Bachelor Project

Czech
Technical
University
in Prague

F3

Faculty of Electrical Engineering
Department of Cybernetics

Calibration of multiple cameras for autonomous driving

Martin Jaroš

Supervisor: Ing. Martin Matoušek, Ph.D.
Field of study: Cybernetics and robotics
May 2021

BACHELOR‘S THESIS ASSIGNMENT

I. Personal and study details

Student's name: **Jaroš Martin** Personal ID number: **474753**
Faculty / Institute: **Faculty of Electrical Engineering**
Department / Institute: **Department of Cybernetics**
Study program: **Cybernetics and Robotics**

II. Bachelor’s thesis details

Bachelor’s thesis title in English:

Calibration of Multiple Cameras for Autonomous Driving

Bachelor’s thesis title in Czech:

Kalibrace více kamer pro autonomní řízení

Guidelines:

Study the state of the art of modelling and calibration of geometric projection model of central perspective cameras (including non-perspective behavior, e.g., radial distortion) suitable for middle- and wide-FOV cameras used in car perception systems (intrinsic calibration).
Study the state of the art of multiple view geometry estimation (extrinsic calibration).
Propose and implement a method for intrinsic and extrinsic geometric calibration of multi-camera moving platform.
The calibration should be done in a controlled indoor environment with the help of calibration targets (coded markers and boards) and known scene constraints (planarity).
Verify the system on a multi-camera platform provided by our laboratory.

Bibliography / sources:

[1] Hartley, Richard and Zisserman, Andrew – Multiple View Geometry in computer vision – Cambridge university press, 2003.
[2] Zhang, Zhengyou – A Flexible New Technique for Camera Calibration – IEEE trans PAMI, 22(11): 1330 – 1334, Dec 2000.
[3] Mei, Christopher and Rives, Patrick – Single View Point Omnidirectional Camera Calibration from Planar Grids – Proc, IEEE International Conference on Robotics and Automation, Roma, Italy, Apr 2007.

Name and workplace of bachelor’s thesis supervisor:

Ing. Martin Matoušek, Ph.D., Robotic Perception, CIIRC

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

Date of bachelor’s thesis assignment: **18.01.2021** Deadline for bachelor thesis submission: ___________

Assignment valid until: **30.09.2022**

Ing. Martin Matoušek, Ph.D.
Supervisor’s signature

prof. Ing. Tomáš Svoboda, Ph.D.
Head of department’s signature

prof. Mgr. Petr Páta, Ph.D.
Dean’s signature

III. Assignment receipt

The student acknowledges that the bachelor’s thesis is an individual work. The student must produce his thesis without the assistance of others, with the exception of provided consultations. Within the bachelor’s thesis, the author must state the names of consultants and include a list of references.

.
Date of assignment receipt

Student’s signature

© ČVUT v Praze, Design: ČVUT v Praze, VIC

Acknowledgements

I would like to thank Ing. Martin Matoušek, PhD. for supervising this work. Especially for providing me with data from laboratory experiments during pandemic restrictions.

Declaration

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

Prague, 21. May 2021

Abstract

Main topic of this work is modeling and calibration of cameras for autonomous driving. It summarizes theory needed for the task of both internal and external calibration. This theory is used for modeling and calibration of middle-FOV and wide-FOV cameras. It includes tangential and equidistant projection models and polynomial and division model of radial distortion. For calibration of wide-FOV cameras are presented methods for initial estimate of camera parameters based on minimization of reprojection error caused by large radial distortion. For both types of camera is proposed method for estimation of relative pose of two cameras based on observation of planar objects by both cameras from different positions. All methods are verified on multi-camera platform in laboratory environment.

Keywords: Camera calibration, camera modeling, Calibration of multiple cameras, autonomous driving

Supervisor: Ing. Martin Matoušek, Ph.D.

Abstrakt

Hlavním tématem této práce je modelování a kalibrace kamer pro autonomní řízení. Práce shrnuje teorii potřebnou pro zjištění interní a externí kalibrace kamer. Tato teorie je využita pro modelování a kalibraci kamer s normálním objektivem i širokoúhlým objektivem typu rybí oko. Modelování zahrnuje ekvidistantní a tangenciální model projekce a polynomiální a podílový model zkreslení. Pro kalibraci širokoúhlých kamer jsou představeny metody pro počáteční odhad parametrů založené na minimilizaci reprojektivní chyby způsobené velkým radiálním zkreslením.Pro oba typy kamer je představena metoda pro odhad relativní pozice dvou kamer, která je založena na pozorování více rovin oběma kamerami z různých pozic. Metody kalibrace jsou verifikovány na pohyblivé kamerové platformě v laboratorních podmínkách.

Klíčová slova: Kalibrace kamer, modelování kamer, kalibrace více kamer, autonomní řízení

Překlad názvu: Kalibrace více kamer pro autonomní řízení

Contents

1 Introduction **1**

1.1 Goals of this thesis 1

1.2 Sources for this thesis 2

1.3 Structure of this thesis 2

2 Theoretical background **3**

2.1 Introduction 3

2.2 Camera model 3

  2.2.1 External calibration matrix . . 4

  2.2.2 Central projections 4

  2.2.3 Radial distortion 6

  2.2.4 Internal calibration matrix . . . 7

2.3 Camera calibration 8

3 Internal camera calibration **9**

3.1 Introduction 9

3.2 Utility function 10

3.3 Middle–FOV cameras 10

  3.3.1 Initial estimation of K 11

  3.3.2 Pose estimation 11

  3.3.3 Distortion model parameters estimation 12

  3.3.4 Overall optimization 13

3.4 Wide–FOV cameras 13

  3.4.1 Initial estimation of K and distortion 13

  3.4.2 Pose estimation 15

  3.4.3 Overall optimization 15

3.5 Experimental results 15

  3.5.1 Middle–FOV camera calibration 15

  3.5.2 Wide–FOV initial parameter estimation 16

  3.5.3 Wide–FOV camera calibration 18

4 Multiple camera calibration **21**

4.1 Introduction 21

4.2 Calibration method 21

4.3 Optimization problem 23

4.4 Experiment with synthetic data 24

4.5 Experiment with real data 24

4.6 Results of experiment with real data 26

5 Technical backgoround 29

5.1 Introduction 29

5.2 Software 29

5.3 Laboratory equipment 30

5.3.1 Multi–camera platform 30

5.3.2 Markers for calibration of multiple cameras 31

6 Conclusion 33

6.1 Future work 33

Bibliography 35

A Attached files 37

B SVD sign ambiguity 39

Figures

2.1 camera block model 3

2.2 Tangential projection 5

2.3 Equidistant projection 6

3.1 Aruco calibration board 9

3.2 Visualisation of projection errors for each point. Error vectors are multiplied by 30 for better visibility 16

3.3 Visualisation of projection errors for each point for method with planes. Errors for vertical lines are yellow. Errors for horizontal lines are red. Error vectors are multiplied by 50 for better visibility. 17

3.4 Visualisation of projection errors for each point for method with homography. Error vectors are multiplied by 30 for better visibility. 18

3.5 Visualisation of projection errors for each point for one way method with homography. Error vectors are multiplied by 30 for better visibility 18

