

Camera Self-Calibration with Known Camera Orientation

Dissertation

zur Erlangung des akademischen Grades
Doktor der Ingenieurwissenschaften
(Dr.-Ing.)
der Technischen Fakultät
der Christian-Albrechts-Universität zu Kiel

Dipl.-Inf. Jan-Michael Frahm

Kiel
2005

1. Gutachter _____ Prof. Dr.-Ing. Reinhard Koch

2. Gutachter _____ Prof. Dr.-Ing. Helmut Mayer

3. Gutachter _____ Dr. Rick Szeliski

Datum der mündlichen Prüfung _____ 30. Mai 2005

Acknowledgments

I would like to express my gratitude to my advisor Prof. Reinhard Koch, for giving me the opportunity to work in his research group. Furthermore I would like to thank him for buying the first orientation sensor and giving the task of using it to me. I also would like to thank him for the fruitful discussions and the help in stressy situations.

I would like to thank Dr. Rick Szeliski and Prof. Helmut Mayer for accepting to be in my reading committee. I thank Dr. Rick Szeliski for the discussions during my internship in his group as well as I thank Prof. Helmut Mayer for the interesting discussions whenever we met at conferences. I am also grateful to the other members of the jury Prof. Rudolph Berghammer and Prof. Peter A. Hoeher.

I also indebted to many of my colleges. I thank Friso Evers-Senne for his continuous software support for the camera interfaces. My colleagues Felix Woelk and Kevin Köser I would like to thank for many fruitful discussions. I thank our system administrator Torge Storm for always fixing my machine and providing enough data space for all my sequences which was really a hard job. Of course I also would like to thank the other members of the group Jan Woetzel, Daniel Grest, Birger Streckel and Renate Staecker for their help, the discussions and providing the exciting working environment.

Last but not least, I would like to express my gratitude to my wife Miriam for always supporting me and my work. I also want to thank my sons Joshua and Noah for suffering under my paper writing. Finally I thank my parents for always supporting my education and my work.

Contents

1	Introduction	1
2	Basic concepts	5
2.1	Camera model	6
2.2	Plane at infinity π_∞	6
2.3	Absolute conic	7
2.4	Absolute dual quadric	8
2.5	Multiple-View geometry	9
2.5.1	Homography	9
2.5.2	Fundamental matrix	12
2.5.3	Essential matrix	15
2.6	Rotation data	16
2.7	Reconstruction	18
2.8	Self-Calibration using the absolute conic	22
2.8.1	Kruppa equations	23
2.9	Self-Calibration with the absolute dual quadric	24
2.10	Critical camera motions for self-calibration	25
3	Previous work on self-calibration	31
3.1	Self-Calibration of a rotating camera	31
3.1.1	Camera with constant intrinsics	32
3.1.2	Camera with some varying intrinsics	33
3.1.3	Camera self-calibration with some known extrinsics	36
3.1.4	Error analysis of calibration from purely rotating cameras	40
3.2	Self-Calibration from general motion	43
3.2.1	Camera with constant intrinsics	43
3.2.2	Camera with varying intrinsics	46
3.2.3	Multi-Camera system with constant inner orientation	47
4	Self-Calibration with known rotation	51
4.1	Rotating camera with known orientation	52
4.1.1	Self-Calibration with fully varying intrinsic parameters	52
4.1.2	Self-Calibration with constant intrinsic parameters	54

4.1.3	Relation to the general self-calibration	55
4.1.4	Evaluation for rotating cameras	57
4.1.5	Error in rotations about the coordinate axes	61
4.1.6	Influence of errors in rotation	67
4.1.7	Comparison to previous approaches	69
4.2	Freely moving camera with known orientation	72
4.2.1	Linear self-calibration with fully varying intrinsic parameters	72
4.2.2	Constraints needed to reach a linear calibration	73
4.2.3	Evaluation for freely moving cameras	74
4.2.4	Influence of errors in rotation	77
4.2.5	Relation to the Kruppa equations	80
4.2.6	Relation to the general self-calibration	80
4.3	Critical motions	83
4.3.1	No rotation at all or rotations about 180°	83
4.3.2	Rotation about one axis	84
4.3.3	Rotation about the x -axis	84
4.3.4	Rotation about the y -axis	85
4.3.5	Rotation about the z -axis	86
5	Experimental results	89
5.1	Results for a rotating camera	89
5.1.1	Algorithm for self-calibration with known orientation	89
5.1.2	Synthetic scene	90
5.1.3	Real image sequence with constant camera parameters	96
5.1.4	Real image sequence with varying parameters	99
5.2	Experimental results for a freely moving camera	102
5.2.1	Algorithm for self-calibration of a freely moving camera	102
5.2.2	Synthetic scene	103
5.2.3	Real image sequence	108
6	Conclusions	111
A	Projective geometry and camera model	113
A.1	Camera model	113
A.1.1	Pinhole camera model	114
A.2	Projective geometry	116
A.2.1	Projective space	117
A.2.2	Transformations in projective space	119
A.2.3	Two dimensional projective space \mathbb{P}^2	120
A.2.4	Conics	123
A.2.5	Three dimensional projective space \mathbb{P}^3	125
A.2.6	Quadric	126
A.2.7	From world to pixel coordinates	128

A.2.8	Projection matrix	129
A.2.9	Radial distortion	131
B	Purely Rotating camera	135
B.1	Self-calibration with varying intrinsics	135
B.2	Self-calibration equations for cameras with zero skew	137
B.3	Self-calibration with constant intrinsics	139
C	Freely moving camera	141
C.1	Freely moving camera	141
D	Sensitivity of rotation calibration	145
D.1	Sensitivity for a purely rotating camera	145
D.2	Sensitivity for a freely moving camera	147
E	Critical Rotations	149
E.1	No rotation of the camera	149
E.2	Rotation about the x -axis	151
E.3	Rotation about the y -axis	152
E.4	Rotation about the z -axis	153

Notation

For a better readability the notation used throughout the thesis will be summarized here. For matrices of any dimension capital letters will be used for example A . If the dimension of the matrix A is of interest then it will be denoted as subscript e.g. $A_{n \times m}$ for a $n \times m$ matrix. The element in row r and column c of the matrix is denoted as $(A)_{r,c}$.

Unless stated differently the indices j , i and k will be used for image indices. The 3×4 projection matrix of a metric camera will always be denoted as P . If it is a projective camera projection matrix it is denoted as P^p .

Further the following notations are used for the different geometric entities:

$\bar{x} = [x, y]^T$	point in the Euclidian space \mathbb{R}^2
$\bar{X} = [X, Y, Z]^T$	point in the Euclidian space \mathbb{R}^3
$\mathbf{x} = [x, y, w]^T$	point in the projective space \mathbb{P}^2
$\mathbf{X} = [X, Y, Z, W]^T$	point in the projective space \mathbb{P}^3
π	plane in projective three dimensional space \mathbb{P}^3
π_∞	plane at infinity
$H_{j,i}^\pi$	homography that maps a point from image j into image i over the plane π
$F_{j,i}$	Fundamental matrix that maps a point from image j into a line in image i
e_i	epipole in image i
K_i	camera calibration matrix of camera i
$R_{j,i}$	rotation between camera j and i
Ω	absolute conic
ω	image of the absolute conic
ω^*	dual image of the absolute conic
\cong	equality up to scale
$\ A\ _F$	Frobenius norm of the matrix A
ϕ	rotation angle of the rotation about the x -axis
Φ	rotation angle of the rotation about the y -axis
Θ	rotation angle of the rotation about the z -axis

Chapter 1

Introduction

One research topic of computer vision is to obtain a 3D-model of a scene from images of the scene. Image based reconstructions are capable of modeling even small details of the observed scene. Due to the achievements in computer graphics and the development of consumer computer graphics cards such reconstructions can be displayed even on a standard computer. This development created a need for automatic model generation for visualization purposes, that can be generated by the scene reconstruction techniques known in computer vision. These models often look more realistic than hand-crafted models used in the past because they exploit the texture information contained in the images. They allow the reconstruction of small details and exhibit a very natural visual appearance that can't be reached by hand-crafted models.

The first approaches for reconstruction from two images used cameras undergoing unknown motion with known intrinsic camera parameters like focal length and principal point were proposed by [LH81, TH84]. Hence the unknown camera motion has to be estimated simultaneously with the reconstruction of the scene. This estimation of scene geometry and camera motion is called **structure from motion**. Later Azerbayejani et al. in [AHP93], Spetsakis and Aloimonos in [SA90] and Szeliski and Kang in [SK93] proposed approaches to use image sequences and a full perspective camera model. A method for structure from motion using factorization and an affine camera model were proposed by Tomasi and Kanade [TK92]. In this case the major remaining problem is the correspondence problem. Given an image of a 3D scene point in one camera image the challenge is to find the image of the same 3D scene point in any other camera image of the scene. This estimation of correspondences is a research topic of its own.

In computer vision research was undertaken during the last decade to weaken the constraints of known camera parameters. This leads to the scene analysis from uncalibrated image sequences often called **uncalibrated structure from motion** and it is still an active research topic. The camera motion in this case can be determined only up to a projective transformation. The earlier work in this field was pioneered by Faugeras in [Fau92] and Hartley in [Har92] for two camera images. Beardsley et al. proposed an approach for reconstruction from multiple images in [BTZ96]. The reconstruction up to a projective transformation has the disadvantage that there is no meaningful measurement of angles and distances possible in the skewed projective space. Hence the knowledge of the projective transformation is required in order to transform the projective reconstruction into a metric reconstruction.

The estimation of this projective transformation from the images themselves or from the

projective reconstruction is the subject of **self-calibration**. It includes the calibration of the intrinsic camera parameters and the transformation from a projective reconstruction into an affine reconstruction by determining the position of the plane at infinity. The major progress in self-calibration occurred in the last thirteen years. The first contribution in this field was the work of Maybank and Faugeras in [MF92] when they proved that it is possible to estimate the constant intrinsic parameters of a camera that moves freely observing an unknown scene. In the following years several self-calibration techniques were proposed by Hartley in [Har94a], Maybank et al. in [FLM92], Heyden et al. in [HÅ96] and Triggs in [Tri97] that improved the performance of the achieved results and introduced new self-calibration concepts like the absolute quadric. In the late nineties researchers proposed techniques to overcome the limitation of constant intrinsic parameters for the camera, e.g. Heyden et al. in [HÅ97, HÅ98] and Pollefeys et al. in [PKG98, PKG99]. All these approaches use the image data alone and sometimes additional constraints for camera motion, constraints on scene structure respectively constraints on the camera calibration like a limited number of varying intrinsic parameters. Since the self-calibration approaches have to rely on the available image content they may suffer from degeneracy and uncertain data. Therefore the self-calibration approaches tend to be sensitive to noise. After the introduction of the first approaches for uncalibrated structure from motion it was realized that there are some critical motion sequences for the camera self-calibration which lead to errors during the reconstruction process.

Often the images are not the only source of information that is available. Fortunately, in many applications additional information like orientation is available. This is delivered by other sensors like the rotation information of the motor of surveillance cameras (video conferencing cameras) or arm positions of cameras mounted on robots. Cameras in cars use inertial sensors to measure orientation or TV studios often already measure the orientation of the camera with in-door tracking systems. Future cars will be equipped with fixed or even rotating or zooming cameras for driver assistance, where at least partial orientation and translation information is available. This additional information is not exploited to improve camera calibration by most of the self-calibration approaches mentioned above. There were only a few approaches proposed in the early nineties that incorporated the known camera orientation. Basu in [Bas93] simultaneously with Du and Brady in [DB93] introduced a self-calibration for a rotating camera in an active vision system. These approaches needed an explicit control of the camera motion. Stein introduced an approach that uses arbitrary camera motion to calibrate the intrinsic camera parameters in [Ste95]. All of these approaches that use the orientation assume to have a camera with constant intrinsic parameters that rotates about its optical center.

The self-calibration of the intrinsic camera parameters using available orientation data of the camera is the major topic of this thesis. The proposed approach does not require any active control of the camera nor constant intrinsic parameters. First of all an approach for the self-calibration of a camera rotating about its optical center with known orientation will be proposed. Additionally a self-calibration technique for a freely moving camera with known orientation will be introduced. Accordingly, the motion of the camera can be arbitrary due to the fact that an automatic model switching as proposed by Torr in [Tor97, Tor98] can choose the right self-calibration approach in dependence of the camera motion.

Furthermore the known orientation provides information that helps to detect critical motion sequences for reconstruction. Hence using the camera orientation improves the reliability of scene reconstruction from image sequences. The benefits of using the proposed self-calibration methods in conjunction with scene reconstruction has already been published in [FK03a]. Basically using the proposed self-calibration leads to a direct estimation of an Euclidian reconstruction without any intermediate projective reconstruction. This provides a more constrained and stable model generation process. The next section will summarize the main contributions of this work.

Main contributions of the thesis

The main contributions of this thesis can be summarized as follows:

- A self-calibration approach for rotating cameras with known orientation will be introduced. This approach does not require any constraints on the intrinsic camera parameters in contrast to all previously proposed approaches. Furthermore it is the only self-calibration that is linear in the intrinsic camera parameters. This approach was already published in [FK03b]. The extension of this approach using the statistical properties of the errors in order to get a more reliable self-calibration was published in [FK03c].
- For a rotating and freely moving camera with known orientation a self-calibration approach was introduced in [FK03b]. This technique is also linear in the intrinsic camera parameters. To achieve a linear solution one constraint on the intrinsic camera parameters is needed, e.g. one component of the principal point. This self-calibration can be done even if it is not possible to establish a consistent reconstruction for all involved cameras because it only uses pairwise image relations. An improvement of this technique for self-calibration without any constraint on the intrinsic camera parameters that further exploits the statistical properties of the errors was published in [FK03c].

Overview of the thesis

The basic concepts used throughout the thesis and the relations between two different images of a static scene are discussed in chapter 2. Furthermore it will introduce the different concepts of self-calibration and reconstruction. The general technique of calibrated structure from motion is introduced and afterwards the uncalibrated structure from motion is described. The absolute conic and the absolute quadric as frequently used entities for self-calibration are also discussed in chapter 2. The following chapter 3 will give a detailed overview of the previous work in self-calibration.

In chapter 4 the new linear calibration approaches for cameras with known orientation are introduced. After a detailed description of the new approaches the relation to the previous work is discussed. Furthermore the robustness of the proposed approaches is measured using synthetic data and analyzed analytically. Afterwards the critical camera rotations for self-calibration with the novel approaches are analyzed. In chapter 5 the new approaches are

evaluated using synthetic data and real image sequences with ground truth data. Chapter 6 will summarize the achievements of this work.

Chapter 2

Basic concepts

The main subject of this thesis is camera self-calibration by exploitation of the known orientation of the camera. In this chapter the basic concepts of multi-view geometry will be introduced as well as the basic geometric entities exploited by the various self-calibration techniques. A general knowledge about projective geometry, conics in the two dimensional projective space \mathbb{P}^2 , quadrics in the three dimensional projective space \mathbb{P}^3 and the camera model in these spaces is needed for the understanding of this chapter. These entities are explained in detail in appendix A.

After a short overview of the used notation for points, camera projection matrices and the camera calibration matrix in section 2.1 the sections 2.2 to 2.4 will introduce special entities of the projective space that are often exploited for self-calibration. Afterwards section 2.5 will investigate the relations between multiple images of the same static scene. There are two types of image relations between two different views of the same scene. These relations are the homography, explained in section 2.5.1, and the Fundamental matrix introduced in section 2.5.2. Depending on the camera motion and the scene structure one of these two relations is present between two different views.

The first major work on self-calibration (denoted as auto-calibration in [HZ03]) of a camera by simply observing an unknown scene was presented in [MF92, FLM92]. It was proven that self-calibration was theoretically and practically feasible for a camera moving through an unknown static scene with constant but unknown intrinsics. Since that time various methods have been developed. There are two major types of self-calibration methods. The first type of methods uses constraints on the image of the absolute conic. The second major type of methods determines the intrinsic camera parameters and the position of the plane at infinity by placing constraints on the dual absolute quadric. Section 2.7 will introduce the general problem of reconstructing a 3D scene from a sequence of images. It will show the limitations introduced for the reconstruction from uncalibrated cameras. The following sections 2.8 and 2.9 will introduce the two major types of self-calibration methods to overcome the limitations of the reconstruction from uncalibrated cameras. The last section 2.10 of this chapter will discuss the image sequences that are critical for self-calibration caused from specific camera motions during the image taking.

2.1 Camera model

This section will introduce the notation used to describe the camera projection and the intrinsic camera parameters. A more detailed explanation is given in appendix A.2.7 and appendix A.2.8.

The projection of a 3D scene point $\bar{X} \in \mathbb{R}^3$ in homogenous coordinates $\mathbf{X} = [X, Y, Z, W]^T \in \mathbb{P}^3$ onto an image by a camera may be modeled by the equation

$$\mathbf{x} = P\mathbf{X}. \quad (2.1)$$

where the image point \mathbf{x} in projective coordinates is $\mathbf{x} = [x, y, w]^T \in \mathbb{P}^2$ and P is the 3×4 camera projection matrix (see appendix A.2.8). The matrix P is a rank-3 matrix which is the camera projection matrix in metric spaces given by

$$P = K[R^T | -R^T\bar{C}], \quad (2.2)$$

and is called metric. The rotation matrix R and the vector $\bar{C} \in \mathbb{R}^3$ represent the Euclidian transformation between the camera and the world coordinate system. These parameters are called **extrinsic parameters** since they describe the camera pose in the world coordinate system. The **intrinsic parameters** of the camera are contained in the camera calibration matrix K which is an upper triangular matrix

$$K = \begin{bmatrix} f & s & u \\ 0 & a \cdot f & v \\ 0 & 0 & 1 \end{bmatrix}, \quad (2.3)$$

where f is the focal length of the camera expressed in pixel units. The aspect ratio a of the camera is the ratio between the size of a pixel in x-direction and the size of a pixel in y-direction. The principal point of the camera is $[u, v]^T$ and s is a skew parameter which models the angle between columns and rows of the CCD-sensor. More details about the camera calibration are given in appendix A.2.7.

In projective spaces \mathbb{P}^n two points \mathbf{x} and \mathbf{y} are equal if there exists a scale ρ such that

$$\mathbf{x} = \rho\mathbf{y} \text{ with } \rho \neq 0. \quad (2.4)$$

This is called equality up to scale and will be denoted by \cong throughout this thesis. The following sections will introduce some special geometric entities of the three dimensional projective space \mathbb{P}^3 that are often used for self-calibration.

2.2 Plane at infinity π_∞

The plane at infinity π_∞ is the extension of the three dimensional space \mathbb{P}^3 that is not contained in the Euclidian space \mathbb{R}^3 . As explained in appendix A.2.2 points on the plane at infinity π_∞ stay on the plane at infinity under affine point transformations in \mathbb{P}^3 . Please note that a point is moved on the plane at infinity by an affine transformation but it does not leave it. Since the plane at infinity π_∞ is described by the vector $[0, 0, 0, 1]^T \in \mathbb{P}^3$ it has the normal vector $n_\pi = [0, 0, 0]$ and its distance d_∞ to the origin is infinity as explained in appendix A.2.5.

Parallelism of planes can be defined with the plane at infinity similar to parallelism of lines in \mathbb{P}^2 which exploits the line at infinity (see appendix A.2.3).

Theorem 1 (Parallelism of planes) *Two planes are parallel if and only if the line of intersection is on the plane at infinity π_∞ .*

This can be proved similar to the proof of the theorem of the parallelism of lines given as theorem 9 in appendix A.2.3. Furthermore for a line parallel to a plane the following constraint is sufficient.

Theorem 2 (Line parallel to a plane) *A line and a plane are parallel if and only if the point of intersection between the line and the plane is on the plane at infinity π_∞ .*

With these two theorems it is easy to understand that affine transformations in \mathbb{P}^3 preserve parallelism of planes and likewise the parallelism of planes and lines because the intersection points remain on the plane at infinity π_∞ .

2.3 Absolute conic

The absolute conic Ω_∞ is a point conic that lies on the plane at infinity π_∞ . For the plane at infinity at position $\pi_\infty = [0, 0, 0, 1]^T$ (often called metric framework) the absolute conic is defined by the two following equations

$$\left. \begin{array}{l} X^2 + Y^2 + Z^2 \\ W^2 \end{array} \right\} = 0. \quad (2.5)$$

These equations hold for points $\mathbf{X} = [X, Y, Z, W]^T \in \mathbb{P}^3$ that are on the absolute conic Ω_∞ . This can be simplified for point $\mathbf{X} = [X, Y, Z, 0]^T \in \pi_\infty$ on the plane at infinity to

$$[X, Y, Z] I_{3 \times 3} [X, Y, Z]^T = 0. \quad (2.6)$$

Hence for all points on the plane at infinity in a metric framework the conic matrix C_∞ of the absolute conic is given as the identity matrix $I_{3 \times 3}$. Further from (2.6) it can be seen that the absolute conic does not contain real points on the plane at infinity. Conics that do not contain real points are called **proper virtual conics**. Proper virtual conics are non-degenerate conics that does not contain any real point on the defining plane.

A useful property of the absolute conic Ω_∞ is that it is a fixed conic under any similarity transformation. This is formalized by the next theorem.

Theorem 3 (Absolute conic is fixed under similarity transformations) *The absolute conic Ω_∞ is fixed under any projective transformation if and only if it is a similarity transformation. That means*

$$H^{-T} \Omega_\infty H^{-1} = \Omega_\infty$$

Proof: It is mandatory that the point transformation H is an affine transformation because only affine transformations leave the points on the plane at infinity π_∞ on the plane at infinity. For the remainder of the proof only the transformation of points on the plane at infinity is used. Then the conic matrix is $\Omega_\infty = I_{3 \times 3}$. Using the structure of an affine transformation (A.26) and the transformation rule for conics (A.22) leads to $A_{3,3}^{-T} I_{3 \times 3} A_{3,3}^{-1} = I_{3 \times 3}$. This requires an orthogonal matrix $A_{3,3}$ so the transformation has to be a similarity transformation. \square

This invariance under similarity transformations means that for a moving and rotating camera the image of the absolute conic only depends on the camera calibration. Hence this invariance to the motion of the camera provides an omnipresent calibration object which can be used to compute the intrinsic camera parameters. For example in the real world the moon is invariant against translation. E.g. if one moves on a straight street the moon is always at the same position. Hence the moon can be seen as a translation invariant object because its image position does not depend on the camera translation. There is no example for rotation invariant objects in the real world since it can't contain any real point as shown above. So far, it was shown that the image of the absolute conic in the camera is always at the same position because the translation and rotation of the camera doesn't matter for its image position.

2.4 Absolute dual quadric

In the previous section the absolute conic Ω_∞ was introduced. This section will focus on the dual of the absolute conic Ω_∞ . It is a degenerate quadric on the plane at infinity called the absolute dual quadric Q_∞^* . The absolute dual quadric Q_∞^* is formed by planes tangent to the absolute conic. So the rim of these planes forms the absolute conic Ω_∞ on the plane at infinity. The corresponding point quadric is called the absolute quadric Q_∞ and was first introduced by Semple and Kneebone in [SK52].

The quadric matrix of the absolute dual quadric Q_∞^* is in a metric framework given as

$$Q_\infty^* = \begin{bmatrix} I_{3 \times 3} & 0 \\ 0 & 0 \end{bmatrix}. \quad (2.7)$$

It can be easily seen that the conic has rank three. Hence the absolute dual quadric is a degenerate quadric. An absolute dual quadric transformed with a projective point transformation H and use of the transformation rule for dual quadrics (A.33) leads to

$$(Q_\infty^*)^p = H Q_\infty^* H^T = H \begin{bmatrix} I_{3 \times 3} & 0 \\ 0 & 0 \end{bmatrix} H^T. \quad (2.8)$$

From (2.7) and (2.8) it can be seen that the absolute dual quadric can be represented through a symmetric 4×4 matrix. According to the full rank of the projective transformation H , the transformed quadric $(Q_\infty^*)^p$ has still rank 3 and is still a symmetric matrix. A symmetric 4×4 matrix has generally ten independent parameters. The invariance to scale of the quadric matrices reduces the number of degrees of freedom by one. Hence for the absolute dual quadric there are eight degrees of freedom left due to the rank 3 constraint.

If the projective transformation H is a similarity transformation the absolute dual quadric Q_∞^* remains constant.

Theorem 4 (Absolute dual quadric is constant under similarity transformations)

If and only if the projective transformation H is a similarity transformation the absolute dual quadric Q_∞^ is fixed under the transformation H*

$$Q_\infty^* = HQ_\infty^*H^T.$$

The proof of this theorem is similar to the proof of theorem 3 about the fixed absolute conic under similarity transformations.

It can be seen from (2.7) that the plane at infinity is the null vector of the absolute dual quadric Q_∞^* . This can be verified by using the canonic form of the conic from (2.7) and the quadric equation (A.31).

The next section will investigate the relations between two consecutive images of a static scene.

2.5 Multiple-View geometry

This section will introduce the relation between multiple views of the same 3D scene. This relation depends on the camera motion. For a camera rotating about its optical center (purely rotating camera) a mapping between the two images can be computed from the camera orientations and the camera calibrations. For a point in the first camera the mapping gives the corresponding image point in the second camera. The details about this mapping process are given in section 2.5.1. For a freely moving camera that observes a general 3D scene more knowledge is required, otherwise only the projection of the projection ray into the second image can be determined. In order to get a point-to-point mapping, knowledge about the depth dependent shift (parallax) is needed. Section 2.5.2 explains this in detail.

2.5.1 Homography

For a purely rotating camera the optical center is fixed in space over time, because the projection ray of a scene point is given by the line that contains the optical center and the 3D scene point. Hence each projection ray is constant for a rotating camera. The corresponding points in two images are intersections of the same projection ray with a rotated plane as illustrated in figure 2.1(a).

This section introduces the homography as a projective transformation in \mathbb{P}^2 . It is a plane-to-plane mapping which means that the homography maps a point from one image plane onto the corresponding point in the second image plane. In general it is a bijective mapping between two planes. Interpreting the homography as a mapping between planes introduces a further application for homography mapping. The homography is a mapping between an arbitrary scene plane and the image plane. This plane-to-plane mapping does not limit the camera to a pure rotation. Hence the camera is allowed to move and rotate freely because all points in the first camera can be mapped over the scene plane into the second image. Algebraically this is simply the concatenation of the homographies between the first image and the scene plane and the scene plane and the second image. This is shown in figure 2.1(b).

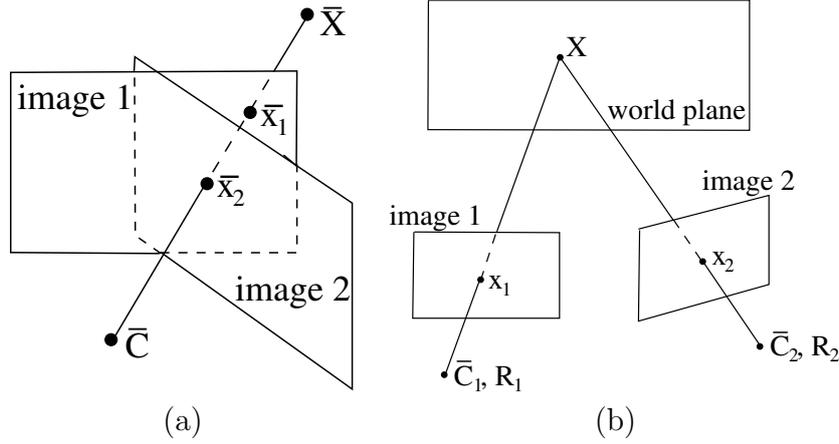


Figure 2.1: Mapping of point \bar{x}_1 from camera image 1 into point \bar{x}_2 in camera image 2. (a) Homography mapping for a rotating camera. (b) Mapping with the homography over a world plane.

To deduce the structure of the homography $H_{j,i}^\pi$, the 3D scene point \bar{X} on the world plane π is projected into both cameras j and i (shown in figure 2.1(b)). Without loss of generality the coordinate system of camera j is assumed to be identical with the world coordinate system which implies that $R_j = I_{3 \times 3}$ and $\bar{C}_j = 0_{3 \times 1}$. The second camera is at position \bar{C}_i and has orientation R_i . According to (2.2) the camera projection matrices are

$$P_j = K_j [I_{3 \times 3} | 0_{3 \times 1}] \text{ and } P_i = K_i [R_i^T | -R_i^T \bar{C}_i].$$

A 3D scene point $\bar{X} = [X, Y, Z]^T$ projects into the image point \mathbf{x}_j in the first image and the image point \mathbf{x}_i in the second image. Using the projection matrices P_j and P_i the image points \mathbf{x}_j and \mathbf{x}_i are given by

$$\mathbf{x}_j = K_j [X, Y, Z]^T \text{ and } \mathbf{x}_i = K_i R_i^T [X, Y, Z]^T - K_i R_i^T \bar{C}_i. \quad (2.9)$$

The transformation that maps the image point \mathbf{x}_j from the first image over plane π into the second image at point \mathbf{x}_i is the homography $H_{j,i}^\pi$:

$$\mathbf{x}_i \cong H_{j,i}^\pi \mathbf{x}_j. \quad (2.10)$$

The image position for a 3D scene point \bar{X} that lies on plane π is developed in dependence on the plane parameters n_π and d_π in the following. An image point $\mathbf{x}_j = [x, y, 1]^T$ in the first camera has the 3D Euclidian world coordinates

$$\bar{X}_j = K_j^{-1} [x, y, 1]^T = K_j^{-1} \mathbf{x}_j \in \mathbb{R}^3. \quad (2.11)$$

Together with the optical center $\bar{C}_j = 0_{3 \times 1}$ of the camera it defines the projection ray r_j

$$r_j(\rho) = [0_{3 \times 1}^T, 1]^T + \rho [\bar{X}_j^T, 0]^T \in \mathbb{P}^3$$

for the image point. The corresponding 3D scene point \mathbf{X} on the plane π of \mathbf{x}_j is the intersection point of the ray r_j and the plane π . For the point $\mathbf{X} \in \mathbb{P}^3$ on the plane π exists a ρ_j such

that $\mathbf{X} = r_j(\rho_j)$ fulfills the plane equation (A.24) in \mathbb{P}^3 . Furthermore for each point \bar{M} on the plane $n_\pi^T \bar{M} = -d_\pi$ holds, where n_π is the normal vector of the plane. Accordingly the scale ρ_j of \mathbf{X} is given as

$$\rho_j = -\frac{d_\pi}{n_\pi^T \bar{X}_j} \quad (2.12)$$

Consequently the intersection point \mathbf{X} on the plane π has the coordinates

$$\mathbf{X} = r_j(\rho_j) = \begin{bmatrix} d_\pi \bar{X}_j \\ -n_\pi^T \bar{X}_j \end{bmatrix}. \quad (2.13)$$

The projection of this point into the second camera with equation (2.9) delivers

$$\mathbf{x}_i = d_\pi K_i R_i^T \bar{X}_j + K_i R_i^T \bar{C}_i n_\pi^T \bar{X}_j,$$

with \bar{X}_j from (2.11) it delivers

$$\begin{aligned} \mathbf{x}_i &= d_\pi K_i R_i^T K_j^{-1} \mathbf{x}_j + K_i R_i^T \bar{C}_i n_\pi^T K_j^{-1} \mathbf{x}_j \\ \Leftrightarrow \mathbf{x}_i &= d_\pi K_i \left(R_{j,i} + \frac{R_{j,i} t_{j,i} n_\pi^T}{d_\pi} \right) K_j^{-1} \mathbf{x}_j, \end{aligned} \quad (2.14)$$

where $R_{j,i} = R_i^T R_j$ is the rotation between the cameras i and j . The translation between camera j and camera i is denoted as $t_{j,i} = \bar{C}_i - \bar{C}_j = \bar{C}_i - 0_{3 \times 1}$. According to (2.10) the homography $H_{j,i}^\pi$ is given as

$$H_{j,i}^\pi = K_i \left(R_{j,i} + \frac{R_{j,i} t_{j,i} n_\pi^T}{d_\pi} \right) K_j^{-1}, \quad (2.15)$$

The homography $H_{j,i}^\pi$ of a purely rotating camera is a special case of the homography in (2.15) because of the fixed camera center. Consequently the camera translation $t_{j,i}$ is zero, which results in the homography $H_{j,i}^\infty$

$$H_{j,i}^\infty = K_i R_{j,i} K_j^{-1} \quad (2.16)$$

for a purely rotating camera. The homography $H_{j,i}^\infty$ in (2.16) is called the infinite homography and maps over the plane at infinity.

Homography estimation from images The homography can be estimated from the images alone. Most estimation techniques use image correspondences given as point matches, optical flow etc. Using those correspondences the mapping (2.10) of a point from one image into another image is performed yielding a constraint for the homography. Due to the two degrees of freedom of a point in the two dimensional projective space \mathbb{P}^2 equation (2.10) provides two constraints for the homography. Any projective transformation in \mathbb{P}^2 has eight degrees of freedom as explained before. A counting argument leads to the conclusion that four independent point correspondences are needed to estimate the homography.

The constraint for a homography $H_{j,i}$ provided by a point correspondence $(\mathbf{x}_j^k, \mathbf{x}_i^k)$ between the point \mathbf{x}_j^k in camera j and a point \mathbf{x}_i^k in camera i can be written with (2.10) and (A.10) as

$$\mathbf{x}_i^k \times \left(H_{j,i} \mathbf{x}_j^k \right) = 0_{3 \times 1} \quad (2.17)$$

where \times is the cross product of two vectors in \mathbb{R}^3 . This can be written as

$$\begin{bmatrix} 0_{1 \times 3} & -w_i^k (\mathbf{x}_j^k)^T & y_i^k (\mathbf{x}_j^k)^T \\ w_i^k (\mathbf{x}_j^k)^T & 0_{1 \times 3} & -x_i^k (\mathbf{x}_j^k)^T \\ -y_i^k (\mathbf{x}_j^k)^T & x_i^k (\mathbf{x}_j^k)^T & 0_{1 \times 3} \end{bmatrix} \begin{bmatrix} (H_{j,i})_{1,\{1,2,3\}} \\ (H_{j,i})_{2,\{1,2,3\}} \\ (H_{j,i})_{3,\{1,2,3\}} \end{bmatrix} = 0_{3 \times 1}. \quad (2.18)$$

This equation system contains two linear independent equations as a consequence of the two degrees of freedom of a point in \mathbb{P}^2 . From four point correspondences a unique solution for the homography can be computed. For more than four point correspondences a least squares solution can be computed. If the data is noise free the estimated homography is the correct homography.

Noise in the input data introduces an estimation error. Accordingly the estimation result is not the exact homography. It is a general projective transformation that is close to the homography but does not fulfill the constraints given by (2.15). Due to the scale invariance of any projective transformation the estimated homography is arbitrary scaled with respect to the homography in (2.15). I.e. the estimated homography is equal up to scale to the homography in (2.15). According to the scale invariance of projective transformations the scale can't be determined.

So far it was introduced that the homography is a bijective point mapping between two planes. This mapping can be estimated using the images correspondences only. The homography depends on the extrinsic and intrinsic camera parameters as expressed in (2.15). The two cases for homography mapping of camera images are a purely rotating camera on the one hand and a freely moving camera that observes a planar scene on the other hand.

The next section will introduce the image relation between two images of a non planar 3D scene captured by a freely moving camera.

2.5.2 Fundamental matrix

In the previous section a bijective transformation between two images was introduced. For a freely moving camera that observes a general 3D scene the coordinates of an image point depend on the depth of the scene point, on the intrinsic camera parameters and on the extrinsic camera parameters. The original depth of the scene point can't be measured in one image. Hence it is not possible to compute a point to point mapping between two images.

However, for each point \mathbf{x}_j in an image j the projection ray can be determined. This projection ray can be projected into a second image i . The projection is a line l_i that corresponds to all possible depths of the 3D scene point. Therefore the corresponding point \mathbf{x}_i in the second image has to lie on this line l_i . This line l_i is called **epipolar line** and is illustrated in figure 2.2. To be more precise the epipolar line l_i is a line segment. It starts with depth zero and ends with the projection of the intersection point of the projection ray and the plane at infinity π_∞ into the second image. The point with depth zero is called the **epipole** \mathbf{e}_i , which is the projection of the camera center \bar{C}_j into the second camera i . Hence all epipolar lines contain the epipole since each projection ray starts at the camera center of the first camera j .

The two camera centers \bar{C}_j and \bar{C}_i and the 3D scene point \mathbf{X} define a plane, called the **epipolar plane**. The epipolar line l_i is the intersection of the epipolar plane and the image plane. This defines a corresponding epipolar line l_j in camera j as the intersection of the epipolar plane and the image plane of camera j . This relation between two images of a freely moving camera observing a general 3D-scenes is called **epipolar geometry**.

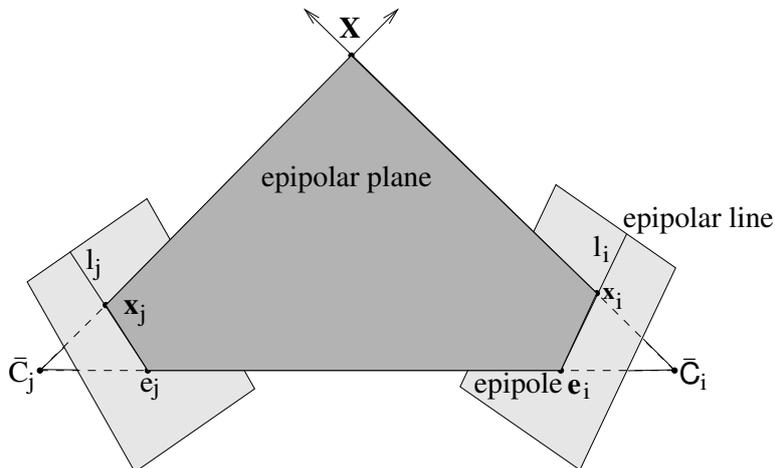


Figure 2.2: Epipolar geometry: A point in one image lies on the corresponding epipolar line in the other image. The epipolar line is the intersection of the epipolar plane with the image plane.

This point-to-line relation can be described by a linear function, which will be deduced in the following. As explained above, the epipole e_i is the projection of the camera center \bar{C}_j into camera i , and all epipolar lines l_i contain the epipole e_i . It is known that a line in \mathbb{P}^2 can be defined by two points (for details see appendix A.2.3). Therefore a second point is needed to define the epipolar line. This point can be constructed as the intersection of the projection ray and an arbitrary plane π in the 3D scene coordinate system. This plane can either be virtual or real. Once the plane π is fixed a homography $H_{j,i}^\pi$ can be used to map the point \mathbf{x}_j into camera i . The point $\mathbf{x}_i = H_{j,i}^\pi \mathbf{x}_j$ in the second image can now be used to define the epipolar line l_i . From the line equation (A.4) in \mathbb{P}^2 it follows

$$l_i = \mathbf{e}_i \times \mathbf{x}_i = (\mathbf{e}_i \times H_{j,i}^\pi) \mathbf{x}_j = F_{j,i} \mathbf{x}_j. \quad (2.19)$$

with

$$F_{j,i} = [\mathbf{e}_i]_\times H_{j,i}^\pi = [\mathbf{e}_i]_\times K_i \left(R_{j,i} + \frac{R_{j,i} (\bar{C}_i - \bar{C}_j) n_\pi^T}{d_\pi} \right) K_j^{-1} \quad (2.20)$$

where $[\cdot]_\times$ is the cross product matrix¹. The matrix $F_{j,i}$ is called the **Fundamental matrix**. It maps the point \mathbf{x}_j in image j onto a line l_i in image i . The choice of the mapping plane π

¹The cross product matrix is the matrix form for the column-wise cross product of a vector $a = [a_1, a_2, a_3] \in \mathbb{R}^3$ with any matrix $B \in \mathbb{R}^{3 \times \cdot}$.

$$[\cdot]_\times = \begin{bmatrix} 0 & -a_3 & a_2 \\ a_3 & 0 & -a_1 \\ -a_2 & a_1 & 0 \end{bmatrix}.$$

does not effect the resulting Fundamental matrix $F_{j,i}$:

$$F_{j,i} = [\mathbf{e}_i]_{\times} H_{j,i}^{\pi_o} = [\mathbf{e}_i]_{\times} H_{j,i}^{\pi_u}$$

for two different planes π_o and π_u . This will be proved in the following. The homography $H_{j,i}^{\pi_o}$ has the structure from (2.15) and the Fundamental matrix $F_{j,i}$ with the homography $H_{j,i}^{\pi_o}$ is then given by

$$F_{j,i} = [\mathbf{e}_i]_{\times} \left(\frac{1}{\rho_{j,i}} K_i \left(R_{j,i} + \frac{R_{j,i}(\bar{C}_i - \bar{C}_j)n_{\pi_o}^T}{d_{\pi_o}} \right) K_j^{-1} \right). \quad (2.21)$$

Without loss of generality the center \bar{C}_j of camera j is the origin of the coordinate system. Then from (2.1) it follows that the epipole \mathbf{e}_i is given as $\mathbf{e}_i = K_i R_{j,i} \bar{C}_i$. With that the Fundamental matrix from (2.21) is

$$F_{j,i} = \frac{1}{\rho_{j,i}} \left([\mathbf{e}_i]_{\times} K_i R_{j,i} K_j^{-1} + \frac{[\mathbf{e}_i]_{\times} \mathbf{e}_i n_{\pi_o}^T K_j^{-1}}{d_{\pi_o}} \right). \quad (2.22)$$

As the nullspace of the cross product matrix $[\mathbf{e}_i]_{\times}$ is \mathbf{e}_i the Fundamental $F_{j,i}$ matrix is

$$F_{j,i} = \frac{1}{\rho_{j,i}} [\mathbf{e}_i]_{\times} K_i R_{j,i} K_j^{-1}. \quad (2.23)$$

It follows that the Fundamental matrix $F_{j,i}$ is independent from the mapping plane of the homography because the mapping plane is not contained in the fundamental matrix. This leads to the following theorem

Theorem 5 (Decomposition of the Fundamental matrix) *Given the Fundamental matrix $F_{j,i}$ between two camera images j and i the three parameter family of decomposition of the Fundamental matrix is*

$$F_{j,i} = [\mathbf{e}_i]_{\times} H_{j,i}^{\pi} \text{ with } H_{j,i}^{\pi} = K_i R_{j,i} K_j^{-1} + \mathbf{e}_i v^T \text{ and } v^T = \frac{n_{\pi}^T}{d_{\pi}} K_j^{-1}, \quad (2.24)$$

where v is a vector from \mathbb{R}^3 that defines the mapping plane π for $H_{j,i}^{\pi}$.

