Institute for Computer Science VII
Robotics and Telematics

Master's thesis

3D Real-time Scanning Using a Projector-based Structured Light System

Sven Jörissen

September 2017

First supervisor: Prof. Dr. Andreas Nüchter
Second supervisor: Prof. Dr. Thomas Kuhn

Abstract

During the last two decades, 3D reconstruction of objects and scenes has become more and more popular. Nowadays, a variety of 3D sensors for different purposes is commercially available. 3D models have a great amount of advantages compared to 2D images, since they give a more detailed reflection of shape and structure of our environment and simplify natural perception. Depending on the desired quality of the reconstruction and the field of application, the cost for 3D sensors lies between a few hundred and millions of Euro.

Structured light is a popular technique for robust, low cost, high quality and fast 3D reconstruction. While typical systems utilise full frame varying pattern such as Gray coding, phase shift or De Bruijn sequences, a smaller amount of systems work with light stripes for shape acquisition. This thesis employs an off-the-shelf projector and industrial camera system for light stripe projection and aims at live reconstruction. Additionally, a self-calibration approach is applied to achieve calibration without the need of additional fixtures. The developed system is then evaluated and compared with a traditional Gray code reconstruction.

The results show, that structured light scanning with a sweeping line is an equally precise alternative to Gray code reconstruction, while providing a denser point cloud. Also, the utilisation of a projector provides an easy and low cost opportunity to test and evaluate the concept, but has drawbacks regarding the generation of lines. To further purse the proposed concept, usage of a laser-line projector is recommended. It is demonstrated, that the self-calibration approach is applicable for the used setup and 3D reconstruction is achieved. In order to improve the quality of the self-calibration result to provide comparable accuracy, the line extraction algorithm must be improved and automatic detection and removal of outliers is to be implemented.

Zusammenfassung

Im Laufe der letzten zwei Jahrzehnte ist die 3D Rekonstruktion von Objekten und Szenen immer beliebter geworden. Heutzutage gibt ein breites Angebot an kommerziell erhältlichen 3D Sensoren für unterschiedliche Anwendungszwecke. 3D Modelle haben verschiedene Vorteile gegenüber zweidimensionalen Bildern, da Form und Struktur der Wirklichkeit genauer abgebildet werden und die Perzeption dieser erleichtert wird. Je nach dem, wie die Anforderungen an Qualität und Auflösung der rekonstruierten Modelle ist und was rekonstruiert werden soll, liegen die Preise für 3D Sensoren zwischen wenigen Hundert und vielen Millionen von Euro.

Die Rekonstruktion mit strukturiertem Licht ist ein beliebtes, robustes, und günstiges Verfahren, das qualitativ hochwertige Ergebnisse liefert. Während klassische Systeme verschiedene Vollbildmuster wie Gray Codes, Phasen Verschiebung oder De Bruijn Sequenzen verwenden, ist eine andere Herangehensweise das Projizieren von einzelnen Lichtstreifen. In dieser Arbeit werden ein handelsüblicher Kinoprojektor und eine Industriekamera dazu verwendet, ebensolche Lichtstreifen zu projizieren und die Szene oder die Objekte in Echtzeit zu rekonstruieren. Zusätzlich wird eine Methode zur Selbstkalibrierung verwendet, um aufwendige, explizite Kalibierung zu ersetzen. Das entwickelte System wird schließlich evaluiert und mit der klassischen Gray Code Rekonstruktion verglichen.

Die Ergebnisse zeigen, dass Scannen mit einer über die Szene wandernden Linie ähnliche Präzision wie klassische Gray Code Rekonstruktion liefert, jedoch dichtere Punktwolken erreicht werden. Die Verwendung eines Projektors ist eine einfache und günstige Möglichkeit, ebensolche linienbasierten System zu simulieren und zu evaluieren, hat aber deutliche Nachteile bezüglich der eigentlichen Liniengenerierung. Um das vorgeschlagene Konzept weiter zu verfolgen wird daher die Verwendung eines laserbasierten Linienprojektors empfohlen. Die Arbeit zeigt, dass die Methodik zur Selbstkalibrierung für den Aufbau und das gestellte Problem anwendbar und anschließend eine 3D Rekonstruktion möglich ist. Um die Qualität der Selbstkalibrierung soweit zu verbessern, dass eine zu anderen, auf strukturiertem Licht basierenden Rekonstruktionsmethoden vergleichbare Genauigkeit erreicht wird, muss der Linienextraktor verbessert werden und automatisches Erkennen und Entfernen von Fehldetektionen und Ausreißern implementiert werden.

Contents

Chapter 1

Introduction

1.1 Background and Motivation

3D reconstruction is for several years now one of the top research interests. Traditional imaging devices capture our three dimensional environment only as two dimensional images, which exacerbates the perception of complex objects and scenes. Nowadays, various techniques for 3D reconstruction are available (cf. Sec. 2.1). The introduction of the Microsoft Kinect^{TM} in 2010 led to a dramatic increase in researchers looking into 3D scanning, since for the first time it was possible to get a 3D sensor for a fraction of the cost of professional devices (e.g. LIDAR). One of the most noticed publications is KinectFusion [New11] from 2011. Also people working with computer-aided design (CAD) gained a new way of easily scanning objects for further processing or 3D printing. 3D reconstruction is also an important part for Virtual Reality (VR) applications.
Since the range of the Kinect^{TM} sensor is short and the quality of the resulting point cloud relatively low, a lot of researchers are using advanced structured light scanner, consisting of an active stereo camera-projector, as a high quality and, compared to other 3D sensors, cost effective alternative. A classical high quality structured light system uses spacial varying or coded patterns for scene reconstruction (cf. Sec. 2.2). Another type of structured light scanner uses a laser projector (MEMS-mirror/Galvano-based mirror or simple line laser), that projects a known pattern for reconstruction. Regardless of the configuration, precise intrinsic (parameters of lens and sensor configuration) and extrinsic (rotation and translation between the devices) calibration (cf. Sec. 3.2.2) are a necessity for high quality results. Intrinsic calibration is performed rather easily, robust and typically does not change without modifying the physical state of camera and/or projector. In contrast, extrinsic calibration is a time-consuming process, which needs to be repeated every time the relation of the devices to each other, e.g., when moving the system to another place, changes. It is, however, possible to fix the extrinsic calibration by mounting the devices on a rig, but this fixes also the field of view and measurement range.
A structured light scanner with a simple line laser is currently developed at INESCTEC (Institute for System and Computer Engineering, Technology and Science) in Porto, Portugal for the ¡VAMOS! project [VAM]. ¡VAMOS! is funded by the EU Horizon 2020 program, which aims to develop a *Viable Alternative Mine Operating System*. It is well known, that abandoned mines all

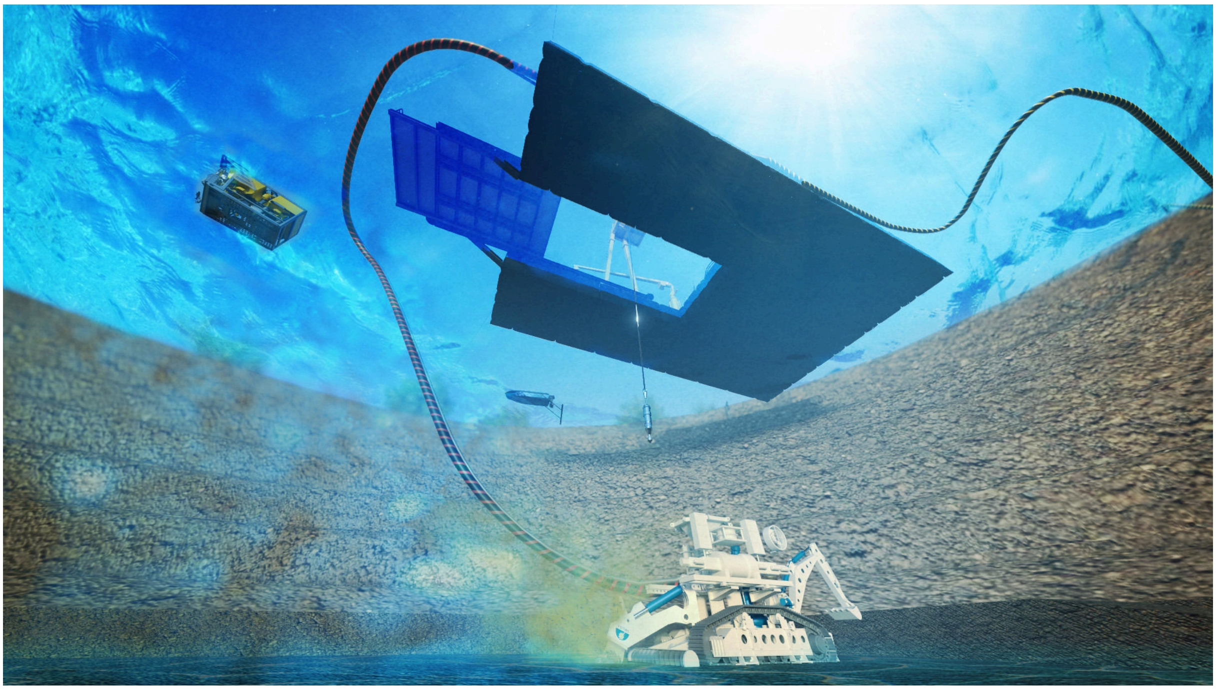

Figure 1.1: Simulation of the ¡VAMOS! project. Rendering was done by DAMEN Dredging Equipment.

over Europe, which were closed many years ago due to the lack of viable mining techniques, still contain a vast amount of mineral resources in a certain depth. Over time, these open-pit mines were, due to natural causes, filled with water. To avoid the costly dewatering and maintaining process for up-to-date, conventional mining techniques, ¡VAMOS! aims at building a prototype underwater, remotely controlled mining machine. This technique can be used to re-open such abandoned mines for further mineral extraction, access mines which are limited by stripping ratio, hydrological or geotechnical problems and also reduces the environmental impact when opening new mines. A simulation of the system is illustrated in Fig. 1.1. For control and navigation, the goal is to provide a virtual reality (VR) interface. The already mentioned structured light system, which will be mounted on the front top of the excavator, has the purpose to scan the environment and create a detailed, 3D map of the surroundings for the VR interface.
A prototype of the scanner is shown in Fig. 1.2 (Left). The device uses a laser line projector, which is mounted on a 1DoF rotational axis and a camera, both fixed on a rig. Additional LEDs (white dots on green surface) are used to overcome the inadequate lightning conditions under water to retrieve colour information. As shown later in this thesis (cf. Sec. 3.2.4), for such a system it is necessary to know the intrinsic parameters as well as the parameters of the laser planes to recover depth information from two dimensional pictures. Therefore, all desired plane positions need to be calibrated. The current procedure for calibrating one laser line position, as shown in Fig. 1.2 (Right) works as follows. The chessboard pattern is moved trough the Field of View (FoV) of the camera. For each position, one image is taken. For all images (between 50 and 60), the chessboard plane is determined and the points extracted from the line on the chessboard plane are reconstructed in 3D. All reconstructed points are then used to fit the laser

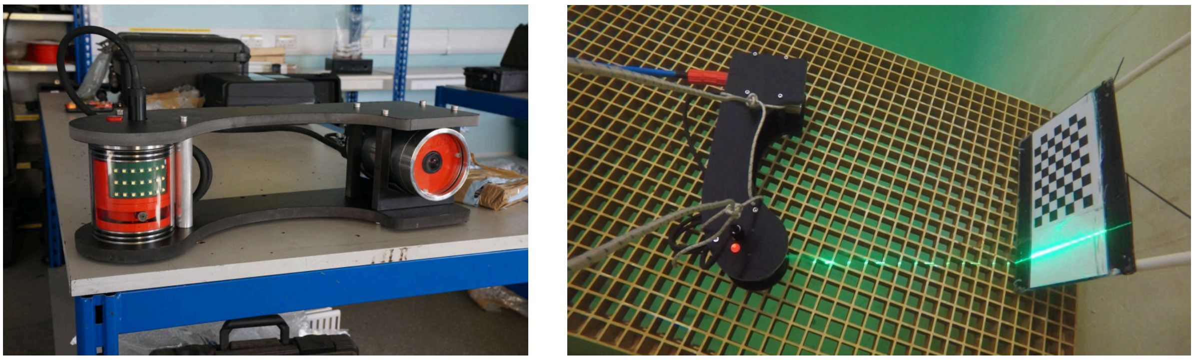

Figure 1.2: Left: Structured light scanner, developed by INESCTEC. Right: Calibration procedure

plane and determine the plane parameters. This procedure is repeated for a subset of all plane positions (around 10). The remaining planes are estimated by interpolating between the measured planes.

As shown, the calibration of a single laser line with a 2D planar pattern is a tedious procedure, which needs multiple images of different positions of the pattern for each single position of the laser line. This procedure could be improved with a 3D calibration pattern (cf. Sec. 4.2.1), which reduces the number of necessary pictures for each laser line position to one. Nevertheless, it is still a time-consuming process, which needs an additional calibration fixture and requires a considerable amount of manual work.

Two more benefits come to mind, when thinking about the ¡VAMOS! project. When developing systems such as the INESCTEC structured light scanner, the system is changed and rebuilt many times before the final goals are achieved. So every time when parts are mounted or unmounted, the calibration needs to be renewed. Additionally, to perform calibration with auxiliary tools such as a chessboard or a 3D pattern under water is far more elaborate and time-consuming. Hence, in order to simplify the process, save time and therefore money, it is desirable to use a self-calibration procedure, which does not require a special geometry of the scene or additional calibration patterns.

With regard to the structured light scanner currently developed by INESCTEC, the method proposed in this thesis is only applicable, if the system would be extended by another laser line projector, which is mounted perpendicular to the currently available projector. This is necessary, since the method requires the system to include a metric constraint in order to be able to perform the calibration on arbitrary scenes.

1.2 Goals and Outline

The goal of this thesis is to design and implement a structured light scanner for object reconstruction. The system is based on a projector-camera pair, providing a simple and cost-effective way for traditional full frame spatially coded structured light scanning as well as the possibility to simulate laser-based projectors, which would be beyond the financial scope of this thesis. Additionally, a self-calibration approach is applied and tested. The intended scanning process can be describe as a light plane sweeping across the scene. The self-calibration process will be performed automatically with regard to the desired light plane patterns to ensure the precise determination of each light plane, but is also applicable for explicit extrinsic calibration of the projector-camera pair and therefore applicable to be used with traditional structured light scanning methods. The scene reconstruction is performed live during the acquisition, which meets the real-time requirement. Finally, an evaluation on the quality of the self-calibration result and the acquired scenes is performed and discussed.

The thesis is outlined as follows:

Chapter 2 provides an overview of the state-of-the-art on all major 3D reconstruction techniques, as well as different structured light scanner systems and self-calibration methods for structured light and laser-based systems.

Chapter 3 introduces the technical detail of the proposed system and introduces the theoretical background on projector and camera calibration, line extraction and the plane parameter estimation necessary. Finally, the methodology for the approach is illustrated.

Chapter 4 describes the experiments carried out in order to analyse the performance of the developed system and software both qualitatively and quantitatively. Further on, the results are discussed, compared and occurring effects are examined.

Chapter 5 gives a short overview of the results obtained during the experiments and summarises the achieved goals, the scientific contribution and the findings. Finally, a short motivation for future work and potential improvements are listed.

Chapter 2

State of the Art

2.1 3D Reconstruction Techniques

An attempt to classify 3D reconstruction techniques is shown in Fig. 2.1. This section gives a short introduction into the depicted categories.

Contact-based Reconstruction

Non-destructive shape acquisition was one of the first 3D reconstruction techniques developed in the early second half of the 20^{th} century. Using a mechanical probe mounted on a high degree of freedom robotic arm to measure multiple points on the surface of a object relative to a world coordinate system provides detailed information about the geometry and shape of the object. For complex objects, this process is very time consuming until enough points have been scanned to provide a dense point cloud for the reconstruction. Nowadays, Coordinate Measuring Machines (CMM), as shown in Fig. 2.2 Top Left, are mainly used in industrial environments for verification of the shape and dimensions of manufactured parts.

Destructive Reconstruction is typically a process of cutting an object into thin slices or grinding away layer by layer to reconstruct shape and internal structure. The grinding machine depicted in Fig. 2.2 was used by W. J. Sollas in 1904 to investigate fossils by serial section [Sol04]. While destroying the object for the reconstruction, a great amount of detail and otherwise non accessible information, such as internal colour, can be acquired. To preserve the specimen, this procedure has been replaced in the recent years by digital imaging methods [Cun14].

Contact-less Reconstruction

Passive devices do not emit any kind of electromagnetic radiation onto the object or scene and use just the information provided by natural lighting conditions.

Depth from stereo is a technical realisation of the human pair of eyes. The typical setup consists of two cameras, mounted on a rig, which acquire images simultaneously (cf. Fig. 2.3 Top Left). In the ideal case, the two cameras are perfectly identical and are mounted in a way such that the two sensors lie in the same plane. Since this configuration is not feasible in reality, the system is

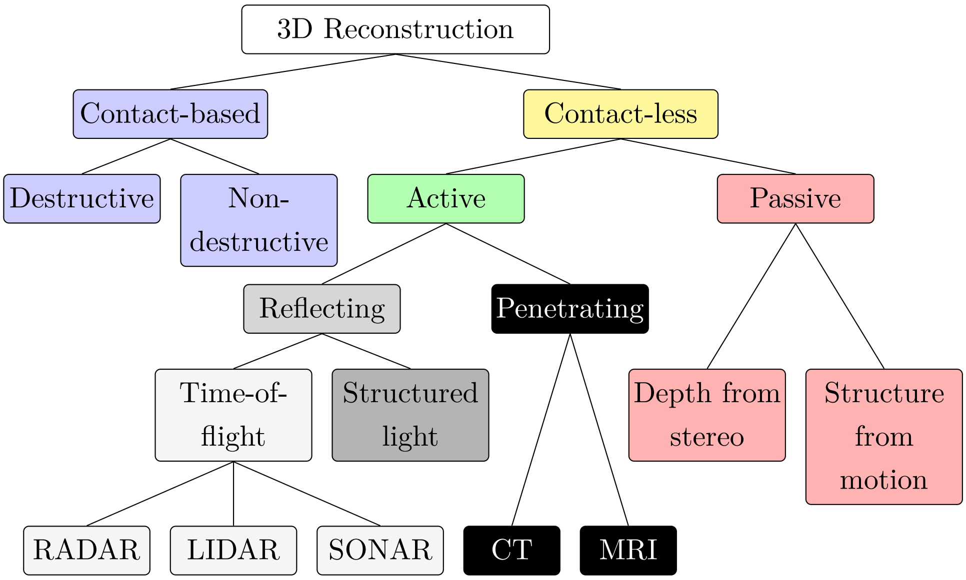

Figure 2.1: A classification of 3D reconstruction techniques

described with the epipolar geometry shown in Fig. 2.3 Top Right. If a 3D point P is found in both pictures as the 2D pixel p_L and p_R (also called correspondences, typically determined with feature detection algorithms such as SIFT/SURF), the plane generated by the three points is called epipolar plane. The line connecting the optical centres C_L and C_R of the cameras (blue) is called baseline. The intersections between the epipolar plane and the sensor plane are called epipolar lines (red). All epipolar planes for different 3D points P include the baseline. The intersection between the baseline and the epipolar lines are called epipoles e_L and e_R. Both images are then rectified as in Fig. 2.3 Bottom Left, which is basically a simulation of the ideal stereo setup previously mentioned. Notice, that all epipolar lines are parallel. The rectified images are then used to calculate a depth image such as Fig. 2.3 Bottom Right. The rectification is performed, because it eases the triangulation necessary for the depth calculation, since the 2D correspondence determination becomes a 1D correspondence search along the epipolar lines.

Structure from motion (SfM) is an extension to stereo reconstruction techniques, since multiple images from different viewpoints are typically acquired while moving the camera around an object or through the scene. Correspondences between images are used to estimate the camera path. The tracking of features and the estimation of the camera poses is a simultaneous minimisation problem and therefore not trivial to solve. Once all camera poses are estimated, depth information is again calculated using triangulation. Different approaches try to estimate the poses one by one (incremental SfM), all at once (global SfM) or different subsets after another (out-of-core SfM). Typical SfM software toolboxes are *openMVG* [Pie] (open-source), *VisualSFM* [Wu,13] (free, mostly closed-source) and *Agisoft Photoscan* [Agi] (commercial).

Active reconstruction techniques operate in different electromagnetic spectra, which are either reflected by the object or propagate through the object.

Magnetic Resonance Imaging (MRI) and *Computed Tomography* (CT) are de facto standard methods for medical diagnosis. While MRIs create a strong magnetic field around an object and detect the emissions of hydrogen nuclei in the body triggered by radio frequency pulses, CTs scan the object slice by slice using x-rays, see Fig. 2.4 Top Left. MRIs themselves generate a 3D model and the 2D CT scans are often used to generate a 3D model of the object. For detailed information on both technologies see [Han09] and [Gol07].

Time-of-flight systems use either sound waves (SONAR), radio waves (RADAR) or light (both visible and invisible) (LIDAR) to irradiate surfaces and measure the time between emission and detection of the reflection of a signal to compute depth information.
When talking about 3D reconstruction with RADAR, a typical field of application is the reconstruction of urban environments and vegetation using airborne *Synthetic Aperture Radar* (SAR) as in [Bal03] and [Kir98], although different applications for smaller objects have been explored [Coo08]. A typical 3D model acquired by RADAR is shown in Fig. 2.4 Top Right.
SONAR for 3D reconstruction is commonly used in underwater applications due to the limiting effects of water on electromagnetic waves in the visible and invisible spectra, such as absorption, diffusion and scattering. Multi-beam SONARs emit acoustic pressure waves and collect the reflected echo, which gives a 2D image. Those images are then used to reconstruct the scene in 3D as in Fig. 2.4 Bottom Left. Various publications on underwater imaging using SONAR are found in [Zer96], [Coi09] and [Son16].
LIDAR use the same principle as RADAR and SONAR, but emit pulsed laser light. Application for such sensors lies e.g. in 3D reconstruction for geodesy, archaeology, geography and forestry. An exemplary point cloud captured with a LIDAR is shown in Fig. 2.4 Bottom Right. A LIDAR was used in this work to determine the angle between two walls and compare it with the reconstruction results of structured light systems (cf. Sec. 4.2.3.

Structured light scanner are the second type of active, reflecting 3D reconstruction systems. A short survey on structured light techniques and their applications is given in the next section.

2.2 Structured Light

The projection of a specially designed and well known 2D spatially varying pattern onto a scene and the 3D reconstruction based on the distortion of the 2D pattern in an image and the extrinsics between the projector and the image sensor is known as structured light scanning, sometimes also called active stereo imaging. Over the years, multiple structured light techniques have been developed, which differ in design of the 2D projection pattern, see Fig. 2.5 Right. The simplest classification differs between multi shot and single shot imaging. Single shot techniques are able to capture dynamic scenes up to the frame rate of projector and camera, whereas multi shot techniques need a static scene and more time, according to the chosen pattern type, but often surpass single shot reconstruction in quality. The generation of a typical set for multi shot binary coded scanning is illustrated in Fig. 2.5 Left Middle. A good overview on structured light

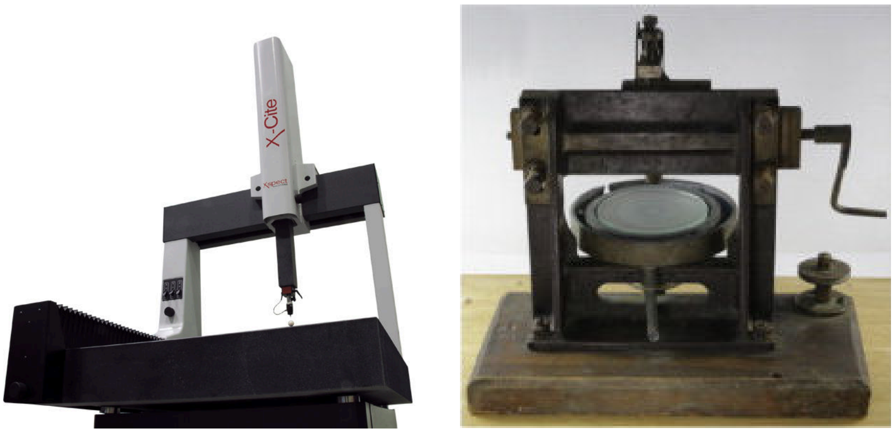

Figure 2.2: Contact-based reconstruction techniques. Left: Coordinate Measuring Machine (CMM) [WEN], Right: Grinding machine used by Sollas in the *First Paleontological Tomographic Studies* [Cun14].