3.6 Visualisation of projection errors for each point. Error vectors are multiplied by 30 for better visibility 19

4.1 Diagram of transformations between cameras and positions . . . 22

4.2 Set of images where optimization converged. Upper right is image from camera a at position 1. Upper left is image from camera b at position 1. Bottom left is image from camera a at position 2. Bottom right is image from camera b at position 2. 26

5.1 Camera rig used for experiments 30

5.2 Camera rig used for experiments top view. Mounted equipment consists of 2 Middle-FOV cameras in the front of rig (left in image). 3 Wide-FOV cameras on sides and rear (right in image). 31

5.3 View of laboratory with markers taped on walls and floor. For experiment served as planar objects mainly floor and right wall. 31

B.1 Results of homography transformation with opposite sign . 39

Tables

3.1 Projection error and resulting focal length, image centre and polynomial model parameters 16

3.2 Projection error and resulting focal length, image centre and division model parameters 16

3.3 Resulting focal length and image centre 17

3.4 Resulting parameters of polynomial distortion 17

3.5 Projection errors 17

3.6 Projection error and resulting focal length, image centre and polynomial model parameters 19

3.7 Projection error and resulting focal length, image centre and division model parameters 19

4.1 Results of optimization which converged to desired solution 26

4.2 Results of optimization for data gathered on 1 drive across laboratory, n is number of measurement and following are value of cost function and difference between result and reference solution 27

Chapter 1

Introduction

Geometric calibration of multiple cameras mounted on a vehicle is a necessary prerequisite for vision-based perception system for autonomous driving. Camera perception systems are widely used in field of autonomously driven vehicles. Advantage of camera as perception system is that it collects large amount of data about surroundings of vehicle. With recent development of machine learning, especially of convolutional neural networks, these data can be segmented and classified effectively. Information about road, other vehicles, pedestrians, traffic signs and many other things can be extracted from the data. Image data are combined with depth information, which is obtained either by Lidar or by reconstruction from multiple camera views. Generally there are many types of cameras with different technical parameters and physical properties, that are used. In this place calibration of cameras comes in use, so that connection between image and real physical world can be made. This means that measurements like distances between objects and dimensions of objects can be done. Also information from multiple cameras and Lidar sensors can be fused. 3D reconstructions from cameras can become more imporatant for autonomous driving after study from Cornell university [1] suggested it as cheaper alternative for Lidar sensors. This work focuses both on internal and external calibration of multiple cameras.

1.1 Goals of this thesis

First goal of this thesis was to study theory of mathematical modeling and calibration, which can be applied on multi-camera moving platform in laboratory that consists of both middle-FOV and wide-FOV cameras. Second goal was to use this theory to propose method for intrinsic and extrinsic calibration of those cameras. Last goal was to verify these methods in laboratory environment using correspondences given by coded markers. Further motivation that exceeds scope of this work is to adapt these methods to use real-world objects instead of markers for calibration. This would mean that calibration could be done outside of laboratory without preparation.

1.2 Sources for this thesis

This work mainly builds on theory presented in Multiple view geometry by Hartley and Ziessermann [2]. Calibration of middle-FOV camera was performed according to Zhang's [3] procedure. For calibration of wide-FOV camera was studied paper by Ch. Mei and T. Rives [4] but in the end their method wasn't used. Regarding calibration of multiple cameras a paper by E. Malis and M. Vargas [5] about decomposition of homography induced by plane was studied and some ideas from this paper were tested.

1.3 Structure of this thesis

Theoretical concepts of camera modeling needed for calibration such as projection models and radial distortion are described at Chapter 2. Chapter 3 is about methods for calibration of camera internal parameters. Existing methods and some new approaches for calibration are introduced and tested on middle–FOV and wide–FOV cameras. Chapter 4 describes method for calibration of relative pose of two cameras, which is based on observing multiple planes from two views by both cameras. Chapter 5 describes technical background of this thesis. Final chapter is conclusion of this work.

Chapter 2

Theoretical background

2.1 Introduction

This chapter presents theory that is used for modeling of camera projection and for calibration of camera model parameters.

2.2 Camera model

In this work camera is modeled from perspective of geometrical optics. Camera model is mapping between 3D world and 2D image. It consists of rigid transformation of world points into camera coordinates, central projection, distortion model and affine transformation between camera internal 2D coordinate system and image coordinate system. Block scheme of camera model can be seen at Figure 2.1.

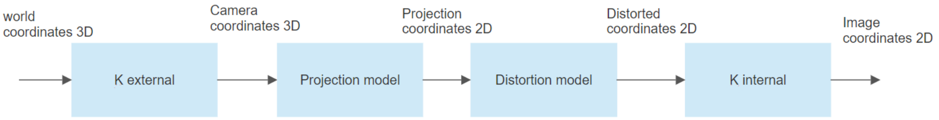
world coordinates 3D
Camera coordinates 3D
Projection coordinates 2D
Distorted coordinates 2D
Image coordinates 2D
K external
Projection model
Distortion model
K internal

Figure 2.1: camera block model

2.2.1 External calibration matrix

External camera calibration matrix [2] $K_{ext} \in \mathbb{R}^{3\times 4}$ represents rigid transformation between 3D world coordinates and 3D camera coordinates. It contains parameters, which describe rotation R and translation t between world and camera coordinates systems. This information can be interpreted as camera position and orientation in space. Matrix can be written as

$$K_{ext} = \begin{pmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{pmatrix} = \begin{bmatrix} R \,|\, t \end{bmatrix}, \tag{2.1}$$

where $t = -Rc$ and c is position of camera centre in world coordinates. The transformation is performed as matrix multiplication

$$\begin{pmatrix} x_c \\ y_c \\ z_c \end{pmatrix} = K_{ext} \begin{pmatrix} x_w \\ y_w \\ z_w \\ 1 \end{pmatrix}. \tag{2.2}$$

Coordinates with c subscript are in camera coordinate frame. Coordinates with w subscript are in world coordinate frame.

2.2.2 Central projections

We are considering central projection model only. A camera projection is mapping between 3D points and their 2D projection. We are using two types of projection – tangential and equidistant. In this section uppercase letters X, Y, Z are used for 3D points and lowercase letters u, v are used for 2D projections. Further we define

$$R = \sqrt{X^2 + Y^2} \tag{2.3}$$

$$r = \sqrt{u^2 + v^2} \tag{2.4}$$

$$\alpha = \arctan 2(R, Z), \tag{2.5}$$

where R is distance between optical axes and 3D point, r is radius of point in image plane and α is angle between ray from camera centre to 3D point and optical axes.