The projection to the image point \mathbf{x}_i of the 3D scene point \mathbf{X} into camera i is always on the line l_i . It can be expressed algebraically with the line equation (A.11) as

$$\mathbf{x}_i^T l_i = \mathbf{x}_i^T F_{j,i} \mathbf{x}_j = 0. \quad (2.25)$$

The constraint is known as the **epipolar constraint**. From the epipolar constraint it can be seen that the Fundamental matrix has rank 2 because the epipole is the right null space of the Fundamental matrix $F_{j,i}$. This results from the nullspace of the cross product matrix $[\cdot]_{\times}$. The Fundamental matrix has seven degrees of freedom because a general 3×3 matrix has nine degrees of freedom; one is removed by the scale invariance and another by the rank 2 constraint.

The mapping of points from camera i into lines in camera j is given by the transposed Fundamental matrix $F_{i,j} = F_{j,i}^T$ and can be deduced from the epipolar constraint (2.25) by

$$(\mathbf{x}_i^T F_{j,i}) \mathbf{x}_j = (F_{j,i}^T \mathbf{x}_i)^T \mathbf{x}_j = l_j^T \mathbf{x}_j.$$

The epipole \mathbf{e}_j in camera j is the left nullspace of the Fundamental matrix $F_{j,i}$. After introducing the basic properties of the Fundamental matrix the computation of the Fundamental matrix will be described in the next paragraph.

Estimation of the Fundamental matrix The epipolar constraint (2.25) provides one linear constraint for the Fundamental matrix $F_{j,i}$ from a point correspondence in the two images. Since the Fundamental matrix has seven degrees of freedom seven independent correspondences are sufficient to compute the Fundamental matrix. From the epipolar constraint (2.25) the following linear equation can be extracted for a correspondence $(\mathbf{x}_j, \mathbf{x}_i)$

$$\underbrace{[x_i \mathbf{x}_j^T \quad y_i \mathbf{x}_j^T \quad w_i \mathbf{x}_j^T]}_{A_f} \mathbf{f}_{9 \times 1} = 0. \quad (2.26)$$

Hereby $f = [(F_{j,i})_{1,1}, (F_{j,i})_{1,2}, \dots, (F_{j,i})_{3,3}]$ is the vector that contains the entries of $F_{j,i}$ and the point in camera i is given by $\mathbf{x}_i = [x_i, y_i, w_i]^T$. For seven point correspondences a matrix A_{seven} that contains the A_f for the correspondences as rows can be constructed. The Fundamental matrix is then computed as the nullspace of A_{seven} . The 7×9 matrix A_{seven} has rank seven. Then the nullspace is two dimensional and every matrix $F = \lambda \mathbf{F}_1 + (1 - \lambda) \mathbf{F}_2$ is a solution of the linear equation system, where \mathbf{F}_1 and \mathbf{F}_2 are the Fundamental matrices corresponding to the two null-vectors of A_{seven} . Exploiting the rank 2 and the resulting condition $\det(F) = \det(\lambda \mathbf{F}_1 + (1 - \lambda) \mathbf{F}_2) = 0$ leads to the characteristic polynomial. The solution for the correct Fundamental matrix is therefore the root of the polynomial.

The next section will introduce the Essential matrix as a specialization of the Fundamental matrix. The Essential matrix exploits a known camera calibration to add more constraints for the point-line mapping.

2.5.3 Essential matrix

The Essential matrix $E_{j,i}$ is a Fundamental matrix for normalized coordinates and has the same properties as the above introduced Fundamental matrix. It maps normalized image points from image j to a line in image i . For normalized image points the calibration matrices K_i and K_j are the identity matrices. Then from (2.23) it follows that the Essential matrix $E_{j,i}$ is

$$E_{j,i} = [e_i]_{\times} R_{j,i} = R_{j,i}^T [R_{j,i} t_{j,i}]_{\times}. \quad (2.27)$$

It follows that the Essential matrix has only five degrees of freedom. The rotation $R_{j,i}$ and the translation $t_{j,i}$ have both three degrees of freedom. The overall scale ambiguity reduces the degrees of freedom by one. In addition to the Fundamental matrix the eigenvalues of the Essential matrix have to fulfill the following constraint:

Theorem 6 (Eigenvalues of the Essential matrix) *A matrix in $\mathbb{R}^{3 \times 3}$ is an Essential matrix if and only if its singular values are one double singular value and the third singular value is zero.*

A proof for this constraint is given in [HZ03].

After introducing the relations between multiple views of a 3D scene the next section will discuss the basic properties and representations of rotations.

2.6 Rotation data

The subject of this thesis is the camera calibration from known orientation. The different representations of orientation data will be introduced in this section. Some comparisons of the different representations will also be given. In the following it will not be distinguished between orientation and rotation.

According to Eulers rotation theorem every rotation can be described by a rotation about the three axes x , y and z of the coordinate system. Hence the camera rotation can always be described by a rotation about the three axes x , y and z . Then the rotation is an isometric transformation in \mathbb{R}^3 and it can be interpreted as projective transformation in \mathbb{P}^2 . In the following the rotation is usually used as a projective transformation that maps points of \mathbb{P}^2 .

A parameterization is called **fair parameterization** if it does not introduce additional numerical sensitivities than inherent in the problem itself [HT99]. This measurement for the parameterization is used here to measure the quality of the parameterization (representation) of the rotation.

Euler angles This is probably the best known parameterization of a rotation. Every rotation R can be described as a concatenation of rotations about the axes x , y and z . The rotation R_x , R_y and R_z for a right handed coordinate systems are given by

$$R_x(\phi) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \phi & -\sin \phi \\ 0 & \sin \phi & \cos \phi \end{bmatrix}, R_y(\Phi) = \begin{bmatrix} \cos \Phi & 0 & \sin \Phi \\ 0 & 1 & 0 \\ -\sin \Phi & 0 & \cos \Phi \end{bmatrix},$$

$$R_z(\Theta) = \begin{bmatrix} \cos \Theta & -\sin \Theta & 0 \\ \sin \Theta & \cos \Theta & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

where R_x , R_y and R_z are the matrices for rotation about the axes x , y and z . Then the rotation R can be written as product of these three rotation matrices.

Due to the fact that the matrix multiplication is not commutative one has to choose the "right" order of the matrices to get the wanted rotation matrix R . The non commutativity here reflects the fact that the result of a sequence of rotations depends on the order of the rotations. This disadvantage of the representation of the rotation makes it difficult to use this representation as a sequence of rotations about the axis x , y and z . A more detailed overview of the Euler angles can be found in [WW92]. The Euler angles are not a fair parameterization of the rotation.

Summarizing it was seen that it is possible to represent an arbitrary rotation as a concatenation of rotations about the axes of the coordinate system and the resulting parameterization is not fair.

Eigenvalues and eigenvectors of rotation matrices Each rotation in \mathbb{R}^3 can be represented as an orthonormal² 3×3 matrix. For the orthonormal matrices the computation of the

² A is an orthonormal matrix if and only if $A^{-1} = A^T$. This shows that the transposed is the inverse matrix. The orthonormal matrices are a group under multiplication. I.e. for two orthonormal matrices A and B the

eigenvalues and eigenvectors is easier than in the general case [Wil65]. For a rotation matrix it is known that the eigenvalue structure is:

$$\text{eigval}(R) = [1, e^{i\phi}, e^{-i\phi}], \quad (2.28)$$

where $\phi \in [-\pi, \pi]$ [Pea83, Fra83]. According the identity matrix $I_{3 \times 3}$, that can be interpreted as a zero rotation with $\phi = 0$, then (2.28) leads to the triple eigenvalue 1. Once the eigenvalues are computed the eigenvector can be calculated as the solution of the equation system

$$(R - \lambda I)x = 0, \quad (2.29)$$

where λ is an eigenvalue and x is the corresponding eigenvector. It was shown that the rotation matrix always has one eigenvalue one and two complex eigenvalues that are conjugated.

Axis angle representation In general every rotation can be described by a rotation about an axis \bar{r} by an angle ϕ . This representation gives a description of the rotation that is unique up to a scale by one or minus one. I.e. the axis and angle can be scaled by -1 without changing the transformation. This is easy to see because a scale of the rotation axis by minus one inverts the direction of the rotation. Therefore the rotation angle has to be change its sign to describe the same rotation as before.

The rotation axis \bar{r} of a rotation has to be invariant under this rotation. That means it is an eigenvector corresponding to the eigenvalue 1. In (2.28) it was introduced that every rotation R has an eigenvalue one. So the corresponding eigenvector is the rotation axis. The angle of the rotation about the axis \bar{r} is contained in the complex eigenvalues

$$\lambda = e^{i\phi} = \cos \phi - i \sin \phi$$

where ϕ is the angle of rotation about the axis \bar{r} [Fra83]. The rotation matrix $R(r, \phi)$ for a given axis $\bar{r} = [x, y, z]$ with $\|\bar{r}\| = 1$ and an rotation angle ϕ is given by [Ale98]

$$R(r, \phi) = \begin{bmatrix} x^2 + (1 - x^2) \cos \phi & xy(1 - \cos \phi) - z \sin \phi & xz(1 - \cos \phi) + y \sin \phi \\ xy(1 - \cos \phi) + z \sin \phi & y^2 + (1 - y^2) \cos \phi & yz(1 - \cos \phi) - x \sin \phi \\ xz(1 - \cos \phi) - y \sin \phi & yz(1 - \cos \phi) + x \sin \phi & z^2 + (1 - z^2) \cos \phi \end{bmatrix}. \quad (2.30)$$

From (2.30) it follows that the rotation is ambiguous to a scale by -1 of the angle ϕ and the axis \bar{r} . To get the axis \bar{r} and the angle ϕ from a given rotation matrix R it is not needed to compute all eigenvalues. The trace of R is according to (2.30) given by

$$\begin{aligned} \text{trace}(R) &= x^2 + (1 - x^2) \cos \phi + y^2 + (1 - y^2) \cos \phi + y^2 + (1 - y^2) \cos \phi \\ &= 1 + 2\cos \phi, \end{aligned}$$

with $\|\bar{r}\| = 1$. The angle ϕ can be computed as

$$\cos \phi = \frac{1}{2} (\text{trace}(R) - 1). \quad (2.31)$$

product AB is also orthonormal.

The sign of the angle ϕ can't be determined with (2.31) and is chosen without loss of generality as a positive angle. Then the rotation axis \bar{r} has to be computed properly. With (2.30) it is

$$\begin{aligned}(R)_{3,2} - (R)_{2,3} &= 2x \sin \phi \\ (R)_{1,3} - (R)_{3,1} &= 2y \sin \phi \\ (R)_{2,1} - (R)_{1,2} &= 2z \sin \phi.\end{aligned}$$

The correct axis \bar{r} is

$$\bar{r} = \frac{1}{2 \sin \phi} \begin{bmatrix} (R)_{3,2} - (R)_{2,3} \\ (R)_{1,3} - (R)_{3,1} \\ (R)_{2,1} - (R)_{1,2} \end{bmatrix}. \quad (2.32)$$

For $\phi = 0$ the rotation axis can be chosen arbitrarily because the resulting rotation is the identity.

It was discussed that the representation of a rotation through an axis and an angle provides a unique representation. The ambiguity of a scale of -1 does not effect the representation because the resulting rotation matrix is the same. Further in [SN01] it was shown that the axis angle representation is a fair parameterization.

Quaternions Quaternions are another fair parameterization of the rotation. The quaternions parameterize the rotation as a vector of length four. The quaternions are linked to the axis angle representation directly by

$$h = \left[\cos \frac{\phi}{2}, \sin \frac{\phi}{2} \bar{r}^T \right]^T, \quad (2.33)$$

where h is the quaternion that describes the same rotation as $R(\bar{r}, \phi)$ in (2.30). In [SN01] Schmidt and Niemann showed that the quaternions are in general comparable to the axis angle parameterization. For details of the quaternions it is referred to [Fau93]. More different rotations and comparisons of the different representations are given in [Kui99, Stu64, Shu93, Sho85, Gol80]

In the following a rotation about at least the x -axis and the y -axis of the camera coordinate system is denoted as a general rotation. We don't assume a rotation about the optical axis (equal to the z -axis) of the camera coordinate system because this special rotation is only a rotation inside the image plane and therefore it gives no information for the focal length and the aspect ratio of the camera self-calibration.

The previous sections introduced the relations between two different images of a static scene in dependence of the camera motion as well as the various representations of orientation. In the following section the principles of scene reconstruction from image sequences are explained.

2.7 Reconstruction

Over the last fifteen years, a lot of techniques were developed to reconstruct a scene only using the image information [MF92, FLM92, PKG98, HÅ96, HÅ97, Bou98, RC02]. This section

describes the general problem of reconstruction of a 3D scene given a sequence of images of an uncalibrated camera, often called uncalibrated structure from motion. To perform the reconstruction, the 3D points have to be computed from their projections into the images. The projection of the 3D scene point \mathbf{X} is given by equation (2.1)

$$\mathbf{x} \cong P\mathbf{X}$$

with the camera projection matrix P . The projection equation is not invertible because the projection matrix P itself is not invertible. The reconstruction algorithms solve the inversion of the projection by exploiting multiple images of the scene. Most reconstruction algorithms have the following common structure.

1. Estimation of point correspondences between the images. Standard matching and tracking techniques are used here [LK81, HS88].
2. Computation of the Fundamental matrix from point correspondences between the images. In this step the estimation techniques introduced in section 2.5.2 are applied.
3. Computation of an initial set of projectively skewed camera projection matrices exploiting the estimated Fundamental matrices.
4. Computation of the skewed 3D scene points from the projection matrices of step 3 and the point correspondences in the images of step 1 with triangulation. Computation of the skewed projection matrices of all cameras.
5. To reach a metric reconstruction a self-calibration of the estimated cameras has to be computed and the projective skew of the scene is obtained. Afterwards the scene is transformed into a meaningful Euclidian scene [Tri97, PKG98, HZ03].
6. Global optimization of the previously estimated cameras and scene points for example with bundle adjustment [TMHF00].

The step 5 is sometimes not performed in the existing approaches or other optimization schemes performed during step 4. The topic of this section is mainly step 3 which computes an initial pair of camera projection matrices P_1 and P_2 for the first two images of the image sequence. Afterwards the projection matrices P_i with $i > 2$ for all other images of the sequence are computed according to this first camera pair. The computation of the projection matrices P_i with $i > 2$ will be discussed at the end of this section.

Reconstruction from image sequences of uncalibrated cameras occurs for example for reconstructions from image sequences without any additional information. I.e. the external and internal calibration of the cameras are unknown. So the only source of information is the image point in pixel coordinates as projection of the 3D scene point. Without the camera calibration it is not possible to compute the projection ray of an image point in the real world which is required for triangulation. Assuming an identity camera calibration matrix, the ray corresponding to the pixel position can be determined together with the optical center of the camera. According to the ignored camera calibration the ray is skewed with the internal camera

calibration of the camera. However the skewed rays of two or more cameras can be used to triangulate a 3D scene point. Consequently the whole achieved reconstruction will be skewed with a projective transformation.

The Fundamental matrix $F_{1,2}$ provides a relation between the image planes of two uncalibrated cameras as shown in section 2.5.2. It depends on the relative translation $\Delta\bar{C}_{1,2}$ and the relative orientation $\Delta R_{1,2}$ between the cameras. Moreover it depends on the intrinsic camera parameters of the involved cameras. When trying to find two consistent camera projection matrices given a Fundamental matrix, without loss of generality the first camera's coordinate system is assumed to be identical with the world coordinate system. Then the cameras can be chosen as [HZ03]

$$P_1^p = [I_{3 \times 3} | 0] \text{ and } P_2^p = \left[[\mathbf{e}_2]_x F_{1,2} + \mathbf{e}_2 \mathbf{a}_2^T \mid \lambda \mathbf{e}_2 \right] \quad (2.34)$$

where \mathbf{e}_2 is the epipole in the second image, $\mathbf{a}_2 \in \mathbb{P}^2$ is an arbitrary vector and $\lambda \in \mathbb{R} \setminus \{0\}$ a non-zero scale factor for the length of the camera translation. Please note, that the vector \mathbf{a}_2 represents the ambiguity of the mapping plane of the homography contained in the first three rows and columns of the projection matrix. This ambiguity is a result of the ambiguous decomposition of the Fundamental matrix as explained in theorem 5.

The skewed projection matrices P_1^p and P_2^p from equation (2.34) can be used to triangulate the 3D points $\tilde{\mathbf{X}}$ from the image correspondences ([HZ03] page 318). These 3D scene points $\tilde{\mathbf{X}}$ are called projectively skewed because the triangulation used a set of skewed rays. The projective skew depends on the internal camera parameters and on a_2 .

So far an initial set of camera projection matrices P_1^p and P_2^p was determined and afterwards a projectively skewed scene $\tilde{\mathbf{X}}$ was triangulated. The cameras are ambiguous in the mapping plane of the homography and the scale of the camera translation vector as mentioned above. In the following the transformation between the projectively skewed cameras and the corresponding metric cameras will be investigated.

It is known for a given Fundamental matrix F that for any two camera pairs (P_1^p, P_2^p) and $((P_1^p)', (P_2^p)')$ which have the same Fundamental matrix F , a non-singular transformation exists such that [HZ03]

$$P_i = P_i^p H = [M_i \mid \lambda e_i] H \quad (2.35)$$

with $H \in \mathbb{R}^{4 \times 4}$. Accordingly there is a transformation $H_{4 \times 4}$ which transforms P_1^p and P_2^p from equation (2.34) into the corresponding metric cameras $P_1 = [K_1 \mid 0_{3 \times 1}]$ and $P_2 = [K_2 R_2^T \mid -K_2 R_2^T C_2]$. From the ambiguous scale λ of the reconstruction in (2.34) it follows that the entry $(H)_{4,4}$ in $H_{4 \times 4}$ can be set to one. From the transformation of P_1^p into P_1

$$[K_1 \mid 0_{3 \times 1}] = [I_{3 \times 3} \mid 0_{3 \times 1}] H_{4 \times 4} = \begin{bmatrix} (H)_{1,1} & (H)_{1,2} & (H)_{1,3} & (H)_{1,4} \\ (H)_{2,1} & (H)_{2,2} & (H)_{2,3} & (H)_{2,4} \\ (H)_{3,1} & (H)_{3,2} & (H)_{3,3} & (H)_{3,4} \end{bmatrix}, \quad (2.36)$$

where $(H)_{m,n}$ is the element in row m and column n of the matrix $H_{4 \times 4}$. It follows that $H_{4 \times 4}$ is of the form

$$H_{4 \times 4} = \begin{bmatrix} K_1 & 0_{3 \times 1} \\ \tilde{v}^T & 1 \end{bmatrix}. \quad (2.37)$$

Looking at the projection equation (2.1) for the projectively skewed 3D scene points

$$\mathbf{x} \cong P^p \tilde{\mathbf{X}} \cong P^p H_{4 \times 4} H_{4 \times 4}^{-1} \tilde{\mathbf{X}} = P \mathbf{X}, \quad (2.38)$$

leads to the observation that $H_{4 \times 4}^{-1}$ can be used to map the projectively skewed scene $\tilde{\mathbf{X}}$ into the corresponding metric scene \mathbf{X} often denoted as a metric frame. The mapping has to ensure that the plane at infinity $\tilde{\pi}_\infty$ is mapped to $\pi_\infty = [0, 0, 0, 1]$. Looking at the position of the plane at infinity in the projectively skewed reconstruction gives

$$\tilde{\pi}_\infty = H_{4 \times 4}^{-T} \pi_\infty = \begin{bmatrix} K_1^{-T} & -K_1^{-T} \tilde{v} \\ 0_{1 \times 3} & 1 \end{bmatrix} \pi_\infty = \begin{bmatrix} -K_1^{-T} \tilde{v} \\ 1 \end{bmatrix}. \quad (2.39)$$

So the vector $[K_1^{-T} \tilde{v}, 1]$ encodes the plane at infinity in the computed projective reconstruction. It follows that the projective skew can be written as

$$H_{4 \times 4} = \begin{bmatrix} K_1 & 0_{3 \times 1} \\ v^T K_1^T & 1 \end{bmatrix}, \quad (2.40)$$

where $v = -K_1^{-1} \tilde{v}$ is the normal of the plane at infinity in the projective reconstruction. The normal vector v depends on the chosen \mathbf{a}_2 in equation (2.34) as well as on the camera motion. If the vector v is the zero vector then the plane π_∞ is still at its canonical position and the reconstruction $\tilde{\mathbf{X}}$ is called an affine reconstruction.

The above described method provides an initial set of cameras for the first two images. The remainder of this section will discuss how the camera projection matrices for all other images of the image sequence can be computed. In the following two different types of methods for the computation of these camera projection matrices will be discussed.

The most common method introduced by Beardsley et al. [BTZ96] is to use the 3D scene points \mathbf{X} respectively $\tilde{\mathbf{X}}$ to compute all other cameras using (2.1). Beardsley used the computed scene points as a known calibration pattern for all further cameras. Once a new camera is estimated the new points are triangulated and used to estimate the position of the next camera. This method can be used for all above mentioned cases of known and unknown extrinsic and intrinsic camera parameters. Please note that the projective skew $H_{4 \times 4}$ is constant for all cameras used to reconstruct the projectively skewed scene points $\tilde{\mathbf{X}}$, because the projectively skewed 3D points $\tilde{\mathbf{X}}$ are used to estimate the new camera projection matrices. Therefore (2.38) is valid for all projection matrices of the image sequence. This fact will later be used to self-calibrate the cameras and it will be referred to as a globally consistent frame.

Another approach introduced by Avidan and Sashua in [AS01] exploited the Trifocal-tensor to compute the camera projection matrices without any 3D scene information (for details about the Trifocal tensor see [HZ03, Moo98]). The technique uses the properties of the Trifocal-tensor to estimate a decomposition of the pairwise Fundamental matrices with a common mapping plane for all homographies from the decompositions. Hence the arbitrary mapping plane is fixed for all homographies. This decomposition can be used to compute the camera projection matrices of the remaining images of the sequence. The technique further guarantees that these projection matrices are all projectively skewed with the same projective skew $H_{4 \times 4}$ without using any 3D scene information.

This section gave an overview about the different situations for reconstruction. The computed scene points \mathbf{X} respectively $\tilde{\mathbf{X}}$ were characterized and the structure of the estimated camera projection matrices was denoted.

The next section will introduce the self-calibration using constraints on the image of the absolute conic to determine the projective skew matrix $H_{4 \times 4}$.

2.8 Self-Calibration using the absolute conic

In this section the absolute conic is introduced as a geometric property that can be used for self-calibration. The self-calibration uses the fact that the absolute conic is invariant under camera motion as explained in section 2.3. Self-calibration addresses the case of reconstruction from unknown intrinsic camera parameters. From the previous section it is known that in general the estimated camera projection matrices are projectively skewed with a homography $H_{4 \times 4}$ from (2.40). It depends on the internal parameters of the first camera K_1 and the normal $[v^T, 1]^T$ of the plane at infinity. The subject of self-calibration is to determine these unknowns which is equal to estimate $H_{4 \times 4}$. The input of each self-calibration is either a sequence of projectively skewed camera projection matrices or a sequence of images.

The Euclidian cameras P_i corresponding to the cameras P_i^p of the projective reconstruction from (2.38) which are estimated from the Fundamental matrix with (2.34) are given by $P_i = [M_i | \mathbf{e}_i] H_{4 \times 4} = P_i^p H_{4 \times 4}$. Looking at the first three rows and columns of the projection matrices leads to the translation independent expression

$$K_i R_i^T \cong (M_i + \mathbf{e}_i \mathbf{v}^T) K_1, \quad (2.41)$$

Rearranging equation (2.41) leads to

$$R_i^T \cong K_i^{-1} (M_i + \mathbf{e}_i \mathbf{v}^T) K_1.$$

This formulation of the general self-calibration problem is transformed to a rotation independent equation using the property $RR^T = I$ of the rotation matrix into

$$K_i K_i^T \cong (M_i + \mathbf{e}_i \mathbf{v}^T) K_1 K_1^T (M_i + \mathbf{e}_i \mathbf{v}^T)^T.$$

Then the basic equations for self-calibration are given by

$$\omega_i^* \cong (M_i + \mathbf{e}_i \mathbf{v}^T) \omega_1^* (M_i + \mathbf{e}_i \mathbf{v}^T)^T, \quad (2.42)$$

where ω^* is the dual image of the absolute conic (DIAC)

$$\omega_l^* = K_l K_l^T. \quad (2.43)$$

I.e. ω_l^* is the dual (line) conic mapped from the plane at infinity into the image l . It can be deduced by applying the transformation rule of conics (A.22) to the absolute conic Ω_∞ from equation (2.6) on the plane at infinity. The homography H_l^∞ that maps points on the plane at infinity into the image of camera l is given by $H_l^\infty = K_l R_l^T$, where K_l is the camera calibration

matrix of camera l and R_l is the rotation of camera l . Accordingly the mapping of the absolute conic Ω_∞ from equation (2.6) into image l is given by

$$\omega_l = (H_l^\infty)^{-T} \Omega_\infty (H_l^\infty)^{-1} = K_l^{-T} R_l^T I_{3 \times 3} R_l K_l^{-1} = K_l^{-T} K_l^{-1},$$

which is called the image of the absolute conic (IAC). The dual image of a non-degenerate conic is given by the inverse of the conic matrix (see appendix A.2.4). Consequently the DIAC ω_l^* is the inverse of the IAC ω_l which is equivalent to (2.43).

Equation (2.42) relates the unknown entries of the dual image of the absolute conic (DIAC) with the known entries of the projectively skewed camera projection matrices. Further it can be seen in (2.42) that the DIAC is a symmetric 3×3 matrix which has six degrees of freedom. Due to the overall scale ambiguity equation (2.42) provides five linear independent equations. The dual image of the absolute conic (DIAC) (see section 2.3) is more attractive for self-calibration because it directly contains the intrinsic parameters of the cameras.

From equation (2.42) it can be seen that there are only eight unknown parameters left in (2.43), five intrinsic parameters in the first camera and three parameters in \mathbf{v} . All other parameters are provided by the known camera projection matrices. Different proposed self-calibration methods introduced different constraints for ω_1^* from the constraints on the camera calibrations K_i with $i = 2, \dots, n$.

In (2.42) the homography $(M_i - \mathbf{e}_i \mathbf{v}^T)$ is used to map the DIAC from camera 1 into camera i . If this homography is the homography H_i^∞ that maps over the plane at infinity, then equation (2.42) leads to the so called infinite homography constraint

$$\omega_i^* \cong H_i^\infty \omega_1^* (H_i^\infty)^T. \quad (2.44)$$

If $\det(H_i^\infty) = 1$ is chosen, then the unknown scale factor is canceled out and (2.44) is Euclidian equality.

2.8.1 Kruppa equations

The Kruppa equations [Kru13] are chronological the first proposed equations for self-calibration. Therefore this calibration approach is shortly discussed here. This self-calibration approach was introduced by Maybank and Faugeras in [MF92]. The Kruppa equations exploit the correspondence of epipolar lines that are tangent to the image of the absolute conic ω . The image of the absolute conic in the first image is ω and for a second camera with the same internal calibration it is also ω . Now two epipolar lines l_1 and l_2 tangent to the IAC ω in the first camera can be found. These epipolar lines can be used to define a degenerate point conic ω_p . This degenerate point conic ω_p can be defined using the point line duality together with the dual image of the absolute conic ω^* . Accordingly two points on the epipolar lines l_1 and l_2 -like the epipole \mathbf{e} in the first view- and the DIAC ω^* define this conic ω_p as

$$\omega_p = [\mathbf{e}]_x \omega^* [\mathbf{e}]_x. \quad (2.45)$$

It follows that every point \mathbf{x} on l_1 or on l_2 lies also on the point conic ω_p . Similarly a point conic ω'_p for the second image is defined using the epipolar lines l'_1 and l'_2 corresponding to l_1

and l_2 . These lines are given as the intersection of the epipolar plane of l_1 respectively l_2 and the image plane of the second camera. The epipolar lines correspond under the homography H^∞ . Then the point conics correspond also under the homography H^∞ :

$$\omega'_p \cong (H^\infty)^{-T} \omega_p (H^\infty)^{-1}.$$

Using the definition (2.45) of the point conic leads to

$$\begin{aligned} [\mathbf{e}'_x \omega^* [\mathbf{e}'_x]_x] &\cong (H^\infty)^{-T} [\mathbf{e}]_x \omega^* [\mathbf{e}]_x (H^\infty)^{-1} \\ [\mathbf{e}'_x \omega^* [\mathbf{e}'_x]_x] &\cong F \omega^* F^T. \end{aligned} \quad (2.46)$$

Equation (2.46) are the Kruppa equations as originally introduced by Viéville in [VL95] into Computer Vision. Once the homogenous scale is eliminated a set of equations quadratic in the unknowns of ω^* can be obtained [VL95, HZ03]. This formulation of the Kruppa equations is in generally not easy to apply. There were several formulations introduced later that are easier to apply for self-calibration [Har97a, ZF96, Luo92, LF97a].

There are several difficulties in solving the Kruppa equations. The constraints provided by the Kruppa equations are weaker than the constraints provided by the infinite homography constraint (2.44) [HZ03]. However if only two views are given then the Kruppa equations give exactly the constraints available for the DIAC ω^* of (2.43). If there is no rotation between the cameras the Kruppa equations provide no constraints for the DIAC because the Fundamental matrix F is then equivalent to the cross product matrix of the epipole $[\mathbf{e}'_x]$. Due to these limitations they are rarely used for the self-calibration of cameras.

The next section will introduce a more tractable concept for self-calibration which is used instead of using the Kruppa equations.

2.9 Self-Calibration with the absolute dual quadric

This section will introduce another class of self-calibration techniques using the absolute dual quadric (see section 2.4). The absolute dual quadric Q_∞^* is very attractive because it simultaneously represents the plane at infinity π_∞ and the absolute conic Ω_∞ on the plane at infinity. The projection of the absolute quadric Q_∞ into the image is given as

$$\omega^* \cong P Q_\infty^* P^T, \quad (2.47)$$

where P can be any projectively skewed camera projection matrix [Tri97]. The main benefit is that it correlates the projection of the absolute dual quadric with the dual image of the absolute conic ω^* from (2.43). In the previous sections constraints on the camera calibration matrices were transferred into constraints on the dual image of the absolute conic (DIAC). Equation (2.47) offers now the possibility to get constraints for the absolute quadric using the DIAC. The DIAC itself directly depends on the camera calibration matrices of the concerned cameras. The major problem for using (2.47) is that it is only an equality up to scale. These scales can be eliminated by using ratios of components and cross-multiplication as constraints

$$(\omega^*)_{i,j} (P Q_\infty^* P^T)_{k,l} - (\omega^*)_{k,l} (P Q_\infty^* P^T)_{i,j} = 0. \quad (2.48)$$

Triggs showed in [Tri97] that 5 of the resulting 15 equations are linear independent. This can be seen because there are 6 degrees of freedom for (2.47) contained in the DIAC minus one for the scale ambiguity. The number of unknowns is 13 because there are 8 degrees of freedom in the absolute dual quadric Q_∞^* as explained below and 5 degrees of freedom in the DIAC ω^* .

For the remainder of this section the basic properties of the absolute dual quadric will be discussed. The most important property for self-calibration is that in a Euclidian space the absolute dual quadric is given as shown in (2.7). For a projective space skewed by a transformation $H_{4 \times 4}$ from (2.40) (see section 2.7) the absolute dual quadric is

$$H_{4 \times 4}^{-T} Q_\infty^* H_{4 \times 4}^{-1} = (Q_\infty^*)^p.$$

From (2.8) it follows that the absolute dual quadric is positive or negative semidefinite [HZ03] depending on the sign of the projective scale because the eigenvalues are scaled by the projective scale [GL89]. From equation (2.7) and equation (2.8) it can be seen that the vector π_∞ representing the plane at infinity in the projective reconstruction is in the null-space of $(Q_\infty^*)^p$.

After determining the absolute dual quadric from the image data the rectifying homography $H_{4 \times 4}$ can be determined by the Eigenvalue decomposition [GL89]. It decomposes any symmetric matrix S into $S = UDU^T$, where U is an orthonormal matrix and D is a diagonal matrix that contains the eigenvalues of S . Please note that this decomposition gives the exact $H_{4 \times 4}$ only for noise free data. For noisy data it does not enforce the known constraints for $H_{4 \times 4}$ and Q_∞ . Therefore these constraints have to be enforced in a second nonlinear optimization step to get the transformation $H_{4 \times 4}$ more precisely. Afterwards the rectifying homography $H_{4 \times 4}$ can be used to get Euclidean reconstruction and metric camera projection matrices.

The next section will inspect the limitations of the introduced self-calibration techniques in dependence of the camera motion during the image taking.

2.10 Critical camera motions for self-calibration

Shortly after introducing the first self-calibration approaches it was realized that not all image sequences offer the possibility to determine the intrinsic and extrinsic camera parameters uniquely. For example for sequences with pure translation it is not possible to get Euclidian reconstruction, only affine reconstruction can be computed [Stu97a, Stu96]. Such image sequences are called Critical Motion Sequences (CMS). Several researchers inspected special cases of CMS in [Tri97, Zel96, PGO97]. The work of Sturm [Stu97a, Stu96] provided a complete list of CMS for constant intrinsic parameters. The case of some varying intrinsic parameters or special camera configurations were discussed in [Stu99, KTÅ00]. This section will give an overview about the different types of CMS. At first the CMS for constant intrinsics are discussed because this is clearly a subset of the CMS for cameras with varying intrinsics. Afterwards the work for cameras with some varying intrinsics namely the focal length is reviewed.

Whether the motion sequence is critical or not does not depend on the calibration of the camera during the image acquisition. For the analysis here the calibration matrix K can be assumed to be the identity matrix. The calibration of the cameras is not possible if there are other proper virtual conics (PVC) apart from the absolute conic Ω that have a fixed projection

into the images because the constraints for self-calibration only exploit these properties of invariance. In that case it is not possible to distinguish between the images of the PVCs and the image of the absolute conic. The achieved calibration is therefore ambiguous. In his work [Stu97a, Stu96], Sturm looked for the image sequences that keep a specific PVC invariant. These specific PVCs can be divided into a few different classes. From these classes Sturm derived a set of critical motion sequence classes. The next paragraphs will show the results in more detail.

Potential absolute conics on π_∞ : At first the PVCs on the plane at infinity are inspected. Sturm derived the constraint that a rotation that keeps the eigenspace of the PVC on the plane at infinity constant also leaves the projection invariant [Stu96]. This constraint is also a sufficient constraint for the CMSs [Stu97b]. Therefore only the eigenspaces of the conics have to be analyzed. In general there are three types of conics on the plane π_∞ depending on the eigenvalue structure.

The first type has a triple eigenvalue. This class contains only the absolute conic itself. Therefore this class of PVCs is not critical for self-calibration. The second type has one double and one single eigenvalue. The eigenspace of this class corresponds to a plane and a line orthogonal to the plane³. Looking at the transformations that conserve the eigenvectors of these PVCs leads to all rotations about the line and a rotation of 180° about a line in the plane coincident to the line. The third type of conics has distinct eigenvalues. This represents three mutually orthogonal lines. Then only rotation by 180° about any of the three lines defined by the eigenvectors leaves the PVC invariant. Please note that all transformations of the latter case are covered by the transformations of the second types of conics with one double eigenvalue and one single eigenvalue because the rotation in the plane is a superset of the rotation about an axis. In both cases the identity transformation represents a transformation that lets the PVC invariant. The above types of PVC that have an invariant image projection under the described transformations gives the first class of CMS.

- **CMS-Class 1** Motion sequences for which all relative rotations are either rotations by an arbitrary amount about some line l or by 180° about lines perpendicular to l .

Further Sturm identified several subclasses of the **CMS-Class 1** with more potential absolute conics due to more restrictive motions in the subclasses

- **CMS-Class 1.1** Motion sequences for which all rotations are by an arbitrary angle about an axis parallel with an arbitrary but fixed line. This refers to an arbitrary translating camera where all rotation axes are parallel.
- **CMS-Class 1.2** Motion sequences for which all rotations are by 180° about mutually orthogonal axes.
- **CMS-Class 1.3** Motion sequences for which all rotations are by 180° about a specific axis.

³The eigenspaces of different eigenvalues of a symmetric matrix (like the PVC) are orthogonal [GL89].

- **CMS-Class 1.4** Motion sequences with no camera rotation at all (i.e. pure camera translation).

These characterizes the possible classes of ambiguities for proper virtual conics on the plane at infinity. The next possibility is that there is a PVC which is not on the plane at infinity but has an invariant image in all cameras like the absolute conic.

Potential absolute conics not on π_∞ : The previous discussion only looked at the camera orientation because the translation of the camera is not important for the image of a conic on the plane at infinity. For conics on any other plane the projection also depends on the position of the camera. The deduction of an algebraic description was done in [ZF96] but it is not clear if this description holds for more than three cameras. Therefore Sturm started with a specific camera pose and orientation. Then he looked for cameras that have the same image of the PVC. To get all classes of CMS in this case one has to inspect all possible PVCs⁴. This provides a geometric description of the classes of CMS instead of an algebraic description. Sturm found four additional classes of CMS by proving that the possible class of transformations is given by the planar reflections of the image of the PVC in the symmetry planes of the PVC [Stu97b]. A detailed description of this group of transformations can be found in [Stu97b]. Sturm deduced that for a CMS the camera centers can only lie on two equally sized circles for a circle. For an ellipse there are only eight points left where the camera center can be placed. These eight points lie also on two equally sized circles as for the circle as image of the potential absolute conic. Therefore these circles represent the group of possible camera center positions for CMSs. Usage of this deductions leads to the following groups of CMS.

- **CMS-Class 2** For a circle as projection of an elliptic cone in the reference view infinitely many positions are possible that lie on two circles and at each position four orientations are possible. This is the class containing the subclass of orbital motions of the camera.
- **CMS-Class 3** For an ellipse as projection of an elliptic cone in the reference view infinitely many positions are possible but all positions lie on a degree twelve curve [Stu97b]. Further at each position there are four choices of the orientation of the camera.
- **CMS-Class 4** For an ellipse as projection of a circular cone in the reference view only four positions for the camera are possible. At each position the camera may rotate freely about the main axis of the cone and about an axis perpendicular to the main axis of the cone.
- **CMS-Class 5** For a circle as projection of an absolute cone in the reference view two positions with any rotation of the camera are possible.

Sturm identified some more subclasses in [Stu97a].

⁴Generally there are three different types of conics possible: the PVC has one triple eigenvalue, it has one double eigenvalue and one single eigenvalue then it is a circular cone or it has three distinct eigenvalues; then it is an elliptic cone. The class of proper virtual conics only contains circles and ellipses[Stu97b].

Further Sturm provided a table of the most practical relevant situations from these classes. For a better understanding this table is reproduced in table 2.1. It shows the number of possible camera positions $\#C$ in the first column, the number of possible camera orientations $\#R$ and the number of ambiguous potential absolute conics $\#\Omega$. The only class of CMS where the calibration of the intrinsic parameters is possible is the **CMS-Class 5** where the transformation is not contained in any of the other classes.

Description	$\#C$	$\#R$	$\#\Omega$
CMS-Class 1.1	∞^3	∞	∞
planar motion	∞^2	∞	∞
pure translations	∞^3	1	∞^5
orbital motions	$2 \times \infty$	1	∞^2
pure rotations	1	∞^3	∞^3

Table 2.1: Practically important classes of CMS. $\#C$ is the number of possible camera positions, $\#R$ is the number of ambiguous camera orientations at each position and $\#\Omega$ is the number of potential absolute conics.

Further in [Stu97a] it was found that for ambiguous reconstruction the reconstructed scene and camera motion are always in the same class of CMS. This is important for the detection of these situations because it is often not decidable before the reconstruction whether a CMS is given. With this result the decision can be made after the reconstruction.

For cameras with varying intrinsic parameters all the above classes of CMS are still the same. In [Stu02] Sturm gave a complete catalog of CMS for cameras with varying intrinsic parameters. The work in [Kah99, Stu99] also discussed the CMS for cameras with varying focal length but they didn't give a complete list of CMS and their results are subsets of the results of [Stu02]. Here the extensions of the classes found in [Stu02] will be given.

For a camera with varying focal length mainly two new classes (denoted as **CMSvarF-Class** of camera motions appear. Both are results of PVCs lying not on the plane at infinity.

- **CMSvarF-Class 6** Collinear optical centers of the cameras with coinciding optical axes except from two cameras that allowed to have arbitrary optical axes orientation. This is illustrated in figure 2.3 (a).
- **CMSvarF-Class 7** The optical centers of the cameras are on an ellipse/hyperbola pair. The optical axis of the cameras are always collinear to the tangent of the ellipse/hyperbola at the optical center. This is illustrated in figure 2.3 (b).

One of the major conclusions is that the calibration of two cameras with varying focal lengths is ambiguous. It is well known that for this case there is a two fold ambiguity for the absolute conic.

Summary: This chapter introduced the basic concepts used throughout the thesis. The problem of uncalibrated structure from motion for scene reconstruction was introduced in section 2.7. The result of uncalibrated structure from motion was a projectively skewed 3D scene. To upgrade this reconstruction in a meaningful Euclidian reconstruction the projective skew has

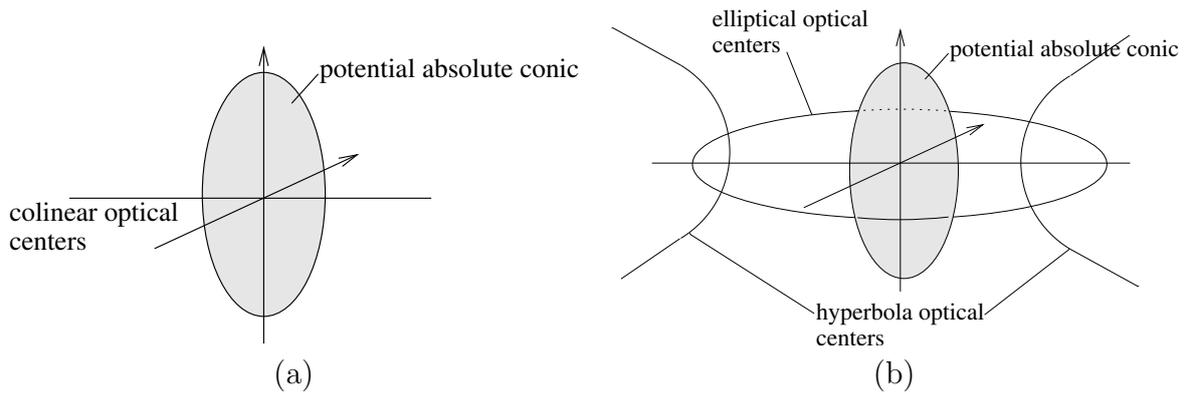


Figure 2.3: Additional CMSs for cameras with varying focal length. (a) **CMSvarF-Class 6**: Collinear optical centers with coinciding optical axes, (b) **CMSvarF-Class 7**: Optical centers of the cameras are on an ellipse/hyperbola pair. (redrawn from [KTÅ00])

to be determined. Accordingly the structure of the projective skew was deduced. Afterwards the entities absolute conic, dual absolute quadric were introduced. They are often used for self-calibration which determines the projective skew transformation. The last section discussed the analysis about the camera motions that are critical for the reconstruction and the estimation of the projective skew. In the next chapter the previous work done in the self-calibration will be discussed.

