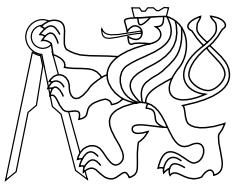
CENTER FOR
MACHINE PERCEPTION



CZECH TECHNICAL
UNIVERSITY IN PRAGUE

MASTER'S THESIS

Camera Rig Calibration

Stanislav Steidl

stanislav.steidl@gmail.com

January 9, 2018

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

This work was supported by EU Structural and Investment Funds,
Operational Programe Research, Development and Education
project IMPACT No. CZ.02.1.01/0.0/0.0/15
003/0000468.EU-H2020 and by EU project LADIO No. 731970.

Center for Machine Perception, Department of Cybernetics
Faculty of Electrical Engineering, Czech Institute of Informatics, Robotics,
and Cybernetics, Czech Technical University
Jugoslávských partyzánů 3, 160 00 Prague 6, Czech Republic
phone: +420 2 2435 4139, www: http://cmp.felk.cvut.cz

CTU
CZECH TECHNICAL UNIVERSITY IN PRAGUE

MASTER‘S THESIS ASSIGNMENT

I. Personal and study details

Student's name: **Steidl Stanislav** Personal ID number: **393096**

Faculty / Institute: **Faculty of Electrical Engineering**

Department / Institute: **Department of Cybernetics**

Study program: **Open Informatics**

Branch of study: **Computer Vision and Image Processing**

II. Master’s thesis details

Master’s thesis title in English:

Camera Rig Calibration

Master’s thesis title in Czech:

Kalibrace soustavy kamer

Guidelines:

1. Review the state of the art in camera rig calibration based on bundle adjustment [1-6] including previous methods available at the CTU (available internally).
2. Suggest an improvement of the calibration methods allowing to work with multiple different cameras in the rig and relaxing the need of previous internal camera calibration.
3. Implement the improvement and investigate the performance on real data from automotive industry.

Bibliography / sources:

[1] R. Hartley, A. Zisserman. Multiple View Geometry in Computer Vision. Cambridge University Press, , 2004.
[2] B. Triggs, P. F. McLauchlan, R. I. Hartley, A. W. Fitzgibbon. Bundle Adjustment - A Modern Synthesis. Workshop on Vision Algorithms 1999: 298-372
[3] S. Agarwal and K. Mierle, et al. Ceres Solver (http://ceres-solver.org/)
[4] CamOdoCal: Automatic Intrinsic and Extrinsic Calibration of a Rig with Multiple Generic Cameras and Odometry (https://github.com/hengli/camodocal)
[5] Bo Li, Lionel Heng, Kevin Köser and Marc Pollefeys. A Multiple-Camera System Calibration Toolbox Using A Feature Descriptor-Based Calibration Pattern Github. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2013
[6] The Kalibr calibration toolbox. (https://github.com/ethz-asl/kalibr)

Name and workplace of master’s thesis supervisor:

doc. Ing. Tomáš Pajdla, Ph.D., Applied Algebra and Geometry, CIIRC

Name and workplace of second master’s thesis supervisor or consultant:

Date of master’s thesis assignment: **13.09.2017** Deadline for master's thesis submission: **09.01.2018**

Assignment valid until: **17.02.2019**

doc. Ing. Tomáš Pajdla, Ph.D.	doc. Ing. Tomáš Svoboda, Ph.D.	prof. Ing. Pavel Ripka, CSc.
Supervisor’s signature	Head of department’s signature	Dean’s signature

Acknowledgements

I would like to express my sincere gratitude to my thesis advisor doc. Ing. Tomáš Pajdla, Ph.D. for his support, advice and overall guidance which allowed me to finish this thesis. I would also like to express my thanks to Ing. Čeněk Albl for his patience while introducing me to his bundle adjustment toolbox. I must also thank to Ing. Martin Matoušek, Ph.D. who explained me how to handle the Up-Drive data.

Last but definitely not least, I would like to express special thanks to my family to which I owe for the possibility to study and which supported me during my studies and work on this thesis.

Author statement for undergraduate thesis

I declare that the presented work was developed independently and that I have listed all sources of information used within it in accordance with methodical instructions for observing the ethical principles in the preparation of university theses.

Prohlášení autora práce

Prohlašuji, že jsem předloženou práci vypracoval samostatně a že jsem uvedl veškeré použité informační zdroje v souladu s Metodickým pokynem o dodržování etických principů při přípravě vysokoškolských závěrečných prací.

Prague, date .... signature

Abstract

A camera calibration is a crucial task in almost any task involving 3D scene reconstruction. With the growth of autonomous industry, the amount of cameras rigidly mounted on a single vehicle or other autonomous robots grows also. All these cameras must be calibrated. This thesis proposes a method of a Rig calibration based on $2D \leftrightarrow 3D$ correspondences, which is capable of complete calibration of all the rigidly connected cameras at once.

The state of the art of partial sub-problems and rig calibration itself is reviewed and improvements are proposed. The innovative concept of the Rig of rigs structure is examined and implemented with respect to a real calibration task. Multiple camera systems mounted on real vehicles are calibrated. The results of the calibration show that the proposed Rig calibration method can calibrate cameras.

If the intrinsic calibration is provided then the reprojection errors may be found in subpixel area. The Rig calibration method is capable of full calibration, including the intrinsic calibration. The quality of a full calibration is dependant on the calibration data. The experiments show that the reprojection error of the full calibration do not excess 5 pixels even for the Trifocal camera system, which is hard to calibrate due to the narrow field of view of the cameras.

The condition of having $2D \leftrightarrow 3D$ correspondences restricts the use of the method. We believe that the extension to the calibration without measured $3D$ points based only on the image correspondences should be possible. Such extension would widen the field of possible applications greatly.

Keywords: computer vision, camera calibration, multiple camera rig, bundle adjustment , intrinsic calibration, automotive

Abstrakt

Kalibrace kamer je zásadní součástí prakticky každé úlohy věnující se 3D rekonstrukci scén. S růstem autonomního průmyslu, roste zároveň i počet kamer, jež jsou připevněny na jednotlivých vozidlech nebo jiných autonomních robotech. Všechny tyto kamery musí být zkalibrovány. Tato práce navrhuje metodu Kalibrace soustavy kamer jež je na základě $2D \leftrightarrow 3D$ korespondencí schopna kalibrace všech pevně spojených kamer najednou.

Současná řešení jednotlivých podproblémů i samotné kalibrace soustavy kamer jsou zkoumány a na jejich základě je navrženo vylepšení. Inovativní koncept struktury soustavy soustav je prozkoumán a a implementován s přihlédnutím k reálnému problému. Vícekamerové systémy připevněné na reálných vozidlech jsou pomocí něj zkalibrovány. Výsledky kalibrace ukazují, že navržená metoda Kalibrace soustavy je schopna kalibrovat kamery.

Pokud jsou dodány interní kalibrace kamer, reprojekční chyba se pohybuje v podpixelovém měřítku. Metoda kalibrace soustavy kamer je také schopna kompletní kalibrace, včetně interní kalibrace. Kvalita kompletní kalibrace je závislá na kvalitě kalibračních dat. Experimenty ukazují, že reprojekční chyba kompletní kalibrace nepřesahuje 5 pixelů a to ani pro Trifokální kamerový systém jež je obtížné zkalibrovat kvůli úzkému zornému poli.

Podmínka použití $2D \leftrightarrow 3D$ korespondencí limituje použitelnost metody. Věříme, že rozšíření na kalibraci předem bez změřených 3D bodů, tedy pouze z korespondencí mezi snímky, je možná. takové rozšíření by výrazně rozšířilo možnosti aplikace této metody.

Klíčová slova: počítačové vidění, kalibrace kamer, soustava kamer, vyrovnání svazků, interní kalibrace, automobilový průmysl

Contents

1 Introduction **2**
- 1.1 Contribution . . . 2
- 1.2 Structure . . . 2

2 State of the art review **4**
- 2.1 Camera models . . . 4
- 2.2 Kalibr radtan camera model . . . 5
- 2.3 Camera Calibration . . . 7
- 2.4 Bundle Adjustment . . . 7
- 2.5 Camera rig . . . 8
- 2.6 Relative pose in rigid rigs . . . 8
- 2.7 N-Camera calibration pipeline . . . 9
  - 2.7.1 Mandatory prior knowledge . . . 10
  - 2.7.2 Provided pipeline . . . 10

3 Proposed solution **13**
- 3.1 Rig of rigs . . . 13
- 3.2 Additional requirements . . . 13
- 3.3 Camera rig structure . . . 14
- 3.4 The detailed description of the proposed solution . . . 15
  - 3.4.1 Main rig calibration . . . 15
  - 3.4.2 Secondary rig calibration . . . 17
  - 3.4.3 Merging the Main and the Secondary rigs . . . 18
- 3.5 Up-Drive camera rig calibration . . . 19
  - 3.5.1 Overview of calibrated systems . . . 19
  - 3.5.2 Rig problem formulation . . . 21
  - 3.5.3 Pipeline structure . . . 21
  - 3.5.4 Observations . . . 21

4 Up-Drive Implementation **22**
- 4.1 The algorithm structure . . . 22
- 4.2 User manual . . . 22
  - 4.2.1 The calibration environment . . . 22
  - 4.2.2 configuration files . . . 23
- 4.3 Documentation . . . 23
  - 4.3.1 Matlab implementation . . . 24
  - 4.3.2 C++ implementation . . . 24

5 Experiments **26**
- 5.1 Environment . . . 26
  - 5.1.1 Cameras . . . 26
  - 5.1.2 Vehicle and Cameras . . . 26
  - 5.1.3 Calibration room . . . 27
  - 5.1.4 Data sessions . . . 27

5.1.5 Frame acquisition . . . 27
5.1.6 Data set . . . 28
5.2 The performed experiments . . . 29
5.2.1 Goal of task specific experiments . . . 29
5.2.2 Goal of the complex experiment . . . 29
5.2.3 Task specific experiment report . . . 29
5.2.4 Complex experiment report . . . 30

6 Results **33**
6.1 The Absolute pose and the Rig calibration comparison . . . 34
6.2 Comparison of the data sets . . . 40
6.3 3D scene reconstruction . . . 40

7 Conclusion **43**
7.1 Future improvements and open questions . . . 44

Bibliography **45**

1 Introduction

The problem of camera calibration is a crucial task in many scene perception tasks. Camera calibration is a mandatory step in measurements and scene reconstruction problems. Cameras showed up to be a cheap and valuable asset in many fields of robotics often with unreachable potential among the other sensors [1]. Therefore, it is not surprising that multiple cameras are being used to enhance task specific perception or provide the general understanding of the surroundings. This led up to a scale where having 4 to 8 cameras on a single body is no excess. With huge deployment of cameras in almost any field concerning autonomous behaviour or automatic reasoning, there is an increased demand on the speed of calibration while keeping the high standards of assisted camera calibrations. Such demands are obviously not easy to meet. One of the approaches would be to calibrate all the cameras attached to a single rigid body at once. Such idea aims to speed up the process of calibration especially in terms of required assistance. The improvement may be achieved exploiting the fact that cameras are not independent to each other, meaning that there exists a rigid rig connecting all of the cameras and thus the relative pose of those cameras does not change in time.

This thesis is providing solution for such auto-calibration with focus on the calibration of autonomous vehicles, especially cars. The solution is capable of full calibration without using any calibration boards nor other specific actions other than the vehicle's movement. The data for calibration are acquired by parking a car into parking slot with 3D measured markers on the walls.

1.1 Contribution

The contribution of proposed work is mainly in two areas. The proposed solution is simplifying the process of calibration of complex camera systems by allowing the calibration of all the cameras at once, while the structure of proposed algorithm is keeping a decent abstraction to cover various different camera systems. The proposed algorithm is very flexible in incorporation of any prior knowledge about the camera system.

The second important advantage of simultaneous calibration of multiple cameras which are mounted on a rigid rig is that proposed solution allows to utilize the correlations between their movement in time to enhance the precision of calibration in general but also works as greedy algorithm for cameras which would be uncalibratable as stand-alone cameras.

1.2 Structure

The thesis is divided into several chapters. First the basics of computer vision, the more advanced concepts, the state of the art review of the problem and key subproblems, which must be handled, may be found in Chapter 2. Chapter 3 starts with a general overview of proposed solution, then presents the solution in the full depth and finishes with an example how may the solution be used. The technical details of implementation may be found in Chapter 4. The experiments which helped shaping of

the algorithm and which shows the quality of the proposed solution are explained in Chapter 5 and their results discussed in Chapter 6. The open questions possible future improvements, advantages and disadvantages are discussed in Chapter 7.

2 State of the art review

The cornerstones of proposed solution are selection of proper camera model, its calibration, relative poses of cameras connected to rigid rig and bundle adjustment. All of those mentioned are discussed in this chapter.

A great first step to the state of the art is the book Multiple View Geometry in Computer Vision by Hartley and Zisserman [2]. It provides a comprehensive overview of overall computer vision problems, geometry and known minimal form solutions. The book gives great insight into single view geometry as well as into two-view Geometry (also known as Stereo vision). As authors admits: "The research in the area of auto-calibration is still quite active and better method than those described in this chapter may yet be developed" (Chapter 19 [2]). The potential problem of [2] is that it does almost exclusively works with the pinhole cameras only. Such approach provides a great simplification of general cases which makes the book easy to understand. The simplification comes with a considerable cost, the simplified version is often far-fetched from the real applications. Especially, concerning the field of wide-angle cameras which are nowadays, in case of automotive, the most used cameras i.e. to monitor the surroundings of a vehicle. As been indicated, the main focus of [2] is in explaining the basics and the background of computer vision geometry. It also provides some insight regarding N-view camera calibration based on planes and theirs hommographie which is not suitable for this thesis.

2.1 Camera models

There are multiple ways to model and estimate the camera parameters. This section presents the basic concept of camera. The camera model used in this work is more complicated and is described in Section 2.2.

Whenever a camera takes a photography, it can be viewed as a projection of a 3D scene into a 2D plane. Such projection may be expressed in a matrix form $P \in \mathbb{R}^{3\times4}$:

$$\lambda u = PX$$

where $u \in \mathbb{R}^2$ is the resulting projection in pixels, $X \in \mathbb{R}^3$ is the observed object in scene coordinates and the λ is a scale factor. The matrix P is also known as the *Projection matrix*. The Projection matrix may be decomposed into following forms:

$$P = K \begin{bmatrix} R & t \end{bmatrix} = KR \begin{bmatrix} I & -C \end{bmatrix}$$

where $K \in \mathbb{R}^{3\times3}$ is a regular *calibration matrix*, $R \in \mathbb{R}^{3\times3}$ is a *Rotation matrix*, $t \in \mathbb{R}^{3\times1}$ is a *camera translation vector* in camera coordinates and $C \in \mathbb{R}^{3\times1}$ is an *optical center* in the scene's coordinate system. The rotation matrix R and either translation of the camera t or camera center C (depends on coordinate system of the context) are also known as the *extrinsic parameters* or the *camera pose*, where the calibration matrix K is composed of the *intrinsic parameters*. Figure 2.1 shows the geometric meaning of described parts of the camera.

The projection matrix represents a simple camera model known as *Standard projective camera model*. The standard projective camera model may be understood as a

description of the set of rays X which intersects in camera center C. The 2D mapped image points u are captured at the intersection of ray X and an arbitrary plane π perpendicular to optical axis. The calibration matrix K may be understood as both description of lenses through which the rays travels to the plane and are modifying them by its optical properties, and the difference between scene's and π's coordinate system. The intersection of an optical axis and the plane π is known as the *center of projection*, denoted as π_0. The distance of a plane π to a camera center C defines the scale which is proportional to focal length f of the camera. For more details concerning the camera models see [2].

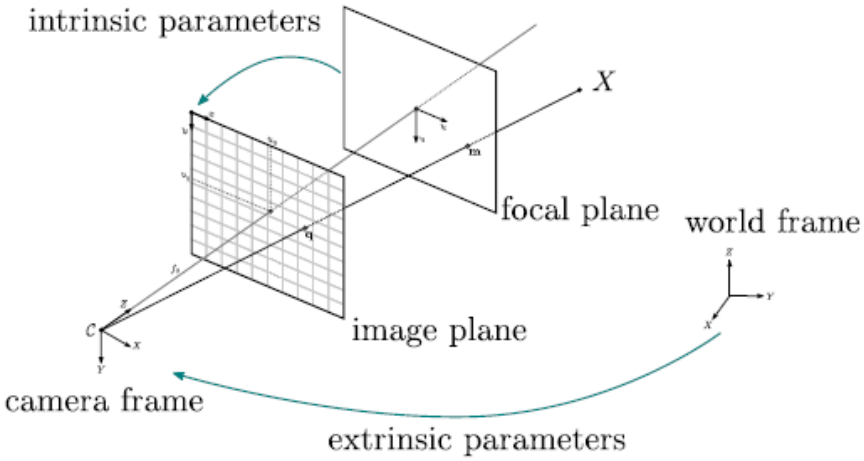


Figure 2.1 A geometric meaning of the standard projective camera model's parameters [3].