Tangential projection

The first projection we used is tangential projection [2], Figure 2.2. This projection can be used for modeling cameras with rectilinear or almost rectilinear lenses with low radial distortion. This projections maps 3D points to their projections on image plane perpendicular to optical axis. Projection is given by formula

$$r = tan(\alpha), \tag{2.6}$$

which leads to

$$\begin{pmatrix} u \\ v \end{pmatrix} = \frac{1}{Z}\begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} \frac{X}{Z} \\ \frac{Y}{Z} \end{pmatrix}. \tag{2.7}$$

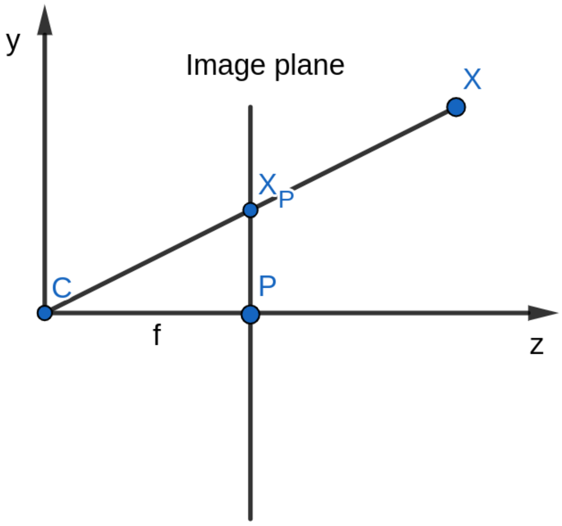

Figure 2.2: Tangential projection

Equidistant projection

The second projection used in this work is equidistant projection [6], Figure 2.3. This projection can be used for modeling cameras with lenses with significant radial distortion (e.g. fisheye lenses). This projection maps 3D points to surface of unit sphere with centre identical to camera centre. Advantage of this projection is that it can project points from all around the camera so it can model even spherical lens. Projection is given by formulas

$$r = \alpha \tag{2.8}$$

and

$$\begin{pmatrix} u \\ v \end{pmatrix} = \frac{\alpha}{R} \begin{pmatrix} X \\ Y \end{pmatrix}. \tag{2.9}$$

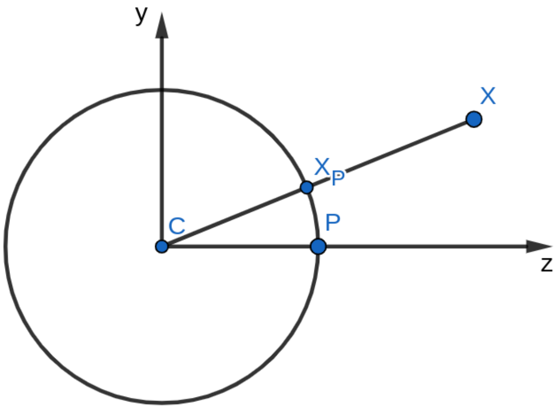

Figure 2.3: Equidistant projection

2.2.3 Radial distortion

Process of image creation cannot be fully described by projection, because lenses generally perform more complex projection. This behaviour causes distortions in images. In this work only radial distortion is considered. For each point on image plane the radial distortion can be described as a function of radius of this point $f(r)$. Radius of the distorted point is

$$r_d = rf(r). \tag{2.10}$$

Generally the function $f(r)$ has unknown parameters that need to be estimated.

Polynomial distortion model

The first method is polynomial model [2].The Function $f(r)$ is approximated by polynomial (2.11). The distortion is described by 3 parameters k_1, k_2, k_3.

$$P(r) = 1 + k_1 r^2 + k_2 r^4 + k_3 r^6 \tag{2.11}$$

For the purposes of camera calibration an inverse function for the polynomial distortion is needed. It is approximated by polynomial (2.12). Coefficients of this polynomial are obtained by following procedure by Drap and Lefèvre [7]

$$Q(r) = 1 + b_1 r^2 + b_2 r^4 + b_3 r^6 \tag{2.12}$$

$$b_1 = -k_1 \tag{2.13}$$

$$b_2 = 3k_1^2 - k_2 \tag{2.14}$$

$$b_3 = 8k_1 k_2 - 12k_1^3 - k_3. \tag{2.15}$$

Division distortion model

Second method for distortion estimation is division model [8]. Advantage of this model is its easy invertibility and only one parameter to be estimated. Forward transformation (distortion) is

$$r' = \frac{2}{1-\lambda} r \tag{2.16}$$

$$r_d = \frac{r'}{1 + \sqrt{1 + \lambda \frac{r'^2}{d_n^2}}}, \tag{2.17}$$

where λ is parameter of distortion, r_d is radius of distorted point, d_n is radius of circle in image that is not changed. The inverse transformation is

$$r = \frac{1-\lambda}{1 - \lambda \frac{r^2}{d_n^2}} r_d \tag{2.18}$$

$$\tag{2.19}$$

2.2.4 Internal calibration matrix

Internal calibration matrix [2] $K_{int} \in \mathbb{R}^{3\times 3}$ represents affine transformation between camera internal coordinates and image coordinates. It contains parameters which describe properties of the camera sensor. These are scales of axes α_x, α_y, skew s and position of the image centre x_0, y_0. The matrix can be written as

$$K_{int} = \begin{pmatrix} \alpha_x & s & x_0 \\ 0 & \alpha_y & y_0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{2.20}$$

Affine transformation is performed as matrix multiplication

$$\begin{pmatrix} u_{im} \\ v_{im} \\ 1 \end{pmatrix} = K_{int} \begin{pmatrix} u_d \\ v_d \\ 1 \end{pmatrix}. \tag{2.21}$$

2.3 Camera calibration

Calibration of camera is a process of determining unknown parameters of camera model. This process is based on taking pictures of object with known geometrical properties and significant points which can be detected. These detected points are reprojected by camera model. Parameters of model are determined by minimization of reprojection error. Formally it can be expressed as

$$p^* = \arg\min_p f(x, p). \tag{2.22}$$

$f(x, p)$ is objective function, x is set of coordinates of detected points and p is set of camera model parameters.

Chapter 3

Internal camera calibration

3.1 Introduction

This chapter describes methods of internal camera calibration. These methods provide modular approach to camera calibration. Depending on situation different models of cameras and distortion can be used. These methods were verified by several experiments. For all experiments calibration board with chessboard pattern was used. Example of the calibration board is at Figure 3.1. Results of these experiments are at the end of this chapter.

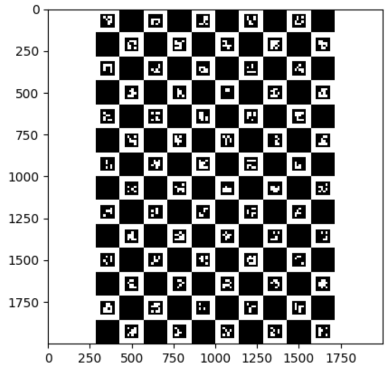
0
250
500
750
1000
1250
1500
1750
0
250
500
750
1000
1250
1500
1750

Figure 3.1: Aruco calibration board

3.2 Utility function

In Section 2.3 we introduced calibration as an optimization problem. The specific instance of utility function that is used in this chapter is

$$f_{proj,dist} = \sqrt{\frac{1}{\sum_{i=1}^{n} |X_i|} \sum_{i=1}^{n} \sum_{x \in X_i} \| x - projection_{proj,dist}(x_{grid}, R_i, t_i, K, P_{dist}) \|^2}. \tag{3.1}$$

Expression (3.1) represents RMS of projection error over all detections. Depending on situation it is $f_{tan,poly}$ for combination of tangential projection and polynomial distortion, $f_{tan,div}$ for combination of tangential model and division distortion, $f_{eq,poly}$ for combination of equidistant model and polynomial distortion, $f_{eq,div}$ for combination of equidistant model and division distortion. X_i is set of detections for image i, x_{grid} is grid point corresponding to detection x, R_i,t_i is pose of camera relative to grid for image i, K is internal calibration matrix, P_{dist} are parameters of the distortion model.