Chapter 3

Previous work on self-calibration

This section will introduce the previous work on self-calibration of cameras. The self-calibration approaches are divided in two main groups. The first group is the self-calibration of a camera that rotates about its optical center. It is often called a purely rotating camera. The second group of approaches solves the self-calibration of a translating and rotating camera. It will be denoted as a freely moving camera in the following. For both of these groups there exist only a few approaches that exploit known orientation of the camera. This chapter is organized as follows: First the approaches to calibrate a purely rotating camera will be discussed. Afterwards the approaches for the self-calibration of a freely moving camera are discussed.

3.1 Self-Calibration of a rotating camera

There are several approaches available for the self-calibration of a purely rotating camera with unknown orientation [Har95, SR97, dAHR01, Har94b, dARH99, Har97b, HHdAR99, SH98, SH99]. This section will discuss in detail the approach from Hartley introduced in [Har94b, Har97b] for the self-calibration of a purely rotating camera with constant internal parameters. It is the most popular approach for this situation. The work of Agapito et al. proposed in [dARH99, dAHR01] is the first and mostly referred approach for self-calibration of a purely rotating camera with varying intrinsic parameters. This technique will be discussed here as a representative approach for the class of self-calibration algorithms for purely rotating cameras with varying intrinsics. Further some approaches for calibrating a purely rotating camera with constant internal parameters and known orientation will be introduced. The latter group of algorithms compares directly to the proposed approach of this thesis, which is a self-calibration of a camera with varying intrinsic parameters exploiting the known orientation of the camera. Hence a purely rotating camera with constant intrinsics is a special case that is covered by the proposed approach of this thesis as published in [FK03b, FK03c]. The error analysis of the above self-calibration approaches were discussed by Wang et al. in [WKSX02, WKSX04] and will be summarized in section 3.1.4.

The techniques for the self-calibration of purely rotating cameras rely on the accurate estimation of the homographies between the given camera images. In order to compute an accurate homography that maps the points with high accuracy it is normally not sufficient to use the

least squares solution of (2.18) for detected feature (e.g. corners) correspondences. In (2.18) the mapping error is used as measurement. The least squares approach assumes a Gaussian distribution with zero mean for the mapping errors. Many corner detection methods in fact give a biased estimation of the corner position that is not Gaussian distributed [SM95, MM01]. Hence the computed least squares solution is often disturbed. To avoid this lack of precision it is recommended to use a direct matching approach like [BAHH92, SR97, dAHR01, SHK98]. The best results will be achieved with a robust cost function like the GNC cost function proposed by Blake and Zisserman in [BZ87] or another from [BR96] and an outlier process [BR96] that is induced by the cost function. The GNC function is a piecewise polynomial function that converges to an upper threshold. Accordingly it avoids the fitting of the solution to outliers. More details about the cost function are given in [BR96]. In the following the use of these well estimated homographies is required in order to get accurate calibration results.

3.1.1 Camera with constant intrinsics

Historically the first self-calibration method for a purely rotating camera with unknown but constant intrinsics was first developed by Hartley in [Har94b, Har97b]. This method used the infinite homography constraint (IHC) (2.44) to calibrate the camera.

Hartley used the fact that for a purely rotating camera the estimated homography is equivalent to the infinite homography H^∞ . Hartley uses the infinite homography H^∞ estimated from the images themselves to compute the IHC (2.44) for a camera with constant intrinsics

$$\omega^* = H_i^\infty \omega^* (H_i^\infty)^T, \quad (3.1)$$

where $\omega^* = K^T K$ is the dual image of the absolute conic from section 2.8. It is equivalent to using the standard calibration equations (2.43) based on the dual image of the absolute conic (DIAC). Hartley deduced from the independent components of the DIAC ω^* that (3.1) provides five linear independent equations. He showed that one image pair does not provide a sufficient number of constraints to solve for the DIAC ω^* uniquely. If only one camera image pair is used for the calibration there is a one parameter family of calibrations as solution of (3.1) left. Using the assumption of zero skew for the camera leads to a unique solution. However using a camera image triplet provides new constraints so that the DIAC ω^* can be calculated as least squares solution of (3.1).

To compute the calibration K of the camera from ω^* Hartley proposed to apply the Cholesky decomposition [PFTV88] on the dual image of the absolute conic ω^* . Using the Cholesky decomposition to get K from ω^* requires that the DIAC is positive-definite or negative-definite. For noise free data this is enforced by construction. Hartley denoted that for noisy data it is not guaranteed because the linear computation of ω^* using (3.1) does not enforce the positive definiteness of the DIAC ω^* . If the DIAC ω^* is not positive definite the calibration of the camera is not possible at all. Hartley mentioned in [Har94b] that in practice this only occurs for gross errors for the points of the correspondences used to compute the homography H^∞ .

To get a camera calibration that is as accurate as possible Hartley proposed a nonlinear optimization of the achieved calibration using (3.1) as cost function and the linear solution as initialization. For example the mapping error of the correspondences can be minimized using

equation (2.10) and the structure (2.16) of the infinite homography. Moreover this offers the possibility to apply further constraints on the calibration of the cameras like known aspect ratio etc. Hartley measured the performance of the proposed method and showed that the obtained results are rather stable in the range of usual accuracies of point matching. In his experiments Hartley also found that some of the components like the focal length and the aspect ratio are estimated very robustly¹. The estimation of the other three components skew and principal point are not as stable as the estimation of focal length in his experiments. This was also observed in my experiments and seems to hold for most types of self-calibrations.

In many applications the camera can't be restricted to have constant intrinsic parameters. For instance surveillance cameras have to zoom in order to recover more details. The next section will discuss one approach to calibrate a rotating and zooming camera.

3.1.2 Camera with some varying intrinsics

The approach of Hartley [Har94b] was extended for purely rotating cameras with varying intrinsic parameters by Agapito et al. in [dAHR98, dARH99, dAHR01]. In [dARH99] the infinite homography constraint (2.44) for cameras j and i with varying intrinsic parameters

$$K_i K_i^T = \rho_{j,i}^2 H_{j,i}^\infty K_j K_j^T (H_{j,i}^\infty)^T \quad (3.2)$$

was used for self-calibration. The scale $\rho_{j,i}$ is eliminated by normalizing the infinite homographies in a manner that $\det(H_{j,i}^\infty) = 1$. Then from (3.2) a set of linear equations for the entries of the DIAC was extracted. Hartley already showed in [Har94b] that this set of equations provides five linear independent constraints for the entries of the DIAC. Accordingly the number of intrinsic parameters that are allowed to vary is limited.

Let U be the number of unknown intrinsics of the first camera and V the number of intrinsics which may vary in subsequent images, the following constraint for the number of unknown intrinsic parameters of the cameras has to hold:

$$U + V(n - 1) \leq (n - 1) \cdot 5 \quad (3.3)$$

where n is the number of camera images. It can be seen from (3.3) that the number of varying intrinsics has to be strictly less than 5, i. e. at least one intrinsic parameter of the camera has to be fixed.

For every set of cameras that fulfills (3.3) unique DIACs $\omega_i^* = K_i K_i^T$ can be estimated using (3.2). In [dAHR98, dAHR01] Agapito et al. proposed to use the Frobenius norm of the difference of the mapped dual image of the absolute conic $H_{j,i}^\infty \omega^* (H_{j,i}^\infty)^T$ and the dual image of the absolute conic ω_i^*

$$\left\| \frac{\omega_i^*}{\|\omega_i^*\|_F} - \frac{H_{j,i}^\infty \omega_j^* (H_{j,i}^\infty)^T}{\|H_{j,i}^\infty \omega_j^* (H_{j,i}^\infty)^T\|_F} \right\|_F \quad (3.4)$$

¹Robust will throughout the thesis used in the sense of insensitive to disturbances caused by noise in the data.

as optimization criterion, where $\|\cdot\|_F$ is the Frobenius norm. It is also used to eliminate the scale $\rho_{j,i}$ in (3.2). The calibration of the cameras can be estimated using a nonlinear optimization with the cost function (3.4). This leads to a solution for the dual image of the absolute conic ω^* .

The calibrations K_i can be computed afterwards using the Cholesky decomposition of the estimated DIACs ω_i^* . This computation of the camera calibration has the same problem regarding the positive definiteness of the DIAC ω_i^* as described by Hartley in [Har94b] and already explained in section 3.1.1. Once the calibration is computed the rotations for each view can be estimated from the homographies by using the structure of the infinite homography (2.16)

$$K_i^{-1} H_{j,i}^\infty K_j = \frac{1}{\rho_{j,i}} R_{j,i}, \quad (3.5)$$

where the scale $\rho_{j,i}$ can be derived using the orthogonality of $R_{j,i}$.

The nonlinear optimization process that uses the cost function (3.4) needs an initialization close to the real solution in order to converge to the global minimum of the cost function (3.4), which is the correct DIAC ω^* . Using the infinite homography constraint (3.2) does not provide an easy way to apply constraints for the intrinsic parameters of the cameras. Zisserman et al. proposed in [ZLA98] to use the image of the absolute conic (IAC) ω

$$\omega = K^{-T} K^{-1} = \begin{bmatrix} \frac{1}{f^2} & \frac{-s}{af^3} & \frac{sv}{af^3} - \frac{u}{f^2} \\ \frac{-s}{af^3} & \frac{1}{(af)^2} + \frac{s^2}{a^2 f^4} & \frac{su}{af^3} - \frac{s^2 v}{a^2 f^4} - \frac{v}{(af)^2} \\ \frac{sv}{af^3} - \frac{u}{f^2} & \frac{su}{af^3} - \frac{s^2 v}{a^2 f^4} - \frac{v}{(af)^2} & 1 + \frac{(sv-afu)^2}{a^2 f^4} + \frac{v^2}{(af)^2} \end{bmatrix}$$

to apply constraints on the intrinsic parameters. The important property is that using the IAC gives linear constraints for a zero skew of the cameras:

- For zero skew s the component $(\omega)_{1,2} = 0$.
- Zero skew and square pixels the aspect ratio a is one, that leads to $(\omega)_{1,1} = (\omega)_{2,2}$. This is called the square pixel constraint.
- For zero skew and a principal point component in the x -axis which is zero the component $(\omega)_{1,3}$ becomes zero. Similarly for a principal point component in the y -axis that is zero $(\omega)_{2,3}$ is zero.

Each of the above constraints provides one linear equation for the IAC. Then with a maximum of six image pairs the IAC of the first image can be computed. Agapito et al. introduced in [dARH99, dAHR01] a technique that used the above constraints on the image of the absolute conic ω (IAC) to apply constraints on the IAC of the first image by

$$\underbrace{K_1^{-T} K_1^{-1}}_{\omega_1} = \frac{1}{\rho_{j,1}^2} (H_{j,1}^\infty)^{-T} K_j^{-T} K_j^{-1} (H_{j,1}^\infty)^{-1}. \quad (3.6)$$

The calibration of the first camera can be computed using the Cholesky decomposition of ω_1 . All other IACs can be calculated using (3.6) and the known IAC in the first image. Here

again the problem occurs that the estimated DIAC is not enforced to be positive definite. Whenever the estimated DIAC is not positive definite it is not possible at all to compute a camera calibration. Agapito et al. didn't refer to this problem in detail.

The linear calibration (3.6) does not cover all possible constraints for the calibration of the camera; for example a constant intrinsic parameter can't be modeled linearly using (3.6). Anyway Agapito et al. proposed to use this linear initialization step even if the cameras only roughly fulfill the constraints. It was mentioned that it still computes a solution close to the true calibration. Then this linear solution was used to initialize the nonlinear optimization process using the infinite homography constraint (3.2) in [dARH99, dAHR01].

During the nonlinear optimization the error function (3.4) is minimized. It measures the algebraic error in dependence on the estimated homographies $H_{j,i}^\infty$. To get a more accurate and more reliable calibration Agapito et al. proposed to refine the result based on a statistic cost function that takes the correspondences into account. This statistical cost function is the likelihood of the image points position given the rotation and calibration of the cameras. It depends on the mapping error of the point correspondences in the images, the previously estimated camera rotations $R_{j,i}$ and the estimated camera calibration matrices K_i .

The point location during the optimization is modeled by a Gaussian noise process. The estimated feature positions used for the homography computation are assumed to be disturbed by Gaussian noise η_{pixel} with zero mean and variance σ_{pixel} . Then the measured feature position \bar{x} in the image is

$$\bar{x} = \hat{x} + \eta_{\text{pixel}}, \quad (3.7)$$

where \hat{x} denotes the true feature position. The generating projection ray $\mathbf{x} \in \mathbb{P}^2$ of this point can be determined using the estimated camera calibration and the computed rotation from (3.5) of each camera. This leads to a maximum likelihood estimate (MLE) of

$$MLE = \arg \min_{K_i, R_i, \bar{X}_j} \sum_{i=1}^n \sum_{j=1}^{\#\text{points}} \left\| \bar{x}_j \simeq K_i R_i^T \mathbf{x}_j \right\|, \quad (3.8)$$

for the camera calibrations K_i , the camera rotations R_i , and a set of feature projection rays \bar{X} , where \simeq denotes minus after homogenization of the points. Agapito et al. stated that this estimation sometimes didn't lead to the correct camera calibration and neither to the correct rotation. This is mainly caused by the ill conditioned estimation of the principal point.

To avoid these problems Agapito et al. proposed to incorporate prior knowledge and use a maximum a posteriori approach. This exploits constraints on the principal point of the camera. The used knowledge is that the principal point probably lies close to the center of the image. The assumed noise model is a Gaussian distribution whose mean is the image center. With that a maximum a posteriori estimation of the probability of the intrinsics is given by

$$\text{MAP}_{pp} = \text{MLE} + \sum_{i \in \text{cameras}} (c^i - pp_{\text{prior}})^T \begin{bmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_y^2 \end{bmatrix}^{-1} (c^i - pp_{\text{prior}}). \quad (3.9)$$

where σ_x and σ_y are the standard deviations of the Gaussian noise model of the principal point, pp_{prior} is the prior knowledge about the principal point and c^i is current estimate of the principal

point of camera i . Minimizing the cost function (3.9) leads to better and more reliable results for the camera calibration and the orientations of the cameras.

The next subsections will discuss the approaches for camera self-calibration using additional constraints.

3.1.3 Camera self-calibration with some known extrinsics

In the previously discussed approaches only the image data was used except from (3.9) which incorporates the orientation of the camera explicitly. In (3.9) the orientation is contained explicitly but it was not the intention to use any previous knowledge of the camera orientation. There exist only a few approaches that combine image analysis and partially known extrinsics like the camera orientation for self-calibration in literature. This lack of attention is somewhat surprising since this situation occurs frequently in a variety of applications: cameras mounted in cars for driver assistance, robotic vision heads, surveillance cameras or PTZ-cameras for video conferencing often provide rotation information. In [DB93, Bas93, BR97, Ste95] the calibration of cameras with constant intrinsics and known rotation were discussed. They use nonlinear formulations of the optimization problem in the Euclidian space \mathbb{R}^2 namely the image plane coordinate system. Hence all these approaches have to use nonlinear optimization techniques in order to estimate the intrinsic camera parameters. Stein additionally estimated the radial distortion of the camera in [Ste95] by using the known orientation. Du and Brady used a special camera motion to estimate the radial distortion of the camera [DB93].

More often, calibrated cameras are used in conjunction with rotation sensors to stabilize sensor drift [NF02]. The next paragraphs will discuss the approaches from [DB93, Bas93, BR97]. The technique introduced by Stein in [Ste95] will afterwards be discussed in more detail since it is more general than the approaches in [DB93, Bas93, BR97].

Du and Brady approach: In [DB93] Du and Brady introduced a self-calibration approach for an active vision head with constant intrinsic camera parameters. They exploited the known orientation of the camera and the correspondences of at least two feature points.

The used model of the calibration assumes that the skew is zero. The provided self-calibration method uses the mapping of an image point under the known camera rotation. This leads to two nonlinear equations in the unknown camera parameters focal length, aspect ratio and principal point for each point. Hence two features provide a sufficient number of constraints to solve the equations proposed by Du and Brady. For more available feature correspondences Du and Brady propose to use a nonlinear least squares optimization. Starting from the nonlinear equations Du and Brady deduce a nonlinear calibration method that incorporates the motion of the features in order to directly compute the camera parameters from three tracked points. This method uses a rotation about the x -axis and afterwards a rotation about the y -axis. These rotations about one axis at a time require an active vision system.

Furthermore the radial distortion is estimated by Du and Brady in [DB93]. The estimation uses the known orientation, the previously computed calibration and a set of points in the image. It is not sufficient to use only two point correspondences to compute a radial distortion. After detecting several points in the image the camera is rotated. Then the two feature positions

are predicted based on the pinhole camera model, the estimated focal length, the computed aspect ratio and the estimated principal point. Next the observed error between the predicted feature position and the observed feature position can be used to compute the radial distortion of the camera.

In [DB93] it was mentioned that the features have to be detected with an accuracy of less than 0.2 pixel for an image of size 512×512 pixel in order to get a reliable camera calibration. In general it was seen in [DB93] that the estimation of the principal point is an ill-posed problem. The accuracy of the proposed approach was not measured in detail.

Basu approach: In [Bas93, BR97] Basu introduced concurrently with Du and Brady another self-calibration approach for a camera in an active vision system. The cameras intrinsic parameters were assumed to be fixed and the skew of the camera was ignored in the model. In contrast to Du and Brady in [Bas93] Basu used points on lines to estimate the flow between two images. This has the advantage of a more stable feature for the tracking and usually a higher precision of the feature position in the direction perpendicular to the direction of the line. The drawbacks are the problems caused by the aperture problem² [Stu11, Hil84, Hil87].

The self-calibration equations for the focal length developed by Basu rely on small rotations to estimate focal length and aspect ratio. The usage of these small rotations requires a high precision for the orientation measurements. In contrast to the focal length estimation the estimation of the principal point uses large rotations of up to 180° . Further this estimation of the principal point used the previously determined focal length and aspect ratio. By exploiting these large rotations for the principal point estimation Basu tries to overcome the poorly conditioned principal point estimation.

In order to self-calibrate the camera Basu developed two separate nonlinear equations for the intrinsic camera parameters. Each equation is only valid for the rotation about one axis at a time. Accordingly one equation models the rotation about the x axis. The second equation models the rotation about the y -axis. Depending on the axis of camera rotation the equation that exploits the rotation about this axis is chosen for the self-calibration. Therefore the approach of Basu requires precise rotations about the x , y and z -axis of the camera coordinate system. With a freely rotating camera this approach can't be used to calibrate the camera because a rotation about one axis at a time can't be assured. Further Basu gave an error analysis for the proposed approach in [BR97]. In order to simplify the self-calibration Basu linearized the nonlinear equations with several multiple step techniques which estimate the parameters independently. For example first of all the focal length was determined with small rotations. Afterwards the principal point was estimated with larger rotations using the previously determined focal length.

The approach was evaluated with synthetic data to measure the noise robustness. The computed results gave reasonable precisions as shown in [BR97]. Unfortunately not all data was given to compute relative errors afterwards.

So far the approaches of Du, Brady from [DB93] and Basu in [Bas93, BR97] were explained.

²Aperture problem denotes the fact that the motion of a line is locally ambiguous because each sensor (algorithm) has a finite field of view: it "looks" at the world through something like an aperture.

These approaches were the first self-calibration approaches that exploited the known camera orientation. The approaches in [DB93] and in [Bas93, BR97] use an active vision system to get well defined camera rotations, because their calibration equations use rotations about one axis at the time. The achieved quality of the calibrations is reasonable. The main disadvantage is that both techniques can't be applied for the calibration of a freely rotating camera due to the fact that the approaches rely on the rotation about one axis at a time. The next introduced approach of Stein [Ste96, Ste95, Ste93] overcomes this limitation by using general rotations.

Stein approach In [Ste95] Stein introduced a nonlinear calibration for a camera with constant intrinsic parameters and known orientation. The proposed method used the known rotation of the purely rotating camera to determine the intrinsic camera parameters, the radial distortion of the camera and an improved camera orientation. The technique exploits the projection equation (2.16) for a rotating camera and the radial distortion (A.45). The used model for the radial distortion does not assume that the principal point (u, v) of the camera is identical to the center (u^r, v^r) of the radial distortion. This leads to the radial distortion model

$$\begin{aligned} x^u &= u^r + (x - u^r) (\kappa_1 r_d^2 + \kappa_2 r_d^4 + \dots) \\ y^u &= v^r + (y - v^r) (\kappa_1 r_d^2 + \kappa_2 r_d^4 + \dots) \end{aligned} \quad (3.10)$$

with the squared radius r_d^2 with respect to the center (u^r, v^r) of the radial distortion given by

$$r_d^2 = (x - u^r)^2 + (y - v^r)^2.$$

After the correction of the nonlinear radial distortion the linear projection model (2.16) for a rotating camera is used as mapping function between the cameras. Eliminating the homogenous scale for the projection equation (2.16) this leads to

$$\begin{aligned} \hat{x}_i &= \frac{r_{1,1}x_j^u + r_{2,1}y_j^u + r_{3,1}}{r_{1,3}x_j^u + r_{2,3}y_j^u + r_{3,3}} \\ \hat{y}_i &= \frac{r_{1,2}x_j^u + r_{2,2}y_j^u + r_{3,2}}{r_{1,3}x_j^u + r_{2,3}y_j^u + r_{3,3}} \end{aligned} \quad (3.11)$$

as mapping equation for a point (x_j, y_j) in the first camera j to a point (x_i, y_i) in the second camera i . After compensating the radial distortion in the features the sum of the squared projection errors in the image

$$E = \sum_{i,j \in \{\text{correspondences}\}} ((\hat{x}_i - x_j)^2 + (\hat{y}_i - y_j)^2) \quad (3.12)$$

is used as cost function for a nonlinear optimization. Using this cost function Stein estimated all intrinsics of the camera, the first two coefficients κ_1, κ_2 of the radial distortion function (3.10) and the orientation of the camera. The initialization of the nonlinear optimization process was not discussed in [Ste95] and the convexity of the used cost function (3.12) wasn't proved.

Stein exploited the known camera orientation for the calibration of a purely rotating camera. The camera panned and tilted in order to compute a calibration. Stein explained that both

rotations are mandatory to estimate all camera parameters. This result was later shown from Sturm in [Stu97a, Stu96] as discussed in section 2.10. Steins arguments described the CMS-Class 1.1 of section 2.10 on page 26 of Sturm. It is the class of the camera rotations about only one axis.

The results of the nonlinear calibration of Stein in [Ste95] show that this calibration achieves very precise calibration results for the internal camera parameters. A summary of Steins measurements for different cameras is shown in table 3.1, the range represents the range of the achieved accuracies for the intrinsics of the different cameras used. These results were reached using a rotation that was measured with an error less than $\pm \frac{1.22^\circ}{60}$, which is a very precise measurement of the rotation. The estimated coefficients of the radial distortion have significantly larger standard deviations than the focal length, the aspect ratio and the principal point. Analyzing the undistortion function showed that the undistortion still works good because the error after compensation of the radial distortion is usually less than 0.5 pixels for a 512×485 pixel image. This is due to the fact that an error in κ_1 in (3.10) can be to some degree compensated with κ_2 .

The achieved calibration result probably depends on the initialization of the nonlinear optimization step.

parameter	standard deviation
focal length	0.08%-0.17%
aspect ratio	0.05%-0.11%
principal point	1.4%-1.8%

Table 3.1: Standard deviation of the estimated parameters using the method proposed by Stein [Ste95]. These results based on rotation with an error of less than $\pm \frac{1.22^\circ}{60}$. The values are computed from the results shown in [Ste95]

In [Ste95] Stein also analyzed the influence of translation on the calibration that occurs if the camera is not exactly rotated about its optical center. This analysis is fully covered by the later proposed more general analysis of Wang et al. in [WKSX04, WKSX02]. The details of this analysis can be found in section 3.1.4.

Concluding, all approaches for self-calibration of a purely rotating camera with known orientation assume constant intrinsic parameters. They all use the motion of a feature between consecutive images to compute the camera parameters. The approach of Stein [Ste95] showed the best results for the intrinsic camera parameter estimation. It uses a nonlinear optimization to compute the calibration. Neither was shown by Stein that the cost function is convex nor was the initialization given in detail. Steins approach is able to use general rotations to calibrate the internal parameters. The approach of Basu linearized the nonlinear calibration equations. This linearization used a multi-step technique due to the ability of the active vision systems to produce rotations that minimize the error in a specific intrinsic parameter. Du and Brady developed a solution for the nonlinear equations that relies on rotation about one axis at a time in order to compute an intrinsic camera calibration. Hence the approaches of Du, Brady and Basu can't use general rotations for calibration.

3.1.4 Error analysis of calibration from purely rotating cameras

During the last years some error analysis was done for the self-calibration of purely rotating cameras. The work that gives the most general overview about sources of errors is the work from Wang et al. in [WKSX04]. Wang et al. investigated the errors caused from rotating the camera not exactly about its optical center since then the cameras optical center is not fixed. In this case often not the desired infinite homography is estimated. Some of these error sources were also investigated earlier by Stein in [Ste95]. Further Agapito et al. did some experiments on the effects of radial distortion on self-calibration in [dAHR01]. This section will summarize the above mentioned results to give an overview about the sources of errors for the self-calibration of a purely rotating camera apart from the critical motion sequences from section 2.10.

Errors from homography: At first the errors caused from the homography used are discussed. These were investigated from Wang et al. in [WKSX04]. The calibration of a rotating camera always registers the images with respect to the first image. This can be seen from the above given approaches from sections 3.1.1-3.1.3 which all use this registration, namely the estimated homographies, as input value of the self-calibration. It is assumed in all these approaches that the estimated homography is the infinite homography H^∞ from (2.16) that maps over the plane at infinity and has the structure

$$H_{j,i}^\infty = K_i R_{j,i} K_j^{-1},$$

where $R_{j,i}$ is the rotation between camera j and i . This assumption is violated if the camera doesn't rotate exactly about its optical center. In this case the estimated homography is not always the infinite homography since if the camera observes a dominant plane π in the scene the estimated homography uses this plane as mapping plane. Hence the estimated homography does not have the structure of the infinite homography (2.16). This homography contains a translational component for the camera like in (2.15):

$$H_{j,i}^\pi = K_i \left(R_{j,i} + \frac{\tilde{t}_{j,i} n_\pi^T}{d_\pi} \right) K_j^{-1} \text{ with } \tilde{t}_{j,i} = R_{j,i} t_{j,i} = [t_x, t_y, t_z]^T.$$

This homography $H_{j,i}^\pi$ is estimated even if an "averaged" scene plane is observed and the cameras will be registered against this plane π_a . This averaged scene plane can be seen as an average of all points on the planes present in the scene, and results from the overdetermined estimation problem for the homography. If the calibration techniques still assume to have the homography $H_{j,i}^\infty$ they implicitly force the computed calibration to fulfill

$$\tilde{K}_i \left(R_{j,i} + \frac{\tilde{t}_{j,i} n_\pi^T}{d_\pi} \right) \tilde{K}_j^{-1} = K_i R_{j,i} K_j^{-1}, \quad (3.13)$$

where \tilde{K}_i and \tilde{K}_j are the estimated calibrations of the cameras i and j . Wang et al. investigated the error of the solution in dependence of the number of estimated parameters for a camera with constant intrinsic parameters.

First of all Wang et al. investigated the estimation of the focal length for known aspect ratio, skew and principal point. In [WKSX04] this is denoted as one parameter estimation. Then the camera calibration matrix can be reduced to

$$K = \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

In case of a general homography instead of an infinite homography the estimation of the focal length using (3.13) leads to

$$\begin{aligned} & \underbrace{\begin{bmatrix} (R_{i,j})_{1,1} + t_x n_x & (R_{i,j})_{1,2} + t_x n_y & f * ((R_{i,j})_{1,3} + t_x n_z) \\ (R_{i,j})_{2,1} + t_y n_x & (R_{i,j})_{2,2} + t_y n_y & f * ((R_{i,j})_{2,3} + t_y n_z) \\ \frac{(R_{i,j})_{3,1} + t_z n_x}{f} & \frac{(R_{i,j})_{3,2} + t_z n_y}{f} & (R_{i,j})_{3,3} + t_z n_z \end{bmatrix}}_{\tilde{K}_i \left(R_{j,i} + \frac{t_n^T}{d_\pi} \right) \tilde{K}_j^{-1}} \\ = & \underbrace{\begin{bmatrix} (R_{i,j})_{1,1} & (R_{i,j})_{1,2} & f * (R_{i,j})_{1,3} \\ (R_{i,j})_{2,1} & (R_{i,j})_{2,2} & f * (R_{i,j})_{2,3} \\ \frac{(R_{i,j})_{3,1}}{f} & \frac{(R_{i,j})_{3,2}}{f} & (R_{i,j})_{3,3} \end{bmatrix}}_{K_i R_{j,i} K_j^{-1}} \end{aligned} \quad (3.14)$$

It can be seen that the error is distributed row-wise over the homography. Exploiting the properties of the distribution function it is possible to deduce a closed form solution for the focal length that is independent of one component of the camera translation vector $\tilde{t}_{j,i}$ [WKSX04]. This shows that one can use a priori knowledge of the error source to reduce the resulting error. For example a computation that is invariant against t_y can be chosen if it is known that the translation error mainly occurs in the y -component of the camera motion. From equation (3.14) it can be seen that the disturbances introduced by the translation are decreasing if the rotation is increasing. This can be used to improve the quality of the calibration by using images with larger rotation in between to estimate the homography. Such a homography then provides a calibration with a smaller calibration error. Further Wang et al. showed in [WKSX04] that for a fronto parallel plane and a panning camera an error free computation of the focal length is possible.

Wang et al. also investigated the estimation of the focal length and the aspect ratio for known principal point and skew. This leads to an exact solution for the focal length for a fronto parallel plane and a panning camera. The aspect ratio can not be computed error free in any case. For a general rotation prior knowledge can be used for the estimation of the focal length with a minimal resulting estimation error. Moreover Wang et al. deduced results for the estimation of aspect ratio and the principal point. In this case it is in general not possible to compute error free intrinsics as a closed form solution even not in the special cases of a fronto parallel plane or a panning camera. Concluding Wang et al. showed that the a priori knowledge about the error distribution can be used to minimize the estimation error caused by the camera motion and scene planes. In some special cases Wang et al. showed that the right choice of the equations used to compute the intrinsics avoids errors in the computed intrinsics.

Errors from radial distortion: Apart from the errors caused by the general homography instead of the infinite homography for real cameras one has to deal with errors caused from the radial distortion of a real camera. In contrast to the error in the homography this effect is a nonlinear disturbance. In general a radial distortion adds an offset (A.45) depending on the pixel position to each pixels location. Moreover radial distortion changes with the focal length of the camera. Agapito et al. investigated the effects of radial distortion in [dAHR01] by experiments on synthetic data and on a real image sequence. The results will be summarized here. If the homography is estimated from correspondences without correcting the effects of the radial distortion the estimated intrinsics and also the extrinsics are strongly disturbed. The disturbance were not analyzed in detail by Agapito et al. in [dAHR01]. Please note that the direction of the disturbance mainly depends on the sign and absolute value of the radial distortion. The conclusion of the experiments in [dAHR01] was that the radial distortion has to be corrected ahead of the calibration. This correction can be done by image based techniques like proposed in [Kan00, TM00].

There is a lot of work on estimating the radial distortion from images. For image sequences with a camera with varying intrinsics it is obvious that each image has its own set of radial distortion parameters. This leads to additional parameters for each image. An estimation of one set of radial distortion parameters for each image is not useful because it leads to unstable estimations [dAHR01] for the radial distortions and therefore also for the intrinsics. In [SK99] Sawhney and Kumar modeled the radial distortion as constant to avoid instabilities. They provided a direct solution to determine the radial distortion. This solution can be used if the radial distortion is known as approximately constant a priori. This is often the case if a small change in focal length is expected. Another solution was introduced by Tordoff and Murray in [TM00]. They propose to approximate the value of the first coefficient κ_1 of the radial distortion function (A.45) for a small group of images with a cubic polynomial of the focal length

$$K_1 = \alpha_0 + \alpha_1 f + \alpha_2 f^2 + \alpha_3 f^3. \quad (3.15)$$

Then only the four coefficients $\alpha_0, \dots, \alpha_3$ of the polynomial have to be estimated. This avoids the use of at least one additional parameter per camera and lead to good results in the test [dAHR01]. It is a clear result that radial distortion has to be eliminated by the self-calibration process like in [Ste95, TM00] or the radial distortion has to be corrected beforehand. There are many techniques available to estimate the radial distortion [Ste96, KWT87, Kan00] for example Kang uses snakes to compute the radial distortion from a single image in [Kan00].

So far a review of the existing approaches of calibrating a purely rotating camera and an overview over the error sources was provided. It was shown that all existing approaches have some limitations. The approaches [dAHR01, dAHR98, dARH99, Har97b, Har94b] for calibrating the intrinsic and extrinsic parameters of the camera have the drawback of using the dual image of the absolute conic. With noisy data this leads sometimes to the problem of a not positive definite DIAC and in those cases the calibration can't be computed. Further, in the approach [dAHR01, dAHR98, dARH99] proposed by Agapito et al. the number of varying intrinsic parameters is limited by (3.3). The few existing approaches [DB93, Bas93, BR97, Ste95] for calibration exploiting a known camera orientation use nonlinear techniques to calibrate a

camera with constant intrinsics. There is no approach available that is able to calibrate a camera with varying intrinsic parameters and known orientation. The next section will discuss the approaches for self-calibration of a freely moving camera.

3.2 Self-Calibration from general motion

The calibration of a freely moving camera was first investigated by Maybank and Faugeras in [MF92]. The presented approach proved that it is possible to calibrate a freely moving camera with constant intrinsic parameters that observes a static scene. Afterwards there were many approaches for camera self-calibration from unknown general motion and constant intrinsics proposed in [FH92, HÅ96, PG97b, Luo92, LF97b, Zel96, Har94a, Tri97, PGO96, PG97a]. These approaches applied various constraints to reach a self-calibration. In section 3.2.1 some methods for self-calibration of a camera with constant intrinsic parameters will be introduced. Later researchers payed attention to the self-calibration of a freely moving camera with varying intrinsic parameters [HÅ97, PKG98]. These approaches will be discussed in section 3.2.2. They all use the images of the cameras as only source of information.

For a freely moving camera in my knowledge only one approach that exploits known extrinsic parameters namely known camera orientation is available. The work of Zomet et al. [ZWS01] uses the known relative orientation of fixed coupled cameras to self-calibrate the cameras. This technique will be discussed in section 3.2.3.

Another class of self-calibration methods ensures the chirality of the cameras and the points. The first approach was proposed by Hartley in [Har98]. Later Nistér improved Hartleys approach in [Nis04]. Further there are some approaches that apply structural constraints on the scene to achieve a self-calibration of the camera. An interesting approach was recently proposed by Rother and Carlsson [RC02]. They jointly estimate Fundamental matrices and homographies from a moving camera that observes simultaneously the scene and some reference plane. The homography induced by the reference plane generates constraints that are similar to a rotation sensor and self-calibration can be computed linearly. However, some structural constraints on the scene are necessary, while the proposed approach of this thesis applies constraints to the camera only.

3.2.1 Camera with constant intrinsics

There were several methods for the self-calibration of a freely moving camera proposed over the past twelve years. These methods basically form three different classes. The first class factorizes the projection matrix and ensures the constraint that the camera calibration is constant like in [Har94a]. The second class uses the epipolar geometry to get constraints on the absolute conic on the plane at infinity. The advantage of these methods is that only the pairwise Fundamental matrices are needed and no 3D scene information is required. The third class requires 3D scene information. These methods are reconstructing a projectively skewed scene and the corresponding camera matrices. From these reconstructions the self-calibration is computed using constraints on the absolute conic or the absolute quadric. In [Tri97] Triggs proposed two

ways to use the dual image of the absolute quadric from (2.43).

Factorization of the projection matrix A method that uses the QR-decomposition of the projection matrix was proposed by Hartley in [Har94a]. Hartley derives from (2.35) by using only the first 3×3 part of the equation: The following constraint

$$P_i^p H_{4 \times 3} \cong K_i R_i^T \quad (3.16)$$

with $H_{4 \times 3}$ as the first three columns of the skewing homography $H_{4 \times 4}$

$$H_{4 \times 3} = \begin{bmatrix} K \\ (Kv)^T \end{bmatrix}$$

from (2.40). The unknowns in (3.16) are the calibration matrix K that is assumed to be constant for all cameras and the normal v of the plane at infinity. Hartley proposed to compute K_i and R_i through the QR-decomposition. For a camera with constant intrinsic parameters the computed calibration K should be similar to all K_i computed by the QR-decomposition. Consequently the following constraint should be fulfilled:

$$K \cong K_i \quad \forall i \in \text{cameras.} \quad (3.17)$$

For the estimated K this equality is up to scale according to the unknown scale of the camera projection matrix. This scale in (3.17) can be computed as the ratio of the traces of K and K_i . Then Hartley proposed to minimize the following error $\epsilon(K, v)$ for all cameras

$$\epsilon(K, v) = \|\rho_i K^{-1} K_i - I_{3 \times 3}\|_F, \quad (3.18)$$

where ρ_i is the scale to achieve an Euclidian equality in (3.17). This method doesn't exploit the rotation invariant calibration pattern of the image of absolute conic or the dual absolute quadric. The techniques that use the constraints on the absolute conic or the dual of the absolute quadric all use rotation invariant equations for self-calibration. An introduction to these methods will be given in the next two paragraphs.

Self-calibration using the Kruppa equations: As mentioned before, the approach of Maybank and Faugeras in [MF92] is historically the first self-calibration approach. This approach uses the Kruppa equations (2.46) from section 2.8.1 to apply constraints on the dual image ω^* of the absolute conic. The scale factor in (2.46) was eliminated by using cross-ratios comparable to those in (2.48). Thus from the Kruppa equation in [MF92] two linear independent equations were deduced. Maybank and Faugeras proposed the following minimization criterion for the error ϵ

$$\epsilon = \left\| \frac{\omega^*}{\|\omega^*\|_F} - \frac{F\omega^*F^T}{\|F\omega^*F^T\|_F} \right\|_F \quad (3.19)$$

for all cameras. This formulation is independent from the position of the plane at infinity. This is an advantage because it is not needed to ensure a consistent framework for the self-calibration process, i.e. the pairwise Fundamental matrices can be used to calibrate the cameras. According

to the invariance of the Fundamental matrix to projective skew the position of the plane at infinity is not contained in (3.19). This property is helpful if it is not possible to register all views into one scene. This advantage becomes a limitation in case of noise. Then the plane at infinity as a regularization term that ensures a consistent framework for images is not available. Due to the lack of a consistent frame as regularization this method tends to fit the calibration of each image pair in a different way [Pol99].

So far, it was explained that self-calibration techniques which use the Kruppa equations only needs the pairwise Fundamental matrices. This is helpful in situations where the computation of a global consistent registration is difficult. The next paragraph introduces a calibration that requires a globally consistent framework in order to calibrate the cameras.

Self-calibration using the dual of the absolute quadric: The absolute quadric as a geometric entity for self-calibration was introduced by Triggs in [Tri97]. The proposed method used the fact that the dual absolute quadric is projected into the dual image of the absolute conic as shown in (2.47). The scale factor in (2.47) is eliminated by using the cross-ratios from (2.48) for computations that provide five linear equations [Tri97], leaving thirteen unknowns that have to be estimated. The DIAC has five unknowns as explained in section 2.3. The number of degrees of freedom of the dual of the absolute quadric was deduced in section 2.4 as eight.

To solve the self-calibration equations provided by (2.48) Triggs proposed two methods. The first solving method reduces the error $\epsilon(\omega^*, Q^*)$ as the residual error of (2.48). The nonlinear optimization process additionally ensures the rank 3 constraint of the absolute dual quadric Q^* . The limitation here is that the self-calibration process has to be initialized.

The second proposed approach for the solution of (2.48) is a quasi linear method for self-calibration. Triggs used the fact that the equations (2.48) are linear in the entries $\omega_i^i Q_i^*$, where \cdot_i denotes the i -th coefficient of the vector that contains the coefficient of the conic or quadric. The dual image of the absolute conic has six coefficients parameterizing the conic and the dual absolute quadric has 10 coefficients like every quadric. This gives sixty different products of these coefficients and therefore fifty nine unknowns due to the scale ambiguity of conics and quadrics. Writing the equations in (2.48) as a scalar product leads to fifteen linear independent equations. Consequently four images are needed to estimate the dual image of the absolute conic and the absolute quadric. The estimation does not enforce the rank three constraint of the dual absolute quadric. Accordingly the rank three constraint is enforced afterwards by setting the smallest singular value of the computed quadric to zero (closest rank three approximation).

Triggs reported that the non-linear method should be preferred for self-calibration, even despite the issues of initialization. Due to the fact that the nonlinear approach is faster, more robust and more accurate.

This section introduced some approaches for the self-calibration of a freely moving camera with unknown but constant intrinsic camera parameters. The next section will introduce approaches that need weaker constraints on the camera parameters.

3.2.2 Camera with varying intrinsics

Techniques for the self-calibration of a freely moving camera with varying focal length were introduced by Heyden and Åstrom in [HÅ97] and Pollefeys et al. in [PGP96]. Pollefeys et al. extended the self-calibration in [PKG98, Pol99] for a camera that has only a zero skew.

