CLIPS-development of a novel camera and laser-based indoor positioning system

Author(s):
Tilch, Sebastian

Publication Date:
2012

Permanent Link:
https://doi.org/10.3929/ethz-a-007597344

Rights / License:
In Copyright - Non-Commercial Use Permitted
CLIPS – Development of a Novel Camera and Laser-Based Indoor Positioning System

A dissertation submitted to

ETH ZURICH

for the degree of

Doctor of Sciences

presented by

SEBASTIAN TILCH

MSc ETH, ETH Zurich

born 24.04.1984

citizen of Germany

accepted on the recommendation of

Prof. Dr. Hilmar Ingensand, examiner
Prof. Dr.-Ing. habil. Thomas Wunderlich, co-examiner

2012
Abstract

The emergence of new optical indoor positioning technologies facilitates a variety of new applications in the guidance of individuals, machines or tools in every indoor environment. Once an individual or object is equipped with a mobile camera, its position can be determined instantaneously with accuracies up to the millimetre range. However, most optical indoor positioning systems are at a development stage. The aim of this work is the development of an automatic, inexpensive and mobile Camera and Laser-Based Indoor Positioning System (CLIPS), which is capable of providing continuous positions and orientations of a mobile camera with respect to a stationary laser projector for each indoor environment. Additionally, high precision mechanics, sophisticated set-ups and additional infrastructure like coded targets on the wall are avoided to simplify its application and to decrease the set-up time to 10 minutes or less.

The CLIPS system consists of a laser projector and a mobile camera. With CLIPS, the projector is designed to project a reference field of 16 red and 36 green laser spots on any surface in an indoor environment. If the camera captures the reference field, the camera's position and orientation can be estimated with respect to the projector by means of stereo image processing and resection. Because of the radial arrangement of the red laser pointers, no metric information is provided to estimate the absolute camera orientation. Therefore, the metric scale has to be introduced separately. The introduction of a metric scale is realized by additional eccentric arranged green laser pointers with a known offset of about 25 cm to the projector origin.

To estimate accurately the camera pose with respect to the projector, both devices must be calibrated. Here, the camera calibration with the interior orientation and lens distortion is automatically estimated according to Brown's model (Brown, 1968). The major challenge lies in the projector calibration. For the reconstruction of the laser bundle, the projector has a static set-up in front of a surface where the laser beams are projected on. The surface is then subsequently shifted. At each location, three-dimensional coordinates of the laser spots are determined by a theodolite measurement system such as Leica Axys or photogrammetric measurements. For each laser pointer, a line with an initial point and a direction vector can be derived via a principle component analysis. Additionally, the apparent intersection point is estimated. Finally, the laser beam directions are described by spherical coordinates and a small offset vector to the apparent intersection point.
Before a camera pose can be estimated, the recognition of each laser spot and its assignment to the corresponding laser beam becomes necessary to generate corresponding point pairs for the camera pose estimation. Firstly, regions of interest are determined via a simple colour channel combination. Disturbing sources like lamps or illumination by daylight are excluded and purely red or green areas remain. Having determined these regions of interest (ROI's) for the red and green laser spots, a template matching is applied to evaluate the region's shape and intensity distribution. If a region is labelled as a laser spot, the centroids are derived by weighted centroid estimation. The identification is solved via a colour-coded approach by exploiting the pattern of the red and green laser spots.

The pose estimation is a two-staged approach. Due to the abundance of approximate values for the first camera location, a simple initialization in a previously defined octant of the projector coordinate system becomes necessary. In this octant, a set of approximate values is generated and refined by a least squares adjustment and finally, the correct pose is selected. Since the relative orientation can be estimated only up to scale, the metric scale is determined by the additional eccentric arranged green laser spots. Afterwards, 3d spot coordinates are calculated via intersection. For consecutive camera locations, approximate values for the camera pose are predicted via Kalman filtering. These approximate values are refined by a least squares resection, to obtain the camera pose. Although, the metric scale is incorporated by a resection, the metric scale is separately determined for each camera pose to refine the 3d laser spot coordinates.

Currently, CLIPS is able to provide camera positions and orientations with an update rate of 10 Hz. With the current instrumental realization, the camera pose can be estimated with accuracies in the cm-range. To increase the positioning accuracy to the mm-range, alternative approaches must be considered like the application of laser distance meters or the measurement of a scale bar. Further, the reproducibility of a camera position is already in the sub-mm range.
Zusammenfassung


Laserstrahlen werden durch sphärische Koordinaten angegeben. Zusätzlich werden die XY-Offsets der Laserstrahlen zum Projektorzentrum bestimmt.


Derzeit liefert CLIPS Kamerapositionen- und Orientierungen mit einer Aktualisierungsrate von 10 Hz. Die aktuelle Realisierung des Projektors erlaubt wegen der kleinen Basis der grünen Laserstrahlen nur eine Positionierungsgenauigkeit im cm-Bereich. Die Präzision liegt dagegen im sub-mm Bereich. In weiterführenden Arbeiten können andere Ansätze für die Genauigkeitssteigerung, wie die Verwendung von Laserdistanzmessern oder die Verwendung von Massstäben bedacht werden.
Contents

Abstract .............................................................................................................................. I
Zusammenfassung ............................................................................................................. III
Contents ............................................................................................................................ V
1 Introduction ..................................................................................................................... 1
  1.1 Aim of this Work ...................................................................................................... 1
  1.2 Motivation ............................................................................................................... 1
  1.3 Outline .................................................................................................................... 2
2 Categorization of Optical Indoor Positioning Systems ................................................. 3
  2.1 Reference from Building Models ............................................................................ 4
  2.2 Reference from Images .......................................................................................... 5
  2.3 Reference via Established Camera Frameworks .................................................... 6
  2.4 Reference from Deployed Targets ......................................................................... 7
  2.5 Reference from Projected Targets ......................................................................... 8
  2.6 Systems without Reference .................................................................................. 10
3 Architecture of CLIPS .................................................................................................. 11
  3.1 Camera and Camera Lens ...................................................................................... 12
  3.2 Laser Pointer .......................................................................................................... 13
  3.3 Projector ................................................................................................................ 13
  3.4 Controller ............................................................................................................. 14
4 Calibration .................................................................................................................... 15
  4.1 Camera Calibration ............................................................................................... 15
    4.1.1 Interior Orientation ......................................................................................... 16
    4.1.2 Lens Distortions .............................................................................................. 16
  4.2 Laser Calibration ................................................................................................... 19
  4.3 Projector Calibration ............................................................................................. 20
    4.3.1 Calibration Set-Up ......................................................................................... 20
    4.3.2 Reconstruction of the Laser Beams ............................................................... 21
4.3.3 Definition of the Projector Coordinate System ................................................. 21
4.3.4 Stability of the Calibration .............................................................................. 23
5 RECOGNITION AND IDENTIFICATION OF LASER SPOTS ........................................ 25
5.1 Camera Preferences ......................................................................................... 26
5.2 Laser Spot Recognition ..................................................................................... 28
  5.2.1 Determination of Regions of Interest ............................................................ 28
5.2.2 Spot Detection ............................................................................................... 34
5.3 Laser Spot Identification .................................................................................. 41
  5.3.1 Time Coded Pattern ..................................................................................... 41
5.3.2 Colour Coded Pattern ................................................................................... 43
6 CAMERA POSE ESTIMATION .............................................................................. 45
  6.1 Parameterization of the Camera Pose ................................................................ 46
  6.2 Least-Squares-Adjustment of the Camera Pose .................................................. 47
    6.2.1 Least-Squares Adjustment Based on Epipolar Geometry ................................ 47
    6.2.2 Least-Squares Adjustment Based on Collinearity Constraints ...................... 49
6.3 Generation of Approximate Values .................................................................... 50
6.4 Introduction of the Metric Scale ......................................................................... 51
6.5 Estimation of 3D Laser Spot Coordinates ........................................................... 52
6.6 Critical Configurations ...................................................................................... 53
    6.6.1 Critical Configurations for Spatial Resection ................................................ 53
    6.6.2 Critical Configurations for the Estimation of the Metric Scale ...................... 55
6.7 Implementation .................................................................................................. 55
    6.7.1 Initial Phase .................................................................................................. 55
    6.7.2 Measurement Phase ..................................................................................... 56
7 SYSTEM EVALUATION ......................................................................................... 59
  7.1 Evaluation Set-Up .............................................................................................. 60
  7.2 Accuracy and Precision ..................................................................................... 61
    7.2.1 Accuracy ...................................................................................................... 61
    7.2.2 Precision ...................................................................................................... 63
    7.2.3 Sensitivity Analysis .................................................................................... 64
7.3 Time Statistics .................................................................................................... 67
7.4 Performance Summary ....................................................................................... 67
8 CONCLUDING REMARKS ................................................................................ 69
  8.1 Conclusion .......................................................................................................... 69
    8.1.1 Concept ........................................................................................................ 69
    8.1.2 Calibration ................................................................................................... 69
1 Introduction

1.1 Aim of this Work

The aim of this thesis was the development of an automatic, inexpensive Camera and Laser-Based Indoor Positioning System (CLIPS) capable of providing continuous three-dimensional positions in the mm-range of a mobile camera for each indoor environment. The device shall not require cost intensive, high-precision mechanics, or sophisticated set-ups, as claimed in Mautz (2009). In particular, the work incorporates an acceleration of the image processing and camera pose estimation to enable high position update rates. Further, a solution for the introduction of a metric scale must be found and implemented. Finally, the performance and limitations of CLIPS must be assessed.

1.2 Motivation

The emergence of new optical indoor positioning technologies facilitates a variety of new applications in the guidance of individuals, machines or tools in potentially every indoor environment. Once an individual or object is equipped with a mobile camera, its position can be determined instantaneously with accuracies in all ranges, i.e. from millimetre to meter level. Accordingly, the determined position can be used for pedestrian navigation in areas of healthcare, disaster management, shopping or conference guiding. In industrial environments, at construction sites or in logistics, optical indoor positioning systems offer a flexible and favourable method for the location and guidance of machines, tools or workers. For example, the camera’s position and orientation can be used to control a robot’s location in its working volume (Boochs et al., 2010; Breuckmann, 2012), to determine the position of a forklift with respect to the goods (Trebilcock, 2011) or to guide tools precisely to the focus of their application.

In comparison to geodetic 3D positioning systems like total stations, laser trackers or iGPS, optical indoor positioning systems have the potential to provide automatic, flexible and inexpensive solutions for the continuous positioning and navigation of mobile devices in every indoor environment. Additionally, they are driven by the idea to obviate sophisticated set-ups and complicated mechanics in order to ease the operator’s work. Although, most optical indoor positioning systems are at a development stage, “Cameras are becoming a dominating technique for positioning …. The success of optical methods originates from improvement and miniaturization of actuators ... and particularly advancement in the technology of detectors ... In parallel there has been an increase in data transmission rates and computational capabilities as well as profound development of algorithms in image processing.” (Mautz, 2012).
The academic research community as well as companies are called to push the development and to satisfy the demands of a growing indoor positioning market (Tilch and Mautz, 2012). In this context, the idea underlying CLIPS was born. First described by Dr. Mautz (2009), it was proposed the Swiss National Science Foundation (SNF). Therein, the continuous pose estimation of a mobile camera with respect to a laser projector is proposed. Such an optical indoor positioning system, which provides camera positions and orientations, facilitates a range of applications. For example, if a camera is attached to a forklift and pointed to the ceiling, which is the projection surface, the forklift’s position in a warehouse can be determined to navigate it on the shortest route to the goods. Similar applications for shopping carts are conceivable. Trajectories of shopping carts can be recorded and evaluated to optimize customer guidance and the placement of goods (Schäfer, 2003; Wunderlich, 2006). If the camera is attached to a headset, its position and orientation can be used to support head motion tracking for simulations (NaturalPoint, 2012) or to compensate head motion in medical imagery (Hossbach, 2012). In industrial areas, CLIPS can be used as a tactile measurement tool for reverse engineering and inspection tasks, if a tip is attached to the camera. If the camera is combined with a tool, CLIPS can be also used to guide workers at a construction site to a target location.

1.3 Outline

This work incorporates eight chapters about different aspects of the development of the optical indoor positioning system CLIPS. Subsequent to the first introductory Chapter, a categorization of optical indoor positioning systems is given. Therein, different types of current optical systems are presented. In the third Chapter, the basic concept of CLIPS, its construction, and the particular components are described. The fourth Chapter is dedicated to the calibration of the camera and the projector. It contains calibration methods for both parts as well as their calibration results. Chapter 5 is devoted to the detection and identification of the red and green laser spots in the camera images. Hereby, different methods for the spot detection are investigated and compared with each other. Subsequent to the spot detection, two different spot identification algorithms are presented. For both aspects, the best algorithm is chosen and its implementation in CLIPS is briefly described. In Chapter 6, the estimation of the camera pose is addressed. It deals with the provision of approximate values for the relative orientation and their refinement by least squares minimisation techniques. Additionally, the focus is laid on the introduction of a metric scale to provide positions in a metric system. At the end of Chapter 6, a flow chart of the implemented pose estimation algorithm is given. The system’s performance according to the parameters accuracy, precision and update rate is evaluated in Chapter 7. Chapter 8 closes this work with conclusions regarding to the optical indoor positioning system CLIPS and gives an outlook for future works and developments.
2 Categorization of Optical Indoor Positioning Systems

This section focuses on optical methods, which are becoming an interesting technique for positioning and navigation. For a coarse classification of optical indoor positioning systems, it is sufficient to distinguish between those systems that determine the pose of a mobile camera and those systems that determine the position of objects using at least one static camera. A comprehensive overview of older works in the area of optical robot navigation is given by Dezouza and Kak (2002). They distinguished three navigation classes based on the provided environmental model. This model could be a map, a building model, or neither.

A more recent study of current optical indoor positioning systems has been carried out by Mautz and Tilch (2011) and Mautz (2012). Pursuant to these studies, the manner by which reference information is obtained is regarded as a decisive characteristic of these systems. Reference information means any kind of information that is necessary to establish a relationship between the relative positional changes of a moving object, (e.g., a camera) and a superior coordinate system. Such a coordinate system can be represented by a room, a building model or a frame of deployed targets. This chapter is a revision of the study that has been carried out by Mautz and Tilch (2011), and distinguishes 6 different classes, as shown in Figure 2.1.

![Figure 2.1: Classification of optical indoor positioning systems.](image)

Despite of the differences among these systems, they share a common key problem to solve. The camera pose or the position of an object in three-dimensional space has to be determined whereas the only observation is a 2D measurement of image coordinates on a CCD sensor.
2.1 Reference from Building Models

The use of building models and floor plans containing objects and features with semantic and geometric information can be one possible solution for the introduction of reference information. This includes, for example, windows, doors, chairs, tables or any other building interior and their position information. The basic idea is to extract objects and features in the image data and match them with the objects and features in the building models or floor plans so that the camera pose can be inferred. The advantage of such approaches is that there is no need for additional sensors like beacons or installations like reference points. However, the building models with their objects and features need to be up-to-date since the building interior can be varied over time.

Kohoutek et al. (2010) presents a method to estimate the pose of a Range Imaging Camera (RIM) in a room by means of a digital spatio-semantic interior building model, based on CityGML (City Geography Markup Language). Thereby, the 3D point cloud that is obtained from the RIM camera is compared with the CityGML model of the room, where each room is modelled with its openings, permanent fixed objects and furniture. The pose estimation of the RIM camera is separated in three steps. First, the objects like chairs, tables, windows or doors and their geometry need to be extracted out of the 3D point cloud. Having these objects, a coarse pose can be estimated by comparing the objects and their geometry with the room interior of the CityGML model. Thus, the room where the camera is located can be identified. In the third and last step, the fine camera pose is determined on a dm-level based on a technique that combines trilateration and spatial resection.

Another way to provide a building model is given by the laser scanning of the indoor environment, as presented in Böhm (2007). Once the indoor environment is represented by a point cloud, distinctive feature points with 3d coordinates can be extracted. In this manner, a Harris corner detector is applied on the intensity images and the corresponding coordinates from the point cloud are assigned. If these feature points are detected in the images, the camera pose can be determined by spatial resection.
A similar approach of camera pose estimation is given by Hile and Borriello (2008). Contrary to Kohoutek et al. (2010), they make use of an off-the-shelf phone camera and a floor plan instead of a range imaging camera and a CityGML model. Extracted features in the camera images are matched with corresponding features on a floor plan, and the location of the phone camera can be computed by using this correspondence. To simplify the finding of the correspondence between the image features and the floor plan, the search area in the floor plan needs to be restricted. For this purpose, a Wi-Fi location system is used for coarse camera pose estimation. Finally, location based information can be overlaid.

Kitanov et al. (2007) present a mobile robot, having the ability for self-localization in an indoor environment. For this purpose, the scene is captured by a camera, which is mounted on the robot. Subsequent to the image acquisition, image features are extracted and compared with an underlying 3D model of the indoor environment. As a result of this comparison, the pose of the robot is obtained. Additionally, position finding is supported by an odometer to stabilize the robustness.

Schlaile et al. (2009) determine the relative orientation of consecutive images of a UAV-mounted (Unmanned Aerial Vehicle) camera to aid its IMU (Integrated Measuring Unit). Since the global scale factor is not determinable with relative orientation techniques, the reference has to be separately introduced. In this case, edges are extracted in the images and compared with an edge model of the indoor environment. Finally, the camera pose is adjusted until the edges in the images and the edges in the model fit together.

### 2.2 Reference from Images

In this class of optical indoor positioning systems, a sequence of consecutive and previously recorded images represents the reference information along a certain and well known route. The pose estimation relies on a comparison between this sequence and an image that is captured in a second run. The drawback of such systems lies in the additional learning run that is required to generate the sequence. Moreover, comparison is complicated by blur effects due to sensor movement. However, it is a simple and robust method to estimate a camera’s pose.

![View Sequence](image1.png) ![Current View](image2.png)

Figure 2.3: Reference introduction via an image sequence. In order to estimate the camera pose a matching between the images of the sequence and the current view has to be carried out.
Ido et al. (2009) present a view-based indoor navigation approach for a mobile robot utilizing a view sequence for pose estimation. This view sequence can be regarded as a set of consecutive images that have been previously captured along a certain route. This is called a “recording run.” Additionally, relational information like “Forward,” “Left” or “Right” is attached to each image of this sequence. This information can be used to guide the robot along the route. In the “autonomous run,” the robot captures front view images of his environment while he is following the route. Each image is compared with the images in the view sequence via template matching. The similarity measure between the image of the “autonomous run” and an image of the view sequence is the normalized correlation coefficient. If this coefficient is above a certain threshold then the robot uses the additional information to define his next action. For example, if the information is “Forward,” then the robot is moving forward. If it is “Left,” then it is turning left. However, the presented approach relies on a sequence “…for a one-way navigation…” (Ido et al., 2009) of the robot. To overcome this restriction in navigation, a view sequence of omnidirectional views can be used, as proposed by Matsumoto et al. (2003).

2.3 Reference via Established Camera Frameworks

In this class of optical indoor positioning systems, the coordinate system is established by a framework of at least two digital video cameras. Given the cameras’ location and orientation, the pose of a mobile object can be estimated via tracking and the exploitation of multiple-view photogrammetry methods. Mostly, these objects consist of active targets like infrared LEDs. This technique is applied for example in motion capture studios or in areas of industrial production to precisely locate sensor mounts like heads of industrial robot arms. The accuracy of the pose estimation lies in the range of 1 mm and better. However, mostly a preliminary, sophisticated and static installation is a presumption to achieve this accuracy level.

For instance, the indoor positioning system of Boochs et al. (2010) uses an established camera framework to precisely locate the head of an industrial robot arm. For this purpose, four digital cameras are mounted at the corners of a 4 x 4 metres frame which is located two metres behind the robot. The mutual location and orientation of the cameras is given, and they thus establish the superior coordinate frame to which the pose of the robot’s head relates. This set of cameras permanently observes a spherical target with evenly distributed LED’s and which is mounted on the head. The indicated accuracy is in the range of 1 mm and better. A considered application is the utilization of the head as a sensor platform for optical sensors like 3D scanners or multispectral cameras.