3.3 Middle–FOV cameras

In this section methods for middle–FOV cameras are introduced. Middle–FOV camera is usually rectilinear with low radial distortion so tangential projection was chosen to model it. For distortion both division and polynomial model were used. Calibration was done in the way of Zhang's flexible camera calibration [3]. Multiple images with different position of calibration board were used. First internal calibration matrix K is estimated directly from multiple views of calibration board. With known K pose of camera R, t relative to calibration board is obtained for each image. Then radial distortion is estimated after pose. With known initial values projection error is optimized by a numerical optimization method method.

3.3.1 Initial estimation of K

Zero skew parameter and fixed aspect ratio was considered for calibration so matrix ω was estimated as

$$\omega = \begin{pmatrix} \omega_1 & 0 & \omega_2 \\ 0 & \omega_1 & \omega_3 \\ \omega_2 & \omega_3 & \omega_4 \end{pmatrix}. \tag{3.2}$$

Linear constraints given by homographies between detected image points and grid point coordinates were used for estimation of ω. Homography for each image can be written as

$$H = \begin{bmatrix} h_1 & h_2 & h_3 \end{bmatrix}. \tag{3.3}$$

Each homography than adds 2 constraints [2] on ω

$$h_1^T \omega h_2 = 0 \tag{3.4}$$

$$h_1^T \omega h_1 = h_2^T \omega h_2. \tag{3.5}$$

These constrains form homogeneous system of linear equations $A\omega = 0$. This system is solved by total least squares method using SVD of matrix A. Finally the matrix K was computed from

$$\omega^{-1} = KK^T \tag{3.6}$$

by Cholesky decomposition.

3.3.2 Pose estimation

With known matrix K pose of camera in relation to calibration board can be obtained using Zhang's method [3]. Transformation between board and image coordinates can be written as

$$s \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = K \begin{bmatrix} r_1 & r_2 & r_3 & t \end{bmatrix} \begin{pmatrix} x \\ y \\ 0 \\ 1 \end{pmatrix} \tag{3.7}$$

$$H = K \begin{bmatrix} r_1 & r_2 & t \end{bmatrix}. \tag{3.8}$$

For each image, pose can be than extracted from homography as follows

$$r_1 = \lambda K^{-1} h_1 \tag{3.9}$$

$$r_2 = \lambda K^{-1} h_2 \tag{3.10}$$

$$r_3 = r_1 \times r_2 \tag{3.11}$$

$$t = \lambda K^{-1} h_3 \tag{3.12}$$

$$\lambda = \left(\frac{\|K^{-1}h_1\|+\|K^{-1}h_2\|}{2}\right)^{-1}. \tag{3.13}$$

Now we have pose in form of translation vector and rotation matrix

$$R = \begin{bmatrix} r_1 & r_2 & r_3 \end{bmatrix} \tag{3.14}$$

which as computed does not necessarily fulfill condition $R^T R = I$ so it is approximated by rotation matrix Q [9] that is the nearest matrix fulfilling mentioned condition in the sense of Frobenius norm.

3.3.3 Distortion model parameters estimation

Last step of internal calibration is to estimate parameters of a distortion model. This part is different for polynomial and division model.

polynomial model

Polynomial model parameters were obtained using least squares method [10] from equations

$$k_1 r^2 + k_2 r^4 + k_3 r^6 = r_d - r. \tag{3.15}$$

Undistorted points are computed as projection of gridpoints and distorted points are computed using inverse transfomation K from image detections. Each pair of points generates one equation for their radii.

division model

Initial estimate of division model parameters was done by minimizing projection error. Parameters obtained in previous sections were used. Initial value for division model were set to $\lambda = 0$. Estimated value was obtained by minimization of objective function (3.1) with respect to λ.

3.3.4 Overall optimization

Last procedure is overall optimization to refine estimated parameters. This procedure has two parts. One part is optimization of R and t for each image with fixed K and distortion parameters. Other part is optimization of parameters K and distortion parameters with fixed R and t. Objective function for this optimization is (3.1). These two parts can be iterated until convergence.

3.4 Wide–FOV cameras

In this section methods for wide–FOV cameras are introduced. Wide–FOV camera has curvilinear mapping so equidistant projection was chosen to model it. For distortion were used both polynomial and division model. Methods from previous section couldn't be used because of high radial distortion. For initial estimate of parameters we tested different methods. In all methods fixed aspect ratio and zero skew in internal calibration matrix are assumed.

3.4.1 Initial estimation of K and distortion

This section presents methods which use minimization of reprojection error to estimate internal parameters. First method is based on plane fitting. Second method is based on estimation of homography between grid coordinates and rays that corresponds to detected points in images.

Calibration with planes

In this approach the fact that corners of calibration chessboard lie in lines in real 3D world and that each of these lines can be viewed as intersection of chessboard plane with plane that goes through camera centre is used. It is considered that each of the corners is at least in one vertical and one horizontal line. Also detected positions of these corners in image are known and can be assigned to their grid position on chessboard, therefore sets of points that lie in one line are known. Inverse distortion and projection is applied on each of these sets and 3D rays are obtained. Those 3D rays are fitted with plane that goes through camera centre. Each ray gives one constraint on the plane and these constraints form homogeneous system of linear equations $An = 0$. Normal vector n of this plane is found by total least squares using SVD of matrix A. Now dot product is used to project 3D rays to this plane. Then projection model and distortion is applied to obtain image positions of these plane points. Now euclidean distances between original detections and these projected points are measured. RMS of these distances over all points and all lines is minimized with respect to scale α_x and image centre x_0, y_0. Image centre is initialized with $\frac{image\,width}{2}, \frac{image\,height}{2}$. Scale was initialized by guess based on image observation. Then with these parameters fixed the error is optimized with respect to distortion parameters. Distortion parameters are initialized to 0.

Calibration with homography

Second approach was to estimate homography between 3D rays and grid points. As in previous method 2D detections were transformed to 3D rays. Homography between those 3D points grid points was estimated. 3D rays and corresponding grid points make constraints on homography. These constraints form homogeneous system of equations $Ah = 0$. Homography is obtained by total least squares using SVD of matrix A. Grid points were transformed by homography and projected to image. Error is computed the same way as in previous section. The parameters α_x, x_0, y_0 and distortion parameters were found as in previous method.

One way calibration with homography

This approach is similar to previous one, however it does not transform image detections to 3D rays instead it directly projects grid points to image. Homography is in this case part of the optimization and it was initialized with identity matrix. This method was intended to avoid inversion of distortion model which is not precise. Optimization is needed to be bounded, otherwise it does not converge.

