



FAKULTÄT  
FÜR INFORMATIK  
Faculty of Informatics



Technical Report  
CVL-TR-9

# Five methods to calibrate overlapping surveillance cameras<sup>1</sup>

Timo Kropp

Computer Vision Lab  
Institute of Computer Aided Automation  
Vienna University of Technology

May 16, 2012

## Abstract

Calibration of surveillance cameras requires using appropriate methods to handle different limitations, such as large distortion of the camera lenses, because of a wide field of view. Another limitation is blur in the image which can decrease the sharpness because of fixed focus lenses. To deal with these limitations a fully automatic calibration method from Svoboda is compared with a specialized technique for internal calibration and with a method for external calibration. The former is the Matlab Calibration Toolbox which uses a planar checkerboard pattern as a calibration object. The latter calculates the external camera parameters with a-priori knowledge of the internal parameters. Finally Svoboda's method is compared to another method which estimates the rotation of cameras with the use of an inertial sensor. Through experimental evaluation it will be shown that the specialized methods for internal and external calibration will outperform the fully automatic calibration method from Svoboda, both in terms of accuracy and robustness under the stated limitations.

---

<sup>1</sup>has been funded by the Vienna Science and Technology Fund (WWTF) through project ICT08-030.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Internal Camera Parameters</b>	<b>3</b>
2.1	Methodology . . . . .	4
2.1.1	Svoboda Multi-Camera Self-Calibration Framework . . . . .	5
2.1.2	Matlab Calibration Toolbox . . . . .	6
2.2	Experiments Svoboda Calibration Framework . . . . .	7
2.2.1	Setup . . . . .	7
2.2.2	Results . . . . .	7
2.2.3	Discussion . . . . .	9
2.3	Experiments Matlab Calibration Toolbox . . . . .	10
2.3.1	Setup . . . . .	10
2.3.2	Results . . . . .	12
2.3.3	Discussion . . . . .	13
2.4	Conclusion of internal calibration . . . . .	14
<b>3</b>	<b>External Camera Parameters</b>	<b>15</b>
3.1	Methodology . . . . .	16
3.1.1	Svoboda Framework . . . . .	16
3.1.2	Inertial Sensor . . . . .	16
3.1.3	Essential Matrix . . . . .	17
3.2	Experiments Svoboda Framework . . . . .	17
3.2.1	Setup . . . . .	17
3.2.2	Results . . . . .	17
3.2.3	Discussion . . . . .	20
3.3	Experiments Inertial Sensor . . . . .	21
3.3.1	Setup . . . . .	21
3.3.2	Results . . . . .	22
3.3.3	Discussion . . . . .	22
3.4	Experiments Essential Matrix . . . . .	23
3.4.1	Setup . . . . .	24
3.4.2	Results . . . . .	24
3.4.3	Discussion . . . . .	27
3.5	Conclusion of external calibration . . . . .	27



# Chapter 1

## Introduction

Hoiem showed that object detection from one camera can be significantly improved if knowledge of the geometry is used [3]. This knowledge can be acquired through calibration of the camera. Other applications of camera calibration are, 3D tracking with one camera [5], multi-camera tracking [4] or video surveillance systems [1].

The purpose of this work is to find a suitable calibration method in the context of the latter application. In this application the cameras can have large non-linear lens distortion because of a wide-angle field of view. With a fixed focus lens the image sharpness can be decreased because of blur from regions which are out of focus. The calibration method has to overcome these limitations and be able to calibrate multiple cameras internally and externally. The internal calibration includes the estimation of the focal length, the pixel aspect ratio, the principle point and the radial lens distortion of each camera regarding to the pinhole camera model [2]. The external parameters are the rotation and translation of each camera in a common three dimensional coordinate frame in metric space. It is assumed that the cameras have a slight or full overlapping field of view.

In order to be able to make a comparison, two state-of-the-art calibration methods for each calibration task are selected, which are experimentally evaluated and compared. The evaluation focuses on the robustness and maximal possible accuracy of those methods. The first method is a calibration framework, which was developed by Svoboda [8]. This method is selected because it can estimate the entire calibration of multiple overlapping synchronized cameras simultaneously. Furthermore it is not influenced by blur of regions which are out of focus, because a point light source is used as a calibration object. A change of the camera focus will also change the size of the projected light source but not its center which is used for the calibration. Apart from that the radial distortion is estimated.

Svoboda's method has to perform against one commonly used approach for internal calibration. This approach uses a calibration object of a chessboard pattern. It is called the Matlab Calibration Framework and is based on the method developed by Tsai [9]. This method is not invariant to changes of the camera focus. For that reason the minimal size of the chessboard pattern is experimentally evaluated.

The Matlab Calibration Framework lacks the external calibration of multiple cameras and therefore another method, which estimates the essential matrix to get the rotation and translation of multiple cameras [2], is taken. This method uses a point light source

as a calibration object similar to the technique from Svoboda. Thus it overcomes the limitation of image blur. The robustness and accuracy of this method is compared to the results from the approach of Svoboda for the rotation and translation.

Both methods of the external calibration have the limitations of a calibration object, which must be visible in at least two views. Therefore a third method is additionally evaluated. This approach is basically different to the others because it uses an inertial sensor and is therefore not image based. The direct advantage is that the cameras do not have to be overlapped and it is not influenced by the limitations of image distortion and image blur. Also for this method the robustness and maximal accuracy will be compared.

For each of the stated methods one experiment is performed to estimate the robustness and maximal possible accuracy. Svoboda's method is evaluated in two separated experiments, because it calculates the internal and external calibration in one run. One experiment is used for the evaluation of the internal parameter estimation and the other for the external parameter estimation.

The experiments are realized with CCD Cameras from Sony. The type of sensor and lens used can be typically found in common surveillance cameras at the time of writing this report in 2012. The sensor has a resolution of 640 pixels in x and 480 pixels in y direction and it is assumed that the lens has a non-linear radial distortion. For all experiments there is no ground truth available. That is why the variance of each method is analyzed which provides information about the maximal possible accuracy.

It will be shown that the Matlab Calibration Toolbox clearly outperforms Svoboda's method in estimating the internal camera parameters. Svoboda's method has a maximal error of 99 pixels of the estimated focal lengths compared to the Matlab Calibration Toolbox which only has an error of 2 pixels. The accuracy of the estimated principle point gives a similar view. The first method has an error of less than 1.4 pixels and the latter method has a maximal error of 59 pixels. The experimental evaluation of the external parameters shows that Svoboda's method is outperformed by the specialized method again. Here the calculated rotation matrices from Svoboda have a variation of up to 38 degrees and the camera centers have a variation of up to factor 2 of their actual relative position. The specialized method which estimates the essential matrix results in a residual error of maximal one ninth of a pixel, which is lower than the reprojection error from Svoboda.

This report is split into two major sections. Chapter 2 deals with the evaluation of the two methods of the internal camera calibration and Chapter 3 includes the evaluation of the remaining methods for external camera calibration. Each chapter contains a section about the methodology, two respectively three sections, which explain the experiments and a final conclusion.

# Chapter 2

## Internal Camera Parameters

This chapter requires a well defined camera model that is used for the internal calibration process and regarding to that model certain needed parameters are estimated using state of the art methods [7] and [8].

Based on central projection of points, the pinhole camera model (Figure 2.1) is used as a fundamental concept. A 3D point  $\mathbf{X} = (X_1, X_2, X_3, W)^T \in \mathbb{P}^3$  is projected onto the image plane at point  $\mathbf{x} = (x_1, x_2, w)^T \in \mathbb{P}^2$  through  $\mathbf{x} = (fX_1 + X_3p_x, fX_2 + X_3p_y, X_3)^T$  whereby  $p_x$  and  $p_y$  defines the principle point (Figure 2.1) and  $\mathbb{P}^n$  is the n-dimensional projective space. If  $\mathbf{x}$  is converted from homogenous coordinates to Euclidian coordinates then  $\mathbf{x} = (fX_1/X_3 + p_x, fX_2/X_3 + p_y)^T$ .

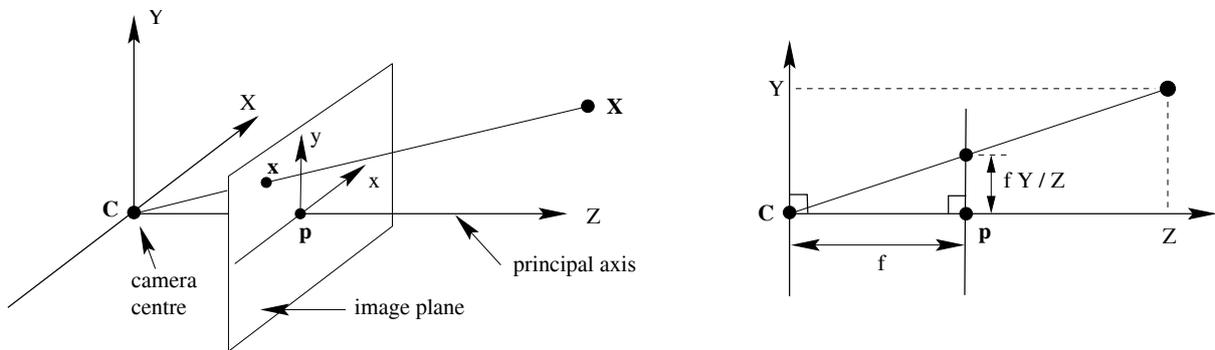


Figure 2.1: Pinhole camera geometry: 3D point  $X$  is projected on the image plane at position  $x$  (Source [2])

Not every camera has squared pixels so a small adaption to this simple model has to be made. Therefore the focal length is determined with 2 parameters, one in  $x$  and one in  $y$ -direction. Now the camera model is represented as the linear mapping between the homogenous coordinates:

$$\begin{pmatrix} x_1 \\ x_2 \\ w \end{pmatrix} = \begin{bmatrix} f_x & s & p_x & 0 \\ 0 & f_y & p_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{pmatrix} X_1 \\ X_2 \\ X_3 \\ W \end{pmatrix} \quad (2.1)$$

whereby  $w$  and  $W$  are the homogenous parts of the vectors and  $s$  is called skew and is assumed to be zero.