Here the approach of Pollefeys et al. [PKG98, Pol99] will be discussed because it is the algorithm that represents the state of the art in self-calibration of a freely moving camera with varying intrinsic parameters. Pollefeys et al. are using the projection (2.47) of the absolute quadric into the images. Looking at the number of degrees of freedom of the absolute quadric in (2.47) gives fifteen unknown parameters, namely eight for the dual absolute quadric Q^* , six for the dual image of the absolute conic ω^* and one for the scale. To get a metric framework with seven degrees of freedom (six for a similarity transformation, one for the scale) one needs at least eight constraints. This leads to the a constraint on the minimum length of the image sequence that is needed to self-calibrate the cameras. It is given by

$$n\mathcal{K} - (n - 1)\mathcal{F} \geq 8 \quad (3.20)$$

where \mathcal{K} is the number of known intrinsics and \mathcal{F} is the number of fixed intrinsic parameters. Hence even one known parameter like zero skew is sufficient to calibrate the camera. Some examples for the minimal required length of the sequence in dependence on the known parameters are shown in table 3.2.

constraint	known	fixed	#required images
zero skew	s		8
Fixed aspect ratio and zero skew	s	a	5
aspect ratio known and zero skew	s, a		4
unknown focal length	s, a, u and v		2
constant intrinsics		f, s, a, u and v	3

Table 3.2: Length of sequence needed for self-calibration in dependence of the number of given constraints for a camera with varying intrinsic parameters. Cite from [PKG99].

Pollefeys et al. deduced a minimal parameterization of the dual absolute quadric Q_∞^* in (2.47) as

$$Q_\infty^* = \begin{bmatrix} \omega^* & -\omega^*v \\ -v^T\omega^* & v^T\omega^*v \end{bmatrix}, \quad (3.21)$$

where v is chosen like in (2.40). Thus $[K^Tv, 1]$ is the normal vector of the plane at infinity. Using this minimal parameterization it is possible to use (2.47) for self-calibration. This provides all unknown entries of the projective skew $H_{4 \times 4}$ from (2.35) of the reconstruction and the cameras. Therefore estimating this parameterization of the dual absolute quadric leads to a

metric reconstruction. Pollefeys proposed to use a nonlinear least squares optimization to solve (2.47). The proposed optimization criterion is

$$\min \sum_{i=1}^n \left\| \frac{\omega_i^*}{\|\omega_i^*\|_F} - \frac{P_i Q_\infty^* P_i^T}{\|P_i Q_\infty^* P_i^T\|_F} \right\|_F, \quad (3.22)$$

where $\omega_i^* = K_i K_i^T$ is the dual image of the absolute conic with the calibration matrix K_i of the i -th camera. The camera projection matrix of the i -th camera is denoted as P_i and $\|\cdot\|_F$ is the Frobenius norm. The dual image of the absolute conic ω_i^* and the projection of the dual absolute quadric are normalized to have Frobenius norm one in order to eliminate the unknown scale in (2.47). Pollefeys et al. further provided a linear estimation technique if only the focal length is known. Then only two images are needed to estimate the varying camera parameters.

The provided self-calibration method was tested in [PKG99, Pol99, PKG98] on synthetic and real sequences. It can be seen in [PKG99] that the nonlinear self-calibration gives reasonable results for low noise in the pixel position (less than 1 pixels standard deviation) only. The model used for this measurement estimates the focal length and the unknown principal point. For higher noise in the pixel position the error increases dramatically. This originates in the low redundancy of the equations extracted from (2.47) used for self-calibration. Pollefeys et al. proposed the presented linear approach for use with a higher noise in the pixel position. It only estimates the focal length and assumes a known aspect ratio together with a principal point set to the image center. This linear approach is much more robust because there is more redundancy that can be used to improve the reliability of the estimation. However due to the assumptions made for this approach it has a higher error for low image pixel noise than the nonlinear approach. Pollefeys et al. also measured the performance of their algorithm on a real image sequence. It was observed that the estimation of only the focal length is rather stable. If the principal point is also estimated than the estimation is not very stable. For example for a camera with constant intrinsic parameters the estimated principal point moves by up to 100 pixel in an image with resolution 768×576 . An improved self-calibration technique as extension of this method was later proposed by Pollefeys et al. in [PGV⁺04]. It exploits the ranges and uncertainties of the internal camera parameters to weight the self-calibration equations. Accordingly the self-calibration process is more robust against noise in the input data.

Summarizing, Pollefeys et al. introduced in [PKG99, Pol99, PKG98] a well accepted approach for the self-calibration of a freely moving camera with varying intrinsic parameters. The only assumption needed is a vanishing skew of the camera. Pollefeys et al. evaluated the approach and proposed to use the nonlinear calibration for a feature localization noise of less than 1 pixel. For higher noise an estimation of the focal length only was proposed.

3.2.3 Multi-Camera system with constant inner orientation

Zomet et al. proposed in [ZWS01] a self-calibration approach for a multi-camera system. They estimated the projective skew matrix $H_{4 \times 4}$ from (2.40) in case of varying intrinsic parameters of the cameras of the multi-camera system. The assumption that has to be valid for the

multi-camera system is that the orientation of the camera coordinate systems remains fixed. This approach is related to the topic of this thesis because it uses a pre-calibrated known orientation of the cameras to each other for self-calibration. In order to calibrate the extrinsic parameters of the multi-camera system beforehand Zomet et al. used a calibration pattern. For this extrinsic calibration of the multi-camera system all intrinsic parameters of the cameras are set to constant values. Finally the multi-camera system is calibrated using one of the approaches for self-calibration of a freely moving camera from section 3.2.1. Once this is done the projective skew $H_{4 \times 4}$ from (2.40) for the cameras is known and the relative orientation of the cameras to each other can be computed from the rectified camera projection matrices. This orientation then has to remain fixed in order to be used for self-calibration. After this pre-calibration step the intrinsic parameters of the cameras are allowed to vary. The multi-camera system can be used afterwards to get an image sequence of an unknown scene. The intrinsic camera parameters of the cameras during the image capturing of new images can be estimated from the known orientation as explained in the following.

For self-calibration Zomet et al. used a similar approach like Hartley in [Har94a]. Hartley proposed to deduce the self-calibration equations from (3.16) with the matrix $H_{4 \times 3}$ containing the first three columns of the projective skew matrix $H_{4 \times 4}$ from (2.40). In contrast to Hartley in [Har94a] Zomet et al. interpreted the matrix $H_{4 \times 3}$ as a general 4×3 matrix with twelve degrees of freedom. However Zomet et al. deduced

$$P_i H_{4 \times 3} R_i \cong K_i \quad (3.23)$$

from equation (3.16). With a known rotation R_i equation (3.23) is linear in the intrinsic parameters of the camera K_i and the matrix $H_{4 \times 3}$. The projective scale ρ_i can be modeled explicitly to achieve an Euclidian equality. This leads to

$$P_i H_{4 \times 3} R_i = \rho_i K_i. \quad (3.24)$$

The known entries of the camera calibration matrix K_i are used to get calibration equations from each known camera projection P_i . Zomet et al. used twelve unknown entries in $H_{4 \times 3}$ and n unknown scales ρ_i , where n is the number of cameras. Hence (3.24) imposes three constraints on the matrix $H_{4 \times 3}$ and one on the scale ρ_i if all parameters are allowed to vary. From a counting argument it follows that four images are needed to calibrate the camera with varying intrinsic parameters. Zomet et al. deduced the following constraint for the number of cameras needed to self-calibrate the cameras of the multi-camera system

$$n + 12 \leq hn + 1, \quad (3.25)$$

where h is the number of known intrinsic parameters.

Zomet et al. evaluated the proposed approach to get an Euclidian reconstruction, regarding the estimation of the projective skew matrix $H_{4 \times 3}$ for cameras with varying intrinsics. The resulting Euclidian reconstruction computed from the projective reconstruction by applying $H_{4 \times 3}$ as described in section 2.7 delivered reasonable results, i.e. the achieved Euclidian reconstruction has reprojection errors of less than one pixel.

It can be seen that the used parameterization of the projective skew $H_{4 \times 3}$ is not optimal because it is interpreted as a general 4×3 matrix. Thus it has twelve degrees of freedom instead of the eight degrees of freedom contained in the projective skewing transformation. Zomet et al. always take the unknown scale into account for the estimation which introduces one additional parameter for each new camera. As long as only the known entries with value zero of the calibration matrices are used this is not needed. In this case each camera provides only three instead of four constraints but it doesn't introduce a new unknown parameter.

Summary: In this chapter the previous work in the field of camera self-calibration was discussed. There are only a few approaches that incorporate a known orientation of the camera. For purely rotating cameras only non-linear approaches were proposed to get a calibration of the camera using the known orientation. For a multi-camera system with constant inner orientation Zomet et al. proposed a linear approach to get the projective skew which is closely related to the self-calibration.

The approaches for self-calibration of purely rotating cameras proposed from Hartley in [Har97b, Har94b] and Agapito et al. [dAHR98, dARH99, dAHR01] suffer from noise in the pixel position. Sometimes this causes that the calibration can't be estimated due to the fact that the Cholesky decomposition can't be applied. The calibration approaches for a freely moving camera with varying intrinsic parameters proposed by Pollefeys et al. also tend to be unreliable under noise in the pixel position. To avoid this only the focal length is estimated as a rough approximation for the camera calibration.

Chapter 4

Self-Calibration with known rotation

In this chapter novel techniques on the use of available external orientation information for camera self-calibration will be introduced. The proposed approaches were published in [FK03b, FK03c]. These techniques address both cases of camera motion, namely purely rotating cameras and freely moving cameras. The new techniques for self-calibration exploit a known camera orientation to reach a linear self-calibration. At a first glance requiring orientation data seems to strongly restrict the field of application but in many situations this orientation information is already available. In robotic applications the orientation of the camera mounted on the robot is easy to compute from the known position of the robot. Furthermore in cars the use of cameras became very popular over the last years. In a modern car the orientation of the car can be computed from the car sensor data [Zom91, GWF03]. Hence the camera position can be calculated from the car orientation. In surveillance applications the used cameras are typically pan-tilt-zoom cameras that also provide orientation information for the camera, e.g. the orientation can be computed from the known control information of the camera. In video-conferencing normally the orientation of the camera is known but the internal camera parameters are unknown because the cameras are zooming in order to get a better image of the currently interesting person. The requirement of known rotation is less restrictive than any requirement for knowledge about the scene because it can be measured by a simple sensor (often also a cheap one) mounted at the imaging device. In fact modern video cameras have gyroscopes to measure jiggling of the camera. These sensors can also be used to measure the orientation of the camera.

For the above mentioned broad field of applications the new technique simply uses the available information about the extrinsic camera parameters. The known rotation will lead to a linear self-calibration that requires a lower number of camera images than required for the previously introduced approaches.

If the rotation of a camera is measured by a sensor that is mounted on the camera or by any external sensor the coordinate systems of the sensor and the camera are generally not identical. Hence a coordinate transformation between the sensor coordinate system and the camera coordinate system has to be applied to use the measured information in the camera coordinate system. This transformation is called "hand-eye-calibration". Several solutions for this problem have been proposed in the robotics research, for instance [TL89, HD95, Dan99]. In the following a known "hand-eye-calibration" is assumed or is computed beforehand using one

of the above approaches. Further, due to the miscellaneous delays during the data recording and the different sampling rates of the camera and the orientation data in a real system, the data has to be aligned in time, too. In [CI02, FK03b] techniques were proposed to do this alignment using the data itself. Nowadays there are some orientation sensors available that have a hardware solution for the alignment in time which promises to be more precise than the estimated alignments. For these reasons it is assumed that the alignment in time between orientation data and camera data is given.

In the next section 4.1 the novel linear self-calibration technique for a purely rotating camera will be presented. Some basic measurements of the robustness¹ of the new approach will be discussed in sections 4.1.4 to 4.1.6. The relations between the novel approach and the existing previous approaches for purely rotating cameras from chapter 3 will be discussed in the section 4.1.7. Afterwards a novel linear self-calibration approach for a freely moving camera will be introduced in section 4.2. The noise robustness of this approach is measured in sections 4.2.3 and 4.2.4. In sections 4.2.5 respectively 4.2.6 the novel approach will be compared to previous approaches for self-calibration of a freely moving camera. Finally in section 4.3 the critical rotations for the new approaches will be analyzed.

4.1 Linear self-calibration of a rotating camera with known orientation

In this section a linear self-calibration approach for a purely rotating camera with known orientation will be introduced. This linear approach exploits given rotation information to reduce the number of images required for self-calibration and to overcome the limitation on the number of varying intrinsics of the general self-calibration as in Agapitos approach from [dAHR98], explained in section 3.2.2 (see also section 2.8 and section 2.9). Moreover the problems of the existing approaches [dAHR01, Har94b, Ste95] caused by noise that lead to the fact that sometimes the calibration can't be computed also will be overcome by the novel approach. At first the general case of a camera with fully varying intrinsic parameters will be investigated. Varying camera parameters occur for zooming cameras and for multi-camera systems since then normally the intrinsic parameters can't be assumed to be constant. Afterwards the specialization for a rotating camera with fixed internal parameters is discussed. A basic evaluation of the linear self-calibration technique on synthetic point data is done in subsection 4.1.4 and the influence of errors in the measured rotation is analyzed analytically in section 4.1.6.

4.1.1 Self-Calibration with fully varying intrinsic parameters

The image planes of two cameras with the same optical center are correlated by a homography as explained in section 2.5.1. Therefore each point from camera j can be mapped to the image plane of camera i using the homography $H_{j,i}^\infty$ that maps over the plane at infinity. The structure of the infinite homography $H_{j,i}^\infty$ is given in equation (2.16). According to the structure of the

¹Robustness measures the sensitivity of the estimated result to noise in the input data.

homography, the following equation can be deduced from (2.16):

$$K_i R_{j,i} = \rho_{j,i} H_{j,i}^\infty K_j. \quad (4.1)$$

In equation (4.1) the projective scale $\rho_{j,i}$ is explicitly modeled to achieve an Euclidian equality in contrast to the equality up to scale in (2.16). In the following Euclidian equality will be needed for the computations and introduces one additional parameter - the scale of the homography - for each image pair. An important property of (4.1) is that it is linear in the calibration matrices K_j and K_i for known scale $\rho_{j,i}$. The contained relative rotation matrix $R_{j,i}$ between camera j and camera i can be computed from the known orientations of the cameras. The infinite homography $H_{j,i}^\infty$ between the cameras can be estimated from the images themselves (see section 2.5.1). For these reasons the homography $H_{j,i}^\infty$ and the rotation $R_{j,i}$ in the following will be used as known transformations. The scale $\rho_{j,i}$ of the homography $H_{j,i}^\infty$ in (4.1) is normally unknown as discussed in section 2.5.1. Hence the remaining unknowns in (4.1) are the intrinsics in K_i , K_j and the unknown scale $\rho_{j,i}$. By introducing a scaled calibration matrix \tilde{K}_i for the second camera

$$\tilde{K}_i = \frac{1}{\rho_{j,i}} K_i, \quad (4.2)$$

equation (4.1) becomes linear in K_j and \tilde{K}_i even for unknown scale $\rho_{j,i}$ exploiting the known homography $H_{j,i}^\infty$ and the known relative rotation $R_{j,i}$. The structure of the calibration matrix \tilde{K}_i from (2.3) provides a simple method to compute the scale $\frac{1}{\rho_{j,i}}$ directly from \tilde{K}_i because it is equal to the last diagonal element of \tilde{K}_i . Accordingly the calibration matrix K_i can be computed directly from \tilde{K}_i by dividing it by its last diagonal element. Therefore the scaled calibration matrix \tilde{K}_i is used in the following instead of using the calibration matrix K_i and the scale $\frac{1}{\rho_{j,i}}$ separately. Using (4.2) in (4.1) leads to

$$\tilde{K}_i R_{j,i} = H_{j,i}^\infty K_j \quad (4.3)$$

which is linear in the intrinsics of the camera j and in the elements of the scaled calibration matrix \tilde{K}_i . Apart from some degenerate cases equation (4.1) provides nine linear independent equations (B.1) for each a camera pair. The nine calibration equations (B.1) resulting from equation (4.3) are given in appendix B.1. An analysis of the linear independency for a general rotation $R_{j,i}$ is given in appendix B.1 and the degenerate cases will be investigated in section 4.3. The unknowns in (4.3) are the five intrinsics contained in K_j , the five intrinsics of K_i and the scale $\rho_{j,i}$ contained in \tilde{K}_i . This is a total of 11 unknowns in the nine equations. If there are no constraints available for the intrinsics of the cameras, (4.3) has no unique solution for a single pair of cameras j and i .

With two additional constraints for the intrinsics and the scale $\rho_{j,i}$ of the homography the solution becomes unique, since nine unknown parameters have to be determined and equation (4.3) still provides nine linear independent constraints.

A useful constraint to get a unique solution for the self-calibration from a single camera pair is the assumption of zero skew. This is often used for CCD-cameras which fulfill this naturally. It provides one constraint for each camera. The remaining set of nine equations (B.7) is still

linearly independent. A proof is given in appendix B.2. Accordingly the self-calibration can be computed from a single image pair with known relative rotation.

For a third camera k with given rotation $R_{j,k}$ between camera j and camera k nine additional linear independent equations are provided by equation (4.3) together with the homography $H_{j,k}^\infty$. The resulting two sets of linear independent equations can be used to calibrate all three cameras concurrently with a linear calibration approach. The camera triplet (j, i, k) with estimated homographies $H_{j,i}^\infty$ and $H_{j,k}^\infty$ together with the known rotations $R_{j,i}$ and $R_{j,k}$ delivers with (4.3) a set of equations

$$\begin{aligned}\tilde{K}_i R_{j,i} &= H_{j,i}^\infty K_j, \\ \tilde{K}_k R_{j,k} &= H_{j,k}^\infty K_j.\end{aligned}\tag{4.4}$$

The camera triplet has 17 unknowns and (4.4) provides at most 17 independent equations. That a third camera provides additional linear independent equation can be deduced from the block structure of the linear equation system (B.2). Following with general relative rotations $R_{j,i}$ and $R_{j,k}$ a unique solution can be computed for the intrinsics and scales $\rho_{j,i}$ and $\rho_{j,k}$ of a camera triplet (j, i, k) . The self-calibration can be solved even for fully varying intrinsic camera parameters of all concerned cameras. Camera j can be seen as a reference view for this calibration since both homographies refer to it.

At a first glance it seems that introducing a reference camera j is a limitation, which as matter of fact is not. Assuming to know the homographies $H_{j,i}^\infty$ and $H_{j,k}^\infty$ in (4.4) does not require the computation of both homographies directly. If only pairwise homographies can be computed, the homography $H_{j,k}^\infty$ can be calculated from the pairwise homographies $H_{j,i}^\infty$ and $H_{i,k}^\infty$. The homography $H_{j,k}^\infty$ to the reference camera is then given by the concatenation of $H_{j,i}^\infty$ and $H_{i,k}^\infty$ as $H_{j,k}^\infty = H_{j,i}^\infty H_{i,k}^\infty$.

The next section will discuss the special case of constant intrinsic parameters for a purely rotating camera.

4.1.2 Self-Calibration with constant intrinsic parameters

In many applications a camera with constant intrinsic parameters is used for image capture. For example in surveillance applications cameras with constant intrinsics over a certain time period are used. If it is known that the intrinsics are constant for all frames or a sequence of consecutive frames, this knowledge provides five constraints for each additional camera. Accordingly only the five intrinsic parameters of the first camera remain as unknown intrinsic parameters. For cameras with constant intrinsics the self-calibration can therefore be computed from a single camera pair. The remainder of this section shows how to detect the constant intrinsics followed by the deduction of the simplified self-calibration problem.

In order to exploit the five additional constraints provided by constant intrinsic camera parameters it is useful to detect this constancy automatically. For constant intrinsic camera parameters K according to (2.16) the homography $H_{j,i}^\infty$ can be written as

$$H_{j,i}^\infty = K R_{j,i} K^{-1}.\tag{4.5}$$

Hence the homography $H_{j,i}^\infty$ is a conjugated rotation matrix². The structure of the eigenvalues of a rotation matrix [Pea83, Fra83, HZ03] is known to be

$$\text{eigval}(R) = [1, \cos(\phi) + i \sin(\phi), \cos(\phi) - i \sin(\phi)], \quad (4.6)$$

where ϕ is the angle of rotation about the rotation axis given by the eigenvector corresponding to the eigenvalue one as earlier explained in 2.6. It is also known that a scaled matrix has the scaled eigenvalues of the original matrix [Pea83]. Consequently the eigenvalues of the scaled homography $\rho_{i,j}H_{j,i}^\infty$ have the structure

$$\text{eigval}(H_{j,i}^\infty) = \rho_{i,j}[1, \cos(\phi) + i \sin(\phi), \cos(\phi) - i \sin(\phi)]. \quad (4.7)$$

It follows with (4.7) that the scale $\rho_{i,j}$ is equivalent to the absolute value of each eigenvalue of $H_{j,i}^\infty$

$$\rho_{i,j} = |\rho_{i,j} \cdot 1| = |\rho_{i,j} (\cos(\phi) + i \sin(\phi))| = |\rho_{i,j} (\cos(\phi) - i \sin(\phi))|.$$

Exploiting the structure of the eigenvalues of the rotation matrix (4.6) with (4.7) it can be decided whether the camera calibration was constant or not. This decision uses the fact that for a camera with varying intrinsics the eigenvalues do not have the structure as shown in (4.7).

How to exploit these five additional constraints on the intrinsics and the known scale will be discussed in the following. Equation (4.1) for a camera pair with constant intrinsic parameters and the known scale $\rho_{j,i}$ is equal to

$$KR_{j,i} = H_{j,i}^\infty K. \quad (4.8)$$

This is linear in the five intrinsics of the cameras j and i . The resulting calibration equations are given in appendix B.3.

From a counting argument it follows that the solution of (4.8) is now unique for a single camera pair, because there are only five unknown intrinsic camera parameters left. Equation (4.8) provides a set of nine self-calibration equations (B.10) whereof five are linear independent. The linear independency of the provided equations (B.10) is shown in B.3. Due to the five linear independent equations delivered by (4.8) it follows that the calibration can be estimated from a single camera pair with constant intrinsic parameters and known orientation.

So far the novel linear self-calibration equations were developed by exploiting the structure of the homography. The next section will show that the self-calibration with the dual image of the absolute conic (DIAC) provides the same calibration equations.

4.1.3 Relation of the linear self-calibration to the general self-calibration

The above introduced novel self-calibration approach used the structure of the homography to deduce the self-calibration equations for the intrinsic camera parameters. In this section it will

²Matrices A and B are conjugated if $A = CBC^{-1}$ for some invertible matrix C . The conjugated matrix A has the same eigenvalues as B and the eigenvectors of B are transformed with C [GL89].

be shown that using the derivation of the the DIAC (as discussed earlier in section 2.8) will provide the same self-calibration equations as extracted in (4.3).

Without loss of generality the camera center of a purely rotating camera can be assumed to be at the origin of the coordinate system. The last column of each projection matrix is accordingly zero for all cameras and the projection matrix has the form $P_i = [M_i|0]$. Hence for a purely rotating camera equation (2.41) can be written as

$$K_i R_i^T \cong \rho_i (M_i - e_i v^T) K_1$$

with the camera center at the origin this is equal to

$$\tilde{K}_i R_i^T = M_i K_1, \quad (4.9)$$

with \tilde{K}_i defined in (4.2) as the scaled camera calibration matrix of camera i . For the self-calibration using the DIAC like [dAHR01, Har95] without known rotation this equation is transformed into a rotation invariant equation (2.43). The transformation used the orthogonality $RR^T = I_{3 \times 3}$ of the rotation. Afterwards the structure of the DIAC is used to compute the camera calibration with the Cholesky decomposition. For known orientation of the cameras the transfer to a rotation invariant problem is not necessary. The sub-matrix M_i of the projection matrix P_i is the mapping from points on the image plane of camera i into the first cameras image plane, which is chosen as the world coordinate system. Hence the sub-matrix M_1 is equal to $I_{3 \times 3}$. Therefore (4.9) is

$$\tilde{K}_i R_i^T = M_i M_1^{-1} K_1 = H_{1,i}^\infty K_1, \quad (4.10)$$

where $H_{1,i}^\infty$ represents the homography that maps the points from the image plane of the first image onto the image plane of the i -th camera over the plane at infinity π_∞ . Equation (4.10) is equal to the novel self-calibration approach (4.3) deduced using the structure of the homography $H_{1,i}^\infty$. It provides the same set of linear constraints for the intrinsics of the cameras.

The novel self-calibration equation (4.10) can be transformed into an equation for the dual image of the absolute conic

$$\tilde{K}_i \tilde{K}_i^T = \omega_i^* \cong H_{1,i}^\infty K_1 K_1^T (H_{1,i}^\infty)^T \quad (4.11)$$

by multiplying each side by its transposed and using the equality from (4.3). This transformation can be seen as a linear combination of the constraints provided by (4.10). Equation (4.11) is equivalent to the infinite homography constraint (2.44) introduced in section 2.8. In contrast to the novel approach the infinite homography constraint does not exploit the known rotation.

Hartley proposed in [Har94b] to use (4.11) to estimate the DIAC and afterwards to apply the Cholesky decomposition to calculate the camera calibration. Moreover the Cholesky decomposition is used to invert the linear combination of the constraints from (4.10) in (4.11). Employing the Cholesky decomposition introduces the problems with the positive definiteness of the dual image of the absolute conic as earlier explained (see section 3.1.1). The proposed novel approach overcomes these problems because it computes the intrinsics directly from (4.10) respectively (4.3).

So far the novel self-calibration equations were introduced and the number of constraints provided by these equations was shown. The next section shows measurements of the robustness in dependence on the noise in the camera orientation and the noise of the correspondences used to compute the homographies.

4.1.4 Evaluation for rotating cameras

The robustness of the proposed calibration in presence of noise in the data will be shown by the measurements in this section. To measure the noise robustness of the calibration of a rotating camera with fully varying internal parameters the above linear approach will be tested with synthetic point data and a detailed analysis of the results will be given. Tests of the proposed technique with real data will be shown in section 5.1.

A camera rotating about its optical center at the origin of the coordinate system is simulated for the test with synthetic point data. The camera rotates about x -axis and y -axis up to six degrees. Rotation about the optical axis was not simulated since it does not provide any information for the estimation of the focal length and the aspect ratio of the camera. For the simulation a uniformly distributed point cloud of 100 points in front of the camera was projected into the image plane of the rotating camera. Figure 4.1 shows a sketch of the setup. For all computations six different views with different internal camera parameters are used. The 3D scene points \bar{X} are uniformly distributed in a cube and projected into image points \bar{x} in the camera images of size 512x512 pixel.

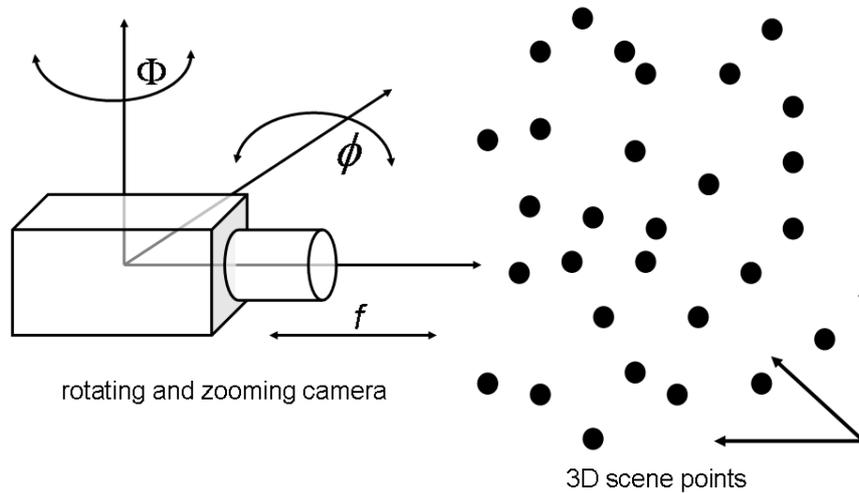


Figure 4.1: Synthetic scene for measuring the noise robustness of the linear calibration

The locations \bar{x} of the projected points are disturbed by Gaussian noise with a maximum standard deviation σ_{pixel} of up to one pixel. The noise level was chosen because the often used

corner detectors [HS88, Smi92, Smi96] or feature tracker [LK81] reach that precision [SM95]. Additionally the known orientation of the camera is also disturbed by Gaussian angular noise with a standard deviation σ_{angular} of up to one degree for each axis of the coordinate system, which is approximately the range of angular noise of consumer rotation sensors like the InterSense InertiaCube² and the XSense MTB 9. For higher noise levels the calibration was tested in [FK03c, FK03b]. The estimation was repeated 1000 times for each noise level in order to estimate the standard deviation and the mean value of the intrinsic parameters.

During the measurements the standard deviation σ_{pixel} of the pixel noise and standard deviation σ_{angular} of the rotational noise were both varied. The homographies $\tilde{H}_{1,i}^\infty$ with $i = 2, \dots, 6$ between the first camera and all other cameras are estimated from point correspondences by least squares estimation using the linear estimation algorithms of section 2.5.1. To achieve a better condition number³ of the resulting linear equation system and reliable calibrations a normalization as proposed in [Har95] was employed. Applying this normalization to the data reduced the condition number of the system matrix in (B.2), for example in this setup the condition number was reduced by four orders of magnitude.

The measurements of relative errors and the standard deviations of these relative errors for the first camera with the camera calibration matrix K_1 set as

$$K_1 = \begin{bmatrix} 415 & 166 & 240.64 \\ 0 & 456.5 & 245.76 \\ 0 & 0 & 1 \end{bmatrix}$$

are shown in figure 4.2 to figure 4.6. The measured errors and standard deviations of the other cameras are comparable to these results.

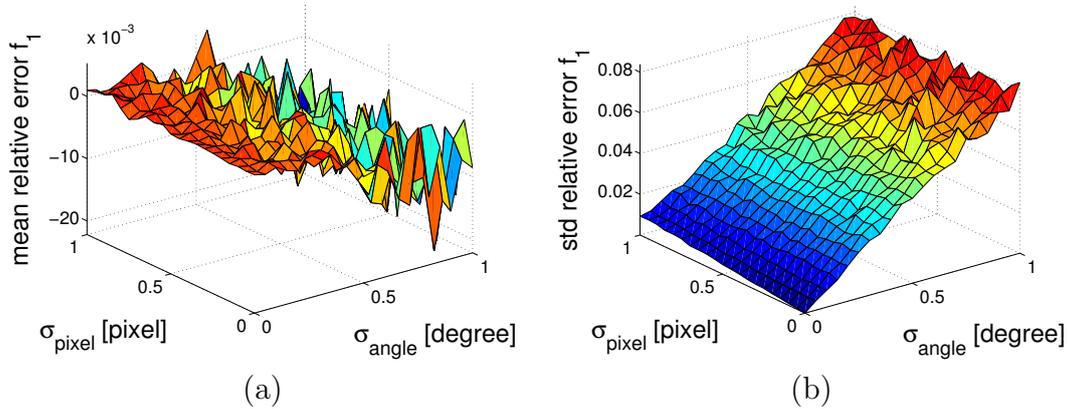


Figure 4.2: Noise robustness measurements with Gaussian angular noise of a standard deviation of up to 1 degree and Gaussian pixel noise with a standard deviation of up to one pixel of the corresponding points: (a) relative error of the mean of the estimated focal length $f_1 = 415$, (b) standard deviation of the relative error of the estimated focal length f_1 .

From figure 4.2 (a) it can be seen that the mean relative error of the estimated focal length is robust against noise in the data for the tested range of Gaussian noise with standard deviations of $\sigma_{\text{pixel}} \in [0, 1]$ and $\sigma_{\text{angular}} \in [0^\circ, 1^\circ]$. The relative error is always smaller than 2.8%. A slight

³The condition number is the ratio of the largest singular value of a matrix to its smallest. Large condition numbers indicate a nearly singular matrix that causes numerical problems for the inversion of the matrix.

systematic decrease of the error (bias⁴) of the estimated focal length can be observed. It is a result of the angular noise of the rotation about the optical axis. A discussion of the bias will be given in section 4.1.5. In figure 4.2 (b) it can be seen that the standard deviation of the relative error increases mainly depending on the standard deviation σ_{angular} of the angular noise. The influence of the noise of the location of the corresponding pixels is much smaller because the homographies $\tilde{H}_{1,i}^\infty$ with $i = 2, \dots, 6$ can still be estimated very precisely.

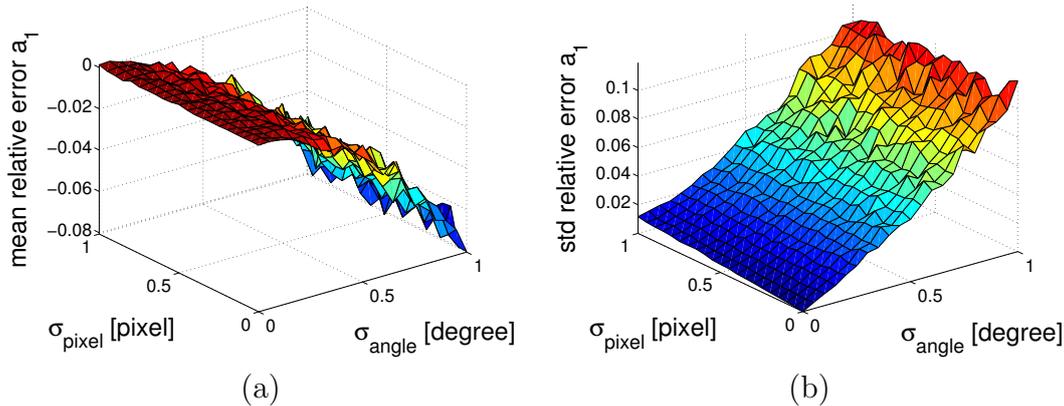


Figure 4.3: Noise robustness measurements with Gaussian angular noise of a standard deviation of up to 1 degree and Gaussian pixel noise with a standard deviation of up to one pixel of the corresponding points: (a) relative error of the mean of the estimated aspect ratio $a_1 = 1.1$, (b) standard deviation of the relative error of the estimated aspect ratio a_1 .

In figure 4.3 (a) the mean relative error of the aspect ratio is shown. It can be seen that the estimation is rather stable for a standard deviation of the pixel noise $\sigma_{\text{pixel}} \in [0, 1]$ and a standard deviation of the angular noise of approximately $\sigma_{\text{angular}} \in [0, 0.5^\circ]$. Then the relative error of the mean estimated aspect ratio is always lower than 2.3%. According to the higher sensitivity to angular noise the estimation of the aspect ratio requires a higher accuracy of the orientation sensor than the focal length estimation. The requirement of higher accuracies results from the dependence of the error of the aspect ratio a on the error of the focal length f and on the product of the focal length and the aspect ratio af . The consumer orientation sensors like InterSense InertiaCube² or XSense MTB 9 provide the mandatory precisions for differential measurements of less than half a degree. Similar to the focal length the mean value of the aspect ratio shows a bias, which leads to an underestimation of the aspect ratio (see figure 4.3 (a)). The reason for this bias is the same as for the focal length, namely the angular noise of the rotation about the optical axis. The standard deviation of the relative error slightly increases with the angular noise similar to the focal length. Again due to the precise homography estimation the relative error depends mainly on the angular noise.

The mean relative error of the skew of the camera in figure 4.4 shows similar tendencies as the mean relative error of the estimated aspect ratio and focal length. For angular noise with a standard deviation $\sigma_{\text{angular}} \in [0, 0.5^\circ]$ the mean relative error is less than 2.7%. For larger standard deviations of up to 1° of the angular noise the error increases up to 8.7%. Similar

⁴A bias is a false association that results from the failure to account for some skewing or influencing factor, or a tendency for the observed results to deviate from the ground truth results. Bias distorts results in a particular direction.

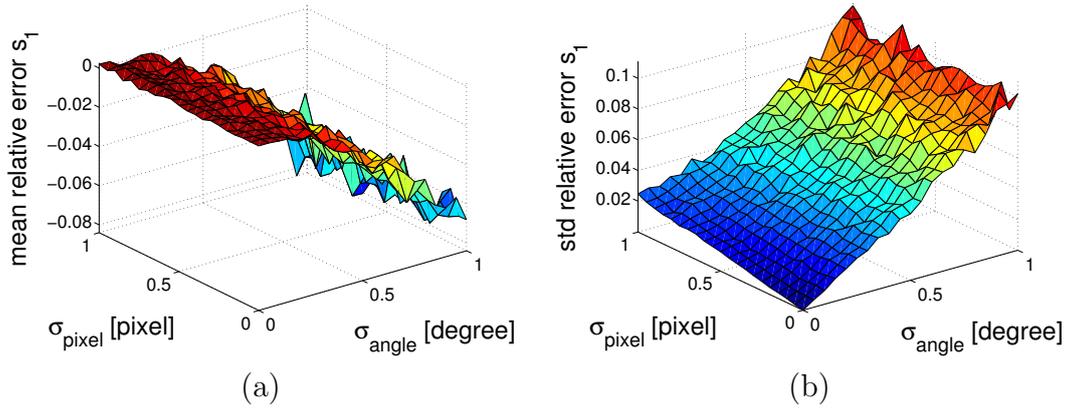


Figure 4.4: Noise robustness measurements with Gaussian angular noise of a standard deviation of up to 1 degree and Gaussian pixel noise with a standard deviation of up to one pixel of the corresponding points: (a) relative error of the mean of the estimated skew $s_1 = 166$, (b) standard deviation of the relative error of the estimated skew s_1 .

to focal length and aspect ratio the mean estimated value shows a bias caused by the angular noise of the rotation about the optical axis. As a result of the poorly conditioned estimation problem of the skew and the principal point, the standard deviation of the estimated skew s increases more than the standard deviation of the focal length in figure 4.2.

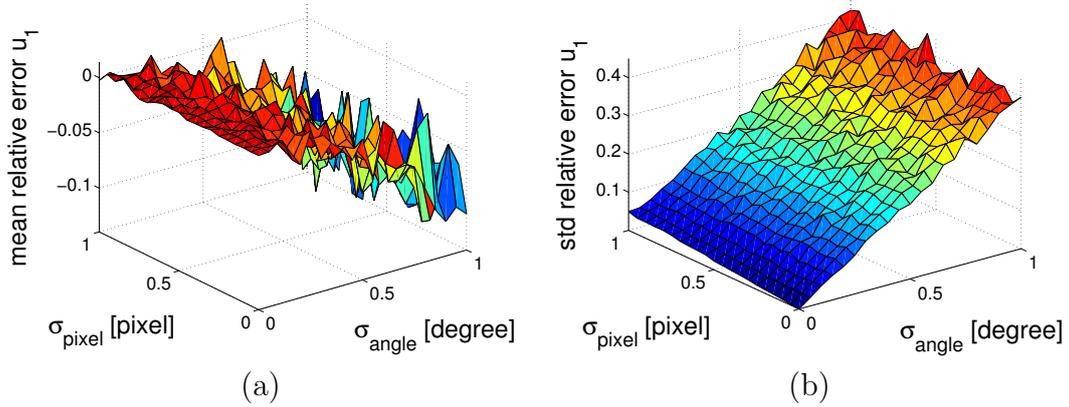


Figure 4.5: Noise robustness measurements with Gaussian angular noise of a standard deviation of up to 1 degree and Gaussian pixel noise with a standard deviation of up to one pixel of the corresponding points: (a) relative error of the mean of the estimated x -component of the principal point $u_1 = 240.64$, (b) standard deviation of the relative error of the estimated x -component of the principal point u_1 .

The results for the estimation of the principal point (u_1, v_1) are shown in figure 4.5 and figure 4.6. It can be seen that the estimation for both components is rather stable for a standard deviation of the angular noise of $\sigma_{\text{angular}} \in [0, 0.5^\circ]$. The mean relative error for these measurements is always less than 5.5% for u_1 and less than 0.5% for v_1 . The measurement for the y -component is even more stable for larger standard deviations of the angular noise. For the x -component again the bias caused by the angular noise of the rotation about the optical axis is observable. The higher sensitivity to noise of the x -component is caused from the joint estimation of this parameter with the skew which can be deduced from the self-calibration equations (B.1).

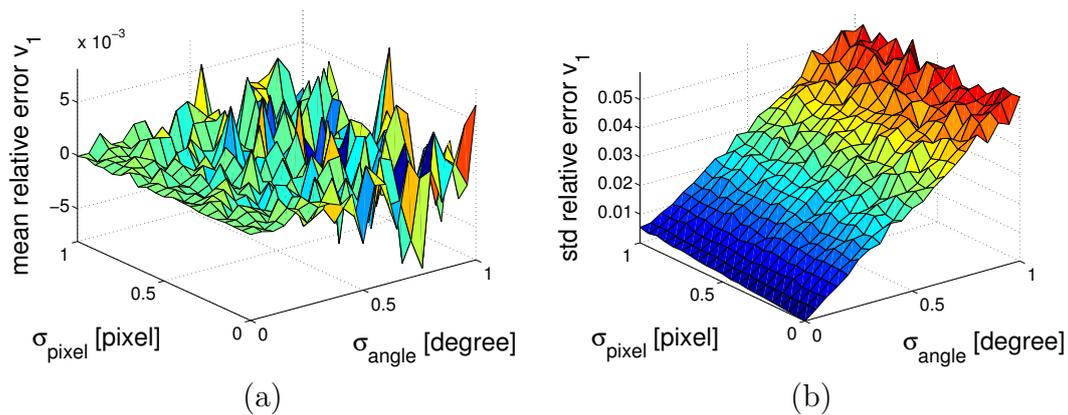


Figure 4.6: Noise robustness measurements with Gaussian angular noise of a standard deviation of up to 1 degree and Gaussian pixel noise with a standard deviation of up to one pixel of the corresponding points: (a) relative error of the mean of the estimated y -component of the principal point $v_1 = 245.76$, (b) standard deviation of the relative error of the estimated y -component of the principal point v_1 .

So far it was seen from the measurements that the influence of the orientation noise is much larger than the error caused from noisy correspondences, since the absolute rotation angle between the cameras is in the range of the noise (six degree camera rotation with standard deviation of up to one degree for the noise) and the homographies still are estimated precisely. The error of orientation sensors like the InertiaCube² from InterSense and the MTB 9 from XSense is normally in the range below half a degree for differential rotations. Hence a reliable calibration can be computed from these measured orientations using the proposed self-calibration approach. In most situations the homography estimation provides mapping errors of less than one pixel. Accordingly the proposed calibration with (4.4) is robust.

As seen in the measurements for the focal length, the aspect ratio, and the skew the above estimation is biased. The next section will deduce the source of this bias.

4.1.5 Error in rotations about the coordinate axes

The above shown (figure 4.2 to 4.6) robustness measurements had a bias for most of the intrinsic parameters. It can be seen that the standard deviation σ_{angular} of the angular noise has the major impact on the bias. Accordingly this impact will be analyzed in this section.

In the following the influence of the noise of the rotations about the axes x , y and z of the coordinate system will be investigated for each axis separately. The first step of the analysis will be performed for two cameras of the above mentioned synthetic setup. These two cameras have a general relative rotation (about x -axis and y -axis) between each other and observe a point cloud of 100 points in front of both cameras. In contrast to the above tests all camera parameters except those of the focal lengths were assumed to be known during the computation. The homography between both images was again estimated from point correspondences using least squares. During the tests the known camera orientation was only disturbed in one direction at a time by Gaussian noise with standard deviation σ_{angular} of up to 3° . For each noise level one thousand tests were made and the mean estimated focal length of each noise level was

computed. The computed mean focal lengths of the cameras are plotted against the standard deviation of the angular noise in figure 4.7 (a) to (c).

