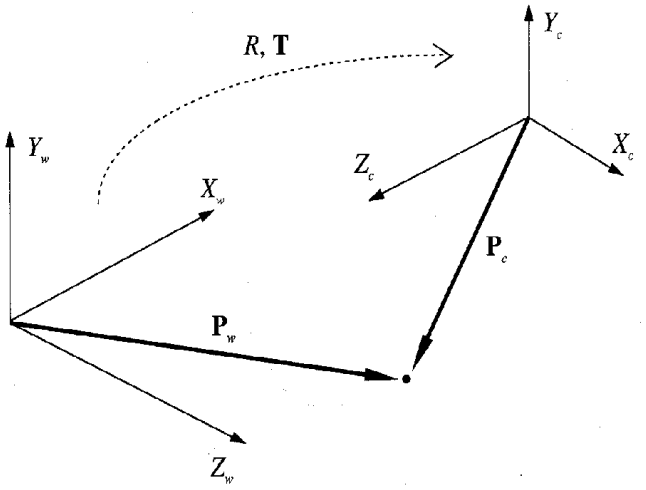


Figure 3.2 The relationship between world and camera reference frames (Figure taken from [51]).

between the coordinates can be described as

$$\begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} = R \begin{bmatrix} x_w \\ y_w \\ z_w \end{bmatrix} + T. \tag{3.2}$$

3.1.3 Distortion Model

Lens distortion model used in this work is a standard polynomial model, that was introduced by Brown in [22]. The way how the model is implemented in Camera Calibration Toolbox is described in details on the project page [19].

It represents both radial and tangential distortions by a 5-vector of coefficients k_c. Undistorted x_n and distorted x_d points coordinates are related as (where $r = x_n^2 + y_n^2$)

$$\begin{pmatrix} x_d \\ y_d \end{pmatrix} = \underbrace{(1 + k_{c_1} r^2 + k_{c_2} r^4 + k_{c_5} r^6) \begin{pmatrix} x_n \\ y_n \end{pmatrix}}_{\text{radial distortion}} + \underbrace{\begin{pmatrix} 2k_{c_3} x_n y_n + k_{c_4}(r^2 + 2x_n^2) \\ 2k_{c_4} x_n y_n + k_{c_3}(r^2 + 2y_n^2) \end{pmatrix}}_{\text{tangential distortion}}. \tag{3.3}$$

Effects of the distortions are visualized in Figure 3.3. Effects of radial (dr) and tangential (dt) distortion are illustrated. The points marked as *ideal* and *distorted* denote the projected point positions without and with the effects of the distortions. Note that the displacement caused by radial distortion (dr) is usually much larger than the one caused by tangential distortion (dt) [19].

It is often not necessary (possibly even wrong) to estimate higher order coefficients of the distortion model. When too complex model is selected the optimization algorithm can substitute the effects of the focal length by using stronger distortions. General information together with recommendations on selection of the distortion model can be found in [19].

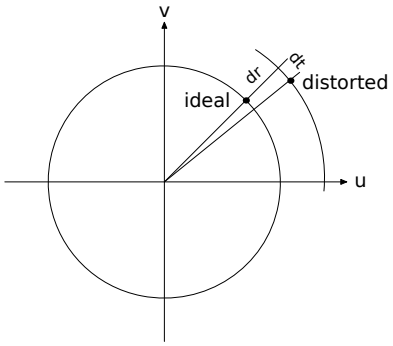


Figure 3.3 Image distortions introduced by camera optics. Effects of radial (dr) and tangential (dt) distortion are illustrated. The points marked as *ideal* and and *distorted* denote the projected point positions without and with the effects of the distortions. Figure is taken from [52].

3.1.4 Image Formation

Taking an image is a process that links points $X_i = \begin{bmatrix} x & y & z \end{bmatrix}^{\mathrm{T}}$ in 3D space to 2D points $x_i = \begin{bmatrix} u & v \end{bmatrix}^{\mathrm{T}}$ on the image plane. We denote this process as *projection* and define it by relation

$$\alpha \begin{bmatrix} x_i \\ 1 \end{bmatrix} = P \begin{bmatrix} X_i \\ 1 \end{bmatrix}, \tag{3.4}$$

where α is a scale factor and P is a camera projection matrix formed from combination of intrinsic and extrinsic camera parameters to be

$$P = KR[I \mid -\mathbf{C}]. \tag{3.5}$$

3.2 Stereo Vision Systems

Different principles of optical depth measurment are shown in Figure 3.4. Usually, stereo vision systems have two cameras separated in space. They are used to obtain different views of the scene. Correspondig points of the scene are extracted and their mutual shift is used to calculate the point distance.

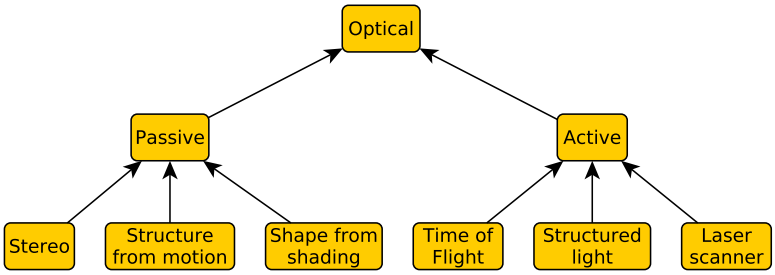


Figure 3.4 Different depth sensing principles (Based on lecture notes of Marc Pollefeys).

Example of two avalible integrated stereo camera solutions is given in Tab. 3.1.

Theorethical uncertainity associated with the stereo reconstruction is illustrated in Fig. 3.5. Since the physical size of the pixel is greater than zero, there will always be an uncertainity regieon with typicall diamon shape.

Table 3.1 Examples of modern integrated stereo cameras

Camera Model	Kinect	Bumblebee 2
Manufacturer	Microsoft Corp.	Point Grey Res.
Resolution [px]	640 x 480	648 x 488
Method	Stereo (active)	Stereo
Baseline [cm]	7.5	12
Field of View [deg]	57 x 43	66 x 66
FPS [s^{-1}]	30	48
Range [m]	0.4 – 8	not specified

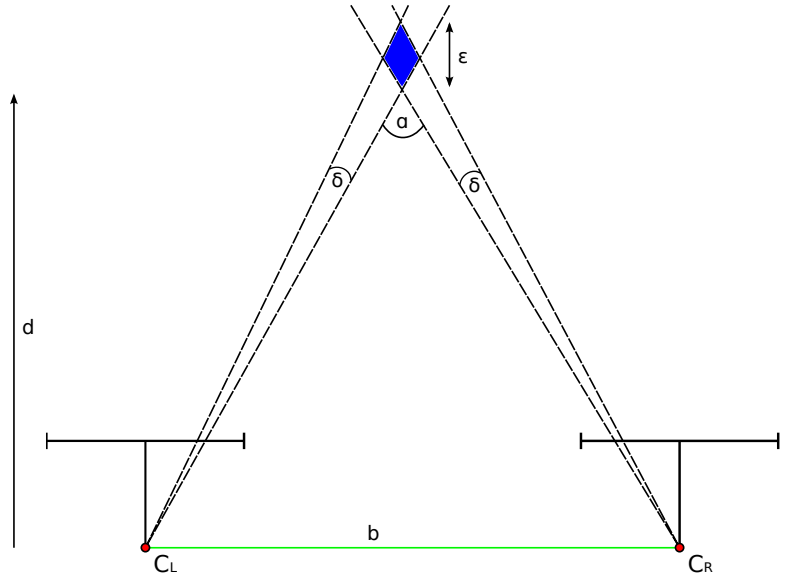


Figure 3.5 Uncertainty associated with stereo reconstruction. The blue region represents the uncertain position of the reconstructed point.

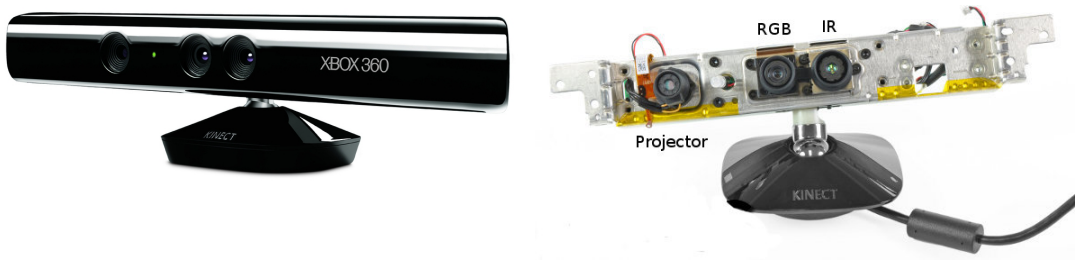
3.3 Microsoft Kinect

To be precise, Microsoft Kinect device (developed by PrimeSense Ltd.) uses a method called *active stereo* (or *structured light scanning*) where one camera from the stereo pair is replaced by a projector transmitting a known pattern onto the object. The IR-camera and the IR-projector form a system that is geometrically the same as a standard rectified stereo pair of cameras.

Originally, it was intended as a controller for Microsoft XBox 360 game console but shortly after release it became popular among computer vision community. It is an integrated device consisting of two cameras (RGB and IR) and one laser-based IR projector (shown in figure 3.6). Basically, the IR laser source projects a constant pattern[1] onto the target, which is then detected by the IR camera. Changes in the captured pattern image are used to calculate the distance for each pixel. Basic technical parameters are listed in table 3.1.

The camera sensors were identified in [36]. The IR camera uses the MT9M001C12STM CMOS sensor [14] from Aptina with resolution 1280 x 1024 pixels of size 5.2 μm. The

[1] Images of IR laser projected points as well as identification of some repeating patterns can be found in [48]. This principle is in literature referenced as Structured light [55].



a) Kinect (Courtesy of Microsoft Corp.) **b)** Kinect without chassis (Courtesy of iFixit)

Figure 3.6 Microsoft Kinect.

RGB camera is the MT9M112 CMOS sensor [15] also from Aptina with resolution 1280 x 1024 pixels of size 2.8 μm. Both cameras operate in 2 x 2 binning mode to allow faster FPS rate.

Output from the device consists of 3 types of images:
- Depth image 640 x 480, 11-bit values,
- IR-camera intensity image 640 x 480, 8-bit values,
- RGB-camera image 640 x 480 (1280x1024), 3 x 8-bit values.

An example of output images is shown in Figure 3.7.

3.3.1 Depth Sensing

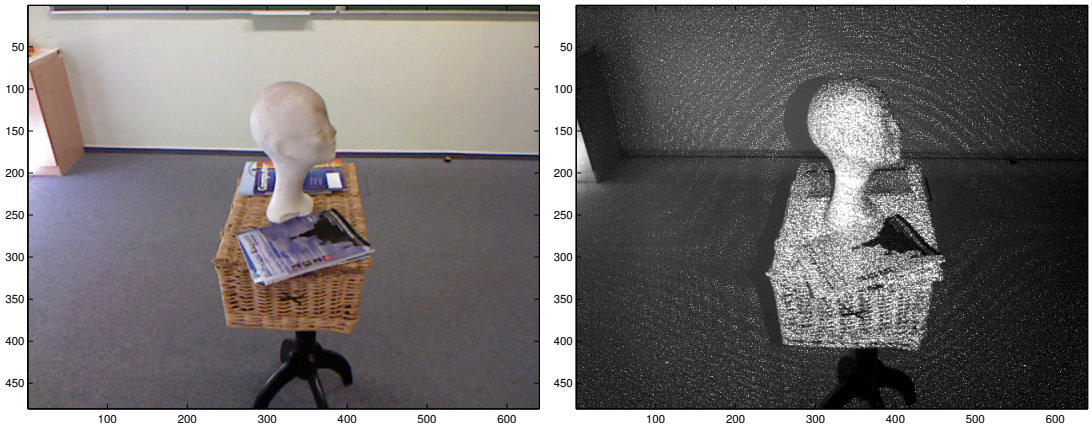
Detailed description of how the system really works is not available for public, and therefore the later is only an educated guess (based on experiments and available 3rd party references).

As a depth information an array of 640 x 480 11-bit unit-less integer values is returned. The pattern projected by the IR-laser is captured by the IR-camere (see Fig. 3.9a) and local difference to the preprogramed pattern image is calculated to yeald the distance measurment. If we have a look at the raw depth image 3.9b, there is a 8 px wide empty band. As is discussed in [1] the band is most probably caused by a correltation window of size 9 x 9, that is used for the depth calculation.

The lowest raw depth number we measured is 415, that corresponds to approx. 48 cm. The highest number is 1070 representing approx. 1500 cm. Points without depth information are filled with value 2047. As a result of experiment 3.3.1 raw depth values with corresponding measured distances are shown in Figure 3.11.

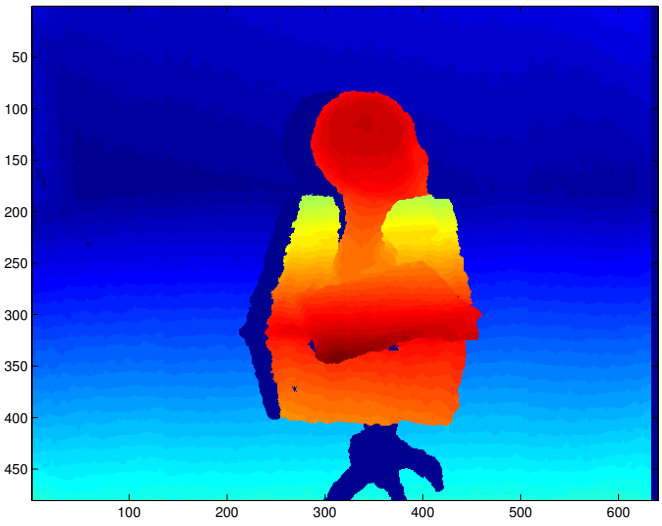
Estimation of Depth Resolution

In order to determine its estimated depth resolution the depth sensing camera was mounted on precise optical bench. At first, two flat targets were used: a square target placed perpendicularly to the optical bench and a rectangular target tilted in approx. 45° from the bench axis were captured at increasing distances between 48 cm and 160 cm from the camera with 1 cm step. At each position two depth images were taken - to allow having learning and testing sets of data separated. The second part of the experiment was done in a bigger hall for distances between 160 cm and 1500 cm with 10 cm step (again 2 images at each distance were taken). This time the target was a square part of the wall. The setup is shown in Figure 3.10. For each image small square



a) RGB-camera image

b) IR-camera image



c) Depth image

Figure 3.7 Example of Kinect output images.

area in the image center was extracted and robustly fitted by a plane. This was done to reject the outliers.

If we have a closer look (at Figure 3.12), it can be easily seen that the raw depth values are quantized with size of the step depending on the actual distance from the camera. To evaluate the size of the quantization step for each distance of the measurement we took all distances (from straight and tilted targets) calculated by the Kinect[2]. From this set we consider only sorted, unique distance values. If we now calculate differences between subsequent values, we obtain the size of the quantization step. The result of this calculation is show in Figure 3.12. In the plot 3.12a the measurement of the tilted plane was available - that provided a very dense set o measured distances. The chart also shows standard deviation and maximum of the depth values difference at the given distance. In the second Figure 3.12b only the flat plane was captured with 10 cm intervals. The quantization can be determined only on the distances where it gets higher than the distance step (approx. 5.5 m and more).

To summarize the experiment, the size of the quantization step was found to be a

[2]The actual metric distance was calculated from raw value using 3.6 with the constants determined by procedure from chapter 4.1.3.

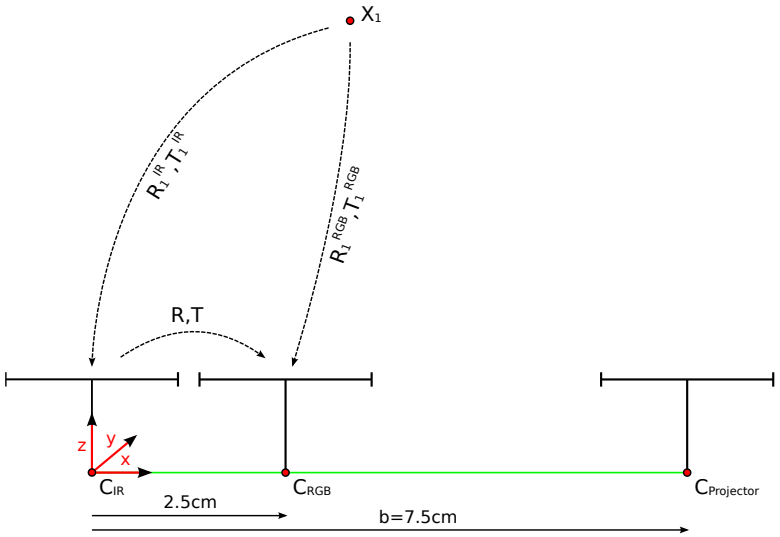


Figure 3.8 Kinect geometrical model.

function of distance with estimated relation (for distance d in meters):

$$q_{\text{step}}(d) = \begin{cases} 1 + 6\ d\ [\text{mm}] & \text{for } 0.5 < d < 1.6 \text{ m} \\ 100 + 50\ d\ [\text{mm}] & \text{for } d > 5.5 \text{ m} \end{cases}$$

Distance Model

Illustration of disparity based depth measurement is shown in Figure 3.14. Camera centers are detonated as C_L and C_R, b denotes the baseline between them and f is the focal length. Relationship between point depth z and distances x_L, x_R that are given by intersection of the rays going through the point X and the image plane can be derived (in details shown in [51]) using *similar triangles* as

$$\frac{b}{z} = \frac{b + (x_R - x_L)}{z - f}$$

and after substituting $d = x_R - X_L$ rearranged to

$$d = \frac{bf}{z}.$$

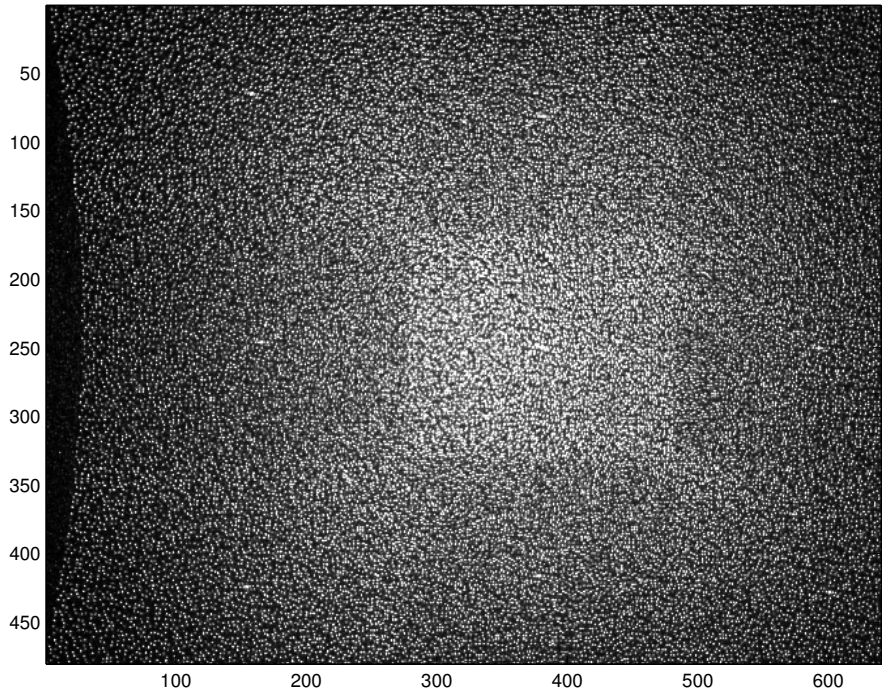
The assumption is that raw disparity r returned by Kinect can be related to distance d by first order polynomial $d = c_1 r + c_0$. For actual distance the depth model is

$$z = \frac{bf}{c_1 r + c_0}, \tag{3.6}$$

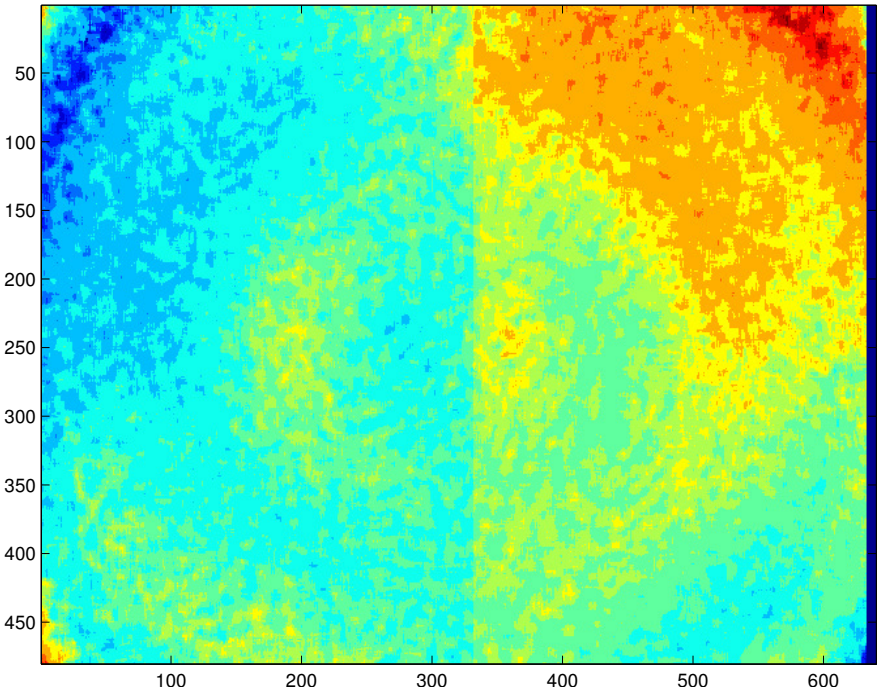
where b is baseline between IR-laser projector and IR-camera (7.5 cm), f is the focal length of the IR-camera and c_1, c_0 are coefficients of the model.

3.3.2 Error Distribution in Real Scene

In order to investigate the stability of the depth measurement in real-world environment a common room was with furniture, flat walls, wooden wardrobes and glass in the doors



a) IR-camera image



b) Raw depth iamge

Figure 3.9 Kinect IR and raw depth images of a flat wall. Note the pattern in the deph image (more in section 4.1.3).

a) Close range 48 - 160 cm **b)** Long range 160 - 1500 cm

Figure 3.10 Experimental setup for estimating of Kinect depth resolution. Distance measured by Kinect is compared with result from measure tape.

was captured (see Figure 3.15). The scene is quite diverse and so rather complicated for reconstruction.

The room was captured 50 times (with the Kinect attached on a tripod) and for each pixel position a standard deviation over all images was calculated. The histogram of occurrences is shown in Figure 3.16. The histogram shows that there are many pixel depth values that did not vary or vary slightly ($\sigma < 30$ mm).

To have better understanding of how the areas with higher deviations are distributed we can have a look at Figure 3.17. The bright regions represent the parts of the image with high uncertainity. These are mostly located on objects edges. The dark red areas mark the pixels where the depth information could not be calculated.

From the analysis of the scene we can conclude that the effect appears on sudden distance changes. We modeled this effect as a second derivative (difference) from a single depth image. If you compare the spatial distribution of the bright places on Figures 3.18 and 3.17 they coincide. This areas can be removed using simple thresholding without loss of any useful information.

3.4 Time-Of-Flight Cameras

Time-of-flight range cameras are relatively modern devices, that capture digital images together with distance information. Such approach can produces data with high FPS, less computational demand and using a very compact devices. However it also suffers from several shortcomings as low spatial resolution, necessity of active illumination etc. Several examples of state-of-the-art 3D-TOF cameras are given in table 3.2. More detailed description can be found in [24, 56].