Let be

$$K = \begin{bmatrix} f_x & s & p_x \\ 0 & f_y & p_y \\ 0 & 0 & 1 \end{bmatrix} \quad (2.2)$$

then

$$\mathbf{x} = K[I \mid \mathbf{0}]\mathbf{X} \quad (2.3)$$

The matrix  $K$  is called camera calibration matrix and it includes all intrinsic parameters of the calibration process.

This linear model with homogenous coordinates of projective mapping from 3D points to the 2D image plane of a camera is not sufficient to describe the non-linear properties which are introduced by the lens [2]. To keep the model simple a radial distortion is assumed. Each normalized image point  $(x_1, x_2)$  is distorted to another position  $(x_{d1}, x_{d2})$  under

$$\begin{pmatrix} x_{d1} \\ x_{d2} \end{pmatrix} = L(r) \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \quad (2.4)$$

where  $L$  is a function of the radial distance  $r = \sqrt{x_1^2 + x_2^2}$  to the distortion center which is assumed to be identical to the principle point [9]. Normalized image points refer to image points that were multiplied with the inverse of  $K$ , thus its coordinates are measured in units of the focal length. The function  $L$  can be approximated with the Taylor expansion  $L(r) = 1 + \kappa_1 r + \kappa_2 r^2 + \kappa_3 r^3 + \dots$ . In the following experiments only the second, fourth and sixth order is taken into account. This is  $\kappa_2$ ,  $\kappa_4$  and  $\kappa_6$  from the Taylor expansion.

This chapter deals with the evaluation of two different approaches that are used to estimate the focal length, principle point and radial distortion of perspective cameras in real world scenarios.

## 2.1 Methodology

In the following experiment, two different methods for internal camera calibration will be evaluated. To maximize the performance of each method a special setup is established. The aim is to find at least one method which can determine the internal camera parameters up to a variance of 5 pixels for the estimated focal length and the principle point.

The first evaluated method is one part of the Svoboda Multi-Camera Self-Calibration Framework [8]. One advantage is that all cameras can be calibrated simultaneously. In the case of surveillance cameras though there is another advantage which is even more important. As stated in Chapter 1 an assumption is made that surveillance cameras are typically focused to infinity and therefore objects which are close to the lens appear blurry in the image. Because the Svoboda Framework detects a projected point light source, the blur effects only the size of the blob but not the position of its center. Thus the detection accuracy is not significantly influenced by the blur.

Matlab’s Calibration Toolbox is used as the second method. In this calibration process each camera has to be calibrated separately but its potential lies in the fact that it detects rectangular structures from a chessboard pattern. The angles of viewing vectors can be determined with the calibration matrix  $K$  and this information is reflected by the rectilinear structure on the calibration pattern. This fact is promising with regards to the desired accuracy with a maximum error of 1% for the estimated focal length.

### 2.1.1 Svoboda Multi-Camera Self-Calibration Framework

Svoboda’s idea of his framework is to calibrate not only one camera but many in one batch without the requirement of a large calibration object. Instead only a freely moving bright spot is required [8]. The outcomes are not only internal camera parameters including radial distortion, but also the orientation and position of all the cameras. With the assumption that columns and rows of the camera sensors are orthogonal, the framework can stratify projective structures to Euclidean ones. Thus complete camera projection models are estimated through the detection of the moving bright spot.

The algorithm consists of the following steps. First the bright spot is projected on each camera image if it is visible in that camera view. With a 2D Gaussian fitting, used as a point spread function, sub-pixel precision for the blob detection of the bright spot can be reached independently in each image. Next epipolar constraints are applied in pairs to each point and are used for their validation. With a rank-4 factorization projective motion and shape are estimated. Finally parameters of a non-linear distortion model are calculated with iterative refinement [8].

To give a brief mathematical background of the idea behind that algorithm,  $m$  cameras and  $n$  object points  $\mathbf{X}_j = [X_j, Y_j, Z_j, 1]^T, j = 1, \dots, n$  are assumed. According to the pinhole camera model (see Chapter 2) the following equation holds

$$\lambda_j^i \begin{pmatrix} u_j^i \\ v_j^i \\ 1 \end{pmatrix} = \lambda_j^i \mathbf{u}_j^i = P^i \mathbf{X}_j, \quad \lambda_j^i \in \mathbb{R}^+ \quad (2.5)$$

where  $u_j^i$  and  $v_j^i$  are the measured pixel coordinates of the  $j$ -th point in the  $i$ -th camera.  $\lambda$  is a scaling factor to keep the homogenous part equal to 1 of the vector  $\mathbf{u}$ . The corresponding 3D points are  $\mathbf{X}_j$  and  $P^i$  is the camera matrix of the  $i$ -th camera which includes the internal and the external parameters of that camera

$$P = K[R | \mathbf{t}] \quad (2.6)$$

where  $K$  is the camera calibration matrix,  $R \in \mathbb{R}^{3 \times 3}$  is the rotation matrix and  $\mathbf{t} \in \mathbb{R}^3$  is the translation vector [2].  $R$  and  $\mathbf{t}$  are referring to the external parameters of a camera (see Chapter 3 for more details).

To estimate projection matrices  $P^i$  and scale factors  $\lambda_j^i$  Svoboda et al. plugged all points and camera projections into one matrix  $W_s$

$$W_s = \begin{bmatrix} \lambda_1^1 \begin{bmatrix} u_1^1 \\ v_1^1 \\ 1 \end{bmatrix} & \cdots & \lambda_n^1 \begin{bmatrix} u_n^1 \\ v_n^1 \\ 1 \end{bmatrix} \\ \vdots & & \vdots \\ \lambda_1^m \begin{bmatrix} u_1^m \\ v_1^m \\ 1 \end{bmatrix} & \cdots & \lambda_n^m \begin{bmatrix} u_n^m \\ v_n^m \\ 1 \end{bmatrix} \end{bmatrix} = \begin{bmatrix} P^1 \\ \vdots \\ P^m \end{bmatrix}_{3m \times 4} [\mathbf{X}_1 \cdots \mathbf{X}_n]_{4 \times n} \quad (2.7)$$

$$W_s = P\mathbf{X} \quad (2.8)$$

The matrix  $W_s$  is called the scaled measurement matrix [8] and  $P$  and  $X$  relates to the projective motion and the projective shape, respectively. If  $\lambda_j^i$  are known and there is a minimum of point measurements then  $W_s$  has rank 4 and can be factored into  $P$  and  $X$  [8]. Through that factorization  $P$  and  $X$  are estimated up to a  $4 \times 4$  projective transformation. Under certain assumptions as stated above this can be calculated too to obtain Euclidean results. Svoboda et al. also proposed a way to estimate the projective depth  $\lambda_j^i$ . It uses the calculation of the fundamental matrix from the cameras pairwise.

### 2.1.2 Matlab Calibration Toolbox

Instead of calculating every parameter of the camera projection matrix like Svoboda et al. does, the Calibration Toolbox from Matlab focuses on estimating the focal length, principle point and radial distortion. The same perspective projection model for the camera is used as stated in Chapter 2.

The underlying idea is to estimate a homography between a planar calibration object and the image plane. Therefore there has to be a well detectable structure on that plane. One solution is a chessboard pattern with evenly distributed corners of each cell. With a corner detector their positions in the image can be estimated.

There exists a  $3 \times 3$  homography because of their coplanarity that depends on the relative position of the camera and the plane and the camera's intrinsic parameters [7]. If it is assumed that the calibration plane is the plane  $Z = 0$  then the following equation holds

$$H \cong KR \begin{bmatrix} 1 & 0 & \\ 0 & 1 & -\mathbf{t} \\ 0 & 0 & \end{bmatrix} \quad (2.9)$$

Where  $R$  is the rotation matrix of the camera regarding the calibration object and  $-\mathbf{t}$  is the translation vector of the camera to the calibration object.  $K$  is the camera calibration matrix defined in Chapter 2. The homography  $H$  can be calculated from four or more point correspondences. If the metric structure of the plane is known then  $H$  can be decomposed in its parts which are shown in Equation 2.9. The desired parameters are included in  $K$ . To estimate the homography  $H$  an over-determined equation system of the measured point coordinates has to be solved by least square minimization. To make that process more stable in terms of accuracy more than one view of the calibration object

should be used [10]. In Chapter 2.3.2 it is shown that about 10 different views give stable results.

The calibration toolbox concurrently calculates the camera matrix  $K$  and the parameters of the radial distortion model. Therefore the radial distortion parameters are initialized with zero and in an iterative procedure the reprojection error of the chessboard corners are minimized by gradient descent. In each step of the iteration  $K$  is refined with the previous estimated distortion parameters.

## 2.2 Experiments Svoboda Calibration Framework

In this chapter the calibration framework from Svoboda [8] will be experimentally evaluated in terms of robustness and accuracy regarding the internal camera parameters. In the following sections the environment of the experiments is depicted, the results are detailed and then discussed with a final conclusion.

### 2.2.1 Setup

Calibrating multiple cameras simultaneously with the Self-Calibration Framework from Tomas Svoboda requires that the cameras are synchronized for taking images at the same time and all targeting the same scene. The calibration object has to be visible and detectable in at least 3 cameras at each timestamp when frames are captured. To establish these demands the four Sony cameras were mounted on tripods, placed next to each other and directed in the same direction so that each viewing volume is overlapping with the other three. An external hardware trigger is connected to each camera and is used to cause the recording of images by generating periodical events.

Due to a suggestion from Svoboda [8] a standard laser pointer is used to generate a small spot at each camera image with an average diameter of 3 pixels. This calibration device is moved manually through the overlapping viewing volume. To improve the calibration results the projected laser light on each camera sensor has to cover as much space as possible of the image plane [8].

For each camera the focal length, principle point, non-linear radial distortion (2nd order) and the standard derivation of the reprojection error in x and y direction is estimated.

There are four different sets of image of a moving bright spot. Each set is evaluated separately and contains 2700 images (set 1) or 500 images (set 2-4).

### 2.2.2 Results

The first step of the Svoboda Multi-Camera Self-Calibration Framework detects the spots in each image for each camera with subpixel accuracy [8]. Figure 2.2 shows one sample image from camera 1 and image set 1 with the detected center.