3.4.2 Pose estimation

Now with known estimate of matrix K, the camera pose can be estimated. First inversion of equidistant model is applied to detected points to transform them to rays. Homography between grid points in homogeneous coordinates and these rays is estimated. Pose is then extracted in similar way as for middle–FOV camera

$$r_1 = \lambda h_1 \tag{3.16}$$

$$r_2 = \lambda h_2 \tag{3.17}$$

$$r_3 = r_1 \times r_2 \tag{3.18}$$

$$t = \lambda K^{-1} h_3 \tag{3.19}$$

$$\lambda = \left(\frac{\|h_1\|+\|h_2\|}{2}\right)^{-1}. \tag{3.20}$$

Rotation matrix is approximated to fulfil condition $R^T R = I$.

3.4.3 Overall optimization

This part is same as for middle–FOV camera. Only the projection model is equidistant. Objective function is expression (3.1). Both polynomial and division model of distortion are optimized with equidistant projection.

3.5 Experimental results

In this section are presented results of experiments.

3.5.1 Middle–FOV camera calibration

Tables 3.4 and 3.5 presents results for calibration with polynomial and division distortion model and corresponding projection errors respectively. Figure

3.5 presents visualisation of projection errors for calibration with polynomial model.

$\mathbf{Err}_{rms}$	α_x	$\mathbf{x}_0$	$\mathbf{y}_0$	$\mathbf{k}_1$	$\mathbf{k}_2$	$\mathbf{k}_3$
1.07	2767.45	1156.25	1035.61	-0.0046	-2.6635	0.6738

Table 3.1: Projection error and resulting focal length, image centre and polynomial model parameters

$\mathbf{Err}_{rms}$	α_x	$\mathbf{x}_0$	$\mathbf{y}_0$	$\mathbf{d}_n$	λ
1.37	2764.83	1156.62	1035.61	100	-0.0052

Table 3.2: Projection error and resulting focal length, image centre and division model parameters

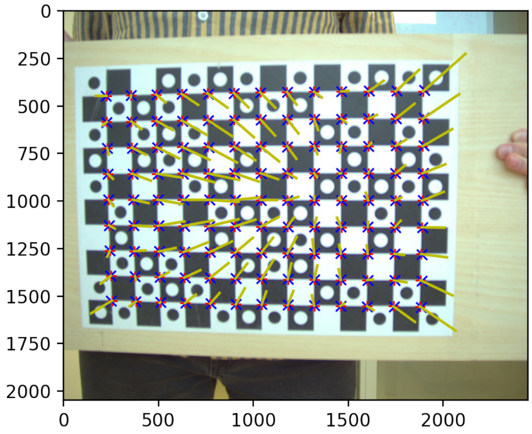

Figure 3.2: Visualisation of projection errors for each point. Error vectors are multiplied by 30 for better visibility

3.5.2 Wide–FOV initial parameter estimation

This section presents results of initial parameter estimation for wide–FOV cameras. Estimated were focal length, image centre and polynomial distortion model parameters. Tables 3.1 – 3.3 presents in order estimated scale of axes and image centre, parameters of distortion model and projection errors. Figures 3.2 – 3.4 presents visualization of projection errors for each method in the same order as in tables.

method	α_x	$\mathbf{x}_0$	$\mathbf{y}_0$
planes	839.46	1260.87	961.79
homography	827.70	1257.05	958.67
one way homography	841.22	1271.97	963.62

Table 3.3: Resulting focal length and image centre

method	$\mathbf{k}_1$	$\mathbf{k}_1$	$\mathbf{k}_1$
planes	0.0053	0.0060	-0.0051
homography	0.0083	0.0067	-0.0055
one way homography	-0.0102	0.0214	-0.0089

Table 3.4: Resulting parameters of polynomial distortion

method	Err_{rms}	Err_{rms}(with disotortion model)
planes	1.03	0.50
homography	4.36	1.36
one way homography	2.95	1.90

Table 3.5: Projection errors

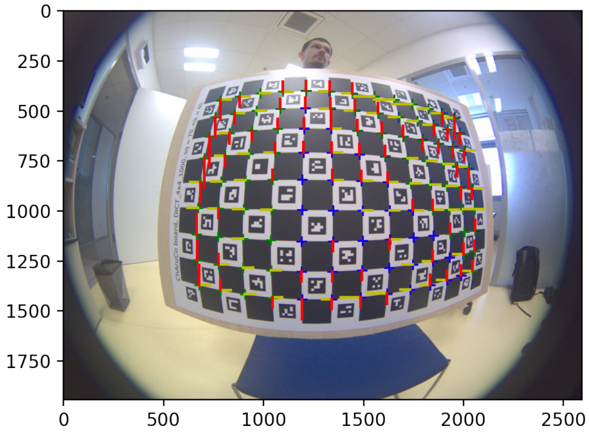

Figure 3.3: Visualisation of projection errors for each point for method with planes. Errors for vertical lines are yellow. Errors for horizontal lines are red. Error vectors are multiplied by 50 for better visibility.

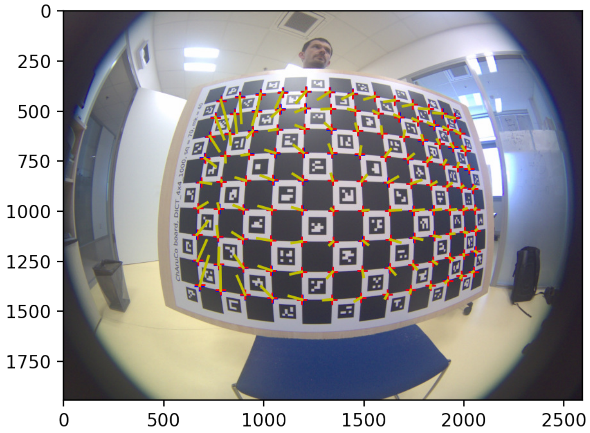

Figure 3.4: Visualisation of projection errors for each point for method with homography. Error vectors are multiplied by 30 for better visibility.

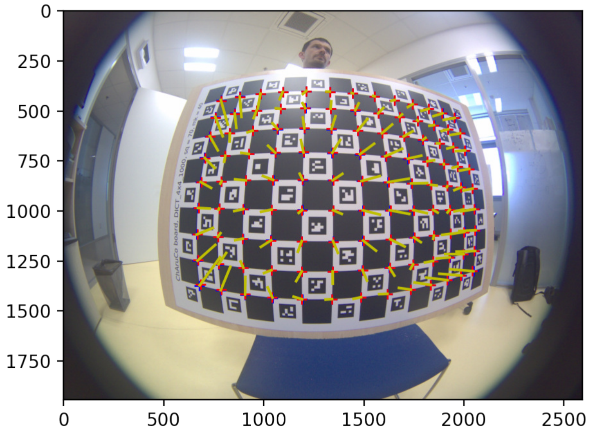

Figure 3.5: Visualisation of projection errors for each point for one way method with homography. Error vectors are multiplied by 30 for better visibility

3.5.3 Wide–FOV camera calibration

Tables 3.6 and 3.7 presents results for calibration with polynomial and division distortion model respectively corresponding projection errors. Figure 3.6 presents visualisation of projection errors for calibration with polynomial model.