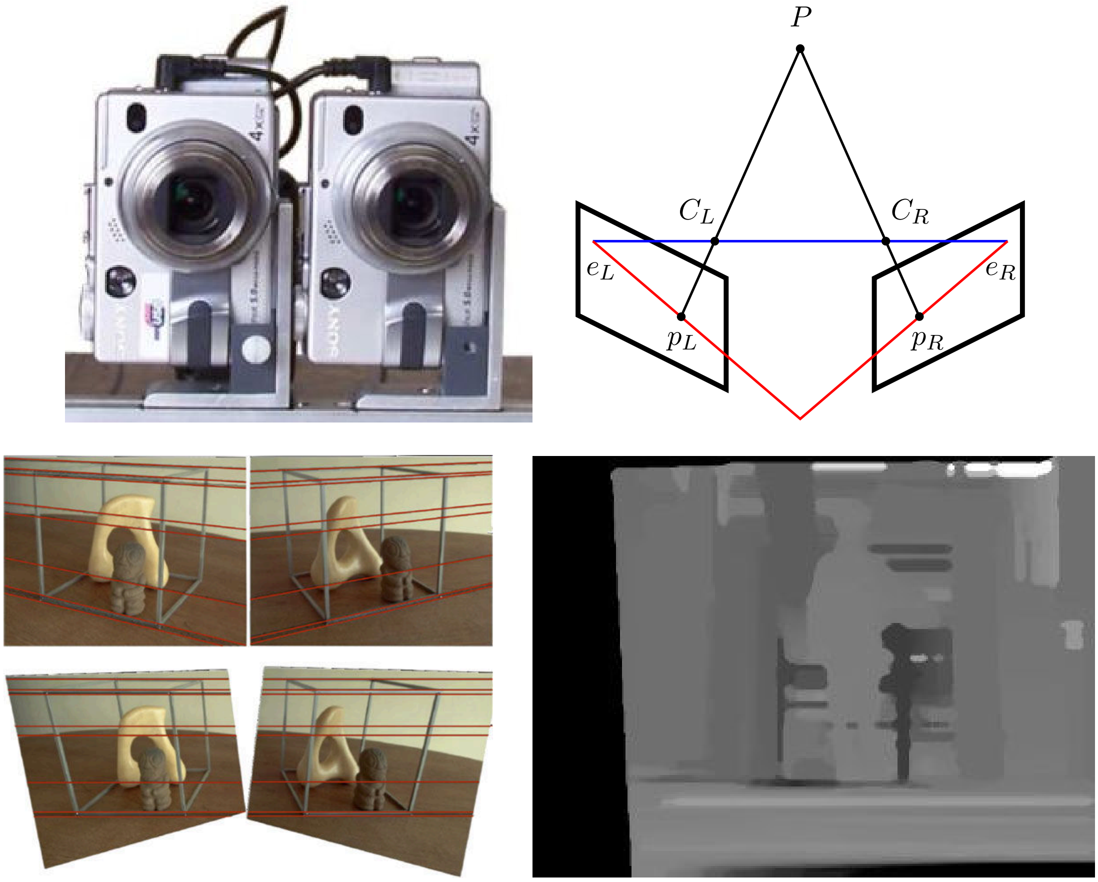

Figure 2.3: Depth from stereo. Top Left: Stereo camera rig ([Her12], p.57), Top Right: Epipolar geometry, Bottom Left: Stereo images before and after rectification, the red lines are epipolar lines [Zan16], Bottom Right: Depth image([Her12], p.57).

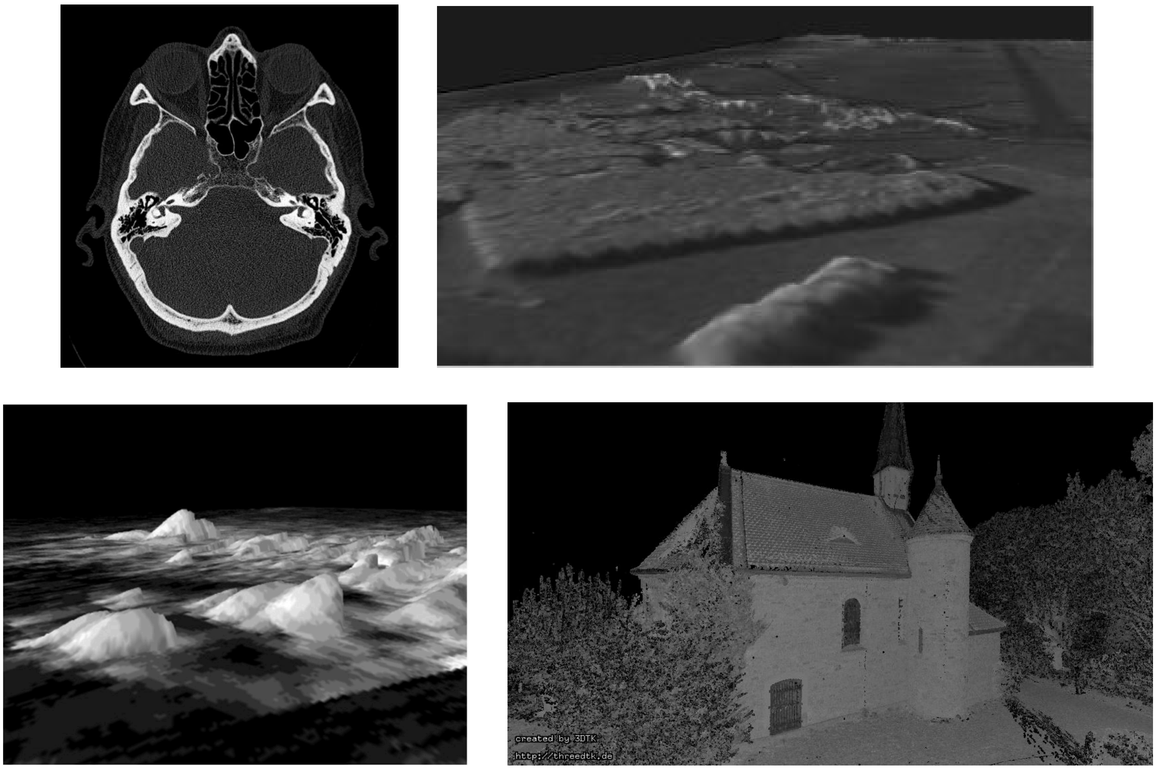

Figure 2.4: Active 3D reconstruction. Top Left: CT scan of a human scull [Hel], Top Right: 3D model generated from RADAR [Kir98], Bottom Left: 3D underwater reconstruction from SONAR [Son16], Bottom Right: 3D model of a chapel acquired with LIDAR.

techniques is given in Jason Geng's *Structured-light 3D surface imaging: a tutorial* [Jas11].
At first glance, a system that projects only a single line or a specific configuration of lines (usually different colours when using multiple lines for better discriminability) is not a classical structured light scanner, but since the plane parameters for each line need to be known for the 3D reconstruction from the deformation of said lines in the pictures, it is also classified as such. The system presented in this work uses exactly this principle. Another type of structured light techniques uses spot patterns, either arranged in lines [Mat97] or in grids [De 04] to reconstruct a scene, see Fig. 2.5 Left Bottom. Probably one of the first mentions of structured light was in Will and Pennington's *Grid Coding* article in 1971 [Wil71]. A nowadays known all-in-one structured light sensors is the aforementioned Kinect$^{\mathbf{TM}}$. Compared to the previously presented systems, the *Kinect* sensor uses a for the human eye invisible light source in the infrared spectrum. Other application for structured light are 3D perception for airbag safety in cars, cf. [Bov02] and even underwater reconstruction as in [Nar05] and [Bru11].

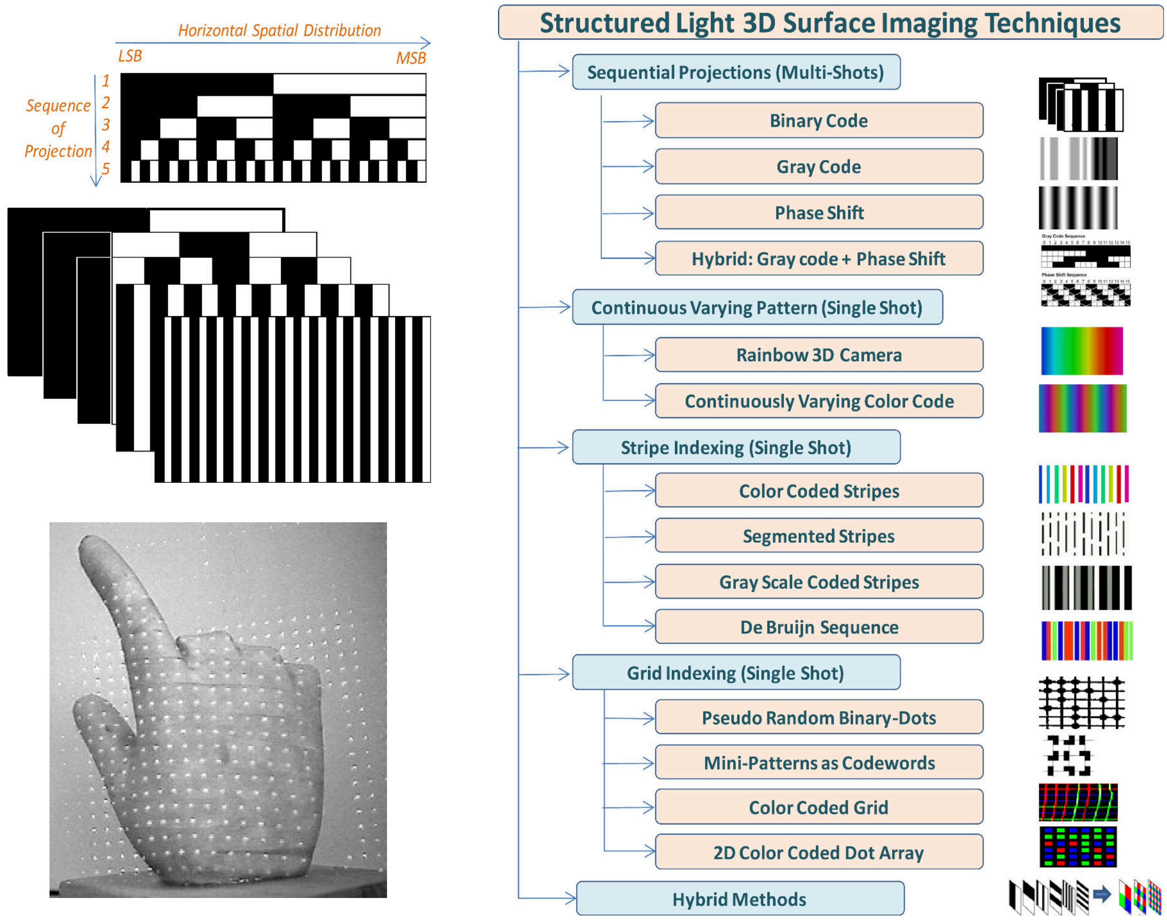

Figure 2.5: Structured light techniques. Left: Sequential binary code generation (Top) [Jas11], structured light grid spot pattern (Bottom) [De 04], Right: Classical structured light 3D surface imaging techniques [Jas11].

2.3 Self-Calibration

Calibration of structured light systems is essential for an accurate 3D reconstruction of the scene. The process typically involves either separate or combined calibration of intrinsic parameters for both camera and projector as well as extrinsic parameters. Those procedures are typically long and need to be repeated every time the physical configuration of the system is changed, as already mentioned in Sec. 1.1

A self-calibration process for structured light systems was proposed by Furukawa and Kawasaki [Fur05]. The setup consists of a calibrated camera and an uncalibrated video projector and uses uncalibrated stereo for the reconstruction. An additional laser pointer is attached to the projector to determine the scaling parameters. Correspondences between the projection and the image are determined for multiple positions of either a moving camera or a moving projector, resulting in a multi-image 3D reconstruction. With this method, it is also possible to move camera and projector in an alternating fashion, giving an even bigger field of view and therefore coverage

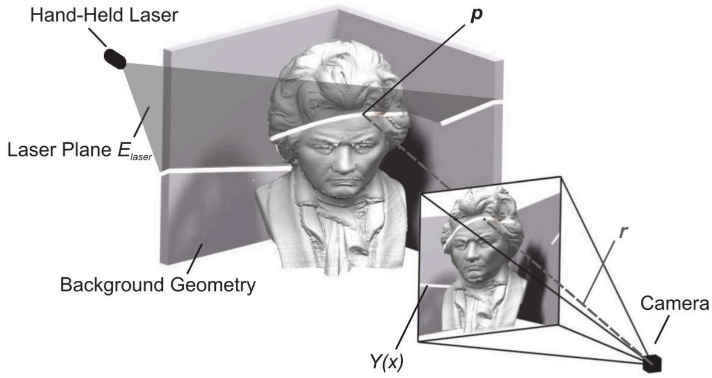

Figure 2.6: The principle of the Davidscanner [Win06].

of the scene. The method requires a precalibration of the camera and ignores the projector intrinsics.
Another method proposed by Aliga and Xu uses uncalibrated projectors and cameras to generate a multi-view 3D point cloud based on a photo-geometric approach [Ali]. First, an uncalibrated photometric stereo procedure using the projectors as a diffuse light source is performed, then a geometric modelling using the previously estimated surface, approximate lightning directions and reprojection equations is applied for self-calibration. Furthermore, the poses of the projectors with respect to the object's centre are initialised based on a uniformly-distributed subset of object points and optimised in a second step. The estimated poses are used for scale recovery. Finally, both reconstructions are combined using the high resolution of the photometric solution and the precise shape of the geometric solution.
Also worth mentioning are self-calibration techniques for laser line scanning.
Jonkinen introduced a calibration technique for a fixed setup of laser line projector and camera, which requires and initial estimation for the extrinsic parameters [Jok99]. Using multiple profile maps from different viewpoints, the estimate is refined with weighted least squares matching.
Winkelbach et al. proposed a self-calibration method for a hand-held laser line projector by placing the object in front of a corner with two known planes [Win06]. The principle of the approach is depicted in Fig. 2.6, the system became popular as the *Davidscanner* and was acquired by Hewlett-Packard (HP).
Furukawa and Kawasaki presented a method for self-calibration of a hand-held laser line projector and a fixed camera without any constraints to the geometry of the scene [Fur06]. However, a laser line projector with a known metric configuration such as known angels between the planes is required in order to perform a 3D reconstruction. The latter approach is used in this work and discussed in detail in Sec. 3.2.4.

2.4 Extraction of Curvilinear Structures

The extraction of curvilinear structures (sometimes also lines or curves) or edge detection is an important task in image processing. Various algorithms have been presented over the years, this section gives a short overview.

In general, line extraction algorithms can be divided into *pixel-precision* and *sub-pixel-precision*. For the detection of a single line, the simplest *pixel-precision* extraction method is to find the maximum value on a Gray-scale picture and use it as centre for a transversal profile. It is also possible to select the brightest pixel in each row and column and connect them to a line. These methods, however, rely on a high contrast between the desired line and the background (this can be achieved via background subtraction) and can only extract one line per image.
The Gray-gravity method (GGM) [LZ12] and the improved Gray-gravity method [LZHL17] are fast and *sub-pixel precise* line extraction approaches, but work only, if the laser line intersects each image column only once. Another simple an fast approach on line extraction was proposed in [GS95], which uses curve fitting based on an Laplacian or Gaussian Operator. This has the advantage, that no system of polynomial equations needs to be solved and the algorithm is therefore fast, but the lines are assumed to have a symmetrical profile, which is often not given. A more detailed overview on line extraction algorithms is given in [LZHL17].
This work uses Steger's line algorithm [Ste98a], since it is highly robust and accurate. Although Steger's approach is relatively slow compared to other methods, it was selected in order to have more flexibility concerning the number of lines in one image and because of the quality of the extracted lines. An introduction into the theory behind the method is given in Sec. 3.2.3 and also measures applied to speed up the original approach are presented.

Chapter 3

Theoretical and Technical Background

This chapter describes the technical details and the methodology applied in this thesis. First, the hardware components are introduced and the experimental setup is described, then the applied methodology and the theoretical models are presented and finally their implementation into the software and the subprograms are discussed in detail.

3.1 Experimental Setup

3.1.1 General Overview

The basic experimental setup is that of a typical structured light system and shown in Fig. 3.1. The scene (A) is illuminated with a projector (B) and captured with a camera (C). Both projector and camera are fixed in their position and therefore with respect to each other. Once calibrated, the system is either applicable for spatially varying structured light (e.g. Gray code or phase shift) or the simulation of line-laser-based structured light.

The proposed system simulates such a laser-based system and works as follows:

The camera is calibrated with the method presented in Sec. 3.2.2 - Camera. Determination of the plane parameters is either done implicitly via the extrinsic calibration according to Sec. 3.2.2 - Projector or explicitly using the self-calibration technique described in Sec. 3.2.4.

As discussed later on, the self-calibration requires metric constraints between some planes and a number of intersections of laser curves. To this end, the vertical lines used for the final scanning process, additional horizontal lines (for the metric constraints, cf. Sec 3.2.4) and random lines to connect the whole system are projected onto scene one after another. An overlay of a reduced set of those lines is depicted in Fig. 3.2. The intersections between lines are marked with green circles.

After successful determination of all plane equations, the actual scanning takes place using only the vertical lines. This is similar to a laser-based projector sweeping a line over the scene. For each line position, a picture is taken and the extracted line is reconstructed in 3D using the corresponding plane parameters. Additionally, the camera intrinsic parameters are necessary to

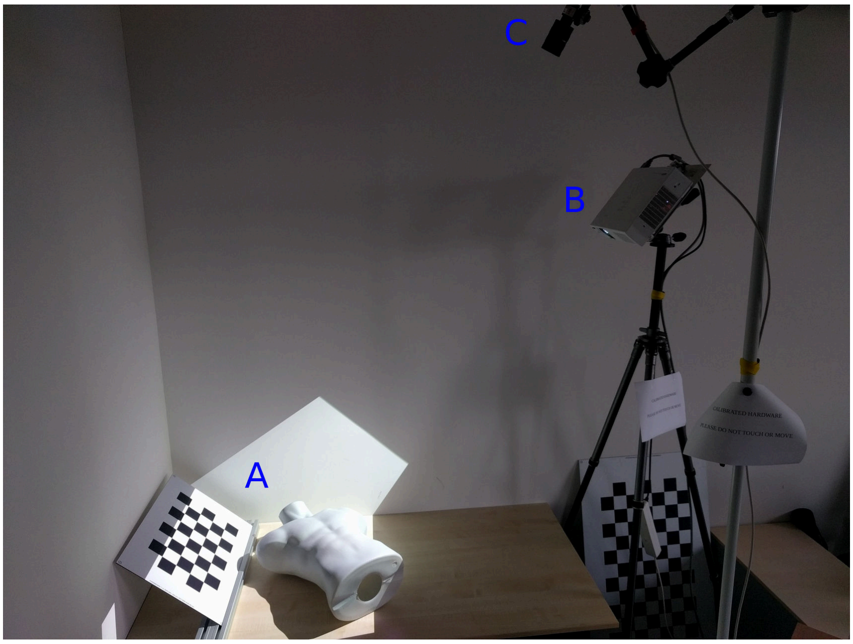

Figure 3.1: Typical setup used for this work. A: Illuminated scene, B: Projector, C: Industrial camera.

undistort the acquired image.
With this concept, the rate of the system constrained by the maximum frequency of the camera, the maximum frequency of the projector and the duration of the related computations.

3.1.2 Hardware

The camera is an IDS 3.2 MP UI-5270CP Rev.2 [IDS] with an 1/1.8" Sony IMX265 Global Shutter sensor with a maximum frame rate of 36 fps at full resolution, which is connected via Gigabit Ethernet. The lens is a Fujifilm HF16SA-1 with a focal length of 16 mm. The projector is a Sanyo PLC XU115 with a native resolution of 1024×768 at a maximum frame rate of 60 Hz. Although the usage of an off-the-shelf consumer video camera is also conceivable, industrial cameras have various advantages.
Manufacturers of industrial cameras provide an application programming interface (API), which provides definitions, protocols and tools for control of and interaction with the camera. This leads to a wider range of adjustments and manual control, which are possible via software, such that the user is able to operate the camera directly with respect to the real time requirements. Another useful feature is the selection of an area of interest (AOI). This reduces the image size, which speeds up the image processing, while increasing the maximum frame rate possible. Additionally, irrelevant parts of the image are masked out and therefore do not consume any processing power.
When using consumer cameras for live streaming, typically the USB standard 2.0 is used, which has a maximum signalling rate in high speed mode of 480 Mbit/s. Most industrial cameras use either USB 3.1 (10 Gbit/s in SuperSpeed+) or Gigabit Ethernet. While USB 3.1 is faster than

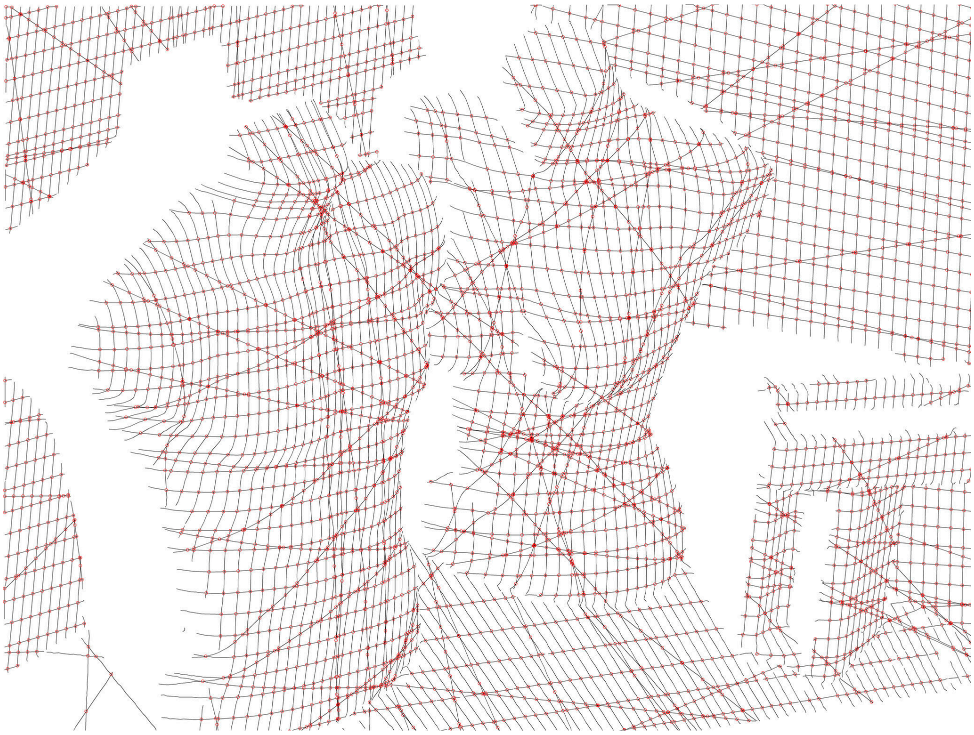

Figure 3.2: Calibration grid of extracted lines, the detected intersections are marked with green circles.

