**Institute of Electrical Measurement and Measurement Signal Processing**
**Institute of Lightweight Design**
Graz University of Technology
A-8010 Graz, Austria

---

Master Thesis

# Image-based Measurement of Relative Motions between Railway Vehicle Carbodies

Author:
**Robert Hödl, BSc.**

Supervisors:
**Ao. Univ.-Prof. Dipl.-Ing. Dr.techn. Axel Pinz**
**Dipl.-Ing. Dr.techn. Thomas Thurner**

January, 2013

# Abstract

The test and validation process of connecting components between railway vehicles comprises the determination of the performed relative motions between two adjacent carbodies during operation. Existing measurement systems have the drawback of being extensive and time consuming regarding installation, measurement and analysis. This thesis is concerned with the feasibility study and prototype development of a robust and cost-efficient image-based measurement system, which is capable of tracking the relative motions between railway vehicle carbodies.

First, the thesis defines the operational requirements of an image-based measurement system. Then a suitable marker and an accompanying target design as well as a robust tracking method are introduced. Further, two appropriate pose estimation algorithms are determined and chosen for evaluation. Moreover, an adequate measurement setup relating to the specified carbody motion model is developed and the corresponding prerequisites are described. At last, these considerations enable the definition an appropriate optical imaging system.

The feasibility of the designed image-based measurement system is investigated by extensive experiments conducted on a laboratory scale and using a full-scale test rig. Detailed evaluation of the uncertainties is carried out which allows to derive important implications concerning the measurement setup. From the results, it is apparent that the proposed system meets the specified requirements and is capable of measuring the relative motions within the defined limits, but to a certain extent remains susceptible to inaccuracies in the measurement setup and changing lighting conditions. Recommendations for a specific pose estimation method and further enhancements to increase robustness are given. The suitability of the system is finally verified in the course of a test ride on board of a high-speed train.

The proposed image-based measurement system contributes a novel, genuine alternative to conventional methods applied to the particular task of measuring the relative motions between two railway vehicle carbodies. It perfectly fulfils the technical and economic requirements.

# Kurzfassung

Zur Auslegung von verbindenden Komponenten zwischen Wagenkästen, werden in der Entwicklung von Schienenfahrzeugen umfangreiche Tests und Prüfungen durchgeführt. Im Zuge dieser Tests ist es auch erforderlich die Relativbewegungen zwischen zwei zusammenhängenden Wagenkästen zu ermitteln. Aktuell eingesetzte Messverfahren haben den Nachteil, einerseits sehr zeit- und kostenaufwendig bei Installation und Inbetriebnahme, sowie andererseits komplex in der Auswertung zu sein. Im Rahmen dieser Arbeit wird deshalb die Realisierbarkeit eines bildgestützten Systems zur Messung von Relativbewegungen zwischen Wagenkästen überprüft. Eine prototypische Umsetzung soll sowohl die Eignung als auch die Robustheit und Wirtschaftlichkeit eines solchen Systems verifizieren.

In dieser Arbeit werden zuerst die grundlegenden Anforderungen und Einsatzbedingungen eines bildgestützten Messverfahrens festgelegt. Weiters werden eine geeignete Messmarke samt einhergehender Anordnung auf einem planaren Zielobjekt und eine zugehörige Trackingmethode vorgestellt, sowie zwei verschiedene Algorithmen zur Poseschätzung präsentiert. Passend zu dem vorgegebenen Modell der Wagenkastenbewegungen wird ein eigens entwickelter Messaufbau und die dafür notwendigen Voraussetzungen beschrieben. Aus diesen Erkenntnissen is abschließend ein adäquates optisches Aufnahmesystem ableitbar.

Die Realisierung und Umsetzung des entwickelten Systems wird in verschiedenen Experimenten, im Labor und auf einem Prüfstand, untersucht. Die ausführliche Evaluierung der Messunsicherheiten erlaubt wichtige Rückschlüsse auf den Messaufbau. Aus den Resultaten der Experimente ist schließlich ersichtlich, dass das präsentierte Verfahren in der Lage ist, die gesuchten Relativbewegungen innerhalb der geforderten Genauigkeiten zu messen. Bis zu einem gewissen Grad bleibt das Verfahren jedoch anfällig in Bezug auf Ungenauigkeiten im Messaufbau sowie auf sich ändernde Lichtverhältnisse. Weiters werden Empfehlungen zur Auswahl eines Algorithmus zur Poseschätzung sowie Verbesserungen bezüglich der Robustheit erarbeitet. Die prinzipielle Eignung des entworfenen Systems wird im Rahmen einer Messfahrt an Bord eines Hochgeschwindigkeitszuges nachgewiesen.

Das vorgestellte bildgestützte Verfahren zur Messung von Relativbewegungen zwischen Wagenkästen stellt eine neuartige Alternative zu den konventionell eingesetzten Methoden dar, da es die gestellten Anforderungen hinsichtlich technischer Eignung und Wirtschaftlichkeit bestens erfüllt.

**This thesis is dedicated to my family**
**for their unconditional love**
**and endless support.**

`5c9320d21e24f4875bb4baa185151f45`

# Statuatory Declaration

I declare that I have authored this thesis independently, that I have not used other than the declared sources / resources, and that I have explicitly marked all material which has been quoted either literally or by content from the used sources.

Date

Robert Hödl

# Contents

# 1. Introduction

## 1.1. Motivation and Problem Statement

The subject of this thesis is settled in the domain of railway systems engineering particularly in the field of rolling stock and was conducted in close cooperation with the Siemens AG Austria - Rail Systems Division[1]. As described in [29] almost every modern railway train is composed of several railway vehicles (coaches or wagons). These vehicles are connected by coupling-systems which transmit tensile and compressive forces. Furthermore there are various other elements attached between these vehicles which transfer energy, information and in the case of the corridor connections, allow the transition of passengers and goods.

During operation the railway vehicles perform relative translational and rotational motions so that all connecting elements are subject to high dynamic loads. In order to guarantee secure and fail-safe operation, these components are designed in a way that all maximum displacements or loads stay within given limits defined by certain vehicle specifications.

The design and precise dimensioning of such vehicles and its components is based on extensive simulation and testing. The underlying models and assumptions, used in the simulations, are validated by conducting tests and experiments. A part of this validation process is the measurement of relative motions between carbodies under real-life conditions. The results of such measurements allow a much more precise estimation of the motions and therefore a more accurate dimensioning of the connecting components between the carbodies. Up to now the measurement of these motions relies on measurements from conventional displacement-sensor systems (e.g. draw wire sensors) which have the drawback of being extensive and time consuming regarding installation, measurement and analysis.

Since the economic constraints of passenger and freight railway services are moving the industry to facilitate the engineering of more cost-optimised products and methods, an alternative solution to the established measurement system is required. In his work *von Flottwell* [29] attempted to calculate such relative motions by using accelerometer

---

[1] http://www.siemens.com/rail-systems/

data from carbodies (see Section 2). Despite various approaches, *von Flottwell* does not succeed in his efforts, hence initially proposing the development of a measurement system based on imagery-data.

The feasibility study and prototype development of such a robust and cost-efficient image-based measurement system, which is capable of tracking the relative motions between two adjacent railway vehicle carbodies is the main objective of this work. In order to achieve this goal, it is necessary to define the operational requirements of an image-based measurement system as well as to demonstrate and quantify its feasibility in a real world scenario by means of a prototype implementation.

## 1.2. Railway Vehicle Carbody Motions

In common railway vehicles, four basic components are necessary for the simulation of running behaviour and modelling motions between carbodies: Track, wheelsets, bogie frame and the carbody itself. The assembly of the wheelsets and the bogie frame including the primary and secondary suspension stages is commonly referred to as the bogie. Usually two bogies are attached to one carbody. An example is shown in Figure 1.1.

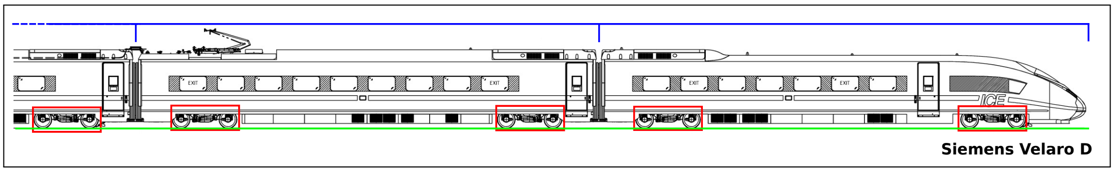


Figure 1.1.: A section of a side view of a Siemens Velaro D serves as an example to identify the main components of modern railway vehicles used in simulation: wheelsets and bogie frame marked with red boxes (referred to as bogie), carbody (blue partitions), rails as parts of the track (green line)

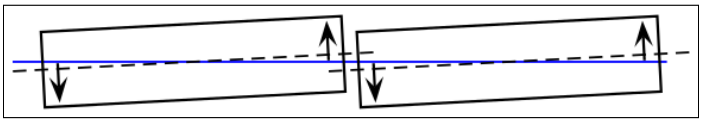
The causes of carbody motions are described in [45] and [29] as a combination of different factors: On the one hand, the vehicle follows the elastic movements of the wheelsets and the suspension as parts of the bogies. These movements are primarily affected by the railway track geometry (e.g. tangent track, curves, transition curves, track cant). In [45] this effect is called *snaking mode*. Due to the rather large curve radii, the occurring frequencies are specified to be smaller than 2 Hz.

On the other hand, the interaction between the rails and the wheelsets causes an oscillating movement called *wheelset hunting* [63]. According to [29] this movement stabilises itself in the character of a damped sinusoidal oscillation with frequencies in the range of 6 to 8 Hz.

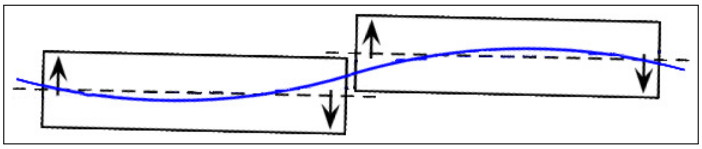
Furthermore, [29] states that the sinusoidal oscillation of the bogie is affected by irregularities of the track. Track irregularities can be classified into five categories: track gauge, vertical, cross level, lateral and twist irregularities [47]. Additionally, they entail a considerable wear and therefore a polygonisation (out-of-roundness) of the wheels. The track irregularities as well as the occuring polygonisation cause distortions in the rolling motions of the wheels, which induces heavy dynamic loads on the bogie. Before acting on the carbody these dynamic accelerations are damped by the primary and secondary suspension units of the bogies. The accelerations transmitted to the carbody are crucial in terms of riding comfort and are specified to be higher than 10 Hz.

In [29], it is concluded that the relevant frequencies concerning the evaluation of relative motions between carbodies lie within the range of 0 to 10 Hz. The highest amplitudes are expected to be in the range from 0 to 2 Hz.

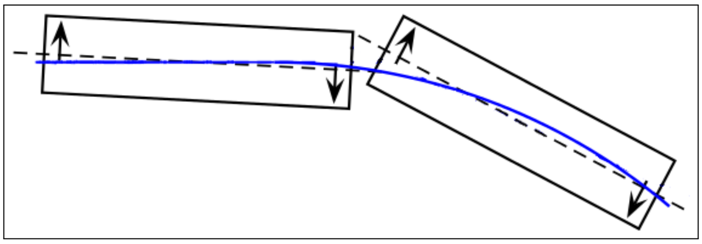
Despite the influences acting on each single vehicle, the relative motions between two carbodies are closely related to the positioning of the vehicles towards each other. This positioning is in turn dependent on the present track geometry as well. Figure 1.2 is adapted from [29] and gives an overview of possible railway vehicle arrangements and their coinciding relative motions affected by the track geometry.



(a) tangent track



(b) s-curve

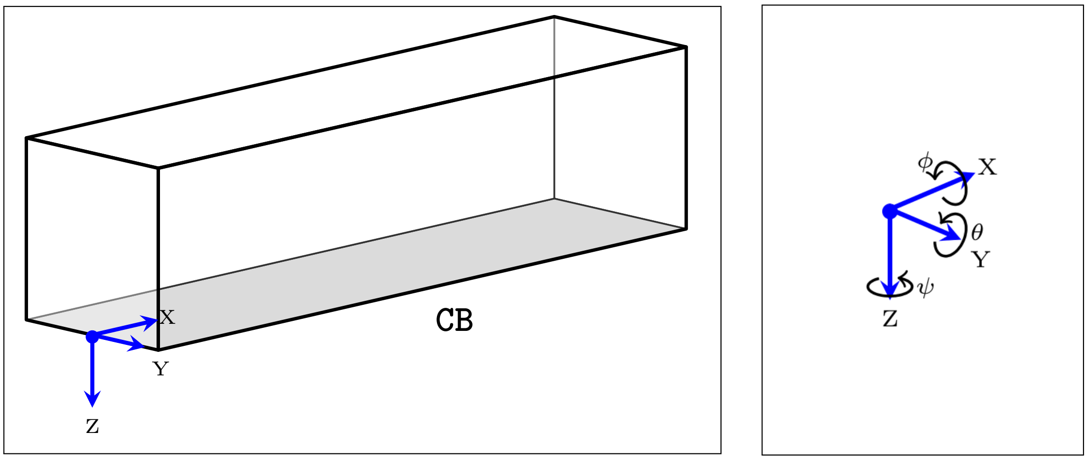


(c) transition curve

Figure 1.2.: (a), (b) and (c) show typical railway track configurations. Railway vehicle arrangements and therefore relative motions between carbodies are mostly dependent on track geometry. Black rectangles symbolise the single railway vehicles whereas the solid blue line denotes the centerline of the railway track. The black dashed line represents the centerline of the vehicles while the arrows indicate the direction of the relative motions. (adapted from [29])

As described by *Hanneforth and Fischer* [35], railway vehicle carbodies perform translational and rotational oscillations. These are classified into six fundamental oscillations (for- and aft oscillation, lateral oscillation, bouncing, pitching, rolling, nosing). However, since the displacement excitations are induced outside the centre of gravity of the carbodies, the fundamental oscillations mostly occur as coupled oscillations (e.g. hunting, swaying, galloping).
A three-dimensional, right-handed Cartesian coordinate system (see Figure 1.3) is used in order to characterize the dynamic behaviour of carbodies with six degrees of freedom (6-DoF). The following list denotes the occurring **relative linear and angular displacements between two carbodies** and defines the corresponding coordinate system commonly used in railway systems engineering. Furthermore, the maximum relative displacements as specified in [42] (see Appendix A) are given:

- **Translation along longitudinal axis**
  - x-axis / in driving direction
  - relative max. displacement[2]: +60/-70 mm
- **Translation along lateral axis**
  - y-axis / perpendicular to driving direction
  - relative max. displacement: ±900 mm
- **Translation along vertical axis**
  - z-axis / vertical to the driving direction
  - relative max. displacement: ±350 mm
- **Rotation about longitudinal axis**
  - roll ($\phi$)
  - relative max. displacement: ±5 °
- **Rotation about lateral axis**
  - pitch ($\theta$)
  - relative max. displacement: ±4 °
- **Rotation about vertical axis**
  - yaw ($\psi$)
  - relative max. displacement: ±15 °



(a) (b)

Figure 1.3.: The coordinate system as commonly used in railway systems engineering. (a) indicates the carbody structure as a semi-transparent cuboid with the origin of the coordinate system for measuring the relative motions located at the carbody ending. The light-grey polygon indicates the carbody ground floor. (b) close-up view on the defined coordinate system with the individual rotations denoted explicitly.

---

[2] value specifically for Siemens Velaro D (SAP-PLM: A6Z00001260505 002)

## 1.3. Requirements, Constraints and Limitations

The following requirements and constraints arising from the problem definition were defined in [42] (see Appendix A) and are considered throughout this work:

R(a) **Accuracy of measurements**
The accuracies which the system is required to meet are specified with a maximum error of ±**10 mm** and ±**0.6**° in the spatial domain by using a minimum sample frequency of **25 Hz** in the time domain.

R(b) **The carbody is considered as a rigid body**
Regarding the given task a carbody can be considered as a permanently rigid body. This means that all twisting and bending deformations [64] within the structure of a carbody itself are assumed to be very small and therefore can be neglected.

R(c) **Measurements are performed inside the carbody**
Although it is possible to conceive a system installed on the outside, it is required that the measurements are conducted within the carbody. This is mainly owed to the fact that most modifications to the exterior of railway vehicles require additional approval by responsible authorities. However, designing the measurement system to be installed within the carbodies, has the advantage of better controllable environmental conditions (lighting, climate, *etc.*) as well as the permanent physical accessibility.

R(d) **Measurements are feasible during passenger service**
The measurement system has to be designed in a way that the system is operational regarding on-board passengers (mostly technicians during test rides). This limits the system to utilize only non-hazardous techniques and for example excludes the implementation of laser-ranging technology or the application of strong infrared (IR) radiation.

R(e) **Applicable for different types of railway vehicles (portability)**
The system should be designed so that it is generically applicable to all different kinds of railway vehicles with the only precondition being that there exists (at least partially) unobstructed visibility[3] from inside one carbody to the other. This requirement also implies that the system needs to be portable from one measuring point to another (intra- and inter-train).

---

[3] e.g. in the presence of a gangway connection, the doors stay open during measurement

R(f) **Limitations in resources, operation and evaluation efforts**
Considering the already mentioned cost-optimisation, the system is also required to make subtle use of its resources in terms of equipment and working hours. A design has to be chosen which facilitates a short installation, set-up and evaluation time, retaining the specified portability (R(e)) and reducing working hours.

R(g) **Real-time capabilities are desirable but not necessary**
Basically, it is not required that the system is capable of computing the relative motions in real-time. With respect to requirement R(f), the applied algorithms should be as computationally efficient (computation time, memory and persistent storage) as possible. Finding an optimal balance between speed and accuracy is required.

## 1.4. Basic Setup and Measurement

In order to optically measure the relative motions between railway vehicle carbodies, the following basic measurement setup is proposed:

- The imaging device and the tracked feature points are installed inside (R(c)) two separate but directly consecutive railway vehicles, in such a way that the imaging device located in one of the vehicles is able to track existing features within the other.
- Due to the elongated appearance of railway vehicles and the narrow corridor connections between them, it is furthermore necessary to place the imaging device at the very end of one vehicle. This enlarges the imaging device's field of view (FOV) within the tracked vehicle, therefore maximising the visibility of the feature points.

A schematic illustration is shown in Figure 1.4.

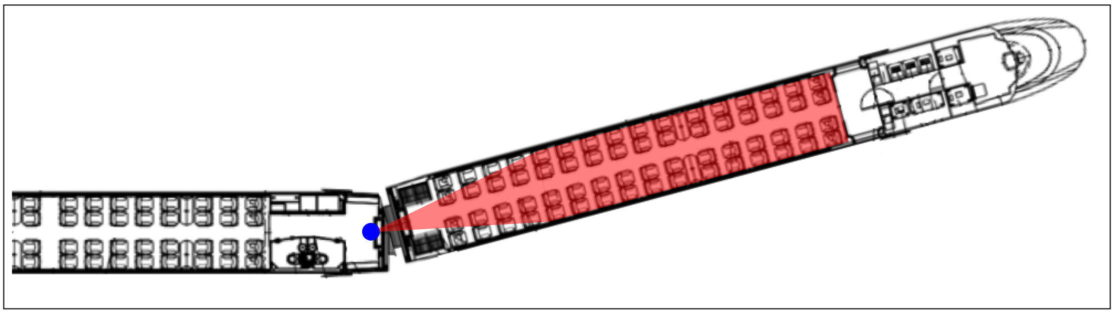


Figure 1.4.: A section of a plan view of a Siemens Velaro D illustrates the proposed basic measurement setup. The imaging device is placed at the very end of one railway vehicle (blue dot) maximising its field of view (red region) within the other, enabling the system to calculate the relative motions.

This proposed basic setup ensures that the system is able to calculate the required relative motions. A thorough model will be introduced in Section 3.4.1.

## 1.5. Outline of the Thesis

This thesis is organized as follows:

Chapter 2 gives an overview of state of the art techniques for measuring rigid body motions, as well as discussing the previous work done by *von Flottwell* in [29]. Since motion can be described as a change in position and orientation over time, the term *pose estimation* is introduced. Furthermore, the advantages of an image-based measurement system are discussed.

Chapter 3 presents all image-processing steps needed in order to measure rigid body motions from imagery data. First, an in-depth description of the essential marker tracking process is given and the final choice for a special type of marker as well as a suitable target design is explained. Second, different types of image-based pose estimation methods and associated theory are discussed. The two algorithms chosen for evaluation are reviewed in detail. Finally, a model which relates the pose estimation methods and the relative motions between railway vehicle carbodies is designed.

Chapter 4 describes a suitable image acquisition system in detail, by first defining the necessary requirements and then introducing an applicable approach to geometric camera calibration. Furthermore, the utilised implementation of a suitable camera calibration procedure is presented accompanied by detailed instructions for camera calibration in practice.

Chapter 5 presents the different experiments conducted in the laboratory and on the test rig in order to evaluate the developed measurement system. The obtained results are discussed, with respect to the achieved measurement uncertainty of the applied methods, in order to demonstrate the system's capabilities, as well as its possible limitations. Furthermore, details about an optimal target design applying a sufficient number of fiducials are examined. Finally, the results of the test ride on an operational high-speed train are presented.

Chapter 6 summarises the presented work and emphasizes the findings and their significance. Additionally, possible improvements and optimizations of the presented approach are discussed. Finally, a best practice method containing comprehensive recommendations for real world operation is derived.

# 2. Related Work

This chapter surveys previous work and presents various related approaches and techniques in the field of (rigid body) pose estimation. The term *pose estimation*[4] is frequently used within the field of computer vision. It is defined as the task to **specify the position and the orientation of an observed object** with respect to a given coordinate system. The process of tracking the pose of a rigid object over time is also referred to as *3D tracking* [53] and can be considered as equivalent to measuring the motion of an object.

Within the scientific community, the term relating to pose estimation changes with the scope of application. For example, in the field of mechanics the process of estimating the position and orientation of an end-effector is called *forward kinematics* [20], whereas in the domain of wireless sensor networks it is commonly referred to as *localization and orientation* [67].

In general, the various existing approaches for estimating the position and orientation can be distinguished by the applied underlying sensor technology and the corresponding field of application. The most basic types are: magnetic systems, mechanical systems, wireless systems based on radio frequency and optical systems using imagery data. Of course, hybrid solutions exist as well. An overview and discussion about the particular benefits and drawbacks of the individual approaches is given below.

Before examining the different approaches according to sensor type, a short review of the previous work done in [29] is presented, since this present thesis is building on top of it. In [29], *von Flottwell* initially defined the problem statement and proposed an image-based measurement approach which now is the research topic of this thesis.

### Pose Estimation from Accelerometer Data

In a first attempt, *von Flottwell* [29] investigates the possibilities to calculate the relative motions between two carbodies solely from accelerometer data. In a second attempt also the angular velocity measured by a gyroscope is taken into consideration. The chosen approach is obvious from the fact that the accelerometer and gyroscope data from within

---

[4] also: *6-DoF pose estimation*, *2D-3D pose estimation*, *camera pose estimation* or *space resectioning*

the carbodies is readily available since it is recorded in the course of riding comfort evaluations.

The underlying theoretical model and corresponding algorithm of a gyroscope-free inertial navigation system (INS) are described in [87]. For a feasible configuration of six accelerometer sensors it is proven that "*one can compute the linear and angular motions*".

Despite the extensive efforts undertaken, in the end the attempts in [29] do not succeed as a result of several severe problems. The most serious of them being that the double integration of acceleration data to estimate motion is very inaccurate due to the implicit integration drift inherent with sensor noise and bias. Other problems are missing acceleration data for specific axes (since the accelerometers were only installed for riding comfort evaluations) or the unknown cant of the track which affects the measurement of the angular velocity.

The approach's virtual lack of success is the reason why the development of an image-based measurement system is motivated.

### Pose Estimation from Electromagnetic systems

In electromagnetic systems, the observed object is tagged by one or more passive antennas. The systems are able to calculate the position and the orientation by analysing the relative magnetic flux between three orthogonal antennas, on both the transmitter and each receiver. In general, these systems are suitable for pose estimation or putting it in the words of [41]:

"*Electromagnetic sensing is the position-tracking technology behind [...] the products that support virtual reality and motion capture for computer animation.*"

An in depth description of such a system can be found in [70]. Although commercially available systems such as *MotionStar Wireless 2* reach a high degree of accuracy (15 mm RMSE and 1 ° RMSE at 3 m range as specified in [19]), the motion estimation systems, based on electromagnetism, suffer from great drawbacks. The main disadvantage is that the measuring accuracy is strongly affected by nearby metallic objects. The limited range also appears to be a major drawback. Since the main parts of railway vehicles are made up of steel, these kind of systems do not seem to be applicable for measuring motions inside railway vehicles.

## Pose Estimation from Mechanical systems

In the field of mechanical engineering the classical approach of pose estimation for the end-effector of a robot is to calculate the so-called *forward kinematics* from the specified values of the manipulators. For example the platform with six degrees of freedom (called Stewart/Gough-Platform) described in [84] was originally designed for flight simulators. The pose of the platform is controlled by and calculated from the given lengths of all six legs installed.

Another example is the hydraulic bogie test rig designed in [68]. A specialised six-cylinder layout is used to apply dynamic loads onto a bogie and thus simulating different riding conditions. Since a real-world instance of this test rig was used for experimental evaluations, a thorough description is given in Chapter 5.4.

The error estimations presented in [68] lead to the conclusion that the pose estimation of such a mechanical system can be calculated very accurately (±3 mm and ±2 °). A standard procedure for measuring the motions between carbodies using a mechanic system is the installation of a displacement-sensor system consisting of a set of string potentiometers and calculating its forward kinematics. The fundamentals of such a system are described in [4]. However, the compulsory needs of a power-supply, control-devices for synchronisation and data recording, as well as the extensive assembly work for sensor installation and validation have to be taken into consideration.

## Pose Estimation from Wireless systems

For the position estimation of a target node within a wireless sensor network, the signal parameters emitted by the network nodes are evaluated. The most common algorithms rely on one or more of the following methods: angle-of-arrival (AOA), time-of-arrival (TOA) and radio-signal-strength (RSS) [31]. Most of the approaches published solely focus on estimating the exact position of a target node. The question of the orientation of a node is often neglected. But there are methods as the one described in [67] which propose an enhanced "*localization and orientation scheme*".

Regarding the achieved accuracy (±220 mm and ±8 ° with 5 beacons at 0.024 nodes/m2) by using AOA, the presented wireless system of [67] does not seem to be sufficient regarding the necessary measuring accuracy (see requirement R(a)).

### Vision Based Pose Estimation

In a survey by *Lepetit and Fua* [53], the benefits of image-based pose estimation in contrast to the other technologies mentioned are described as that:

"*Vision has the potential to yield non-invasive, accurate and low-cost solutions to this problem, provided that one is willing to invest the effort required to develop sufficiently robust algorithms.*"

Originating from the field of photogrammetry, numerous approaches, methods and algorithms have been developed over the years. Therefore, pose estimation is arguably among the most extensively studied topics in computer vision. It is applied in a vast variety of tasks in different domains like robotics and visual servoing [90, 13], navigation [62], space flight [49], medical surgery [24] and augmented reality [1].

A closer examination of projects with a comparable type of problem definition, reveals that image-based systems are able to reach a very high measurement accuracy. The system described by *Kelsey et al.* [49] tries to determine the relative pose of spacecraft during proximity operations. The presented *Model-Based Pose Refinement Algorithm* was tested using a stereo camera setup. The achieved accuracy, at a range up to 1.5 m, was ±4.3 mm and ±0.86 °.

Another work, with a related task, is presented by *Manz et al.* in [57]. It aims to track the leading vehicle out of a moving vehicle within a convoy and determine its relative 6-DoF pose. The prior knowledge of the vehicle's geometric shape and appearance is fused with a dynamic vehicle model using particle filtering. The presented algorithm reaches an accuracy of approximately 500 mm RMSE and 1.74 ° RMSE deviation from the ground-truth (up to 35 m distance) measured by RTK-DGPS[5]. These results are quite remarkable considering the harsh environmental conditions (complex lighting conditions, partial occlusions) the algorithm has to deal with.

Commercially available systems like the *AICON3d - MoveInspect WheelWatch* claim to reach a measurement accuracy of up to ±0.1 mm and ±0.015 ° at 0.5 m distance from the object [32]. The system measures a car wheel's 6-DoF movement using a special carbon fibre wheel adapter as an optical target.

Concerning the present thesis, the major challenges that need to be addressed are the highly dynamic environment considering the variable lighting conditions and the fast motions during train ride. Also the limited spatial capacities (inside railway vehicle), the large displacements and the considerably long distance between the imaging device and the feature points are difficulties that need to be solved.

---

[5] Real time kinematic - Differential GPS

# 3. Rigid Body Pose Estimation

There exist manifold approaches how pose estimation is realised using image-based measurement. The choice mainly depends on the target application. A short discourse on the different possibilities and rationales for the final chosen approach is given below:

- **Single view vs. stereo view**
  Basically, as described in [7], the application of a stereo-based approach for pose estimation can be regarded as "*more robust and more accurate*" compared to a monocular-vision system. A major benefit of a stereo vision system is that the camera and scene geometry can be directly computed from image point correspondences only (see [37] Part II). For single view systems, this is not possible without additional information.

  Although several other pose estimation systems such as the already mentioned methods [49, 32] utilize a stereo-setup, this kind of approach is according to *Lepetit and Fua* [53] considered "*less popular*" because it suffers from a number of drawbacks. First, when applying a stereo-based approach for image-based measurement, an elaborate calibration is required. Second, the time-synchronisation between the multiple views is compulsory. If unsynchronised, the different views would not readily represent the same points in space at the same moments of time. Furthermore, the computational overhead produced by the process of robust feature matching has to be considered. Altogether, the application of a stereo-rig would inevitably lead to an increase in installation and evaluation efforts conflicting with the specified requirements (see R(f), R(e)).

  Since state of the art pose estimation methods based on single view geometry like [54, 52, 48, 76] apparently provide sufficient accuracy with respect to the given problem statement, the proposed measurement system within this thesis is restricted to monocular view.

- **Model-based vs. non-model-based**
  In the context of image-based measurement, it seems much more appropriate to apply a model-based pose estimation approach, which makes use of *a priori* information like object geometry, shape and texture. Non-model-based approaches such as optical flow [80] or structure from motion [22], basically do not maintain a geometrical model of the tracked object. A model-based approach is clearly favoured since within railway vehicles sufficient *a priori* information is available or can be explicitly installed prior to the measurements, leading to increased accuracy when compared to non-model based techniques.

- **Markerless vs. marker-based**
  Model based pose estimation is based on datasets of 2D-3D correspondences. The feature of an object in 3D is matched to its 2D correspondence in the acquired image. Since, within markerless pose estimation methods, the correspondence generation is realised automatically, using feature detectors like DoG/SIFT [56] or MSER [59], there also occur a certain amount of outliers (i.e. false matches). These outliers have to be removed from the correspondence datasets (e.g. by applying a RANSAC-Scheme [28]) because of their huge impact on the quality of measurements.

  According to *Lepetit and Fua* [53] the advantages of pose estimation based on special markers (also called fiducials) are that:
  "*[markers] constitute image features easy to extract, and they provide reliable, easy to exploit measurements for the pose estimation*" and furthermore conclude that:
  "*practical vision-based 3D tracking systems still rely on fiducials because this remains the only approach that is sufficiently fast, robust, and accurate.*"

  In order to address the outlier problem and to ensure that the correspondence generation from image features provides a maximum of measurement accuracy, it is clearly favoured to utilise a marker-based tracking method using designated fiducial markers.

Summarizing the above, the decision is derived that a **single-view, model-based pose estimation approach using fiducial markers** is the method of choice. In literature, there exist a variety of approaches, like [2, 48, 43] applying the same pose estimation scheme.

This results in the overall processing pipeline for pose estimation and system overview presented below.

## 3.1. System Overview

The vision-based system for measuring the relative motions between carbodies is composed of five main components (see Figure 3.1): First, the camera is rigidly attached to the carbody in the back (`CB_01`), the fiducial markers are mounted in suitable locations of the carbody in front (`CB_02`) and the arrangement of the railway vehicles has to be in an initial state. The internal parameters of the camera need to be known (calibration). Second, the train is put into operation and a continuous stream of images is acquired. Third, the mounted fiducials are tracked throughout the image stream, storing their corresponding image locations. Fourth, from the generated 2D-3D correspondences the pose of `CB_02` is estimated relative to the camera in `CB_01` on a per-frame basis. Fifth, this finally enables the system to calculate the sought-after relative motions from the continuous pose estimates.

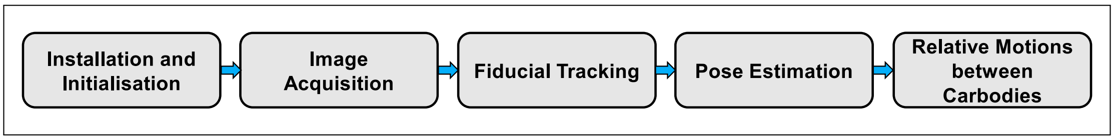


Figure 3.1.: Illustration of the five-step processing pipeline for measuring the relative motions between railway vehicle carbodies.

In such a pose estimation scheme, the selection of a type of fiducial marker which meets the specified requirements as well as the choice for a suitable tracking method are of utmost importance. The next section presents the different possibilities regarding the choice of fiducial markers as well as corresponding suitable tracking methods.

## 3.2. Marker Tracking

At the beginning of the development of a marker-based tracking system, two important questions arise: First, a suitable type of a marker that meets the requirements provided has to be chosen. Second, a robust tracking algorithm precisely adapted to the proposed type of marker, the environmental conditions, speed and accuracy has to be defined. The following sections provide an overview of the possibilities and explain the decisions reached within the process of marker selection and the finding of an appropriate tracking algorithm. Then a discussion is presented on the influence of perspective distortion on the measurement accuracy with respect to the chosen fiducial marker. Finally, in-depth implementation details are given.

### 3.2.1. Fiducial Markers

In contrast to natural landmarks, the primary purpose of *fiducial markers*[6] is to provide reliable, robust and accurate retrieval of known real-world positions within 2D image data. Since the fiducial markers are mounted within their environment in such a way that their exact 3D positions are known up to a certain accuracy, they provide a set of 2D-3D correspondences defined as $\{\mathbf{x}_i \leftrightarrow \mathbf{X}_i\}$.

In order to support robust and accurate retrieval from the image data, the chosen type of marker requires to provide a maximum of tolerance with respect to changing lighting conditions and partial occlusions. Furthermore, it should be taken into consideration that the occurring effects of perspective distortion due to the expected translations and rotations of the carbodies influence the measurement accuracy.

As opposed to many augmented reality (AR) applications, where the **detection** of markers in natural environments (e.g. *Claus and Fitzgibbon* [16], *Kato et al.* [48]) is a major issue, it is not taken into consideration for the given problem statement. It is sufficient to assume that the positions of the markers in the initial image of a sequence are previously known or labelled manually at the beginning of the measurement.