At the final step, the bundle adjustment tries to minimize the reprojection error. The final result is shown in Figure 2.3. The outcome is that about 60% of the measured points can be reconstructed with a visual verifiable low reconstruction error relative to the image

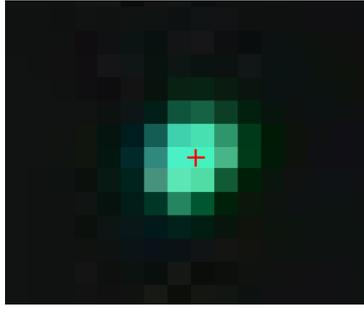


Figure 2.2: Crop of an image with the projected laser light (green) and detected center with subpixel accuracy (red cross)

size (see Figure 2.3). Also some points are reprojected to the outlying part of the image plane. These were detected by the other 3 cameras.

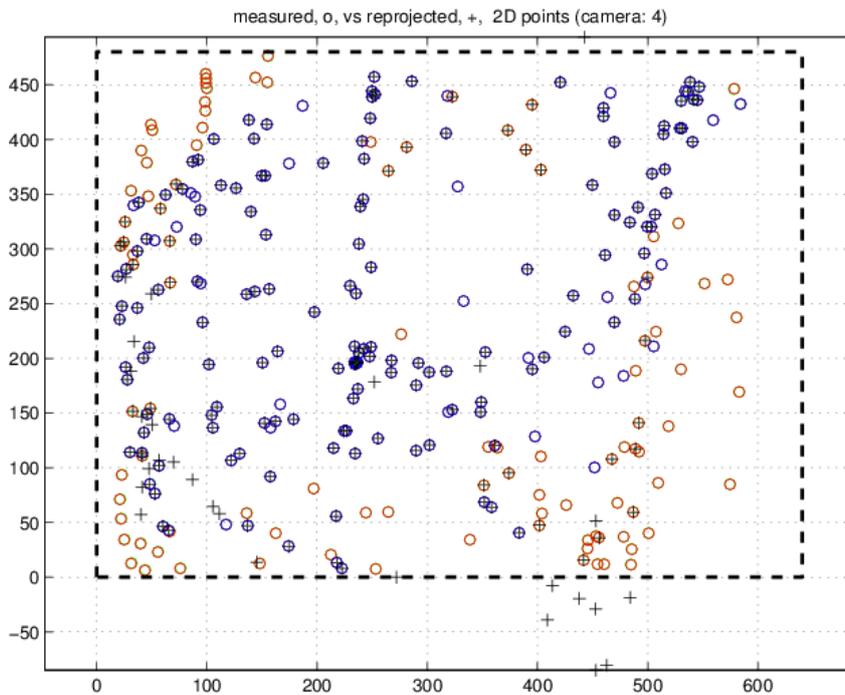


Figure 2.3: Detected points classified as outliers (red circles), detected points detected as inliers (blue circles) and reprojected points (black crosses) on the image plane of a camera. The black dotted rectangle is the camera image border

The estimated focal length varies for camera 1 respective to the different image sets in a range from 616 to 711 pixels in the x and y direction. Similar results point out for the other cameras, concrete 617 to 704 pixels for camera 2, 616 to 715 pixels for camera 3 and 649 to 706 pixels for camera 4. The maximal difference between the focal length in x and in y direction is over all cameras for all image sets 1.543 pixel and the minimum is 0.001 pixel (Table 2.1). Therefore from this method it can be assumed that the sensing elements of the camera has and aspect ratio of 1 in our model.

Another estimated parameter is the principle point. The calculated principle points for each camera are in the range of 304 to 363 pixels in x direction and for y it's in 222 to 255 pixels (Table 2.1). These results are closed to the half of the image height and width which are  $640/2 = 320$  pixels and  $480/2 = 240$  pixels respectively .

To minimize the reprojection error the non linear lens distortion is taken into account. A non linear radial distortion model is used and the estimated parameter (2nd order) is in the range of  $-0.042$  to  $-0.002$  for all four cameras (Table 2.1).

The minimized standard derivation of the reprojection error could be reduced for each camera with each image set to a value less than 0.241 pixel in x and 0.244 pixel in y within 10 iterations of the algorithm (Table 2.1).

		Focal Length		Principle Point		distortion	Error (Std)	
		x	y	x	y	2nd order	x	y
Camera 1	Set 1 (2700)	710.266	710.407	338.833	241.069	-0.040	0.141	0.142
	Set 2 (500)	688.973	689.000	331.740	240.487	-0.027	0.150	0.153
	Set 3 (500)	616.208	616.320	332.320	236.395	-0.017	0.123	0.133
	Set 4 (500)	672.457	672.558	315.238	235.606	-0.029	0.136	0.140
Camera 2	Set 1 (2700)	705.016	704.838	341.536	240.487	-0.029	0.142	0.190
	Set 2 (500)	684.546	683.884	326.574	241.922	-0.016	0.161	0.201
	Set 3 (500)	617.219	617.097	306.711	254.094	-0.002	0.134	0.178
	Set 4 (500)	672.774	673.104	300.832	241.554	-0.021	0.125	0.179
Camera 3	Set 1 (2700)	714.818	714.959	305.236	222.204	-0.041	0.155	0.168
	Set 2 (500)	686.983	687.054	304.113	227.193	-0.018	0.151	0.167
	Set 3 (500)	616.567	616.547	334.030	247.854	-0.017	0.241	0.204
	Set 4 (500)	663.426	663.767	298.112	229.038	-0.011	0.176	0.174
Camera 4	Set 1 (2700)	705.572	705.247	306.523	225.640	-0.042	0.134	0.180
	Set 2 (500)	682.637	682.062	313.167	230.410	-0.009	0.118	0.186
	Set 3 (500)	649.833	651.375	362.671	255.181	-0.020	0.189	0.244
	Set 4 (500)	672.231	672.232	320.715	235.085	-0.036	0.113	0.154

Table 2.1: Results of internal calibration with Svoboda Calibration Framework and all 4 Sony cameras. With each setup the cameras were calibrated simultaneously.

### 2.2.3 Discussion

While the results have a low reprojection error of less than 0.25 pixels, the estimated focal lengths of all four cameras in one set of images varies significantly less than the estimated focal lengths of one camera in all four different sets. This indicates that the results are not unbiased from this method. This effect can not be identified for the estimated principle point, but from the results it is still not possible to make a unique prediction for each camera due to its variance.

As a consequence the Svoboda Multi-Camera Self-Calibration Framework does not produce results which are accurate enough in each determined internal parameter for the purpose of this work. The variation within the different image sets is too large with a range of 99 pixels for the focal length. The difference of the largest and smallest value of

the determined principle point is 59 pixels for x and 33 pixels for y which is even a large uncertainty under the initial premises (see Chapter 1).

There is a correlation between the estimated parameter of the radial distortion and the focal length. The higher the focal length the lower is the value of the distortion model. Despite of the large variation of the estimated values for the parameter the reprojection error remains always smaller than a quarter of a pixel. This indicates that Svoboda's method is overfitting the data.

Besides the assumption of overfitting, another reason for the large variation of the focal length and principle point, with respect to the different image sets, the decomposition of the actual camera matrix  $P$  plays a significant role (see Chapter 2.1.1).

## 2.3 Experiments Matlab Calibration Toolbox

As well as the experimental evaluation of Svoboda's calibration framework this chapter covers another experimental evaluation. In this case it is dealing with the performance and robustness estimation of the Matlab calibration toolbox. Finally the two experiments give an opportunity to conclude which one is preferable regarding the environment. First the setup, the results and the discussion of the second experiment are presented.

### 2.3.1 Setup

To estimate the internal parameters of the surveillance camera, with the Matlab Calibration Toolbox, a planar calibration object is needed. A structure similar to a chessboard has to be printed onto the surface of that object (Figure 2.4).

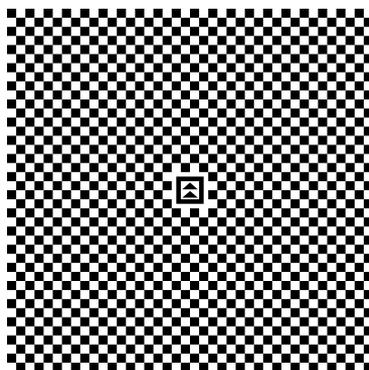


Figure 2.4: Pattern of calibration object

It is assumed that the focus of the lens of the camera is fixed and is set to infinite. Then the depth of sharpness is maximal but objects near to the camera are blurred in the image. The actual amount of blurriness depends not only on the objects distance to the camera but also on the diameter of the aperture and the focal distance too. Because of that the calibration object has to be at a certain distance to the camera, so that the corners of the chessboard pattern have a minimum sharpness and a minimum size to be detectable by the method.