A very similar but more mobile optical metrology system has been developed by the Norwegian manufacturer METRONOR AS (2007). Here, the system consists of two cameras and a mobile probing unit with embedded active targets. During the measurement phase, the cameras are mounted on tripods, and track the mobile probing unit with an accuracy of better than sub-millimetre. Thus are determined the six degrees of freedom (6 DOF) of the probing unit. A possible application is given by Breuckmann GmbH (2012). This firm combined its fringe projector stereoSCAN 3D with Metronor DUO. This results in the portable fringe projector naviSCAN 3D attached, (e.g., on an industrial robot arm in order to cover a larger measurement area). A possible application is the inspection of surfaces. Similar systems are, for example, the HandyPROBE or the MetraSCAN system from CREAFORM (2012).
2.4 Reference from Deployed Targets

Most systems establish the relationship to a superior coordinate system by the deployment of artificial targets on walls, ceilings or any interior object. Thus deployed, the target is well known in the superior system. In contrast to natural features in images, artificial targets do not suffer under varying illumination conditions. Mostly, they are made of infrared retro-reflective materials facilitating and simplifying their detection. Additionally, also non-reflective targets with a high contrast relative to the environment are used. Furthermore, the target’s shape is well-known, which additionally increases the robustness of target detection. Such systems typically encode their ID information with a barcode, an arrangement of concentric sectors, or an arrangement of coloured dots (see Figure 2.4).

![Figure 2.4: Illustration of coded targets. They either encode the ID information with a barcode, an arrangement of concentric sectors, or a coloured dot configuration.](image)

One decisive drawback of coded targets can be seen in their physical deployment. At first, they have to be installed at best-suited locations on walls and ceilings. Then, a relationship to the superior system must be established by determining their coordinates within this system. Although a one-time process, it is time consuming. Summing up, however, the deployment of coded targets allows for robust and accurate detection and thus for robust and accurate camera pose estimation. This usefully explains; why the majority of commercial optical indoor positioning systems can be found in this class.

Maye et al. (2006) present a commercial optical odometry-based indoor positioning system using an optical mouse sensor, an electronic compass and IR reflective cat-eye targets on the ground. Whereas translational changes can be determined with the mouse sensor, orientation has to be estimated separately by the compass. Because of the systematic error of the mouse sensor, the path error accumulates over time. As a consequence, the error has to be corrected by setting the sensors back on track. For this reason, IR reflective targets are deployed at well-known positions on the ground. Possible applications can be seen in locating shopping carts, robots or the assistance of disabled persons. Practical tests have revealed problems with linoleum surfaces. The positioning accuracy is about 1% of the whole path length.

Lee and Song (2007) deploy infrared reflective targets on the ceiling to determine a robot’s pose. These targets have a right-angle triangular shape organized in 9 sectors. The corner sectors serve for the orientation estimation whereas the other six sectors encode the target’s ID. Only sectors that are used for orientation estimation and identification are coated with an infrared reflective material. In order to detect the targets in the images, the scene is captured twice, once with IR illumination and once without IR illumination. By taking the difference of both images, the targets can be detected. Therefore, the measuring rate is reduced by the factor 2. Additionally, a pan-tilt motion of the camera makes it possible to capture a larger scene by stitching nine single images to one large image. The reported accuracy is in the sub-decimetre domain.
Mulloni et al. (2009) developed an indoor positioning system using an off-the-shelf camera phone and 2D barcodes. The targets, each of them coded with a unique ID, are deployed, for example, on walls, posters or other objects in the indoor environment. The location of each target is marked on a map and stored in a marker location database. If the user captures a target, the six degrees of freedoms (6 DOF) can be estimated with centimetre level accuracy. Additionally, location-based information and/or the route to a destination can be displayed. This system was successfully deployed on several conferences as a conference guide tool and is now a commercial product.

Trebilcock (2011) describes the optically-based real-time locating system Sky-Trax from Sky-Trax Inc. It was developed to track lift trucks and pallets in a warehouse of Genco ATC, a third-party logistics provider in the US. For this reason, 2D bar codes are installed on the ceiling in a grid pattern. These targets are captured by a camera that is mounted on top of a lift truck. Via image processing methods, the exact location and speed of each truck can be estimated. Additionally, each bar code contains a dark L shape to facilitate the heading estimation of the vehicle. The reported accuracy is between one inch and one foot.

Similar to the Sky-Trax system, the StarGazer system of Hagisonic (2012) deploys coded targets on the ceiling of a room. However, the targets reflect IR light and encode the information by a 3 x 3 or 4 x 4 point pattern. An IR LED ring light on the sensor platform illuminates the scene which is captured by the IR camera. If a target is captured in an image, the pose of the camera can be determined within sub-decimetre accuracy. The intended application is the indoor localization of mobile robots or intelligent hoovers (vacuum cleaners, U.S.).

AICON 3D Systems (2011) developed an optical metrology system based on a camera and a deployed reference field of control points. For stationary applications, (e.g., in pre-calibrated rooms or bays), the reference targets are installed on walls and the ceiling. Mobile applications can be facilitated by using target panels. Like prior approaches, the robustness of target recognition is enhanced by using IR reflective targets. In consequence, a field of IR LED’s around the camera lens illuminates the targets. Additionally, the camera unit is equipped with a measuring tip allowing for tactile measurements of object points with sub-millimetre accuracy. This system is used for example in automobile crash experiments for the pre- and post-crash documentation.

2.5 Reference from Projected Targets

The projection of reference points or patterns spares the physical deployment of targets in the environment, making this method economical. For some applications, the mounting of reference markers is undesirable or not feasible. Optionally, infrared light can be projected to attain unobtrusiveness to the user. In contrast to systems relying only on natural image features, the detection of projected patterns is facilitated due to their distinct colour, shape and brightness. The principle of an inverse camera (or active triangulation) can be exploited where the central light projection replaces the optical path of a camera. The main disadvantage of active light-based systems is that both the camera and the light source require direct view on the same surface.
The work of Habbecke and Kobbelt (2008) presents an optical system for three-dimensional reconstruction of indoor scenes. It consists of a digital video camera mounted on a tripod and a mobile projector. This projector is equipped with 20 arbitrarily aligned laser pointers casting "...an irregular set of (laser) rays into a scene..." (Habbecke & Kobbelt, 2008). The camera's location is fixed because of the tripod. For the depth-map reconstruction, the laser spots are captured by the video camera while the projector device is moved through the scene. Via a least-squares adjustment, the projector pose can be found with respect to the video camera. Given the projector’s pose, the three-dimensional coordinates of the laser spots can be determined in the object space by triangulation. Thereby, the reported average update rate is about 7 Hz.

Another system that uses projected targets was developed by Köhler et al. (2007), and is called TrackSense. Hereby, a grid of projected laser lines is captured by the digital camera. Additionally, projector and camera are mounted on the same rig and are calibrated with respect to each other. Therefore, the relative orientation between the camera and the projector is known. Once the scene is captured by the camera, the lines and their intersections can be detected in the image with sub-pixel accuracy. Using these intersections, pairs of corresponding points can be established with help of the known relative orientation, and, thus, the three-dimensional coordinates of the grid intersections can be determined via intersection. Having these 3D coordinates, planes can be identified by a RANSAC (Random Sample Consensus) algorithm. If only one plane (wall) can be identified, only the distance of the TrackSense unit to the wall can be determined. With two walls, the two-dimensional position and orientation with respect to the ground plane can be estimated. Full 3D information about the pose and the orientation of the TrackSense unit can only be obtained if three orthogonal planes are observed. The indicated accuracy is about 4 cm in position and 2 degrees in orientation with a precision of 3 cm in position and 1 degree in orientation.

A similar setup is chosen by Popescu et al. (2006) with their ModelCamera system. This system consists of a digital video camera and a laser pointer that projects a 7 x 7 laser point pattern in the camera’s field of view by means of a diffraction grid. Both the camera and the laser projector are mounted on a rig. Additionally, the relative orientation between the camera and the projector is known through a previous calibration. But unlike Köhler et al. (2007), the main...
goal is not the pose estimation but the derivation of a three-dimensional model, such as a couch in an indoor environment. For that reason, 49 (7x7) depth values are obtained in each frame by intersection of the laser beams and the image rays of the corresponding laser spots. A complete model is derived by the colour and depth registration of all frames. The model quality is indicated to be in a range of 0.2 mm to 1 cm.

The CLIPS system, which is the subject of this thesis, can also be categorized into the group of Reference from Projected Targets. It consists of a projector and a mobile camera. The reference information for the camera pose estimation is obtained by the projection of a laser spot pattern on any surface. Through a one-time high precision calibration the 3D directions of these laser beams are known. Thereby, the laser device represents an inverse camera. The main functions of the projector can be summarized as, (a) the projection of flexible reference points and (b) the simulation of a second camera. When laser spots have been captured by a mobile camera, the relative orientation of the camera to the projector is determined by exploiting the concepts of stereo photogrammetry. As a major advantage, no high-precision mechanics or sophisticated set-ups are required, making the system a low-cost, mobile and easy-to-use device for high precision positioning.

A commercial indoor positioning system that exploits projected targets is the NorthStar system by Evolution Robotics (2012). It was developed to locate robotic hoovers or shopping carts in indoor environments. Here, a projector unit projects multiple infrared spots, each with a unique code, on the ceiling or any other surface. Once the spots have been captured by the camera mounted on the moving vehicle, its position and heading can be directly measured with sub-decimetre accuracy.

2.6 Systems without Reference

Systems that do not use any kind of reference information with respect to a superior coordinate frame are mostly used to detect changes in position and orientation of objects. For example, changes in the pose of a camera sensor, or position changes of an object observed by the camera.

A very simple indoor positioning system is proposed by Tappero (2009). He uses a CMOS image sensor with a resolution of 356 x 292 pixels, which is mounted on the ceiling and pointing downwards. The idea is based on the calculation of difference images whereas the first image shows only the room without the to-be-located object. For every consecutive frame, the difference image is calculated by subtracting the first image from the current frame. Image regions where grey values greatly differ from zero are evaluated, finding the region which represents most likely the object that has to be located. The reported position accuracy is in the domain of sub-metre.

The work of Muffert et al. (2010) describes how a trajectory of an omnidirectional camera system, based on relative orientation of consecutive images, can be obtained. The focus lies on the relative motion because of the missing absolute relationship to a reference system. But, without any kind of reference information, the pose estimation errors accumulate over time which leads to a drift of the current path with respect to the true trajectory. This effect is known for every kind of sensors, which merely rely on dead reckoning. The reported accuracy of the yaw angle lies between 0.1 gon and 0.2 gon for an observation period of 40 seconds.
3 Architecture of CLIPS

The fundamental idea of CLIPS is the real-time pose determination of a mobile camera with respect to a stationary projector to realize sub-mm accuracy in 3D position and sub-second accuracy in the rotation angles.

The projector is designed to project a reference field of 16 red and 36 green laser spots on any surface in an indoor environment.

A half-sphere with 16 red laser pointers is arranged in such a way that the lasers have a common “virtual” origin. This origin is defined by the intersection of all red laser beams. The origin is also the zero-point of the local coordinate system. Thus, the projection of the red spots can be modelled by a central projection, similar to the mapping model of a pin-hole-camera. Additionally, the direction of the red laser rays is precisely known through a previous calibration. Once the mobile camera has captured the red laser spots, its relative pose with respect to the projector can be estimated by exploiting the concepts of stereo photogrammetry. Figure 3.1 illustrates the basic concept of CLIPS.

Since the relative camera orientation only allows determining the camera position up to scale, a separate introduction of a metric scale becomes necessary. Therefore, the projector was upgraded by a scale cross with a diameter of 50 cm to allow eccentric mounting of additional laser pointers. Three particular arms are equipped each with a green laser pointer, which is parallel to the projector’s z-axis. The fourth arm has two green laser pointers, to dissolve the laser spot pattern symmetry.
Their offset vectors to the origin and their direction vectors are precisely known through a one-time calibration. Once the green laser spots have been identified in the image, the metric scale can be derived.

The core component of CLIPS is the projector device that mounts 16 red and 5 green laser pointers. Via two chipsets, these lasers can be individually controlled. Two power supply units provide the necessary current for the red and green laser pointers. The applied digital camera is an industrial CCD camera from Allied Vision Technologies (AVT), which is equipped with a standard lens. Via FireWire 1394a, the camera is connected to the laptop on which the pose estimation algorithms are running.

The remainder of this chapter is organized according to the components of CLIPS. In Section (3.1), the digital camera AVT GUPPY GF 080 C and the Pentax camera lens are described. Section (3.2) offers an overview of the applied red and green lasers. The Sections (3.3) and (3.4) are dedicated to the projector and the laser control unit.

### 3.1 Camera and Camera Lens

According to the basic concept of CLIPS, the camera can be regarded as a measuring tool. The applied digital camera is a GUPPY GF 080 C camera from Allied Vision Technologies.

| Table 3.1: Sensor Parameters of AVT GUPPY GF 080 C (AVT, 2009) |
|-----------------------------|------------------|
| **Sensor**                  | IT CCD ICX204 AK |
| **Width**                   | 1032 pixel       |
| **Height**                  | 778 pixel        |
| **Pixel size (mm)**         | 0.00465          |
| **Colour Pattern**          | Bayer layout (RGGB) |

The communication with the PC as well as the power supply is managed via a 1394a (Firewire) interface.

For the camera lens, a Pentax unit with an adjustable f-number between 1.2 and 16 has been chosen. The f-number denotes the ratio between the focal length and the aperture’s diameter. It is indirectly proportional to the number of photons, which are falling onto the sensor. For a high depth of field, an aperture with \( f \)-number = 11 has been chosen. A low sensor exposure time has been chosen to reduce the influence of ambient light, to prevent image blur, and to increase the frame rate. Unfortunately, a low exposure time leads to flickering laser spots in the video stream, which can affect the accuracy of centroid estimation. Pixels require a certain amount of photons to register a colour value. If the exposure time is too low, periphery pixels don’t register enough photons for delivering a colour value. For a strong image signal and a reduction of spot flickering, the sensor gain has been increased. Thereby, the image signal but also the noise is multiplied. Therefore a moderate gain of 380 has been chosen. Finally, focus and aperture can be locked to guarantee stable calibration parameters of the camera.
3.2 Laser Pointer

Laser light is of coherent nature, temporally as well as spatially. Temporal coherency means that emitted wave packages have the same wavelength. Spatial coherency describes the fact that two emitted wave packages show a constant phase difference.

Thus, wave packages and fronts can be considered to be parallel. With these excellent properties, lasers are natural selections for the projection of monochromatic spots. For that reason, we decided to use a collection of red and green lasers for reference field projection. With the red lasers a reference field is projected to enable the pose estimation of the digital camera. The green lasers are solely used to introduce the system scale.

Operating with lasers always raises questions of eye safety. Because of the large projector opening angle of 90°, the operator might glimpse into a beam. To guarantee eye safety, CLIPS was designed with lasers of class 2 (1mW power). An overview of the different laser classes is provided in the appendix. The characteristics for both, the red and the green laser are summarized in Table 1.

<table>
<thead>
<tr>
<th>Manufacturer</th>
<th>Red Laser-LEDs</th>
<th>Green Laser-LEDs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manufacturer</td>
<td>Zwei Brüder Optik</td>
<td>Picotronic</td>
</tr>
<tr>
<td>Model</td>
<td>V9 Laser</td>
<td>LFD532-1-3</td>
</tr>
<tr>
<td>Major Semi Axes (mm)</td>
<td>1.5</td>
<td>1.5</td>
</tr>
<tr>
<td>Minor Semi Axes (mm)</td>
<td>1</td>
<td>1.5</td>
</tr>
<tr>
<td>Divergence for Major Semi Axis (mrad)</td>
<td>n/a</td>
<td>0.5</td>
</tr>
<tr>
<td>Divergence for Minor Semi Axis (mrad)</td>
<td>n/a</td>
<td>0.5</td>
</tr>
<tr>
<td>Intensity (mW)</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

According to Table 1, the red spots show an elliptical shape whereas the green spots have a circular shape. For both lasers, the intensity distribution is similar to a 2d Gaussian distribution with the maximum in the spot centre and a decreasing gradient to the spot boundary. However, two phenomena, namely diffraction and interference influence the spot appearance such that it is not possible to project real focused spots. Due to the pinhole in the laser coating, arriving parallel wave fronts will be diffracted. Together with interference, this effect leads to a beam divergence and circular diffraction pattern around the spot. Additionally, subjective speckle pattern can occur caused by interferences of reflected waves in the image plane.

3.3 Projector

As previously described, the central idea of CLIPS is the real-time pose estimation of a mobile camera with respect to a stationary projector. This projector enables the projection of a reference field of laser spots on any indoor surface and, in addition the simulation of a second camera. Thereby, the projector consists of two parts.

The first part is a half-sphere with a diameter of 12 cm and 16 radial arranged boreholes. All holes have an approximate common origin in the centre of the half-sphere and are assembled with red laser pointers. Assuming a spherical coordinate system with its origin in the centre of the sphere, the boreholes are evenly distributed over two circles of latitude at 45° and at 75°.
The circle at latitude 45° contains 12 boreholes whereas the remaining 4 boreholes are distributed over the circle at latitude 75°, as shown in Figure 3.2.

![Figure 3.2: Mount for the red and green laser pointers. On the left, the mount for the red lasers is shown. The two images on the right illustrate the scale cross.](image)

At the beginning of Chapter 3, it was stated that the metric scale cannot be established by the previously mentioned half-sphere. For that reason, a second cross-shaped projector with a diameter of about 50 cm has been constructed (see Figure 3.2). Each particular arm contains 9 boreholes that are arranged in a 3-by-3-matrix structure with a grid distance of about 17 mm. These boreholes can be assembled with green laser pointers.

Both parts can be combined such that their centres are vertically aligned. Additionally, the scale cross has a 5/8 inch screw thread in its centre in order to set up the whole projector on a geodetic tripod.

### 3.4 Controller

Each laser pointer can be independently switched on and off via programmable Basic stamps from Parallax Inc. One basic stamp is used for the controlling of the red lasers and five stamps are used for the controlling of the green lasers. All stamps are mounted at the rear side of the scale cross, as shown in Figure 3.3. Via a Serial-to-USB interface, the stamps are connected to the computer. The programs for the laser control have been written in a simple Basic dialect and can be uploaded to the stamps’ EEPROM (Electrically Erasable Programmable Read Only Memory) – a memory that can be erased and reprogrammed.

![Figure 3.3: Projector Controller. The rear side of the scale cross serves as a mount for the basic stamp controller.](image)

The basic stamps also guarantee the power supply. The red laser controller requires an external 12 Volt power supply, and the green laser controllers require a 5 Volt power supply.
4 Calibration

Calibration is a crucial step in areas of metrology and production, since it influences the system performance of metrology devices significantly. The International Organization for Standardization (2007) defines the term calibration in ISO/IEC Guide 99:2007 as...

"...operation that, under specified conditions, in a first step, establishes a relation between the quantity values with measurement uncertainties provided by measurement standards and corresponding indications with associated measurement uncertainties and, in a second step, uses this information to establish a relation for obtaining a measurement result from an indication."

Practically speaking, calibration can be considered as the process to determine the deviation of the current state of a measuring device from a given model. In our case, it means the determination of the deviation of a digital camera from the physical model of a pin-hole-camera, as described in Luhmann (2010). Further, the precise directions of the laser beams and their common intersection point must be known in order to apply the mathematical model of a central projection.

In the first Section (4.1), the calibration of the digital camera according to Brown's model is presented. Here, the focus lies on geometric camera calibration and not on radiometric issues like intensity decreases to the image borders. In the second Section (4.2), deviations from the ideal laser spot intensity distribution due to diffraction, and varying surface properties are discussed. Finally presented are the calibration of the projector, the calibration parameters, and their stability (4.3).