The feature of incorporating an individual tracking-code or payload (e.g. described in [48]) into the markers is not required within this work. Although this feature is important to AR-applications, to make the markers distinguishable, there is no risk of inter-marker confusion, since the identification of the fiducials is simply achieved through the *a-priori* known spatial relationships.

Furthermore, the following discussion mainly focuses on the application of so-called 2D *passive markers* which do not actively emit light and are, in general, distinguished by

---

[6] also called *fiducials* or *artificial landmarks*

their geometric shape. The advantage of passive markers is that they can be easily manufactured using standard consumer printers which is in good accordance with requirements R(e) and R(f). In contrast, *active markers* mostly make use of coloured or IR LEDs and require a much more elaborate manufacturing including electronic circuitry and batteries. In first place, this additional effort does not seem to be necessary. Section 5.7 includes a brief discussion about possible enhancements to the tracking system.

In recent works containing comparative studies [53, 26, 65, 72, 75], a variety of fiducial markers is presented. Basically, the markers can be classified into two main categories based on their main geometric appearance: square and circular fiducials. Figure 3.2 presents an overview of popular fiducial markers.

(a) AR Toolkit (b) ARTag (c) Intersense (d) Fourier Tag (e) Rune Tag

Figure 3.2.: Examples of different fiducial markers which primarily differ in their geometric appearance. Square: (a) AR Toolkit [48], (b) ARTag [26]; Circular: (c) InterSense [60], (d) Fourier Tag [91]; Special: (e) Rune Tag [8] consists of a circular set of sizeable dots. Notice the different approaches of assigning an individual code or payload to each marker. This common feature is not required within this work.

Since all the fiducial markers introduced in Figure 3.2 are designed as black and white patterns (including Fourier Tag (3.2d) assuming 1-bit payload encoded as either black or white), they are represented with a maximum contrast within the image data. This design contributes to robust segmentation results even under varying lighting conditions. The use of colour information, as for example proposed by *Cho and Neumann* in [15], is disregarded since it would exclude the utilisation of a monochrome imaging device from the outset.

An advantage of squared fiducials is that one can retrieve four feature points from solely one single marker, which (as discussed later in Section 3.3.5) is sufficient to compute the exact pose. The general pose estimation approach for square markers as proposed by [48, 26], is to first determine the image location of the marker. This is done by extracting the edge boundaries and exploiting the fact that under perspective projection a square yields a quadrilateral. Finally, the four vertices or corners are determined from the intersection of the edges and thereby the pose of the square target can be computed.

When applying squared fiducials in image-based measurement systems, it is important to note an effect, denoted by *Fiala* [27] as so-called *vertex jitter*, which is described as "*the noise in the marker corner positions*". In his thesis, *Brandner* [12] explains this issue by the fact that "*some or all of the [corner] detectors introduce a systematic bias to the estimation result*" moreover concluding that "*it is not possible to determine the true corner location [...]*".

Compared to this drawback, various sources recommend the application of cicular fiducials when focusing on the accuracy of measurements. Again, *Brandner* [12] states that "*circular blob features are widely used in computer vision for applications that require robust and accurate determination of point positions in 3D space*". In their survey, *Lepetit and Fua* [53] also endorse the application of circular feature points mentioning that "*most of the professional solutions use circular or spherical fiducials [...]*". Furthermore, for example, *Heikkilä* [39], *Li et al.* [55] and *Abad et al.* [3] propose camera calibration approaches based on circular fiducials.

Circular fiducials are preferred because their basic features remain invariant in the case of translations and the roll-rotation. Their centroid can be recovered with sub-pixel accuracy by assuming that the centroid of the projected ellipse coincides with the true centre of the circular target. This assumption is proven wrong in the case of general perspective projection. In [12], an in-depth description of this effect is given, basically stating that "*blob features introduce a systematic bias when being mapped under perspective distortion*", nevertheless confirming that there exist possibilities to correct this effect. However, the discussion in Section 3.2.4 reveals that for a small relative target size and a limited range of rotation angles, this bias becomes negligibly small.

When aiming at the design of a highly accurate pose estimation system, it is clearly essential to obtain bias free data and best precision from 2D-3D correspondence point recovery. This is why the utilisation of circular fiducials is clearly favoured over squared markers. Summarizing the findings from above, it is proposed to apply a **2D, passive, high-contrast, circular marker** for straightforward production and easy handling as well as enabling a high degree of accuracy.

The 2D circular fiducials presented in Figure 3.2, *InterSense* and *Fourier Tag*, integrate coding schemes for unique inter-marker recognition which is not essential for this work. Therefore it is recommended to employ a circular marker introduced by *Gatrell et al.* [30] called *Concentric Contrasting Circle (CCC)*. An example of a CCC is given in Figure 3.3.

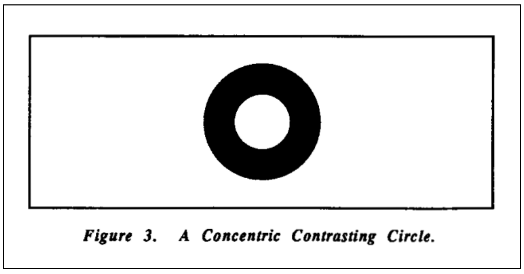


Figure 3.3.: Example of a Concentric Contrasting Circle (CCC) taken from the original work by *Gatrell et al.* [30]. A CCC is formed by a black ring on white background. The ratio of the radii between the outer and the inner circle is defined with $\frac{r_{outer}}{r_{inner}} = \frac{2}{1}$

The application of a CCC fiducal marker brings the following advantages:

- **Robustness and accuracy**
  The general appearance of a CCC promises a stable and highly discriminable representation of the marker against the background. Furthermore, the application of two coinciding circular features provides a high degree of accuracy. Moreover, for CCCs there exist methods to recover the true centroid under general perspective distortion. Different suitable methods for centroid retrieval are presented in [51], [39] and [3].
- **Easy production and handling**
  The CCCs can easily be fabricated in large numbers using standard vector graphic tools and consumer printers. Varying the size is nearly effortless.
- **Robust against partial occlusions**
  As shown in [50], if necessary the centres of a CCC can be retrieved even under partial occlusion up to 50%.
- **Detection strategies exist**
  Although not directly demanded, there exist robust detection strategies for CCCs as demonstrated in [46]. This would facilitate an auto-setup procedure without manual initialisation of the tracking system.

A drawback in the application of CCCs still remains: One CCC fiducial can only provide one 2D-3D correspondence. It is essentially necessary to mount multiple markers in the scene. How this is implemented in the best possible way is discussed in the next section.

### 3.2.2. Target Design

Multiple markers are mounted in the scene in order to provide a sufficiently large set of 2D-3D correspondences. Feasible solutions are presented, for example in the works by *Claus and Fitzgibbon* [16] and *Hoff et al.* [44]. They recommend to arrange multiple markers on a planar surface to form one target. The method of acquiring multiple 2D-3D correspondences from one planar target with multiple (square or circular) fiducials, is also commonly used for camera calibration (see [93, 88, 10]).

This approach has the advantage of greatly simplifying the installation and setup process. As already discussed for a model-based pose estimation system, it is crucial that the positions of the fiducials in the world reference are known prior to the measurements. Instead of mounting single CCC markers in the scene and determining the exact world-positions one by one, it is far more accurate and much faster to apply several markers onto one target. Since the alignment of the markers on the target is predefined, only the position of the multi-fiducial target has to be determined during the installation process.

Applicable designs of multi-fiducial targets are illustrated in Figure 3.4. The design consists either of five or seven CCCs arranged in a regular pattern. The 5-CCC target is similar to the design proposed in [16]. However, instead of using four dots arranged in a square with a unique code in the centre, it is composed of five CCCs which resembles the common die face. In addition, the 7-CCC target was designed to make full use of the amount of space offered by standardised ISO A-series paper size (aspect ratio of $1 : \sqrt{2}$) and to provide extra markers at approximately the same inter-marker distance.

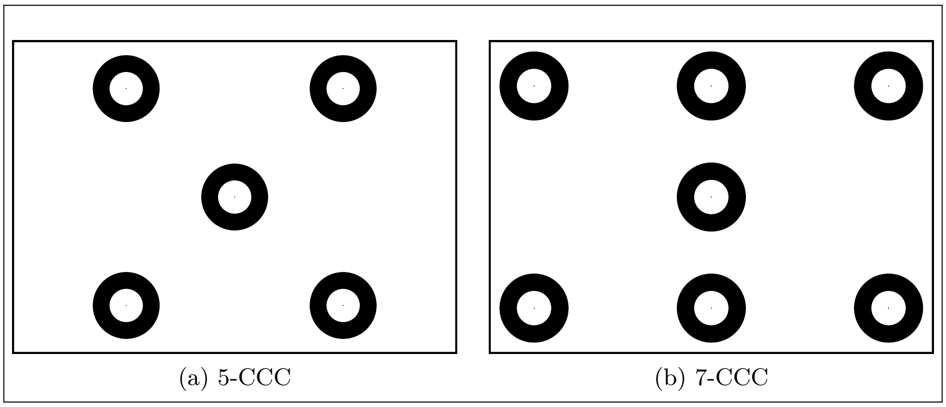

(a) 5-CCC
(b) 7-CCC

Figure 3.4.: Illustration of two designs of multi-fiducial targets. (a) 5-CCC: uses five CCCs arranged in an equally distributed pattern (b) 7-CCC: taking advantage of the full amount of space provided by ISO A-series paper sizes by expanding and adding two extra markers.

### 3.2.3. Tracking Method

In general, 2D tracking aims at determining the exact position of a moving object within continuous image data (video sequences). In their survey, *Yilmaz et al.* [92] list various approaches for tracking rigid and non-rigid objects in 2D images. The approaches are distinguished by the general representation of an object's position in the image, which for example relies on: points, primitive geometric shapes, silhouette and contour, templates, skeletal models or active appearance models.

When searching for most accurate 2D-3D correspondences, one is always interested in a point representation of the tracked object, in this case the centroid of the CCC fiducial marker. In the context of camera calibration, *Heikkilä* [38] proposes two methods for centroid estimation: ellipse-fitting and centre of gravity (CoG) calculation. He finds that, generally iterative ellipse-fitting procedures reach a higher degree of accuracy than CoG estimation. This is truly important in the sense of camera calibration, but the difference in terms of subpixel accuracy only impacts on the second decimal place [38, p. 101]. This rather small gain in measurement accuracy does not account for the much more elaborate and computationally intensive ellipse fitting algorithms.

In their work about accurate fiducial registration, *Bose and Amir* [9] denote the CoG calculation approach as *centroid method*. They argue that the benefits of the centroid method are its simple implementation, the subpixel accuracy provided and the fast execution time. According to *van Assen et al.* [5], in an input image $I$, the CoG $\hat{\mathbf{c}}$ over a search window $S$ is calculated by:

$$\hat{\mathbf{c}} = \left( \frac{\sum\limits_{x,y\in S} xI(x,y)}{\sum\limits_{x,y\in S} I(x,y)}, \frac{\sum\limits_{x,y\in S} yI(x,y)}{\sum\limits_{x,y\in S} I(x,y)} \right) \tag{3.1}$$

with $I(x,y)$ denoting the grey-level at the corresponding image pixel.

This centroid calculation approach is identical to the one described by *Cheng* [14] in his well known revision of the *mean shift algorithm* (using a unit kernel), from which numerous robust real-time tracking algorithms like [17, 11] were derived. For example, the *Continuously Adaptive Mean Shift* (CAMSHIFT) procedure introduced by *Bradski* [11] was originally designed to track the movements of faces. The implications of these considerations justify the application of the CoG as a robust and accurate estimate for the centroid of circular fiducials.

The opportunity to apply the proven CoG calculation also reveals another advantage of a marker based tracking system: In combination with the limited displacements of the carbody motions (specified in Section 1.2), the marker based approach greatly supports the accuracy and robustness as well as accelerates the task of the tracking algorithm. This is substantiated by the fact that, if mounted in a suitable way, the specifically chosen markers are not expected to greatly change in size and appearance during the test ride, which facilitates continuous CoG calculation. Recursive procedures and update strategies, as proposed by other tracking algorithms (e.g. CAMSHIFT), are assumed to be unnecessary.

These reduced demands on a suitable tracking solution allow the application of a low-level but fast and accurate approach. On the basis of the tracking methods described for the InterSense marker [60] and the CCC fiducial [30], the following tracking method is proposed:

Given an image sequence $\{I_t \mid t = 0, \dots, N-1\}$, repeat the following procedure for every element of the set of fiducial points $\{\mathbf{X}_i \mid i = 0, \dots, K-1\}$ in the scene (implementation details are described in Section 3.2.5):

1. **Initialisation**
   In image $I_0$, assign a rectangular search window $S_{0,i}$ centred at a seed point $\mathbf{s}_{0,i}$, for fiducial $\mathbf{X}_i$ which is extracted from image $I_0$. The only condition being that the full CCC is located within the search window.

2. **Local preprocessing**
   As a step prior to further processing, remove the noise within the search window $S_{t,i}$ by applying a filter operation. Due to its capabilities to remove certain types of random noise, a median filter [81, p. 129] is chosen to remove video compression artefacts.

3. **Image binarization**
   In order to get an elementary representation of the structures, segment the region of search window $S_{t,i}$ into a foreground and a background. Thus an adaptive threshold $T_{t,i}$ is chosen to determine if, depending on the greyscale intensity value, a pixel $g(\mathbf{x})$, $\mathbf{x} \in S$ belongs to the foreground or to the background, resulting in a binarised window ($S_{t,i} \in [0, 1]$). Further enhancements are achieved by applying morphological operations.

4. **Identify components and determine their CoG**
   Within the binarised search window identify connected components, in both fore- and background, by using a four- or eight-connected neighbourhood relation. Then separately for every connected component $C_j$ determine its CoG ($\hat{\mathbf{c}}_j(x, y)$ with $x, y \in C_j$) by applying Equation 3.1.

5. **Determine CCC candidate**
   Appoint potential CCC candidates by evaluating the following criteria: size, area ratio and the coincidence of centroids of two contrasting components (i.e. one black and one white component are located within a predefined distance). If two connected components fulfil all the mentioned criteria they are appointed as CCC candidate.

6. **Set new seed point and iterate**
   For image $I_t$, the salient centroid $\hat{\mathbf{c}}_{t,i}$ equals $\mathbf{x}_{t,i}$, which establishes the sought-after 2D-3D correspondence $\{\mathbf{x}_{t,i} \leftrightarrow \mathbf{X}_i\}$ between the world and the image point of the fiducial. Iterate the tracking algorithm by setting the centroid as new origin $\mathbf{s}_{t+1,i} = \mathbf{x}_{t,i}$ of the search window $S_{t+1,i}$. Repeat steps 2 to 6 until $t = N$.

For a more detailed description of the single steps of the tracking algorithm see Section 3.2.5.

In the best case, the tracking algorithm results in a list $L$ of size $(N \times K)$ which, for every frame in the sequence contains the corresponding image location of the fiducials $\{\mathbf{x}_{t,i} \leftrightarrow \mathbf{X}_i\}$. This list of correspondences provides the fundamental basis for all further pose estimation calculations. An example of intermediate results during one iteration of the tracking procedure is illustrated in Figure 3.5.

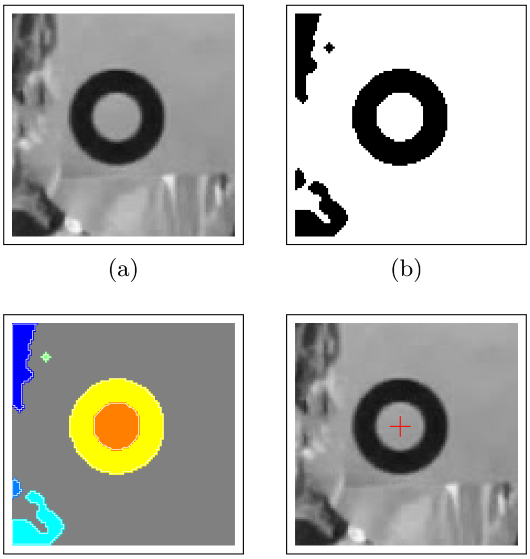
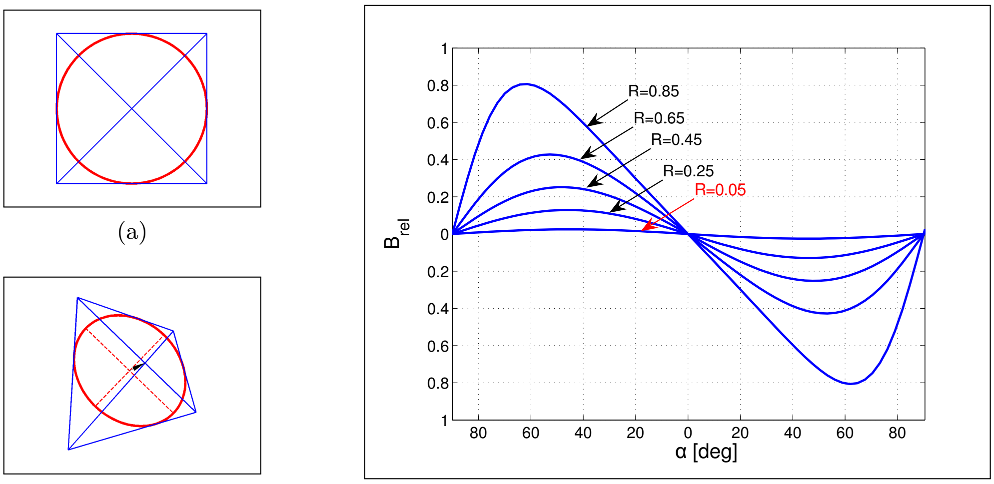

(a) (b)
(c) (d)

Figure 3.5.: Example of the iterative steps of the tracking algorithm. (a) input search window (b) binarized search window gained by applying P-tile thresholding and morphological operations (note the background clutter) (c) different connected components identified by neighbourhood relations (d) final result of CoG calculation on CCC candidate.

### 3.2.4. Influence of Perspective Distortion

As already mentioned, a systematic bias is introduced when estimating the centroid of a circular fiducial from its (elliptical) perspective projection in a 2D image. In his thesis, *Brandner* [12] provides an in-depth investigation of this effect, which he states depends on the geometric setup of the sensor (i.e. the angle between the principal axis of the camera and the normal vector of the target plane) and the size of the projected fiducial in the image. The illustrations and the graph depicted in Figure 3.6 are entirely adapted from [12]. Figures 3.6a and 3.6b illustrate the deviation when estimating the CoG of a circular blob under perspective projection. The bias in position between the estimated and the true centre is clearly visible.

In [12], the bias is normalised by relating it to the target size, in order to investigate the distortion effect at varying viewing angles ($\alpha \in [-90\,°, 90\,°]$) as well as for different target sizes. This relative bias is defined as $B_{rel} = B(\mathbf{c})/R_{\alpha=0}$ with $B(\mathbf{c}) = \mathbf{c} - \hat{\mathbf{c}}$ being the deviation between the estimation of the blob centre $\hat{\mathbf{c}}$ and the true centre location $\mathbf{c}$. The term $R_{\alpha=0}$ denotes the radius of the target as seen from position $\alpha = 0$ when the sensor of the camera and the target are in parallel. Furthermore, the radius $R$ is normalised with respect to the image-width. For example a normalised radius $R = 0.1$ extends to 10% of the image width.



(a)

(b)

(c)

Figure 3.6.: (a) At viewing angle $\alpha = 0$ the result of the CoG calculation and the intersections of the enclosing square coincide. (b) Under perspective projection a bias in centroid estimation is introduced when using CoG calculation. (c) The relative bias is illustrated at different viewing angles for different normalised sizes of the target blob. It shows that the deviation strongly increases with size and viewing angle. Illustrations and graph taken from [12].

The relation between the relative bias $B_{rel}$ and the viewing angle $\alpha$, for different relative sizes of the target, is depicted in Figure 3.6c. The graph shows that the deviation strongly increases with the viewing angle and the target size up to a certain maximum. It is reasonable to assume that, within the image acquisition system designed for the measurement of carbody motions (see Chapter 4), the single CCC-fiducials will extend to no more than 5% ($R = 0.05$) of the image width or height. In Figure 3.6c the function of a target with relative size equal to 5% is marked with a red arrow. One can see that the relative bias for a target of that size is very small. The maximum of $|B_{rel}| = 0.025$ is reached at an angle of $\pm 45\,^\circ$. This relative bias becomes even smaller when considering that the maximum relative rotational displacements of a railway vehicle carbody are specified within $\pm 15\,^\circ$ (see Section 1.2). This assumption holds if the setup between camera and targets is chosen adequately.

The conclusion of this discussion is that the influence of the bias introduced by CoG calculation due to perspective distortion can be neglected regarding the specific task this image-based measurement system is designed for. This decision is based on the investigations by *Brandner* [12], which clearly show that for a small relative target size and the limited range of rotation angles, the bias becomes negligibly small.

### 3.2.5. Implementation Details

The initialisation of the tracking algorithm (step 1) is accomplished by manually assigning a seed point to each fiducial in the scene. The size of the search window $S_i$ depends on the geometric resolution of the marker in the image and the expected physical motions (these may differ from test ride to test ride) in combination with the sampling frequency. As the geometric resolution of all markers within the same multi-fiducial target can be considered as approximately equal, it is sufficient that for each target a constant window size is defined. Appropriate side lengths $m \times n$ for the search windows are determined by empirical evaluations. Generally, it is suitable to assume that the side lengths lie in the range of two to three times the diameter of the fiducial in the image.

The image preprocessing in step 2 is applied in order to normalise the image data for further processing. The major distortion which needs to be handled, arises from artefacts introduced by video compression. As stated in [21] conventional lossy video compression algorithms like the MPEG-4 coding standard introduce so-called *blocking* and *mosquito* artefacts. An example is depicted in Figure 3.7a.

As can be seen from [21] and the example given in Figure 3.7b, a median filter is an appropriate choice to reduce the influence of compression artefacts. The median filtering replaces the intensity value of a pixel by the median of its ordered neighbourhood and is performed by default using a $3 \times 3$ neighbourhood.

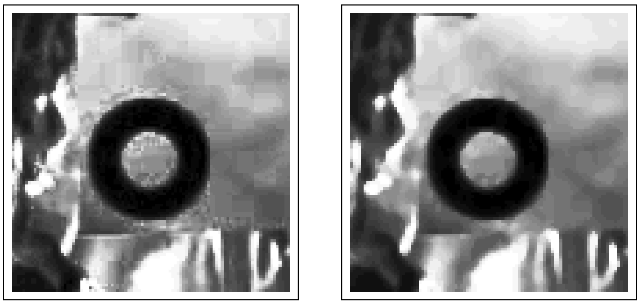

(a) (b)

Figure 3.7.: Reducing artefacts introduced by video compression using a median filter. For better visibility, histogram equalisation is applied to emphasise the effects. (a) Blocking artefacts are visible in uniform regions. Strong mosquito artefacts occur at the boundaries of the black circle (b) Artefacts are removed by applying a median filter operation.

For step 3 of the tracking algorithm, the threshold $T_{t,i}$ is denoted with index variables, since the expected variable and non-uniform lighting conditions during a test ride render the usage of a fixed threshold as obsolete. The value of $T_{t,i}$ has to be continually determined. A suitable method for adaptively determining a threshold is called *P-tile thresholding*[7] [81, p. 179]. It makes use of the prior knowledge about the expected ratio between foreground and background (i.e. black circle on white background) in the search windows $S$. In a normalised histogram, the probability $p$ of a greyscale intensity $g$ is the number of corresponding pixels $n_g$ as a fraction of the total number of pixels in the search window: $p(g) = n_g/n$.

Based on the cumulative histogram $c(g)$ it is possible to choose a threshold $T$ such that approximately $p$ percent of the search window area have an intensity value less than $T$. Using the definition of the cumulative histogram

$$c(g) = \sum_{i=0}^{g} p(g), \tag{3.2}$$

the P-tile threshold corresponds to the intensity value $g = T$ with $c(T) \simeq p/100$, where the cumulative sum of normalised pixel intensities is closest to the demanded percentile $p$.

---

[7] P-tile as short for percentile.

In order to obtain a further simplified and less detailed representation of the structure within the binarised search window, a morphological opening operation is applied. As stated in [34, p. 662] the opening operation (defined as erosion followed by dilation; $S \circ B = (A \ominus B) \oplus B$ ) smoothes contours and eliminates small islands and sharp peaks. A disk-like structuring element $B$ is applied to preserve the circular appearance of the CCC fiducial.

In step 5, from all the connected components found in step 4, the CCC candidate is determined by evaluating the following sequence of criteria:

1. **Centroid location**
   The coincidence of the centroids from two contrasting connected components is evaluated first, since it promises to provide the highest entropy. This feature is common to CCC fiducials but unlikely in other structures in the scene. Since the CoG calculation results in subpixel accuracy, the choice for maximum distance of $d \leq 2$ with $d = \|\hat{\mathbf{c}}_j(x,y) - \hat{\mathbf{c}}_k(x,y)\|$ and $j \neq k$ seems appropriate.

2. **Size**
   The relative size of the outer ring of the CCC, compared to the size of search window, is likely to correspond to the percentage of foreground $p$ as used for the computation of the adaptive threshold.

3. **Area ratio**
   In accordance with [30], the ratio of the radius of the outer ring to the radius of the inner ring is defined as $2 \times r_{inner} = r_{outer}$. Thus, the ratio between the areas is tested for: $4 \times A_{inner} = A_{outer}$.

When the appropriate CCC candidate is selected, recalculate the CoG based on the outer ring with the encircled area filled up to remove possible uncertainties arising from the segmentation process. The result is a robust and accurate estimation of the salient centroid $\hat{\mathbf{c}}_{t,i}$.

Generally, due to the constrained movements of a single marker regarding its neighbours on the same target, a reliable strategy for handling occlusions, unexpected fast motions or a marker sliding outside the field of view can be implemented: Simply reconstruct the most-likely image position of a lost marker from the knowledge about positions of other markers on the target. If, within the reconstructed search window, a suitable CCC candidate can be appointed again, reinitialise the common tracking procedure.

## 3.3. Pose Estimation Algorithms

As already mentioned, the term pose estimation relates to specifying the position and the orientation of an observed object with respect to a given coordinate system. In their review of different pose estimation methods *Haralick et al.* [36] state that the foundations of this problem were first addressed and solved in the field of photogrammetry by *Grunert* in 1841. Its introduction into the domain of computer vision dates back to the 1970s and originates from the task of digital camera calibration [71] as a fundamental requirement to most image-based measurement systems (see Section 4.2 for an in-depth discussion).

Generally, in the field of computer vision the coined term *2D-3D pose estimation* refers to the procedure where an imaging device is capturing 2D measurement data and the corresponding world or object model data is given in 3D. Putting it more precisely, 2D-3D pose estimation is the process of estimating the relative position and orientation of a 3D object with respect to a reference camera coordinate system. An overview of this basic problem statement is given in Figure 3.8.

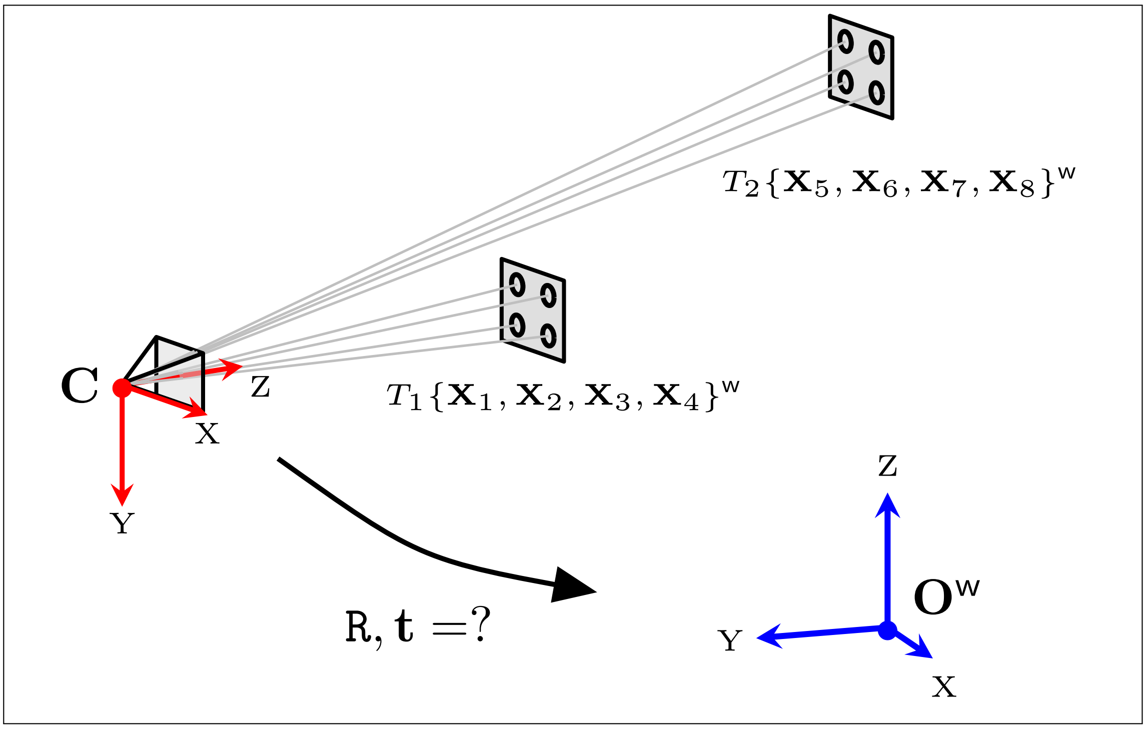


Figure 3.8.: Basic pose estimation problem statement. Targets $T_1$ and $T_2$ provide a set of 3D points $\{\mathbf{X}_1 \dots \mathbf{X}_8\}$ in world coordinates. The pose of the camera corresponds to a translational ($\mathbf{t}$) and rotational ($\mathtt{R}$) transformation which aligns the world and the camera reference frames. It is estimated from the 2D projections of the 3D points on the image plane. $\mathbf{O}^{\mathsf{w}}$ and $\mathbf{C}$ denote the origins of the respective coordinate systems.

The algorithms developed for 2D-3D pose estimation are generally based on perspective or projective camera models. The mathematical representations of camera models provide an essential theoretical background and are therefore presented in the next section.

### 3.3.1. Camera Representation

This section introduces required mathematical representations of camera models according to *Hartely and Zisserman* [37, p. 153]. First the basic *pinhole camera* as a simplistic but essential model as well as associated terms and components are defined. Then a specialisation called *general finite projective camera* and its accompanying modifications are presented.

#### Pinhole Camera Model

The pinhole camera model uses the concept of *central projection* to transform 3D points onto a 2D plane. The *camera centre* (also *optical centre*) $\mathbf{C}$, is the centre of projection and is defined as the origin of a Cartesian coordinate system. The *image plane* is represented as $Z = f$, with $Z$ called the *principal axis* which is orthogonal to the image plane and intersecting at the origin. The intersection of the principal axis with the image plane is called *principal point*. The *principal plane* is the plane through the camera centre and parallel to the image plane. The pinhole camera model and its components are illustrated in Figure 3.9:

The transformation from a 3D point with coordinates $\mathbf{X} = (X, Y, Z)^\mathsf{T}$ onto the 2D image plane is described by a central projective mapping, where the straight line between $\mathbf{X}$ and the centre of projection (denoted with $\mathbf{C}$ in Figure 3.9) meets the image plane ($\mathbb{R}^3 \mapsto \mathbb{R}^2$):

$$(X, Y, Z)^\mathsf{T} \mapsto (fX/Z, fY/Z)^\mathsf{T} \tag{3.3}$$

Expressed in terms of matrix multiplication using homogeneous coordinates, this linear transform is written as:

$$\begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \mapsto \begin{pmatrix} fX \\ fY \\ Z \end{pmatrix} = \begin{bmatrix} f & & & 0 \\ & f & & 0 \\ & & 1 & 0 \end{bmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \tag{3.4}$$

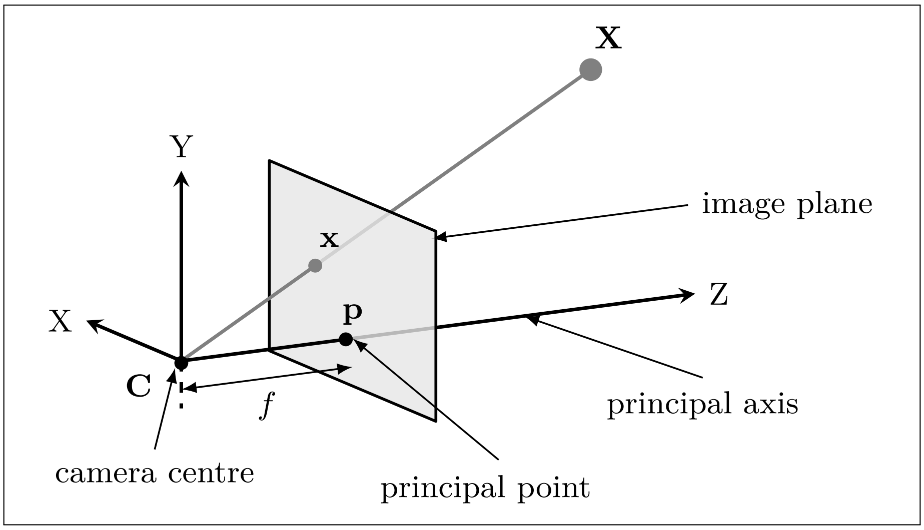


Figure 3.9.: Basic pinhole camera model describes the central projection of a world point $\mathbf{X}$ onto the image plane as image point $\mathbf{x}$. The distance from the camera centre to the image plane is denoted with $f$ the *focal length*.

The $3 \times 4$ homogeneous matrix in this equation is called *camera projection matrix* further on denoted with $\mathtt{P}$. For the 4-vector world point $\mathbf{X}$ and its 3-vector projection on the image plane $\mathbf{x}$ Equation 3.4 can be compactly written as:

$$\mathbf{x} = \mathtt{P}\mathbf{X} \tag{3.5}$$

where $\mathtt{P} = \mathtt{diag}(f, f, 1)\ [\mathtt{I} \mid \mathbf{0}]$ with $\mathtt{I}$ being the identity matrix. Generally, the origin of coordinate system of the image plane does not coincide with the principal point. Thus the following extension called *principal point offset* is defined:

$$\begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \mapsto \begin{pmatrix} fX + Zp_x \\ fY + Zp_y \\ Z \end{pmatrix} = \begin{bmatrix} f & & p_x & 0 \\ & f & p_y & 0 \\ & & 1 & 0 \end{bmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \tag{3.6}$$

where $(p_x, p_y)^\mathsf{T}$ are the coordinates of the principal point expressed in the image coordinate system. In Equation 3.6 the $3 \times 4$ homogeneous matrix is called *camera calibration matrix* (further on denoted with $\mathtt{K}$), which allows the expression in a compact form as:

$$\mathbf{x} = \mathtt{K}\ [\mathtt{I} \mid \mathbf{0}]\mathbf{X}^\mathsf{c} \tag{3.7}$$

The superscript $\mathsf{c}$ in $\mathbf{X}^\mathsf{c}$ emphasises the fact that the 3D world point $\mathbf{X}$ is expressed in the *camera coordinate frame* which is congruent with the Cartesian coordinate system defined in the beginning. In practice, the 3D world points are expressed in a different coordinate system, commonly known as *world coordinate frame*.

