Registration of range images using geometric features

Master's Thesis in Geomatics

Geoscience & Remote Sensing
Faculty of Civel Engineering
Delft University of Technology
Mekelweg 4, 2628 CD Delft, The Netherlands

Hoe-Ming Wong
homing85@gmail.com

5th June 2012

TU Delft Delft University of Technology

Graduation

Author
Hoe-Ming Wong (homing85@gmail.com)
Title
Registration of range images using geometric features
MSc presentation
5th June 2012

Graduation Committee

Prof.dr. Massimo Menenti (Graduation Professor)	Delft University of Technology
Dr.ir. Ben Gorte (Supervisor)	Delft University of Technology
Dr. Alexander Bucksch (Supervisor)	The Georgia Institute of Technology
Dr. Hugo Ledoux (Co-reader)	Delft University of Technology
Dr.ir. M.J.P.M. Lemmens (External examiner)	Delft University of Technology

Abstract

Measuring our environment in 3D in real-time has never been so close to the public before. Range cameras, also called 3D cameras, are sensors capable of providing range and amplitude images in real-time. Each pixel in an amplitude image gives information of the amount of reflected light, while each pixel in the range image gives the metric radial distance from the optical center of the camera to a surface in the field of view. The range image can be directly transformed into a 3D point cloud providing instant 3D information in the field of view of the sensor.

The focus in the thesis is on computing one 3D point cloud of the scanned environment from a series of range images. The images are acquired by moving the range camera slowly through a static indoor environment.

The process to 'stitch' all the point clouds into one consistent point cloud is called registration. Existing methods for registration of range images are mostly based on the following two techniques:

a. Iterative Closest Point (ICP) - two point clouds a and b are registered iteratively to transform the points of b into the coordinate system of a. The calculation is complete if the summed distances between the point cloud is minimized.

b. Image feature tracking techniques - like Scale Invariant Feature Transform (SIFT) can find unique feature points from the amplitude image and track them within a series of images. Direct feature point correspondences between subsequent images are mapped onto the corresponding feature points from the range image. Three corresponding 3D points not in a plane are required to calculate the transformation between two point clouds.

The feature tracking techniques reduce the point sets used to compute the transformation to unique point correspondences. Disadvantages: Low registration accuracy because less points are considered. Sensitive to lightning condition because it depends on the amplitude image and the image features may not be reliable because they are not 3D. Advantage: It is fast because it reduces the points for registration significantly and it provides direct point correspondence. The ICP algorithm takes all points into account for determining the transformation. Disadvantages: Slow, mainly because all points are considered. False registration can occur with ICP, if the overlapping parts of the two point clouds to be registered are not determined accurately. Advantage: Accurate registration.

Both techniques were not designed to register range images, but are a convenient basis to tackle the problem of real-time registration of range image series. The combination of the mentioned techniques proved to be more robust and adequately fast than a single technique.

In this thesis a novel method is developed specialized to register range image series. The chosen approach combines the advantages of both current techniques. To reduce the point set to more meaningful points for the registration, geometric feature points are detected within the 3D point cloud of the range images. An adapted ICP algorithm is used to register the geometric features points. The correspondence finding process takes feature characteristics into account, to increase the robustness to find the correct correspondence.

The geometric features discussed in this thesis are physical edges like edges of a table, characterized by its 3D shape. The geometric features are detected in the range image. The feature points have a distinct topology with the neighboring points, they are either concave or convex. The physical edges are captured discontinuous if a part of the edge is occluded from the camera perspective and continuous if the edge is completely visible. The geometric features are viewpoint independent to ensure consistency in the location. Viewpoint independent means the features do not change in 3D with respect to environment when viewed from a different camera perspective.

The classification method to find the geometric features takes the measurement errors caused by the camera and/or environment into account. In particular, the systematic error caused by lens distortion and distance-related error is corrected based on a correction model derived from repeated measurements in a test setup. The non systematic errors such as random noise is minimized by median filtering and floating points caused by the effect of discontinuous edges are detected and removed. Additionally remaining errors are identified in the registration results.

In the feature point extraction the distance measurement performance of the range camera is taken into account. This is done by finding the minimal detectable feature size for each distance. The minimal detectable feature size is the minimum required surface size at a certain distance before the range camera can measure a correct distance value. This is derived from a test setup with repeated measurements of a 3D Siemens Star.

In the correspondence matching step of the adapted ICP method, edge information is included for finding the "closest" and "similar" points as corresponding pairs. This results in more reliable matches and more robust registration.

The quality of the registration result is dependent on the scene properties and the camera movements. The reflectivity of surfaces in the scene has showed to be of significant influence in the registration results.

Preface

First I would like to express my gratitude to both my supervisors Alexander Bucksch and Ben Gorte for their unlimited support and enthusiasm for my topic. Giving me freedom to explore and useful insights in the different problems I have faced during the project. Thank Ben Gorte and Sisi Zlatanova for igniting my interest for the range camera in the master course "3D GIS". Together with my supervisors I was able to find my path in creating a valuable thesis research.

Furthermore I would like to thank my graduation professor Massimo Menenti for the support and insights. Thank Hugo Ledoux for being my co-reader and providing me some useful tips. My deepest gratitude for Joshua Weitz and Alexander Bucksch, for the hospitality and support during my research period at Georgia Institute of Technology. Special thanks goes to Alexander, I enjoyed the time working together in Atlanta and the great conversations we had. Last but not least thank my family and my friends who have been there to support me.

Hoe-Ming Wong

Delft, The Netherlands
5th June 2012

Contents

Preface **v**

1 Introduction **3**
1.1 Background . . . 3
1.2 Problem statement . . . 6
1.3 Research objectives . . . 8
1.4 Methodology . . . 9
1.5 Outline of the thesis . . . 10

2 Theory and Literature review **11**
2.1 Range camera . . . 12
2.1.1 Swissranger SR4000 . . . 12
2.2 Error and calibration . . . 16
2.2.1 Systematic errors . . . 17
2.2.2 Random errors . . . 19
2.2.3 Calibration techniques . . . 21
2.3 Registration approaches . . . 24
2.3.1 Iterative closes point: ICP . . . 25
2.3.2 Efficient ICP . . . 26
2.3.3 Registration approaches with range camera . . . 26
2.4 Summary and conclusion . . . 29

3 Calibration and 3D resolution analysis **31**
3.1 Lens distortion correction . . . 32
3.1.1 Result lens distortion calibration . . . 33
3.2 Depth distortion calibration . . . 37
3.2.1 Distance-related error . . . 37
3.2.2 Fixed pattern noise . . . 40
3.2.3 Result systematic error correction . . . 41
3.3 Non-systematic error correction . . . 43
3.4 Minimal detectable surface size . . . 45

4 Registration method with Geometric features **49**
4.1 Geometric feature classification . . . 49
4.1.1 Geometric features: Viewpoint independent 3D edges . . . 50
4.1.2 The classification method . . . 53
4.1.3 The classification method: Algorithm design . . . 54
4.1.4 The classification method: Limiting case . . . 61

4.2 Geometric feature registration method 62
4.2.1 Input parameters 65

5 Results **67**
5.1 Registration results 67
5.1.1 Scene 1: Box on a black table nearby a corner. 69
5.1.2 Scene 1: Box on a white table nearby a corner. 72
5.1.3 Scene 2: Part of room with bookshelf and sofa 74
5.1.4 Scene 3: Freeform object 76
5.1.5 Input parameters 79
5.2 Comparison with existing ICP approaches 80

6 Conclusions **81**
6.1 Method performance 81
6.2 Limitations 82
6.2.1 Limits by sensor performance 82
6.2.2 Limits by geometric features: Viewpoint independent edges 82
6.2.3 Limits by the adapted ICP algorithm 83
6.3 Answers to research questions 83
6.4 Future work 84

Nomenclature

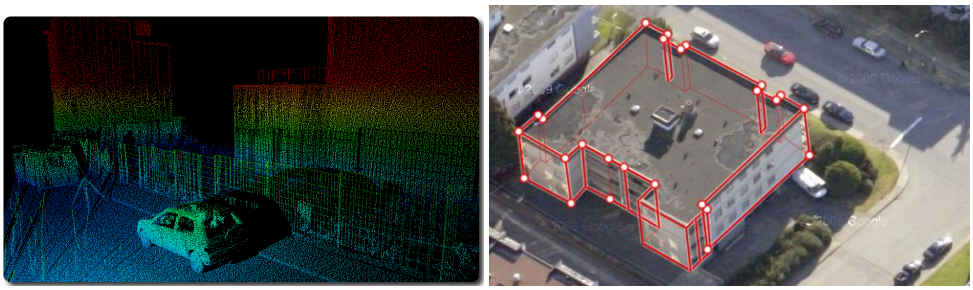
abbreviation	**Description**
$2D$	Two dimensional
$3D$	Three dimensional
CCD	Charge-coupled device
$CMOS$	Complementary Metal Oxide Semiconductor
ESM	Efficient Second-order Minimization
FOV	Field of view
$IFOV$	Instantaneous field of view
ICP	Iterative Closest Point
KLT	Kanade-Lucas-Tomasi Feature tracker
$LIDAR$	Light Detection And Ranging
$RANSAC$	Random Sample Consencus
RMS	Root Mean Square

Chapter 1

Introduction

1.1 Background

Perception of depth is one of the great abilities of the human sensory system. Depth perception allowed us to describe the world in three dimensions(3D). Throughout history the concept 'spatial information' was formed by describing the physical location of objects and the relationships between the objects. Different studies have emerged related to acquiring, enhancing and applying spatial information. Surveying, cartography, geodesy and photogrammetry made it possible to describe, navigate, monitor and analyze every location on the earth surface. The physical locations of objects described on cartographic maps can now be modeled in 3D. The electronic/digital era accelerated the developments in acquisition and processing of 3D information to such an extent, that almost every surface around us is recorded and represented as 3D models. The 3D surface data for these models is commonly acquired by laser scanners or photogrammetric system (See figure 1.1). 3D data acquisition and processing is expensive, mostly done by Geomatics experts and applied in a certain professional field (Oil industry, Military, Civil services etc). With the arrival of a new generation of sensors, called range cameras, scanning the environment in 3D has come closer to the public than ever before.



(a) A laser scanned data set (b) Photogrammetric data model

Figure 1.1: 3D points from real environment

3D models of the real environment are based on sets of sampled points on object surfaces, so-called point clouds. The points are expressed in 3D coordinates or distances in the coordinate system of the measurement instrument. The coordinates describe the physical location of a sample in the environment.

As mentioned before most 3D point clouds of environments are now scanned by laser scanners and photogrammetric systems:

1. Laser scanner (LIDAR) is an active system based on mainly two principles:

   - Time of flight: Distance computed based on physical characteristics of light: Frequency, amplitude and speed of light.

   - Triangulation: Compute the distance based on triangle: moving mirror, a sensor and light emitter.

Using one of the two principles, the light is emitted in a pattern through the scene. Together with the knowledge of the scanner's pose and movement, distances to points in the scene are measured one by one at high speed (up to 90000 p/s (Faro, 2012)) (See figure 1.1).

2. Photogrammetric system is a passive system:
Depth information is retrieved from two photographs taken from different perspectives. With the intrinsic camera parameters known, matching enough corresponding points from the two images, the 3D coordinates of points in the scene and of the camera position can be computed (See figure 1.1).

Both methods do not capture the depth information of a whole scene in an instantaneous way. But they were also not designed for this capability. The laser scanner is used to capture large scenes up to distances of several km in one scan. As the scanned data can contain millions of points, processing the data is time consuming and expensive. The photogrammetric process is also time consuming, suffers from correspondence problems and is dependent of light conditions.

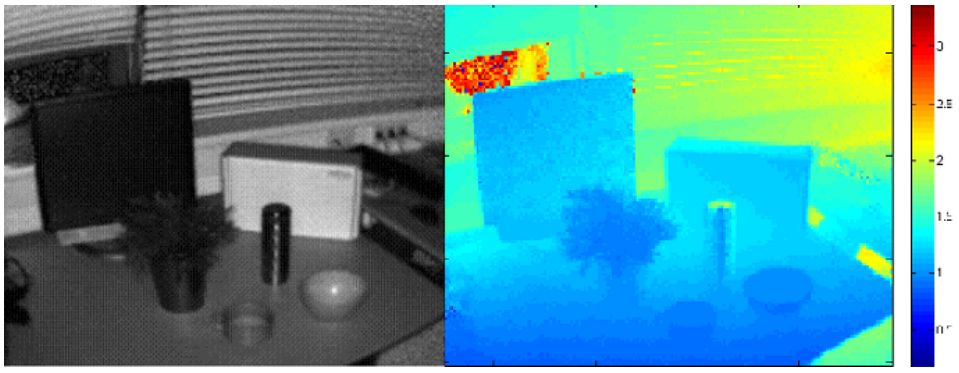


Figure 1.2: Left: Amplitude image. Right: Range image

A range camera is an active system that measures distances within the field of view of the camera instantaneously and it measure the strength of the returned signal, provided in a range and amplitude image with a resolution of 176x144 (See figure 1.2). The range values are the radial distances from the camera optic center, the amplitude values reflects the strength of the returned signal. It acquires images at a relative fast rate of 30 frames per second, such that it can detect real-time movements in the 3D scene. It is designed for small indoor environments (range : 1-5 meters) with less external light sources. The data packages are small sized (1MB-2MB) compared to laser scanned data and photos (Gigabytes).

There are different techniques to determine the distance with range camera. Two most often seen techniques are:

Structured light: The system projects a pattern of light into the scene, by measuring the displacment of the specific pattern in the scene the distances can be computed. For example the X-box Kinect sensor is based on structured light principle (Scharstein and Szeliski, 2003).

Time of flight (ToF): Distance computed based on physical characteristics of light. The range camera SR4000 compute distances by measuring the phase of the returned active light (Lange, 2000). Note that most laser scanner measure directly the time of flight, while only short range systems uses the phase to measure distance.

The focus in this thesis is on acquiring and registering 3D point clouds from real environments with the Swissranger SR4000 ToF camera. Range cameras compact size, high acquisition rates and low cost label create new possibilities for different applications. Current research on range camera data can roughly be divided into two categories:

1. To detect 3D descriptions of moving objects and patterns in real-time: Objects (Hagebeuker, 2007),moving hands (Ren et al., 2011), moving humans (Knoop et al., 2006), facial expressions (Böhme et al., 2009). The camera should then be in a fixed position or it has to track its own position in order to distinguish its own movements from the dynamic scene.

2. To track 3D object surface and patterns in real-time: Reconstruct camera own trajectory, 3D static scene reconstruction. The camera should be moved with care through a static scene.

Currently, many new methods are developed to achieve the tasks mentioned above. This thesis will contribute to advancements in the second category.

In this thesis a new approach is developed to reconstruct a 3D point clouds from the range images of the SR4000, inspired by current approaches and the idea to map an indoor environment real-time.

1.2 Problem statement

The process from data acquisition, data processing to the application, reconstruct one 3D point cloud of the scanned environment is evaluated. In order to fit within the time constraint, this is done with some restrictions and extra focus on certain topics. The problem description, research questions and objectives is described in this section.

The small physical size, fast acquisition rate and small data output of the range camera, make it possible to investigate the whole mapping process:

1. Data acquisition: Scan the object or environment by moving the range camera.
2. Calibration: Correct the systematic errors present in the data.
3. Filtering: Remove points that are considered not use-able.
4. Registration: The process to align two points clouds.

The range camera is moved by hand through the scene, with the following assumptions:

- The scene is static: There are no other moving objects.
- The scene is within the limitations of the cameras performance.

The range and amplitude measurement provided by the range camera are in the form of images. The range image contains for each pixel the radial distance to a physical surface in the field of view. The amplitude image contains for each pixel the strength of the returned signal. The errors in the images are described by systematic and random errors. Random errors have a zero arithmetic mean for repeated measurements, while systematic errors show a deviation, so-called bias from the mean. This implies that the systematic errors can be corrected based on analysis of the repeated measurements in a test set-up. The systematic errors are minimized in the calibration process and a part of the random errors is removed in the filter process. To reconstruct one 3D point cloud from the scanned scene, the range images are registered together in one coordinate system, such that it represent the real scene in 3D points.

The registration problem is not new and discussed since 1992, when a successful algorithm is designed by (Besl and McKay, 1992) to register two 3D point clouds, the Iterative Closest Point (ICP). The purpose is to find for two corresponding point sets the best transformation parameters (rotation and translation) that maps one set of points onto the other. The criteria for the "best" in ICP algorithm is to minimizes the sum of squared distances between the points in an iterative process.

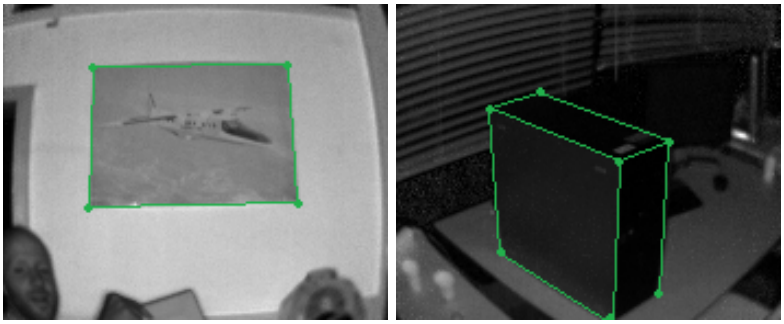
Since the introduction of the range camera, new adapted registration approaches have been developed. The approaches can be divided in three classes by means of their required input:

a The ICP algorithm or variants of it to register the range images. Data input: Range images

b Detect point correspondences by image feature tracking techniques on the subsequent amplitude images. Map the point correspondences to the 3D points in the range image. Use the correspondences to directly solve the transformation parameters for registering the range images. Data input: Amplitude images

c A combination of *a* and *b* or an algorithm that combines the data input range and amplitude image for registration. This is popular for range camera, because it exploits both the available range and amplitude data and covers the weaknesses of *a* and *b*. Data input: Range and amplitude images.

The key issue noticed in method *b* is that image features are 2D in nature, the features are derived from the texture information in the 2D image space. The 2D image is a simplification of the 3D space. Image features are reliable only if the camera does not translate in 3D, but this does not hold when the camera moves. The detected image features that are actually defined by their 3D physical shape like corners of a table or edge of a leaf cannot be tracked correctly (See figure 1.3b). Although there are still 2D features available to track, like an edge of a poster on a wall that is not described by its 3D physical shape (See figure 1.3a).



(a) Poster on wall (b) A computer case

Figure 1.3: 2D feature and 3D feature

Image features reduce the amount of points considered for registration. Direct point correspondences from image features solves the transformation parameters quickly, but it lacks robustness due to the 2D nature of image features. The idea to find 3D geometric features to assist the registration process has emerged. Image features could still be used but only the ones that are not 3D physically related.

In this thesis I have developed a new approach for registration of range images by using geometric features. The novel aspect of the method is to extract geometric features from the range images and use the geometric features to register the range images in an adapted ICP. The geometric features do not provide direct point correspondences, therefore the transformation parameters have to be determined in an iterative way. The method takes as input only range images.

The method takes the range camera specific performance and error characteristics into account. Because the error is of high influence on the registration results, all the error sources have to be handled and addressed. In the geometric feature classification method, camera specific relations derived from a data analysis related to the noise and resolution is used. The method is thereby adaptive to the performance of the range camera.

1.3 Research objectives

The main research question:

Is it possible to do continuous registration of range images based on geometric features with the SR4000 Swiss ranger?

The following sub questions are important related to the mapping process:

1. How to manage the error in range images?
2. What are the geometric features and how to segment them from the range data?
3. How is the novel method performing relative to a current registration approach ?

The new registration approach is developed with the following objectives:

- Able to cope with the range camera specific performance.
- Able to produce robust results given a set conditions.

The main goal of the research is to improve the registration process of range images by a developing a new registration approach based on key elements of existing methods and knowledge from Geomatics. The chosen approach uses geometric features to exploit the spatial information available in the range image, this improves the robustness and achieve reasonable speed compared to existing methods. The method should be adaptive to the range camera performance, by taking the noise behavior and resolution into account.

1.4 Methodology

The geometric feature registration method developed in this thesis is based on key advantages of existing approaches: ICP-based and Image feature-based registration. Extraction of useful geometric features from the range data reduces the amount of input data for the ICP process. The geometric features are 3D (viewpoint independent) and there therefore invariant to 3D movement of the camera. The main steps taken to design the method are as follow:

1. First a literature study of existing calibration, processing to registration approaches with range camera is performed to gain knowledge of state of art techniques. Leading to the conclusion that the chosen approach "Registration of range images with geometric features in an adapted ICP" is new and promosing.

2. The range data is analyzed for its resolution in 3D by a test setup with a 3D Siemens Star, in which the minimal surface required for the camera to measure a correct distance is determined. The systematic errors lens distortion, distance-related error and fixed noise pattern is corrected based on a correction model. The lens distortion is corrected by a standard Photogrammetric calibration. The correction model for distance-related error and fixed noise pattern error is derived from a test setup where repeated measurements of a white wall from variable distances is performed. Remaining random errors: floating edges are detected and removed by existing techniques and noise is minimized by median filtering.

3. Geometric feature is defined as physical edges visible in the range image such as an edge of a table. They are viewpoint independent, as these features do not change in location when viewed from a different perspective. A classification method is designed to segment the geometric feature points in the range image taken into account the result of data analysis.

4. Use the set of geometric feature points in an adapted ICP to find the optimal transformation parameters between two point clouds (range images). The main difference from normal ICP is the addition of a geometric feature attribute in the correspondince finding process. Parameter access to the method allow the possibility to fine-tune the registration.

5. From the registration results the performance of the method is evaluated. The evaluation leads to suggestions for future improvements.

1.5 Outline of the thesis

The thesis is organized in six chapters. Following the introduction, in chapter 2 a literature review is presented with state-of-art theory discussed for the range camera, the calibration and filter process and the existing registration approaches. In Chapter 3 results of data analysis for two test set-up is given and the calibration and filtering process of the SR4000 is presented. In Chapter 4 the geometric features are defined and the classification method to segment geometric feature points is described. In Chapter 5 the geometric feature points are used in an adapted ICP and the registrations results are evaluated. In Chapter 6 conclusions are drawn from the registration results, limitations are identified, the advantages and disadvantages of the method is discussed future works is proposed.

Chapter 2

Theory and Literature review

In this chapter a literature review is presented focused on the state-of-the-art achievements on "Continuous 3D mapping with range cameras". The general process involves data acquisition, calibration, filtering and application. The goal of this chapter is to get familiar with the essential background and to identify the problems arising with the 3D mapping process.

This chapter leads to the conclusions that: 1. More spatial information in range images can be exploited in the mapping process, 2. The main unsolved errors are random in nature, 3. Current registration approaches with a single range camera are not robust enough for consistent real-time 3D mapping. Remarking that the title of the thesis is "Continuous registration of range images with geometric features.", this chapter also underlines the innovation of the thesis w.r.t existing approaches.