4.1 Camera Calibration

Generally, an image sensor is an array of photon sensitive picture elements that are arranged in a matrix of columns and rows. Via a camera lens system, a real-world scene is mapped on the sensor. Thus, the basic mapping process is based on a pinhole camera model with deviations due to manufacturing imperfections, like uneven or oblique sensors, or distortions of the camera lens system. Since the primary task in photogrammetric image analyses is the accurate measurement of image coordinates, these deviations from the ideal central projection of a pin-hole-camera and the parameters of the inner orientation have to be considered. For that reason, a geometric calibration is required. A detailed overview of geometric camera calibration is given in Luhmann (2010).
Additionally, lens aberrations like chromatic and spherical aberration, coma, astigmatism and lens distortion can occur. However, special camera lenses, lens systems, or the use of apertures can significantly minimize these lens aberrations. Since the applied lens system shows small aberrations and since it is additionally equipped with an aperture, these kinds of aberration won’t be further considered for photogrammetric camera calibration.

4.1.1 Interior Orientation

The inner orientation of a camera denotes the focal length $f$ and the location of the principal point $P(x_0, y_0)$. By means of this point and the focal length $f$, an image coordinate system can be defined with its centre in $P$. The x-axis is defined to be parallel to the column-axis and the y-axis is antiparallel to the row-axis of the sensor. The z-axis completes this system to a right-handed coordinate system. Thereby, the principal point doesn’t have to coincide with the image centre due to an uneven or oblique image sensor. In this case, the principal point corresponds to the orthogonal projection of the projection centre on the image plane. The relationship between the pixel coordinate system and the image coordinate system is given by

\[
\begin{pmatrix}
  x_c \\
  y_c \\
  z_c
\end{pmatrix} =
\begin{pmatrix}
  s & 0 & -x_0 + dx \\
  0 & -s & -y_0 + dy \\
  0 & 0 & -f
\end{pmatrix}
\begin{pmatrix}
  u - u_{\text{max}} / 2 \\
  v - v_{\text{min}} / 2 \\
  1
\end{pmatrix}
\]

with the image coordinates $(x, y, z)^T$, the pixel coordinates $(u, v, 1)^T$, the maximum number of pixels $(u_{\text{max}}, v_{\text{max}})$ in u- and v-direction and the pixel size $s$. The units of the principal point coordinates $(x_0, y_0)$, the focal length $f$ and the pixel size $s$ are given in millimetres.

4.1.2 Lens Distortions

Subject to photogrammetric camera calibration are lens distortions due to decentered lenses, oblique or uneven image sensors.

Radial Symmetric Distortion

Due to radial symmetric distortion, straight lines in a scene appear as curves in images. Therefore, two kinds of barrel distortion and pincushion distortion can be distinguished. Depending on the radial distance $r$ of an image point to the principal point $P_0$ the radial displacement

\[
dr_{\text{rad}} = \sum_{i=1}^{n} K_i \cdot r^{2i+1}
\]

of an image point can be expressed by a power series with the distortion parameters $K_i$, as shown in Brown (1968), (1971). Conventionally, the first three high-order terms $K_1$, $K_2$ and $K_3$ are sufficient to approximate the radial distortion (Luhmann, 2010). Thereby, the sign of parameter $K_1$ defines the distortion type as a) a pincushion distortion if $K_1 > 0$ or b) a barrel distortion if $K_1 < 0$. 
The corresponding corrections, to obtain the undistorted image coordinates, can be calculated as follows:

\[
dx_{rad} = x \frac{dr_{rad}}{r}, \quad dy_{rad} = y \frac{dr_{rad}}{r}.
\]  

(4.3)

Figure 4.1 illustrates the displacements in x- and y-direction. The green dots denote the undistorted image points that experience a displacement into the directions of the red vectors resulting in the distorted blue dots. Herein, a pillow distortion of the applied camera lens can be observed.

**Decentring Distortion**

In Brown (1968), decentring distortion is described as a systematic mapping error due to de-centred lens elements. Thereby, the two components of a) asymmetric radial distortion and b) tangential distortion can be distinguished and expressed by the mathematical equations

\[
dx_{dec} = B_1(r^2 + 2x^2) + 2B_2xy[1 + B_3r^2]...
\]

\[
dy_{dec} = B_2(r^2 + 2x^2) + 2B_1xy[1 + B_3r^2]...
\]

(4.4)

with the coefficients \(B_i\). Conventionally, coefficients with order 3 or higher are neglected.

Figure 4.1 illustrates the displacements in x- and y-direction. It can be seen, that the undistorted image points (green) experience a shift to the right respectively a shift to the right bottom of the image, which results in the distorted spots (blue). Furthermore, the influence of
the image radius \( r \) on the shifts can be seen. The farther the image point is located from the principal point the larger is the corresponding shift.

In practical experiments, Fraser (1997) observed a high correlation between the distortion parameters \( B_1, B_2 \) and the location of the principal point \( P \). Therefore, he stated to suppress the parameter estimation for the decentring distortion.

**Affinity and Shearing**

According to Fraser (1997), not only the camera lens system contributes to image distortion but also the image sensor. Due to the non-orthogonality of the image axis and a different scaling, affinity and shearing effects can occur. In Luhmann (2010), the correction terms are given by

\[
\begin{align*}
    dx_{aff} &= C_1 x + C_2 y \\
    dy_{aff} &= 0.
\end{align*}
\]  

Figure 4.1 illustrates the displacement vectors (red) that describe an affine distortion of the image.

**Total distortion**

The accurate measurement of image coordinates in sub-pixel space is one of the most important working steps in the photogrammetric processing chain, since the quality of the derived data highly depends on the quality of the measured coordinates. Furthermore, the epipolar constraint for camera pose estimation is only valid for undistorted images. Thus, all distortions have to be considered in the coordinate measurement. All distortions can be summarized to the total distortion

\[
\begin{align*}
    dx &= dx_{rad} + dx_{dec} + dx_{aff} \\
    dy &= dy_{rad} + dy_{dec} + dy_{aff}
\end{align*}
\]  

with its components in \( x \)- and \( y \)-direction.

In Figure 4.1, the total distortion for the applied camera-lens-system is shown. The dominant influence of the radial distortion can be seen. A comparison between radial distortion and total distortion hardly reveals any difference. As a consequence, only the determination and the correction of the radial symmetric distortion are necessary for the applied camera-lens-system.
4.2 Laser Calibration

Because of the coherent nature of laser light, spots are monochromatic with 2D Gaussian intensity distribution. Thereby, the intensity decreases with increasing distance to the spot centre. In Figure 4.2, the assumed intensity distribution and an exemplary laser spot from an image are shown.

![Figure 4.2: Comparison of a laser spot with an ideal 2D Gaussian intensity distribution (left) and the intensity distribution of a captured laser spot (right).](image)

However, the intensity distribution also deviates from the assumed intensity distribution due to varying surface properties, as roughness and absorbance or due to the laser incidence on the projection surface. A thorough investigation of laser-surface interactions has been carried out by Ingensand (2006) and Schäfer (2011).

Mostly, monochromatic light is backscattered to all directions when it hits a material. "In addition, ... materials (can be) invaded by the laser ray and the laser ray is also refracted and reflected in the material itself" (Ingensand, 2006). Here, speckle patterns occur due to the interference of the diffuse backscattered laser waves. With the application of a wide opened aperture, this effect can be minimized or even eliminated, as stated in Bauer (1991). Since this requirement is contrary to the previously claimed small aperture for a high depth of focus (see Chapter (3.1)), not the smallest aperture F16 but a medium aperture with F11 is chosen.

Since surface backscattering is mostly diffuse, the spot intensity depends on the incident angle $\theta$ between the laser ray and the normal vector to the surface. Lambert's cosine law for isotropic reflection expresses this effect:

$$I_{\text{reflected}}(\lambda) = I(\lambda) \cdot k_d(\lambda) \cdot \cos(\theta)$$  \hspace{1cm} (4.7)

Here, $k_d(\lambda)$ denotes the diffuse reflection coefficient and $I(\lambda)$ denotes the incident laser light intensity. Due to the cosine, the reflected intensity has its maximum for an incidence angle of $\theta = 0^\circ$. With an increasing incidence angle, the reflected intensity decreases until it vanishes with $\theta = 90^\circ$. Assuming a projection of the laser spots on a planar surface such that the projector’s $z$-axis and the normal vector of the projection surface coincide, the reflected intensity for laser spots of the inner ring decreases by factor 0.97, whereas the intensity for the second ring laser spots decreases by factor 0.71. As a consequence, spots on the second ring have a lower intensity than spots on the inner ring.
Further, parallel wave fronts of laser light are diffracted when they pass the laser pointer opening. The intersection of the resulting, diffracted wave fronts with the projection surface leads to characteristic diffraction patterns, as illustrated in Figure 4.3.

![Illustration of diffraction](image)

**Figure 4.3: Illustration of diffraction.** Parallel wave fronts are diffracted by an aperture (A). The intersection of the new wave with a projection surface (S) results in the characteristic diffraction pattern. Right: Image of diffraction effects for the applied red lasers at a distance of 52.

All previously mentioned effects shape the laser spot intensity distribution. Due to the fact that intensities serve as weights for the centroid estimation in Section (5.2.2.3), laser spot centres are influenced by the laser-surface interaction. However, these interactions vary for each spot and for each surface material. Therefore, it is not straightforward to make any rules about the influence on the centroid estimation for all laser spots. For that reason, intensity deviations due to surface properties or diffraction are not further considered.

### 4.3 Projector Calibration

Although the laser projector has been manufactured accurately, the geometry of the laser beams deviates considerably from the nominal geometry. This is due to the unknown position of the laser unit within the coating. During the manufacturing process, the diodes are plugged into the coating without alignment, and, thus, the directions of the laser beams deviate from the nominal directions. For that reason, a precise calibration of the laser beam directions becomes necessary. This can be carried out via a one-time calibration. The following section summarizes a more detailed description of the projector calibration in Tilch (2009) and Tilch and Mautz (2012).

#### 4.3.1 Calibration Set-Up

The reconstruction of the laser bundle is based on a one-time calibration of the projector. The projector has a static set-up in front of a surface on which the laser beams are projected, as shown in Figure 4.4. The surface is then subsequently shifted. At each location, three-dimensional coordinates of the laser spots are determined by a theodolite measurement system or photogrammetric measurements.

![Calibration set-up](image)

**Figure 4.4: Illustration of the calibration set-up.** On the left, the reference field with the coordinated reference points can be seen. The black plane in the middle is the movable wooden plane with the projected laser spots. The projector is closely located on the right side of plane.
As a result of deploying multiple surfaces, each laser beam is represented by a set of 3D spot coordinates

\[ \mathbf{L}_{i,j} = (X_{L,i,j}, \ Y_{L,i,j}, \ Z_{L,i,j})^T \]  

which can be used for its 3D reconstruction. Index \( i \) denotes the number of the laser beam, whereas index \( j \) represents the projection surface. For the sake of robustness and reliability of the beam reconstruction, three different points for each laser beam have been measured.

### 4.3.2 Reconstruction of the Laser Beams

Based on the 3D coordinates of the laser spots, each beam is reconstructed by a set of three positions (since each beam was measured at three distances). In order to reconstruct a certain laser beam, a so-called principal component analysis is applied on the corresponding set of laser spots. With this kind of analysis, the inner precision of the data – in this case a set of 3D laser spots – is expressed by a set of uncorrelated variables, which are called principal components (Böker, 2008). For that reason, the eigenvalues and eigenvectors of the covariance matrix

\[
\Sigma = \begin{pmatrix}
\sigma_x^2 & \sigma_{xy} & \sigma_{xz} \\
\sigma_{yx} & \sigma_y^2 & \sigma_{yz} \\
\sigma_{zx} & \sigma_{zy} & \sigma_z^2
\end{pmatrix}
\]  

with their variances \((\sigma_x^2, \sigma_y^2, \sigma_z^2)\) and co-variances \((\sigma_{xy}, \sigma_{xz}, \sigma_{yz}, \sigma_{zx}, \sigma_{zy})\) have been estimated. Böker (2008) has shown that the eigenvalues can be interpreted as variances. Therefore, the eigenvector with the largest eigenvalue denotes the direction with the largest variation in the data. Assume that this direction vector represents the corresponding laser beam.

Then, each laser beam

\[ l_i; \mathbf{x} = \mathbf{a}_i + \lambda \ast \mathbf{b}_i \]  

can be defined by the unit direction vector \( \mathbf{b}_i = (b_{x,i}, b_{y,i}, b_{z,i})^T \) and the centre of gravity \( \mathbf{a}_i = (a_{x,i}, a_{y,i}, a_{z,i})^T \) of the corresponding laser spots. Here, the number of the laser beam is denoted by the index \( i \).

### 4.3.3 Definition of the Projector Coordinate System

The geometric model of an inverse camera consists of a projection centre \( \mathbf{O}_p \), as well as the laser beam directions with azimuth \( \text{Az}_i \) and zenith angle \( \theta_i \) and their offsets \( \Delta_{\text{offset}} \) to the projection centre.

To reconstruct the laser beam bundle, their apparent intersection point \( \tilde{\mathbf{O}}_p \) is determined via a least squares adjustment of the observation equation

\[ \mathbf{e}_i = (I_{3 \times 3} - \mathbf{b}_i \mathbf{b}_i^T) (\tilde{\mathbf{O}}_p - \mathbf{a}_i). \]  

where \( I_{3 \times 3} \) denotes an identity matrix and \( \mathbf{e}_i \) denotes the residual vector. Due to the laser beams’ skewness, point \( \tilde{\mathbf{O}}_p \) represents the point with the closest distances to all laser beams.
The term \((I_{3x3} - b_i b_i^T)\) corresponds to a design matrix \(A_i\) and \((I_{3x3} - b_i b_i^T)a_i\) to an observation vector \(o_i\). Thus, the observation equation for a laser beam \(i\) can be rewritten as

\[
\varepsilon_i = A_i \tilde{O}_p - o_i. \tag{4.12}
\]

Given \(n\) laser beams and 3 coordinate unknowns of \(O_p\), a design matrix \(A_i\) and an observation vector \(o_i\) for each laser beam \(i\) can be composed to a total design matrix

\[
A = \begin{pmatrix} A_1 \\ \vdots \\ A_n \end{pmatrix}
\]

with dimension \((3n \times 3)\) and a total observation vector

\[
o = \begin{pmatrix} o_1 \\ \vdots \\ o_n \end{pmatrix}
\]

of length \(3n\). Hereby, variable \(n\) denotes the total number of laser beams. The terms (4.12) to (4.14) lead to a set of observation equations

\[
\varepsilon = A \tilde{O}_p - o \tag{4.15}
\]

with the unknown intersection point \(\tilde{O}_p\). According to the least squares approach, the unknown coordinates of \(\tilde{O}_p\) can be determined by:

\[
\tilde{O}_p = (A^T A)^{-1} (A^T o) \tag{4.16}
\]

Having the intersection point \(\tilde{O}_p\), the laser bundle is shifted into the origin of the coordinate system \(O_p\), which represents the projector centre.

Subsequently, the laser bundle is oriented such that the mean of all red direction vectors coincide with the \(z\)-axis. First, the main projection direction

\[
r = \frac{\sum_{i=1}^{n} b_i}{n \cdot |\sum_{i=1}^{n} b_i|} \tag{4.17}
\]

of the red laser beams and its azimuth \(Az\) and zenith angle \(\theta\) is determined. Having calculated both angles, the laser beam bundle is transformed with the corresponding rotations

\[
a'_i = R_y R_z (a_i - \tilde{O}_p) \tag{4.18}
\]

\[
b'_i = R_y R_z b_i
\]

around the \(z\)- and \(y\)-axis with the rotation matrices.
Perpendicular to the z-axis, the orthogonal projection of laser beam $i = 1$ onto the xy-plane is defined as the direction of the x-axis. For its definition, a third rotation

$$
R_y = \begin{pmatrix}
\cos(\theta) & 0 & -\sin(\theta) \\
0 & 1 & 0 \\
\sin(\theta) & 0 & \cos(\theta)
\end{pmatrix}
$$

$$
R_z = \begin{pmatrix}
\cos(Az) & \sin(Az) & 0 \\
-\sin(Az) & \cos(Az) & 0 \\
0 & 0 & 1
\end{pmatrix}
$$

around the z-axis is required where $\alpha$ denotes the current azimuth angle of laser beam $i = 1$. The y-axis completes a right-handed coordinate system. Afterwards, azimuth angles $Az_i$ and zenith angles $\theta_i$ of the reconstructed beams and the xy-offsets

$$
\Delta_{\text{offset},i} = \frac{a''_i}{b''_i}
$$

to the projection centre are determined. The orientation angles of the beams are estimated with an accuracy of about 0.007 degrees whereas the offsets are determined with an accuracy of 0.08 mm.

### Stability of the Calibration

Variations in temperature, displacements due to transport, and technical revisions change the directions of the laser beams. Temperature variations cause volume extension of the projector and the laser pointers, which lead to a displacement of the laser beams. Transport and outer physical forces displace laser diodes within the coatings. Thirdly, lasers with malfunctions have to be replaced from time to time. As a consequence, the directions and the offsets of the beams change. Figure 4.5 shows the xy-offsets of the laser beams for two independent calibrations.

Figure 4.5: Offset displacements of the red laser beams between two different calibration epochs. The red dots represent the offsets for the calibration (24/02/2011) whereas the blue dots represent the offsets for a later calibration (28/05/2012). Corresponding offsets are connected with a blue line.
The largest $xy$-displacement of $1\, \text{mm}$ is for laser beam $i = 16$. Not only the offsets, but also the directions are affected by the factors mentioned previously. According to Figure 4.6, most of laser beams experienced a counter clockwise rotation in azimuthal direction in the order of 30 $\text{minutes}$ and a decrease of the zenith angle in the order of 2 $\text{minutes}$.

Figure 4.6: Variation of the direction angles (azimuth and zenith) of the current calibration (28/05/12) with respect to a previous calibration in (24/02/11).
5 Recognition and Identification of Laser Spots

Before a camera pose can be estimated, the recognition of each laser spot and its assignment to the corresponding laser beam becomes necessary to generate corresponding point pairs for the pose estimation.

Approaches for the recognition of laser spots in digital images have been published in Riner (2009), Tilch (2010), Seitz (2011) and Tilch and Mautz (2011). The spot recognition has to cope with varying laser spot appearances due to light sources of all kinds and materials of different absorptivity. As a result, some spots cannot be recognized or image regions are falsely labelled as spots. The occurrence of such errors and misclassifications must be avoided. For that reason, the spot recognition is required to work in a robust and reliable manner. Subsequent to the laser spot recognition, the centres have to be determined with sub-pixel accuracy, since the accurateness of laser spot centres is the main driver for the camera pose accuracy.

Algorithms for the spot identification have been presented in Tilch (2010) and in Tilch and Mautz (2011). To allow the operator a flexible camera handling, the spot identification has to cope with fragmentary captured spot patterns. Furthermore, the identification has to cope with slightly distorted spot patterns due to irregular projection surfaces as well as with falsely recognized spots. Finally, all previous mentioned steps have to be processed in real-time to enable high position update rates.

The remainder of this Chapter is organized as follows. In Section (5.1), camera preferences are discussed that help to simplify the spot recognition. The recognition itself and possible solutions are described in Section (5.2). Here, the recognition of the laser spots is divided in two parts. In the first subsection, the extraction of regions of interest for the laser spots is discussed. The definite spot detection by means of difference images or template matching is shown in the second subsection. The last Section (5.3) is dedicated to the identification of the laser spots.
5.1 Camera Preferences

For spot recognition, it is important that spots can be clearly distinguished from the background. Disturbing sources like lamps, illumination by daylight or other areas with a high portion of red or green must be excluded.

Lens filters are not suitable for supporting this exclusion. A red filter blocks wavelengths for the green and blue channel, and, thus, only red laser spots are visible. Otherwise, a green filter blocks wavelengths for the red and blue channel, and, thus, only green laser spots are visible. For the pose estimation and the scale introduction both red and the green spots must be visible in the images. Thus, this application of lens filters has been dismissed from consideration.

Robust spot recognition requires a high-contrast background. Additionally, laser spots must have a high saturation of either red or green for a clear distinction from light sources that show high intensity values in all colour channels.