2.2 Kalibr radtan camera model

The correct choice of a camera model is a crucial component of any CV's geometry problem. The most basic camera model is the pinhole camera model described by projection matrix P. Such camera model is suitable for rough approximation. The most usual case is that the camera lenses do contain some form of a distortion which must be taken into account in order to achieve the acceptable accuracy. Based on Brown, 1971 [4] there are many different variations [5], [6] or [7] which are modeling not just the radial distortion of lenses but also the tangential distortion caused by the imprecise positioning of a perception layer (usually image sensors CMOS or CCD [8]). The distortion is usually modeled as a second order Taylor approximation. Therefore, there are 2 parameters for both distortions,

$$\begin{bmatrix} t_1 & t_2 & r_1 & r_2 \end{bmatrix}$$

which are applied to the result of standard projection afterwards (see Algorithm 1, line 6,7). Work [7] claims that even though modeling using the standard projection enriched with distortion is often precise enough, it is not sufficient for the wide-angle and fish-eye cameras. It is necessary to add an extra *mirror* parameter $\xi \in [0, 1]$ where $\xi = 0$ represents a standard projection and $\xi = 1$ is a parabolic distortion. In this thesis the model of [9] will be used. It is an extension of [7] and work as shown in Algorithm 1

Algorithm 1 Kalibr [9] radtan projection function

1: **function** u = P(X)
2: $Y = K'[R - RC]\begin{bmatrix} X \\ 1 \end{bmatrix} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$
3: $r_z = (z + \xi \left\| Y \right\|)^{-1}$
4: $x_z = x r_z$
5: $y_z = y r_z$
6: $d = t_1(x_z^2 + y_z^2) + t_2(x_z^2 + y_z^2)^2$
7: $p = \begin{bmatrix} x_z + x_z d + 2r_1 x_z y_z + r_2(x_z^2 + y_z^2 + 2x_z^2) \\ y_z + y_z d + 2r_2 x_z y_z + r_1(x_z^2 + y_z^2 + 2y_z^2) \end{bmatrix}$
8: $u = \frac{1}{\lambda}\begin{bmatrix} f_u & 0 & c_u \\ 0 & f_v & c_v \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} p \\ 1 \end{bmatrix}$
9: **end function**

where $\begin{bmatrix} f_u & 0 & c_u \\ 0 & f_v & c_v \\ 0 & 0 & 1 \end{bmatrix}$ may be interpreted as a calibration matrix K with rectangular pixels, R and C are the rotation and camera translation respectively. The calibration matrix K' contains another set of camera calibration parameters. This ambiguity of multiple calibration matrices may be solved by setting K' to identity. The purpose of having another calibration matrix is to model the standard cameras which have an additional add-on lenses (i.e. fisheye lenses). The K' is also important to simplify the mapping between different camera models. In presented thesis, K' is fully ambiguous and thus will be treated as identity unless stated otherwise. X is a 3D scene point which is being projected to the image pixel coordinates u. The process of projection may be understood as three consequential step as shown in Algorithm 2.

Algorithm 2 Kalibr [9] radtan projection

1: **function** u = P(X)
2: align world coordinates to orthogonal camera coordinates using R and C
3: undistort ξ, r and t
4: project the undistorted point p to image coordinates using K
5: **end function**

To summarize the camera variables:

- **extrinsic calibration** requires to find 3 parameters of 3D translation t, 3 unknowns of 3D rotation R, total 6 parameters
- **intrinsic calibration** consist of 5 parameters $\begin{bmatrix} \xi & f_u & f_v & c_u & c_v \end{bmatrix}$
- **distortion calibration** consist of 4 parameters $\begin{bmatrix} t_1 & t_2 & r_1 & r_2 \end{bmatrix}$

In total such camera model contains 15 variables while assuming rectangular pixels. Or 21 if K' is incorporated as upper triangular calibration matrix.

2.3 Camera Calibration

Obtaining the camera parameters given the set of measurements is known as camera calibration. There are multiple approaches to camera calibration regarding the complexity of calibration. As a model of camera may be divided into different parts (intrinsic, extrinsic) there are different methods to calibrate a specific part of camera given the rest of the parameters. In case of a standard pinhole camera model there is a set of well defined P-n-P (*Perspective-n-Point*) problems with known minimal solutions [2]. However, in cases where other than the pinhole camera model is deployed, the methods needs to be adjusted as well. Thus, for cases with unknown distortions, the P-n-P methods are not suitable. Luckily, there is a solution of an absolute pose problem with unknown radial distortion [10] based on solving polynomial equations, which provides fast computation suitable for a model estimation using the RANSAC [11]. This approach is used to estimate the intrinsic calibration of the cameras in this work.

The calibration algorithms may be also divided by amount of assistance which is needed. The difference is most significant especially in the intrinsic calibration where the most common [12], [13] approach is to capture the images of known and precise calibration board under various angles and positions which cover the whole field of view. The calibration is estimated based on hommographies of the boards in the images. The quality of this approach is usually good and dependant on how well does the images cover the field of view. The main drawback of this approach is the acquisition of the board images. It must to be done manually and it turns out to be quite expensive, especially time-wise if done properly. Obviously, there is a great demand on the simplification of this process. The answer to this demand may cover self-calibration methods. In general, self-calibration is a harder problem since there are no guaranteed hommographies. As [13] claims, that the current state of the art self-calibration methods perform worse than the board calibrations ones. The most common approach to self-calibration is to establish correspondences between the images of specific a scene and based on those correspondences to propose an initial model which is possibly refined using Bundle Adjustment methods [14].

2.4 Bundle Adjustment

The Bundle Adjustment (BA) is an optimization technique. It is a process to find the Maximum Likelihood estimation of all the scene constrains given the data measurements. More formally:

$$\min_{P_{ij},X_k} \sum_{i,j,k} \|P_{ij}X_k - u_{ijk}\|^2 \tag{2.1}$$

where P_{ij} stands for a camera i at a frame j, X_k represents the k^{th} 3D object and u_{ijk} stands for an observation of X_k in an image taken by the camera i at the timeframe j which is being calibrated. The BA methods do iteratively descent to local optima of the task using well-known gradient-descent techniques such as Levenberg-Marquardt [15]. Therefore, it can provide the local optimization only.

The current computer vision problems are usually very large and a simple gradient descent methods which requires the computations of inverses of these large, problem defining matrices would be unfeasible in practical use. The state of the art BA tools are more sophisticated and use set of clever techniques to utilize the specific form of the data structure.

Almost every state of the art scene reconstruction methods do use some form of BA where most of them uses it as the final tool to refine the results.

A great introduction to the problematic of BA is summarized in [16] and [2] also. The [2] provides comprehensive information about what BA actually is. The focus of [16] is in explaining how to actually handle the optimization. The [16] is a survey of the theory and the methods and thus does not provide a a closed form solution. The survey is quite thorough and justifies usage of specific selections such as Levenberg–Marquardt [15] iteration technique or the benefits of Cholesky decomposition.
The efficient implementation of the BA is quite complex task with high demands on mathematical insight as well as the programming skills. The BA used in this work is therefore an extension of existing BA toolbox [17]. Core of the bundle adjustment is provided by the Ceres solver [18], which is a general optimization solver designed by Google. The BA of rig calibration proposed in presented thesis is an extension of [17] original design.

2.5 Camera rig

In the most general case, the calibration problem consists of N independent cameras and therefore, it models them as individuals moving in the world independently as well. However in many cases, those cameras are placed on a firm rig which does not change during the time. Therefore, we can say that those cameras are bound by fixed relative pose. If the relative pose is known and so is exterior calibration of single camera, the rest of cameras may be derived as well. Therefore, instead of finding N different camera poses at every time $t_1 \dots t_m$ (which is $N \times M$ poses total), it is sufficient to find N relative poses and rig pose at every time $t_1 \dots t_m$ (which is $N+M$ poses total only). However, that is not the only reason to choose this approach, such problem formulation does also sort of "put all eggs into one basket". Which means that all cameras do influence all other cameras, so if there are a few cameras with insufficient amount or bad quality data the other cameras can still provide enough support to find reasonable result and thus calibrate otherwise an uncalibratable camera. Although it is important to keep in mind that such "badly conditioned" cameras do decrease precision of the "well conditioned" ones and if there is more than critical amount of them, the whole rig calibration may fail. As may be seen in fig. 2.2, the relative pose of cameras to the rig base does not change, thus only the rig base to the world coordinate system needs to be updated.

2.6 Relative pose in rigid rigs

The core motivation to express cameras mounted on a rigid body as a a single entity is to remove ambiguity caused by multiple cameras sharing the same movement in time. That makes sense from multiple point of views, first is the simplification of computations, where instead of computing nm poses, it is necessary to compute just $n + m$ which in larger scale has a considerable impact. The second good reason is to make the estimation more robust by finding the consensus among all the m mounted cameras at given timeframe.

To utilize the structure of a rig, a formal representation is needed. The proposed solution of [19] is used in this work. A shift of coordinate system relative to the pose

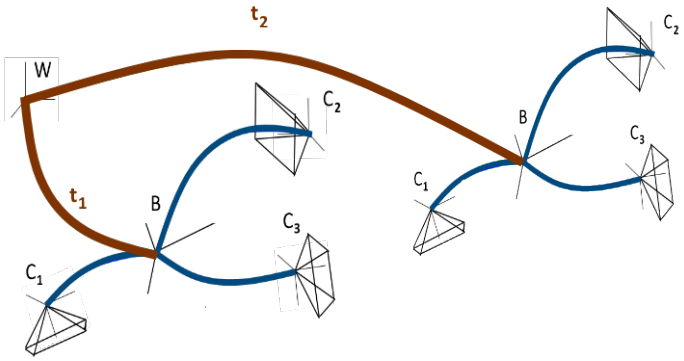


Figure 2.2 An example of a rig containing 3 cameras. World coordinate system is marked as W, Rig's Base coordinate system B, camera coordinate system C_i. World to Rig Base transformation (brown) in time t_1 and t_2. Rig Base to camera transformation (blue).

of two arbitrary cameras may be expressed as an similarity transformation $T \in \mathbb{R}^{4\times4}$

$$T = \begin{bmatrix} \lambda R & C \\ 0^T & 1 \end{bmatrix} \tag{2.2}$$

where $R \in \mathbb{R}^{3\times3}$ rotation matrix and $C \in \mathbb{R}^{3\times1}$ is the translation. $\lambda \in \mathbb{R}$ is a scaling factor, if both the cameras share the magnitude of coordinate system then $\lambda = 1$ and the T is called *Euclidean transformation*. Such notation allows simple concatenation of multiple transformations by simple multiplication of T. That is extremely convenient when the camera rig is introduced. It is essential to retrieve the absolute pose of cameras efficiently as it is usually performed many times. Obtaining the T from $\{R, C\}$ is quite straightforward, however to decompose T into $\{R, C\}$ may be more challenging. The [19] provides a comprehensive description of manipulation with T.

To illustrate the concept of transformations see an example of a rig containing 3 cameras in Figure 2.3. The camera rig exterior calibration is defined as set of $B2C_i$ transformations, and the $W2B$ transformation. To retrieve the pose of an arbitrary camera C_i of the rig, which corresponds to transformation T from world coordinate system to C_i's orthogonal coordinate system, the simple concatenation of $W2B$ and $B2C_1$ transformations only is needed.

$$W2C_i = W2B \cdot B2C_i$$

It is important to note that the calibration K is not involved in this transformation as only the exterior pose of cameras is handled. The rig system replaces the extrinsic parameters in Figure 2.1 only.

2.7 N-Camera calibration pipeline

Calibration of multiple cameras at once is a method how to speed up the process of calibration. Many has been done in this field and this thesis is using the implementation provided by the thesis advisor, doc. Ing. Tomáš Pajdla, Ph.D.

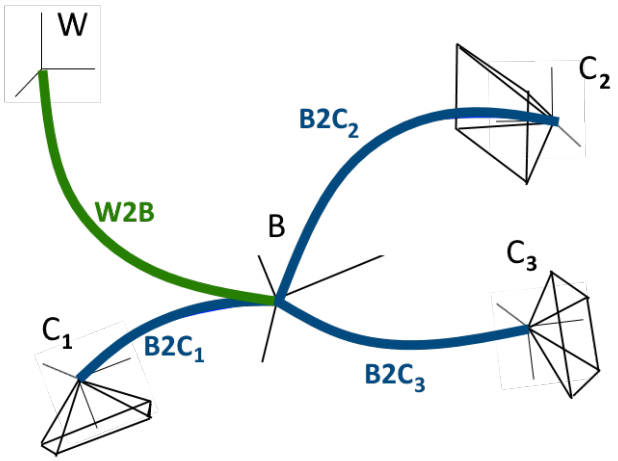


Figure 2.3 An example of a rig containing 3 cameras. World coordinate system is marked as W, Rig's Base coordinate system B, camera coordinate system C_i. World to Rig Base transformation (green). Rig Base to camera transformation (blue).

2.7.1 Mandatory prior knowledge

The following things are necessary to use the N-camera calibration algorithm. The most obvious and critical knowledge are the images. It is expected to have sufficient amount of images from all cameras which are being calibrated. The images are organized into so called Time–frames groups. There are, of course, standard demands on images from Computer vision point of view like low-noise, good resolution etc. On the other hand, it is not necessary to have images from all cameras in any single Timeframe group (for instance, if a car is driving into calibration room backwards, the front cameras cannot see any calibration markers and thus they add nothing to the calibration precision and therefore, they are not necessary at all).

2.7.2 Provided pipeline

In this section the entire pipeline of calibration will be described. Besides some initialization steps like correspondence detection, the algorithm can be divided into 2 main steps.

Single camera calibration

A camera calibration consists of 2 consecutive steps. The division is important in case when the intrinsic calibration is given.

Internal calibration of a camera This step solves absolute pose problem using RANSAC [11] for each $C_1 \dots C_n$ camera, $I_1 \dots I_m$ image pair independently.

Even though the cameras may be attached to rigid body hence they do share properties with respect to camera poses, what they do not share are the internal parameters. Therefore, every camera must be treated independently. The internal calibration is an initial step and thus there is no pose or any other information given. In such circumstances the problem is a classical absolute pose N-view problem [2]. The camera model

used is a Division model [20] which is similar to the Kalibr radtan model.

External calibration of a camera This step uses internal calibration to estimate camera pose from image correspondences. Resulting in absolute pose estimation with given initial calibration. In this state of the intrinsic calibration for every camera is known. Similarly to the intrinsic calibration, the cameras are treated as independent to each other. To estimate the camera pose, whole camera resection must be computed. Due to complicated undistortion of Kalibr model, it is not efficient do solve the resection problem [2] in this form, therefore as an initial guess of camera pose a classical pinhole camera model is used instead of the kalibr model. After that the estimated pinhole pose is combined with given intrinsic parameters and as such each pair of camera and image is then Bundle Adjusted alone with complete gauge freedom regarding the camera. The reprojection error is measured afterwards and outliers are removed. The threshold must be very generous, to remove true outliers only. at this point a relatively precise calibrations are available for every main camera in every positions.

Camera rig formulation

The rig is constructed based on the strongest relative pose between the Main cameras. The relative poses are evaluated based on the quality of reprojection errors of the observations in that area. Such pairwise camera connections forms a fully connected simple graph. The camera rig is formed by a minimum spanning tree of the graph. At this point, the camera positions are known as well as their intrinsic calibration. Given the synchronization of the frames it is possible to estimate a rig connecting all the cameras for a single time-frame by their relative pose transformations. A Transformation T consist of a translation t and a rotation R. It can be written in a matrix form:

$$T = \begin{bmatrix} R & -Rt \\ 0^T & 1 \end{bmatrix} \tag{2.3}$$

R is a 3x3 rotation matrix and t is column 3-vector.

To T be valid transformation between camera C_i and C_j, it must hold following:

$$T_{W,j} = T_{W,i} \cdot T_{i,j} \tag{2.4}$$

where $T_{W,x}$ is a transformation from the world's coordinate system W to coordinate system of camera x (x marks an arbitrary camera).