3.4.1 Depth Sensing

The object distance R is calculated by measuring a round-trip time t that light (at speed $c \approx 3\ 10^8\ m/s$) needs to travel to the object and back to the camera

$$R = \frac{ct}{2}. \tag{3.7}$$

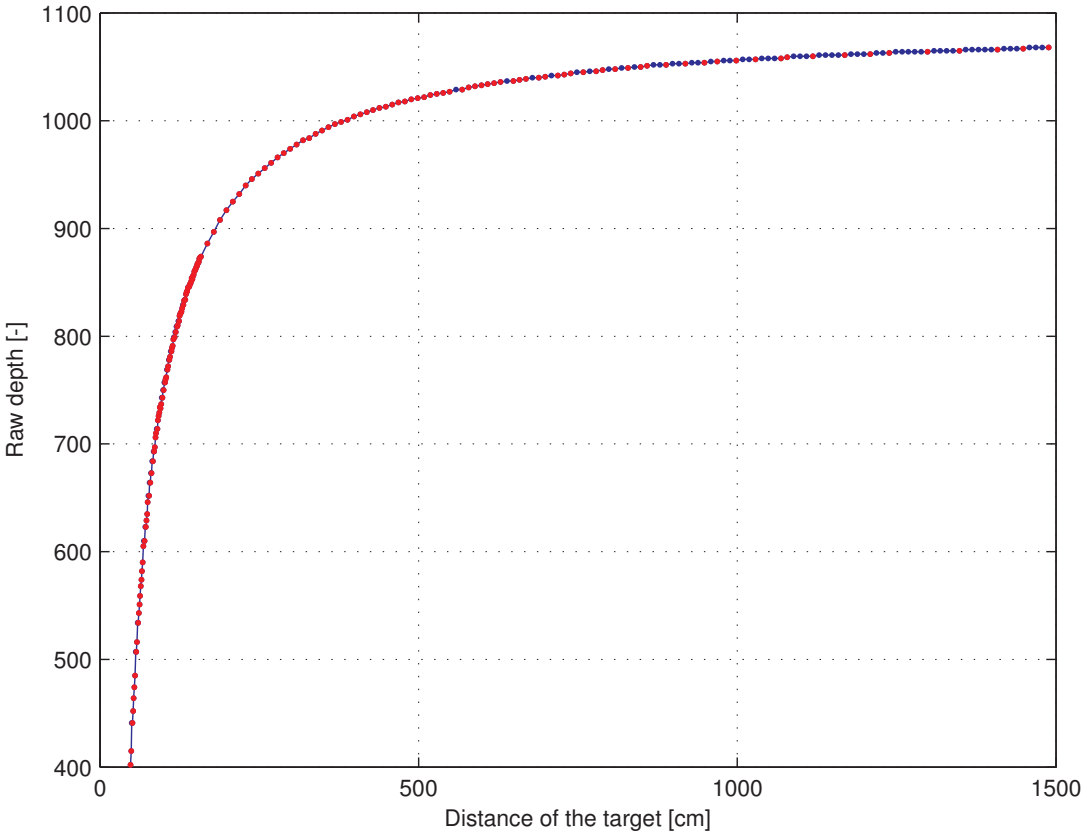
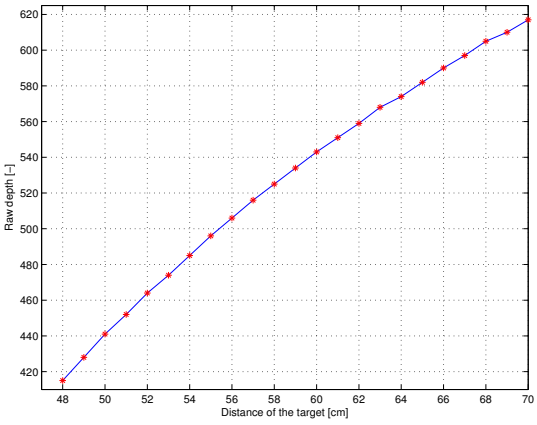
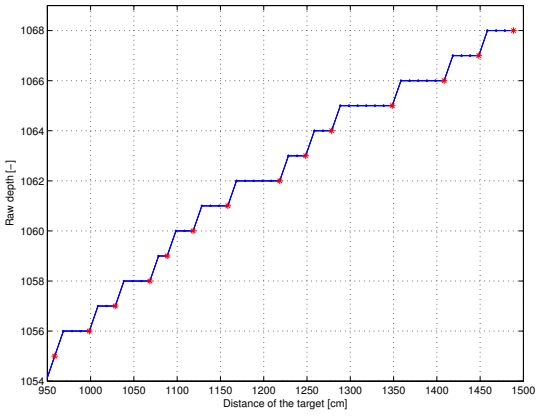


Figure 3.11 Kinect raw depth values with corresponding real distances (full range). Red dots mark the points at which the raw depth values changed since the last measurement.

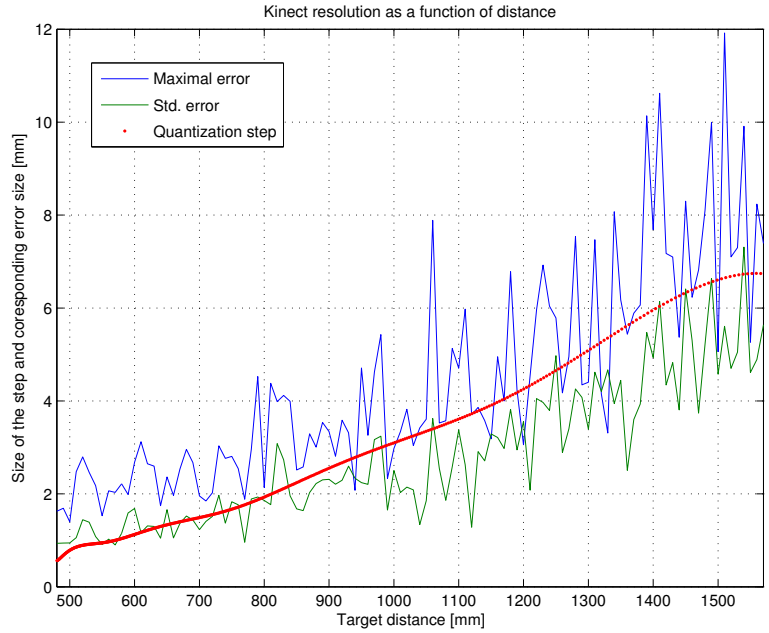


a) First 20 cm of measured distance (approx. 0.1 cm per one quantization step)

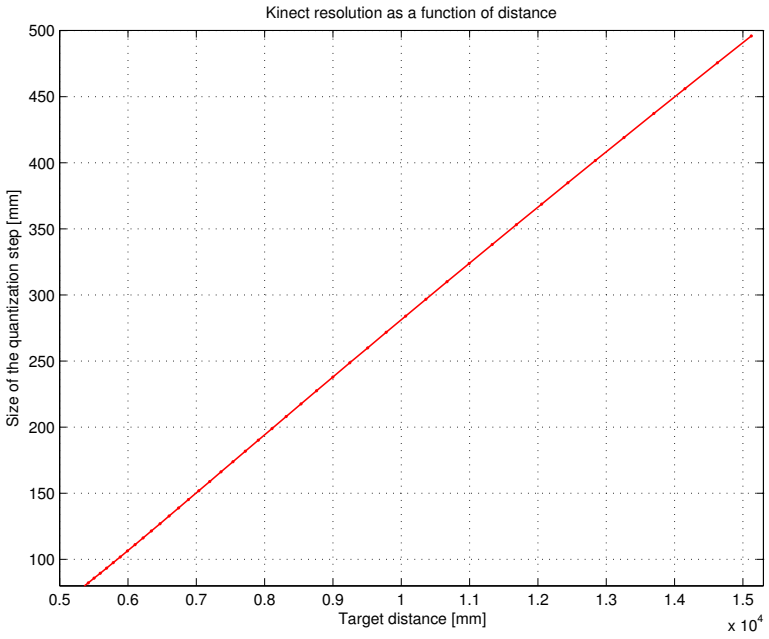


b) Last 500 cm of measured distance (approx. 40 cm per one quantization step)

Figure 3.12 Kinect raw depth values with corresponding real distances (beginning and end of the measured distance interval). Note the difference quantization in step.



a) Close range 48 - 160 cm



b) Long range 500 - 1500 cm

Figure 3.13 Sizes of the quantization step as a function of target distance.

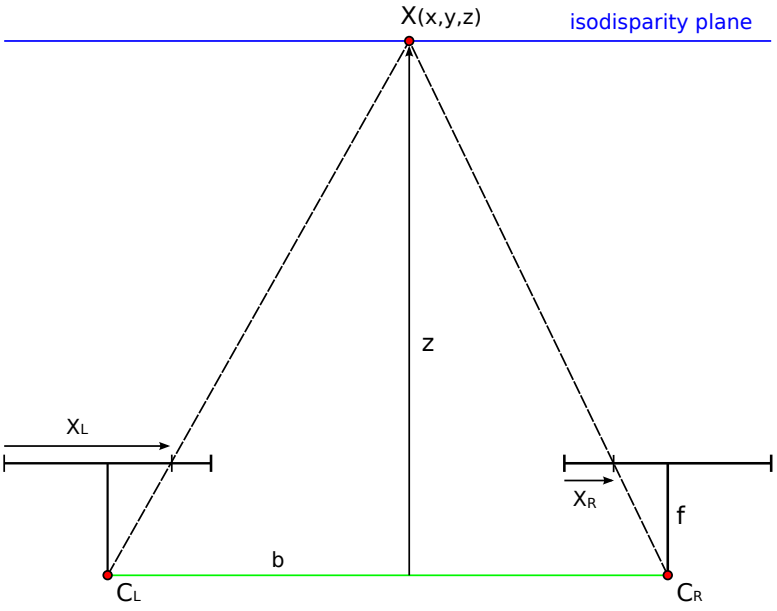


Figure 3.14 Illustration of disparity based depth measurement.

Figure 3.15 Scene for evaluation of error distribution in real-world environment.

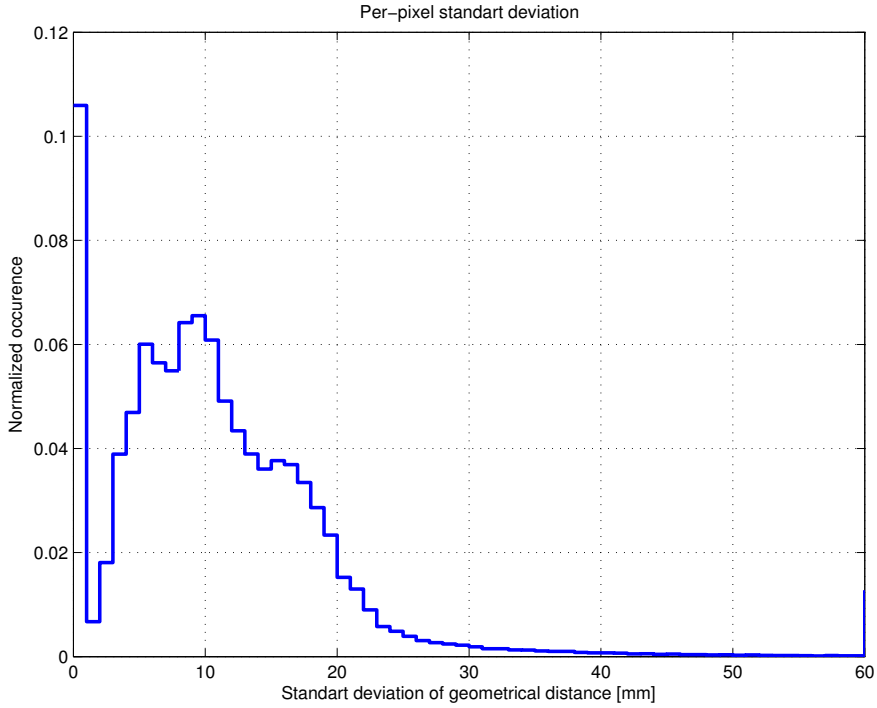


Figure 3.16 Scene for evaluation of error distribution in real-world environment.

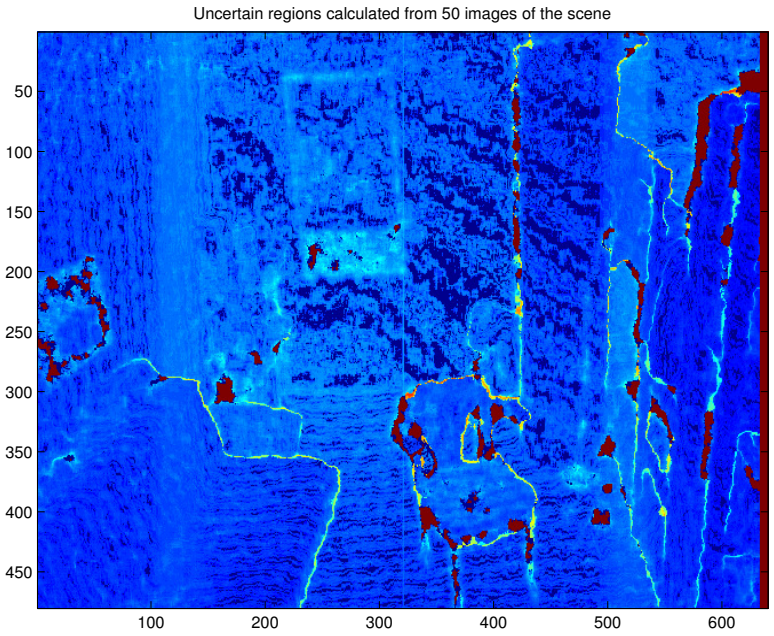


Figure 3.17 Uncertain regions calculated as pixel-wise standard deviation from 50 images of the scene.

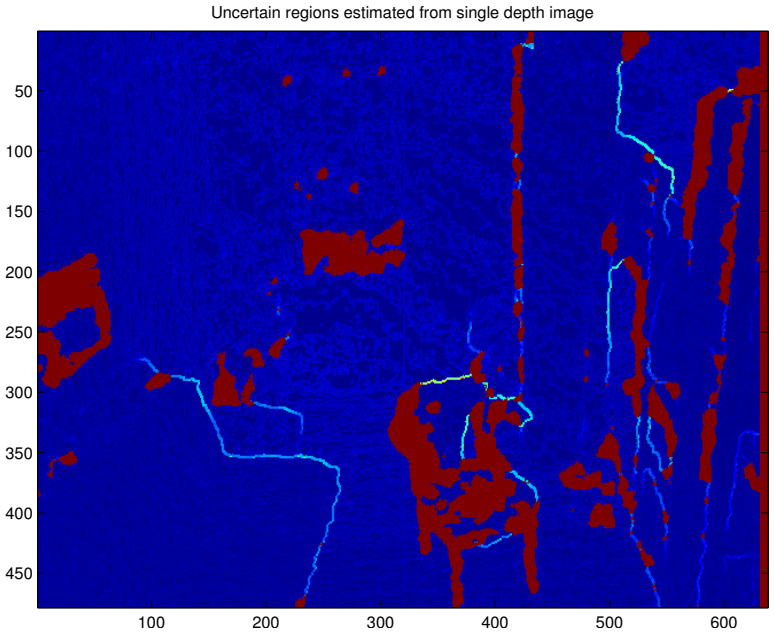


Figure 3.18 Uncertain regions estimated from a single depth image.

This can be measured either directly, which would require precise timing in order of picoseconds, or by using a continuously-modulated wave. The latter method has much lower requirements on the hardware components (e.g. frequency bandwidth, signal generation, timing base) and is principle limited only by the modulation frequency f_{mod} [24]. From now on we will focus only on continuously-modulated wave method, that is nowadays used in majority of commercial TOF cameras [33]. Basic idea is that the phase difference ϕ between the transmitted and received infra red signal is measured in each pixel. Because the modulation frequency f_{mod} is precisely known we can compute the distance as

$$R = \frac{c}{2f_{\text{mod}}} \left(\frac{\phi}{2\pi} + 2N\pi \right), \tag{3.8}$$

where $N = 1, 2, \ldots$ represent the the reflection from further objects for which the distance can not be uniquely determined i.e. *distance aliasing*. This leads to conclusion that the non-ambiguity range has a upper bound directly related to the modulation frequency by

$$R = \frac{c}{2f_{\text{mod}}}, \tag{3.9}$$

that for common $f_{\text{mod}} = 20$ MHz yields maximum range $R_{max} = 7.5\ m$ [34].

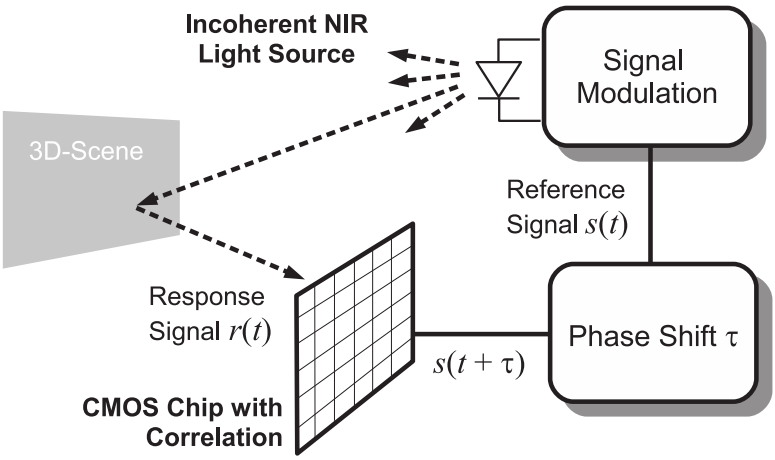


Figure 3.19 Illustration of PMD/ToF-measurement principle. Figure is taken from [42].

Table 3.2 Several examples of state-of-the-art 3D-TOF cameras

Camera Model	SR4000	C70	CamCube 3.0	D Imager	ZC-1000
Manufacturer	Mesa Imaging	Fotonic	PMD Tech.	Panasonic	Optex
Resolution [px]	176 x 144	160 x 120	200 x 200	160 x 120	160 x 120
Modulation [MHz]	29 / 30 /31	44	19 / 20 / 21	50	20 / 30
Illumination type	IR LED	IR Laser	IR LED	IR LED	IR LED
Field of View [deg]	44 x 35	70 x 50	40 x 40	60 x 44	70 x 55
FPS [s^{-1}]	50	75	40	30	60
Range [m]	0.8 – 5	0.1 – 7	0.3 – 7	1.2 – 9	0.5 – 4

3.4.2 Error Sources

Random Errors

The topic of random errors is threated extensively by Lange *et. al.* in [40]. We can briefly state that the major sources are

- electron shot noise,
- multiple-ways reflection,
- light scattering.

Systematic Errors

Several types of systematic errors were so far described in literature [45, 39].

- distance-related (wiggling) errors - caused by imperfections of the NIR LED; the transmitted signal is not harmonic which is however assumed and required by the method [45].
- intensity-related errors -as a result of physical propertied of the CCD detector the depth measurement is influenced by the total amount of incident photons,
- pixel fixed-pattern noise - fixed offsets of particular pixel caused by imperfections of the detector
- flying pixels - when an area with with different depths is observed by single pixel the phase shift calculation process can introduce artifacts.

3.5 Swissranger SR-4000

Swissranger SR-4000 is a compact TOF camera developed by Mesa Imaging AG company from Switzerland (Figure 3.21a).

Output from the device consists of 3 types of images (example is shown in Figure 3.20):

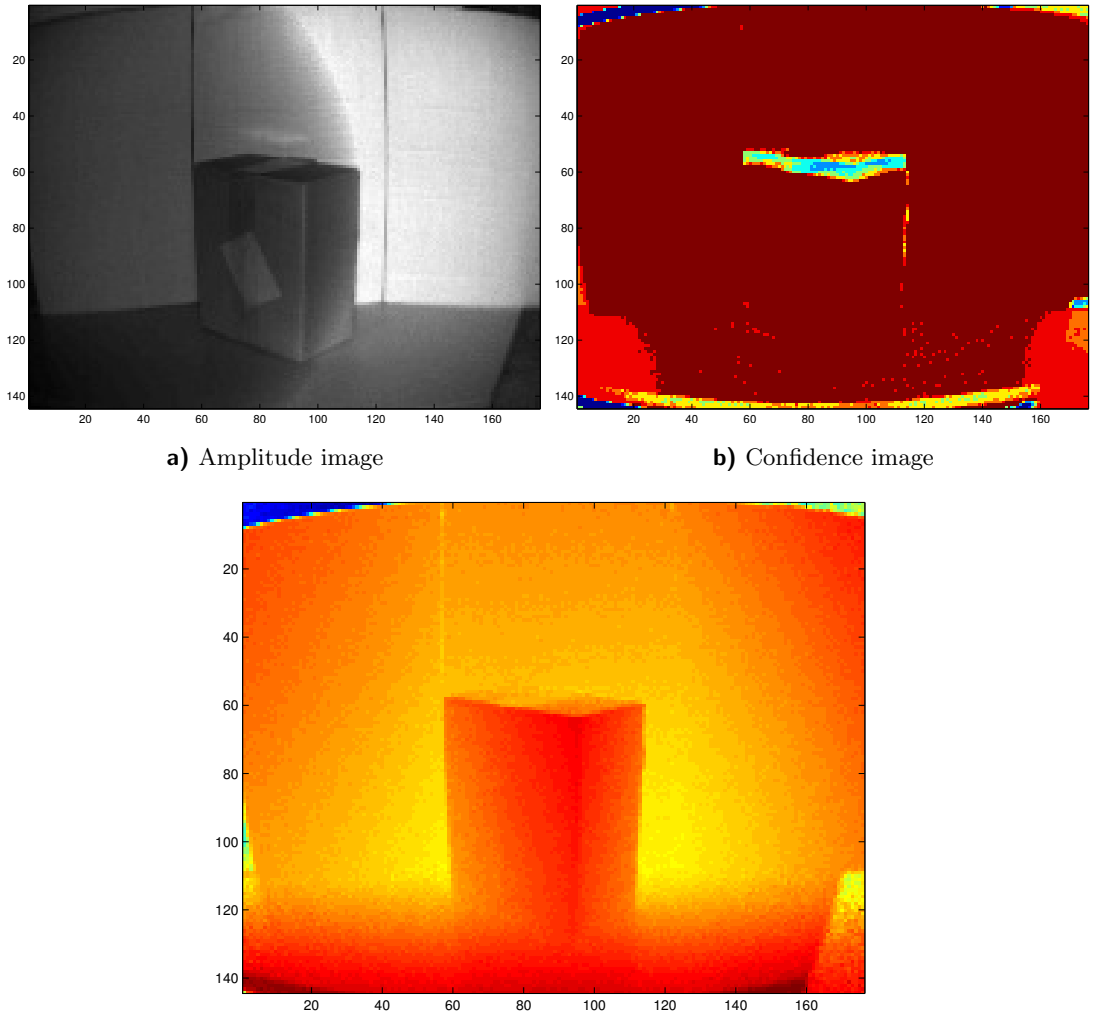
- depth image,
  - calibrated distance in Z direction, 176 x 144, float with 4 decimal points,
  - calibrated distance in X direction, 176 x 144, float with 4 decimal points,
  - calibrated distance in Y direction, 176 x 144, float with 4 decimal points,
- intensity image 176 x 144, 16-bit values,
- confidence map 176 x 144, 16-bit values, greater values representing higher confidence.

The software *SR 3D View*[3] provides distance data output in Cartesian coordinates, expressed in meters. During the transformation the data are corrected for effects of radial distortion of the optics.

Coordinate System

The directions of principle axis of the coordinate system are shown in Figure 3.21b. It should be noted that the coordinate system used by the camera software is not the same as we have considered so far. To transfer data from the Swissranger reference frame to the common perspective projection frame all data should be rotated around Z axis by π and translated. The paramters of the conversion were determined during the device calibration in section 4.2.1. Illustration 3.22 shows relationship between coordinate

[3] Available from the company web page `http://www.mesa-imaging.ch/`.



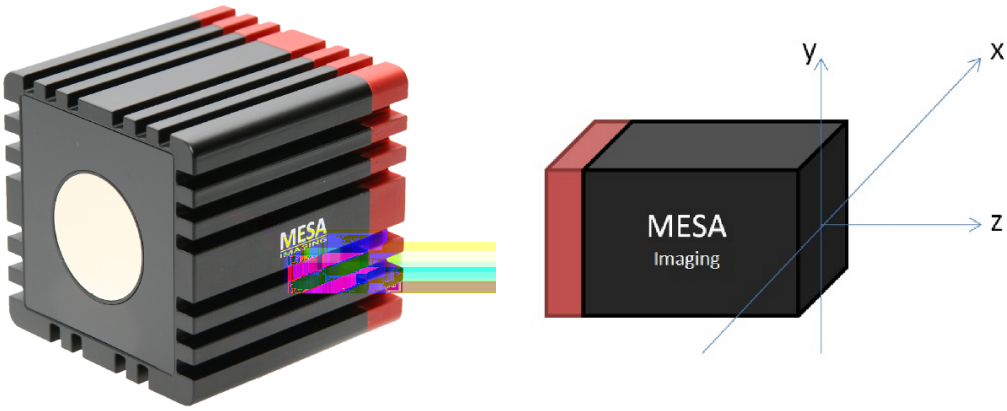
a) Amplitude image **b)** Confidence image

c) Depth image

Figure 3.20 Example of Swissranger SR-4000 output images.

system of TOF camera (shown dashed and denoted with x', y', z') and conventional reference frame used in this work (detonated with C, x, y, z).

Information about practical aspects of using the device can be found in User manual [35].



a) SR-4000 **b)** Used reference frame

Figure 3.21 TOF-camera Swissranger SR-4000 together with definition of the used coordinate system (Courtesy Mesa Imaging AG).

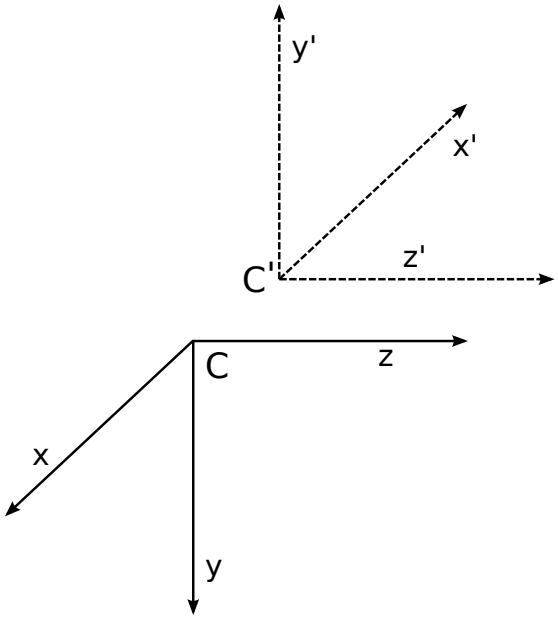


Figure 3.22 Geometrical model of SR-4000 TOF camera. Illustration shows relationship between coordinate system of TOF camera (shown dashed) and conventional reference frame used in this work.

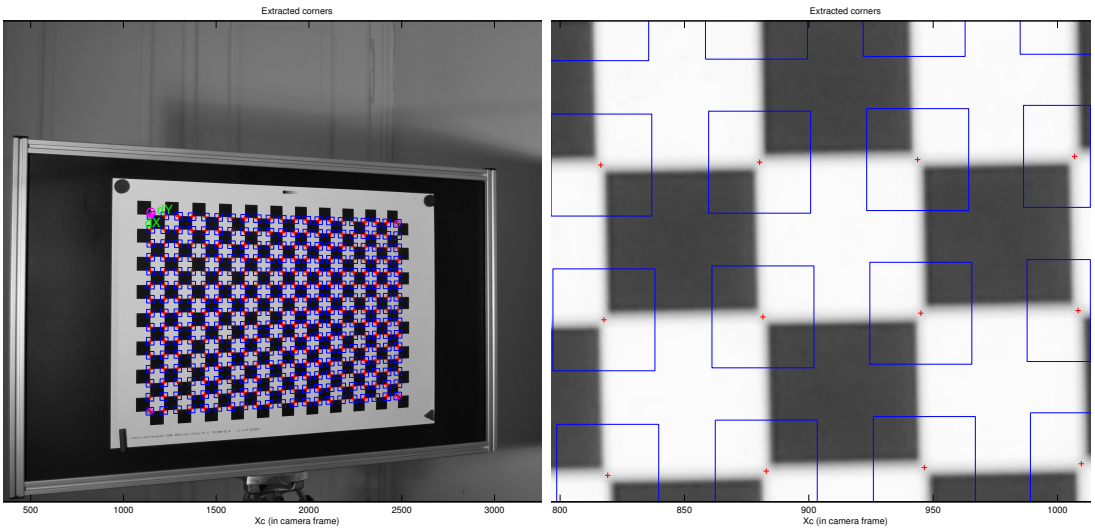
4 Calibration Procedure

4.1 Calibration of Kinect Device

For operating the device software package from Nicolas Burrus[1] with *libfreenect* drivers was used.