With an optimal interaction of sensor exposure time, gain, and white light balance the spot recognition can be effective. The sensor exposure time drives the time span for the illumination of the camera sensor, and, therefore, it has an influence on the number of photons impinging the sensor. Additionally, the frame rate depends on the exposure time. The shorter the exposure time, the higher is the achievable frame rate. Higher exposure times cause blooming effects in areas of bright light sources. The gain can be regarded as a dimensionless factor to enhance the incoming signal, but also may increase the noise at the image sensor. Using cameras with automatic exposure, the sensor gain will be automatically adjusted to prevent overexposure to a certain extent. Too high gain values cause effects similar to blooming. With white light balance, the mapping of true colours can be achieved and colour casts can be prevented. The Automatic White Light Balance (AWB) analyses white areas and adjusts the colour temperature such that white areas appear white.

Illumination conditions are different for every indoor environment and vary over time. Thus, it is difficult to find a certain set of parameter values, which are valid for every indoor environment. Nevertheless, an attempt has been made to find the best possible parameter set that is valid for most indoor environments. The effect of sensor exposure time and gain has been investigated by a series of images showing the same scenery with two red laser spots and a lamp, as illustrated in Figure 5.1. In the first row, the exposure time has been increased from 10 milliseconds to 40 milliseconds. The gain value has been adjusted automatically. In the second row, the gain has been increased from 0 to 460 with an automatic adjustment of the sensor exposure time. For both cases, the overexposure around the lamp increases from left to right and superimposes the red laser spots.
Figure 5.1: Influence of sensor exposure time (shutter) and sensor gain on the image brightness.

The results of the complete investigation are shown in Figure 5.2. On the top, gain and frame rate have been adjusted automatically in dependency to the sensor exposure time. Obviously, gain and frame rate decrease with an increasing exposure time. Beyond 30 milliseconds, the frame rate decreases significantly below 30 frames per second. On the bottom, the gain has been varied with an automatic adjustment of exposure time and frame rate. Contrary to the illustration on the top, the frame rate seems to be constant. From a gain value of 640, the exposure time decreases from 81 milliseconds to 71 milliseconds to prevent sensor overexposure.

Figure 5.2: Mutual impact of sensor exposure time and gain Left: Plot of the variable sensor exposure time against the sensor gain and frame. Right: Plot of the variable gain against the sensor exposure time and frame rate.

For the CLIPS project, the camera parameters, exposure time and gain are adjusted according to the investigation results. The exposure time is set to 40 milliseconds, which corresponds to an image acquisition rate of 30 Hz. As a consequence, disturbing effects of overexposure in areas of light sources and motion blurring can be minimized. According to Figure 5.2, the gain value is adjusted automatically to a value of 680. This leads to overexposed areas such as in the second row of Figure 5.1. Therefore, the automatic adjustment of the sensor gain is disabled and a moderate gain of 200 is chosen. Only white balance will be automatically adjusted.
5.2 Laser Spot Recognition

Laser spot recognition can be regarded as image segmentation whereby spots represent the image foreground that has to be separated from the remainder scene. For that reason, key features for a unique spot description need to be determined. Several approaches exploit features like colour, colour distribution, shape or size of a spot. Methods range from simple operations like threshold operations in the specific colour channel, the calculation of difference images or the combination of channels to more sophisticated methods like template matching and classification strategies. In this work, recognition is distinguished in approaches to determine regions of interest (ROI) and approaches for a definite spot detection. Hereby, the outcome of the ROI algorithms can be exploited to speed up the second type of spot recognition techniques.

Approaches for the determination of ROIs are discussed in the first Section, followed by a description of spot detection techniques in the second Section.

5.2.1 Determination of Regions of Interest

Some pattern detection techniques, like template matching, reveal high computational costs when applied on the whole image. Consequently, template matching is not well suited for real-time applications. To minimize the computation costs, regions of interest (ROI) are defined that have a high probability to contain laser spots. Such regions show high intensities in the red or green channel and have circular or elliptical shapes. Both properties, intensity and shape can be exploited for rapidly reducing the search space. For this work, the following techniques are considered:

- Threshold Operations
- Background Estimation by Means of Morphological Filters
- Combination of Colour Channels

5.2.1.1 Threshold Operations

Assume that laser spots are silhouetted against their environment with a high intensity distribution in the correspondent colour channel. To separate the laser spots from the image background, a simple threshold operation on the correspondent channel can be applied. According to Burger and Burge (2006), a threshold operation

\[
I'(u,v) = \begin{cases} 
0, & \forall I(u,v) < t \\
1, & \text{else}
\end{cases}
\]  

is regarded as a 1-bit quantisation of a grey value image \(I(u,v)\) into a binary image \(I'(u,v)\). Thereby, all grey values from 0 to 255 are compared with threshold level \(t\) and are mapped to the binary values 0 and 1. Hereby, the selection of an optimal threshold level to minimize the involved loss of information is a vast field in computer vision and several approaches have been proposed. A good overview of threshold selection algorithms is given in an exhaustive survey of Sezgin and Sankur (2004). In this work, the scope has been limited to two simple methods that have been proposed by Zack, Rogers and Latt (1977) and Otsu (1979), and only the basic concepts of the investigated approaches are described. In Figure 5.3, the test image for the threshold operation is shown.
Figure 5.3: Test image for threshold operations. For a better visibility, the brightness has been enhanced by 80%.

The triangle algorithm of Zack, Rogers and Latt (1977) determines the first and last non-zero value and the highest peak of the histogram. Furthermore, a line is defined by the histogram's maximum and the last non-zero value. The optimal threshold is considered to be the grey value with a maximum distance to the line plus a fixed offset, as shown in Figure 5.4. The algorithm's name “triangle” relates to the triangular shape that is defined by the maximum peak B, last non-zero value C and threshold $t^*$. The result of the triangle method for the red channel is shown in Figure 5.4. It contains red and green laser spots as well as lamps with large illuminated surrounding areas. Since maximum and non-zero value search as well as distance calculations are simple operations, the triangle approach requires only 0.02 s for each colour channel. However, this approach is best suited for images with a homogeneous background, as stated in Zack, Roger, and Latt (1977). Therefore, this approach isn't suited for the spot recognition in CLIPS.

Figure 5.4: Results of the triangle thresholding. Left: Histogram of the red channel. A, B and C denote the first non-zero value, the maximum value and the last non-zero value of the histogram. The red line indicates the threshold $t^*$. Right: Results of the threshold selection according to the triangle algorithm.
Otsu (1979) proposes to separate the normalized histogram in two classes by exploiting only their class occurrence probabilities, their class means and class variances. To separate both classes, he proposes to select an optimal global threshold $t^*$ that maximizes their inter-class-variance. The effect of Otsu’s method is illustrated in Figure 5.5. The first image shows a clipping of the colour image. It contains red and green laser spots and lamps. The following binary images represent quantized grey value images for the single colour channels red ($t^* = 99$), green ($t^* = 125$) and blue ($t^* = 126$). The threshold operation for all three channels requires about 0.8 seconds respectively, 0.26 seconds for each channel.

![Figure 5.5](image)

**Figure 5.5: Results of the threshold selection according to Otsu’s method.** From left to right: The images show clippings of the original colour image and the quantized grey value images that represent the single colour channels red ($t=99$), green ($t = 125$) and blue ($t = 126$). For a better visibility, the brightness of the colour image has been increased. In the binary images, the dotted circles point up red laser spots whereas the solid lined circles indicate the presence of green laser spots.

Despite the fact that laser spots show high intensity peaks in the corresponding colour channels, two of eight red laser spots are missing in the quantized red colour channel. Surprisingly, green laser spots are also visible in binary images of the red and blue channel. This is contrary to the assumed monochromaticity of laser light. A closer look to the spot centres in the colour image reveals white areas that indicate high saturations in all colour channels. Possible explanations are blooming effects or varying surface properties, which cause scattering and absorption.

### 5.2.1.2 Background Estimation by Means of Morphological Operators

An elegant method in spot recognition is based on the background estimation by means of morphological operators. Once the background is known, it can be subtracted from the original image. The results should only contain the laser spots.

Morphological operators are defined to change image structures in binary or grey value images with a previously defined structure element $H(i,j)$. A structure element can be regarded as a small matrix with binary or real values that is shifted over the original image. The structure element’s centre is denoted as hot spot. In the following, morphological operators only for grey value images are investigated. A detailed description for binary morphological operators can be found in Burger and Burge (2006).
The elementary morphological operations are called Erosion and Dilation. Whereas Erosion deletes image structures smaller than the structure element, Dilation leads to a growth of image structures. According to Burger and Burge (2006), both morphological grey value operations can be defined as follows. The erosion

\[(I \ominus H)(u, v) = \min_{(i,j) \in H} \{I(u + i, v + j) - H(i, j)\}\]

is defined as the minimum difference between the values of the structure element and the image sample. Contrarily, dilation

\[(I \oplus H)(u, v) = \max_{(i,j) \in H} \{I(u + i, v + j) + H(i, j)\}\]

sums up the values of the structure element and the image sample and determines the maximum value. The resulting value is assigned to the hot spot position in the resulting image.

Through combination of both operations, two frequently used composed operations called Opening and Closing can be defined. Opening is defined as the sequence of Erosion and Dilation, whereas Closing is defined by a sequence of Dilation and Erosion. Hereby, Opening eliminates structures (i.e., laser spots), which are smaller than the structure element. Structures larger than the structure element (i.e., lamps) still remain. In Figure 5.6, an opening on a grey value image and the difference between the original image and the estimated background is shown. For that reason, a disk shaped structure element is chosen. The first image contains laser spots and lamps. The second image shows the estimated background via Opening. A subtraction of the estimated background from the original image separates the laser spots that reveal significant peaks in the difference image. Subsequently, they can be labelled as red or green spots. This method works well but requires a processing time of about 0.5 seconds per frame. Thus, only a measurement rate of about 2 Hz is achievable.

![Figure 5.6](image.png)

Figure 5.6: Illustration of laser spot recognition by means of morphological operators. Left: The graphic shows the original grey value image. The narrow peaks indicate laser spots, whereas stretched maxima belongs to fluorescent lamps. Centre: The image shows the estimated background as a result of an opening operation with a disk shaped structure element. Narrow peaks with a diameter smaller than the structure element are eliminated. Lamps with their linear structure still remain. Right: A subtraction of the estimated background from the original image eliminates the linear light sources and reveals significant peaks at the spot positions.
5.2.1.3  Colour Channel Combination

Assume that lamps and the sun emit a wave spectrum of different wavelengths that appears mostly white. Furthermore, let us assume that these light sources and the illuminated areas show a similar intensity distribution in the red and green colour channel. Red or green laser spots show high contribution in both colour channels, but only the corresponding colour channel is fully saturated.

Under these assumptions, disturbing light sources like lamps, sunlight or illuminated areas can be eliminated by a simple colour channel combination. Hereby, mathematic operations like addition, subtraction, division and multiplication of colour channels are considered for possible colour channel combinations. In case of CLIPS, the subtraction

\[
\begin{align*}
   dI_{RG} &= I_R - w_G \times I_G \\
   dI_{GR} &= I_G - w_R \times I_R
\end{align*}
\]

of the green channel \( I_G \) from the red channel \( I_R \), and vice versa deliver good results to detect red and green ROIs. The factors \( w_G \) and \( w_R \) weigh the corresponding colour channel and can be optimised in relation to the current scene illumination. The optimal values for both factors \( w \) were 2. Additionally, a threshold operation is carried out that delivers a binary image where white areas are considered as ROIs. The threshold is given as a percentage of the maximum value in \( dI_{RG} \) and a threshold of 50% of the maximum value in \( dI_{GR} \) works best. The dependency between the current illumination and the optimal factors \( w_G \) and \( w_R \) hasn't been found yet. Subsequent to the application of the threshold, the morphological close operator is applied to merge adjacent white pixels to white regions.

The overall computation costs for both channel operations account for about 0.08 seconds, which facilitates frame rates of about 12.5 Hz. An exemplary extraction of red ROIs is given in Figure 5.7.
Figure 5.7: Illustration of the channel combination (top to bottom). The first image shows a part of the original colour image. The two consecutive images represent the corresponding colour channels Red and Green. Comparing the red and green channels reveals parts of green spots in the red channel and vice versa. To determine all red regions of interest, the green channel is subtracted from the red colour channel. The result is shown in the fourth image, where the green spots are eliminated. A subsequent threshold operation converts the result image in a binary image, which can be used as a mask (e.g., for Template Matching). For a better visibility, the brightness has been enhanced to 80%.
5.2.2 Spot Detection

All previous described methods deliver images that contain candidates for red or green laser spots. Nevertheless, not all regions represent laser spots such as lamps or objects with high red or green channel saturation. Therefore, algorithms are required to find out correctly whether a region can be labelled as a red or green laser spot. The first method subtracts the scene background. In the result, only red and green laser spots remain. This method does not require ROI information and works for static measurements very well. In contradistinction, the second method of template matching makes use of the ROIs to accelerate and refine the evaluation. Thereby, a template of a laser spot is compared with the ROIs and a similarity measure is derived.

5.2.2.1 Calculation of Difference Images

The calculation of difference images is regarded as the simplest approach for laser spot recognition. Thereby, the recognition is based on the difference of two images from the same scene, whereby one image \( I_{FB} \) contains information about the fore- and the background, and the other image \( I_B \) solely represents the background. The digital camera captures the scene twice, one capture with all lasers switched on, the other with all lasers switched off. Taking the difference of both images

\[
dI = I_{FB} - I_B
\]

eliminates the background from the image. As a result, only the red and green laser spots remain. Having recognized the laser spot regions, they can be assigned as members of the red or green region class. Red regions show a higher portion in the red channel than in the green channel, and vice versa for green regions. This approach delivers satisfactory results due to the fact that disturbing light sources and illuminated areas can be successfully excluded.

The measurement rate is halved due to the fact that two consecutive images are required for spot recognition. This fact doesn’t exclude the suitability for real-time applications. Assuming a frame rate of 30 Hz, the measurement rate decreases to 15 Hz, which is adequate for real-time measurements. The bottleneck lies in the constraint that the camera pose and illumination conditions must not change between two consecutive images. Due to small movements of the camera hand, the camera pose changes slightly and the constraint cannot be applied. In case of small pose variations, the spatial and radiometric correlation between both consecutive images decreases. As a consequence, small shifts or rotations of the camera appear in the difference image, as shown in Figure 5.8. Herein, a small camera shift becomes perceivable in some kind of colour seams in the area of the lamp. The appearance of these colour seams and shifts complicates the recognition or even prohibits the application of difference images.
Furthermore, the camera must be triggered such that two consecutive images contain the background or the back- and foreground. This requires synchronization between the camera and the projector that switches the laser on and off.

As long as static measurements are concerned, this method is preferable due to its simplicity and robustness with respect to light sources. If kinematic measurements are desired, as in CLIPS, then this method should not be applied.

### 5.2.2.2 Template Matching

According to Omachi and Omachi (2007), template matching is one of the most important techniques in image analysis and signal processing. In Rosenfeld (1969), the basic concept of template matching is presented and described as a technique to determine how well the template \( g(u, v) \) and the covered part of the image match together. Hereby, the template \( g(u, v) \) is shifted over the image \( f(u, v) \) and is compared against the covered area. A similarity measure, like the normalized cross correlation

\[
NCC = \frac{\iiint f(u,v)g(u+i,v+j)dudv}{(\iiint f^2dudv)0.5} \tag{5.6}
\]

is an indicator for the similarity of both patterns, as described in Rosenfeld (1969). Especially, the stability of the NCC against intensity variations in the image and the standardised co-domain with values between -1 and 1 make the NCC advantageous in comparison to other similarity measures like a) sum of differences, b) maximum difference and c) sum of quadratic distances, as described in Burge and Burger (2006).
Figure 5.9: Illustration of template matching. A laser spot template is shifted over the image and compared with the region covered by the template. Then a similarity measure is derived to indicate whether a region matches to a laser spot.

According to the correlation theorem, the nominator of equation (5.6) can be efficiently estimated using the Fourier transform. In Uenohara and Kanade (1997), the Fourier transform of the correlation function

\[ F[f * g] = F[f]F^*[g] \]

(5.7)

is expressed as the product of the Fourier transform \( F[f] \) with the complex conjugate Fourier transform \( F^*[g] \). Taking the inverse Fourier transform delivers the correlation coefficient. To obtain a normalized correlation coefficient in the range between \( \text{NCC} = -1 \) and \( \text{NCC} = 1 \), the denominator needs to be estimated separately. For that purpose, both functions \( f(u, v), g(u, v) \) can be normalized prior the Fourier transform.

Template matching involves high computational costs, as stated in Uenohara and Kanade (1997). To reduce these costs, template matching has been applied with several modifications. In Vanderbrug and Rosenfeld (1977), a coarse-to-fine approach is presented to reduce the computational costs. Grün (1985) presented an adaptive least squares extension of correlation. Grey value differences between template and image patch are minimized through an affine transformation between both patches. However, this method was developed for aerial images and requires approximations for the unknowns, and is, therefore, not suited for CLIPS. Yoshimura and Kanade (1994) extract “eigenimages” from a set of rotated patterns, and generate a template by a linear combination of the “eigenimages.” Image patch and template are compared against each other via normalized cross correlation. Additionally, image pyramids are used to reduce the computational costs. In Uenohara and Kanade (1997), all reference images are “…expressed in terms of a linear combination of a finite set of orthonormal basis” (Uenohara and Kanade, 1997) derived from the eigenvectors of the reference images. The NCC is estimated subsequent to a projection of the image patch into the previously defined orthonormal basis. For a more efficient NCC computation, the Fast Fourier Transform is applied. Omachi and Omachi (2007), however, approximate templates with Legendre polynomials and calculate the NCC.
Basically, all these template-matching approaches determine an NCC to evaluate the coincidence between template and image patch. They either differ in their mathematical template representation, whether they are using image pyramids, or if they determine the correlation coefficient via Fourier transformation to reduce the computational costs.

In case of CLIPS, function $f(u, v)$ represents an image of the scene and $g(u, v)$ denotes a laser spot template. To reduce computational costs, only previously extracted regions of interest (ROI’s) are evaluated. Instead of using a set of templates that covers various spot appearances, a template $g(u, v)$ for each region is automatically generated by means of the region’s statistics. Both functions $f(u, v)$, $g(u, v)$ are convolved via Fourier transformation and the correlation coefficient is derived. The algorithm is as follows.

The ROIs, which are represented in a binary image, adjacent white pixels are connected to regions, evaluating the 8th neighbourhood. For each region, the corresponding intensity values $I(u, v)$, their standard deviation $\sigma$, the bounding box and the weighted centroid are determined. Hereby, the bounding box defines the image patch $f(u, v)$ in the red or green colour channel and the template size. Note that normalized intensity values are required for the correlation function. Therefore all intensity values must be divided by the value 255. Afterwards, the template

$$g(u, v) = \exp\left(\frac{(u-u_i)^2+(v-v_i)^2}{\sigma^2}\right)$$

(5.8)

in the same size of the image patch $f(u, v)$ is generated. For that reason, a two-dimensional Gaussian intensity distribution with the ROI’s standard deviation $\sigma$ is assumed. The coordinates of the weighted centroid are denoted as by $u_i$ and $v_i$. The domain of the image coordinates $u, v$ is limited by the region’s bounding box. Both functions are convolved exploiting the Fourier Transform as described in equation (5.7), and the maximum correlation coefficient for the corresponding ROI is derived. In case of CLIPS, all ROIs with a maximum correlation coefficient greater than 0.3 are labelled as laser spots.

5.2.2.3 Centroid Estimation

As a result of point recognition, spots can be described as a region of pixels with a certain intensity distribution, as illustrated in Figure 5.10.