The rig is considered as constructed when all the cameras are added to the structure. A single camera may be added to the structure, by a transformation from any other camera which is already in the graph. Each step of this gluing algorithm has to choose from multiple transformations (example: if structure contains 3 cameras and 4^{th} is to be added, then there are 3 possible transformations to append it). The choice may be random but it makes sense to use the "strongest" connection between two cameras C_i,C_j which will be discussed later. Having m different frames it is possible to estimate $R_1 \dots R_m$ of such rigs.

This problem may also be reformulated into a graph as follows. The nodes of the graph G represents the cameras where the edges represents the relative poses. If a camera pair (C_i, C_j) is present at frame F_k then there is an edge e_{ijk} connecting nodes C_i, C_j and its value is equal to quality of the connection. G is fully connected but not simple, since any camera pair C_i, C_j may and should occur in multiple frames. The task

of an arbitrary rig construction is equivalent to the task of finding the spanning tree of G. If the edge values are truly representing the quality then the minimal spanning tree is also the most optimal camera rig given the measurements. See the example of graph construction in Figure 2.4. The values in the example are based on Table 2.1.

In general, the spanning tree obtained as been described is usually not optimal. There are two main reasons. The first is the imprecise measurement which is always present in such tasks. The second is the missing valid true quality metrics to evaluate the values of edges. It is not possible given the observation without knowledge of its error to find the ground truth pose. If there was, then all such computation would be meaningless. The missing metrics of quality is enforcing usage of heuristics to determine which edge to choose. The value of edge e_{ijk} is a minimum of estimated inliers between camera $C_{i,k}$ and camera $C_{j,k}$. The tiebreaker in situations where two different edges share same amount of inliers is the maximal reprojection error. The threshold of inliers is situation dependent. This heuristics might be perceived as simple and naive, on the other hand, more of good quality measurements shall provide more precise results, or at least it is expected to some extend. As been said the obtained rig is most likely not optimal.

	F_1		F_2	
	inliers	max. reproj. Error	inliers	max. reproj. Error
C_1	8	3	10	2
C_2	4	3	7	2
C_3	15	1	8	5
C_4	12	1	13	2

Table 2.1 An example of data for the graph construction.

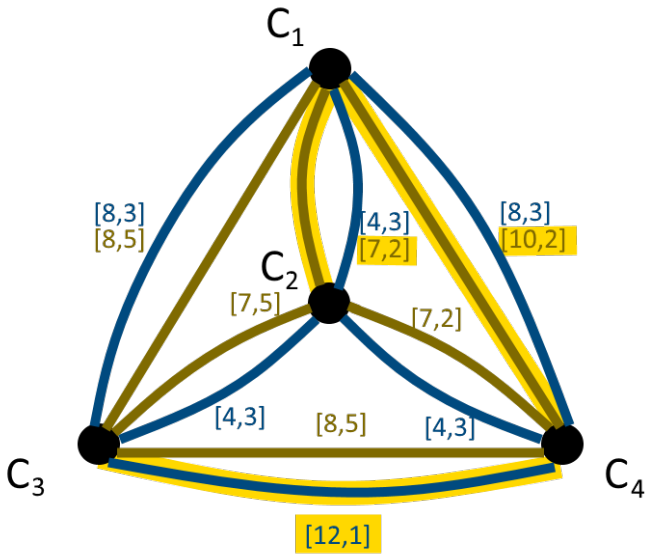


Figure 2.4 An example of graph G (based on table 2.1) of 4 cameras $C_1 \dots C_4$ in two frames (blue) F_1 and (brown) F_2. The resulting spanning tree and thus the rig structure is formed by the yellow edges.

3 Proposed solution

The proposed solution is an extension of Section 2.7 with focus on utilization of rig structure of the calibrated task.

At the beginning the overall concept and the idea behind the solution is presented. Once the idea is clear the description of generic design is provided. And finally, the concrete design of automotive industry specialized setup is described.

3.1 Rig of rigs

The rig of cameras described in Section 2.5 is still quite general. In many cases it is possible to divide cameras into smaller convenient groups. This might be especially useful in cases where it is expected that some sub-group of cameras is likely to be ill conditioned (e.g. specialized cameras, limited field of view). If such cameras can be identified and removed, the remaining cameras would form a sub-rig which is likely to be precise. Let such sub-rig be defined as the **Main rig** and its members be referred as the **Main cameras**. If the Main rig consists of well conditioned cameras only and then the calibration is expected to be rather simple to compute. However, the remaining cameras must be calibrated as well. The remaining cameras are divided into non-overlapping groups (trivial group with single member is also valid) and each group has assigned one or more members of the Main rig. Such groups of remaining cameras and its assigned Main cameras forms a **Secondary rigs** and its members are referred as the **Secondary cameras**.

The Main cameras do provide the stability to a Secondary rig, note that it is also possible that the whole Main rig is part of a Secondary one. Having at least one member of the Main rig in a Secondary one also ensures that if the Main rig and a Secondary rigs are estimated independently, there is always a relation between them which allows to join the relative poses and merge those rigs together. Therefore, cameras that are members of the Main and a Secondary rig are referred as **Connecting cameras**. The cameras which are members of Secondary rig only are referred as **Slave cameras**. Once the rigs are estimated and joined together into single rig, the rig is referred as the **Merged rig**. The details of merging will be discussed later, but the important thing is that the resulting Merged rig's pose is estimated based on the Main rig, thus from well conditioned cameras only and the error of ill conditioned ones is therefore not propagated to the other cameras.

An example of Main rig and Secondary rig structure may be found in Figure 3.1. The Connecting camera shows the relation of the otherwise independent rig structures.

3.2 Additional requirements

The pipeline requirements do not change, however the synchronization of cameras in the time-frames becomes crucial. This is a critical condition and the algorithm is expected to perform only as well as the cameras are synchronized.

Usual case is that there is more that is known about cameras that are being calibrated. Some of such knowledge might be used to enhance the results. Note that following

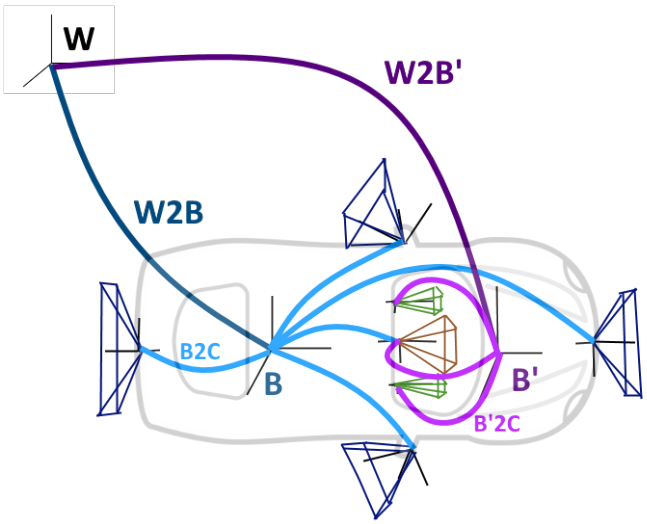


Figure 3.1 An example of the Main (blue) rig and the Secondary rig (magenta) application on a vehicle from the bird's eye perspective. The connecting camera (brown) is a member of both rigs. The slave cameras are green and the Main cameras are blue, with the exception of Connecting camera which is brown.

knowledge is not necessary condition but is highly recommendable.

Great improvement might be achieved if cameras are structured as a rig of rigs (see Section 3.1). Such description is always desirable and close to mandatory for more complex calibration tasks.

Similarly to the N-camera calibration tool, the internal calibration of arbitrary amount of cameras can be utilized. The Rig Calibration algorithm is only as good as are the input images. Hence manufacturer's internal calibration is often at least as good as estimate from any experiment where the quality is not guaranteed. It is assumed that the 2D $\leftrightarrow$ 3D correspondences are given.

3.3 Camera rig structure

Before the calibration method is described, it is necessary to formalize the task. Given:

set of n unknown cameras	$C = \{C_i\};\ i = 1..n$
set of m time-frames	$F = \{F_j\};\ j = 1..m$
set of q 3D objects	$X = \{X_k\};\ k = 1..q$
set of nm images taken by C at F	$I = \{I_{i,j}\};\ i = 1..n; j = 1..m$
set of nmq observations of X in images I	$u = \{u_{i,j,k}\};\ i = 1..n; j = 1..m; k = 1..q$
set of cameras C belonging to rig B^l	$C_B^l \subseteq C;\ l = 1..r$
s.t.	$\bigcup_{l=1}^{r} C_B^l = C$
	$C_B^1 \cap C_B^l \neq \emptyset;\ l = 2..r$

The rig B^1 stands for the Main rig and $B^l; \forall l > 1$ may be understood as the Secondary rigs. The task is to find the following unknowns:

set of n camera intrinsic parameters	$C_i^{in};\ i = 1..n$
set of nm camera extrinsic parameters	$C_{i,j}^{ex};\ i = 1..n; j = 1..m$
set of m rig poses	$B_j^0;\ i = 1..m$
set of m world to rig coord. system transformations	$W2B_j^0;\ j = 1..m$
set of n rig to camera coord. system transformations	$B^0 2C_i;\ i = 1..n$
set of ml rig poses	$B_j^l;\ j = 1..m; l = 1..r$
set of ml world to rig coord. system transformations	$W2B_j^l; j = 1..m; l = 1..r$
set of nl rig to camera coord. system transformations	$B^l 2C_i;\ l = 1..r; C_i \in C_B^l$

The camera parameter's details may be found in Section 2.2. The rig 0 may be understood as the Merged rig. The merged rig may be obtained trough computation of the Main rig and all the Secondary rigs. It is obvious that the combination of $W2B_j$ and $B2C_i$ covers the camera extrinsic parameters $C_{i,j}^{ex}$. Therefore, once the rig representation is computed, the $C_{i,j}^{ex}$ are no longer updated and must be calcilated from the rig representation.

3.4 The detailed description of the proposed solution

This section describes the improvements proposed to enhance the quality of rig calibration.

3.4.1 Main rig calibration

Calibration of the Main rig provides the exterior calibration of the final Merged rig, therefore the demands on quality of the Main rig are higher, compared to the Secondary rigs.

Internal calibration of Main cameras

At this stage, every camera is perceived as a single camera with given set of images $I_1 \dots I_m$ and its 2D to 3D correspondences. The procedure of single camera calibration consists of multiple steps. The very first step is to acquire absolute pose estimation same as in Section 2.7. Given the results of previous step there are m proposals of internal calibration. The experiments showed that median values of internal parameters are the most stable and thus are used as internal parameters. Now, the proposals are converted to the Kalibr radtan models. The Camera candidates $C_1 \dots C_m$ are then optimized using BA. They are bundled first independently as C_i camera absolute pose with gauge restriction applied to external parameters which may not change. Secondary a complete system of $C_1 \dots C_m$ is bundled together. The gauge freedom of bundle adjustment is set to share the intrinsic parameters of cameras. After this procedure the intrinsic parameters for every camera are estimated.

Construction of the initial Main rig

The main rig is constructed in similar way. The difference may be found in the used data structure. In this case the subset consisting of Main cameras only is used.

Local Optimization of the Main rig

Local optimization is performed by BA. The general BA has been described in Section 2.4, therefore, this section focuses on the data structure and what the bundled problem looks like.

The BA consist of camera descriptions C, observations u in images taken by these cameras and 3D objects X which are being observed. In case of this optimization, the u to X correspondences are given and they are the source of possible correction which BA may provide and thus neither of them is being optimized. The desired output of BA is a rig structure, which consists of the relative poses to all the Main cameras, the intrinsic parameters of every camera and the pose of the rig in every frame. Therefore, every camera consist of 27 parameters:

- Six parameters of world-to-rig transformation $W2B$. Three parameters are the rotation R_{W2B} in Rodrigues' formula representation, and the other triplet is the translation t_{W2B} from world origin to rig's base.
- Six parameters of rig-to-camera transformation $B2C$. Three parameters are the rotation R_{B2C} in Rodrigues' formula representation, and the other triplet is the translation t_{B2C} from rig's base to bundled camera.
- Six parameters of upper triangular calibration matrix $K = \begin{bmatrix} k_1 & k_2 & k_3 \\ 0 & k_4 & k_5 \\ 0 & 0 & k_6 \end{bmatrix}$.
- Five Kalibr radtan model parameters $\begin{bmatrix} \xi & f_u & f_v & c_u & c_v \end{bmatrix}$.
- Four distortion parameters $\begin{bmatrix} t_1 & t_2 & r_1 & r_2 \end{bmatrix}$.

If n denotes the number of frames and m denotes the number of cameras in the rig, there are n times m cameras where each consists of 27 parameters. Therefore, there are **27nm** camera parameters to be optimized.

If the scene is composed of k different 3D objects X, there are also **3k** parameters representing 3D coordinates of all the objects.

The total amount of observations u is not straightforward to enumerate because it is unknown how many of X are being observed at a specific camera and time. The maximum of observed objects in specific image is always k which is the total number of objects in the scene. Therefore, there are nmk observations at most. Every observation consists of x and y pixel coordinate, so there are Ω **(2nmk)** observations.

To summarize the size of the problem, there are $\mathbf{P = 27mn + 3k + \Omega\ (2nmk)}$ parameters P in total.

From the rig description, it is clear that there are many ambiguities since the $W2B$ transformation of a specific frame is presented m times while it is a single entity. Furthermore as been stated before, the 2D $\leftrightarrow$ 3D correspondences are not being optimized and thus are constant. There comes the gauge freedom manipulation to link parameters where needed:

- The *W2B* transformation is shared between all the cameras in the same frame.
- The all other camera parameters are shared among the same camera in every frame.

- 3D objects X may not change.
- 2D observations u may not change.

Such constrains simplify the problem greatly and the remaining variables may be expressed as

$$V = 6n + 21m$$

where n is number of frames and m is the number of the main cameras.

To evaluate the BA iterations, it is necessary to measure quality of given state of iteration. The evaluation is computed as a norm of residuals caused by reprojection errors across all the observations:

$$e_i = \sqrt{(u_{0i} - u_i)^2 + (v_{0i} - v_i)^2} \tag{3.1}$$

where $\begin{bmatrix} u'_{0i} \\ v'_{0i} \end{bmatrix}$ is the observation and $\begin{bmatrix} u_i \\ v_i \end{bmatrix}$ is the reprojection of a 3D object X_i.

3.4.2 Secondary rig calibration

The secondary rigs are also composed of multiple cameras and the main difference compared to the Main rig is that calibration of the Connecting cameras is given prior to the computation. That may be utilized on multiple stages. The general procedure is very similar and therefore only the differences are presented.

Application of optional prior knowledge

An approximate structure of the secondary rig might be known prior the computation. In that case it makes sense to use this approximate structure instead of estimating it from scratch. That may be particularly convenient if a task-specific cameras with a narrow field of view are used.

For better understanding it is recommended to see an example Section 3.5.1 which is an example setup of so called Trifocal camera [1] which is quite common in nowadays autonomous vehicles.

Calibration of the Slave cameras

In general case there is no other option than solving the absolute pose problem same as with the Main cameras. Both the intrinsic and exterior calibrations are computed using the same pipeline as the main cameras. If there was any information known prior the calibration and the exterior calibration is known, that the absolute pose is computed in same manner with the only exception of locking the exterior calibration and searching in RANSAC the intrinsic part only. The exterior calibration is thus not needed and only the camera-wise BA is performed.

Construction of the initial Secondary rig

The secondary rig is analogous to the Main rig. Yet there is a difference in the construction of the initial rig. The Main rig was constructed based on minimum spanning tree of camera graph. In case of the secondary rig, the experiments showed that the edges connecting two slave cameras are less stable then if the connection is between

one connecting camera and one slave camera. Based on this observation the edges connecting two arbitrary slave cameras are removed and the minimum spanning tree is computed afterwards.

Local optimization of the Secondary rig

The initial secondary rig is treated exactly same as the initial main rig (see Section 3.4.1). Multiple additional approaches were proposed and experimentally tested, but none yielded sufficient improvement in any terms and thus are not used.

3.4.3 Merging the Main and the Secondary rigs

At this stage there are multiple rigs in the space. Prior to the BA, the rigs were dependant to each other due to the connecting cameras. However, after BA it is no longer guaranteed that it is the case any longer and the connecting camera's absolute pose in different rigs may not overlay perfectly. As been stated in the general overview, the Main rig is considered as more reliable and thus its pose is taken as the true pose.