In Figure 2.1 the general mapping process is illustrated. First the theoretical background of the instrument for data acquisition is explained, then a section about possible data errors and their handling to improve data quality is presented. Then different existing methods concerning 3D point cloud registration will be presented. Finalizing this chapter by concluding the problems of state currently existing 3D mapping methods. An overview of the state of technology is presented in this chapter

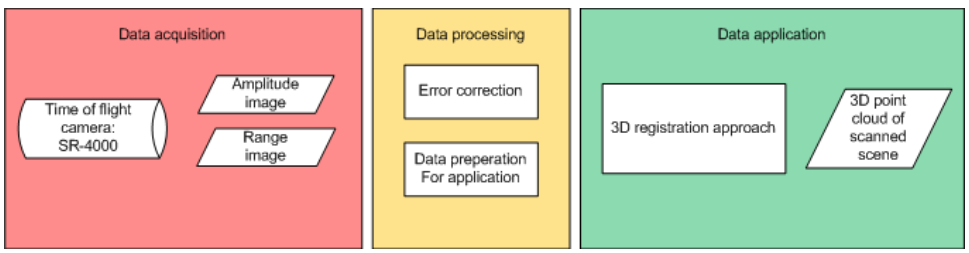


Figure 2.1: General Mapping Flow Diagram

2.1 Range camera

The category of sensors that produce images with distance information stored in each pixel are generally called range cameras. The depth information in the field of view of the camera is acquired instantaneous in a stream of images. The acquisition speed is multiple frames per second, implying that real-time 3D information of the scene is retrieved. Different techniques can be used to determine the distance for each pixel, the type of range camera is usually named after this technique.

Common used measuring principles are:

- Time-of-flight, based on physical property of light (discussed next section)
- Structured Light, based on stereo vision (Scharstein and Szeliski, 2003)

The type of range camera used in the thesis is a Time-of-flight (ToF) camera, the Swissranger SR4000. The underlying technique, data properties and functionality the SR4000 is described next.

2.1.1 Swissranger SR4000

The Swissranger SR4000 is a ToF camera (See Figure 2.2). The main characteristics of the SR4000 are, a field of view (FOV) of 43.6° x 34.6°, image resolution of 176x144, operating range is 1-5 meter and acquisition speed up to 30 f/s. The strong points of the range camera are the speed and the instantaneous 3D scene capturing ability. Complete specification is described in the datasheet (SR4000-DataSheet, 2011).

Figure 2.2: The Swissranger SR4000

The instantaneous field of view (IFOV) of a single detector (pixel) is not given. There are two options: 1. The IFOV would be different for each pixel, if the optic is designed to compensate slant view distortion, meaning the pixels are all equal in sizes. 2. The IFOV is assumed to be regular, the same for all pixels if the optic is not designed to compensate slant view distortion. The pixel size is then dependent of the pixel coordinate.

Time of flight principle

The distance measurement is relying on the physical properties of light. The phase difference of the modulated light emitted from the sensor to the surface objects in the field of view and back to the sensor is measured. With the speed of light given in standard atmospheric conditions together with the measured

phase, the distance to the surface object can be derived from the phase difference. See Figure 2.3 for a schematic overview of the Time of Flight principle. It should be noted that the principle "Time of flight" is different for laser altimeter.

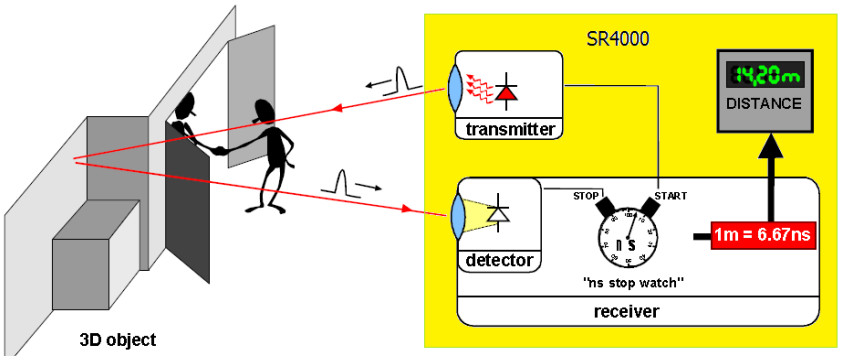


Figure 2.3: Time of flight principle (Lange, 2000)

With an array of near-infrared light sources (LEDS) the SR4000 emits a signal modulated in amplitude, into the field of view. After the signal is reflected back from a surface, the phase delay between the emitted and returned signal can be resolved by looking at the modulation. The phase difference is measured in each pixel within the array image sensor (CCD/CMOS).

The distance is determined as a fraction of one full cycle of the modulated signal (see relation 2.1), where c is the speed of light and f is the modulation frequency. The operating range is dependent of the modulation frequency, this model operates at a modulation frequency of f=30 MHZ, assuming the speed of light c= 300000000 [m/s], the operating range is then about D=5 meter.

$$D = \frac{c}{2f} \tag{2.1}$$

Demodulation is achieved by correlating the original modulation signal $g(t)$ with the returned signal $s(t)$, this procedure is called cross correlation. With t the real time and τ the delayed time with respect to t. The cross correlation function is shown in equation 2.2 as $c(t)$ (Lange, 2000).

$$c(\tau) = s(t) \otimes g(t) = \lim_{T \to \infty} \frac{1}{T} \int_{-\frac{T}{2}}^{-\frac{T}{2}} s(t) \cdot g(t+\tau) dt \tag{2.2}$$

The returned signal $s(t)$ is composed of an offset signal(background noise) and a sinusoidal modulated signal. The original emitted signal $g(t)$ is sinusoidal as well. The phase difference is ϕ. A is the modulation ampltitude. The equations are given below:

$$\begin{aligned} s(t) &= 1 + A \cdot cos(\omega \cdot t - \phi) \\ g(t) &= cos(\omega \cdot t) \end{aligned} \tag{2.3}$$

The SR4000 evaluates the cross correlation function in 4 fixated phase delays $w\tau_0 = 0°$,$w\tau_1 = 90°$, $w\tau_2 = 180°$ and $w\tau_3 = 270°$. A schematic of the signal is illustrated in Figure 2.4 with the relations A, B and ϕ.

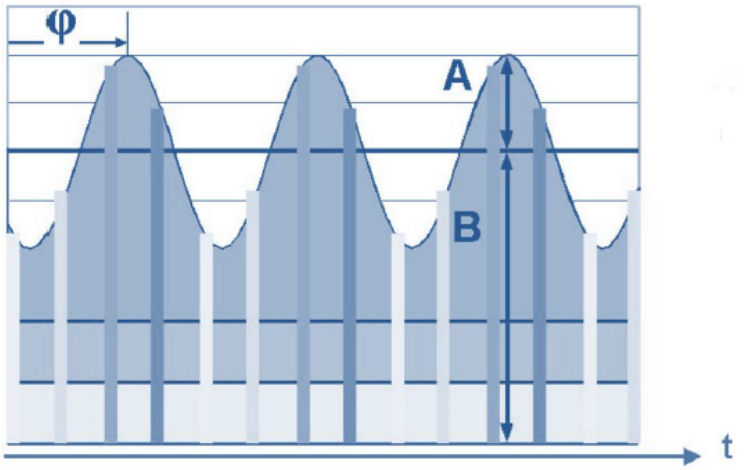


Figure 2.4: Time of flight sampling of returned signal (SR4000Manual, 2011)

$$\begin{aligned} \phi &= atan\frac{c(\tau_3) - c(\tau_1)}{c(\tau_0) - c(\tau_2)} \\ B &= \frac{c(\tau_0) + c(\tau_1) + c(\tau_2) + c(\tau_3)}{4} \\ A &= \frac{\sqrt{[C(\tau_3) - C(\tau_1)]^2 + [C(\tau_0) - C(\tau_2)]^2}}{2} \end{aligned} \tag{2.4}$$

Signal B is the mean of the total incoming light on the sensor, this includes the background noise and modulated signal. The B signal is used to plot the intensity image. By high reflective material the offset B indicates whether or not saturation occurs in the solid state image sensor. The amplitude A is a direct measure of the accuracy. The phase delay ϕ is directly proportional to the target distance D. The distance is computed as follow:

$$D = L \cdot \frac{\phi}{2 \cdot \pi}, with\ L = \frac{c}{2 \cdot f_m} \tag{2.5}$$

The L is the non-ambigious operating range of the camera with c the speed of light and f_m the modulation frequency. This method is known as the four steps method, Lange (2000).

Data properties

The SR4000 provides continuous images up to 30 frames per seconds. The following output can be retrieved:

1. Amplitude image: Strength of returned signal per pixel.

2. Range image: Spherical radial distance per pixel.

3. XYZ-format: The calibrated spherical distances transformed into Cartesian coordinates. (only provided by software)

4. Confidence map: Indication of reliability of the measurement for each pixel. (only provided by software)

In the context of this thesis we use the manufacturer provided calibration as a reference to compare our custom calibration in Chapter 3.

A summary of the general data specification is given in table 2.1 below:

Absolute accuracy	+/- 10 mm
Repeatability central pixels	4 to 7 mm
Repeatability outer pixels	7 to 14 mm

Table 2.1: General data specification

The absolute accuracy is the difference between the mean value and the distribution of the real value, also called the standard deviation or the Root Mean Squared error (RMS). This value describes the precision of the range camera. The repeatability is the standard deviation of the noise region present in the data, the noise increases with increasing distance from the image center.

Integration time

The SR4000 allows access to the integration time, this can enhance the accuracy performance of the SR4000. The integration time has influence on 3 aspects, a compromise of the three is eventually needed for a certain setting (See table 2.2).

Acquisition speed	The frame rate at which the images are acquired is inversely proportional to the integration time. Smaller integration time means: - Higher frame rate. - Less movement artifacts from moving objects. - minimizing the time budget needed for the overall measurement.
Measurement quality	The noise of the measurement is inversely proportional to the integration time: Longer integration time allows capturing a larger amount of reflected light, which in turn reduces noise.
Type of measured object	- The amount of light reflected from the measured objects is proportional to their reflectivity (in the infrared domain). Thus to reach the same measurement quality (repeatability), a higher integration time needs to be set for low reflective objects than for high reflective objects. - the light intensity decreases with the scare of the traveled distance. Farther away, a higher integration time is needed.

Table 2.2: The influence of integration time on measurement quality

The integration time is left on default mode, this means an optimal integration time is set by the driver software of the camera.

Filters from the camera

Filters are available from the camera's driver. They help reducing errors to a certain extend. The following default filters are available:

1. Median filter
2. Convert gray mode: Compensate for the light intensity send out by the sensor itself.
3. Adaptive Neighborhood Filter: Hardware-implemented noise filter. It reduces noise while preserving detail such as edges or small structures.

In this thesis only the default filters are used in the data acquisition process.

2.2 Error and calibration

The range data acquired with range cameras contains error. The difference between a measured value and the actual value is a quantification of the error. The error sources can be group into two main categories:

systematic errors they are predictable and can be corrected based on a relation that is derived from repeated measurements in a fixed test set-up. They do not have a null arithmetic mean, therefore have a constant bias with respect to the mean value.

non-systematic error also called **random errors**. The random errors on contrary are unpredictable, scattered around the true value. They have a null arithmetic mean.

A brief summary of the errors present in the data and their possible correction is given based on various related publications. The focus of the thesis is not on improving the calibration process, but to understand the errors present in the range data. Only the calibration methods that are directly applicable based on the analysis done in chapter 3 is used in the thesis.

2.2.1 Systematic errors

Systematic errors are that have a fixed behavior in measurements repeated in time. The measured distances show a constant bias with the true distance. The bias can be modeled to correct the error.

Lens distortion
Lens distortion is caused by the physical interaction of light through a lens. This is characterize by a radial distortion and a tangential distortion effect. In Figure 2.5(left) the red line is curved by the lens distortion, but in reality it is a straight line.

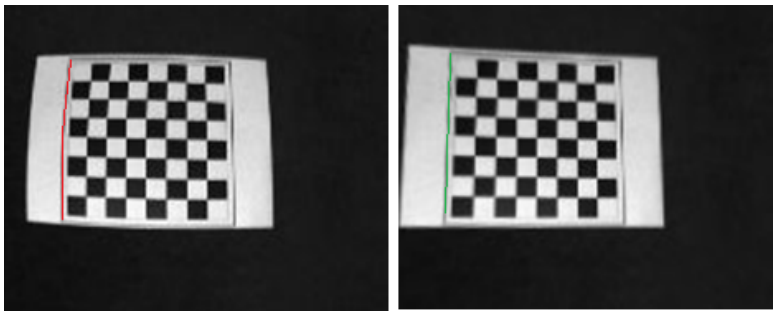
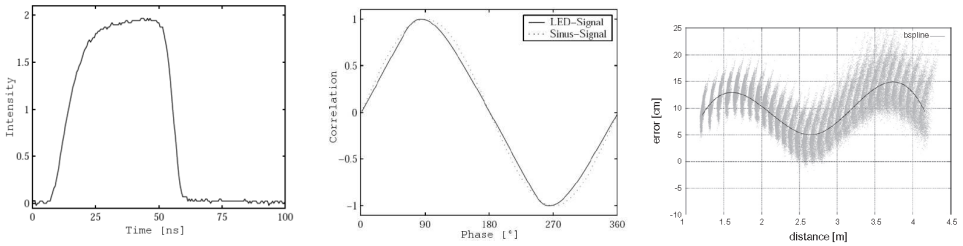


Figure 2.5: Lens distortion illustration: Left distorted. Right undistorted.

Wiggling error, circular distance error, distance-related error
In theory the signal for modulation interferometer is assumed sinusoidal, but if in reality the signal is not sinusoidal (See Figure 2.6. This will result in systematic bias in the computed phase delay and distance.



(a) Left: Returned sinusoidal signal. Right: Difference with ideal sinus signal (Fuchs and May, 2007)

(b) Distance deviation (Lindner et al., 2010)

Figure 2.6: Distance-related error

Amplitude related error

Amplitude related error is caused by nonlinearities of the pixel's electronic components. The arrival of different numbers of photons at a constant distance results in different distance measurements (Lindner and Kolb, 2007) (See Figure2.7). The accuracy of the distance measurement relies on the amount of returned Near Infrared Red active light. This is dependent of the surface IR-reflectivity and surface orientation. When there is a reasonable signal to noise ratio, the distance measurement can only be influenced if the signal shape is distorted. Assume the reflectivity has a systematic influence on the shape of the signal, such that it produce biased distance measurement. Note the roughness of the surface also shapes the returning signal.

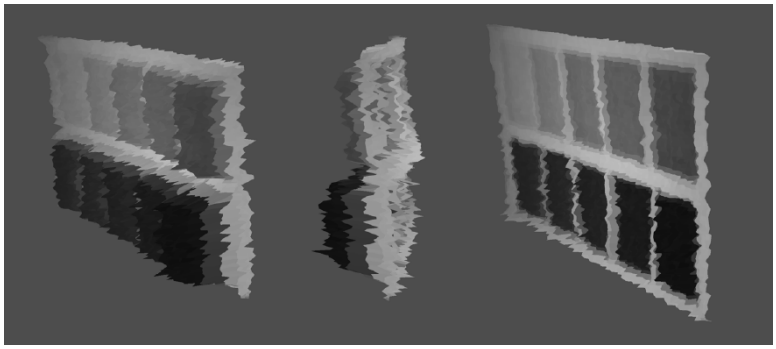


Figure 2.7: Distance deviation due to variable reflectivity of surfaces in the same plane. On the right side a calibration method is applied in (Lindner and Kolb, 2007) to reduce the effect.

Inhomogeneous image illumination

Depends on the configuration of the LEDS, the optics, and the cameras field of view. The illumination decreases in radial direction from the center as a result of two effects (See Figure 2.8):

- The inhomogeneous scene illumination through the sensor's active emitters.
- The vignette effects induced by the sensors optics.

Note that this error contains contribution of systematic and nonsystematic errors in different proportion.

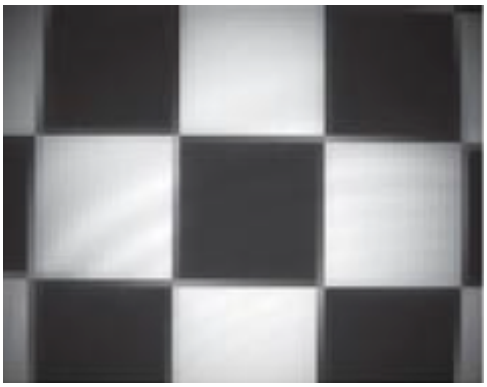


Figure 2.8: Decreasing illumination in radial direction from the center (May et al., 2009)

Fixed pattern noise
Fixed pattern noise is a unique error depending of hardware properties (May et al., 2009):

- Each pixel has an individual characteristic due to an imperfect manufacturing process and different material properties in each CMOS gate, which yields pixel-individual fixed measurement offset.
- The triggering of each pixel depends on the position on the chip and its distance to the signal generator.

The systematic errors can be corrected by applying a correction model. The error function is determined from a series of repeated measurements. Such an error function can correct for several errors.

2.2.2 Random errors

The errors that do not have a fixed behavior in repeated measurements. The random errors cannot be modeled.

Noise
The noise inherent to the hardware limits the performance of ToF camera. Lange (2000) subdivided this noise in three classes:

1. Photon shot noise: Describes the statistical Poisson-distributed nature of the arrival process of photons and the generation process of electron-hole pairs.
2. Photocharge conversion noise: Includes all noise sources(reset noise, flicker noise, thermal noise, shot noise of dark current) that disturb the "optical" information in the process chain of converting the optically generated electron hole pairs into an analogous output signal. They all increase with temperature.

3. Quantization noise: Caused in the analog-to-digital conversion.

Figure 2.9: Noise illustration

Multipath
A signal not directly reflected back to the sensor give a false distance measurement. Typically this is caused by multiple subsequent reflections from a concave area, e.g a corner(See Figure 2.10).

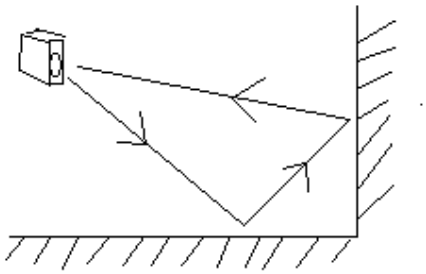


Figure 2.10: Multipath illustration

Mixed pixel or floating pixel
The footprint of the reflected signal is an area consisting of surface at different distances (See Figure 2.11).

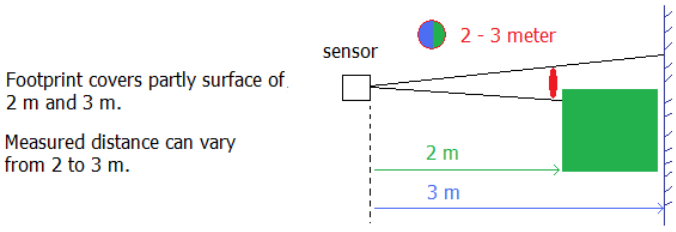


Figure 2.11: Floating or mixed pixel illustration

Saturated pixel
The amplitude of the signal is too high or low caused by either:

1. The objects is out of range requirement (1- 5 meters)
2. Too much incoming light caused by external light sources
3. High reflective property of the objects surface.

Internal scattering
Artifacts created by multiple internal reflection of the received signal occurring between the camera lens and the image sensor. This occurs when the weak signal from far objects is affected by the strong scattering signal from foreground objects.

Mixed pixel and saturated pixels are caused under specific conditions. Although it cannot be corrected globally, but it can be removed given the conditions the error is detectable.

Also explained by Lange (2000) the background illumination(external light sources) influences the accuracy. Optical filter that only transmit the spectrum of modulated light is applied to reduce the background illumination. The closer the distance to the object, the more "active" emitted light is returning to the sensor, thus increasing the accuracy.

2.2.3 Calibration techniques

In Zhang (2000) and in Heikkila and Silven (1997) a calibration procedure for determining lens distortion parameters and interior orientation parameters is explained. The calibration test set-up only requires a camera and a grid patterned surface like a chessboard with known geometry. The procedure is always a closed form solution for determining the interior and distortion parameters of the camera followed by a nonlinear refinement.

In Kahlmann et al. (2006) the accuracy of the distance measurements is investigated. In the test-setup accurate distance tracking devices were used to obtain a ground truth. The influence of different integration time settings and different reflectivity was tested. The final distance correction model is based on a look-up table for different integration times. The bias to the ground truth was estimated by linear and cosine functions, but the results were not better than the look-up table. The look-up table was based on measurements of one central pixel. In an ideal case the distance calibration should be done for each individual pixel, as each pixel can be seen as an individual sensor. An alternative way is to approximate a Fixed Pattern Noise matrix by measuring a flat wall, but take into account that this is done for one fixed integration time.

In Lindner and Kolb (2006) a distance calibration for the wiggling error and fixed pattern noise developed. By measuring a flat surface with a grid pattern (chessboard) two main calibration steps were taken. First a global distance adjustment for the entire image is applied by using a cubic B-spline fit to account

for the global error, secondly a per-pixel distance calibration is applied after the global correction. Further attempt has been made in Lindner et al. (2008) to improve the distance accuracy by enhanced demodulation schemes. But the advantages were not clear compared to the constant or linear per pixel adjustment of the original demodulation scheme. In Lindner et al. (2010) the amplitude related error is investigated more in depth. The calibration procedure involves planar chessboard with different levels of reflectivity. The approach provides a significant improvement in correction of the amplitude related error.

In the publication Fuchs and May (2007) and Fuchs and Hirzinger (2008) a calibration procedure using distance and amplitude measurements is developed. Done in a test set-up with an external positioning system (robot-arm). First a number of amplitude and depth images of a chessboard is taken for intrinsic camera calibration. Second the images are used to create a depth correction model taken into account the wiggling error, amplitude error and fixed pattern noise. This method has shown to provide significant data improvement.

In recent studies different calibration approaches have been suggested to correct the systematic errors. In these approaches the correction models are always dependent of quantities that can be measured or computed.

It can be concluded that for most of the systematic errors, a calibration is available. To correct the errors, a measurement set-up is configured. In table 2.3 below a summary of the errors and the possible corrections is given.

Systematic error	**Correction model**
Lens distortion	Correction: Calibrated by a photogrammetric process, by which a set of camera parameters that describe the mapping between 3D object coordinates and 2D image coordinates are determined. There are different closed-form solutions and nonlinear minimization solutions for the calibration process. In the thesis a calibration process is used as described in Heikkila and Silven (1997).
Distance related error	Correction: Performing repeated measurements of a white plane from different distances to the plane to determine the deviations to a best fitting plane. Model the distance related error.
Amplitude related error	Correction: Performing repeated measurements of a plane with different reflectivity properties. Model the Amplitude related error.
Inhomogeneous lightning	Correction: This type of error cannot be modeled seperately from the other errors in the repeated measurements, but can be seen in the distance related error.
Fixed pattern noise	Correction: Performing repeated measurements of a white plane. The error is included per pixel in the deviations to a best fitting plane.
Random error	**Correction**
Noise	Correction: All can be reduced or eliminated by signal processing techniques or cooling except the photon shot noise. Photon noise cannot be suppressed and is the theoretical limitation of all photo detectors. Lange (2000)
Multipath	Cannot be corrected
Mixed pixel	Can be detected but not corrected
Saturated pixel	Can be detected but not corrected
Internal scattering	Cannot be corrected

Table 2.3: Summary of errors

Implemented calibration techniques
In the thesis the systematic errors caused by lens distortion, distance-related, fixed pattern noise and the non-systematic "mixed pixel" error and noise is handled.

The systematic error can be corrected based on an correction model (discussed in chapter 3). The lens distortion is corrected with a photogrammetric calibration process. For the distance related error and fixed pattern noise error a measurement test set-up is created to model the error and create a correction model.