![Figure 5.10: Illustration of a laser spot's shape and intensity distribution in a grey value image. The intensity distribution can be considered as a function of the pixel coordinates $(u,v)$.](image-url)
The first approach estimates the geometric centre
\[ \mathbf{x}_i = (u_i, v_i)^T, \quad u_i = \sum_u \frac{u}{n}, \quad v_i = \sum_v \frac{v}{n} \]  
(5.9)
of a laser spot region \(i\), where \(u_i\) and \(v_i\) denote the coordinates of the spot centre \(\mathbf{x}_i\). The variable \(n\) denotes the number of pixels of a spot region. For the example given in Figure 5.10, the centre coordinates are \(\mathbf{x} = (4.59, 4.63)^T\) pixel. Here, the intensity distribution is neglected in the centroid estimation, which may lead to small deviations from the true spot centre. In order to avoid such deviations, a weighted average that considers the intensity distribution for the centroid estimation can be calculated. For that reason, pixels with a higher intensity should have larger weights
\[ \omega(u, v) = \frac{I(u, v)}{\sum_u \sum_v I(u, v)} \]  
(5.10)
to the centroid estimation than pixels with lower intensities. For that purpose, the weighted average
\[ u_i = \sum_u \sum_v \omega(u, v) * u, \quad v_i = \sum_u \sum_v \omega(u, v) * v \]  
(5.11)
over all pixel coordinates is taken. For the example in Figure 5.10, the centre coordinates are \(\mathbf{x} = (4.56, 4.58)^T\) pixel. In comparison to the geometric gravity centre of our example in Figure 5.10, the centre from the weighted centroid estimation is slightly shifted to the upper left. In this project, the spot centres are calculated via weighted centroid estimation.
5.2.2.4 Comparison and Implementation

The described detection algorithms have been evaluated according to computational costs and the detection reliability. On the one side, rapid methods facilitate real-time detection of laser spots but maybe have rather poor results. Lamps and other disturbances provoke false detections. On the other side, reliable methods deliver good results but require a lot of computational resources. As a consequence, a trade-off between computational costs and detection reliability needs to be found. A suitable method to increase the detection reliability is given by the template matching approach, as stated in Section 5.2.2.2. It evaluates the shape and radiometric information of a potential spot candidate and delivers a similarity measure for the final decision. The spot candidates are delivered by one of the described ROI extraction approaches. To find the optimal detection algorithm, each ROI extractor is combined with a consecutive template matching and is evaluated considering the computational costs and the reliability in spot detection. Additionally, a colour channel combination without consecutive template matching is evaluated for spot recognition and compared against the combined methods.

![Graphs showing computational costs and success rate against gain](image1)

Figure 5.11: Comparison of spot detection approaches. Computational costs and the success rate are evaluated with respect to the camera’s gain (TM = Template Matching).

The evaluation uses an image set of 65 images, showing the same scene under different illumination conditions. Thereby, the impacts of gain and shutter time on computational costs and detection success rate are evaluated, as shown in Figure 5.11 and Figure 5.12. The detection success rate

\[
SR = \frac{n_L^2}{n_A \cdot n_V}
\]  

is defined as the ratio between the square number of correctly detected laser spots \(n_L\) and the product of all detected spots \(n_A\) with all laser spots \(n_V\) that are visible in the scene.
Recognition and Identification of Laser Spots

Figure 5.12: Comparison of spot detection approaches. Computational costs and the success rate are evaluated with respect to the camera's shutter time (TM = Template Matching).

Alternatively, Equation (5.12) can be described as the product of two ratios. The first ratio \( n_L/n_A \) denotes the percentage of correctly detected laser spots. The second ratio \( n_L/n_V \) represents the percentage of how many spots of the visible pattern are recognized.

Both the increase of gain and the increase of shutter time lead to brighter images and as a consequence to an increase of extracted ROIs. Since each ROI is evaluated by a consecutive template matching, the computational costs grow. A comparison of all approaches shows that detection success rate \( SR \) decreases and the computational costs increase for the 'Threshold' and 'Morphological Operators' methods. In case of CLIPS, the costs for both approaches are higher than 0.3 s, even with good illumination conditions and higher success rates. In contrast, the 'Colour Channel Combination' methods reveal quite constant computational costs between 0.08 s and 0.1 s. With template matching, success rates for red spot detection are higher than 0.6 and for the green spots they are higher than 0.4. Only for the channel combination without template matching, the success rates for the red spots decrease significantly and approach values in a range between 0.2 and 0.4.

Since colour combination with template matching provides the best results for a wide range of illuminations, considering the spot detection rate (\( SR > 0.4 \)) and mostly constant computational costs that are smaller than 0.1 s, this approach has been implemented for the spot detection in CLIPS.
5.3 Laser Spot Identification

Since pose estimation for CLIPS is based on the co-planarity and co-linearity constraint, an assignment of the detected laser spots to their corresponding laser beams is required.

For the identification, the projected laser pattern is used. If the projector is aimed at a planar surface, the basic pattern consists of 16 red laser spots that are distributed over two concentric rings. The first ring has four laser spots whereas the second ring contains twelve red laser spots. However, this pattern is of symmetric nature and thus identification techniques are required, which can break the pattern symmetry.

![Laser Spot Identification Diagram](image)

Figure 5.13: Illustration of laser spot identification algorithms. Left: Time-coded laser spot identification. Right: Colour-coded laser spot identification.

During the CLIPS project, two different approaches have been developed. The first method (Tilch, 2010) evaluates a sequence of two projection patterns and can be considered as a time-coded approach. For a single camera pose, both patterns need to be captured and compared. In Tilch and Mautz (2011), the second approach has been presented. Hereby, a colour-coded pattern of red and green laser spots is evaluated by the exploitation of three pattern properties – colour, membership to one of the rings and membership to a line defined by the spots on the inner ring. Both approaches are illustrated in Figure 5.13.

5.3.1 Time Coded Pattern

In the previous version of CLIPS (Tilch, 2010), the laser spot identification is based on a sequential projection of two unique patterns that are captured by the static camera. Thereby, either one laser spot or all laser spots are sequentially switched on.

Taking an image of a single laser spot, all spots in the other image can be easily identified. For that purpose, the single laser spot in the one image is assigned to the corresponding laser beam and labelled with ID #1. Assuming the same spot location in the other image identifies the same spot therein. Additionally, the gravity centre

$$\mathbf{g} = \left(\mathbf{u}_g, \mathbf{v}_g\right)^T, \quad \mathbf{g} = \frac{\sum_{i=1}^{n} \mathbf{x}_i}{n}$$

of all visible spots in the image is calculated, where \(\mathbf{x}_i\) denotes the centre coordinates of spot \(i\).
Afterwards, the directions

\[ \delta_i = \arctan \left( \frac{v_g - v_i}{u_g - u_i} \right) \]  

from the gravity point to the laser spots \( i \) are calculated. The direction of the identified spot \( i = 1 \) defines the reference direction and serves as an orientation unknown. All remaining laser spots can be identified by a counter-clockwise numeration with respect to the previous defined reference direction. For that purpose, the resulting intermediate angles

\[ \alpha_i = \delta_i - \delta_1 \]

are determined and sorted in ascending order. For the inner ring, the order of the sorted spot represents the corresponding beam number. For the outer ring, the order of the sorted spot is an index to a lookup table with the corresponding beam number, as shown in Table 5.1.

**Table 5.1: Look-Up-Table for the laser spot identification.**

<table>
<thead>
<tr>
<th>Order</th>
<th>Inner Ring</th>
<th>Outer Ring</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>8</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>11</td>
<td>12</td>
<td>13</td>
</tr>
<tr>
<td>14</td>
<td>15</td>
<td>16</td>
</tr>
</tbody>
</table>

To prevent confusion between the spots of the inner and outer ring, the identification is carried out for both rings separately, based on a distance analysis. It is assumed that the four spots of the inner ring have the four closest distances to the gravity point \( g \). The remaining spots are assigned to the second ring.

Due to the fact that this approach is based on a counter-clockwise numeration according to the laser’s angle to the reference direction, all lasers need to be captured. If only one laser is missing then the next laser in the counter-clockwise sequence would be identified as the missing one. Therefore, the camera has to be aimed at the laser spots such that the whole pattern is captured. Further, a static camera setup is required to assure the coincidence of both images. Without this coincidence, the assignment of the spot in the “single-spot” image to the corresponding spot in the “all-spots” image becomes impossible. As a consequence, only static pose estimations can be carried out.

Both requirements lead to a complicated handling of the camera. For that reason, this approach is not further considered for the CLIPS project.
5.3.2 Colour Coded Pattern

To enable real-time pose estimation, the spot identification has to be solved for each image, even if just a part of the captured pattern is visible in the images. This can be managed with additional eccentric green laser spots to introduce the metric scale and dissolve the pattern symmetry of the red laser spots, as illustrated in Figure 5.13. Here, two assumptions have been made.

First, it is assumed that all red spots of the inner ring and the green spots are visible in the image. Second, at least four red spots of the second ring are visible and form a concentric ring around the spots of the first ring. This becomes necessary to distinguish between red spots of the first and second ring.

To dissolve the pattern symmetry, two of the five additional green laser spots are closely arranged such that they present a small point group. By means of this point group, the closest red laser spot of the inner ring can be determined and labelled with its ID $i = 3$. The remaining three spots of the inner ring are labelled clockwise with their IDs $i = \{2, 1, 4\}$. Similarly, the remaining green laser spots are labelled.

Having identified the red spots of the inner ring and the green spots, the red spots of the second circle can be identified. For that purpose, the distance to a line, which is defined by two of four red spots of the inner ring, is evaluated. One spot is denoted as the initial point and the second point is required to define the line’s direction. Now, the closest distances of each outer-ring red laser spot to each line is calculated. All these distances can be composed in the distance matrix $D$, where each row contains the distances of all spots to one line and each column contains the distances of a spot to all lines. For each column the minimum distance of a spot to its nearest line is found. This way, two points of the outer ring can be associated with each of these six lines. To dissolve this ambiguity, both spots are distinguished into a spot with a closest or farthest distance to the line’s initial point. Finally, all spots of the outer ring can be labelled.

In case of CLIPS, this approach facilitates the spot identification for each image in a hundredth of a second, even with fragmentary spot patterns. Thus, the fast identification in conjunction with loosened capturing constraints facilitates the real-time pose estimation and has been implemented in the CLIPS project.
6 Camera Pose Estimation

Once the red and green laser spots have been identified in the image, the camera pose with respect to the projector can be estimated. In this thesis, the term camera pose is used to describe the camera’s position and orientation with respect to the projector. The camera pose is expressed by the relative orientation parameters and a separately estimated metric scale that represents the length of the basis vector between the projector centre and the camera projection centre. The separate estimation becomes necessary, since the arrangement of the red lasers does not provide metric information for the pose estimation.

To estimate the camera pose, a relationship between the unknown camera pose and the measured image coordinates must be established. In this thesis, the relationship is represented by non-linear functional models, which are based on the co-planarity constraint or collinearity constraint (Luhmann, 2010). To determine the unknown camera pose, these non-linear functional models have to be solved.

In this thesis, the pose estimation algorithm is split into an initial phase and a working phase. During the initial phase, a set of possible relative camera orientation parameters is refined by a least-squares-adjustment. The set that achieves the best result is selected to represent the relative orientation of the first camera position. Afterwards, the metric scale is estimated from the green lasers’ offset, their direction information and image measurements. Finally, the three-dimensional spot coordinates are calculated via intersection of the lasers and the image rays. This way, additional information in form of 3d laser spot coordinates is provided to enable the camera pose estimation via a least-squares resection. Although resection delivers the camera pose in the projector coordinate system, the metric scale is estimated independently and three-dimensional laser spot coordinates are calculated for every camera position. Together with previously determined laser spot coordinates, a new set of averaged and refined spot coordinates can be derived. As a consequence, the accuracy of the laser spot coordinates and, therefore, the accuracy of the camera position improves during the measurements.

The remainder of this chapter is organized as follows. In the first Section (6.1), a definition of the camera pose parameterization is given. Afterwards in Section (6.2), the least squares refinement of the approximate values is described. In Section (6.3), a strategy for the provision of approximate values is given. Section (6.4) is dedicated to approaches, which enable the estimation of the metric scale. In Section (6.5), the estimation of 3d laser spot coordinates is briefly described. Finally, the overall developed algorithm and its implementation is explained in Section (6.6).
6.1 Parameterization of the Camera Pose

The camera pose with respect to the projector can be expressed by a translation and a rotation of the camera coordinate system in three-dimensional space. Thereby, the translation vector

$$b_c = (b_x, b_y, b_z)^T$$

represents the vector between the projector centre to the camera projection centre. Due to the fact that the length of this vector cannot be determined via relative pose estimation, it is here defined with the unit length with \(\|b_c\| = 1\). To obtain the metric position of the camera projection centre

$$O_c = (X_c, Y_c, Z_c)^T$$

$$= m \cdot b_c$$

the metric scale \(m\) must be estimated. Several expressions for the camera's rotation exist. A common parameterization, especially in photogrammetry, is given in the form of three rotational angles \(\omega, \varphi\) and \(\kappa\). Here, positive angles define a counter-clockwise rotation around the corresponding axes (x, y, and z). Using the matrix notation, each rotation can be expressed by one of the following 3-by-3 matrices:

$$R_1(\omega) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_\omega & -s_\omega \\ 0 & s_\omega & c_\omega \end{pmatrix}, \quad R_2(\varphi) = \begin{pmatrix} c_\varphi & 0 & s_\varphi \\ 0 & 1 & 0 \\ -s_\varphi & 0 & c_\varphi \end{pmatrix}$$

$$R_3(\kappa) = \begin{pmatrix} c_\kappa & -s_\kappa & 0 \\ s_\kappa & c_\kappa & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

The complete rotation can be defined by the sequence

$$R = R_1(\omega)R_2(\varphi)R_3(\kappa)$$

$$= \begin{pmatrix} c_\varphi c_\kappa & c_\varphi s_\kappa & -c_\omega s_\varphi s_\kappa - s_\omega c_\varphi \\ c_\omega c_\varphi s_\kappa - s_\omega c_\varphi c_\kappa & c_\omega c_\varphi c_\kappa + s_\omega c_\varphi s_\kappa & s_\varphi \\ s_\omega s_\kappa - c_\omega s_\varphi c_\kappa & s_\omega s_\kappa c_\phi + c_\omega s_\varphi s_\kappa & c_\omega c_\varphi \end{pmatrix}$$

$$= \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix}$$

of all single rotations.

The cosinus function \(c_{\text{angle}}\) and the sinus function \(s_{\text{angle}}\) are abbreviations for the terms \(\cos(\text{angle})\) and \(\sin(\text{angle})\), as shown in Kraus (2007). Hereby, the rotation sequence plays an important role, since matrix multiplication is not of commutative nature. A different sequence with the same single rotations defines a completely different overall rotation. Detailed information about rotations and sequences are given in Kraus (2007) and McGlone et al. (2004).
An explicit expression of the camera rotation angles is given by

\[
\omega = \arctan\left(\frac{-r_{23}}{r_{33}}\right), \varphi = \arctan\left(\frac{-r_{12}}{\sqrt{r_{11}^2 + r_{12}^2}}\right) \quad \text{and} \quad \kappa = \arctan\left(\frac{-r_{12}}{r_{11}}\right). \tag{6.5}
\]

Further, the following terms apply. If only the camera position is addressed, it will be denoted by the position vector \( \mathbf{O}_c = (X_c, Y_c, Z_c)^T \). The orientation of the camera is denoted by three orientation angles \( (\omega, \varphi, \kappa) \) or the corresponding rotation matrix \( \mathbf{R}(\omega, \varphi, \kappa) \). The camera pose refers to both the position and the orientation of the camera.

### 6.2 Least-Squares-Adjustment of the Camera Pose

To solve the camera pose, the unknown camera pose parameters have to be optimized by taking all available reference information into account. In Geodesy, local Gauss-Newton optimization approaches are used to solve such optimization problems (Mautz, 2001). To solve the functional model for the unknown camera pose, it has to be linearized around approximate values of the camera pose, which are corrected until improvements become insignificant. This is a challenging task, since approximate values for the first camera pose are missing.

#### 6.2.1 Least-Squares Adjustment Based on Epipolar Geometry

According to Luhmann (2010), the camera pose can be estimated by solving the co-planarity constraint

\[
\Delta = \mathbf{d}_p^T (\mathbf{b}_c \times \mathbf{d}_c) = 0 \tag{6.6}
\]

where \( \mathbf{d}_p \) denotes the direction vector of the red laser beams. Vector \( \mathbf{b}_c \) denotes the vector between the projector centre \( \mathbf{O}_p \) and camera projection centre \( \mathbf{O}_c \). Vector \( \mathbf{d}_c \) represents the vector of the image projection. The co-planarity constraint is the mathematical expression of the epipolar geometry, where the vectors \( \mathbf{d}_p, \mathbf{b}_c \) and \( \mathbf{d}_c \) have to lie in a plane. This relationship is illustrated in Figure 6.1.

![Figure 6.1: Epipolar Geometry (see Luhmann, 2010)](image)

Since relative orientation algorithms only describe the mutual orientation of image pairs, the model coordinates are given an arbitrary scale. In this case, the x-component of basis vector \( \mathbf{b}_c \) is set to \( b_{cX} = 1 \).
The co-planarity constraint describes the geometrical relationship between two convergent images and can be rewritten as the determinant

\[
\Delta = \begin{vmatrix}
1 & d_{p,x} & d_{c,x} \\
b_{c,y} & d_{p,y} & d_{c,y} \\
b_{c,z} & d_{p,z} & d_{c,z}
\end{vmatrix} = 0.
\] (6.7)

The projection centres \(O_p, O_c\) and the laser spots define an epipolar plane, where the vector

\[
d_c = R_c \, x_c
\] (6.8)

represents the projection vector \(x_c\) of a laser spot in the image rotated into the object coordinate system using the rotation matrix \(R_c\). Due to the fact that the \(x\)-component of basis vector \(b_c\) is set to one, only five independent unknowns \(b_{c,y}, b_{c,z}, \omega, \varphi\) and \(\kappa\) remain. Now, the task of solving the co-planarity constraint leads to a nonlinear estimation problem. For this reason, the observation equation has to be brought in to a linearized form and good approximate values for the five RO unknowns are required. The linearized form of the model equation reads according to Luhmann (2010)

\[
v_\Delta = \frac{\partial \Delta}{\partial b_{c,y}} db_{c,y} + \frac{\partial \Delta}{\partial b_{c,z}} db_{c,z} + \frac{\partial \Delta}{\partial \omega} d\omega + \frac{\partial \Delta}{\partial \varphi} d\varphi + \frac{\partial \Delta}{\partial \kappa} d\kappa.
\] (6.9)

The derivatives in equation (6.9) are listed in the appendix. Accordingly, the equation system can be written as

\[
v = A \ast du - l
\] (6.10)

where matrix \(A\) contains the derivatives in equation (6.9). Vector \(du = (db_{c,y}, db_{c,z}, d\omega, d\varphi, d\kappa)^T\) denotes the additive corrections for the approximate values and vector \(l\) denotes the shortened observations. Claiming the square sum of the residuals \(v^T v = \min\), enables to estimate the additive corrections

\[
du = (A^T A)^{-1} (A^T l).
\] (6.11)

Since one observation equation expresses the geometric relationship of one corresponding point pair of both images, at least five corresponding point pairs are required to solve the equation system and to estimate the RO unknowns. For this reason, initial approximate values are refined by an iterative least-squares adjustment until the improvements become insignificant. The accuracy of estimated RO unknowns can be extracted from the co-variance matrix

\[
K_{uu} = \sigma_0^2 (A^T A)^{-1}
\] (6.12)

\[
= \sigma_0^2 Q_{uu}
\]

where

\[
\sigma_0^2 = \frac{v^T v}{n - 5}
\] (6.13)

represents the variance of unit weight. Matrix \(Q_{uu}\) denotes the co-factor matrix and \(n\) denotes the number of corresponding point pairs.
6.2.2 Least-Squares Adjustment Based on Collinearity Constraints

Another way to solve the camera pose is given by the least-squares resection, which is based on the collinearity constraint. This constraint claims that the vector between the camera projection centre \( \mathbf{O}_c = (X_c, Y_c, Z_c) \) and given three-dimensional laser spot coordinates \( \mathbf{L} = (X_l, Y_l, Z_l) \) is collinear to the vector between the camera projection centre and the image measurements \( \mathbf{x}_c = (x_c, y_c, -f)^T \). This relationship is illustrated in Figure 6.2.