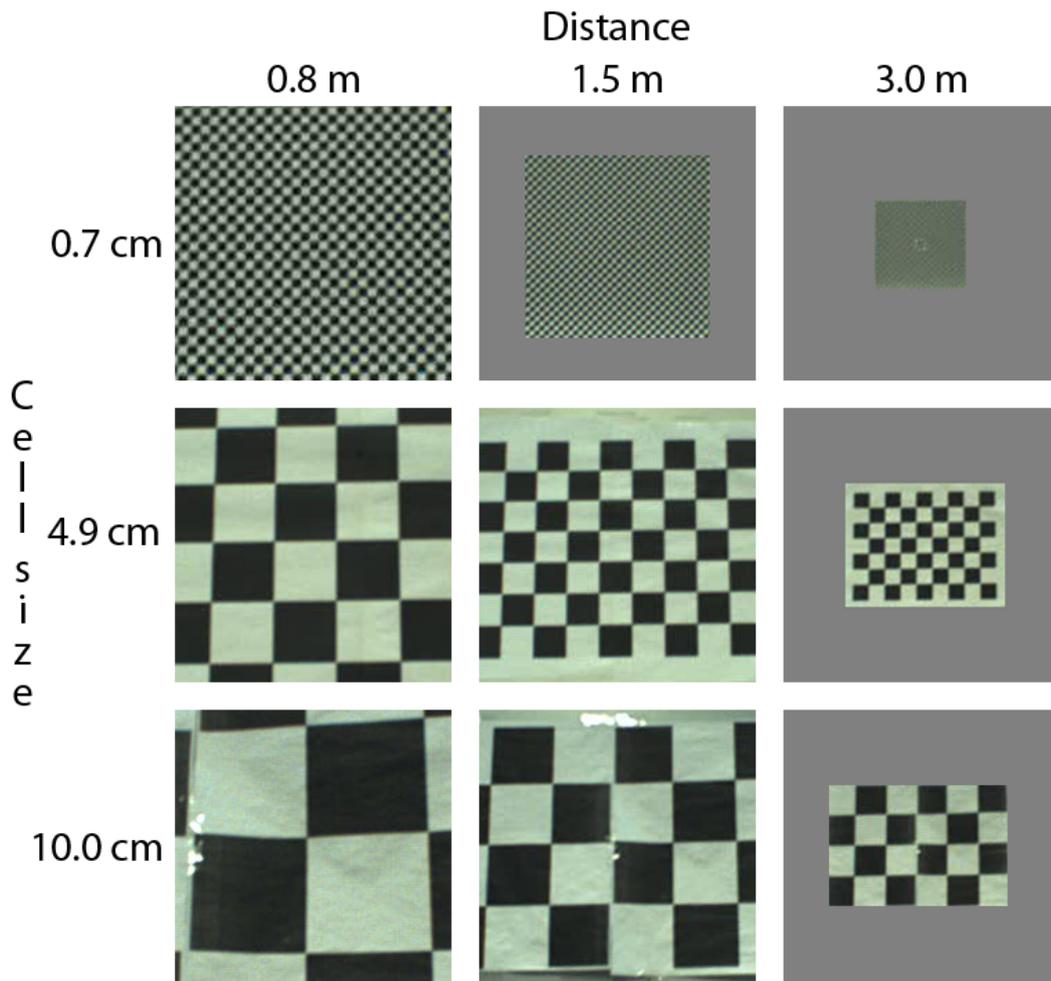


Figure 2.5: Images from Sony Camera with different cell sizes and distances to camera

To estimate the minimum size of the squares with regards to the distance from the camera different, values were tested. Figure 2.5 shows that a cell size of the chessboard pattern of 0.7 cm is too small even at a low distance of 0.8 m because the corners are blurred and cannot be detected with the Matlab Calibration Toolbox. The sharpest results are given by the cell size of 10.0 cm. But even with a size of 4.9 cm the corners are clearly distinguishable. From this it follows that the cell size can have a minimum size of 4.9 cm in both width and height so as to be detectable even at a distance of 3 meters from the camera.

For the setup a calibration object was constructed with a cell size of 5.0 cm of the chessboard pattern. The marker contains 40 cells in width and height. Due to limitations of available printers the final chessboard pattern has a width of 140 cm (28 Cells) and 200 cm height (40 Cells) (Figure 2.5).

To estimate the focal length, the principle point and the non linear lens distortion for each of the four cameras 35 images of the calibration object from different angles were taken. The stability of the resulting values estimated by the Toolbox is evaluated while

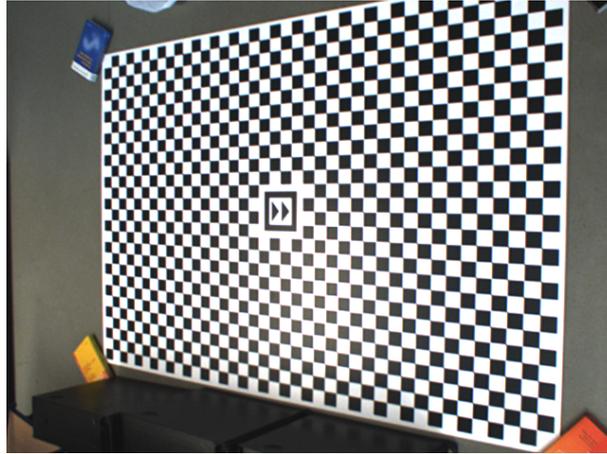


Figure 2.6: Calibration object used to estimate the internal camera parameters

for each camera the 35 images are divided into three sets of 12, 12 and 11 unique images and are processed separately.

### 2.3.2 Results

After reading all captured images for the calibration process from the memory, the corners of the chessboard pattern are recognized. For more than 90% of the images all visible corners were detected. The accuracy of all detections is similar to Figure 2.7 shown below.

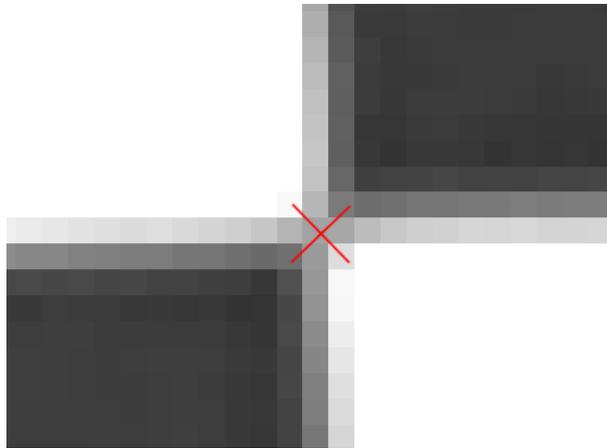


Figure 2.7: Crop of chessboard pattern with one corner and its detected position (red cross)

The resulting focal length in x and y direction is in the range of 685 and 692 pixels. All cameras are equipped with the same sensor and have the same lens mounted. This leads to the expectation of equal focal lengths and fits with the actual estimated variation of 7 pixels. Apart from the fourth, for all cameras, each of its evaluated subset results to a focal length in x and y with a maximal difference of 0.753 pixel (camera 1, all, focal

length  $y$ ) regarding to all images. The biggest difference of camera 4 is 2.094 pixel for its focal length (camera 4, set 1, focal length  $x$ ) (Table 2.2).

The variation of the focal length in  $x$  and  $y$  direction for all calculated values is less than 1 pixel. For each camera there is a set which has a difference less than 0.1 pixels (camera 1, set 3; camera 2, set 3; camera 3, set 3; camera 4, all). Therefore it is assumed that the physical sensor elements of the cameras are squared (Table 2.2).

Similar to the estimated focal length the estimated principle point for each camera varies in each subset in maximal 1.5 pixels (camera 4, set 2) and at least 0.008 pixels (camera 1, set 2) regarding to all images of the respective cameras. Between each camera the principle point varies 23 pixels in  $x$  and 7 pixels in  $y$  direction (Table 2.2).

For all four cameras the estimated coefficients of the radial distortion model are less than  $|0.14|$  (2nd, 4th and 6th order) and not smaller than  $|0.11|$  (2nd and 4th order) and  $|0.003|$  (6th order) (Table 2.2).

The standard deviation of the reprojection error in  $x$  and  $y$  direction, which is minimized by the Calibration Toolbox, is in all cases less than 0.19 pixels. The only outlier is test set 3 of camera 1 with about 0.27 pixels in  $y$  (Table 2.2). This slightly worse result has no significant effect on the estimated parameters for this image set.

		Focal Length		Principle Point		distortion			Error (Std)	
		$x$	$y$	$x$	$y$	2nd	4th	6th order	$x$	$y$
Camera 1	All	686.587	686.875	327.407	242.669	-0.119	0.155	0.024	0.188	0.155
	Set 1	-0.500	-0.753	-0.966	-0.343	0.004	-0.038	0.099	0.142	0.149
	Set 2	0.140	0.434	-0.008	0.194	-0.001	0.001	-0.005	0.140	0.157
	Set 3	-0.063	-0.118	0.082	0.200	-0.003	0.041	-0.113	0.272	0.156
Camera 2	All	685.576	685.832	316.282	245.691	-0.119	0.158	0.033	0.157	0.168
	Set 1	0.152	-0.001	-0.334	-0.852	0.000	-0.002	0.010	0.141	0.153
	Set 2	0.263	-0.024	-0.278	-0.577	0.004	-0.034	0.073	0.155	0.224
	Set 3	0.163	0.221	0.158	0.857	-0.004	0.035	-0.079	0.169	0.132
Camera 3	All	686.604	686.688	305.781	239.747	-0.121	0.166	-0.004	0.138	0.133
	Set 1	-0.540	-0.403	0.083	0.086	0.002	-0.017	0.033	0.136	0.112
	Set 2	0.194	-0.006	-0.259	-0.230	0.000	0.002	-0.015	0.143	0.140
	Set 3	0.220	0.187	-0.031	-0.219	-0.003	0.023	-0.035	0.135	0.145
Camera 4	All	689.305	689.559	320.393	241.054	-0.122	0.177	-0.024	0.140	0.152
	Set 1	2.094	2.031	-0.012	-1.348	-0.003	0.022	-0.055	0.127	0.133
	Set 2	-0.137	0.802	1.462	-0.234	0.003	-0.036	0.085	0.118	0.167
	Set 3	1.202	0.624	0.038	-0.470	0.003	-0.016	0.031	0.152	0.138

Table 2.2: Results of internal calibration with the Matlab Calibration Toolbox for all 4 Sony Cameras. The results of each set except the error are the absolute difference to the result from the overall set (all).

### 2.3.3 Discussion

Assuming this approach is unbiased then the focal length is estimated with an accuracy of  $\pm 2$  pixel. This conclusion is made from the previous results because for independent image sets of at least 11 images the calculated focal lengths never differ more than 2 pixels. Consequently the Matlab Calibration Toolbox has robust outcomes in calculating the focal length in  $x$  and  $y$  direction.

A similar statement can be made for the remaining parameters which are estimated during the calibration process. For each camera the principle point varies for all independent image sets less than 1.0 pixel except for two cases with 1.3 pixels and 1.4 pixels. For the width of 640 pixels this is less than 0.15% and for the height of 480 pixels it is less than 0.21%. These values underline the previous concluded robustness of the implementation.

The assumed radial distortion model is stable in the results for the second order values too. The deviation of each independent set compared to the overall set in each camera is less than 3.4%. For the remaining higher order it increases to a maximum of 26.5%. Only the difference of the sixth order is up to 470% (Camera 1 Set 1 and 3). This apparently huge variance can be disregarded because the absolute difference is less than 0.12 pixels for the sixth order and less than 0.038 pixels for the fourth order. The overall error of maximal 0.28 is higher than the error in the distortion model.