As depicted in Figure 3.10 these two coordinate frames are related by a translation $\mathbf{t}$ and a rotation $\mathtt{R}$.

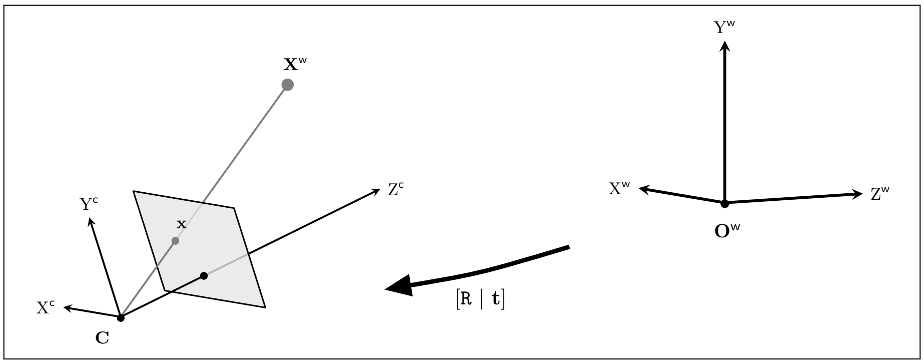


Figure 3.10.: A camera coordinate frame and a world coordinate frame are aligned by a translational and rotational transformation denoted with $\mathtt{R}, \mathbf{t}$.

If the camera centre in the world coordinate frame is expressed with $\mathbf{C}^{\mathrm{w}}$ and $\mathtt{R}$ represents the orientation of the camera coordinate frame, then the mapping of a point $\mathbf{X}$ from the world to the camera coordinate frame is expressed as:

$$\mathbf{X}^{\mathrm{c}} = \begin{bmatrix} \mathtt{R} & -\mathtt{R}\mathbf{C}^{\mathrm{w}} \\ 0 & 1 \end{bmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} = \begin{bmatrix} \mathtt{R} & -\mathtt{R}\mathbf{C}^{\mathrm{w}} \\ 0 & 1 \end{bmatrix} \mathbf{X} \tag{3.8}$$

inserting this into Equation 3.7 yields:

$$\mathbf{x} = \mathtt{KR}\,[\mathtt{I} \mid -\mathbf{C}^{\mathrm{w}}]\mathbf{X} \tag{3.9}$$

Equation 3.9 represents the general mapping as provided by the basic pinhole camera model. The parameters within $\mathtt{K}$ are called *intrinsic camera parameters* whereas $\mathtt{R}$ and $\mathbf{C}^{\mathrm{w}}$, which relate the camera position and orientation to a given world coordinate frame, are termed as *extrinsic camera parameters*. Finally the camera projection matrix $\mathtt{P}$ can be decomposed as:

$$\mathtt{P} = \mathtt{K}\,[\mathtt{R} \mid \mathbf{t}] \tag{3.10}$$

with $\mathbf{t} = -\mathtt{R}\mathbf{C}^{\mathrm{w}}$, in order to omit the explicit representation of the camera centre. The matrix $[\mathtt{R} \mid \mathbf{t}]$ is often referred to as *camera pose* or *exterior orientation*.

### General Projective Camera Model

In order to account for digital camera sensors, the *finite projective camera model* describes the image coordinates as number of pixels per unit distance $m_x$ and $m_y$. In the (unlikely) case of non-square pixels the focal length is defined separately for horizontal and vertical pixel dimensions with $\alpha_x = fm_x$ and $\alpha_y = fm_y$. The principal point is specified respectively as: $x_0 = p_x m_x$ and $y_0 = p_y m_y$. Additionally, a *skew* parameter $s$ is introduced considering non-perpendicular horizontal and vertical image axes. The skew is usually assumed to be 0 for real cameras. With these modifications the following camera calibration matrix $\mathtt{K}$ is considered:

$$\mathtt{K} = \begin{bmatrix} \alpha_x & s & x_0 \\ & \alpha_y & y_0 \\ & & 1 \end{bmatrix} \tag{3.11}$$

A finite projective camera with $\mathtt{P} = \mathtt{K}[\mathtt{R} \mid \mathbf{t}]$ and $\mathtt{K}$ of form 3.11, has eleven degrees of freedom: five for $\mathtt{K}$ with $(\alpha_x, \alpha_y, x_0, y_0, \text{s})$ and six for the rigid transformations $[\mathtt{R} \mid \mathbf{t}]$.

Representing the finite projective camera by an arbitrary $3 \times 4$ homogeneous matrix of rank 3 (i.e. remove restrictions resulting from the finite projective camera model), yields the so-called *general projective camera* with:

$$\mathbf{x} = \mathtt{P}\mathbf{X} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{bmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \tag{3.12}$$

This representation is commonly used as a starting point to estimate $p_{11} \ldots p_{34}$ from point correspondences (see Section 3.3.3).

At this point, attentive readers may have noticed the relation between railway vehicle car body motions and image-based pose estimation: Like the rigid transformation $[\mathtt{R} \mid \mathbf{t}]$, the pose between two railway vehicle carbodies is determined by translational and rotational displacements. Under the assumption that a suitable basic measurement setup is chosen (see Section 1.4), the rigid transformation $[\mathtt{R} \mid \mathbf{t}]$ between the world and the camera coordinate frame has the essential capabilities to describe the pose between two adjacent carbodies.

Different approaches on how pose estimation is calculated are presented in the next section.

### 3.3.2. Pose Estimation from Point Correspondences

So far, the measurement system designed in the previous sections enables the pose estimation of a rigid railway vehicle carbody from a set of 2D-3D point correspondences. For the sake of completeness, it should be mentioned that pose estimation can be accomplished by using higher order entities as well (e.g. lines). Such methods are reviewed by *Rosenhahn* in [74] who at the outset states that, pose estimation is generally calculated from point correspondences and that "*many fundamental and important publications can be found for this class*".

According to *Lepetit et al.* [54], under the assumption that the intrinsic camera parameters are known, existing methods for pose estimation from point correspondences can be classified into two types of approaches: iterative and non-iterative ones. On the one hand, iterative approaches as presented by *DeMenthon and Davis* [23] or *Schweighofer and Pinz* [76] are stated to be accurate, but slow and error-prone in case of poor initialisation. On the other hand, non-iterative methods are considered much faster at the cost of a slight reduction in accuracy.

In the course of this thesis, two non-iterative approaches are examined more closely:
First, the **Direct Linear Transformation (DLT)** described by *Hartley and Zisserman* [37] as a well established and popular method for pose estimation is elected, introduced and evaluated.
Second, an algorithm called **Efficient Perspective-n-Point Camera Pose Estimation (EPnP)** presented by *Lepetit et al.* [54] as a representative of PnP-type methods (see Section 3.3.4) is chosen and examined the same way.

These two methods are selected because the DLT is considered as a straightforward and proven method and EPnP is regarded as a state-of-the-art approach. Both methods promise to be robust and sufficiently fast. Furthermore, extensively tested implementations are available in MATLAB[8] for prototyping as well as in high-level languages such as C++, regarding a later application in a productive environment.

An introduction to both DLT and EPnP and their algorithmics is given in the next sections.

---

[8] http://www.mathworks.com

### 3.3.3. DLT - Direct Linear Transformation

According to [88] the DLT method originates from a work first published as early as in 1971. Based on the set of 2D-3D point correspondences $\{\mathbf{x}_i \leftrightarrow \mathbf{X}_i\}$ for one image, retrieved by the tracking algorithm described in Section 3.2.3, the DLT algorithm requires to find the camera matrix $\mathtt{P}$, as specified in Equation 3.12, such that $\mathbf{x}_i = \mathtt{P}\mathbf{X}_i$ for all $i$. The DLT method applied to pose estimation is described in [37, p. 178] the following way:

For each correspondence $\mathbf{x_i} \leftrightarrow \mathbf{X_i}$, Equation 3.12 is expressed using the vector cross product as:

$$\mathbf{x}_i = \mathtt{P}\mathbf{X}_i \quad \Leftrightarrow \quad \mathbf{x}_i \times \mathtt{P}\mathbf{X}_i = 0 \tag{3.13}$$

Then the following notation for $\mathtt{P}$ is introduced:

$$\mathtt{P} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{bmatrix} \overset{!}{=} \begin{bmatrix} \mathbf{P}^{1\mathsf{T}} \\ \mathbf{P}^{2\mathsf{T}} \\ \mathbf{P}^{3\mathsf{T}} \end{bmatrix}$$

with that, rewriting Equation 3.13 using $\mathbf{x}_i = (x_i, y_i, w_i)^\mathsf{T}$ gives:

$$\mathbf{x}_i \times \mathtt{P}\mathbf{X}_i = \begin{pmatrix} y_i\mathbf{P}^{3\mathsf{T}}\mathbf{X}_i - w_i\mathbf{P}^{2\mathsf{T}}\mathbf{X}_i \\ w_i\mathbf{P}^{1\mathsf{T}}\mathbf{X}_i - x_i\mathbf{P}^{3\mathsf{T}}\mathbf{X}_i \\ x_i\mathbf{P}^{2\mathsf{T}}\mathbf{X}_i - y_i\mathbf{P}^{1\mathsf{T}}\mathbf{X}_i \end{pmatrix} = \begin{bmatrix} \mathbf{0}^\mathsf{T} & -w_i\mathbf{X}_i^\mathsf{T} & y_i\mathbf{X}_i^\mathsf{T} \\ w_i\mathbf{X}_i^\mathsf{T} & \mathbf{0}^\mathsf{T} & -x_i\mathbf{X}_i^\mathsf{T} \\ -y_i\mathbf{X}_i^\mathsf{T} & x_i\mathbf{X}_i^\mathsf{T} & \mathbf{0}^\mathsf{T} \end{bmatrix} \begin{pmatrix} \mathbf{P}^1 \\ \mathbf{P}^2 \\ \mathbf{P}^3 \end{pmatrix} = \mathbf{0} \tag{3.14}$$

Since from these three equations only two are linear independent [86], the third row is omitted and the system results in:

$$\underbrace{\begin{bmatrix} \mathbf{0}^\mathsf{T} & -w_i\mathbf{X}_i^\mathsf{T} & y_i\mathbf{X}_i^\mathsf{T} \\ w_i\mathbf{X}_i^\mathsf{T} & \mathbf{0}^\mathsf{T} & -x_i\mathbf{X}_i^\mathsf{T} \end{bmatrix}}_{\mathtt{A}_i} \begin{pmatrix} \mathbf{P}^1 \\ \mathbf{P}^2 \\ \mathbf{P}^3 \end{pmatrix} = \mathbf{0} \tag{3.15}$$

where each $\mathtt{A}_i$ is a $2 \times 12$ matrix. For a set of $n$ image to world point correspondences, $2n$ equations are stacked, which accumulates to a linear equation system $\mathtt{A}\mathbf{p} = 0$ where $\mathtt{A}$ is of size $2n \times 12$. The camera matrix $\mathtt{P}$ is then calculated by solving this set of equations.

Since the camera matrix $\mathtt{P}$ has 11 degrees of freedom, a minimum of $5\frac{1}{2}$ point correspondences is required to obtain an exact solution for the camera matrix $\mathtt{P}$. In the overdetermined case with an arbitrary number of correspondences ($n \geq 6$), $\mathtt{P}$ is estimated by minimizing the algebraic error of $||\mathtt{A}\mathbf{p}||$ subject to the constraint $||\mathbf{p}|| = 1$ using *Singular*

*Value Decomposition* (SVD) [37, p. 585]. The final result is the last column of $\mathtt{V}$ in the SVD: $\mathtt{UDV}^{\mathsf{T}} = \mathtt{A}$.

From the obtained camera matrix $\mathtt{P}$, the intrinsic camera parameters $\mathtt{K}$ and the exterior orientation $[\mathtt{R} \mid \mathbf{t}]$ are computed by decomposing $\mathtt{P}$ using *RQ-decomposition*. For an in-depth description of SVD and RQ-decomposition see [37, p. 578].

### Remarks

In their survey, *Lepetit and Fua* [53] suggest to separately estimate the intrinsic camera parameters. They state that, by using a calibrated camera and estimating only the position and orientation $[\mathtt{R} \mid \mathbf{t}]$ it is possible to obtain far more reliable results.

When applying the DLT algorithm, the normalisation of the underlying data is crucial regarding the achieved measurement accuracy. Putting it in the words of *Hartley and Zisserman* [37]: "*[Data normalisation] must not be considered optional*". It is recommended that the points $\mathbf{x}_i$ are normalized by translating them so that their overall RMS distance from the origin of the coordinate system is $\sqrt{2}$.

A drawback of the DLT algorithm is that, in contrast to other algorithms, it minimizes an algebraic error which, in terms of the estimated geometric relationships, is not meaningful. But as stated in [37] this drawback can be overcome by an good choice of normalisation.

The advantages of the DLT algorithm compared to other pose estimation methods are that it is fairly simple to calculate, that there are no initial parameter approximations required and since it represents a direct solution, no problems with convergence will occur.

### 3.3.4. EPnP - Efficient Perspective-n-Point Camera Pose Estimation

As its name implies the EPnP algorithm [54] evolves from the methods which solve *Perspective-n-Point* (PnP) type of problems. This term was coined by *Fischler and Bolles* [28] in 1981. Basically, under the precondition that the intrinsic camera parameters are known, they formulate the PnP-problem as:

"*Given the spatial locations of n control points (3D world points), and given the angle to every pair of control points from an additional point called the Center of Perspective (C), find the lengths of the line segments joining C to each of the control points.*"

The easiest way to describe the PnP-problem is by introducing its most basic case called P3P. The P3P-problem is applied to the smallest subset of points ($n = 3$) which results in a finite number of solutions. All other PnP-problems comprise the P3P-problem as a special case. Acccording to *Quan and Lan* [69] the P3P-problem is computed as follows: Given the intrinsic parameters of a camera and the camera centre $\mathbf{C}$ as well as three 2D-3D point correspondences, each pair of correspondences $\{\mathbf{x}_i \leftrightarrow \mathbf{X}_i\}$ and $\{\mathbf{x}_j \leftrightarrow \mathbf{X}_j\}$ gives a constraint on the unknown distance between camera centre and world points: $c_i = ||\mathbf{CX}_i||$ and $c_j = ||\mathbf{CX}_j||$. Using this triangular configuration, the following equation is derived using the law of cosines:

$$d_{i,j} = c_i^2 + c_j^2 - 2c_ic_jcos(\alpha_{i,j}) \tag{3.16}$$

with the angle $\alpha_{i,j}$ spanned by the straight lines $\overline{\mathbf{CX}_i}$ and $\overline{\mathbf{CX}_j}$ where $\alpha_{i,j}$ can be measured by a calibrated camera and the distance $d_{i,j} = ||\mathbf{X}_i\mathbf{X}_j||$ is determined by the known spatial relationship between the points. For three point correspondences with three unknown distances, this results in a fourth degree polynomial equation system which has a maximum of four solutions. To remove this ambiguity, an additional correspondence is added and the equation system is solved by taking subsets of three of the four points and calculating the common solution.

As stated in [69] and [54], this approach is common practice in computer vision (e.g. [28, 52]) but results in the drawback that several fourth-degree polynomials need to be solved. Furthermore, the agreement on a common solution is probably difficult due to noisy data since no use of redundant information is made. P4P approaches as for example proposed by *Quan and Lan* [69], are able to obtain a unique solution. But the computed redundant polynomial equation systems suffer from an increase in complexity, since the expansion of the equations produces additional parameters.

In order to address these problems, the central idea of the efficient $O(n)$ implementation of the EPnP algorithm is to express the coordinates of the given $n$ 3D points as a weighted sum of four virtual control points. All further calculations are then performed only on these control points which for a larger number of $n$ yields a much smaller number of unknown variables which reduces complexity and therefore accelerates further computations.

The general parametrisation within the EPnP algorithm looks as follows: In order to distinguish more clearly, points expressed in the image, camera and world coordinate system are denoted with a $\{\mathsf{i}, \mathsf{c}, \mathsf{w}\}$ superscript respectively. First 4 control points $\mathbf{C}_j^{\mathsf{w}}$ with $j = 1 \dots 4$ are defined. Then each 3D world point $\mathbf{X}_i^{\mathsf{w}}$ is expressed by the weighted sum of the control points. This yields:

$$\mathbf{X}_i^{\mathsf{w}} = \sum_{j=1}^{4} \theta_{i,j} \mathbf{C}_j^{\mathsf{w}}, \qquad \text{with} \quad \sum_{j=1}^{4} \theta_{i,j} = 1 \tag{3.17}$$

The same relation is assigned to the camera coordinate system:

$$\mathbf{X}_i^{\mathsf{c}} = \sum_{j=1}^{4} \theta_{i,j} \mathbf{C}_j^{\mathsf{c}} \tag{3.18}$$

Assuming the camera calibration matrix $\mathtt{K}$ to be known and $\mathbf{x}_i^{\mathsf{i}} = (x_i^{\mathsf{i}}, y_i^{\mathsf{i}}, 1)^{\mathsf{T}}$ with $i = 1 \dots n$ being the 2D projections of the world points $\mathbf{X}_i^{\mathsf{c}}$ as well as expanding $\mathbf{C}_j^{\mathsf{c}} = (x_j^{\mathsf{c}}, y_j^{\mathsf{c}}, z_j^{\mathsf{c}})^{\mathsf{T}}$ for each control point, gives:

$$\forall i: \quad w_i \begin{pmatrix} x_i^{\mathsf{i}} \\ y_i^{\mathsf{i}} \\ 1 \end{pmatrix} = \mathtt{K}\mathbf{X}_i^{\mathsf{c}} = \mathtt{K} \sum_{j=1}^{4} \theta_{i,j} \mathbf{C}_j^{\mathsf{c}} = \begin{bmatrix} \alpha_x & 0 & x_0 \\ & \alpha_y & y_0 \\ & & 1 \end{bmatrix} \sum_{j=1}^{4} \theta_{i,j} \begin{pmatrix} x_j^{\mathsf{c}} \\ y_j^{\mathsf{c}} \\ z_j^{\mathsf{c}} \end{pmatrix} \tag{3.19}$$

with $w_i$ representing scalar projective parameters.

Furthermore, from the last row of Equation 3.19 it can be seen that $w_i = \sum_{j=1}^{4} \theta_{i,j} z_j^{\mathsf{c}}$. Thus, for each reference point, two linear equations are obtained from substituting this expression into the remaining other two rows:

$$\sum_{j=1}^{4} \theta_{i,j} \alpha_x x_j^{\mathsf{c}} + \theta_{i,j}(x_0 - x_i^{\mathsf{i}}) z_j^{\mathsf{c}} = 0 \tag{3.20a}$$

$$\sum_{j=1}^{4} \theta_{i,j} \alpha_y y_j^{\mathsf{c}} + \theta_{i,j}(y_0 - y_i^{\mathsf{i}}) z_j^{\mathsf{c}} = 0 \tag{3.20b}$$

When, for all $n$ point correspondences, the obtained Equations 3.20a and 3.20b are stacked, this results in a linear equation system of the form:

$$\mathbf{M}\mathbf{x} = \mathbf{0} \tag{3.21}$$

which is solved to receive the final solution $\mathbf{x} = (\mathbf{C}_1^{\mathsf{c}\mathsf{T}}, \mathbf{C}_2^{\mathsf{c}\mathsf{T}}, \mathbf{C}_3^{\mathsf{C}\mathsf{T}}, \mathbf{C}_4^{\mathsf{c}\mathsf{T}})$.

### Remarks

As stated by its authors [54], in contrast to the DLT algorithm the normalisation of the image data is not necessary within the EPnP algorithm, since Equations 3.20a and 3.20b do not depend on data from the image referential system.

In their work, *Lepetit et al.* [54] conduct extensive experiments and prove the superiority of the EPnP algorithm over other state-of-the-art iterative and non-iterative methods. They show that their approach is only slightly less accurate but much faster (calculations are performed in constant time $O(n)$) and robust than the other methods. Moreover, a Gauss-Newton optimisation procedure is proposed to further increase the accuracy, while still maintaining the $O(n)$ performance.

The efficient $O(n)$ implementation is indeed a great advantage of the EPnP algorithm over other methods, but will not play such a significant role, since it is only affected by a large number of $n$ point correspondences, which is not the case for the present measurement system designed (see Section 3.3.5).

### 3.3.5. Necessary Number of Correspondences

As discussed before, for the general DLT algorithm a number of $n = \lceil 5^1\!/\!_2 \rceil = 6$ point correspondences is necessary to determine a unique solution. For a planar target as proposed in Section 3.2.2, the DLT algorithm is applied to estimate the so-called *2D homography*. Given a number of 2D-2D point correspondences $\{\mathbf{x}_i \leftrightarrow \mathbf{x}_i'\}$, the 2D homography basically determines the transformation between two planar surfaces (i.e. the planar target and the image plane). Due to the constrained problem formulation of the homography estimation, only $n \geq 4$ point correspondences are required to determine a unique solution [37].

According to [54], the EPnP algorithm is applicable for all planar and non-planar configurations with $n \geq 4$ point correspondences.

These findings are in very good accordance with the target design proposed in Section 3.2.2. Even in the case of a single 5-CCC multi-fiducial target, a solution for the pose can be estimated from an overdetermined equation system by the DLT and the EPnP algorithm.

The discussion within this chapter clearly shows that the DLT and the EPnP algorithms are perfectly suitable for the given task of pose estimation between two carbodies. This is the reason why both algorithms are tested and evaluated, presenting in-depth details in the Experimental Evaluation Section 5. The only precondition for the application of the EPnP algorithm is that it requires the intrinsic camera parameters ($\mathtt{K}$) to be estimated beforehand.

## 3.4. Measuring Relative Motions between Railway Vehicle Carbodies

The previous chapters introduced all the necessary preconditions, tools and algorithms which basically enable the estimation of the pose between two railway vehicle carbodies. This chapter rounds up this process by giving a detailed description of the underlying carbody motion model as well as an in-depth explanation on the measurement setup and the vital step of correct initialisation.

### 3.4.1. Railway Vehicle Carbody Motion Model

In the designed measurement system, the motions between two adjacent railway vehicle carbodies are derived from the pose (i.e. the translational and rotational displacements) on a per-frame basis. Since the frames are recorded continually at a predefined rate, the pose estimates from the whole sequence of images provide a series of motions over time. This section introduces the underlying carbody motion model and the proposed measurement setup.

An example of a coupled oscillation between two carbodies is used to define the model for relative carbody motions. It is presented in Figure 3.11 with detailed close-up views depicted in Figure 3.12.

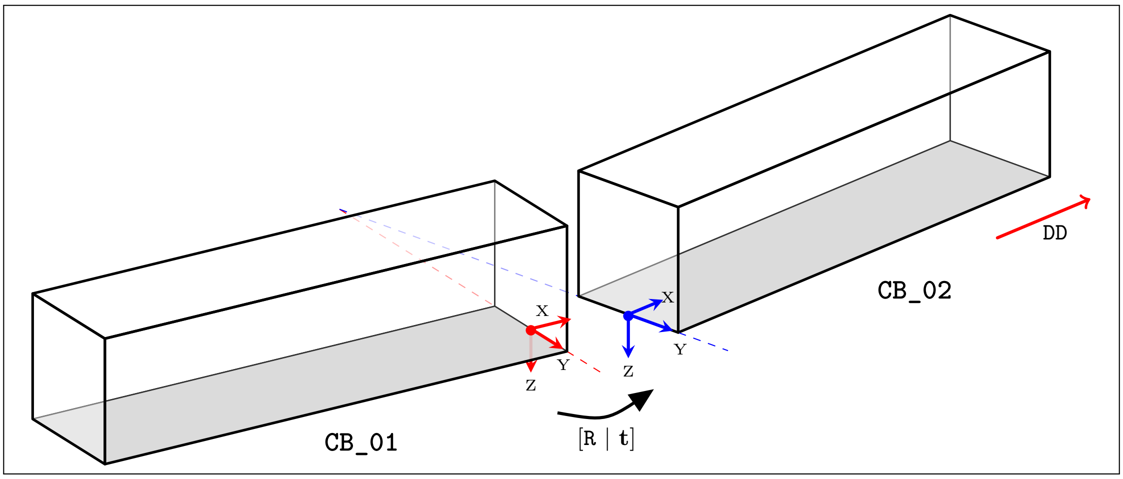
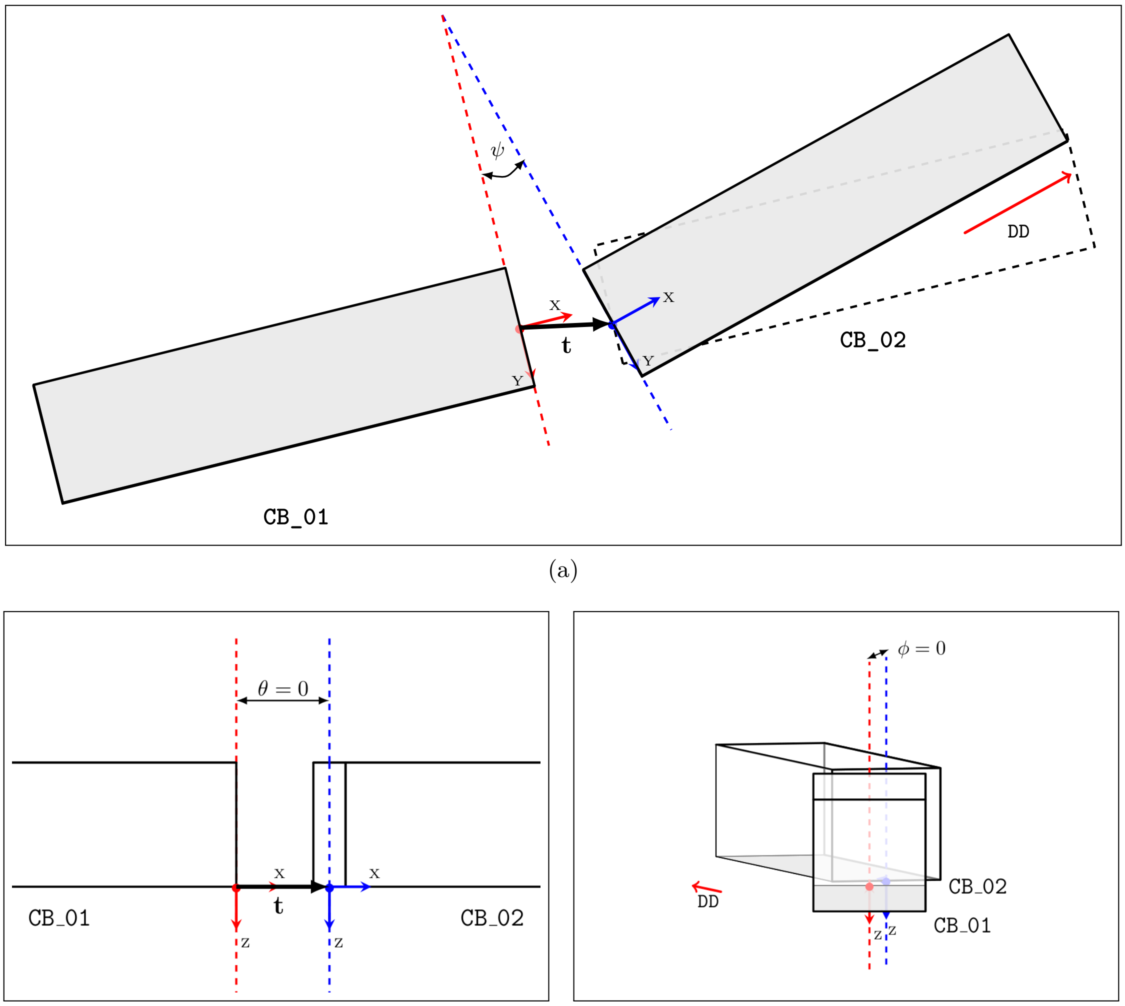


Figure 3.11.: The model for relative carbody motions comprises two adjacent carbodies (denoted with `CB_01` and `CB_02`). In principle, the measurement system seeks to estimate the rotation and translation [R | **t**] between the two reference frames (red and blue) of the carbodies. In this illustration `CB_02` is rotated and translated compared to `CB_01`. The driving direction is denoted with `DD`.

The coupled oscillation exemplified in Figure 3.11 is called *hunting* which is composed of a translation along the y-axis and a rotation about the z-axis (yaw-angle). From the origins of the reference frames marked, it is apparent that the relative motions are measured **between two directly adjacent carbody endings**.



(a)

(b)

(c)

Figure 3.12.: The carbody motion model in detail: (a) illustrates a plan view with the yaw-angle $\psi$ and the translation vector **t** drawn explicitly. The dashed rectangle denotes a referential (initial) position. `CB_02` is rotated by $\psi = 15\,^\circ$ about the z-axis (yaw-rotation) and translated ⅙ of the carbody width into the y-direction. (b) depicts a side view. The angle $\theta = 0$ denotes that no pitch rotation is observed. (c) shows a view from behind, along the driving direction. Likewise, no roll angle is observed ($\phi = 0$).

Figure 3.12a shows the hunting movement as a simultaneous lateral and angular displacement. For better visibility, an initial carbody pose is indicated by the dashed black rectangle. Although zero in this case, from Figures 3.12b and 3.12c the definitions of the roll ($\phi$) and pitch ($\theta$) angles are visible.

### 3.4.2. Measurement Setup and Installation

A setup which directly measures the pose between the two carbodies would imply to track targets within `CB_01` and `CB_02` at the same time, which is not suitable as discussed in Section 1.4. For the designed optical measurement system, it is meaningful to indirectly measure the displacements of `CB_01` and `CB_02` by relating the camera reference frame and the coordinate system of `CB_01`. The instance that the imaging device is rigidly attached to `CB_01`, enables to define this relation by a constant translational and rotational offset. The approach of estimating the pose between the camera and an external reference frame is in good accordance with the basic pose estimation problem statement as illustrated in Figure 3.8.

The proposed measurement setup is presented in Figure 3.13. It illustrates the transformation of the origin of the coordinate system `CB_01` to the camera centre. The orientation of the coordinate system is considerably changed and it is not necessarily axially symmetric aligned with the original reference frame. Furthermore, Figure 3.13 shows that the pose between `CB_01` and `CB_02` is estimated between the rigidly attached imaging device and an arbitrary number of targets (in this case two) installed in `CB_02`.

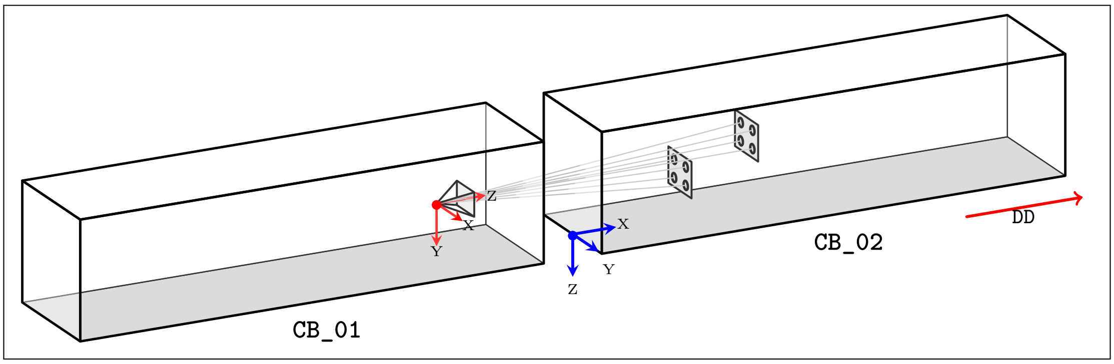


Figure 3.13.: Proposed measurement setup with the two carbodies in an initial pose assumed at time $t = 0$. The pose is estimated between the rigidly attached camera in `CB_01` and the targets installed in `CB_02`. Note that the orientation of the original coordinate system of `CB_01` and the camera are not necessarily axially symmetric aligned. This constant angular misalignment is compensated by calculation later on.

For the pose estimation, the exact knowledge of the 3D coordinates of the targets, given in the reference frame of `CB_02` is essential. Therefore it is recommended (but not compulsory) to mount the targets in a way that they are automatically axially symmetric aligned with the reference frame of `CB_02`. This can be easily accomplished by using suitable existing structures and furnishings within the carbody. This approach facilitates a short installation time since it is only necessary to determine the translational displacement of a target to the origin of the carbody reference frame. The elaborate determination of the angular deviations of each planar target can therefore be avoided.

### 3.4.3. Initialisation and Relative Pose Estimation

Figure 3.13 also illustrates another important specific aspect: The arrangement of the railway vehicles is in an **initial pose** which is essential to all further measurements. This initial pose in which the carbodies and their respective reference frames are exactly physically aligned, is a demanded precondition of the measurement system. This requirement is necessary because without it, the measurements would suffer from an unknown bias. From now on, it is considered that the first frame in the sequence of images, $I_0$, depicts an arrangement in which the concerned carbodies are in an initial pose.

As described in the introduction (see Section 1), the key objective of the designed system is to measure the **relative** motions between the carbodies. This means that from the initial pose only the relative change of displacements between two carbodies is determined (i.e. only the deviation ($\Delta$) between the designated initial pose and further poses in the sequence of images is calculated).

When applying the proposed measurement setup, the parameters $[\mathtt{R} \mid \mathbf{t}]$ retrieved from the pose estimation algorithms described in Section 3.3, express the rotational and translational displacements in terms of the camera reference frame. As the camera reference frame is most likely not axially symmetric aligned, it is necessary to compensate for this deviation since the sought-after motions need to be expressed in terms of the carbody reference frame.

As defined before, the first frame in the sequence of images yields the initial pose with $[\mathtt{R} \mid \mathbf{t}]_{I_0} = [\mathtt{R_0} \mid \mathbf{t_0}]$. It describes the translational and rotational offset of the camera from the reference frame in `CB_02`. This initial pose is used in order to compensate for the arbitrary displacement of the imaging device and to calculate the relative pose between $[\mathtt{R_0} \mid \mathbf{t_0}]$ and the pose in image $I_t$ with $[\mathtt{R_t} \mid \mathbf{t_t}]$.

The calculation of the relative poses is accomplished as follows:

$$\Delta \mathtt{R}_t = \mathtt{R}_0^{-1} \mathtt{R}_t = \mathtt{R}_0^{\mathsf{T}} \mathtt{R}_t \tag{3.22a}$$

$$\Delta \mathbf{t}_t = \mathtt{R}_0^{-1} (\mathbf{t}_t - \mathbf{t}_0) = \mathtt{R}_0^{\mathsf{T}} (\mathbf{t}_t - \mathbf{t}_0) \tag{3.22b}$$

with $\mathtt{R}^{-1} = \mathtt{R}^{\mathsf{T}}$ because $\mathtt{R}$ is a rotation matrix. If in Equation 3.22a $\mathtt{R}_t$ is corrected by the initial deviation, all further displacements directly yield relative rotations between `CB_01` and `CB_02`. In Equation 3.22b, the relative translations $\Delta \mathbf{t}_t$ are calculated by first correcting the rotational camera offset and then subtracting the initial translational displacement $\mathbf{t}_0$.