![Figure 6.2: Camera pose estimation via resection.](image)

Having three-dimensional laser spot coordinates \( \mathbf{L} = (X_l, Y_l, Z_l) \), the corresponding image measurements \( \mathbf{x}_c = (x_c, y_c, -f)^T \) can be expressed as a function

\[
\mathbf{x}_c + \nu \mathbf{x} = F(X_c, Y_c, Z_c, x_0, f, dx, X_l, Y_l, Z_l)
\]

\[
\mathbf{y}_c + \nu \mathbf{y} = F(X_c, Y_c, Z_c, y_0, f, dy, X_l, Y_l, Z_l)
\]

(6.14)

of camera pose \( (X_c, Y_c, Z_c, \omega, \varphi \) and \( \kappa \), interior camera orientation \( (f, x_0, y_0, dx, dy) \) and three-dimensional laser spot coordinates, as shown in Luhmann (2010). The geometric relationship between the camera image and the laser spot coordinates is expressed by

\[
x_c = x_0 - \frac{r_{11}(X_l - X_c) + r_{21}(Y_l - Y_c) + r_{31}(Z_l - Z_c)}{r_{13}(X_l - X_c) + r_{23}(Y_l - Y_c) + r_{33}(Z_l - Z_c)} + dx
\]

(6.15)

\[
y_c = y_0 - \frac{r_{12}(X_l - X_c) + r_{22}(Y_l - Y_c) + r_{32}(Z_l - Z_c)}{r_{13}(X_l - X_c) + r_{23}(Y_l - Y_c) + r_{33}(Z_l - Z_c)} + dy
\]

where the factors \( r_{11} \ldots r_{33} \) denote the components of the rotation matrix in equation (6.4). Due to the fact that a calibrated camera with known interior camera orientation parameters is used, unknowns are limited to the coordinates of the camera projection centre \( \mathbf{O}_c = (X_c, Y_c, Z_c) \) and the camera orientation angles \( \omega, \varphi \) and \( \kappa \).
The linearization of equation (6.15) with approximate values of the unknowns is as followed:

\[
\begin{align*}
\nu x_i &= \left( \frac{\partial x_c}{\partial X_c} \right)^0 dX_c + \left( \frac{\partial x_c}{\partial Y_c} \right)^0 dY_c + \left( \frac{\partial x_c}{\partial Z_c} \right)^0 dZ_0 + \left( \frac{\partial x_c}{\partial \omega} \right)^0 d\omega + \left( \frac{\partial x_c}{\partial \varphi} \right)^0 d\varphi + \left( \frac{\partial x_c}{\partial \chi} \right)^0 d\chi + (x_{c,i} - x_c^0) \\
\nu y_i &= \left( \frac{\partial y_c}{\partial X_c} \right)^0 dX_c + \left( \frac{\partial y_c}{\partial Y_c} \right)^0 dY_c + \left( \frac{\partial y_c}{\partial Z_c} \right)^0 dZ_0 + \left( \frac{\partial y_c}{\partial \omega} \right)^0 d\omega + \left( \frac{\partial y_c}{\partial \varphi} \right)^0 d\varphi + \left( \frac{\partial y_c}{\partial \chi} \right)^0 d\chi + (y_{c,i} - y_c^0)
\end{align*}
\]

(6.16)

The derivations of the observation equations according to the unknowns are given in the appendix. Similar to Section (6.2.1), the equation system (6.16) can be expressed by a matrix-vector product and solved according to equation (6.11). With six unknowns, at least three identified laser spots are required to solve the equation system. As previously stated, good approximate values for the unknowns must be provided.

### 6.3 Generation of Approximate Values

To solve the equation systems in Section (6.2), the introduction and the iterative improvement of approximate values become necessary. However, the provision of approximate values is in some cases a non-trivial problem. Hereby, it is useful to make use of every available prior information like 3d laser spot coordinates, assumptions about the first camera pose, the history of previous camera poses as well as the projector’s and camera’s geometry.

Since approximate values are missing for the first camera pose, a Monte-Carlo strategy is followed where camera positions are simulated on the grid points of a tessellated sphere (Cronk et al., 2006). Contrary to Cronk et al. (2006), the solution space for the camera pose is limited by the initialization of the camera in a previously selected octant of the projector coordinate system. Starting from any initial values of the orientation angles (usually chosen to be zero), every set of approximate values is improved by a least squares adjustment. The correct solution is found by comparing the residual vectors and taking geometric constraints (i.e., the camera is always located in front of the laser points) into account. Once camera poses have been estimated, the pose history can be exploited to predict the next approximate values via Kalman filtering (Kalman, 1960; Maybeck, 1979; Welch and Bishop, 2006).

![Figure 6.3: Approximate values are generated by simulation of different camera positions on a tessellated sphere (see Cronk et al., 2006).](image)
6.4 Introduction of the Metric Scale

In Chapter (2), different approaches for reference introduction are described. Popular approaches exploit 3d building models or maps, deploy coded targets and scale bars, or project their reference information. In case of CLIPS, the idea is to project the reference information with laser beams. For the current configuration, a bundle of red laser beams establish an inverse camera to estimate the camera’s relative orientation with respect to the projector. However, relative orientation only contains information about the orientation angles (ω, φ, κ) and the basis vector \( b_c \) up to scale. Thus, a separate estimation of a metric scale becomes necessary. For this purpose, green laser pointers were added to the laser projector in eccentric positions with the eccentric vector \( b_g \) as shown in Figure 6.4. The mount for the additional green laser beams is called Scale Cross, as described in Chapter (3). Since the direction \( d_g \) of the additional laser pointer and the base line \( b \cdot b_g \) to the laser rig are known from a one-time calibration, the system scale becomes estimable. Note that the vectors \( b_g, d_g, b_c \) and \( d_c \) are used as unit vectors in the following derivations.

![Figure 6.4: Introduction of metric information in CLIPS.](image)

The known length of the baseline between the laser-projector and the additional green laser pointer is denoted with \( b \). An expression for the scale can be derived by intersecting plane \( \Pi \) spanned by the vectors \( d_g \) and \( d_c \) with the straight line \( l \) defined by the base vector \( b_c \). The plane can be expressed by the Hessian normal form

\[
\Pi: (d_g \times d_c) \cdot X - (d_g \times d_c) \cdot b_g = 0 
\]

(6.17)

where the cross product \( (d_g \times d_c) \) represents the normal vector, the base vector \( b \cdot b_g \) represents an initial point of plane \( \Pi \) and \( X \) is an arbitrary point of the plane. The straight line

\[
l: X = a + m \cdot b_c 
\]

(6.18)

is defined by the direction of vector \( b_c \) and initial point \( a = (0, 0, 0)^T \), which represents the projector centre. \( X \) is any point of the line and \( m \) is the metric scale.
Inserting (6.18) in (6.17) leads to

$$m \ast (d_g \times d_c) b_c - b \cdot (d_g \times d_c) b_g = 0. \quad (6.19)$$

After solving (6.19) for the metric scale $m$ we obtain

$$m = b \frac{(d_g \times d_c) b_g}{(d_g \times d_c) b_c}, \quad (6.20)$$

where $m$ is equal to the length of the baseline between camera and projector. According to equation (6.2), the metric scale is introduced by determining the camera position by $O_c = mb_c$. Maximal 36 green eccentric laser pointers can be assembled such that the metric scale can be estimated 36 times independently. Currently, only 5 additional green laser spots are mounted on the scale cross, to simplify the spot recognition. The contribution of each scale arm can be expressed by the $SQM$ (Scale Quality Measure)

$$SQM = \left( \frac{||b_g \times d_g||^2 \times ||b_c \times d_g||}{||b_g \times d_g|| \ast ||b_c \times d_g||} \right)^2, \quad (6.21)$$

which corresponds to the angle between the plane defined by the vectors $b_c, d_g$ and the plane defined by the vectors $b_g$ and $d_g$. If these two planes are parallel then the distance quality measure becomes zero, since the cross product vanishes. Furthermore, the ratio of equation (6.21) is raised by the power of two, which enhances the values for good geometries and diminishes it for unfavourable geometries.

Due to the fact that 5 green laser beams are applied, the corresponding $SQM$ can be used for a weighted average of the metric scale. Given $n$ green laser beams ($j = 1 \ldots n$), with the corresponding metric scales $m_j$ and scale quality measures $SQM_j$, the weighted average of the metric scale is calculated as follows:

$$\hat{m} = \frac{\sum_{j=1}^{n} m_j \ast SQM_j}{\sum_{j=1}^{n} SQM_j}. \quad (6.22)$$

Finally, the resulting metric scale $\hat{m}$ provides the metric information for basis vector $b_c$.

### 6.5 Estimation of 3D Laser Spot Coordinates

Once, the metric scale $\hat{m}$ of a camera pose is determined, 3d laser spot coordinates $L_i = (X_{i,i}, Y_{i,i}, Z_{i,i})^T$ in the projector coordinate system can be derived by intersection. Since laser spot coordinates are calculated for every camera pose $k$, the averaged laser spot coordinates

$$\bar{L}_{i,k} = (\bar{X}_{i,k}, \bar{Y}_{i,k}, \bar{Z}_{i,k})^T$$

$$= (1 - \frac{1}{k})\bar{L}_{i,k-1} + \frac{1}{k}L_{i,k} \quad (6.23)$$

for the current camera pose $k$ are recursively calculated on basis of the previous average $\bar{L}_{i,k-1}$ at camera pose $(k - 1)$ and the calculated laser spot coordinates $L_{i,k}$ at camera pose $k$. As a result, the accuracy of the laser spot coordinates $\bar{L}_{i,k}$ improves during the measurements. Thereby, it is recommended to determine the laser spot coordinates $L_{i,k}$ from camera poses around the laser spot pattern, to obtain the best coordinate accuracy.
6.6 Critical Configurations

For the spatial resection and the introduction of the metric scale, critical configurations exist, which complicate or disable the camera pose estimation. In this section, critical configurations for the spatial resection and the metric scale are discussed.

6.6.1 Critical Configurations for Spatial Resection

The problem of critical configurations for spatial resection is described in detail by Gotthardt (1940, 1974), Wunderlich (1943) and Killian (1990). Thereby, three different kinds of critical configurations can be distinguished:

- All control points/laser spots lie on a line.
- A circular cylinder, which is defined by only three available control points/laser spots and the camera projection centre.
- Three or more control points/laser spots and the camera projection centre lie on a horopter curve.

The first case can be easily explained. If all laser spots are members of a line, a circle around the line can be defined such that the line is perpendicular to the circle and crosses the circle’s centre. Additionally, the projection centre of the camera is a member of the circle, as illustrated in Figure 6.5. If the camera projection centre is moved on the circle, the angles between the image rays to the laser spots do not change. Thus, the circle defines infinite possibilities for the location of the camera projection centre. In case of CLIPS, this critical configuration can be excluded due to the projection of an extensive laser spot pattern.

Figure 6.5: Critical configuration - All laser spots lie on a line.
The second case can occur, if only three laser spots are identified and their circumcircle
defines a circular cylinder, whose curved surface area contains the camera projection centre, as
shown in Figure 6.6. If four or more laser spots are identified and if they lie on a plane, the
occurrence of this critical configuration can be excluded according to Gotthardt (1940) and
Wunderlich (1943). Currently, this case can be also excluded for CLIPS since at least four red
laser spots of the inner circle and four red laser spots of the outer circle are required for a
proper spot identification.

![Figure 6.6: Critical configuration: All available laser spots lie on a circular cylinder.](image)

According to Gotthardt (1974) and Killian (1990), a third critical configuration exists if
four or more reference points are members of a horopter curve, as shown in Figure 6.7. Here, the
horopter curve represents the cutting line if a circular cylinder is intersected with an orthogonal
cone. If the camera projection centre is a member of the horopter curve and the projection
centre is diametral to the horopter’s asymptote, then a critical configuration exists. Since for a
given cylinder various horopters can exist, which intersect in the camera projection centre,
critical configurations cannot be excluded entirely for CLIPS, as illustrated in Figure 6.7. This
scenario can occur, (e.g., in tunnel environments or barrel-vaults) where all laser spots lie on the
circular curved ceiling.

![Figure 6.7: Critical configuration: All laser spots lie on horopter curves (red & green), which intersect in the camera projection centre.](image)
6.6.2 Critical Configurations for the Estimation of the Metric Scale

In principle, a single eccentric green laser pointer can be used for the determination of the metric scale. But, if the camera is located in the plane spanned by base vector $\mathbf{b}_g$ and base vector $\mathbf{d}_g$ then plane $\Pi$ contains line $l$ such that there is no intersection and the determination of the metric scale becomes ill posed, as illustrated in Figure 6.8. An alternative explanation can be derived by equation (6.20). If the camera is located in the plane spanned by vector $\mathbf{b}_g$ and $\mathbf{d}_g$, the denominator and nominator of equation (6.20) becomes zero. As a consequence, the metric scale becomes undeterminable.

![Figure 6.8: Critical configuration for the estimation of the metric scale.](image)

In order to avoid such geometric instabilities, all green laser pointers are used for the metric scale estimation, as shown in Section 6.4. However, a critical configuration remains. If the camera is located in the z-axis of the projector, the metric scale estimation becomes ill-posed even if all green laser pointers are used.

6.7 Implementation

Depending on the image number, the overall pose estimation algorithm is divided in an initial phase ($k = 1$) and a measurement phase ($k > 1$). For the pose evaluation, three suitable tests are available, which are applied during an initial phase or for the measurement phase. These tests incorporate restrictions to the solution space, evaluations of the overall geometry and comparisons to previous measurements. Hereby, the geometry evaluation and the comparison are based on three-dimensional laser spot coordinates, which are obtained by an intersection of laser beams and image rays. In the following sections, the pose evaluation for the initial phase and the measurement phase is described. Finally, a flow chart of all applied camera pose algorithms is given.

6.7.1 Initial Phase

Due to the lack of a priori approximate values for the first camera pose, a set of suitable values is generated according to the algorithm of Cronk et al. (2006), as shown in Section (6.3). Once each set of generated approximate values is refined by a least squares adjustment, as described in Section (6.2.1), the identification of the correct parameter set is required.
To simplify the evaluation of all refined parameter sets, the first camera location is assumed in a previously defined octant of the projector coordinate system. Hereby, every octant can be used for the initialization of the camera pose. In case of CLIPS, the 8th octant of the projector coordinate system is chosen. For the remaining parameter sets, the 3d laser spot coordinates are obtained via intersection of laser beams and image rays, according to Section (6.5). Having estimated the laser spot coordinates, their location is evaluated with respect to the camera and the projector. For this reason, each device is represented by a plane defined by the projection centre \( \mathbf{O}_p = (0, 0, 0)^T \) respectively \( \mathbf{O}_c \) and the main viewing direction \( \mathbf{n}_p = (0, 0, 1)^T \) respectively \( \mathbf{n}_c = (r_{13}, r_{23}, r_{33})^T \). Inserting laser spot coordinates \( \mathbf{L}_i \) in the Hessian normal form delivers the point-plane distance

\[
ppd_{i,p/c} = \mathbf{n}_{p/c} \mathbf{L}_i - \mathbf{n}_{p/c} \mathbf{O}_{p/c}. \tag{6.24}
\]

According to Weisstein (2012), the point-plane distance turns positive, if a laser spot is located in the half-space determined by the direction of unit normal vector \( \mathbf{n}_{p/c} \). If all point-plane distances are positive, all laser spots are located in front of the camera and the projector. In this case, the parameter set is selected as a correct one – otherwise it is rejected. If more than one parameter set remain, the final selection criteria is given by the smallest Helmert point error

\[
\sigma_H = \sqrt{\sigma_x^2 + \sigma_y^2 + \sigma_z^2}. \tag{6.25}
\]

Here, the variances \( \sigma_x^2, \sigma_y^2, \sigma_z^2 \) represent the variances of the camera position and can be extracted from the diagonal elements of co-variance matrix \( \mathbf{K}_{uu} \) in Equation (6.12).

### 6.7.2 Measurement Phase

Once a camera pose is estimated, approximate values for further camera poses are predicted via Kalman filtering. Since 3d laser spot coordinates are available, a resection according to Section (6.2.2) can be used for the camera pose estimation. The small offsets of the red laser beams \( \Delta_{offset_i} \) of equation (4.21) are considered by adding them to the corresponding spot coordinates of the red lasers. In contrary to the initial phase, only one set of approximate values is refined by least-squares adjustment. Thus, the identification of the correct solution from several parameter sets like in the initial phase is not required. However, the final camera pose may be corrupted by falsely recognized laser spots. Such camera poses have to be detected and rejected in a two-staged evaluation.

Like in the initial phase, new 3d laser spot coordinates \( \mathbf{L}_{ik} \) are derived by intersection. Afterwards, their location with respect to the camera and the projector is evaluated. If all laser spots are located in front of the camera and the projector, they will be compared with the averaged laser spot coordinates \( \bar{\mathbf{L}}_{i,k-1} \) of previous camera poses. If the maximum distance

\[
dL_i = \max_{i = 1 \ldots 16} (\| \bar{\mathbf{L}}_{i,k-1} - \mathbf{L}_{i,k} \|) \tag{6.26}
\]

is smaller than 50 mm, the camera pose is accepted. Otherwise, the camera pose is rejected. The relatively high limit has been chosen empirically to facilitate faster camera movements and measurement rates. In Figure 6.9, the overall pose estimation algorithm for the CLIPS system is shown.
Implementation

Figure 6.9: Pose Estimation Algorithm
7 System Evaluation of CLIPS

According to Mautz (2012), an indoor positioning system can be evaluated from different perspectives. The first perspective addresses user requirements like Accuracy, Costs or Update rate, whereas the evaluation according to technical aspects represents the complementary second perspective, with such as Technology, Principle or Application. Of course, there are many further parameters, which must "...be weighted against each other" (Mautz, 2012). Parameters according to user requirements and technical parameters are shown in Figure 7.1.

The objective of CLIPS is to develop an inexpensive measurement tool that enables automatic, continuous and accurate positions of a mobile camera with respect to a static projector in all indoor environments. In accordance with the evaluation scheme in Figure 7.1, the following user requirements for applications in construction and metrology can be derived:

- Accuracy: mm-level in position, sub-degree in orientation
- Coverage Area: scalable
- Cost: ca. 2000 CHF
- Infrastructure: no requirement for additional infrastructure
- Output data: Six degrees of freedom (6DOF, 3D Camera Position + Orientation)
- Update Rate: Continuously (>10 Hz)
- Availability: In every indoor environment
- Scalability
- Number of Users: unlimited

Figure 7.1: Evaluation Criteria for indoor positioning systems. Left: User requirements according to Mautz (2012). Right: Technical Parameters according to Mautz (2012).
Furthermore, the following technical parameters can be identified:

- Technology: Optical
- Measured quantity: Directions (eventually additional distances)
- Basic measuring principle: Active triangulation by means of an inverse camera
- Positioning algorithm: Relative orientation on basis of epipolar geometry, Resection
- Signal: Visible light
- Application: Construction and metrology
- Coordinate reference: Primarily local, after transformation global

To verify the requirements with respect to the current performance parameters, namely accuracy and update rate, measurements and simulations have been conducted and evaluated. The remaining performance parameters are fulfilled as given in the listing.

The remainder of this Chapter is organized as follows. In Section (7.1), the evaluation set-up is described and an overview of the applied PC hardware is given. The second Section (7.2) is dedicated to the assessment of accuracy and precision of CLIPS. It is based on a comparison of two laser spot coordinate sets, which are determined by a theodolite measurement system (TMS) and by CLIPS itself. The precision is analysed by means of static measurements. Afterwards, a sensitivity analysis is conducted to estimate the influence of image measurements and the accuracy of the laser offsets on the camera positioning accuracy. The sensitivity analysis is presented in Section (7.2). Time statistics about the measurement are analysed in Section (7.3). The chapter concludes with a summary of the performance parameters and a comparison with current positioning systems is given in Section (7.4).

7.1 Evaluation Set-Up

If a probe is attached to the camera, a proven evaluation method is the comparison of tip coordinates at a reference target with the target's 3d coordinates. However, a calibrated probe has not been attached to the camera during the CLIPS project, and thus, only the position of the projection centre and the orientation angles are estimated. This implies the following problem. The camera's projection centre is located inside the camera case, such that a relationship between the projection centre and any coordinated reference target cannot be established. For that reason, another test must be conducted to evaluate the positioning accuracy of CLIPS.