The $B2C$ transformations of the slave cameras which are valid with respect to the Secondary rig's base must be reformulated with respect to the base of the Main rig. Due to the convenient formulation of the transformations, it is possible to chain the transformation by simple multiplication. The example Figure 3.2 shows the process of connection of the Main and single Secondary rig.

For every Secondary rig, there is one of its Connecting cameras used as a Base camera C_B of the rig. Since the Connecting camera is also part of the Main rig then there is a known transformation $B2C_B$. The transformation of an arbitrary Slave camera C_S may be derived as follows:

$$B2C_S = B2C_B \cdot C_B2B' \cdot B'2C_S \tag{3.2}$$

Given:

$$C_B2B' = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{3.3}$$

$$B2C_S = B2C_B \cdot B'2C_S$$

Where $B2X$ denotes transformation from the Base of the Main rig to a camera X and $B'2Y$ denotes transformation from Base of the Secondary rig to a camera Y. If there are multiple secondary rigs, they are processed iteratively. Since the Slave cameras are members of at most one Secondary rig, it is clear that every Slave camera is added exactly once.

The consequence of such possibly nonzero shift of extrinsic parameters without correction of corresponding intrinsic parameters increase the reprojection error of the Secondary rigs. On the other hand, after this step the ambiguities of duplicated Connecting cameras were removed. Due to the increased error, such rig is clearly not optimal.

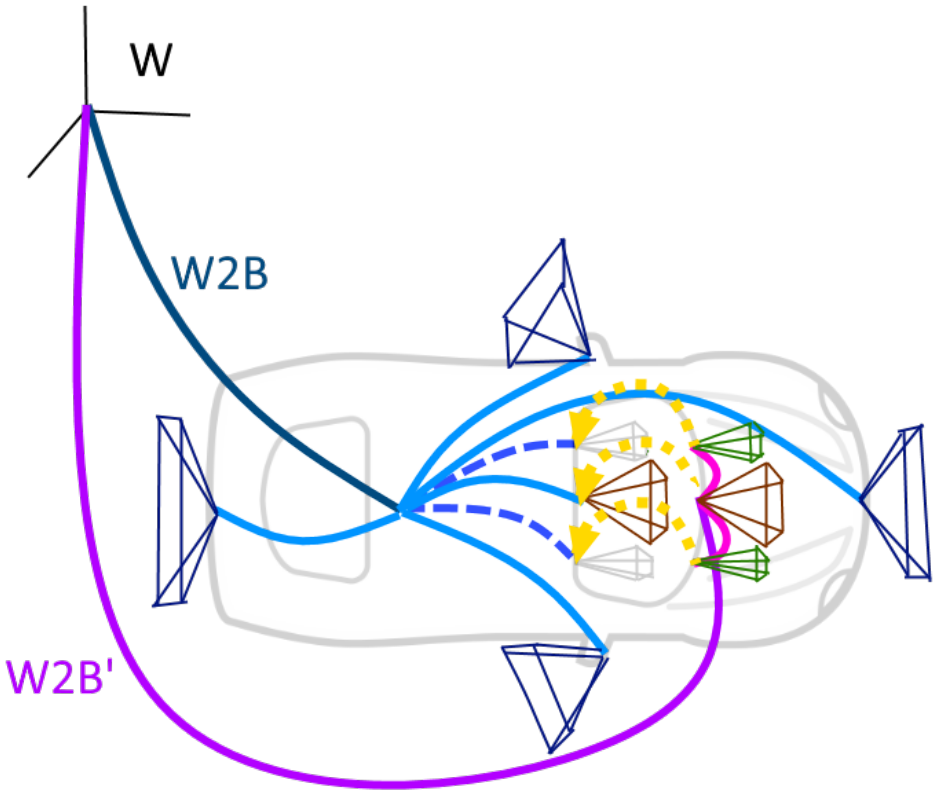


Figure 3.2 A visualization of merge of (blue) the Main rig and the Secondary rig (magenta) trough the Connecting camera (brown). The shift (yellow dotted) of the cameras corresponds to the C_B2B' transformation. The new connections (dashed blue) of the Slave cameras (green) are established.

Local optimization of the Merged rig

The optimization is analogical to the BA of Main rig. See Section 3.4.1 for more details. The expected result is that the Main cameras does not change, while the Slave cameras. On the other hand, the Secondary cameras shall adjust its exterior pose due to the performed Merge shift.

3.5 Up-Drive camera rig calibration

The proposed solution was employed to calibrate multiple four wheel vehicles with similar camera calibration task. The vehicle was equipped with a Topview camera system and a Trifocal camera system. A brief description of the Topview and Trifocal cameras is provided in Section 3.5.1 and more technical details in the Chapter 5. This part describes how is the proposed Rig calibration system employed on an actual task.

3.5.1 Overview of calibrated systems

Trifocal camera system

The Trifocal camera is composed of three different cameras which are usually located close to each other in the area of rear view mirror on the windscreen. (see Figure 3.3.) First camera is a close to 180° field of view and thus is rather stable to the well-known

degenerated cases such as having all the correspondences on a single plane [2]. That makes it a great candidate to be a Connecting camera and lets denote it as the the Master camera. The second and third cameras are quite similar to each other, both with relatively narrow field of view and thus volatile to the single plane problem. Let these two cameras mark as slave cameras. The experiments showed that if the exterior calibration of the Connecting camera is used as the initial estimate of the exterior poses of both the slave cameras the overall stability of auto-calibration has grown rapidly while the precision does not drop and in average it does even improve. Then, the absolute pose problem shrinks to the intrinsic calibration problem only, which is far simpler.

Figure 3.3 An example of a Trifocal hardware.

Topview camera system

The Topview camera system consist of four cameras with close to 180° degree field of view. The experiments showed that larger field of view corresponds with greater stability of results. They are located at each side of a vehicle and together they provide complete 360° view of surroundings (see Figure 3.4).

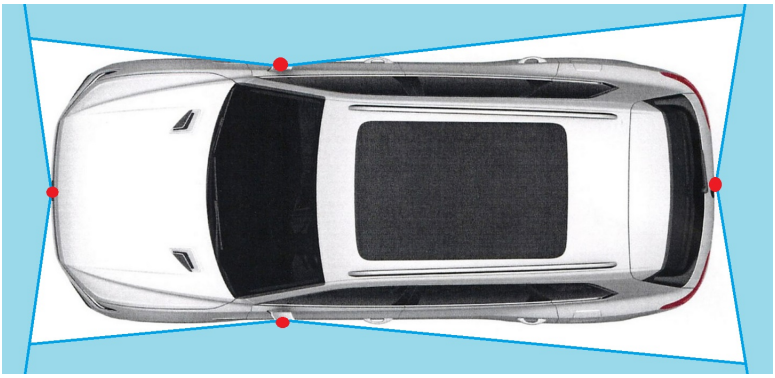


Figure 3.4 An example of a field of view (blue) of a Topview system (red).

3.5.2 Rig problem formulation

There are seven cameras total, five of them are wide-angle cameras (four Topview and the master Trifocal camera) which forms the Main rig. The master Trifocal is the only connecting camera. The Trifocal camera forms the secondary rig. Since all the cameras are now part of some rig and there exist a connection between the Main rig and the Secondary rig (in this case only one Secondary rig), trough the Connecting camera, the problem formulation is valid and complete.

3.5.3 Pipeline structure

Main rig – It is not uncommon that there is a known intrinsic calibration provided by the manufacturer. However the provided calibration is often too imprecise and need to be enhanced. Such intrinsic calibrations may be used to skip the intrinsic parts of the pipeline and the imprecision will be improved in the BA steps. In the Up-Drive case there are two camera systems, where the Main rig contains cameras of both of them. In case that only a subset of the Main cameras is missing the intrinsic calibration of the missing part is computed only. The Main rig calibration is processed as it is described in the previous chapter.

Secondary rig – At this stage the Connecting camera's parameters are found. The Up-Drive Secondary rig is unique in the sense that the cameras are very close to each other (≤ 7 [cm]) and thus it is unlikely that there would be a different locally optimal pose other than the true pose. That gives a freedom to use the Connecting camera's extrinsic parameters as the extrinsic parameters of both the slave cameras. If the intrinsic parameters for the Trifocal cameras are also known, the Secondary camera-wise calibration is complete. The construction of rig of cameras with identical pose is trivial, yet the structure of gauge freedom does slightly change. This step is not the final BA of the whole pipeline, and thus it may be used to optimize only the intrinsic parameters of the Slave cameras, while locking the exterior parameters.

Merged rig – The Merged rig construction is simple and straightforward since the Slave cameras still share the camera pose with the Master camera. The Merged rig is bundled afterwards with full freedom, as been described in previous part.

3.5.4 Observations

It is important to note that the differentiation into the Main rig and the Secondary rig might seem artificial and a single rig would be sufficient, well, in case of known intrinsic calibration of the Trifocal camera it truly is. On the other hand, if the Trifocal camera's intrinsic calibration is not known then the differentiation is vital for the success. The narrow field of view of the Slave cameras makes the absolute pose estimation very unstable and the result is too dependant on perfect calibration data. If the reliable pose of the Master camera is found in the Main rig and it is known that the cameras are close to each other, it allows to prune the search space of possible Slave camera's parameters. The separation of Trifocal camera allows a multi-step Bundle Adjustment and thus robustify the intrinsic calibration of Slave cameras which would not be possible other-wise. Also note that the used heuristics that cameras are close to each other may look different based on the situation. A constrain that cameras must lie on a line or other simple shape may be enforced as well which provides a powerful and universal tool to robustify camera calibrations with non-optimal calibration data.

4 Up-Drive Implementation

The algorithm is implemented in Matlab with combination of C++. It is a part of Camera rig calibration toolbox (CRCT) which is yet to be published. This Chapter explains the program's structure, presents a user guide to explain how to operate with the algorithm and complete documentation.

4.1 The algorithm structure

The algorithm itself is structured into logical parts which corresponds to the description in previous Chapter 3. The parts are structured into the sets of consecutive functions where after every step of the algorithm a partial solution is saved. Such setup has multiple advantages. It simplifies the experimental work where any part can be enforced to be used with the exactly same input given to it and thus it is possible to compare the results independently on previous partial results which may differ due to RANSAC and other stochastic parts. The other advantage is the possibility to repeat only the stochastic parts to verify other results in case the user feels fit to it. Also, in case the full rig calibration is not needed then the partial results may be utilized.

There are multiple possible views on the algorithm. The algorithm will be described from two major points of view. The first view is a view of potential user who would like to use the implementation himself. Since the solution is presented in its generic form there might be a confusion how to do the setup for a specific task. Hence, the user manual is presented. The other possible view is a view of a curious programmer who is interested how were the technical challenges solved. The first view may be found in Section 4.2 and the second in Section 4.3.

4.2 User manual

The calibration itself is rather complicated task, hence to pretend that it is a single button algorithm for generic purposes would be a mistake. Every step of the pipeline has certain amount of parameters which needs to be set in order to work properly. On the other hand, it is the only thing that is required to be updated by the user in order to use the algorithm. This section is explaining the meaning of the parameters and where to find them. This section also describes the environment of the calibration which is expected in order to retrieve the results and where it is possible to find the partial results. The calibration may be used trough the crct interface script, which provides a calibration walk-trough GUI.

4.2.1 The calibration environment

There are two files which contain the parameters of the calibration. It is the *crctpar.m* and *crctparini.m*. The crctparini file contains general settings shared a for a group of specific tasks. It contains general values, expected paths up to folders. The *crctpar* is a the settings for the specific task, it provides the names of the intrinsic calibrations if they are known, which cameras to bundle and from which frames. The details of these

setting will be discussed later. It is expected that every calibration creates a following folder structure. At the top level there are 3 sub-folders called *internal, external, rig* and a file *crctpar.m*. The internal and external folder are very similar. If the same images are used for both intrinsic and exterior calibration the folders regarding the images may be found it is not necessary to duplicate the information. The internal and external folders contains:

- The intrinsic parameters of calibrated cameras. The sub-folder structure containing the intrinsic parameters is specified in the *crctparini.m* file.The actual filenames are stored in *crctpar.m*.

- The *detections* folder which contains the detection results, the 2D $\leftrightarrow$ 3D correpsondencies.

- The *allims* folder containing the images for the intrinsic calibration.

This may be done manually or by the provided setupTool which is a part of the Appendices.

4.2.2 configuration files

There are two configuration files **crctpar** and **crctparini** which allows to control almost any stage of the algotihm. The hierarchy is that the crctpar settings always overrides the crctparini. Therefore, the first is presented the crctparini and the crcrpar follows afterwards.

Crctparini

The initialization file is rarely to be updated, and is recommended to perform the updates on the level of crctpar only. The crctparini specifies all the parameters needed for the computation, such as object detection parameters, plot settings or bundle adjustment configuration selections. If the crctparini file is modified then it should always be a structural change in case of new type of task is to be solved. for instance a calibration task of 4 camera drone system, which will same for different types of drones with 4 equipped cameras.

Crcrpar

At the level of crcrpar a particular task's settings is modified. That includes names of calibration files as strings or Rig of rig structure specified by a bitmap mask. Crctpar also specifies the details, which may vary, that is which cameras of the system are actually being calibrated or whether the intrinsic calibration is provided or is to be computed.

4.3 Documentation

The algorithm is implemented in two different languages and thus is divided into corresponding chapters.

4.3.1 Matlab implementation

The implementation with the exception of Bundle Adjustment is exclusively in Matlab. The rig calibration is relying on an implementation of core functions provided by thesis advisor, doc. Ing. Tomáš Pajdla, Ph.D. The implementation of the rig calibration is divided into multiple functions which may be divided into multiple sets: initiation set, core set, utils and tools set.

The initiation set is not necessary to run the algorithm, however simplifies the procedure of setup of all the dependencies (i.e. setupNewExperiment).

The utils and tools set consist of the support tools, like figure plotting functions, or data interface between the program data structure and used functions or task specific functions (like the function div2radtan which converts the division model to the Kalibr radtan model).

The core set contains multiple functions which are directly called from the crct script. Every such function corresponds to a step of the proposed solution. An exception may be the crcrCTrifocal function which is an camera-wise calibration tailored with respect to the heuristics of Trifocal camera system. The implementation follows the description of provided in Chapter 3.

Data Structure

The calibration data is held by single global variable with multiple fields. The cameras and their frames are forming a matrix where $\{i, j\}$ element contains the calibration of ith camera in j^th frame.

4.3.2 C++ implementation

The C++ part of this work is a plugin to the well-designed CMPBA bundle adjusting tool based on [17]. The CMPBA basically an interface layer to the Ceres optimization solver. The CMPBA s designed to help solve the BA tasks where the Camera, 3D objects and their observations are involved. It also allows to adjust the gauge freedom of the task. The CMPBA may be used on a new task with specific conditions (camera model, gauge freedom settings) by providing the following functions:

- **The reporjection error function** which calculates the residuals of reprojection.
- **The shared parameters function** which solves the shared parameter constrains during the inicialization.
- **The locked parameters function** which specifies which parameters may be optimized.

There are 3 different reprojection functions: *KalibrRadtanReprojectionError, KalibrRadtanReprojectionErrorCameraDetails, RigKalibrRadtanReprojectionError*. Where all of them solves the different bundling task. The CMPBA is designed to using the C++ structures with overloaded operator() which solves the reprojection problem and updates the residuals, see Figure 4.1. an instance of the struct is created for every observation which does not change, during the optimization the operator function takes a set of the current iteration's camera parameters and the 3D point as the input and compares the reprojection with the constant observation. In case of this thesis, the 3D points are constant as well.

```
130  struct RigKalibrRadtanReprojectionError {
131      RigKalibrRadtanReprojectionError(double observed_x, double observed_y)
132          : observed_x(observed_x), observed_y(observed_y) {
133      }
134
135      template <typename T>
136      bool operator()(const T* const eaxW2B, // 0
137          const T* CW2B,  // 3
138          const T* const eaxB2C, //6
139          const T* CB2C, //9
140          const T* K, //12
141          const T* k, //18
142          const T* r, //23  //total 0..26 = 27
143          const T* const point,
144          T* residuals) const { ... }
194
195      double observed_x;
196      double observed_y;
197  };
198
199  #endif
```

Figure 4.1 A header of the RigKalibrRadtanReprojectionError function

Kalibr radtan reprojection error

The function is designed to solve the reprojection problem of the Kalibr radtan camera. The camera is in the single block, hence the computation is faster.