The non-systematic error cannot be corrected, but can be detected and minimized. Removing noise involves a certain smoothing procedure applied to the

dataset. The most reliable method mentioned in literature is the median filter. The median filter exhibits the benefit to be edge preserving, which is important in geometric 3D data. The mixed pixels can be detected based on local angle technique (May et al., 2009) or local point distribution technique (Rusu and Cousins, 2011) and removed.

2.3 Registration approaches

In this section the registration problem is explained and the literature is reviewed. State of the art techniques are discussed and a novel approach is proposed "Registration of range images based on geometric features".

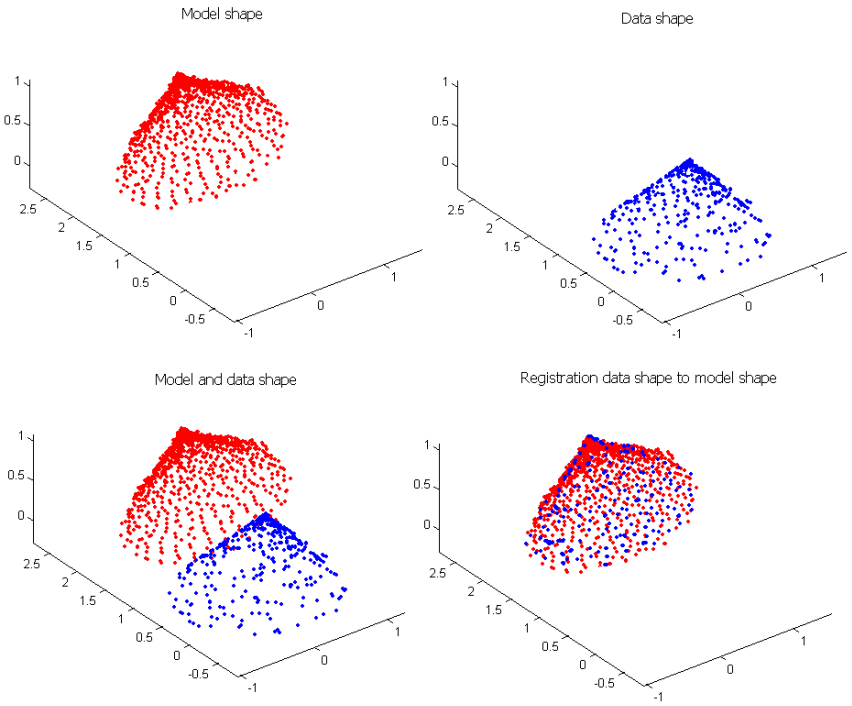


Figure 2.12: Registration problem

The registration problem is described as follow (See Figure 2.12): There are two 3D point sets describing tro overlapping parts of the same surface. Each set of points is described in its own (sensor) coordinate system. The objective is to transform (rotation and translation) one set of points into the coordinate system of the other set of points, such that the combined set of points describes the scene a signle coordinate system in the same coordinate system. The two sets of points are not exactly corresponding, meaning the points are distributed differently, but describe the same surface area. If the 2 sets of points were exactly corresponding, the absolute distance between the two sets of points after

transformation would be zero. In our case the two set of points are scanned by a sensor system, the point density is determined by the resolution of the sensor system.

The main problems in registration are the following:

- How to determine the transformation?
- How to verify the transformation is correct?
- How to determine the points representing the same surface, if both point clouds contain some number of sample points on different surfaces?

The registration problem is a long discussed problem in computer vision and mathematics, key developments found in literature are described in the following sections in chronological order.

2.3.1 Iterative closes point: ICP

In Besl and McKay (1992), a well-known general algorithm to register two 3D shapes is developed. To register the data shape with N points to a model shape with X primitives, the optimal rotation and translation is estimated by minimizing the distance between the shapes. Assume the data shape is measured in a sensor coordinate system. With X primitives is meant that the model shape can be a set of points, lines, curves, triangles, parametric surfaces described in a model coordinate system.

The ICP algorithm is stated as:

a *Compute the closest point in the model for each point in the data shape.*
b *Using the corresponding closest points, compute the transformation parameters for registration using Singular Value Decomposition (SVD) or Quaternion's.*
c *Apply the registration, transform the data shape using the transformation parameters*
d *Do the steps a-c again with the latest transformed data shape, stop the iteration when the change in mean squared distance between model and data shape is below a chosen threshold.*

The shortcoming of the standard ICP algorithm are:

- Sensitive to outliers.
- Time consuming because all points considered
- Initial alignment of the two shape is needed.
- The data shape is assumed to overlap completely with the model shape.

- It may not converge to the desired global minimum, but find a non-optimal local minimum.
- Does not work when there is no orientable surface in the scene, for example only a flat wall does not work.

2.3.2 Efficient ICP

After the ICP algorithm of Besl and McKay (1992), many variants have been proposed to improve different phases of the ICP algorithm. Rusinkiewicz and Levoy (2001) give a review of different variants and evaluated on their running time to achieve an correct registration (alignment). From this comparison the best variants for the following six stages of the ICP algorithm is described here:

1. **Selection** of some point set of both model and data set:
   Random sampling for selecting point sets
2. **Matching** the points:
   Projection-based algorithm to generate point correspondences
3. **Weighting** the corresponding point pairs:
   Constant weighting of the point pairs
4. **Rejection** certain pair of points:
   Distance threshold for rejecting pairs
5. Assign **error metric** based on point pairs:
   Point-to-plane-error metric
6. **Minimize** the error metric:
   Standard "Select-match-minimize ICP iteration"

The combination of the best variants of ICP has delivered a refined version of ICP where the following problems are minimized:

- Non-overlapping area is minimized by distance threshold.
- Not all points are considered, results in data reduction.
- Point-to-plane metric is evaluated to converge fastest.

2.3.3 Registration approaches with range camera

Next to the 3D registration problem, 2D image registration is researched thoroughly in computer vision to compute 3D models from several 2D images. For the large 3D point clouds acquired from aerial laser scanner or terrestrial laser scanner, the registration process is done by using control points (known coordinates) in the scanned area. The ICP is applied afterwards to refine the registration, this is often a manual process and the duration of the process was not an issue. The use of range camera for real-time mapping an environment

introduced some specific challenges to the registration problem.

The challenges can be summarized in the following points:

1. To enable real-time mapping, the registration approach should perform at real-time speed.
2. The registration approach should be robust and fully automatic.
3. Having no additional sensors or control points, the position of the scanner(camera) is estimated solely based on the range images.

Since the introduction of the range camera novel registration approaches specific for real-time mapping have emerged. In robot navigation, Simultaneous Localization And Mapping (SLAM) has become an important topic. SLAM enables a robot to map the 3D environment and to determine the robot (camera) position in real-time with respect to the environment (odometry). In Davison (2003) it is described that a robot should be able to navigate autonomously with a range camera in an indoor environment without additional sensors. A summary of the recent approaches developed for 3D mapping using only a ToF camera is found in May et al. (2009). Two new approaches additional to the classical ICP method are discussed: The image feature based registration and the range-amplitude combined approach.

Image feature based registration
Point correspondences are found based on image feature tracking between subsequent amplitude image. Two well-known image feature tracking techniques can be used: Scale Invariant Feature Tracker (SIFT) and the Kanade-Lucas-Tomasi Feature Tracker KLT. The point correspondences found in image space are then applied to the 3D points from range image, taking into account that those 3D points may be mixed pixels or outliers. The 6 transformation parameters can be directly calculated, if at least three correct corresponding 3D point pairs are available.

The image feature approach has the following key points:

- It is less computationally heavy than the ICP approach.
- Image feature tracking techniques depend on the amplitude image, which is sensitive to external light sources and reflectivity of surfaces.
- Does not work when there is no significant texture information.
- Can only give reliable 2D features. 2D images are projections of the 3D scene, because of that KLT and SIFT are dependent on the camera orientation. 3D features are not reliably tracked.

Range and Amplitude image combined registration approach
In Malis (2007) is a Efficient Second-order Minimization(ESM) method is described to find an optimal transformation between two subsequent data sets. It is implemented to utilize both amplitude and range images.

In Henry et al. (2010) a combined approach using image features (SIFT) and

3D shape alignment (ICP) is used for the registration.

It is shown in both cases that by applying both range and amplitude images, the robustness of the registration approach is increased.

The three main registration approaches are summarized in table 2.4 below:

ICP-based	High computational cost. Very sensitive to non overlapping parts, that has highly variable depth structure. Shows most accurate registration results.
Image Feature based	Low computational cost. No large overlap between images is required, but distinct image features. Registration result is of lower in accuracy compared to ICP approach, because it is based on a small subset of points (Image features). The registration is likely to be distorted by depth errors. Result depends on amount of texture in amplitude, which is relative low in range camera.
Combined approach	Good registration, but is less accurate as the ICP approach, mainly because of sparse and possible wrong tracked 2D features.

Table 2.4: Table

The three registration methods are a good representation of the current available approaches. By combining the ICP and Image Feature based methods a more robust and efficient approach can be developed. Now that the majority of registration approaches has been discussed, the following concluding points form my motivation to develop a novel approach in this thesis:

1. Accuracy: ICP-based method is the best algorithm in terms of accurate transformation. It depends on the point distribution and amount of noise.

2. Robustness: Combining ICP-based and Image Feature-based methods gives reasonable robustness. It fails in cases when the image feature tracking is not able to find reliable 2D features.

3. Running time: Image Feature-based methods is fast, because of the reduced data set of image features and the point correspondence information.

From these points the following idea emerged:
The ignored problem in using image features techniques for the registration problem is that image features are 2D. The registration problem is 3D but image features are 2D in nature. More strictly saying 2D features stays the same in 2D space, and the 2D image is a projection of the 3D scene. Meaning the 2D image features can be tracked correctly, if the camera is only rotating and not changing in position (fixed 3D coordinates). The image features cannot

be tracked correctly if the camera is translating in 3D, because of the orientation of 3D physical objects changes with camera perspective. The problem can be ignored in image application, because it is not a 3D problem. But in the registration process it cannot be ignored. Currently the false 2D features are filtered by statistical methods combining the knowledge of the corresponding 3D coordinates.

It is shown in literature that image features techniques improved the speed and robustness of the general registration process. Having discussed the ignored problem of using 2D features in the 3D registration problem, the idea to use 3D features to be detected from the range image has emerged. The chosen approach developed in this thesis is based on the idea to detect 3D geometric features in the 3D point cloud from the range image, which is 3D invariant. Using these geometric features in an adapted ICP will improve the overall registration process by maintaining a speed similar to the combined ICP-ImageFeature approach but with increased robustness.

Geometric feature detection techniques
In Lo and Siebert (2009) a method 2.5D SIFT is developed based on the image feature tracking techniques Scale Invariant Feature Transform (SIFT). It finds geometric features from range images in similar way as SIFT. Concluding that 2.5D SIFT produce more reliable features matches than SIFT, but the principle is based on SIFT, which is not designed to deal with 3D data.

In Novatnack and Nishino (2008) a method to find scale dependent geometric features is described. It gives a scale dependent/invariant local 3D shape descriptor to the range image. For the registration the correspondences is determined based on the similarity of the average angle between corresponding normals. RanSAC is used to determine the best transformation that maximizes the area of overlap between two 3D point sets. Random sampling is performed from the coarsest scale to the finest scale, because in each scale different geometric features can be found. With condition that both range images have the same scale, only one successful case was demonstrated.

Attempts to classify 3D geometric features in range images have been made. But none where designed with the goal "Real-time mapping with the range camera". In this thesis a new method to do the registration of range images with geometric features is designed. The method is described in chapter 4.

2.4 Summary and conclusion

Summarizing the key points of the consulted literature related to "Real-time 3D mapping with range cameras", with topics related to the data acquisition, calibration, filtering and registration.

Calibration and filtering process
In the calibration process the errors are identified and corrected. The errors are systematic or random of nature.

- Systematic errors in the range data can be minimized by modeling the error.
- Non-systematic error cannot be corrected by modeling the error as they appear 'random'. However some can be detected locally by measurable quantities.

The systematic error sources have to be corrected to give more reliable measurements. In chapter 3 a test set-up is made to correct the systematic errors: distance-related error and fixed noise pattern error. The non-systematic errors noise is minimized by median filtering and the mixed pixel is detected and removed based on local angle technique (May et al., 2009) and local point distribution technique (Rusu and Cousins, 2011).

Registration process
In the registration process the 3D point clouds from range images are subsequently mapped over each other, resulting in one 3D point cloud such that it represents the scanned scene.

- The approach is using a combination of an adapted ICP and image feature correspondence tracking is to be most robust to data errors and efficient in time.
- The image feature tracking detects 2D features that may be 3D in nature. This results in wrong feature correspondences.
- 3D geometric features are not yet used in the registration process, but have the potential to improve the process.

In efficient ICP methods all 3D points are considered equal in importance and used to find the correct transformation parameters in an iterative way. The image feature tracking technique needs only 3 point correspondences to solve the transformation parameters, if the 3 points are correct. The combination of image feature tracking and ICP as an approach improves the robustness and speed of the registration. Image features are high level information entities providing direct point correspondences in the 2D image space, but they are not able to distinguish a 3D feature from a 2D feature. This give rise to the idea to use 3D geometric features for improving the general registration problem with range camera's. The 3D geometric feature points are segmented from the 3D point cloud of the range image.

Geometric feature points do not give unique point correspondences between subsequent images like in the Image-feature tracking approach. Therefore it uses an adapted ICP method, to determine the optimal transformation from the segmented subset of 3D point clouds, given by the 3D geometric image features

Chapter 3

Calibration and 3D resolution analysis

In this chapter errors specific to the swissranger SR-4000 are investigated. As explained in the literature review in chapter 2, the systematic errors can be corrected. The errors "lens-distortion", "distance-related" and "fixed noise pattern" are corrected in this chapter. The errors are contained in the distance values, measured for each pixel in the range image. By modeling the deviation between the acquired distance and an assumed 'true' distance, a correction model is created for each systematic error. The systematic errors are modeled in two measurement experiments.

First a common photogrammetric calibration process is applied to correct the lens distortion. This is done in a test-setup by taking images of a chessboard from different perspectives. Second a test-setup is made to correct the systematic distance-related and fixed noise pattern error in the measured distances of the range image. Repeated measurements of a flat wall are taken from different distances to the wall. A correction model for the systematic errors is made based on the analysis of the measurements.

Non-systematic errors cannot be modeled, but they can be minimized by median filtering. Mixed pixels can be detected and removed by local angular filtering.

The 3D radial resolution analysis is to measure the ability of the range camera to give correct distance measurements in relation to the object surface size. This is necessary because the 2D standard resolution does not say anything about depth resolution. In the 2.5D range image the Z-value may be incorrect for example at a mixed pixel. A wrong distance that occurs at the edges, which can be detected and removed. It is important to know the performance of the range camera in the 3D dimensions: At what distance and at what size of object surface area can it measure a correct distance value. A third test-setup is made to measure a cut-out Siemens star at different distances. Analysis of the measurements described the "minimal detectable feature size" as function of the distance.

3.1 Lens distortion correction

Lens distortion is caused by the physical interaction of light through a medium in the shape of a lens. There are mainly two effects:

1. The circular effects are caused by the circular shape of a lens. In Figure 3.1 the lens effect is clearly visible.
2. "de-centering effect" is caused by manufacturing error, where the exact mounting of the lens is not at the center of the image.

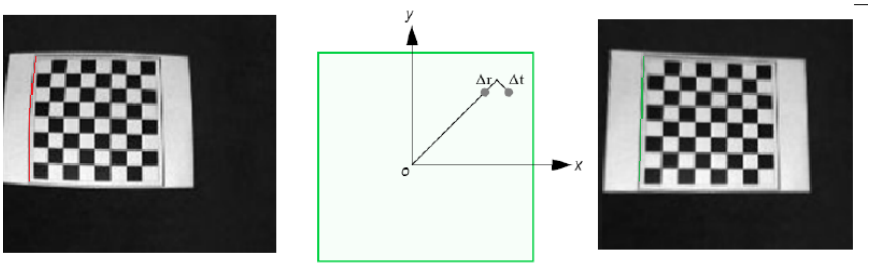


Figure 3.1: Left:Lens distorted. Right: Lens distortion corrected

The lens distortions is corrected by a photogrammetric calibration process, in which the 3D object space is related to the 2D image space taking into account the lens distortion (See figure 3.2).

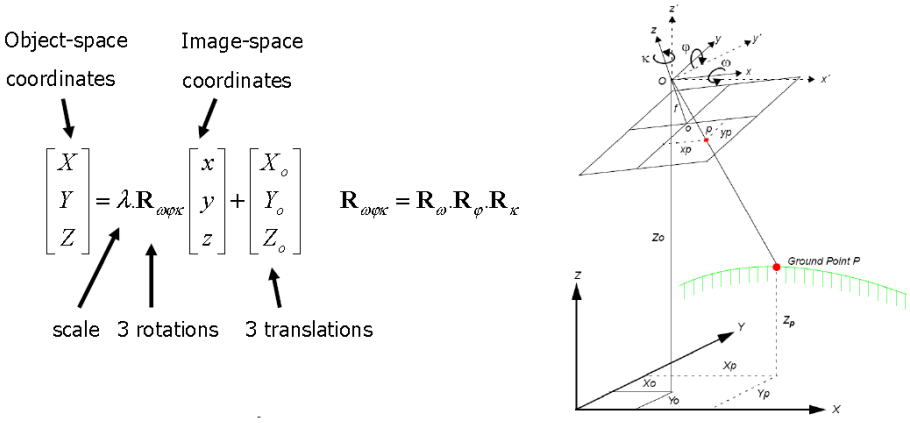


Figure 3.2: Object space and image space (Khoshelham, 2011)

The relation between the 3D coordinates of object space and the 2D coordinates in the image space is described in the collinearity equation (Khoshelham, 2011).

$$
\begin{aligned}
x - x_0 + \Delta x &= -f \times \frac{m_{11}(X - X_o) + m_{12}(Y - Y_o) + m_{13}(Z - Z_o)}{m_{31}(X - X_o) + m_{32}(Y - Y_o) + m_{33}(Z - Z_o)} \\
y - y_0 + \Delta y &= -f \times \frac{m_{21}(X - X_o) + m_{22}(Y - Y_o) + m_{23}(Z - Z_o)}{m_{31}(X - X_o) + m_{32}(Y - Y_o) + m_{33}(Z - Z_o)}
\end{aligned} \tag{3.1}
$$

Here x and y are the image space pixel coordinates. The x_0 and y_0 are the principal point (the center coordinates of the image coordinate system). The capitalized X Y Z are the corresponding 3D object space coordinates, with X_C Y_C Z_C the translation to image space coordinate system. The focal length f is expressed in pixel units, the m contains all the three rotational parameters ϕ,ω and κ. The distortion of lens is corrected by the terms Δx and Δy , they describe the radial and de-centering effect.

$$
\begin{aligned}
\Delta x &= \Delta x_r + \Delta x_d \\
\Delta y &= \Delta y_r + \Delta y_d
\end{aligned} \tag{3.2}
$$

Radial effect:

$$
\begin{aligned}
\Delta x_r &= \bar{x}\frac{\Delta r}{r} \quad \text{, where} \quad \bar{x} = (x - x_0) \\
\Delta x_r &= \bar{x}\frac{\Delta r}{r} \quad \text{, where} \quad \bar{x} = (x - x_0)
\end{aligned} \tag{3.3}
$$

$$
\Delta r = k_1 r + k_2 r^3 + k_3 r^5 + k_4 r^7
$$

De-centering effect:

$$
\begin{aligned}
\Delta x_d &= P_1(r^2 + 2\bar{x}^2) + 2P_2\bar{x}\bar{y} \\
\Delta y_d &= P_2(r^2 + 2\bar{y}^2) + 2P_1\bar{x}\bar{y}
\end{aligned} \tag{3.4}
$$

In the photogrammetric calibration process, the lens distortions parameters are estimated from a set of measurements of a chess board from different perspectives. The calibration procedure is done by using a Matlab calibration toolbox, which is similar to the work of Heikkila and Silven (1997) and Zhang (2000). The set of measurements consist of images of a chessboard with known geometry from different perspectives. Looking at the collinearity equations described earlier, there are in total 11 unknowns to be estimated: 3 rotations, 3 translations, 1 focal length, k1 , k2, P1, P2. The main focus is to get the focal length in pixels and the lens distortion parameters. It is assumed that the lens distortion can be modeled by k1, k2, P1, P2. Two equations and 11 unknowns means at least 6 tie-points with known coordinates in 2D image space and 3D object space is required. To model the lens distortion as accurate as possible, measurements from different perspectives are taken.

3.1.1 Result lens distortion calibration

The following focal length, principal point location and lens distortion parameters are estimated from 18 sequenced images of a chessboard from 18 different perspectives.

With a pixel error in column and row = [0.14401 0.15657]

Focal length = 249.4446 [pixels]

$k1$ = **0.8472** $k2$ = **2.0688** $P1$ = **0.0072** $P2$ = **-0.0044**

In Figure 3.3 two sample images of the chessboard is shown. The tie-points with known coordinates in image coordinates and object coordinates are the corner locations of the black and white squares in the chessboard. For every image in total 9x7=63 corners are known, in both image space and objects space. The calibration process to solve the lens distortion parameters is done with 63 points for each of the 18 different perspectives.

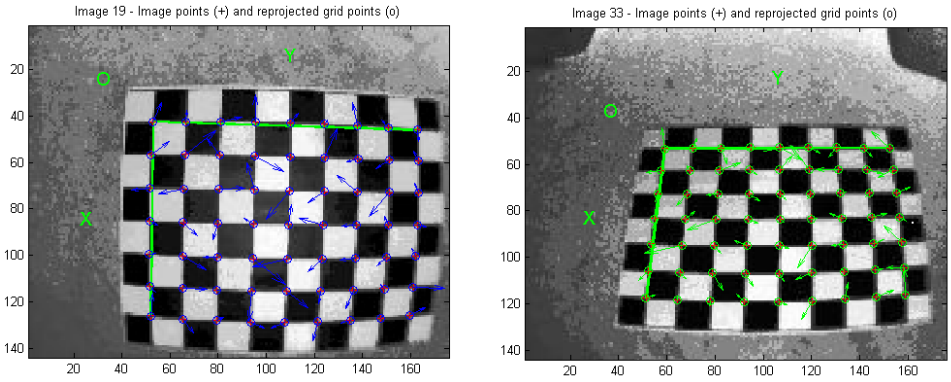


Figure 3.3: 2 sample images from the 18 images used

In Figure 3.4 is shown the pixel error per point detected in the chessboard images. The respective error in pixel units is given. It can be concluded that the precision is at sub-pixel level.

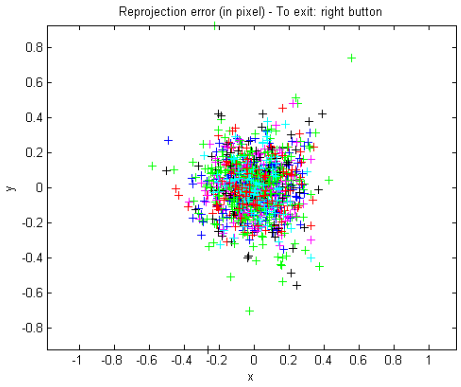


Figure 3.4: The error of pixels in x and y

The exterior camera orientation parameters are retrieved along the calibration process, that is the camera position with respect to the chessboard. See Figure

3.5.