Ethernet, the maximum length of the cable is restricted to a few metres. Ethernet, on the other hand, allows cables up to 100 m and is applicable for multiple connections compared to the Point-to-Point USB connection.
Another advantage of industrial cameras is the additional trigger input. This provides the opportunity for an easy synchronisation of the camera with the pattern projector for faster and more precise picture handling. Although triggering is not used in this work, since the projector does not support it, the feature is useful for future work.

3.2 Methodology

This section introduces the applied models, approaches and algorithm used for camera, calibration and line extraction.

3.2.1 Camera Model

To describe the projection between the three dimensional world and the two dimensional image plane, a model for the camera needs to be introduced. To represent the physical camera, a set of parameters is defined in the following sections.

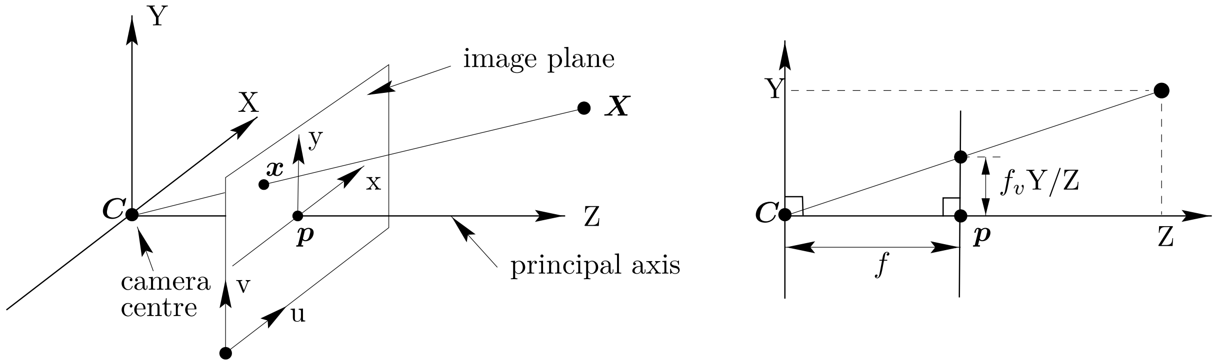

Figure 3.3: Pinhole camera model (modified from [Har04], p.154).

The Pinhole Camera

To approximate the camera projection, the pinhole camera model as introduced in [Har04] is used. A sketch of the model is depicted in Fig. 3.3. The principle axis Z is defined as the axis, which originates in the centre of the camera $\boldsymbol{C}$ and is perpendicular to the image plane. The point, where the principle axis and the image plane intersect is called principle point $\boldsymbol{p}$, the distance between the camera centre and the principle point is called focal length f. The coordinate system of the image plane is depicted with u and v. Using the intercept theorem, the projection of a point $\boldsymbol{X}$ in the world coordinate system to a point $\boldsymbol{x}$ onto the image plane can be described as a projection from three-dimensional euclidean space into two-dimensional euclidean space, $\mathbb{R}^3 \to \mathbb{R}^2$, following

$$(X, Y, Z)^T \mapsto \left(\frac{f_u X}{Z}, \frac{f_v Y}{Z}\right)^T = (x, y)^T \tag{3.1}$$

with the focal lengths f_u and f_v in u and v direction, respectively. Per definition, the focal length f is fixed. Due to imperfections of the sensor (pixels are not perfectly square) however, it is more precise to assume two different focal lengths (or scaling factors) in u and v direction. Eq. (3.1) is only valid, if the origin of the image plane and the principle point p coincide. If the origin of the image plane differs from the location of p, an additional offset must be considered, extending equation Eq. (3.1) to

$$(X, Y, Z)^T \mapsto \left(\frac{f_u X}{Z} + u_0, \frac{f_v Y}{Z} + v_0\right)^T = (u, v)^T \tag{3.2}$$

with $(u_0, v_0)^T$ as the location of the principle point with respect to the image plane origin. Expressing equation Eq. (3.2) with homogeneous coordinates yields

$$\begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \mapsto \begin{pmatrix} f_u X + Z u_0 \\ f_v Y + Z v_0 \\ Z \end{pmatrix} = \begin{bmatrix} f_u & 0 & u_0 & 0 \\ 0 & f_v & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}. \tag{3.3}$$

The matrix

$$A = \begin{bmatrix} f_u & 0 & u_0 & 0 \\ 0 & f_v & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \tag{3.4}$$

is then called *camera calibration matrix* and contains the physical parameters of the camera.

Geometrical Distortions

Due to imperfections in design and assembly of the lenses as well as physical limitations, distortions in the picture are present. These distortions result in misplaced projections of the 3D-points onto the image plane. Therefore, the pinhole camera model is extended using radial and tangential distortion parameters. Radial distortion appears, when parallel incoming rays do not intersect in the same focal point. This leads to effects called *barrel* and *pincushion* distortions, as shown if Fig. 3.4, Left. Tangential distortion is caused, if sensor and lens are not perfectly parallel assembled, leading to an offset between the principle point and the intersection between the lens normal and the sensor, see Fig. 3.4, Right. To deal with those distortions, correction parameters are introduced. Radial distortion is corrected using

$$\begin{aligned} x_{distorted} &= x(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) \\ y_{distorted} &= y(1 + k_1 r^2 + k_2 r^4 + k_3 r^6), \end{aligned}$$

where $r = \sqrt{x^2 + y^2}$ is the distance of the point to the *principle point*. Tangential distortion is corrected using

$$\begin{aligned} x_{distorted} &= x + [2p_1 xy + p_2(r^2 + 2x^2)] \\ y_{distorted} &= y + [p_1(r^2 + 2y^2) + 2p_2 xy]. \end{aligned}$$

In OpenCV [ea17], the parameters are then stored in a 1×5 matrix

$$d_{coeffs} = (k_1 \quad k_2 \quad p_1 \quad p_2 \quad k_3). \tag{3.5}$$

The *camera calibration matrix* and the radial and tangential distortion parameters are typically subsumed under the term intrinsic parameters or simply intrinsics. The determination of those parameters is described in Sec. 3.2.2 - Camera.

3.2.2 Camera and Projector Calibration

The Calibration of the camera is performed using *3DTK – The 3D Toolkit's* [N$^+$17] calibration tool, which is based on the OpenCV implementation of Zhang's method [Zha00]. For explicit calibration of the projector intrinsics and determine the extrinsic parameters for comparison with Grey code reconstruction, Moreno's method [Mor12] is applied. A short overview of the two methods follows.

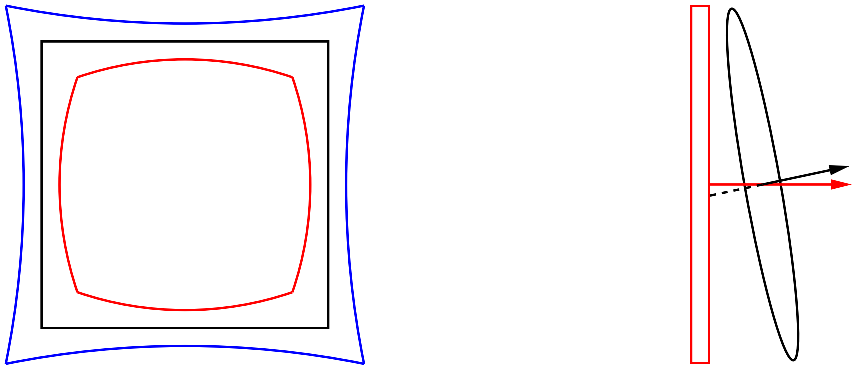

Figure 3.4: Geometrical Distortions. Left: Barrel (red) and pincushion (blue) distortion. Right: Sensor (red) and lens (black) are not parallel.

Camera

Zhang's approach is a widely adopted method for camera calibration due to its robustness and simplicity. It makes use of a slightly modified version of the previously described pinhole model with radial and tangential distortion. The camera intrinsics is given by

$$\mathbf{A} = \begin{bmatrix} f_u & \gamma & u_0 \\ 0 & f_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \tag{3.6}$$

with u_0 and v_0 as the camera's principle point with respect to the image plane origin, f_u and f_v as focal lengths (or scaling factors) in the image u and v directions and γ as the image skew.
A 3-D point $\boldsymbol{M} = [X, Y, Z]^T$ and its 2D projection $\boldsymbol{m} = [u, v]^T$ are then related as

$$s\widetilde{\boldsymbol{m}} = \mathbf{A}[\mathbf{R} \;\; \boldsymbol{t}]\widetilde{\boldsymbol{M}}, \tag{3.7}$$

with s as scaling factor, $\mathbf{R}$ and $\boldsymbol{t}$ as extrinsic rotation and translation between the camera coordinate system and the world coordinate system and $\widetilde{\boldsymbol{m}}$ and $\widetilde{\boldsymbol{M}}$ as the homogeneous vectors of $\boldsymbol{m}$ and $\boldsymbol{M}$.
It is now assumed, the model plane is on $Z = 0$ of the world coordinate system. This changes Eq. (3.7) into

$$\begin{aligned} s\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} &= \mathbf{A}\begin{bmatrix} \boldsymbol{r_1} & \boldsymbol{r_2} & \boldsymbol{r_3} & \boldsymbol{t} \end{bmatrix}\begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \\ &= \mathbf{A}\begin{bmatrix} \boldsymbol{r_1} & \boldsymbol{r_2} & \boldsymbol{t} \end{bmatrix}\begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} \end{aligned}$$

with $\boldsymbol{r_i}$ as the i^{th} column of $\mathbf{R}$. $\mathbf{A}\begin{bmatrix}\boldsymbol{r_1} & \boldsymbol{r_2} & \boldsymbol{t}\end{bmatrix}$ is called homography $\mathbf{H}$, giving

$$s\widetilde{\boldsymbol{m}} = \mathbf{H}\widetilde{\boldsymbol{M}}. \tag{3.8}$$

Denoting $\mathbf{H} = \begin{bmatrix}\boldsymbol{h_1} & \boldsymbol{h_2} & \boldsymbol{h_3}\end{bmatrix}$ as the resulting 3×3 matrix gives with Eq. (3.8) and the orthonormality constraint of $\boldsymbol{r_1}$ and $\boldsymbol{r_2}$

$$\boldsymbol{h_1}^T\mathbf{A}^{-T}\mathbf{A}^{-1}\boldsymbol{h_2} = 0 \tag{3.9}$$

$$\boldsymbol{h_1}^T\mathbf{A}^{-T}\mathbf{A}^{-1}\boldsymbol{h_1} = \boldsymbol{h_2}^T\mathbf{A}^{-T}\mathbf{A}^{-1}\boldsymbol{h_2}. \tag{3.10}$$

Eq. (3.9) and Eq. (3.10) are two constraints on an 8 DoF homography. One approach to solve this problem is analytical:

$$\mathbf{B} = \mathbf{A}^{-T}\mathbf{A}^{-1} \equiv \begin{bmatrix} B_{11} & B_{12} & B_{13} \\ B_{12} & B_{22} & B_{23} \\ B_{13} & B_{23} & B_{33} \end{bmatrix} \tag{3.11}$$

is symmetric and therefore defined by the 6D vector

$$\boldsymbol{b} = [B_{11}, B_{12}, B_{22}, B_{13}, B_{23}, B_{22}]^T. \tag{3.12}$$

Using

$$\boldsymbol{h_i}^T\mathbf{B}\boldsymbol{h_j} = \boldsymbol{v_{ij}}^T\boldsymbol{b} \tag{3.13}$$

with $\boldsymbol{h_i}$ and $\boldsymbol{h_j}$ as the i^{th} and j^{th} column vector of $\mathbf{H}$ and

$$\boldsymbol{v_{ij}} = [h_{i1}h_{j1}, h_{i1}h_{j2} + h_{i2}h_{j1}, h_{i2}h_{j2}, h_{i3}h_{j1} + h_{i1}h_{j3}, h_{i3}h_{j2} + h_{i2}h_{j3}, h_{i3}h_{j3}] \tag{3.14}$$

to rewrite Eq. (3.9) and Eq. (3.10) gives

$$\begin{bmatrix} \boldsymbol{v_{12}}^T \\ (\boldsymbol{v_{11}} - \boldsymbol{v_{11}})^T \end{bmatrix} \boldsymbol{b} = \mathbf{0} \tag{3.15}$$

for one image. Combining n images of the model plane by stacking each equation as in Eq. (3.15) gives

$$\mathbf{V}\boldsymbol{b} = 0 \tag{3.16}$$

with the $2n \times 6$ matrix $\mathbf{V}$. Eq. (3.16) has, for $n \leq 3$, a unique solution for $\boldsymbol{b}$ up to a scaling factor, which is also known as the eigenvector of $\mathbf{V}^T\mathbf{V}$ associated with the smallest eigenvalue. With the expanded version of Eq. (3.11), all parameters of $\mathbf{A}$ are then determined as

$$\begin{aligned} v_o &= (B_{12}B_{13} - B_{11}B_{23})/(B_{11}B_{22} - B_{12}^2) \\ \lambda &= B_{33} - [B_{13}^2 + v_0(B_{12}B_{13} - B_{11}B_{23})]/B_{11} \\ f_u &= \sqrt{\lambda/B_{11}} \\ f_v &= \sqrt{\lambda B_{11}/(B_{11}B_{22} - B_{12}^2)} \\ \gamma &= -B_{12}f_u^2 f_v/\lambda \\ u_0 &= \gamma v_0/f_v - B_{13}f_u^2/\lambda \end{aligned} \tag{3.17}$$

With the parameters from Eq. (3.17) and Eq. (3.8), the extrinsic parameters are calculated as

$$\begin{aligned}\boldsymbol{r_1} &= \lambda \mathbf{A}^{-1}\boldsymbol{h_1}\\ \boldsymbol{r_2} &= \lambda \mathbf{A}^{-1}\boldsymbol{h_2}\\ \boldsymbol{r_3} &= \boldsymbol{r_1} \times \boldsymbol{r_2}\\ \boldsymbol{t} &= \lambda \mathbf{A}^{-1}\boldsymbol{h_3}\end{aligned} \tag{3.18}$$

with $\lambda = 1/\|\mathbf{A}^{-1}\boldsymbol{h_1}\| = 1/\|\mathbf{A}^{-1}\boldsymbol{h_2}\|$.
Since the analytical solution is basically obtained by minimising an algebraic distance, using maximum likelihood interference can improve the result by minimising the reprojection error

$$\sum_{i=1}^{n}\sum_{j=1}^{m}\|\boldsymbol{m}_{ij} - \hat{\boldsymbol{m}}(\mathbf{A}, \mathbf{R}_i, \boldsymbol{t}_i, \boldsymbol{M}_j)\|^2 \tag{3.19}$$

where $\hat{\boldsymbol{m}}(\mathbf{A}, \mathbf{R}_i, \boldsymbol{t}_i, \boldsymbol{M}_j)$ is the projection of the 3D point $\boldsymbol{M}_j$ in image i as related in Eq. (3.7), n is the number of images and m is the number of points on the model plane. Minimising the non-linear optimisation problem in Eq. (3.19) is done using the Levenberg-Marquardt (LM) Algorithm. The solutions in Eq. (3.17) and Eq. (3.18) for intrinsics and extrinsics are typically used for the initial guess required for the LM-Solver.
The calibration of this ideal camera needs now to be extended for the geometrical distortions described in section 3.2.1. Zhang's calibration - compared to the previously introduced OpenCV implementation - considers only the first two radial parameters k_1, k_2. Denoting (x, y) and $(\widetilde{x}, \widetilde{y})$ as the undistorted and distorted normalised image coordinates, respectively, their relationship denotes as

$$\begin{aligned}\widetilde{x} &= x + x[k_1(x^2 + y^2) + k_2(x^2 + y^2)^2]\\ \widetilde{y} &= y + y[k_1(x^2 + y^2) + k_2(x^2 + y^2)^2].\end{aligned} \tag{3.20}$$

With (u, v) and $(\tilde{u}, \tilde{v})$ as the ideal and observed pixel image coordinate, respectively, and their relationship $\widetilde{u} = u_0 + f_u\widetilde{x}$ and $\widetilde{v} = v_0 + f_v\widetilde{y}$, Eq. (3.20) yields

$$\widetilde{u} = u + (u - u_0)[k_1(x^2 + y^2) + k_2(x^2 + y^2)^2] \tag{3.21}$$

$$\widetilde{v} = v + (v - v_0)[k_1(x^2 + y^2) + k_2(x^2 + y^2)^2]. \tag{3.22}$$

k_1, k_2 are estimated using the following procedure. Rewriting Eq. (3.21) and Eq. (3.22) as

$$\begin{bmatrix}(u - u_0)(x^2 + y^2) & (u - u_0)(x^2 + x^2)^2\\ (v - v_0)(x^2 + y^2) & (v - v_0)(x^2 + x^2)^2\end{bmatrix}\begin{bmatrix}k_1\\ k_2\end{bmatrix} = \begin{bmatrix}\widetilde{u} - u\\ \widetilde{v} - v\end{bmatrix} \tag{3.23}$$

gives two equations for each pixel. To estimate $\boldsymbol{k} = [k_1, k_2]^T$, the equation $\mathbf{A}\boldsymbol{k} = \boldsymbol{d}$ for $2mn$ equations (m points in n images) has the least-square-solution

$$\boldsymbol{k} = (\mathbf{D}^T\mathbf{D})^{-1}\mathbf{D}^T\boldsymbol{d}. \tag{3.24}$$

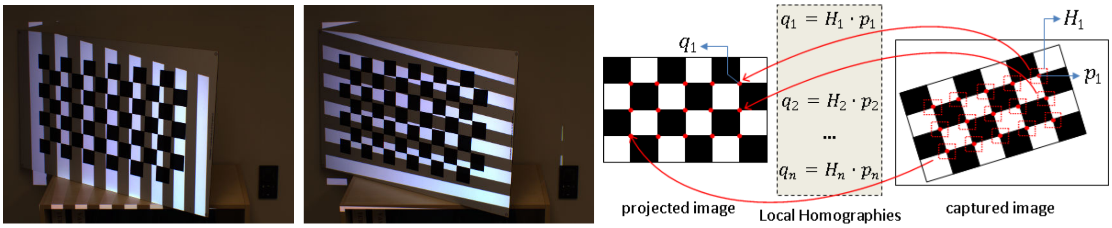

Figure 3.5: Projector-camera calibration. Sample image from Gray code sequence vertical (left) and horizontal (middle). Right: System of local homographies.

Finally, Eq. (3.19) is extended as

$$\sum_{i=1}^{n}\sum_{j=1}^{m}\|\boldsymbol{m}_{ij} - \hat{\boldsymbol{m}}(\mathbf{A}, k_1, k_2, \mathbf{R}_i, \boldsymbol{t}_i, \boldsymbol{M}_j)\|^2 \tag{3.25}$$

including the radial distortion parameters to refine the complete set of intrinsic parameters. Again, the Levenberg-Marquardt Algorithm is used for minimisation. Eq. (3.25) gives therefore a solution for the calibration of a pinhole camera with radial distortion.

Projector

Moreno's approach uses an active stereo system consisting of camera and projector. Both devices are modelled by a pinhole camera with two radial and two tangential distortion parameters as previously introduced. In order to perform the calibration, a chessboard - similar to the one used in Zhang's method - needs to be illuminated with a Gray code sequence. According to the author, the provided software requires a minimum of three datasets to perform the calibration. Giving the datasets, camera and projector intrinsics are first calibrated independent from each other, then the extrinsics are determined. Camera calibration is performed using the OpenCV camera calibration. In order to calibrate the projector like a camera with OpenCV, the chessboard corners must be translated into the projector coordinate system. To this end, the complementary Gray code sequence for both rows and columns is projected, decoded and direct and global light component are estimated. Additionally, local homographies for each chessboard corner are found considering all valid points centred around said corner, also called patch, by minimising

$$\hat{\mathbf{H}} = \underset{H}{argmin} \sum_{\forall p}\|\boldsymbol{q} - \mathbf{H}\boldsymbol{p}\|^2, \tag{3.26}$$

with $\mathbf{H} \in \mathbb{R}^{3\times3}$, $\boldsymbol{p} = [x, y, 1]^T$ as the image pixel coordinate of a point in the patch and $\boldsymbol{q} = [col, row, 1]^T$ as the decoded projector pixel for that point. A chessboard corner $\bar{\boldsymbol{p}}$ in the camera coordinate system is then translated into $\bar{\boldsymbol{q}}$ in the projector coordinate system applying the respective homography $\hat{\mathbf{H}}$ as

$$\bar{\boldsymbol{q}} = \hat{\mathbf{H}}\bar{\boldsymbol{p}}. \tag{3.27}$$

Finally, the translated chessboard corners are used to calibrate the projector intrinsics with the same procedure used for the camera.
Moreno then uses OpenCV's *stereoCalibrate()* function to calculate rotation $\mathbf{R}$ and translation

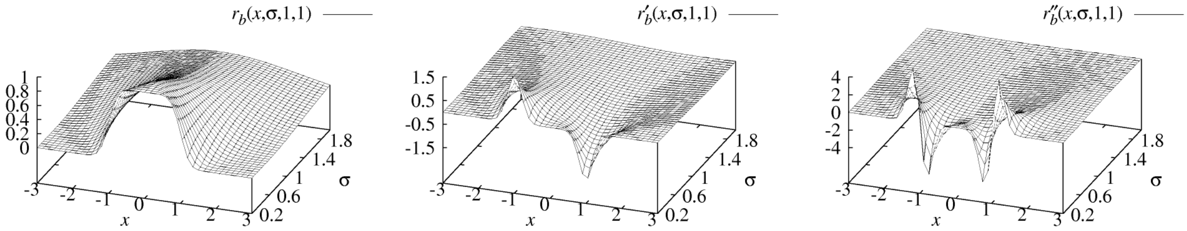

Figure 3.6: Scale-space behaviour of a bar-shaped profile f_b convolved with the Gaussian smoothing kernel for $x \in [-3, 3]$ and $\sigma \in [0.2, 2]$ [Ste98a].

$\boldsymbol{t}$ from the camera coordinate system to the projector coordinate system using the previously determined intrinsics and the chessboard parameters.

3.2.3 Extraction of Curvilinear Structures

The extraction of curvilinear structures, often simply referred to as lines, is important for all kinds of structured light applications, since it is the first of many steps and errors are carried on and amplified throughout the whole process. To this end, it is important to have a well performing and precise line extraction. As already depicted in Ch. 2, Steger's algorithm ([Ste98a]) is the de facto standard for high quality line extraction. This section introduces the algorithm and presents the features used for the implementation.