$\mathbf{Err}_{rms}$	α_x	$\mathbf{x}_0$	$\mathbf{y}_0$	$\mathbf{k}_1$	$\mathbf{k}_2$	$\mathbf{k}_3$
0.34	815.26	1324.03	981.99	0.0149	0.0051	-0.0048

Table 3.6: Projection error and resulting focal length, image centre and polynomial model parameters

$\mathbf{Err}_{rms}$	α_x	$\mathbf{x}_0$	$\mathbf{y}_0$	$\mathbf{d}_n$	λ
0.75	815.57	1324.03	981.99	1	0.00003

Table 3.7: Projection error and resulting focal length, image centre and division model parameters

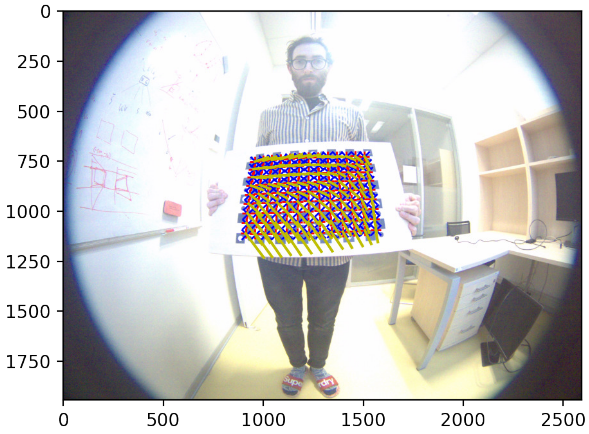

Figure 3.6: Visualisation of projection errors for each point. Error vectors are multiplied by 30 for better visibility

Chapter 4

Multiple camera calibration

4.1 Introduction

In this chapter we proposed method for estimation of relative pose of two cameras mounted on a single frame with no necessity for overlap in FOV. This method is based on observing multiple planes visible in both cameras from different positions. Afterwards are presented results of experiments on data captured in laboratory environment.

4.2 Calibration method

This method estimates relative pose of two cameras by observing multiple planes from different positions. Supposing that world coordinates frame origins at first camera, we have two camera matrices

$$P_a = K_a \left[I | \, 0 \right] \tag{4.1}$$

$$P_b = K_b \left[M \, | \, m \right], \tag{4.2}$$

where M represents relative rotation between two cameras and m represents translation between those cameras. Goal of this method is to estimate M and m. The estimation can be done by solving system of generally non-linear equations induced by viewing multiple planes from two different position. Relation between two views of one plane is given by homography transformation[11]

$$H = R - tn^T/d \tag{4.3}$$

where R, t represents transformation between the viewing positions and $\pi^T = [n^T|d]$ is representation of the plane with respect to coordinates system of camera in the first position.

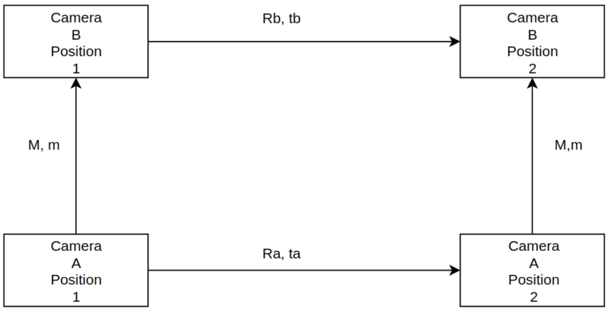

Figure 4.1: Diagram of transformations between cameras and positions

Transformations between camera a and b are derived using coordinates of points in different coordinate systems, a, b identify camera and 1, 2 identify position. Figure 4.2 illustrates this situation.

$$x_{a2} = R_a x_{a1} + t_a \tag{4.4}$$

$$x_{b1} = M x_{a1} + m \tag{4.5}$$

$$x_{b2} = R_b x_{b2} + t_b \tag{4.6}$$

$$x_{b2} = M x_{a2} + m \tag{4.7}$$

By substitution we get equation

$$R_b M x_a + R_b m + t_b = M R_a x_a + Mt + m. \tag{4.8}$$

From this equation we get relations for rotations and translations of camera a and camera b between position 1 and 2.

$$\begin{aligned} R_b &= M R_a M^T \\ t_b &= M t_a + m - R_b m \end{aligned}$$

Transformation of plane π_a to coordinates of camera b is derived from

$$n_a^T x_{a1} + d_a = 0 \tag{4.9}$$

$$n_b^T (M x_{a1} + m) + d_b = 0 \tag{4.10}$$

From (5.10) we get

$$n_b^T M x_{a1} + n_b^T m + d_b = 0 \tag{4.11}$$

Now by comparison of (5.9) and (5.11) we get relations

$$\begin{aligned} n_b &= M n_a & (4.12)\\ d_b &= d_a - m^T M n_a & (4.13)\\ \pi_b^T &= [n_b^T | d_b] & (4.14) \end{aligned}$$

for plane in coordinates of camera b.

For one plane observed by two joint cameras and motion we get set of equations

$$\begin{aligned} H_a &= R_a - \frac{t_a n_a^T}{d_a} & (4.15)\\ H_b &= R_b - \frac{t_b n_b'^T}{d_b'} & (4.16)\\ n_b &= M n_a & (4.17)\\ d_b &= d_a - m^T M n_a & (4.18)\\ R_b &= M R_a M^T & (4.19)\\ t_b &= M t_a + m - R_b m. & (4.20) \end{aligned}$$

H_a and H_b are homographies. For each plane observed from both cameras we get additional two homographies and corresponding equations. Now we have system of equations with 3 DOF forR_a, 3 DOF t_a, 3 DOF for$M$, 3 DOF for m, 4 DOF for n, d. This means 18 DOF and 18 equations (homography is defined up to scale). Each added plane adds 4 DOF for n, d, 2 DOF for homography and another 18 equations. So for at least two planes we get an overdetermined system of equations.

4.3 Optimization problem

Above described system of equations was solved numerically by minimization of error, which is difference between image position of points transformed by homography and their detected position induced by movement of rig. Homographies are computed as in equations (4.15),(4.16) for each plane.

Objective function for one plane can be written as

$$f_1(X_{a1}, X_{a2}, X_{b1}, X_{b2}, R, t, M, m, \pi_1) \tag{4.21}$$
$$err_a = ||proj_{eq,poly}(H_a\, proj^{-1}_{eq,poly}(X_{a1})) - X_{a2}|| \tag{4.22}$$
$$err_b = ||proj_{eq,poly}(H_b\, proj^{-1}_{eq,poly}(X_{b1})) - X_{b2}|| \tag{4.23}$$
$$f_1 = (err_a, err_b). \tag{4.24}$$

For two planes objective function is

$$f = f(X_{a1}, X_{a2}, X_{b1}, X_{b2}, R, t, M, m, \pi_1, \pi_2) \tag{4.25}$$
$$f = (f_1, f_2) \tag{4.26}$$

Output of this function is vector of difference in position for each detected point. This function is minimized with respect to R, t, M, m, π_1, π_2. Rotation matrices are represented by axis and angle for purposes of optimization. Arguments X_{a1},X_{a2},X_{b1},X_{b2} are detections of markers. For minimization was used least squares method. Initial values for optimization were set as rough estimate for expected calibration. Rotation between cameras was initialized to $90°$ around y axis and translation vector was set to $(1, 0, -1)^T$. Normal vectors of planes were initialized as estimate from image view and scalars d were initialized to 1.