Both equations provide the relative pose of the two carbodies for all images in the recorded sequence and the defined initial pose. Of course it applies for the initial pose that $\Delta[\mathtt{R} \mid \mathbf{t}]_0 = [\mathtt{I} \mid \mathbf{0}]$.

To determine the final relative angular values $\Delta\{\phi, \theta, \psi\}_t$, the rotation matrix $\Delta\mathtt{R}_t$ is decomposed, as described in [66, p. 70], by applying an inverse of roll, pitch and yaw transform by first defining the following notation for $\Delta\mathtt{R}_t$:

$$\Delta\mathtt{R}_t = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \tag{3.23}$$

which then is decomposed into the relative angular values using:

$$\Delta\phi_t = atan2(r_{21}, r_{11}) \tag{3.24a}$$

$$\Delta\theta_t = atan2(-r_{31}, cos(\phi)r_{11} + sin(\phi)r_{21}) \tag{3.24b}$$

$$\Delta\psi_t = atan2(sin(\phi)r_{13} - cos(\phi)r_{23}, -sin(\phi)r_{12} + cos(\phi)r_{22}) \tag{3.24c}$$

This proposed measurement setup and its initialisation offers two important advantages. First, it is possible to choose an arbitrary position (although of course basic constraints apply) for the imaging device without knowledge of its exact position and orientation. These values are determined implicitly from the initial carbody pose. This simplifies the setup procedure and enables to quickly adapt the system to possibly changing environmental conditions, as well as maintaining the required portability. Second, the multi-fiducial planar targets need to be mounted in one carbody only, which also facilitates reduced installation efforts.

# 4. Optical Image Acquisition System

Most generally, an optical imaging system for image-based measurement is composed of a camera (basically consisting of housing, sensor and lens), mounting (e.g. tripod), internal or external storage device, power supply and devices providing illumination. This chapter describes the several requirements and constraints which are taken into account regarding the given specific task and motivates the choice of a suitable imaging system.

## 4.1. Requirements

For the prototypical realisation of the whole measurement setup described in this document, the requirements differ in a way that there is a significant shift towards higher portability which in a later stage of development does not apply. This higher portability regarding the imaging system ensures faster responses of the measurement setup to unexpected challenges and changes in the environment. Certainly, for an implementation in a productive environment, a system meeting altered and more mature requirements needs to be realised.

### 4.1.1. Camera, Image Sensor and Lens

In order to realise a highly portable imaging system it is considered beneficial that the camera can operate autonomously and independent of provided infrastructure such as electric power, external memory or additional control units like a PC or laptop.

Another important question which arises, is the choice for an appropriate digital image sensor with a suitable technical specification. As defined in Section 1.3, the carbody motions are required to be sampled at a minimum sample frequency of 25 Hz = 25 frames per second (fps). Thus, the utilised image sensor has to provide at least 25 fps (R(a)). In combination with a sensor capable of sensing images in a progressive scan mode (i.e. acquiring a full frame at a time), these requirements suffice to avoid motion-induced blur in the recorded images.

Additionally, a high sensitivity of the image sensor is considered important, since it allows the recording of regularly illuminated images and at the same time using a small aperture setting. The small aperture setting in turn increases the *depth of field*, which is the ability to keep objects in focus at various distances. Of course, a high quality, wide-aperture lens and a good illumination contribute to this effect.

Furthermore, an increase in the accuracy of measurements is expected as the geometrical resolution of the recorded images increases. Therefore, a high spatial resolution of the sensor is considered beneficial. The required spatial resolution of the image sensor is determined by the knowledge about the field of view (FOV), the working distance (i.e. distance between camera and target) and the size of the smallest feature which needs to be detected by the system.

The working distance is assumed up to a maximum of $w = 10000$ mm (half the length of a carbody). The FOV is approximated by using the maximum angular displacement of the carbodies into yaw direction with: $2 \times max(\Delta\psi) = 2 \times 15^\circ = 30^\circ$. Using the working distance and the common aspect ratio of 3 : 2 this approximately yields a FOV of size $5400 \times 3600$ mm$^2$. Which, with the ignored lateral displacements and a necessary tolerance is rounded up to FOV = $7000 \times 4600$ mm$^2$.

From the target design introduced in Section 3.2.2, it is clear that the smallest feature which needs to be detected is represented by the inner circle of the CCC. It has, if printed on ISO-A3 sized paper, a diameter of about $d^{\mathrm{w}} = 30$ mm. According to [53], for accurate centre of gravity (COG) calculation the minimum size of a target is in the range of $d^{\mathrm{i}} = \lceil 4 - 5 \rceil = 5$ pixel. Using these considerations, the required sensor resolution (SR) in both dimensions can be derived as follows:

$$\mathrm{SR} = d^{\mathrm{i}} \times \left(\frac{\mathrm{FOV}}{d^{\mathrm{w}}}\right) = 5 \times \left(\frac{\{7000, 4600\}}{30}\right) = \{1166, 766\} \simeq 1200 \times 800 \text{ pixel}^2 \qquad (4.1)$$

This estimation indicates that the image recording needs to be done at least equivalent to the standardised HD720 format [25] with a spatial resolution of $1280 \times 720$ pixel$^2$.

From a specific sensor model and its format (SF), it is possible to calculate a suitable focal length in order to select an appropriate optical lens for the camera system. Assuming an APS-C[9] sized sensor with (SF = $23.5 \times 15.6$ mm$^2$) the necessary focal length is calculated as follows:

---

[9] http://www.dpreview.com/learn/?/key=sensor%20sizes

$$f = \frac{\mathrm{SF} \times \mathrm{w}}{\mathrm{FOV}} = \frac{\{23.5, 15.6\} \times 10000}{\{7000, 4600\}} \simeq 33 \text{ mm} \tag{4.2}$$

Since the FOV and the image sensor share the same aspect ratio, $f$ is approximately the same in both dimensions.

As a result of the discussion above it was decided to utilise a high-end digital consumer camera for image acquisition during various experiments and test rides (see Chapter 5). In particular a SONY NEX-5N[10] was chosen, as it fulfils all of the mentioned requirements.

The NEX-5N is a mirrorless interchangeable lens camera with an APS-C sized sensor and capable of recording in HD1080p/50. This means it has the capability of capturing frames with $1920 \times 1080$ pixel$^2$ at 50 fps using progressive scan. The image data is processed using the *Advanced Video Coding High Definition* (AVCHD) [82] standard. The supplied zoom lens is the SEL1855 with $f = 18 - 55$ mm. Because of the NEX-5N's small body and integrated battery and memory card, it is able to operate entirely autonomous.

A disadvantage of NEX-5N is that, according to the manual [83] only sequences up to maximum of 29 minutes can be recorded because of automatic termination to prevent overheating of the sensor. For the prototypical realisation of this measurement setup this issue is ignored, since experiments and test rides are expected to last shorter than that. Furthermore, it is considered a drawback that the NEX-5N compresses the imagery data according to the AVCHD standard. This introduces artificial compression artefacts which make special pre-processing necessary as discussed in Section 3.2.5.

In a productive environment it is certainly preferable to employ an industrial (i.e. more professional, higher quality and better ruggedised) imaging device. A probably good choice is the DMK 23G445 from *The Imaging Source*[11]. It provides suitable technical specifications (monochrome images with $1280 \times 960$ pixel$^2$ at 30 fps using progressive scan) and records raw imagery data but in contrast is dependent on external power supply and storage.

In the case of the DMK 23G445, which implements the *GigE Vision* interface standard [6], the power supply and connection to an external storage and control unit is achieved by using a standard Ethernet cable (up to 100 m).

---

[10] http://www.sony.co.uk/hub/nex-compact-system-camera/range/nex-5n

[11] http://www.theimagingsource.com/en_US/products/cameras/gige-cmos-ccd-mono/dmk23g445

The so-called *Power over Ethernet* (PoE) technology provides both power supply and data connection using a single cable. It therefore allows manageable and low cost integration into a productive environment. Figure 4.1 shows both camera systems introduced.


(a) (b)

Figure 4.1.: (a) SONY NEX-5N with the SEL1855 zoom-lens. (b) The Imaging Source - DMK 23G445 with unknown lens. The depicted sizes are not proportional.

### 4.1.2. Illumination

Suitable illumination is the one of the most important aspects in a machine vision system. Usually railway vehicles, in particular passenger coaches, are equipped with artificial ambient lighting inside the carbodies. Thus during a test ride, even at night basic visibility of the mounted targets is guaranteed.

In the paper of *Heikkilä* [39] about camera calibration, one can see from the discussion on systematic errors that, even small changes in illumination may severely affect performance of vision algorithms. So an essential intent of artificial illumination is to keep the ambient lighting conditions as constant as possible and to reduce effects on the targets like shadows and glares. As stated in [61] common types of light sources include LEDs or halogen dampers.

For the first prototypical evaluation it is not considered necessary to install additional illumination devices during test rides, because the existing lighting within the carbodies is regarded as sufficient. This decision is also founded on the fact that even in a productive measurement system, it does not seem applicable to install artificial light sources due to the expected large working distances of up to 10 m and the associated extra efforts and costs.

In the case that proper illumination cannot be sufficiently achieved by the existing lighting, it is recommended to employ active targets rather than extra light sources. A prototype of a low-cost and minimal effort active CCC target with integrated green LED powered by a 9-volt battery is pictured in Figure 4.2a. The corresponding circuit diagram is shown in Figure 4.2b.

(a)

(b)

Figure 4.2.: (a) Active Target: CCC on A4-sized paper with green LED manufactured as proof of concept. Consists of: plate, printed CCC target, green LED, resistor, battery and cabling. Estimated costs of parts: $\leq 4$ Euro. (b) Circuit diagram of the active target. $R_1 = 600\Omega$, $U_1 = 9V$, $D_1 =$ NICHIA NSPG520AS[12].

From the discussions on the requirements of the optical imaging system, it is concluded that for a first evaluation the application of a consumer camera (specifically SONY NEX 5N) is highly recommended. This is primarily due to its capabilities of operating autonomously. An other finding is that in a first step artificial lighting is not considered necessary. For the operation of the measurement system in a productive environment, it is further recommended to apply an industrial imaging device and (if necessary) to make use of active targets.

---

[12] http://www.alldatasheet.com/datasheet-pdf/pdf/240293/NICHIA/NSPG520AS.html

## 4.2. Geometric Camera Calibration

The main objective of geometric camera calibration is to determine the intrinsic camera parameters (i.e. parameters of the camera calibration matrix $\mathtt{K} : \alpha_x, \alpha_y, x_0, y_0, s$). Sometimes it is additionally referred to as estimating the position and orientation of a camera with respect to a world coordinate frame (extrinsic parameters $[\mathtt{R} \mid \mathbf{t}]$). **The whole process of camera calibration is therefore closely related to pose estimation.**

The reason why this topic is discussed at this place is that the finite projective camera model presented in Section 3.3.1 is not necessarily sufficient, in the way that it does not account for possible distortions of the camera lenses. Thus, geometric camera calibration is also about modelling lens distortions.

Furthermore, as described in Section 3.3.2, camera calibration is conceived as a separate step before the actual pose estimation measurements and is therefore regarded as an integral part of the image acquisition system.

### 4.2.1. Camera Calibration Methods

Several approaches to geometric camera calibration are presented in literature. According to *Zhang* [93] calibration methods can be classified into two main categories: *object based calibration* and *self calibration*. In object based calibration, the known geometry of 2D or 3D rigid objects is used to estimate the camera parameters. Whereas within the self calibration approach the constraints to estimate the parameters are gained by just moving the camera through a static scene.

Following the same arguments concerning markerless vs. marker-based pose estimation (see introduction to Chapter 3) as well as according to [93], from self calibration methods it is not possible to "*always obtain reliable results*". Therefore, regarding this image-based measurement system an object based calibration approach is clearly preferred.

In their survey, *Remondino and Fraser* [71] present a good comparison on recent object based calibration methods. They state that frequently applied methods are those of *Tsai* [88], *Heikkilä and Silvén* [40] and *Zhang* [93]. Common to all three methods is that they rely on the basic pinhole camera model and include concepts for modelling lens distortions.

Basically all calibration methods comprise the following four steps: acquisition of images by moving a calibration target (e.g. chequerboard pattern) in front of the camera, detection and localisation of features in the images, identification of features from the known calibration pattern and computation of camera parameters by using the established feature correspondences (adopted from [46]).

From a practical point of view, the key difference between the individual approaches to camera calibration is that there exist proposals relying on 3D calibration structures (e.g. [33, 12] and [37, p. 289]) and methods which make use of simple planar targets (e.g. [88, 93]). The latter, are generally considered "*more flexible*" [53] and "*popular for their practical convenience*" [46]. The method of *Heikkilä and Silvén* [40] is able to handle both 2D and 3D calibration targets.

Based on the implications of this discussion, a suitable implementation of a geometric camera calibration procedure is presented in the next section.

### 4.2.2. Camera Calibration Implementation

A suitable, constantly maintained and over the years extensively tested implementation for camera calibration is provided by *Bouguet* [10] called *Camera Calibration Toolbox*. It is available in MATLAB[13] and as C++ implementation as part of the OpenCV library[14]. The MATLAB version of the toolbox provides an easy-to-use graphical user interface for more practical convenience and comes with an in-depth documentation.

The actual calibration procedure realised in the toolbox is based on the methods proposed by *Zhang* [93] (initialisation phase, parameter estimation) and *Heikkilä and Silvén* [40] (intrinsic camera model with enhanced lens distortion model).

The implemented method of *Zhang* [93], first recovers the intrinsic camera parameters by homography estimation using a closed-form solution. Then it minimises the reprojection error by using a Maximum Likelihood estimation based on the *Levenberg-Marquardt algorithm* [58] to refine all retrieved camera parameters.

The Camera Calibration Toolbox applies this scheme of error minimisation to estimate all parameters which are involved in the process of camera calibration (i.e. intrinsic parameters, lens distortion parameters and extrinsic parameters).

### 4.2.3. Modelling Lens Distortion

As already mentioned the introduced finite projective camera model is not considered sufficient when high accuracy is required. It has to be extended by additional terms in order to model possible lens distortions. According to *Weng et al.* [89] such distortions occur due to imperfect lens shape and improper lens and camera assembly (e.g. lens surfaces not strictly collinear).

---

[13] http://www.vision.caltech.edu/bouguetj/calib_doc/index.html

[14] http://docs.opencv.org/2.4.3/modules/calib3d/doc/camera_calibration_and_3d_reconstruction.html

The specific distortions considered by the model of *Heikkilä and Silvén* [40] are: *radial distortion*, *decentering distortion* and *thin prism distortion*. All three types are expressed by certain radial and tangential components in the model. Figure 4.3 illustrates selected effects of different types of distortions.

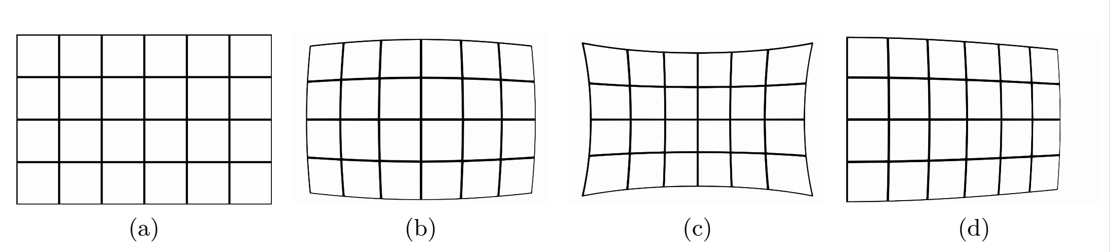

(a) (b) (c) (d)

Figure 4.3.: Examples of different effects of lens distortions: (a) no distortion; (b) negative radial distortion (barrel effect); (c) positive radial distortion (pincushion effect); (d) thin prism distortion caused by slight tilt of lens elements or image sensor.

For this model, a 3D point in the camera reference frame $\mathbf{X}^\mathsf{c} = (X, Y, Z)^\mathsf{T}$ is projected onto the image plane by applying the central projective mapping of the pinhole model (Equation 3.3). This yields a normalised but distorted point coordinate $\mathbf{x}_d$ (as described in [10]):

$$\mathbf{x}_d = \begin{pmatrix} x_d \\ y_d \end{pmatrix} = \begin{pmatrix} fX/Z \\ fY/Z \end{pmatrix} \tag{4.3}$$

with defining $r = ||\mathbf{x}_d|| = \sqrt{x_d^2 + y_d^2}$, the undistorted point coordinate $\mathbf{x}_u$ is expressed as:

$$\mathbf{x}_u = \begin{pmatrix} x_u \\ y_u \end{pmatrix} = (1 + k_1 r^2 + k_2 r^4 + k_3 r^6)\mathbf{x}_d + \mathbf{t} \tag{4.4}$$

where $k_1 \dots k_3$ model the radial distortion and $\mathbf{t}$ being tangential distortion vector with:

$$\mathbf{t} = \begin{pmatrix} 2p_1 x_d y_d + p_2(r^2 + 2{x_d}^2) \\ p_1(r^2 + 2{y_d}^2) + 2p_2 x_d y_d \end{pmatrix} \tag{4.5}$$

The undistorted point coordinate $\mathbf{x}_u$ and the **final pixel coordinates** $\mathbf{x}_p = (x_p, y_p, 1)^\mathsf{T}$ are then related by applying the following transformation

$$\mathbf{x}_p = \begin{pmatrix} x_p \\ y_p \\ 1 \end{pmatrix} = \mathtt{K} \begin{pmatrix} x_n \\ y_n \\ 1 \end{pmatrix} \tag{4.6}$$

with $\mathtt{K}$ being the finite projective camera matrix of form 3.11. From these equations one can see that the initial set of intrinsic camera parameters $\alpha_x, \alpha_y, x_0, y_0, s$ is extended by the five distortion coefficients $k_1, k_2, k_3, p_1, p_2$.

### 4.2.4. Camera Calibration in Practice

The following instructions should offer comprehensive guidance for camera calibration in practice.

1. **Preliminaries**
   Calibrate camera on-site, before test ride to ensure similar lighting conditions as during measurement. Use large enough, torsion-free calibration target. Example of a chequerboard pattern is given in Figure 4.4.

2. **Focus**
   Set focus of lens so that the depth of field best matches the working distance between camera and multi-fiducial targets.

3. **Image acquisition**
   Move calibration target in front of camera and acquire images from different viewing angles and positions of the calibration target. Select 10-15 format-filling images which do not exhibit motion blur.

4. **Calibration procedure**
   Use the Camera Calibration Toolbox to automatically label corners of the chequerboard pattern and start calibration routine. It returns the described extended intrinsic camera parameters as well as an uncertainty estimation.

5. **Reprojection error**
   The toolbox automatically calculates the uncertainties for the results of the calibration. Although *Heikkilä* [39] proclaims an achievable accuracy of 0.02 pixel, compared to the evaluations in [85] an error of $\leq 0.5$ pixel is considered sufficient.

6. **Reduced distortion model**
   According to *Bouguet* [10], it is justifiable (due to high quality of modern lenses) to apply a reduced distortion model by not considering the $6^{th}$ order radial coefficient and the complete tangential distortion component (i.e. $\{k_3, p_1, p_2\} = 0$). If applicable, adjust the toolbox respectively.

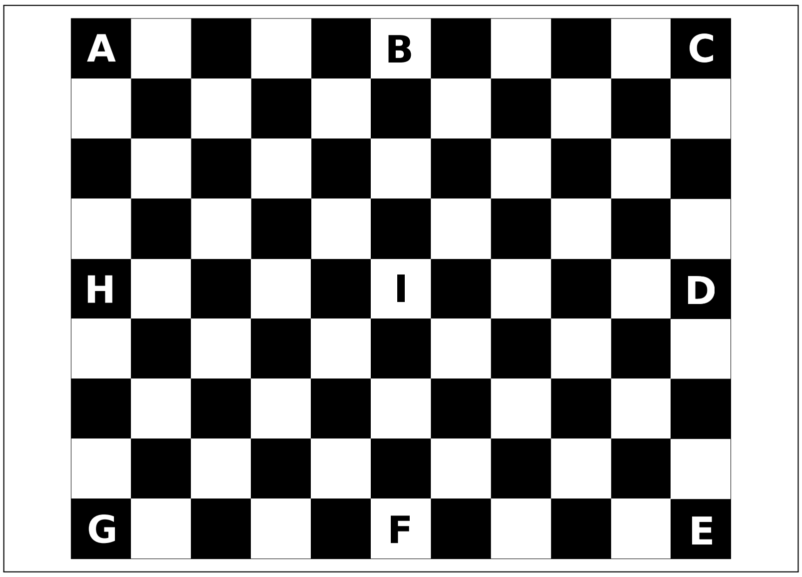


Figure 4.4.: Example of a planar calibration target using a chequerboard pattern. It is designed to be used with ISO A-series paper sizes. The letters are used for identifying always the same designated origin even under rotation.

# 5. Experimental Evaluation

This chapter presents the extensive experiments and evaluations conducted in order to test, evaluate and verify the feasibility of the designed measurement system for optically tracking relative motions between railway vehicle carbodies.

## 5.1. Methodology

The evaluation is divided into three main parts, as the experiments are realised using three distinct measurement setups, each employing a different target object tracked. The first series of experiments is conducted at a laboratory scale in order to evaluate the basic functionality of the measurement system. For the second set of experiments, the 6-DOF motions of the carbodies are simulated using a full-sized hydraulic bogie test rig. The third experiment is conducted in a real world environment, on board a running high speed train.

For all the experiments the image acquisition system as described in Section 4.1.1 is used to record imagery data. The SONY NEX-5N is mounted on a suitable tripod and calibrated according to the guidance in Section 4.2.4. During the experiments, single images or video streams are used to record the motions of the target object.

After image acquisition the data is processed according to the steps described in Section 3. First the markers on all present multi-fiducial targets are tracked by the described tracking algorithm. After extracting the subpixel-accurate image location of each fiducial, the pose of the carbody is estimated from the given 2D-3D correspondences. This is accomplished by applying the DLT and/or EPnP algorithms as presented in Section 3.3. Then the sought-after relative pose is calculated by differencing the pose from image $I_0$ (initial position) and the one in the current image. This process is repeated for every acquired image.

The prototypical realisation of this processing toolchain is implemented in MATLAB. In contrast to the specified requirement R(g), it is not capable of computing the relative motions in real-time since image acquisition and further computations are processed sequentially. In a productive system a change towards parallel processing is considered necessary.

## 5.2. Implementations of Pose Estimation Algorithms

The specific implementations for the pose estimation algorithms are taken from:

- **DLT**
  A MATLAB implementation of the DLT algorithm is part of the Camera Calibration Toolbox (see Section 4.2.2). The function for computing the pose of the camera is realised in the file: `compute_extrinsic.m`.

  Download:
  http://vision.caltech.edu/bouguetj/calib_doc/download/toolbox_calib.zip

  Remark:
  This implementation of the DLT algorithm comes with an additional iterative refinement process based on the Levenberg-Marquardt algorithm (see `compute_extrinsic_refine.m`). This refinement is applied to all pose estimation calculations.

- **EPnP**
  This MATLAB implementation is directly taken from the homepage of its creator. The function which estimates the pose is implemented in the file `efficient_pnp.m` and including the Gauss-Newton optimisation in `efficient_pnp_gauss.m` respectively.

  Download:
  http://cvlab.epfl.ch/software/EPnP/EPnP_matlab.zip

  Remark:
  In accordance with the usage described in [54], the standard EPnP procedure is applied in the case of coplanar fiducials (i.e. all lie in one plane), whereas in the non-coplanar case the EPnP with Gauss-Newton optimisation is utilised.

## 5.3. Experiments in Laboratory

The experiments carried-out in the laboratory serve as a proof of concept to demonstrate basic functionality and feasibility of the proposed measurement system. Therefore these experiments were conducted on a small-scale basis, with constrained motions performed by the target object while acquiring single images instead of whole image streams. The employed basic measurement setup is described in the next section.

### 5.3.1. Measurement Setup

The measurement setup for the series of experiments conducted in the laboratory, hereinafter referred to as Experiment 1, consists of a solid rail-mounted base frame for one multi-fiducial target, with the rails permitting uniform linear displacements over a length of up to 860 mm. Furthermore the frame can pivot about the vertical axis (yaw-rotation ($\psi$)) from $0^\circ$ to $360^\circ$.

For each experiment the camera is operated in single-image mode. In order to simulate continuous linear movements, a sequence of images is captured recording the target in sequential equidistantly spaced positions. By defining the target in frame $I_0$ to be in the initial pose (see Section 3.4.3), the final relative pose estimations are calculated with respect to this position. Figure 5.1 pictures the measurement setup with the rail system holding the manually moveable base frame.

Precise reference data of the uniform linear displacements is recorded using a laser rangefinder (Bosch GLM 150). The uncertainty of the reference data is roughly estimated by accumulating the uncertainty specified by the manufacturer of the laser rangefinder ($\pm 1$ mm under typical conditions [73]) and an assumed misalignment of $\pm 1^\circ$ between the measured and the actual direction of displacement (yielding $\sim$860.13 mm instead of 860 mm). This amounts to an upper bound of uncertainty of $\leq 1.2$ mm.

This measurement setup facilitates the following experiments which are described in the subsequent sections:

- Experiment 1a: simulation of translational movement along x-axis
- Experiment 1b: simulation of translational movement along y-axis
- Experiment 1c: simulation of rotation about z-axis (yaw-rotation)

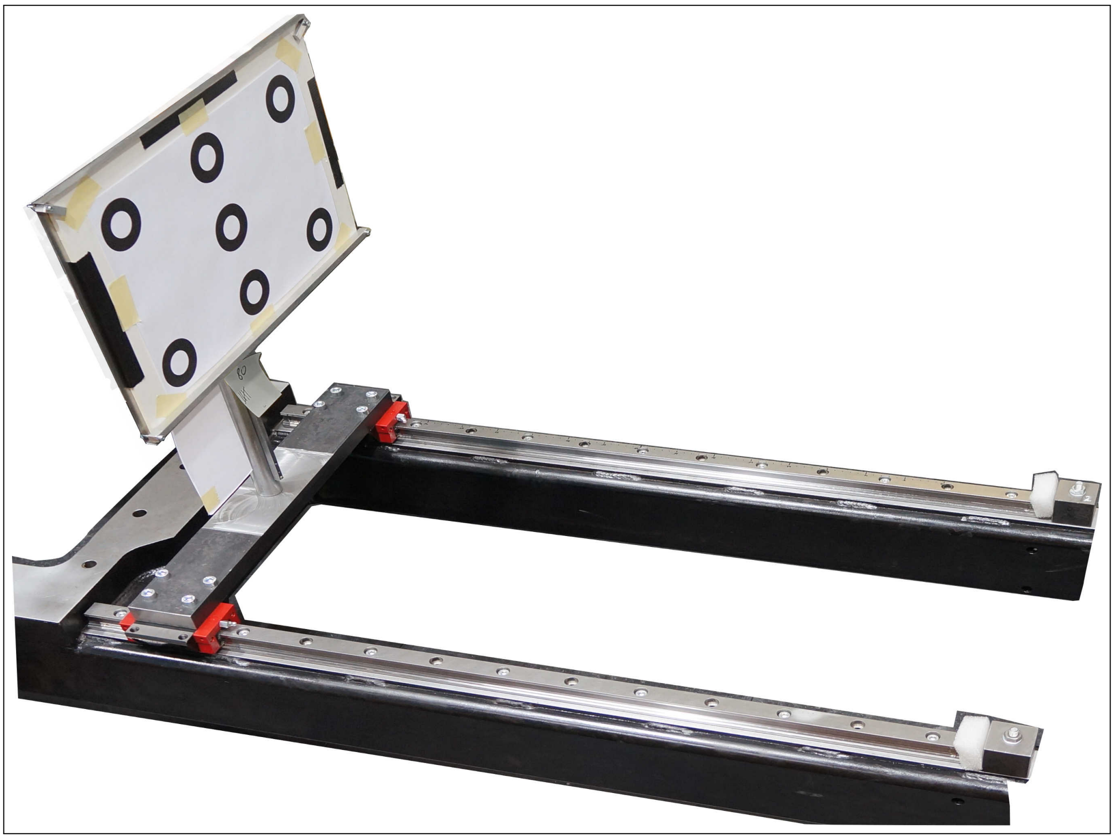


Figure 5.1.: Measurement setup used to conduct the series of experiments in laboratory (referred to as Experiment 1). Setup consists of solid rail system and transverse mounted, manually moveable base frame holding rotatable multi-fiducial 7-CCC target.

### 5.3.2. Experiment 1a

For this experiment, the 7-CCC target, the rail-system and the camera were arranged in a row, enabling the simulation of a uniform target motion along the x-axis (see Figure 5.2) of the carbody motion model presented in Section 3.4.1.

The linear movement was performed over a range of 800 mm in steps of 10 mm. To ensure a better verification, two images were acquired per step yielding a total of 160 measurement points. The reference data for the movement was measured along the x-axis only. All translations and rotations along other DoF are expected to be zero.

Although the measurement setup was adjusted very carefully, the presence of a small rotational misalignment between the axis of the rail system and the coordinate system of the initial pose is inevitable. This effect introduces a certain bias to the calculated translational displacements. As this misalignment is of constant nature, it is computationally compensated for in an extra post-processing step (using a constant angle-factor).

The measurement setup is subject to the following parameters:

- Distance to target: 4500 mm (at $I_0$)
- Image size: $4912 \times 3264$ pixel$^2$
- Focal length ($f$): 55 mm
- Aperture value: $f/22$
- Exposure time: 1/50 sec

(a) $I_0$ (b) $I_{159}$

Figure 5.2.: (a) and (b) show frames recorded at beginning ($I_0$) and end ($I_{159}$) of linear target displacement along x-axis performed in Experiment 1a. The tracking results are superimposed with green crosses.

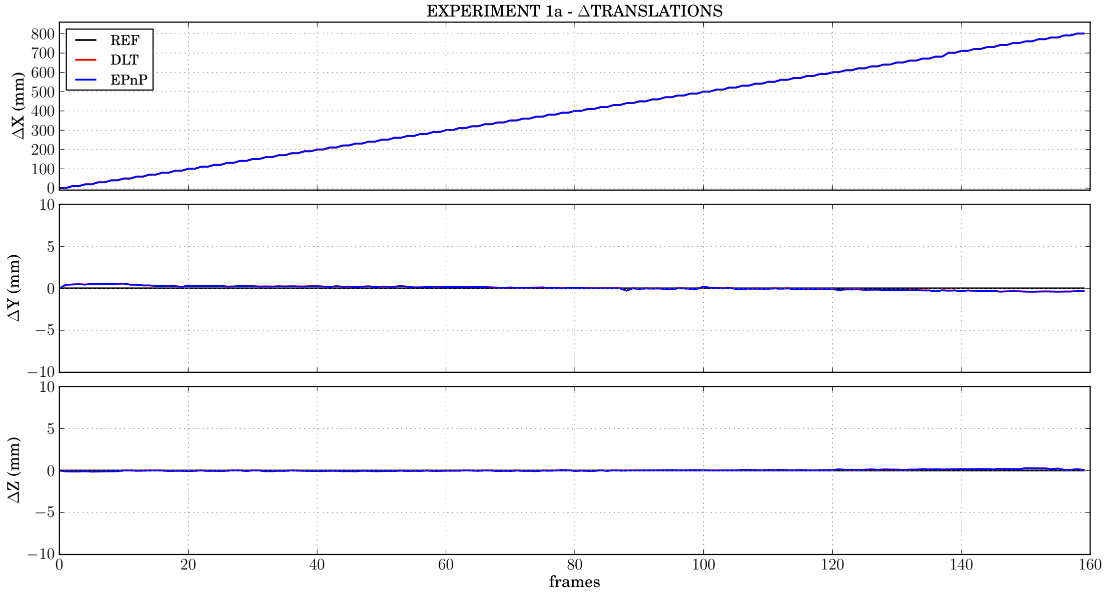


Figure 5.3.: Relative translations performed by target object in Experiment 1a represented by the individual axes {ΔX,ΔY,ΔZ}. The deviations of displacements in other than the main direction {ΔY,ΔZ} are very low (note the different scale). The ΔX-signal is staircase-shaped because of two images recorded per target position.

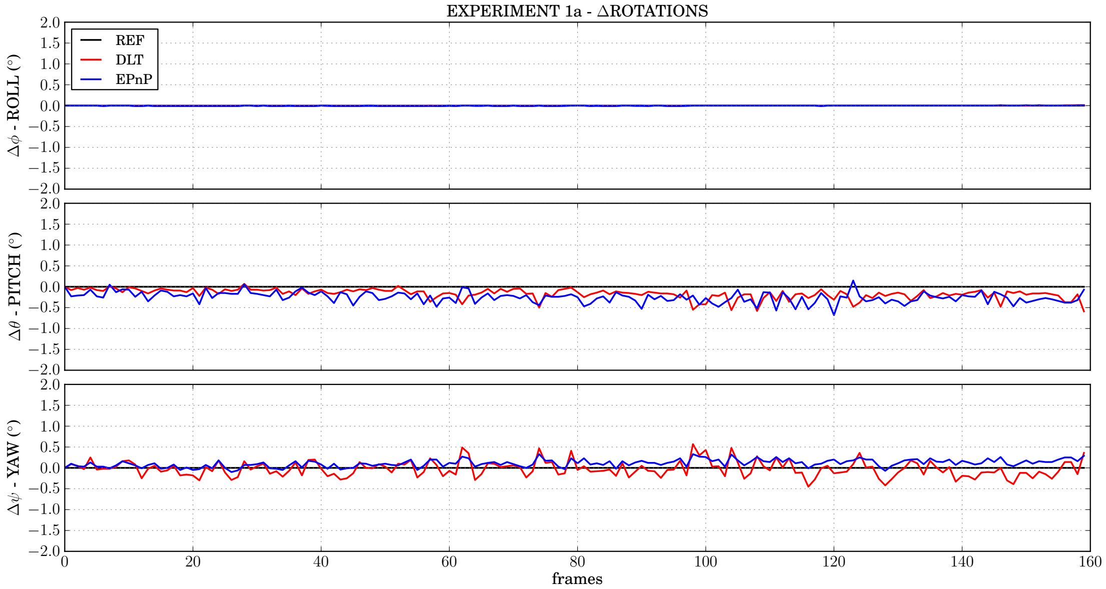


Figure 5.4.: Relative rotations performed by target object in Experiment 1a represented by the individual angles of rotations $\{\Delta\phi,\ \Delta\theta,\ \Delta\psi\}$. The high variance of the $\{\Delta\theta, \Delta\psi\}$ angles is most probably due to minor input changes and numerical issues in pose estimation methods as well as inaccuracies in the measurement setup (rail system, poor initialisation).