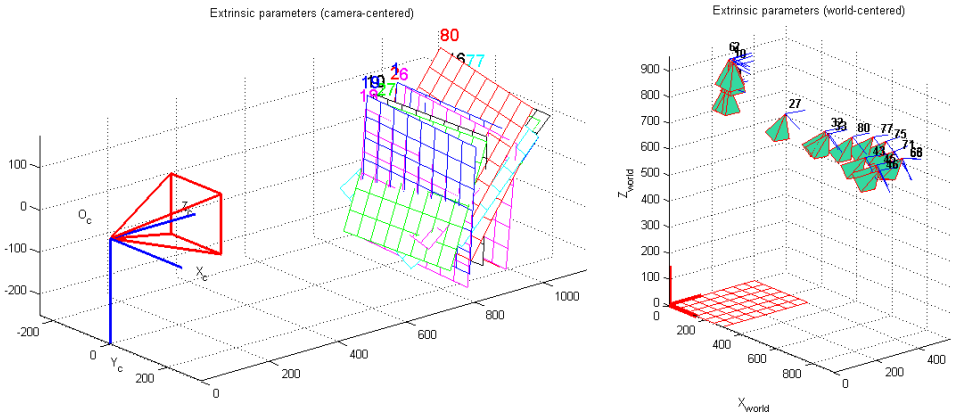


Figure 3.5: The camera position w.r.t to images

The result after calibration is clearly visible comparing Figure 3.6 and Figure 3.3.

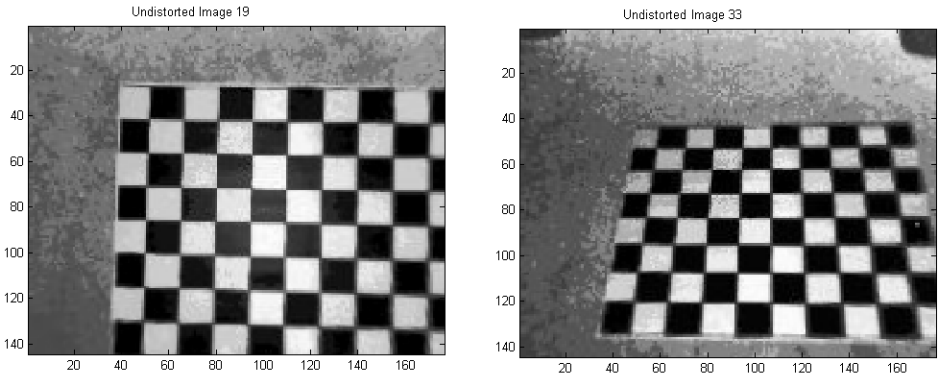


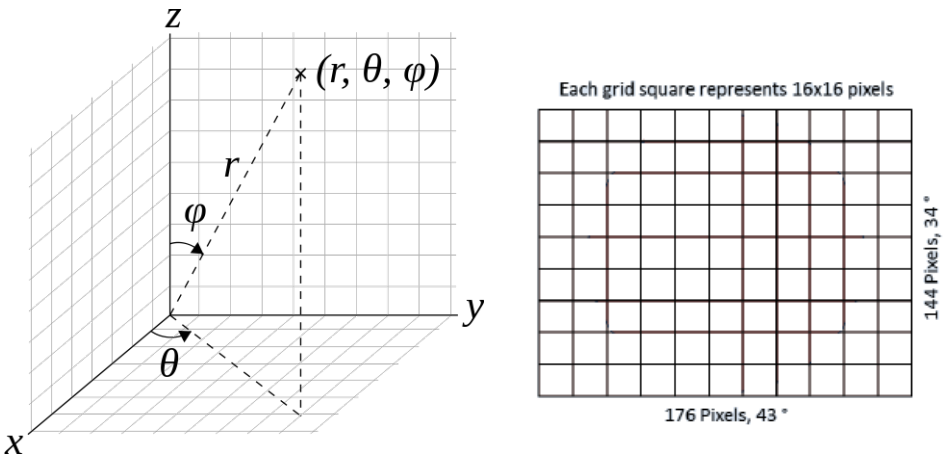
Figure 3.6: The calibrated 2 sample images

Note that the lens distortion exists only in the image plane. After estimation of the lens distortion parameters the range image can be transformed into to 3D coordinates, with the X and Y coordinates corrected for lens distortion. The range image covers a field of view (FOV) 43°x34° (See Figure **??**). Each value in the range image represent the spherical radial distance between the lens optical center and the sampled point of object surface. To transform the radial distance

to X, Y, and Z, the following relations are used.

$$\begin{aligned} x_{\text{undistorted}} &= \frac{X}{Z} \\ y_{\text{undistorted}} &= \frac{Y}{Z} \\ Z &= R cos\phi \end{aligned} \tag{3.5}$$

The $x_{\text{undistorted}}$ and $y_{\text{undistorted}}$ are related to X and Y through a normalization with Z. Given the radial distance for each pixel, the X, Y, and Z can be computed if the ϕ is known for each pixel coordinate. As discussed in Chapter 2 section 2.1.1, depending on the optic design the instantaneous field of view (IFOV) is constant for all pixels or the pixel size is constant for all pixels. Because the optic design is not clear from literature, it is assumed that the IFOV is constant for each pixel. The depth calibration in the next chapter will take the possible distortions into account. A fixed IFOV means that the angles are constant for each pixel. With a FOV of 43°x34° over 176x144 pixels, the fixed angles are the $\alpha = \frac{43°}{176}$ and $\beta = \frac{34°}{144}$. The ϕ can be calculated for each pixel (row,column). Because the angle ϕ for each pixel can be computed, if α and β is known for each pixel (see Figure 3.7). The spatial resolution (pixel size and shape) is then dependent of pixel location.



(a) Spherical

(b) Field of view of SR4000

Figure 3.7: The fixed angle ϕ related to the fixed regular sensor array

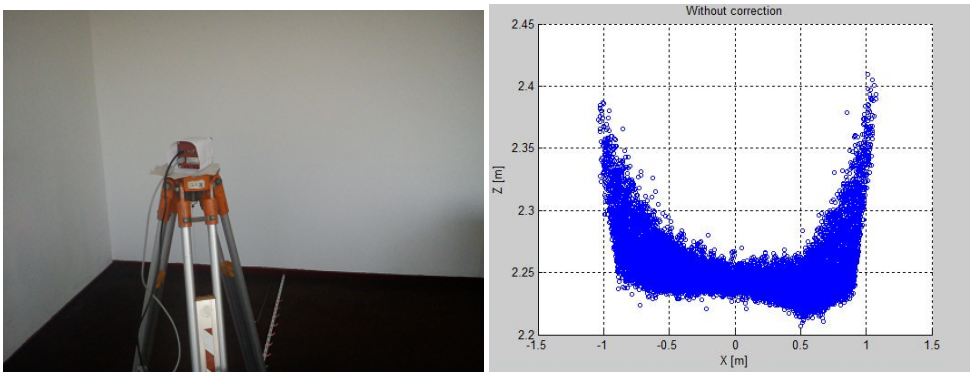
The resulting 3D coordinates are corrected for lens distortions. In the next section the Z-value is analyzed and corrected for errors.

3.2 Depth distortion calibration

Here we will describe how to deal with the systematic error in the measured distance values of the range image. There exist an error in the range image not covered by the conventional photogrammetric calibration process, which in this case only accounts for the distortions in the 2D image space by the lens effect. The distance measured in the range image contains systematic and non-systematic errors, discussed in Chapter 2. To simplify the case only the transformed Z-value from the range image is analyzed and corrected for the systematic error. In this thesis a test set-up is made to account for the following two systematic errors:

1. Distance-related error.
2. Fixed pattern noise error

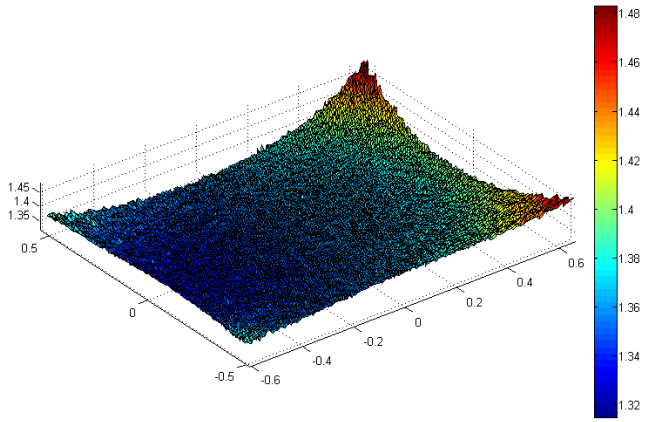
3.2.1 Distance-related error



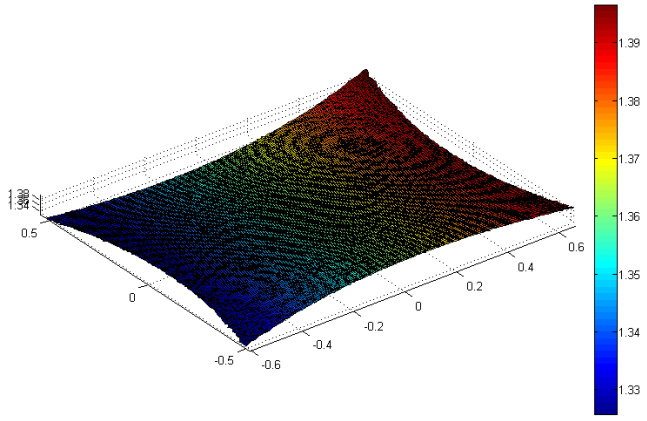
(a) Test Setup (b) Points of scanned wall in X-Z plane

Figure 3.8: Test Configuration

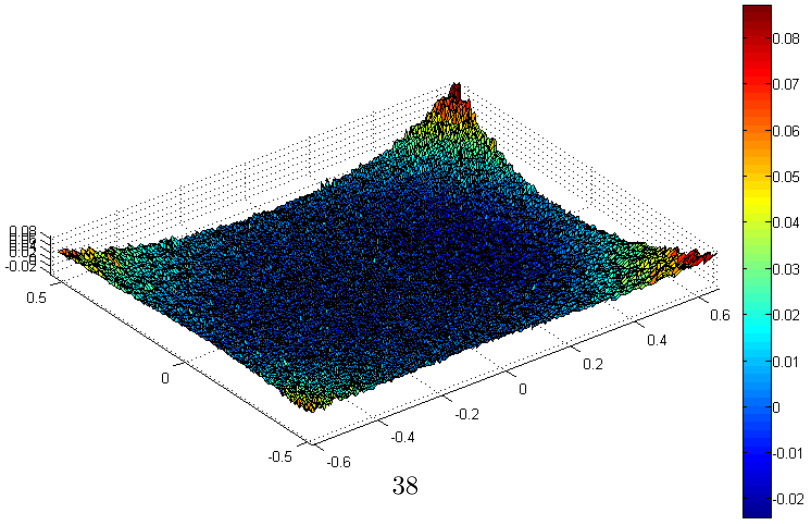
In the test-setup a flat wall is measured from variable distances covering a large part of operating range of the camera. Test set-up shown in figure 3.9. 10 measurements are made from variable distances to the wall, with the distance interval of 10 cm from 1 meter to 3.5 meters. In total 250 range images were made by the range camera fixed on a tripod. Assume the wall is perfect flat. By fitting a plane through the measured point cloud and remove the fitted plane from the measured values. The deviations from the plane show a systematic behavior of the error present in the measurements, they are most probably caused by the distance-related error and fixed noise pattern error. The measurements are of a white wall, showing no variability in reflectivity, the error caused by variable reflectivity "amplitude-related error" is not considered here.



(a) XYZ measured values of wall



(b) XYZ fitted plane



(c) Deviation from plane

Figure 3.9: Plane deviations per pixel

A trend is existing in the deviations, the pixels in the middle tend to be closer to the camera than the pixels on the outer boundary (See Figure 3.10). This effect is accounted in the so-called fixed pattern noise error. The plane deviations with increasing distance to the wall is shown in Figure 3.10 as a plot of the "Standard deviation (RMS) of the plane deviations" versus "the distance to wall".

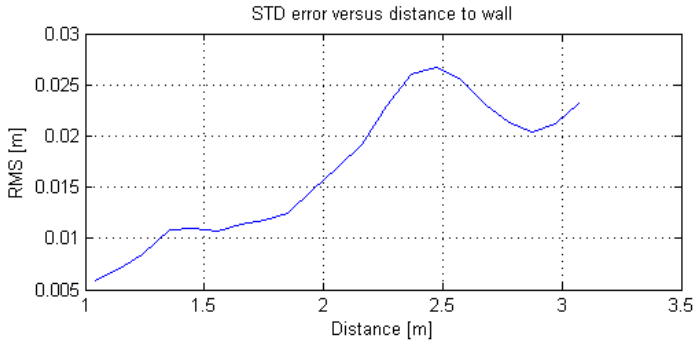
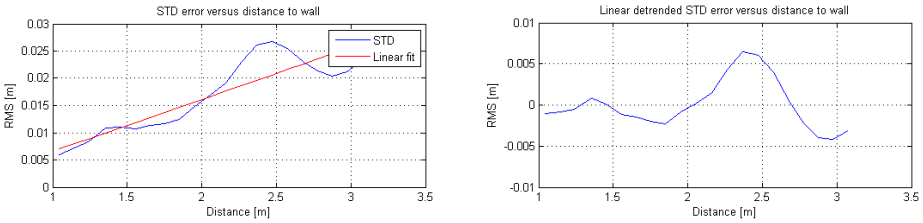


Figure 3.10: Standard Deviation of the Error versus distance to wall

The RMS is increasing with the distance to the wall. This is the noise behavior of the range camera. It is a limitation of the camera that cannot be corrected systematically. However the distance-related error is visible by its sinusoidal or wave-like form, this trend in the noise can be corrected.

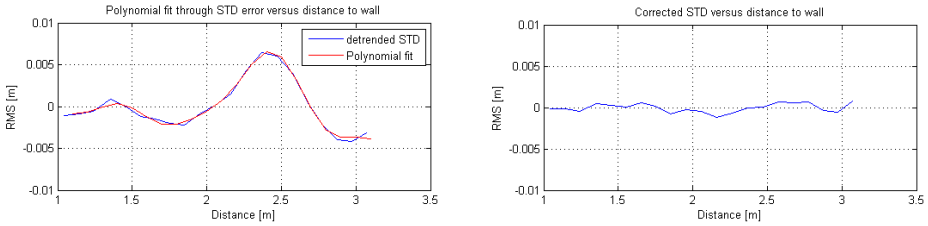


(a) STD

(b) Linear detrended STD

Figure 3.11: Linear detrend the standard deviation

The distance-related error can be minimized globally, because it is consistent for all pixels. First the noise behavior is linear detrended see Figure 3.11.



(a) Linear detrended STD

(b) Wiggling corrected STD

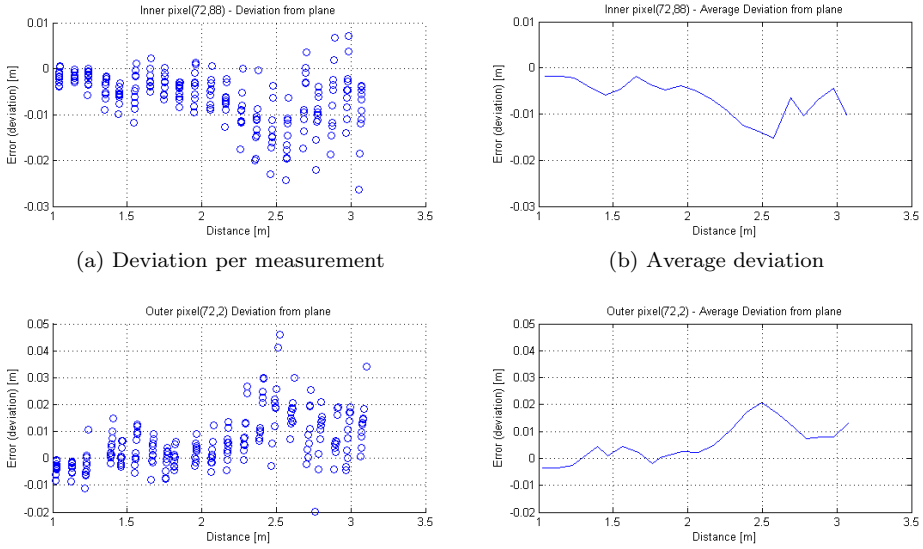
Figure 3.12: Correction for the wiggling error

Then a Fourier approximation of the curve is applied. See Figure 3.12 after removing the fitted trend, the remaining error does not show the sinusoidal effect anymore. An alternative is to make an look-up table, to correct for the distortion. The correction model is the fitted curve as function of the Z value.

$$E_{distance-related}(Z) = a_0 + \sum_{i=1}^{9} a_1 \cos(\frac{2\pi Z_1 \frac{i-1}{2}}{P}) + \sin(\frac{2\pi Z_1 \frac{i-1}{2}}{P}) \qquad (3.6)$$

The sinusoidal behavior is removed from the noise, the correction improved the noise model accuracy from 0.006 meter to a 0.001 meter.

3.2.2 Fixed pattern noise



(a) Deviation per measurement

(b) Average deviation

(c) Deviation per measurement

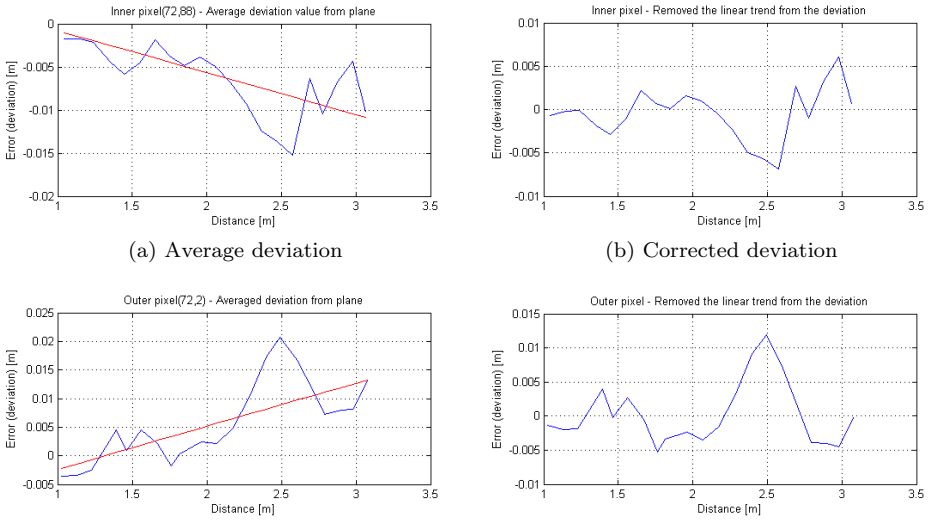
(d) Average deviation

Figure 3.13: Pixel dependent behavior of plane deviations

The different plane deviation behavior related to the pixel position (row, column) is part of the fixed pattern noise error. In Figure 3.13, the average deviations

from plane of a center pixel and an outer pixel is shown. The pixels in the center tend to deviate negative from the plane, they appear to be closer to the camera. The pixels on the outer boundary have depth values greater than in the center. The trend in deviation of the specific pixel is shown in Figures 3.14a and 3.14c

The fixed pattern noise (FPN) error is dependent of pixel. The error is accounted per pixel by a linear fit. The equation is dependent of the pixel location and distance to wall. In figure 3.14 a correction is applied to an inner and outer pixels.



(a) Average deviation

(b) Corrected deviation

(c) Average deviation

(d) Corrected deviation

Figure 3.14: Corrected deviation for each pixel

The systematic correction is applied by removing the linear trend from the measured Z-value. The linear trend per pixel is only dependent of Z-value value:

$$E_{r,c}(Z) = \alpha Z + \beta \tag{3.7}$$

3.2.3 Result systematic error correction

Two systematic errors in the measured Z-value are detected and corrected.

1. One is a linear trend in depth deviation dependent of pixel position and Z-value wall.

2. Second the sinusoidal behavior in the standard deviation of plane deviation versus Z-value.

The remaining noise cannot be corrected as it has a zero arithmetic mean. It is a limitation of the range camera. The fixed noise pattern and distance-related errors are minimized. The improvements are visible in the standard deviation of plane deviations in Figure 3.15.

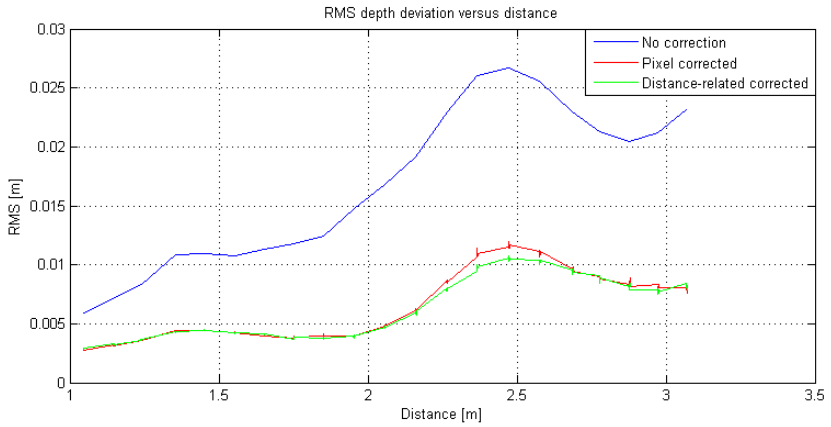


Figure 3.15: The RMS after correction

The error in depth is minimized by a factor of 2.5. This shows the necessity to correct the systematic error in Z-value. It is assumed that the noise effect also exist in X and Y direction, but in this thesis this is not handled. In Figure 3.16 the point cloud of a wall with and without correction is shown.

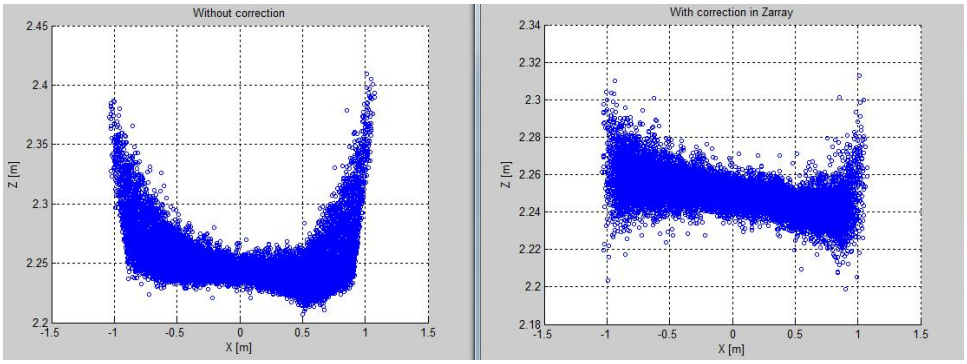
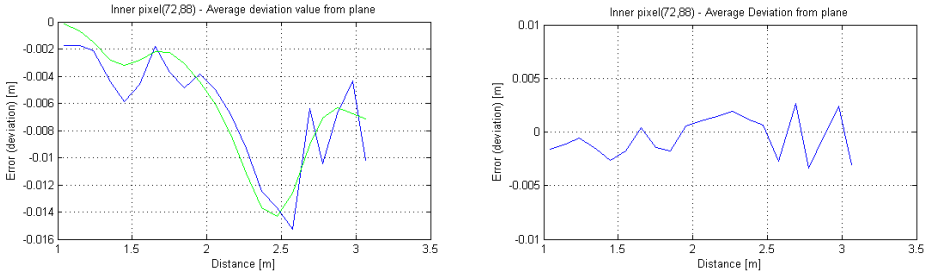


Figure 3.16: The depth value corrected

In Figure 3.17 the inner pixel (72,88) is corrected for the pixel and distance related error.