It can be seen that the angular noise of a rotation about the x -axis doesn't influence the mean value of the estimated focal length. An error of a rotation about the y -axis only slightly influences the estimated focal length. The major source of error seems to be the error of the rotation about the z -axis (optical axis of the camera). It will be analyzed in detail in the following.

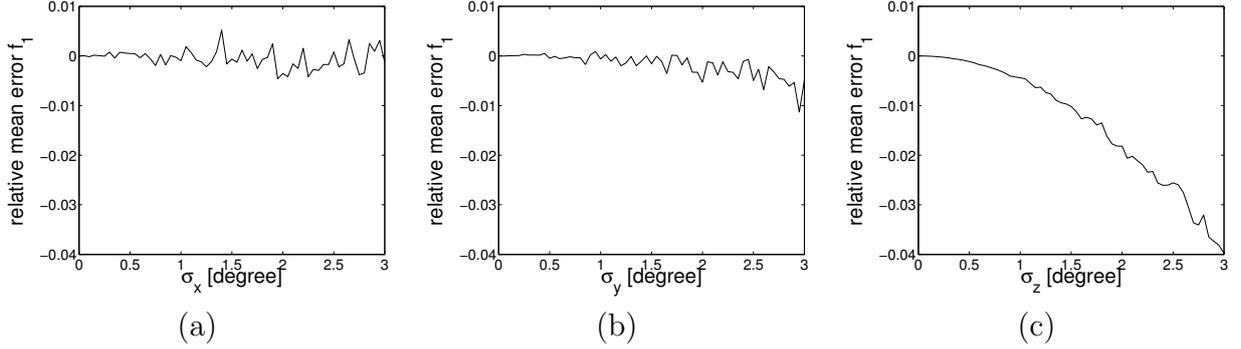


Figure 4.7: Relative error in the focal length depending on the standard deviation σ_x , σ_y and σ_z of the angular noise in degree of the given rotation. (a) angular noise for the rotation about the x -axis, (b) angular noise for the rotation about the y -axis, (c) angular noise for the rotation about the z -axis.

To deduce the noise sensitivity the estimated focal length from two views of a purely rotating camera is investigated. To show the influence of angular noise of the rotation about the optical axis on the focal length estimation all intrinsic parameters of the cameras are assumed to be zero except those of the focal length and the aspect ratio of one (square pixels). Accordingly the calibration can be computed from an image pair with known orientation. The true focal lengths f_j and f_i of the two cameras provide the calibration matrices of the two views

$$K_j = \begin{bmatrix} f_j & 0 & 0 \\ 0 & f_j & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad K_i = \begin{bmatrix} f_i & 0 & 0 \\ 0 & f_i & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Exploiting the structure (2.16) of the infinite homography $H_{1,2}^\infty$ between both camera image planes gives

$$H_{j,i}^\infty = K_i R_z K_j^{-1} = \begin{bmatrix} \frac{f_i}{f_j} \cos \Theta & -\frac{f_i}{f_j} \sin \Theta & 0 \\ \frac{f_i}{f_j} \sin \Theta & \frac{f_i}{f_j} \cos \Theta & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad (4.12)$$

where R_z is the rotation matrix of a rotation about the optical axis with an angle Θ . In the following it is assumed that there is no rotation about the x -axis and the y -axis.

The above described self-calibration approach (4.3) computes the intrinsic parameters con-

tained in the estimated calibration matrices \hat{K}_j and \hat{K}_i

$$\hat{K}_j = \begin{bmatrix} \hat{f}_j & 0 & 0 \\ 0 & \hat{f}_j & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad \hat{K}_i = \begin{bmatrix} \hat{f}_i & 0 & 0 \\ 0 & \hat{f}_i & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

In presence of noise rotation matrix $\hat{R}_{1,2}$ of the disturbed rotation is given by

$$\hat{R}_{1,2} = \begin{bmatrix} \cos(\Theta + \Delta\Theta) & -\sin(\Theta + \Delta\Theta) & 0 \\ \sin(\Theta + \Delta\Theta) & \cos(\Theta + \Delta\Theta) & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

where $\Delta\Theta$ is the angular noise added to the true rotation angle Θ of the rotation about the optical axis of the camera. The linear equation system (B.2) for the computation of the intrinsic camera calibration leads to the following estimation problem

$$\begin{bmatrix} (H_{j,i}^\infty)_{1,1} & -\cos(\Theta + \Delta\Theta) \\ (H_{j,i}^\infty)_{2,2} & -\cos(\Theta + \Delta\Theta) \\ (H_{j,i}^\infty)_{1,2} & -\sin(\Theta + \Delta\Theta) \\ (H_{j,i}^\infty)_{2,1} & \sin(\Theta + \Delta\Theta) \end{bmatrix} \begin{pmatrix} \hat{f}_j \\ \hat{f}_i \end{pmatrix} = \underbrace{\begin{bmatrix} \frac{f_i}{f_j} \cos \Theta & -\cos(\Theta + \Delta\Theta) \\ \frac{f_i}{f_j} \cos \Theta & -\cos(\Theta + \Delta\Theta) \\ \frac{f_i}{f_j} \sin \Theta & -\sin(\Theta + \Delta\Theta) \\ -\frac{f_i}{f_j} \sin \Theta & \sin(\Theta + \Delta\Theta) \end{bmatrix}}_A \begin{pmatrix} \hat{f}_j \\ \hat{f}_i \end{pmatrix} = 0. \quad (4.13)$$

It is an overdetermined estimation problem which can be solved using least squares. The least squares solution of (4.13) is according to [GL89] the minimum of

$$\begin{aligned} & \begin{pmatrix} \hat{f}_j \\ \hat{f}_i \end{pmatrix}^T A^T A \begin{pmatrix} \hat{f}_j \\ \hat{f}_i \end{pmatrix} = \epsilon \\ & \hat{f}_j^2 \left(\frac{f_i}{f_j} \right)^2 - 2\hat{f}_j \hat{f}_i \left(\frac{f_i}{f_j} \right) \cos(-\Delta\Theta) + \hat{f}_i^2 = \epsilon. \end{aligned} \quad (4.14)$$

The derivations according to \hat{f}_j respectively to \hat{f}_i are leading to

$$\begin{aligned} \hat{f}_j &= \hat{f}_i \left(\frac{f_j}{f_i} \right) \cos(-\Delta\Theta) = f_j \left(\frac{\hat{f}_i \cos(-\Delta\Theta)}{f_i} \right) \\ \hat{f}_i &= \hat{f}_j \left(\frac{f_i}{f_j} \right) \cos(-\Delta\Theta) = f_i \left(\frac{\hat{f}_j \cos(-\Delta\Theta)}{f_j} \right) \end{aligned} \quad (4.15)$$

as the minimum of (4.14) for \hat{f}_j respectively \hat{f}_i . Consequently the mean value \bar{f}_j respectively \bar{f}_i of several estimations of \hat{f}_j respectively \hat{f}_i using a rotation about the optical axis that is disturbed by Gaussian noise is given as

$$\begin{aligned} \bar{f}_j &= \frac{f_j \hat{f}_i}{|M| f_i} \sum_{m \in M} \cos(-\Delta\Theta^m) \\ \bar{f}_i &= \frac{f_i \hat{f}_j}{|M| f_j} \sum_{m \in M} \cos(-\Delta\Theta^m), \end{aligned} \quad (4.16)$$

where M is the set of measurements, $|M|$ delivers the number of elements of M and $\Delta\Theta^m$ is the angular disturbance of the m -th measurement. It can be seen from (4.16) that the mean value \bar{f}_j contains a systematic error factor $\epsilon_{bias} = \sum_{m \in M} \hat{f}_i \cos(-\Delta\Theta^m)$. In absence of noise the factor ϵ_{bias} is equal to one and has no influence. If noise is present it is always less than one due to the symmetry of the cosine even if the distribution of the angular error $\Delta\Theta$ has zero mean. Hence the mean value \bar{f}_j respectively \bar{f}_i of the focal length is systematically underestimated in the presence of angular noise for the rotation about the optical axis. Further it can be seen that the bias in (4.16) depends on the error in the rotation angle Θ itself and not on the absolute value of the rotation. Hence the influence of the error does not depend on the signal to noise ratio. For an estimation of all camera intrinsics concurrently a similar analysis can be done, that shows that the aspect ratio of the camera is also biased by angular noise of the rotation about the optical axis.

From equation (4.16) it can be seen that a unique solution for \hat{f}_j and \hat{f}_i is not possible for a pure rotation about the optical axis. Accordingly this is a critical rotation for the proposed self-calibration method. A detailed investigation of critical rotations for the proposed method will be given in section 4.3. The above analysis showed the bias introduced by an error of the rotation measurement about the optical axis. To compute a unique solution for the self-calibration an additional general rotation is needed. In general a rotation about the x -axis or the y -axis won't compensate the error introduced by the measurement error in the rotation about the optical axis.

The test with synthetic point data from the previous section 4.1.4 was repeated to validate the results of the above analysis for the mean value of the estimated parameters, where the angular noise for the rotation about the optical axis was set to zero. The results for the robustness measurements are shown in the figures 4.8 to 4.12.

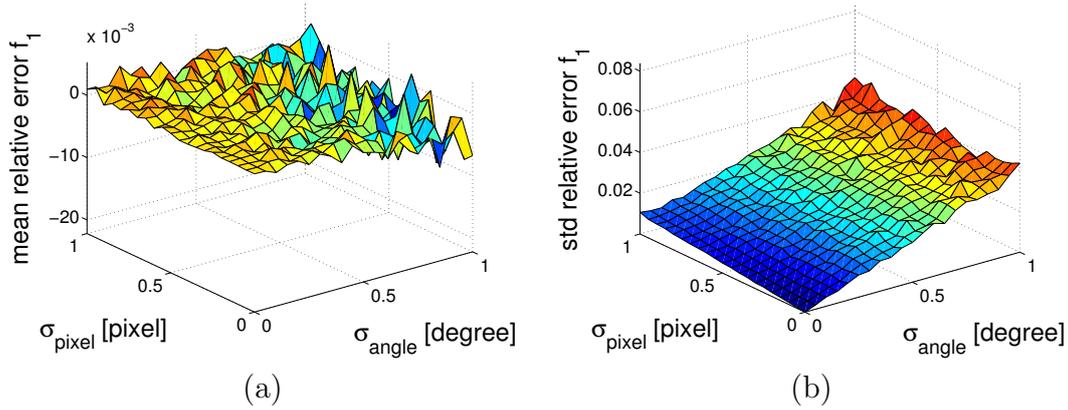


Figure 4.8: Noise robustness measurements with Gaussian distributed angular noise for rotation about the x -axis and y -axis of a standard deviation of up to 1 degree and Gaussian distributed pixel noise of the corresponding points of a standard deviation of up to one pixel: (a) relative error of the mean of the estimated focal length $f_1 = 415$, (b) standard deviation of the relative error of the estimated focal length f_1 .

It can be seen from figure 4.8 that the mean relative error of the focal length is strongly reduced compared to the measurement with angular noise for the rotation about the optical axis shown in figure 4.2. The axes have the same scale as in the original measurement 4.2.

According to the elimination of the bias the mean relative error of the focal length was reduced from 2.8% to 1%. As expected the variance of the results is also slightly reduced because the general noise level is lower according to only two noisy rotation angles instead of three.

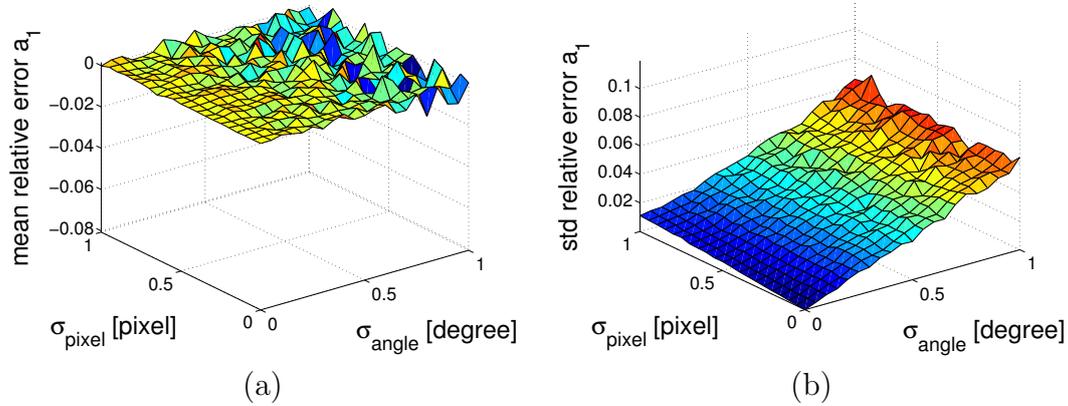


Figure 4.9: Noise robustness measurements with Gaussian distributed angular noise for rotation about the x -axis and y -axis of a standard deviation of up to 1 degree and Gaussian distributed pixel noise of the corresponding points of a standard deviation of up to one pixel: (a) relative error of the mean of the estimated aspect ratio $a_1 = 1.1$, (b) standard deviation of the relative error of the estimated aspect ratio a_1 .

The results for the estimation of the aspect ratio a are shown in figure 4.9. It can be seen that the mean relative error of the estimated aspect ratio is also reduced as result of the eliminated angular noise of the rotation about the optical axis. The maximal relative error is reduced from 2.3% to 0.6%.

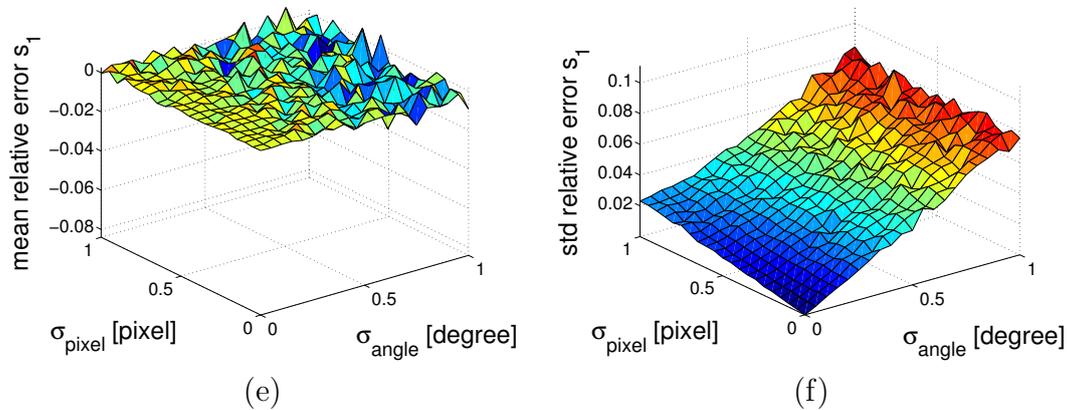


Figure 4.10: Noise robustness measurements with Gaussian distributed angular noise for rotation about the x -axis and y -axis of a standard deviation of up to 1 degree and Gaussian distributed pixel noise of the corresponding points of a standard deviation of up to one pixel: (a) relative error of the mean of the estimated skew $s_1 = 166$, (b) standard deviation of the relative error of the estimated skew s_1 .

The computation of the skew of the camera is also improved in absence of angular noise for the rotation about the optical axis as shown in figure 4.10. The highest mean relative error is reduced from 8.7% with angular noise to 2% without angular noise in z .

The measurements for the principal point are given in figure 4.11 and figure 4.12. The previously mainly disturbed x -component of the principal point u_1 is well estimated without

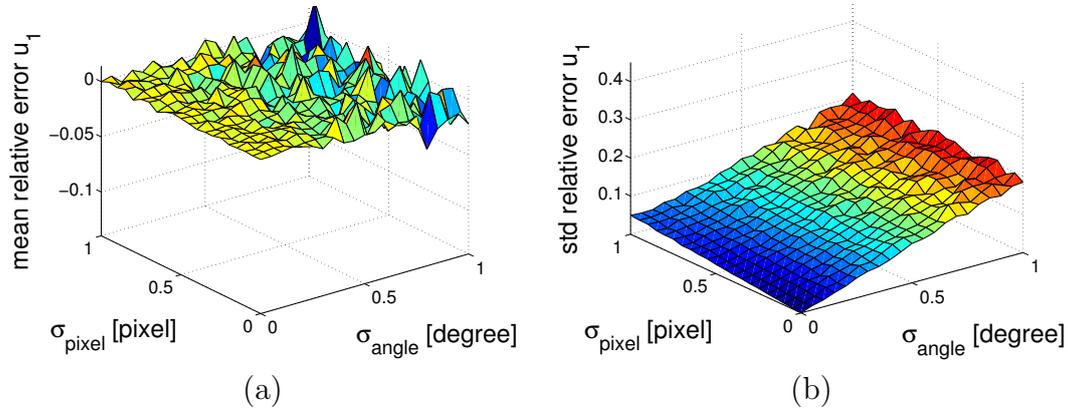


Figure 4.11: Noise robustness measurements with Gaussian distributed angular noise for rotation about the x -axis and y -axis of a standard deviation of up to 1 degree and Gaussian distributed pixel noise of the corresponding points of a standard deviation of up to one pixel: (a) relative error of the mean of the estimated x -component of the principal point $u_1 = 240.64$, (b) standard deviation of the relative error of the estimated x -component of the principal point u_1 .

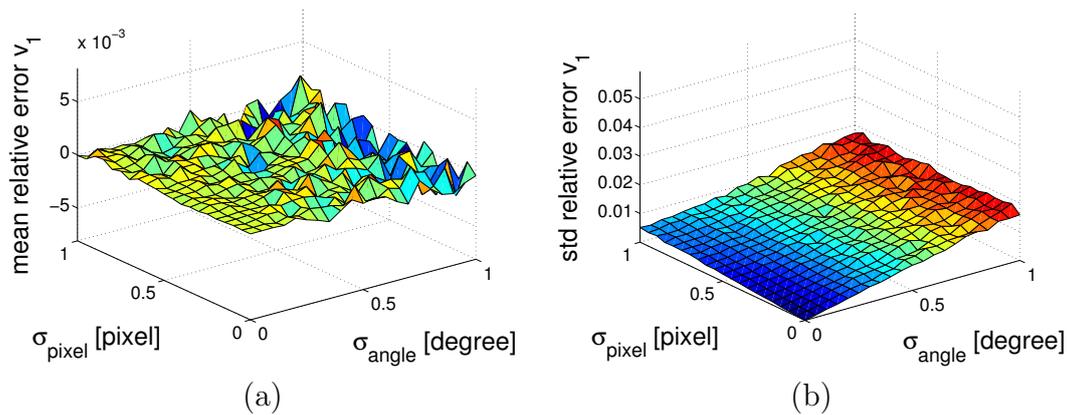


Figure 4.12: Noise robustness measurements with Gaussian distributed angular noise for rotation about the x -axis and y -axis of a standard deviation of up to 1 degree and Gaussian distributed pixel noise of the corresponding points of a standard deviation of up to one pixel: (a) relative error of the mean of the estimated y -component of the principal point $v_1 = 245.76$, (b) standard deviation of the relative error of the estimated y -component of the principal point v_1 .

angular noise. The maximal error is reduced to 1.3% compared to 3% for angular noise with a standard deviation of up to $\sigma_z \in [0, 0.5^\circ]$. The mean relative error for the y -component is reduced from 0.5% with angular noise in z to 0.1 % without angular noise.

So far it was shown that the uncertainty for the rotation angle about the optical axis influences the quality of the estimated camera calibration. The analytical deduction was proved with experiments on synthetic point data that showed that the angular noise introduced a bias into the estimated intrinsic camera parameters.

4.1.6 Influence of errors in rotation

As seen in the evaluation of the previous section the source of error with the biggest influence on the error of the estimated intrinsics is the still the rotation error even without noise in the rotation about the optical axis. Consequently further investigation on this error source will be done in this section. The remaining sources of rotation errors are the rotations about the x -axis and the y -axis. In the following the rotation $R_{i,j}$ between both cameras is set as a rotation about the x -axis and the y -axis for the following analysis

$$R_{j,i} = R_x R_y \text{ with } R_x(\phi) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \phi & -\sin \phi \\ 0 & \sin \phi & \cos \phi \end{bmatrix}, R_y(\Phi) = \begin{bmatrix} \cos \Phi & 0 & \sin \Phi \\ 0 & 1 & 0 \\ -\sin \Phi & 0 & \cos \Phi \end{bmatrix}. \quad (4.17)$$

The angle ϕ is the rotation angle of the rotation about the x -axis and the angle Φ is the rotation angle of the rotation about the y -axis. For simplification again a camera with constant and known parameters except from the focal length is assumed. Accordingly the focal length can be estimated from a single camera pair. Without loss of generality the camera calibration matrices K_j and K_i are given by

$$K_j = \begin{bmatrix} f_j & 0 & 0 \\ 0 & f_j & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad K_i = \begin{bmatrix} f_i & 0 & 0 \\ 0 & f_i & 0 \\ 0 & 0 & 1 \end{bmatrix}. \quad (4.18)$$

The homography $H_{j,i}^\infty$ between the two cameras j and i can be computed with (2.16) exploiting the camera calibration matrices K_j and K_i from (4.18) and the rotation $R_{j,i}$ from (4.17) by

$$H_{j,i}^\infty = K_i R_{j,i} K_j^{-1}. \quad (4.19)$$

This homography $H_{j,i}^\infty$ is error free and has scale $\rho_{j,i}$ equal to one. The measured rotation angles $\hat{\phi}$ and $\hat{\Phi}$ have a disturbance by angular errors $\Delta\phi$ respectively $\Delta\Phi$. Hence the disturbed rotation matrix $\hat{R}_{j,i}$ used for self-calibration is given by

$$\hat{R}_{j,i} = R_x(\phi + \Delta\phi) R_y(\Phi + \Delta\Phi), \quad (4.20)$$

with $R_x(\cdot)$ and $R_y(\cdot)$ from equation (4.17). Consequently the self-calibration equations (4.3) are equal to

$$\hat{K}_i \hat{R}_{j,i} - H_{j,i} \hat{K}_j = 0_{3 \times 3}, \quad (4.21)$$

In equation (4.21) the matrices \hat{K}_j and \hat{K}_i are the estimated calibration matrices given by

$$\hat{K}_j = \begin{bmatrix} \hat{f}_j & 0 & 0 \\ 0 & \hat{f}_j & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \hat{K}_i = \begin{bmatrix} \hat{f}_i & 0 & 0 \\ 0 & \hat{f}_i & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad (4.22)$$

with \hat{f}_j and \hat{f}_i denoting the estimated focal lengths. From the calibration equation (4.21) a linear equation system

$$A_H[\hat{f}_j, \hat{f}_i] = b_H \quad (4.23)$$

is deduced similar to appendix B.1. Afterwards an analytical solution of the equation system was calculated in order to analyze the resulting error. The details of the achieved solution are explained in appendix D.1.

In figure 4.13 the resulting relative error is plotted for two different angular disturbances $\Delta\phi$ respectively $\Delta\Phi$. For the chart a usual range of rotations in image sequences of perspective cameras of $\phi \in [0.5^\circ, 20^\circ]$ and $\Phi \in [0.5^\circ, 20^\circ]$ was chosen. For greater rotations the correspondence problem becomes more difficult to solve.

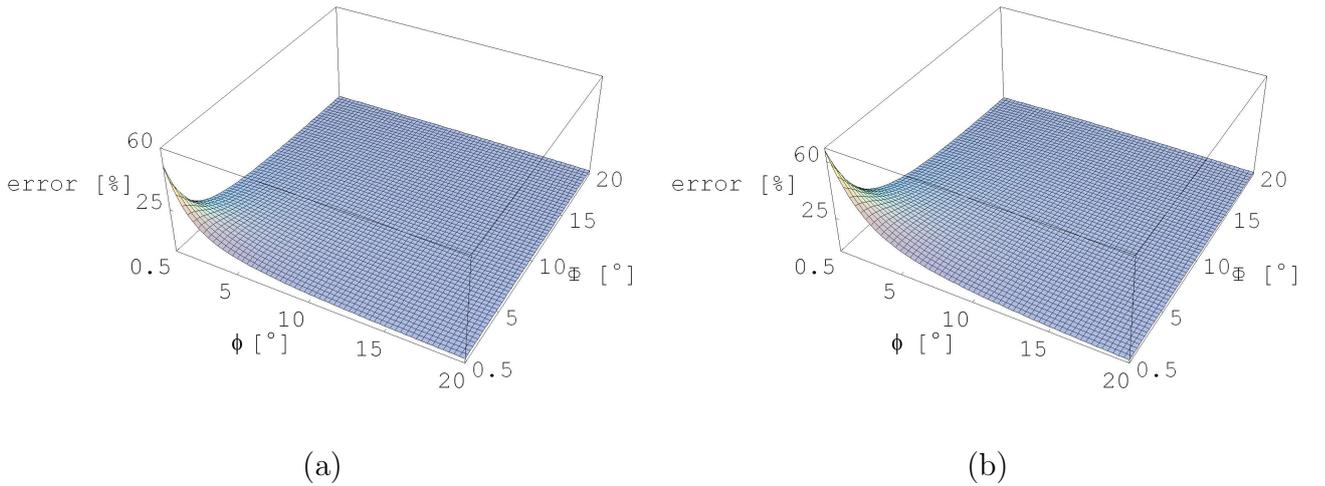


Figure 4.13: Relative error of the estimated focal length \hat{f}_j in dependence on the rotation angles about x -axis and y -axis. (a) Results for an angular disturbance of half a degree, (b) error for angular disturbance of one degree.

Figure 4.13 shows the relative error of the estimated focal length for an angle disturbance $\Delta\phi = \Delta\Phi \in \{0.5^\circ, 1^\circ\}$ in dependence on the angle of the absolute camera rotation. It can be seen from the plotted errors in figure 4.13 that for small rotations the error caused by a disturbance is larger than for large absolute rotations. This was expected beforehand because for larger absolute rotations the absolute rotation disturbance of 0.5° respectively 1° becomes less important. The value of the calibration error increases with the absolute value of the disturbance.

This result leads to the observation that a minimum absolute rotation is required in order to get small errors in the estimated intrinsic parameters. The needed minimum absolute rotation for a desired error level depends mainly on the present disturbance of the rotation

measurements. Furthermore it can be deduced that the sign of the calibration error depends mainly on the sign of the measurement error (see appendix D.1). Accordingly the required minimum absolute rotation can be reduced, if multiple estimations of the intrinsic parameters with rotation measurements are performed and the rotation error distribution has a mean value of zero. Multiple measurements can be computed for image sequences by varying the neighbors used for calibration.

So far the new proposed self-calibration approach was introduced and evaluated in terms of noise robustness. It was shown that the proposed approach is robust against noise during the homography estimation and robust against noise in the measured rotation if the absolute rotation is large enough. The next section will show the relation of the new approach to the previous work in self-calibration of purely rotating cameras.

4.1.7 Comparison to previous approaches

The relations of the proposed approach to the previously introduced approaches will be discussed in this section. The properties of the techniques of Stein, Basu, Du and Brady already explained in section 3.1.3 will be compared to the proposed self-calibration approach. All these approaches only cover the special case of constant intrinsic parameters from section 4.1.2. Furthermore the proposed novel self-calibration approach will be compared to the methods of Hartley and Agapito discussed in section 3.1. The important difference is that the new method requires the knowledge of the camera orientation. As mentioned before for many applications this is not a strong requirement because this information is often available. Therefore the new approach is also compared to these approaches but note that the approaches of the sections 3.1.1 and 3.1.2 can be applied to a more general field of applications.

Comparison to Du and Brady In contrast to the approach of Du and Brady in [DB93] the novel approach neither requires that the camera calibration is constant nor that the skew of the camera is zero. The equations developed by Du and Brady based on points are nonlinear in the focal length, the aspect ratio and the principal point. The proposed approach leads to linear calibration equations in all five degrees of freedom of the intrinsic camera parameters.

The measurements of Du and Brady showed that the feature positions have to be estimated with a precision of about 0.2 pixel in order to compute a reliable self-calibration. As seen in section 4.1.4, where the proposed approach was evaluated, larger errors in the feature positions can be accepted in order to compute a reliable self-calibration.

The proposed approach needs at least four correspondences for self-calibration because these are needed to compute the homographies. The technique of Du and Brady required only two correspondences in order to calibrate the camera. Finally the approach of Du and Brady can't be applied to arbitrary rotating cameras because it explicitly exploits rotations about one axis at a time.

Comparison to Basu The novel proposed approach of this thesis does not apply any constraint on the camera intrinsics. Basu in [BR97] assumed a constant intrinsic calibration of the camera and a zero skew of the camera. Furthermore the approach of Basu relies on rotations

about one axis at a time. Hence it can't be applied for arbitrarily rotating camera like the novel approach.

The approach of Basu estimates the focal length and the aspect ratio from small rotations. The proposed approach becomes unstable for rotations with very small angles off less than 0.5° . This can be seen from the proofs of the linear independency of the provided equation in appendix B.1 and the analysis of the influence of the errors in the measured rotation.

Comparison to Stein Stein assumed in [Ste95] constant intrinsic parameters of the camera. From the known orientation and the mapping between the images in Euclidian space Stein deduced a nonlinear cost function for the self-calibration. The proposed approach (4.3) used the projective space to linearize the mapping between the image planes. Consequently a set of linear self-calibration equations was deduced instead of nonlinear [Ste95].

The presented measurements of Stein showed a high precision and reliability given high accuracies in the orientation measurements. With the setup described earlier in section 4.1.4 similar conditions were simulated for the new approach (4.3). The orientation error of all three axes was set to $\pm \frac{1.22^\circ}{60}$ and the pixel noise was chosen as about one pixel. The achieved results are compared in table 4.1 to Steins results. It can be seen that the new approach shows better or at least comparable results. Later during the evaluation of the proposed approach in chapter 5, all shown measurements use larger errors in orientation due to the fact that only consumer sensors are used.

parameter	standard deviation Stein	standard deviation new
focal length	0.08%-0.17%	0.01%
aspect ratio	0.05%-0.11%	0.02%
principal point	1.4%-1.8%	0.01%-1.8%

Table 4.1: Standard deviation of the estimated parameters using the method [Ste95] proposed by Stein. For these results a rotation with an error of less than $\pm \frac{1.22^\circ}{60}$ is measured. The third column shows the standard deviation on synthetic data with comparable values of angular noise and pixel noise.

Comparison to Agapito and Hartley In table 4.2 the new approaches from section 4.1.1 and section 4.1.2 are compared with the existing approaches of Agapito [dAHR01] and Hartley [Har94b]. It can be seen from table 4.2 that the proposed self-calibration approach always needs a lower number of images to compute a camera calibration than the previously proposed approaches of Hartley [Har94b] and Agapito et al. [dAHR01]. Additionally this is a linear solution instead of a nonlinear. Accordingly the camera can be calibrated faster and if more frames are available the additional equations can be used to get a least squares solution which is expected to be better due to the higher redundancy in the measurements. For constant intrinsic parameters Hartley proved in [Har94b] that three views are always needed to self-calibrate the camera. In table 4.2 it is shown that the known orientation provides a unique solution even for two views. In the proof in [Har94b] Hartley used the QR-decomposition [GL89] to compute a decomposition of the first three rows and columns of the projection matrix. The

resulting decomposition is ambiguous as shown in [Har94b]. It follows that two cameras are not sufficient to compute a unique solution of the self-calibration problem. If the orientation is known it can be used to compute the calibration by multiplication with the inverse rotation matrix. A decomposition by the inverse of the known rotation was also used by Zomet et al. in [ZWS01] to avoid the nonlinear QR-decomposition. In that case it also leads to a linear calibration technique yet for a freely moving camera.

Another major advantage is that the computation of the calibration is independent of the positive definiteness of the dual image of the absolute conic. Consequently a lower influence of noisy data is provided by the novel approach in contrast to the existing approaches, which are not able to calibrate the camera, if the dual image of the absolute conic is not estimated to be positive definite (see section 3.1.1). Avoiding the rotation invariant dual image of the absolute conic leads together with the known rotation to a unique intrinsic calibration.

The new approach further does not demand that the number of varying internal camera parameters hold any requirement like (3.3) for the approach of Agapito et al. [dAHR01]. In [HÅ98] Heyden and Åström proved that all self-calibration approaches with unknown intrinsic and extrinsic camera parameters need at least one known intrinsic parameter. Later they relaxed this to one fixed but unknown parameter in [HÅ99]. Even if all internal parameters are allowed to vary the new approach is able to calibrate the cameras in contrast to all existing previous approaches as advantage of the known extrinsics.

Constraints	number of views	
	unknown rotation	known rotation
all parameters vary	-	3
skew zero	6	2
principal point known	4	2
skew known, and aspect ratio known	3	2
constant intrinsics	3	2

Table 4.2: The number of views required under various constraints on the intrinsic camera parameters in order to self-calibrate a purely rotating camera linearly. For unknown rotations the number of views is given as required by the technique of Agapito et al. from section 3.1.2. If the camera parameters are constant the number of views needed by the approach of Hartley [Har94b] is also three. For known rotation the number of views needed by the novel approach from section 4.1.1 is denoted.

Summary: This section proposed a new linear self-calibration technique for a purely rotating camera. It exploited the fact that the images of a rotating camera are correlated by a homography. The structure of this homography was used to deduce the novel linear self-calibration equations. Further this approach was compared to the existing approaches, displaying the advantages of the new approach. However the second interesting type of camera motion in which the camera moves and rotates can't be handled by this approach. In the next section the self-calibration of a moving and rotating camera will be introduced.

4.2 Self-Calibration of a freely moving camera with known rotation

In this section the self-calibration of a moving and rotating camera with known rotation will be introduced. Knowledge about the translation of the camera will be not required for the proposed self-calibration. The novel approach will utilize the rotation data to achieve a linear self-calibration which requires a lower number of images than in the general case. The image relation between two frames of a freely moving camera is described by the Fundamental matrix as shown in section 2.5.2. It maps a point in one image to a line in the other image. The novel approach will exploit the structure of the Fundamental matrix for self-calibration and its invariance under projective skew (see section 2.5.2). The next section will introduce the new approach. The following section 4.2.3 will evaluate the proposed approach with synthetic point data. Section 4.2.4 will analyze the influence of errors in the measured rotation on the computed calibration. Finally the novel approach will be compared to the previously discussed approaches from section 3.2.

4.2.1 Linear self-calibration of a freely moving camera with known rotation and fully varying intrinsic parameters

The general structure of the fundamental matrix from equation (2.20) is now exploited to derive a set of self-calibration equations for a freely moving camera. According to theorem 5 from section 2.5.2 the general decomposition of a Fundamental matrix $F_{j,i}$ is given by

$$F_{j,i} = \frac{1}{\rho_{j,i}} [e_i]_{\times} H_{j,i}^{\pi},$$

where $\rho_{j,i}$ is the unknown scale of the Fundamental matrix which can be estimated from the images themselves. The homography $H_{j,i}^{\pi}$ can be any member of a three parameter family described in theorem 5

$$H_{j,i}^{\pi} = H_{j,i}^{\infty} - e_i v^T \text{ with } v \in \mathbb{R}^3,$$

where $H_{j,i}^{\infty}$ is the infinite homography that maps over the plane at infinity, e_i is the epipole in camera i and v is a vector of the \mathbb{R}^3 that encodes the normal of the mapping plane π . Consequently the decomposition of the Fundamental matrix $F_{j,i}$ is ambiguous in the mapping plane π . Due to the invariance of the Fundamental matrix under projective skew (as explained in section 2.5.2) without loss of generality in the following $v = [0, 0, 0]^T$ will be used. It is equivalent to simply use $H_{j,i}^{\pi} = H_{j,i}^{\infty}$ as the homography induced by the plane at infinity. With (2.16) and (2.20) it follows that

$$F_{j,i} = \frac{1}{\rho_{j,i}} [e_i]_{\times} K_i R_{j,i} K_j^{-1} \quad (4.24)$$

is a valid decomposition of the Fundamental matrix $F_{j,i}$. Moreover it can be reordered to

$$\frac{1}{\rho_{j,i}} [e_i]_{\times} K_i R_{j,i} = F_{j,i} K_j. \quad (4.25)$$

Given the relative rotation $R_{j,i}$ between camera j and camera i , knowing the Fundamental matrix $F_{j,i}$ and its scale $\frac{1}{\rho_{j,i}}$, equation (4.25) is linear in the intrinsics of camera i and the intrinsics of camera j . Usually the scale $\frac{1}{\rho_{j,i}}$ of the estimated Fundamental matrix is unknown (see section 2.5.2). Applying the scaled intrinsics \tilde{K}_i of camera i as defined in (4.2) leads to the following calibration equations

$$[e_i]_{\times} \tilde{K}_i R_{j,i} = F_{j,i} K_j, \quad (4.26)$$

which are linear in the scaled intrinsics of camera i and the intrinsics of camera j for a known rotation $R_{j,i}$ and an estimated Fundamental matrix $F_{j,i}$. Please note the relationship to equation (4.3). One can see that (4.26) is an extension of (4.3) which contains the unknown camera translation t in the epipole e_i . According to the rank two of the Fundamental matrix and the cross product matrix $[\cdot]_{\times}$, the matrices $[e_i]_{\times} \tilde{K}_i R_{j,i}$ and $F_{j,i} K_j$ have also rank two. For this reason (4.26) provides only six linearly independent equations for the scale $\frac{1}{\rho_{j,i}}$ and the intrinsics of the cameras contained in K_j , K_i given a known general rotation. The self-calibration equations (C.1) resulting from equation (4.26) are given in appendix C.1. A proof of the linear independency of the equations is given in appendix C.1. Furthermore it can be seen that similar to the Kruppa equations from section 2.8.1, the position of the plane at infinity is not contained in (4.26). Consequently it shares the advantage of only requiring the pairwise Fundamental matrices with the Kruppa equations (3.19).

To calibrate the cameras i and j for each camera five intrinsic parameters and one scale for the estimated Fundamental matrix $F_{j,i}$ have to be determined. From a counting argument it follows that the solution for a single camera pair is not unique if less than five constraints for the scale $\frac{1}{\rho_{j,i}}$ and the intrinsics of the cameras j and i are available, since only six linear independent equations are provided for the eleven unknowns of the first camera pair. Employing an additional camera k introduces five additional intrinsic camera parameters contained in K_k and one additional unknown scale $\frac{1}{\rho_{j,k}}$. So any additional camera provides the same number of constraints as unknown parameters. For this reason it is not possible to reach a unique solution for the calibration equation (4.26) without at least five additional constraints on the intrinsics or the scale. In order to reach a unique self-calibration sets of known parameters will be discussed in the next section.

4.2.2 Constraints needed to reach a linear calibration

The sets of constraints for achieving a unique solution for the self-calibration problem of a freely moving camera are investigated now. A constraint for a parameter has one of two types: the value of the parameter is known for each camera or it's known that the parameter is fixed for all cameras yet its value is unknown. Each parameter known for all n cameras gives n constraints for the self-calibration. A parameter known as fixed provides one constraint less than the number of cameras thus for n cameras $n - 1$ constraints. For a unique solution of the self-calibration problem therefore the following constraint has to hold

$$f(n - 1) + kn \geq 6n - 1,$$

where f is the number of unknown but fixed parameters and k is the number of known parameters.

Known principal point For a known principal point, or even for only one known component of the principal point a unique solution of the self-calibration problem can be found. Applying the known principal point to the calibration equation (4.26) delivers

$$[e]_{\times} \left(\tilde{K}_i + \tilde{C}_i^{u,v} \right) R_{j,i} = F_{j,i} \left(K_j + C_j^{u,v} \right) \quad \text{with } C^{u,v} = \begin{bmatrix} 0 & 0 & -u \\ 0 & 0 & -v \\ 0 & 0 & 0 \end{bmatrix}, \quad (4.27)$$

where \tilde{C} is the constraint matrix scaled by $\frac{1}{\rho_{j,i}}$. As in equation (4.26) the calibration is linear in the three unknown intrinsics focal length, aspect ratio and skew of camera j and the intrinsics scaled by $\frac{1}{\rho_{j,i}}$ of camera i and the scale $\frac{1}{\rho_{j,i}}$. The constrained self-calibration problem (4.27) also provides six linear independent equations for each camera pair. Again from a counting argument it follows that a unique solution of the constrained self-calibration problem can be computed from an image triplet, because there are three unknowns left for the first camera and each additional image pair introduces four new unknowns, namely the scale, the focal length, the aspect ratio, and the skew of the additional camera.

Known skew and principal point The latter paragraph showed that for the first camera pair seven unknowns remain but (4.26) provides only six linear independent equations. To reach a calibration from a camera pair only one additional constraint is needed. The most common constraint for CCD-Cameras is the zero skew. Then the focal length and aspect ratio can be estimated directly from a single camera pair exploiting the knowledge of the orientation of the cameras.

Known skew, known aspect ratio and known principal point As mentioned above with known principal point and known or zero skew the cameras can be calibrated using a single camera pair. Sometimes only the focal length of the camera is of interest; then the aspect ratio is also known like for video cameras (PAL⁵ $a = 1.067$, NTSC $a = 1.09$). For this special set of constraints the introduced approach as well as the general approach without known rotation deliver a linear solution of the self-calibration problem [HZ03].

So far the self-calibration technique for freely moving cameras with known orientation was introduced. In the next section the noise robustness of the proposed approach will be measured using synthetic data.

4.2.3 Evaluation for freely moving cameras

To measure the noise robustness of the proposed linear calibration technique for freely moving cameras synthetic point data with known noise and ground truth information will be used in

⁵Short for Phase Alternating Line, the dominant television standard in Europe. The United States uses a different standard, NTSC short for National Television System Committee. Details can be found in [ITUb, ?, ITUa].

this section. Six cameras were positioned on a sphere observing a uniformly distributed 3D point cloud of 100 points inside the sphere (see fig 4.14). The 3D points were projected into the camera images of size 512x512 pixel. Afterwards the image points were disturbed with Gaussian pixel noise of a standard deviation σ_{pixel} of up to 1 pixel in each coordinate and zero mean. For numerical reasons the image points were normalized as proposed in [Har95] during the computation of the Fundamental matrices $F_{j,i}$. The Fundamental matrices were estimated by computing a least squares solution of (2.26). The Fundamental matrices $F_{j,i}$ were used for the self-calibration in order to measure the robustness against noise in the the pixel position of the features. The known orientation of the cameras was also disturbed by Gaussian distributed angular noise with a standard deviation σ_{angular} of up to one degree and zero mean. The results for the linear calibration of the first camera with camera calibration matrix K_1 set as

$$K_1 = \begin{bmatrix} 415 & 41.5 & 0 \\ 0 & 456.5 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

applying the constraint of a known principal point $[u, v] = [0, 0]$ are shown in figures 4.14 to 4.17. The achieved results for the remaining five cameras are comparable to the shown results for the first camera.