4.1.1 Photogrammetric Calibration

Camera photogrammetric calibration is a vital procedure in extraction of precise 3D information from captured images. Basically it is used to determine unknown variables in projection matrix (i.e. the intrinsic and extrinsic parameters). Moreover the calibration procedure is required to estimate coefficients of the distortion model (as was discussed in chapter 3). In widely used approach the planar target with calibration chessboard (example shown in figure 4.1) is imaged in different orientations in the camera's fields of view. Reference points (i.e. corners) are extracted with sub-pixel resolution from the images and used for estimating the projection matrix of the camera. This method is described in details by Zhang in [58]. It is a standard procedure implemented in many camera calibration software packages (e.g. OpenCV, Camera Calibration Toolbox for Matlab). Interested reader can refer to [49, 26].



a) Detected corners positions **b)** Detail of precise corner detection

Figure 4.1 Calibration chessboard with corners extracted using Calibration Toolbox.

In this work Jean-Yves Bouguet's Camera Calibration Toolbox for Matlab was used [19]. Detailed procedure is described in Proc. 4.1. For the experiment a chessboard with 20 mm squares providing 315 corners was used.

[1] Software Kinect RGB Demo is available at http://nicolas.burrus.name/index.php/Research/KinectRgbDemoV5 currently in i version 0.5.

Procedure 4.1 Capturing of the Calibration data

1: Turn on the halogen lamp and cover Kinect IR-laser projector. Capture image of the calibration chessboard by the RGB and IR-camera on the Kinect.
2: Capture image of the calibration chessboard by both DSLR cameras.
3: Turn off the halogen lamp and remove covering of the projector.
4: Capture more (3-5) depth images of the calibration target. Form one resulting depth image by taking median of values at same pixel position over all images.
5: Change position and orientation of the chessboard and repeat the procedure .

Corner Extraction

Photogrammetric calibration procedure starts with extracting corners coordinates of the calibration grid. It should be noted that IR-camera image is by default dark and covered by bright spots from the IR-laser projector and thus does not contain enough details for precise extraction of the the corners. Automatic corner detection function of the Calibration toolbox (that also enables sub-pixel detection) in this case often could not find right corner positions. This was successfully solved by using very small corner detection window and also by using strong illumination of the calibration target by halogen lamp while the IR-laser pattern projector was covered[2]. Comparison of images taken with and without IR-laser projector and additional light source is shown in Figure 4.2. On image 4.2c it is almost impossible to distinguish the corner coordinates.

Intrinsic Parameters and Distortion Model Coefficients

Parameters of intrinsic calibration matrix of the camera were found using the Calibration toolbox to be the following (see results for IR and RGB camera in Table 4.1 and 4.2 respectively). Since the image sensors are known (see section 3.3) to have square pixels, we can add constraint $f_x = f_y$ to the calibration procedure.

Table 4.1 Intrinsic parameters of Kinect IR camera

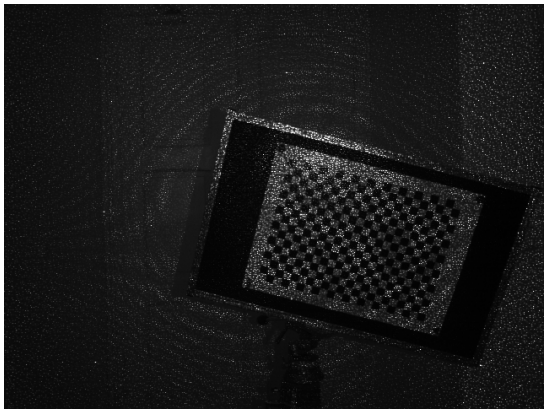
Focal length		Principal point		Distortion coefficients				
f [px]	f [mm]	x_0 [px]	y_0 [px]	k_{c_1}	k_{c_2}	k_{c_3}	k_{c_4}	k_{c_5}
585.6	6.1	316	247.6	-0.1296	0.45	-0.0005	-0.002	N.A.

Table 4.2 Intrinsic parameters of Kinect RGB camera

Focal length		Principal point		Distortion coefficients				
f [px]	f [mm]	x_0 [px]	y_0 [px]	k_{c_1}	k_{c_2}	k_{c_3}	k_{c_4}	k_{c_5}
524	2.9	316.7	238.5	0.2402	-0.6861	-0.0015	0.0003	N.A.

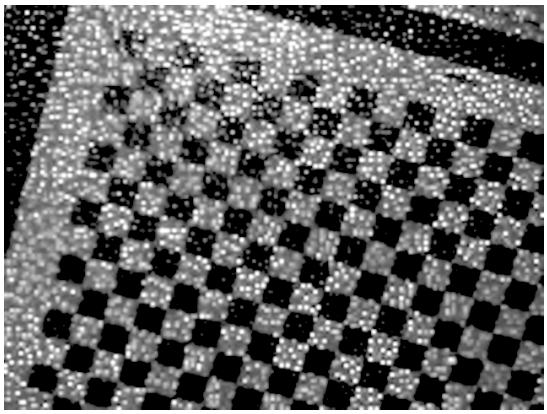
To visualize the impact of the distortions Figures 4.3 and 4.4 shows its effect on each pixel of the image. Arrows represent the direction of the pixel displacement induced by the lens distortion. The red numbers denote the actual size of the error. The cross indicates the image center, and the circle the location of the principal point.

[2]This was suggested by Alex Trevor from Kinect group at http://www.ros.org/ and by Nicolas Burrus from [23].



a) Illuminated by IR-laser projector

b) Illuminated by halogen lamp, projector is covered



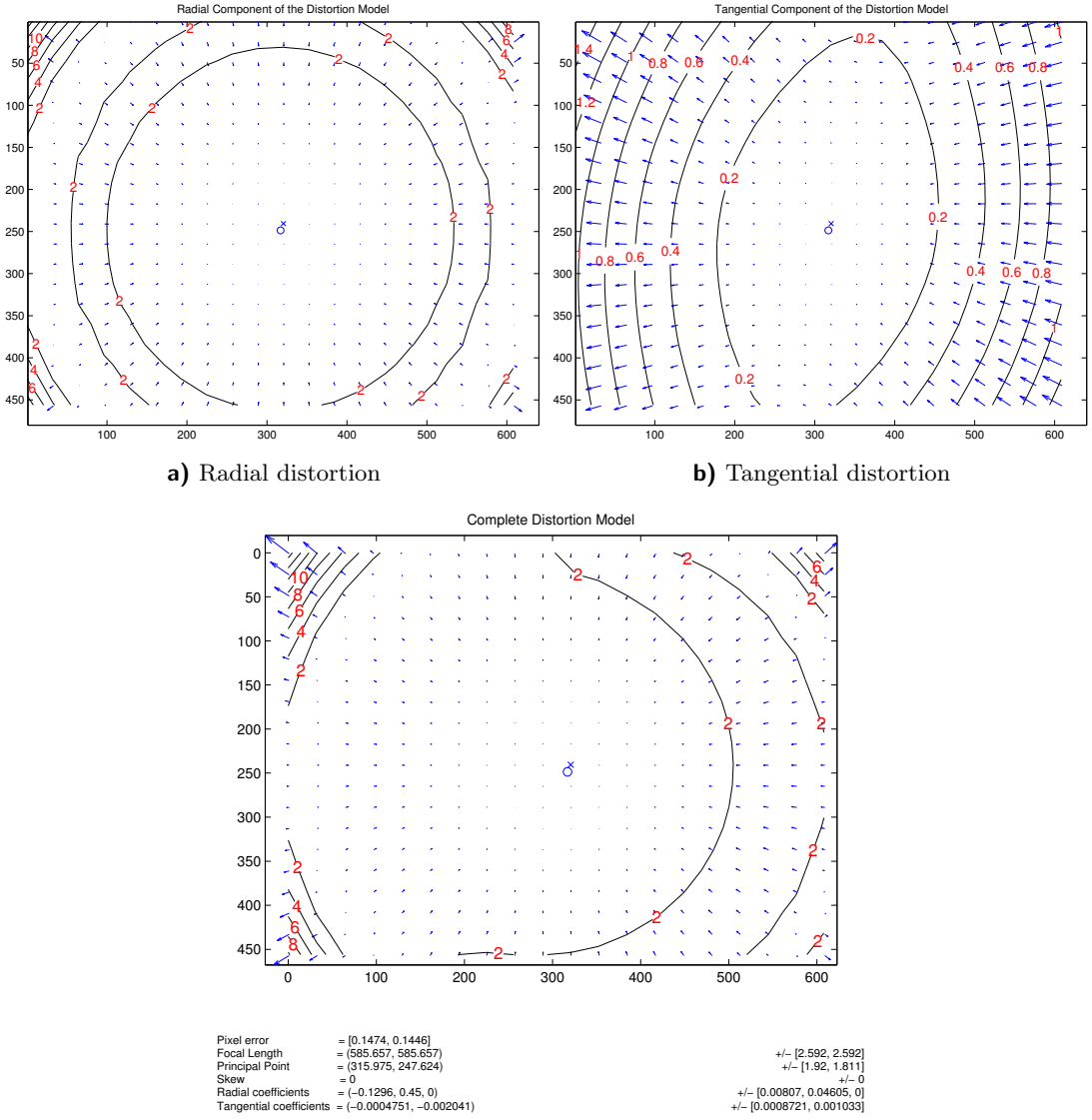
c) Illuminated by IR-laser projector - detail with enhanced contrast

d) Illuminated by halogen lamp, projector is covered - detail

Figure 4.2 Comparison of images taken with and without IR-laser projector and additional light source.

If we have a closer look at 4.3a, we can see that the displacement of 2 pixels is the maximum for most of the image, however in the image corners the radial distortion effect is much stronger (about 8 pixel). The tangential factor of the distortion (shown in 4.3b) is much smaller in comparison with the radial part of the model. The total effect of the nonlinear distortion can be seen in 4.3c.

Having such small tangential coefficient of the distortion model leads to idea to disregard this therm from the calibration optimization procedure. We tested this approach (see Fig. 4.5), but with the resulting calibration matrix the reconstruction accuracy at our experiment 4.1.3 was lower.



a) Radial distortion

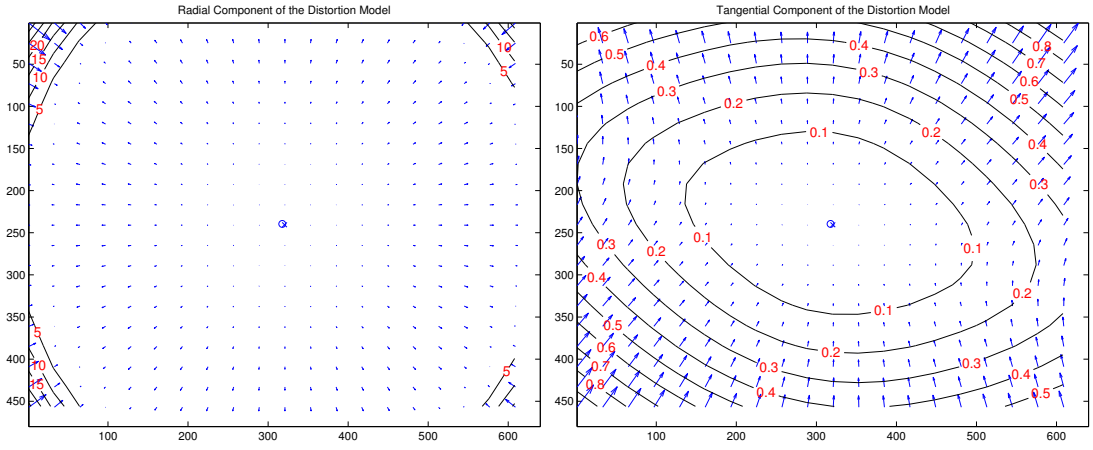
b) Tangential distortion

c) Complete distortion model

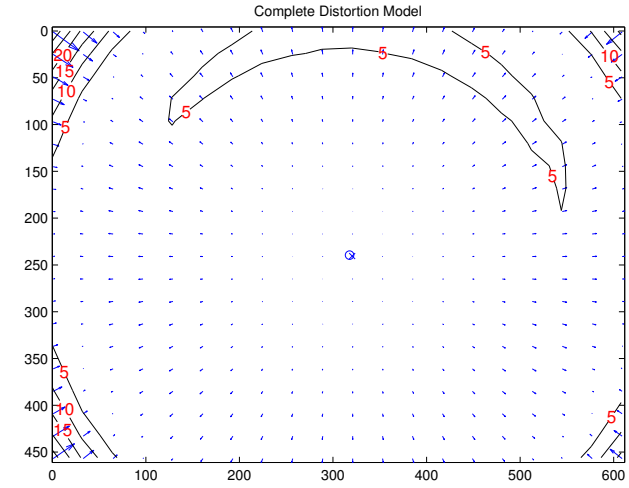
Figure 4.3 Visualization of distortion effects estimated during photogrammetric calibration of Kinect IR camera.

The distortion model of the RGB-camera (shown in Figure 4.4) is very similar with the note that the error associated with the radial component in Figure 4.4a is approx. twice bigger and therefor we can conclude that the optics of this camera is worse.

It was found, that estimating first four coefficients of the distortion model (3.3) yields



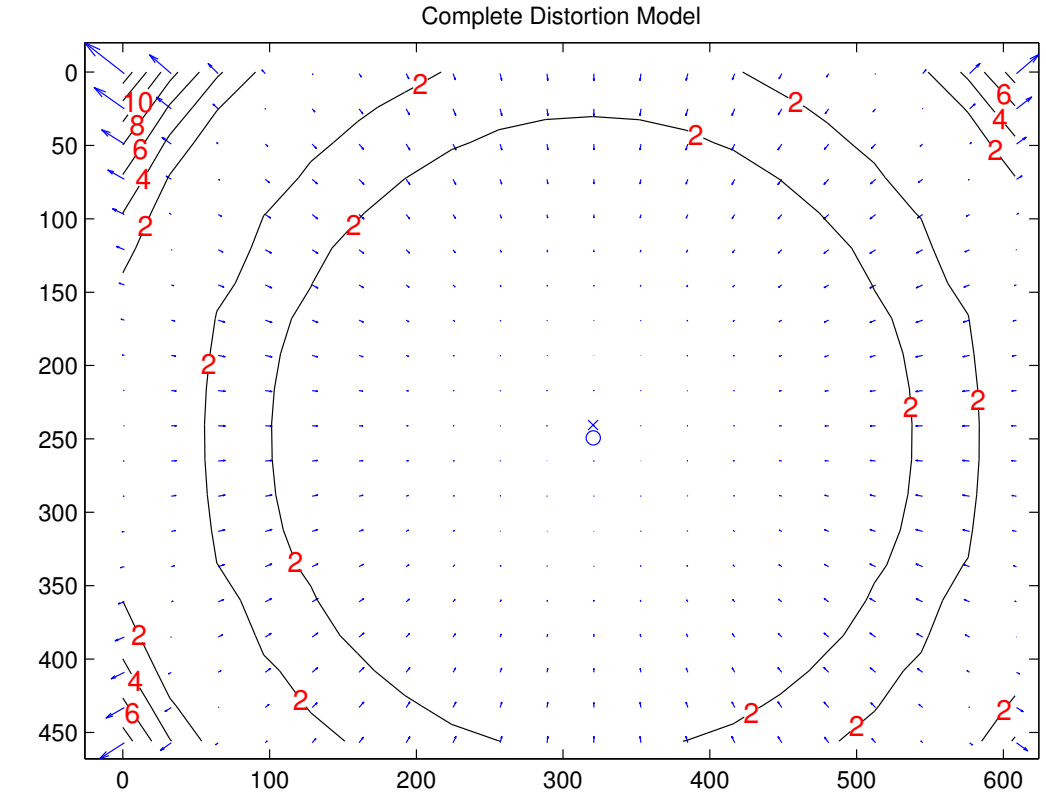
a) Radial distortion **b)** Tangential distortion



Pixel error	= [0.1683, 0.1652]	
Focal Length	= (524.002, 524.002)	+/− [2.8, 2.8]
Principal Point	= (316.698, 238.489)	+/− [1.849, 1.989]
Skew	= 0	+/− 0
Radial coefficients	= (0.2402, −0.6861, 0)	+/− [0.01135, 0.05657, 0]
Tangential coefficients	= (−0.001479, 0.0003259)	+/− [0.001734, 0.001571]

c) Complete distortion model

Figure 4.4 Visualization of distortion effects estimated during photogrammetric calibration of Kinect RGB camera.



Pixel error	= [0.1481, 0.1445]	
Focal Length	= (585.418, 585.418)	+/– [2.601, 2.601]
Principal Point	= (319.504, 248.344)	+/– [0.8395, 1.006]
Skew	= 0	+/– 0
Radial coefficients	= (–0.1262, 0.4297, 0)	+/– [0.008005, 0.04555, 0]
Tangential coefficients	= (–0, –0)	+/– [0, 0]

Figure 4.5 Complete IR-camera distortion model with the tangential part neglected.

good results. When the distortion parameters are determined we can compensate the images for these effects. In this step all the images (IR, Depth and RGB) are corrected using corresponding set of coefficients.

Extrinsic Calibration

Position and orientation of the calibration chessboard is determined as a part of the procedure. Exploiting the fact that during the procedure images of calibration chessboards were captured simultaneously by both cameras the relative rigid transformation between them (see Fig. 3.8) can be determined. Theses parameters are optimized using Calibration toolbox.

Using this knowledge we can form a common coordinate system and estimate rigid body transformations (R, T) to coordinate systems of the other camera. For simplicity we have chosen the reference frame to coincide with the coordinate system of the IR-camera[3] and so the extrinsic parameters for IR-camera are

$$R_{IR} = I_{3x3}, \quad C_{IR} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}. \tag{4.1}$$

For the RGB-camera the set of extrinsic parameters was found to be

$$R_{RGB} = \begin{bmatrix} 0.9999 & 0.0093 & 0.0039 \\ -0.0092 & 1.0000 & -0.0030 \\ -0.0039 & 0.0029 & 1.0000 \end{bmatrix}, \quad C_{RGB} = \begin{bmatrix} 24.8273 \\ -0.1076 \\ 4.1667 \end{bmatrix}. \tag{4.2}$$

We can express rigid body transformation between the reference frame of IR and RGB-cameras by the rotational part

$$R = \begin{bmatrix} 0.9999 & -0.0092 & -0.0039 \\ 0.0093 & 1.0000 & 0.0029 \\ 0.0039 & -0.0030 & 1.0000 \end{bmatrix}, \tag{4.3}$$

and the translation part (in millimeters)

$$T = \begin{bmatrix} -24.8273 \\ 0.1076 \\ -4.1667 \end{bmatrix}. \tag{4.4}$$

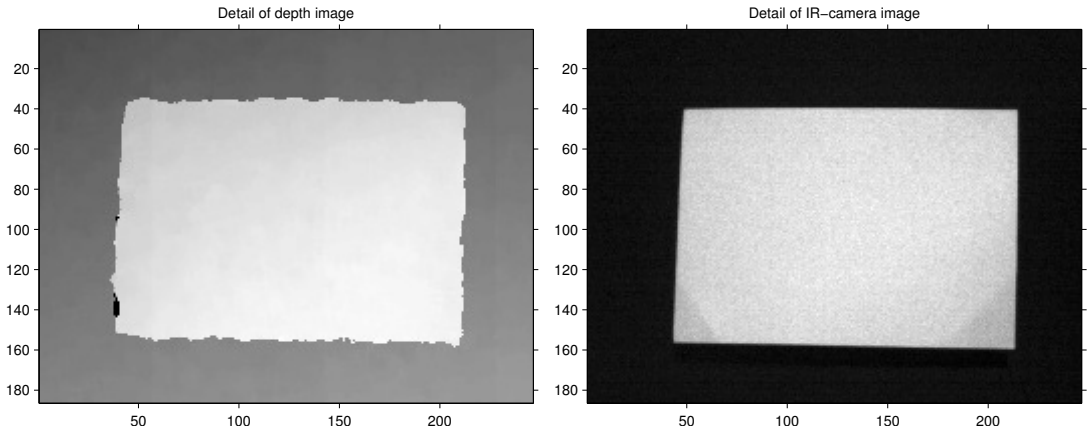
4.1.2 Raw Depth Data Processing

I order to find the distance model parameters of the camera the raw depth values need to be preprocessed. The raw depth data contain a quantized information (as was explained in Section 3.3.1) and so averaging depth value at same position over a set of many images would not make much improvement in sense of noise attenuation. Taking a median of 3-5 raw depth images was tested to be completely sufficient in removing random wrong measurements. The depth measurement was found unstable on objected edges. This does not need to be an issue for the calibration procedure since the raw depth values used for he calibration are situated far from the edges.

[3] This was advantageous due to the fact that this camera produces two sets of data - IR intensity image and depth measurements.

IR to Depth-camera Image Misalignment Correction

When IR and Depth-camera images were compared a constant pixel coordinates displacement was found. To determine the size of the shift a rectangular white paper target was attached approx. 8 cm from dark flat background and captured both by depth and IR-camera. Details of this images are shown in Figure 4.6.



a) Depth image **b)** IR-camera image

Figure 4.6 White paper target was attached approx. 8 cm from dark flat background and captured both by depth and IR-camera.

The misalignment is visualized in Figure 4.7a where the IR-camera image is shown in black and depth image was proceed using edge detector and is shown in white. In order to find the size of the misalignment, a cross-correlation between images was performed; this allowed to determine value of the displacement [29]. To ensure that just the shape of the target will be considered by the cross-correlation algorithm both images were transformed to binary (black & white) with the threshold chosen to distinguish between the target and the background [28]. The results of several experiments are shown in Table 4.3. It should be noted that the obtained result is only approximate mostly due to the unstable depth measurement on object edges. Example of several images before and after alignment process is shown in Figure 4.7.

Table 4.3 IR to Depth-camera pixel position misalignment values

Image	x_{off}	y_{off}
1	2.8	3
2	2.9	2.7
3	3	2.8
4	3.4	3.1
Average	3.025	2.9

The size of the displacement was estimated as a mean value of all experiments. Such value suggest on using correlation window of size 7 x 7 pixels in the depth calculation process . This is in contrary with our previous assumption of window with size 9 x 9 pixels (refer to section 3.3.1). Both shift size values (i.e. $x_{\text{off}} = 3$, $y_{\text{off}} = 3$ and $x_{\text{off}} = 4$, $y_{\text{off}} = 4$) were tested on our reconstruction experiment (see section 4.1.3) with the first producing lower reconstruction errors - that shows our estimated result to be more

reliable.

The Depth and IR-camera pixel coordinates are related as

$$\begin{pmatrix} x_{\mathrm{D}} \\ y_{\mathrm{D}} \end{pmatrix} = \begin{pmatrix} x_{\mathrm{IR}} - 3 \\ y_{\mathrm{IR}} - 3 \end{pmatrix}. \tag{4.5}$$

Object Intensity Effect

Since TOF cameras are known to be sensitive on object color intensity (see chapter 4.2.2) we have investigated, if Kinect also suffers from similar issues. A planar A4 paper target used in the experiment is shown in Figure 4.8a. It has six squares printed in black with different intensity levels (approx. 10%, 25%, 50%, 75% and 100% black and white).

The distribution of depth values in squares with different color intensity can be seen in Fig. 4.9. The result show that the data are quantized and symmetric around the mean value. To asses the possible effect more systematically, the target was attached to a flat wall and captured 12 times from different distances (approx. 70-140 cm) with the camera axis placed perpendicularly to the plane of the target. To each image a plane was fitted and subtracted from the distance image and only these *residuals* were used for further processing. The reason for this step is that measurements from different distances can be used and also this procedure compensates the error of not perfectly perpendicular placement of camera w.r.t. to target. Using the IR-camera images positions of the squares were identified for each measurement - example can be seen in Figure 4.8b.