(a) Deviation and correction fit

(b) Corrected deviation

Figure 3.17: Depth correction pixel (72,88)

The influence by the reflectivity of the surface will not be corrected (amplitude related error), it can be corrected by a more elaborate calibration process described in Lindner et al. (2010).

3.3 Non-systematic error correction

The non-systematic errors cannot be corrected, because it has a null arithmetic mean. However it can be minimized and may be detectable. The noise is reduced by median filtering, a filtering technique without smoothing effect and preserving edges. The mixed pixel error, which is often assumed to be random, can be localized on discontinuous edges. The mixed pixel gives an wrong distance measurement, because its footprint covers partly surfaces of different distances (See Figure 3.18).

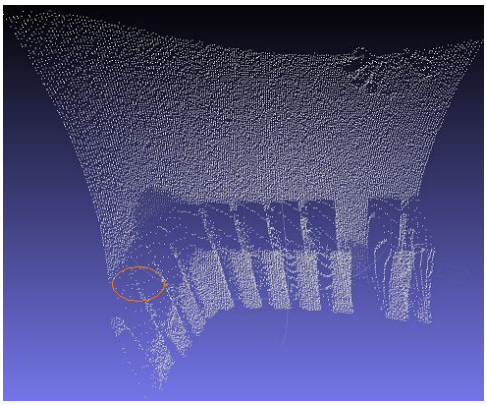
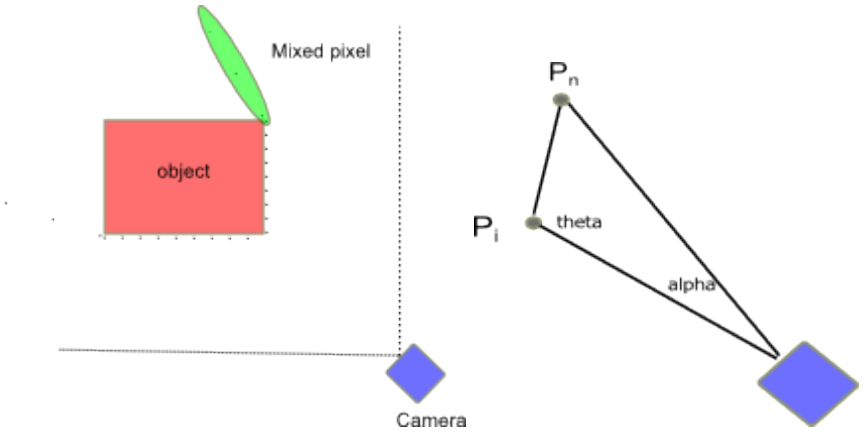


Figure 3.18: Wrong distance information on discontinuous edges

It can be detected by an local angle technique in May et al. (2009), also called the jump edge filtering.



(a) The mixed pixels

(b) The Triangle between camera and 2 neighboring points

Figure 3.19: Local angular filter

$$
\begin{aligned}
L_{P_n-camera} &= \sqrt{X_n^2 + Y_n^2 + Z_n^2} \\
L_{P_i-P_n} &= \sqrt{(X_n - X_i)^2 + (Y_n - Y_i)^2 + (Z_n - Z_i)^2} \\
\frac{L_{P_n-camera}}{\sin\theta} &= \frac{L_{P_n-P_i}}{\sin\alpha} \\
\theta &= \arcsin\frac{P_n \sin\alpha}{L_{P_i-P_n}}
\end{aligned}
\tag{3.8}
$$

Here θ is the maximum allowable angle between two points. For every point in the range image, the θ with the neighboring 8 points is checked. If more than 6 points have an $\theta > 120°$, the points is considered an mixed pixel. The local angle filtering is sensitive to noise and therefore applied after median filtering. In Figure 3.20 the mixed pixels are filtered out.

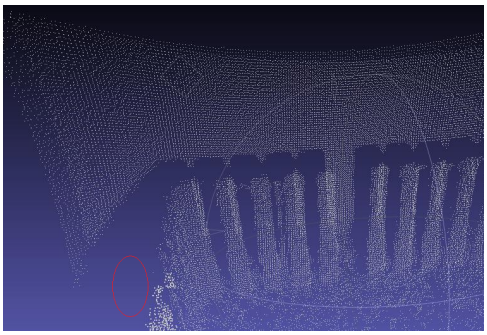


Figure 3.20: Filtered mixed pixel

3.4 Minimal detectable surface size

In this analysis the ability of the range camera to measure correct distance depending on the object surface area is derived. In theory the resolution is the surface area covered by a pixel at a certain distance. Now it is assumed that the minimum surface area required to measure a correct distance is the same as the standard resolution. But this is not true as the standard resolution only applies to 2D images, where the depth is not relevant. The pixels in the range image give the radial distance, where after calibration the image (row, column and radial distance) is transformed into real 3D Cartesian coordinates. Now not only the resolution in X and Y is important to know, but also in Z direction. The noise in Z-direction is already identified and minimized in the previous section. It is important to identify the minimum surface area to measure correct distances. This to assure the geometric features detected in Chapter 4 should be based on an minimum surface area with valid distance measurements.

A test set-up is made to derive the minimal detectable surface size as function of the distance to object surface. The object surface may have different shapes,but for the test a flat surface area is assumed. This surface is approximatly perpendicular to the Z-axis of the camera. In the test setup a cut-out Siemens star is used to assess the "3D resolution". The Siemens star is originally created to test the resolution of optical instruments (See Figure 3.23). The black and white part touches each other only in the center. The center of the Siemens star is visible in our eyes, but the sensor can only resolve the center of the Siemens star to a certain degree from the center (See Figure 3.23). The resolution limit of optical devices are tested based on the minimal resolvable detail from the Siemens star center by checking the distance from the center where the white and black strokes are not mixed up.

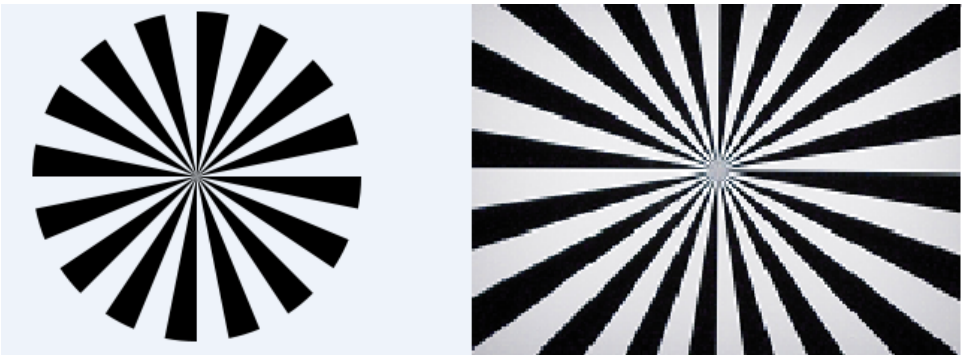
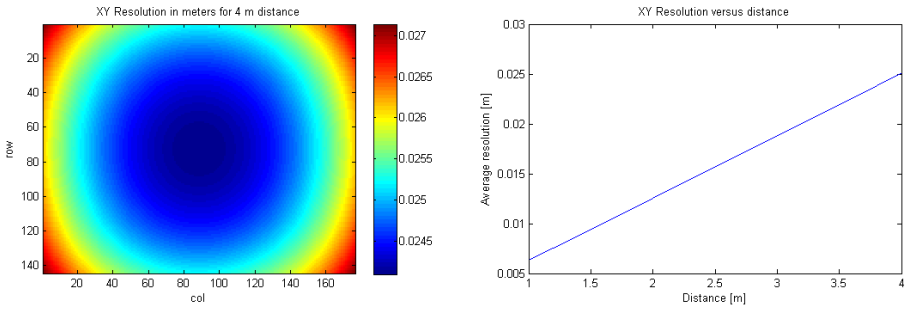


Figure 3.21: Left: Standard Siemens Star. Right: Resolution limit visible in center.

The surface area represented by each pixel can be determined by the 2D resolution of the camera. Knowing the camera geometry, the projection of each pixel into object space is fixed and determined by its pixel position and distance to object surface in the scene (See figure fig:theo1). This holds for the amplitude and range image.

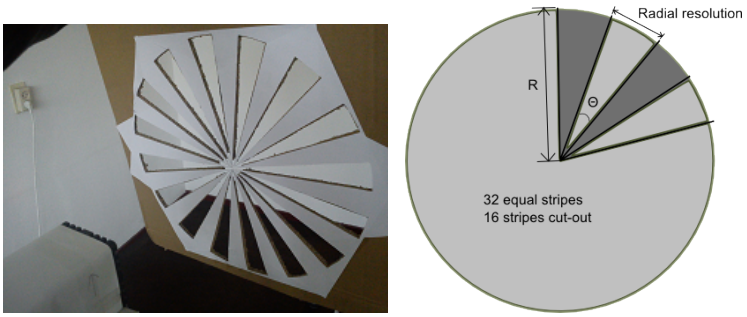


(a) Theoretical resolution at 4 meter (b) Theoretical average resolution per distance

Figure 3.22: Theoretical resolution of range camera SR4000

The main interest is to investigate is whether or not a correct range is returned for each pixel for variable surface area. Each patch of the cut-out Siemens star in figure 3.23a has a decreasing width in the direction to the center. By tracking from what radial distance of the center, correct distance measurements are returned by the range camera. The radial resolution is determined with the setup shown in(See figure 3.23b). This radial resolution is called the "Minimal detectable feature size". A test setup is made to check the 'minimal detectable feature size' as function of distance to object. The minimal detectable feature size, is actually the radial resolution at which the range camera is able to return a correct range measurement.

A manual detection in the range images acquired from the adapted Siemens Star at different distances is used to model the relation "Minimal detectable feature size" as function of distance (see relation 3.9). Measurements are taken from distances from 1 to 3 meters to the 3D Siemens star (See figure 3.23a). The minimal detectable feature size is described by the radial resolution, that is the minimal detectable width of one patch of the Siemens Star (See figure 3.23b).

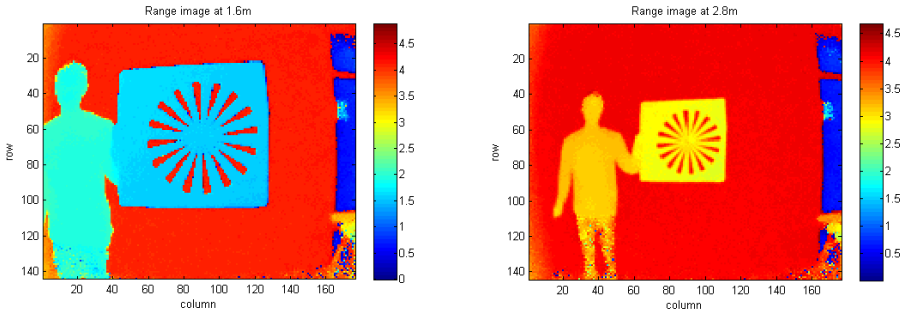


(a) Radial grated cardboard for testing the true resolution of the range image (b) Minimal detectable feature size in Radial resolution

Figure 3.23: Test set-up Siemens Star for range resolution

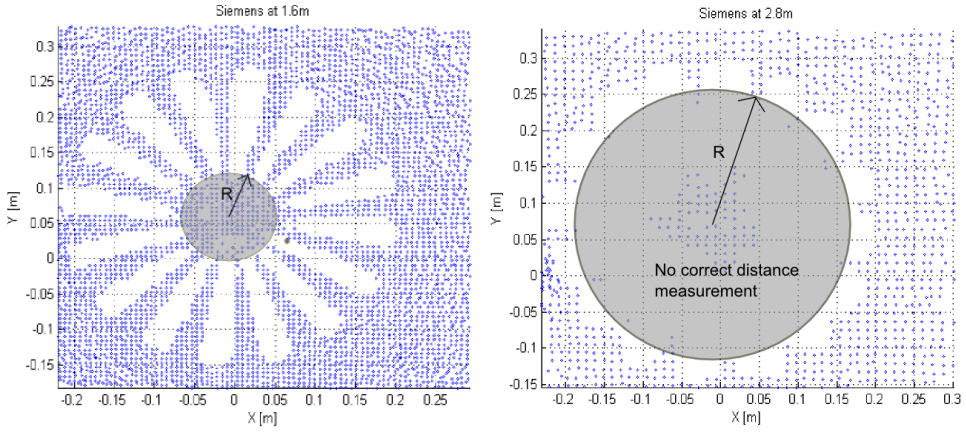
$$\theta = (\frac{2\pi}{32})$$
$$\text{Radial resolution} = \theta \times R \tag{3.9}$$

The radial resolution is a measure of the "minimal detectable feature size" at a certain distance. The radial resolution increases with the distance to the object, in this case the Siemens star. In Figures 3.24 and 3.25 are shown plots of the points detected on the range image of the Siemens star at a distance of 1.6m and a distance 2.8m. The plot shows only the points detected in the Z-value region 1.55m - 1.65m and 2.75 - 2.85m of the cardboard area. The other points are not considered correct distant measurements or not relevant to the experiment.



(a) Range image at 1.6 meter (b) Range image at 2.8 meter

Figure 3.24: Range images at different distance



(a) Detected points of Siemens star at 1.6 meter (b) Detected points of Siemens star at 2.8 meter

Figure 3.25: The detected points on Siemens star at different distances

The further away from the Siemens Star, the smaller the ratio points per square meter P/m^2 are. The process to detect the correct distance measurement of

the Siemens star is done manually by noting the radial distance from the center to the first correct measurement, the radial resolution is then calculated by equation (3.9). In two experiments 20 range images were made in total from a distance to wall of 0.8m to 3.3m. For each image 8 radial distances of the first correct distance measurement were noted down. The following relation of Minimal Detectable Size in Radial resolution is retrieved from the experiment (See figure 3.26).

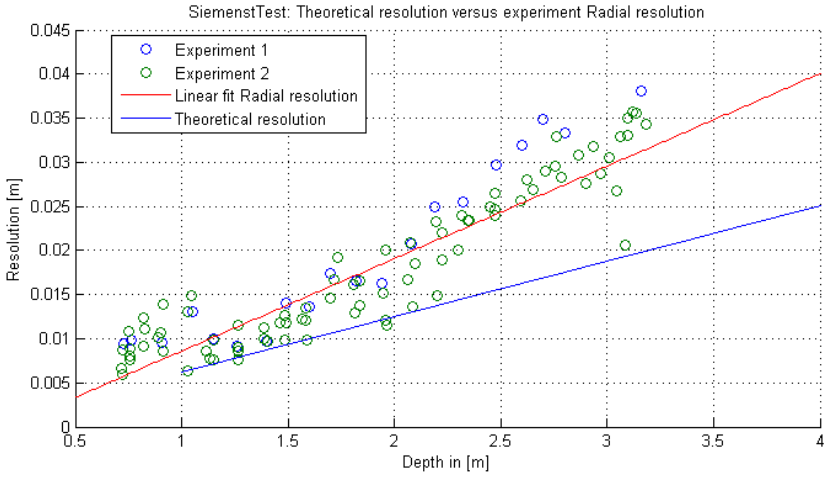


Figure 3.26: Radial resolution compared with Theoretical resolution

The required surface area for the range camera to detect a correct distance is approximatly 30% larger than the theoretical resolution. The radial resolution is the average "Minimal detectable feature size" per distance. With the linear relation an approximation of the minimal detectable feature size per distance is given. This relation will be used in the classification method described in chapter 4, it forms the limiting boundary for the SR4000 to detect features of certain sizes.

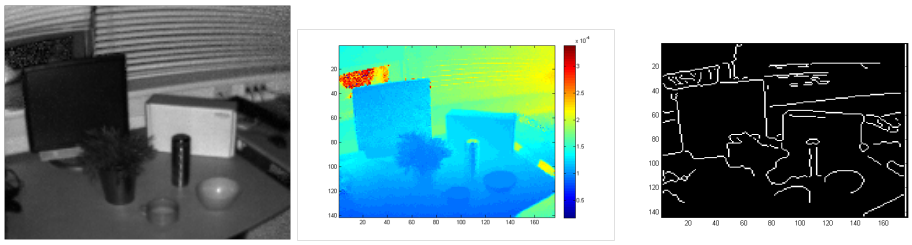
Chapter 4

Registration method with Geometric features

In this chapter a novel method is described for "registration of range images with geometric features". The geometric features addressed are physical 3D edges and corners. To characterize edges and corners in the range image, geometric descriptors are used. The geometric descriptor takes into account the cameras noise and resolution properties. In the end an adapted ICP is described to use the geometric features for registration of the range images.

4.1 Geometric feature classification

Geometric feature is a broad term. In the context of geometry every feature that can be described by a function of spatial coordinates is a geometric feature. The focus is on easy recognizable 3D geometric features seen in our daily indoor environment: Edges and corners. In Figure 4.1 a frame is shown of a common scene in an office.



(a) Amplitude image office scene (b) Range image office scene (c) Canny edge detection

Figure 4.1: Common scene in office

Edge detection is a well developed area in the field of image processing. A simple way is using gradient operator where significant change in amplitude value in neighboring pixels is an indication of an edge. An example of the popular Canny edge operator is illustrated in figure 4.1. The identified edges

are locally dependent on amplitude values, which is sensitive to the lightning condition and the reflective properties of the surface. The way the edges are identified in the amplitude image has no connection with the physical 3D geometry, while most of the detected edges are characterized by their 3D physical shape. More 3D geometry information of these edges needs to be extracted. All the edges characterized by their 3D physical shape are referred as geometric features. The physical edges and corners are to be detected from the 3D points of the range image. Conventional 2D edges detection techniques are not used because they do not utilize the 3D information of the range image, but only the range values in the 2D image space. The novel method takes into account the full 3D information of the range image. The edges in 3D are the boundary of a surface, or more specific a surface curved along a line. Some important requirements have to be considered about 3D edges and its nature in the point clouds of range images:

Scale Depending in which scale you are looking at an edge, it may not even appear as an edge. Like you don't see the earth is spherical when you are standing on it. To describe geometric features, knowledge of the feature size is required. The discussed physical edges in the range image of the swissranger can typically be described in the scale of centimeters.

Surface The physical edges are part of a surface and the surface is described by points in the range image. To describe the most simple surface plane, at least three 3D points are required. To describe an curved surface that may be an edge, a minimum of 4 points is required. The size of the surface area indicates the feature size (scale).

Orientation The 3D points represent the physical locations of a surface area and is described in the coordinate system of the range camera. If the camera position changes, the orientation of the 3D points describing the same surface changes. What does not change is the physical location of the objects in the scene with respect to each other. The geometric features found in the range images are always dependent of the camera orientation. It is therefore important to find geometric features in the range images that are independent of camera position.

Range camera limitation Range image data have limits in 3D resolution and accuracy. The relations for noise and minimal detectable feature size found in Chapter 3 forms a limiting boundary for the camera to detect physical edges of smaller sizes.

These characteristics are taken into account in the edge classification method. First the different types of 3D edges are discussed next.

4.1.1 Geometric features: Viewpoint independent 3D edges

There are different types of 3D edges visible in the range image, but which ones are to be used and not to be used as geometric features. In Figure 4.2 a range

image is shown with four edges highlighted.

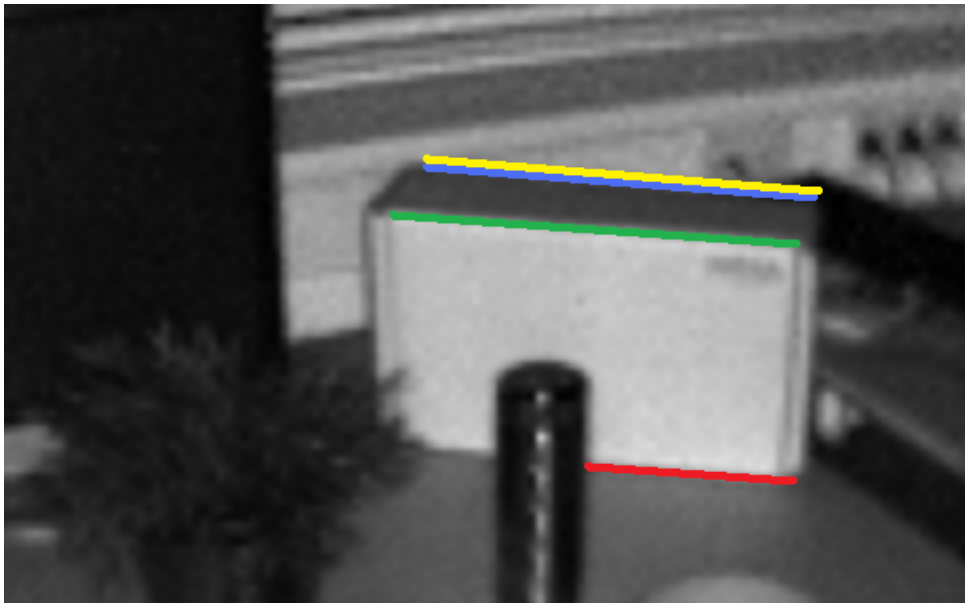


Figure 4.2: Different edge types: Yellow is a concave and discontinuous edge. Blue is a convex and discontinuous edge. Green is a convex and continuous edge. Red is a concave and continuous edge

The following attributes are found to be of interest in describing the different edges:

Concave or Convex edge Convex and concave are distinct geometric features in 3D. They describe whether or not a curved surface bends inward or outwards. In Figure 4.2 a **concave** edge is displayed in **red** and a **convex** edge is displayed in **green**.

Continuous edge The physical edges are visible from the camera's perspective. The edges are not half occluded. **Continuous** edges are the **red** and **green** edges in Figure 4.2. A continuous edge is either concave or convex.

Discontinuous edge Only half of the edge is visible from the camera's perspective. The other half is occluded, this part is referred as the shadow edge. A **discontinuous** edge is colored **blue** and **yellow** in Figure 4.2. The edge on the foreground (blue) is a true physical edge, the shadow edge (yellow) is not a physical edge. A discontinuous edge is always convex(blue) and concave(yellow) edge together.

Following the **orientation** requirement useful edges are the ones that do not change in location when the camera position is changed. These are referred to as viewpoint independent edges. The concave and discontinuous edge, the yellow edge in Figure 4.2 does not fulfill this requirement. This is the shadow

edge, an edge created in the shadow of the front edge (blue line) with respect to the camera (light source). The shadow edge moves with the camera perspective, like the edge of a shadow moves with the light source movement. The remaining three types of edges, the red, blue and green ones are viewpoint independent edges. They are characterized by the following attributes:

1. **The continuous and concave edge (red)**.
2. **The continuous and convex edge (green)**.
3. **The discontinuous and convex edge (blue)**.

Next is described the method to segment the points part of the viewpoint independent edges.