Figures 5.3 and 5.4 show the relative translational and rotational displacements of the linear target movement, subdivided into the six individual DoF (i.e. {ΔX,ΔY,ΔZ} and $\{\Delta\phi, \Delta\theta, \Delta\psi\}$). Each diagram shows the values of the DLT and EPnP pose estimation algorithms and the provided reference data (REF). The main direction of displacement of the target is into axial x-direction, with no other major translation or rotation measured.

From Figure 5.3 it can be seen that the measurements of the translations are very accurate. Only small deviations to the reference data are perceivable (e.g. $\text{MAX}_{\text{DLT}}(\Delta\text{X})=$ 1.5 mm, see Table 5.1). The staircase-shape visible in the signal of the x-translation ($\Delta$X) is due to the two images recorded per target position.

The first notable characteristic about the measured rotations presented in Figure 5.4 is that the roll angle ($\theta\phi$) can be determined most accurately. The perceivable variance in the signals of $\Delta\theta$ and $\Delta\psi$ is influenced by three major factors:
First, it is evident that for both pose estimation algorithms, the slightest change in the subpixel image location of one or more fiducials primarily affects the calculation of $\Delta\theta$ and $\Delta\psi$ angles. In contrast, the image positions of all fiducials need to be affected in order to induce deviations of the $\Delta\phi$ angle. From the differences in measurements between two consecutive frames in which the target remained fixed (no staircase-effect visible), it can be concluded that the DLT and EPnP algorithms are susceptible to minor input changes. These input changes result from different initialisations of the CoG method during the tracking procedure (see Section 3.2.3).
Second, the particular configuration, with only one target and the small distances between the fiducials in the image, amplifies the mentioned effect. In this case the pose estimation methods are influenced by numerical issues, where even very small changes to the image locations result in large deviations.
Third, the frames where both DLT and EPnP show the same magnitude of deviation, apparently indicate the presence of a certain amount of play (flexibility) between rails and sliding base frame. Thus, the deviations at least partially arise from inaccuracies introduced by the rail system.

The small but constant offset of both DLT and EPnP in the measurements of $\Delta\theta$, is considered as a consequence of a poor initialisation, where the value calculated from the pose in the first frame ($I_0$) is not representative for all subsequent ones.

For a better comparison, Figure 5.5 shows the deviation of the $\Delta$X signal of DLT and EPnP to the reference data. Compared to the results of $\Delta$Y and $\Delta$Z, $\Delta$X exhibits the largest deviation, but is still within the limits defined in requirement R(a). It is important to observe that the deviation over the whole range does not increase with the distance of the target from the camera.

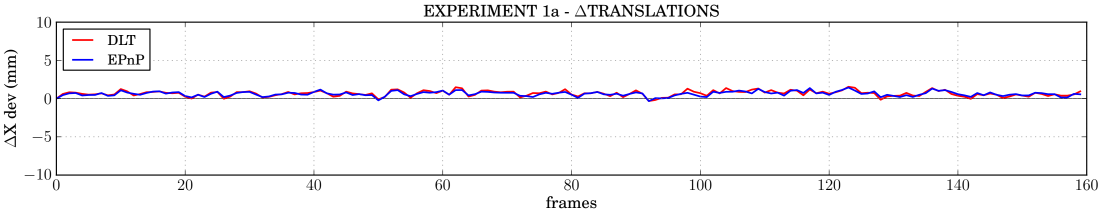


Figure 5.5.: Deviations of the $\Delta$X signal for both DLT and EPnP from the reference data. Note that the deviation does not increase with the distance of the target from the camera.

### 5.3.3. Experiment 1b

For this experiment the target, the rail system and the camera were arranged in a way to simulate a uniform relative translation of the carbody into the y-direction of the defined coordinate system. This time, the translation was performed back and forth over the range of 860 mm in steps of 20 mm.

As in Experiment 1a, two images were acquired per position yielding 176 measurement points. The reference data for the displacement was measured along the y-axis only, with the same estimation of uncertainty as valid for Experiment 1a. All translations and rotations along other DoF are expected to be zero. Furthermore, the initial rotational misalignments of the measurement setup (rail-system vs. initial pose) were corrected by a constant angle-factor after completing the measurement in an additional post-processing step.

Figure 5.6 illustrates the configuration of the setup: the camera was installed alongside the rail system with the projection of the target being located in the image centre in $I_0$ and at the image border in $I_{88}$.

The measurement setup is subject to the following parameters:

- Distance to target: 3600 mm (at $I_0$)
- Image size: $2448 \times 1624$ pixel$^2$
- Focal length ($f$): 35 mm
- Aperture value: $f/22$
- Exposure time: 1/40 sec

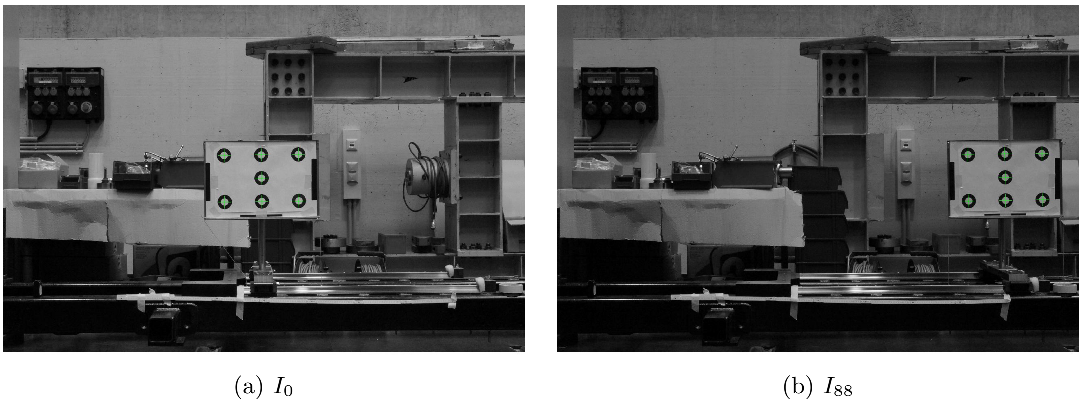

(a) $I_0$ (b) $I_{88}$

Figure 5.6.: (a) and (b) show frames recorded at beginning ($I_0$) and maximum ($I_{88}$) of linear target displacement along y-axis. Although using less distance between camera and target, compared to Experiment 1a, the appearance of the target is smaller due to the shorter focal length applied. The tracking results are superimposed as green crosses.

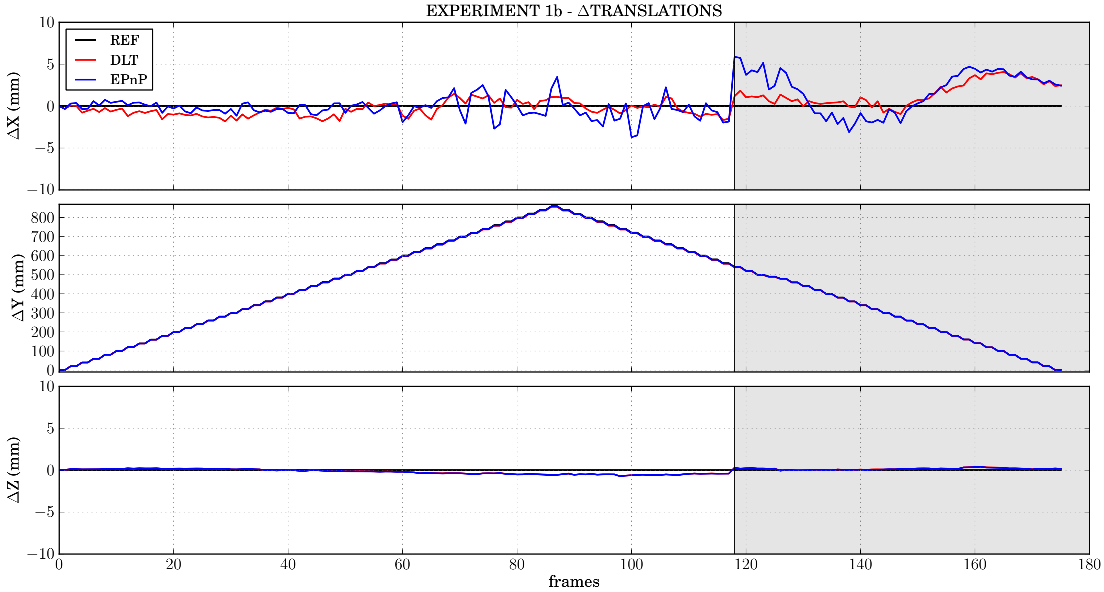


Figure 5.7.: Relative translations performed by target object in Experiment 1b represented by the individual axes {ΔX,ΔY,ΔZ}. The ΔY-signal is staircase-shaped because of two images recorded per target position. Outliers from frame 118 onwards are result of battery change (greyed out).

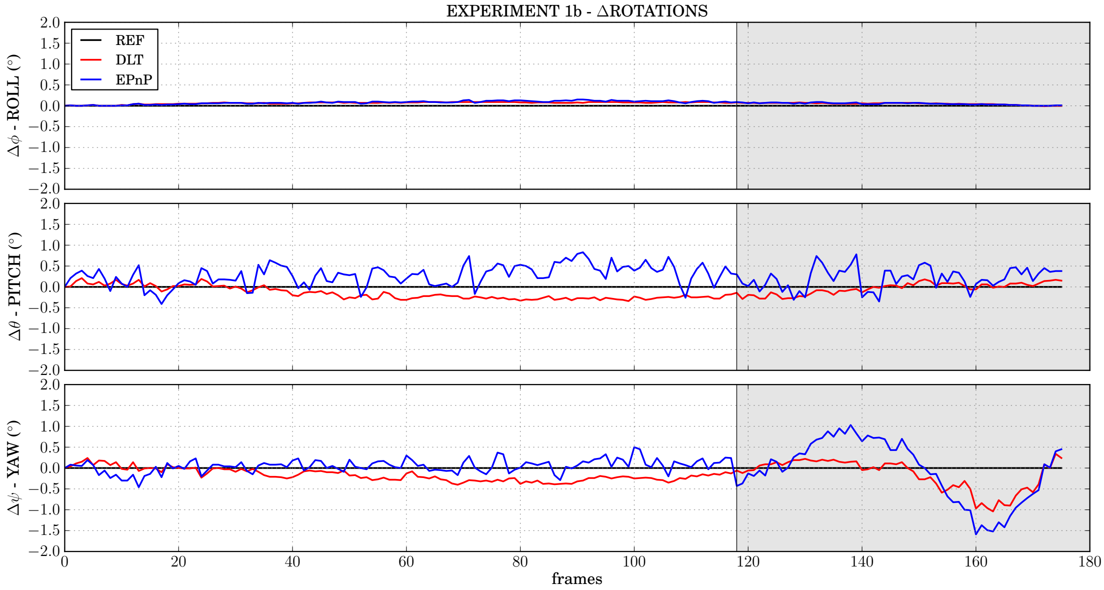


Figure 5.8.: Relative rotations performed by target object in Experiment 1b represented by the individual angles of rotations {$\Delta\phi$, $\Delta\theta$, $\Delta\psi$}. Outliers from frame 118 onwards are result of battery change (greyed out).

As expected, from Figure 5.7 it is apparent that the target's main direction of displacement is performed along axial y-direction. Figures 5.7 and 5.8 show larger deviations of the measurements from the reference data compared to Experiment 1a. An obvious reason for this effect lies in the reduced resolution of the images and the smaller appearance of the target in the image data due to the shorter focal length applied.

The considerable outliers in the measurements at and after frame 118, visible in the signals of $\Delta X$, $\Delta Z$ and $\Delta\psi$, are result of a battery change which was necessary during the experiment. Although not unmounted from the tripod, it recognizably changed the orientation of the camera.

Furthermore, in this experiment the pose calculations of the EPnP algorithm generally exhibit larger deviations compared to the results of the DLT algorithm (e.g. $\mathrm{STD}(\Delta\theta)_{\mathrm{DLT}} = 0.15\,°$, $\mathrm{STD}(\Delta\theta)_{\mathrm{EPnP}} = 0.26\,°$; see Table 5.2). A reason for this effect is that for a planar target, as in the present case, the EPnP algorithm without the Gauss-Newton optimisation is applied (conforms with usage described in [54]; application of optimisation procedure leads to severe outliers in resulting pose estimates).

For a better comparison Figure 5.9 shows the deviation of the $\Delta Y$ measurements of both pose estimation algorithms to the reference data.

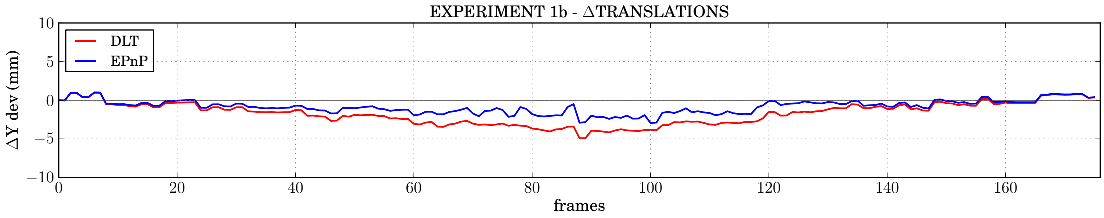


Figure 5.9.: Deviations of the $\Delta Y$ signal for both DLT and EPnP from the reference data. The slightly increasing deviation most likely caused by not properly aligned device for measuring the reference data.

Although the signals $\{\Delta X, \Delta Y, \Delta Z\}$ exhibit larger deviations, these are still within the limits defined in requirement R(a). The deviations of $\Delta\theta$ and $\Delta\psi$ are outside the required tolerance. Both effects are best explained with uncertainties in the pose estimation procedures and inaccuracies in the rail system as part of the measurement setup. The given data does not provide evidence for any significant systematic bias in neither the camera calibration procedure, the measurement setup nor the pose estimation algorithms. Even the slight increase in deviations of the reference data in $\Delta Y$ (see Figure 5.9) is most probably caused by a small misalignment of the laser rangefinder while measuring the reference data.

### 5.3.4. Experiment 1c

For this experiment the same arrangement in the measurement setup as in Experiment 1b was chosen. This time the behaviour of the measurement system while applying rotations to the tracked object is investigated. For this experiment the target was pivoted around its z-axis (yaw-rotation ($\Delta\psi$)) from its initial position at $0\,°$ up to $\pm 25\,°$ in steps of $5\,°$. The trajectory with a total of 42 measurement points performed is: $0\,°, 25\,°, 0\,°,\ -25\,°, 0\,°$.

Since, for this case the laser rangefinder was not considered to provide suitable reference data, the values were read manually after rotating the target to the given measurement points. Thus, it is important to understand that, in this experiment **the reference data can only be used as an indicator for possible deviations**. The maximum uncertainty for the manually obtained reference data is estimated with $\pm 2\,°$.

In contrast to the other two experiments, the origin of the target object's coordinate system is adjusted to be located at the centre of rotation of the base frame.

The measurement setup is subject to the following parameters (same as in Experiment 1b):

- Distance to target: 3600 mm (at $I_0$)
- Image size: $2448 \times 1624$ pixel$^2$
- Focal length ($f$): 35 mm
- Aperture value: $f/22$
- Exposure time: 1/40 sec

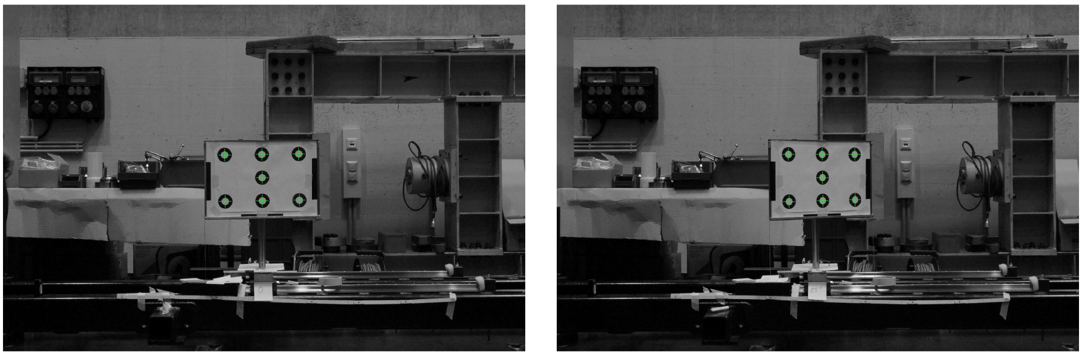

(a) $I_0$ (b) $I_{11}$

Figure 5.10.: (a) and (b) show frames recorded at $0\,°$ ($I_0$) and at maximum rotation of $25\,°$ ($I_{11}$) of target rotation about the z-axis. The tracking results are superimposed as green crosses.

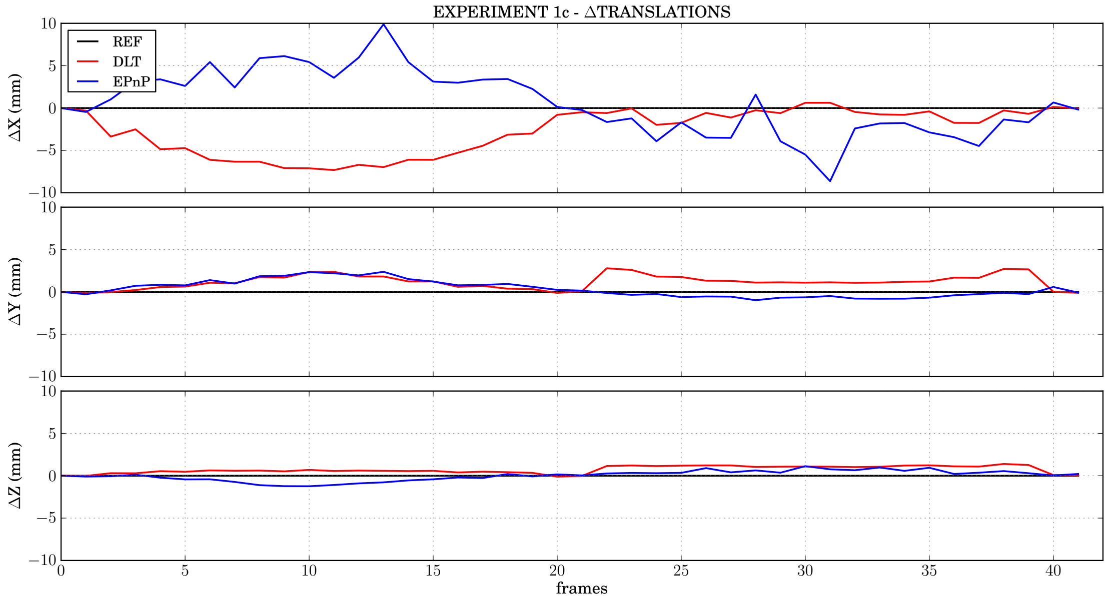


Figure 5.11.: Relative translations performed by target object in Experiment 1c represented by the individual axes {ΔX,ΔY,ΔZ}. The high deviation in the ΔX-signal are most likely due to a misalignment of the target axis and the axis of rotation of the base frame. Note the divergent solutions of DLT and EPnP in ΔX.

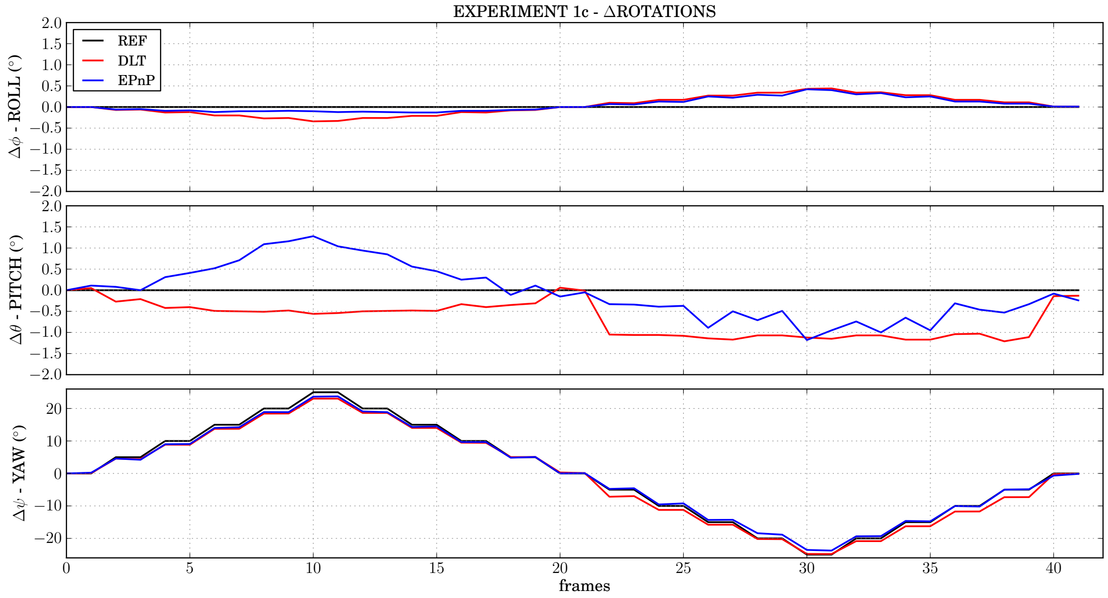


Figure 5.12.: Relative rotations performed by target object in Experiment 1c represented by the individual angles of rotations $\{\Delta\phi,\ \Delta\theta,\ \Delta\psi\}$. Note the divergent solutions of DLT and EPnP in $\Delta\theta$.

Figures 5.11 and 5.12 exhibit significant deviations of the measurements in Experiment 1c. To a great extent, these differences arise from a misalignment of the target coordinate system and the axis of rotation of the base frame holding the target. It shows that even the smallest misalignment in the setup can lead to considerable deviations under rotations up to ±25 °.

In the measurements of $\Delta X$ and $\Delta\theta$ the pose estimates of DLT and EPnP result in divergent solutions. This may be explained by pose ambiguities in critical configurations introduced when employing a planar target. This influence on measurements is investigated in detail by *Schweighofer and Pinz* [76].

For a better comparison Figure 5.13 shows the deviation of the $\Delta\psi$ measurements of both pose estimation algorithms to the reference data.

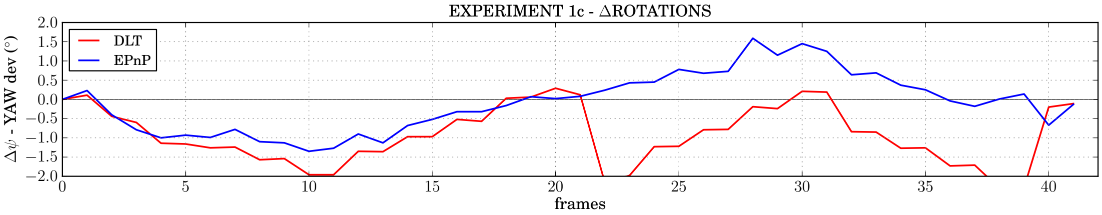


Figure 5.13.: Deviations of the $\Delta\psi$ signal for both DLT and EPnP from the reference data. Note that the corresponding reference data is determined manually.

The lack of reliable reference data and the obvious misalignment of the target object coordinate system with the base frame's centre of rotation do not allow for any causal interpretation of the measurements. The deviations from the reference data visible in all figures of Experiment 1c (Figures 5.11, 5.12, 5.13) only serve as a coarse indicator for the actual deviations.

Nevertheless, one can observe that the resulting measurements are by no means at random, but clearly show the expected main rotation about the z-axis of the target. Furthermore, it is visible, that the deviation is not necessarily increasing with growing rotational displacement, as the deviation around −25 ° (frame 30) of the DLT algorithm is very small.

Moreover, Experiment 1c reveals the important fact that, for an identical input of a degenerate configuration the two employed pose estimation algorithms show distinct results.

### 5.3.5. Uncertainty Analysis

The evaluation of the measurement uncertainty of the experiments in the laboratory is conducted by using the common measures of the maximum error (MAX), the root-mean-square error (RMSE), the arithmetic mean (MEAN) and its corresponding standard deviation (STD). The uncertainties of DLT and EPnP are presented in Tables 5.1, 5.2 and 5.3 for each experiment and each DoF respectively. Red values in the MAX column indicate values of signals which exceed the limits defined in requirement R(a). Reasons for these exceedances and the corresponding uncertainties of the reference data are given in the respective discussions above.

**In general, it can be concluded that the estimation of uncertainties for Experiment 1 confirms the basic feasibility and a sufficient accuracy of proposed measurement system**.

| **EX 1a** | **MAX** | **RMSE** | **MEAN** | **STD** |
|---|---|---|---|---|
| (mm) | ΔX / ΔY / ΔZ | ΔX / ΔY / ΔZ | ΔX / ΔY / ΔZ | ΔX / ΔY / ΔZ |
| **DLT** | 1.5 / 0.6 / 0.3 | 0.8 / 0.2 / 0.1 | 0.7 / 0.0 / 0.0 | 0.4 / 0.2 / 0.1 |
| **EPnP** | 1.5 / 0.6 / 0.3 | 0.7 / 0.2 / 0.1 | 0.6 / 0.0 / 0.0 | 0.3 / 0.2 / 0.1 |
| (°) | $\Delta\phi$ / $\Delta\theta$ / $\Delta\psi$ | $\Delta\phi$ / $\Delta\theta$ / $\Delta\psi$ | $\Delta\phi$ / $\Delta\theta$ / $\Delta\psi$ | $\Delta\phi$ / $\Delta\theta$ / $\Delta\psi$ |
| **DLT** | 0.01 / 0.59 / 0.57 | 0.01 / 0.20 / 0.19 | 0.00 / -0.16 / -0.02 | 0.00 / 0.12 / 0.19 |
| **EPnP** | 0.01 / 0.68 / 0.33 | 0.01 / 0.28 / 0.14 | 0.00 / -0.25 / 0.11 | 0.00 / 0.13 / 0.09 |

Table 5.1.: Uncertainty of measurements for Experiment 1a.

| **EX 1b** | **MAX** | **RMSE** | **MEAN** | **STD** |
|---|---|---|---|---|
| (mm) | ΔX / ΔY / ΔZ | ΔX / ΔY / ΔZ | ΔX / ΔY / ΔZ | ΔX / ΔY / ΔZ |
| **DLT** | 4.1 / 4.9 / 0.7 | 1.4 / 2.2 / 0.3 | 0.2 / -1.7 / -0.1 | 1.4 / 1.4 / 0.3 |
| **EPnP** | 5.9 / 2.9 / 0.7 | 2.0 / 1.2 / 0.3 | 0.4 / -0.9 / -0.1 | 1.9 / 0.9 / 0.3 |
| (°) | $\Delta\phi$ / $\Delta\theta$ / $\Delta\psi$ | $\Delta\phi$ / $\Delta\theta$ / $\Delta\psi$ | $\Delta\phi$ / $\Delta\theta$ / $\Delta\psi$ | $\Delta\phi$ / $\Delta\theta$ / $\Delta\psi$ |
| **DLT** | 0.09 / 0.34 / 1.04 | 0.07 / 0.19 / 0.30 | 0.06 / -0.11 / -0.16 | 0.03 / 0.15 / 0.25 |
| **EPnP** | 0.15 / 0.83 / 1.59 | 0.08 / 0.36 / 0.44 | 0.07 / 0.26 / 0.01 | 0.04 / 0.26 / 0.45 |

Table 5.2.: Uncertainty of measurements for Experiment 1b.

| **EX 1c** | **MAX** | **RMSE** | **MEAN** | **STD** |
|---|---|---|---|---|
| (mm) | ΔX / ΔY / ΔZ | ΔX / ΔY / ΔZ | ΔX / ΔY / ΔZ | ΔX / ΔY / ΔZ |
| **DLT** | 7.3 / 2.8 / 1.4 | 3.7 / 1.4 / 0.8 | -2.7 / 1.2 / 0.7 | 2.7 / 0.8 / 0.4 |
| **EPnP** | 9.9 / 2.4 / 1.2 | 3.8 / 1.0 / 0.6 | 0.6 / 0.3 / 0.0 | 3.8 / 1.0 / 0.6 |
| (°) | $\Delta\phi$ / $\Delta\theta$ / $\Delta\psi$ | $\Delta\phi$ / $\Delta\theta$ / $\Delta\psi$ | $\Delta\phi$ / $\Delta\theta$ / $\Delta\psi$ | $\Delta\phi$ / $\Delta\theta$ / $\Delta\psi$ |
| **DLT** | 0.44 / 1.21 / 2.31 | 0.22 / 0.78 / 1.19 | 0.03 / -0.66 / -0.92 | 0.22 / 0.42 / 0.76 |
| **EPnP** | 0.42 / 1.28 / 1.59 | 0.17 / 0.63 / 0.76 | 0.05 / -0.04 / -0.08 | 0.16 / 0.64 / 0.77 |

Table 5.3.: Uncertainty of measurements for Experiment 1c. Shown greyed out because the data is not statistically meaningful due to lack of reliable reference data.

## 5.4. Experiments on Test Rig

A series of experiments (hereinafter referred to as Experiment 2) were performed on a hydraulic bogie test rig (Stewart/Gough-Platform), to test and simulate the proposed measurement system in a real-world application. During regular operation, this test rig is used to conduct endurance and life-cycle tests. With the platform and its specialised six-cylinder layout (see Figure 5.14), it applies dynamic loads onto the bogie, thus simulating different riding conditions. For this specific task the platform is able to perform the 6-DoF motions of a carbody and therefore acts as a perfect target object for the proposed measurement system. Furthermore, reliable and accurate reference data about pose of the platform can be retrieved.

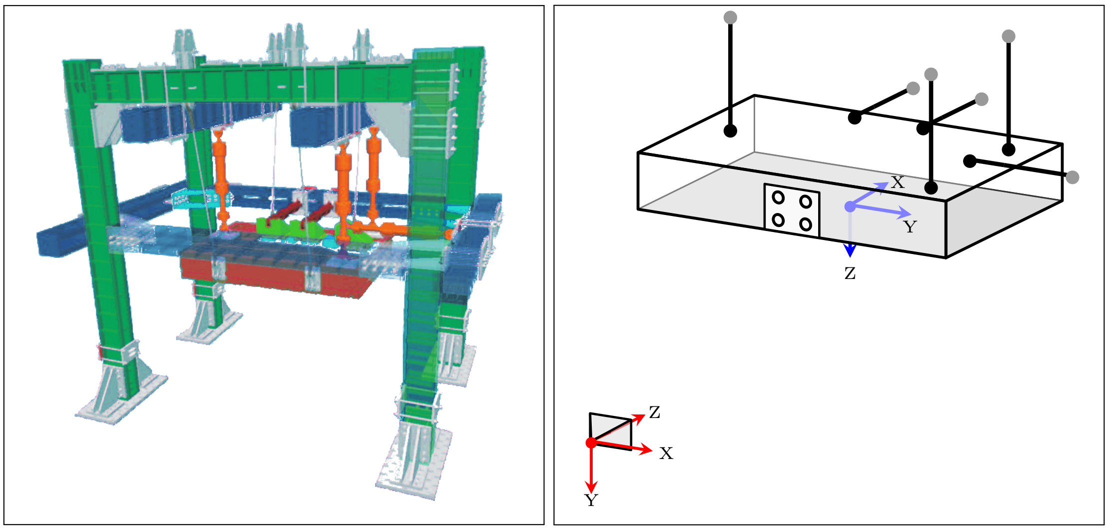
In this section presenting Experiment 2, first the applied basic measurement setup is introduced (see Section 5.4.1) and important details about the available reference data and synchronisation are given (see Section 5.4.2). Then the results of the conducted three test runs with different settings and parameters, are presented (see Sections 5.4.3, 5.4.4 and 5.4.5). Thereafter, from the gained data of the experiments, extensive evaluations are made in order to make reliable predictions about the influence of the applied number of fiducials and targets on the measurement uncertainty of the system (see Sections 5.4.6 and 5.4.7).

The following list presents an overview of the conducted experiments and evaluations:

- Experiment 2a: Test run using multiple targets.
- Experiment 2b: Same as Experiment 2a but with different displacements.
- Experiment 2c: Test run with one planar target with larger appearance.
- Evaluation 2a: Uncertainty of DLT and EPnP, subject to increasing number of fiducials.
- Evaluation 2b: Demonstrates change in uncertainties when applying non-coplanar input data.

### 5.4.1. Measurement Setup

For Experiment 2 the measurement setup consists of the mentioned hydraulic bogie test rig with various multi-fiducial targets attached to its movable platform and the image acquisition system introduced in Section 4. Figure 5.14a shows a CAD model of the whole bogie test rig. Figure 5.14b illustrates the movable platform and the defined reference frames between which the simulated relative carbody motions are measured.



(a) (b)

Figure 5.14.: (a) CAD Model of the bogie test rig. Consists of a structure carrying the platform (red) and the attached six cylinders (red/orange). By continuously changing the positions of the cylinders the platform performs different 6-DoF motions. Image adapted from [68]. (b) illustration of the basic measurement setup. Camera facing towards platform with a mutli-fiducial target attached to front. The origin of the reference frame is defined to be at the platform's centre of the bottom surface.

The camera was placed on the ground in about 5 m distance to the platform. For each experiment the motions of the platform are recorded by the camera in a continuous video stream in the AVCHD standard using the HD1080p/50 format (i.e. full frames with $1920 \times 1080$ pixel$^2$ at 50 fps).

For Experiments 2a and 2b multiple ISO A3 sized 7-CCC targets (three coplanar targets in the front, one spatially displaced at the bottom back) were attached to the platform. For Experiment 2c one large target (ISO A1 sized) was attached to the front in order to evaluate the system using larger fiducials.

As proposed in Section 4.2 the camera calibration was done on-site. For setting up the calibration target as well as mounting the multi-fiducial targets so-called KAPA boards[15] were used.

---

[15] http://www.kapaplatten.de/produkte/kapa_mount.html

KAPA boards are special lightweight, aluminium-reinforced boards with a rigid foam core (see [18]). The sandwich construction makes them particularly high-strength and torsionally resistant but still enables easy processing (e.g. cutting to size). These KAPA boards are perfectly suitable for mounting the multi-fiducal targets as well as holding the calibration target.

Figure 5.15 shows the measurement setup with the image acquisition system, the mounted multi-fiducial targets and the calibration target used.