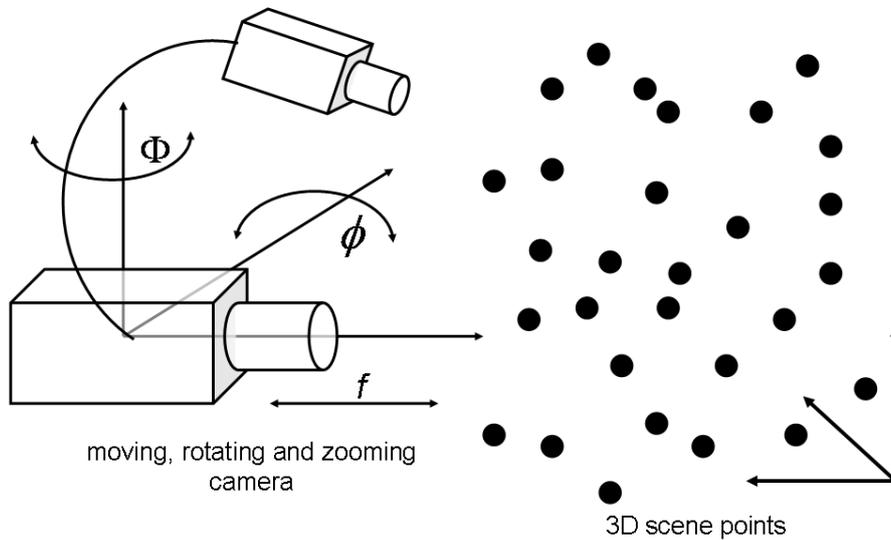


Figure 4.14: Synthetic scene for measuring the noise robustness of the linear calibration of a freely moving camera.

In figure 4.15 (a) the relative error of the mean estimated focal length is shown. It can be seen that the mean focal length is estimated very accurately in the tested range of noise for pixel position and for angular noise. The maximal measured error is 1.5% of the ground truth focal length. The maximal standard deviation of the estimated focal length is about

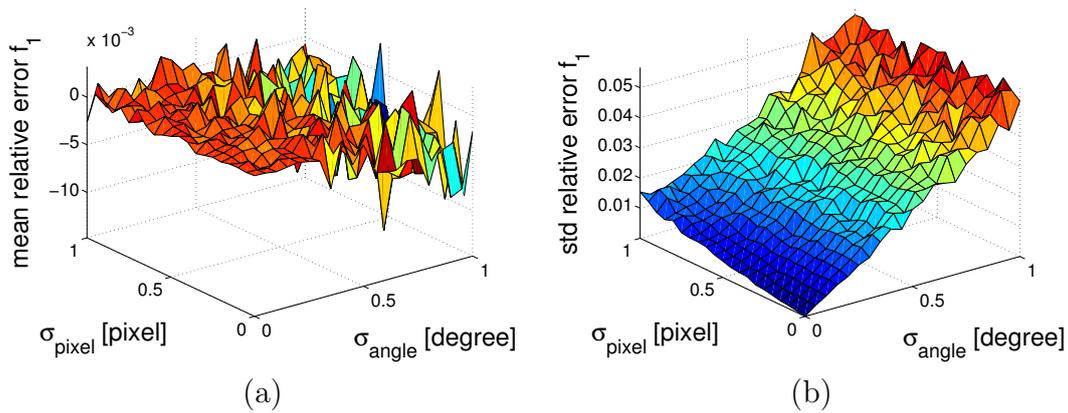


Figure 4.15: Noise robustness measurements with Gaussian angular noise for rotation about the x -axis and y -axis with a standard deviation of up to 1 degree and Gaussian pixel noise of the corresponding points with a standard deviation of up to one pixel: (a) relative error of the mean of the estimated focal length $f = 415$, (b) standard deviation of the relative error of the estimated focal length.

5.7% for the achieved calibrations as shown in figure 4.15 (b). Hence the estimation of the focal length is rather stable in the chosen range of noise that is typical for consumer rotation sensors. The achieved mean value and standard deviation of the focal length are slightly better than for a calibration of a purely rotating camera. At a first glance this is surprising. It can be explained by the applied known principal point which does not introduce additional numerical instabilities as in the case of a purely rotating camera.

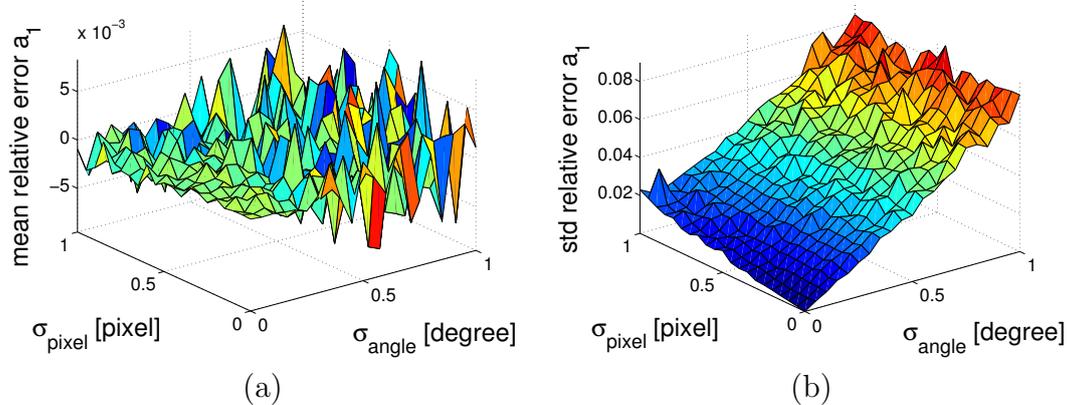


Figure 4.16: Noise robustness measurements with Gaussian angular noise for rotation about the x -axis and y -axis with a standard deviation of up to 1 degree and Gaussian pixel noise of the corresponding points with a standard deviation of up to one pixel: (a) relative error of the mean of the estimated aspect ratio $a = 1.1$, (b) standard deviation of the relative error of the estimated aspect ratio a .

The estimated relative error of the mean aspect ratio is shown in figure 4.16 (a). It can be seen that the estimation of the aspect ratio is rather stable in the tested noise range for angular noise and pixel noise. The maximal relative error is about 8% of the original aspect ratio. The standard deviation of the relative error is with up to 9% slightly higher than for the focal length. Thus like for the calibration of a purely rotating camera the estimation of the aspect ratio gives reasonable results but is not as robust as the estimation of the focal length.

So the cumulated errors of the focal length f and the product of focal length and aspect ratio af also cause a higher sensitivity to noise for the aspect ratio. Similar to the measurements of the focal length the estimated aspect ratio is slightly better than for a purely rotating camera. Again the applied constraint of known principal point improves the estimation.

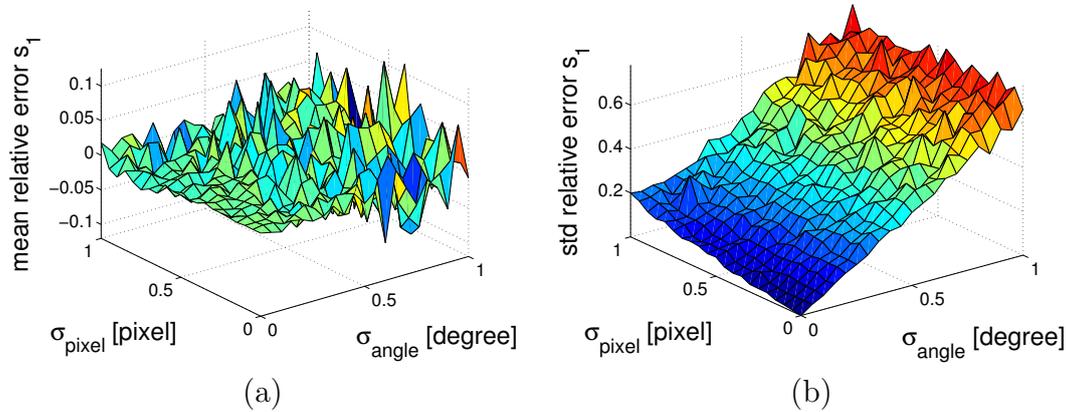


Figure 4.17: Noise robustness measurements with Gaussian angular noise for rotation about the x -axis and y -axis with a standard deviation of up to 1 degree and Gaussian pixel noise of the corresponding points with a standard deviation of up to one pixel: (a) relative error of the mean of the estimated skew $s = 41.5$, (b) standard deviation of the relative error of the estimated skew s .

The measurements for the estimated skew of the camera are shown in figure 4.17. It can be seen that the estimation of the skew is rather stable for an angular noise of σ_{angle} less than half a degree and a pixel noise of σ_{pixel} less than half a pixel. In this range of noise the relative error is less than 5% and the standard deviation is less than 30%. These estimation results are not as stable as the estimation of the focal length due to the small skew of a tenth of the focal length of the synthetic camera.

So far it was seen that the angular error has the major impact on the standard deviation of the estimated intrinsics. Consequently the next section will analyze the sensitivity of the focal length on errors in the rotation data in more detail.

4.2.4 Influence of errors in rotation

As seen in the evaluation of the previous section the source of error with the biggest impact is the error in the measured rotation and this will be analyzed in this section. To investigate the influence of rotation errors the achieved calibration of a single camera pair is used. The most important intrinsic parameter for self-calibration is the focal length of the camera. Accordingly the following analysis will focus on the focal length, all other camera parameters will be set as known and fixed. Then it is possible to compute a unique solution for the focal lengths from a single camera pair and without loss of generality the camera calibration matrices K_j and K_i are given by

$$K_j = \begin{bmatrix} f_j & 0 & 0 \\ 0 & f_j & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad K_i = \begin{bmatrix} f_i & 0 & 0 \\ 0 & f_i & 0 \\ 0 & 0 & 1 \end{bmatrix}. \quad (4.28)$$

In the following it is assumed that the rotation $R_{i,j}$ between both cameras is a rotation about the x -axis and the y -axis given by

$$R_{j,i} = R_x R_y \text{ with } R_x(\phi) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \phi & -\sin \phi \\ 0 & \sin \phi & \cos \phi \end{bmatrix}, R_y(\Phi) = \begin{bmatrix} \cos \Phi & 0 & \sin \Phi \\ 0 & 1 & 0 \\ -\sin \Phi & 0 & \cos \Phi \end{bmatrix}, \quad (4.29)$$

where ϕ is the rotation angle about the x -axis and Φ is the rotation angle of the rotation about the y -axis. The rotation about the optical axis is set to be zero due to the fact that it does not provide linear constraints on the focal lengths (explained in more detail in section 4.3.5). The analysis uses a general epipole $\mathbf{e}_i = [e_1, e_2, e_2]$. Hence the Fundamental matrix between the two cameras j and i can be computed with (2.23) and (4.28) as

$$F_{j,i} = [\mathbf{e}_i]_{\times} K_i R_{j,i} K_j^{-1}. \quad (4.30)$$

For this Fundamental matrix the scale $\rho_{j,i}$ is equal to one. Please note, the Fundamental matrix $F_{j,i}$ is an exact transformation without any error. The measured rotation angles $\tilde{\phi}$ and $\tilde{\Phi}$ are assumed to be disturbed by an angular error $\Delta\phi$ respectively $\Delta\Phi$. Consequently the disturbed rotation matrix $\hat{R}_{j,i}$ is given by

$$\hat{R}_{j,i} = R_x(\phi + \Delta\phi) R_y(\Phi + \Delta\Phi), \quad (4.31)$$

with $R_x(\cdot)$ and $R_y(\cdot)$ from equation (4.29). Hence the self-calibration equations (4.26) are equal to

$$[\mathbf{e}_i]_{\times} \hat{K}_i \hat{R}_{j,i} - F_{j,i} \hat{K}_j = 0_{3 \times 3}, \quad (4.32)$$

with the estimated calibration matrices \hat{K}_j and \hat{K}_i given by

$$\hat{K}_j = \begin{bmatrix} \hat{f}_j & 0 & 0 \\ 0 & \hat{f}_j & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \hat{K}_i = \begin{bmatrix} \hat{f}_i & 0 & 0 \\ 0 & \hat{f}_i & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad (4.33)$$

where \hat{f}_j and \hat{f}_i denote the estimated focal lengths. Together with equation (C.2) the linear equation system $A_F [\hat{f}_j, \hat{f}_i]^T = b_F$ can be deduced in the entries of $F_{j,i}$, \mathbf{e}_i and the disturbed rotation $\hat{R}_{j,i}$. To analyze the error in the focal length estimation that results from $\Delta\phi$ respectively $\Delta\Phi$ the solution of $A_F^T A_F [\hat{f}_j, \hat{f}_i]^T = A_F^T b_F$ was computed analytically. This solution for the unknown focal lengths \hat{f}_j and \hat{f}_i is now given in the rotation angles $\tilde{\phi}$ and $\tilde{\Phi}$, the angular errors $\Delta\phi$ respectively $\Delta\Phi$, the epipole \mathbf{e}_i and the components of the Fundamental matrix $F_{j,i}$. Due to the high complexity of the analytic solution it can be found in appendix D.2. To show the effects of angular error in figure 4.18 the relative error of the estimated focal length \hat{f}_j is plotted for two different angular disturbances in dependence on the rotation angles ϕ and Φ with an epipole $\mathbf{e}_i = [1, 1, 1]^T$. The analysis showed that the epipole does not influence the general properties of the error (see appendix D.2).

In figure 4.18 the relative error of the estimated focal length f_j is plotted depending on the true rotation angles $\phi \in [0.5^\circ, 20^\circ]$ and $\Phi \in [0.5^\circ, 20^\circ]$. The range was chosen as a typical range

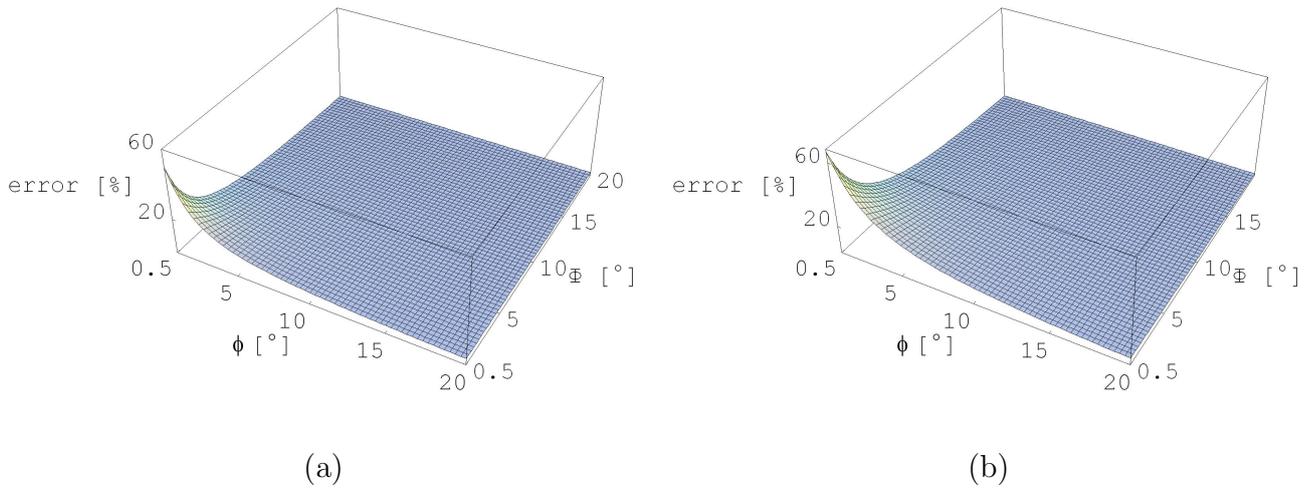


Figure 4.18: Relative error of the estimated focal length \hat{f}_j in dependence on the rotation angles about x -axis and y -axis. (a) Results for an angular disturbance of half a degree, (b) error for angular disturbance of one degree.

of rotations possible in image sequences of perspective cameras. Please, note that the error functions of the self-calibration for a purely rotating camera from section 4.1.6 look similar to the shown errors of the freely moving camera. As seen during the deduction of the calibration equations (4.26) of a freely moving camera it is also similar to the calibration equations (4.3) of the purely rotating camera. Accordingly it is not surprising that the errors are similar.

From figure 4.18 it can be seen that for small rotations the error caused by a disturbance is larger than for large rotations. This is plausible because for larger absolute rotations the absolute rotation disturbance of 0.5° respectively 1° becomes less important. Furthermore it can be seen that the larger the rotation disturbances $\Delta\phi$ and $\Delta\Phi$ are, the larger is the calibration error.

From the analysis it follows that in dependence on the present rotation measurement error small rotations between the cameras of a sequence will lead to larger calibration errors than larger rotations. Hence in applications it is helpful to use absolute rotations that are as large as possible. The minimal required rotation depends on the present rotation measurement error and the desired precision of the self-calibration. Furthermore the analysis discussed in section D.2 showed that the mean value is not disturbed by the angular error. The insensitivity of the mean value of multiple estimates was also observed in the experiments of the previous section. Therefore multiple estimation from independent measurements will reduce the error of the calibration. For images it is usually possible to calculate multiple measurements by varying the reference frame and the neighbors used for calibration.

So far the performance of the proposed approach and the influence of noise to results of the novel linear self-calibration approach were evaluated. The next sections will show the relations to the previously proposed approaches.

4.2.5 Relation of self-calibration of freely moving cameras to the Kruppa equations

The linear calibration for freely moving cameras showed that the calibration is possible without computing the camera motion, except from the known camera orientation. The derived self-calibration equations (4.26) do not require a consistent frame for all cameras. I.e. it is not needed to find the same position of the plane at infinity in the Fundamental matrices since this plane is not contained in the equations as result of the invariance against projective skew of the Fundamental matrix.

The advantage of only employing the pairwise Fundamental matrices without requiring any global consistent frame is shared with the Kruppa equations from section 2.8.1. In contrast to the Kruppa equations (2.46) the new approach is more robust against noise, and the second major advantage is that it is a linear approach. The advantages of the new approach result from avoiding to switch to a rotation invariant formulation. In the following it will be shown that the novel approach can be transformed into the Kruppa equations.

The Kruppa equations also exploited the epipolar geometry to deduce constraints on the camera calibration through the dual image of the absolute conic. They were introduced for cameras with constant intrinsic parameters. Hence in the following a constant intrinsic camera calibration is used. Equation (4.26) provides the equality of $[e_i]_x \tilde{K}_i R_{j,i}$ and $F_{j,i} K_j$. Multiplying each side in (4.26) by its transposed implies the following equation

$$[e_i]_x \tilde{K}_i R_{j,i} R_{j,i}^T \tilde{K}_i^T [e_i]_x^T = F_{j,i} K_j K_j^T F_{j,i}^T.$$

Eliminating the scale $\rho_{j,i}$ contained in \tilde{K}_i then provides

$$\Leftrightarrow [e_i]_x \underbrace{K_i K_i^T}_{\omega_i^*} [e_i]_x^T = F_{j,i} \underbrace{K_j K_j^T}_{\omega_j^*} F_{j,i}^T, \quad (4.34)$$

which are exactly the Kruppa equations (2.46) introduced in section 2.8.1 for a camera with constant intrinsic parameters. So the Kruppa equations can be seen as a quadratic formulation of the novel approach (4.26). Accordingly it is plausible that the Kruppa equations provide quadric constraints in the unknowns of the dual image of the absolute conic in contrast to the linear constraints of the proposed approach (4.26).

The next subsection will show the relation between the novel approach and the general self-calibration equation derived from the dual image of the absolute conic.

4.2.6 Relation of the linear self-calibration to the general self-calibration

In the previous sections 4.2.1 and 4.2.5 the self-calibration approach only applied constraints to the image pair relation and not to the position of the plane at infinity π_∞ . Using the structure of the Fundamental matrix to self-calibrate the cameras had the advantage of not requiring a global consistent frame for all computations which is required for most self-calibration techniques as explained in section 2.7. In this section it will be shown how to use the position of the plane at

infinity as a common entity without any 3D scene. This will provide a nonlinear self-calibration approach for freely moving cameras with known orientation.

Employing the general self-calibration given by the absolute conic as described in section 2.8 constraints can be applied to the DIAC ω^* and the plane at infinity π_∞ . The estimated projective skew matrix $H_{4 \times 4}$ contains the calibration of the first camera and the normal vector of the plane at infinity π_∞ . Once determined the calibration of all cameras can be computed from the projectively skewed projection matrices using matrix multiplication. The self-calibration equation (2.41)

$$K_i R_i^T \cong (M_i - e_i v^T) K_j$$

uses the projectively skewed camera matrix. In order to compute the camera projection matrices of a camera pair j and i from the Fundamental matrix $F_{j,i}$ it is assumed that camera j is identical with the world coordinate system. Then for a given Fundamental matrix $F_{j,i}$ the projectively skewed projection matrices P_j^p and P_i^p for the corresponding cameras j and i are given by [HZ03]

$$P_j^p = [I_{3 \times 3} | 0] \text{ and } P_i^p = \left[[e_i]_\times F_{j,i} + e_i a_i^T \mid \lambda e_i \right] \text{ with } a_i \in \mathbb{R}^3, \quad (4.35)$$

where the vector a_i determines the mapping plane of the homography. Exploiting the structure of the projection matrix from equation (4.35) equation (2.41) can be written as

$$K_i R_i^T = \frac{1}{\rho_{j,i}} \left(([e_i]_\times F_{j,i} + e_i a_i^T) - e_i v^T \right) K_j,$$

where $[v^T, 1] \in \mathbb{P}^3$ encodes the plane at infinity. This is similar to equation (3.16) Hartley used for self-calibration with QR-decomposition in [Har94a]. Equation (2.41) can be reordered to

$$\left(([e_i]_\times F_{j,i} K_j + e_i a_i^T K_j) - e_i \tilde{v}^T \right) R_i \cong \tilde{K}_i \quad \text{with } \tilde{v} = K_j^T v. \quad (4.36)$$

It can be seen that the known entries with value zero of the K_i with $i = 2, \dots, n$ provide three constraints for the eight unknowns of K_j , \tilde{v} and the three unknowns in a_i . It is not mandatory to estimate the scale $\frac{1}{\rho_{j,i}}$ because it doesn't affect the constraints provided by the zero entries. Due to the fact that each image provides three constraints and introduces three unknowns in a_i a self-calibration can not be reached. Using techniques like the threading of Fundamental matrices proposed from Avidan and Shashua in [AS01] helps to overcome this problem of three new parameters for each image. The proposed technique uses the trifocal tensor to ensure a consistent mapping plane for all cameras. Then the three parameters in a_i are determined. Another way to determine a_i in advance is the employment of a projective reconstruction of the scene. The reconstructed projective scene can be exploited to fix the position of the plane at infinity for all cameras and the mapping plane encoded by a_i of the cameras.

For a fixed mapping plane of the homographies of all cameras each camera provides three constraints in the eight unknowns of K_j and \tilde{v} in (4.36). Using these constraints four cameras are needed to calibrate the first camera and to compute the normal vector of the plane at infinity using \tilde{v} . The number of cameras needed to calibrate has to fulfill the relation

$$f(n-1) + (k+3)n \geq 8.$$

The number of cameras needed to calibrate for different sets of constraints is given in table 4.3 also for a fixed mapping plane of the homography ensured by a threading of Fundamental matrices or by a projective reconstruction. If the consistent mapping plane for the homographies can't be ensured at least one constraint for the intrinsic camera calibration is needed.

It can be seen that the approach with the consistent mapping plane needs a smaller number of images to calibrate the cameras. Further it also computes the full projective skew matrix $H_{4 \times 4}$ as introduced in section 2.7 without requiring any 3D information of the scene if the mapping plane is constructed by the threading of Fundamental matrices from [AS01].

Constraints	number of views		
	absolute quadric	(4.26)	known rotation (4.36)
all parameters vary	-	-	4
skew known and not zero, all other parameter vary	-	5	3
principal point known, all other parameters vary	4*, 5	3	3
principal point known, skew zero	3	2	2
principal point known, skew known, and aspect ratio known	2	2	2

Table 4.3: The number of views required under various conditions in order to self-calibrate linearly. There may be multiple solutions for the cases marked with an asterisk. Second column denotes the number of images for the self-calibration using the structure of the Fundamental matrix from section 4.2.1. In the third column the number of images for the self-calibration approach (4.36) using the approach of this section are denoted.

A comparison in terms of the number of views needed for self-calibration between the approaches from section 3.2.2 and the new approach from section 4.2.1 which uses the structure of the Fundamental matrix is shown in table 4.3. It can be seen that the number of views required for the new linear self-calibration with known rotation (4.26) is lower than the number of views required for the self-calibration using the dual of the absolute quadric.

The previously introduced approach in (4.36) is similar to the calibration introduced by Zomet et al. in [ZWS01] discussed in section 3.2.3. The self-calibration equations (4.36) can be written as matrix equation

$$\underbrace{[[e_i] \times F | e_i]}_{P_i} \underbrace{\begin{bmatrix} K_j \\ \tilde{v}^T \end{bmatrix}}_{H_{4 \times 3}} R_i = \tilde{K}_i, \quad (4.37)$$

which is equivalent to equation (3.23) proposed by Zomet et al. to compute the calibration of the cameras of a multi-camera system in [ZWS01]. Hartley suggested equation (3.16) in [Har94a] as

self-calibration equation which is also equivalent to equation (4.37). The difference is that (4.37) explicitly incorporates the pairwise Fundamental matrices for the self-calibration. Moreover Zomet et al. used a general 4×3 matrix in [ZWS01] that does not utilize the knowledge of the possible projective skew. Implicitly the knowledge is enforced in their approach by employing the first camera $P_1 = [I_{3 \times 3} | 0]$ as constraint; so (4.37) becomes

$$K_j = \frac{1}{\rho_j} K_j.$$

However due to the suggested least squares solution for noisy data the solution will not hold these constraints. Enforcing this constraint requires the structure of $H_{4 \times 3}$ as denoted in (4.37). Consequently the problem as formulated by Zomet is over parameterized with four additional parameters.

So far two new self-calibration methods were proposed for cameras with known orientation. The major error sources of this approaches were identified and analyzed analytically. The next section will discuss the critical camera rotations for the proposed approaches.

4.3 Critical camera motions for calibration with known rotation

In section 2.10 the critical motion sequences for self-calibration were discussed. The CMS classes for the estimation of a metric reconstruction also apply for the new introduced approach. Hence this section will focus on the degenerate cases when not only the projective reconstruction is ambiguous but also the intrinsic camera parameters can't be estimated uniquely. These classes of rotation can be found by analyzing the properties of the rotation matrix. For each class the special constraints on the intrinsic camera parameters using equations (B.1) respectively (C.1) can be deduced from the properties of the rotation matrix. The results for both calibration techniques are equivalent in terms of the critical rotations; therefore the deduced results are valid for both calibrations. A summary of the results of the next section 4.3.1 to 4.3.5 is shown in table 4.4.

In the following sections the self-calibration of a single camera pair will be investigated in order to deduce the provided constraints. According to the block structure of the equations systems (B.1) and (C.1) the analysis applies also for each additional camera pair.

4.3.1 No rotation at all or rotations about 180°

It is clear that the first class of critical rotations of the camera consists of the rotations about any axis of 180° and no rotation of the camera at all. In this case the resulting camera rotation is similar to an identity matrix with respect to the number of non-zero elements. Hence the calibration equations (B.1) respectively (C.1) do not provide a sufficient number of constraints to compute a unique self-calibration. This class of rotations is contained in the CMS-classes 1.2 and CMS-classes 1.3 introduced by Sturm in [Stu96] and was already explained in section 2.10.

Applying an identity matrix in the equations (B.1) and (C.1) leads to a set of six linear independent equations in the resulting equation system (E.2). These equations deliver constraints on the principal point v in y direction for cameras with varying intrinsics and non-zero skew. Moreover constraints on the ratio of focal lengths, the aspect ratios, the skews and the principal points in x direction of the involved cameras are provided. The focal lengths as well as the aspect ratios can't be determined at all. Consequently it follows for this class of rotations for a zooming camera that does not rotate or rotates about 180° about any axis, it is only possible to estimate the principal point in y . For a freely moving camera a similar set of five linear independent constraints can be deduced.

The estimation of the principal point is poorly conditioned and the most cameras have skew zero. Accordingly it is not expected that the estimations exploiting the proposed approaches for self-calibration of cameras with known orientations are practically feasible. A more detailed deduction for results mentioned above is given in appendix E.1.

type of rotation	f	s	u	af	v
no rotation	-	-	-	-	x
x -axis rotation	-	x	x	x	x
y -axis rotation	x	-	x	-	x
z -axis rotation	-	x	x	-	x

Table 4.4: Critical rotations for self-calibration of cameras with known orientation. The first column describes the type of degenerate rotation. The following columns denote for each intrinsic camera parameter whether it can be estimated "x" or not "-".

4.3.2 Rotation about one axis

The class of rotations about one axis is the CMS-class 1.1 from section 2.10. Sturm found that in this case several conics on the plane at infinity exist that have an invariant projection into the images. Accordingly the achieved self-calibration of the camera is ambiguous. A rotation about an axis in general position is not critical for the proposed approach since it still provides enough constraints for the self-calibration. The only exception are rotations about the axes of the camera coordinate system. The following paragraphs will discuss these rotations about the different axes of the coordinate system.

4.3.3 Rotation about the x -axis

Self-calibration of a camera that rotates about the x -axis of the camera coordinate system is only able to estimate the product af of aspect ratio a and focal length f as well as the principal point $[u, v]$ and the skew s of the cameras. A rotation matrix R_x for a rotation about the x -axis with angle ϕ has the following structure

$$R_x(\phi) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \phi & -\sin \phi \\ 0 & \sin \phi & \cos \phi \end{bmatrix}.$$

Exploiting the rotation R_x for self-calibration in the equations (B.1) respectively (C.1) delivers self-calibration equations that only constrain the principal point $[u, v]$ and the product of aspect ratio a and focal length f and the skew s of the cameras. Furthermore a constraint on the ratio of the focal lengths of the cameras is provided. A detailed deduction is given in appendix E.2. The resulting linear equation system for two purely rotating cameras has seven linear independent equations. For a freely moving camera similar constraints can be deduced from (4.26), where five of them are linear independent. Consequently a unique self-calibration can't be computed from a single camera pair for the remaining parameters which are the focal length in y direction af , the principal point $[u, v]$ and the skew s without any additional constraints on these parameters.

However such a critical rotation can be detected from the structure of a given rotation matrix automatically. Hence the computation of the self-calibration can take into account that only constraints for the focal length in y direction af and the principal point $[u, v]$ and the skew s are available.

4.3.4 Rotation about the y -axis

Similarly to a rotation about the x -axis a rotation about the y -axis with angle Φ

$$R_y(\Phi) = \begin{bmatrix} \cos \Phi & 0 & \sin \Phi \\ 0 & 1 & 0 \\ -\sin \Phi & 0 & \cos \Phi \end{bmatrix},$$

only provides constraints on the focal length f and the principal point $[u, v]$. More details about the deduction of these results can be found in appendix E.3. Two additional linear constraints are provided, one for the ratio of the focal lengths in y -direction and one for the ratio of the skews of the involved cameras. Furthermore it can be seen that a rotation about the y -axis still provides nine linear independent equations for the self-calibration of a single camera image pair. It seems to be surprising that a rotation about the y -axis provides more linear independent equations than a rotation about the x -axis. It is a result of the modeling of the skew in the camera calibration matrix that is biased to the x -direction. The skew provides two linear independent constraints for the calibration equations (B.1).

However, again it is possible to identify the new set of the constraints from the rotation matrix itself. Thus a self-calibration technique can be adapted to only estimate the focal length f and the principal point $[u, v]$. Then a unique self-calibration for the remaining parameters of the camera can be estimated from a single camera pair. For cameras with zero skew a camera triplet is needed for a unique solution.

An analogous analysis can be done for a freely moving camera that rotates about the y -axis. It delivers six linear independent constraints on the focal length f , the principal point $[u, v]$, the ratio of the focal lengths in y -direction af of the cameras and the ratio of the skews of the cameras. Hence it is possible to estimate the focal lengths and the principal points of the involved cameras from a camera triplet for a freely moving camera.

4.3.5 Rotation about the z -axis

For a camera that purely rotates about its optical axis the rotation matrix is given by

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

for a rotation angle of Θ . Employing the rotation R_z in the self-calibration equations (B.1) respectively (C.1) provides seven linear independent equations. These equations model constraints on the principal point and the skew of the camera. Hence these parameters can be computed after an automatic detection of this type of rotation using the structure of the rotation matrix. The resulting set of equations provides a sufficient number of constraints to achieve a unique solution of the self-calibration for a purely rotating camera.

For a freely moving camera a comparable set of constraints on the the principal point and the skew of the camera can be deduced from the self-calibration equations (C.1). Six of these constraints are linear independent. Accordingly a unique solution for the remaining parameters of a camera triplet is possible.

So far it was shown that the only critical rotations for the calibration of the camera are rotations about 180° about any axis, no rotations at all and the rotation about one of the axis of the coordinate system. Rotations about only one constant general axis are not critical. As seen in the proofs for the linear independency of the self-calibration equations provided by (4.4) and (4.26) the sinus of the rotation angle numerically influences the linear independency of the equations. Hence the self-calibration becomes ambiguous with small angle rotations due to numerical problems.

Summary: This chapter proposed two new self-calibration algorithms for cameras with known orientation. The first self-calibration approach was developed to calibrate purely rotating cameras with known orientations. The robustness of the approach was measured with synthetic point data. These measurements showed that the algorithm gave reliable calibrations for pixel noise with a standard deviation of up to one pixel and angular noise with a standard deviation of up to one degree. Furthermore the error of the measured rotations was identified as major source of error. Accordingly it was analyzed in detail. As results of the analyzes it could be seen that it is helpful to use large rotations between the images and multiple estimations of the intrinsic parameters. The proposed approach was compared to the previously proposed self-calibration techniques for the self-calibration of rotating cameras. It shows that one of the major advantages of the novel approach is the lower number of images required for self-calibration. Another advantage that the proposed technique overcomes the limitation on the number of varying intrinsics.

The second proposed self-calibration approach can be applied to calibrate freely moving cameras with known orientation. The noise robustness of this approach was measured with synthetic point data by varying the pixel noise of the image features with pixel noise of a standard deviation of up to one pixel and angular noise with a standard deviation of up to one degree. It was seen that the calibration technique gave reasonable results in the tested range of noise. According to the achieved results the major source of error was also the error

in the measured rotation. Consequently the influence of it was analyzed analytically. Similar to a rotating camera the analysis delivers that result that large absolute rotations and multiple estimates of the intrinsics are helpful. Furthermore the novel approach was compared to the existing approaches for the self-calibration of a freely moving camera. The comparison showed as the major advantage of the novel approach that the number of images required for self-calibration can be reduced. The next chapter will measure the performance of the proposed self-calibration algorithms on image sequences of synthetic scenes and real scenes.

Chapter 5

Experimental results

The proposed approaches from section 4.1.1 for self-calibration of a rotating camera with known orientation and from section 4.2.1 for self-calibration of a freely moving camera with known rotation will be evaluated in this chapter. First of all image sequences of synthetic scenes will be employed to measure the noise robustness of the proposed techniques. Afterwards both approaches will be tested on image sequences of real scenes. For the image sequences of the real scenes the intrinsic parameters of the cameras are calibrated manually beforehand. The next section will discuss the results for the calibration of a purely rotating camera with known orientation. It will be followed by section 5.2 which measures the performance of the self-calibration of freely moving cameras.

5.1 Results for a rotating camera

First of all an image sequence of a synthetic scene is exploited to evaluate the performance of the new linear self-calibration method of a purely rotating camera with known orientation. Afterwards the new technique is used to calibrate a camera that took an image sequence of a real scene. The next section will describe in detail the algorithm exploited for self-calibration from image sequences.

5.1.1 Algorithm for self-calibration with known orientation

For a self-calibration of an image sequence of a purely rotating camera incorporating the known camera orientation the following steps are needed.

1. Correspondence estimation
2. Robust computation of the homographies
3. Estimation of the intrinsic camera parameters from the homographies together with the known camera orientation.
4. Combination of multiple estimations of the same intrinsic parameter if necessary.

These steps are all performed for the self-calibrations shown in the following sections. The next paragraphs will discuss the details of each step performed.

Correspondence estimation The correspondences were generated with the KLT-tracker [LK81, BM04]. It uses the image brightness and the image gradient to find the feature position in image i given the position of the feature in image j . As result a position of the feature in image i with sub-pixel accuracy was provided. For the experiments the implementation of Birchfield was used.

Robust computation of the homographies The correspondences generated with the KLT-tracker [LK81] also contain mismatches, i.e. correspondences that do not describe the position of the projection of the same scene point in image j and image i . To estimate a homography these wrong correspondences should not be included. Therefore RANSAC¹ is used to determine the right correspondences [FB81]. A least squares estimation for the homography on the correct correspondences was calculated. Afterwards a nonlinear optimization with Levenberg-Marquardt for the homography and the feature positions was performed. The minimized cost function measures the mapping error of the features and the difference of the estimated and the detected feature position as proposed by Torr [Tor98].

Estimation of intrinsics The estimated homographies were employed together with the known relative rotations to estimate the intrinsic parameters of the cameras. The self-calibration was performed with the novel approach from section 4.1.

Combination of multiple estimates The homography for an image pair can be computed between each pair of cameras as long as correspondences between the images were known. Accordingly often several calibrations for a single camera are available. Consequently the more stable mean value of all these estimations is computed.

So far, the used self-calibration algorithm for an image sequence was introduced. The next section will discuss the achieved results for an image sequence of a synthetic scene.

5.1.2 Synthetic scene

In order to have exact ground truth data for the evaluation of the proposed approach the synthetic scene shown in figure 5.1 (a) is used to generate an image sequence of a purely rotating and zooming camera.

The sequence used for evaluation simulates a camera that rotates about its optical center with a pan of 0.5° and a tilt of 0.2° between two consecutive images of the generated image sequence. The camera center was chosen to be in front of the scene and the camera calibration matrices K_i with $i = 1, \dots, 46$ of the cameras were

$$K_i = \begin{bmatrix} f_i & 0 & 320 \\ 0 & 1 \cdot f_i & 240 \\ 0 & 0 & 1 \end{bmatrix}, \quad (5.1)$$

where the focal length varies as shown in figure 5.1 (b). It simulates a panning and tilting camera which zooms out of the scene to get a better overview. The image size for the image

¹Random Sampling Consensus

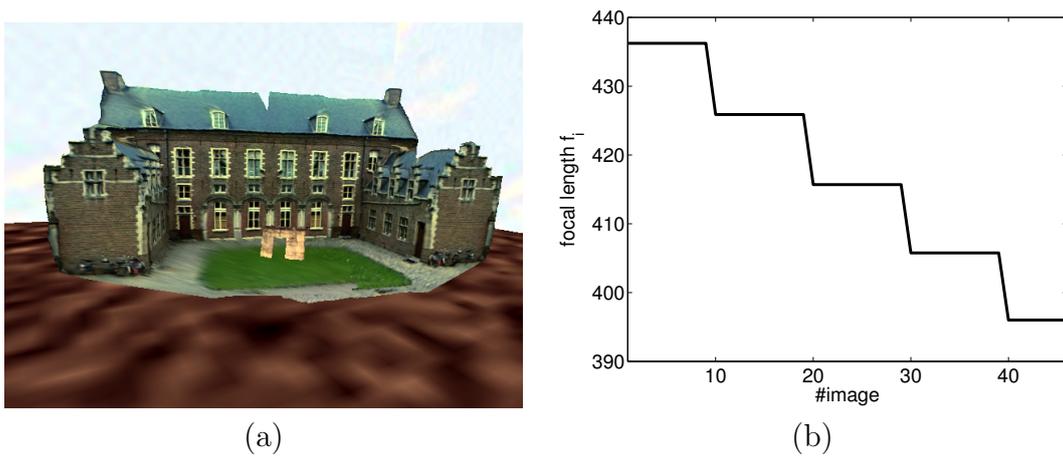


Figure 5.1: (a) Overview of the synthetic scene used to generate the image sequences for evaluation of the proposed approach. (b) Focal length of the cameras used for the image sequence of a synthetic scene in order to measure the performance of the proposed approach.

sequence is 640×480 pixel and the principal point is at the image center $[320, 240]$. The skew was chosen as zero in order to simulate a typical digital camera that naturally fulfills the zero skew constraint. In figure 5.2 two images of the image sequences are shown to illustrate the view of the camera and the used zoom. In the following the mean relative error is the error of the mean value of the multiple estimates of one intrinsic parameter averaged over all frames with respect to the ground truth value. The shown standard deviations give are the standard deviations of the multiple estimations. Accordingly the standard deviation can become larger than the maximal relative error.



Figure 5.2: First image of the image sequence of the synthetic scene on the left and the last image of the sequence on the right.

The resulting image sequence of the synthetic scene is processed using the algorithm explained in section 5.1.1. The following calibration incorporates the known rotation of the above described cameras. According to the skew which is equal to zero it is possible to estimate the remaining intrinsics: the focal lengths, the aspect ratios and the principal points from a single image pair with known relative orientation. Hence for each camera several calibrations to different neighbors were computed. The mean value and the standard deviations of these

multiple calibrations are shown in the figures 5.3 to 5.5 and summarized in table 5.1.

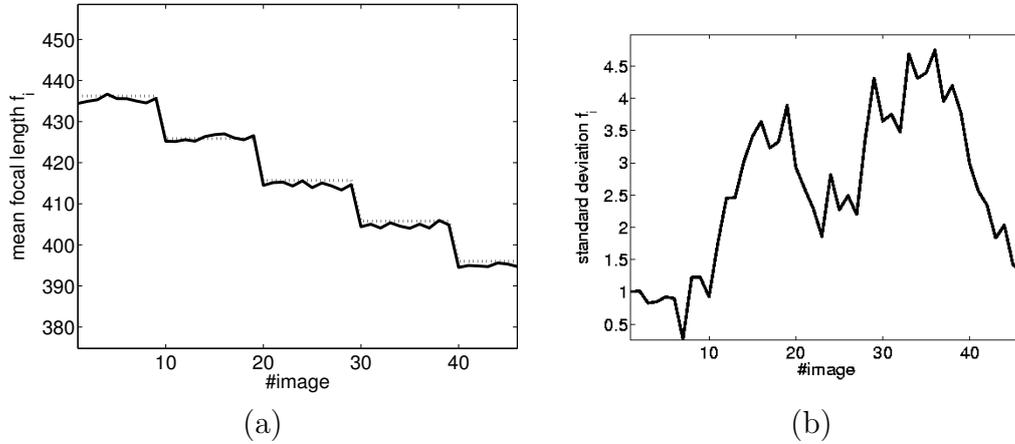


Figure 5.3: (a) Estimated focal length (solid line) of a rotating camera for an image sequence of the synthetic scene. The ground truth focal length is shown as dotted line. (b) Standard deviation of the estimated focal length.