Detection of Lines in 1D

When detecting lines with height h and width $2w$, different line profiles need to be considered. The profile of a line is the distribution of intensity along a perpendicular cut to the line direction. The position of the line in 1D is then the point with the highest intensity, which is the point where the profile reaches the height h. Different profiles are symmetrical and asymmetrical bar shaped, Gaussian or parabolic. For simplicity, this introduction to Steger's line extraction algorithm focuses on symmetrical bar shaped profiles.
A function to describe the symmetrical bar shaped profile is given by

$$f_b(x) = \begin{cases} h, & |x| \leq w \\ 0, & |x| > w \end{cases} \tag{3.28}$$

Now assuming for the moment a synthetic line $z(x)$ with a parabolic profile, the position of the line can easily be detected by determining the point, where $z'(x) = 0$. An additional criterion for the selection of salient lines is $z''(x) \ll 0$ for bright lines on dark background or $z''(x) \gg 0$ for dark lines on bright background. These principles for parabolic profiles do not work for bar shaped profiles, since they do not have one distinct peak, as the derivative from Eq. (3.28) $f_b(x)' = 0$ for all x, . Additionally, the obtained lines from real images would not be accurate for any kind of line profile, since noise leads to false detections. Therefore, most image processing algorithms first remove the noise by smoothing the image using the Gaussian smoothing kernel, and then extracting the lines depending on the derivations of the image. This two step approach

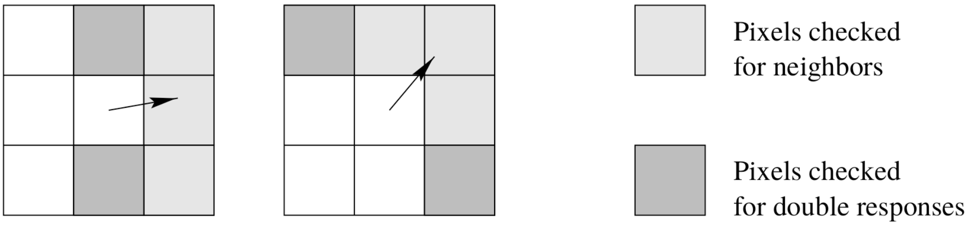

Figure 3.7: Illustration of neighbouring pixels for a given line direction [Ste98b].

is facilitated by convolving the image directly with the derivatives of the Gaussian smoothing kernel, which are given as

$$\begin{aligned} g_\sigma(x) &= \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}} \\ g'_\sigma(x) &= \frac{-x}{\sqrt{2\pi}\sigma^3} e^{-\frac{x^2}{2\sigma^2}} \\ g''_\sigma(x) &= \frac{x^2-\sigma^2}{\sqrt{2\pi}\sigma^5} e^{-\frac{x^2}{2\sigma^2}}. \end{aligned} \tag{3.29}$$

The responses for the convolution, as shown in Fig. 3.6 are then given by

$$\begin{aligned} r_b(x,\sigma,w,h) &= g_\sigma(x) * f_b(x) \\ &= h(\phi_\sigma(x+w) - \phi_\sigma(x-w)) \\ r'_b(x,\sigma,w,h) &= g'_\sigma(x) * f_b(x) \\ &= h(g_\sigma(x+w) - g_\sigma(x-w)) \\ r''_b(x,\sigma,w,h) &= g''_\sigma(x) * f_b(x) \\ &= h(g'_\sigma(x+w) - g'_\sigma(x-w)) \end{aligned} \tag{3.30}$$

with

$$\phi_\sigma(x) = \int_{-\infty}^{x} e^{-\frac{t^2}{2\sigma^2}}\,dt. \tag{3.31}$$

For the bar shaped profile, from Eq. (3.30) and Fig. 3.6 can be seen, that the first derivative vanished at $x = 0$ for all $\sigma > 0$ just as it does for a parabolic profile, but the second derivative does not have its minimum for small σ. However, $\partial/\partial\sigma(r''_b(0,\sigma,w,h)) = 0$ is true, if

$$\sigma \geq \frac{w}{\sqrt{3}} \tag{3.32}$$

and has therefore the maximum negative response for $\sigma = \frac{w}{\sqrt{3}}$. With this restriction, after the convolution of an image with the Gaussian Kernels, bar shaped lines can be detected using the same principles as for parabolic profiles.

Detection of Lines in 2D

To describe a 2D line $s(t)$, often a 1D line profile as previously introduced is used. As discussed, this profile typically vanishes for the first directional derivative and has a large absolute value for

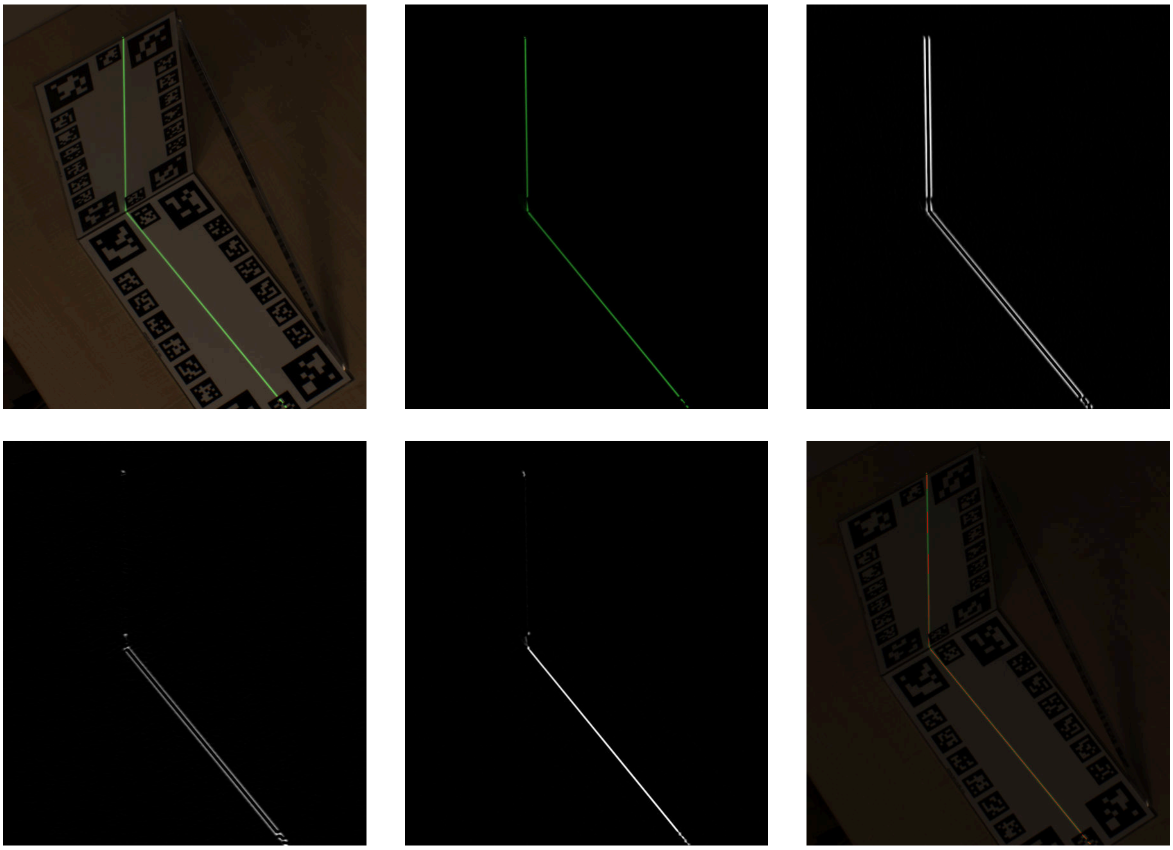

Figure 3.8: Extraction of curvilinear structures. From top left to bottom right: Input picture, picture after background subtraction, r_{xx}, r_{yy}, r_{xy}, detected points.

the second directional derivative. To compute the direction of the line locally for each point, the image $I(x, y)$ is convolved with the discrete two-dimensional Gaussian partial derivative kernels. The eigenvectors and values of the Hessian matrix

$$H(x,y) = \begin{pmatrix} \frac{\partial^2 g_\sigma(x,y)}{\partial x^2} & \frac{\partial^2 g_\sigma(x,y)}{\partial x \partial y} \\ \frac{\partial^2 g_\sigma(x,y)}{\partial x \partial y} & \frac{\partial^2 g_\sigma(x,y)}{\partial y^2} \end{pmatrix} * I(x,y) = \begin{pmatrix} r_{xx} & r_{xy} \\ r_{xy} & r_{yy} \end{pmatrix} \tag{3.33}$$

are calculated, where r_{xx}, r_{xy} and r_{yy} are the second partial derivatives. The direction perpendicular to the line $n = (n_x, n_y)$ with $\|(n_x, n_y)\|_2 = 1$ is then the eigenvector corresponding to the biggest eigenvalue. To determine whether the first directional derivative along (n_x, n_y) vanishes, a quadratic polynomial is used. The location of this point P along n is determined using a Taylor expansion, where the maximum or minimum is given by

$$(p_x, p_y) = (tn_x, tn_y) \tag{3.34}$$

with

$$t = -\frac{r_x n_x + r_y n_y}{r_{xx} n_x^2 + 2 r_{xy} n_x n_y + r_{yy} n_y^2} \tag{3.35}$$

where r_x and r_y are the first partial derivatives. In order to be a point on the line and lie within the boundaries of a pixel, $(p_x, p_y) \in [-0.5, 0.5] \times [-0.5, 0.5]$ must hold. The point is determined with sub-pixel accuracy and is only valid, if the value of the maximum absolute eigenvalue is above a threshold, characterising salient lines.
The points then need to be linked together to form lines. For tracing a line along the extracted ridge points, two thresholds for the second derivative are introduced. The upper threshold defines the minimum value for a point to be considered a starting point for a line segment, the lower threshold defines the minimum value for a point to be considered a line point. This additionally removes noise and makes sure, that only salient lines are extracted. The procedure works as follows.
The starting point is selected as the pixel with the highest second derivative. Depending on the direction of the pixel, (n_x, n_y), three neighbouring pixels are considered as as possible next point, ie. for Fig. 3.7 Left, the current pixel direction lies within $[-22.5°, 22.5°]$, for Fig. 3.7 Right, the current pixel direction lies within $[22.5°, 67.5°]$. For each neighbouring pixel, the distance $d = \|p_2 - p_1\|_2$ and the orientation difference $\beta = |\alpha_2 - \alpha_1|$ to the current pixel is calculated. The next line point is the one, that minimises $d + c\beta$, where $c = 1$ is used. To cover the whole line, this procedure is carried out in positive and negative direction of the starting point. Points are added to the line until one of three conditions is reached. Either no more neighbouring pixels are found, or a pixel is detected, which has already been assigned to another line or the remaining pixels have a second directional derivative below a user-specified (lower) threshold. In the second case, the point is marked as a junctions and the line, that contains the point, is split into two separate lines. Once a line is completed, a new starting point is selected and the routine starts over. New starting point are selected, as long as the second directional derivative lies above another user-selected (upper) threshold. Finally, the line normals at all points are oriented in a way, such that all normals point to the right side of the line. The right side of the line is determined by orienting the normal of the starting point to the right with respect to the image. Double responses as in Fig.3.7 (e.g. if another not yet processed line intersects almost perpendicular with the current line) are recognised by comparing the direction of both points with the normal direction of the current point. If they have roughly the same orientation, the points are marked as processed and do not influence the current line.
Since for this approach, exactly one line is projected per image, no junctions are found and the procedure is only illustrated for the sake of completeness. In order to make the desired lines more visible, background subtraction is applied to every image. The extracted points are further on filtered comparing the colour information based on HSV colour space with predefined colours from a look-up table for speed improvement. This additionally reduces noise and makes it easy to simply extend the approach for a projection of multiple, different coloured lines per image while making the lines discriminable.
Since the convolution of the image with the derivatives of the Gaussian kernel is computationally expensive and also increases for higher resolutions, the image is divided into kernel-size patches and convolved if and only if the patch contains at least one point which fulfils the previously specified colour segmentation to reduce computational time. The laser line extraction is illustrated in Fig. 3.8, showing the input image, the image after background subtraction, the second directional derivatives for the Hessian matrix and the detected line points.

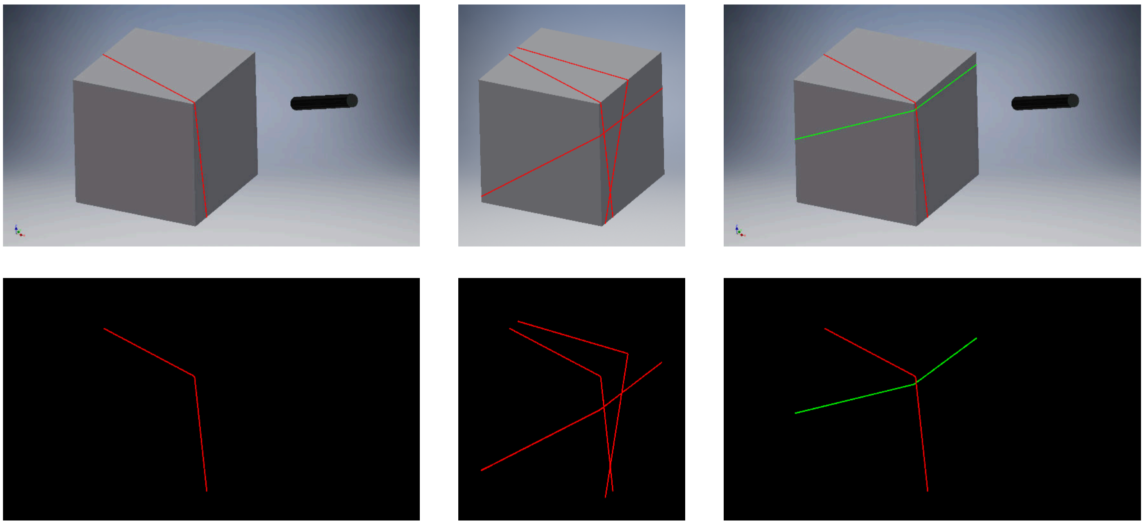

Figure 3.9: Configuration of active vision system. Top: Single laser plane (Left), multiple consecutive laser planes overlayed (middle) and crosshair-laser configuration (right). Bottom: Detected laser curve polylines.

3.2.4 Plane Parameter Estimation

Furukawa and Kawasaki proposed a method to estimate the plane parameters for an active vision system (laser line projector and precalibrated camera) in [Fur06] and later extended the approach such that the intrinsic parameters of the camera are additionally estimated [Fur09]. The implementation for this work follows the approach presented in the second paper, but uses a pre-calibrated camera.
Starting with a single laser line, the problem itself is described as follows:
The laser line projector is moved in 3D space, illuminating the scene with respect to a fixed camera. The plane of one laser line (or curve) π as in Fig. 3.9 (Left) can be represented as

$$\pi : aX + bY + cZ + 1 = 0. \tag{3.36}$$

Using the projective pinhole camera model from Eq. (3.2) and rearranging the terms yields

$$\pi : a\left(\frac{x - u_0}{f_u}\right) + b\left(\frac{y - v_0}{f_y}\right) + c = -\frac{1}{Z}. \tag{3.37}$$

Therefore, with known plane parameters $(a, b, c)^T$, the camera intrinsics matrix and the 2D coordinate $(x, y)^T$, the corresponding 3D point can easily be calculated as

$$\begin{aligned} Z &= -\frac{1}{a\left(\frac{x-u_0}{f_u}\right) + b\left(\frac{y-v_0}{f_y}\right) + c} \\ X &= Z\frac{x - u_0}{f_u} \\ Y &= Z\frac{y - v_0}{f_v}, \end{aligned} \tag{3.38}$$

with the Z coordinate is negative, since the Z-axis is assumed to be directed backwards from the camera. Eq. (3.38) are commonly known as light sectioning equations.
For two curves π_i and π_j with the intersection point $(x_{ij}, y_{ij})^T$ as in Fig. 3.9 (Right), combining the respective Eq. (3.37) as

$$a_i\left(\frac{x_{ij}-u_0}{f_u}\right)+b_i\left(\frac{y_{ij}-v_0}{f_y}\right)+c_i=-\frac{1}{Z}=a_j\left(\frac{x_{ij}-u_0}{f_u}\right)+b_j\left(\frac{y_{ij}-v_0}{f_y}\right)+c_j \tag{3.39}$$

yields

$$(a_i-a_j)\left(\frac{x_{ij}-u_0}{f_u}\right)+(b_i-b_j)\left(\frac{y_{ij}-v_0}{f_y}\right)+(c_i-c_j)=0. \tag{3.40}$$

Eq. (3.40) is homogeneous and contains the differences of the plane parameters. Therefore, both set of plane parameters have the same indeterminacies, a scalar s and an offset vector $\boldsymbol{o}$. Two equations can be found as

$$\boldsymbol{a_i}=(a_i,b_i,c_i)^T=s(a'_i,b'_i,c'_i)^T+\boldsymbol{o} \quad \text{and} \quad \boldsymbol{a_j}=(a_j,b_j,c_j)^T=s(a'_j,b'_j,c'_j)^T+\boldsymbol{o}, \tag{3.41}$$

where $\boldsymbol{a'_i}=(a'_i,b'_i,c'_i)^T$ and $\boldsymbol{a'_j}=(a'_j,b'_j,c'_j)^T$ are solutions for $\boldsymbol{a_i}$ and $\boldsymbol{a_j}$ up to scale. Given N curves with M intersections, all equations such as Eq. (3.41) are combined in a homogeneous linear system as

$$\mathbf{L}\boldsymbol{p}=0 \tag{3.42}$$

with the 3N-dimensional vector

$$\boldsymbol{p}=s\boldsymbol{A}+\boldsymbol{O}, \tag{3.43}$$

with $\boldsymbol{A}=(\boldsymbol{a'_1},\dots,\boldsymbol{a'_N})$ and $O=(\boldsymbol{o},\dots,\boldsymbol{o})$ as in Eq. (3.41), and the M×3N matrix $\mathbf{L}$ containing $(\pm x_{ij}-u_0/f_u)$, $(\pm y_{ij}-v_0/f_v)$ and (± 1) at the appropriate positions to form the corresponding homogeneous linear equations as Eq. (3.40). The solution for $\boldsymbol{p_i}$ ($0 \leq$ i $\leq$ N) is found as

$$\boldsymbol{p_i}=s(a'_i,b'_i,c'_i)+\boldsymbol{o} \tag{3.44}$$

having 4-DOF indeterminacy with a scalar s and an offset vector $\boldsymbol{o}$, if the system is solvable and has no degenerate conditions ($\geq$ 4-DOF indeterminacy, i.e. if all intersections for one curve a collinear) and is called *projective solution.* A trivial solution for $\boldsymbol{p}$ is obviously the zero vector, so the system Eq. (3.42) is solved using Singular Value Decomposition (SVD) under the constraint $\|\boldsymbol{p}\|=1$. It is worth mentioning, that the 4-DOF indeterminacy of this general solution can be described as a 4 parameter homography, that transforms the 3D points and plane parameters.
To estimate the plane parameters up to scale, a metric constraint needs to be obtained. The simplest modification is to introduce orthogonality. Although orthogonality in the scene (e.g. orthogonal walls) would be sufficient as a metric constraint (cf. Sec. 2.3, Davidscanner), the method should be applicable to scenes regardless of their geometrical features. Therefore, a crosshair-laser-line configuration as in Fig. 3.9 (Right) is used.
Given a set of pairs of orthogonal planes, an error function for each pair i, j from the set of all

$C_v = (ij)|(\pi_i \perp \pi_j)$ is defined as

$$E(\boldsymbol{o}) = \sum_{(ij)\in C_v} \cos^2 \theta_{ij}(\boldsymbol{o}) \tag{3.45}$$

$$= \sum_{(ij)\in C_v} N(a_i, b_i, c_i, \boldsymbol{o})^T N(a_j, b_j, c_j, \boldsymbol{o})^2, \tag{3.46}$$

where θ_{ij} is the angle between the two planes and $N(\cdot)$ is the normal of the plane computed using plane parameters and the offset vector. A vector $\hat{\boldsymbol{o}}$ is found using non-linear optimisation, which minimises the error function as

$$\hat{\boldsymbol{o}} = \underset{\boldsymbol{o}}{argmin} E(\boldsymbol{o}). \tag{3.47}$$

This solution is called *metric solution* and calibrates the system up to scale.
As already mentioned, using a crosshair-laser-line configuration, it is not possible to recover scale. By adding an additional line-laser, which projects a plane parallel to one of the already existent planes and acquiring a set of parallel planes, the error function for each pair of parallel planes from the set $C_p = (kl)|(\pi_k \parallel \pi_l)$ is extended as

$$E(\boldsymbol{o}) = \sum_{(ij)\in C_v} \cos^2 \theta_{ij}(\boldsymbol{o}) + \sum_{(kl)\in C_p} \sin^2 \theta_{kl}(\boldsymbol{o}) \tag{3.48}$$

$$= \sum_{(ij)\in C_v} N(a_i, b_i, c_i, \boldsymbol{o})^T N(a_j, b_j, c_j, \boldsymbol{o})^2 \tag{3.49}$$

$$+ \sum_{(kl)\in C_p} \|N(a_k, b_k, c_k, \boldsymbol{o})^T N(a_l, b_l, c_l, \boldsymbol{o})\|^2. \tag{3.50}$$

However, this parallel constraint cannot be achieved with the setup used in this work and is therefore just noted for the sake of completeness.

3.3 Software

To simplify the operation of the developed software and show real time data, a graphical user interface (GUI) as shown in Fig. 3.10 was designed using Qt 5 [Qt]. This section gives a short introduction into the functionality of the developed software and introduces the most important external libraries.

3.3.1 Functionality

The top row contains two buttons to connect to and disconnect from the camera, a list which lists all the available cameras and a refresh button to update the list. The main window is divided into two columns. The left column contains three tabs to control camera, projector and the reconstruction. The right column contains two tabs to show the live image and the reconstructed point cloud. Detailed views of the software are found in Appendix A.1.
If a camera is connected, the live video stream is displayed on the right side. The camera tab

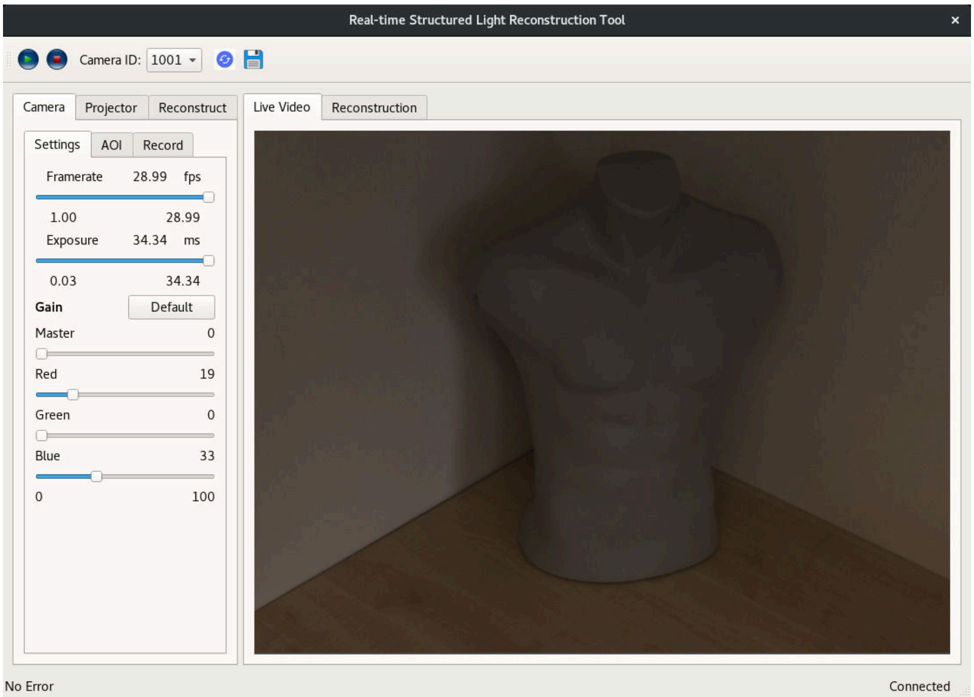

Figure 3.10: Interface of the *Real-time Structured Light Reconstruction Tool*

is divided into three sections. The first window provides slider for the adjustment of frame rate and exposure as well as the control of the master gain and colour adjustment (Fig. A.1).
The area of interest (AOI) is selectable via the second window, after reducing width or height, the x or y position of the selected window can be changed with respect to the upper left corner (Fig. A.2).
Although the camera provides additional functionalities such as different image encodings, changing of the pixel clock and automatic setting of white balance, brightness and contrast, only the very basic settings were implemented in this work, since they are sufficient for the application. It is worth mentioning, that all automatic adjustments are disabled during the initialisation of the camera in order to have a steady and comparable image quality.
The last window concerning the camera provides tools for video and image recording, as well as a live line extraction overlay to tune the parameters of the laser line extractor (Fig. A.3).
The projector tab provides various options to project different pattern, such as a static vertical line (Fig. A.4), a randomly translating and rotating cross (Fig. A.5) and a Lissajous figure (Fig. A.6). Additionally, the line width as well as the projector update frequency are adjustable.
The reconstruction tab provides the controls for both pre-calibrated and self-calibration reconstruction.
When using an intrinsic and extrinsic pre-calibration determined with Moreno's software, the file first needs to be loaded, the plane equations are determined and the scene can be reconstructed using a vertical sweeping line. Additionally, a set of Gray coded images can be acquired (Fig. A.7) and later on used for reconstruction using pre-calibrated line scanning (Fig. A.8).