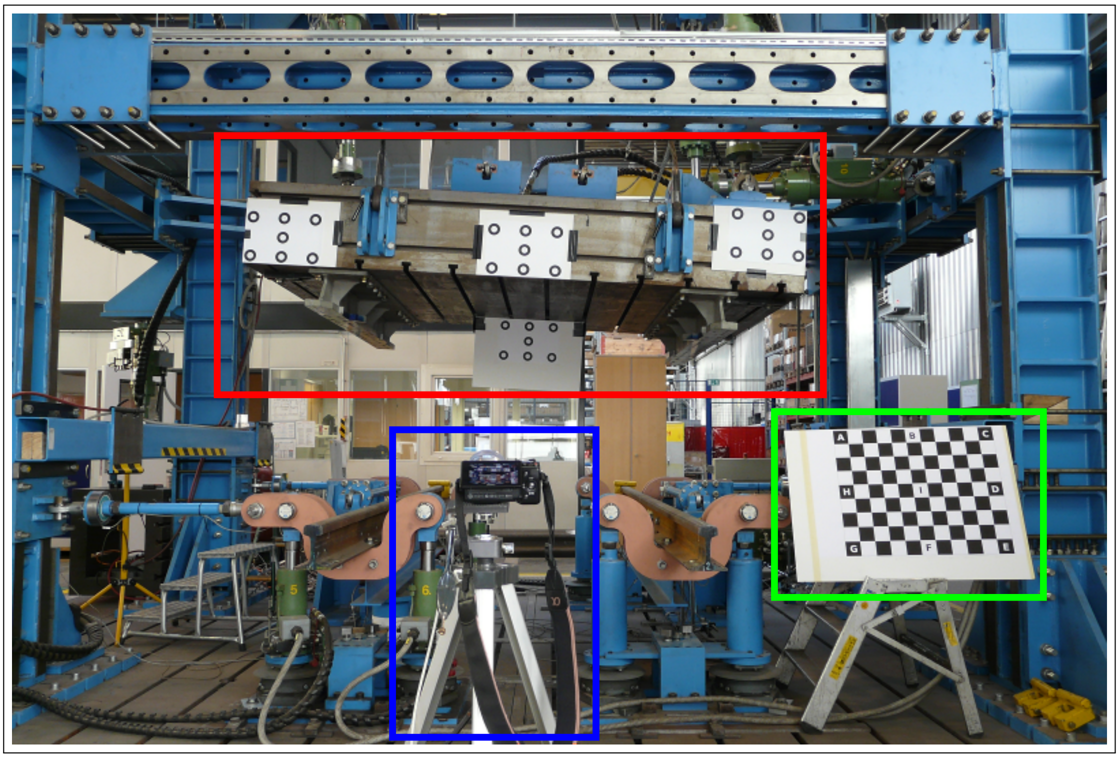


Figure 5.15.: Basic measurement setup used in Experiment 2: the movable platform with three targets mounted in front and one in the back (red box), the calibration target (green box) and the tripod mounted camera in the foreground (blue box). All targets are applied to KAPA boards for torsion-free mounting. Note that all experiments were conducted with the test rig unloaded (i.e. no bogie present).

### 5.4.2. Reference Data and Synchronisation

The reference data for the current pose of the platform is retrieved from the control system of the test rig. It calculates the pose directly from the positioning of the cylinders measured by additional displacement-sensors attached to each hydraulic cylinder. According to [68] the maximum error (MAX) of pose calculations of the test rig lies in the range of ±3 mm and ±2 °.

The data of the test rig is sampled at a frequency of 50 Hz. This sampling rate is very convenient regarding the task of synchronisation because the chosen video recoding mode samples images at 50 fps likewise.

The inherent drift of the internal clocks of the two measurement systems is disregarded, because the clocks are considered to be sufficiently accurate and the test runs are expected to last shorter than 20 minutes. Thus only the beginning of the reference data and the image data has to be aligned.

Precise initial time alignment of the reference data and measurements is a very crucial part of the evaluation since it strongly impacts the uncertainty of measurement.

The time alignment is based on a simple image subtraction followed by a summation of absolute grey level differences (sum of absolute differences (SAD), see Equation 5.1). It is used to calculate a basic metric between the first ($F_0$) and the subsequent frames ($F_t$ with $t = 0 \ldots 1200$ (1200 estimated)) in the recorded sequence.

$$\mathrm{SAD} = \sum_{x,y \in F} |F_0(x,y) - F_t(x,y)| \tag{5.1}$$

The specific frame at which the sum of differences first exhibits the most significant change (the metric gets larger) is determined to be the start of the test run (i.e. where the platform first moves) and defined to be $I_0$ in the tracking process. It is marked with a blue square in Figure 5.16.

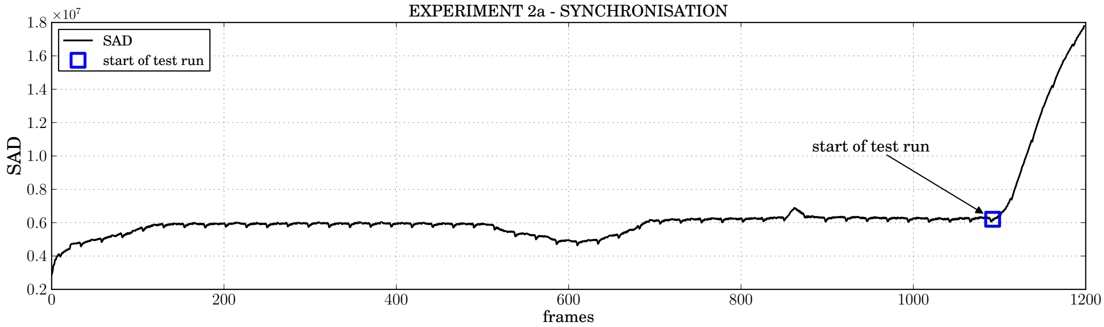


Figure 5.16.: Shows the result of the image differencing algorithm for Experiement 2a. The start of the test run is easily identified at frame 1092 (blue square).

The high values and the regular pattern visible in the SAD signal arise from the data compression in the images. The local maxima and minima are most probably due to small adjustments by the test rig's control system before the start of the actual test run.

Because all the experiments on the test rig were conducted unloaded (i.e. no bogie present), for security reasons the simulated motions were carried out very slow. Therefore, it is not reasonable to evaluate the acquired data at its maximum sampling frequency of 50 Hz. In order to simulate a train ride at normal speed, it is decided to **evaluate the acquired measurement data at only 1Hz (1fps)**. This also has the advantage of greatly reducing the amount of data and processing time during various experiments and evaluations.

### 5.4.3. Experiment 2a

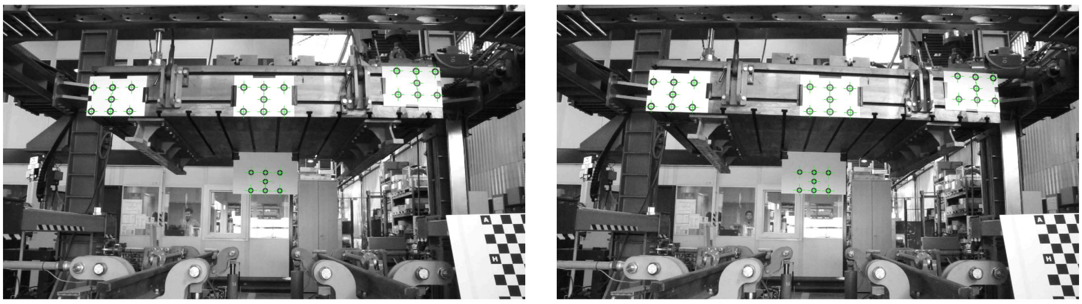
For this experiment, in total four ISO A3 7-CCC targets were attached to the platform (three coplanar in the front and one at the bottom back). The whole test run lasted about 17 minutes (exact: 986 sec = 986 measurement points) in which the platform simulated motions into all 6-DoF (illustrated in Figures 5.17 and 5.22). At the beginning of the sequence each DoF was tested separately, describing the trajectory by the following sequence: $\Delta$Z,$\Delta$X,$\Delta$Y,$\Delta\phi$,$\Delta\psi$,$\Delta\theta$ ($I_0 - I_{711}$). Afterwards a coupled oscillation (i.e. combined DoF $\Delta$Y + $\Delta\phi$) called swaying was tested ($I_{712} - I_{986}$). All available $28 = 7 \times 4$ fiducials are used to estimate the pose in this experiment.

The maximum displacements carried out by the platform in Experiment 2a are:
$\Delta$X: ±100 mm, $\Delta$Y: ±120 mm, $\Delta$Z: $-200/+150$ mm
$\Delta\phi$: ±5 °, $\Delta\theta$: ±5 °, $\Delta\psi$: ±6 °

Although the test rig was not able to simulate the maximum range of expected motions as specified in Section 1.2, it is still legitimate to evaluate the basic feasibility of the proposed measurement system.

The specific measurement setup is subject to the following parameters:

- Distance to platform origin: 4601 mm (at $I_0$)
- Image size: 1920 × 1080 pixel$^2$
- Focal length ($f$): 18 mm
- Aperture value: $f/22$
- Exposure time: 1/50 sec



(a) $I_0$ (b) $I_{563}$

Figure 5.17.: (a) shows frame recorded at beginning ($I_0$) and (b) during maximum angular yaw ($\Delta\psi$) displacement ($I_{563}$) performed in Experiment 2a. The tracking results are superimposed with green crosses.

As in Experiment 1, the performed relative motions are illustrated by their single DoF trajectories (REF) and the calculated pose from DLT and EPnP in Figures 5.18 and 5.20. For a better comparison, the figures with the occurring deviations of the both pose estimation algorithms are depicted directly beneath the corresponding displacements (see Figures 5.19 and 5.21).

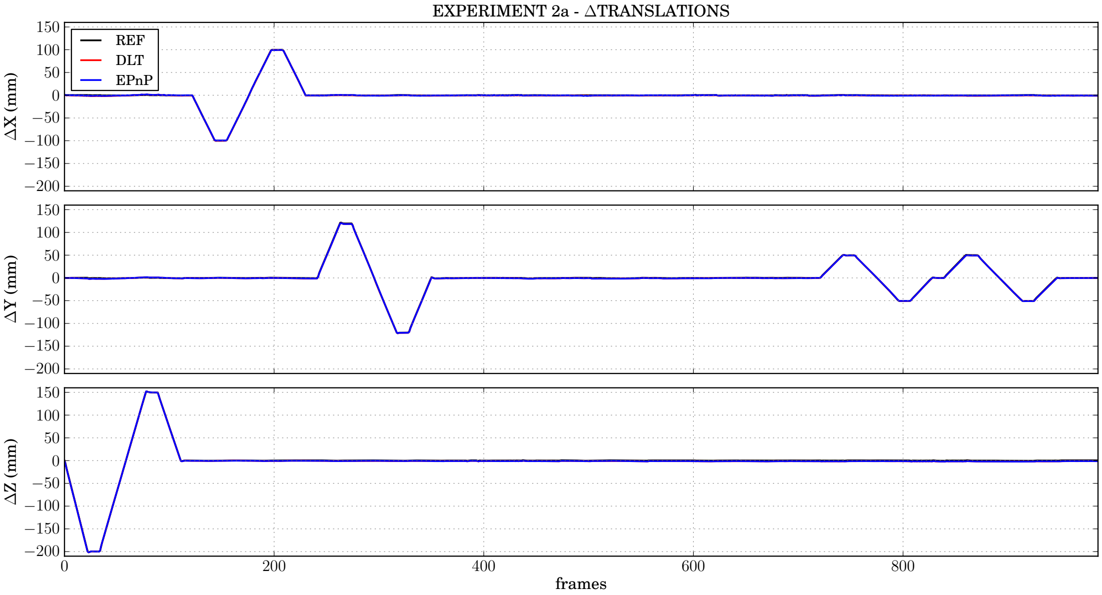


Figure 5.18.: Relative translations performed by platform in Experiment 2a represented by the individual axes {ΔX,ΔY,ΔZ}.

Figure 5.19.: Deviations of the {ΔX,ΔY,ΔZ} signals for both DLT and EPnP from the reference data. Step response like deviations result from erroneous synchronisation. Linear deviations indicate misalignment of the reference frames or imprecise camera calibration.

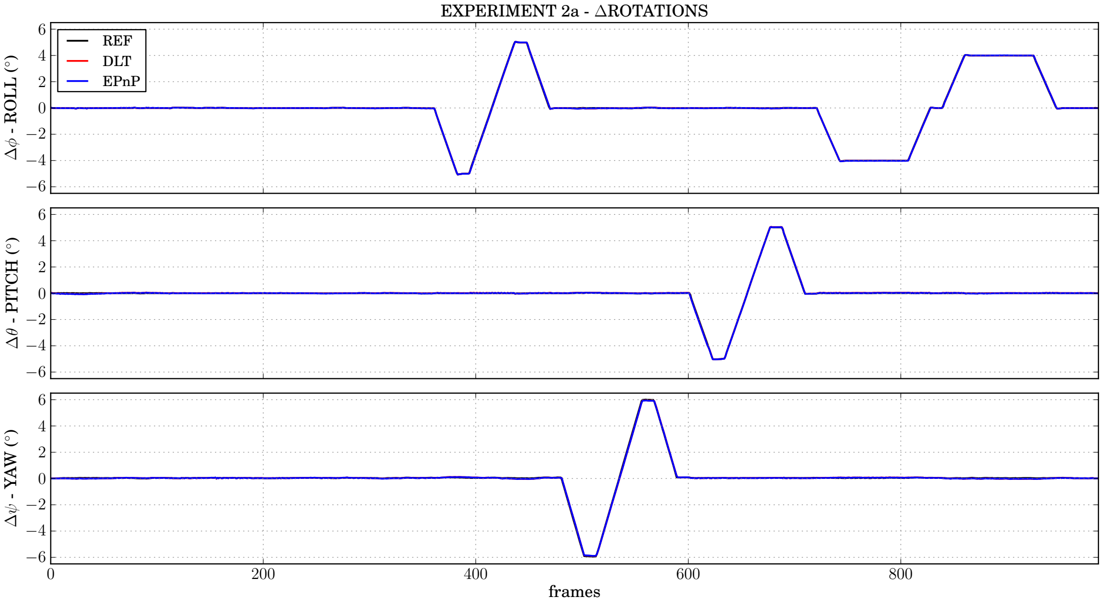


Figure 5.20.: Relative rotations performed by platform in Experiment 2a represented by the individual angles of rotations $\{\Delta\phi,\, \Delta\theta,\, \Delta\psi\}$.

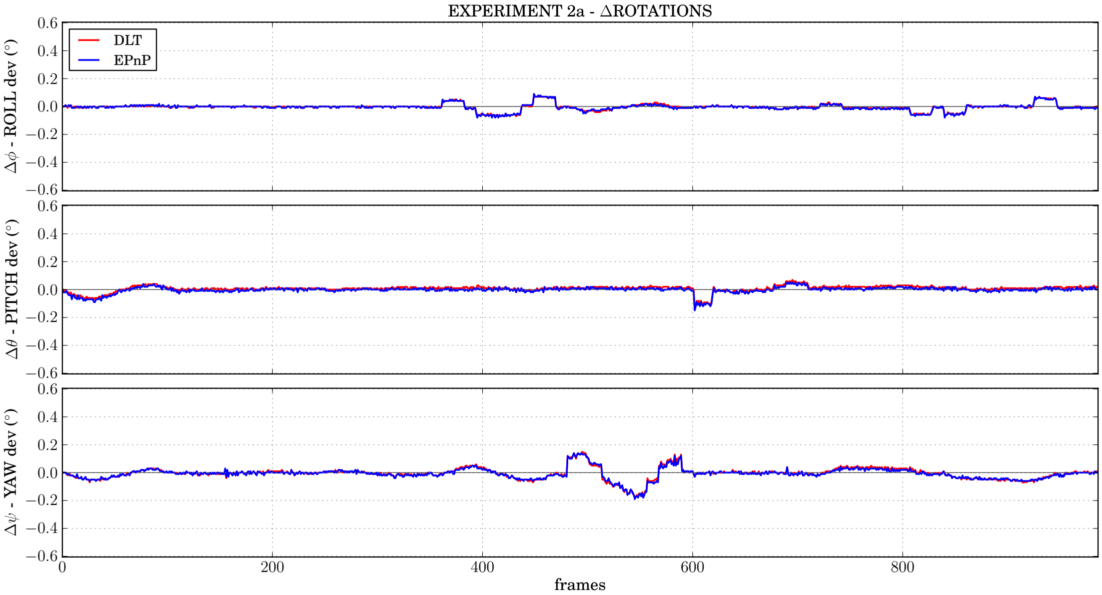


Figure 5.21.: Deviations of the $\{\Delta\phi,\, \Delta\theta,\, \Delta\psi\}$ signals for both DLT and EPnP from the reference data. Step response like deviations result from erroneous synchronisation. Linear deviations indicate misalignment of the reference frames or imprecise camera calibration.

From Figures 5.18 and 5.20 the performed trajectories of the relative motions are clearly observable. As only the blue line from EPnP is visible in the figures (it is plotted last), it is evident that the obtained deviations of both DLT and EPnP are generally very small compared to the performed linear and angular displacements.

The achieved results are only true when all available 28 fiducials are applied in the pose estimation process. Evaluation 2a (Section 5.4.6) examines the system's behaviour using different numbers of fiducials and evaluates the performance of DLT and EPnP under these conditions.

The deviations visible in Figures 5.19 and 5.21 show that these are not normally distributed but rather exhibits a systematic bias. The step response like deviations (e.g. in intervals ΔZ dev ($I_0 - I_{112}$), $\Delta\phi$ dev ($I_{360} - I_{470}$)), suggest some kind of constant offset and therefore strongly indicate that these deviations are due to inaccurate synchronisation. Whereas, in the same interval of the sequence, the linear deviations depicted in {ΔX dev, ΔY dev, $\Delta\theta$ dev, $\Delta\psi$ dev} are likely to arise from a small misalignment of the coordinate systems because of either inaccuracies in mounting of the targets or a poorly determined initial pose. Another reason for these deviations in measurements can be an insufficiently accurate camera calibration. This is not very likely because the deviations concerned are of linear order.

At this place, problems with the pose estimation algorithms itself may be excluded insofar as the results of DLT and EPnP only differ to a small extent (e.g. $\max(\Delta X_{DLT} - \Delta X_{EPnP}) \leq 1.2$ mm).

The slowly but constantly increasing deviation of both DLT and EPnP in the ΔZ dev signal is explained by the fact that, in contrast to the image acquisition system, the base of the test rig is mounted on air bearings to avoid external influence on the measurements. Obviously these air bearings change their state in the course of the test run. This type of relative motion is not covered by the reference data.

Table 5.4 presents the uncertainty of the results achieved in Experiment 2a. The maximum deviations for all DoF are very well within the defined limits of R(a).

| **EX 2a** | **MAX** | **RMSE** | **MEAN** | **STD** |
|---|---|---|---|---|
| (mm) | ΔX / ΔY / ΔZ | ΔX / ΔY / ΔZ | ΔX / ΔY / ΔZ | ΔX / ΔY / ΔZ |
| **DLT** | 2.0 / 1.8 / 2.2 | 0.7 / 0.8 / 1.4 | -0.5 / -0.6 / -1.3 | 0.4 / 0.5 / 0.6 |
| **EPnP** | 1.7 / 1.8 / 2.2 | 0.5 / 0.8 / 1.4 | -0.3 / -0.5 / -1.2 | 0.4 / 0.5 / 0.6 |
| (°) | $\Delta\phi$ / $\Delta\theta$ / $\Delta\psi$ | $\Delta\phi$ / $\Delta\theta$ / $\Delta\psi$ | $\Delta\phi$ / $\Delta\theta$ / $\Delta\psi$ | $\Delta\phi$ / $\Delta\theta$ / $\Delta\psi$ |
| **DLT** | 0.09 / 0.14 / 0.18 | 0.02 / 0.02 / 0.04 | 0.00 / 0.01 / -0.01 | 0.02 / 0.02 / 0.04 |
| **EPnP** | 0.09 / 0.14 / 0.19 | 0.02 / 0.02 / 0.04 | -0.01 / 0.00 / -0.01 | 0.02 / 0.02 / 0.04 |

Table 5.4.: Uncertainty of measurements for Experiment 2a.

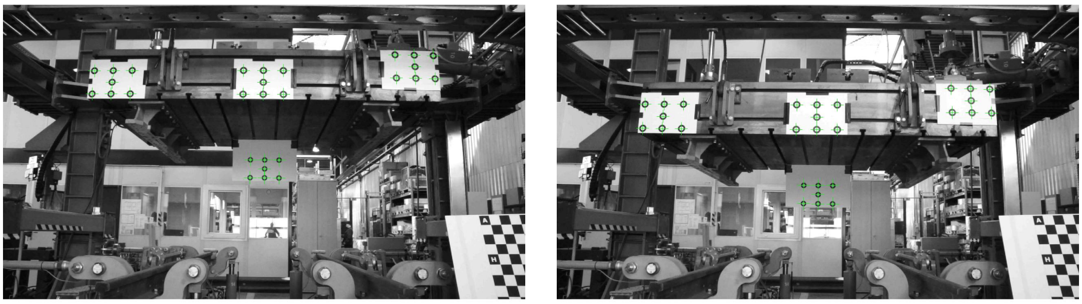
### 5.4.4. Experiment 2b

The measurement setup for Experiment 2b was identical to the one of Experiment 2a, except for the different maximum displacements carried out by the platform. The recorded test run lasted about 1015 sec yielding 1015 measurement points. Again, all available $28 = 7 \times 4$ fiducials are used to estimate the pose in this experiment.

The maximum displacements (different from Experiment 2a) carried out by the platform in Experiment 2b are:
$\Delta$X: ±70 mm, $\Delta$Y: ±150 mm, $\Delta$Z: ±150 mm
$\Delta\phi$: ±3 °, $\Delta\theta$: ±3 °, $\Delta\psi$: ±5 °

The measurement setup is subject to the following parameters (same as in Experiment 2a):

- Distance to platform origin: 4601 mm (at $I_0$)
- Image size: $1920 \times 1080$ pixel$^2$
- Focal length ($f$): 18 mm
- Aperture value: $f/22$
- Exposure time: 1/50 sec



(a) $I_{30}$ (b) $I_{85}$

Figure 5.22.: (a) and (b) show frames recorded at the corresponding maximum displacements during z-translation ($I_{30}$ and $I_{85}$) performed in Experiment 2b. The tracking results are superimposed with green crosses.

Because in Experiment 2b the platform performed the same trajectory as in Experiment 2a, only the figures illustrating the deviations from the reference data are presented. See Figures 5.23 and 5.24 on the next page.

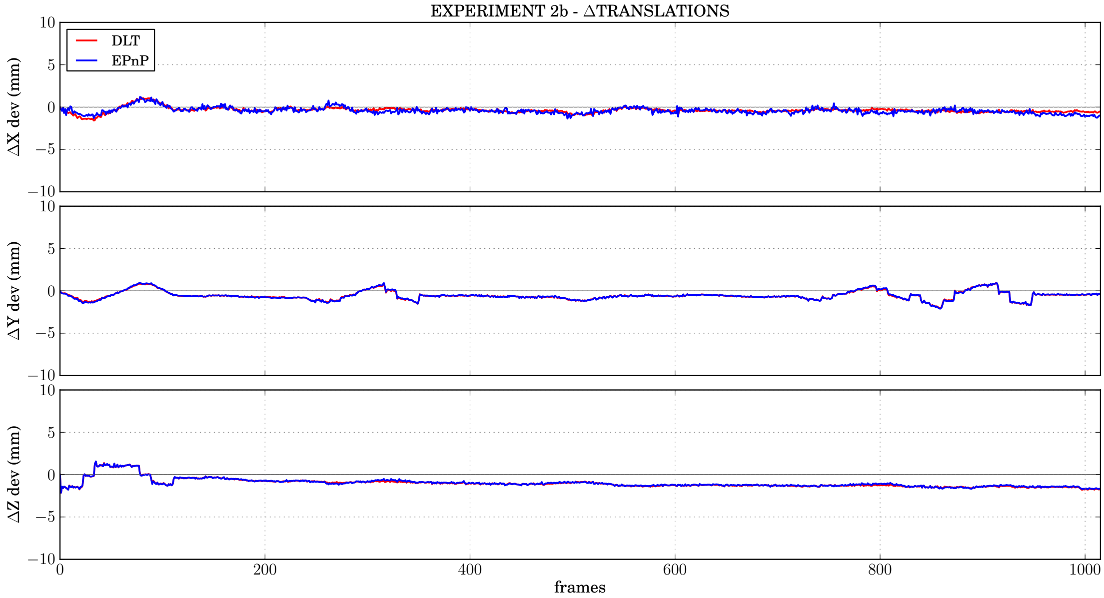


Figure 5.23.: Deviations of the {ΔX,ΔY,ΔZ} signals for both DLT and EPnP from the reference data in Experiment 2b.

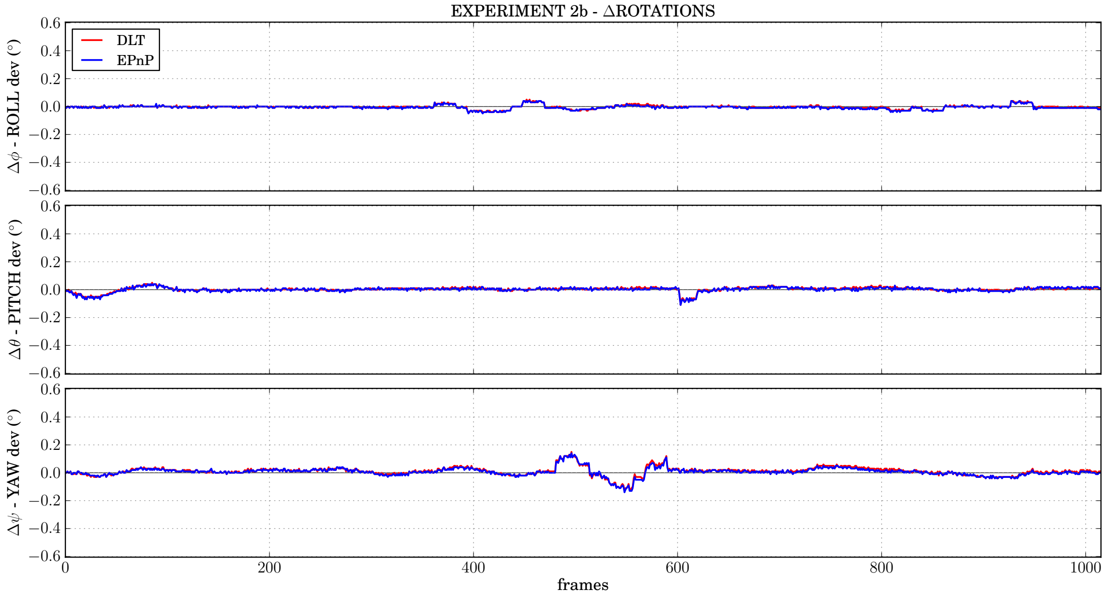


Figure 5.24.: Deviations of the {$\Delta\phi$, $\Delta\theta$, $\Delta\psi$} signals for both DLT and EPnP from the reference data in Experiment 2b.

From the deviations depicted in Figures 5.23 and 5.24, it is obvious the calculation of the synchronisation for Experiment 2b was more precise since the step response like deviations are smaller. The rest of the depicted deviations is very similar to the results achieved in Experiment 2a. This strongly indicates the exact repeatability of results gained from different measurements. Table 5.5 presents the evaluation of uncertainty of measurement for Experiment 2b.

| **EX 2b** | **MAX** | **RMSE** | **MEAN** | **STD** |
|---|---|---|---|---|
| (mm) | $\Delta$X / $\Delta$Y / $\Delta$Z | $\Delta$X / $\Delta$Y / $\Delta$Z | $\Delta$X / $\Delta$Y / $\Delta$Z | $\Delta$X / $\Delta$Y / $\Delta$Z |
| **DLT** | 1.6 / 2.0 / 2.1 | 0.5 / 0.7 / 1.2 | -0.4 / -0.6 / -1.0 | 0.3 / 0.5 / 0.6 |
| **EPnP** | 1.4 / 2.1 / 2.1 | 0.6 / 0.8 / 1.1 | -0.4 / -0.6 / -1.0 | 0.4 / 0.5 / 0.6 |
| (°) | $\Delta\phi$ / $\Delta\theta$ / $\Delta\psi$ | $\Delta\phi$ / $\Delta\theta$ / $\Delta\psi$ | $\Delta\phi$ / $\Delta\theta$ / $\Delta\psi$ | $\Delta\phi$ / $\Delta\theta$ / $\Delta\psi$ |
| **DLT** | 0.05 / 0.11 / 0.14 | 0.01 / 0.02 / 0.03 | 0.00 / 0.00 / 0.01 | 0.01 / 0.02 / 0.03 |
| **EPnP** | 0.05 / 0.11 / 0.14 | 0.01 / 0.02 / 0.03 | -0.01 / 0.00 / 0.01 | 0.01 / 0.02 / 0.03 |

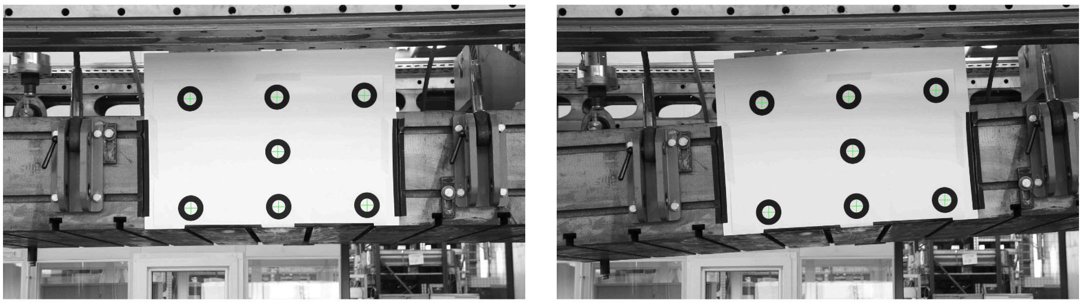
Table 5.5.: Uncertainty of measurements for Experiment 2b.

### 5.4.5. Experiment 2c

The measurement setup for Experiment 2c differs from the previous experiments by the application of only one but therefore larger ISO A1-sized 7-CCC target and the change towards a bigger focal length in order to further increase the appearance of the fiducials. The recorded test run yields 965 measurement points with the platform performing the identical trajectory as for Experiment 2b.

The measurement setup is subject to the following parameters:

- Distance to platform origin: 5665 mm (at $I_0$)
- Image size: 1920 × 1080 pixel$^2$
- Focal length ($f$): 55 mm
- Aperture value: $f/5.6$
- Exposure time: 1/50 sec



(a) $I_0$ (b) $I_{870}$

Figure 5.25.: (a) and (b) show frames recorded at the beginning of the sequence ($I_0$) and during the simulated coupled-oscillation (swaying) at $I_{870}$. The tracking results are superimposed with green crosses.

Because in Experiment 2c the platform performed the same trajectory as in Experiment 2b, only the figures illustrating the deviations from the reference data are presented. See Figures 5.26 and 5.27 on the next page.

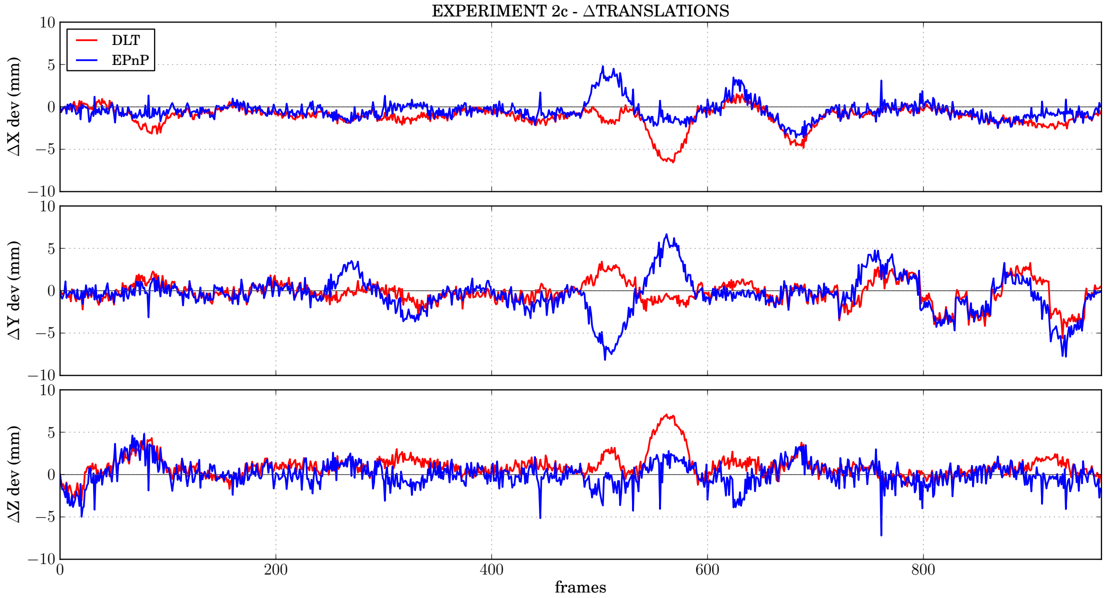


Figure 5.26.: Deviations of the {ΔX,ΔY,ΔZ} signals for both DLT and EPnP from the reference data in Experiment 2c.

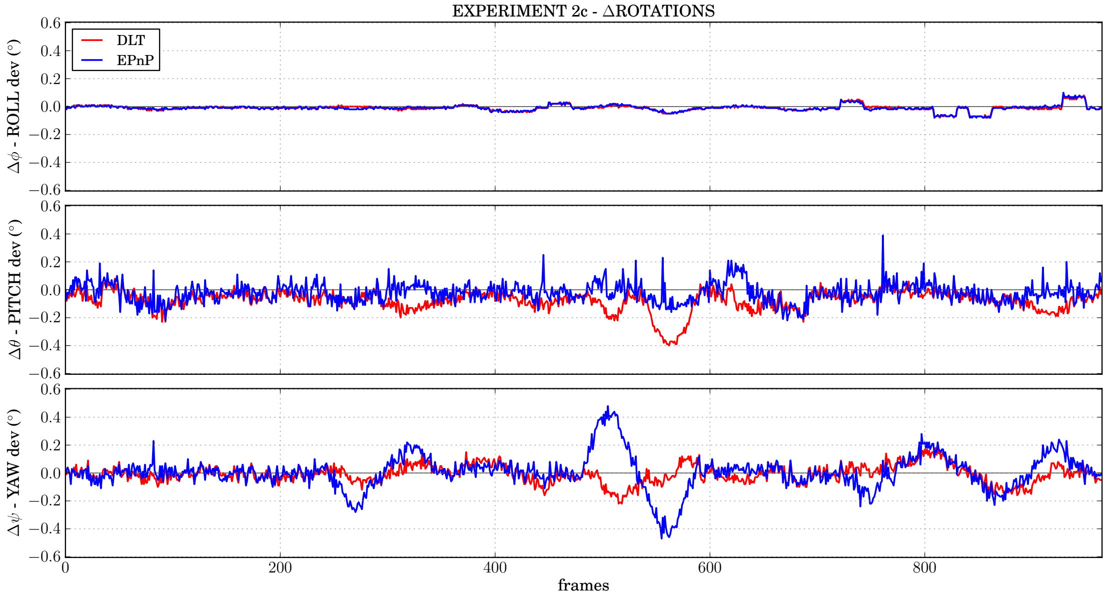


Figure 5.27.: Deviations of the $\{\Delta\phi,\ \Delta\theta,\ \Delta\psi\}$ signals for both DLT and EPnP from the reference data in Experiment 2c.