During the experiment only squares with intensities 50%, 75% and 100% were considered because the others with lower intensities were hard to identify on most images. Data from squares on corresponding positions from image of depth residuals were extracted and visualized in Figure 4.10 in form of a normalized histogram. For all intensity levels the data are approximatively normally distributed with almost zero mean (see Table 4.4) and thus without implication on any trend or dependency between contrast and measured depth. It can be seen in image 4.8b that the areas with darker color

Table 4.4 Mean values of of residuals of plane fitting in areas with different contrast

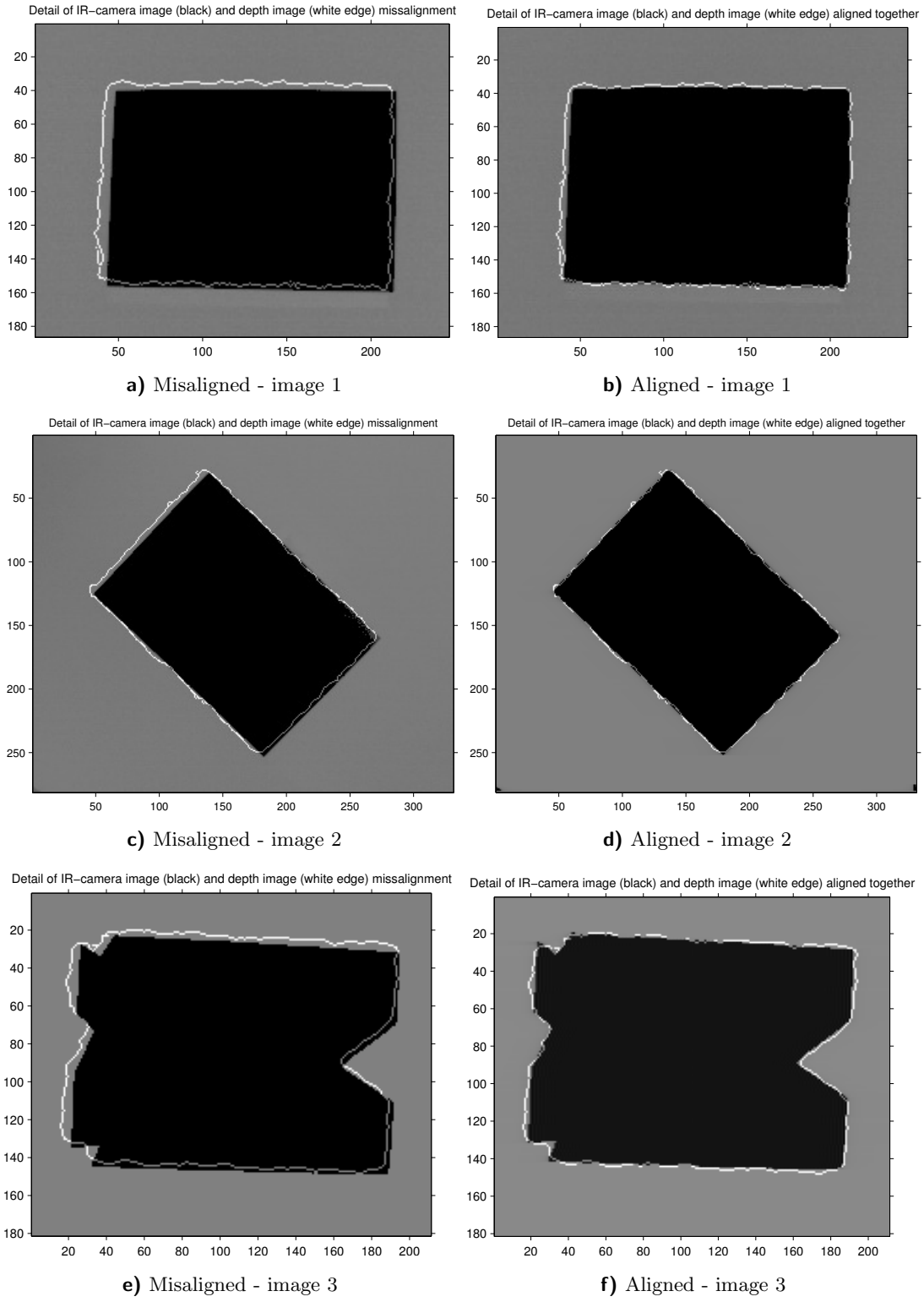
Contrast	μ [mm]
Black 100%	-0.0707
Black 75%	-0.0858
Black 50%	-0.0893
White	0.0180

have worse reflection of the IR dots. The ability of Kinect to calculate right disparities on dark (high intensity) areas is probably lower which can correspond to smaller peak with zero mean in histogram 4.10.

4.1.3 Distance Model Calibration

Raw Depth Extraction

In order to continue with the calibration, depth values at corresponding pixel positions as were the corners selected during the photogrammetric calibration (in Section 4.1.1)



a) Misaligned - image 1

b) Aligned - image 1

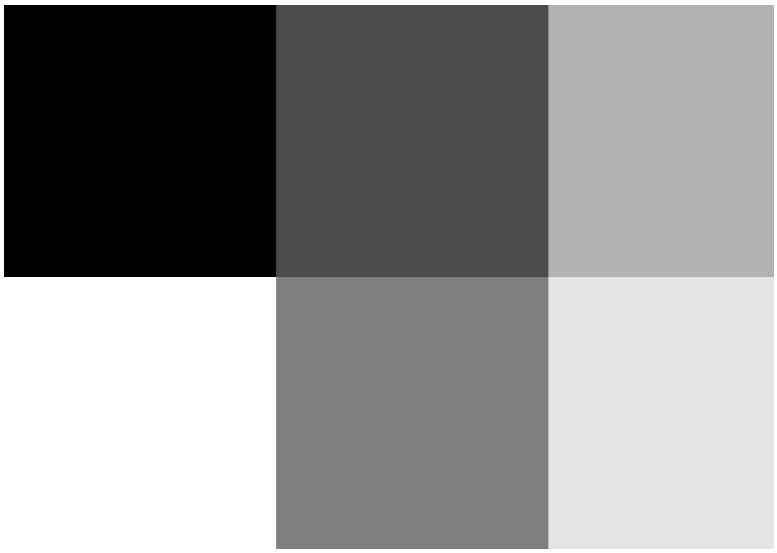
c) Misaligned - image 2

d) Aligned - image 2

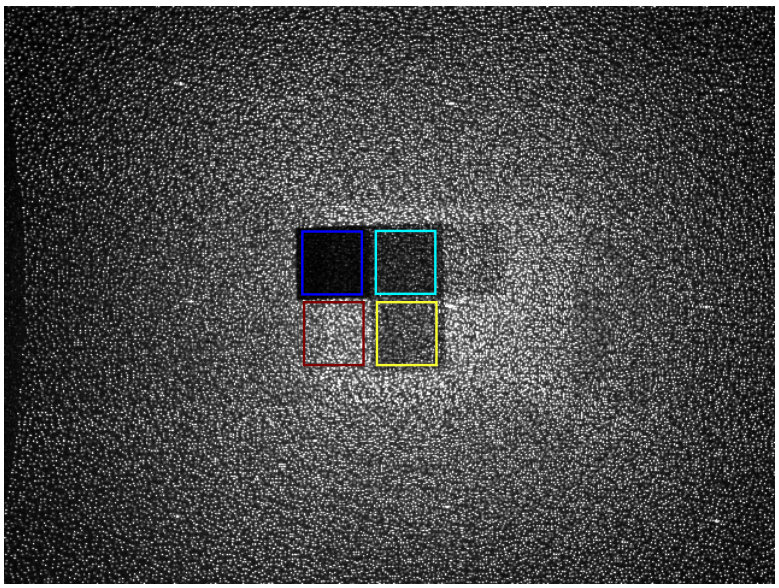
e) Misaligned - image 3

f) Aligned - image 3

Figure 4.7 Illustration of IR to Depth-camera pixel position misalignment and its correction. The IR image is shown in black and the depth image is represented by its white edge.



a) Original pattern



b) IR-camera image with evaluated areas

Figure 4.8 Planar paper target with six black squares with different intensity levels used for investigating of effects of object contrast on the measurement.

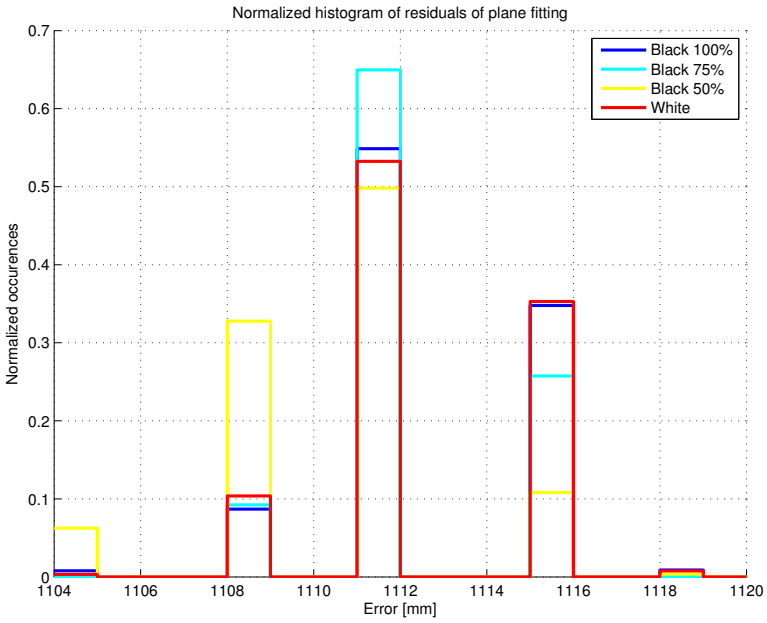


Figure 4.9 Histogram (normalized) of distances at target with different color intensity.

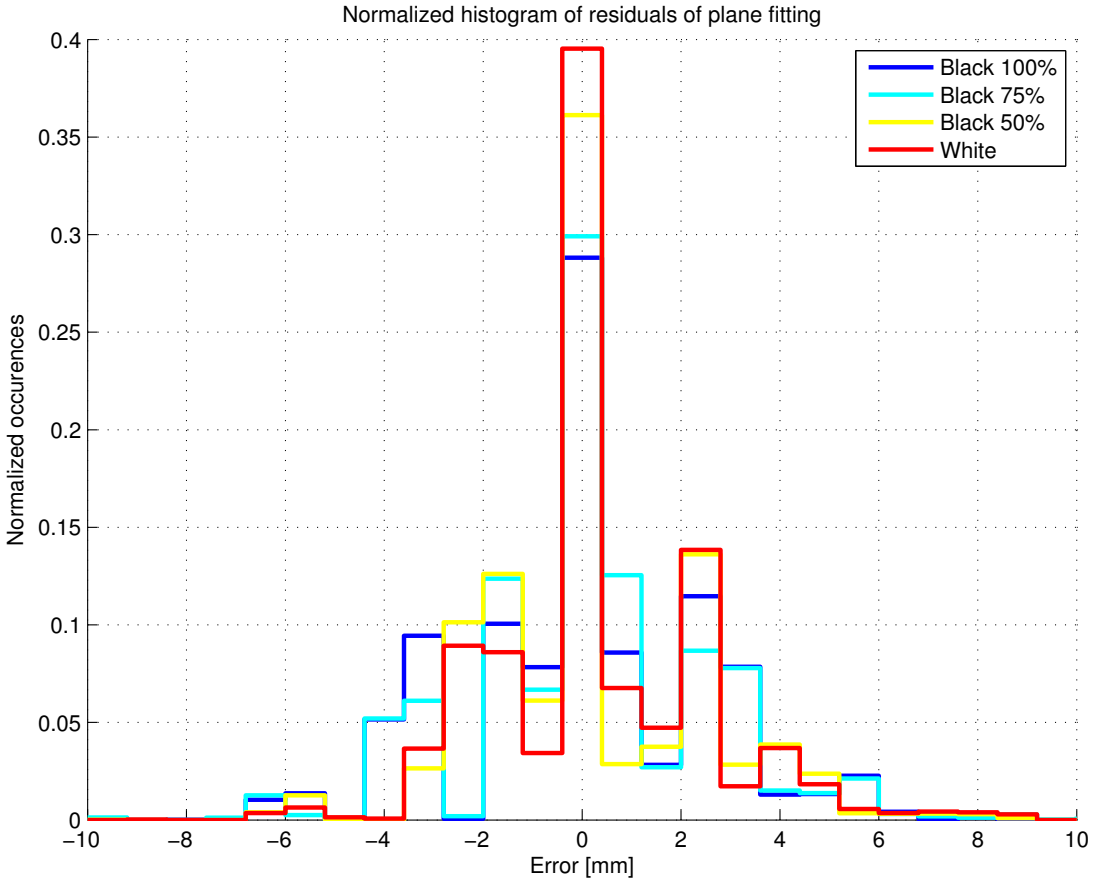
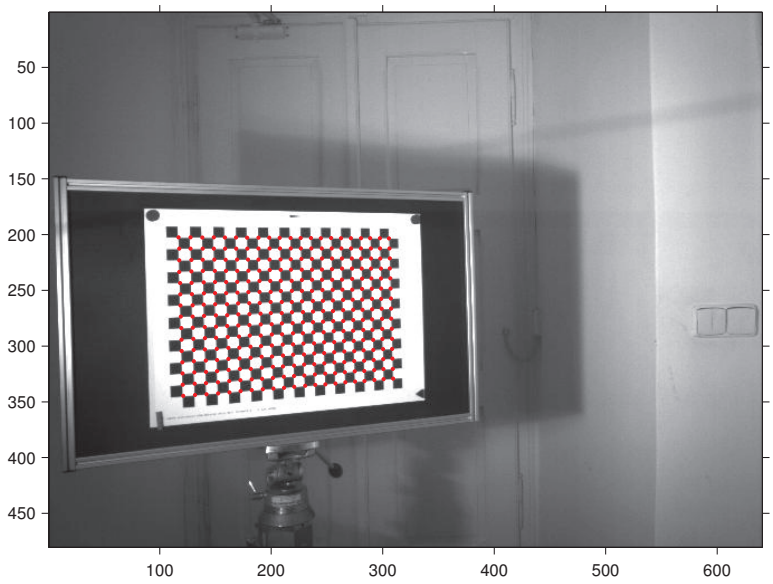
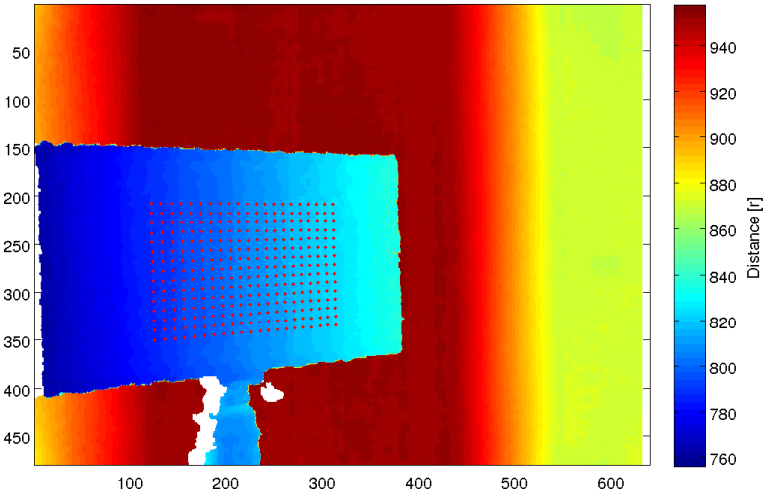


Figure 4.10 Histogram (normalized) of residuals of plane fitting in areas with different contrast.



a) IR-camera intensity image, detected corners



b) Depth image, corners where depth is extracted

Figure 4.11 Extraction of depth values at same pixel positions as were the corners (marked as red dots) selected during the photogrammetric calibration.

need to be extracted. Prior to the extraction the pixel coordinate need to be shifted according Equation 4.5. An example is shown in Figure 4.11. In case that some values were not calculated by Kinect, due to the local reflection, they can be interpolated from neighboring values (since they all lie on a plane) - but these values should be then excluded from the distance model calibration.

Depth Model Fitting

The distance calibration[4] can be accurately done after all preceding corrections were applied. For all control points on calibration chessboard a 3D position is determined with high level of confidence by photogrammetric calibration.

For each such point $X(x,\ y,\ z)$ from reconstructed calibration grid we also have the corresponding raw r distance value from depth image. Using the depth model (given by Equation 3.6) we can relate the third distance coordinate z of the point X (further as X_z)to the perpendicular distance z' that is represented by

$$z' = \frac{fb}{c_1 r + c_0}.$$

For all available measurement z should be equal to z' and so we can rearrange the equation to

$$c_1 r + c_0 = \frac{fb}{X_z},$$

where unknown coefficients of the model $c_1,\ c_0$ can be determined using least-square fit.

In our calibration experiment the distance model was calibrated using *even* measurements from total of 14 different poses of calibration chess board (with 315 corner positions detected on each). This yields a total number of $2205 = 7 \cdot 315$ points with corresponding distance measurements that were used in fitting procedure.

Table 4.5 Distance model parameters that were found using least-square fit

Parameters found by the fit		Parameters of the model	
c_1	c_0	f [mm]	b [mm]
-0.0013	1.4389	6.0908	75

Evaluation of Reconstruction Performance

The reconstruction performance was evaluated in terms of geometrical distance between points reconstructed during photogrammetric calibration (considered as ground truth) and points reconstruct from Kinect depth measurement. The error distribution for all calibration chessboards is shown in Figure 4.12. The local mean is marked by red line. Note that different colors are used to distinguish points that belong to different images used during the calibration. Numerically the performance is evaluated in Table 4.6.

[4]Distance calibration in sense of finding unknown parameters $c_1,\ c_0$ of distance measurement model as described in equation 3.6.

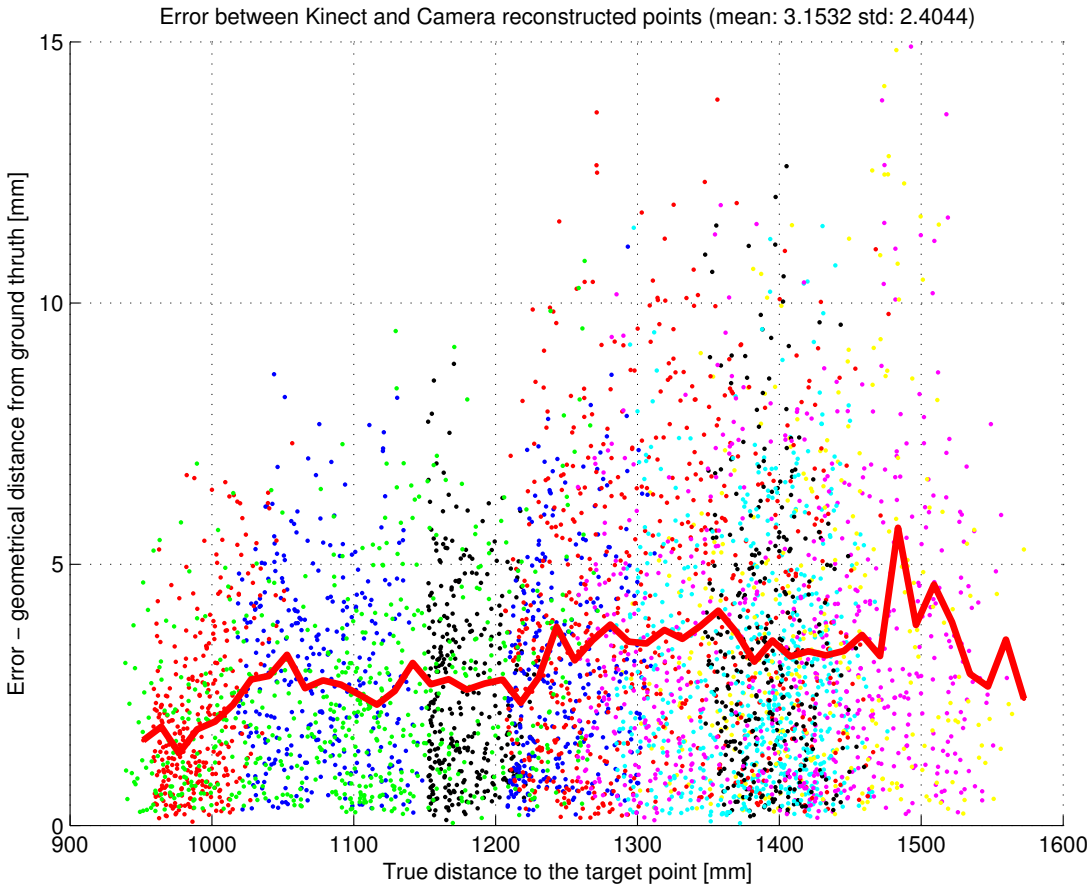


Figure 4.12 Reconstruction error of calibrated Kinect device. The solid red line marks the local mean.

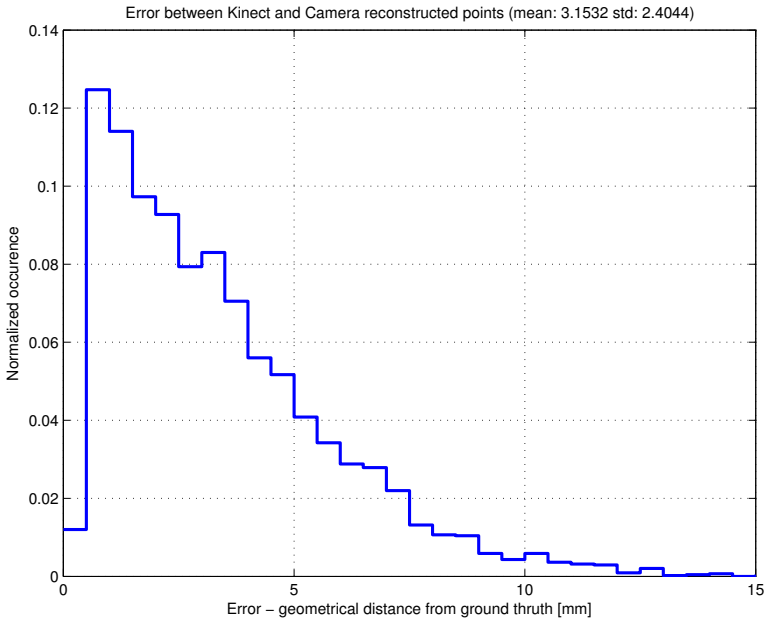


Figure 4.13 Normalized histogram of reconstruction errors of calibrated Kinect device.

Table 4.6 Reconstruction performance of calibrated Kinect device. The precision and accuracy are evaluated in terms of geometrical distance between points reconstructed during photogrammetric calibration (considered as ground truth) and points reconstruct from Kinect depth measurement. Measurement numbers printed in bold denote the images used for depth model calibration. Each chessboard represent 315 control points.

Cal. chess.	Geometrical error [mm]		
num.	μ	σ	max
1	2.9691	1.7897	8.6367
2	2.1822	1.6120	7.3188
3	1.9903	1.4242	7.3007
4	2.6243	1.7817	8.8375
5	2.9192	2.1073	11.4719
6	3.6014	2.6925	14.9086
7	3.8469	2.9848	14.8437
8	4.7584	2.9656	13.8947
9	2.9384	1.9913	11.0805
10	3.6551	2.6856	13.6486
11	2.9166	2.1006	10.8053
12	3.5056	2.6491	12.6191
13	2.8826	2.3462	11.4418
14	3.3545	2.4754	11.8759
Total	3.1532	2.4044	14.9086

Comparison of Different Distance Models of Kinect Camera

Several procedures for computing the measured distance from Kinect raw values were published. We have compared them on our calibration data. The results of the comparison are shown in Table 4.7 and in Figure 4.14. Our method was trained on one half of the testing data (see section 4.1.3). The method from Nicolas Burrus [23] is a division model (needed to be corrected for a constant shift of 55 mm to match our coordinate system). Stéphane Magnenat proposed a model [43] with tangent function. The ROS model is similar to our but with only one degree of freedom [1]. The OpenNi model is a division model used in new Kinect drivers [46]. It should be noted that if constants of the models were tuned directly for our device they could possibly exhibit better performance.

Table 4.7 Comparison of different available distance models. The methods are evaluated on a set of 4410 points.

Method	Geometrical error [mm]		
source	μ	σ	max
Our	3.1532	2.4044	14.9086
ROS	3.4578	2.6490	17.2457
Nicolas Burrus	4.2321	3.0504	17.9869
Stéphane Magnenat	4.3220	2.9477	18.3392
OpenNi	4.9816	3.1456	18.0974

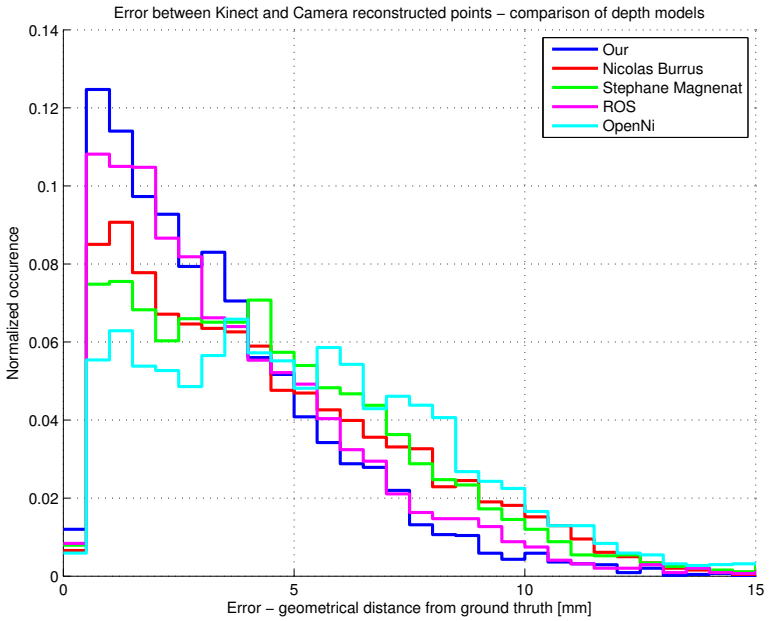
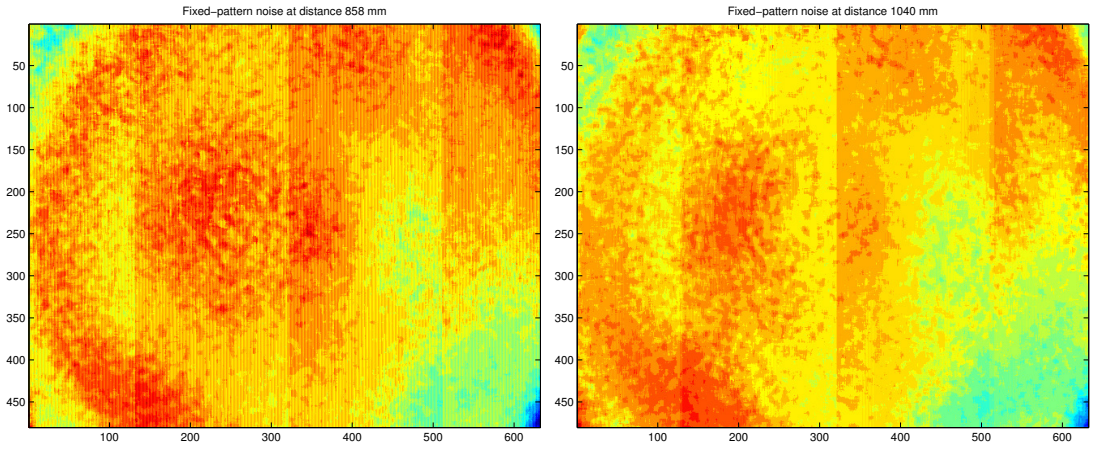


Figure 4.14 Comparison of different available distance models.