The self-calibration tab is of similar design. First, the camera calibration needs to be loaded. Second, the desired lines must be acquired (Fig. A.9) and the calibration needs to be performed, or the plane equations can be loaded from a previous calibration. Finally, the reconstruction works as for the pre-calibration.
During the scanning process, the lines are extracted, reconstructed and coloured in real time. The resulting point cloud is continuously updated and displayed using the PCLVisualizer. Once the whole scene has been reconstructed, the point cloud is saved in a file for further analysis and processing.

3.3.2 External Libraries

Aside from Qt for GUI design, the software makes use of different well known mathematical and computer vision libraries. This section introduces and shortly describes the most important libraries utilised and explains their application in the software.

Ceres

The Ceres Solver [AM+] is a library provided by Google for modelling and solving of optimisation problems. For non-linear least-square problems, Ceres provides Levenberg-Marquardt, Powell's Dogled and Subspace dogled methods. To reduce errors before starting the solver, Ceres is able to automatically and numerically differentiate the input. A wide range of companies use Ceres in their software, such as Google itself in Street View for pose estimation and panorama generation in Android, Willow Garage for the solving of SLAM problems and the SfM tool OpenMVG [Pie] for bundle adjustment, to name just a few. In this work, Ceres is used to calculate the *metric solution* described in Sec. 3.1.

Eigen

Eigen [GJ+10] is a C++ template library for linear algebra. It includes headers for matrix and vector operations, various linear solvers and algorithms. Eigen serves as library for many projects such as the aforementioned Ceres Solver, the Space Trajectory Analysis project at ESA, the Robotic Operating System (ROS) by Willow Garage and the Point Cloud Library (PCL). This thesis uses Eigen's Bidiagonal Divide and Conquer SVD for solving the projective reconstruction problem described in Sec. 3.1.

IDS Software Suite

One advantage of industrial cameras compared to consumer cameras is, as already mentioned in Sec. 3.1.2, their API. The *IDS Software Suite* includes, aside from the API, also demo and configuration applications. The *IDS Camera Manager* (3.11, Left) is a tool to manage and configure all connected cameras. It also displays information such as device ID and serial number. The *uEye Demo* (3.11, Right) is a sample program to demonstrate the functionality of the camera.

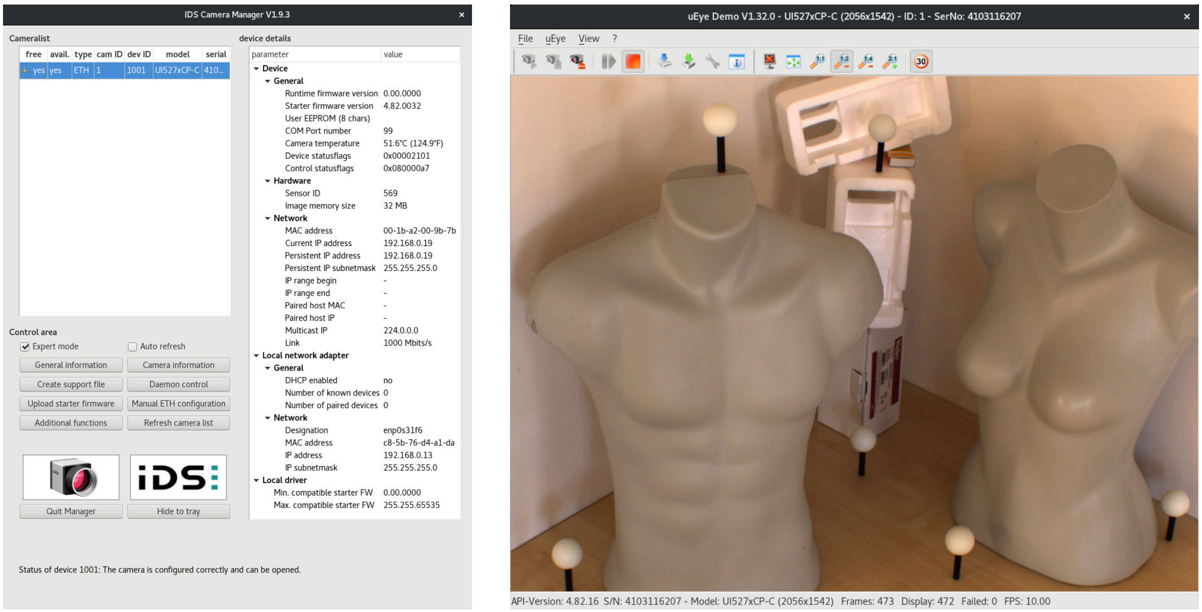

Figure 3.11: IDS software suite. IDS camera manager (left), uEye cockpit (right)

OpenCV

The Open Source Computer Vision Library (OpenCV) aims at real-time computer vision and machine learning. Due to the cross platform (Windows, Linux, Max OS, Android) and language (C, C++, Python, Matlab, Java) support, it is one of the most employed libraries in this area. This work uses OpenCV's image processing tools for image handling, camera calibration (adapted in 3DTK's calibration tool) and intrinsic parameter utilisation. Moreno's structured light software (cf. Sec. 3.2.2 - Projector) additionally relies on the stereo calibration tool for extrinsic calibration.

PCL

"The Point Cloud Library (or PCL) is a large scale, open project for 2D/3D image and point cloud processing. The PCL framework contains numerous state-of-the art algorithms including filtering, feature estimation, surface reconstruction, registration, model fitting and segmentation. These algorithms can be used, for example, to filter outliers from noisy data, stitch 3D point clouds together, segment relevant parts of a scene, extract keypoints and compute descriptors to recognize objects in the world based on their geometric appearance, and create surfaces from point clouds and visualize them – to name a few." [PCL]
As already mentioned, the PCLVisualizer is used for the live reconstruction, PCL itself provides the necessary interface for a communication and data exchange between the internal point cloud reconstruction and the PCLVisualizer.

3.3.3 Additional Software

3DTK - The 3D Toolkit

"The 3D Toolkit provides algorithms and methods to process 3D point clouds. It includes automatic high-accurate registration (6D simultaneous localisation and mapping, 6D SLAM) and other tools, e.g., a fast 3D viewer, plane extraction software, etc. Several file formats for the point clouds are natively supported, new formats can be implemented easily." [N+17]
In this work, 3DTK is used for camera calibration and point cloud viewing.

CloudCompare

"CloudCompare is a 3D point cloud (and triangular mesh) processing software. It has been originally designed to perform comparison between two dense 3D points clouds (such as the ones acquired with a laser scanner) or between a point cloud and a triangular mesh. It relies on a specific octree structure dedicated to this task. Afterwards, it has been extended to a more generic point cloud processing software, including many advanced algorithms (registration, re-sampling, colour/normal/scalar fields handling, statistics computation, sensor management, interactive or automatic segmentation, display enhancement, etc.)." [Clo17]
In this work, CloudCompare is used for point cloud viewing, registration, plane and sphere fitting and statistical analysis of the acquired point clouds.

Chapter 4

Experiments and Discussion

In order to evaluate the performance of the system, a number of tests is conducted. This section introduces the applied experiments and presents and discusses their evaluation. First, the reconstruction methods applied are presented. A qualitative analysis compares the Gray code reconstruction with the line sweeping. In a quantitative analysis, the plane parameters are evaluated using a explicit determination obtained with a 3D calibration fixture. Shape reconstruction is compared to objects of known size and two reconstructed planes are compared to a ground truth. Finally, the results of the self-calibration approach are discussed and potential improvements are proposed.

4.1 Qualitative Analysis

To get a first impression of the achievable results, a qualitative analysis is performed. To this end, a scene is reconstructed using three approaches based on spatially varying Gray code, pre-calibrated line sweeping and self calibrated line sweeping. A picture of the scene used for evaluation and self-calibration is shown in Fig. 4.1.

The **Gray code** reconstruction is performed using the software provided alongside Moreno's paper, which was previously discussed (cf. Sec. 3.2.2 - Projector).
After calibration of the system, 42 images of the scene need to be acquired to obtain a point cloud. The number of necessary images depends on the the resolution of the projector. Since the projector used in this work has a width resolution of $2^{10} = 1024$ pixel, a 10-bit Gray code is used. The pixel columns of the projector are consecutively illuminated in an alternating fashion (non-inverted and inverted) according to the column values of the Gray code. The same procedure is done for the projector pixel rows. This gives 20 images with horizontal and 20 images with vertical coding (for a visualisation of how to generate the patterns from a Gray code sequence, cf. Fig. 2.5, Top Left). Additionally, one completely illuminated image and one image under natural lighting conditions are taken in order to estimate the direct and global light components.
The result shown in Fig. 4.2, Left, is a coloured point cloud containing approximately 400000 points. The theoretical maximum number of points achievable with this method is precisely

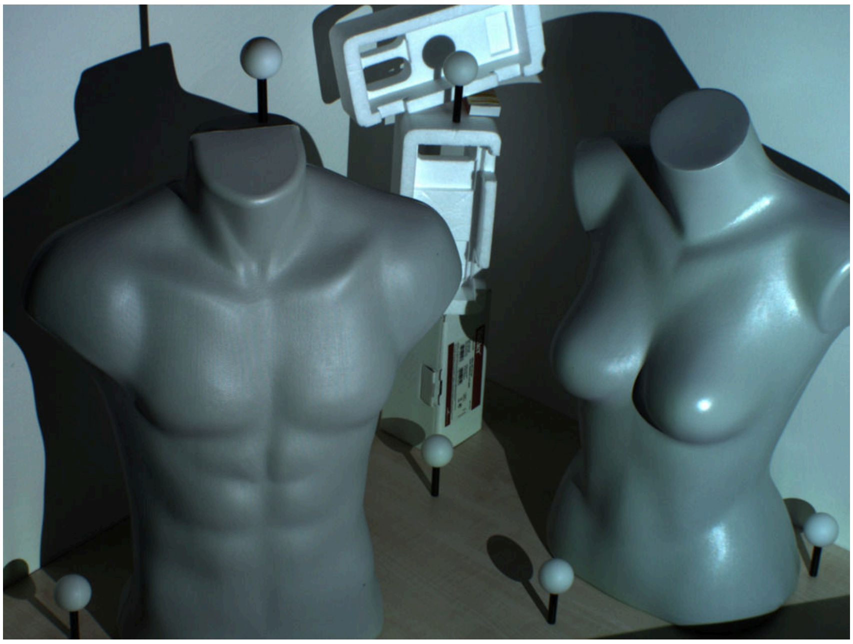

Figure 4.1: Image of the scene used for reconstruction.

786432 (1024 x 768), if every pixel of the projector is visible in the image and decodable. Although noise is present due to pattern decoding errors, the quality of the point cloud itself is satisfying. The geometry of the scene is completely determined including scale and considers the distortion effects of both projector and camera.

The **pre-calibrated sweeping line** method is based on the same calibration as the first method. The lines are produced by illuminating the respective pixel columns of the projector, which simplifies the reproducibility of each line position. The plane equations are determined in a two step process. First, each plane is spanned by two vectors pointing from the center of the projector to the virtual top and bottom pixel of the current column using an inverted pinhole camera model. The normalised cross products of those vectors, i.e. respective the plane parameters in the projector coordinate system, are then transformed into the camera coordinate system using the extrinsic calibration. Compared to Moreno's software, for reasons of simplicity this technique does not consider the projector distortion. The occurring effects will be discussed later in this section. The reconstructed point cloud is shown in Fig. 4.2, Right, and contains around 1.4 million points. Compared to the upper point limit of the Gray code method, the maximum number for this approach is limited by the resolution of the camera, which is 3.17 million pixel. This is due to the fact, that the lines are extracted in the higher resolving camera image and therefore interpolated regarding the projector resolution.
The point cloud contains less outliers (compare Fig. A.10 and A.11), but does not reproduce some areas at all, such as the right part of the belly of the female torso or parts of the bottom surface in the lower left corner. This is, however, not a problem of the reconstruction method itself, but rather a matter of the adjustment of the line extraction thresholds. Due to the low

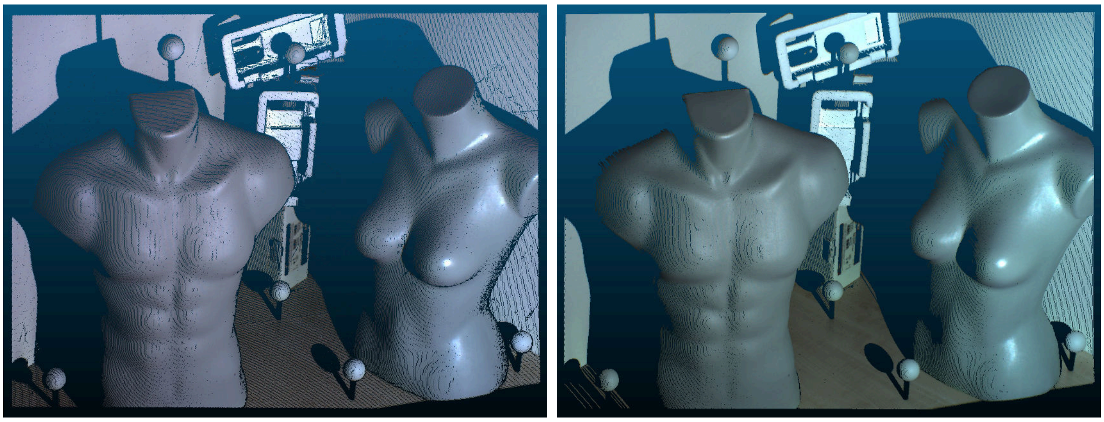

Figure 4.2: Scene reconstruction. Left: Using spatially varying Gray code. Right: Using pre-calibrated line scanning, for bigger images see A.10 and A.11.

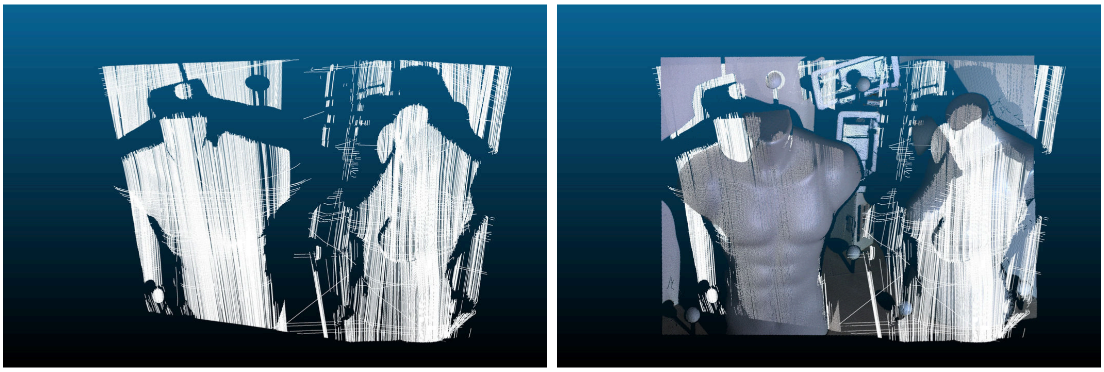

Figure 4.3: Scene reconstruction using self calibrating line scanning. Left: Resulting point cloud, Right: Comparison with Gray code reconstruction.

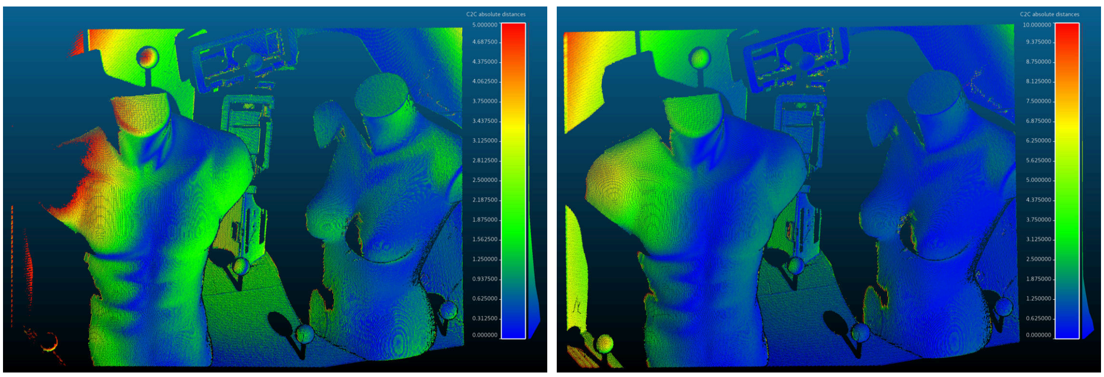

Figure 4.4: Point-to-point distance of Gray code reconstruction and line scanning, the error is given in mm.

RMS [mm]	$\bar{d}$ [mm]	σ [mm]
1.700	1.762	3.201

Table 4.1: Scene Registration Results. The values are computed from the registration of the Gray code reconstruction with the line scanning point cloud. Mean distance $\bar{d}$ and standard deviation σ are computed from the point-to-point error.

intensity of the projected lines, some projections cannot be reconstructed without introducing too much noise into the clearly visible lines. Therefore, the line extraction threshold need to be adjusted carefully.

The **self-calibrating sweeping line** method works the same as the prior technique, but utilises the self calibration technique described in Sec. 3.2.4. As shown in Fig. 4.3, the scene is reconstructed, but the geometry of the scene is distorted. Due to the metric inaccuracies, automatic scaling of the point cloud using fiducial markers is not possible and was done by hand for Fig. 4.3, Right. The performance of the self-calibration as well as potential improvements are discussed in Sec. 4.3.

Due to the errors introduced by the self-calibration technique, for further analysis and evaluation of the system, only the first two approaches are considered.
In order to compare the quality of the whole scene, the Gray code reconstruction is registered with the line scanning. The results of the registration are shown in Tab. 4.1. Fig. 4.4 shows the Gray code reconstruction, coloured with the absolute error to the line scanning. Points with an error of over 5 mm (Left) and 10 mm (Right) are not visible. Regarding all 430,057 points, 2,009 points have an error of over 10 mm, which is 0.4671 %. Although the error values are absolute numbers, the comparison is still qualitative, since neither reconstruction can be considered as ground truth. However, the RMS of the registration of 1.700 mm is an indicator, that the two reconstructions are of comparable quality in wide areas of the point cloud. The mean distance of all point-to-point errors 1.762 mm confirms this observation. Note, that for the registration and therefore computation of the RMS, only a subset of all points is used. The point-to-point distance, and so $\hat{d}$ and σ, are computed for all points. The distribution of the errors with respect to the position shows, that the omission of the projector distortion appears to have an influence on the reconstruction with the line scanning, particularly in the left part of the scene. This also explains the relative high standard deviation of the error distribution of 3.201 mm.
Apart from these disparities, the overall quality of the reconstruction methods is comparable and does, except for the edge regions of the projection, hardly differ. The density of the point cloud obtained with the line sweeping is higher, due to the previously mentioned restrictions.

4.2 Quantitative Comparison of Reconstruction Techniques

A quantitative comparison of the reconstruction is performed by evaluating the determined plane parameters, scanning objects with a precisely known shape and comparing reconstructed planes to a ground truth. To this end, a 3D calibration fixture is used for plane parameter evaluation,

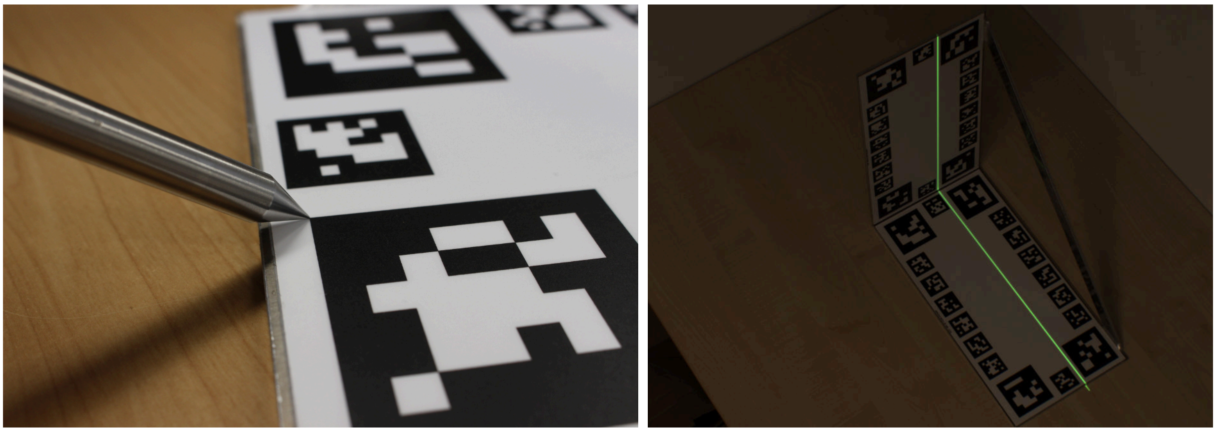

Figure 4.5: Explicit plane parameter determination using a 2D calibration fixture.

two types of specimen are used for metric evaluation. The influence of the projector distortion is analysed, a precise 3D model of the experimental environment is obtained and the results are discussed.

4.2.1 Plane Parameter Evaluation

To evaluate the precision of the plane parameters determined with the pre-calibration, a two dimensional calibration fixture, as mentioned in the introduction, is used. The calibration pattern consists of two perpendicular planes, each containing an arrangement of AprilTag fiducial markers [Ols11]. AprilTags are an alternative to chessboards as fiducial markers for calibration and plane estimation, which provide the advantage, that not all markers must be visible during the capturing, since the tags are detected independent from each other. The position of the AprilTags is determined using a mechanical probe, mounted on a Kuka KR16 industrial manipulator (repeatability ± 0.04 mm), see Fig. 4.5, Left. With the three dimensional positions of the AprilTags relative to each other, the plane parameters for both pattern planes are determined from the image with high accuracy.
For each projected light curve, an image is taken where the line intersects most of the white area of the calibration pattern (Fig. 4.5, Right). The determination of the plane parameters is performed by means of the extracted line points and the calibration fixture planes applying the same principle used for the already mentioned Davidscanner [Win06]. For each set of plane parameters determined with pre-calibration and the explicit calibration, the angular difference between the plane normals and the error of the difference of the distance from plane to origin is calculated. The results are plotted in Fig. 4.6.
The absolute error of the distance to the origin is less than 0.7 mm, yielding with respect to the distance a maximum deviation of 0.1 %. The angular difference of the plane normals in general lies below 0.55 %. However, it is noticeable, that for the three leftmost planes, the angular difference is large compared to the rest. This supports the previous observation made during the qualitative analysis, that the projector distortion has a bigger influence on the scene reconstruction in the left part of the scene.

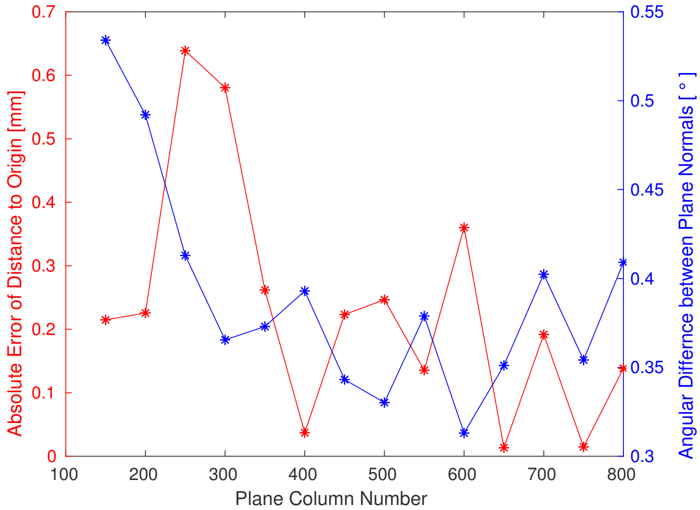

Figure 4.6: Plane parameter evaluation using a 2D calibration fixture.