4.1.2 The classification method

A novel method is developed to extract geometric feature points from the range images. These points are part of viewpoint independent edges. As described in the last section the view point independent edges are:

1. The continuous and concave edge.
2. The continuous and convex edge.
3. The discontinuous and convex edge.

The method takes as input a range image and it gives a set of 3D points as an output, which represents the viewpoint independent edges. The set of 3D points have the following properties:

1. Part of the viewpoint independent 3D edges.
2. Not a mixed pixel
3. Not on the boundary of a range image

The classification method consists of two main steps:

1. Create the data layers:
   (a) The mixed pixels: D_{mix}
   (b) The discontinuous edge area: D_{dis}
   (c) The boundary of image: D_b
   (d) The descriptor to detect concavity and convexity: D_{des}
2. To segment the geometric feature points: Perform a sequence of overlay operations with the data layers

$$
\begin{aligned}
D_{edges_{\text{unfiltered}}} &= [\|D_{des} > 2\| - [D_{dis} \otimes (D_{des} < -2)]] \\
D_{edges_{\text{filtered}}} &= D_{edges_{\text{unfiltered}}} \otimes D_{mix} \otimes D_b \\
D_{X_{\text{edges}}} &= D_{edges_{filtered}} \otimes D_X \\
D_{Y_{\text{edges}}} &= D_{edges_{filtered}} \otimes D_Y \\
D_{Z_{\text{edges}}} &= D_{edges_{filtered}} \otimes D_Z
\end{aligned}
\tag{4.1}
$$

The data layers are all derived from the input range image, therefore they all have the size 174x144. The $\otimes$ operation is an elementwise multiplication. The D_X,D_Y and D_Z are the 3D coordinates transformed from the radial distance in the range image. How the classification method was designed and how the data layers were derived is explained in the following section.

4.1.3 The classification method: Algorithm design

The algorithm design is based on the requirements discussed in Chapter 4.1 for detecting 3D edges in range images. The design choices for the algorithm implementation is explained and visualized.

The method to detect 3D edges has to take into account the following requirements:

- Surface: 3D edges are part of surface. A minimum of 4 points is required to detect them, because an 3D edge can be described by two triangles sharing an edge.
- Scale and size: The size of the surface required to detect the 3D edge indicates the scale of the edge. Necessary to know because it is directly related to the capability of the scanner.
- Range camera limits: Noise and Minimal detectable surface area has to be considered in the process, because it determines what can be detected and what not.
- Orientation: This requirement is included by detecting only viewpoint independent edges, they are invariant with respect to orientation.

The edges can be identified by the attributes:

- Concave or convex
- Continuous or discontinuous

This resulted in the following design choices illustrated in the table 4.1 below:

Requirement	**Algorithm implementation**
Surface	A 3x3 kernel is used, the surface is always described by 9 points.
Scale	The surface area covered by the 9 points indicate the scale
Range camera limits	The effect of noise and minimal detectable feature size is incorporated in the kernel operation
Edge attributes	1. The concavity and convexity of the surface area covered by the geometric descriptor D_{des}: Operation with the 9 points in 3x3 kernel. 2. Whether an edge is continuous or discontinuous is covered by the overlay operations with the geometric descriptor D_{des} and the discontinuous edge area layer D_{dis}.

Table 4.1: Requirements implemented in classification algorithm

Next the design choices for the algorithm implementation and the algorithm working principle are explained in detail.

Surface by 9 points: 3x3 kernel
The surfaces in the scene are described by the range image. To describe a 3D edge at least 4 points is required. The pixels in the range image contains the radial distances, which are transformed into Cartesian coordinates X,Y and Z with the camera as origin and the XY-plane parallel to the image plane. The X,Y and Z are stored 2D arrays. The size of the arrays the image size. The range image of the SR4000 is a structured array of the size 176x144. Given this data structure, a common operation is to use a kernel to classify new features (see figure 4.3). Making advantage of the array structure, a kernel of 3x3 is used to describe each point in terms of its 8 neighboring points.

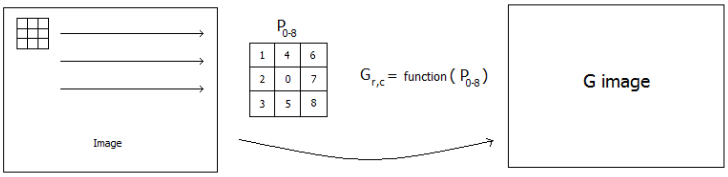


Figure 4.3: Geometric C-descriptor

Edge attributes
To detect the points related to 3D edges in the range image, all points has to be classified in terms of the edge attributes:

1. Concave or convex
2. Continuous or discontinuous

Edge attribute: Concave or convex
The kernel operation that give a measure whether a point is part of a concave or convex surface is called the C-descriptor (See figure 4.4). The kernel size is 3x3, the level of concavity or convexity of a point is determined by comparing the Z-value with its neighboring 8 points and assign 1,0, or -1.

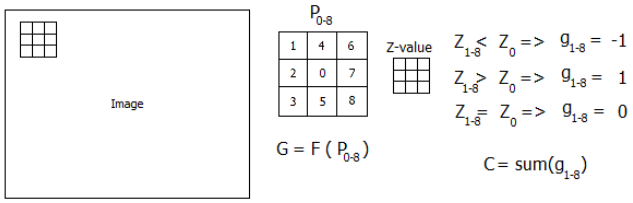


Figure 4.4: Geometric C-descriptor

The C-descriptor give an additional attribute to each point, that is the measure of "concavity or convexity". It is based on the "concave-convex" principle explained in Figure 4.5.

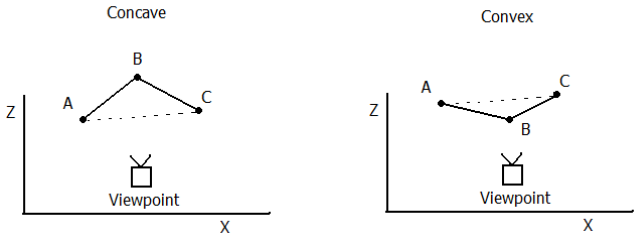


Figure 4.5: Geometric C-descriptor describing a discontinuous edge

The most simple case of concavity or convexity is explained by a line of 3 points in the $Z-X$-plane. Point B is concave (bending outward), if the point is on the left side of the line formed by the points A to C. Point B is convex (bending inward) when it is on the right side of the line A to C. Describing a surface in terms of how much it bends in or outward can be done by checking the point in the middle of the kernel with respect to the cross-side neighboring points in the Z image. whether it is higher or lower than the considered point. This can be done with the 8 neighboring points. The summation of the neighboring points is a measure how concave or convex the points are (See figure 4.6 and 4.7)). Negative G means a concave point and positive G is a convex point.

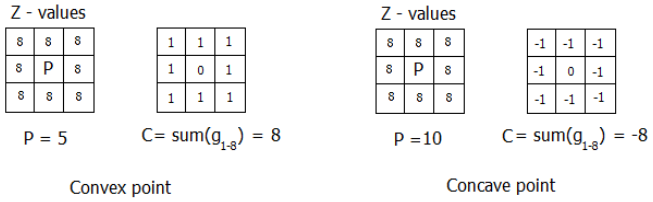


Figure 4.6: C-descriptor describing a continuous surface

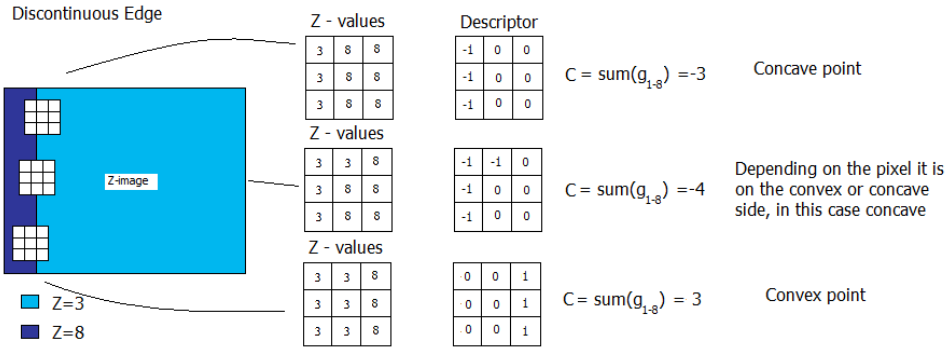


Figure 4.7: Geometric C-descriptor describing a discontinuous surface

There is no clear boundary between an 3D edge and a corner. But if the value is high or maximal like in Figure 4.6, then the points are considered as corners.

The C-descriptor create the data layer that is referred in chapter 4.1.2 as D_{des}.

Edge attribute: Continuous or discontinuous edges
To detect whether or not a point is part of a discontinuous surface is done by an additional G-descriptor.

- Continuous edges: The physical edge is visible and described completely by the range image.
- The discontinuous edge: The physical edge is partly occluded and not completely described by the range image, the other part is occluded from the cameras perspective.

The method to discriminate the continuous and discontinuous edges is to look at the gradient magnitude of the Z-value image, that is the magnitude in change of the Z-value between the points. With the gradient, the discontinuous edges can be detected if they show a change larger than minimal twice the standard deviation of the nominal gradient(See figure 4.8).

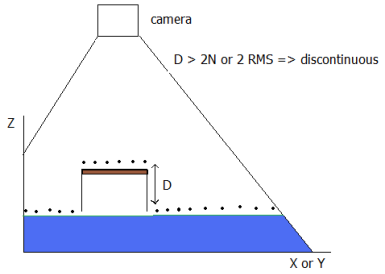


Figure 4.8: Discontinuous edge

The nominal gradient is the "standard deviation of depth changes between points of a continuous surfaces". It is as function of the Z-value, similar to the noise relation derived in chapter 3. The only difference is in noise the deviation from plane is modeled and for gradient the deviation from the average depth changes. This relation basically tells a point is discontinuous if the change in depth with the surrounding points are above twice the standard deviation 2σ, which means 95% of the cases.(See figure 4.8).

As discussed earlier the concave side of the discontinuous edge is not viewpoint independent. This side of the edge is removed by the following overlay operations:

$$D_{\text{viewpoint independent}} = \|D_{des} > 2\| - [D_{dis} \otimes (D_{des} < -2)] \tag{4.2}$$

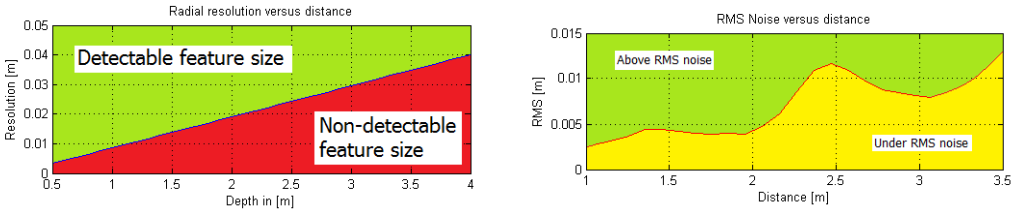
The value 2 is used to distinguish edges from planar surfaces. With the C-descriptor and the G-descriptor all the useful viewpoint independent edges are extracted.

Range camera limits

There are two limitations of the camera analyzed in the thesis. They are also considered in the process to segment feature points.

1. Minimal detectable surface: The size of detectable surfaces is determined by the relation found in the Siemens star tests in chapter 3: The radial resolution (See Figure 4.9) is a limitation of the range camera. Features smaller described by a surface smaller than the radial resolution give no reliable distance measurement.

2. Noise: The noise present in the range measurement is limiting the accuracy. This means a difference of neighboring Z-values under the noise level is not considered a reliable difference.



(a) Radial resolution

(b) Noise

Figure 4.9: Limitation of the SR4000

A reliable distance can only be measured above the range camera radial resolution limit. The descriptor is based on 9 points in the kernel that span a certain surface area. The size of surface area can be roughly determined by the width and length spanned in the row and column direction of the 3x3 kernel. The width of the kernel in columns-direction is the sum of distances between the 3 points in the center row. The length of the kernel is the sum of distances between the 3 points in the center column (See Figure 4.10). To account for the radial resolution, the width and length should be at least twice as big as the radial resolution for the respective distance to object (Z-value).

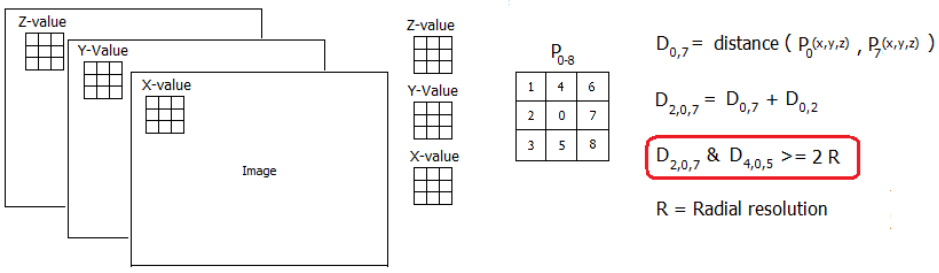


Figure 4.10: Descriptor's surface area

To account for the noise, the difference in Z-value between point 0 and its neighboring points should be at least twice the standard deviation of noise (See Figure 4.11).

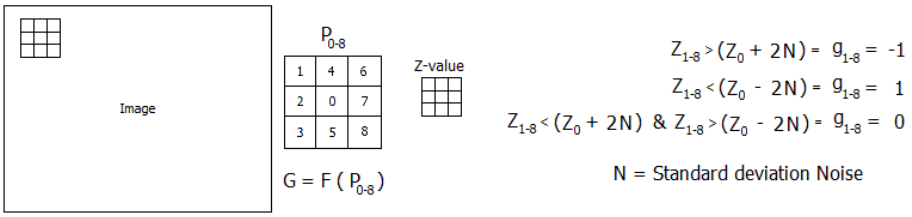


Figure 4.11: Descriptor's noise boundary

Result: Classified geometric feature points
The different edges detected in the range images with the described method is illustrated in Figure 4.12. These edges are invariant in 3D and therefore useful geometric features for the registration.

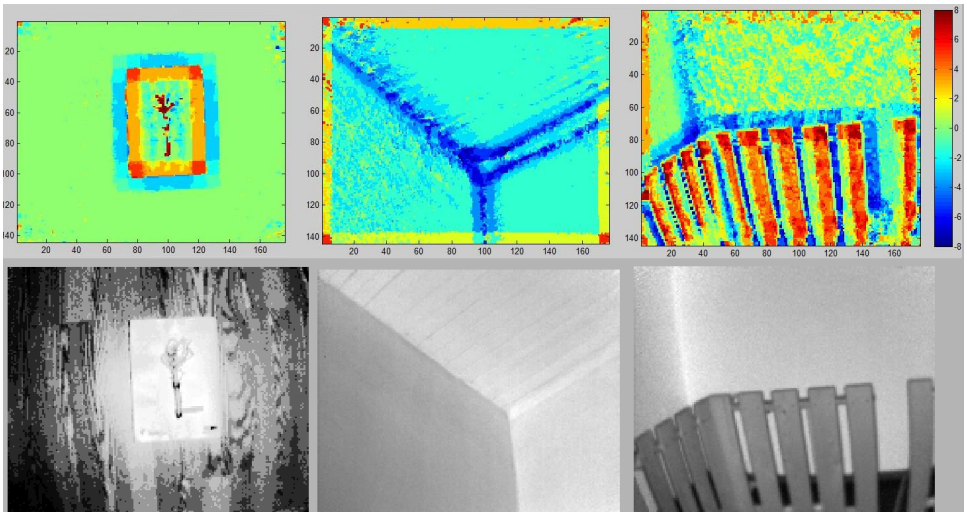
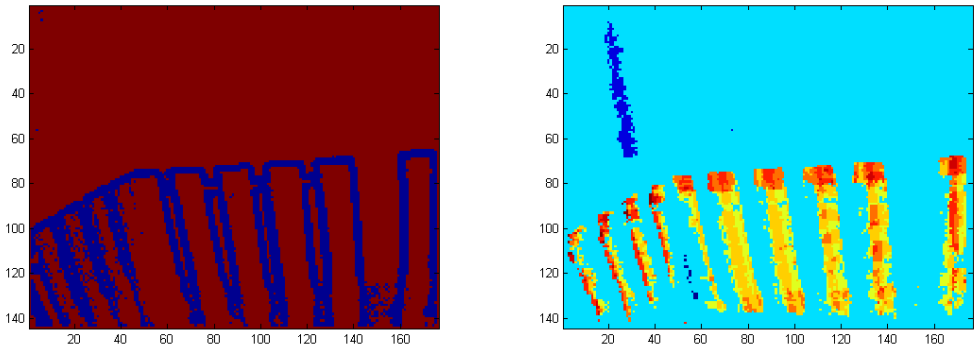


Figure 4.12: Geometric descriptor: Left: A book with a root on top. Middle: A corner of ceiling and walls. Right: Set-up with skeleton of a bed.

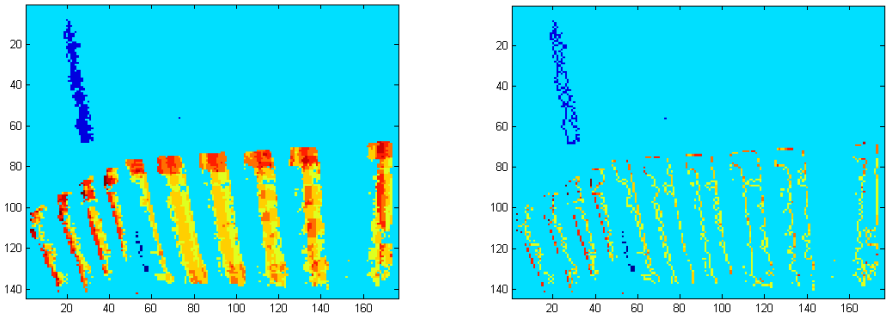
The edges are characterized by the C-descriptor as large magnitude values. Concave edges are negative values displayed as the blue pixels in the image. Convex edges are positive, displayed as red pixels. Discontinuous edge is detected by the G-descriptor. The concave discontinuous edges are removed. This is done using the G-descriptor to detect the discontinuous edges and filter out the concave side of the discontinuous edges(See Figure 4.13).



(a) Discontinuous edges

(b) Viewpoint independent edges

Figure 4.13: Filtering concave discontinuous edges



(a) Viewpoint independent edges

(b) Thinned edges

Figure 4.14: The most important physical edges

The points closer by the physical edges are more valuable than the points farther away. When all the edges are identified, a morphological operation is performed to remove the interior of every clustered points, leaving only the edges of the "edges" in the image (See figure 4.14). It is important to use points closest to the real edges in the registration process, because points on a plane are not the viewpoint independent.

4.1.4 The classification method: Limiting case

A case of viewpoint dependent edges is not solved in the current classification. In Figure 4.15 is shown the extreme cases of viewpoint independent and viewpoint dependent edges. Both cases have discontinuous edges, but the convex physical edge part of the discontinuous edge was assumed to be viewpoint independent. This is not the case for a sphere or cylinder, where the physical edge is changing in location depending on the viewpoint, it is a viewpoint dependent edge.

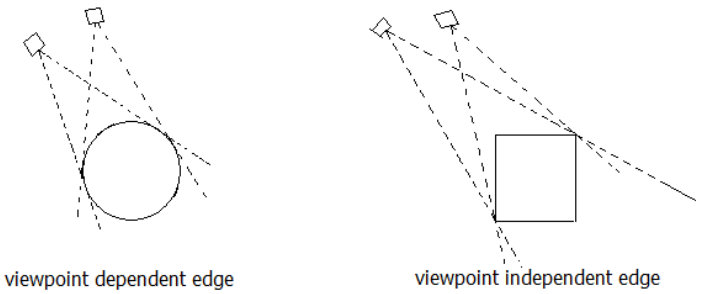


Figure 4.15: Physical edge

In the current method is these viewpoint dependent physical edges are not removed.

4.2 Geometric feature registration method

The points of viewpoint independent edges are extracted with the classification method discussed last chapter. These points are used in an adapted ICP to register the 3D point clouds from subsequent range images into one large point cloud, representing of the scanned 3D scene. The registration process takes 2 subsequent range images at a time and calculates the transformation parameters. Additionally the edge attribute level of concavity of the 3D points are used to assist the correspondence finding criteria in the ICP process. The level of concavity is given by the C-descriptor. The original ICP code is written by (Mia, 2005)

The main steps of the adapted ICP algorithm, referred as GM-ICP:

a With as input two subsequent reduced point sets of geometric features, where point set P_a is the reference model and point set P_b is to be registered to P_a.

b Find corresponding point in P_a for each point in P_b by finding the minimum d in the normalized "closest similar point" relation:

$$\begin{aligned} d(P_a, P_b) &= \alpha \cdot \frac{D^2}{D^2_{\max}} + \beta \cdot \frac{\Delta C^2}{\Delta C^2_{\max}} \\ D^2 &= (x_a - x_b)^2 + (y_a - y_b)^2 + (z_a - z_b)^2 \\ \Delta C^2 &= (C_a - C_b)^2 \\ \text{Correspondence} &= d(P_a, P_b)_{\min} \end{aligned} \tag{4.3}$$

Where D is the metric distance between the points in P_a and P_b, normalized by $D_{\max}$.The C is the edge attribute concavity. The scale factors α and β determines the weight of importance between the two, this is on default 1. The d is unitless, with maximum value 2.

c Reject point pairs by a distance threshold.

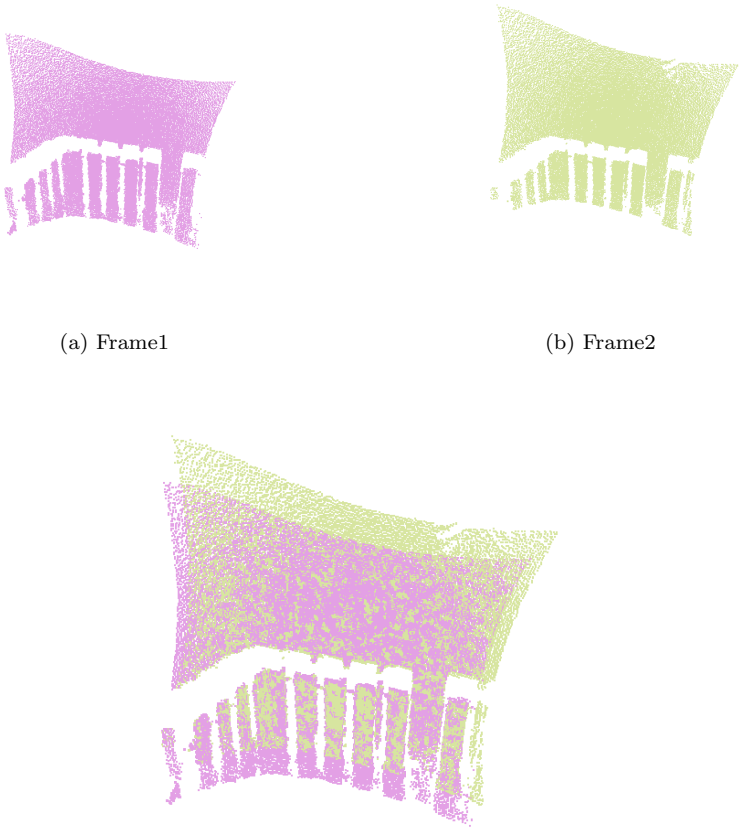
d Compute transformation parameters based on the correspondences by Singular Value Decomposition(SVD).

e Perform the registration and do steps **a** to **d** again until the following minimization criteria is satisfied:

$$\begin{aligned} e &= \frac{\sum \|P_a - P_b\|}{N_{\text{points}}} \\ \|\delta e\| &< Resolution/10000 \end{aligned} \tag{4.4}$$

The value of concavity is included in the correspondence searching. The correspondence searching process will deliver point pairs that are close in metric distance and similar in concavity. More spatial information is exploited from the range image, by including the edge attribute in the ICP process and robustness is increased.