Fixed-pattern Noise Correction

If we observe Kinect depth images of a flat target covering the whole field of view a fixed pattern emerges.

To better visualize it a wall was captured from 18 different distances (0.7 - 1.3 m) and a each resulting depth map a plane was fitted. The residuals of plane fitting are shown in Figure 4.15.



a) Distance 86 cm **b)** Distance 104 cm

Figure 4.15 Residuals of plane fitting showing the fixed-pattern noise on depth images from different distances.

We can see that visually the pattern is relatively similar. For better visualization we show in Figure 4.16 residual values for even measurements on 250 horizontal scan-line (middle of the image). The plot shows that there exist a trend starting in negative values in left part of the image, then rising to positive values and ending again in

negative values. The local mean is shown as a solid red line. The higher distance (last two) show the tendency to differ from the mean more than the other measurements. This effect is known also from conventional CCD cameras [47] - where it is caused by manufacturing imperfections. Since in the case of Kinect camera the patterns 4.15 corresponds to the patterns on the IR-camera image (see 3.9a) the probable cause is the imperfection of the projected IR-laser pattern.

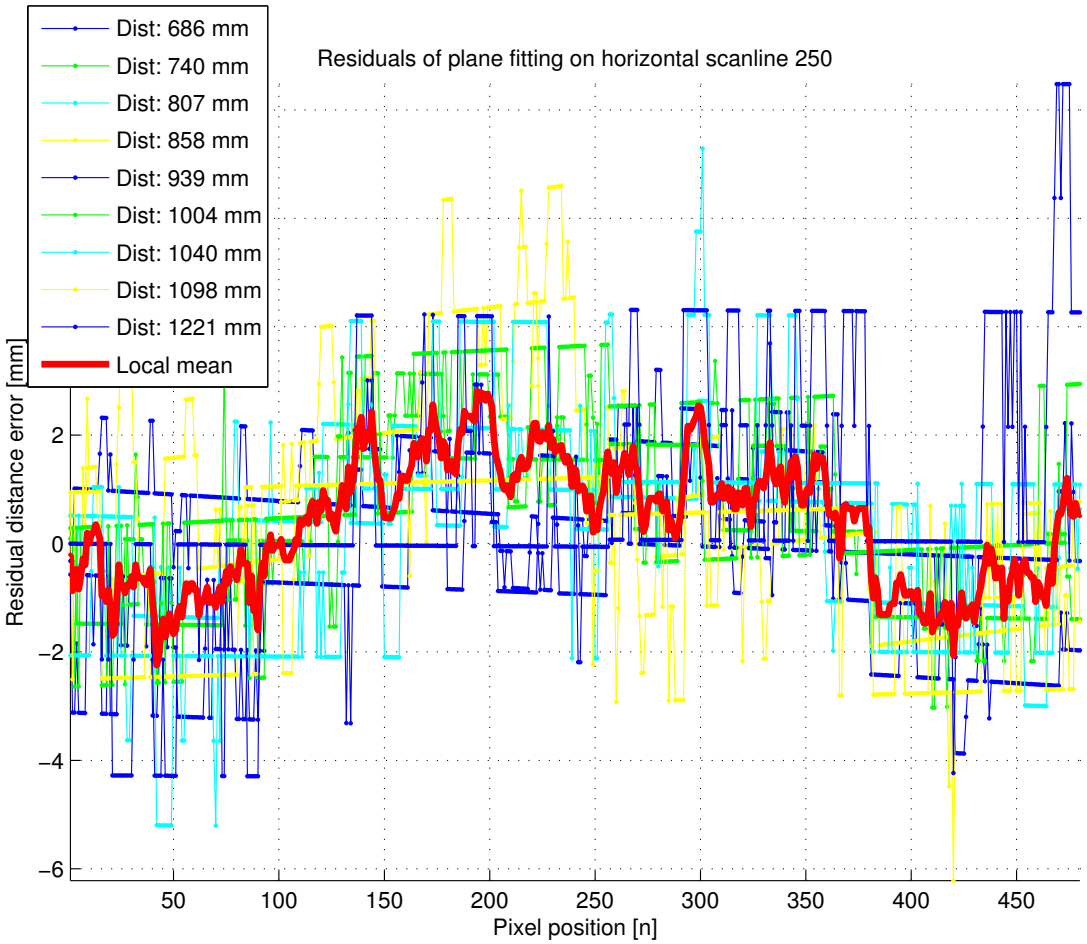


Figure 4.16 Residuals of plane fitting on 250 horizontal scan-line (middle of the image). The local mean is shown as a solid red line.

In order to compensate for such effect several techniques were discussed in the literature e.g. in [13]. As a prove of concept and because the distance range was relatively small and the residuals are close to the local mean we can form a *correction table* as a pixel-wise mean of the residual images. Such correction table can be then simply added to newly captured data in order to mitigated the effect of the fixed-pattern noise.

In order to evaluate the method the correction table was calculated only from residuals of even images. The correction table was then applied (added to) both odd and even depth image and the standard deviation of the result was compared in the Table 4.8. Because the target was a flat plane we would assume the the depth values should all be the same i.e. to have small deviation from the mean. After applied the correction both the training and the testing depth images showed lower standard deviation.

To further evaluate the effect of this correction we used the experiment described in

Table 4.8 Evaluation of Fixed-pattern noise correction. The standard deviation of depths of a flat target is lower after application of the correction

	Standard deviation [mm]	
Dataset	Original σ	Corrected σ
Training	2.1846	1.5438
Testing	1.9791	1.3403

chapter 4.1.3 and applied the correction table to the distance measurements. As can be seen in Table 4.9 and in Figure 4.17 the the mean and standard deviation of the error improved by approx. 0.25 mm. The maximal deviation increased at several points for approx. 0.5 mm (3 mm at one point). The comparison was done on 4410 points located in different parts of the field of view.

According to the results the fixed-pattern noise can be corrected to achieve better performance. It is probable that a more sophisticated correction method will produce better results.

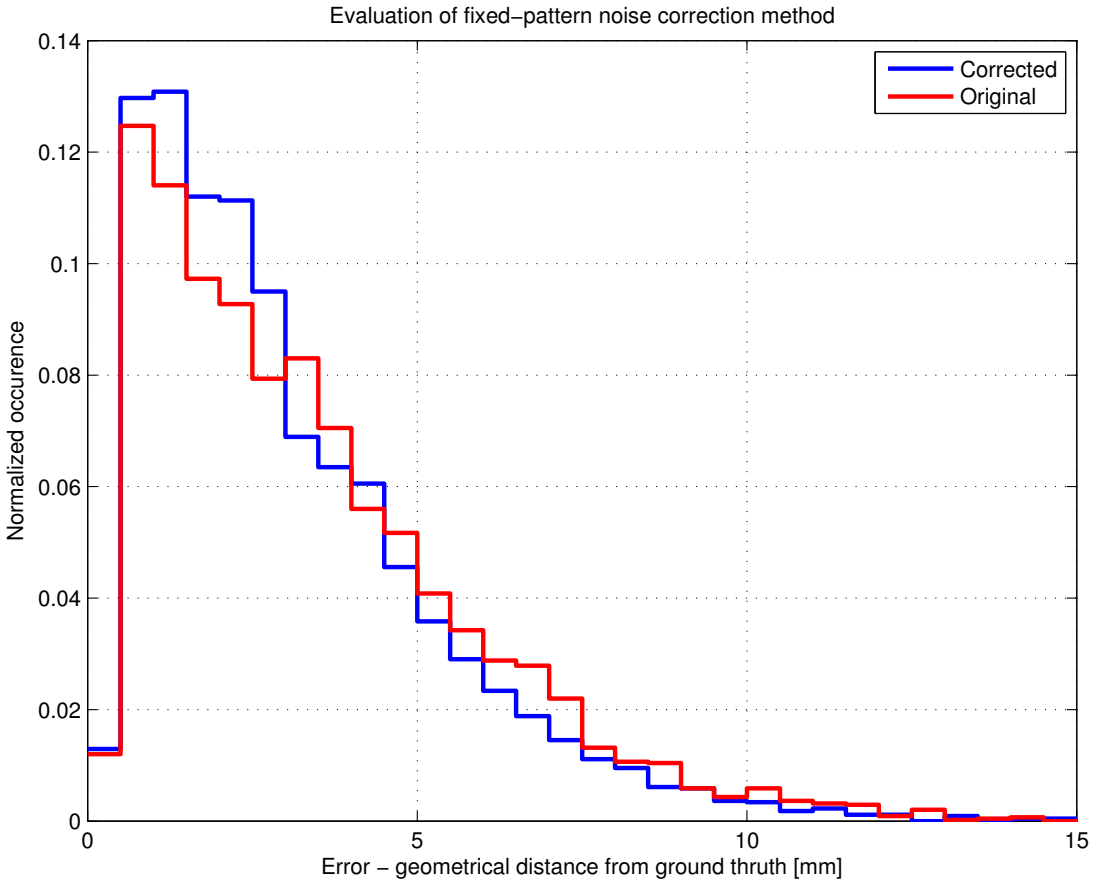


Figure 4.17 Evaluation of the effects of fixed-pattern noise correction. The plot shows normalized histogram of reconstruction errors of calibrated Kinect device with and without fixed-pattern noise correction.

4.1.4 Complete Calibration Procedure

Tasks, hat we described so far can be now combined together to form a complete calibration procedure. Schematically the necessary steps are shown in Figure 4.18. The outputs are the intrinsic parameters and the distance model. Both need to be determined only once on each device.

Table 4.9 Reconstruction performance of calibrated Kinect device with fixed-pattern noise correction. The compare the effect refer to Table 4.6. Each chessboard represent 315 control points.

Cal. chess.	Geometrical error [mm]		
num.	μ	σ	max
1	2.1619	1.4760	7.8440
2	2.0003	1.3540	8.0831
3	2.1831	1.5858	7.9050
4	2.4644	1.6784	8.0095
5	2.5760	1.8667	9.4688
6	3.6092	2.8506	18.2025
7	3.5812	2.8558	14.6706
8	3.9356	2.7118	11.7871
9	2.5706	1.8295	11.1875
10	2.7195	1.9964	9.8377
11	2.7926	2.0264	10.7308
12	3.4708	2.5656	12.1418
13	2.9984	2.1776	13.8976
14	2.5969	1.9401	9.0422
Total	2.8329	2.1962	18.2025

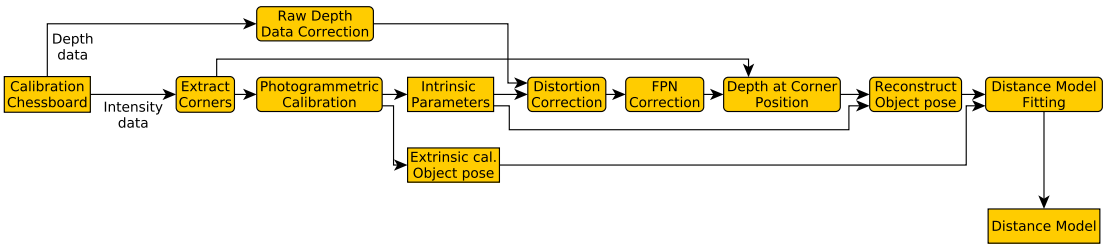


Figure 4.18 Overview of the calibration procedure.

4.1.5 Forming a Metric Point Cloud

For a point $x(u,\ v)$ on the image plane we can project a ray $\vec{x}$ from the camera center going through the point x using inversion of camera intrinsic matrix K^{-1} as

$$\vec{x} = K^{-1} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix}. \tag{4.6}$$

The scale factor z provided by the Depth sensing camera allow us to reconstruct the position of 3D point X in camera reference frame. This situation is illustrated in Figure 4.20.

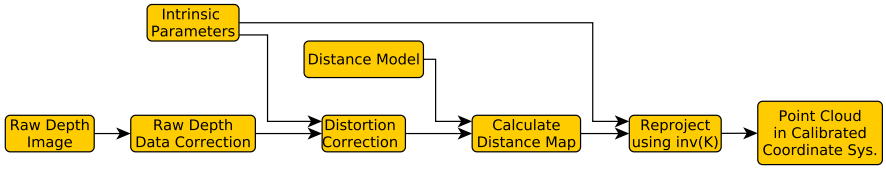


Figure 4.19 Overview of how the point cloud in calibrated reference frame is produced.

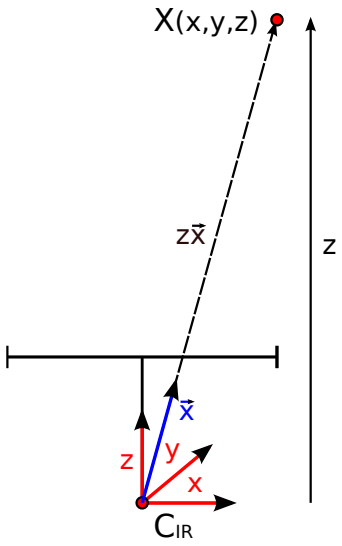


Figure 4.20 Reconstruction of point 3D position on projected ray using know point distance.

The overview of the process can be seen in Fig. 4.19.

4.1.6 Coloring the Point Cloud

Having acquired RGB and Depth-camera images simultaneously a $2\frac{1}{2}$D reconstruction in form of a colored point cloud can be made.

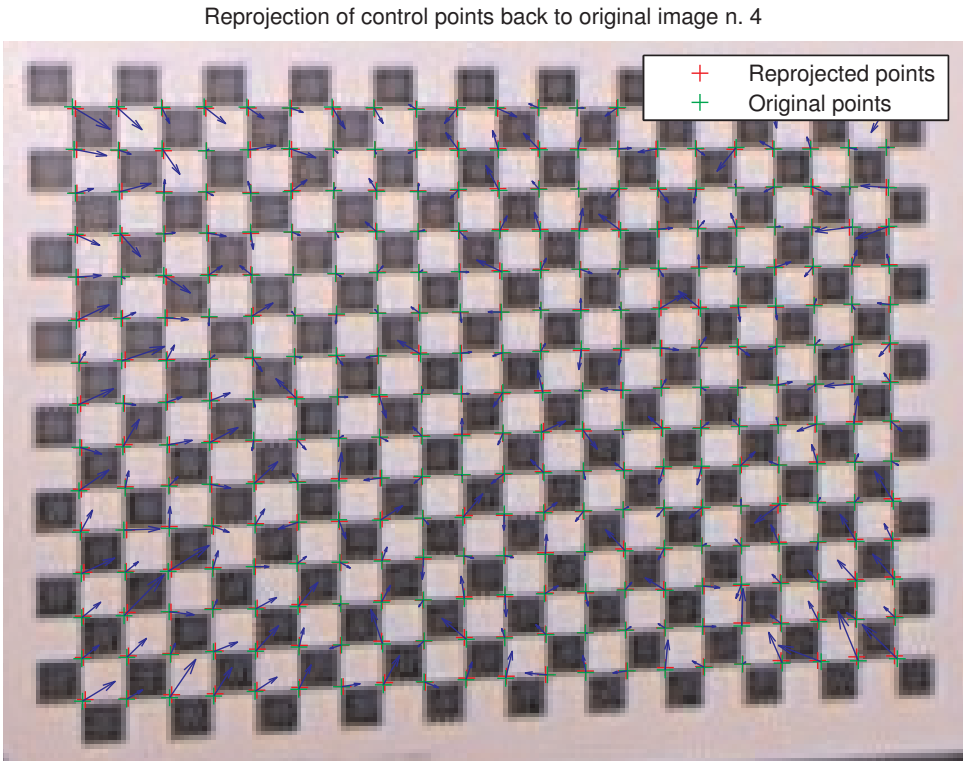
First we evaluate wherever the estimated transformation between both cameras is precise enough. We start with reconstructing the grid control points from the depth image. The resulting point cloud is transformed using the estimated transformation to the RGB-camera reference frame. Then we can project the points to the RGB camera image plane using its calibration matrix (result on one calibration chessboard is shown in Fig. 4.21). To asses the precision of the transfer the reader should have a look to Fig. 4.22. The mean absolute pixel error is 0.25, the x-y mean is almost zero. We can conclude that the transformation was determined with sufficient precision.

For each pixel where the depth information was retried a point in Depth-camera reference frame can be reconstructed. This cloud of points is then transformed to coordinate system of the RGB-camera and projected back onto the image plane. The procedure is more formally described in Alg. 4.2. For further details the reader should refer to [41].

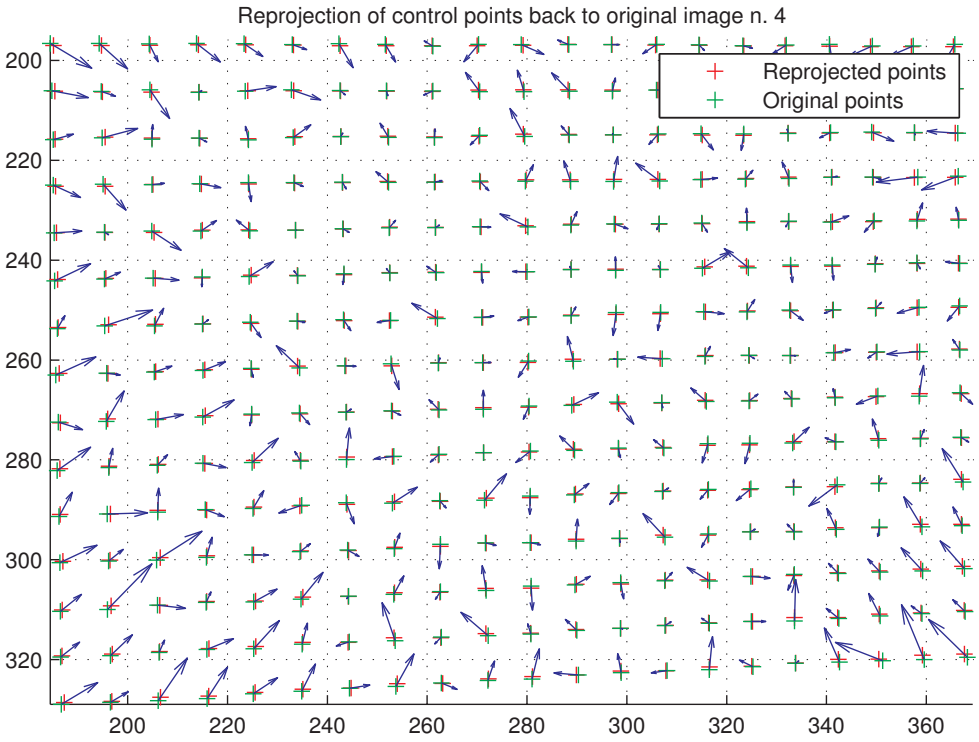
Capability of reconstructing complex 3D surface was tested using the procedure described in Algorithm 4.1.6. As an example a paper box shown in Figure 4.24a was captured. The result of the scene reconstruction can be seen in Figure 4.24b - note that the text *Granko* is visually well reconstructed, even though the whole model consists only of colored points and no a priori knowledge that the points lie on a plane was used.

Hidden Surface Removal

Depth and RGB-camera are physically located at different positions and so their view directions are different. This may lead to occlusion when some parts of the scene captured by the Depth camera can not be seen in image of RGB-camera (see Figure 4.23). For these regions a wrong color information can be assigned. In order to prevent such mis-mapping a method called Z-buffering described in [41, 57] was utilized.

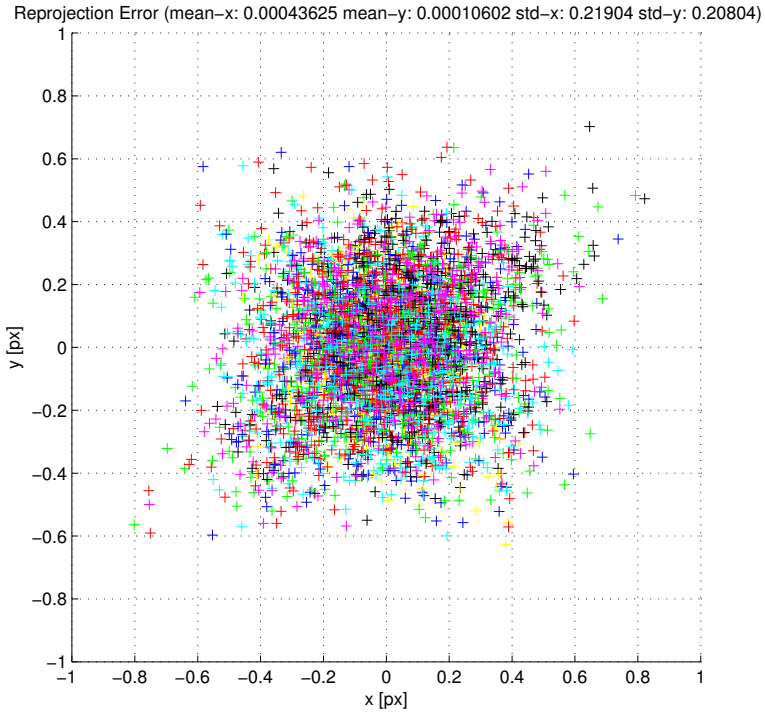


a) Re-projection on image

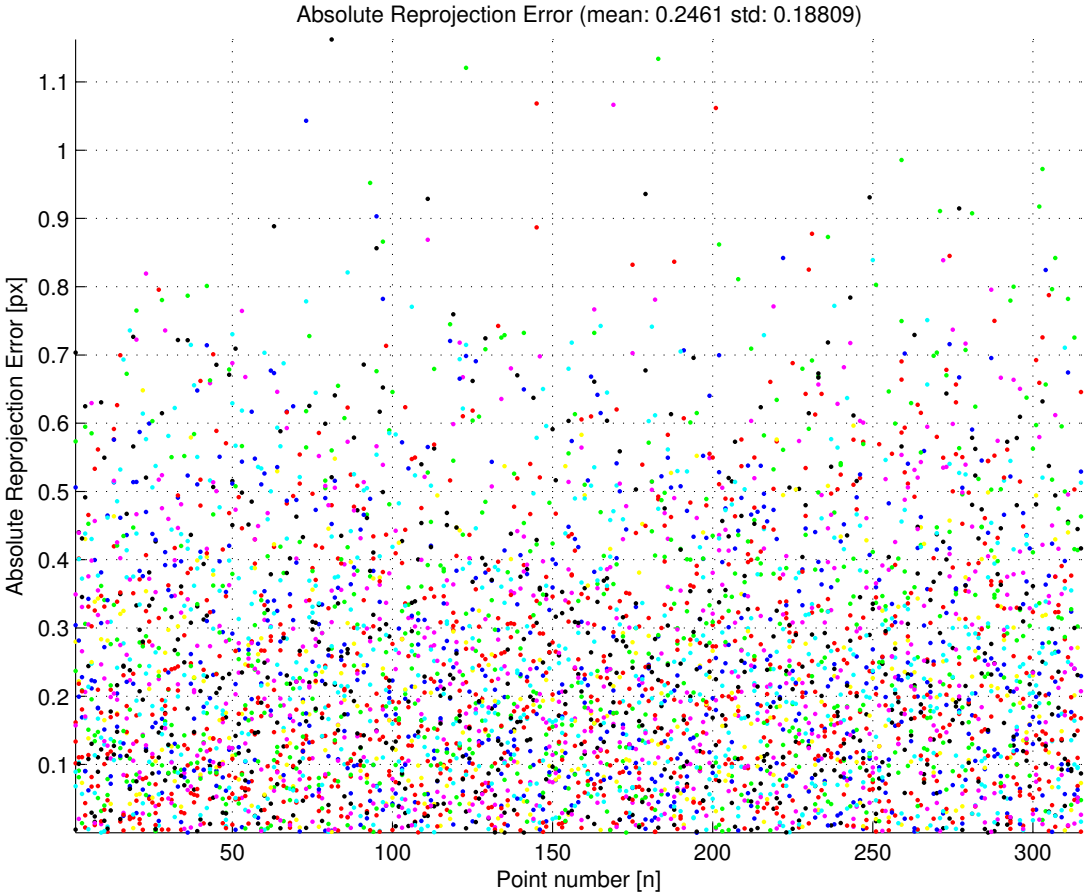


b) Re-projection without underlying image - shown for better readability

Figure 4.21 Re-projection of point cloud reconstructed from Depth-sensing camera after it was transformed to RGB-camera reference frame (example of one image). The blue arrows mark the direction of the error.



a) Re-projection error in x an y-direction



b) Re-projection error - absolute distance

Figure 4.22 Re-projection error of point clouds reconstructed from Depth-sensing camera after it was transformed to RGB-camera reference frame. The errors were calculated on 4410 points.

Procedure 4.2 Scene Reconstruction - Coloring the Point Cloud

1: Re-project depth image to form point cloud in Depth camera reference system X_D^D.

2: Using known rigid transformation transfer the point cloud to RGB-camera reference system $X_D^{RGB} = RX_D^D + T$.
3: Project X_D^{RGB} to RGB-camera image plane $x = P_{RGB}X_D^{RGB}$ using RGB-camera projection matrix P_{RGB}.
4: For each pixel determine the geometrical distance of all points projected on it. Only the closet point is kept for further processing, other are neglected to avoid incorrect color assignment from hidden surfaces (see paragraph 4.1.6).
5: Resulting points can be then matched with corresponding pixel colors and form a *colored 3D point cloud*.