As shown in Figures 5.26 and 5.27, it is reasonable that the pose calculations are subject to larger deviations and higher variance when using only seven compared to the 28 fiducials in the previous experiments.

Remarkable is the fact that for only seven CCC fiducials the EPnP algorithm exhibits much larger deviations than DLT. An explanation for this effect is the application of the EPnP without the iterative refinement step (planar target; discussed in Experiment 1b (Section 5.3.3)).

Furthermore it is apparent that the largest systematic deviations in the interval $I_{480} - I_{590}$ (during yaw-rotation) are visible for both algorithms which even return distinct results. Probable reasons for that may be an inaccurate calibration, problems during the tracking process due to changing lighting conditions or the occurrence of a critical configuration (for the latter see discussion in Experiment 1c (Section 5.3.4)).

Table 5.6 compares the uncertainty of the DLT pose estimation results of Experiment 2c (ISO A1 sized) with the ones of Experiment 2b (when using the centred ISO A3 target to calculate the pose). It shows that the application of larger fiducials does not significantly contribute to a smaller measurement uncertainty.

| **EX2b / EX2c** | **MAX** | **RMSE** | **MEAN** | **STD** |
|---|---|---|---|---|
| (mm) | $\Delta X$ / $\Delta Y$ / $\Delta Z$ | $\Delta X$ / $\Delta Y$ / $\Delta Z$ | $\Delta X$ / $\Delta Y$ / $\Delta Z$ | $\Delta X$ / $\Delta Y$ / $\Delta Z$ |
| **DLT**(EX 2b) | 5.1 / 6.0 / 6.6 | 1.4 / 1.6 / 2.2 | -0.1 / -0.7 / -0.9 | 1.4 / 1.4 / 2.0 |
| **DLT**(EX 2c) | 6.6 / 5.6 / 7.1 | 1.7 / 1.4 / 1.7 | -1.1 / -0.2 / 1.0 | 1.2 / 1.4 / 1.4 |
| (°) | $\Delta\phi$ / $\Delta\theta$ / $\Delta\psi$ | $\Delta\phi$ / $\Delta\theta$ / $\Delta\psi$ | $\Delta\phi$ / $\Delta\theta$ / $\Delta\psi$ | $\Delta\phi$ / $\Delta\theta$ / $\Delta\psi$ |
| **DLT**(EX 2b) | 0.11 / 0.38 / 0.32 | 0.03 / 0.10 / 0.08 | -0.02 / -0.01 / 0.01 | 0.03 / 0.10 / 0.08 |
| **DLT**(EX 2c) | 0.09 / 0.40 / 0.22 | 0.02 / 0.11 / 0.07 | -0.01 / -0.08 / 0.01 | 0.02 / 0.07 / 0.06 |

Table 5.6.: Comparison of uncertainties for Experiment 2b with Experiment 2c. Although fiducials have much larger appearance, the results of Experiment 2c do not show any significant improvement in measurement accuracy. The results of the EPnP algorithm are omitted.

From the results in Table 5.6, it is concluded that it is not the size but that rather the amount of applied fiducials greatly impacts the measurement accuracy. This is in good accordance with [53] which states that "*larger targets give similar, but not better results*". Thus, the question arises about a suitable number of fiducials with which the measurement system meets the required accuracy. This important property of the measurement setup is investigated in the next section (Evaluation 2a).

### 5.4.6. Evaluation 2a

From the discussion in Section 3.3.5 and with the results of Experiments 2a and 2b it is shown that a suitable number of fiducials applied in measurements lies in the range of four and the maximum of 28 divided into four 7-CCC targets.

The purpose of this evaluation is to examine the system's accuracy with different numbers of fiducials applied to estimate the pose using the DLT and the EPnP algorithm. For this purpose, a numbering scheme (depicted in Figure 5.28) is related to the targets and fiducials mounted on the platform. Each fiducial receives an unique identifier which corresponds to the particular number in the increasing order it is used for the pose calculations. The number of an individual fiducial ($F_K$) can be calculated by:

$$F_K = (7 \times (\mathrm{T_N} - 1)) + \mathrm{O_M} \tag{5.2}$$

with $\mathrm{T_N}$ being the target number and $\mathrm{O_M}$ the order of the fiducials on the target.

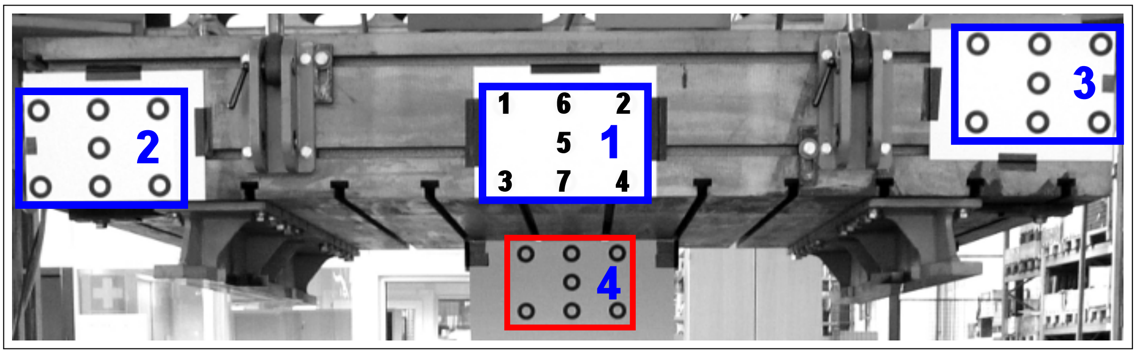


Figure 5.28.: Numbering scheme used in evaluation. Target numbers ($T_N$) are depicted in blue. Black digits ($O_M$) superimposed on target 1 exemplify the order of fiducials per target. The red rectangle indicates that target 4 is displaced into the back of the platform.

**For this evaluation, the measurement setup, parameters and data of Experiment 2a is used.** The pose is estimated the same way by both DLT and EPnP but this time the number of fiducials is continuously increased (following the proposed numbering scheme) while monitoring the development of the MAX and RMSE deviations.

As the expectation is that the rotational deviations show a similar behaviour, only the results for the translational deviations are illustrated in Figure 5.29.

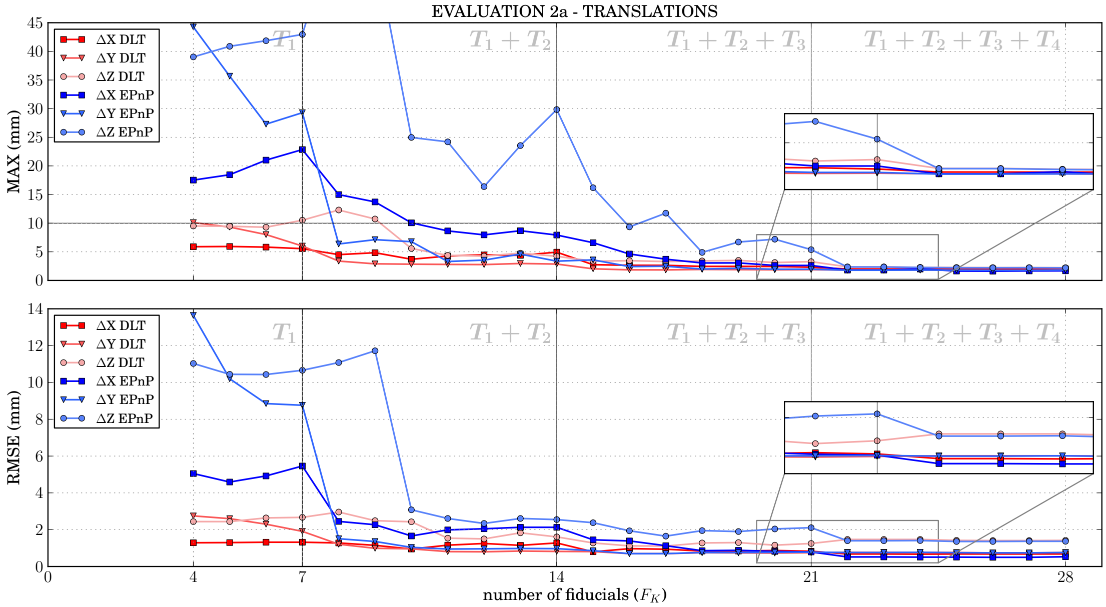


Figure 5.29.: MAX and RMSE deviations of the {ΔX,ΔY,ΔZ} signals for both DLT and EPnP from the reference data when applying an increasing number of fiducials. Intervals with fiducials belonging to one target are denoted explicitly. EPnP deviations are generally higher but with slightly better results when using fiducials on all four targets.

Figure 5.29 depicts the development of the MAX and RMSE translational deviations for DLT and EPnP algorithms while applying an increasing number (4–28) of fiducials. The target numbers ($T_N$) are displayed within the corresponding intervals.

Generally it is visible that, the results of the EPnP algorithm exhibit a higher uncertainty until using 22 fiducials on all four targets, from which onwards the deviations approximately coincide with the ones of the DLT algorithm. For the DLT it shows that when using 2 targets with 11 coplanar points ($F_{11}$) the deviation drops below the required 10 mm maximum deviation (R(a), denoted with a thin black line). The EPnP algorithm fulfils the requirement when using 3 coplanar targets and 18 fiducials.

Furthermore the results show that the uncertainty of both DLT and EPnP still decreases after adding the first fiducial from the spatially displaced target ($T_4$). This fact is magnified by the two zoom-boxes in Figure 5.29. Obviously, the inclusion of information from non-coplanar points reduces the measurement uncertainty. This effect is investigated in the next section (Evaluation 2b).

### 5.4.7. Evaluation 2b

The goal of this evaluation is to provide information about the importance of the application of non-coplanar points for this specific pose estimation task. Therefore the evaluation is conducted using the same preconditions as in Evaluation 2a. Except the numbering scheme is adapted by changing the order of targets evaluated from $T_1, T_2, T_3, T_4 \ \rightarrow \ T_1, \mathbf{T_4}, T_3, T_2$.

The results of this evaluation when the spatially displaced target $T_4$ is applied earlier in the sequence of evaluation are illsutrated in Figure 5.30.

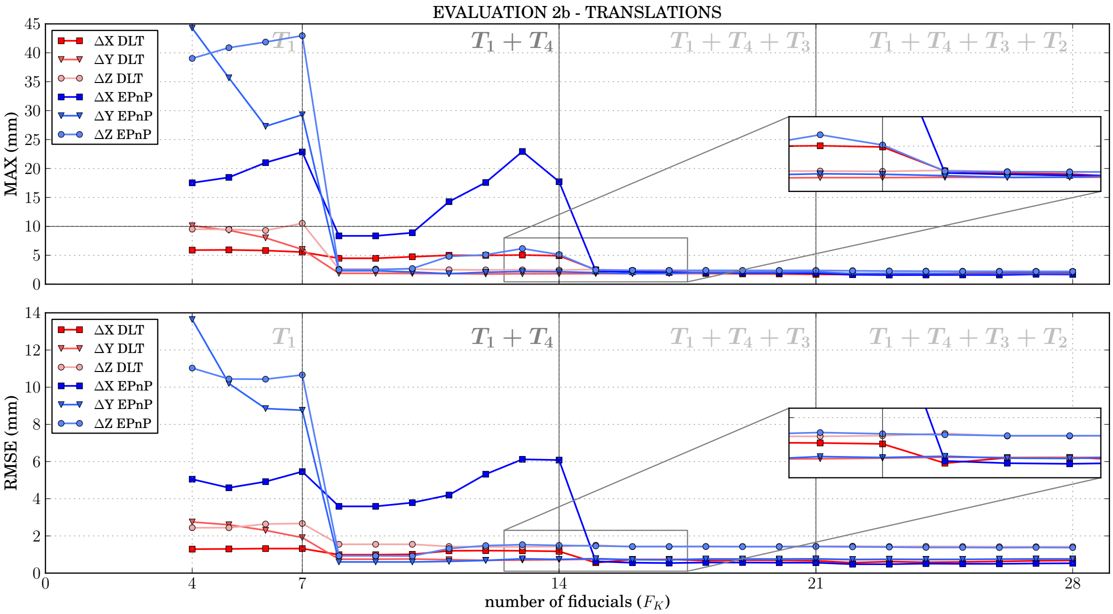


Figure 5.30.: MAX and RMSE deviations of the {ΔX,ΔY,ΔZ} signals for both DLT and EPnP from the reference data when applying an increasing number of fiducials. Interval of $T_2$ is exchanged with $T_4$ to show behaviour when applying non-coplanar points.

From Figure 5.30 it is visible that the accuracy is dramatically increased when applying information from non-coplanar instead of coplanar points (as depicted in Figure 5.29). With the exception of the $\Delta X_{\mathrm{EPnP}}$ signal, all deviations drop to a more or less constant value. When adding further information of $T_3$ at $F_{15}$ only slight improvements in accuracy are measurable (see the zoom-boxes in Figure 5.30).

These findings indicate that it is important to integrate depth information from non-coplanar points into the measurement setup. In order to reach a reasonable measurement accuracy it proves sufficient to apply three 5-CCC targets (introduced in Figure 3.4a), with at least one which is spatially displaced.

These findings are confirmed by the results of the uncertainty evaluation of the DLT algorithm listed in Table 5.7. It provides the evidence that the pose calculations from three 5-CCC targets (simulated using $\{T_1, T_2, T_4\} \setminus \{O_6, O_7\}$) with 15 fiducials can be done at virtually the same accuracy as with four 7-CCC targets using 28 fiducials.

| **EX2a** | **MAX** | **RMSE** | **MEAN** | **STD** |
|---|---|---|---|---|
| (mm) | ΔX / ΔY / ΔZ | ΔX / ΔY / ΔZ | ΔX / ΔY / ΔZ | ΔX / ΔY / ΔZ |
| **DLT**(4×7-CCC) | 2.0 / 1.8 / 2.2 | 0.7 / 0.8 / 1.4 | -0.5 / -0.6 / -1.3 | 0.4 / 0.5 / 0.6 |
| **DLT**(3×5-CCC) | 1.9 / 1.8 / 2.4 | 0.7 / 0.7 / 1.4 | -0.5 / -0.5 / -1.3 | 0.4 / 0.5 / 0.6 |
| (°) | $\Delta\phi$ / $\Delta\theta$ / $\Delta\psi$ | $\Delta\phi$ / $\Delta\theta$ / $\Delta\psi$ | $\Delta\phi$ / $\Delta\theta$ / $\Delta\psi$ | $\Delta\phi$ / $\Delta\theta$ / $\Delta\psi$ |
| **DLT**(4×7-CCC) | 0.09 / 0.14 / 0.18 | 0.02 / 0.02 / 0.04 | 0.00 / 0.01 / -0.01 | 0.02 / 0.02 / 0.04 |
| **DLT**(3×5-CCC) | 0.09 / 0.14 / 0.19 | 0.03 / 0.02 / 0.04 | -0.01 / 0.00 / -0.01 | 0.02 / 0.02 / 0.04 |

Table 5.7.: Comparison of uncertainty of DLT using 4×7-CCC and 3×5-CCC. It shows that the pose calculations can be done at virtually the same accuracy. The results of the EPnP algorithm are omitted.

The presentation of the results of the EPnP algorithm is omitted in Table 5.7, since the results of Evaluation 2a and 2b clearly show a superior performance of the DLT algorithm when applying fewer ($\leq F_{15}$) fiducials for pose calculation. Therefore, depending on the situation in a real world environment, it is proposed to apply DLT or EPnP according to the number of fiducials present in the measurement setup.

It is concluded that the results presented in Experiment 2 prove the basic feasibility of the proposed image based measurement system. Furthermore, it is shown that a measurement setup using three 5-CCC targets and 3D input data is sufficient in terms of measurement uncertainty, overachieving the defined requirements. Moreover, the evaluations in Experiment 2 show that it is seems reasonable to apply either DLT or EPnP for pose estimation depending on the number of fiducials present in the scene.

## 5.5. Evaluation of a Test Ride on a High-Speed Train (Velaro D)

In order to verify and test the feasibility of the proposed image-based measurement system, a test ride was conducted on board the latest high speed train manufactured by Siemens for Deutsche Bahn AG: Siemens Velaro D (see Figure 1.1 and the technical data sheet [79]).

The test ride was performed on a branch line serving as a curved test track which is situated at the *Test and Validation centre, Wegberg - Wildenrath* [78]. This curved test track consists of several successive reverse curves with a given radius of curvature of 150 m. It was traversed by the Siemens Velaro D at a constant but low speed ($\leq$ 10 km/h) while measuring the occurring relative motions between two selected carbodies (pictured in Figure 5.31).

(a)

(b)

Figure 5.31.: (a) shows an outside view of the two selected carbodies of the Siemens Velaro D which are linked by a flexible corridor connection. (b) pictures the carbodies during relative displacement into y-direction.

### 5.5.1. Measurement Setup

The basic measurement setup for the test ride complies with the one proposed in Section 3.4 and illustrated in Figure 3.13. The camera was mounted on a heavy-weight tripod and the targets were attached by using regular adhesive tape. The pose is estimated over time between the rigidly attached camera in `CB_01` and the targets installed in `CB_02`. This enables the system to measure the sought after relative motions.

Altogether, four traversals of the train through the curved test track were conducted. From the corresponding four acquired sequences only the first is chosen for evaluation. This specific test ride lasts about 4545 sec. In order to simulate a train ride at higher speed and to reduce the computational efforts, it is decided to **evaluate the acquired imagery data at only 1Hz (1fps)**. Thus, this yields a total of 4545 measurement points.

The measurement setup is subject to the following parameters:

- Distance to origin of carbody (`CB_02`): 1652 mm (at $I_0$)
- Image size: 1920 × 1080 pixel$^2$
- Focal length ($f$): 18 mm
- Aperture value: $f/11$
- Exposure time: 1/50 sec

(a) $I_0$ (b) $I_{3810}$

Figure 5.32.: (a) and (b) show frames with the carbodies in initial pose ($I_0$) and at maximum displacement into y-direction at $I_{3810}$ (compare to Figure 5.31b). Note the partial occlusion of $T_2$ by the gaiter of corridor connection. The tracking results are superimposed as green crosses.

As proposed by the results of Evaluation 2b (Section 5.4.7) three multi-fiducial targets (one 5-CCC ($T_1$) and two 7-CCC ($T_2$ left and $T_3$ right)) were installed (see Figure 5.32). All of them being installed in a non-coplanar way. Since, due to partial occlusions, the outmost fiducials on $T_2$ and $T_3$ with numbers $F_6, F_8, F_{14}, F_{16}$ are not considered, a total of 15 fiducials is used in the pose estimation process.

### 5.5.2. Measurement Results

The results of the measurements from the test ride are presented in Figures 5.33 and 5.34. Unfortunately **no reference data is available** for the test ride. Therefore only the results of the DLT pose estimation is given.

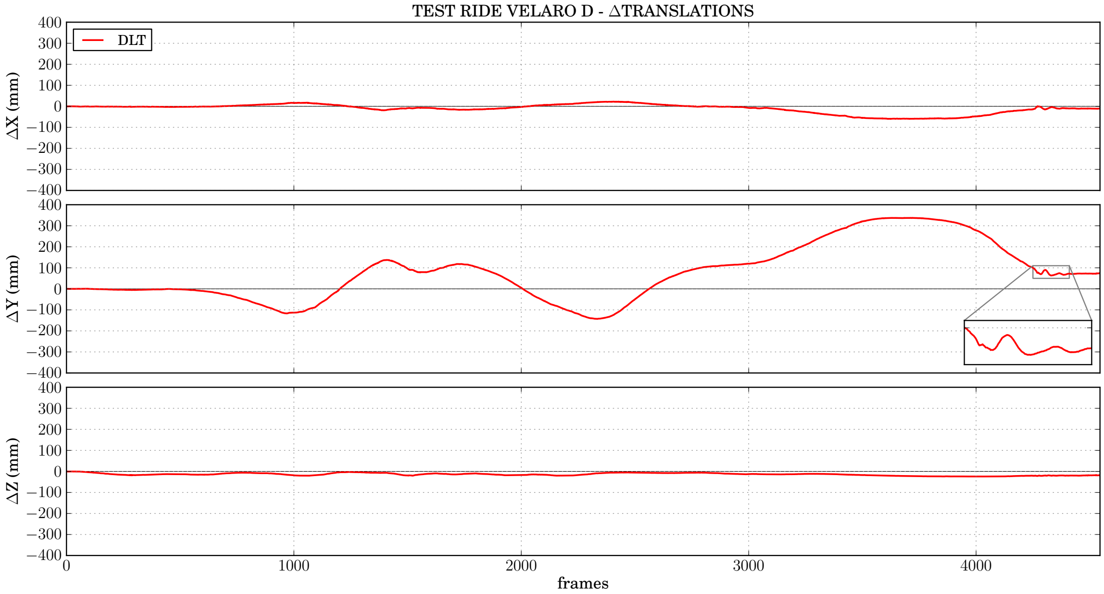


Figure 5.33.: Relative translations of the carbodies measured during the test ride represented by their individual axes {ΔX,ΔY,ΔZ}. The jerk at the end of the test ride is magnified.

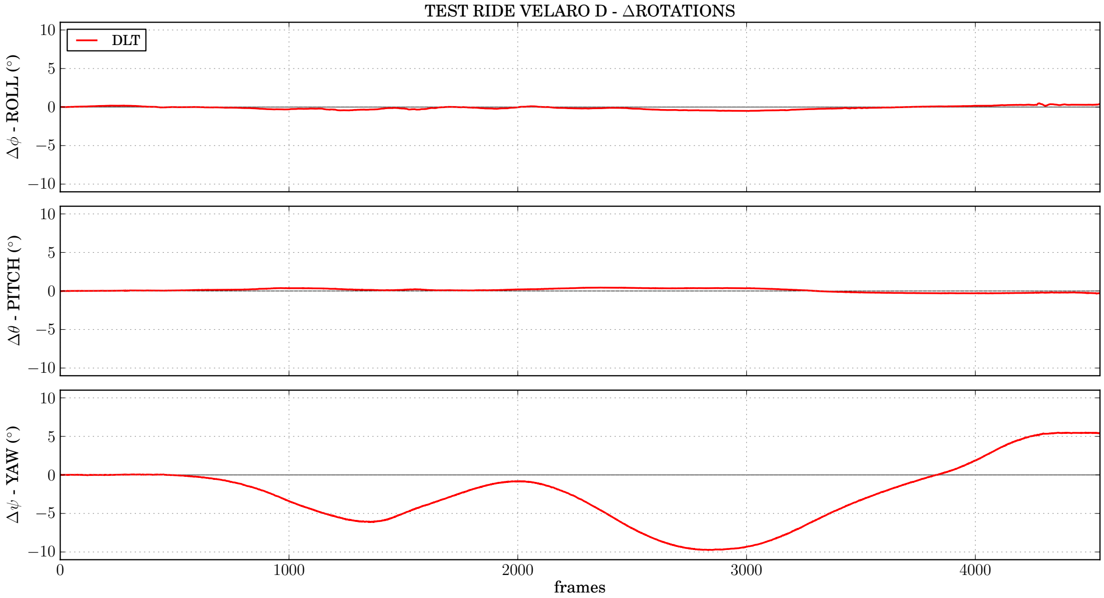


Figure 5.34.: Relative rotations of the carbodies measured during the test ride represented by their individual axes $\{\Delta\phi, \Delta\theta, \Delta\psi\}$.

Even without suitable reference data, the following statements about the measurements are possible:

- First, as expected on the plain test track, more or less no lateral or angular displacements into $\Delta$Z, $\Delta\phi$, and $\Delta\theta$ directions are observed.
- Second, the main displacements measured are into $\Delta$Y and $\Delta\psi$ direction, which relates to a hunting oscillation (see Figure 3.12). This motion is characteristic for the traversal of reverse curves. The braking process at the end of the sequence ($I_{3100} - I_{4400}$) is visible in $\Delta$X signal, indicating that both carbodies approach each other.
- Third, the end of the test ride is clearly visible by the measured jerk of the carbodies as the result of the small sudden change in acceleration when the train finally comes to a halt. It marks the actual end of the test ride. This event is magnified by the zoom-box of the $\Delta$Y signal in Figure 5.33.

All these considerations basically indicate the plausibility of the measurements. It can be considered that the uncertainties of the obtained results are in the same order of magnitude than for the experiments on the test rig in Evaluation 2b (Table 5.7). Furthermore, the results of the test ride and the prototypical realisation prove the technical feasibility and suitability of the proposed image-based measurement system for measuring the relative motions between railway vehicle carbodies.

## 5.6. Computation Time

For the evaluation of the average processing time of one frame, the actual time needed for image acquisition and storage is omitted. As the MATLAB prototype implementation is divided into separate, sequentially processed tracking and pose estimation parts, the evaluation of performance is given correspondingly.

For this evaluation the same data as used for Evaluation 2b (see Section 5.4.7) is applied and executed on a standard 1.85 GHz Dual Core laptop. The results are presented in Table 5.8.

| **Computation Time EV2a** | **TRACKING** | **POSE ESTIMATION** |
|---|---|---|
| **DLT**(3×5-CCC) | 0.125 s (∼ 8 fps) | 0.0062 s (∼ 161 fps) |
| **EPnP**(3×5-CCC) | | 0.0049 s (∼ 204 fps) |

Table 5.8.: Average processing time for one frame by the prototype implementation, given for the tracking and the pose estimation parts separately.

It shows that, the tracking procedure is by far the most computationally complex part in the system as it has to process the image data (size: 1920 × 1080 pixel$^2$). Once the tracking is complete the pose calculations can be done very fast, as expected with slight advantages for the EPnP algorithm. Altogether, the prototype implementation is not capable of processing the data in real time but enables the possibility to partially analyse the measurements on-site, for example in the course of a test ride.

## 5.7. Best Practice Approach

This section provides a summary of the knowledge and experience gained throughout the experiments and the accomplished test ride. The findings are presented in the course of a best practice approach.

1. **Preparations prior to measurement**
    - Print 5-CCC targets on 100 g heavy, white paper and stick to suitably-sized KAPA boards (see Section 5.4.1) trying to achieve an as level and smooth surface as possible.
    - Prepare a planar, chequerboard pattern, camera calibration target (see Figure 4.4).
    - For installation of targets bring tape, water level, laser range finder or other suitable devices.
    - Use heavy-weight tripod to mount camera to definitely avoid vibrations of imaging device.
2. **Installation and Setup**
    - Perform camera calibration on-site using Camera Calibration Toolbox using instructions given in Section 4.2.4.
    - Comply basic measurement setup described in Section 1.4 and illustrated in Figures 1.4 and 3.13.
    - Use suitable existing structures (which are inherently aligned with the carbody coordinate system) to mount the targets. Where possible, choose locations for targets where they are not exposed to sudden changes in lighting conditions or direct sunlight (e.g. as pictured in Figure 5.32 directly after the corridor connection or beneath the ceiling).
    - Determine precise location (and orientation) of 5-CCC targets within carbody reference frame.
    - Check and verify installation of targets using brought devices. The measurement accuracy highly depends on a precise installation.
3. **Configure camera and lens**
    - Configure camera and lens according to environmental conditions. Adjust focal length and aperture value (and other necessary parameters) to lighting conditions and given working distance. A small aperture setting increases the depth of field. A detailed description of parameters is given in Section 4.1.1.

4. **Measurement**

   - Assure that the respective carbodies are in the initial pose (described in Section 3.4.3) when the measurement starts. Mark beginning and end of test ride with clear signal in image data (useful for later processing). During measurement verify the vibration-free installation of camera and targets.

5. **Evaluation**

   - Use the prototype implementation to evaluate measurement data. Transfer parameters gained from camera calibration to evaluation framework. For detailed instructions see the supplied manual.

# 6. Conclusion

This thesis addresses the task of conducting a feasibility study and the accompanying prototypic development of a robust and cost-efficient image-based measurement system, which is capable of tracking the relative motions between railway vehicle carbodies.

The discussion about a robust, fast and accurate tracking system resulted in a clear decision in favour of a 2D, passive, high-contrast, circular marker called Concentric Contrasting Circle (CCC) [30] (see Figure 3.3) as well as a corresponding tracking method based on a centre of gravity (CoG) [5] calculation. Furthermore, it is shown that for this particular type of application the influence of perspective distortion on the accuracy of CoG calculation can be neglected.

The choice for the two presented and evaluated pose estimation algorithms DLT [37, p. 178] and EPnP [54] was primarily founded on the fact that both are considered either as straightforward and "*Gold Standard*" [37] (DLT) or state-of-the-art (EPnP) approaches. Another important reason for this decision was the availability of extensively tested implementations. This feature is regarded especially important for the later application in a productive environment.

The proposed basic measurement setup fully complies with the introduced railway vehicle carbody motion model. Together with the presented compulsory initialisation (initial pose) it facilitates fast installation, configuration and operation of the measurement system.

In case of the optical imaging system it was found beneficial to utilise a SONY NEX 5N because of its capabilities of operating entirely autonomous. Concerning the operation of the measurement system in a productive environment, the recommendations range from the application of an industrial imaging device to the utilisation of active targets. Furthermore, with the Camera Calibration Toolbox [10] a feasible and proven method for doing geometric camera calibration was introduced.

The results of the extensive experiments (laboratory (5.3), test rig (5.4), train (5.5)) and accompanying evaluations generally prove the feasibility and suitability of the designed image-based measurement system. It is shown that three multi-fiducial 5-CCC targets are sufficient to guarantee that the measurement uncertainty stays within the required limits. Furthermore, the evaluations demonstrate the positive impact of non-coplanar targets on the uncertainty in pose estimation. The results also indicate that when a smaller number of fiducials is applied ($\leq 15$) the DLT algorithm is, compared to the EPnP, the more appropriate method of choice for pose estimation.

The proposed image-based measurement system contributes a genuine alternative, considering the introductory mentioned drawbacks of conventional methods applied to the particular task of measuring the relative motions between two railway vehicle carbodies. The conventional methods, relying on draw wire sensors, suffer from the drawback of being extensive and time consuming regarding installation, measurement and analysis. Whereas, the efforts for installation and configuration of the proposed system appear limited. In fact, the only time consuming parts are the calibration of the camera and the precise mounting and alignment of the targets. Furthermore, the proposed system basically allows the measurement independent of any external power supply, network or storage device. Moreover, it certainly reduces evaluation efforts, as the analysis of measurements is possible on-site, in the course of the test ride.

Regarding the computer vision aspects, the major challenges such as the variable lighting conditions and fast motions (motion blur) during train ride, are addressed mainly by the choice of the fiducial and the tracking method. The selected high-contrast CCC fiducial printed on several planar targets in combination with the specifically developed tracking method applying an adaptive threshold enables robust and accurate tracking.

Based on the presented results the following drawbacks, possible enhancements and further research are identified:

- **Inaccuracies in measurement setup**
  The results from Experiment 1 show that the measurement system is in fact very susceptible to inaccuracies in the measurement setup. Especially small deviations from the expected rotational alignment of a target gives rise to a systematic bias. In experiments on the test rig the rectangular shape of the platform greatly supported the installation and alignment of the targets. This is not always the case in a real world environment. In Section 3.4.2 it was recommended to use existing structures and furnishings within the carbody. This approach was applied during the test ride, but is only advisable if it is certain that these structures are axially aligned with the reference frame of the carbody. It has to be investigated how a precise installation of the targets can be guaranteed.

- **Camera calibration from circular fiducials**
  Although for the pose estimation measurements the circular CCC fiducial with a corresponding target design is proposed, in context of the Camera Calibration Toolbox a squared chequerboard pattern is used for camera calibration. It is suggested to investigate a planar target design which is suitable for both pose estimation and camera calibration.
- **Increased robustness and speed of tracking method**
  The accuracy of the system strongly depends on accurate marker tracking results. Although, with the choice for the passive CCC marker and the target design the system is applicable in a regular environment, it is still sensitive to sudden changes in lighting conditions (e.g. the train entering a tunnel). The application of retro-reflective materials or even active markers (see Figure 4.2) has a high potential to considerably increase robustness and speed up the tracking procedure.