Therefore, the laser spot coordinates have been measured via a theodolite measurement system (TMS) and via CLIPS, as proposed in Tilch (2010). The TMS delivers laser spot coordinates, with sub-mm accuracy, which is a magnitude better than the calibrated CLIPS system. For the evaluation set-up, the projector pointed to the ceiling at a distance of about 70 cm. Because of the large projector aperture angle, the laser spot pattern has a diameter of about 140 cm. Next to the projector, a theodolite measurement system, consisting of a Wild T2000 and a Wild T3000 theodolite was established to measure the 3d laser spot coordinates. The overall evaluation set-up is shown in Figure 7.2. By means of a 3d Helmert transformation, the laser spot coordinates from the CLIPS system are compared against the laser spot coordinates from the TMS measurement. This test is conducted for static set-ups as well as for kinematic measurements. For kinematic measurements, the camera is hand-held. Only for static measurements, the camera is mounted on a tripod. For both measurements, about 2500 camera poses have been recorded and finally compared.
In case of static measurements, the precision of the estimated camera pose is analysed. Of special interest is not only the mean position variation but also the principal component which has the largest influence on the camera pose. All laser spots are captured with an AVT Guppy camera, which is described in Section (3.1). With a focal distance of 6 mm and a sensor diagonal of about 6 mm, the applied camera lens from Pentax has a field of view of about 53°. The software for the camera control and the camera pose estimation are implemented in Matlab 2012a and run on a Dell Precision M4600 laptop with a Windows 7 operating system. The laptop’s performance parameters are given in Table 7.1.

<table>
<thead>
<tr>
<th>Performance Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Processor</td>
<td>Intel Core i7-2720QM @ 2.20 GHz</td>
</tr>
<tr>
<td>Cores</td>
<td>8</td>
</tr>
<tr>
<td>RAM</td>
<td>8 GB</td>
</tr>
<tr>
<td>System Type</td>
<td>64 Bit</td>
</tr>
<tr>
<td>Graphic Card</td>
<td>NVIDIA QUADRO 2000M</td>
</tr>
</tbody>
</table>

### 7.2 Accuracy and Precision

The investigation of measurements always requires a distinction between the terms accuracy and precision. Accuracy is defined as a measurement’s bias to the true value, whereas precision describes the reproducibility of a measurement or a value derived from measurements.

#### 7.2.1 Accuracy

To evaluate the accuracy, laser spot coordinates $\hat{L}_i = (X_{L,i}, Y_{L,i}, Z_{L,i})^T$ are determined with a theodolite measurement system with sub-mm accuracy. During the pose estimation of the static camera, laser spot coordinates $\tilde{L}_{i,k}$ are estimated and continuously refined. Having both coordinate sets, they can be compared via a 3d Helmert transformation, as described in Wunderlich (2004). Because of the accurate calibration of projector and camera, residuals of the 3d Helmert transformation between both point sets can be explained only by inaccuracies of the camera pose estimation. Further, the transformation can be considered as a free stationing of the camera. Under this assumption, standard deviations of the 3d Helmert transformation can be used as accuracy estimations of the camera pose.
Figure 7.3: Accuracy of the camera position for three static camera set-ups and one kinematic measurement.

In Figure 7.3, the accuracy for the translation, respectively the camera position of the static and the kinematic measurements are shown. Here, the standard deviations $\sigma_x$, $\sigma_y$ and $\sigma_z$ are smaller than 6 mm. Due to the fact that only laser spots in front of the camera contribute to the pose estimation, standard deviations in z-direction are larger than standard deviations in x- and y-direction. A similar situation exists in GNSS positioning, where only satellites above the horizon are used for position estimation. Additionally it can be seen that standard deviations for the kinematic measurements are smaller than standard deviations for static measurements. In contrary to static measurements, the camera is moved around the projector during kinematic measurements and laser spots are captured from different points of view. This way, independent sets of laser spot coordinates $L_{i,k}$ for each camera pose are calculated, which contribute to the refined laser spot coordinates $\bar{L}_{i,k}$.

Figure 7.4: Accuracy of the camera orientation for three static camera set-ups and one kinematic measurement.

If a tip is attached to the camera, the position accuracy of the tip is not only affected by the camera’s position accuracy, but also by the camera’s orientation accuracy. In Figure 7.4, the accuracy for the rotation angles respectively the camera orientation of the static and kinematic measurements are shown. Here, a similar behaviour as for the camera position accuracy can be observed. Standard deviations for the static measurements are larger than for the kinematic measurements. In case of the static set-ups, standard deviations are smaller than 0.14°. For the kinematic measurements, the standard deviations are about 0.06°. Assume a tip of 10 cm length, which is attached to the camera. In this case, a deviation of about 0.12° for all three directions induces a deviation of the tip position of about 0.2 mm in x-, y- and z-direction.
According to Figure 7.3 and Figure 7.4, an overall positioning accuracy in the mm-range is achievable. Thereby, the camera’s orientation accuracy has a smaller influence on a tip position than the camera’s positioning accuracy. A dependency of the accuracy with respect to the camera height could not be observed.

### 7.2.2 Precision

To evaluate the precision of CLIPS, static measurements have been conducted with 2500 camera poses for each static camera set-up. The standard deviations for all static set-ups are listed in Table 7.2.

**Table 7.2: Standard deviations (1-sigma) of measurement series for different static camera set-ups at height Z.**

<table>
<thead>
<tr>
<th></th>
<th>Z = 22 mm</th>
<th>Z = -720 mm</th>
<th>Z = -1120 mm</th>
<th>Z = -1530 mm</th>
</tr>
</thead>
<tbody>
<tr>
<td>$s_x$ (mm)</td>
<td>0.51</td>
<td>0.68</td>
<td>0.58</td>
<td>0.35</td>
</tr>
<tr>
<td>$s_y$ (mm)</td>
<td>0.63</td>
<td>0.56</td>
<td>0.72</td>
<td>0.41</td>
</tr>
<tr>
<td>$s_z$ (mm)</td>
<td>0.41</td>
<td>0.53</td>
<td>0.57</td>
<td>0.38</td>
</tr>
<tr>
<td>$s_u$ (')</td>
<td>1.1</td>
<td>0.71</td>
<td>0.36</td>
<td>0.44</td>
</tr>
<tr>
<td>$s_v$ (')</td>
<td>0.68</td>
<td>0.29</td>
<td>0.33</td>
<td>0.39</td>
</tr>
<tr>
<td>$s_w$ (')</td>
<td>0.67</td>
<td>0.23</td>
<td>0.12</td>
<td>0.26</td>
</tr>
<tr>
<td>$s_m$ (mm)</td>
<td>0.77</td>
<td>0.95</td>
<td>1.03</td>
<td>0.51</td>
</tr>
</tbody>
</table>

A look at Table 7.2 reveals a standard deviation for the camera position better than 0.7 millimetres and a standard deviation for the camera orientation angles of about one minute and better. It seems that the variation of the camera position ($s_x, s_y, s_z$) decreases, the lower the camera is located.

Of special interest is the parameter of the camera pose, which makes a major contribution to the variation of the camera position. To identify the direction with the largest variation, a principal component analysis is performed on the set of estimated camera positions. It turned out that the direction with the largest variation is almost parallel to the estimated basis vector $b_c$, as shown in Figure 7.5.

**Figure 7.5: Estimated camera positions for the static set-up at camera height Z = -720 mm.** The left image is a lateral view and the right image is from the top. The green line represents the mean direction between the camera projection centre and the projector centre. The red line represents the direction of the largest variation of the camera position.

Due to the fact that the basis vector is multiplied by the metric scale $\hat{m}$, the metric scale is identified as the parameter with the major contribution to the variation of the camera position. In case of static measurements with camera height $Z = -720 \text{ mm}$, the variance in the direction with the largest standard deviation is approximately $0.97 \text{ mm}$, which corresponds to the standard deviation of the distance measure $s_m = 0.95 \text{ mm}$ in Table 7.2.
7.2.3 Sensitivity Analysis

In Saltelli et al. (2010), the term Sensitivity Analysis is described as “...the study of how uncertainty in the output of a model (numerical or otherwise) can be apportioned to different sources of uncertainty in the model input factors...” (Saltelli et al., 2010). Herein, the uncertainty in the model input factors is considered in a normal statistical context. In the previous sections, the metric scale \( \hat{m} \) has been identified as a decisive factor to the system’s measurement accuracy. According to equation (6.51), the metric scale depends on the image ray \( d_c \) to the laser spots, the basis vector \( b_c \) of the mobile camera, the green laser spot’s offset vector \( b_g \), the distance \( b \) of the laser beam to the projector centre and the direction \( d_g \) of the green laser. Especially, the influence of systematic variations in distance \( b \) and the centroid accuracy of the green spots on the metric scale \( \hat{m} \) are of interest. Further, the question arises of how the imaging geometry affects the estimation of the metric scale.

The distance \( b \) between the projector centre and the initial point of the green laser beam can be affected by systematic variations due to temperature effects or a constant bias due to an imprecise determination of \( b \). Now, it is of interest how such variations in \( b \) affect the metric scale \( \hat{m} \) and if a dependency with respect to the mapping geometry exists. According to equation (6.20), the metric scale \( m \) is linear dependent on the laser offset distance \( b \). Thus, it can be assumed that a small systematic variation \( \Delta b \) causes a small systematic variation \( \Delta m \):

\[
\Delta m = \Delta b \left( \frac{d_g \times d_c}{d_g \times d_c} \right) \frac{b_g}{b_c} \tag{7.1}
\]

Resolving equation (6.20) to the fraction term

\[
\frac{m}{b} = \left( \frac{d_g \times d_c}{d_g \times d_c} \right) \frac{b_g}{b_c} \tag{7.2}
\]

the fraction term itself can be expressed by the ratio of metric scale \( m \) and the laser offset distance \( b \).

![Graph showing the relationship between distance and ratio](image)

Figure 7.6: Amplification of the systematic shift in \( b \). Here, a laser-offset distance of \( b = 250 \text{ mm} \) is assumed.

Replacing the fraction term in equation (7.1) by the ratio in equation (7.2), results in a simplified expression of the metric scale variation

\[
\Delta m = \Delta b \frac{m}{b} \tag{7.3}
\]
Obviously, the metric scale variation is only affected by the variation of the laser-offset distance, which is amplified by the ratio of the metric scale and projector and the laser-offset distance. Thus, given a fix variation $\Delta b$ and a fix laser-offset distance $b$, the only influencing factor of the mapping geometry is the metric scale respectively the distance between the projector centre and the camera projection centre. Assuming a laser-offset distance of about $b = 250$ mm, as it is the case for the CLIPS projector, the variation $\Delta b$ is amplified as shown in Figure 7.6. For instance, a distance $m = 2500$ mm between the camera and the projector amplifies $\Delta b$ ten times. An offset variation of $\Delta b = 0.1$ mm results in a metric scale variation of $\Delta m = 1$ mm. Additionally, it can be shown with equation (7.3) that a large laser-offset distance $b$ decreases the impact of the offset variation on the system scale.

![Figure 7.6: Variation of the metric scale for two different static camera set-ups. The solid line represents camera set-ups at a height of $Z = 0$ m, whereas the dashed line represents camera locations at a height of $Z = -2.46$ m. The azimuth angle represents the azimuth of the camera's basis vector $b_z$.](image)

However, the influence of a systematic laser-offset variation does not explain the scale variations in Figure 7.5. Since the camera has a static set-up for the precision evaluation, only a variation of the green laser centroids in the images can cause variations of the metric scale in Figure 7.5. To evaluate the impact of the laser centroids on the metric scale, several camera set-ups have been simulated with a centroid variation of 0.01 pixels respectively 0.000465 mm. For the first evaluation, two static camera set-ups at a height of $Z = 0$ m, $Z = -2.46$ m, a distance to the projector of 2.5 m and only one green laser offset $b_g = (1, 0, 0)^T$ is assumed. The azimuth angle represents the azimuth of the basis vector $b_z$. The resulting variation of the metric scale is shown in Figure 7.7. Obviously, there are discontinuities at azimuth angles of 0° and 180°. At these locations, the camera is in the same plane spanned by the laser beam and green laser's basis $b_g$, and, thus, the metric scale $m$ becomes an ill-posed problem. Since a centroid variation can be regarded as a variation of the image ray direction $d_g$, the impact of a centroid variation depends on the distance to the green laser spot. It means that centroid variations cause slightly larger variations at lower camera heights, for example $Z = -2.46$ m than at higher camera heights, for example, $Z = 0$ m.
Figure 7.8: Metric scale variation $\Delta m$ for a static camera set-up with a distance of 2.5 m to the projector centre. Two green laser offsets and a camera height of $Z = 0$ mm are assumed.

In Figure 7.8, the scale variation for two perpendicular scale-cross arms is shown. The graphs for the counterparts are similar. Compared to each other, the discontinuities of both curves reveal a shift of 90°. To avoid these discontinuities, the weighted average of the different metric scales is used according to Chapter (6.4).

Figure 7.9: Metric scale variation $\Delta m$ of the basis vector length for four different static camera set-ups with a distance of 2.5 m, 3.0 m, 3.5 m and 4.0 m to the projector centre. The camera height is assumed with $Z = 0$ m. Only one green laser offset with $b_g = (1, 0, 0)^T$ is assumed.

For the third evaluation, variations of the basis vector length for different distances between the camera and the projector have been simulated, as shown in Figure 7.9. The camera height is about $Z = 0$ m. In comparison to the variation of the offset length $b$, the centroid variations reveal a larger influence on the metric scale. Having a calibrated offset length $b$ with an accuracy of about 0.1 mm, a distance of 2.5 m between the camera and the projector amplifies the offset variation about 10 times. The resulting distance variation is about 1.0 mm. In contrary, a centroid variation of a hundredth pixel for a camera-projector distance can cause a variation of the metric scale of up to 10 mm. Generally, the scale variations increase with an increasing distance between the projector and the camera, as shown in Figure 7.6, Figure 7.7 and Figure 7.9.
7.3 Time Statistics

Nowadays, kinematic processes like navigation, tracking or rapid stakeouts in engineering and construction require measurement tools like tachymeters, which are able to determine positions or orientation angles with update rates higher than 10 Hz. Therefore, it is an objective of CLIPS, to provide camera poses with update rates of 10 Hz or higher. To evaluate, whether this objective has been achieved, the update rate for each camera pose has been recorded for static and kinematic measurements.

Figure 7.10: Update rate for static measurements (left) and kinematic measurements (right). The horizontal lines represent the mean update rate for the corresponding camera set-up.

In Figure 7.10, the update rates for static and kinematic measurements are shown. Hereby, the time series have been smoothed by a weighted linear local least squares regression, based on a 1st degree polynomial model (Mathworks, 2012). For static and kinematic measurements it can be seen that the update rate decreases from 10 – 12 Hz at the beginning to 3 – 4 Hz at camera position \( k = 2000 \). Afterwards, a jump in the update rate from 3 – 4 Hz to 5 – 6 Hz can be observed. In comparison to static measurements, the update rate of kinematic measurements reveals larger variations due to the fact that loss of sight increases the time span to the next estimated camera pose. However, the same trend as for static measurements can be observed. The overall decrease of the update rate is caused due to Matlab's memory management. The reservation of new memory leads to a leap to 5 – 6 Hz. Additionally, the graphical output of the camera’s track and the image preview decelerate the update rate.

7.4 Performance Summary

At the beginning of this Chapter, the objectives of CLIPS have been stated. Subsequently, the performance of CLIPS has been analysed with respect to accuracy, precision and the measurement rate. In Table 7.3, the performance parameters of CLIPS and two commercial optical measurement tools are given.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>CLIPS</th>
<th>AICON ProCam (Mautz, 2012)</th>
<th>Total Station (Mautz, 2012)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy (mm)</td>
<td>&lt; 1 cm</td>
<td>0.1 mm</td>
<td>2 mm + 5 ppm</td>
</tr>
<tr>
<td>Coverage Area (m)</td>
<td>3–5</td>
<td>Depends on measuring volume (mostly size of a car)</td>
<td>&gt; 2000 m</td>
</tr>
<tr>
<td>Cost (CHF)</td>
<td>2000 (estimated)</td>
<td>n/a</td>
<td>&gt; 10’000</td>
</tr>
<tr>
<td>Update Rate (Hz)</td>
<td>5 – 10</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>Infrastructure</td>
<td>Projector</td>
<td>Coded targets</td>
<td>Prisms</td>
</tr>
<tr>
<td>Output Data</td>
<td>6 DOF</td>
<td>6 DOF</td>
<td>3 DOF</td>
</tr>
<tr>
<td>Line of Sight to</td>
<td>Laser spots</td>
<td>Coded targets</td>
<td>Measured object</td>
</tr>
<tr>
<td>Number of Users</td>
<td>unlimited</td>
<td>unlimited</td>
<td>1</td>
</tr>
</tbody>
</table>
8 Concluding Remarks

8.1 Conclusion

In this work, the development of the optical indoor positioning system CLIPS is divided in the aspects of literature research, the overall system design, the calibration, the laser spot recognition and the pose estimation. Additionally, the system performance has been evaluated with regard to accuracy, precision, sensitivity and update rate.

8.1.1 Concept

Designed as a full 6 DOF (Degrees of Freedom) measuring tool, CLIPS not only provides 3D position information but also 3D orientation with respect to the laser projector at an update rate of 5 Hz to 10 Hz. Since the reference information is projected, there is no need to establish a reference field, for example coded targets. This enables high flexibility for the application of CLIPS. The implementation of additional green laser pointers with a known offset to the projector centre enables distance estimation between the camera and the projector.

8.1.2 Calibration

For accurate measurements, a calibration of the camera and the projector is necessary. Hereby, the camera calibration is carried out automatically with a standard software package such as iWitness. Only the target distribution and the capturing of the targets involve minor manual work (~1 hour). The software automatically estimates the inner camera orientation and lens distortion parameters. In contrast, the projector calibration – as it is now – requires a larger manual effort (~1 day). The projection surfaces must be shifted and for each surface location, 3d spot coordinates must be measured with a theodolite measuring system or with photogrammetric techniques. This procedure takes about one day. Having obtained the 3d spot coordinates for each surface location, the processing with the offset and direction estimation of the laser beams is done automatically within seconds. With the current calibration approach, laser offsets can be determined with an accuracy of 0.08 mm. The directions are determined with an accuracy of 0.007°.
8.1.3 Spot Recognition

An improved camera triggering and the implementation of the new laser spot recognition algorithm enables measurement rates up to 10 Hz compared to rates of 2 Hz before applying these improvements. Here, the recognition benefits from the fusion of the colour channel combination approach with the template-matching algorithm. Although disturbing light sources are eliminated through the colour channel combination, brighter areas bloom laser spots such that distinction between the laser spots and the environment becomes difficult. To reduce blooming effects, a low sensor exposure time and a moderate gain are chosen. Finally, the improved approach enables the identification for each image by exploiting the colour coded spot pattern of red and green laser spots. Until now, an almost planar projection surface is necessary for the spot identification algorithm, due to the small number of projected laser spots.

8.1.4 Camera Pose Estimation

The challenge of CLIPS is the absence of approximate values for the first camera pose. For that reason, the camera pose estimation has to be initialized in the 8\textsuperscript{th} octant of the projector coordinate system. For this octant, a set of approximate values is generated and refined by a least squares adjustment. The best solution is chosen and the 3d laser spot coordinates are derived via intersection. For consecutive camera locations, the camera pose is determined via resection and the 3d laser spot coordinates are refined. In principle, the distance estimation between the camera and projector by means of the green laser spots is functional. However, the projected basis information of about 50 cm is far too small. For an assumed workspace of 5 m small changes in the green lasers’ offsets and other uncertainties become extrapolated tenfold, causing larger uncertainty in the resulting camera position. Unfortunately, a larger base isn’t practicable, and, thus, other approaches to handle the introduction of the metric scale must be considered. A possible approach is given by the measurement of scale bar endings. The comparison between measured and given scale-bar-length enables the application of a scale. Alternatively, laser pointers can be replaced by laser distance meters to obtain 3d laser spot coordinates directly.