4.2.2 Scanning of Objects with a Known Shape

After the evaluation of the plane parameters has shown, that the precision is mostly influenced by the omission of the projector distortion, objects with a known shape are reconstructed using Gray code reconstruction and line scanning. Two different types of objects with a known shape are selected. The results are compared and discussed.

Table Tennis Balls

To guarantee fair and equal conditions for all competitions, table tennis balls are subjected to hard restrictions regarding their manufacturing precision. These regulations are defined by the International Table Tennis Federation (ITTF). A standard 40 mm table tennis ball must have a minimum radius of 19.75 mm and must not exceed 20.25 mm [ITT16]. This deviation of ± 0.25 mm provides a sufficient precise reference for the scanning methods used in this work.
The scene from Fig. 4.1 was captured four times with both scanning approaches, each time under different conditions regarding ambient light quantity and exposure time. The six balls are numbered following the scheme depicted in Fig. 4.7. The radii found in Tab. 4.2 and 4.3 were determined using CloudCompare's sphere fit. The data is visualised in Fig. 4.8. Box diagrams of the scans and a comparison of all values are shown in Fig. B.1 and B.2. The graph's y-axes are equally scaled for better comparability.
The radii determined from the Gray code reconstruction lie (except for two, marked red in Tab. 4.2) within the manufacturing precision. The mean of all 24 determined radii is slightly lower than the desired radius of a table tennis ball. The precision of the Gray code reconstruction lies therefore within the precision of a table tennis ball radius, since the average radius ($19.91 \pm 0.12[mm]$) is an element of the manufacturing precision ($20.00 \pm 0.25[mm]$).
The average radius determined using line scanning reconstruction is $19.75 \pm 0.16[mm]$. When

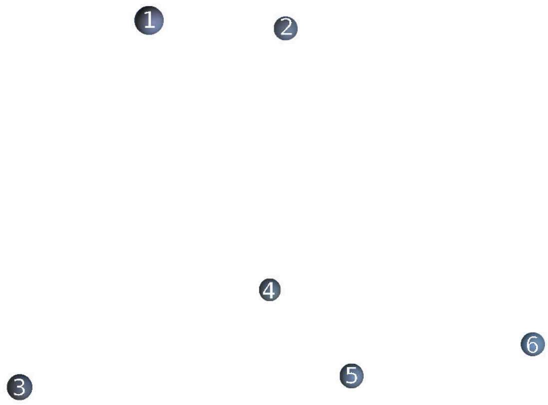

Figure 4.7: Numbering scheme of table tennis balls for radius estimation.

	Ball 1	Ball2	Ball 3	Ball 4	Ball 5	Ball 6	$\bar{r}$/scan	σ/scan
Scan 1	20.016	19.804	19.863	19.996	19.859	20.087	19.93	0.11
Scan 2	19.842	20.039	20.040	19.727	20.035	19.967	19.94	0.13
Scan 3	19.973	19.787	20.016	19.848	19.888	20.078	19.93	0.11
Scan 4	19.903	20.024	19.721	20.152	19.787	20.064	19.94	0.17
$\bar{r}$/ball	19.93	19.91	19.91	19.93	19.89	20.05	19.94	
σ/ball	0.08	0.14	0.15	0.18	0.10	0.06		0.12

Table 4.2: Determined radii of table tennis balls using Gray code reconstruction, all values are given in mm.

analysing these values, it is important to remember that the calibration used is the same as for the Gray code method, but the distortion of the projector is omitted. An additional source of error is the line extraction. Although Steger's line extraction algorithm is known for its high accuracy, it is never guaranteed that the extracted line is precisely at the desired position or perfectly straight, especially since the projector itself can generate only lines as thin as one pixel column and the lines are not continuous in contrast to laser lines. This leads, depending on the distance of the projector to the scene as well as the depth difference of the scene itself to a strong variation in line width and sometimes to big distortions, depending on the incident angle. However, this does not exclusively affect the reconstruction of table tennis balls, but is a general problem when using a projector for line generation (cf. Fig. 4.9).

Torso

A second experiment is conducted using a plastic torso. From a CT scan with 2 mm resolution, a mesh is extracted using the marching cubes algorithm. The mesh contains around 1.5 million triangles. In order not to loose precision, five million points are sampled. The torso is scanned at three different positions regarding the visible area using both Gray code and pre-calibrated

	Ball 1	Ball2	Ball 3	Ball 4	Ball 5	Ball 6	$\bar{r}$/scan	σ/scan
Scan 1	19.631	19.814	19.921	19.746	19.915	19.514	19.76	0.16
Scan 2	19.710	19.876	19.673	19.602	19.537	19.922	19.72	0.15
Scan 3	19.538	19.742	19.479	19.674	19.919	19.950	19.72	0.19
Scan 4	19.826	19.807	19.625	20.038	19.686	19.981	19.83	0.16
$\bar{r}$/ball	19.68	19.81	19.67	19.76	19.76	19.84	19.76	
σ/ball	0.12	0.05	0.18	0.19	0.19	0.22		0.16

Table 4.3: Determined radii of table tennis balls using line scanning reconstruction, all values are given in mm.

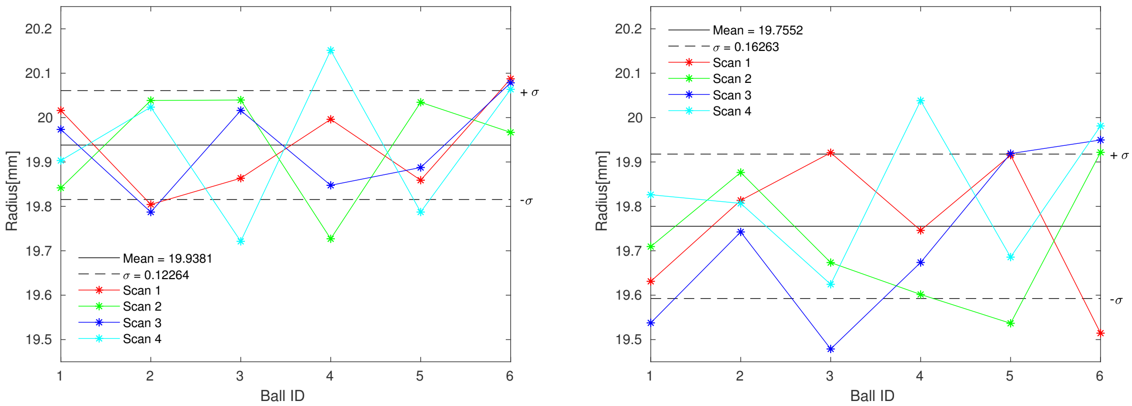

Figure 4.8: Estimated table tennis ball radii. Left: Reconstruction using Gray code, Right: Reconstruction using line scanning.

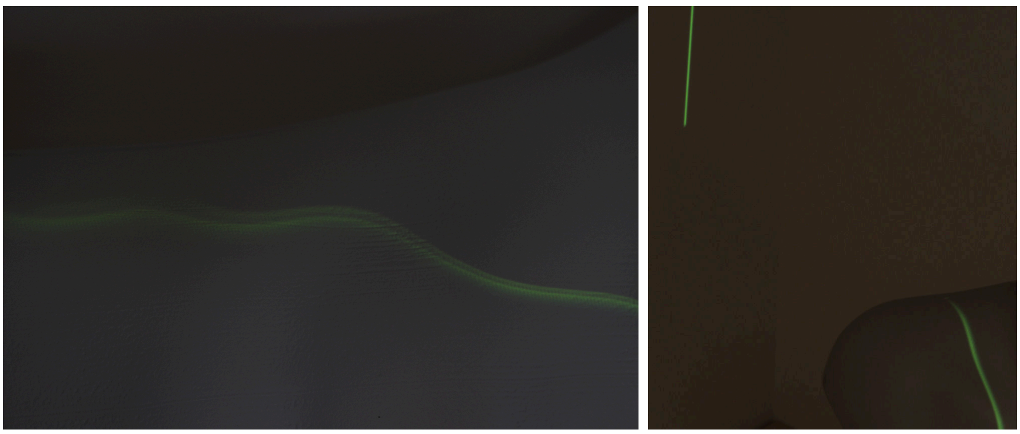

Figure 4.9: Line projection abnormalities. Left: Distortion of a line depending on the incident angle. Right: Poor depth of focus of a projector.

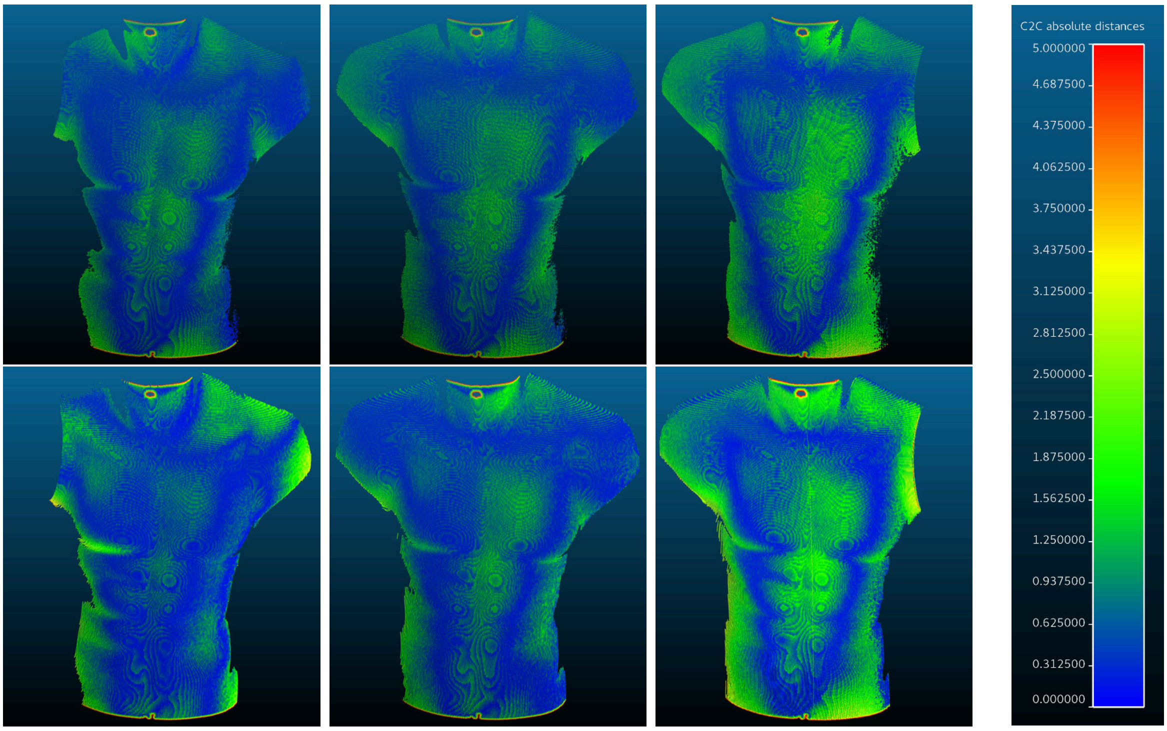

Figure 4.10: Registered reconstruction of a male torso. The the positions are displayed from left to right. top row: Gray code reconstruction, bottom row: line scanning, the scale is given in mm.

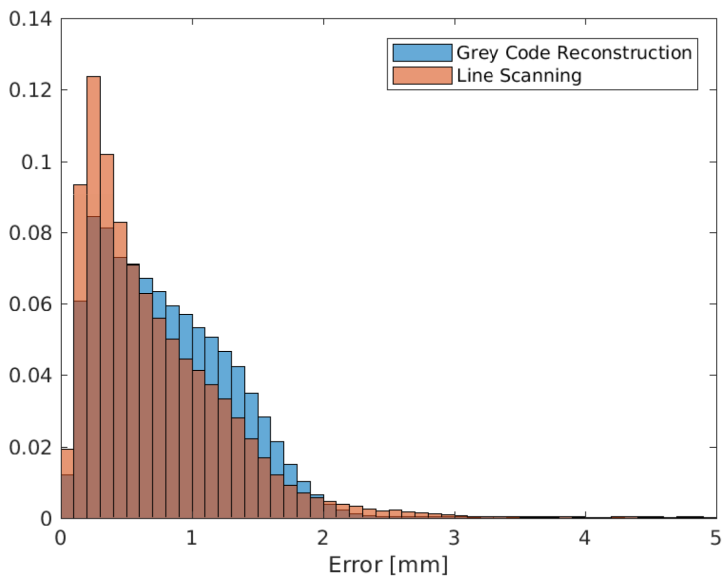

Figure 4.11: Histogram of the point-to-point errors for the middle position.

	Position	Nr. Points	RMS	$\bar{x}$	σ
Gray Code	Left	146,264	1.7580	1.3409	3.9808
	Middle	144,229	1.7978	1.8388	5.8413
	Right	183,102	1.9082	2.0341	5.9338
Line Scanning	Left	493,702	1.7637	1.9893	6.2622
	Middle	445,534	1.7202	1.9568	6.5729
	Right	706.808	1.9117	1.4083	2.7912

Table 4.4: Registration results for torso reconstruction. RMS, mean distance $\bar{x}$ and standard deviation σ are given in mm.

line reconstruction, and then registered with the CT scan. The results of the registration are shown in Tab. 4.4. Fig. 4.10 shows the point clouds for all positions coloured with the absolute point-to-point error in mm. Due to the high sampling of 5 million points on the surface of the CT model and the proportional lower number of points of the scans, the point-to-point distance is a good approximation for the point-to-plane distance.
The histogram in Fig. 4.11 shows a normalised distribution of the point-to-point errors of both methods for the middle position. It is noticeable, that only a few percent of the errors are bigger than 2 mm. This is important, since, due to the CT scan resolution, the reconstructed model contains visible artefacts from interpolation errors while generating the mesh. Therefore, it is fair to say that both applied methods are at least as precise as the reconstruction using a CT scan. From the coloured point clouds in Fig. 4.10 and also the RMS and the mean distance $\bar{x}$ in Tab. 4.4 it is noticeable, that the omission of the projector distortion has less influence on the reconstruction compared to the whole scene. This leads to the conclusion, that the distortion mainly influences the reconstruction on a global scale when reconstructing a whole scene, but is almost not present locally when considering only objects, that are significantly smaller than the field of view.

4.2.3 Plane Fitting

To further examine the influence of the projector distortion and compare both methods to a ground truth, two walls of a room are reconstructed. The room is additionally scanned using a Riegl VZ-400 3D Terrestrial Laser Scanner [Rie]. Although the VZ-400 has an accuracy of only 5 mm, the fitted planes are a good approximation of the reality due to the high number of points per wall (>2,000,000 compared to <200,000 for the applied methods) and the fact, that the planes fitted through the Riegl scans cover the complete walls of the room, while the evaluated reconstruction contain only a small section of the walls due to the limited field of view. Therefore, the determined plane normals and the resulting angle between the walls are considered as ground truth.
It is noticeable, that both reconstruction methods yield an angle different from the reference angle. While the value determined with Gray code reconstruction is slightly bigger, the line scanning value is correspondingly smaller. This means, that the distortion parameters for the

	θ [$^\circ$]	$\Delta\theta$ [$^\circ$]
Gray	90.2323	0.3824
Line	89.5540	-0.2959
Riegl	89.8499	

α [$^\circ$]	β [$^\circ$]
0.9393	0.1266

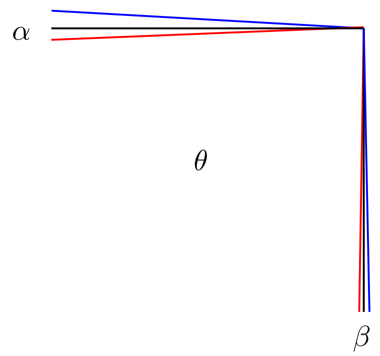

Figure 4.12: Registration results for Gray code reconstruction. Left: Comparison with ground truth. Middle: Angular difference between respective planes of both reconstruction methods. Right: Visualisation of plane fitting.

projector determined by the Moreno method do not model the real distortion accurately. However, when comparing corresponding planes from both applied methods, it is interesting, that the angular difference between both left wall segments is almost $1\,^\circ$, but the difference between both right wall segments is just $0.13\,^\circ$. Therefore, the distortion is more present in the left area of the projected structures and does barely influence the middle and right side of the projection on a large scale, confirming the observations made in previous experiments.

4.3 Discussion of the Self-calibration Reconstruction Results

As already mentioned, the self-calibration approach produces a geometrically distorted point cloud. Fig. 4.13 shows a top view of the scene already depicted in Fig. 4.3. Two additional views (from both left and right side) are found in Fig. A.12 and A.13. The method in general is working for the intended use, but not robustly. As clearly visible, a meaningful comparison with the Gray code reconstruction and the pre-calibrated line scanning is not possible. The reasons for this instability are diverse and will be discussed in this section.

An obvious error source is depicted in Fig. 4.14, where the marked intersections are wrong detections, which are introduced during the line extraction process. The almost horizontal line segment is an imprecise extension of a vertical line and thus, must never intersect with another vertical line. Possible reasons are the thick lines produced by the projector, which appear closely together on the surfaces and may overlap in the camera image due to distortions and diffusion (cf. distortion in Fig. 4.9), or the line tracing, which can produce these imprecise extensions - or artificial deflections - depending on the extraction parameters and the illumination conditions. The resulting erroneous detections influence, if considered valid, the estimation of the plane parameters dramatically, since suddenly lines intersect, which are physically impossible to intersect. The "grid structure" of the projected lines (cf. Fig. 3.2) is, on one hand, necessary, since all vertical lines are later on used for the scanning process and the horizontal lines provide the orthogonality constraints, but on the other hand increases the likelihood of such false intersection detections. This problem can firstly be solved by tagging lines with their orientation and

Figure 4.13: Comparison of self-calibration reconstruction with Gray code (View from top).

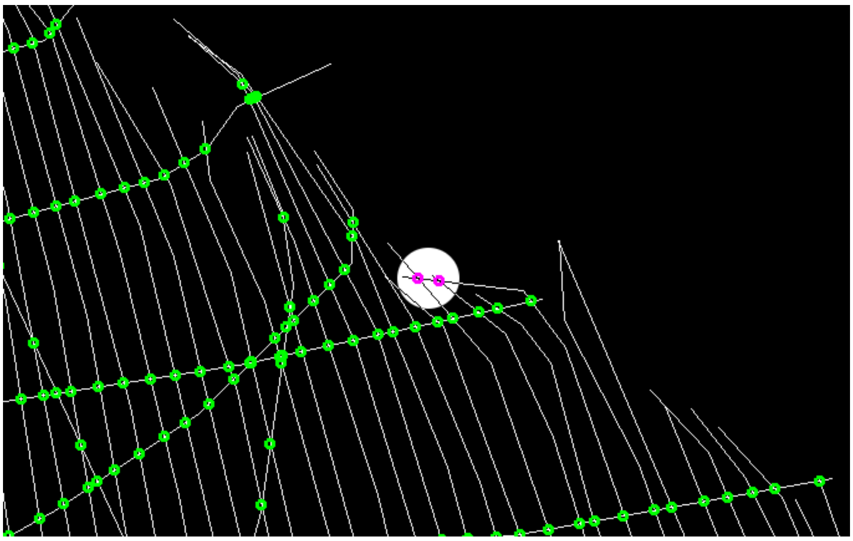

Figure 4.14: Intersection error due to artificial deflections introduced during the line extraction.

introduce a set of rules, which prevent the consideration of such false detections, and secondly, by improving the line tracing such that the generation of deflections is suppressed in the first place.

Due to restrictions of the underlying mathematical problem, not all desired planes can be estimated directly. These restrictions consist mainly of collinear intersections, which lead to degenerate conditions (cf. Sec. 3.2.4). Therefore, a collinearity threshold is applied in order to remove those planes from the initial solution. Planes that cannot be estimated directly are determined by fitting a plane through the intersection points with already solved planes.
The initial solution is highly dependent on noisy detections of intersections like the ones previously described. To improve the solution and reduce the influence of noise, a weighting function can be introduced into the error term. If the intersections are weighted by the 3D distance to the desired plane within the initial solution, the influence of wrong detections and noise can be reduced. Due to the great number of intersections, the problem itself is highly over-determined - in one exemplary case, 1080 valid planes vs.$>$100,000 intersections. This provides the additional possibility to completely remove erroneous intersections in order not to degrade the accuracy of the solution or even prevent the system to be solvable in the first place. This measure provides the opportunity to refine the solution on a local scale without having to re-run the expensive minimisation of the global problem.
In [Fur09], Furukawa and Kawasaki also present a possible improvement. When counting the singular vales of $\mathbf{L}$ from Eq. (3.42), which are below a certain threshold, degenerate conditions can be detected and excluded from the initial solution. However, they also note, that the determination of a satisfying threshold is difficult and considered as future work.

Chapter 5

Conclusion

5.1 Summary

In this thesis, a projector-camera active stereo structured light system for 3D real-time reconstruction was developed. The approach utilises, different to other projector-based systems, a single line for scanning, which is then sweeped across the scene. This was done with two main objectives.
First, to obtain a cost-effective simulation of a laser-based projector and analyse the performance of line sweeping compared to spatially varying full-frame structured light.
Second, to test and evaluate a self-calibrating approach based on Furukawa and Kawasaki ([Fur06], [Fur09]).
The results of the line scan were compared with the reconstruction of a Gray code scan, obtained using a software provided along [Mor12]. In general, the proposed method can compete with the Gray code reconstruction regarding precision and accuracy. Compared to a Gray code reconstruction, the density of the obtainable point cloud is higher, since the maximum number of points is not limited by the projector resolution, but by the camera resolution. This is especially interesting, when using a laser projector, since the lines constructed with a cinema projector need to be interpolated for higher resolutions. The errors introduced into the system on a global scale are mainly caused by the omission of the projector distortion, which leads in the applied case to a wrong determination of the pre-calibrated plane parameters and to lines, that are not perfectly straight. These distortions are not present, when using a laser-based projector. The errors on a local scale are the result of a noisy line extraction, depending of the selected parameters. The problem here is less caused by the line extraction algorithm itself, but more influenced by the imprecise projection of the lines regarding the previously discussed distortions and depth of focus problems, and also the intensity of the line. Again, this problem is solvable by using a laser-based projector.
The developed GUI serves as a tool for camera and projector handling and simplifies the calibration and reconstruction workflow. Additionally, the in real-time reconstructed point cloud is displayed and saved after completion.
Regarding reconstruction rate, the current state of the system cannot compete with a Gray code based reconstruction. The reasons for this is the projector lacking any kind of feedback

or triggering signal, which was already briefly addressed. In order to compensate this and be able to identify the projected line, a manual delay between 300 ms and 500 ms, depending on the total workload of the system, is necessary. By applying a trigger signal to both camera and a potential laser-based line projector, the reconstruction rate of the system is only limited by the maximum framerate of the camera, providing a massive speed improvement.
The self-calibrating approach is able to determine plane equation and reconstruct the scene in 3D, but the reconstruction is geometrically distorted due to different influences previously discussed. The problems were analysed and possible solutions were suggested.
However, despite these effects and minor complications, the proposed system works as planned and can be used for simple, precise and dense reconstruction of objects and small scenes.

5.2 Future Work

In order to improve the quality and performance of the system, improvements and solutions for problems were presented, which are shortly summarised in the following.
The self-calibrating approach was demonstrated to be working, but still needs improvements and further research. The system needs to be made robust to noise and outliers, while making the determination of thresholds more flexible and easy. As a next step ideas provided in Sec. 4.3 need to be mathematically analysed, implemented and tested.
The line extractor needs to be extended, especially regarding the line tracing, in order to reduce noise and erroneous deflections.
To further improve the quality of the extracted lines and therefore the quality of the whole reconstruction as well as the performance of the self-calibration, it is recommended to use a laser-based projector. By projecting thinner, more intensive lines, that are less prone to distortions affected by the incident angle and providing a better depth of focus, a lot of problems encountered during this work can be simplified or even solved with a small modification of the setup.
However, in order to use a laser-based projector while maintaining the flexibility of the setup, it is crucial to improve the self-calibration, since otherwise the setup needs to be calibrated as described in the introduction.
Regarding the structured light scanner currently developed at INESCTEC for the ¡VAMOS! project, two matters are important. First, the systems needs to be extended with another horizontal mounted laser line projector, if the self-calibration approach is to be pursued and second, the self-calibration needs to be robust, since the conditions under water may evoke different effects than the ones observed in air.