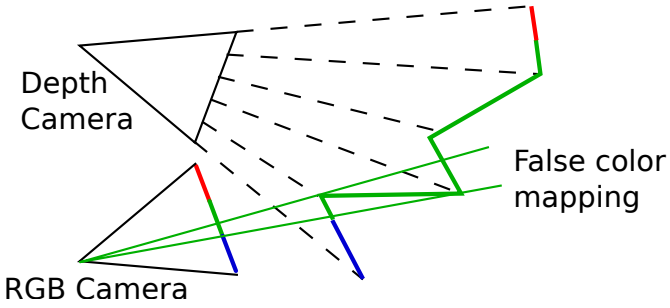


Figure 4.23 False color assignment in shadowed region w.r.t. the RGB sensor. Figure is taken from [41].

The main idea is to determine for each pixel the geometrical distance of all points (from Depth camera) projected onto it and only to closets points will have assigned color and kept for further processing. For the procedure a variable $z_{\text{buff}}(u,\ v)$ which stores the minimal per-pixel z-distance is used. All 3D points are evaluated in such way that their actual 3-space distance $z(u,\ v)$ is checked against $z_{\text{buff}}(u,\ v)$ at corresponding pixel position. If the distance is bigger than the one stored in

$$z_{\text{buff}}(u,\ v) > z(u,\ v),$$

the point is neglected from further processing. However if

$$z_{\text{buff}}(u,\ v) < z(u,\ v),$$

then the new value for

$$z_{\text{buff}}(u,\ v) = z(u,\ v)$$

is assigned and the pixel is matched with corresponding color.

Exporting the Point Cloud

To allow to work further with the reconstructed point cloud we have implemented a function that exports the data in *.ply* file format. The structure of the file (described in details in [21]) is relatively straightforward: a header is followed by list of points with coordinates and corespondent colors in RGB. The *.ply* files can be for example used in powerful 3D modeling software tools Meshlab and Blender [6, 3]. The Figure 4.24b was captured from point cloud visualized in Meshlab software.

a) Image taken by Kinect RGB-camera **b)** Colored point cloud

Figure 4.24 Example of scene reconstruction in form of a colored point cloud.

4.2 Calibration of SR-4000 TOF Camera

4.2.1 Photogrammetric Calibration

For the experiment a chessboard with 28 mm squares providing 88 corners was used. The integration time was selected to 30 ms during all experiments.

Intrinsic Parameters and Distortion Model Coefficients

The amplitude images provided by the SR-4k camera are highly distorted and in very low resolution (see Image 4.25a). Therefore the corner detection procedure must be supervised to ensure the proper corner extraction. Due to the low spatial resolution of the camera sensor a calibration target with bigger squares had to be used.

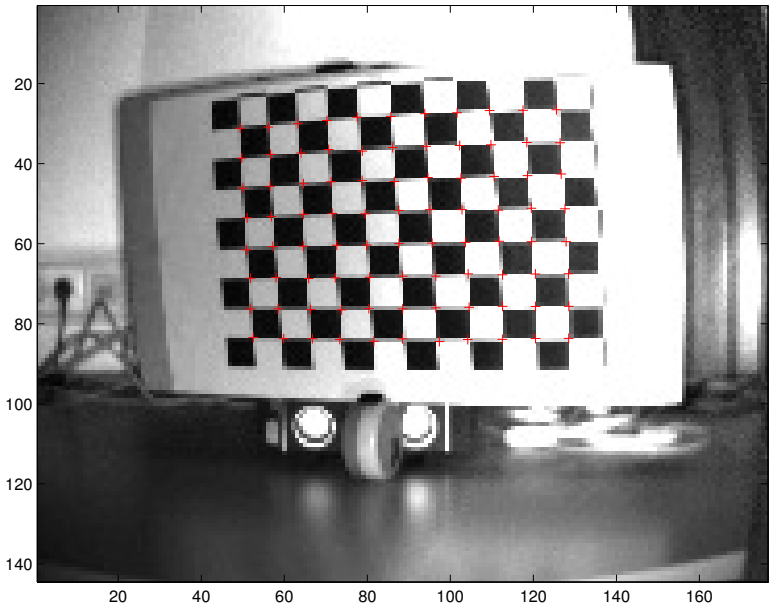
Parameters of intrinsic calibration matrix of the camera were found using the Calibration toolbox to be the following - see Tab. 4.10.

Table 4.10 Intrinsic parameters of Swissranger SR-4000 camera

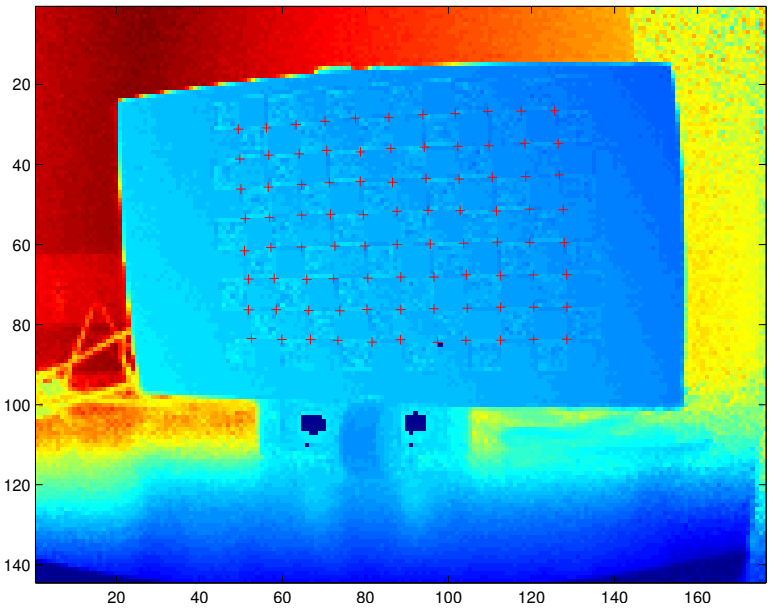
Focal length		Principal point		Distortion coefficients				
f [px]	f [mm]	x_0 [px]	y_0 [px]	k_{c_1}	k_{c_2}	k_{c_3}	k_{c_4}	k_{c_5}
257.6	10.3	91.7	54.6	-0.8148	0.03856	0.05969	-0.02681	N.A.

According to the results the camera image suffers from high radial distortion, which is vizuallized in Figure 4.26[5]. If we have a close look at 4.26a that the displacement of 2 pixels is almost in the center of the camera view. In the image corners the radial distortion effect is very strong (about 15 pixel). The tangential factor of the distortion (shown in 4.26b) is much smaller in comparison with the radial part of the model, but still not negligible.

[5] Arrows represent the direction of the pixel displacement induced by the lens distortion. The red numbers denote the actual size of the error. The cross indicates the image center, and the circle the location of the principal point.

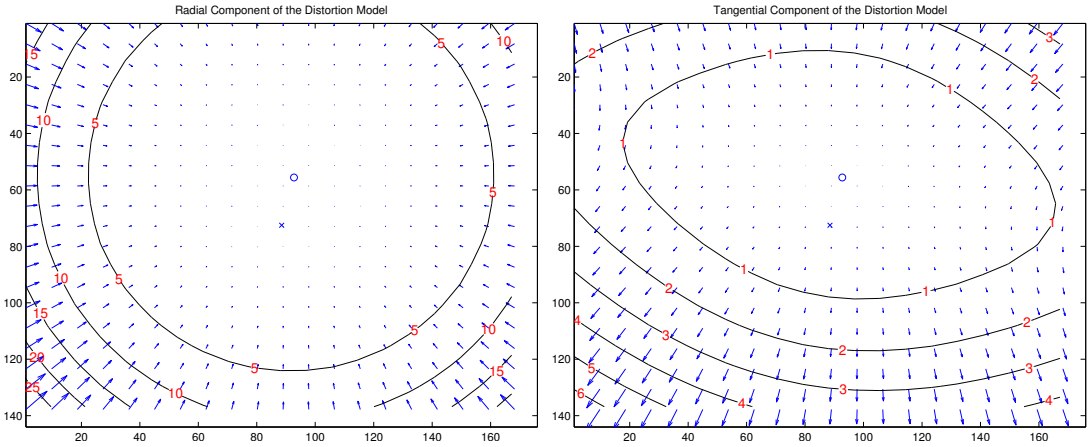


a) Intensity image, detected corners

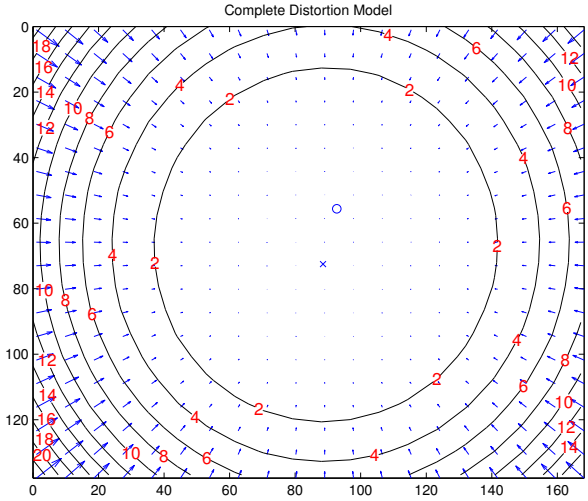


b) Depth image, corners where depth is extracted

Figure 4.25 Extraction of depth values at same pixel positions as were the corners (marked as red points) selected during the photogrammetric calibration.



a) Radial distortion **b)** Tangential distortion



Pixel error	= [0.2695, 0.2678]	
Focal Length	= (254.071, 254.071)	+/– [8.9, 8.9]
Principal Point	= (91.7155, 54.6069)	+/– [6.508, 6.087]
Skew	= 0	+/– 0
Radial coefficients	= (–0.9135, –0.5484, 0)	+/– [0.2032, 2.741, 0]
Tangential coefficients	= (0.04389, –0.008816)	+/– [0.01628, 0.01655]

c) Complete distortion model

Figure 4.26 Visualization of distortion effects estimated during photogrammetric calibration of Swissranger SR-4000 camera.

Extrinsic Parameters

Since the amplitude and depth[6] images are aligned together (see Fig. 4.25) we can extract the point from the depth map at same pixel positions as were the points for the photogrammetric calibration.

The extracted calibrated metric point cloud lies in reference frame, that is different from the one we use in the rest of this work (refer to section 3.5). We rotated the points by π around Z axis. Then we used a procedure ([16]) that finds transformation between 3D points reconstructed from the photogrammetric calibration and the 3D points calculated by the Swissranger camera software. The rotational part of the transformation was very small and therefore we decide to neglect it. For now on only the translation was used. The Camera centers (see Fig. 3.22) were find to be related as

$$C = C' + \begin{pmatrix} 15 \\ 20 \\ 70 \end{pmatrix} [\text{mm}]. \tag{4.7}$$

This results will be probably influenced by the systematic distance error - that was described in section 3.4.2.

4.2.2 Depth Data Correction

Object Intensity Effect Evaluation

The distance measurement of a TOF camera is known to be dependent on amount of light reflected from the object. This effect can be seen on Img. 4.25b, where the black chessboard squares can be clearly distinguished on the flat board.

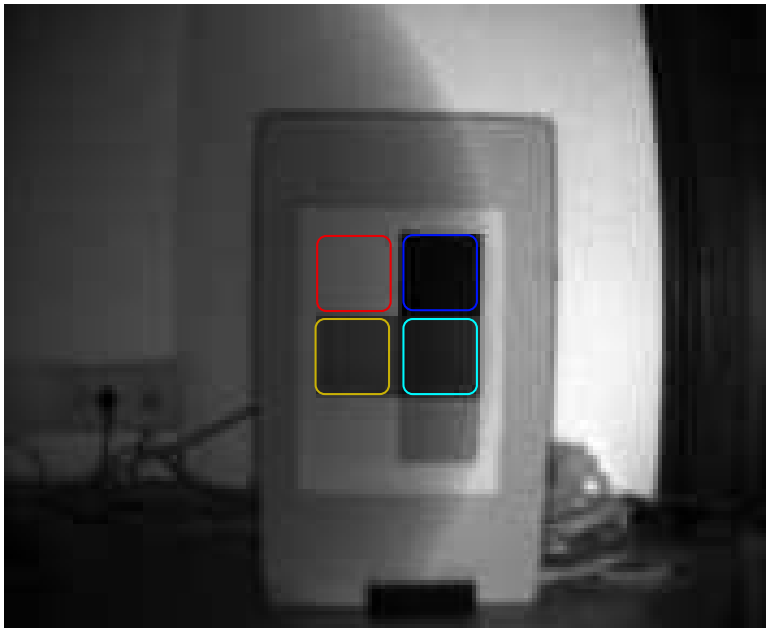


Figure 4.27 Amplitude image with evaluated areas of different intensity.

To asses this error the same method as in section 4.1.2 was used. We extracted the depth from areas illustrated on Fig. 4.27. We first plot just the distribution of the

[6] Z-coordinate distance values of the point cloud are used as depth maps with orthogonal distances.

distances from each square (see Fig. 4.28) on one image. The values are randomly distributed around the mean.

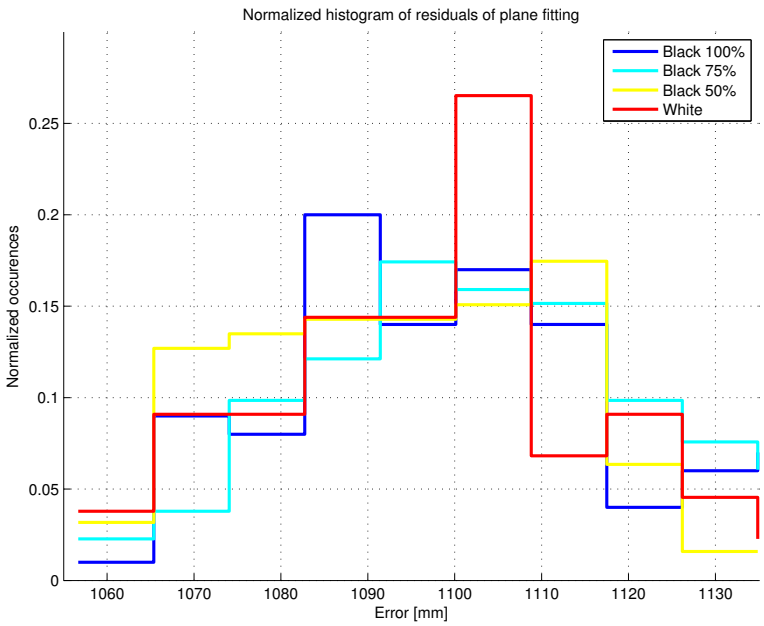


Figure 4.28 Histogram (normalized) of distances at target with different color intensity.

To asses the data in more details 10 measurements from distance range 0.75 - 1.5 m were considered. If we now evaluate the data with same methodology as we the with Kinect, we can see the results in Tab. 4.11 and as a histogram 4.29. The white color is the most suitable for measuring the distance - since the standard deviation of the error is the smallest. We can see that the mean values between black and white colors differs systematical for almost 5 mm. The high uncertainty of the black colored target can be physically justified in a way that the surface with low reflectivity has also low Signal-to-Noise ration.

Table 4.11 Residuals of plane fitting in areas with different contrast

Contrast	μ [mm]	σ [mm]
Black 100%	-4.5665	16.0967
Black 75%	-1.7057	7.2247
Black 50%	-1.0704	4.3669
White	1.0111	2.6693

Systematic Error Evaluation

The reconstruction performance was evaluated in terms of geometrical distance between points reconstructed during photogrammetric calibration of the stereo-pair of DSLR cameras (considered as ground truth) and points reconstruct from TOF camera. The error distribution for all calibration chessboards is shown in Figure 4.30. The local mean is marked by red line. Note that different colors are used to distinguish points that belong to different images used during the calibration. Numerically the performance is evaluated in table 4.12. To determine systematic trend, as described e.g. in [39], would

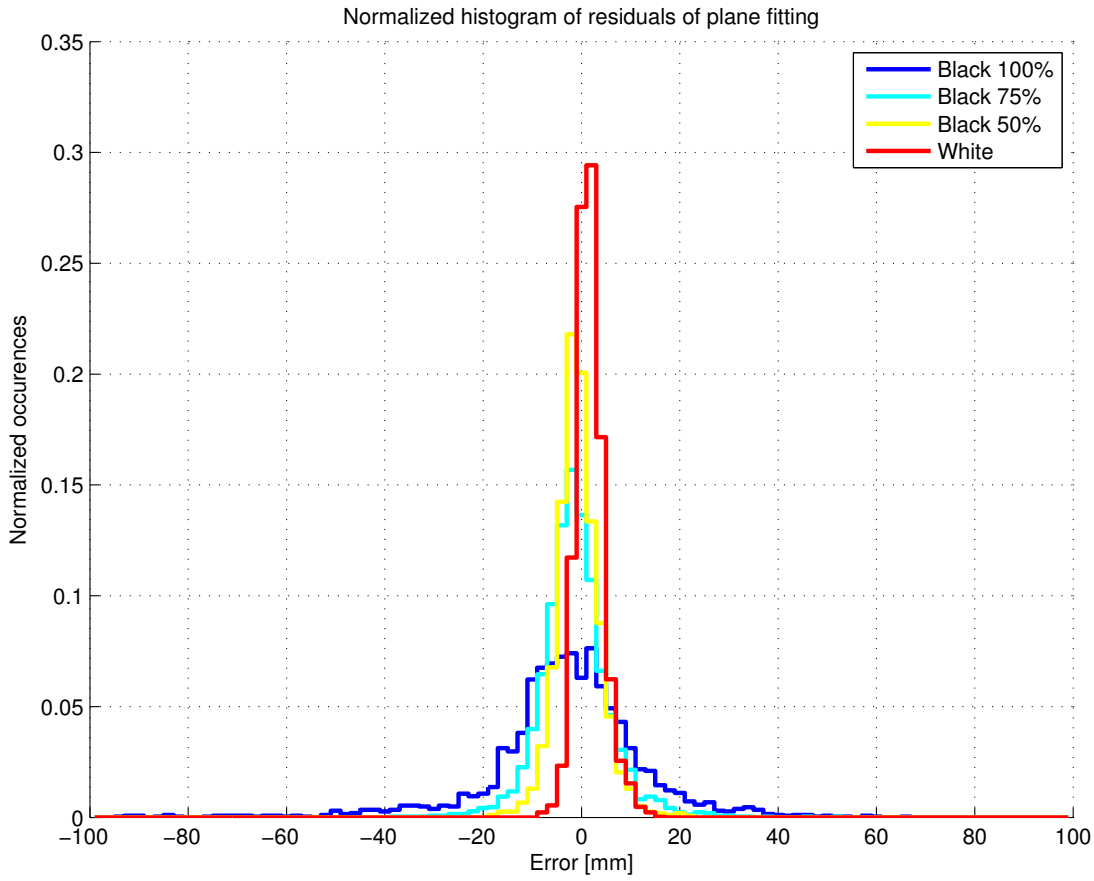


Figure 4.29 Residuals of plane fitting in areas with different contrast.

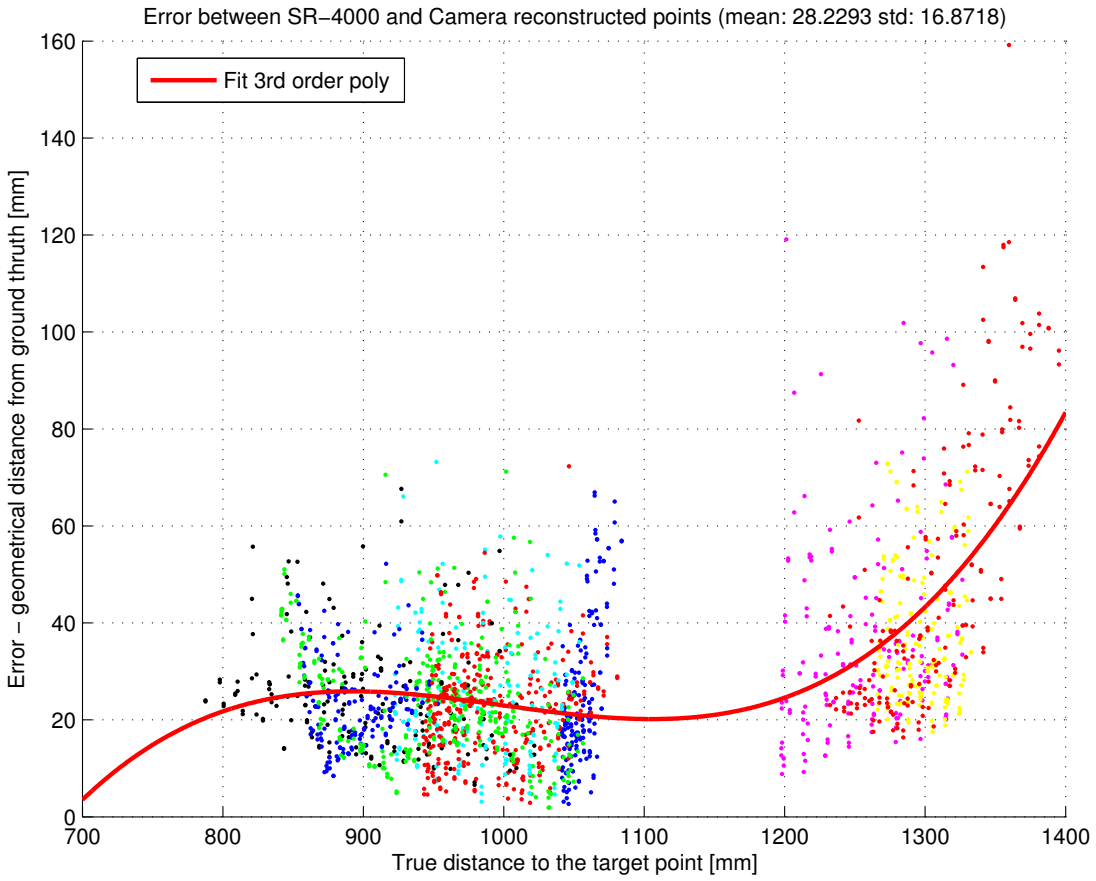


Figure 4.30 Reconstruction error of Swissranger SR-4000 camera. The solid red line marks a 3^{rd} order polynomial fit.

require more reference data. The polynomial fit is just an approximation.

Table 4.12 Reconstruction performance of Swissranger SR-4000 camera reconstruction. Each chessboard represent 88 control points.

Cal. chess.	Geometrical error [mm]		
num.	μ	σ	max
1	23.9560	15.4561	58.5652
2	21.3109	11.7171	70.2227
3	18.6130	8.7992	46.6200
4	21.4955	13.9074	68.1787
5	22.6811	12.3971	62.1742
6	35.0988	13.6793	90.4937
7	38.1913	13.5319	71.9044
8	52.4240	26.2008	133.8451
9	21.2957	11.1472	54.5754
10	20.3825	12.0323	52.1661
11	22.2823	12.2414	48.7043
Total	27.0665	17.4528	133.8451

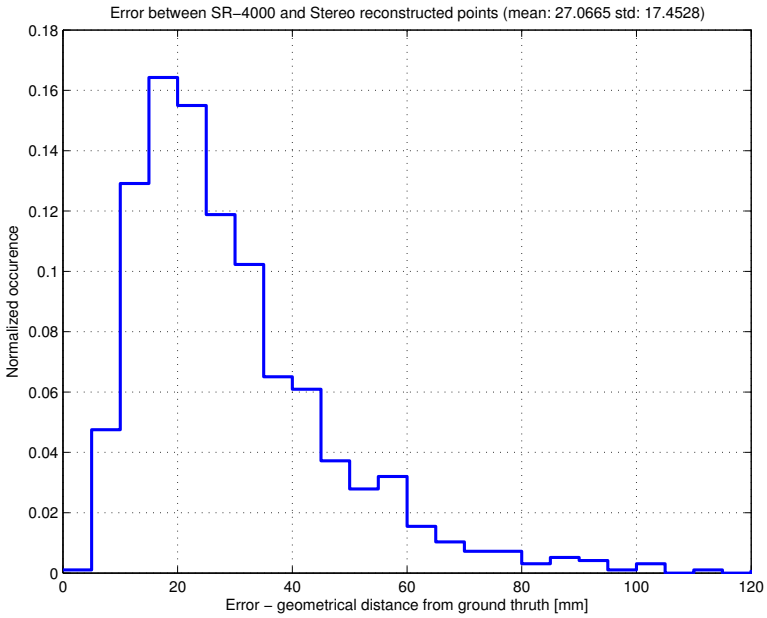


Figure 4.31 Normalized histogram of errors of Swissranger SR-4000 camera reconstruction.

4.3 Calibration of a Depth Sensing Camera and other Cameras

4.3.1 Experimental Setup

A system consisting of two consumer-grade DSLR cameras (Nikon D 60) together with a depth sensing camera (Kinect, SR-4000) was used to simultaneously capture images of calibration chessboard. During the experiment positions of all cameras was fixed and so relationship of all internal coordinate systems can be expressed by a rigid body motion parameters R, T. The focal length on DSLR cameras was set to approx. 24 mm

and the auto-focus was turned off. The distance between both cameras (baseline) was approx. 45 cm and the lenses were pointed slightly (about 5 deg) towards the center. Shutter was operated synchronously by a wireless remote control. The setup is shown in Figure 4.32.

a) Kinect and stereo **b)** Swissranger SR-4000 and stereo

Figure 4.32 Experimental setup - Depth sensing camera and two Nikon D 60 DSLR cameras.