4.4 Experiment with synthetic data

First method was verified with synthetic data. For each plane were chosen 5 points. R, t, M, m, π_1, π_2 were chosen to emulate movement and spatial distribution of rig. Positions of points in both camera coordinate systems were calculated. These points were then projected to image coordinates and served as synthetic detections. Optimization was done on those data and converged to chosen parameters R, t, M, m, π_1, π_2. Afterwards the same task was done, but gaussian noise was added to the data to emulate real detection inaccuracy. Again method converged to expected results with noise variance up to 1 px.

4.5 Experiment with real data

Calibration method was then verified on camera rig in laboratory environment. Camera rig, which can be seen at figure 5.1, was dragged across laboratory room with markers taped on the walls and floor which served as planes for calibration. Two of the rigs cameras (rear and right one in figure 5.2) were synchronously capturing images of surroundings during movement. For calibration were provided these images with detected positions of markers used as correspondences. For experiment was chosen movement of the rig where two planes were visible for all time for both cameras. Internal calibration parameters were obtained in experiments in previous chapter and can be seen in Table 4.6. For calibration were used multiple measurements. For each measurement 4 images were chosen. This images had to fulfill condition that at least two common planes were visible for both cameras at both positions and that each plane had to have at least 4 visible correspondences at images from both positions.

4.6 Results of experiment with real data

Here are shown results of estimation of relative pose of two cameras. Results are presented as difference between measured M, m and data measured by different method at laboratory. Difference of M is presented as angular difference between matrices and difference of m is presented as euclidean distance between vectors. Resulting cost of optimization is also presented. Optimization method has ambiguity for length of vector m. This was solved by measuring real distance between cameras. Obtained vector m is normalized and multiplied by this distance to get translation between cameras with correct scale.

cost	$\mathbf{M}_{err}$	$\mathbf{m}_{err}$
3.37	0.02	0.04

Table 4.1: Results of optimization which converged to desired solution

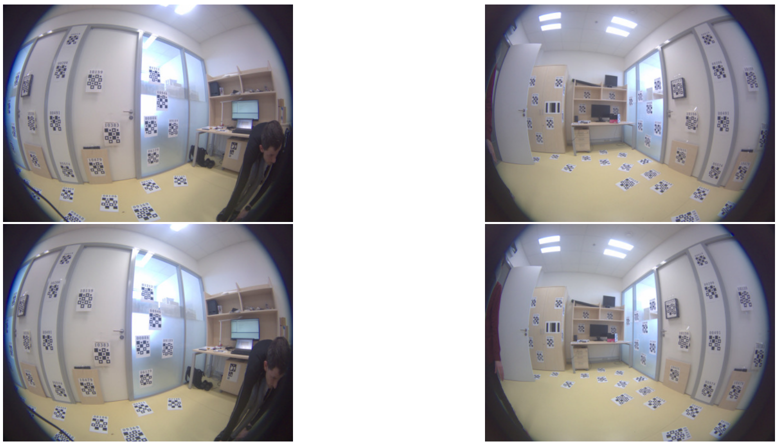

Figure 4.2: Set of images where optimization converged. Upper right is image from camera a at position 1. Upper left is image from camera b at position 1. Bottom left is image from camera a at position 2. Bottom right is image from camera b at position 2.

Results from table 4.1 were achieved on data shown in figure 4.2. Other good results were obtained on data close to data from figure 4.2 in sense of position of the rig in the room. However generally optimization did not converged to expected results.

n	cost	$\mathbf{M}_{err}$	$\mathbf{m}_{err}$
1	5043.1	2.66	0.43
2	5091.3	2.69	0.45
3	96.6	0.10	0.19
4	1367.6	3.11	0.06
5	2597.8	3.10	0.17
6	8.6	0.03	0.19
7	6.44	0.08	0.22
8	1316.7	2.36	0.91
9	6.5	0.09	0.23
10	1677.3	2.65	0.80
11	8.4	0.04	0.23
12	16.6	0.01	0.20
13	20.7	0.01	0.20
14	1690.2	2.92	0.66
15	18.13	0.02	0.31
16	18.95	0.04	1.06
17	3.13	0.01	0.08
18	3.37	0.02	0.04

Table 4.2: Results of optimization for data gathered on 1 drive across laboratory, n is number of measurement and following are value of cost function and difference between result and reference solution

Table 4.2 shows results from calibration done on data from one drive of rig over the laboratory. It can be seen that measured data which actually converged to correct solution had lowest value of cost function (results 17, 18). This is an important observation because it shows that with decreasing value of cost function estimated solution converges to correct solution.

Chapter 5

Technical backgoround

5.1 Introduction

In this chapter will be briefly summarized technical aspects such as laboratory equipment and software used for computations.

5.2 Software

All numerical experiments and data processing were done in python language. For purposes of matrix computations were mainly used functions of numpy library and for numerical computations were used functions of scipy.optimize library. For data visualisation was used matplotlib library.

This work was done distantly on data from laboratory provided by supervisor. For experiments were provided images and detected positions of markers. For purposes of this thesis data were processed to form suitable for experimenting. Experiments were implemented in python as scripts.

5.3 Laboratory equipment

5.3.1 Multi–camera platform

Camera rig is cart with mounted cameras and other sensor equipment. It has wheels attached to its frame so that it could be moved over the floor. Camera calibration was done on both wide–FOV and middle–FOV of the rig. Multiple camera calibration estimated relative position of rear Wide–FOV camera and one of the side cameras.

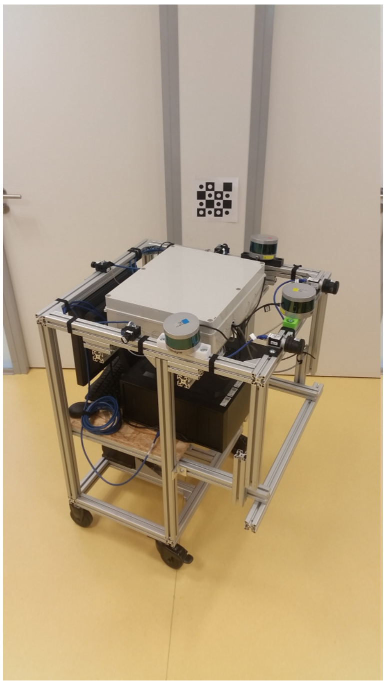

Figure 5.1: Camera rig used for experiments

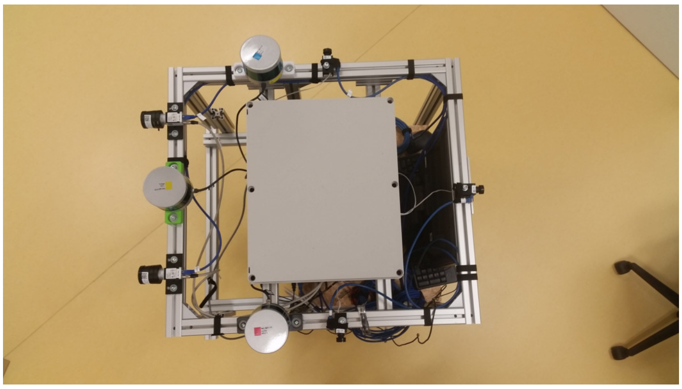

Figure 5.2: Camera rig used for experiments top view. Mounted equipment consists of 2 Middle-FOV cameras in the front of rig (left in image). 3 Wide-FOV cameras on sides and rear (right in image).