## 2.4 Conclusion of internal calibration

Two different calibration methods for estimating internal camera parameters were presented. Each one was independently evaluated experimentally with the use of the same cameras. Both methods require a calibration object which has to be captured from different positions and orientations. Results from the experiments show that the Matlab Calibration Toolbox clearly outperforms the Svoboda Calibration Framework in terms of robustness and the resulting accuracy. The former achieved a variation of 2 pixels for the calculated focal length (Chapter 2.3.3), where the latter has a variation of 99 pixel for the same cameras (Chapter 2.2.3). The results are reliable with regards to the estimated reprojection error which is less than 1/4 pixel in each case. Similar numbers are given for the principle point where the maximal variation is 1.4 pixels with the Matlab Calibration Toolbox and 59 pixels with the Svoboda Calibration Framework. These relatively large differences can result from the fact that the Svoboda Calibration Toolbox is primarily calculating the camera matrix  $P$  which can be in general estimated only up to a projective ambiguity and only with special assumptions (Chapter 2.1.1) stratify to Euclidean structure. Finally the matrix  $K$  with the internal parameters has to be extracted from  $P$ . This process can introduce variant errors. In contrast the Matlab Calibration Toolbox estimates the internal parameters directly from a homography between a planar calibration object and the image.

Concluded, in terms of accuracy and reliability the Matlab Calibration Toolbox is preferable to the Svoboda Calibration Framework for estimating the internal camera parameters. The latter depicts its advantages if multiple cameras need to be calibrated and the time factor is more important than the accuracy of the resulting parameters. If a precision of only 20% percent is needed then Svoboda's Calibration Framework will give expected results.

# Chapter 3

## External Camera Parameters

To complete the pinhole camera model of Chapter 2, 6 additional parameters are added. Such parameters determine the relative position of several cameras to each other and are called external or extrinsic parameters. Three values determine the translation in the x, y and z direction and 3 parameters determine the orientation what is constituted of three angles for each axis in the 3 dimensional spaces (see Figure 3.1).

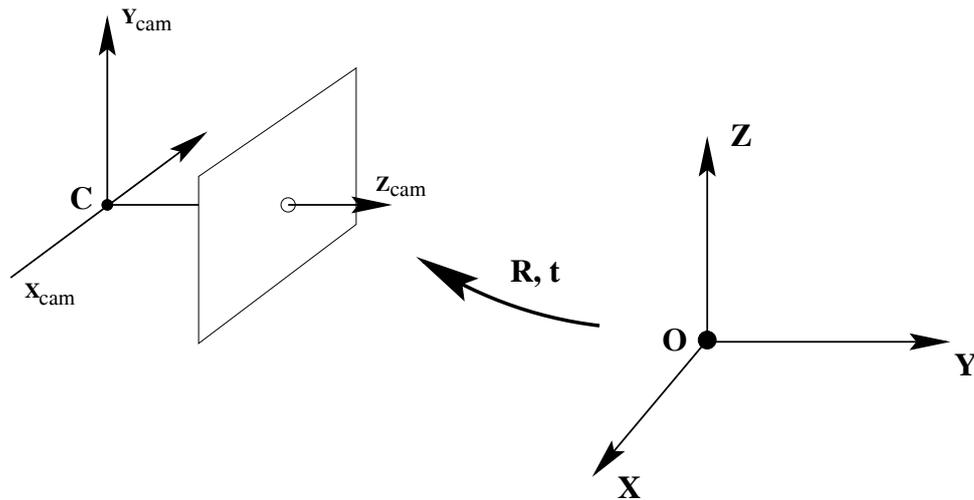


Figure 3.1: Transformation between two coordinate frames to describe external calibration ([2])

Regarding to the camera projection model  $\mathbf{x} = P\mathbf{X}$  with  $\mathbf{x} \in \mathbb{P}^2$  and  $\mathbf{X} \in \mathbb{P}^3$  the camera projection matrix  $P \in \mathbb{R}^{3 \times 4}$  is extended with the translation vector  $\mathbf{t} \in \mathbb{R}^3$  and rotation matrix  $R \in \mathbb{R}^{3 \times 3}$ :

$$P = K[R \mid \mathbf{t}] \tag{3.1}$$

with the camera calibration matrix  $K \in \mathbb{R}^{3 \times 3}$ .

## 3.1 Methodology

This chapter deals with the experimental evaluation of three different methods for external camera calibration of multiple cameras. Two of these methods estimate the rotation matrix and translation vector of several cameras from their images of a calibration target. The third method does not require a calibration object and the images do not need to overlap. Using this latter method, only the rotation matrices are calculated.

### 3.1.1 Svoboda Framework

The camera matrix  $P$  is the central target of Svoboda's Calibration Framework. It is estimated through point projections on several camera images (see Figure 3.2). In addition to the external parameters, the cameras internal parameters are included in  $P$  with  $P = K[R | \mathbf{t}]$ . To extract the rotation matrix  $R$  and the translation vector  $\mathbf{t}$ ,  $P$  is transformed from projective space to Euclidean space with stratification. For more details see Chapter 2.1.1 where the fundamentals of Svoboda's Calibration Framework are explained.

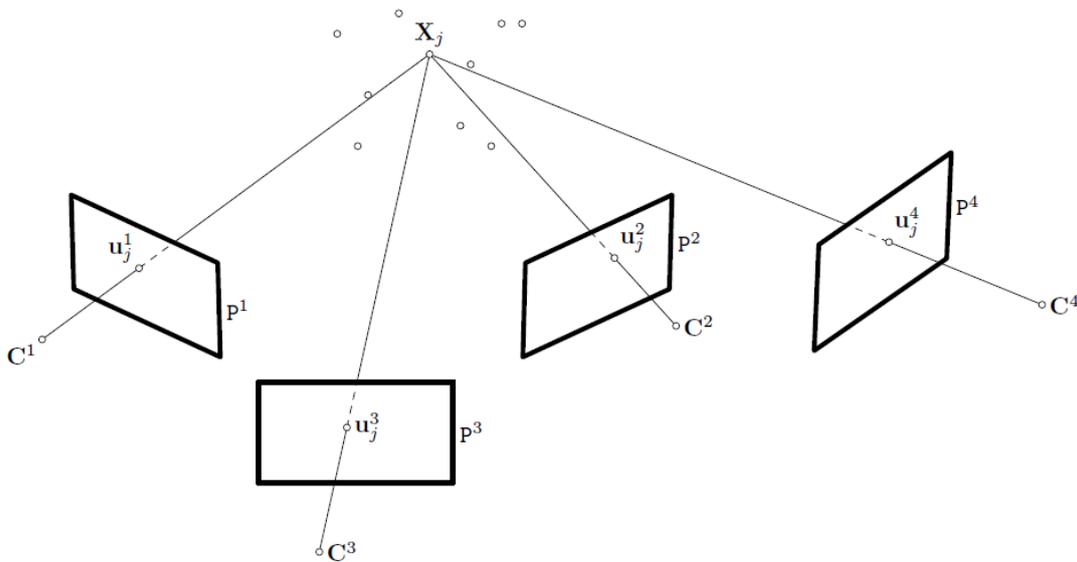


Figure 3.2: Outline of 4 cameras with their centers, images and projected points ([8])

### 3.1.2 Inertial Sensor

With the use of an inertial sensor the intention is to estimate the orientation of each camera within a global coordinate frame. In this way, information concerning the cameras location or even its internal calibration is not a necessary predetermined parameter. In comparison to image based techniques, here the cameras' viewing volumes do not need to overlap. The inertial sensor MTi Xsensor measures the 3 Euler angles specific to gravity and the Earth's magnetic pole.

The top of the camera's body consists of a plain surface. This is used to mount the inertial sensor on top of the body of the cameras. Subsequently the camera's outer body orientation is measured with the inertial sensor.

### 3.1.3 Essential Matrix

The latter of the three methods suitable for estimating the external calibration based on the calculation of the essential matrix which is a particular fundamental matrix [2]. In principle this technique is similar in terms of the required calibration object to Svoboda's Calibration Framework. A free moveable bright spot is needed and its projections are detected synchronously on two camera images. The results of the internal calibration from Chapter 2 are used to improve the estimation of the external camera parameters with the detected points. With the knowledge of the focal length, principle point and radial distortion, the essential matrix can be calculated. From the essential matrix the rotation and translation of the two related cameras can be reconstructed in Euclidean space [2].

## 3.2 Experiments Svoboda Framework

Svoboda's Calibration Framework estimates not only the internal camera parameters but also the external parameters which are the orientation and translation of each camera within a specific coordinate system.

Aside from the evaluation of internal calibration, a second experiment has been developed to estimate the robustness and thus the highest achievable accuracy of the Svoboda Calibration Framework for external camera calibration.

Following the necessary initial setup, the results obtained from four different test scenarios are here discussed.

### 3.2.1 Setup

To evaluate the performance of the Svoboda Calibration Framework of the resulting extrinsic parameters, the same setup is used as for the evaluation of the internal parameters (see Chapter 2.2.1). To give a brief reminder, four cameras capture simultaneously images of the same scene from different positions and viewing angles. A bright spot is manually moved through the working volume and the actual 3D positions are recovered from their projected image locations. Finally the camera matrices  $P$  are estimated from the 3D and 2D positions of the points and then the internal and external parameters are extracted from  $P$ . Given the projective invariance of  $P$  a Euclidean stratification has to be applied before extracting the parameters [8].

### 3.2.2 Results

Using the above setup, four different test runs are evaluated and then compared. Each case consisted of maximal 900 detectable points. After the Euclidean stratification (see

Chapter 2.1.1) each test results in a different Euclidean frame. The origin is placed in the center of the reconstructed points and the orientation and scale is arbitrary. To make the results comparable normalization is applied to the estimated poses. Therefore the origin is moved to the position of camera 1 and the orientation is changed to camera 1 as well. That means  $C_1 = (0, 0, 0)^T$  and  $R_1$  is the  $3 \times 3$  identity matrix. The overall scale can only be approximated. Each test case is scaled to the unit of the mean Euclidean distance of camera 2, 3 and 4 to camera 1. To improve the readability the position and orientation of camera 1 is not included in the tables of the following evaluation.

Reconstructed points and camera poses are returned by the algorithm from Svoboda et al. For all four test cases an image of this reconstruction is shown in Figure 3.3. The 3D position of each entity cannot be distinguished from that image, but from the specific point of view the reader can see the general order of the camera setup which is correctly related to the real setup.