Putting the viewpoint independent edges into the adapted ICP will produce the following results.

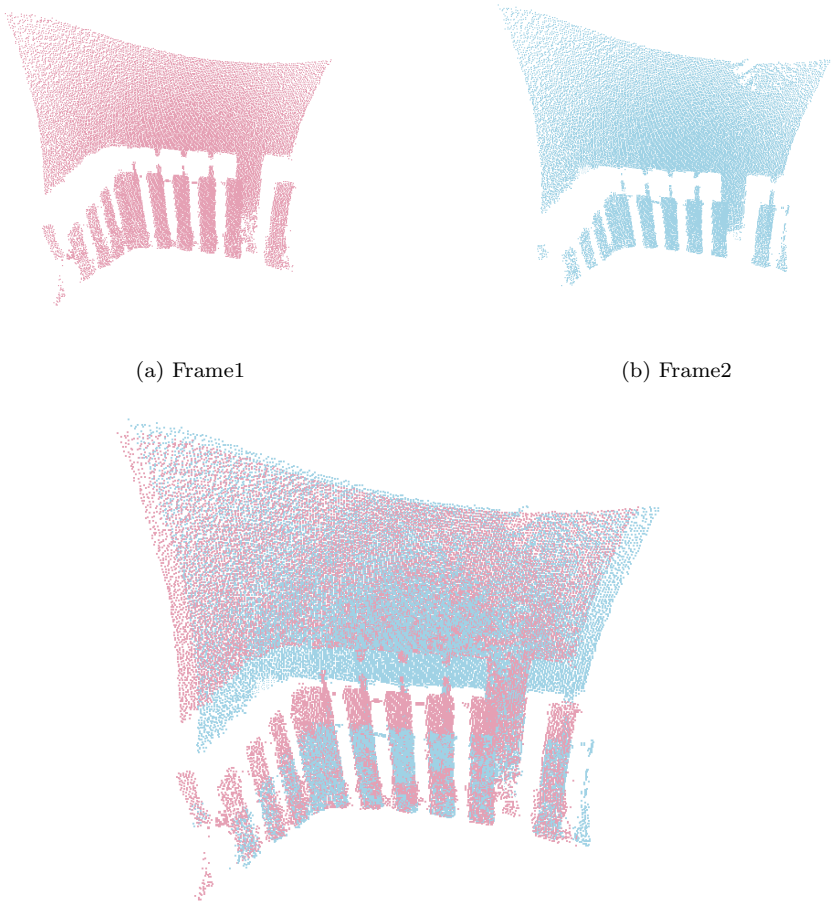


(a) Frame1

(b) Frame2

(c) Registration frame2 into frame1

Figure 4.16: A registration with GM-ICP



(a) Frame1

(b) Frame2

(c) Registration frame2 into frame1

Figure 4.17: A registration with a normal ICP

A successfull registration by GM-ICP is illustrated in Figure 4.16. In Figure 4.17 a failed registration by normal ICP is shown. The main reason for failure in normal ICP is that all points are considered of equal importance, but not all points provide equal spatial information. In the GM-ICP method the geometric features are points from viewpoint independent 3D edges. The points on 3D edges are fixed along the 3D edge, they have only one degree of freedom. The points of an 3D corner are totally fixed in a 3D, they have no degree of freedom. These points are much more valuable than a point on a surface, which has two degree of freedom on the plane. But more important there is no classification between these points in normal ICP.

From the first registration result with the GM-ICP (figure 4.16) is concluded that it more robust that the normal ICP (figure 4.17).

4.2.1 Input parameters

There are 4 input parameters introduced to fine-tune the registration with geometric features. As the method is still experimental, it is necessary to check the influence of the different decisions made in the method. The input parameters are:

1. Noise factor: The factor that multiplies the noise level used in the C-descriptor in the classification method. It is apparently better to have a higher factor when large part of the scene is far away from the camera.
2. Gradient factor: The factor that multiplies the gradient level used in the G-descriptor in the classification method, to determine a discontinuous edge.
3. Excluding factor: The factor that determines the criteria for excluding points in the correspondence finding process in the adapted ICP. It is the "excluding factor" times resolution.
4. Kernel factor: The factor that multiplies the "Minimal detectable feature size", that determines the surface covered by the 3x3 kernel.

Chapter 5

Results

5.1 Registration results

The registration of three different scenes is evaluated by the following points:

Speed Compare the relative speed of registration between standard ICP and the geometric feature ICP (GM-ICP).

Robustness Compare the robustness of the registration process between an standard ICP and the geometric feature ICP.

Error Assess the quality of registration by the accumulated error and by addressing the remaining errors:

1. After each ICP registration there exist an error e_n, expressed by the average distance between the corresponding geometric feature points in the overlapping part of the two registered point clouds $e_n = \frac{\sum (P_{\text{frame1}} - P_{\text{frame2}})}{N_{\text{correspondences}}}$, where $P_{\text{frame1}} - P_{\text{frame2}}$ is an array of absolute distances of the pointpairs, N is the number of point pairs, n is the number of registration. In each registration the transformation parameters are based on two subsequent images (1 and 2). The error created during each registration accumulates over the whole registration process. The total error after registration has a range of $\sum e_n$. This error indicates the maximum possible deviation from the true value. The error is derived from the minimized distance between the geometric feature points of subsequent frames.

2. The quality of registration is determined by the errors. The reflection plays an important role in the range measurement and therefore in the registration. It is caused by external illumination and the materials reflective property.

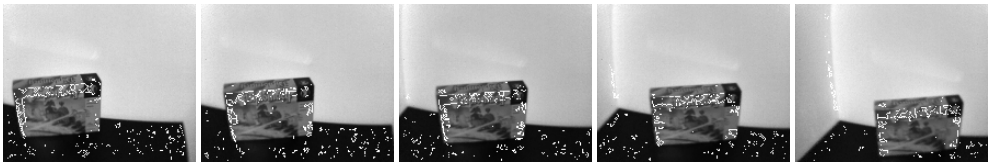
These points are evaluated, because in real-time mapping with range cameras the speed and robustness of the registration algorithm is important. The speed is now evaluated relatively between the new method GM-ICP and normal ICP. The third point is that the errors created per registration is traced, and the error sources causing of a failed registration is addressed. A failed registration is a direct result of the random errors caused by the conditions of the scanned environment or by simply having not enough geometric features. Inaccurate and wrong distance measurements are caused by the inability of the camera to measure consistent correct distances from variable surface reflectivity or interfering illumination.

Not the code is written in Matlab, the speed is therefore not close to real-time, but it gives a good comparison with an existing registration ICP approach in Matlab using only range images.

5.1.1 Scene 1: Box on a black table nearby a corner.

(a) Frame1 (b) Frame2 (c) Frame3 (d) Frame4 (e) Frame5

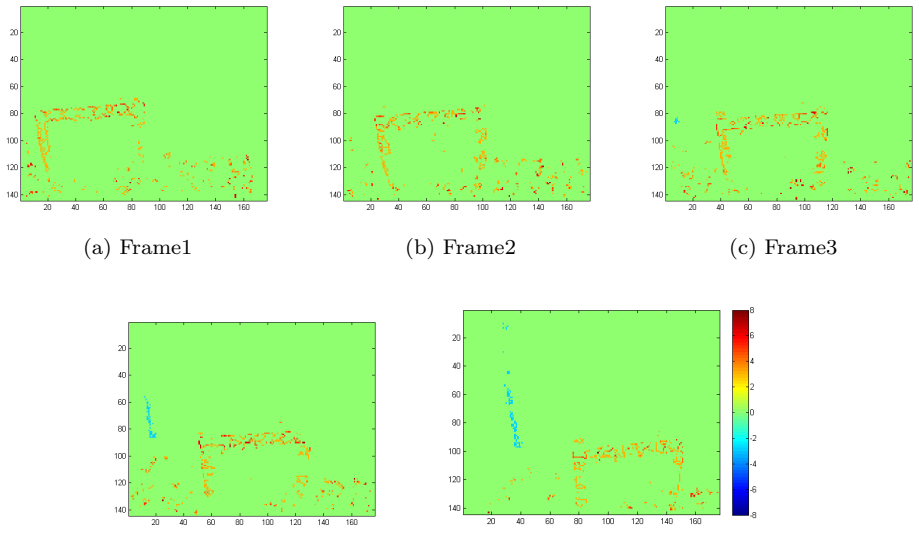
Figure 5.1: Scene 1: Box on dark table with a corner



(a) Frame1 (b) Frame2 (c) Frame3 (d) Frame4 (e) Frame5

Figure 5.2: Scene 1: With the geometric features as white spots

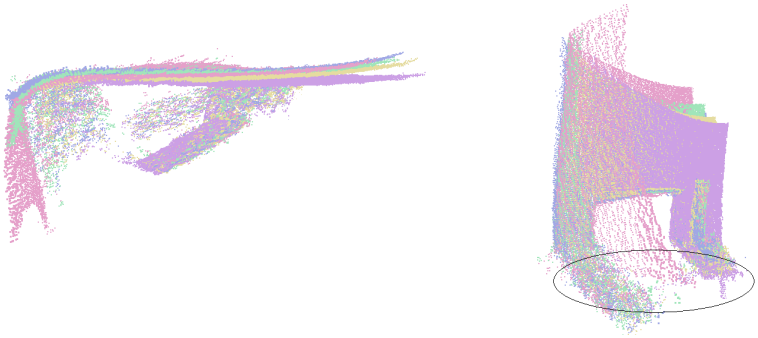
In figure 5.2 can be noticed that there are wrong geometric features detected on the dark table. Low reflective surfaces increases the noise in the distance measurement causing wrong geometric features detection.



(a) Frame1 (b) Frame2 (c) Frame3

(d) Frame4 (e) Frame5

Figure 5.3: Scene 1: Geometric features with the value indicating concave or convex point

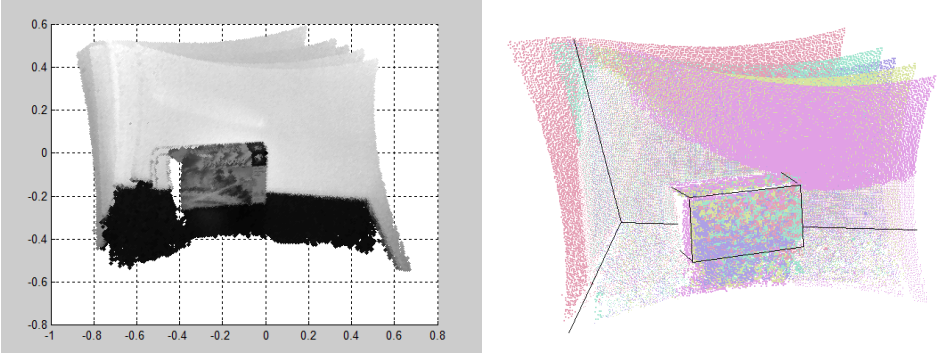


(a) Scene 1:Top-view of the 5 registered point clouds

(b) Scence 1: Side-view of the registered point clouds

Figure 5.4: Scene 1: Top and side view

In figure 5.4b is illustrated the unclear surface of the table. It can be concluded that low reflective surface not only increases noise, but tends to yield higher distance measurements. Nevertheless the registered point cloud correctly registered according to the available range data.



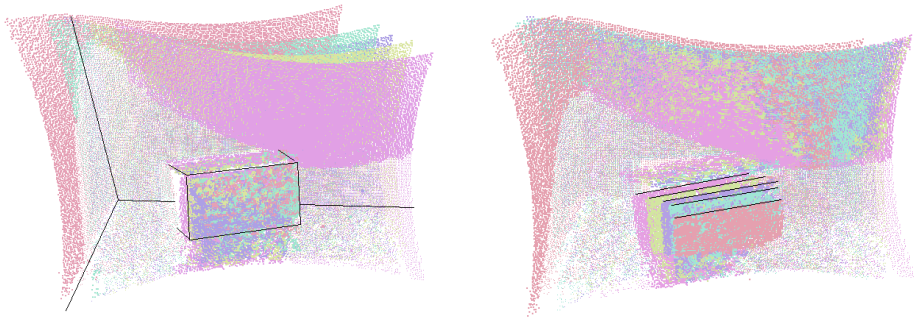
(a) Plot in XY plane of the registered 3D point cloud in Matlab

(b) Pointcloud visualized in Meshlab

Figure 5.5: Scene 1: 5 images registered

Speed: The time required to register 5 images with standard ICP is 172.83 seconds and with GM-ICP it is 62.83 seconds.

- The average time to register 2 subsequent image in standard ICP is 32.81 seconds.
- The average time to register 2 subsequent image in GM-ICP is 14.61 seconds.



(a) GM-ICP registration 5 images visualized in Meshlab

(b) ICP registration 5 images visualized in Meshlab

Figure 5.6: Scene 1: 5 images registered

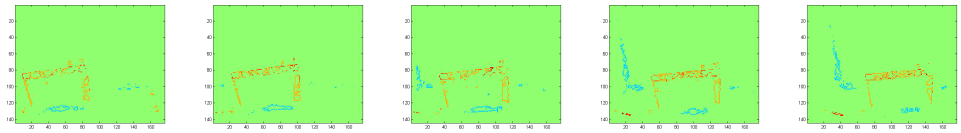
Robustness: It is clear in figure 5.6 that the standard ICP registration is not robust in the subsequent registration. The main indication can be seen in the not overlapped front part of the box, that should have been overlapping. The 5 different colors indicate the 5 subsequent images made. It can be seen that the boundaries of subsequent images is more overlapping in the standard ICP compared to the GM-ICP. This is true because all the points are taken into account in standard ICP and they are of equal importance. The amount of points on the walls is much more than the points on the physical edges of the box. The normal ICP algorithm just overlap the larger amount of points on the wall surface.

Quality of registration
$\sum e_n$ Registering 5 range images accumulates the errors e_1 to e_4. The error after full registration has a range of $\sum e_n$.
e_1=0.019 [m]
e_2=0.024 [m]
e_3=0.027 [m]
e_4=0.017 [m]
This give an accumulated error range of $\sum e_n n$=0.0900 [m].It is the maximum possible deviation from the true value based on the distances between geometric features

In this case the error related to the reflectivity has a large influence to the registration. Normally the concave edge between table and wall should be clearly detected, but in this case it is not. For this point the scene 1 has been redone with the table covered with white paper.

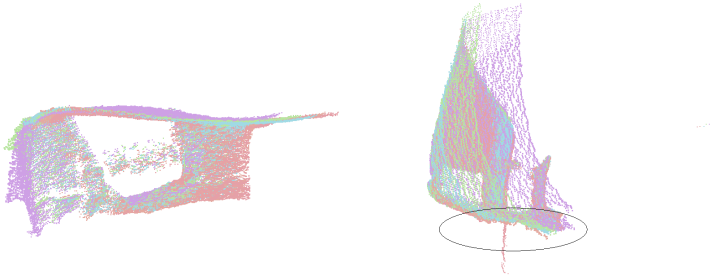
5.1.2 Scene 1: Box on a white table nearby a corner.

(a) Frame1 (b) Frame2 (c) Frame3 (d) Frame4 (e) Frame5



(f) Frame1 (g) Frame2 (h) Frame3 (i) Frame4 (j) Frame5

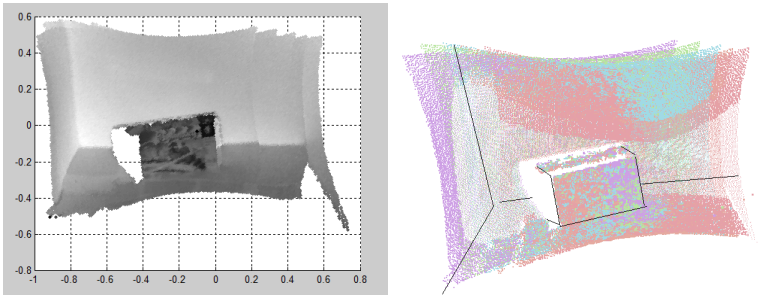
Figure 5.7: Scene 1: Box on white table with a corner



(a) Scene 1:Top-view of the 5 registered point clouds (b) Scence 1: Side-view of the registered point clouds

Figure 5.8: Scene 1: Top and side view

Comparing Figures 5.8b and 5.4b, it can be concluded that the difference between a white and black table is significant.



(a) Plot in XY plane of the registered 3D point cloud in Matlab

(b) Pointcloud visualized in Meshlab

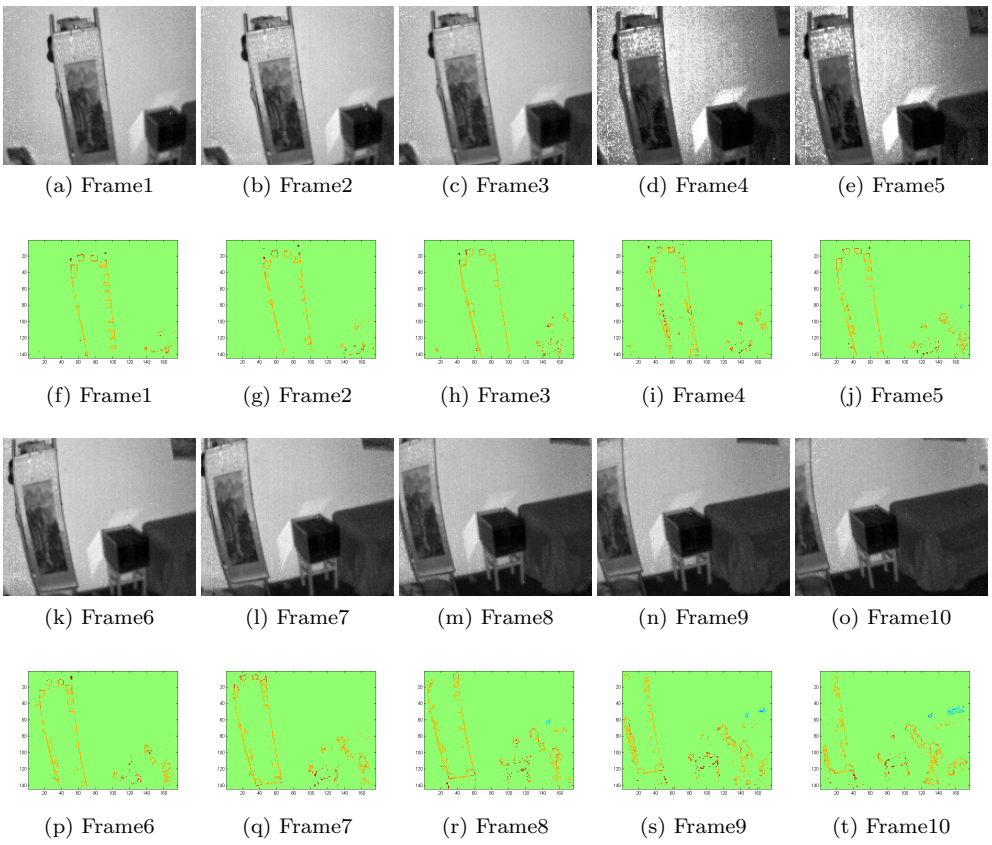
Figure 5.9: Scene 1: 5 images registered

Speed: The 5 registrations were done in 171.73 seconds by ICP and in 62.80 by GM-ICP.

Robustness: The probability of correct registration is increased in the GM-ICP because more correct geometric feature points were detected, see the comparison in figure 5.3a and 5.7f. Wrong geometric feature point detection is caused by reflectivity issue, a correct registration is therefore dependent of the reflectivity of surfaces in the scene.

Quality:Registering 5 range images accumulates the errors e_1 to e_4. The error after full registration has a range of $\sum e_n$.
e_1=0.021 [m]
e_2=0.020 [m]
e_3=0.016 [m]
e_4=0.0087[m]
This give an accumulated error range of $\sum e_n n$=0.067 [m], this is less than the accumulated error with a black table. The edge of table with the wall is now more visible, because the table is covered with white paper. It can be concluded that the increased noise by low reflective surface has great effects on the registration performance and the accumulated error.

5.1.3 Scene 2: Part of room with bookshelf and sofa



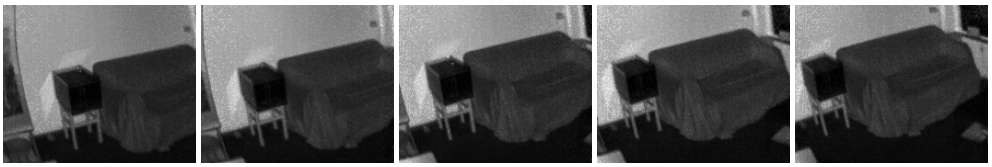
(a) Frame1 (b) Frame2 (c) Frame3 (d) Frame4 (e) Frame5

(f) Frame1 (g) Frame2 (h) Frame3 (i) Frame4 (j) Frame5

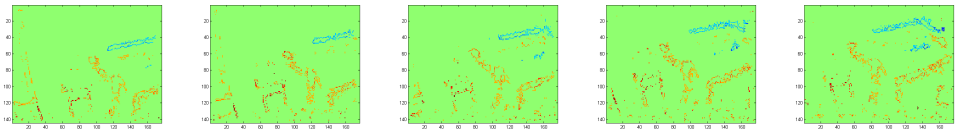
(k) Frame6 (l) Frame7 (m) Frame8 (n) Frame9 (o) Frame10

(p) Frame6 (q) Frame7 (r) Frame8 (s) Frame9 (t) Frame10

Figure 5.10: Scene 2: Amplitude images with the classified geometric features

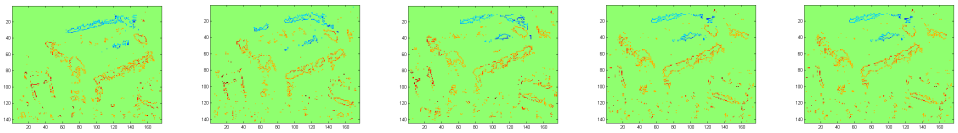


(a) Frame11 (b) Frame12 (c) Frame13 (d) Frame14 (e) Frame15



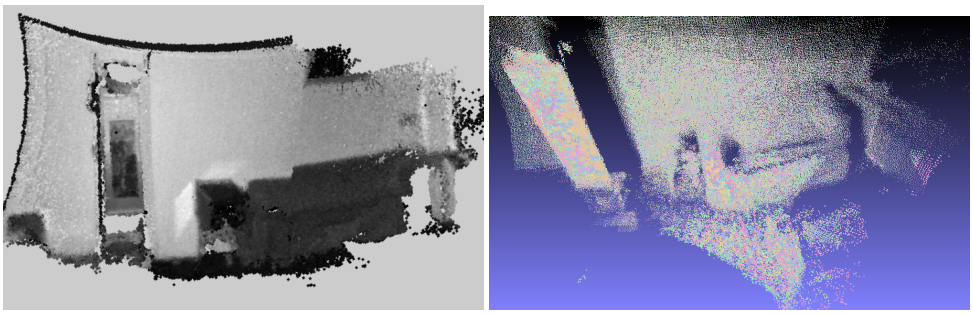
(f) Frame11 (g) Frame12 (h) Frame13 (i) Frame14 (j) Frame15

(k) Frame16 (l) Frame17 (m) Frame18 (n) Frame19 (o) Frame20



(p) Frame16 (q) Frame17 (r) Frame18 (s) Frame19 (t) Frame20

Figure 5.11: Scene 2: Amplitude images with the classified geometric features



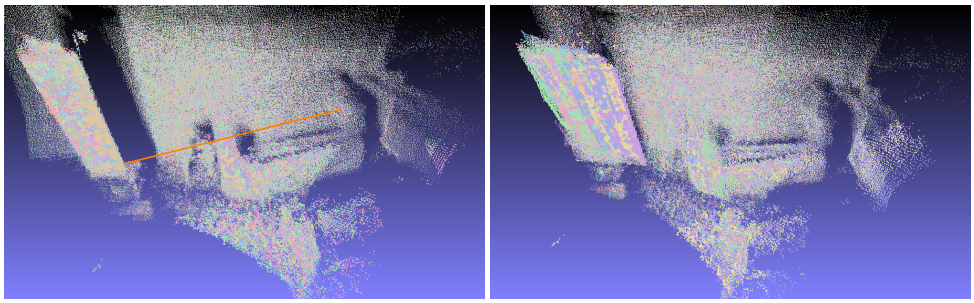
(a) Pointcloud plot in Matlab (b) Pointcloud visualized in Meshlab

Figure 5.12: Scene 2: 20 images registered

Speed: The time required to register 20 images with standard ICP is 674 seconds and with GM-ICP it is 255 seconds.