5.3.2 Markers for calibration of multiple cameras

For calibration correspondences in images are needed. These correspondences were provided by markers which can be detected in image. Markers were taped to walls and floor of laboratory and each of them had unique identification code. According to those codes it can be determined in which plane detected point lies.

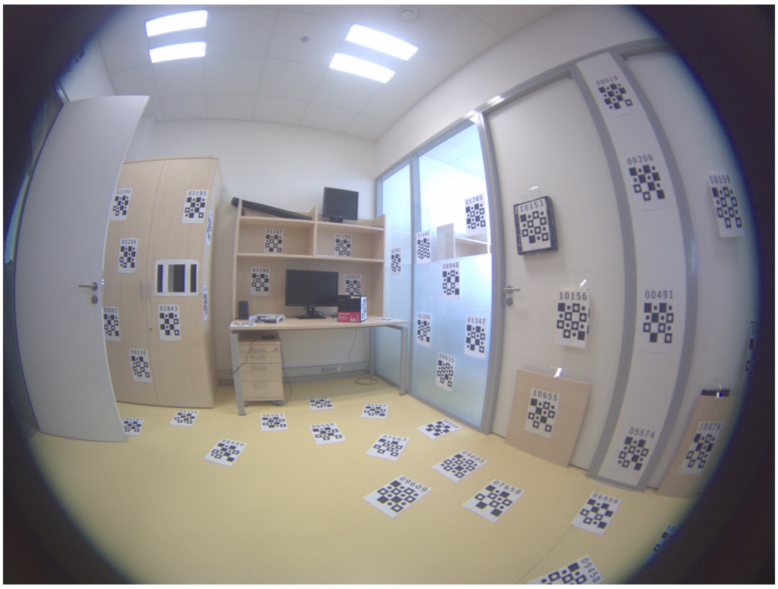

Figure 5.3: View of laboratory with markers taped on walls and floor. For experiment served as planar objects mainly floor and right wall.

Chapter 6

Conclusion

This work presents different approaches to camera calibration and results of their application on camera platform in laboratory environment. tested two different approaches for internal calibration of both middle-FOV and wide-FOV cameras. For wide-FOV cameras were introduced and tested methods for initial estimation of parameters which are based on minimization of reprojection error that is caused by large radial distortion. For each calibration different models of distortion models were tested. These experiments produced results that were further utilized in external calibration.

We introduced method for calibration of relative pose of multiple cameras with non–overlapping FOV. First this method was theoretically derived and than verified on synthetic data and moving camera platform at laboratory. Experiment with synthetic data prooved that this method can produce correct results for calibration even with slightly noisy data. Experiment with real data also produced correct results but these results weren't achieved reliably.

6.1 Future work

It is obvious that calibration method for multiple cameras at present state is not usable in real world applications. Further work can be done on improving reliability of the method. First idea is to use more sets of images in one step of optimization so that result is constrained by more equations. Second idea is to find better initial values for optimization directly from captured data.

Possible improvement for real world application is to adapt this method to use marker–free correspondences.

Bibliography

[1] Y. Wang, W.-L. Chao, D. Garg, B. Hariharan, M. Campbell, and K. Q. Weinberger, "Pseudo-lidar from visual depth estimation: Bridging the gap in 3d object detection for autonomous driving," *Cornell University, Ithaca, NY*, pp. 5–7, 2016.

[2] R. Hartley and A. Zisserman, "Multiple view geometry in computer vision," *Cambridge*, 2007.

[3] Z. Zhang, "A flexible new technique for camera calibration," *Microsoft Research*, 1998.

[4] C. Mei and P. Rives, "Single view point omnidirectional camera calibration from planar grids," *INRIA*, 2004.

[5] E. Malis and M. Vargas, "Deeper understanding of the homography decomposition for vision-based control," *INRIA*, 2007.

[6] P. Sturm, S. Ramalingam, J.-P. Tardif, S. Gasparini, and J. Barreto, "Camera models and fundamental concepts used in geometric computer vision," *Now Publishers*, p. 58, 2011.

[7] P. Drap and J. Lefèvre, "An exact formula for calculating inverse radial lens distortions.," *Sensors, MDPI*, pp. 5–7, 2016.

[8] A. Fitzgibbon, "Simultaneous linear estimation of multiple view geometry and lens distortion," vol. 1, pp. I–I, 2001.

[9] S. Sarabandi, A. Shabani, J. Porta, and F. Thomas, "On closed-form formulas for the 3d nearest rotation matrix problem," *IEEE Transactions on Robotics*, 2020.

[10] Z. Zhang, “A flexible new technique for camera calibration,” *Microsoft Research*, p. 7, 1998.

[11] R. Hartley and A. Zisserman, “Multiple view geometry in computervision,” *Camebridge*, p. 327, 2007.

[12] R. Bra, E. Acar, and T. Kolda, “Resolving the sign ambiguity in the singular value decomposition,” *Sandia National Laboratories*, pp. 7–12, 2007.

Appendix A

Attached files

```
attachment
└─ code
   ├─ calibration.py
   ├─ tangential_projection.py
   ├─ ea_projection.py
   ├─ polynomial_distortion.py
   └─ division_distortion.py
```

Appendix B

SVD sign ambiguity

When estimating solutions of overdetermined systems of equations by SVD, problem called SVD sign ambiguity[12] occurred.

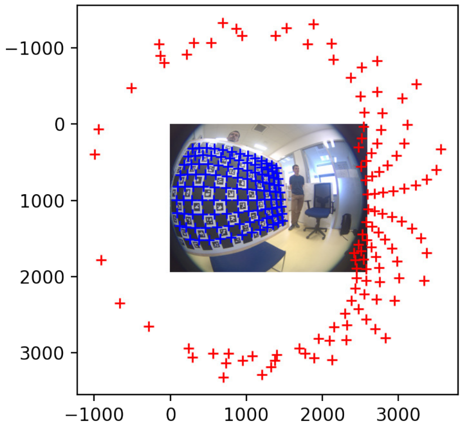

Figure B.1: Results of homography transformation with opposite sign

This problem arose when it was attempted to obtain approximate solution of overdetermined system of linear equations, represented by matrix A, by applying svd on matrix A and taking right singular vector as solution estimate. Image above shows result of using homography with opposite sign and projection of corrupted result. Problem was dealt by checking reprojection error of results given by estimated homography. Homography with opposite sign could be easily recognised because reprojection error was considerably higher (at least 6 orders of magnitude). Homography was than easily corrected by multiplying it by -1.