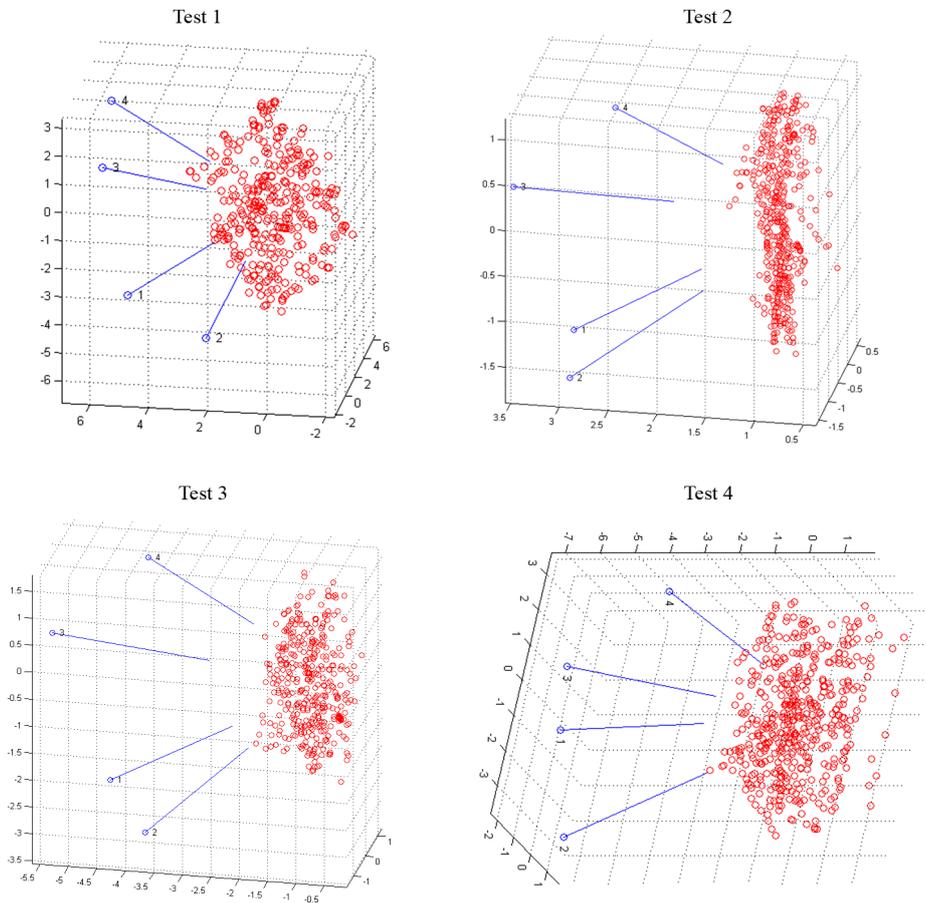
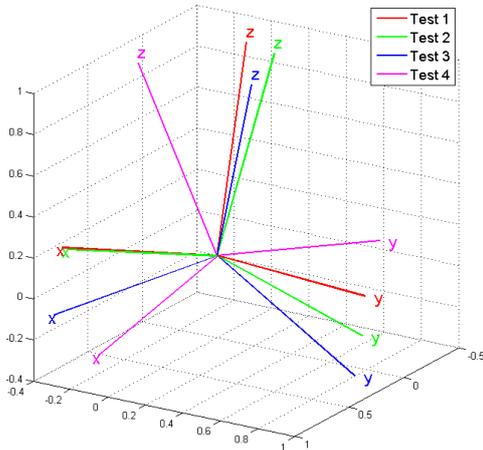


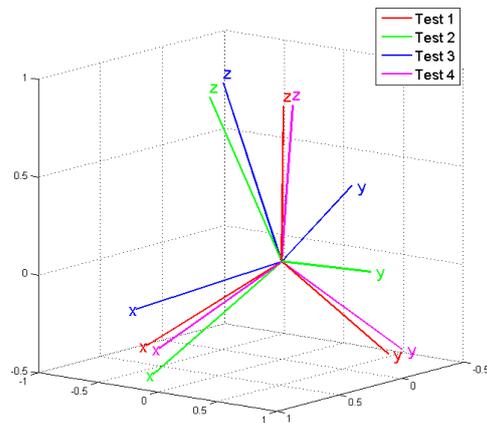
Figure 3.3: Reconstructed points, camera positions and their principle axis from 4 different test cases.

For a more detailed visual impression of the orientation of each camera in each test case Figure 3.4 gives a deeper insight. From these results it is deduced that camera 2 and 3 have similar orientations in each test case in the sense that the different axis x, y and

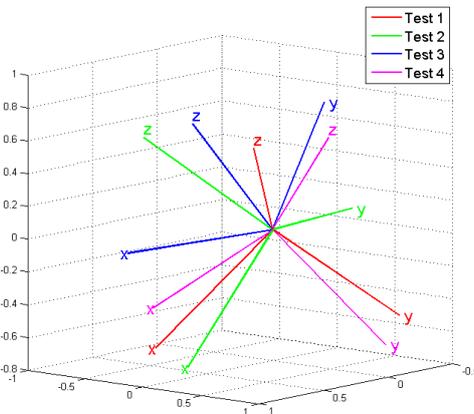
z are clustered. This means that the inner angle between the axis is lower than the angle between two different axes. Only camera 4 plays an exceptional role. Here the y-axis of test case 3 is closer to the cluster of the z-axes than it is to the cluster of the y-axes (see Figure 3.4). From that figure the orientation of the x-axis is the most stable result.



(a) Camera 2



(b) Camera 3



(c) Camera 4

Figure 3.4: Results of the orientations of the cameras from different test cases.

After this qualitative evaluation from visual results of Svoboda's Calibration Framework a quantitative analysis is accomplished. To make the orientations comparable the angle between each axis of the corresponding camera to the axis of the coordinate system is shown. They are called  $\alpha$ ,  $\beta$  and  $\gamma$  for the x-, y- and z-axis.

The angles and camera positions of each test case for the cameras 2 to 4 with the standard deviation of the reprojection error is given in Table 3.1. For all executed tests

the error in the x and y direction is less than 0.233 pixel and thus small enough to be compared.

For the resulting angles  $\alpha$ ,  $\beta$  and  $\gamma$  the range for each camera is given in Table 3.1 which is the absolute value of the difference between the largest and smallest value of all tests. Contrary to the previous conclusion from the qualitative analysis, the orientation of the x-axis does not give the most stable results but instead the z-axis does which relates to the  $\gamma$  angle lies in the smallest range of maximal 9.1 degrees for all cameras. The largest range is formed by the results of camera 4 and the  $\beta$  angle which is 37.6 degrees. The most stable results are camera 2 with 9.5 degrees for  $\alpha$ , 8.9 for  $\beta$  and 6.3 for  $\gamma$ .

Given the normalization of the overall scale for each camera the absolute scale unit is unknown. Thus the camera centers can be compared only in a relative fashion.

The smallest range is in the y-values of camera 2 which is 0.71 and the largest range is in the x-values of camera 4 which is 2.19. The actual value of the former is between  $-0.4$  and  $0.3$  and for the latter it is between  $-1.2$  and  $1.0$ . Therefore the range is about twice as high as the actual position. The overall largest value of a coordinate from the estimated camera centers is 1.0 (camera 4, test 1, x) and the smallest value is  $-1.2$  (Camera 4, test 4, x). Camera 2 and camera 3 have similar ranges for the x and y values in comparison to camera 4. The x-range has a difference of less than 0.5% and the y-range has a difference of less than 11%.

		Angles			Camera centers			Error (Std)	
		$\alpha$	$\beta$	$\gamma$	x	y	z	x	y
Camera 2	Test 1	22.530	16.816	15.119	-0.497	-0.400	-0.246	0.147	0.168
	Test 2	21.069	18.489	16.197	-0.681	-0.063	-0.065	0.140	0.206
	Test 3	19.436	25.718	19.886	-0.154	0.312	0.586	0.140	0.184
	Test 4	13.045	18.684	21.370	0.590	-0.295	-0.168	0.132	0.163
	max - min	9.485	8.902	6.251	1.271	0.712	0.832		
Camera 3	Test 1	18.752	22.250	28.211	0.824	-0.139	-0.409	0.232	0.173
	Test 2	28.156	13.399	26.776	0.474	-0.425	-0.675	0.160	0.165
	Test 3	13.529	33.821	31.022	0.111	-0.938	0.090	0.161	0.175
	Test 4	16.673	22.328	27.918	-0.453	-0.442	0.679	0.183	0.170
	max - min	14.627	20.422	4.246	1.277	0.799	1.354		
Camera 4	Test 1	38.645	31.000	50.227	1.026	0.582	-0.727	0.171	0.158
	Test 2	46.488	20.181	50.352	0.945	-0.966	-0.308	0.127	0.199
	Test 3	8.632	57.746	57.350	-0.521	-1.201	-0.401	0.200	0.223
	Test 4	18.870	45.843	48.243	-1.168	-0.656	0.378	0.144	0.185
	max - min	37.856	37.565	9.107	2.194	1.783	1.105		

Table 3.1: Normalized results of external camera parameters from Svoboda’s Calibration Framework. Each test case relates to the same camera configuration.

### 3.2.3 Discussion

The implemented experiment targets the estimation of the robustness of the calibration method from Svoboda. The absolute accuracy is not determined because there is no

ground truth available for the data, but the maximal precision of the algorithm is evaluated by re-running the algorithm with different data using the same setup. Therefore, similarly to the previous sections the robustness is estimated.

Assuming Svoboda’s method is unbiased, the maximal accuracy for the estimated orientation is not better than 37.9 degrees for the  $\alpha$  angles (Section 3.2.2). This relatively large variation of the resulted orientation indicates that the method is not robust due to different image sets.

The estimated camera center in the coordinate frame is even less accurate. The error is up to twice as high as the actual position (see Section 3.2.2). Despite the relatively small reprojection error of less than 0.24 pixels, Svoboda’s Framework cannot estimate the orientation or the center of four cameras in the given setup with a similar high accuracy. One of the reasons is that the camera matrices  $P$  are optimized over the reprojection error and not the rotation and translation itself. Not until the matrices  $P$  are factorized with a Euclidean stratification of its components. This final step can lead to the large uncertainty observed.