Appendices

Appendix A

Images

A.1 Graphical User Interface

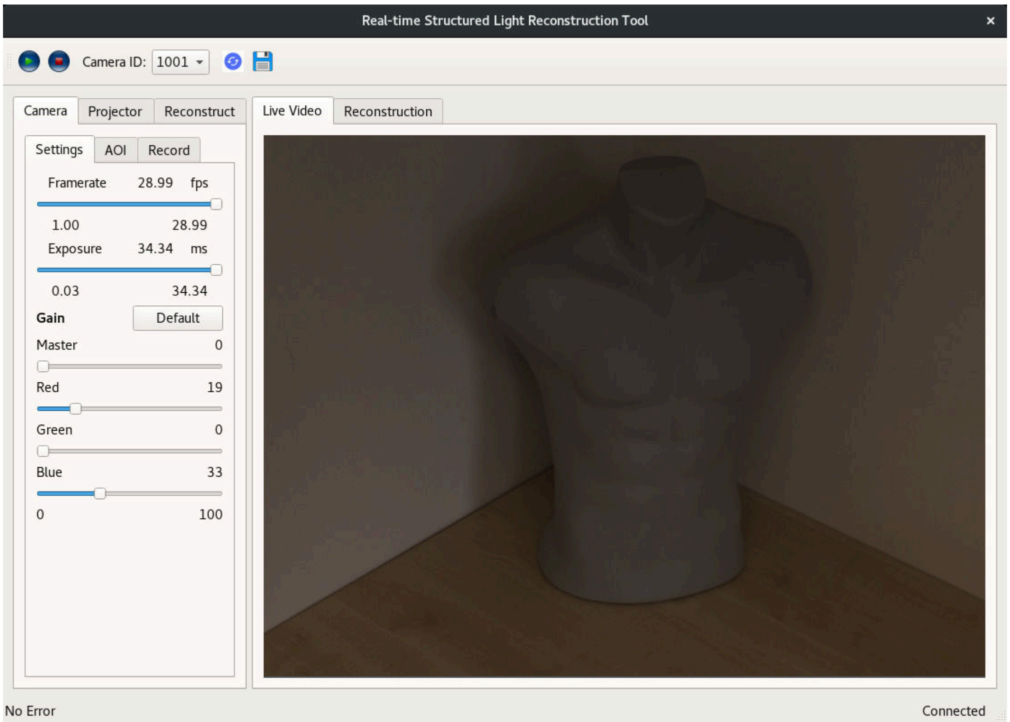

Figure A.1: Basic camera controls.

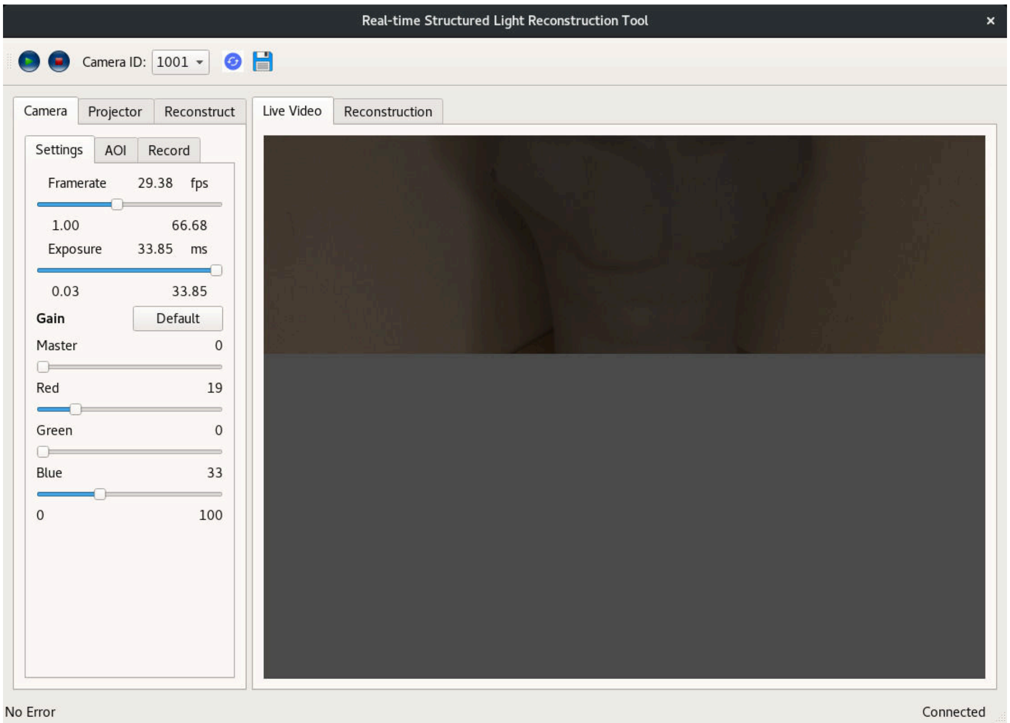

Figure A.2: AOI camera controls.

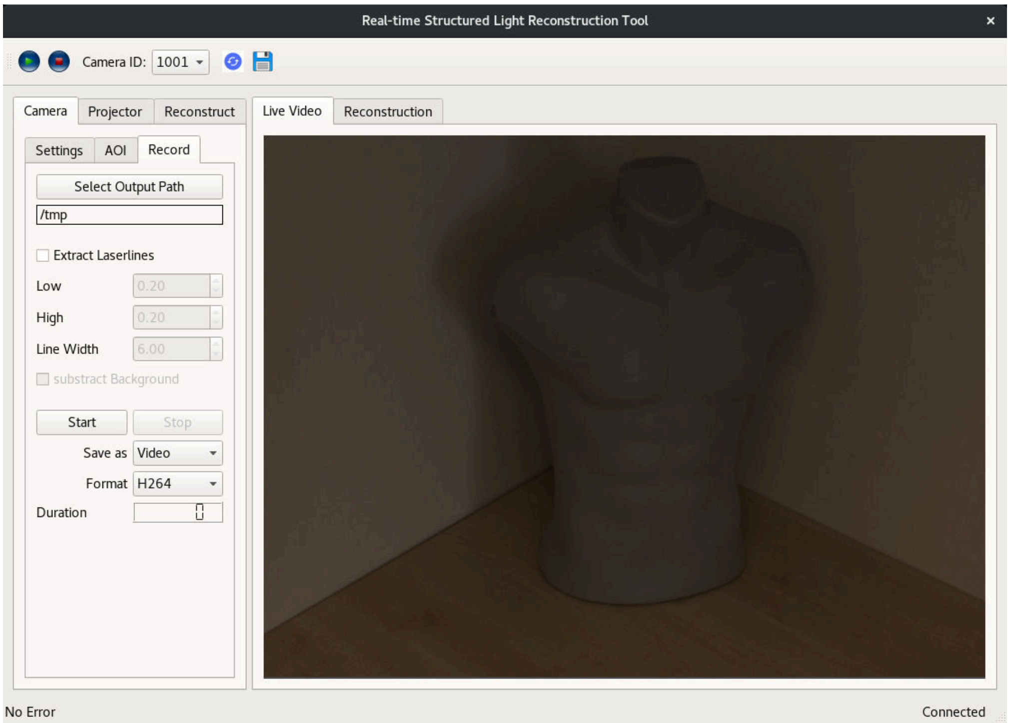

Figure A.3: Tools for picture and video recording.

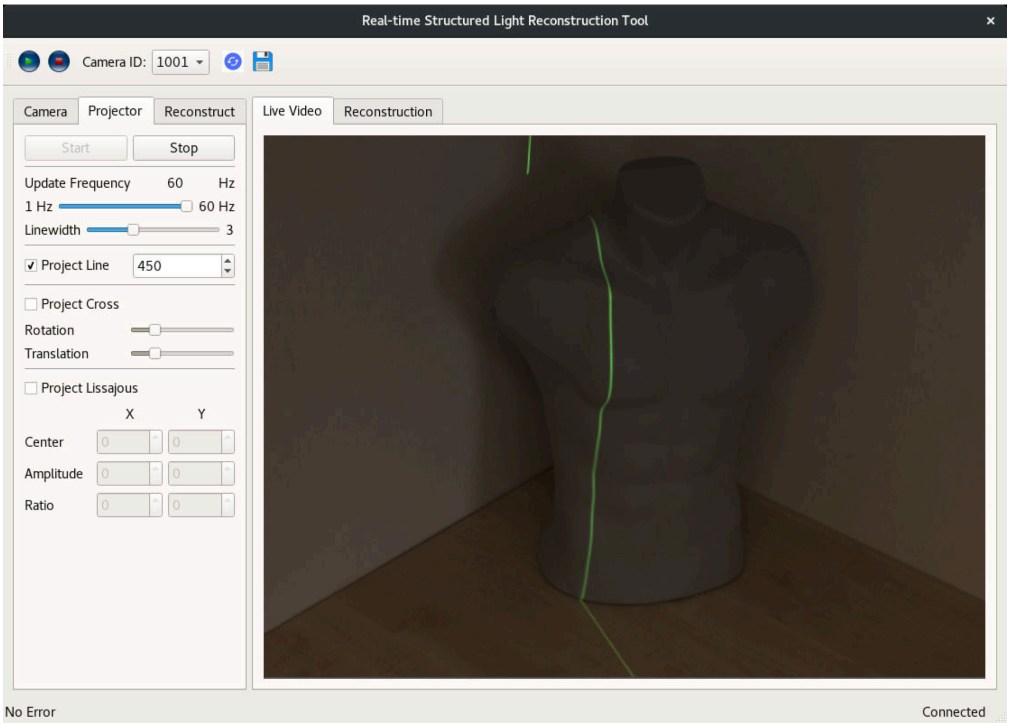

Figure A.4: Projector painting a line in a specified column.

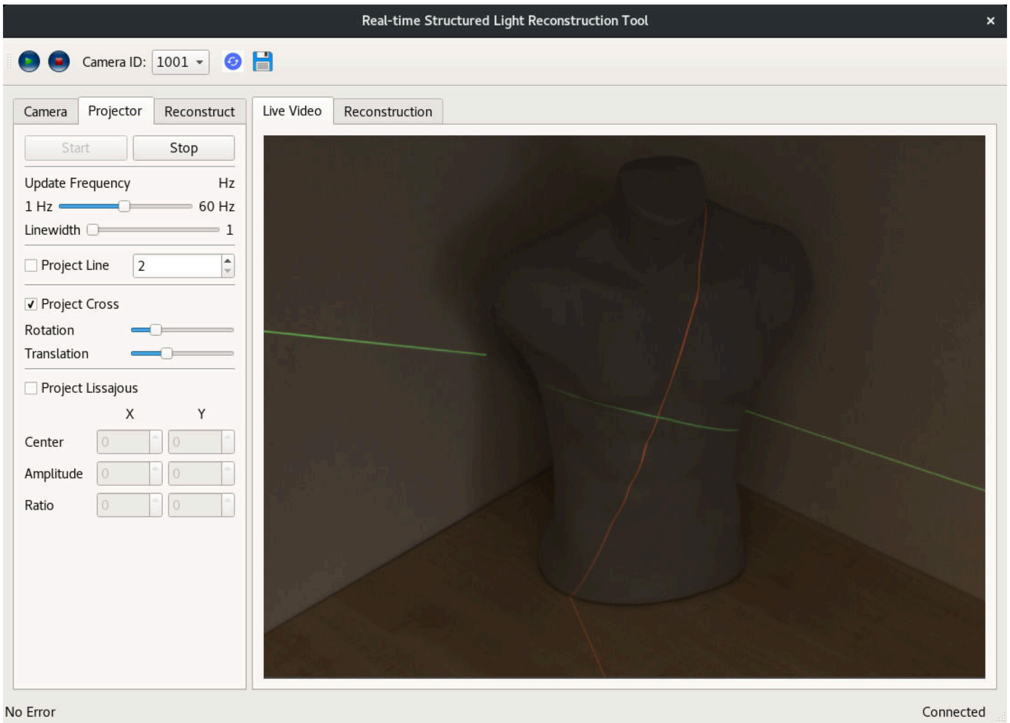

Figure A.5: Projector painting a cross, which moves randomly through the scene.

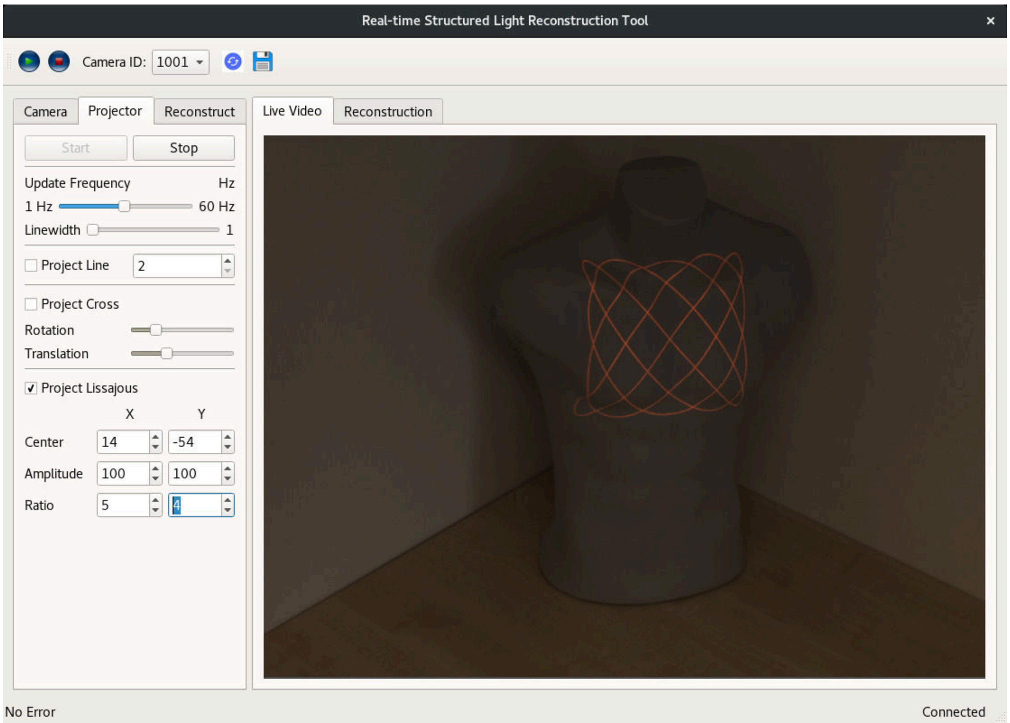

Figure A.6: Projector paining a lissajous figure.

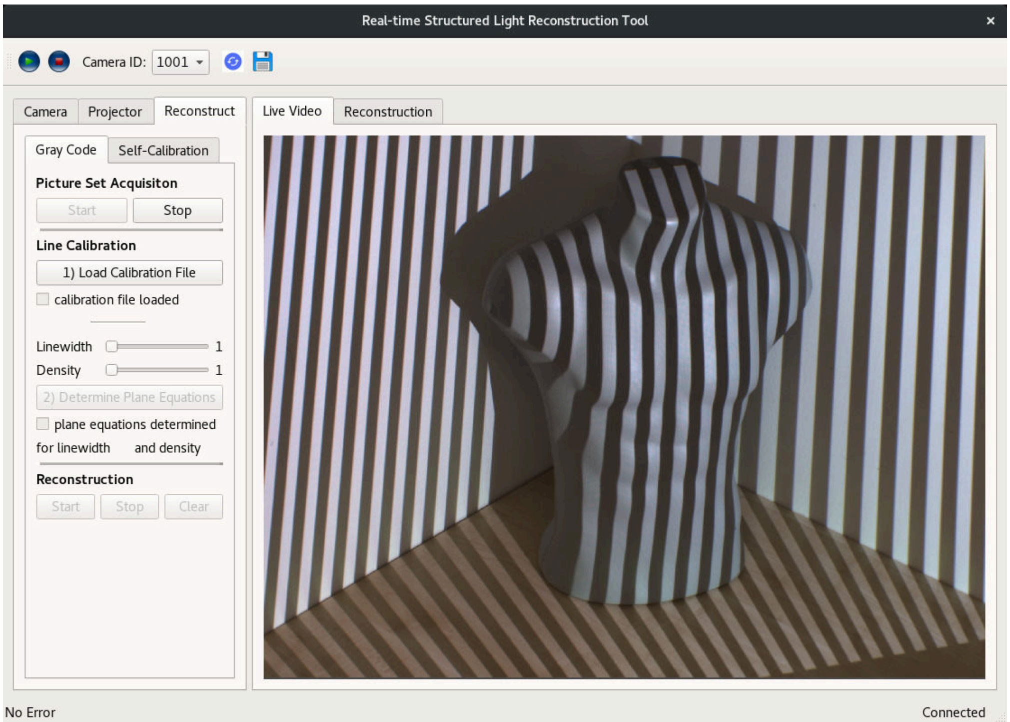

Figure A.7: Gray code projection for scene reconstruction.

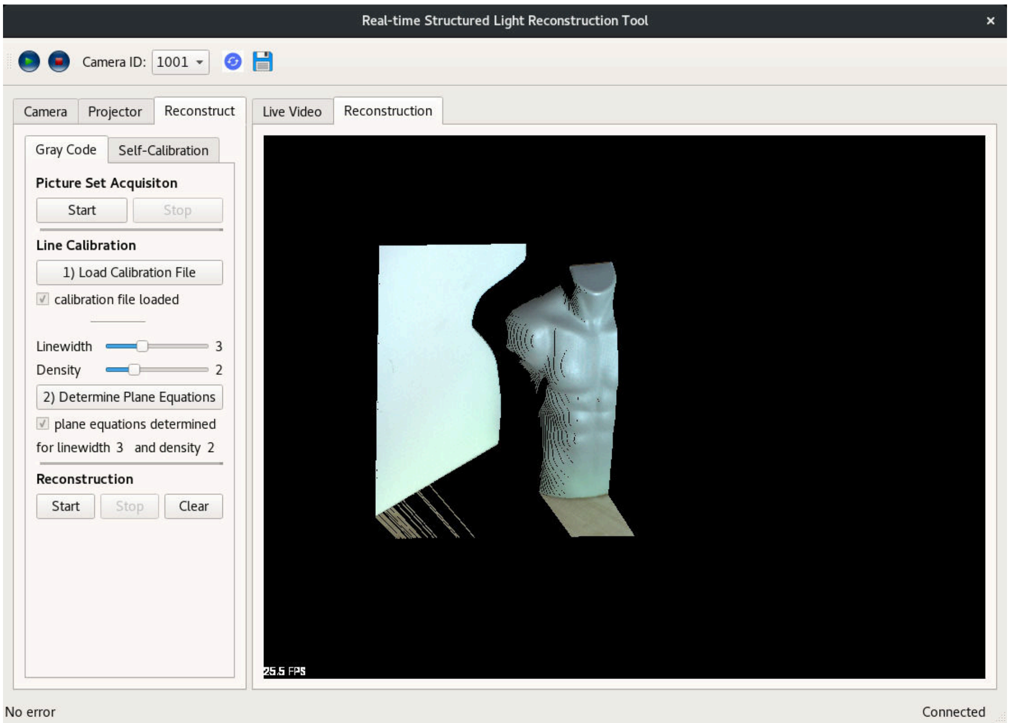

Figure A.8: Partially reconstructed scene using pre-calibrated line scanning.

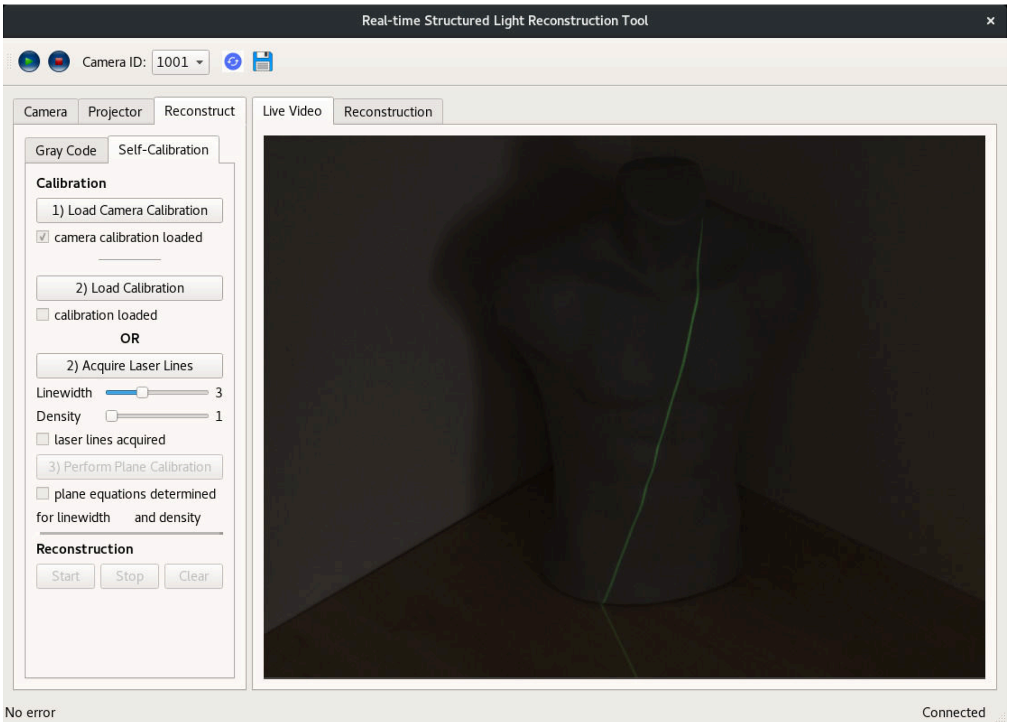

Figure A.9: Line acquisition for self-calibration.