In figure 5.3 the estimated mean focal length for the image sequence of the synthetic scene is presented. It can be seen that the estimation is very accurate. The mean relative error compared to the ground truth focal length is less than 0.5% and the maximal relative error is about 0.6%. Furthermore the standard deviation of the relative errors of the multiple estimates of the focal lengths from the different homographies are less than 1.2%. These measurements show that the proposed approach is robust against noise in the pixel positions of the corresponding points since the estimated point correspondences of the KLT-tracking are naturally disturbed.

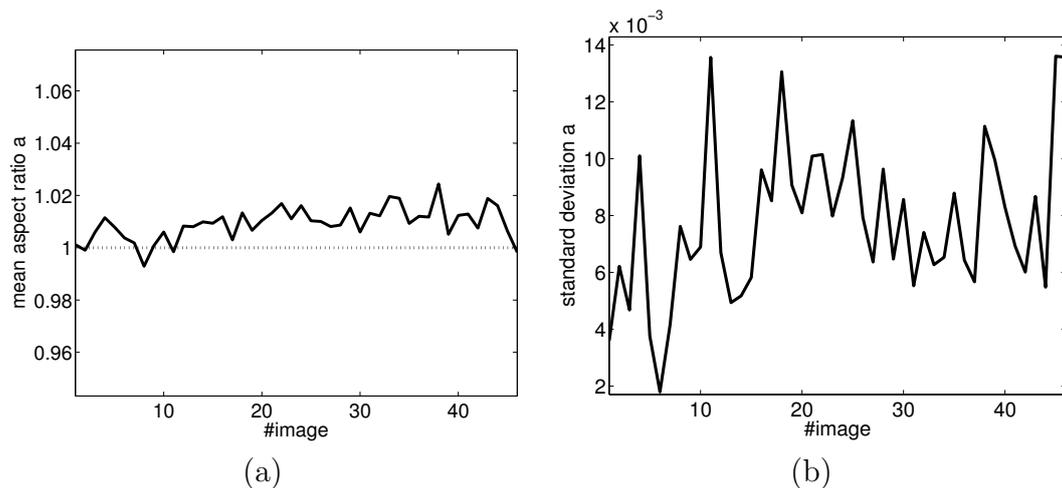


Figure 5.4: (a) Estimated aspect ratio (solid line) of a rotating camera for an image sequence of the synthetic scene. The ground truth aspect ratio is plotted as dotted line. (b) Standard deviation of the estimated aspect ratio.

In figure 5.4 the estimated aspect ratio is shown as well as the standard deviation of the estimated aspect ratios of the different images of the sequence. It can be seen that the aspect

ratio is estimated with a maximal relative error of less than 2.4% compared to the real aspect ratio. The mean relative error is about 1%. The standard deviation of the estimation is less than 1.4% and the mean standard deviation is 0.8%. These results show again that the proposed self-calibration method for cameras with known orientation is robust against error in the positions of the features.

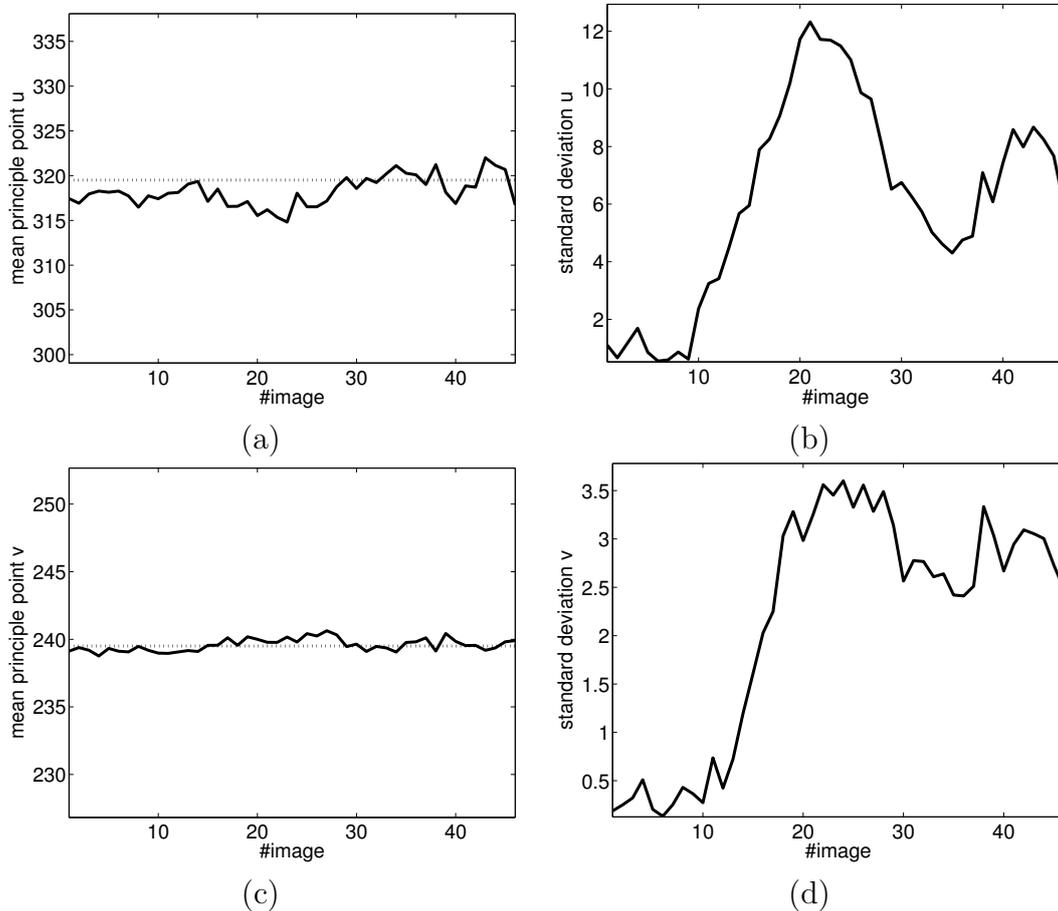


Figure 5.5: (a) Estimated principal point in x -direction (solid line) of a rotating camera for an image sequence of the synthetic scene. The ground truth principal point in x -direction is plotted as dotted line. (b) Standard deviation of the estimated principal point in x -direction. (c) Estimated principal point in y -direction (solid line) of a rotating camera for an image sequence of the synthetic scene. The ground truth principal point in y -direction is plotted as dotted line. (d) Standard deviation of the estimated principal point in y -direction.

In figure 5.5 the mean estimated principal point is shown and the standard deviations for each component are also shown. It can be seen that the principal point is also well estimated. The relative error for the x -component of the principal point is always less than 1.5% and the relative error for the y -component is less than 0.5%. The standard deviations for the relative error of the x -component are lower than 3.8% and for the relative error of the y -component the standard deviations are smaller than 1.6%. These results show that for noise in the feature positions the principal point is estimated robustly.

To summarize the experiments for self-calibration of an image sequence of a synthetic scene

with the proposed linear approach, it was shown that it is robust against noise in the positions of the corresponding features.

Additionally the robustness against errors in the measured rotation was evaluated. For these measurements the ground truth rotation data were disturbed with Gaussian angular noise of the rotation about the x -axis and the y -axis. The angular noise had standard deviation $\sigma_{angular}$ of up to half a degree. The measurements were performed with the same setup as above were done. The figures 5.6 to 5.8 show the estimated intrinsic parameters in dependence on the angular noise $\sigma_{angular}$.

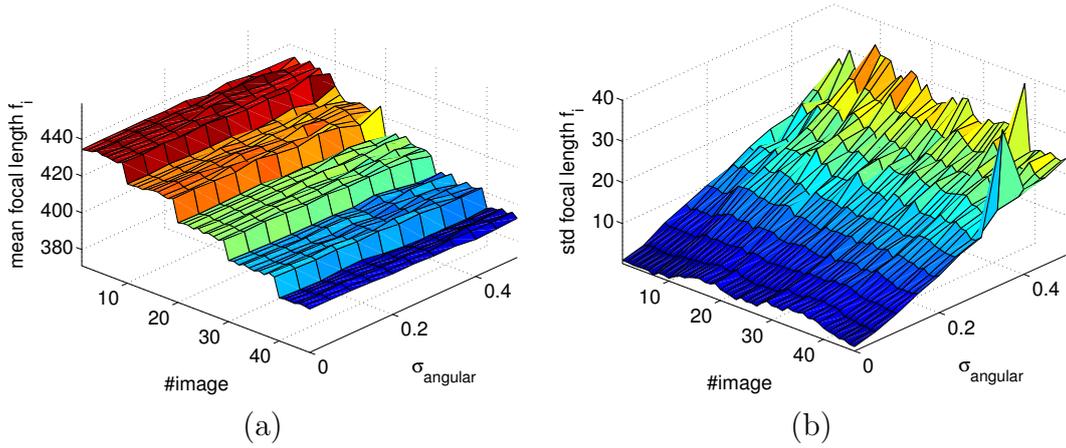


Figure 5.6: (a) Estimated focal length of a rotating camera for an image sequence of the synthetic scene in dependence on angular noise. (b) Standard deviation of the estimated focal length in dependence on angular noise.

The estimated mean focal length is shown in figure 5.6. It can be seen that the estimation is still very accurate even in the presence of angular noise. The mean relative error of the focal length is unchanged about 0.5%. The maximal relative error is in the presence of angular noise increased to 2% from 0.6% in absence of angular noise. The presence of noise can be seen in the increasing standard deviation of the estimated focal length. The standard deviation of 9.4% is approximately eight times higher than before. After the analysis in section 4.1.6 where the influence of rotation errors was analyzed this higher standard deviation was expected because the analysis leads to the result that the mean value is not disturbed and the standard deviation of the error will increase.

In figure 5.7 the estimated aspect ratio is shown as well as the standard deviation of the estimated aspect ratios. It can be seen that the mean relative error of the aspect ratio is nearly unchanged (0.8% compared to 1.0%), the standard deviation is increased in presence of noise from a mean standard deviation of 0.8% to a standard deviation of 8% of the ground truth aspect ratio.

The measurements of the principal point are plotted in figure 5.8. These measurements show the same tendencies as the measurements for the focal length and the aspect ratio. The mean relative error is only slightly changed and the standard deviation is increased in the presence of noise. A summary of these measurements is given in table 5.1.

These measurements of the robustness of the proposed approach against errors in the mea-

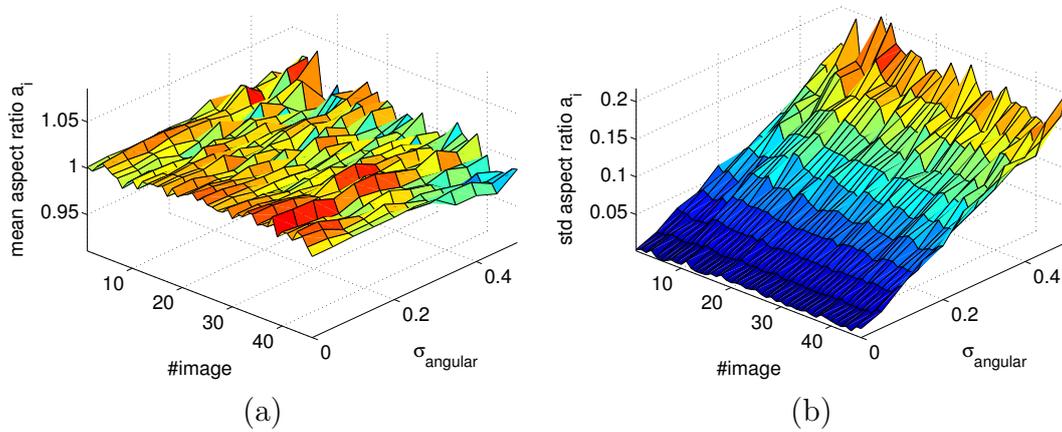


Figure 5.7: (a) Estimated aspect ratio of a rotating camera for an image sequence of the synthetic scene in dependence on angular noise. (b) Standard deviation of the estimated aspect ratio in dependence on angular noise.

parameter	pixel noise		pixel & angular noise	
	relative error	standard deviation	relative error	standard deviation
f	0.5%	1.2%	0.5%	9.4%
a	1.0%	0.8%	0.8%	8.0%
u	1.5%	3.8%	1.5%	6.5%
v	0.5%	1.6%	0.4%	3.0%

Table 5.1: Measured relative errors for the calibration of an image sequence of a synthetic scene captured by a purely rotating camera. The results of the second and third column were computed with undisturbed rotations. The last two columns contain the measurements for present noise in the rotation.

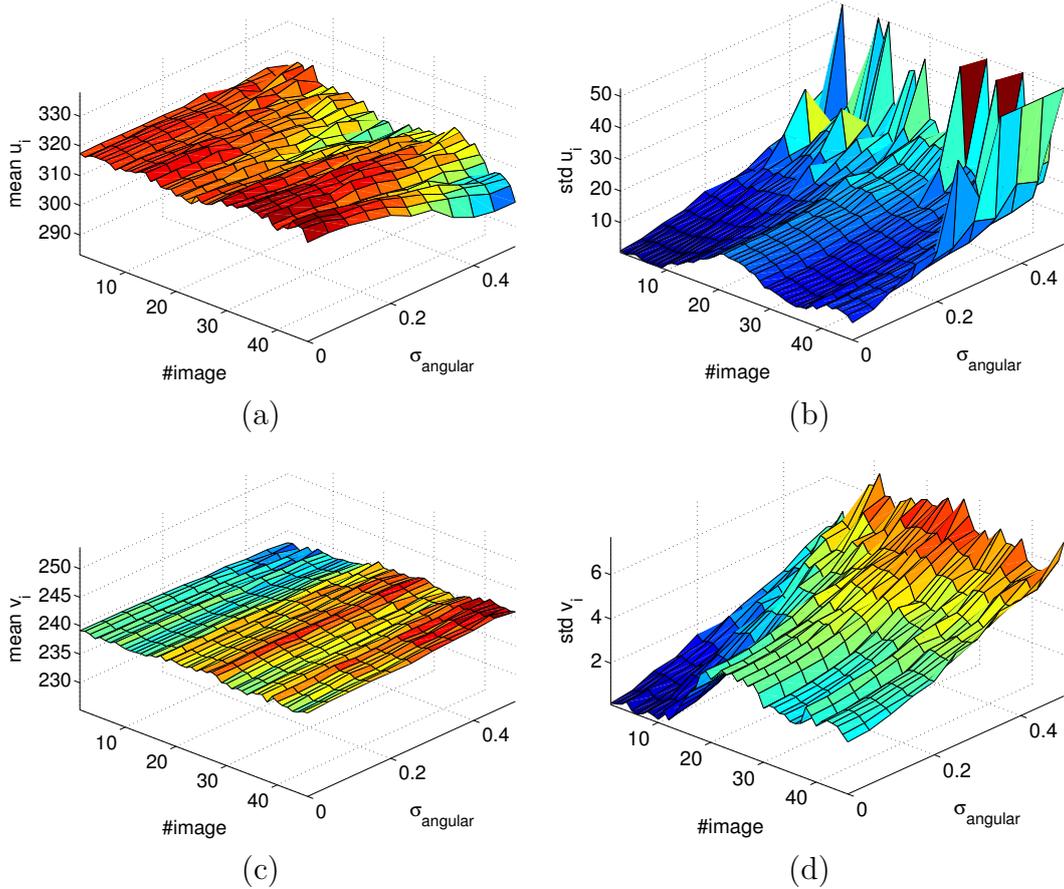


Figure 5.8: (a) Estimated principal point in x -direction of a rotating camera for an image sequence of the synthetic scene in dependence on angular noise. (b) Standard deviation of the estimated principal point in x -direction in dependence on angular noise. (c) Estimated principal point in y -direction of a rotating camera for an image sequence of the synthetic scene in dependence on angular noise. (d) Standard deviation of the estimated principal point in y -direction in dependence on angular noise.

sured rotation show that the mean value of multiple estimations of the intrinsic parameters is robust. Accordingly the above measurements verify the results of the sensitivity analysis of section 4.1.6. After this evaluation of the proposed approach on synthetic data the next sections will measure the robustness of the new self-calibration technique for purely rotating cameras with known orientation on a real image sequence.

5.1.3 Real image sequence with constant camera parameters

The next evaluation was done with a sequence taken by a consumer video conferencing pan-tilt-zoom camera (Sony EVI D-31). The camera is panning and tilting during the sequence consisting of 22 frames. Two frames of the sequence are shown in figure 5.9.



Figure 5.9: Two example images from the sequence of a panning and tilting video-conferencing camera.

The intrinsic parameters of the camera were measured beforehand using the calibration of Zhang [Zha99, Zha00]. It provides an approximate ground truth information for the camera. The manual calibration itself contains measurement errors therefore it is not "real" ground truth but it is useful for the evaluation of the achieved results. The calibration of Zhang requires several images of a planar calibration object in order to compute the intrinsic camera parameters and the radial distortion of the camera. The camera calibration matrix delivered by the manual calibration was given by

$$K_i = \begin{bmatrix} 940 & 0 & 384 \\ 0 & 940 & 288 \\ 0 & 0 & 1 \end{bmatrix}, \quad (5.2)$$

for the used zoom of the camera. The fact that the intrinsic parameters of the camera are constant was not exploited during the self-calibration process, i.e. the general self-calibration equations (B.1) were employed.

The camera rotation was taken from the camera motor data, which means the angles of the camera were computed from the position of the step motor. Therefore the rotation error depends on the positioning accuracy of the pan-tilt head which is in the range of less than 0.5° for each axis.

The radial distortion of the camera can be compensated beforehand by applying image based techniques like proposed in [Kan00, TM00] without any knowledge of the intrinsic calibration of the cameras. Furthermore the following self-calibrations assumed a skew equal to zero in order to calculate calibrations from a single image pair. The computed intrinsic parameters are shown in figure 5.10 to 5.12. For each image the homographies to five other images with a minimum pan angle of 1° and a minimum tilt angle of 1° were computed. Hence for each image 5 to 10 calibrations can be computed using the proposed self-calibration (4.3). The charts for the calibration results show the mean for all calibrations of the estimated intrinsics and the standard deviations of these multiple estimations for each intrinsic camera parameter. A summary of the achieved precisions is given in table 5.2.

In figure 5.10 the results for the focal length estimations are shown. They have a maximal relative error of 5% compared to the manual calibrated focal length and the mean relative

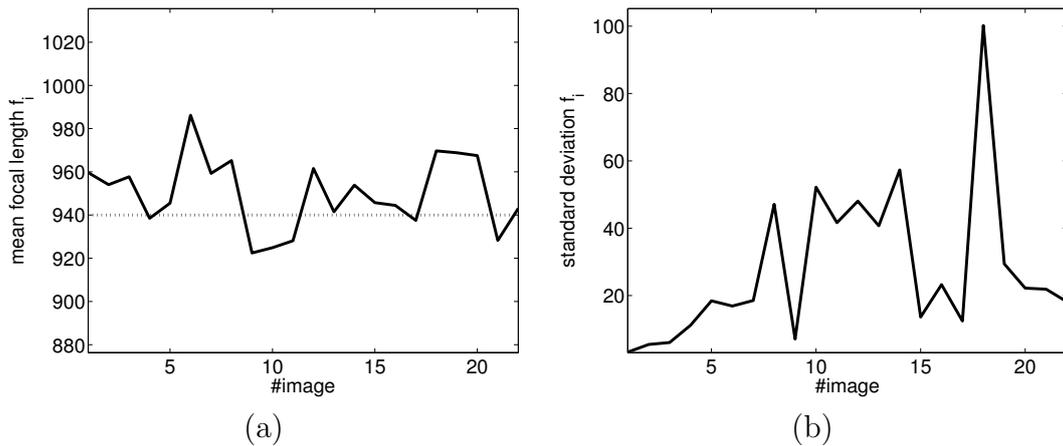


Figure 5.10: (a) Mean focal lengths computed with the self-calibration approach for a rotating camera from section 4.1.1. (b) Standard deviations of the estimated focal lengths.

error is about 1%. The mean standard deviation of all measurements is 2.9%. The maximal standard deviation of about 10.6% is caused by one bad estimation for image 18. Accordingly the results for the estimation of the focal length show that the proposed technique is robust against noise of the rotation and robust in terms of errors that are naturally contained in the estimated homographies.

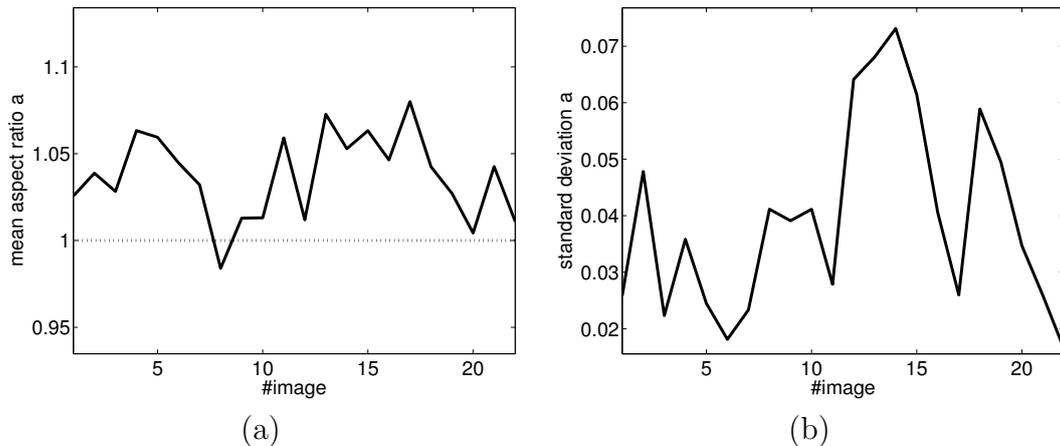


Figure 5.11: (a) Mean aspect ratios computed with the self-calibration approach for a rotating camera from section 4.1.1. (b) Standard deviations of the estimated ratios.

In figure 5.11 the estimated aspect ratio of the cameras is presented. It can be seen that the estimation of the aspect ratio is robust against noise in the rotation of up to 0.5° in each axis and the position error of the features employed for homography estimation. The maximal relative error of the aspect ratio is 8% with respect to the manually calibrated aspect ratio. With 3.7% the mean relative error is slightly higher than the error for the focal length of the camera. The mean standard deviation of the aspect ratio is about 4%. These measurements show that the new linear approach estimates the aspect ratio accurately.

Figure 5.12 shows the results of the self-calibration for the principal point $[u, v]$. As before the x -component is estimated slightly worse than the y -component. The maximal relative error

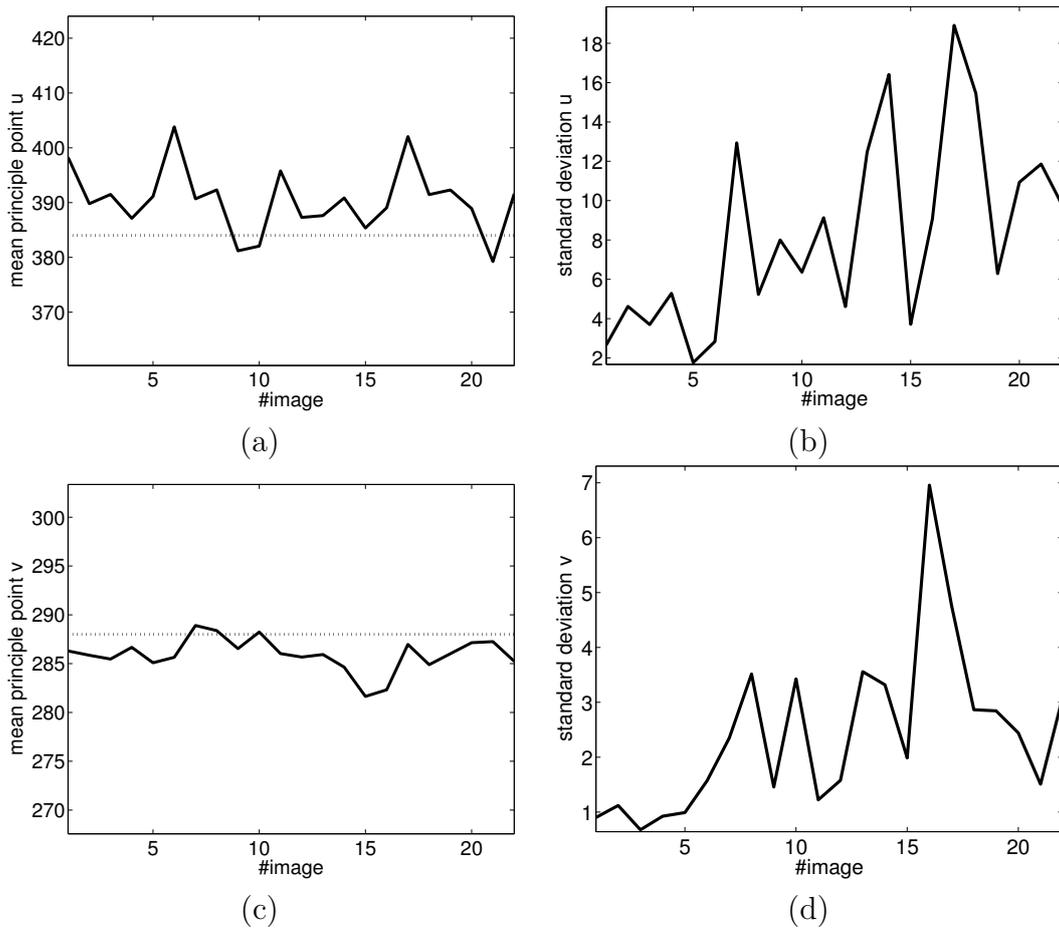


Figure 5.12: (a) Mean principal point in x -direction computed with the self-calibration approach for a rotating camera from section 4.1.1. (b) Standard deviation of the estimated principal point in x -direction. (c) Mean principal point in y -direction computed with the self-calibration approach for a rotating camera from section 4.1.1. (d) Standard deviation of the estimated principal point in y -direction.

for the x -component is 5.5% and for the y -component of the principal point the maximal error is 2.2%. The mean relative error for the x -component is 2% and for the y -component is 1%. A summary of the measurements is given in table 5.2.

Summarizing these measurements on a real scene show that with the new technique reliable calibrations of the cameras can be computed. The error is in a range that is sufficient for the most applications of self-calibration. In the next section the performance of the new method for a camera with varying focal length will be shown.

5.1.4 Real image sequence with varying parameters

In this section the proposed approach will be evaluated with a sequence of a rotating and zooming camera. It was the same pan-tilt-zoom camera used in the previous section. For the used zoom steps the focal length were again calibrated beforehand using the technique of Zhang [Zha99, Zha00] and the radial distortions of the images were compensated beforehand.

parameter	relative error	standard deviation
f	1.0%	2.9%
a	3.7%	4.0%
u	2.0%	1.9%
v	1.0%	0.8%

Table 5.2: Measured relative errors for the calibration of an image sequence of a real scene captured by a purely rotating camera.

Example images of the sequence are shown in figure 5.13.



Figure 5.13: Two example images from the sequence of a panning, tilting and zooming video-conferencing camera.

The manual calibration provided camera calibration matrices K_i

$$K_i = \begin{bmatrix} f_i & 0 & 384 \\ 0 & f_i & 288 \\ 0 & 0 & 1 \end{bmatrix}, \quad (5.3)$$

with $f_i \in [1206, 1085]$ pixel. The ground truth information for the focal length is shown in figure 5.14. The homographies used for calibration were computed between each image and all other images with a difference of at least 1.5° in the pan angle and a difference of at least 1.5° in tilt angle. According to these constraints up to twelve homographies for each image could be estimated. The mean values of the different calibrations from these homographies are shown in the figures 5.14 to 5.16 as well as the standard deviations of the multiple estimations. Table 5.3 summarizes the measurements of the relative errors and the corresponding standard deviations of the intrinsic parameters.

In figure 5.14 the estimated focal lengths are shown together with the corresponding standard deviations. The mean relative error of the focal length is about 4% compared to the manual calibration. The mean standard deviation is about 10.0% of the "true" focal length. These results show again that the mean value of multiple estimations is robust against noise. The reached precision shows that the new technique can be used in a wide range of applications.

Figure 5.15 shows the computed aspect ratio for the image sequence. The mean relative error in aspect ratio is about 7.8% with respect to the manual calibrated aspect ratio. The

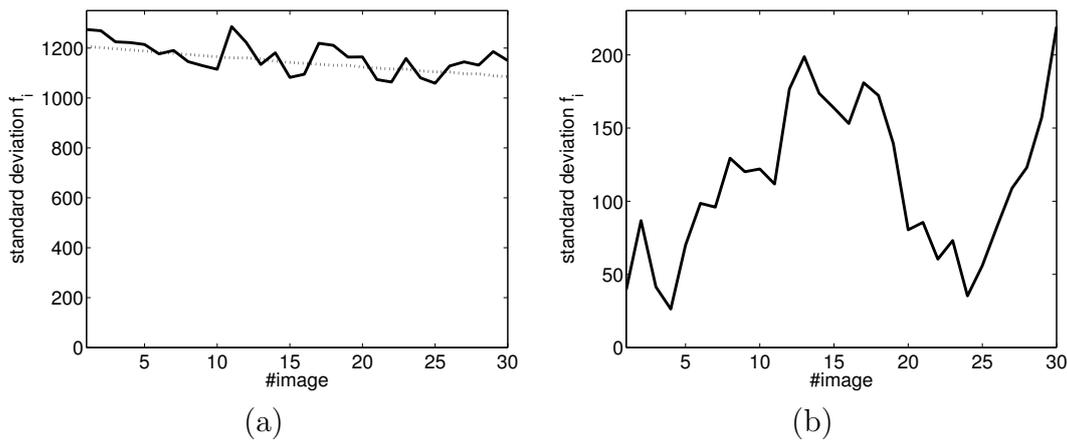


Figure 5.14: (a) Mean estimated focal lengths (solid line) compared to the manually calibrated focal length (dotted line). (b) Standard deviation of the estimated focal length.

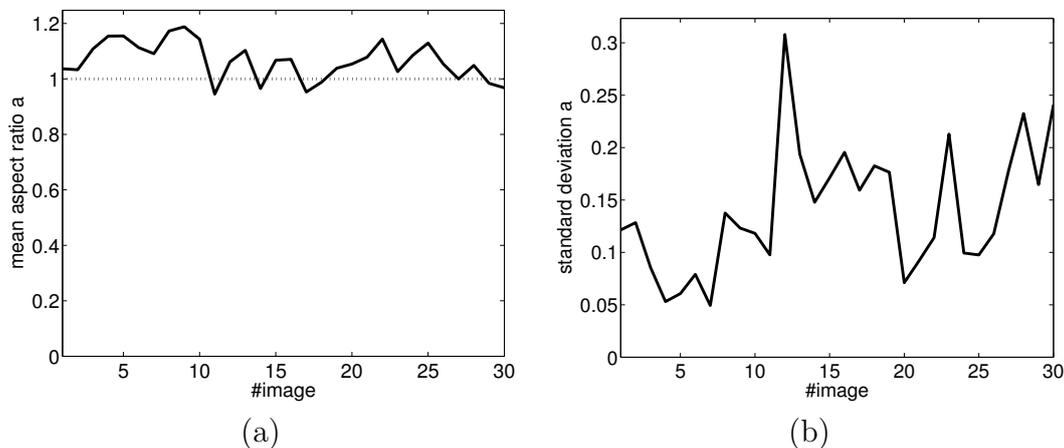


Figure 5.15: (a) Mean estimated aspect ratios (solid line) compared to the manually calibrated aspect ratios (dotted line). (b) Standard deviation of the estimated aspect ratios.

standard deviation of the computed aspect ratio is 14%. These measurements show a reasonable quality of the mean value of the multiple estimations of the aspect ratio.

The estimations for the principal point of the cameras are shown in figure 5.16. It can be seen that the estimated principal point is estimated with a mean relative error of 4.8% in both directions. The mean standard deviations of the estimated principal points are about 8% of the ground truth value. This shows that the new technique delivers a reliable estimation of the principal point. The higher level of the errors compared to the previous section is mainly a result of the homography estimation which had larger reprojection errors than in the previous section.

So far, the calibration of a purely rotating camera was evaluated. It showed that the proposed self-calibration is robust against noise in a range that is reasonable for most applications. The next section will evaluate the calibration of a freely moving camera with known orientation.

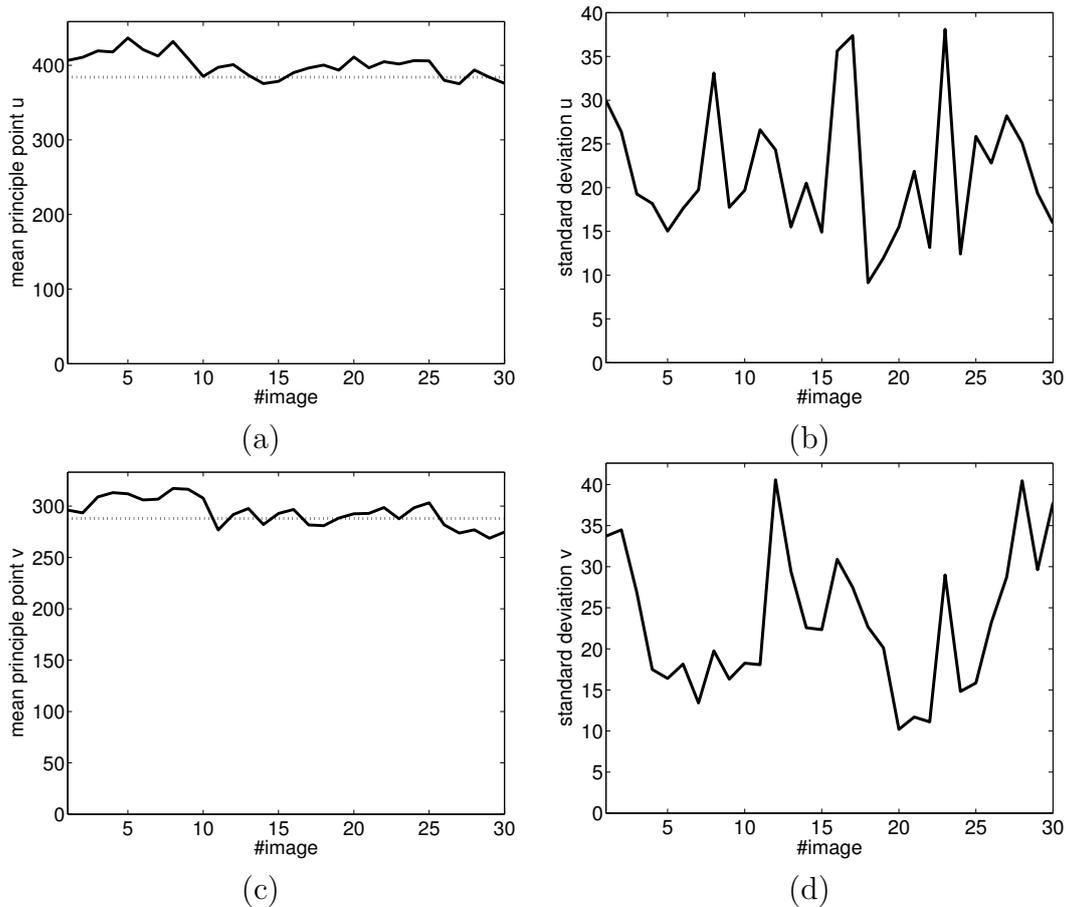


Figure 5.16: (a) Mean estimated principal point in x -direction (solid line) compared to the manually calibrated principal point in x -direction (dotted line). (b) Standard deviation of the estimated principal point in x -direction. (c) Mean estimated principal point in y -direction (solid line) compared to the manually calibrated principal point in y -direction (dotted line). (d) Standard deviation of the estimated principal point in y -direction.

5.2 Experimental results for a freely moving camera

In this section the self-calibration of a freely moving camera with known orientation from section 4.2.1 will be evaluated. First of all an image sequence of a synthetic scene is used to measure the performance of the proposed approach. Afterwards an image sequence of a real scene is used to investigate the performance of the novel technique. The next section will introduce the algorithm applied for self-calibration of a freely moving camera.

5.2.1 Algorithm for self-calibration of a freely moving camera

For a self-calibration of all cameras of an image sequence exploiting the known camera orientation the following steps are required.

1. Correspondence estimation
2. Robust computation of the Fundamental matrix

parameter	relative error	standard deviation
f	4.0%	10.0%
a	7.8%	14.0%
u	4.8%	8.0%
v	4.8%	8.0%

Table 5.3: Measured relative errors for the calibration of an image sequence of a real scene captured by a purely rotating and zooming camera.

3. Estimation of the intrinsic camera parameters by exploitation of the estimated Fundamental matrices and the known camera orientation.
4. Combination of multiple estimations if necessary.

These steps were done for the self-calibrations as shown in the following paragraphs. In the next paragraphs the details of each performed step will be discussed.

Correspondence estimation Similar to the algorithm from section 5.1.1 for a rotating camera the correspondences were generated with the KLT-tracker [LK81, BM04].

Robust computation of the Fundamental matrices In order to avoid an adaption to wrong correspondences of the estimated Fundamental matrix a RANSAC approach is applied. It estimates the Fundamental matrices followed by a nonlinear optimization of the Fundamental matrix. More details about the robust estimation of a Fundamental matrix can be found in [TM97, TZ98].

Estimation of intrinsics The estimated Fundamental matrices were exploited to compute the intrinsic parameters of the cameras. The self-calibration is done with the novel approach from section 4.1 incorporating the known orientation of the camera.

Combination of multiple estimates The epipolar geometry between the different image pairs is estimated with a sufficient number of correspondences for each pair of cameras. Therefore several calibrations are estimated for each camera. According to the more stable mean value of the estimations as seen in section 4.2.3 it is used as the estimated intrinsic parameter. Furthermore this provides measurements for the standard deviation of the estimated parameters. These standard deviations will also be shown during the evaluations.

In the following sections the introduced algorithm is applied for self-calibration of all cameras of the image sequences. The next section will evaluate the performance of the proposed self-calibration method with an image sequence of a synthetic scene.

5.2.2 Synthetic scene

For evaluation the same synthetic scene was used as for the evaluations in section 5.1.2. An image of the scene is shown in figure 5.17. The calibration of the camera varied throughout the

image sequence. The camera calibration matrices of the cameras were given by

$$K_i = \begin{bmatrix} f_i & 0 & 320 \\ 0 & 1 \cdot f_i & 240 \\ 0 & 0 & 1 \end{bmatrix}, \quad (5.4)$$

with $f_i \in [396, 436]$ as shown in figure 5.17 (b).

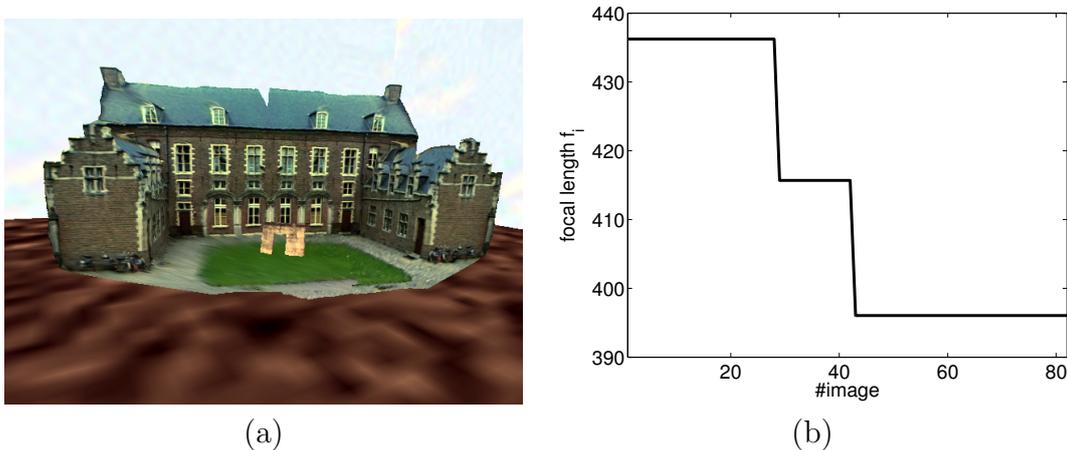


Figure 5.17: (a) Overview of the synthetic scene used to generate the image sequences for evaluation of the proposed approach. (b) Focal length of the cameras used for the image sequence of a freely moving camera of a synthetic scene used to measure the performance of the proposed approach.

The self-calibrations of the cameras of the image-sequence were performed using the algorithm introduced in section 5.2.1. The skew of the camera was chosen to be zero and the principal point was given as the image center in order to estimate the focal length and the aspect ratio of the camera from a single image pair. The results of the self-calibration are shown in figures 5.18 and 5.19. Furthermore the results are summarized in table 5.4.

In figure 5.18(a) the mean estimated focal length of the self-calibration is shown and in figure 5.18(b) the corresponding standard deviation of the multiple measurements are shown. The maximal relative error of the estimated focal length is about 3.5% of the true focal length and the mean relative error is about 2%. The corresponding standard deviation has a maximum of 5.0% and a mean value of less than 0.6%. These estimations of the camera calibrations show that the proposed approach is robust against noise in the correspondences since the achieved errors in the estimated mean focal length are small.

Figure 5.19 (a) shows the estimated aspect ratio for the cameras of the image sequence of the synthetic scene. The estimated aspect ratio has a maximal error of about 9.5% and a mean relative error of 5%. It can be seen that the aspect ratio is for most of the cameras slightly overestimated. The standard deviations of the estimated aspect ratio are shown in figure 5.19 (b). The maximal standard deviation of the aspect ratio is 3.6% and the mean standard deviation of the estimated aspect ratio is 0.6%. These measurements show that the estimation of the focal length and the aspect ratio with the proposed approach for a freely moving camera are robust against noise in the correspondences.