Kalibr radtan reprojection error camera details

The function solves the same problem as Kalibr radtan reprojection error functio in Section 4.3.2. The difference is that the camera is in multiple blocks which slows the computation but allows to modify the gauge freedom of every block independently.

Rig kalibr radtan reprojection error

The function is designed to solve the rig BA. If the parameters $eaxW2B$ (which is a rotation of rig base in Rodrigues' formula) and $CW2B$ (which corresponds to the translation of the rig base) are set to 1_{3x1} and 0_{3x1} respectively, then the result is equal to result obtained by Kalibr radtan reprojection error function in Section 4.3.2.

5 Experiments

This chapter describes the environment of the experiments on which the proposed Camera Rig Calibration algorithm was evaluated. The content of this chapter covers the scene description, camera details, car details with focus on the location of calibrated cameras and frame selection. The execution of the experiment is also part of this chapter, while the actual results may be found in the following Chapter 6.

5.1 Environment

Since the experiments were not performed on synthetic data, the description of the environment is necessary to understand constrains and the context which it provides.

5.1.1 Cameras

In all the experiments, there were seven different cameras. The cameras can be divide into two separate logical groups. One is the Top View group which consist of four wide-angle cameras. The other one is the Trifocal camera group which consist of three cameras. A brief introduction to the camera systems is also in Section 3.5.1.

Trifocal camera

The Trifocal camera is a single device containing three different cameras. It is known that these cameras are oriented in similar direction, thus the optical axes of these cameras are expected to be close to parallel. It is also known that these cameras are close to each other. There is one wide-angle camera (noted as Master camera) and two close to perspective cameras (noted as Slave cameras) with a narrow field of view. Any of the Trifocal cameras provides a raw image files which must be yet demosaiced of the standard Bayer filter [21]. The resolution is around 1 megapixel.

The Topview cameras

The Topview cameras are quite standard RGB cameras with field of view similar to the Master Trifocal camera. the resolution is approx. 1 megapixel (1280×800).

5.1.2 Vehicle and Cameras

All the vehicles used in presented experiment are unspecified cars. Every vehicle has well defined front and read. The vehicles can provide additional measurements, e.g. GPS or local odometry. All of the cameras described in previous section are firmly connected to the testing vehicle and thus they do not change the relative pose in time. The vehicle divides its surroundings into four zones of interest. One at each side, one in the rear and one in the front of vehicle. Each of these zones has its own wide-angle camera. The field of view of its camera does cover the entire zone and has slight overlap to its neighbor zones. Those are the Top View cameras. Two are located on the wing-mirrors and the other two close to license plates. The front zone, as the most important one, is covered by three additional cameras merged into the Trifocal

Camera. It is located close to the rear-view mirror. Those three cameras are focused to different distances. The one focused closest to the vehicle has similar field of view to the TopView cameras. The other two which are focused further have similar field of view to each other. Those cameras are close to be perspective cameras.

For the calibration purposes the important information is:

- The Trifocal cameras are close to each other (≤ 7[cm])
- The Trifocal cameras are expected to be oriented similar way, the possible difference is expected only in pitch direction.
- The Topview cameras share field of view with their neighbors, but not enough to be sufficient for the standalone calibration.
- The Topview cameras are placed to cover every side of the vehicle (front, rear, left, right).
- If a detectable object is in non-occluded focus depth distance in arbitrary direction from the vehicle, at least one of cameras shall detect it.

5.1.3 Calibration room

The calibration room is a single vehicle garage. It is an empty room of cuboid shape of size approximately six times three meters. The walls of calibration room are covered with unique markers (see Figure 5.2) that can be identified and distinguish from each other. There are extra boards added to avoid situations with single plane in the image. These extra boards are located in the four corners of the room and so are on the edges connecting ground to side and the front wall. The rear wall has no such board because the rear wall is the entrance. See Figure 5.1. All extra boards have fixed position and are covered with markers as well. The 3D positions of markers are precisely measured in advance to the experiment. Obtaining the 2D to 3D correspondences in an arbitrary image taken inside the calibration room is rather simple. It is only necessary to detect the IDs of markers in the image and given calibration room look-up-table the correspondences are found. The markers in the image are detected using ellipse fitting, but it is not part of this work. It is assumed that the observations u or the markers are given.

5.1.4 Data sessions

The experiment is composed of two separate sessions. In both scenarios the vehicle slowly drives in and drives out in arbitrary order. There is no difference if the vehicle's starting position is inside the room or outside as long as it drives both in and out. The only difference between session 1 and session 2 is the orientation of the vehicle. In one session the vehicle goes front-first and in the other it goes rear-first.

5.1.5 Frame acquisition

The data set consists of two parts. For practical reasons the calibration of additional sensors is needed. Therefore, there is a given predefined common set of time frames which are mandatory to calibrate. Calibration of additional sensors is not part of this work and does not have any other impact on this work. Besides the given set, there is yet a second part. Since there is not control about the sampling of given set, an additional

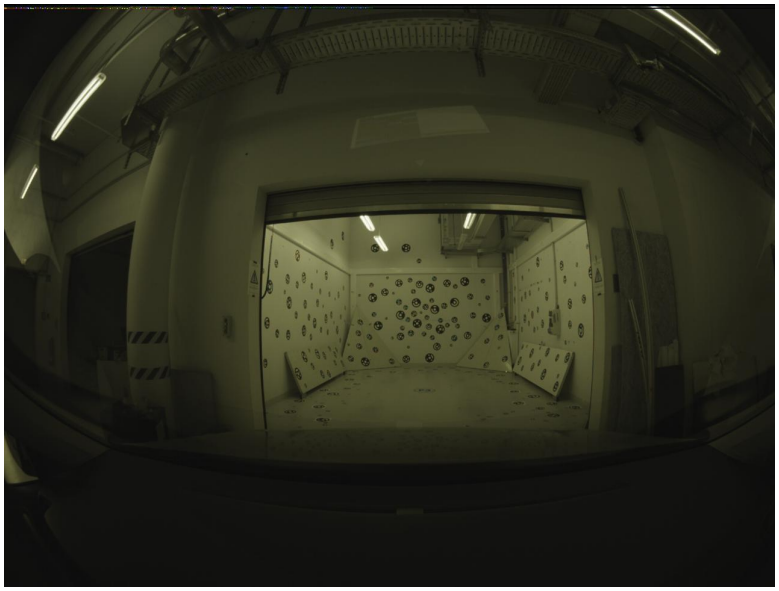


Figure 5.1 A calibration room perceived by the Master Trifocal camera

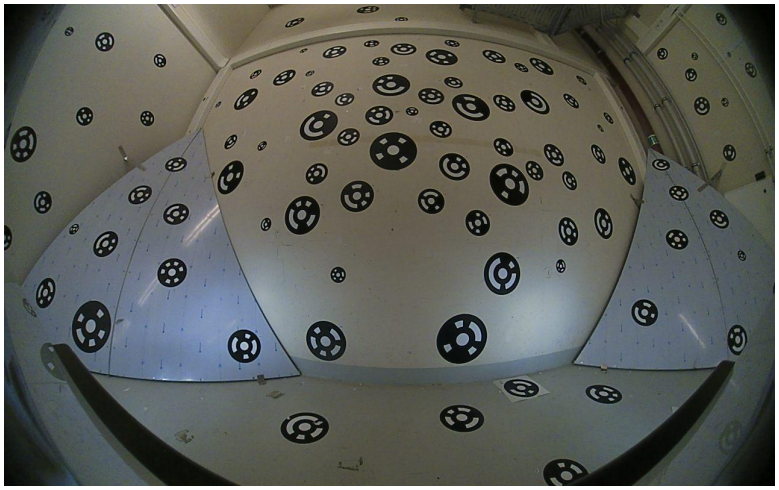


Figure 5.2 A close range image of the markers, shot by the front Topview camera

sampling is needed to guarantee a proper calibration. The sampling method utilizes vehicle's odometry measurements to reconstruct the trajectory and sample equidistantly to cover various poses of calibration. Considering this work only, there is no drawback to not use any given frame and acquire them by additional sampling instead. The given set is mentioned to clarify how is the data set selected.

The frame acquisition turned out to be a tricky part. The camera synchronization is far worse than may be expected, and the difference between the frame-timestamp and first available frame of an arbitrary camera is up to $+/-40$ms. If the vehicle goes with speed of 5km/h, the translation is over 5cm, which might cause immense imprecision, especially in the close range frames.

5.1.6 Data set

The data set is derived from the time frames. The frame acquisition is described in previous chapter. The data set consists of images specified by the time frames across

all the cameras. It is also possible that images for given frame are not available for all the cameras, that is a valid state and solution is obliged handle it.

5.2 The performed experiments

The experiments performed may be divided into two groups. The first may be called Task specific experiments where the goal is to choose the best approach to tackle issues related and to compare multiple proposed solutions to some specific sub-problem. the second is a Complex experiment which shall evaluate how well does the proposed calibration work.

5.2.1 Goal of task specific experiments

There has been made a lot of small independent tests to verify correctness of chosen approach. A great example of such experiment is the experiment regarding heuristics that may be applied to the Trifocal camera system and the relative pose of its cameras. The one option is to lock the camera centers together and pretend that all three cameras share a camera center and may only differ in the rotation. This simplification would lead to increased stability of the solution in cases of the sub-optimal calibration data, where such greedy approach may be the only possibility. It must be validated that this proposed change does not affect the quality of results.

5.2.2 Goal of the complex experiment

The goal is to validate that proposed solution can calibrate cameras. The hypothesis is that the quality of the rig calibration shall outperform the camera independent calibration results. The experiment consists of complex calibration of multiple vehicles where the resulting reporojection errors are compared.

5.2.3 Task specific experiment report

The experiments regarding the Trifocal camera are performed during the Secondary rig computation. It is assumed that the Connecting Master camera has been found. Three different approaches are being evaluated:

- **Fully independent cameras** – The cameras are treated as independent entities which are mounted to the rigid rig.
- **Semi-independent cameras close to each other** – The cameras are treated as independent entities but the initial estimate is shared.
- **Cameras with identical camera center** – The cameras share the camera center.

There is a justification of all the points of view. Treating the cameras as independent is the basic concept which must be examined since it is the most natural and heuristics-free approach. To force cameras to share the camera center is clearly an approximation since no two object may be at the same place given time. The question is whether such approximation is enough precise to be used. The semi-independent cameras approach seem like a middle approach. It is known that the camera centers are close. On the other hand, it is not known whether they are close enough to converge to the ground truth position. In that case the fully independent approach shall be superior. It is important to note that the fully independent cameras may fail to calibrate due to the insufficient calibration data.

Results

As may be seen from the 3D reconstructions in Figure 5.3, Figure 5.4 and Figure 5.5 the semi-independent camera system provides the best results. The fully independent system fails to calibrate seventh camera (magenta). The problem of the identical center is that the translation of the cameras is compensated trough the rotation. Hence, the calibration provides acceptable reprojection errors on the calibration data, however the reprojection error of the more distant objects in real use would be too great.

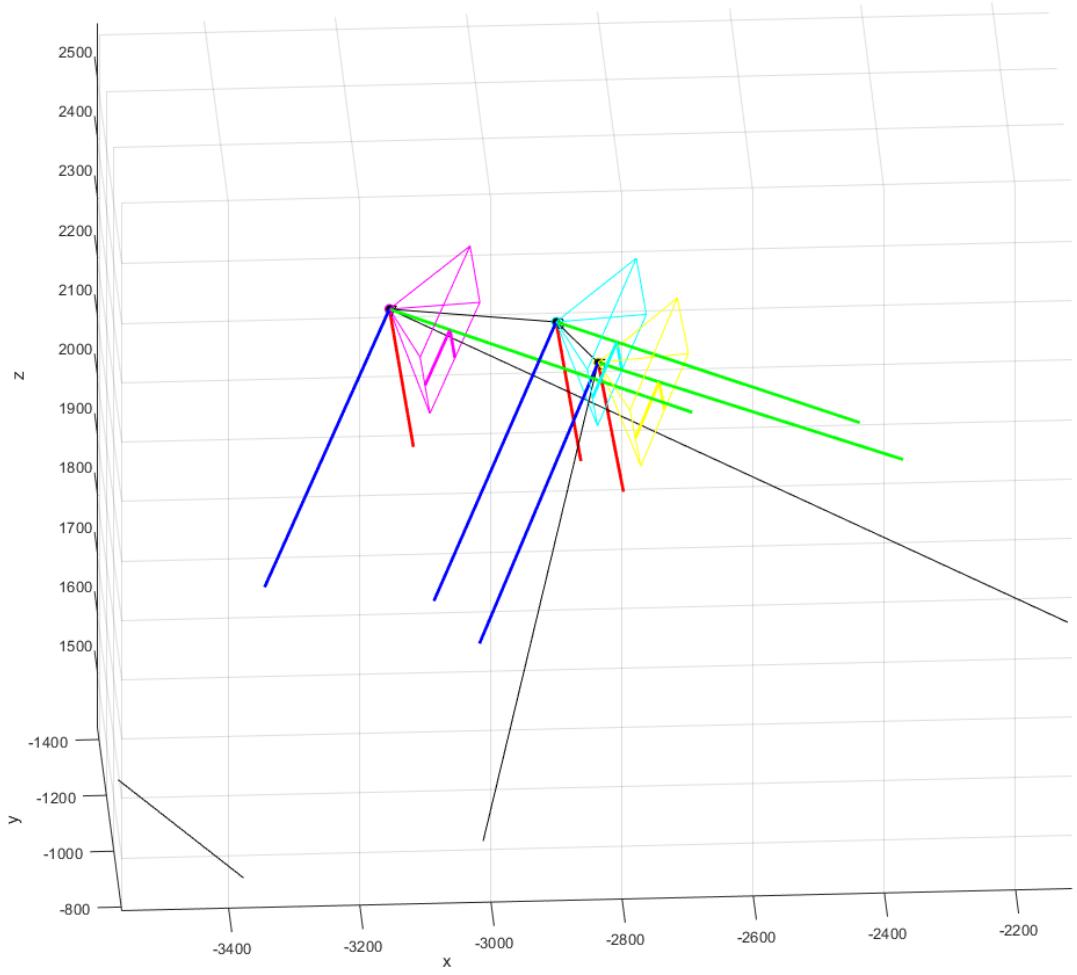


Figure 5.3 Fully independent Trifocal camera system 3D reconstruction

5.2.4 Complex experiment report

The experiment consist of independent calibrations of three different vehicles (*Lady, Summi, Wolle*). Every experiment is composed from two sessions distinguishable by the vehicle orientation, thus one is front-first and the second is rear-first. During the every session 30 equidistant frames is obtained and 16 additional mandatory frames. The Topview cameras have data from both of the sessions where the Trifocal cameras has data just from the front-first. A set of intrinsic parameters for the Topview has been provided in calibration of Lady and Summi vehicles. The Wolle calibration is calibrated without any intrinsic information. In all three cases the Trifocal intrinsic calibration heuristic has been deployed as been described in Section 3.5.1. The main