- The average time to register 2 subsequent image in standard ICP is 33.72 seconds.

- The average time to register 2 subsequent image in GM-ICP is 13.61 seconds.



(a) GM-ICP registration 20 images visualized in Meshlab

(b) ICP registration 20 images visualized in Meshlab

Figure 5.13: Scene 2: 20 images registered

Robustness: It figure 5.13 can be seen that the standard ICP registration is not robust in the subsequent registration at all. Like in scene 1. There are a lot of unsuccessful registrations. The GM-ICP performes well despite of the different

Quality of registration
$\sum e_n$ Registering 20 range images accumulates the errors e_1 to e_19. The error after full registration has a range of $\sum e_n$.

e_1=0.0191 [m] e_5=0.0170 [m] e_9=0.0341 [m] e_{13}=0.0411 [m] e_{17}=0.0380 [m]
e_2=0.0177 [m] e_6=0.0231 [m] e_{10}=0.0444 [m] e_{14}=0.0412 [m] e_{18}=0.0315 [m]
e_3=0.0296 [m] e_7=0.0268 [m] e_{11}=0.0291 [m] e_{15}=0.0317 [m] e_{19}=0.0348 [m]
e_4=0.0375 [m] e_8=0.0386 [m] e_{12}=0.0424 [m] e_{16}=0.0375 [m]

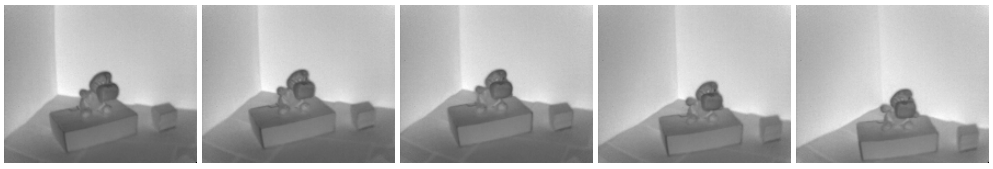
This give an accumulated error range of $\sum e_n n$=0.6153 [m].

A manual measurement was performed in figure to measure the orange line in figure 5.13a. The measurement in the registered point cloud showed 2.19 [m], while the manual measurement showed 2.24 [m], showing a difference of only 0.06[m] . This is in within the expectation of the accumulated error.

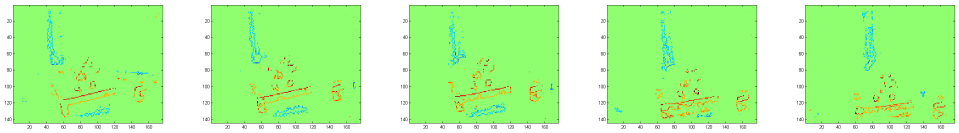
Not all geometric features are detected. The main cause are the amplitude-error and random errors.

5.1.4 Scene 3: Freeform object

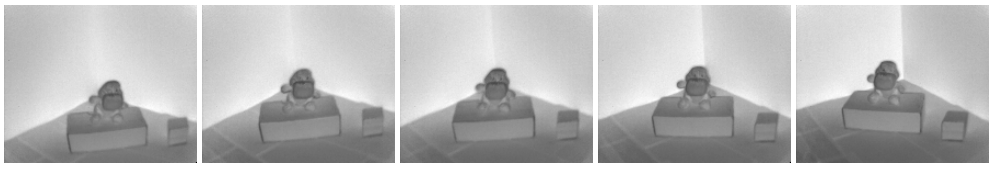
Registration example of an object with viewpoint dependent edges. It is possible to register such an objects if there are sufficient "good" geometric features in the field of view.



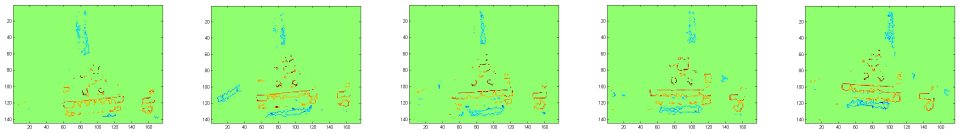
(a) Frame1 (b) Frame2 (c) Frame3 (d) Frame4 (e) Frame5



(f) Frame1 (g) Frame2 (h) Frame3 (i) Frame4 (j) Frame5

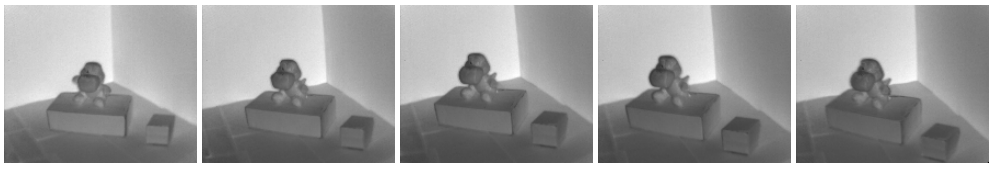


(k) Frame6 (l) Frame7 (m) Frame8 (n) Frame9 (o) Frame10

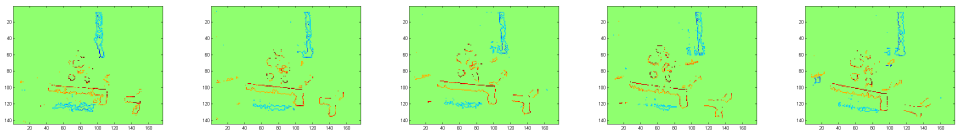


(p) Frame6 (q) Frame7 (r) Frame8 (s) Frame9 (t) Frame10

Figure 5.14: Scene 3: Amplitude images with the classified geometric features

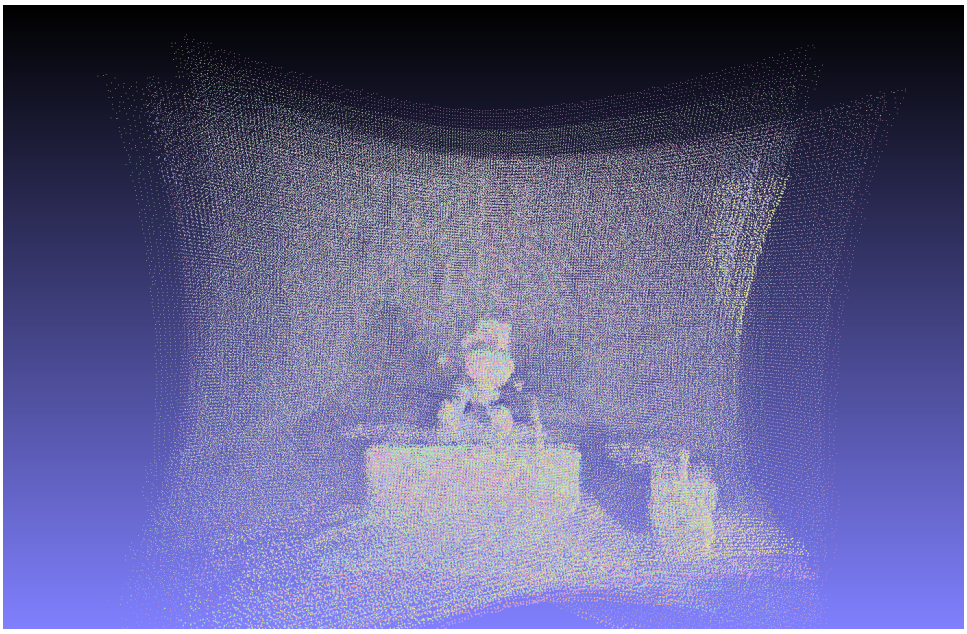


(a) Frame1 (b) Frame2 (c) Frame3 (d) Frame4 (e) Frame5



(f) Frame1 (g) Frame2 (h) Frame3 (i) Frame4 (j) Frame5

Figure 5.15: Scene 3: Amplitude images with the classified geometric features



(a) Frame1

Figure 5.16: Scene 3: Registered point cloud in Meshlab

Speed: The time required to register 15 images with standard ICP is 505 seconds and with GM-ICP it is 191 seconds.

- The average time to register 2 subsequent image in standard ICP is 34.72 seconds.
- The average time to register 2 subsequent image in GM-ICP is 14.65 seconds.

Robustness: The standard ICP registration is performing equal to the GM-ICP method, this is because the walls perpendicular to each other are strong for the normal ICP.

Quality of registration
$\sum e_n$ Registering 15 range images accumulates the errors e_1 to e_15. The error after full registration has a range of $\sum e_n$.

e_1=0.0116 [m] e_5=0.0106 [m] e_9=0.0151 [m] e_{13}=0.0104 [m]
e_2=0.0110 [m] e_6=0.0383 [m] e_{10}=0.0144 [m] e_{14}=0.0108 [m]
e_3=0.0162 [m] e_7=0.0173 [m] e_{11}=0.0120 [m] e_{15}=0.0157 [m]
e_4=0.0122 [m] e_8=0.0125 [m] e_{12}=0.0078 [m]

This give an accumulated error range of $\sum e_n n$=0.2167 [m]. This value is relative large, but in this case the registered point cloud does not show deviation as large as 0.2167m. As described before this value indicates the maximal deviation possible.

It is possible to register freeform object with viewpoint dependent edges (edges of the doll), by having enough viewpoint independent edges in the scene (edges of the box and wall junctions).

5.1.5 Input parameters

The registration were made with the following input parameters.

1. Noise factor is 2.5: A bit more than 2σ.
2. Gradient factor is 10: It is much larger than 2σ, it has effect on removing the discontinuous concave edges. To preserve these edges, this value has to be high.
3. Excluding factor is 20: This means the maximum separation in distance between points is allowed to be 20 times the resolution.
4. Kernel factor is 4: This means 4 times the "Minimal detectable surfaces".

These parameters proved to be working in all scenes. The following conclusions related to the input parameters have been drawn in the process of fine tuning different scenes.

The noise factor is related to the amount of geometric features found in the range image, a larger factor means only large geometric features are found. The gradient factor is related to removing the discontinuous shadow edges, a larger factor means removing all the points that are labeled as shadow edges. Shadow edges may occur because of noise. The Excluding factor is chosen after scanning different scenes, it seems that the movement of the camera, or more specifically the overlap between subsequent frame is influencing this factor. The bigger the kernel factor, the lesser the chance an outlier will be labeled as a geometric feature.

5.2 Comparison with existing ICP approaches

Compared to the common ICP method the data reduction for the computationally heavy task in ICP by using geometric features increased the speed of registration. The process to find corresponding closest points is improved by including the level of concavity, it is now called finding the closest 'similar' points.

Compared to the image-feature based technique the geometric features are based on 3D points and viewpoint independent but does not provide unique corresponding points like in image features. The image features are invariant in image space, but not in 3D space. Image features can provide unique point correspondence in 2D, but is not reliable because 3D features changes with respect to camera viewpoint. Geometric features are 3D invariant.

The additional value to existing methods
The method is an improvement to the existing ICP approaches and a valuable addition to the existing general registration approaches for range images. By using image features and geometric features in ICP, a more robust registration approach can be achieved.

Chapter 6

Conclusions

The registration results are concluded and the limitations of the method are discussed in this final chapter. The answers to the research questions are given and improvements of the method are suggested.

6.1 Method performance

From the registration results it is concluded that the developed method is able to classify points from the viewpoint independent edges in the scene. The registration process with an adapted ICP is performing relative faster and more robust to systematic and non-systematic errors than conventional ICP algorithms.

Main observations:

- The reduced geometric feature point set increased the speed of the ICP algorithm.
- The reduced point sets are located on viewpoint independent edges, these points are fixed along the 3D edge. They are more valuable in geometric sense: A point on a flat surface has two degrees of freedom, a point on a viewpoint independent edge has one degree of freedom.
- The point correspondences are found based on shortest metric distance and the similarity of the level of concavity. The criteria is changed from the closest point to closest "similar" points. Similar by means of of the newly introduced concavity measure. The ability to find correct correspondences compared to the conventional ICP is improved.
- In the classification of geometric feature points the camera performance is taken into account: Noise and the "Minimal detectable feature size" is considered in the classification method.
- The systematic errors "Lens distortion", "distance-related error" and "fixed noise pattern" are corrected. The "Mixed pixel" is detected and removed, while the random noise is reduced by median filtering.

6.2 Limitations

From the registration results it is concluded that the method is sensitive to errors caused by the camera. The noise proved to be of big influence in finding correct geometric features. Because it is caused by different factors like reflectivity of surface material and external light sources, which is difficult to deal with in practice. The method only works when there are enough viewpoint independent geometric feature points found to orient in 3D.

6.2.1 Limits by sensor performance

The spatial resolution or point density is decreasing with the scan distance and thefore the accuracy too. The accuracy is partly determined by the systematic errors. The radial resolution determined in Chapter 3 allows only feature detection above the minimal detectable surface. However it does not exclude features from smaller than the minimal detectable size. The noise boundary relation determined in chapter 3 does not hold true when low reflective material is in the scene. Because the noise relation is derived from an experiment with a high reflective white wall.

The remaining systematic error related to the amplitude is not corrected and can create large deviations, mainly caused by the variable reflectivity of surface material in the scene.

Non-systematic noise effects can produce artifacts, or false data that cannot be filtered in the current method. The reflectivity of surface material contributes both to the systematic amplitude error and the random errors. External light sources cause inferences with the sensors illumination contributing to the random errors. False geometric features detected due to these errors cause poor registration results. Avoiding the conditions where the discussed errors are prone to happen would help.

6.2.2 Limits by geometric features: Viewpoint independent edges

In the method the geometric features are points from viewpoint independent edges. The point is assumed to be fixed along the 3D edge or totally fixed in a 3D corner, they are classified by the level of concavity. This means 2D translation and 2 rotation is solved if one detect the same edge in two point clouds.

The scene is fully described in space by the 6 transformation parameters: 3 rotation and 3 translation, if the one of the following can be found in the range image:

- Minimum two viewpoint independent 3D edges not in the same plane.
- Minimum three viewpoint independent corners not in the same plane.

In the current classification method there are viewpoint dependent edges that cannot be removed yet. These are the edges along a curved surface like a sphere.

They have a negative influence on the registration process.

6.2.3 Limits by the adapted ICP algorithm

Unknown overlap is a problem in the ICP process. Currently it is minimized by selecting the points within a certain distance threshold. As discussed in Chapter 4, this threshold is the resolution times an input parameter. Moving the camera slowly, will reduce the effect of non-overlapping areas.

There should be enough geometric features available in the scene such that unequal distribution of geometric features will not disturb the process. Pointing the camera in the scene to a field of view with a lot of geometric features will help the registration.

6.3 Answers to research questions

Answering all the research questions.
Sub questions:

1. How to manage the error in range images?
   Correct the systematic error by modeling the error in repeated measurements in a test setup. Minimize the noise by median filtering. Detect and remove random error such as mixed pixels. (Chapter 3, section 3.1,3.2)

2. What are the geometric features and how to segment them from the range data?
   The discussed geometric features are points on viewpoint independent edges. They are points that are fixed independent of the camera perspective. The classification method discussed in chapter 4 can extract the geometric feature points from the range image, by an overlay operation on different data layers. The 3D resolution and noise properties of the camera is taken into account in the classification method. (Chapter 4, section 4.1.2)

3. How is the novel method performing relative to a current registration approach ?
   The registration process is faster than the current ICP approaches, because the defined geometric features reduces the input for the ICP process. The registration process is more robust than current ICP approaches, because the correspondence finding process is based on closest similar points, the similarity of the points level of concavity is taken into account. Compared to the image feature tracking based approach it is more correct, because the features are fully 3D. (Chapter 4, section 4.2. Chapter 5)

Main research question

"Is it possible to do continuous registration of range images based on geometric features with the SR4000 Swiss ranger?"

This is proven to be possible under the conditions:

1. The camera is moved steady and slow in a static environment
2. There are sufficient viewpoint independent edges in the field of view to orientate the scene.
3. With a registration error that is increasing per registration of two subsequent point clouds.

The results of the geometric feature registration approach is showing to be promising to register a complete scene. The approach is new and can be used with other range cameras. The method is a valuable addition to existing approaches for the purpose "Real-time mapping with range cameras".

6.4 Future work

- The systematic amplitude-related error can be corrected by analyzing the distance deviations for variable reflectivity of surface object.
- The non-systematic error may be minimized by localizing the source by means of geometry, such as multipath is bound to happen in corners. By researching the external light source influence on the data, a more advanced light filtering technique may be applied on software level or hardware level.
- The registration process can be improved by extracting more types of geometric features and use them in their own class in terms of degree of freedom, such as planes, centre points etc..
- Post processing to reduce the error in the final point cloud by loop closure or other known error relaxation techniques. Loop closure is used in SLAM (Chapter 2, section 2.3.3) to relieve the accumulated error through the registration process, by memorizing and detecting features that have been encountered before.
- Combine image feature tracking technique with geometric feature ICP to attain an even more robust method registration method for range images.

Bibliography

Besl, P. and N. McKay (1992). A method for registration of 3-d shapes. *IEEE Transactions on pattern analysis and machine intelligence 14*(2), 239–256.

Böhme, M., M. Haker, K. Riemer, T. Martinetz, and E. Barth (2009). Face detection using a time-of-flight camera. *Dynamic 3D Imaging*, 167–176.

Davison, A. (2003). Real-time simultaneous localisation and mapping with a single camera. In *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, pp. 1403–1410. Ieee.

Faro (2012). Faro laser scanner. http://www.faro.com/focus/uk.

Fuchs, S. and G. Hirzinger (2008). Extrinsic and depth calibration of tof-cameras. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pp. 1–6. IEEE.

Fuchs, S. and S. May (2007). Calibration and registration for precise surface reconstruction with tof cameras. In *Proceedings of the Dynamic 3D Imaging Workshop in Conjunction with DAGM (Dyn3D*, Volume 1.

Hagebeuker, D. (2007). A 3d time of flight camera for object detection.

Heikkila, J. and O. Silven (1997). A four-step camera calibration procedure with implicit image correction. In *Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on*, pp. 1106–1112. IEEE.

Henry, P., M. Krainin, E. Herbst, X. Ren, and D. Fox (2010). Rgb-d mapping: Using depth cameras for dense 3d modeling of indoor environments. In *the 12th International Symposium on Experimental Robotics (ISER)*.

Kahlmann, T., F. Remondino, and H. Ingensand (2006). Calibration for increased accuracy of the range imaging camera swissrangertm. *Proc. of IEVM*.

Khoshelham, K. (2011). Photogrammetry. Course Notes.

Knoop, S., S. Vacek, and R. Dillmann (2006). Sensor fusion for 3d human body tracking with an articulated 3d body model. In *Robotics and Automation, 2006. ICRA 2006. Proceedings 2006 IEEE International Conference on*, pp. 1686–1691. Ieee.

Lange, R. (2000). *3D time-of-flight distance measurement with custom solid-state image sensors in CMOS/CCD-technology*. Ph. D. thesis, Universitätsbibliothek.

Lindner, M. and A. Kolb (2006). Lateral and depth calibration of pmd-distance sensors. *Advances in Visual Computing*, 524–533.

Lindner, M. and A. Kolb (2007). Calibration of the intensity-related distance error of the pmd tof-camera. In *Proc. SPIE, Intelligent Robots and Computer Vision*, Volume 6764, pp. 67640W.

Lindner, M., A. Kolb, and T. Ringbeck (2008). New insights into the calibration of tof-sensors. In *Computer Vision and Pattern Recognition Workshops, 2008. CVPRW'08. IEEE Computer Society Conference on*, pp. 1–5. IEEE.

Lindner, M., I. Schiller, A. Kolb, and R. Koch (2010). Time-of-flight sensor calibration for accurate range sensing. *Computer Vision and Image Understanding 114*(12), 1318–1328.

Lo, T. and J. Siebert (2009). Local feature extraction and matching on range images: 2.5 d sift. *Computer Vision and Image Understanding 113*(12), 1235–1250.

Malis, E. (2007). An efficient unified approach to direct visual tracking of rigid and deformable surfaces. In *Intelligent Robots and Systems, 2007. IROS 2007. IEEE/RSJ International Conference on*, pp. 2729–2734. IEEE.

May, S., D. Droeschel, D. Holz, S. Fuchs, E. Malis, A. Nüchter, and J. Hertzberg (2009). Three-dimensional mapping with time-of-flight cameras. *Journal of field robotics 26*(11-12), 934–965.

Mia, A. S. (2005). An implementation of the iterative closest point (icp) algorithm. ICP matlab code.

Novatnack, J. and K. Nishino (2008). Scale-dependent/invariant local 3d shape descriptors for fully automatic registration of multiple sets of range images. In *Proceedings of the 10th European Conference on Computer Vision: Part III*, pp. 440–453. Springer-Verlag.

Ren, Z., J. Meng, J. Yuan, and Z. Zhang (2011). Robust hand gesture recognition with kinect sensor. In *Proceedings of the 19th ACM international conference on Multimedia*, pp. 759–760. ACM.

Rusinkiewicz, S. and M. Levoy (2001). Efficient variants of the icp algorithm. In *3-D Digital Imaging and Modeling, 2001. Proceedings. Third International Conference on*, pp. 145–152. IEEE.

Rusu, R. B. and S. Cousins (2011, May 9-13). 3D is here: Point Cloud Library (PCL). In *IEEE International Conference on Robotics and Automation (ICRA)*, Shanghai, China.

Scharstein, D. and R. Szeliski (2003). High-accuracy stereo depth maps using structured light. In *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, Volume 1, pp. I–195. IEEE.

SR4000-DataSheet (2011). *SR4000-DataSheet* (2.0 ed.). Technoparkstrasse 1: MESA Imaging AG.

SR4000Manual (2011). *SR4000Manual* (2.0 ed.). Technoparkstrasse 1: MESA Imaging AG.

Zhang, Z. (2000). A flexible new technique for camera calibration. *Pattern Analysis and Machine Intelligence, IEEE Transactions on 22*(11), 1330–1334.