# Bibliography

[1] F. Ababsa and M. Mallem. Robust camera pose estimation using 2d fiducials tracking for real-time augmented reality systems. In *Proceedings of the 2004 ACM SIGGRAPH international conference*, pages 431–435, New York, NY, USA, 2004. ACM. 2

[2] F. Ababsa and M. Mallem. Robust circular fiducials tracking and camera pose estimation using particle filtering. In *IEEE International Conference on Systems, Man and Cybernetics (ISIC)*, pages 1159 –1164, October 2007. 3

[3] Francisco Abad, Emilio Camahort, and Roberto Vivó. Camera calibration using two concentric circles. In *Proceedings of the International Conference on Image Analysis and Recognition*, pages 688–696, 2004. 3.2.1, 3.2.1

[4] J. Andrade-Cetto and F. Thomas. A wire-based active tracker. *IEEE Transactions on Robotics*, 24(3):642–651, June 2008. 2

[5] H.C. van Assen, M. Egmont-Petersen, and J.H.C. Reiber. Accurate object localization in gray level images using the center of gravity measure: accuracy versus precision. *IEEE Transactions on Image Processing*, 11(12):1379–1384, December 2002. 3.2.3, 6

[6] Automated Imaging Association. GiGE Vision - Video Streaming and Device Control over Ethernet Standard, version 2.0.02, November 2011. 4.1.1

[7] P. Azad, T. Asfour, and R. Dillmann. Stereo-based vs. monocular 6-dof pose estimation using point features: A quantitative comparison. In *Autonome Mobile Systeme (AMS)*, 2009. 3

[8] F. Bergamasco, A. Albarelli, E. Rodola, and A. Torsello. Rune-tag: A high accuracy fiducial marker with strong occlusion resilience. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 113–120, June 2011. 3.2

[9] C.B. Bose and J. Amir. Design of fiducials for accurate registration using machine vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12:1196–1200, December 1990. 3.2.3

[10] Jean-Yves Bouguet. Caltech Camera Calibration Toolbox for MATLAB, last update 2010-07-27. `http://www.vision.caltech.edu/bouguetj/calib_doc/`. last downloaded 2012-11-13. 3.2.2, 4.2.2, 4.2.3, 6, 6

[11] R. Bradski. Computer vision face tracking for use in a perceptual user interface. *Intel Technology Journal Q2*, 1998. 3.2.3

[12] Markus Brandner. *Uncertainty Evaluation in Vision-Based Measurement Systems.* PhD thesis, Institute of Electrical Measurement and Measurement Signal Processing - Graz University of Technology, 2009. 3.2.1, 3.2.4, 3.6, 3.2.4, 4.2.1

[13] F. Chaumette and S. Hutchinson. Visual servo control, Part I: Basic approaches. *IEEE Robotics and Automation Magazine*, 13(4):82–90, December 2006. 2

[14] Yizong Cheng. Mean shift, mode seeking, and clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(8):790–799, August 1995. 3.2.3

[15] Youngkwan Cho and U. Neumann. Multi-ring color fiducial systems for scalable fiducial tracking augmented reality. In *Proceedings of the IEEE Virtual Reality Annual International Symposium*, 1998. 3.2.1

[16] D. Claus and A. W. Fitzgibbon. Reliable fiducial detection in natural scenes. In *European Conference on Computer Vision*, pages 469–480. Springer-Verlag, 2004. 3.2.1, 3.2.2

[17] D. Comaniciu, V. Ramesh, and P. Meer. Real-time tracking of non-rigid objects using mean shift. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 2, pages 142–149, 2000. 3.2.3

[18] ALCAN Composites. Kapa mount - technial datasheet. `http://www.kapaplatten.de/pdf-dokumente/kapa_mount_en.pdf`, July 2007. last downloaded 2012-12-17. 5.4.1

[19] Ascension Technology Corp. Technical description of MotionStar - Realtime Motion Capture. `http://www.ascension-tech.com/docs/products/motionstar_10_04.pdf`, 2004. last downloaded 2012-07-13. 2

[20] John J. Craig. *Introduction to robotics: mechanics and control.* Pearson/Prentice Hall, 2nd edition, 1989. 2

[21] S. Del Corso, C. Miro, and J. Jung. MNR: a novel approach to correct MPEG temporal distortions. *IEEE Transactions on Consumer Electronics*, 49(1):229–236, 2003. 3.2.5, 3.2.5

[22] Frank Dellaert, Steven M. Seitz, Charles E. Thorpe, and Sebastian Thrun. Structure from motion without correspondence. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 557–564, 2000. 3

[23] Daniel F. DeMenthon and Larry S. Davis. Model-based object pose in 25 lines of code. *International Journal of Computer Vision*, 15:123–141, 1995. 3.3.2

[24] Christophe Doignon, Florent Nageotte, Benjamin Maurin, and Alexandre Krupa. *Pose Estimation and Feature Tracking for Robot Assisted Surgery with Medical Imaging*, volume 8 of *Lecture Notes on Electrical Engineering*. Springer-Verlag, May 2008. 2

[25] EBU-European Broadcasting Union. Information Paper on HDTV Formats. `http://tech.ebu.ch/docs/techreports/tr005.pdf`, February 2010. last downloaded 2012-10-15. 4.1.1

[26] M. Fiala. Artag, a fiducial marker system using digital techniques. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 2, pages 590–596, June 2005. 3.2.1, 3.2, 3.2.1

[27] M. Fiala. Designing highly reliable fiducial markers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(7):1317 –1324, july 2010. 3.2.1

[28] Martin A. Fischler and Robert C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 24(6):381–395, June 1981. 3, 3.3.4, 3.3.4

[29] Edward von Flottwell. *Statistische Analyse von Relativbewegungen gekuppelter schienengebundener Fahrzeuge als Vorgabe für Lebensdauertests.* Wilhelm Büchner Hochschule, Darmstadt, 2009. 1.1, 1.2, 1.2, 1.5, 2, 2

[30] L. Gatrell, W. Hoff, and C. Sklair. Robust image features: Concentric contrasting circles and their image extraction. In Stoney W., editor, *Cooperative Intelligent Robotics in Space*, volume 1612, pages 512–520. SPIE - The international Society for Optical Engineering, 1991. 3.2.1, 3.3, 3.2.3, 3, 6

[31] Sinan Gezici. A survey on wireless position estimation. *Wireless Personal Communications*, 44(3):263–282, February 2008. 2

[32] AICON 3D Systems GmbH. Technical description of AICON 3D - Movie Inspect Technology WheelWatch. `http://www.aicon3d.com/fileadmin/user_upload/produkte/de/moveinspect/wheelwatch/pdf/Brochure_WheelWatch_en_04_2012.pdf`, 2012. last downloaded 2012-07-19. 2, 3

[33] Robert Godding. Geometrische Kalibrierung und Orientierung digitaler Bildaufnahmesysteme. `http://www.falcon.de/falcon/pdf/de/aicon/geometric_calibration.pdf`. last downloaded 2012-11-13. 4.2.1

[34] Rafael C. Gonzalez and Richard E. Woods. *Digital Image Processing*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 2nd edition, 2001. 3.2.5

[35] W. Hanneforth and W. Fischer. *Laufwerke*. Transpress Verlag, Berlin, Germany, 1986. 1.2

[36] R.M. Haralick, D. Lee, K. Ottenburg, and M. Nölle. Analysis and solutions of the three point perspective pose estimation problem. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 592–598, June 1991. 3.3

[37] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2nd edition, 2004. 3, 3.3.1, 3.3.2, 3.3.3, 3.3.3, 3.3.3, 3.3.5, 4.2.1, 6

[38] Janne Heikkilä. Accurate camera calibration and feature based 3-d reconstruction from monocular image sequences. Master's thesis, Department of Electrical Engineering - University of Oulu, 1997. 3.2.3

[39] Janne Heikkilä. Geometric camera calibration using circular control points. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 22(10):1066–1077, October 2000. 3.2.1, 3.2.1, 4.1.2, 5

[40] Janne Heikkilä and Olli Silvén. A four-step camera calibration procedure with implicit image correction. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1106–1112, June 1997. 4.2.1, 4.2.2, 4.2.3

[41] Jeffrey Hightower and Gaetano Borriello. Location systems for ubiquitous computing. *Computer*, 34(8):57–66, August 2001. 2

[42] Robert Hödl. Bildgestützte Messung von Relativbewegungen zwischen Wagenkästen – Masterarbeit - Projektvorschlag. internal document, November 2011. 1.2, 1.3

[43] William A. Hoff, Lance B. Gatrell, and John R. Spofford. Machine-vision-based teleoperation aid. *Telematics and Informatics*, 8(4):403–423, 1991. 3

[44] William A. Hoff, Khoi Nguyen, and Torsten Lyon. Computer vision-based registration techniques for augmented reality. In *Intelligent Robots and Computer Vision XV*, pages 538–548, 1996. 3.2.2

[45] Simon Iwnicki. *Handbook of Railway Vehicle Dynamics*. Taylor and Francis Group, LLC, London, United Kingdom, 2006. 1.2

[46] Guang Jiang and Long Quan. Detection of concentric circles for camera calibration. In *Proceedings of the Tenth IEEE International Conference on Computer Vision (ICCV)*, volume 1, pages 333–340, Washington, DC, USA, 2005. IEEE Computer Society. 3.2.1, 4.2.1

[47] Tomas Karis. Track irregularities for high-speed trains - evaluation of their correlation with vehicle response. Master's thesis, KTH Royal Institute of Technology, Stockholm, Sweden, 2009. 1.2

[48] H. Kato, M. Billinghurst, I. Poupyrev, K. Imamoto, and K. Tachibana. Virtual object manipulation on a table-top ar environment. In *Proceedings of the IEEE and ACM International Symposium on Augmented Reality (ISAR)*, pages 111–119, 2000. 3, 3.2.1, 3.2, 3.2.1

[49] J.M. Kelsey, J. Byrne, M. Cosgrove, S. Seereeram, and R.K. Mehra. Vision-based relative pose estimation for autonomous rendezvous and docking. In *Proceedings of the IEEE Aerospace Conference*, 2006. 2, 3

[50] Jun-Sik Kim, P. Gurdjos, and In-So Kweon. Geometric and algebraic constraints of projected concentric circles and their applications to camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(4):637 –642, April 2005. 3.2.1

[51] Jun-Sik Kim and In-So Kweon. A new camera calibration method for robotic applications. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, volume 2, pages 778–783, 2001. 3.2.1

[52] Laurent Kneip, Scaramuzza Davide, and R Siegwart. A novel parametrization of the perspective-three-point problem for a direct computation of absolute camera position and orientation. In *Proceedings of The IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, Colorado Springs, CO, USA, June 2011. 3, 3.3.4

[53] V. Lepetit and P. Fua. Monocular Model-Based 3D Tracking of Rigid Objects: A Survey. *Foundations and Trends in Computer Graphics and Vision*, 1(1):1–89, 2005. 2, 2, 3, 3.2.1, 3.2.1, 3.3.3, 4.1.1, 4.2.1, 5.4.5

[54] V. Lepetit, F. Moreno-Noguer, and P. Fua. EPnP: An Accurate O(n) Solution to the PnP Problem. *International Journal of Computer Vision*, 81:155–166, 2008. 3, 3.3.2, 3.3.4, 3.3.4, 3.3.4, 3.3.5, 5.2, 5.3.3, 6

[55] Liang Li, Hanming Chen, and Guixiang Wang. Camera calibration based on circular markers. In *International Conference on Computational Intelligence and Software Engineering (CiSE)*, pages 1–4, December 2009. 3.2.1

[56] David G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004. 3

[57] M. Manz, T. Luettel, F. von Hundelshausen, and H.-J. Wuensche. Monocular model-based 3d vehicle tracking for autonomous vehicles in unstructured environment. In *IEEE International Conference on Robotics and Automation (ICRA)*, pages 2465–2471, May 2011. 2

[58] Donald W. Marquardt. An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial and Applied Mathematics*, 11(2):431–441, 1963. 4.2.2

[59] J. Matas, O. Chum, M. Urban, and T. Pajdla. Robust wide baseline stereo from maximally stable extremal regions. In *Proceedings of the British Machine Vision Conference*, pages 384–393, 2002. 3

[60] Leonid Naimark and Eric Foxlin. Circular data matrix fiducial system and robust image processing for a wearable vision-inertial self-tracker. In *Proceedings of the 1st International Symposium on Mixed and Augmented Reality*, ISMAR '02, Washington, DC, USA, 2002. IEEE Computer Society. 3.2, 3.2.3

[61] National Instruments. Image Acquisition System Setup. `http://www.ni.com/white-paper/4229/en`, 2011. last downloaded 2012-11-08. 4.1.2

[62] David Nistér, Oleg Naroditsky, and James Bergen. Visual odometry for ground vehicle applications. *Journal of Field Robotics*, 23:2006, 2006. 2

[63] Isao Okamoto. How bogies work. *EJRCF-Japan Railway and Transport Review*, 18, December 1998. 1.2

[64] Anneli Orvnäs. Methods for reducing vertical carbody vibrations of a rail vehicle. Technical report, KTH Engineering Sciences, Dept. of Aeronautical and Vehicle Engineering, Division of Rail Vehicles, Stockholm, Sweden, 2010. R(b)

[65] Charles B. Owen, Fan Xiao, and Paul Middlin. What is the best fiducial? In *The First IEEE International Augmented Reality Toolkit Workshop*, pages 98–105, Darmstadt, Germany, September 2002. 3.2.1

[66] Richard P. Paul. *Robot Manipulators: Mathematics, Programming, and Control*. MIT Press, Cambridge, MA, USA, 1st edition, 1982. 3.4.3

[67] R. Peng and M.L. Sichitiu. Angle of arrival localization for wireless sensor networks. In *Sensor and Ad Hoc Communications and Networks*, volume 1, pages 374–382, September 2006. 2, 2

[68] Martin Pumberger. Parallelkinematik für Funktionsversuch und Betriebsmessung von Fahrwerken. Master's thesis, Graz University of Technology, Institut für Maschinenelemente und Entwicklungsmethodik, Graz, 2005. 2, 5.14, 5.4.2

[69] Long Quan and Zhongdan Lan. Linear n-point camera pose determination. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(8):774–780, August 1999. 3.3.4, 3.3.4

[70] F.H. Raab, E.B. Blood, T.O. Steiner, and H.R. Jones. Magnetic position and orientation tracking system. *IEEE Transactions on Aerospace and Electronic Systems*, AES-15(5):709–718, September 1979. 2

[71] F. Remondino and C. Fraser. Digital camera calibration methods: considerations and comparisons. In ISPRS, editor, *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, volume 36, Dresden, Germany, 2006. 3.3, 4.2.1

[72] Andrew C. Rice. Dependable systems for Sentient Computing. Technical Report UCAM-CL-TR-686, University of Cambridge, Computer Laboratory, May 2007. 3.2.1

[73] Robert Bosch GmbH. GLM Professional 150 | 250 VF – original instructions. `http://www.bosch-professional.com/de/de/ocs/download/pdf/manual/63683/glm-150-professional_manual.pdf`, July 2009. last downloaded 2013-01-23. 5.3.1

[74] Bodo Rosenhahn. *Pose Estimation Revisited.* PhD thesis, Institut für Informatik u. Praktische Mathematik der Christian-Albrechts-Universität zu Kiel, 2003. 3.3.2

[75] Samuel Schulter. Messtechnische Evaluierung von passiven Landmarken. Technical report, ICG - Institute for Computer Graphics and Vision, Graz University of Technology, Graz, Austria, 2012. 3.2.1

[76] G. Schweighofer and A. Pinz. Robust pose estimation from a planar target. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(12):2024–2030, December 2006. 3, 3.3.2, 5.3.4

[77] C. E. Shannon. Communication in the Presence of Noise. *Proceedings of the IRE*, 37(1):10–21, January 1949.

[78] Siemens AG. Wegberg-Wildenrath Test- and Validationcenter – information brochure, June 2009. 5.5

[79] Siemens AG. High-Speed Trainset Velaro D (Class 407) – data sheet. `http://www.siemens.com/press/pool/de/events/industry/mobility/2010-04-velaro-d/datasheet_siemens_velaro-d_en.pdf`, April 2010. last downloaded 2013-01-17. 5.5

[80] C. Silva and J. Santos-Victor. Robust egomotion estimation from the normal flow using search subspaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(9):1026–1034, September 1997. 3

[81] Milan Sonka, Vaclav Hlavac, and Roger Boyle. *Image Processing, Analysis, and Machine Vision*. Thomson-Engineering, 3rd edition, 2007. 2, 3.2.5

[82] SONY Corporation. AVCHD Introduction Handbook. `http://pro.sony.com/bbsccms/assets/files/micro/nxcam/manuals/AVCHD_E_1127-rev3.pdf`, December 2009. last downloaded 2012-10-15. 4.1.1

[83] SONY Corporation. NEX-5N Interchangeable Lens Digital Camera. `http://www.sony.co.uk/support/en/product/NEX-5N/manuals`, 2011. last downloaded 2012-10-15. 4.1.1

[84] D. Stewart. A platform with six degrees of freedom. In *Proceedings of Institution of Mechanical Engineers (UK)*, volume 180, Pt 1, No 15, pages 371–386, 1965-1966. 2

[85] Wei Sun and J.R. Cooperstock. Requirements for camera calibration: Must accuracy come with a high price? In *Seventh IEEE Workshops on Application of Computer Vision*, volume 1, pages 356–361, January 2005. 5

[86] I.E. Sutherland. Three-dimensional data input by tablet. *Proceedings of the IEEE*, 62(4):453–461, April 1974. 3.3.3

[87] Chin-Woo Tan and Sungsu Park. Design and error analysis of accelerometer-based inertial navigation systems. Technical report, Institute of Transportation Studies, UC Berkeley, 2002. 2

[88] Roger Y. Tsai. A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the-shelf tv cameras and lenses. *Journal of Robotics and Automation*, RA-3(4):323–344, August 1987. 3.2.2, 3.3.3, 4.2.1

[89] J. Weng, P. Cohen, and M. Herniou. Camera calibration with distortion models and accuracy evaluation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(10):965–980, October 1992. 4.2.3

[90] Annie Wu, Clint Bidlack, Arun Katkere, and Roy Feague. Vision based object pose estimation for mobile robots. Technical report, University of Michigan, Artificial Intelligence Laboratory, 1994. 2

[91] Anqi Xu and G. Dudek. Fourier tag: A smoothly degradable fiducial marker system with configurable payload capacity. In *Proceedings of the Canadian Conference on Computer and Robot Vision (CRV)*, pages 40–47, May 2011. 3.2

[92] Alper Yilmaz, Omar Javed, and Mubarak Shah. Object tracking: A survey. *ACM Computing Surveys*, 38(4), December 2006. 3.2.3

[93] Zhengyou Zhang. A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11):1330–1334, November 2000. 3.2.2, 4.2.1, 4.2.2

# A. Project Proposal (in German)

Institut für Elektrische Meßtechnik
und Meßsignalverarbeitung
Technische Universität Graz

Bildgestützte Messung von
Relativbewegungen zwischen
Wagenkästen

**Masterarbeit - Projektvorschlag**

Diplomand:
Robert Hödl, BSc.
`robert.hoedl@student.tugraz.at`

Betreuer:
Ao. Univ.-Prof. Dipl.-Ing. Dr.techn. Axel Pinz und
Dipl.-Ing. Dr.techn. Thomas Thurner

02. November 2011

# Inhaltsverzeichnis

# 1 Projektvorschlag

## 1.1 Motivation / Ist Zustand

Wie in [1] beschrieben, führen die einzelnen Einheiten von Schienenfahrzeugen (SFZ) im Betrieb zueinander translatorische und rotatorische Bewegungen aus. Zwei solche Einheiten sind mit verschiedensten Verbindungen gekoppelt. Diese dienen zur Übertragung von statischen und dynamischen Zug- und Stoßkräften sowie von Information und Energie und ermöglichen den Übergang von Personen und Gütern. Die Grenzen der Relativbewegungen zwischen den einzelnen Einheiten sind durch Faktoren wie Gleisgeometrie, Federwege und Toleranzen festgelegt.

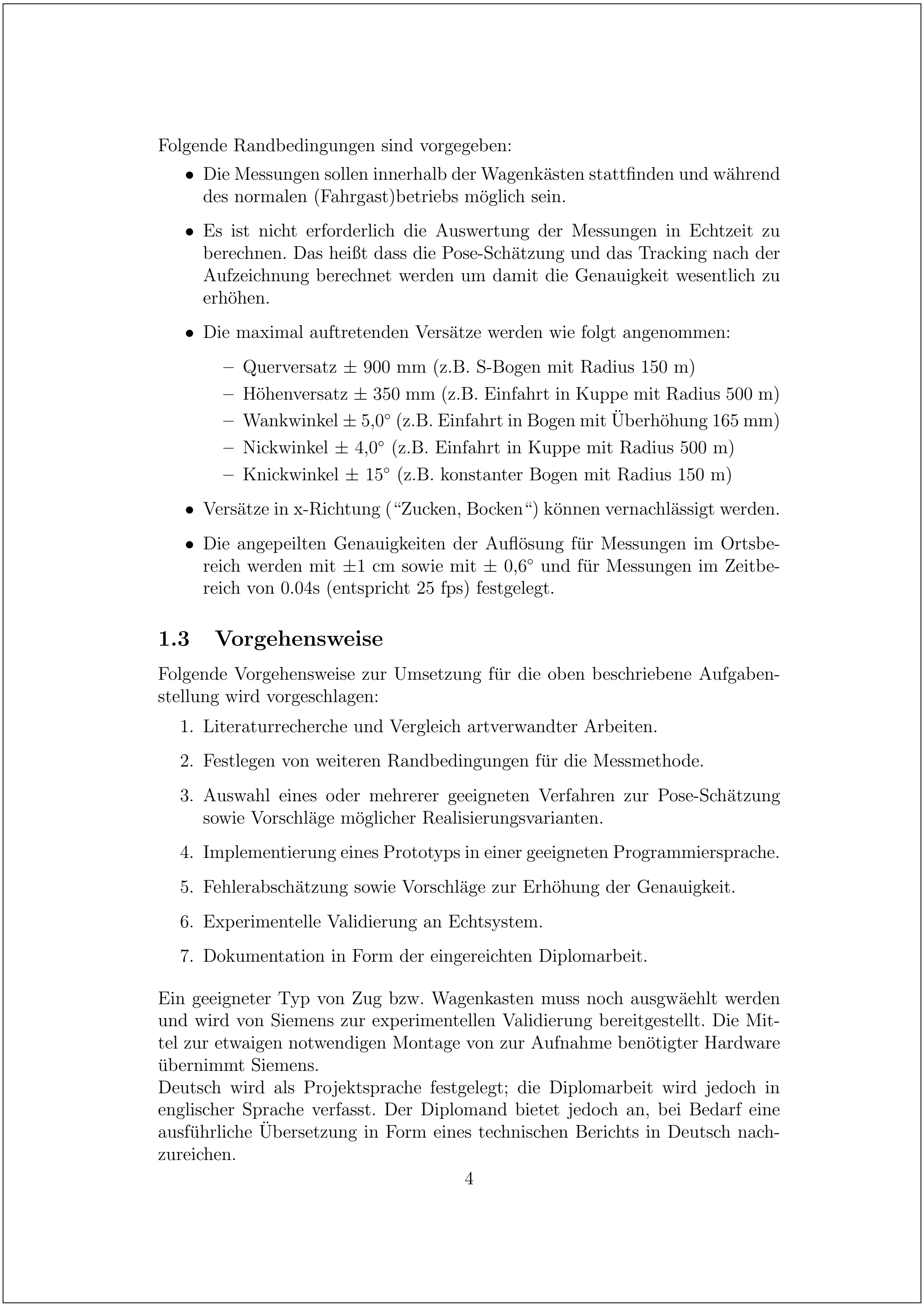
Laut [1] werden zur Zeit die Verbindungselemente gekuppelter SFZ "anhand berechneter Maximalauslenkungen, Versuchen und Erfahrungswerten so dimensioniert, dass für die Grenzstellungen die Dauerfestigkeit garantiert ist". Der Lebensdauerauslegung der Verbindungen liegt ebenfalls eine ausschließlich empirische Abschätzung der Häufigkeitsverteilung von Bewegungsklassen von SFZ zugrunde.

Um künftig einerseits eine den realen Gegebenheiten entsprechende, exaktere Bauteilauslegung (Dimensionierung, Beweglichkeit, Bauraum) zu ermöglichen und um andererseits die Kosten zu reduzieren, soll die statistische Verteilung des Auftretens von Relativbewegungen, eingeteilt in verschiedene Bewegungsklassen im Echtbetrieb, zwischen den einzelnen Einheiten von SFZ gemessen und untersucht werden.

## 1.2 Problemstellung

Da die in [1] ausgewerteten Beschleunigungsdaten keine definitiven Rückschlüsse auf die auftretenden Relativbewegungen zwischen zwei Einheiten von SFZ (speziell Wagenkästen) zulassen, sollen diese Bewegungen nun mittels eines bildgebenden Verfahrens erfasst und ausgewertet werden.

Dazu sollen mittels eines im Innenraum der Wagenkästen angebrachten optischen Aufnahmesystems diese Relativbewegungen (Rotationen und Translationen, 5 bis 6 Freiheitsgrade) aufgezeichnet und nachverfolgt werden und somit die prinzipielle Machbarkeit eines solchen Verfahrens demonstriert werden.



Folgende Randbedingungen sind vorgegeben:

- Die Messungen sollen innerhalb der Wagenkästen stattfinden und während des normalen (Fahrgast)betriebs möglich sein.
- Es ist nicht erforderlich die Auswertung der Messungen in Echtzeit zu berechnen. Das heißt dass die Pose-Schätzung und das Tracking nach der Aufzeichnung berechnet werden um damit die Genauigkeit wesentlich zu erhöhen.
- Die maximal auftretenden Versätze werden wie folgt angenommen:
  - Querversatz ± 900 mm (z.B. S-Bogen mit Radius 150 m)
  - Höhenversatz ± 350 mm (z.B. Einfahrt in Kuppe mit Radius 500 m)
  - Wankwinkel ± 5,0° (z.B. Einfahrt in Bogen mit Überhöhung 165 mm)
  - Nickwinkel ± 4,0° (z.B. Einfahrt in Kuppe mit Radius 500 m)
  - Knickwinkel ± 15° (z.B. konstanter Bogen mit Radius 150 m)
- Versätze in x-Richtung ("Zucken, Bocken") können vernachlässigt werden.
- Die angepeilten Genauigkeiten der Auflösung für Messungen im Ortsbereich werden mit ±1 cm sowie mit ± 0,6° und für Messungen im Zeitbereich von 0.04s (entspricht 25 fps) festgelegt.

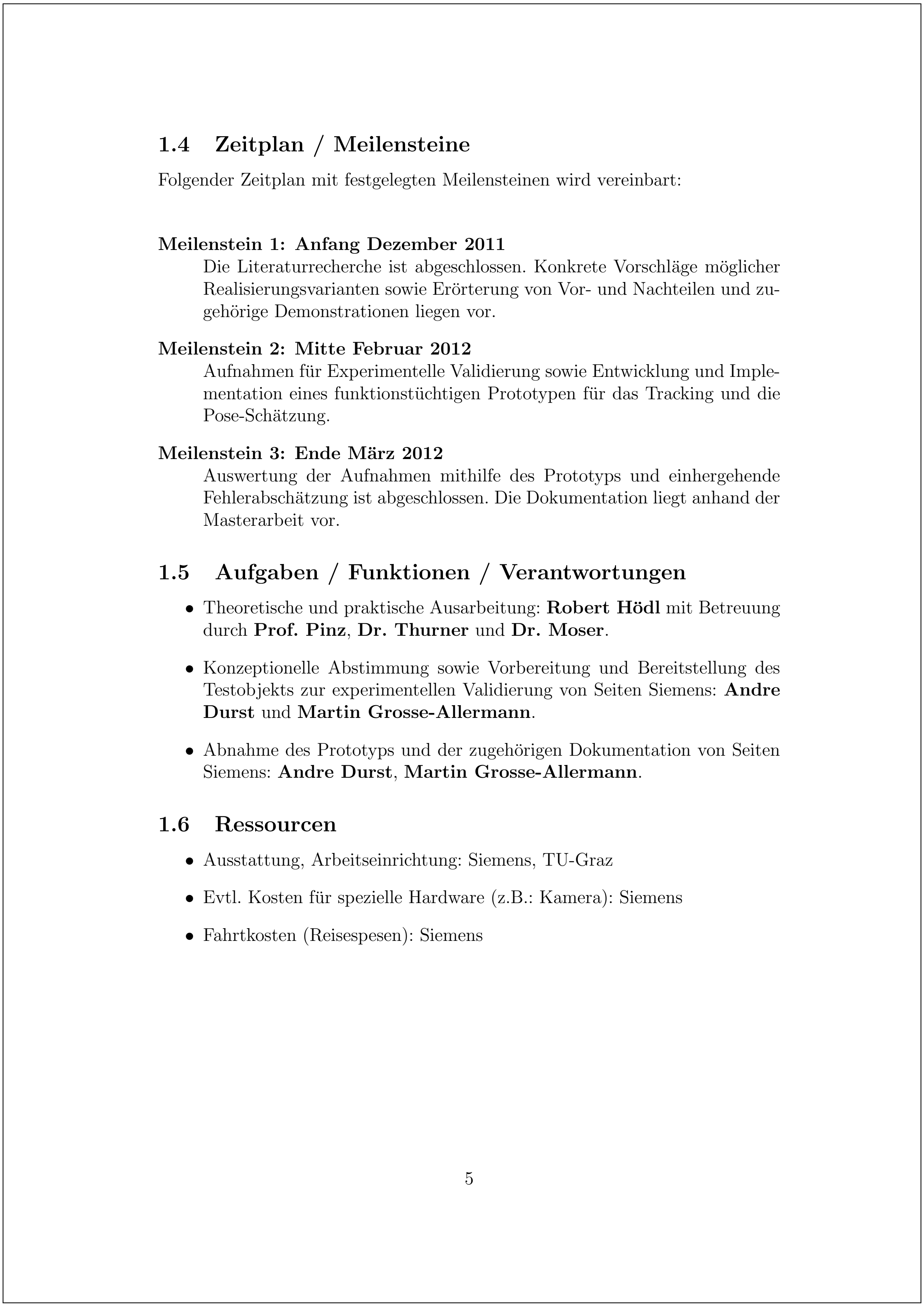
### 1.3 Vorgehensweise

Folgende Vorgehensweise zur Umsetzung für die oben beschriebene Aufgabenstellung wird vorgeschlagen:

1. Literaturrecherche und Vergleich artverwandter Arbeiten.
2. Festlegen von weiteren Randbedingungen für die Messmethode.
3. Auswahl eines oder mehrerer geeigneten Verfahren zur Pose-Schätzung sowie Vorschläge möglicher Realisierungsvarianten.
4. Implementierung eines Prototyps in einer geeigneten Programmiersprache.
5. Fehlerabschätzung sowie Vorschläge zur Erhöhung der Genauigkeit.
6. Experimentelle Validierung an Echtsystem.
7. Dokumentation in Form der eingereichten Diplomarbeit.

Ein geeigneter Typ von Zug bzw. Wagenkasten muss noch ausgwäehlt werden und wird von Siemens zur experimentellen Validierung bereitgestellt. Die Mittel zur etwaigen notwendigen Montage von zur Aufnahme benötigter Hardware übernimmt Siemens.
Deutsch wird als Projektsprache festgelegt; die Diplomarbeit wird jedoch in englischer Sprache verfasst. Der Diplomand bietet jedoch an, bei Bedarf eine ausführliche Übersetzung in Form eines technischen Berichts in Deutsch nachzureichen.

4



## 1.4 Zeitplan / Meilensteine

Folgender Zeitplan mit festgelegten Meilensteinen wird vereinbart:

**Meilenstein 1: Anfang Dezember 2011**
Die Literaturrecherche ist abgeschlossen. Konkrete Vorschläge möglicher Realisierungsvarianten sowie Erörterung von Vor- und Nachteilen und zugehörige Demonstrationen liegen vor.

**Meilenstein 2: Mitte Februar 2012**
Aufnahmen für Experimentelle Validierung sowie Entwicklung und Implementation eines funktionstüchtigen Prototypen für das Tracking und die Pose-Schätzung.

**Meilenstein 3: Ende März 2012**
Auswertung der Aufnahmen mithilfe des Prototyps und einhergehende Fehlerabschätzung ist abgeschlossen. Die Dokumentation liegt anhand der Masterarbeit vor.

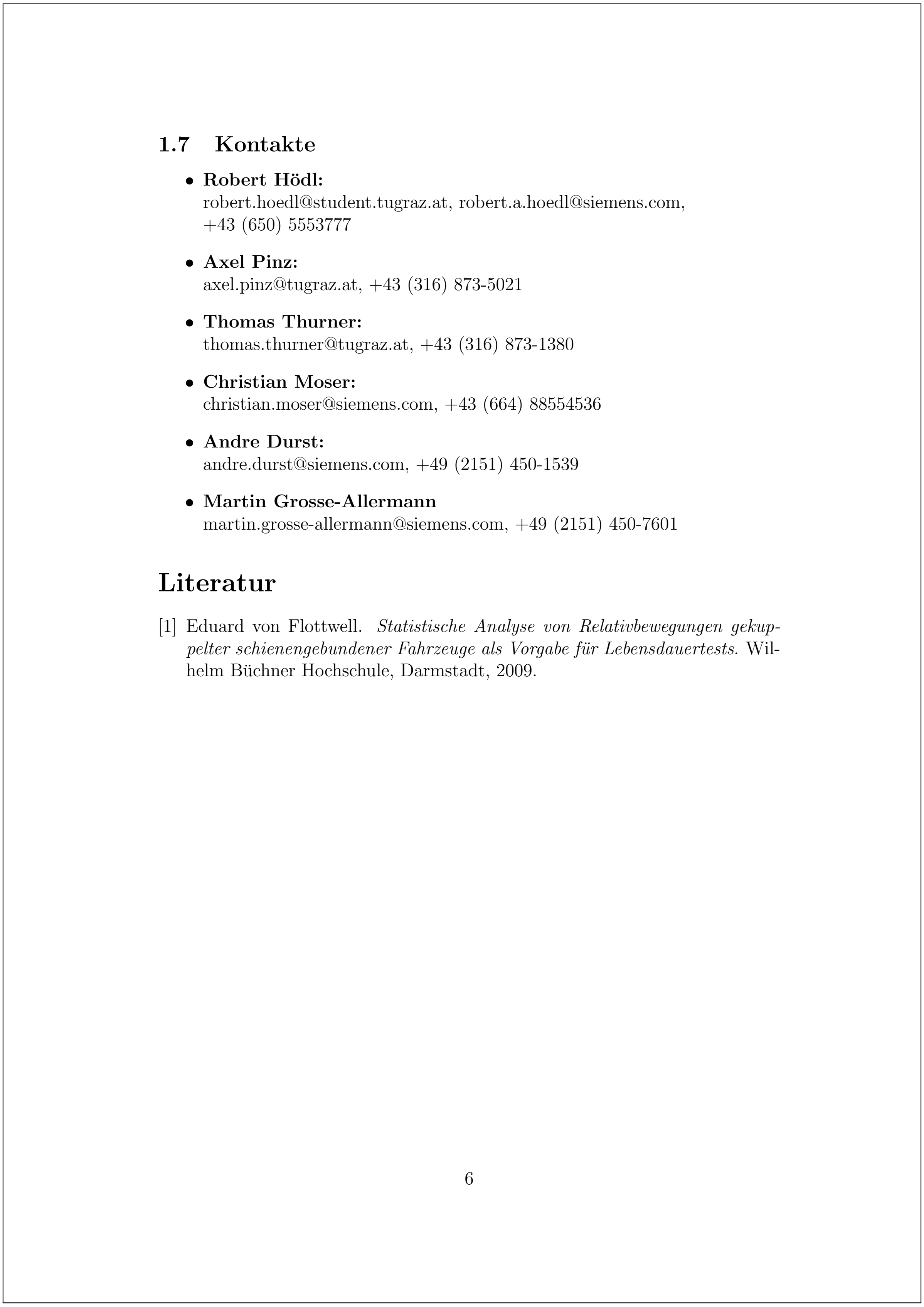
## 1.5 Aufgaben / Funktionen / Verantwortungen

- Theoretische und praktische Ausarbeitung: **Robert Hödl** mit Betreuung durch **Prof. Pinz**, **Dr. Thurner** und **Dr. Moser**.
- Konzeptionelle Abstimmung sowie Vorbereitung und Bereitstellung des Testobjekts zur experimentellen Validierung von Seiten Siemens: **Andre Durst** und **Martin Grosse-Allermann**.
- Abnahme des Prototyps und der zugehörigen Dokumentation von Seiten Siemens: **Andre Durst**, **Martin Grosse-Allermann**.

## 1.6 Ressourcen

- Ausstattung, Arbeitseinrichtung: Siemens, TU-Graz
- Evtl. Kosten für spezielle Hardware (z.B.: Kamera): Siemens
- Fahrtkosten (Reisespesen): Siemens

5



## 1.7 Kontakte

- **Robert Hödl:**
  robert.hoedl@student.tugraz.at, robert.a.hoedl@siemens.com,
  +43 (650) 5553777
- **Axel Pinz:**
  axel.pinz@tugraz.at, +43 (316) 873-5021
- **Thomas Thurner:**
  thomas.thurner@tugraz.at, +43 (316) 873-1380
- **Christian Moser:**
  christian.moser@siemens.com, +43 (664) 88554536
- **Andre Durst:**
  andre.durst@siemens.com, +49 (2151) 450-1539
- **Martin Grosse-Allermann**
  martin.grosse-allermann@siemens.com, +49 (2151) 450-7601

## Literatur

[1] Eduard von Flottwell. *Statistische Analyse von Relativbewegungen gekuppelter schienengebundener Fahrzeuge als Vorgabe für Lebensdauertests.* Wilhelm Büchner Hochschule, Darmstadt, 2009.