8.1.5 Evaluation of CLIPS

For the performance assessment of CLIPS, a comparison between determined 3d laser spot coordinates and measured 3d laser spot coordinates has been carried out. The comparison revealed a current positioning accuracy of CLIPS in the cm-range. The position’s repeatability is in the sub-mm range. Hereby, the largest variation can be observed along the translation vector of the camera. This indicates a major influence of the metric scale on the camera’s position. In turn, the distance estimation is mostly affected by the spot centroid estimation. As an overall conclusion, the current instrumental realization of the CLIPS approach delivers camera positions with cm-accuracy.
8.2 Outlook

8.2.1 Future Work

To improve the spot identification, the accuracy and the workspace, a new CLIPS architecture may be considered.

One approach is the replacement of the laser projector by a diffractive grid that projects a dense and pseudo random spot pattern in every room direction similar to the Kinect system by Microsoft (2012). The advantage of such a pattern is threefold. Since a dense pattern would be projected to most of the surfaces in a room, the flexibility of the camera handling is increased. For almost every camera viewing direction, enough spots can be found to estimate the camera pose. Second, the spot identification is liberated from the constraint to project spots on a planar projection surface. Assuming a dense pattern, the pattern distortion for small pattern sub-samples becomes smaller than for the overall pattern. Thus, spot identification can be supported by the search for pattern areas with weak pattern distortions. Finally, the accuracy of the pose estimation is improved since a dense spot pattern increases the redundancy for the least squares adjustment.

Further, the replacement of the scale cross with its green laser pointers by laser distance meters will improve the positioning accuracy. Knowing the laser directions and the laser offsets with respect to the projector, the 3d spot coordinates can be determined via distance measurements with mm-accuracy. Once the 3d spot coordinates are known, the camera pose can be estimated via resection. Here, an accuracy improvement to mm-positions is expected.

The prediction of the camera pose can be supported by an inertial measurement unit (IMU), which could be attached to the camera. Although IMUs show a drift during runtime, they provide precise information about changes in position and orientation between two subsequent camera poses. By means of these changes, the Kalman filter can be extended to improve the prediction of the next camera pose, and, thus, the provision of better approximate values.

8.2.2 Vision

Generally, the pose of every available camera (e.g., on a tool, headset or smartphone) can be determined. Imagine an architect or engineer on a construction site, who points his smartphone on a spot projection. With a special construction site app, the smartphone’s pose can be determined and construction relevant data can be overlaid on the screen. If a small camera attached to a tool, its position and orientation can be determined and compared with the desired location of its application. Over a small screen, the operator is informed about the current and the desired tool location. This way, the operator is precisely guided to the location where the tool must be applied.

In combination with augmented reality, CLIPS has a large potential for various indoor applications. Having a camera attached to a headset, the head position, orientation and motion can be determined. With a transparent heads-up display in front of the eyes, additional information can be displayed, depending on the operator’s viewing direction. Promising headsets with a camera and a transparent head up display, as with the Google Glass project (Google, 2012), are currently under development. To give an example, let us assume indoor areas at a construction site. The projector will be set-up in a room, such that its pose is known with respect to the room. Once the projector is switched on, the operators head position and
Concluding Remarks

orientation can be determined via the headset camera. According to the head position and orientation, construction relevant data like the run of electric cables or water lines or other data from CAD plans etc. can be displayed on the head up display.

Alternatively, CLIPS can be extended to a Simultaneous Location and Mapping (SLAM) system to generate 3d models of indoor environments. Knowing the camera pose and the camera image for each camera location, a 3d point cloud can be derived via bundle block adjustment. With a dense spot pattern, this point cloud can be derived directly by the determination of 3d spot coordinates. Additionally, texture information can be used to increase the density of the estimated point cloud. Enabling CLIPS to provide 3d point clouds of indoor environments (e.g. at a construction site), it can be used to document the work’s progress.

Finally, CLIPS is not limited to a certain number of operators. Once the projector is set-up in the indoor environment, CLIPS can be used as a tool to provide locations for numerous tools, smart phones or other objects, which are equipped with a camera. Due to this fact, several operators can use CLIPS simultaneously. Thus, CLIPS can be regarded as a multi functionality tool, which enables “total station free surveying” (Prof. Dr. Hilmar Ingensand, personal communication, 2012).
Appendix A - Laser Classes

The use of laser radiation involves risks for skin and eyes. Depending on wavelength, intensity and laser mode, different laser light-object interactions can occur. In the case of eyes the damages range from conjunctivitis to total blindness. Serious damages on the skin may cause cancer. For reasons of eye and skin safety, lasers can be categorized into seven classes, regarding possible risks, wavelengths and power levels, as shown in Table 8.1.

Table 8.1: Laser Classes regarding the risks, the wavelength and the power of the laser (International Electrotechnical Commission, 2009)

<table>
<thead>
<tr>
<th>Laser Class</th>
<th>Risks</th>
<th>Wavelength (nm)</th>
<th>Power (mW)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>eye safe</td>
<td>all</td>
<td>wavelength independent</td>
</tr>
<tr>
<td>1M</td>
<td>similar to class 1 but without optical instruments</td>
<td>302.5 – 4000</td>
<td>wavelength independent</td>
</tr>
<tr>
<td>2</td>
<td>eye safe through lid closure reflex</td>
<td>400 – 700</td>
<td>1</td>
</tr>
<tr>
<td>2M</td>
<td>similar to class 2 but without optical instruments</td>
<td>400 – 700</td>
<td>1</td>
</tr>
<tr>
<td>3R</td>
<td>between class 2M and 3B</td>
<td>&gt; 302.5</td>
<td>5</td>
</tr>
<tr>
<td>3B</td>
<td>harmful if you’re staring directly into the beam</td>
<td>all</td>
<td>500</td>
</tr>
<tr>
<td>4</td>
<td>very harmful for skin and eyes, even diffuse reflexions</td>
<td>all</td>
<td>&gt; 500</td>
</tr>
</tbody>
</table>
Appendix B - Derivatives of the Co-Planarity Constraint

According to Luhmann (2010), the derivatives for the observation equation in (6.9) can be written as followed:

\[ \frac{\partial \Delta}{\partial b_{c,y}} = \begin{bmatrix} 0 & d_{p,x} & d_{c,x} \\ 1 & d_{p,y} & d_{c,y} \\ 0 & d_{p,z} & d_{c,z} \end{bmatrix} \]

\[ \frac{\partial \Delta}{\partial b_{c,x}} = \begin{bmatrix} 0 & d_{p,x} \\ 0 & d_{p,y} \\ 1 & d_{p,z} \end{bmatrix} \]

\[ \frac{\partial \Delta}{\partial \omega} = \begin{bmatrix} 1 & d_{p,x} & 0 \\ b_{c,y} & d_{p,y} & -d_{c,z} \\ b_{c,x} & d_{p,z} & d_{c,y} \end{bmatrix} \]

\[ \frac{\partial \Delta}{\partial \varphi} = \begin{bmatrix} 1 & d_{p,x} & -d_{c,y} \sin(\omega) + d_{c,x} \cos(\omega) \\ b_{c,y} & d_{p,y} & d_{c,x} \sin(\omega) \\ b_{c,x} & d_{p,z} & -d_{c,x} \cos(\omega) \end{bmatrix} \]

\[ \frac{\partial \Delta}{\partial \kappa} = \begin{bmatrix} 1 & d_{p,x} & -d_{c,y} \cos(\omega) \cos(\varphi) - d_{c,x} \sin(\omega) \cos(\varphi) \\ b_{c,y} & d_{p,y} & d_{c,x} \cos(\omega) \cos(\varphi) - d_{c,x} \sin(\varphi) \\ b_{c,x} & d_{p,z} & d_{c,x} \sin(\omega) \cos(\varphi) + d_{c,y} \sin(\varphi) \end{bmatrix} \]

With:

- \( b_c = (b_{c,x}, b_{c,y}, b_{c,z})^T \) : Basis vector between projector and camera
- \( d_p = (d_{p,x}, d_{p,y}, d_{p,z})^T \) : Direction vector of the laser beams
- \( d_c = (d_{c,x}, d_{c,y}, d_{c,z})^T \) : Direction vector of the image rays to the laser spots
- \( (\omega, \varphi, \kappa) \) : Orientation angles of the camera orientation
Appendix C - Derivatives for the Resection Algorithm

According to Luhmann (2010), the derivatives for the observation equation in (6.14) can be written as followed:

\[
\frac{\partial x_c}{\partial X_c} = -\frac{f}{D^2}(r_{13}N_x - r_{11}D)
\]

\[
\frac{\partial x_c}{\partial Y_c} = -\frac{f}{D^2}(r_{23}N_x - r_{21}D)
\]

\[
\frac{\partial x_c}{\partial Z_c} = -\frac{f}{D^2}(r_{33}N_x - r_{31}D)
\]

\[
\frac{\partial x_c}{\partial \omega} = -\frac{f}{D}\left\{\frac{N_x}{D}\left[r_{33}(Y_i - X_c) - r_{23}(Z_l - Z_c)\right] - r_{31}(Y_l - Y_c) + r_{21}(Z_l - Z_c)\right\}
\]

\[
\frac{\partial x_c}{\partial \phi} = -\frac{f}{D}\left\{\frac{N_x}{D}\left[N_x \cos(\kappa) - N_y \sin(\kappa)\right] + D \cos(\kappa)\right\}
\]

\[
\frac{\partial x_c}{d\kappa} = -\frac{f}{D}N_y
\]

\[
\frac{\partial y_c}{\partial X_c} = -\frac{f}{D^2}(r_{13}N_y - r_{12}D)
\]

\[
\frac{\partial y_c}{\partial Y_c} = -\frac{f}{N^2}(r_{23}N_y - r_{22}D)
\]

\[
\frac{\partial y_c}{\partial Z_c} = -\frac{f}{N^2}(r_{33}N_y - r_{32}D)
\]

\[
\frac{\partial y_c}{\partial \omega} = -\frac{f}{D}\left\{\frac{N_y}{D}\left[r_{33}(Y_i - X_c) - r_{23}(Z_l - Z_c)\right] - r_{32}(Y_l - Y_c) + r_{22}(Z_l - Z_c)\right\}
\]

\[
\frac{\partial y_c}{\partial \phi} = \frac{f}{D}\left\{\frac{N_y}{D}\left[N_x \cos(\kappa) - N_y \sin(\kappa)\right] - D \sin(\kappa)\right\}
\]

\[
\frac{\partial y_c}{d\kappa} = \frac{f}{D}N_x
\]
With:

\( f \): Focal length of the camera.

\( N_x, N_y \): Nominators of the co-linearity equations in (6.15).

\( D \): Denominator of the co-linearity equations in (6.15).

\( L = (X_l, Y_l, Z_l) \): 3D Coordinates of a laser spots in the projector coordinate system.

\( (x_c, y_c, z_c) \): Image coordinates of a laser spot.

\( O_c = (X_c, Y_c, Z_c) \): Coordinates of the camera position in the projector coordinate system.

\( (\omega, \varphi, \kappa) \): Orientation angles of the camera orientation.
Publication List

Conferences and Journals

Throughout the CLIPS project, the basic concept of CLIPS and progresses of the development have been presented to the scientific community in oral presentations, demonstrations and publications. The following publications according to CLIPS have been published:


**Student Works**

Furthermore, some aspects of CLIPS have been investigated in two Bachelor theses, one Master Project Work and one Master thesis.


### List of Figures

<table>
<thead>
<tr>
<th>Figure Number</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Classification of optical indoor positioning systems</td>
</tr>
<tr>
<td>2.2</td>
<td>Interior building model, based on CityGML (Kohoutek et al., 2010)</td>
</tr>
<tr>
<td>2.3</td>
<td>Reference introduction via an image sequence</td>
</tr>
<tr>
<td>2.4</td>
<td>Illustration of coded targets</td>
</tr>
<tr>
<td>2.5</td>
<td>Example of projected targets</td>
</tr>
<tr>
<td>3.1</td>
<td>Basic Concept of CLIPS</td>
</tr>
<tr>
<td>3.2</td>
<td>Mount for the red and green laser pointers</td>
</tr>
<tr>
<td>3.3</td>
<td>Projector Controller</td>
</tr>
<tr>
<td>4.1</td>
<td>Illustration of camera lens distortion</td>
</tr>
<tr>
<td>4.2</td>
<td>Comparison of a laser spot with an ideal 2D Gaussian intensity distribution and the intensity distribution of a captured laser spot</td>
</tr>
<tr>
<td>4.3</td>
<td>Illustration of diffraction</td>
</tr>
<tr>
<td>4.4</td>
<td>Illustration of the calibration set-up</td>
</tr>
<tr>
<td>4.5</td>
<td>Offset displacements of the red laser beams between two different calibration epochs</td>
</tr>
<tr>
<td>4.6</td>
<td>Variation of the direction angles between 2 calibrations</td>
</tr>
<tr>
<td>5.1</td>
<td>Influence of sensor exposure time and sensor gain on the image brightness</td>
</tr>
<tr>
<td>5.2</td>
<td>Mutual impact of sensor exposure time and gain</td>
</tr>
<tr>
<td>5.3</td>
<td>Test image for threshold operations</td>
</tr>
<tr>
<td>5.4</td>
<td>Results of the triangle thresholding</td>
</tr>
<tr>
<td>5.5</td>
<td>Results of the threshold selection according to Otsu’s method</td>
</tr>
<tr>
<td>5.6</td>
<td>Illustration of laser spot recognition by means of morphological operators</td>
</tr>
<tr>
<td>5.7</td>
<td>Illustration of the channel combination</td>
</tr>
<tr>
<td>5.8</td>
<td>Illustration of spot detection by means of difference images</td>
</tr>
<tr>
<td>5.9</td>
<td>Illustration of template matching</td>
</tr>
<tr>
<td>5.10</td>
<td>Illustration of a laser spot’s shape and intensity distribution in a grey value image</td>
</tr>
<tr>
<td>5.11</td>
<td>Comparison of spot detection approaches</td>
</tr>
<tr>
<td>5.12</td>
<td>Comparison of spot detection approaches</td>
</tr>
<tr>
<td>5.13</td>
<td>Illustration of laser spot identification algorithms</td>
</tr>
<tr>
<td>6.1</td>
<td>Epipolar Geometry</td>
</tr>
<tr>
<td>6.2</td>
<td>Camer pose estimation via resection</td>
</tr>
<tr>
<td>6.3</td>
<td>Approximate value generation</td>
</tr>
<tr>
<td>6.4</td>
<td>Introduction of metric information in CLIPS</td>
</tr>
<tr>
<td>6.5</td>
<td>Critical configuration - All laser spots lie on a line</td>
</tr>
<tr>
<td>6.6</td>
<td>Critical configuration: All available laser spots lie on a circular cylinder</td>
</tr>
<tr>
<td>6.7</td>
<td>Critical configuration: All laser spots lie on horopter curves (red &amp; green)</td>
</tr>
</tbody>
</table>
List of Figures

Figure 6.8: Critical configuration for the estimation of the metric scale........................................55
Figure 6.9: Pose Estimation Algorithm .............................................................................................57
Figure 7.1: Evaluation Criteria for indoor positioning systems..........................................................59
Figure 7.2: Evaluation Set-Up .............................................................................................................61
Figure 7.3: Accuracy of the camera position for three static camera set-ups and one kinematic measurement..................................................................................................................62
Figure 7.4: Accuracy of the camera orientation for three static camera set-ups and one kinematic measurement...........................................................................................................................62
Figure 7.5: Estimated camera positions for the static set-up at camera height $Z = -720$ mm.........63
Figure 7.6: Amplification of the systematic shift in $b$.....................................................................64
Figure 7.7: Variation of the metric scale for two different static camera set-ups.............................65
Figure 7.8: Metric scale variation for a static camera set-up with a distance of 2.5 m to the projector centre. Two green laser offsets and a camera height of $Z = 0$ mm are assumed........66
Figure 7.9: Metric scale variation of the basis vector length for four different static camera set-ups. The camera height is assumed with $Z = 0$ m. Only one green laser offset is assumed...........66
Figure 7.10: Update rate for static measurements (left) and kinematic measurements (right). The horizontal lines represent the mean update rate for the corresponding camera set-up.......67
List of Tables

Table 3.1: Sensor Parameters of AVT GUPPY GF 080 C (AVT, 2009) ........................................12
Table 3.2: Characteristic parameters of the applied lasers. .........................................................13
Table 5.1: Look-Up-Table for the laser spot identification .............................................................42
Table 7.1: PC Hardware ..................................................................................................................61
Table 7.2: Standard deviations (1-sigma) of measurement series for different static camera set- ups at height Z .........................................................................................................................63
Table 7.3: Performance parameters of CLIPS and comparison to commercial positioning systems. .................................................................................................................................67
Table 8.1: Laser Classes regarding the risks, the wavelength and the power of the laser (International Electrotechnical Commission, 2009) ................................................................67
Acknowledgements

This dissertation is the outcome of the CLIPS project at the Institute of Geodesy and Photogrammetry, ETH Zurich. The project has been funded by the Swiss National Fund (SNF). At this point I would like to thank everyone who has accompanied and supported me throughout the last three years. Without their support, the development of CLIPS and the work on the dissertation would not have been possible.

First, I would like to thank Prof. Dr. Hilmar Ingensand for giving me the unique opportunity to participate in the genesis of a novel positioning system. Particularly valuable to me have been his experiences in every stage of instrumental development, the fruitful discussions, and his visionary thinking. Further, I appreciate that he opened so many doors to the industry, so that I have learned the industrial perspective on instrumental development.

I would like to acknowledge my co-examiner Prof. Dr.-Ing. Thomas Wunderlich, who shared his knowledge and experience for critical configurations in positioning and navigation and made very valuable suggestions to my work.

I am greatly indebted to PD Dr. Rainer Mautz – the inventor of CLIPS. He aroused my interest for this project very early during my studies and encouraged me to work on CLIPS. I appreciate his visionary thinking and enthusiasm for this project, and I enjoyed the many fruitful and long discussions about CLIPS, surveying in general and many other topics. Further, his profound experience in scientific writing was very valuable for the improvement of my own writing skills. I also would like to thank him that he took his time to proof-read this dissertation.

I thank Paul Sorber for the development and implementation of the projector electronics. He took his time whenever new electronic parts were required or if defect hardware had to be replaced. My acknowledgements also apply to Hans-Peter Oesch, who supported the production of the projector from the drawing board to its realization.

I wish to thank my colleagues David Grimm, Tobias Kohoutek, David Novak and Pascal Theiler for the many discussions, for their valuable suggestions and comments, for all the common projects and for their understanding that the work on my dissertation sometimes possessed priority over other tasks.

My deepest gratitude goes to my family, especially to my Mum Heike Tilch and my Brother Erik Tilch who were with me during all ups and downs and who supported me in so many ways that I cannot list them all. I dedicate this work to them.
Curriculum Vitae

Personal Data

Name: Sebastian Tilch
Date of Birth: 24.04.1984
Place of Birth: Leipzig
Nationality: German

Education

2010 - 2012 Doctoral studies at the Institute of Geodesy and Photogrammetry, ETH Zurich.


Working Experience

2010 - 2012 Research and Teaching Assistant
Institute of Geodesy and Photogrammetry, ETH Zurich
Geodetic Metrology and Engineering Geodesy (Prof. Dr. H. Ingensand)

2009 Student Assistant
Institute of Geodesy and Photogrammetry, ETH Zurich
Geodetic Metrology and Engineering Geodesy (Prof. Dr. H. Ingensand)

2006 - 2008 Student Assistant
Institut für Geodäsie, GIS und Landmanagement, TU München
Lehrstuhl für Geodäsie (Prof. Dr. T. Wunderlich)
Fachgebiet Geoinformationssysteme (Prof. Dr. M. Schilcher)

2004 - 2005 Trainee
Engineer’s office Rossipal, Fürstenfeldbruck

2003 - 2004 Trainee
Land surveying office, Fürstenfeldbruck