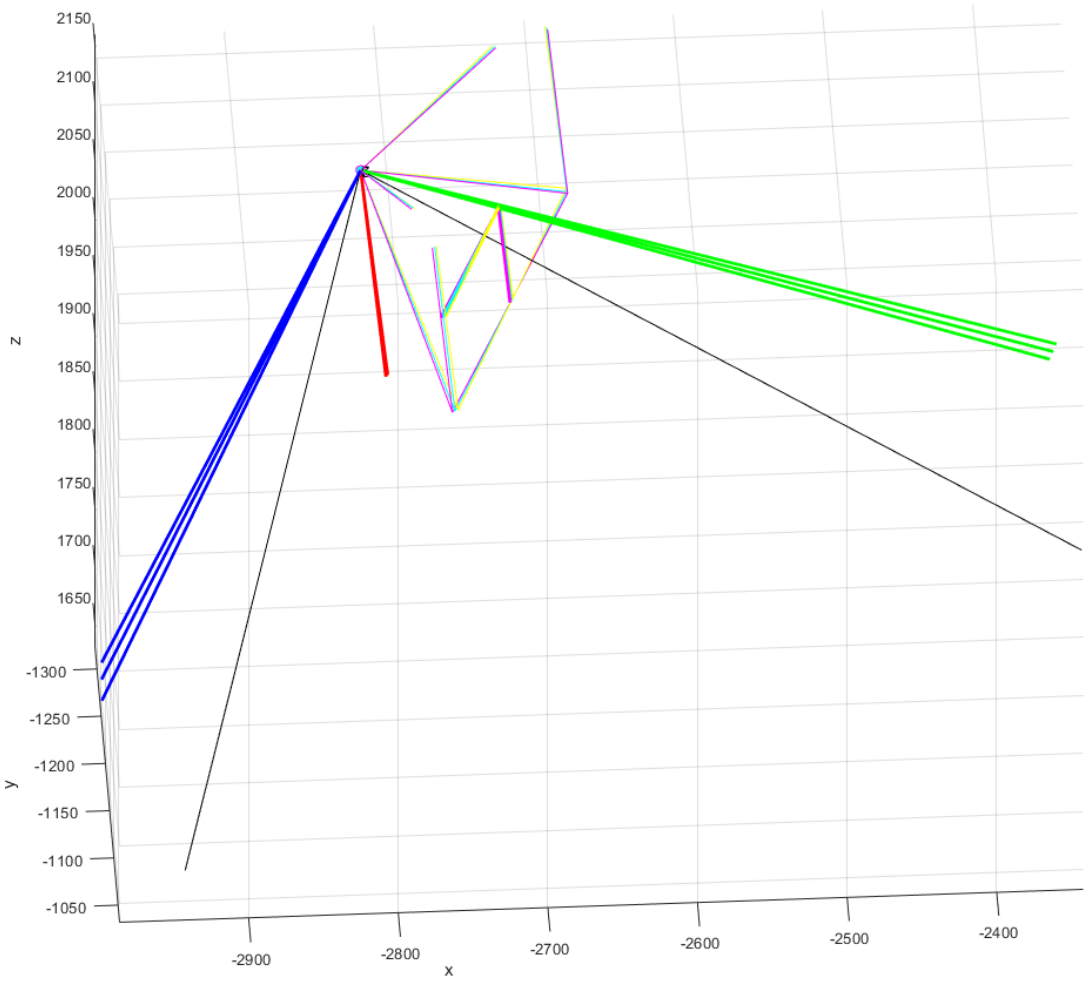


Figure 5.4 3D reconstruction of Trifocal camera system with identical camera center

rig is composed of five cameras (four Topview cameras and fifth is the Master Trifocal camera). The Secondary rig is composed of all three Trifocal cameras. In case Lady and Summi, the intrinsic calibration of the Main rig is computed only for the fifth camera, the Master Trifocal camera. In case of Wolle all five cameras are computed. The exterior calibration is computed either by the provided intrinsic calibration or the one computed in previous step. The camera resection is optimized using the BA afterwards. At this point the camera-wise calibration is complete, hence the reprojections of this stage are measured and compared with the results of rig calibration.

The initial Main rig is established and BA of the Main rig calibration is computed. With known calibration of the Connecting camera, the Secondary rig is estimated. The rigs are merged and again, optimized using the BA. At this point the Rig calibration is complete and the reprojections are measured. The detailed results and the comparison of the reprojections after camera-wise calibration and the rig calibration may be found in Chapter 6.

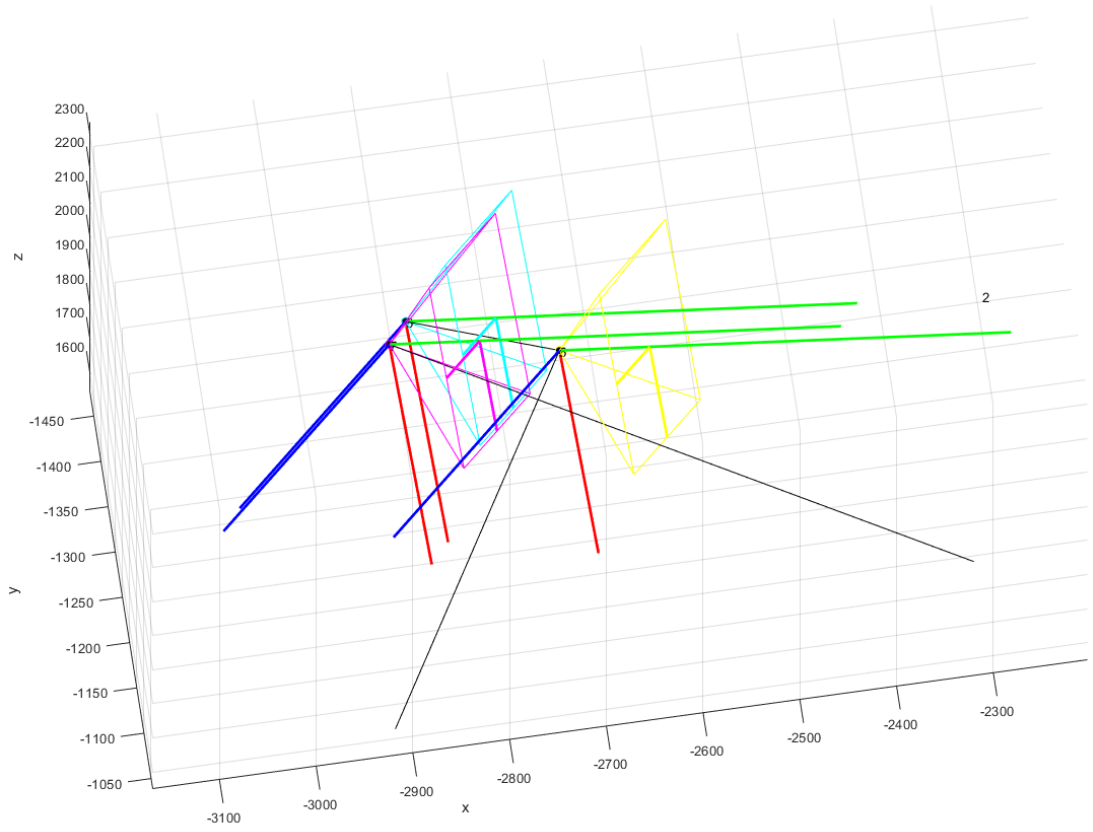


Figure 5.5 3D reconstruction of Trifocal camera system with cameras close to each other. Also the result of the final 3D scene reconstruction of the Summi vehicle.

6 Results

This chapter presents the results of the experiment described in previous Chapter 5. The calibration results of Summi are presented in full scale. The Sedric and Lady results may be found in comparison Figure 6.7 . The cameras may be noted as their ID in the program:

	Topview system				Trifocal system		
Camera ID	1	2	3	4	5	6	7
Camera name	front	left	rear	right	master	1^{st} slave	2^{nd} slave.

Table 6.1 Camera number look-up-table.

As been stated, the main goal of calibration is to find a model which minimizes the residuals of reprojections. The residuals should not excess single digit pixel values and the aim is to get into subpixel level of error. A detail of Summi residuals in a frame may be seen in Figure 6.1 which is a zoom to the lower-left corner of the frame. The

Figure 6.1 Zoom to details of reprojection errors in 2^{nd} slave Trifocal camera, 3^{rd} frame. Yellow dots are the observations u, Cyan dots are the calibration's reprojections. Magenta ray is the error vector scaled by 10x.

zero error is not expected due to the noise, however it is expected that the normalized residuals are evenly distributed around the normalization center. Even though that from the Figure 6.1 it might seem that the radial distortion parameter is missing or is too low, if all the residuals are taken into account it can be seen that it is not the case and there exists reprojections with residuals due to too large radial distortion which is the compensation of these orientation-correlated errors. In Figure 6.2 the residuals of the inliers are displayed. The threshold is very weak and aims to remove only the true outliers. As may be seen the mean values and geomedians of the residuals are close to the origin. Assuming the image noise and other causes of imprecision have a Gaussian

distribution, the overall structure of residual clusters shows that the model corresponds to it and thus is likely to be locally optimal.

The residuals themselves are obviously not enough to verify whether the cameras are well calibrated. If the correspondence structure does not cover the whole field of view of the cameras it is likely that the calibration is not good. Even though the overall reprojection error may be low it may happen that the model is overfitted to specific part of the image and produces immense residuals on the different parts which were not part of the calibration data set. Hence, it is important to study the reprojection in the original images. The reprojections of frames five to eight in both sessions and all its cameras may be seen in Figure 6.3 and Figure 6.4 respectively. From the study of reprojection images it is clear that there is an insufficient coverage in lower part of the images. The only exceptions are the Trifocal slave cameras, which suffer from having images on a single plane. The left and right cameras, for instance, have hardly any markers on the bottom part of the image, partially due to the car occlusion but mainly due to the missing markers on the floor.

6.1 The Absolute pose and the Rig calibration comparison

The results of the rig calibration has been described, however the results of the absolute pose were not discussed. Hence, the comparison of the absolute pose and the rig calibration is provided in this section. It is important to note that the comparison of the absolute poses and camera rig is not strictly fair in terms of the reprojection error to the Rig calibration, The absolute pose is not bounded by any relative restrictions other than shared intrinsic calibration across the frames for any camera. The described freedom provides a space for the overfitting which reduces the reprojection error, on the other hand it is likely to violate the rigid structure. The Topview and Trifocal systems were calibrated under different conditions and thus are separated in following figures.

Figure 6.5 shows the comparison of reprojection errors in all frames that were used to the calibration. The bars presented in figures represents the mean or maximum values of the reprojections across all the observations in specific frame and camera. As may be seen if the camera has no data in a frame, there are no bars, or if the data is too few or in ill conditions the frame fails to calibrate. The figures shows that the absolute pose has a lower level of errors, but fails to calibrate more often. The Rig calibration results show that the average error is increased but still stayed in the range of a sub-pixel precision. Hence, the sub-pixel precision ratio is higher. Another observation is that the overall amount of uncalibrated frames has dropped, due to the support of the other cameras.

Figure 6.6 shows the results of the Trifocal camera which are different than the Top view results. The core difference is the quality in either case, which has significantly dropped. The difference is caused by the missing intrinsic calibration which was provided in case of Topview. the results shows that it is possible do the intrinsic calibration this way, which is simpler and less time consuming to do but the quality drops accordingly to the quality of the data. An interesting observation is that in case of the worse quality of the intrinsic calibration, the rig structure also improves the camera-wise precision by a significant amount.

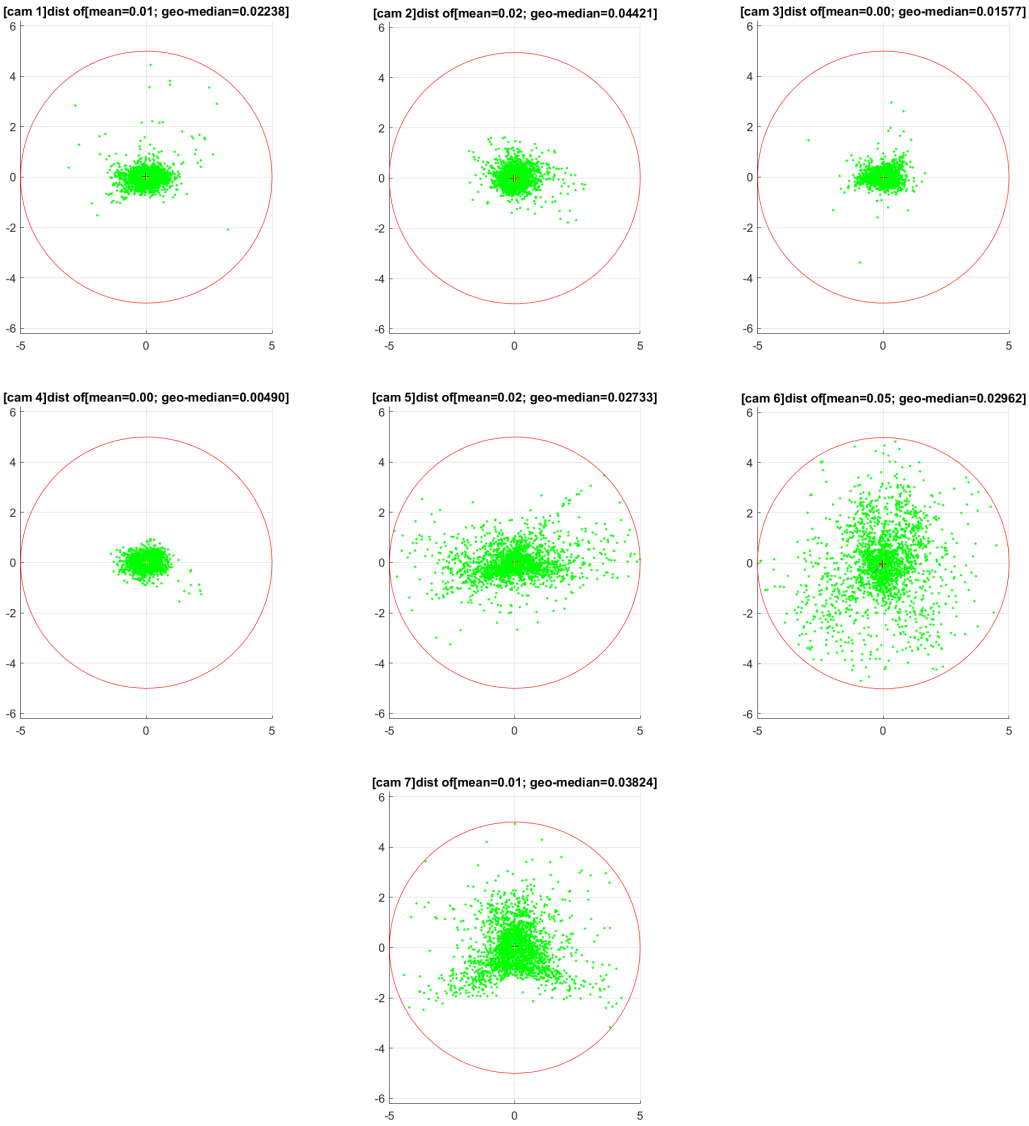


Figure 6.2 Residuals of inlier reprojections in cameras (left to right, top to bottom) in order 1 to 7 (see Table 6.1). The residuals are marked as green dots. Red circle is the threshold of inliers. The red cross denotes the mean residual value across the inliers. Black cross denotes the geometric median. The magenta ray is in direction of the mean residual scaled 10x.

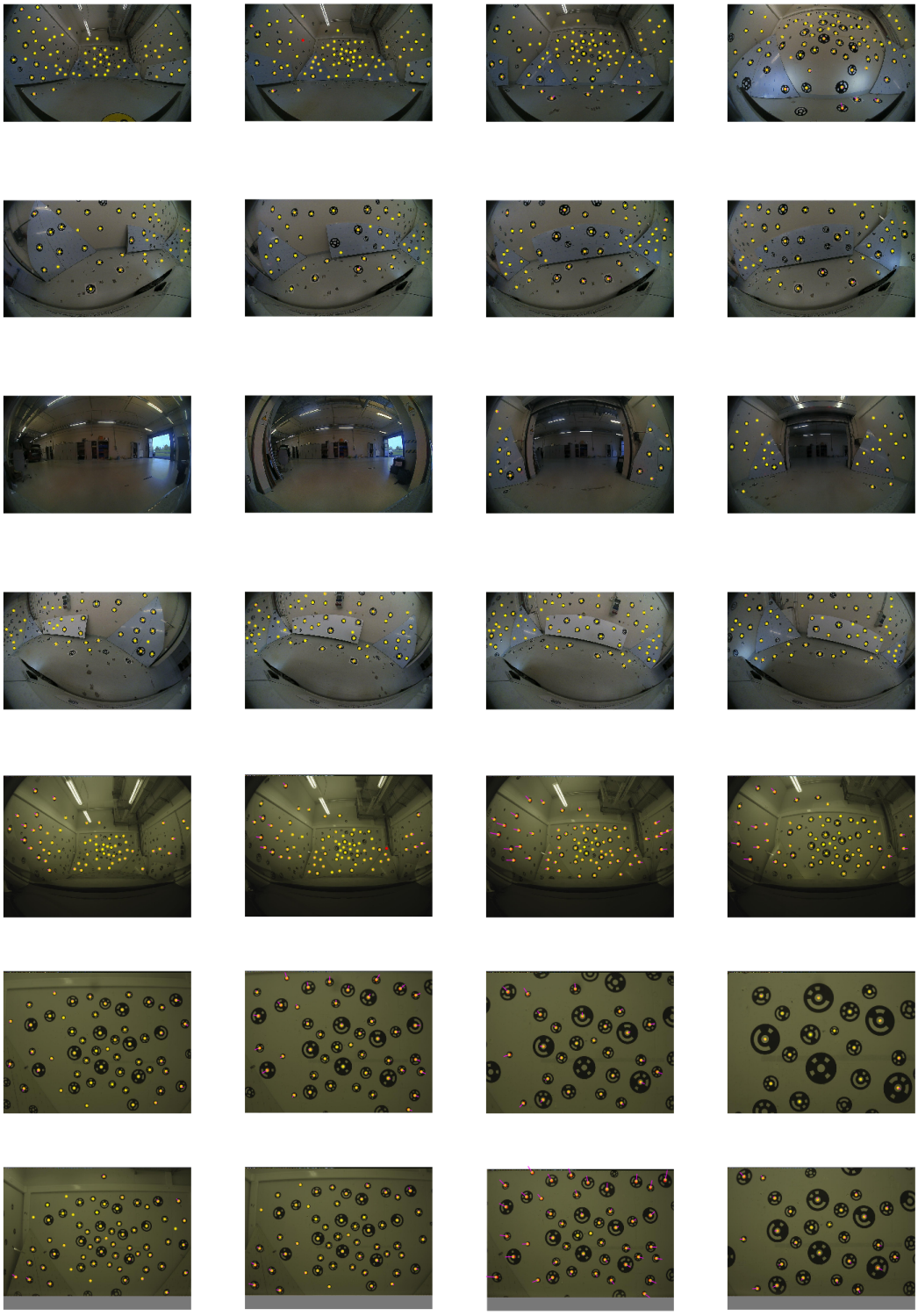


Figure 6.3 Session 1: all 7 cameras (see Table 6.1) (rows) in 4 different time-frames (columns). The Yellow dots marks the reprojection. Red dots denote the observations u. The observations may not be fully seen if occluded by reprojection marker. The magenta rays are the residuals scaled 10x.