4.3.2 Photogrammetric Calibration

Intrinsic Parameters and Distortion Model Coefficients

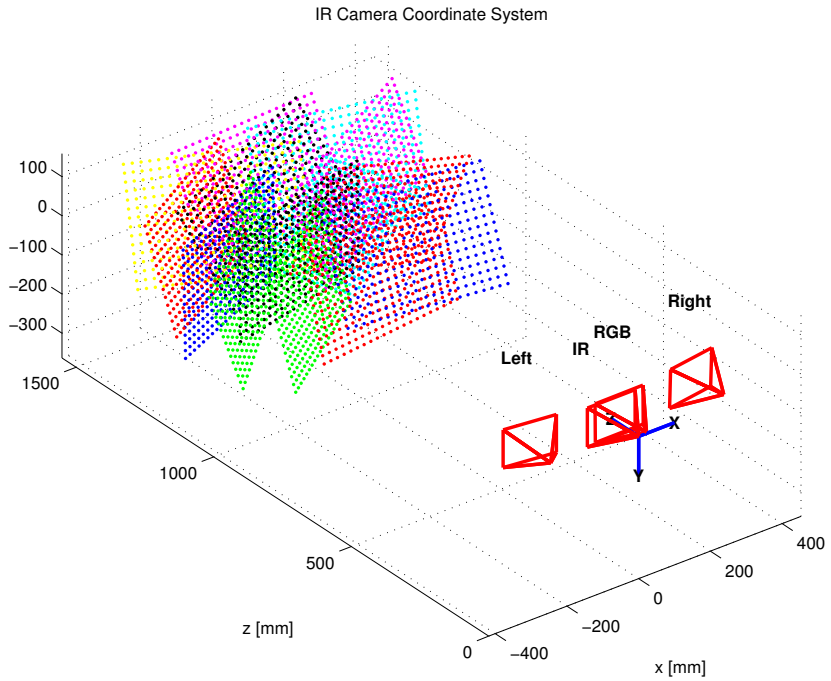
DSLR cameras used for the experiment have high resolution sensors and very good optics. Intrinsic parameters of the cameras used in the experiment with Kinect are shown in Table 4.13.

Table 4.13 Intrinsic parameters of Nikon D 60 DSLR cameras used during the experiment with Kinect

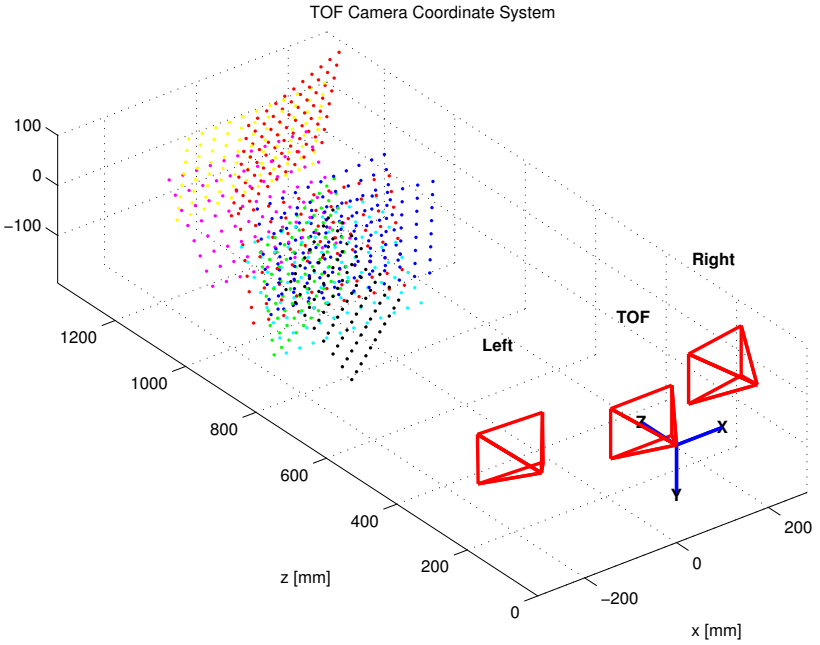
Camera	Focal length		Principal point		Distortion coefficients				
	f [px]	f [mm]	x_0 [px]	y_0 [px]	k_{c_1}	k_{c_2}	k_{c_3}	k_{c_4}	k_{c_5}
Left	4065.3	23.9	1985.8	1042.6	-0.0277	0.0367	-0.0134	-0.0035	N.A.
Right	4019.5	23.7	1772.6	1069.5	–0.0299	0.1251	0.0066	-0.0104	N.A.

4.3.3 Registering Multiple Cameras in Common Coordinate System

Exploiting the fact that during the procedure images of calibration chessboards were captured simultaneously by all cameras the relative rigid transformation between them can be determined. Using this knowledge we can form a common coordinate system and estimate rigid transformations $(R,\ T)$ to coordinate systems of all cameras. For simplicity we have chosen the reference frame to coincide with the coordinate system of the IR-camera (this camera produces two sets of data). The reconstructed experimental setups are shown in Fig. 4.33.



a) Kinect and stereo



b) Swissranger SR-4000 and stereo

Figure 4.33 Experimental setup - reconstructed from photogrammetric calibration. Note that the camera projection planes are not in scale.

4.3.4 Reconstruction Performance Comparison

Stereo Triangulation

As was already described before a major difference of Depth sensing cameras against classical stereo-based systems is that there is no need to find correspondences between images in order to determine the distance. If we disregard this issues and assume that the correspondences between stereo-camera images are exactly known we can make a performance comparison of Depth-cameras and the stereo-pair of cameras.

The corresponding points positions were selected (and stored in file) during the photogrammetric calibration procedure - and so can be now used to reconstruct their 3D positions using *Linear triangulation method*[7]. Function is implemented in Matlab according to [20] with the main idea stated bellow.

Given $P_L = [X_L\ Y_L\ Z_L]$ and $P_R = [X_R\ Y_R\ Z_R]$ to be coordinates of a projected point P in left and right camera reference frame there exist a rigid body transformation (with known parameters $R,\ T$) between them

$$X_L = RX_R + T. \tag{4.8}$$

Projecting the point on left and right image plane yields the coordinate vectors $p_L \doteq P_L/Z_L = [x_L\ y_L\ 1]$ and $p_R \doteq P_R/Z_R = [x_R\ y_R\ 1]$ that can be substituted in 4.8 to get

$$p_L Z_L = Rp_R Z_R + T,$$

which can be rearranged to

$$[-Rp_R\ p_L] \begin{bmatrix} Z_R \\ Z_L \end{bmatrix} = T,$$

and after forming a data matrix $A = [-Rp_R\ p_L]$ can be solved using pseudo-inverse as least square problem

$$\begin{bmatrix} Z_R \\ Z_L \end{bmatrix} = \left(A^{\mathrm{T}} A\right)^{-1} A^{\mathrm{T}} T. \tag{4.9}$$

Performance Comparison of Different Depth Sensing Methods

Having done all the preceding experiments, we can now present a performance comparison of tested depth sensing methods. It needs to be noted that the data were captured during different experiments. The calibration chessboard used for the measurement with SR-4000 had fewer corner points (88 compared to 315). Measurements only from common distance range for all experiments (0.9 - 1.4 m) were used. The error values can be compared in Table 4.15. To illustrate their distribution a comparison of normalized histograms is shown in Fig. 4.35.

We can conclude that, the stereo triangulation (assuming known correspondences) is superior to both depth sensing cameras. Together with Microsoft Kinect they did outperform the SR-4000 TOF camera, that had the mean reconstruction error approx 15 times bigger.

[7] Details on Linear triangulation methods can be found in [32, 31]. Procedure used in this work is detonated as *Linear–LS Method*.

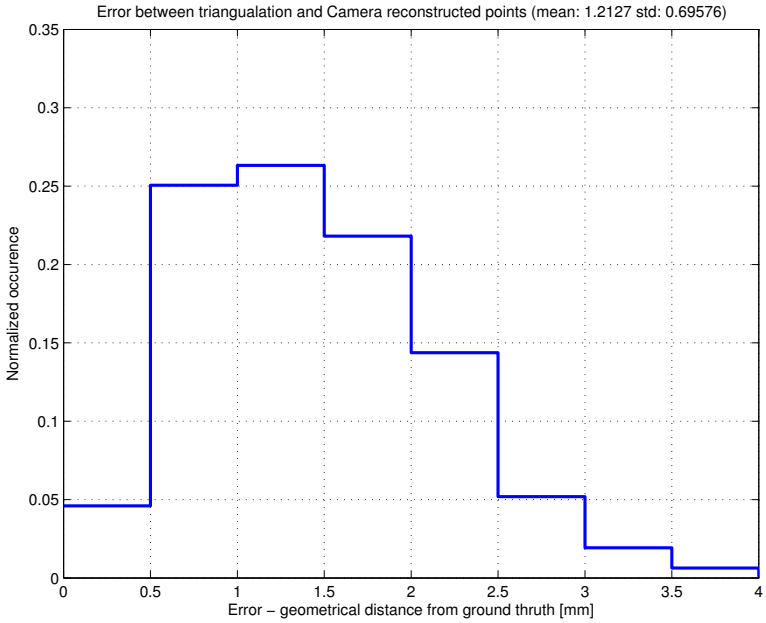


Figure 4.34 Normalized histogram of errors of stereo reconstruction.

Table 4.14 Reconstruction performance of stereo reconstruction. Data from two experiments are shown. During the measurement with Kinect each calibration chessboard consisted from 315 control points. In the experiment with TOF camera each chessboard had 88 control points.

Cal. chess.	Error [mm] - Kinect exp.			Error [mm] - TOF exp.		
num.	μ	σ	max	μ	σ	max
1	0.6500	0.4052	1.7883	0.8462	0.5276	2.0768
2	1.6167	0.4109	2.7909	1.1334	0.7685	3.0346
3	1.3194	0.3608	2.2550	1.9622	0.5067	3.0955
4	1.1518	0.5073	2.0714	1.4338	0.4786	2.5027
5	0.4866	0.2360	1.0579	1.8784	0.8240	3.7682
6	1.2011	0.6222	2.3156	5.1605	0.9953	7.3838
7	1.3555	0.9349	3.4765	1.7491	0.8219	3.3695
8	1.2913	0.5563	2.5659	1.0035	0.5495	3.1617
9	1.5356	1.0647	3.8639	1.8838	1.2930	5.2386
10	1.1726	0.5533	2.8600	2.5658	1.3199	5.6162
11	1.2288	0.5757	2.6381	1.8784	1.0125	4.1155
12	1.7935	0.5490	2.7191			
13	0.8228	0.4524	3.1111			
14	1.3525	0.7098	3.4584			
Total	1.2127	0.6958	3.8639	1.9541	1.4193	7.3838

Table 4.15 Performance comparison of stereo triangulation, Microsoft Kinect and Swissranger SR-4000 depth sensing devices. Note that the data were captured during different experiments. Measurements only from common distance range 0.9 - 1.4 m were used.

Method	Geometrical error [mm]		
	μ	σ	max
Triangulation	1.5701	1.1454	7.3838
Kinect	2.3905	1.6665	8.6367
TOF	27.6148	18.1976	133.8451

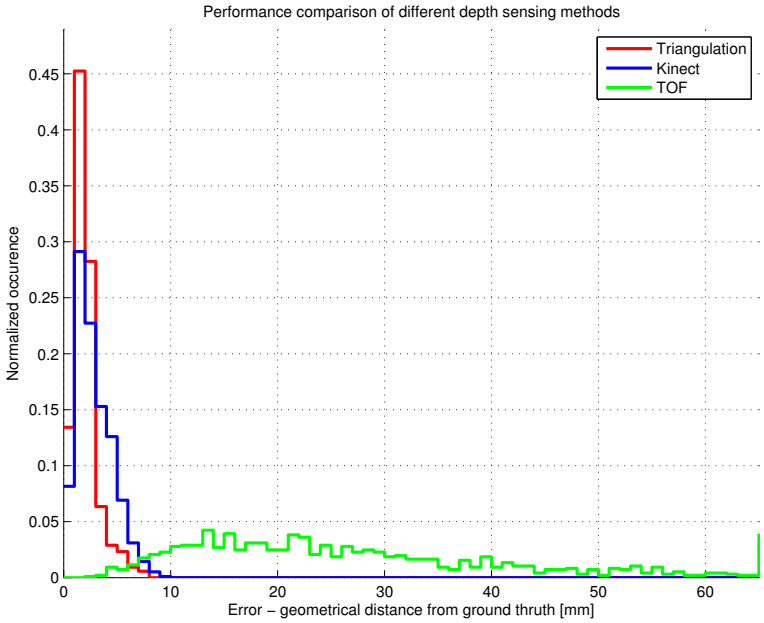


Figure 4.35 Reconstruction error distribution of stereo triangulation, Microsoft Kinect and Swissranger SR-4000 depth sensing devices. Note that the data were captured during different experiments. Measurements only from common distance range 0.9 - 1.4 m were used.

5 Application for Reconstruction

Preceding techniques can viewed as preliminary steps in more complex computer vision tasks. In this chapter we present an examples of such application.

5.1 3D Scene Reconstruction from Kinect RGB-D Camera

Scene 3D model reconstruction from intensity images is a well established method. We used the RGB-camera images from and also the corresponding depth measurements (see Figure 5.1 for example of input data). During the experiment 60 images were captured. For each measurement the color image was undistorted and the distance data were transfer to RGB-camera reference frame. The procedure to remove hidden surfaces (section 4.1.6) was done to form a depth map aligned with the color image. For each image the camera position was found using structure-from-motion method at [4] and dense reconstruction procedure described in [37] was used to create the model.

The point cloud aligned together from all Kinect measurements can be seen in Fig. 5.2.

To asses the advantage of using the additional distance measurement both results, with using only the intensity images and with using both intensity and depth image, are shown in Fig. 5.3. It is clearly visible that the model reconstructed with additional depth information is much better.

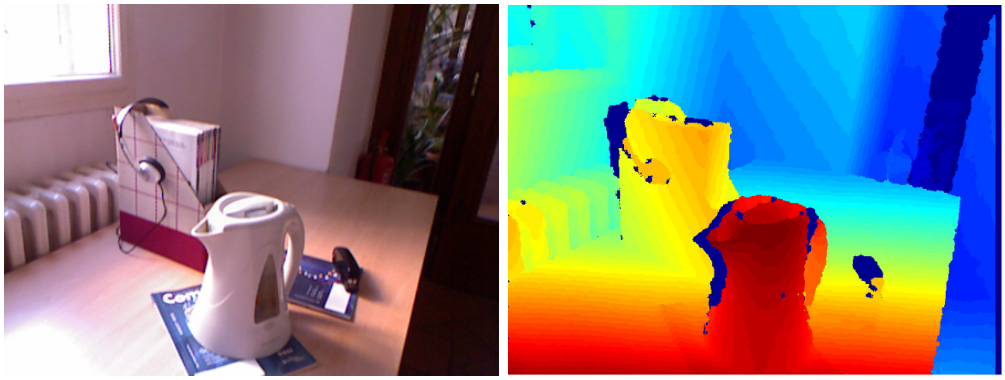
5.2 3D Scene Reconstruction using combined Stereo-pair and Kinect Camera

In this experiment we used setup shown on Fig. 5.5 consisting of Kinect and two DSLR Nikon D-60 cameras.

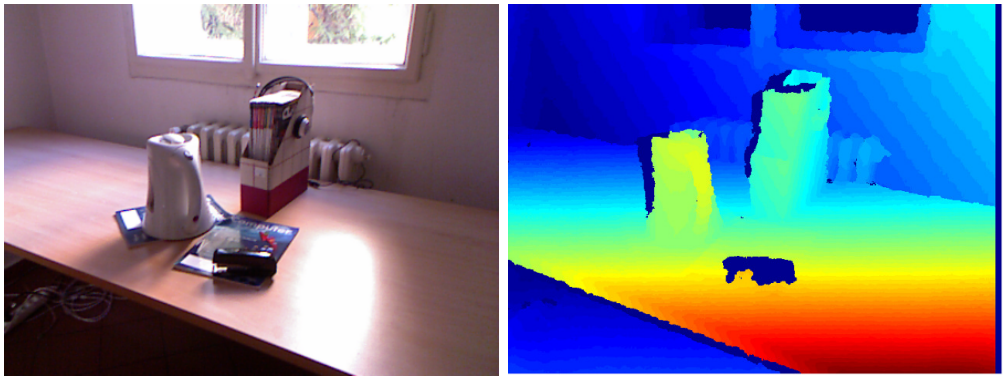
During the experiment 2 x 24 images were captured together with the distance map. For each measurment the color images was undistorted and the distance data were interpolated to match the higher resolution of the cameras and transfered to both reference frames (see Fig. 5.4). The procedure to remove hidden surfaces (section 4.1.6) was done to form a depth map aligned with the color image. For each image the camera position was found using structure-from-motion method at [4] and dense reconstruction procedure described in [37] was used to create the model.

Point clouds reconstructed from Camera and Kinect depth maps are shown in Fig. 5.6. Note that the Kinect point cloud is denser, but not perfectly aligned. This could be probably improved by using ICP algorithm - and it remains as recommended future work.

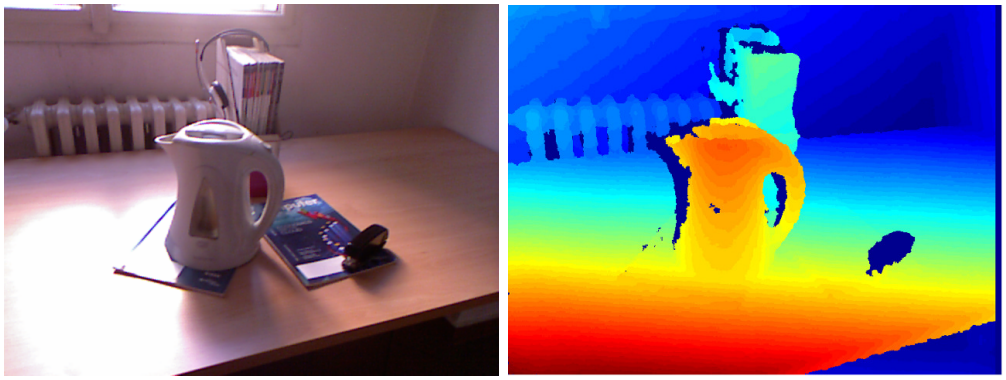
If we have a look at Figures 5.7 and 5.8 the quality of the reconstruction is comparable. Better alignment of the Kinect point clouds would probably improve the reconstruction.



a) RGB image **b)** Depth image

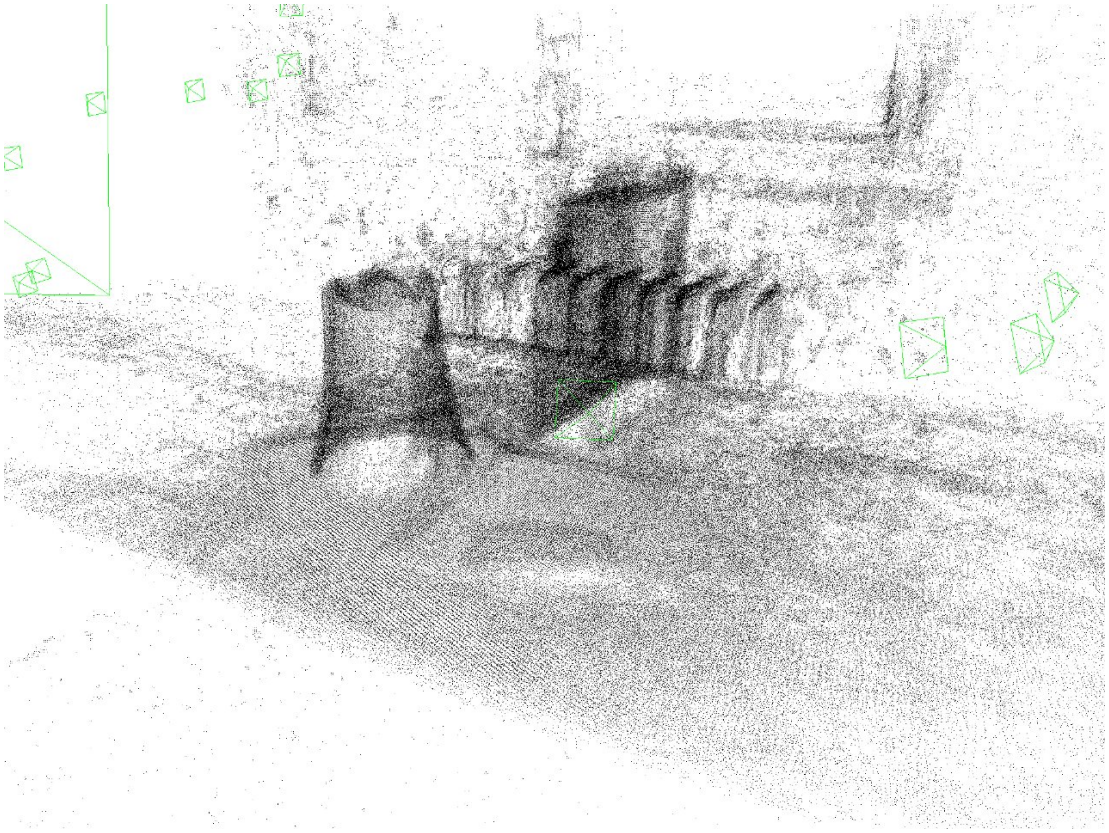


c) RGB image **d)** Depth image



e) RGB image **f)** Depth image

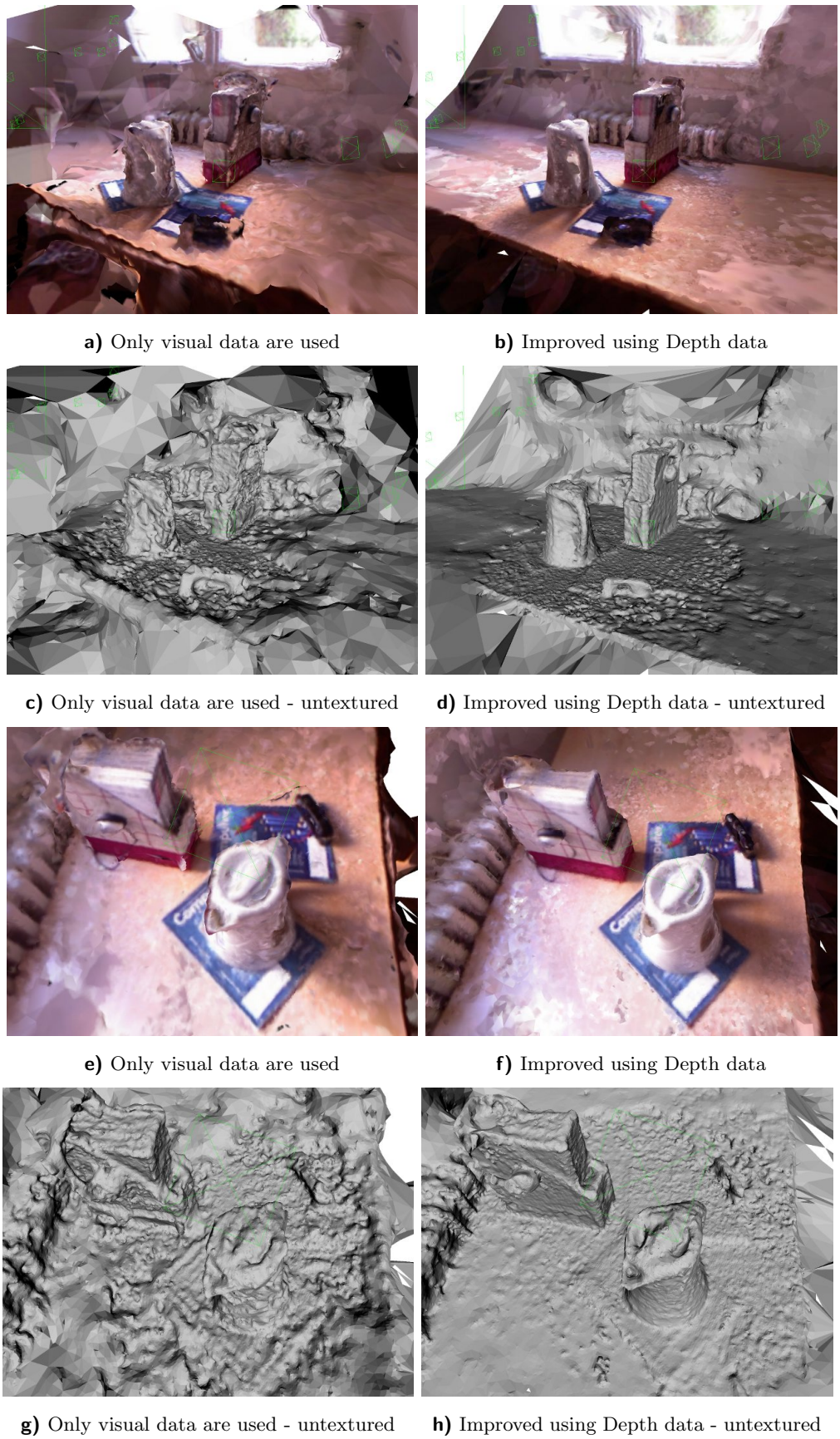
Figure 5.1 Several examples of images from Kinect Depth and RGB-camera that were used for scene reconstruction.



a) View 1

b) View 2

Figure 5.2 Point clouds captured by the Kinect Depth sensing camera matched together during the reconstruction procedure.