### 3.3 Experiments Inertial Sensor

One promising way to get the desirable accuracy is to use an external inertial sensor. The Xsense MTi 3.5 has an accuracy of  $< 1$  degrees in all 3 measured angles (yaw, roll and pitch) according to the datasheet of the producer. These angles are referenced by both gravity and the earth’s magnetic field. The device can only measure its rotation but not the translation to any reference system.

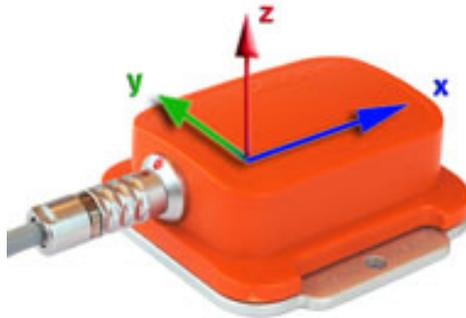


Figure 3.5: Inertial sensor MTi from Xsense with its coordinate system

With the following experiment the general use of the Xsense MTi for determining orientation of surveillance cameras will be evaluated. Due to several uncertainties (see Chapter 3.3.1) a comparison to other image based techniques is not yet feasible.

#### 3.3.1 Setup

To determine the rotation of the surveillance cameras with regards to the pinhole camera model (see Chapter 3) the inertial sensor is mounted on top of each camera body and

manually aligned. If each camera uses the same body and the inertial sensor is placed to it relatively in the same position, then the rotation of each camera body can be determined in the same reference system. One of the main disadvantages is that significant uncertainty occurs by placing the inertial sensor manually on the camera body. Another uncertainty is that the position of the camera sensor relative to the body can be arbitrary. Regarding the pinhole model, the orientation of the camera is determined by the orientation of the camera sensor and the lens and not by the camera body.

### 3.3.2 Results

Three different and independent experiments with the inertial sensor were carried out. Each scene contains 4 cameras, the mean value and standard derivation of the estimated 3 Euler angles for each camera are shown below. During measurement the MTi Xsense returns 100 values for each angle per second. Thus the mean and standard derivations are calculated from at least 800 measurements (see Table 3.2).

For the Roll angle the standard derivation is always less than 0.3 degrees. The results of the Pitch angle are less than 0.09 degrees and for the Yaw angle they are less than 1.02 degrees. The results of the mean values are shown in Table 3.2. The Roll mean is always lower than the Pitch mean and the Pitch mean is always lower than the Yaw mean.

		Mean			Std		
		Roll	Pitch	Yaw	Roll	Pitch	Yaw
Scene 1	Camera 1	0.060	-10.154	-151.759	0.068	0.036	0.226
	Camera 2	-0.692	-24.505	65.488	0.083	0.056	0.076
	Camera 3	-0.224	-4.877	-159.632	0.058	0.037	0.210
	Camera 4	1.516	-0.670	84.921	0.054	0.032	0.164
Scene 2	Camera 1	-0.293	19.391	-98.882	0.110	0.035	0.071
	Camera 2	1.257	17.696	135.803	0.098	0.035	0.083
	Camera 3	0.385	14.915	112.277	0.102	0.039	1.017
	Camera 4	0.915	19.212	156.423	0.201	0.056	0.576
Scene 3	Camera 1	-1.483	18.802	166.392	0.181	0.036	0.237
	Camera 2	-1.693	7.639	163.538	0.083	0.086	0.144
	Camera 3	-0.656	-0.258	-170.325	0.040	0.073	0.037
	Camera 4	0.706	-3.453	-153.053	0.064	0.030	0.056

Table 3.2: Results from three different scenes of external calibration with inertial sensor.

### 3.3.3 Discussion

The expected accuracy of a maximal error of 1 degree was not observed. The standard derivation of the yaw angle of camera 3 in scene 2 is 1.017 degrees. In all other 35 cases it is less than 1 degree. From these results it is followed that the MTi Xsense sensor gives robust results. But regarding the orientation of the camera sensor and lens this method is not unbiased because only the orientation of the camera body is measured. The relationship between the sensor and the body is still unknown.

In cases where cameras are reachable by humans, the use of the inertial sensor gives robust results without the need of an external calibration object or access to the camera image. But in any case only the orientation of the camera body is measured and not the orientation of the sensor itself which is not equivalent. This method is also not suitable for calibration of the entire external parameters. It is only suitable for the orientation, but the position of the camera is never calculated. One advantage of using this method is that the cameras do not have to have an overlapping viewing volume. The orientation of each camera is determined independently.

### 3.4 Experiments Essential Matrix

For the last of the three methods for external calibration, one experiment is used which allows two error measurements to be estimated using four different methods for estimating the essential matrix, which has known calibration of the cameras' internal parameters. With a priori knowledge, no camera calibration from the projected points has to be observed, which can be error-prone for the use of the Euclidean stratification. Without any knowledge about the cameras, only the fundamental matrix can be calculated from scene points and thus the camera positions and rotations can only be reconstructed in projective space which is not satisfactory for the stated problem. Thus the essential matrix is computed. From its values the camera matrices can be retrieved up to scale and four-fold ambiguity in metric space [2]. The correct camera matrices can be chosen from the ambiguity by reconstructing the measured points in 3D space. Only for one configuration all points are lying in front of every camera.

To validate the reliability of the estimated essential matrices the mean reprojection error is used which is defined as:

$$\frac{1}{N} \sum_i^N d(\mathbf{x}_i, \hat{\mathbf{x}}_i) + d(\mathbf{x}'_i, \hat{\mathbf{x}}'_i) \quad (3.2)$$

where  $\mathbf{x}_i \leftrightarrow \mathbf{x}'_i$  are the measured correspondences, and  $\hat{\mathbf{x}}_i$  and  $\hat{\mathbf{x}}'_i$  are the estimated correspondences that satisfy  $\hat{\mathbf{x}}'_i K'^{-T} E K^{-1} \hat{\mathbf{x}}_i = 0$  and  $K$  and  $K'$  are the camera matrices and  $E$  is the essential matrix.  $N$  is the number of measurements and  $d(\mathbf{x}, \mathbf{y})$  is the Euclidean distance between  $\mathbf{x}$  and  $\mathbf{y}$ .

As the second error measurement, the residual error is used and defined as:

$$\frac{1}{N} \sum_i^N d(\mathbf{x}'_i, F\mathbf{x}_i) + d(\mathbf{x}_i, F^T\mathbf{x}'_i) \quad (3.3)$$

with  $F = K'^{-T} E K^{-1}$ . It can be geometrically interpreted as the mean distance from the measured points in the image to their epipolar lines of the corresponding points. Similar to that the reprojection error has also a geometrically interpretation. It describes the mean distance between the measured points and their reprojections from the estimated 3D positions.

The algorithms compared for estimating the essential matrix from measured image points in two views are: first the normalized 8-point algorithm for the fundamental matrix,

from Hartly and Zisserman [2], adapted to estimated the essential matrix, second the 5-point algorithm from Nister [6] for solving the relative pose problem, third and fourth the gold standard algorithm for estimating  $F$  which is the Maximum Likelihood Estimate (MLE) for  $F$ , from Hartly and Zisserman [2] adopted to estimated  $E$ . The solutions of the first two methods are used as initial values for the last method and the results are compared separately.

### 3.4.1 Setup

This method requires an overlap of at least two cameras, that one point can be detected in both views. The idea is that the cameras need to overlap only in a narrow space. In total there are four cameras which are positions at the corners of a parallelogram from top view. In each camera at least one other camera is visible in its upper right or upper left corner (see Figure 3.6). This fact allows to compare the visible position of the epipole later on, which is the camera center with the calculated epipole. The cameras are located in an indoor environment and have a distance between 3 and 5 meters to each other.

To accomplish the experiment a LED-light is moved through the space visible by the cameras. Similar to Svoboda’s Method (see Section 2.1.1) the light dots are captured simultaneously with the cameras.

### 3.4.2 Results

The fundamental technique given the entire estimation of the essential matrix and the methods used is the detection of the image points with sub-pixel precision. This can be compared to the method in Section 2.2.2 where accuracy of at least half a pixel is obtained, which can be concluded from the reprojection error (see Figure 3.6 and Table 3.3).

After applying an undistortion to the measured pixel, the coordinate positions of the project light dots are significantly changed, especially any which are close to the border of the image. Because of the slight overlap of the cameras the correction of the lens distortion plays a significant role for the essential matrix  $E$ , estimated from image points, because there is a shift of more than one pixel of the measured positions with the used lenses and cameras (see Figure 3.6).

From the estimated essential matrix the epipoles can be extracted. They are the left and the right null-space of the essential matrix [2]. For each image point in one camera the corresponding epipolar line in the image of the other camera can be computed with  $E$ . If  $E$  has rank 2 and the two singular values are identical, then all epipolar lines intersect in the epipole [2]. For one calculated essential matrix of camera 1 and 2 all points and their epipolar lines are shown in Figure 3.7. The covered area of the blue and red lines correspondence to the overlapping field of view and also determine its border, because epipolar lines from image points at the border are included. The overall area is relatively small in respective of the entire image (see Figure 3.7 for verification).