A.2 Reconstruction

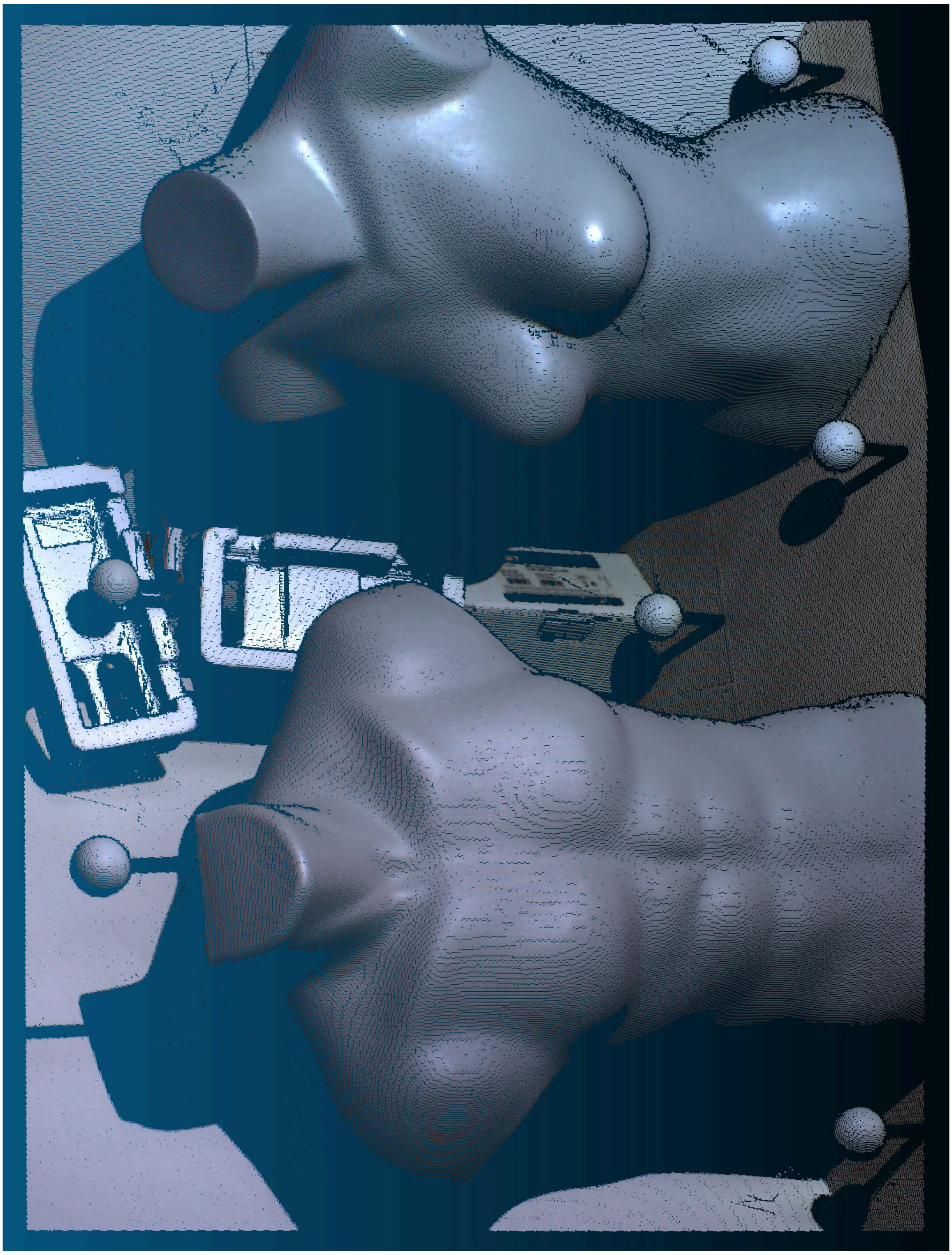

Figure A.10: Scene reconstruction using spatially varying Gray code.

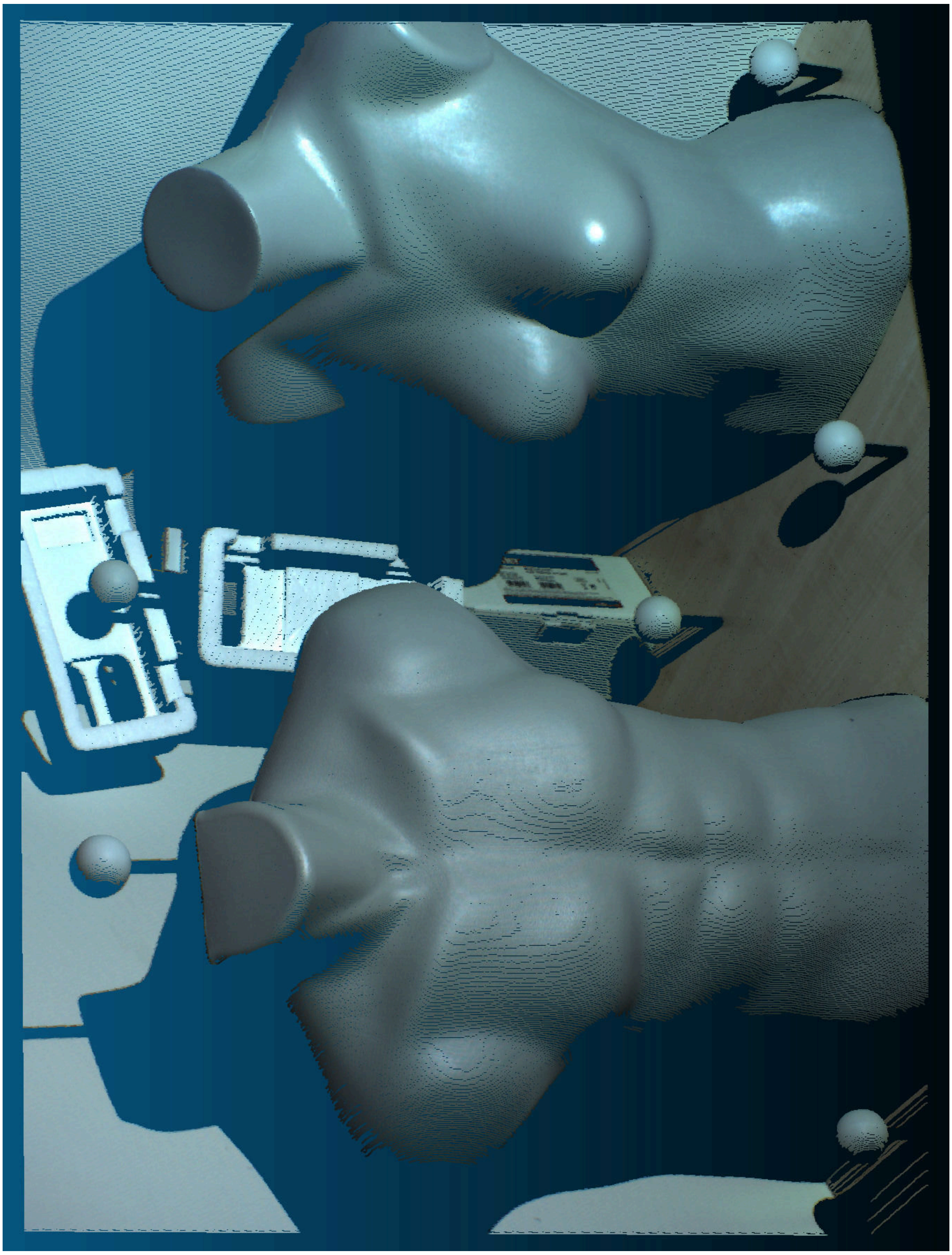

Figure A.11: Scene reconstruction using pre-calibrated line scanning

A.3 Self-calibration

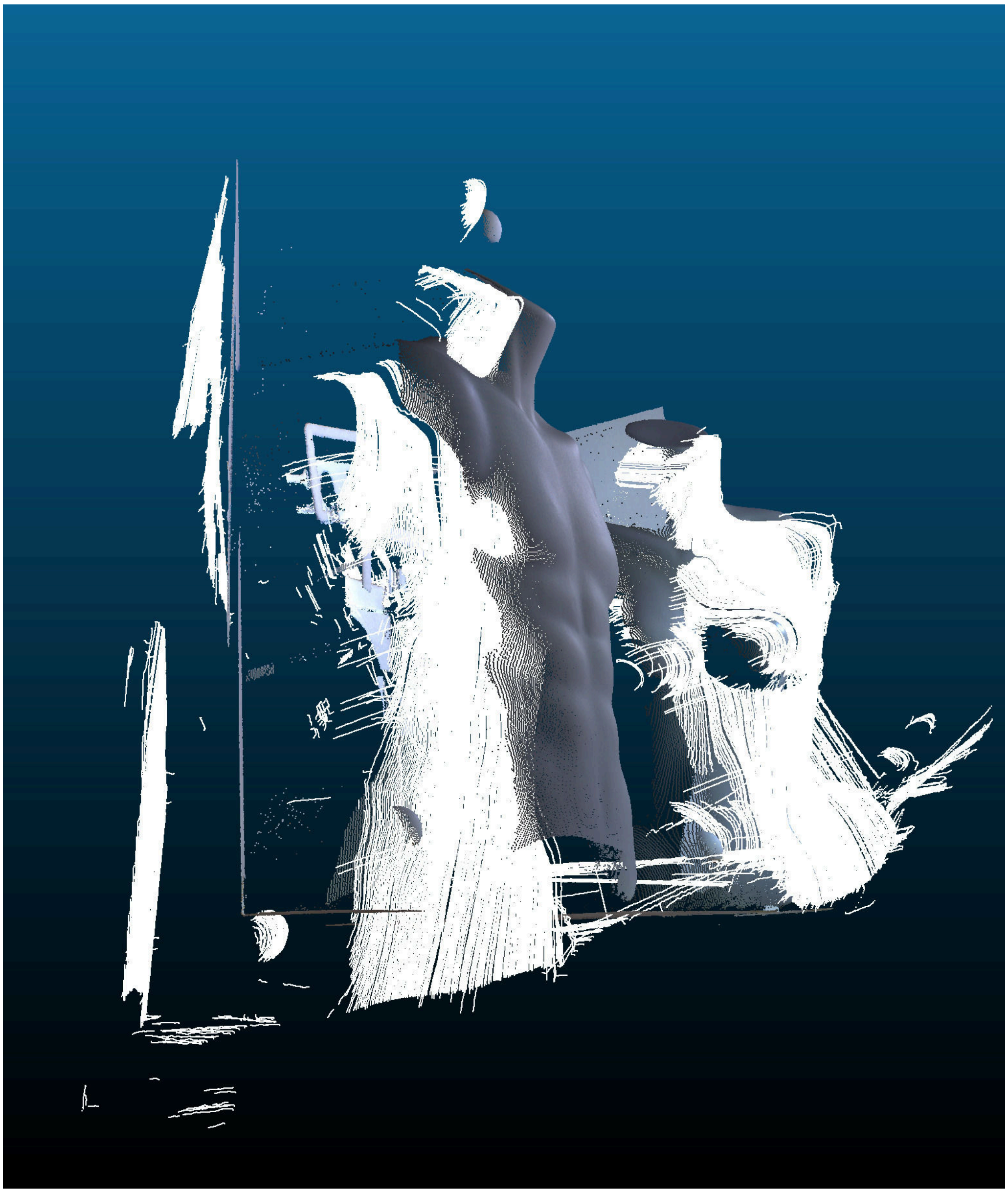

Figure A.12: Comparison of self-calibration reconstruction with Gray code (View from Left).

Figure A.13: Comparison of self-calibration reconstruction with Gray code (View from right).

Appendix B

Diagrams

B.1 Table Tennis Balls

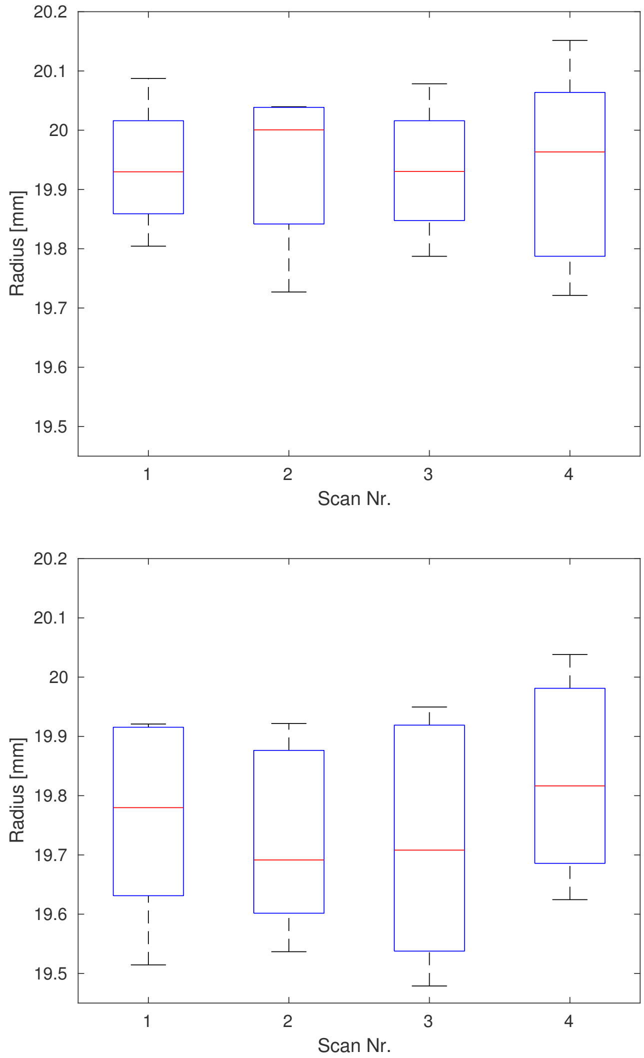

Figure B.1: Box plot of table tennis ball radius estimation. Top: Reconstruction using Gray code, Bottom: Reconstruction using line scanning.

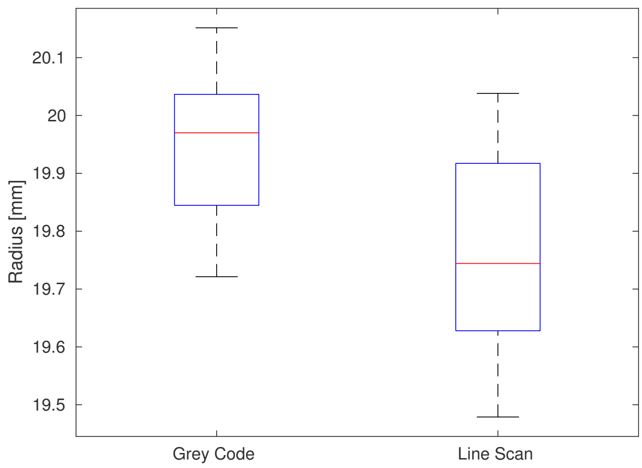

Figure B.2: Box plot comparison of table tennis ball radius estimation.

List of Figures

List of Tables

List of Acronyms

AOI	Area of Interest
API	Application Programming Interface
CAD	Computer-aided Design
CMM	Coordinate Measuring Machines
CT	Computed Tomography
DoF	Degree of Freedom
FoV	Field of View
fps	Frames per Second
GUI	Graphical User Interface
LED	Light-emitting Diode
LIDAR	Light Detection and Ranging
MEMS	Microelectromechanical Systems
MRI	Magnetic Resonance Imaging
PCL	Point Cloud Library
RADAR	Radio Detection and Ranging
ROS	Robotic Operating System
SfM	Structure from Motion
SIFT	Scale-invariant Feature Transform
SONAR	Sound Navigation and Ranging
SURF	Speeded Up Robust Features
SVD	Singular Value Decomposition
VR	Virtual Reality

Bibliography

[Agi] Agisoft PhotoScan. `http://www.agisoft.com/downloads/installer/`.

[Ali] Aliaga, Daniel G and Xu, Yi. A self-calibrating method for photogeometric acquisition of 3D objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010.*

[AM+] Sameer Agarwal, Keir Mierle, et al. Ceres Solver. `http://ceres-solver.org`.

[Bal03] Balz, Timo and Haala, Norbert. SAR-based 3D-reconstruction of complex urban environments. In *IAPRS Workshop on*, 2003.

[Bov02] Boverie, S and Devy, M and Lerasle, F. 3D perception for new airbag generations. *IFAC Proceedings Volumes*, 35(1):19–24, 2002.

[Bru11] Bruno, F and Bianco, Gianfranco and Muzzupappa, Maurizio and Barone, Sandro and Razionale, AV. Experimentation of structured light and stereo vision for underwater 3D reconstruction. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(4):508–518, 2011.

[Clo17] CloudCompare Version 2.8.1. Retrieved from `http://www.cloudcompare.org/`, 2017.

[Coi09] Coiras, E and Groen, J. Simulation and 3D reconstruction of side-looking sonar images. In *Advances in sonar technology.* InTech, 2009.

[Coo08] Cooper, K. B. and Dengler, R. J. and Llombart, N. and Bryllert, T. and Chattopadhyay, G. and Schlecht, E. and Gill, J. and Lee, C. and Skalare, A. and Mehdi, I.and Siegel, P. H. Penetrating 3-D Imaging at 4- and 25-m Range Using a Submillimeter-Wave Radar. *IEEE Transactions on Microwave Theory and Techniques*, 56:2771–2778, 2008.

[Cun14] Cunningham, John A. and Rahman, Imran A. and Lautenschlager, Stephan and Rayfield, Emily J. and Donoghue, Philip C.J. A virtual world of paleontology. *Trends in Ecology & Evolution*, 29(6):347 – 357, 2014.

[De 04] De La Hamette, Patrick and von Waldkirch, Marc and Tröster, Gerhard. Laser triangulation as a means of robust visual input for wearable computers. In *ISWC'04—Proceedings of the 4th IEEE international symposium on wearable computers, Doctoral Colloquium. IEEE Computer Society, Washington DC*, pages 18–20, 2004.

[ea17] Itseez et al. OpenCV - Open Source Computer Vision Library. `https://http://opencv.org/`, 2017.

[Fur05] Furukawa, Ryo and Kawasaki, Hiroshi. Uncalibrated multiple image stereo system with arbitrarily movable camera and projector for wide range scanning. In *3-D Digital Imaging and Modeling, 2005. 3DIM 2005. Fifth International Conference on*, pages 302–309. IEEE, 2005.

[Fur06] Furukawa, Ryo and Kawasaki, Hiroshi. Self-calibration of multiple laser planes for 3D scene reconstruction. In *3D Data Processing, Visualization, and Transmission, Third International Symposium on*, pages 200–207. IEEE, 2006.

[Fur09] Furukawa, Ryo and Kawasaki, Hiroshi. Laser range scanner based on self-calibration techniques using coplanarities and metric constraints. *Computer Vision and Image Understanding*, 113(11):1118–1129, 2009.

[GJ+10] Gaël Guennebaud, Benoît Jacob, et al. Eigen v3. http://eigen.tuxfamily.org, 2010.

[Gol07] Goldman, Lee W. Principles of CT and CT technology. *Journal of nuclear medicine technology*, 35(3):115–128, 2007.

[GS95] Ardeshir Goshtasby and Hai-Lun Shyu. Edge detection by curve fitting. *Image and Vision Computing*, 13(3):169–177, 1995.

[Han09] Hanson, Lars G. Introduction to magnetic resonance imaging techniques. 2009.

[Har04] Hartley, R. I. and Zisserman, A. *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521540518, Second edition, 2004.

[Hel] Hellerhoff. Felsenbein in der Computertomographie. [`https://commons.wikimedia.org/wiki/File:Felsenbein-CT-axial-18.jpg`accessed September 21, 2017.

[Her12] Hertzberg, Joachim and Lingemann, Kai and Nüchter, Andreas. *Mobile Roboter*. Springer-Verlag Berlin Heidelberg, 2012.

[IDS] IDS UI-5270CP Rev. 2. `https://en.ids-imaging.com/store/ui-5270cp-rev-2.html`.

[ITT16] International Table Tennis Federation - Equipment Regulations: Technical Leaflet T3: The Ball. `http://www.old.ittf.com/ittf_equipment/Technical_Leaflets/T3_Ball.pdf`, 2016.

[Jas11] Jason Geng. Structured-light 3D surface imaging: a tutorial. *Adv. Opt. Photon.*, 3(2):128–160, Jun 2011.

[Jok99] Jokinen, Olli. Self-calibration of a light striping system by matching multiple 3-D profile maps. In *3-D Digital Imaging and Modeling, 1999. Proceedings. Second International Conference on*, pages 180–190. IEEE, 1999.

[Kir98] Kirscht, Martin and Rinke, Carsten. 3D Reconstruction of Buildings and Vegetation from Synthetic Aperture Radar (SAR) Images. In *MVA*, pages 228–231, 1998.

[LZ12] Zhihua Lv and Zhiyi Zhang. Build 3d scanner system based on binocular stereo vision. *journal of computers*, 7(2):399–404, 2012.

[LZHL17] Yuehua Li, Jingbo Zhou, Fengshan Huang, and Lijian Liu. Sub-pixel extraction of laser stripe center using an improved gray-gravity method. *Sensors*, 17(4):814, 2017.

[Mat97] booktitle= Robotics and Automation, 1997. Proceedings., 1997 IEEE International Conference on Matthies, Larry and Balch, Tucker and Wilcox, Brian. Fast optical hazard detection for planetary rovers using multiple spot laser triangulation. volume 1, pages 859–866. IEEE, 1997.

[Mor12] Moreno, Daniel and Taubin, Gabriel. Simple, Accurate, and Robust Projector-Camera Calibration. In *Proceedings of the 2012 Second International Conference on 3D Imaging, Modeling, Processing, Visualization & Transmission*, 3DIMPVT '12, pages 464–471, Washington, DC, USA, 2012. IEEE Computer Society.

[N+17] Andreas Nüchter et al. 3DTK – The 3D Toolkit. http://slam6d.sourceforge.net/, June 2017.

[Nar05] Narasimhan, Srinivasa G and Nayar, Shree K. Structured light methods for underwater imaging: light stripe scanning and photometric stereo. In *OCEANS, 2005. Proceedings of MTS/IEEE*, pages 2610–2617. IEEE, 2005.

[New11] Newcombe, Richard A and Izadi, Shahram and Hilliges, Otmar and Molyneaux, David and Kim, David and Davison, Andrew J and Kohi, Pushmeet and Shotton, Jamie and Hodges, Steve and Fitzgibbon, Andrew. KinectFusion: Real-time dense surface mapping and tracking. In *Mixed and augmented reality (ISMAR), 2011 10th IEEE international symposium on*, pages 127–136. IEEE, 2011.

[Ols11] Edwin Olson. AprilTag: A robust and flexible visual fiducial system. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, pages 3400–3407. IEEE, May 2011.

[PCL] PCL - Point Cloud Library. `http://pointclouds.org`.

[Pie] Pierre Moulon et al. OpenMVG. An Open Multiple View Geometry library. `https://github.com/openMVG/openMVG`.

[Qt] Qt - Cross-platform software development for embedded & desktop. `https://www.qt.io/`.

[Rie] Riegl VZ-400 3D Terrestrial Laser Scanner. `http://www.riegl.com/nc/products/terrestrial-scanning/produktdetail/product/scanner/5/`.

[Sol04] W. J. Sollas. A method for the investigation of fossils by serial sections. *Philosophical Transactions of the Royal Society of London. Series B, Containing Papers of a Biological Character*, 196:259–265, 1904.

[Son16] Song, Young-eun and Choi, Seung-Joon. Underwater 3D Reconstruction for Underwater Construction Robot Based on 2D Multibeam Imaging Sonar. *Journal of Ocean Engineering and Technology*, 3(3), Jun 2016.

[Ste98a] Steger, Carsten. An Unbiased Detector of Curvilinear Structures. *IEEE Trans. Pattern Anal. Mach. Intell.*, 20(2):113–125, Feb 1998.

[Ste98b] Steger, Carsten. *Unbiased extraction of curvilinear structures from 2D and 3D images.* Utz, Wiss., 1998.

[VAM] Viable Alternative Mine Operating Systeml. `http://vamos-project.eu`.

[WEN] WENZEL Group GmbH & Co. KG. Manual 3D coordinate measuring machine XCite. `http://www.wenzel-group.com/praezision/en/products/manual/xcite-manual-3d-coordinate-measuring-machine.php?we_ID=3095`.

[Wil71] Will, Peter M and Pennington, Keith S. Grid coding: A preprocessing technique for robot and machine vision. *Artificial Intelligence*, 2(3-4):319–329, 1971.

[Win06] Winkelbach, Simon and Molkenstruck, Sven and Wahl, Friedrich M. Low-cost laser range scanner and fast surface registration approach. *Lecture Notes in Computer Science*, 4174:718, 2006.

[Wu,13] Wu, Changchang. Towards linear-time incremental structure from motion. In *3DTV-Conference, 2013 International Conference on*, pages 127–134. IEEE, 2013.

[Zan16] Zanuttigh, Pietro and Marin, Giulio and Dal Mutto, Carlo and Dominio, Fabio and Minto, Ludovico and Cortelazzo, Guido Maria. *Time-of-Flight and Structured Light Depth Cameras - Technology and Applications.* Springer International Publishing AG Switzerland, 2016.

[Zer96] Zerr, Beno ît and Stage, Bjarne. Three-dimensional reconstruction of underwater objects from a sequence of sonar images. In *Image Processing, 1996. Proceedings., International Conference on*, volume 3, pages 927–930. IEEE, 1996.

[Zha00] Zhang, Zhengyou. A Flexible New Technique for Camera Calibration. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(11):1330–1334, Nov 2000.

Proclamation

Hereby I confirm that I wrote this thesis independently and that I have not made use of any other resources or means than those indicated.

Würzburg, September 2017