a) Only visual data are used

b) Improved using Depth data

c) Only visual data are used - untextured

d) Improved using Depth data - untextured

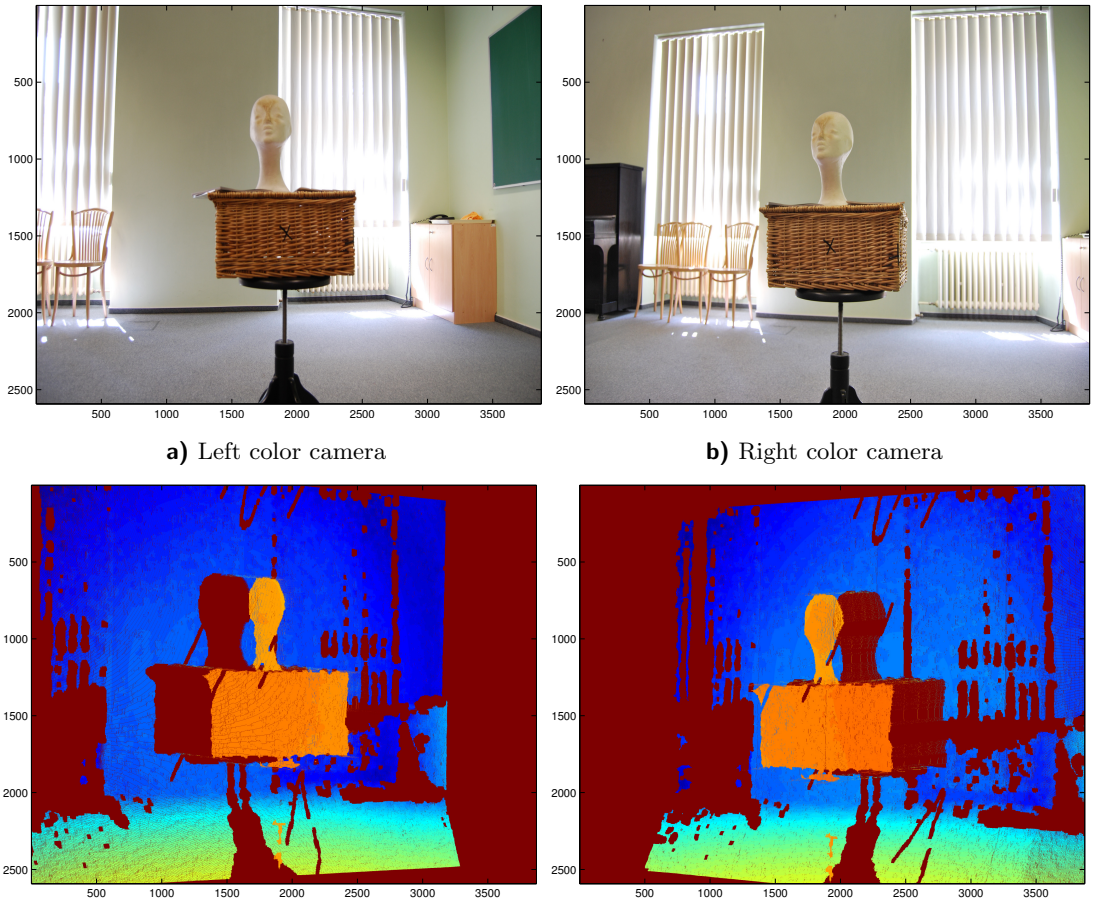
e) Only visual data are used

f) Improved using Depth data

g) Only visual data are used - untextured

h) Improved using Depth data - untextured

Figure 5.3 Scene Reconstruction from RGB-D Camera. The figure shows a comparison of reconstruction quality when the scene is reconstructed only using *structure-from-motion* and the case when the depth information is also available from Depth sensing camera.



a) Left color camera

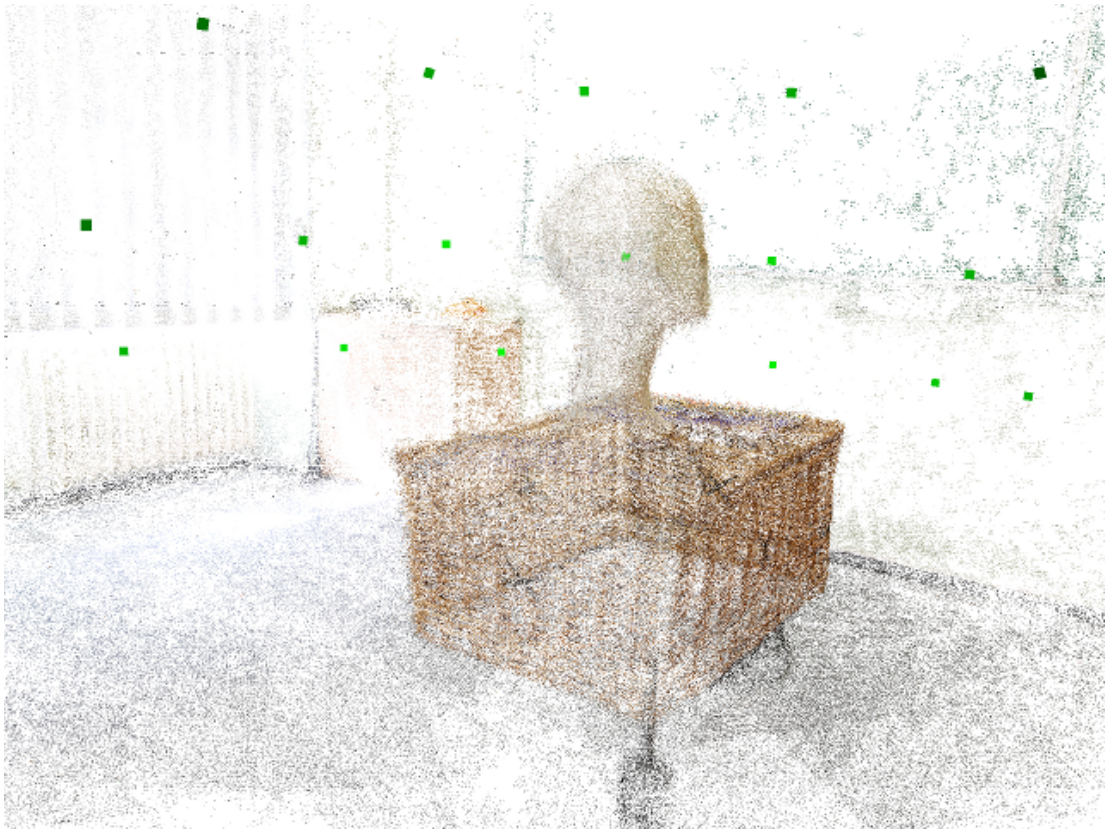
b) Right color camera

c) Depth map in left camera frame

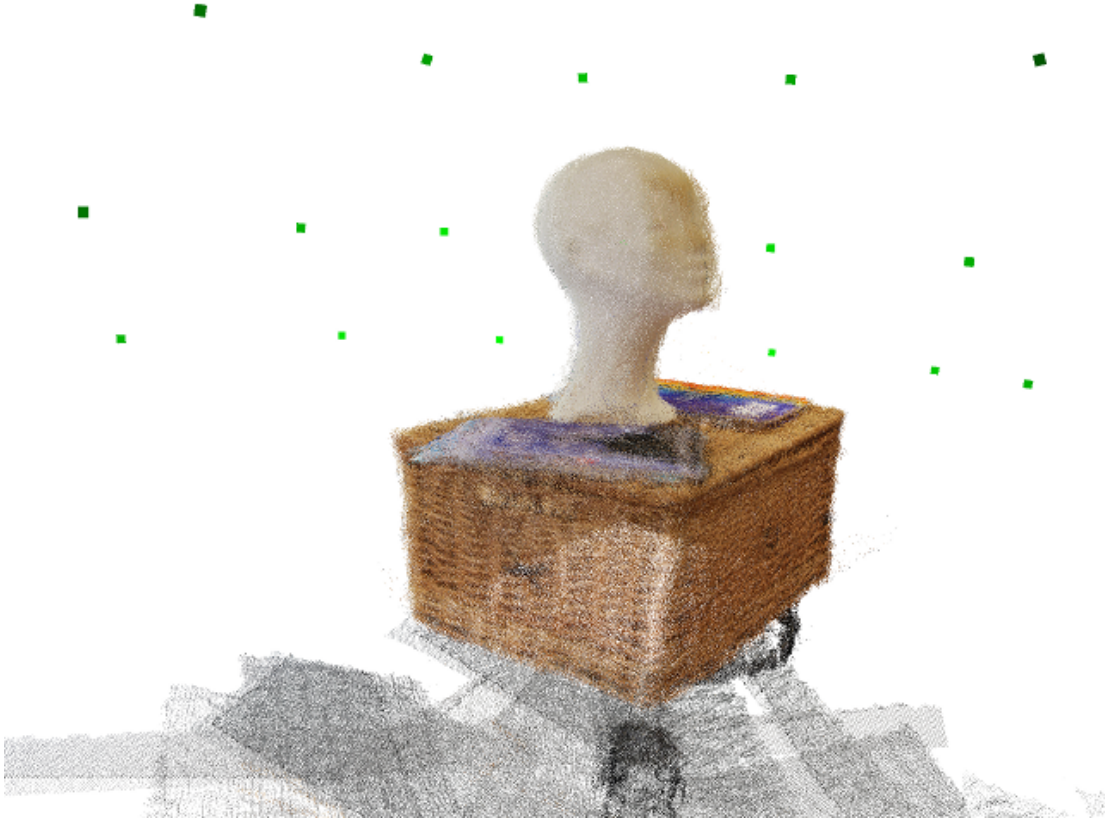
d) Depth map in right camera frame

Figure 5.4 Input data example for the second reconstruction experiment.

Figure 5.5 Object to be reconstructed (a bust) together with the camera setup (combined stereo-pair and Kinect Camera).



a) Camera



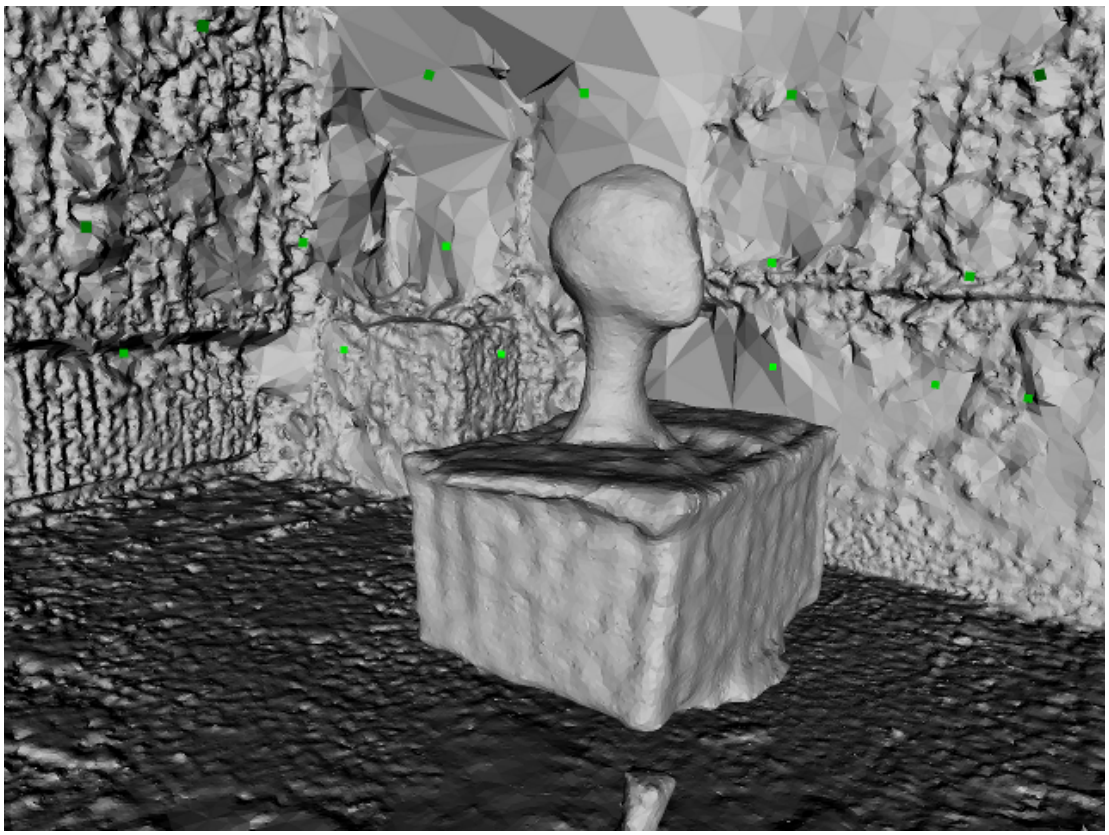
b) Kinect

Figure 5.6 Point clouds reconstructed from Camera and Kinect depth maps. Note that the Kinect point cloud is denser, but not perfectly aligned.

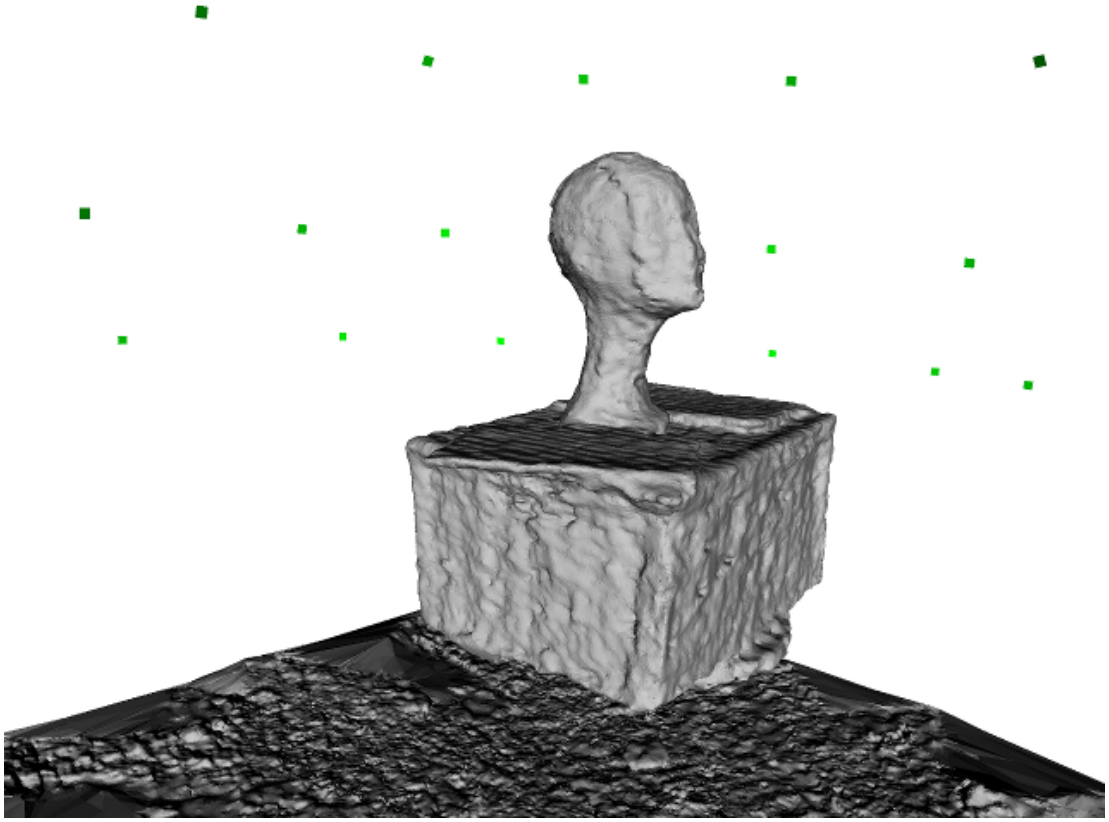
a) Camera

b) Kinect

Figure 5.7 Object reconstructed from Camera and Kinect depth maps maps.



a) Camera



b) Kinect

Figure 5.8 Untextured object reconstructed from Camera and Kinect depth maps.

6 Conclusion

We studied the topic of depth sensing camera calibration. Two 3D Cameras Microsoft Kinect and Swissranger SR-4000, that work on different physical principles, were investigated. The devices were described and subjected to experiments in order to evaluate their performance. Several systematic error sources were identified and we proposed a method to compensate for them. Both cameras can be registered together with other conventional/depth cameras in common coordinate frame.

We proposed a practical solution to Kinect camera calibration. The device was not yet well described in the literature and therefore we presented our evaluation of the reconstruction performance. Error sources and other technical details that have been identified, are discussed. The resulting method produce a calibrated metric point cloud with assigned color from the Kinect RGB camera.

While TOF camera calibration procedures were already widely investigated in the literature, we have tested several of them on our SR-4000 camera. Because the camera directly produces a metric point cloud, we evaluated its accuracy and provided a method to use the device together with other cameras in the same reference frame.

A comparison of reconstruction performance of the 3D cameras and a stereo-pair of cameras was presented in section 4.1.3. If we compare mean geometrical error, the stereo triangulation is superior to both depth sensing cameras ($\mu = 1.6$ mm). The Kinect, that is 1.5 times worse with ($\mu = 2.4$ mm), still outperforms the SR-4000 TOF camera ($\mu = 27.6$ mm).

Finally, we show an application of the depth sensing camera together with conventional color camera in area of complex scene reconstruction. At certain situation (low quality color camera), using of the depth information from Kinect was clearly superior. In the second experiment, where DSLR cameras were used, the visual quality of the reconstruction while using the depth information from Kinect was comparable to the state of the art reconstruction algorithm.

A Content of the Enclosed CD

- Dp_2011_smisek_jan.pdf - diploma thesis pdf report file,
- /data/ - data captured during the calibration,
- /src/ - matlab source code of calibration procedure.

Bibliography

[1] Technical description of kinect calibration. http://www.ros.org/wiki/kinect_calibration/technical. 4, 10, 39

[2] Academics, enthusiasts to get kinect sdk. http://research.microsoft.com/en-us/news/features/kinectforwindowssdk-022111.aspx, April 2011. 4

[3] Blender. http://www.blender.org/, April 2011. 48

[4] Cmp sfm web service. http://ptak.felk.cvut.cz/sfmservice/, April 2011. 62

[5] Hacking the kinect. http://www.ladyada.net/learn/diykinect/, March 2011. 4

[6] Meshlab. http://meshlab.sourceforge.net/, April 2011. 48

[7] Open kinect project. http://openkinect.org/, March 2011. 4

[8] Openni. http://openni.org/, April 2011. 4

[9] Planetary robotics vision ground processing. http://www.provisg.eu/, April 2011. 2

[10] Planetary robotics vision scout. http://www.proviscout.eu/, April 2011. 2

[11] Proviscout: Esa and nasa develop an independent mars exploration system. http://www.wired.co.uk/news/archive/2010-09/23/proviscout, April 2011. 1

[12] Rgb-d: Techniques and usages for kinect style depth cameras. http://ils.intel-research.net/projects/rgbd, April 2011. 4

[13] T. Kahlmann A, F. Remondino B, and H. Ingens. Calibration for increased accuracy of the range imaging camera swissranger tm. 41

[14] Aptina. Mt9m001c12stm datasheet. http://www.aptina.com/products/image_sensors/mt9m001c12stm/, April 2011. 9

[15] Aptina. Mt9m112 datasheet. www.aptina.com/assets/downloadDocument.do?id=519, April 2011. 10

[16] K. S. Arun, T. S. Huang, and S. D. Blostein. Least-squares fitting of two 3-d point sets. *IEEE Trans. Pattern Anal. Mach. Intell.*, 9:698–700, September 1987. 52

[17] F. G. Becerro. External-self-calibration of a 3d time-of-flight camera in real environments, 2008. 5

[18] Christian Beder, Bogumil Bartczak, and Reinhard Koch. A comparison of pmd-cameras and stereo-vision for the task of surface reconstruction using patchlets. In *In Proceedings of the second international ISPRS workshop BenCOS*, 2007. 4

[19] Jean Yves Bouguet. Camera calibration toolbox. http://www.vision.caltech.edu/bouguetj/calib_doc/. 7, 24

[20] Jean Yves Bouguet. Stereo triangulation in matlab. http://www.multires.caltech.edu/teaching/courses/3DP/ftp/98/hw/1/triangulation.ps, April 1998. 59

[21] Paul Bourke. Ply - polygon file format. http://paulbourke.net/dataformats/ply/, April 2011. 48

[22] Duane C. Brown. Close-range camera calibration. *PHOTOGRAMMETRIC ENGINEERING*, 37(8):855–866, 1971. 7

[23] Nicolas Burrus. Kinect calibration. http://nicolas.burrus.name/index.php/Research/KinectCalibration, March 2011. 4, 25, 39

[24] Bernhard Büttgen, Thierry Oggier, Michael Lehmann, Rolf Kaufmann, and Felix Lustenberger. Ccd/cmos lock-in pixel for range imaging: Challenges, limitations and state-of-the-art. *CSEM, Swiss Cen- ter for Electronics and Microtechnology.*, 2004. 14, 20

[25] Filiberto Chiabrando, Roberto Chiabrando, Dario Piatti, and Fulvio Rinaudo. Sensors for 3d imaging: Metric evaluation and calibration of a ccd/cmos time-of-flight camera. *Sensors*, 9(12):10080–10096, 2009. 4

[26] T. A. Clarke and J. G. Fryer. The Development of Camera Calibration Methods and Models. *The Photogrammetric Record*, 16(91):51–66, 1998. 24

[27] S. Fuchs and G. Hirzinger. Extrinsic and depth calibration of tof-cameras. pages 1–6, 2008. 4

[28] Rafael C. Gonzalez and Richard E. Woods. *Digital Image Processing*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 2nd edition, 2001. 31

[29] Manuel Guizar-Sicairos, Samuel T. Thurman, and James R. Fienup. Efficient subpixel image registration algorithms. *Opt. Lett.*, 33(2):156–158, Jan 2008. 31

[30] Uwe Hahne and Marc Alexa. Combining time-of-flight depth and stereo images without accurate extrinsic calibration. *Int. J. Intell. Syst. Technol. Appl.*, 5:325–333, November 2008. 4

[31] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521540518, second edition, 2004. 59

[32] Richard I. Hartley and Peter Sturm. Triangulation, 1994. 59

[33] Rapp Holger. Experimental and theoretical investigation of correlating tof-camera systems, 2007. 4, 20

[34] S. Hussmann, T. Ringbeck, and B. Hagebeuker. A performance review of 3d tof vision systems in comparison to stereo vision systems. http://www.pmdtec.com/fileadmin/pmdtec/downloads/documentation/White_Paper_3D_TOFvs.Stereo.pdf, March 2011. 2, 20

[35] Mesa Imaging. Swissranger sr4000 user manual. http://www.mesa-imaging.ch/dlm.php?fname=customer/Customer_CD/SR4000_Manual.pdf, April 2011. 22

[36] Chipworks Inc. Teardown of the microsoft kinect. http://www.chipworks.com/en/technical-competitive-analysis/resources/recent-teardowns/2010/12/teardown-of-the-microsoft-kinect-focused-on-motion-capture/, April 2011. 9

[37] M. Jancosek and T. Pajdla. Multi-view reconstruction preservingweakly-supported surfaces. 2011. 62

[38] Y.M. Kim, D. Chan, C. Theobalt, and S. Thrun. Design and calibration of a multi-view tof sensor fusion system. pages 1–7, 2008. 5

[39] A. Kolb, E. Barth, R. Koch, and R. Larsen. Time-of-Flight Sensors in Computer Graphics. In M. Pauly and G. Greiner, editors, *Eurographics 2009 - State of the Art Reports*, pages 119–134, CH-1288 Aire-la-Ville, March 2009. Eurographics Association, Eurographics. 4, 21, 53

[40] Robert Lange and Peter Seitz. Solid-state time-of-flight range camera. *JOURNAL OF QUANTUM ELECTRONICS*, 2001. 21

[41] Marvin Lindner, Andreas Kolb, and Klaus Hartmann. Data-fusion of pmd-based distance-information and high-resolution rgb-images. *2007 International Symposium on Signals Circuits and Systems*, 1:1–4, 2007. xii, 45, 48

[42] Marvin Lindner, Ingo Schiller, Andreas Kolb, and Reinhard Koch. Time-of-flight sensor calibration for accurate range sensing. *Computer Vision and Image Understanding*, 114(12):1318 – 1328, 2010. Special issue on Time-of-Flight Camera Based Computer Vision. xi, 4, 20

[43] Stéphane Magnenat. Stéphane magnenat's distance model. http://groups.google.com/group/openkinect/browse_thread/thread/31351846fd33c78/e98a94ac605b9f21, April 2011. 39

[44] Stefan May, David Droeschel, Stefan Fuchs, Dirk Holz, and Andreas Nüchter. Robust 3d-mapping with time-of-flight cameras. In *IROS*, 2009. 5

[45] Stefan May, David Droeschel, Dirk Holz, Stefan Fuchs, Ezio Malis, Andreas Nüchter, and Joachim Hertzberg. Three-dimensional mapping with time-of-flight cameras. *J. Field Robotics*, 26(11-12):934–965, 2009. 4, 21

[46] Kyle McDonald. Openni distance model. http://groups.google.com/group/openkinect/browse_thread/thread/55a6855549ef5559/541a1fb4c046d92f, April 2011. 39

[47] S. Meroli. Noise analysis of particle sensors and pixel detectors. fixed pattern noise and pixel noise measurement. http://meroli.web.cern.ch/meroli/Lecture_Particle_Detector_Noise.html, April 2011. 41

[48] Daniel Reetz and Matti Kariluoma. Looking at kinect ir patterns. http://www.futurepicture.org/?p=116", March 2011. 9

[49] F. Remondino and C. Fraser. Digital camera calibration methods: considerations and comparisons. In *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences,*, volume Vol. XXXVI, Dresden, Germany, 2006. 24

[50] Sebastian Schuon, Christian Theobalt, James Davis, and Sebastian Thrun. Lidarboost: Depth superresolution for tof 3d shape scanning. 2009. 5

[51] Emanuele Trucco and Alessandro Verri. *Introductory Techniques for 3-D Computer Vision*. Prentice Hall PTR, Upper Saddle River, NJ, USA, 1998. xi, 6, 7, 12

[52] J. Weng, P. Cohen, and M. Herniou. Camera calibration with distortion models and accuracy evaluation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 14(10):965–980, 1992. xi, 8

[53] Wikipedia. Opportunity rover. http://en.wikipedia.org/wiki/Opportunity_rover, April 2011. 1

[54] Wikipedia. Spirit rover. http://en.wikipedia.org/wiki/http://en.wikipedia.org/wiki/Spirit_rover, April 2011. 1

[55] Wikipedia. Structured-light 3d scanner. http://en.wikipedia.org/wiki/Structured-light_3D_scanner, March 2011. 9

[56] Wikipedia. Time-of-flight camera. http://en.wikipedia.org/wiki/Time-of-flight_camera, March 2011. 14

[57] Wikipedia. Z-buffering. http://en.wikipedia.org/wiki/Z-buffering, April 2011. 45

[58] Zhengyou Zhang. A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22:1330–1334, 2000. 24