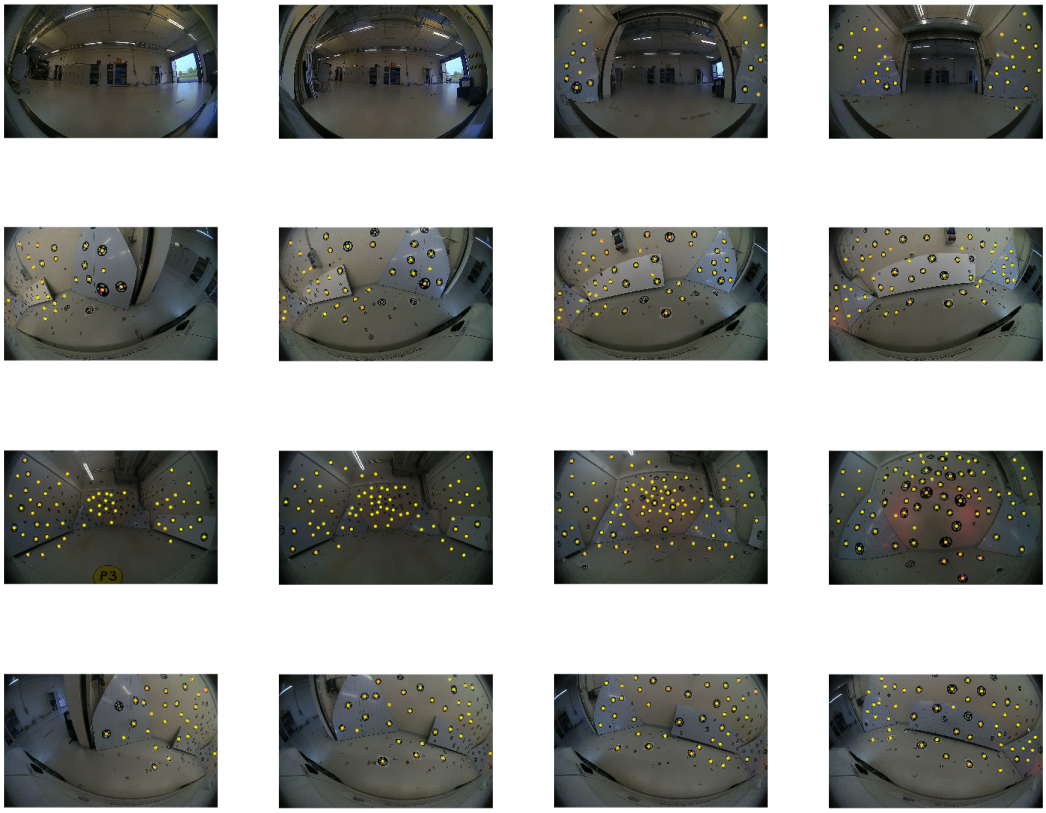


Figure 6.4 Session 2: first four cameras (see Table 6.1) (rows) in four different time-frames (columns). The Yellow dots marks the reprojection. Red dots denotes the observations u. The observations may not be fully seen if occluded by reprojection marker. The magenta rays are the residuals scaled 10x.

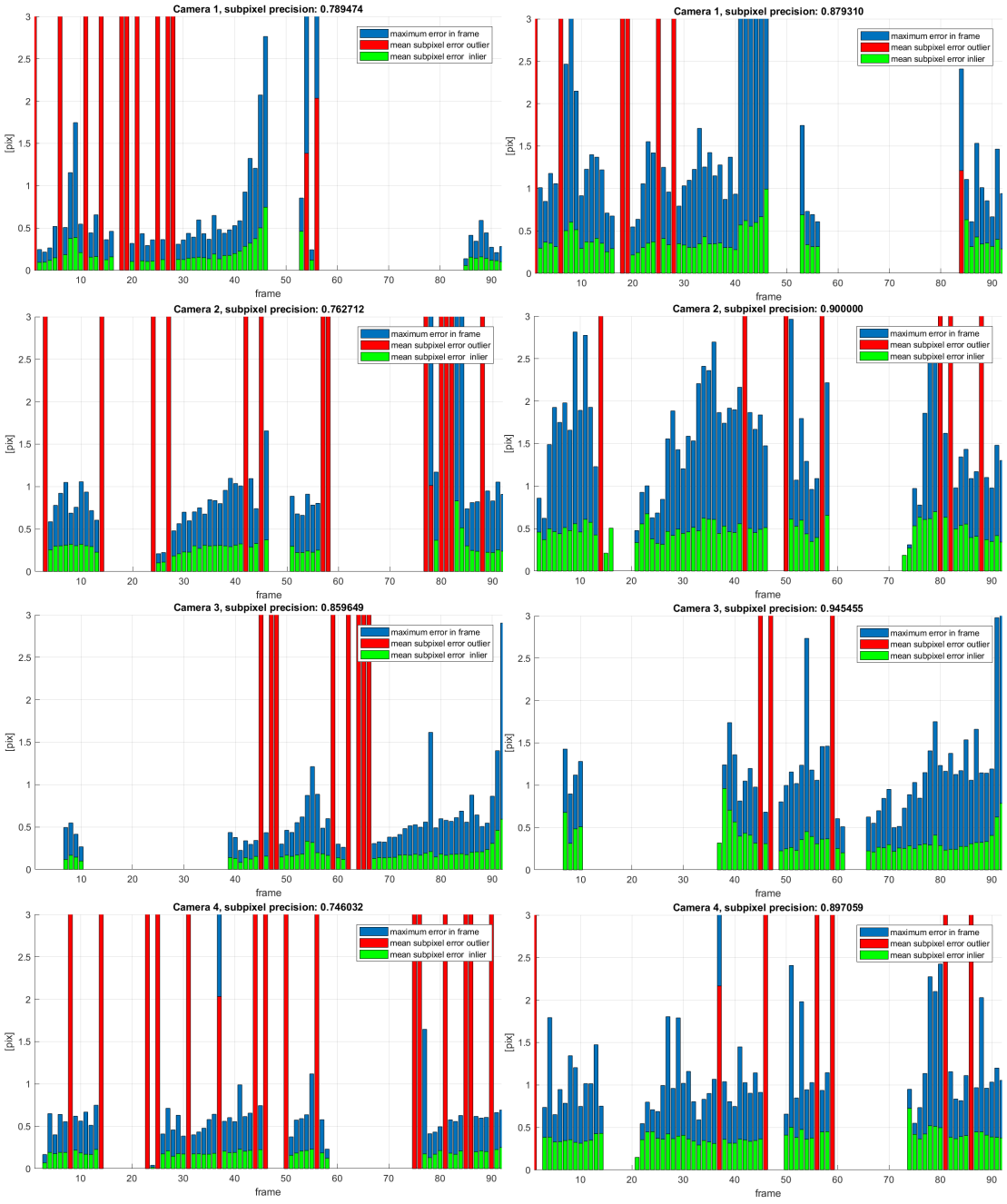


Figure 6.5 The comparison of absolute pose and rig calibration reprojection errors on the Topview camera system. The values are taken across all the cameras in the given frame. The mean error value changes color from red to green if is in the subpixel precision area. The blue shows the maximum error.

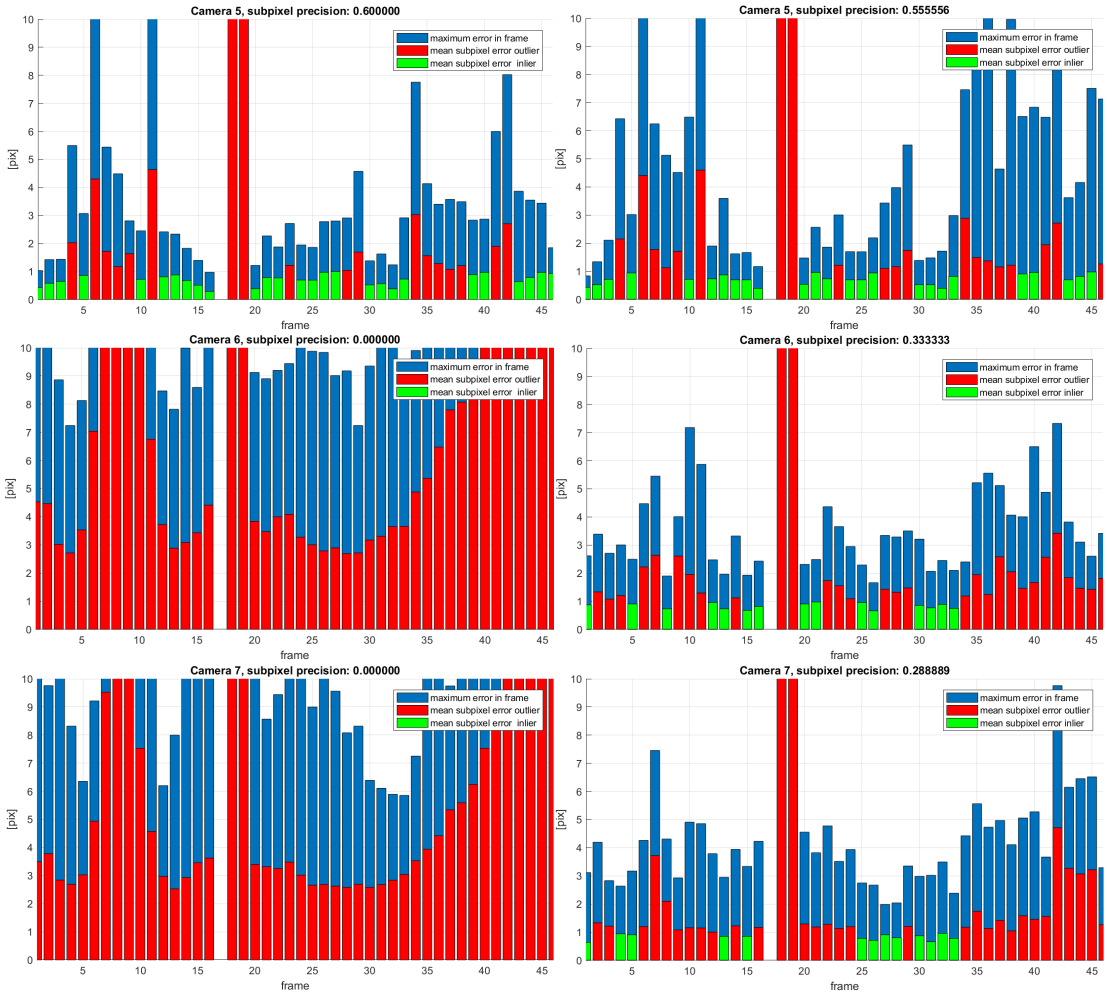


Figure 6.6 The comparison of absolute pose and rig calibration reprojection errors on the Trifocal camera system. The values are taken across all the cameras in the given frame. The mean error value changes color from red to green if is in the subpixel precision area. The blue shows the maximum error.

6.2 Comparison of the data sets

After the Wolle data set, which is the oldest one, was acquired the room with markers was updated. As may be seen at the Figure 6.7, the reprojection error dropped by a significant margin. If the outliers are removed, the precision is bellow 5 pixels even for the Trifocal cameras. Figure 6.8 and Figure 6.9 shows the mean value m across the cameras of a car per frame. The value m is the median reprojection error across all the image's reporojections. Hence we may say that it is the figure of the mean medians of reprojections per the vehicle at frame. Even though the frames in vehicles do not correspond to each other, they were taken in similar environment we may see that the quality of calibration is consistent.

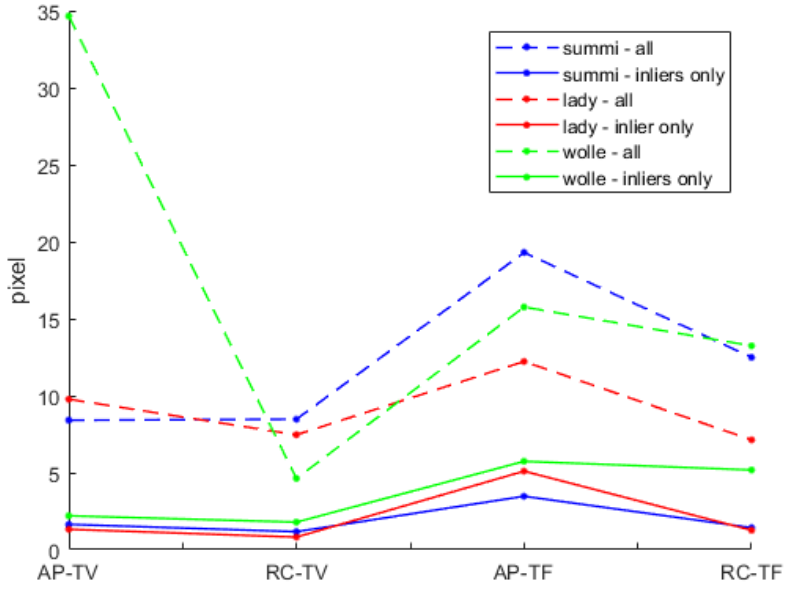


Figure 6.7 Mean reprojection error across all frames and cameras for given vehicle. The values are compared at four consecutive stages of the algorithm. (AP-TV) stands for the Absolute pose calibration of the TopView system. (RC-TV) is the TopView system after the Rig calibration.(AP-TF) is the absolute pose estimation of the TriFocal system based on the Rig calibration of the Topview.(RC-TF) is the rig calibration of the TriFocal system. Dashed are the mean values of all the reprojections, while the solid lines are the inliers only.

6.3 3D scene reconstruction

As the final verification step the 3D scene reconstruction is presented. due to the large amount of frames it is not possible to show all of them. Therefore, only eight frames are presented in Figure 6.10. The reconstructed poses seems reliable. See Figure 6.11 to inspect whether a single rig structure is reliable as well. The Trifocal camera system reconstruction may be also found in Figure 5.5.