The estimated essential matrices between the cameras from 1 to 4 all have a reprojection error of less than 0.7 pixels for all 4 methods. The lowest error has the 5 Point and the MLE 5 algorithm with less than 0.099 pixels, which is a displacement of less than a

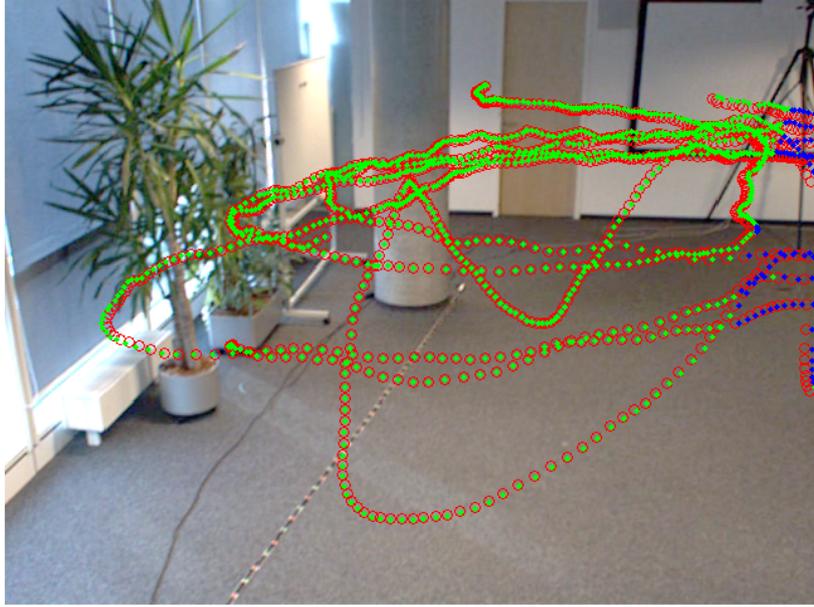


Figure 3.6: Detected points in one camera (green dots), detected points in two cameras (blue dots) and their distorted original position (red circles).

tenth of a pixel. Thus the Maximum Likelihood Estimate (MLE) has no significant change over the 5 Point algorithm (see Table 3.3). If the result from the 8 Point algorithm is used as an initial value for the MLE, then an improvement of more than half a pixel can be obtained if the error from the 8 Point algorithm is not already less than 0.11 pixels (Table 3.3).

The Residual Error mirrors a similar impression as the reprojection error. Here the MLE 5 cannot improve the values returned by the 5 Point algorithm significantly (Table 3.3). The difference is that the error values of the camera 2 and 3 configuration are up to 30 times larger, but the absolute value is still less than 0.13 pixels and for the other 2 camera configurations it is less than 0.004 which is smaller than the precision of the original measured points which was up to 0.5 pixels.

		8 Point	5 Point	MLE 8	MLE 5
Residual Error (Pixel)	Camera 1-2	0.0040	0.0005	0.0010	0.0005
	Camera 2-3	0.1226	0.1166	0.1223	0.1165
	Camera 3-4	0.0029	0.0004	0.0004	0.0004
Reprojection Error (Pixel)	Camera 1-2	0.709	0.098	0.172	0.098
	Camera 2-3	0.106	0.089	0.101	0.089
	Camera 3-4	0.585	0.082	0.081	0.082

Table 3.3: Residual and Reprojection Errors of epipolar geometry from calibrated cameras of different algorithms for estimating the essential matrix

Even a reprojection error in the value range of the measurement precision of the projected points does not guarantee to have the same precision in the resulted camera

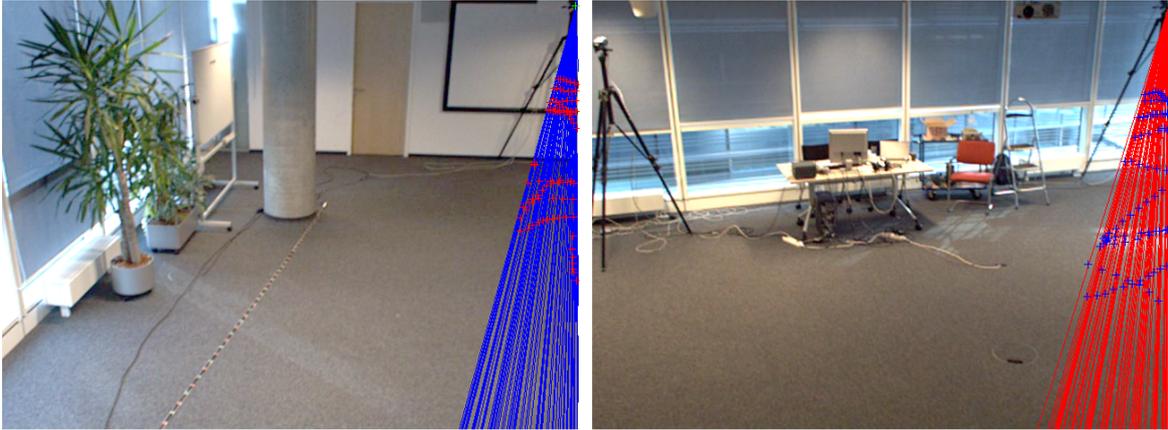


Figure 3.7: Camera 1 (right) and camera 2 (left) with detected points (crosses) and their corresponding epipolar line in the other image (same color).

matrices extracted from the essential matrix [2]. But in order to avoid having any further steps to give more evidences of the correctness of  $E$  without a ground truth, the calculated epipoles can be compared with the images of the camera where the true epipole is located. For the image of camera 3 captured by camera 2 the epipoles are estimated from the 4 algorithm and plotted in Figure 3.8. All estimated epipoles are lying in a range of less than half a pixel. From Figure 3.8 it can be concluded that the true epipole is located no more than 3 pixels away from the estimated ones. For all other 3 cameras the estimated epipoles are positioned always on the visible camera body.

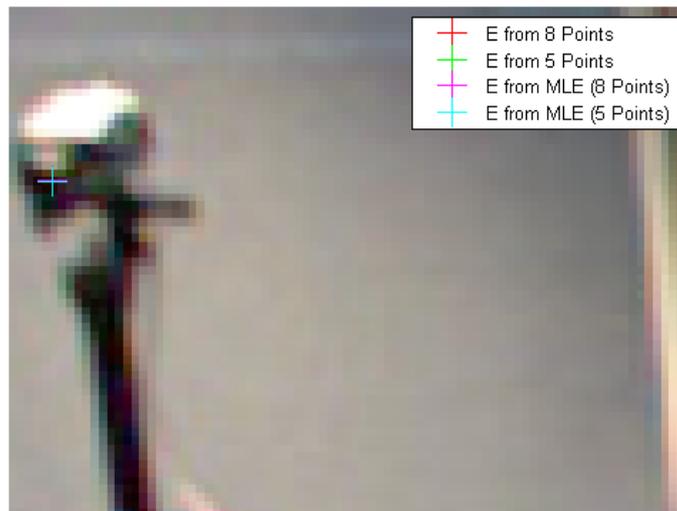


Figure 3.8: Reprojected epipoles of camera 3, viewed from camera 2 (cropped image).

### 3.4.3 Discussion

The given experiment showed that the Maximum Likelihood Estimate can improve the results from the 8 Point algorithm significantly. But the residual error and reprojection error from the 5 Point algorithm have no improvement in the tested range of  $10^{-4}$  pixel. That means that the time consuming MLE algorithm is not needed in the given setup because the 5 Point algorithm from Nister [6] already has the same precision. The most important statement which is followed from this experiment is that even with detected points which are close to only one border of the image and with a covered area of no more than 10% of the entire image, the camera matrices can be estimated in metric space with calibrated cameras as foreknowledge and the obtained reprojection error is less than 0.2. This is in the range of the precision of the measurements from the projected points. The accuracy is confirmed by the projected epipole with an error of maximal 3 pixels.

## 3.5 Conclusion of external calibration

Three different methods for external camera calibration were presented. Two of them estimate the rotation and translation of several cameras in metric space. The other one can only be used for estimating the rotation of camera bodies, but the advantage of this method is that the results are located in a world reference frame regarding to the gravity and the magnetic field of the earth, also there is no need of overlapping field of views. The other two methods require overlapping cameras. The first needs as much overlap as possible with at least 3 cameras and the second can handle slight overlapping cameras but it also needs pre-calibrated cameras. For both methods the reprojection error is in the range of the detection accuracy of the projected points which is maximal half of a pixel. In despite of this error the method from Svoboda [8] is not robust regarding different calibration data point clouds. The experiments showed a variation up to 38 degrees of the estimated rotation matrices. Also the estimated camera centers have a variation up to factor 2 of their actual relative position. If it is assumed that Svoboda's method is unbiased, although the precision of the results are not better than the actual variation from different input data with the same camera setup. More reliable results came from the Essential Matrix method for which already calibrated cameras are used. Therefore reconstruction can be realized in metric space directly without further assumptions. From the results a reprojection error is obtained in the same range as the one from Svoboda's method but here also a second error measurement is evaluated which is the residual error. This allows also for a geometrical interpretation and for all compared methods of estimating the essential matrix, it is always lower than a ninth of a pixel, which is even lower than the reprojection error. Further evidence that the essential matrix method is more robust than Svoboda's method, is that the error of the reprojected epipole is maximal 3 pixels which is 0.5 percent of the image width of 640 pixels.

# Bibliography

- [1] R. T. Collins, A. J. Lipton, T. Kanade, H. Fujiyoshi, D. Duggins, Y. Tsin, D. Tolliver, N. Enomoto, O. Hasegawa, P. Burt, and L. Wixson. A system for video surveillance and monitoring. Technical Report CMU-RI-TR-00-12, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, May 2000.
- [2] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521540518, second edition, 2004.
- [3] D. Hoiem, A. A. Efros, and M. Hebert. Putting objects in perspective. *Int. J. Comput. Vision*, 80:3–15, October 2008.
- [4] K. Kim and L. S. Davis. Multi-camera tracking and segmentation of occluded people on ground plane using search-guided particle filtering. In *ECCV*, pages 98–109, 2006.
- [5] V. Lepetit and P. Fua. Monocular model-based 3d tracking of rigid objects. *Found. Trends. Comput. Graph. Vis.*, 1:1–89, January 2005.
- [6] D. Nistér. An efficient solution to the five-point relative pose problem. *IEEE Trans. Pattern Anal. Mach. Intell.*, 26:756–777, June 2004.
- [7] P. F. Sturm and S. J. Maybank. On plane-based camera calibration: A general algorithm, singularities, applications. In *CVPR*, volume 1, pages 1432–1437, 1999.
- [8] T. Svoboda, D. Martinec, and T. Pajdla. A Convenient multi-camera self-calibration for virtual environments. *PRESENCE: Teleoperators and Virtual Environments*, 14(4):407–422, August 2005.
- [9] R. Y. Tsai. A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the-shelf tv cameras and lenses. *IEEE Journal on Robotics and Automation*, 3(4):323–344, 1987.
- [10] Z. Zhang. Flexible camera calibration by viewing a plane from unknown orientations. In *in ICCV*, pages 666–673, 1999.