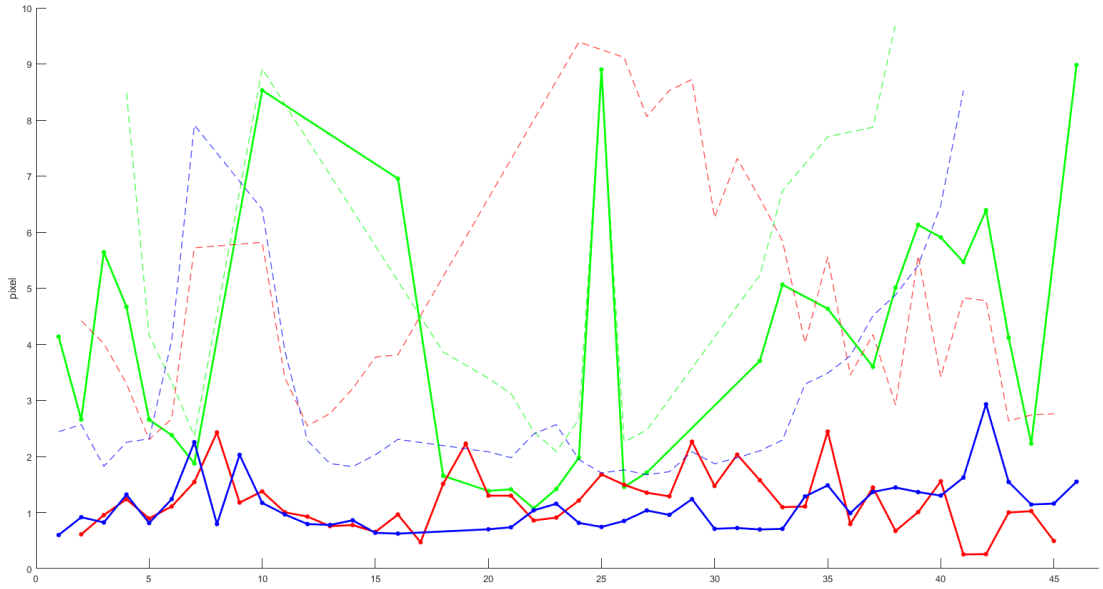


Figure 6.8 Dashed lines are the values of absolute pose. The solid lines are the rig calibration results. (green – Wolle, red – Lady, blue – Summi). Results may be shown in 46 frames of first session, due to Trifocal data not being available in session 2.

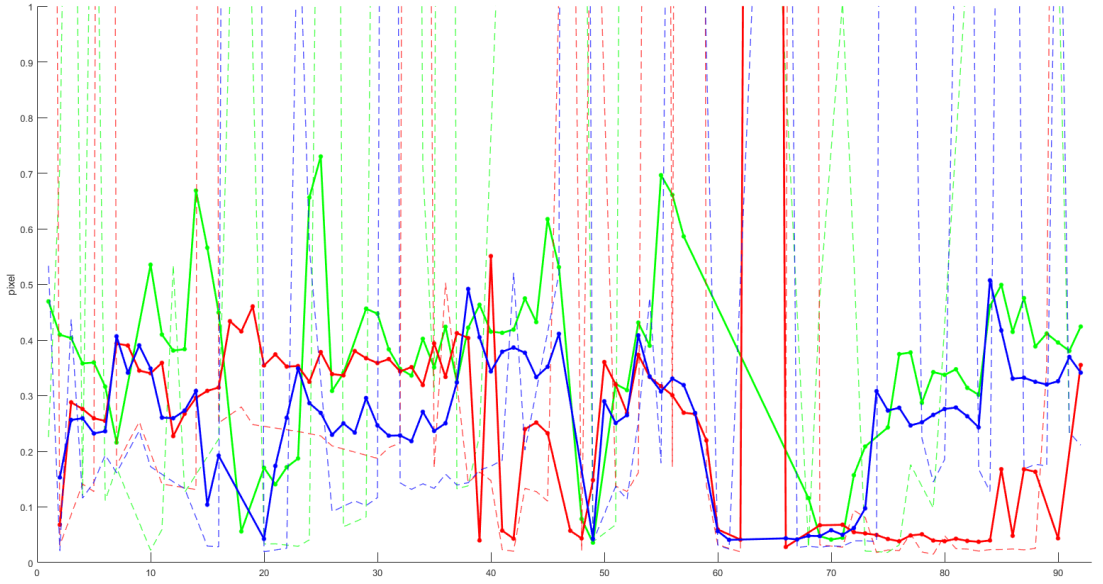


Figure 6.9 Dashed lines are the values of absolute pose. The solid lines are the rig calibration results. (green – Wolle, red – Lady, blue – Summi). The figure in full resolution may be found in the digital appendix.

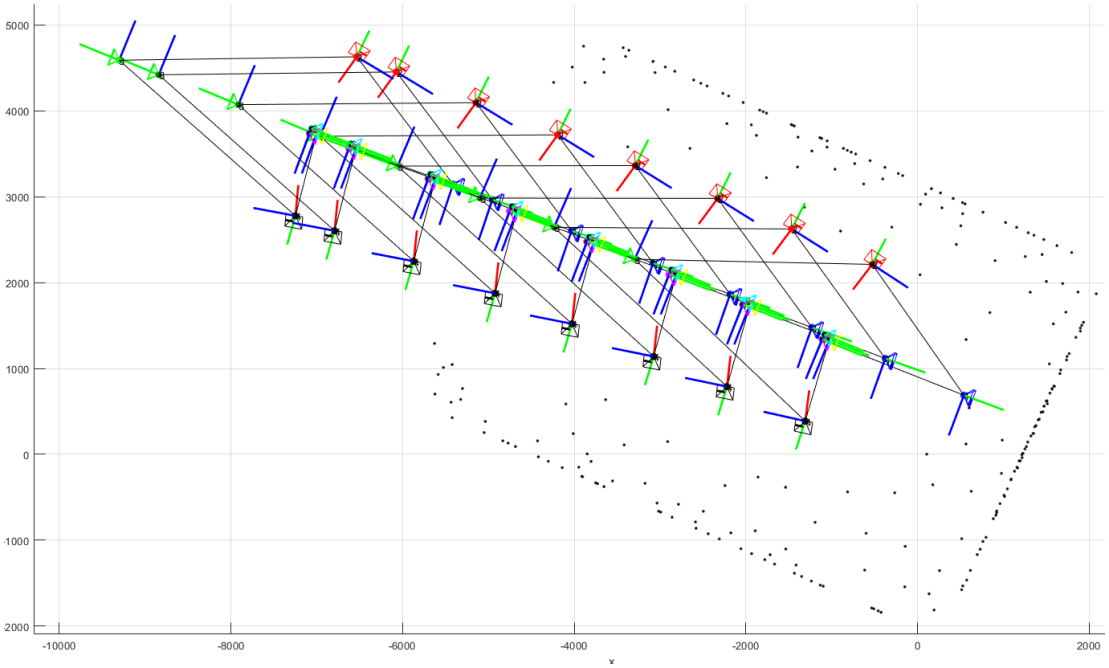


Figure 6.10 The 3D reconstruction of the scene. (black dots) Markers on the walls and floor of the calibration room.

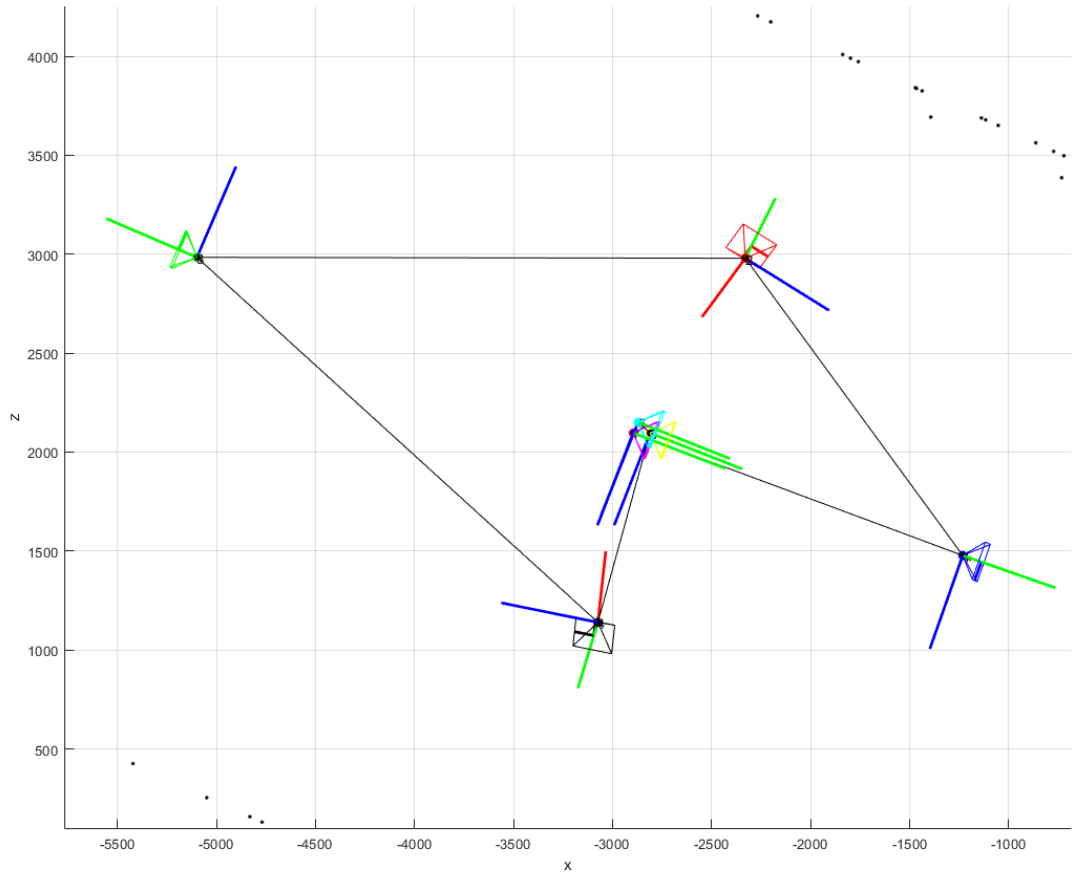


Figure 6.11 A single rig reconstruction of the Summi vehicle. The green ray denotes the optical axis of its camera, see fig. 5.5 to inspect the Trifocal camera reconstruction in detail.

7 Conclusion

In this work the camera rig calibration has been reviewed, implemented and tested on the real calibration problem and non-trivial formulation of the rig structure had been proposed. Due to the non-synthetic data usage, the whole calibration pipeline had to be implemented in order to perform the experiments. The experiments showed that the pipeline is capable of calibration based on given $2D \leftrightarrow 3D$ correspondences. The experiments have also shown that the implemented intrinsic calibration performs not as good as the state of the art calibration methods. The difference in the performance might be caused by the disparity of the calibration data. The calibration data used in this work were not captured with purpose of intrinsic calibration only. The state of the art intrinsic calibration methods rely on the calibration board movement which is incompatible with the provided experimental data, and thus the actual influence could not be examined. Nevertheless, the core focus of this work was the camera rig calibration, mostly from the exterior calibration point of view.

The rig of rigs structure proposed in the thesis was evaluated in experiments performed on the real vehicles. Even though the standalone camera calibration, in case of successful calibration yields lower reprojection errors, the quality is not as good as the quality of rig calibration due to two reasons. Based on the experiments, the rig calibration fails to calibrate in less frames than the standalone calibration, thus is more stable. Even though the ground truth positions of the cameras are not known, the experiments showed that the absolute pose estimation itself cannot guarantee the constant relative poses over the frames due to the over-fitting problem. If a rig was formed as the minimum spanning tree from the absolute poses as a naive form of a rig construction, the reporojection errors were in hundreds of pixels.

A heuristic model of a specialized automotive equipment, the Trifocal camera system, has been proposed. The experiments verified that the proposed Trifocal camera system model is capable of calibration on subpixel or up to 5 pixel precise calibration based on whether the intrinsic calibration is given or is calibrated by the pipeline. The experiments showed that using the rig structure, it is possible to calibrate cameras which would be uncalibratable as stand alone cameras from the provided data.

The usage of real data as the experimental data, caused multiple problems, i.e. the imperfect camera synchronization where a single frame was $40ms$ long which corresponds up to $5cm$ of vehicle translation, which had to be int he end, relaxed. The silver lining is definitively the insight it provided into the problems of real applications.

Based on the observations of the experiments, it can be said that the rig of rigs structure proved as a valid calibration rig concept. The main contribution of rig of rigs is in two core ideas. First is based on the idea that the rig calibration should be divided into multiple steps, and the exterior pose of the rig should be estimated from the cameras with sufficient data only. The second is that for a subset of the cameras a specific position constrains may be enforced, such as all the cameras lies on a single line or are equidistant from the rig center. I.e. the Trifocal system which enforces its cameras to be close to each other.

The purpose of this work is to provide a simplification to the calibration of complex camera systems. It is expected that using a set of complex and heavily specialized calibrations shall outperform the proposed solution but if compared by the expenses

the proposed solution shall outperform such setup while keeping a reasonable standard of quality.

7.1 Future improvements and open questions

At the current state the rig of rigs, does not compensate the real camera positions given the frame timestamps and pretends that the cameras are perfectly synchronized. which is obviously in general case not true and an interpolations of positions would enhance the results. As only the simplest constrain of Trifocal system had been evaluated, it is yet an open question whether the other proposed simple constrains would yield an improvement.

The proposed algorithm expects the knowledge of true 3D points, which is rather constraining factor and limits the usability to predefined scenes. It would be interesting to see whether the input can be acquired just by finding the correspondences among the cameras and then establishing tentative $2D \leftrightarrow 3D$ correspondences. This is an open question which is not possible to answer in general case (depends on various factors i.e. the size of the shared field of view, vehicle speed, image quality) but is worth exploring in task specific situations.

Bibliography

[1] Amnon Shashua. Mobileye - the future of computer vision and automated driving. http://www.mobileye.com/. 2, 17

[2] Richard Hartley and Andrew Zisserman. *Multiple view geometry in computer vision*. Cambridge university press, 2003. 4, 5, 7, 8, 10, 11, 20

[3] Pierre Moulon, Pascal Monasse, Renaud Marlet, and Others. Openmvg. https://github.com/openMVG/openMVG. 5

[4] Duane C Brown. Close-range camera calibration. *Photogrammetric Eng.*, 37(8):855–866, 1971. 5

[5] Jonathan Courbon, Youcef Mezouar, Laurent Eckt, and Philippe Martinet. A generic fisheye camera model for robotic applications. In *Intelligent Robots and Systems, 2007. IROS 2007. IEEE/RSJ International Conference on*, pages 1683–1688. IEEE, 2007. 5

[6] Werner Tecklenburg, Thomas Luhmann, and Heidi Hastedt. Camera modelling with image-variant parameters and finite elements. *Optical*, pages 328–335, 2001. 5

[7] Juho Kannala and Sami S Brandt. A generic camera model and calibration method for conventional, wide-angle, and fish-eye lenses. *IEEE transactions on pattern analysis and machine intelligence*, 28(8):1335–1340, 2006. 5

[8] Dave Litwiller. Ccd vs. cmos. *Photonics Spectra*, 35(1):154–158, 2001. 5

[9] Paul Furgale, Jérôme Maye, Jörn Rehder, and Thomas Schneider. The kalibr calibration toolbox. https://github.com/ethz-asl/kalibr. 5, 6

[10] Zuzana Kukelova, Martin Bujnak, and Tomas Pajdla. Real-time solution to the absolute pose problem with unknown radial distortion and focal length. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2816–2823, 2013. 7

[11] Martin A Fischler and Robert C Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981. 7, 10

[12] B. Li, L. Heng, K. Koser, and M. Pollefeys. A multiple-camera system calibration toolbox using a feature descriptor-based calibration pattern. In *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1301–1307, Nov 2013. 7

[13] L. Heng, B. Li, and M. Pollefeys. Camodocal: Automatic intrinsic and extrinsic calibration of a rig with multiple generic cameras and odometry. In *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1793–1800, Nov 2013. 7

[14] Richard I Hartley. Self-calibration from multiple views with a rotating camera. In *European Conference on Computer Vision*, pages 471–478. Springer, 1994. 7

[15] 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. 7, 8

[16] Bill Triggs, Philip F McLauchlan, Richard I Hartley, and Andrew W Fitzgibbon. Bundle adjustment—a modern synthesis. In *International workshop on vision algorithms*, pages 298–372. Springer, 1999. 8

[17] C Albl and T Pajdla. Constrained bundle adjustment for panoramic cameras. In *Computer Vision Winter Workshop*, pages 1–7, 2013. 8, 24

[18] Sameer Agarwal, Keir Mierle, and Others. Ceres solver. `http://ceres-solver.org`. 8

[19] Sandro Esquivel, Felix Woelk, and Reinhard Koch. Calibration of a multi-camera rig from non-overlapping views. In *Joint Pattern Recognition Symposium*, pages 82–91. Springer, 2007. 8, 9

[20] Andrew W Fitzgibbon. Simultaneous linear estimation of multiple view geometry and lens distortion. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, volume 1, pages I–I. IEEE, 2001. 11

[21] Henrique S Malvar, Li-wei He, and Ross Cutler. High-quality linear interpolation for demosaicing of bayer-patterned color images. In *Acoustics, Speech, and Signal Processing, 2004. Proceedings.(ICASSP'04). IEEE International Conference on*, volume 3, pages iii–485. IEEE, 2004. 26