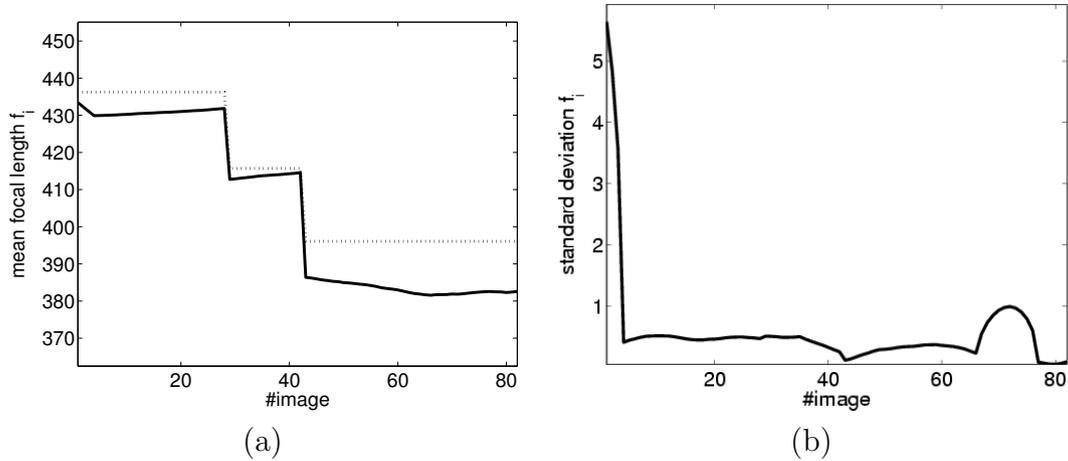


Figure 5.18: (a) Estimated focal length (solid line) of a freely moving camera for an image sequence of the synthetic scene. The ground truth focal length is shown as dotted line. (b) Standard deviation of the estimated focal length.

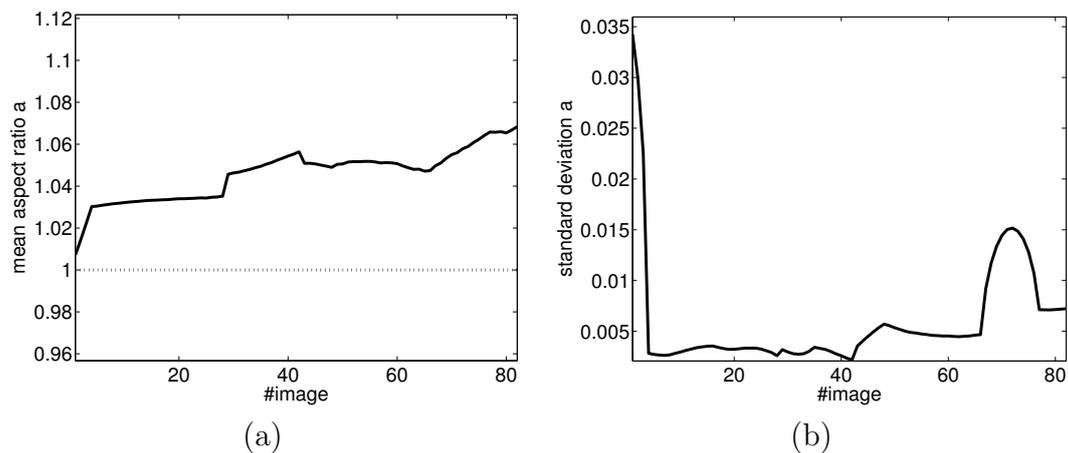


Figure 5.19: (a) Estimated aspect ratio (solid line) of a freely moving camera for an image sequence of the synthetic scene. The ground truth aspect ratio is plotted as dotted line. (b) Standard deviation of the estimated aspect ratio.

To measure the robustness against errors in the measured rotation, angular noise was added to the ground truth rotation. The angular noise had a Gaussian distribution with a variance σ_{angular} of up to one degree. With these noisy rotation data the above measurements were repeated for different standard deviations σ_{angular} . The results of these measurements are shown in figure 5.20 and figure 5.21. A summary of the achieved results is given in table 5.4.

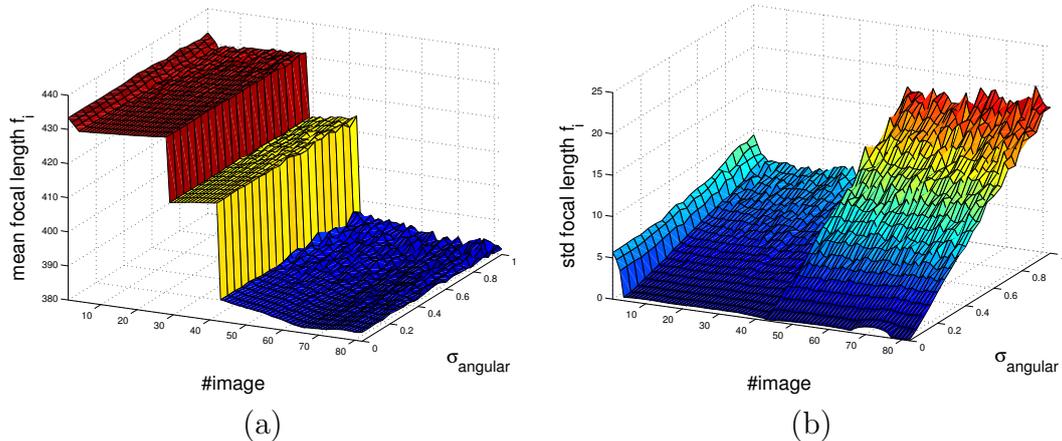


Figure 5.20: (a) Estimated focal length (solid line) of a freely moving camera for an image sequence of the synthetic scene in dependence on the angular noise of the measured rotation. (b) Standard deviation of the estimated focal length in dependence on the angular noise of the measured rotation.

The achieved results for the mean estimated focal lengths are shown in figure 5.20(a) and the correlated standard deviation of the focal length is plotted in figure 5.20(b). It can be seen that the mean value of the estimated focal lengths is robust against the angular noise in the rotation as previously deduced in section 4.2.4. The maximal relative error of the estimated focal length is about 4.3% of the true focal length. This is only slightly higher than for the above measurements with a noise free rotation where the maximal mean relative error was about 3.5%. The mean relative error is nearly unchanged with 2.2% compared with 2% for the measurements with noise free rotation data.

The influence of the noisy rotation is visible in the standard deviation of the estimated focal lengths. The standard deviation is increasing with an increasing noise level of the angular noise. The mean standard deviation of the focal length is now approximately 2.5% compared to 0.6% for the measurements with perfect rotation data.

Figure 5.21 (a) shows the measurements for the mean aspect ratio depending on the angular noise. Similar to the focal length it can be seen for the aspect ratio that the mean estimated value is nearly unchanged by the angular noise. The maximal relative error of the estimated aspect ratio is with 8.3% slightly lower than before. A slightly lower mean relative error of 4.1% was measured. Effects of the angular noise can be seen in the measured standard deviations of the multiple estimates. The maximal standard deviation is increased from 3.6% to 9.8%. Similar the mean standard deviation is increased from 0.06% to 2.9%.

From these measurements it follows that the proposed approach is also robust against errors in the rotations if it is possible to get multiple measurements of each intrinsic parameter.

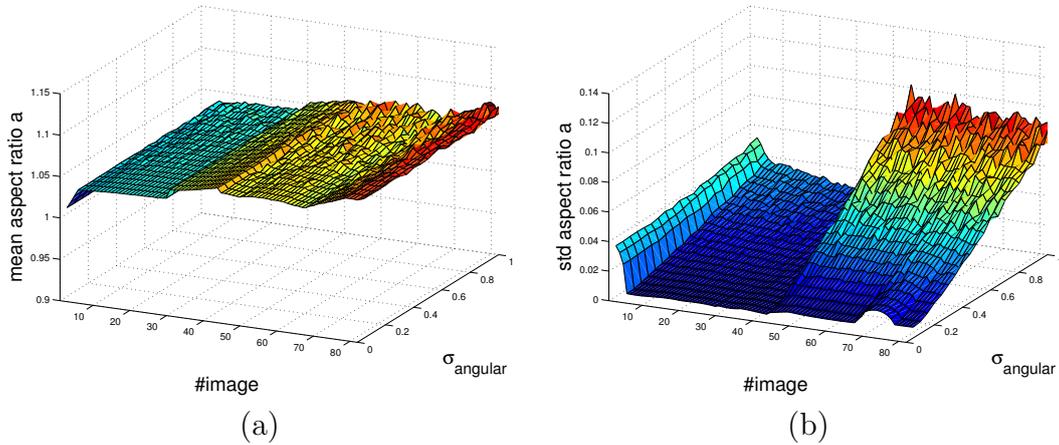


Figure 5.21: (a) Estimated aspect ratio (solid line) of a freely moving camera for an image sequence of the synthetic scene in dependence on the angular noise of the measured rotation. (b) Standard deviation of the estimated aspect ratio in dependence on the angular noise of the measured rotation.

parameter	pixel noise		pixel & angular noise	
	relative error	standard deviation	relative error	standard deviation
f	2.0%	0.6%	2.2%	2.5%
a	5.0%	0.6%	4.1%	2.9%

Table 5.4: Measured relative errors for the calibration of an image sequence of a synthetic scene with a freely moving and zooming camera. The results of the second and third column were computed with undisturbed rotations. The last two columns contain the measurements for present noise in the rotation.

Then the mean value of the multiple measurements can be used as a robust estimation of the parameter value. This was deduced earlier in section 4.2.4 where the influence of the rotation error was analyzed. After this detailed evaluation on image sequences of synthetic scenes in the next section the approach will be tested on an image sequence of a real scene.

5.2.3 Real image sequence

The results of the evaluation of the proposed self-calibration for a sequence of a real scene will be discussed in this section. The pan-tilt-zoom camera Sony EVI D-31 described in section 5.1.3 was used to capture the sequence consisting of 35 images. It was moved and rotated during the sequence. The orientation of the camera was measured using the motor information of the pan-tilt-zoom unit. As approximation of ground truth camera parameters the manual calibration by Zhang [Zha00] was used. The intrinsic parameters of the camera remain constant during the sequence and the camera calibration matrix K was given by

$$K_i = \begin{bmatrix} 940 & 0 & 384 \\ 0 & 940 & 288 \\ 0 & 0 & 1 \end{bmatrix}.$$

The additional constraints from the constant intrinsic parameters were not used for the calibration process. In order to get a unique self-calibration a known principal point at the image center was used as previous knowledge. Two images of the sequence are shown in figure 5.22. The achieved calibrations with the corresponding standard deviations are sketched in figure 5.23 and 5.24. Table 5.5 shows a summary of the the achieved calibration results.



Figure 5.22: Example images of the sequence of a real scene used to calibrate the cameras that captured the sequence.

Figure 5.23 (a) shows the computed mean focal length for the cameras of the image sequence. The mean absolute relative error of the calibration is 3.5% with respect to the approximate ground truth from the manual calibration. The maximal absolute relative error is about 7.9% of the true value. From the standard deviation in figure 5.23 (b) of the multiple estimates of the focal length it can be seen that the estimation is reliable with a maximal standard deviation of 3.3% of the ground truth focal length of 940 pixel and a mean standard deviation of 1.6%.

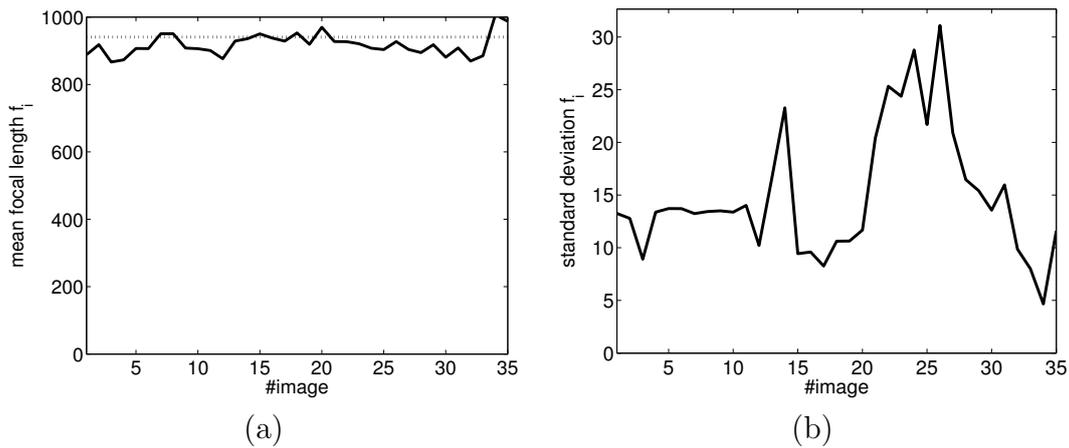


Figure 5.23: (a) Linear self-calibration of the focal length for a sequence of a real scene. (b) Standard deviation of the multiple estimates of the focal length for each frame.

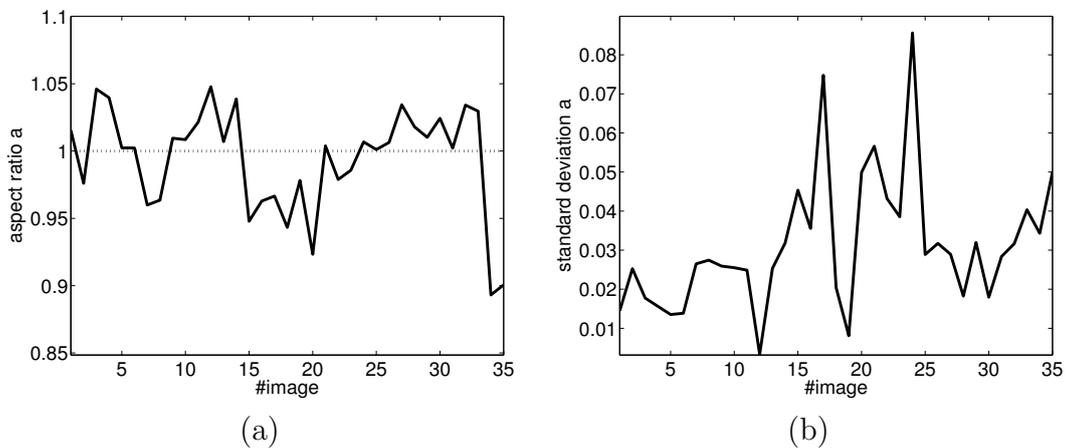


Figure 5.24: (a) Linear self-calibration of the aspect ratio for a sequence of a real scene. (b) Standard deviation of the multiple estimates of the aspect ratio for each frame.

The achieved self-calibrations for the mean aspect ratio are shown in figure 5.24 (a). The mean relative error of the estimated aspect ratio is 3% with respect to the ground truth aspect ratio of 1. The maximal relative error of the computed aspect ratio is 10.7%. The corresponding standard deviations of the multiple estimates of the aspect ratio are plotted in figure 5.24 (b). The maximal standard deviation of the estimated aspect ratio is 8.5% and the mean standard deviation is 3.1%. These self-calibrations showing that the proposed approach provides reliable estimates of the intrinsic parameters of a freely moving camera.

parameter	relative error	standard deviation
f	3.5%	1.6%
a	3.0%	3.1%

Table 5.5: Measured relative errors for the calibration of an image sequence of a real scene captured by a freely moving camera.

This chapter evaluated the achieved self-calibration results of arbitrary moving cameras with known orientation. At first the self-calibration of purely rotating camera with known orientation was investigated with image sequences of a synthetic scene with known ground truth information. The analysis showed that the proposed approach is robust against errors caused by the correspondence and the homography estimation. Afterwards the proposed approach was used to compute the self-calibration of two image sequences of real scenes. For these sequences a reference calibration was measured manually using the approach of Zhang[Zha00]. The evaluation showed that the calibration error of the new approach is small for example the mean absolute error of the focal length was about 4%.

The second proposed approach for the self-calibration of freely moving cameras with known orientation was also tested on sequences of synthetic scenes with known ground truth information and on an image sequence of a real scene. The evaluation on the synthetic scene showed that the proposed method is robust in terms of errors caused by the correspondence generation and the estimation of the Fundamental matrix. The evaluation on an image sequence of a real scene showed that also errors in the measurement of the rotation can be compensated by the proposed technique. The mean absolute error of the estimated focal length was about 3.5%.

These evaluations show that the proposed approach computes reliable self-calibrations for freely moving cameras and for purely rotating cameras. The achieved accuracy of the calibration depends mainly on measurement errors of the camera orientations as deduced in sections 4.1.6 and 4.2.4. Furthermore it could be seen that the measurement errors of common techniques for correspondence estimation are sufficient for accurate self-calibrations. Additional examples for the evaluation of the proposed approach with other sequences can be found in [FK03c, FK03b].

Chapter 6

Conclusions

This dissertation focused on the self-calibration of cameras needed for example in uncalibrated structure from motion in order to compute meaningful metric reconstructions. The need for the reconstructions has increased over the last years with the fast development in computer graphics hardware. The proposed approaches improve this model generation process by providing precise camera calibrations.

First of all the existing approaches for camera self-calibration were analyzed. From this analysis it was seen that the existing approaches do not incorporate additional information and therefore suffer under noisy data. This sometimes leads to the disadvantage that the calibration can't be computed at all. Since in many future applications more information about the camera motion is at hand, it should be used to improve the self-calibration of the camera.

In this contribution two novel self-calibration approaches were proposed that exploit knowledge of external rotation information in conjunction with image-based estimation of homographies and Fundamental matrices. These approaches solve the self-calibration linearly by utilizing the known orientation of the camera which improves the self-calibration performance. Furthermore it was explained that this method in contrast to previous approaches is always able to compute a solution even for noisy data.

The first proposed approach for the self-calibration of purely rotating cameras overcomes the given limitations of previous approaches in the number of varying intrinsic parameters. This was achieved by avoiding rotation invariant calibration objects like the absolute conic or the absolute quadric in order to exploit the known orientation of the camera. It was shown that the novel approach can be transferred into the previous approaches by losing available information due to the transformation. The robustness in terms of noise of the available camera orientation and the estimated homographies was measured using synthetic point data. These measurements showed that the self-calibration is rather stable in a range of noise that is provided by existing techniques for correspondence generation and current consumer rotation sensors. Afterwards synthetic and real image sequences with ground truth data are used to evaluate the performance of the proposed approach. Experiments on this also showed that the new technique delivers good calibration results with measurement errors of less than 4% of the focal length.

For a freely moving camera a second self-calibration approach was proposed that utilizes the known orientation of the camera and the image-based estimated Fundamental matrix. This approach computes the intrinsic camera parameters exploiting the structure of the Fundamental

matrix under the assumption of one available constraint on the intrinsic parameters. Experiments showed that the new technique delivers good calibration results with measurement errors of less than 3.5% of the focal length. The advantage of the new technique is that no global consistent frame is needed in order to calibrate the cameras. Furthermore it was explained how an available consistent frame can be exploited to overcome the limitation of one known intrinsic parameter of the proposed approach.

The proposed approach was embedded in the context of the previously introduced approaches for self-calibration of a freely moving camera. It was deduced that the proposed approach can be transformed into the same equations as previous proposed approaches used by loosing constraints on the rotation of the camera. Synthetic point data was used to measure the robustness of the proposed approach. It showed that the estimated self-calibration is rather stable according to the accuracies provided by existing estimation techniques of the Fundamental matrix and consumer orientation sensors. Finally the performance of the proposed self-calibration was evaluated using synthetic and real image sequences with ground truth information. This evaluation showed that the new technique provided good results for the calibration of freely moving cameras.

Future work The proposed techniques use the homography respectively the Fundamental matrix in order to calibrate the intrinsic camera parameters. Nowadays there are a variety of techniques available that estimate the relations between two consecutive frames of an image sequence in real time. According to the low computational complexity of the proposed approach it is possible to provide real-time camera calibration using the novel approaches. This real-time self-calibration will be a major step into the direction of uncalibrated real-time structure from motion.

It seems natural for a freely moving camera for example in cars or mounted on robots also to incorporate the available approximate translation. Using this translation or maybe only the direction of the motion promises to overcome the left limitation of one known parameter in the proposed self-calibration of freely moving cameras.

Appendix A

Projective geometry and camera model

This appendix introduces the camera model used as well as the geometric spaces exploited to model the camera imaging process. First of all the pinhole camera model is introduced in section A.1 which models the imaging process. The pinhole camera model for the projection of 3D scenes onto the image plane is used in the thesis to describe the camera transformation. Afterwards the projective geometry will be introduced in section A.2 to linearize the camera projection model. Further in section A.2.9 additional nonlinear effects are introduced that are caused by the lenses of real cameras and can't be modeled with the linear pinhole camera model in projective space.

A.1 Camera model

This section introduces the camera model for a perspective camera that is used throughout the thesis. In general the camera can be seen as a device that projects the 3D world onto a two dimensional plane (camera image plane). This image plane can be for example a film in a camera or in case of a digital camera a CCD-sensor¹. A model for the transformation from the three dimensional world into the two dimensional image is the pinhole camera model². It will be described in section A.1.1. The pinhole camera model will be extended in section A.2.7 to model the properties of the imaging process for digital images since then a pixel coordinate system will be needed for computations.

In general a camera is described by two different sets of parameters. The first set consists of the **extrinsic** parameters. They are the position of the camera in the world and the orientation of the camera according to the world coordinate system. The other set of parameters is the **intrinsic** parameters. These parameters model the lens and the digitalization properties like zoom and pixel size.

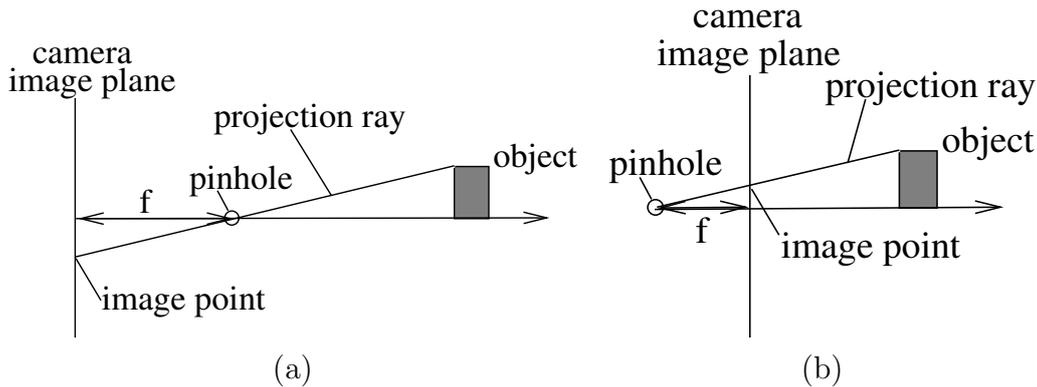


Figure A.1: (a) pinhole camera model for the projection of an object onto the camera image plane, (b) pinhole camera with the image plane in front of the pinhole to avoid mirroring of the objects image.

A.1.1 Pinhole camera model

The pinhole camera is made of an infinitely small hole and an image plane placed behind the hole as shown in figure A.1(a). Accordingly the image of a pinhole camera is produced from a light ray (projection ray) that passes through the pinhole of the camera and intersects the camera image plane. The point of intersection of the projection ray and the camera image plane is the image point of the corresponding 3D scene (object) point.

The pinhole of the camera will further be denoted as the optical center of the camera. The axis perpendicular to the camera image plane that passes through the pinhole is called the optical axis of the camera. The direction of the optical axis is the viewing direction of the camera. The intersection point of the optical axis and the camera image plane is called the principal point of the camera. The distance f between the pinhole and the camera image plane is called the focal length of the camera. The focal length determines the image size of the object. The larger the focal length the larger the image of an object. It can be seen in figure A.1(a) that the object image is mirrored on the image plane. To avoid this the camera image plane is often placed in front of the optical center of the pinhole camera as shown in figure A.1(b). Geometrically there is no difference between both models. In the following it will not be distinguished between them.

Canonical camera In order to model the imaging process of a camera different right-handed coordinate systems are needed

- **World coordinate system** or scene coordinate system is the coordinate system that is used for the scene description or measurements in the scene. It is a coordinate system for a three dimensional space. Usually it is the coordinate system of the real world.

¹CCD is a abbreviation for charge-coupled device. It is an integrated circuit containing an array of linked, or coupled, capacitors that integrates the photons of the light. The value of each array element is related to the gray value of the image.

²The pinhole camera is also known as "camera obscura".

- **Camera coordinate system** is the three dimensional coordinate system that is aligned to the location and orientation of the camera. This coordinate system is used to describe the scene in dependence on the camera position and orientation.
- **Camera image plane coordinate system** is a two dimensional coordinate system in the image plane. This coordinate system is used to describe the coordinates of the image points.

At first the relations between these three coordinate systems are described shortly. The camera is placed at an arbitrary position in the world coordinate system usually also with an arbitrary orientation. This means that the axes of the camera coordinate system are rotated with respect to the axes of the world coordinate system. The camera image plane coordinate systems position and orientation is usually described according to the camera coordinate system. The position of the image plane mainly depends on the focal length of the camera.

The canonical camera simplifies the relations between the three different coordinate systems.

Definition 1 (Canonical camera) *For a canonical camera the world coordinate system is chosen to be identical with the camera coordinate system. Furthermore the axes of the camera image plane coordinate system are identical with the x-axis and y-axis of the camera coordinate system. The image plane is placed at position $z = 1$. The camera image plane coordinate system is chosen in a way that the focal length f is equal to one.*

The projection of a 3D scene point $\bar{X} = [X, Y, Z]^T$ that is in front of the camera ($Z > 1$) is given for a canonical camera by

$$\bar{x} = \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \frac{X}{Z} \\ \frac{Y}{Z} \end{bmatrix}, \quad (\text{A.1})$$

where $\bar{x} = [x, y]^T$ is the image point in the camera image plane coordinate system. The image projection for a canonical camera is illustrated in figure A.2. This projection onto the image plane is called central projection. The central projection introduces perspective shortening in dependence on the depth Z of the scene point \bar{X} . That means that objects near to the pinhole have larger projections than objects of the same size further away from the image plane. The perspective shortening results from the normalization by the depth Z of the scene point \bar{X} in (A.1).

For a camera with a focal length f that is different from one (a not canonical camera), the image point is given as

$$x = f \frac{X}{Z}, \quad y = f \frac{Y}{Z}. \quad (\text{A.2})$$

Until now the imaging process of the camera was modeled for a camera at the origin of the world coordinate system. Usually the camera is moved around while taking an image sequence. In that case the image projection has to take the camera position and orientation into account. The image projection for a camera in general position is introduced in the next paragraph.

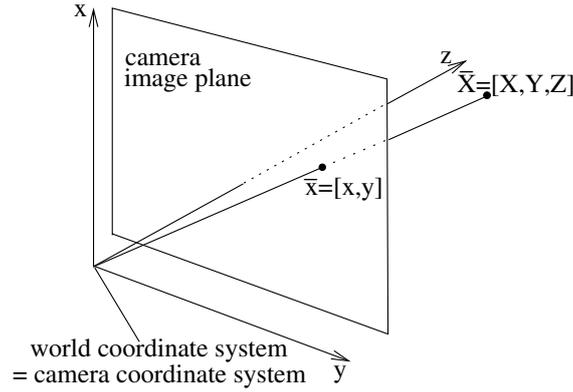


Figure A.2: Projection of the 3D scene point \bar{X} into the image plane of a canonical camera.

Camera in arbitrary pose For a camera in an arbitrary pose the world coordinate system and the camera coordinate system are not identical. Figure A.3 illustrates this situation. The camera is located in the world coordinate system at position $\bar{C} \in \mathbb{R}^3$. The camera coordinate system is the rotated world coordinate system by rotation $R \in \mathbb{R}^{3 \times 3}$. Then the camera coordinate system can be transformed into the world coordinate system with a similarity transformation. The similarity transformation inverts the rotation R and the translation \bar{C} of the camera. According to the ortho-normality of the rotation R the inverse transformation R^{-1} is the transposed rotation matrix R^T . Using the above mentioned similarity transformation with

$$R = \begin{bmatrix} r_{1,1} & r_{1,2} & r_{1,3} \\ r_{2,1} & r_{2,2} & r_{2,3} \\ r_{3,1} & r_{3,2} & r_{3,3} \end{bmatrix} \quad \text{and} \quad \bar{C} = \begin{bmatrix} C_x \\ C_y \\ C_z \end{bmatrix}$$

the projection equations (A.2) are

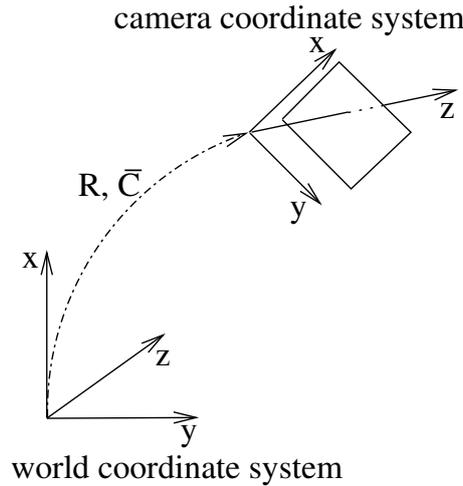
$$\begin{aligned} x &= f \frac{r_{1,1}(X - C_x) + r_{2,1}(Y - C_y) + r_{3,1}(Z - C_z)}{r_{1,3}(X - C_x) + r_{2,3}(Y - C_y) + r_{3,3}(Z - C_z)} \\ y &= f \frac{r_{1,2}(X - C_x) + r_{2,2}(Y - C_y) + r_{3,2}(Z - C_z)}{r_{1,3}(X - C_x) + r_{2,3}(Y - C_y) + r_{3,3}(Z - C_z)} \end{aligned} \quad (\text{A.3})$$

for a camera at position \bar{C} with orientation R . The position \bar{C} and the orientation R of the camera are called external camera parameters because they influence the camera transformation outside the camera coordinate system.

This section introduced the pinhole camera model for the projection of a 3D scene onto the camera image plane. The next section will introduce the projective geometry. The projective geometry will be exploited in section A.2.8 to linearize the camera projection model.

A.2 Projective geometry

This section will introduce the basic geometric primitives used to model the projection of the camera. A camera taking an image of a 3D scene measures the light rays of the scene which

Figure A.3: Camera at position \bar{C} with orientation R

intersect the image plane as explained earlier. A geometric description that models this imaging process is introduced in the next two sections.

A.2.1 Projective space

Each point \bar{x} can be described by its coordinates $\bar{x} = [x, y]^T \in \mathbb{R}^2$ in the image plane. Further the point \bar{x} is also described by the ray r and the plane itself namely as the intersection point \bar{X}_r of the ray r and the image plane. The ray r can be seen as a line in \mathbb{R}^3 . A line in \mathbb{R}^3 is described by a line equation

$$\bar{X}_r \times \bar{D} - \bar{N} = 0 \quad (\text{A.4})$$

where \bar{D} denotes the direction of the line, \bar{X}_r is a point on the line and N is perpendicular to \bar{D} and \bar{X}_l . This representation of a line is called Plücker-coordinates. The ray r consequently has six degrees of freedom, namely the point \bar{X}_r on the line and the direction of the line \bar{D} . The line equation (A.4) is valid for each point \bar{X} on the ray r . Moreover the ray r is a one dimensional subspace of the three dimensional space \mathbb{R}^3 . The line equation (A.4) also holds for a point $\mathbf{x} = [x, y, 1] \in \mathbb{R}^3$ which is the pendant of the 2D point \bar{x} on the camera image plane of a canonical camera as explained in section A.1.1. The point $\mathbf{x} = [x, y, 1]$ can further be interpreted as point in the projective space \mathbb{P}^2 . Points with the last component one are called homogenized points in projective space. In the thesis the projection ray r will be denoted as a line in \mathbb{R}^3 . The description given above shows that the projective two dimensional space \mathbb{P}^2 can be described by the projection rays of a camera. This gives an interpretation of the two dimensional projective space \mathbb{P}^2 as the set of lines passing through the origin. In the following an exact definition of the general projective space will be given.

A more formal definition of the projective space is given as:

Definition 2 (Projective space) *The projective space \mathbb{P} is the set of hyperplanes³ of the*

³Suppose V is a vector space and suppose that L is a non-empty subset of V . If there exists a $v \in V$ such that $L+v = \{v+l \mid l \in L\}$ is a vector subspace of V , then L is a linear manifold of V . The dimension of L is the dimension of $L+v$ and write $\dim L = \dim(L+v)$. If $\dim L = \dim V - 1$, then L is called a hyperplane [Pea83].

vector space V where the vector space V is defined over the commutative body K . The dimension of \mathbb{P} is then given as

$$\dim(\mathbb{P}) = \dim(V) - 1.$$

The elements of $x \in \mathbb{P}$ are called projective points. The point $x = 0_{\dim(V) \times 1}$ is not contained in the projective space \mathbb{P} . Each point x is represented through its spanning vector.

Due to the fact that the point $x = 0_{\dim(V) \times 1}$ is not an element of the projective space \mathbb{P}^n , it is not a vector space. In the following \mathbb{R} will be used as commutative body. According to definition 2 two points $x \in \mathbb{P}$ and $y \in \mathbb{P}$ are equal if and only if

$$x = \rho y \text{ with } \rho \in \mathbb{R} \setminus \{0\}. \quad (\text{A.5})$$

This results from the representation by the spanning vector for a point from the projective space \mathbb{P} . This is often called equality up to scale and is denoted by \cong in this thesis.

The mapping between the Euclidian space \mathbb{R}^n of dimension n and the projective space \mathbb{P}^n of dimension n is given by two mapping functions. The injective function

$$\mathbb{R}^n \rightarrow \mathbb{P}^n : [x_1, \dots, x_n]^T \mapsto [x_1, \dots, x_n, 1]^T \quad (\text{A.6})$$

maps a point from the Euclidian space \mathbb{R}^n into the projective space \mathbb{P}^n . It can be seen that the mapping gives exactly the homogenized point in the projective space because the last component is equal to one. Accordingly it is still in the Euclidian space \mathbb{R}^n . The backward mapping from projective space \mathbb{P}^n into the Euclidian space \mathbb{R}^n is given by

$$\mathbb{P}^n \rightarrow \mathbb{R}^n : [x_1, \dots, x_n, \rho]^T \mapsto \frac{1}{\rho} [x_1, \dots, x_n]^T \quad \forall \rho \in \mathbb{R} \setminus \{0\}. \quad (\text{A.7})$$

This transformation maps the points from the line, described by the projective equality and the point in projective space \mathbb{P}^n , into the intersection point of this line and the Euclidian space \mathbb{R}^n . The backward mapping can't be applied for points with $\rho = 0$ in (A.7) therefore the mapping is not surjective. These points have no representation in the Euclidian space \mathbb{R}^n .

In general the projective space \mathbb{P}^n of dimension n is an extension of the Euclidian space \mathbb{R}^n . The \mathbb{R}^n is extended by the points on the hyperplane at infinity

$$\pi_\infty = \{[x_1, \dots, x_n, 0] \mid x_1, \dots, x_n \in \mathbb{R} \text{ with } \exists x_i \neq 0 \quad i \in \{1, \dots, n\}\}. \quad (\text{A.8})$$

The elements of the hyperplane at infinity π_∞ are called ideal points. These ideal points are not contained in the Euclidian space \mathbb{R}^n . For the two dimensional projective space \mathbb{P}^2 this hyperplane for example is the line at infinity l_∞ .

The earlier explanations showed that the projection ray r can be described by the three degrees of freedom of a line in \mathbb{R}^3 . Furthermore the intersection point of the image plane and the ray r gives an image point. This point is described by its coordinates on the image plane. The image point itself together with the camera center \bar{C} can be used to compute the projection ray r . Hence both entities can be used to describe the same properties of the projective two dimensional space. They are called dual descriptions. In a projective space this can be generalized and leads to an important principle in projective spaces:

Theorem 7 (Duality of points and hyperplanes) *To any theorem in a projective space \mathbb{P} that includes points and hyperplanes exists a dual theorem. The dual theorem is deduced by interchanging the role of points and hyperplanes in the original theorem.*

Please see [Fau93] for details about the duality of points and hyperplanes. The duality provides the advantage that for each theorem in the projective space the dual theorem is proved through theorem 7. Generally both proofs are dual proofs by interchanging the role of points and lines.

After having defined the projective space, the transformations in this space will be introduced in the next section.

A.2.2 Transformations in projective space

There are two types of transformations in a projective space \mathbb{P}^n . The first type is the class of affine transformations. It is characterized by the property that it leaves the position of hyperplane at infinity unchanged. I. e. all points on the hyperplane at infinity stay on the hyperplane at infinity under affine transformations. The second type of transformation is the projective transformation in \mathbb{P}^n . This class of transformation also changes the position of the hyperplane at infinity.

Affine transformations In general an affine transformation is an invertible linear transformation $A_{n \times n}$ from the Euclidian space \mathbb{R}^n into the same space followed by a translation $t_{n \times 1}$ in \mathbb{R}^n . For points of the projective space \mathbb{P}^n this can be written in matrix form as

$$\mathbf{x}_2 = \underbrace{\begin{bmatrix} A_{n,n} & t_{n \times 1} \\ 0_{1 \times n} & 1 \end{bmatrix}}_{T_A} \mathbf{x}_1 \text{ with } \mathbf{x}_1, \mathbf{x}_2 \in \mathbb{P}^n, \quad (\text{A.9})$$

where the point \mathbf{x}_1 is mapped to the point \mathbf{x}_2 by the affine transformation T_A . It is easy to see that this class of transformation T_A transforms the points on the hyperplane at infinity into points on the hyperplane at infinity because the scalar product of a point $\mathbf{x}_1 = [x_1, \dots, x_n, 0]$ on the hyperplane at infinity and the last row of T_A is always zero. Consequently the last component of \mathbf{x}_2 is also zero and \mathbf{x}_2 lies on the hyperplane at infinity.

A subgroup of the affine transformation in the projective space \mathbb{P}^n is the similarity transformation. It contains all transformations that have a isotropically scaled orthogonal linear transformation $A_{n,n}$ in (A.9). The properties of the affine transformations will be discussed later in detail for the projective spaces \mathbb{P}^2 and \mathbb{P}^3 in section A.2.3 and section A.2.5.

Projective transformations The projective transformation T_p in a projective space \mathbb{P}^n is a non-singular matrix of size $(n+1) \times (n+1)$. It is a generalization of the affine transformation in \mathbb{P}^n . This transformation moves points from the hyperplane at infinity into the Euclidian space \mathbb{R}^n , i.e. the last component of ideal points can become different to zero under a projective transformation.

The transformation T_p can be scaled by an arbitrary scale $\rho \neq 0$ without changing the transformation. That means a projective transformation is scale invariant in contrast to an

affine transformation. The affine transformation is not scale invariant because after scaling the component $(A)_{n+1,n+1}$ becomes different from one.

A.2.3 Two dimensional projective space \mathbb{P}^2

Each point \mathbf{x} in the two dimensional projective space \mathbb{P}^2 is given as a vector $\mathbf{x} = [x, y, w]^T \in \mathbb{R}^3 \setminus \{0_{3 \times 1}\}$. In the previous section the two dimensional projective space \mathbb{P}^2 was introduced as the set of lines in \mathbb{R}^3 that go through the origin. Each of these lines can also be described by a point in the \mathbb{R}^2 because a point describes a line together with the origin.

For the two dimensional projective space \mathbb{P}^2 equality up to scale means that a point $\mathbf{x} = [x, y, 1]^T$ on the plane is equal to all points $\mathbf{y} = \rho[x, y, 1]^T$ on the line that is defined by the origin and the point itself namely the projection ray. All points \mathbf{y} on the line as well as \mathbf{x} hold the line equation (A.4). In projective space this relation can be expressed by

$$[a, b, c]\mathbf{x} = 0 \Leftrightarrow [a, b, c]\mathbf{y} = \rho[a, b, c]\mathbf{x} = 0 \text{ with } \mathbf{x} \cong \mathbf{y}, \quad (\text{A.10})$$

where $[a, b, c]$ contains the two degrees of freedom of the projection ray. These degrees of freedom are the two independent ratios $\{a : b : c\}$. They are uniquely determined by the x -component and the y -component of the homogenous point in the image plane. It can be seen that the equality up to scale \cong defines a one dimensional subspace in \mathbb{P}^2 . For example a line can be parameterized as the intersection point $[x_{int}, 0]^T$ of the line with the x -axis and the slope s_l of the line l .

For the two dimensional projective space \mathbb{P}^2 the hyperplanes are equal to lines in \mathbb{R}^3 . According to theorem 7 for each theorem for points and lines in \mathbb{P}^2 a dual theorem exists that interchanges their roles.

A line in the two dimensional space \mathbb{P}^2 is uniquely defined by any two non identical points in \mathbb{P}^2 . Both points \mathbf{x}_1 and \mathbf{x}_2 have to fulfill the line equation

$$0 = ax + by + cw = [a, b, c]\mathbf{x}^T \text{ with } \mathbf{x} \in \{\mathbf{x}_1, \mathbf{x}_2\}. \quad (\text{A.11})$$

Accordingly the line l in the two dimensional projective space \mathbb{P}^2 can be described by the vector $l = [a, b, c]^T$. Like every entity in projective space the equality of lines is defined as equality up to scale. Therefore the lines $l_1 = [a, b, c]^T$ and $l_2 = \rho[a, b, c]^T$ are equal in projective space. Please note that the lines described here are not only the set of projection rays like before. Here the lines include arbitrary lines in any subspace of \mathbb{P}^2 .

Computation of lines from points As explained before a line through the origin has two degrees of freedom namely the two independent ratios. Using equation (A.11) a point provides one constraint on the degrees of freedom of the line. Hence two points uniquely define a line. This is the common principle for defining lines in \mathbb{R}^2 . From two points \mathbf{x}_1 and \mathbf{x}_2 of the projective space \mathbb{P}^2 the line l is computed as the solution of

$$\underbrace{\begin{bmatrix} \mathbf{x}_1^T \\ \mathbf{x}_2^T \end{bmatrix}}_{A_l} l = 0. \quad (\text{A.12})$$

This solution can be computed as the nullspace⁴ of the matrix A_l . This concept for the computation of geometric entities like points and planes as the nullspace of a matrix will be used frequently.

The basic properties of two lines like intersection and parallelism will now be introduced for the projective space.

Theorem 8 (Intersection of lines) *The intersection point \mathbf{x} of two lines l_1 and l_2 is simply given as the crossproduct or vector product of the lines*

$$\mathbf{x} = l_1 \times l_2. \quad (\text{A.13})$$

Proof: The intersection point \mathbf{x} has to lie on both lines l_1 and l_2 . This can be deduced by using the scalar product of \mathbf{x} from (A.13) with the lines

$$(l_1 \times l_2)l_1^T = (l_1 \times l_2)l_2^T = 0,$$

because each line is perpendicular to the cross product of both lines by definition of the cross product. \square

The parallelism of two lines in \mathbb{P}^2 can be decided using the result of theorem 8. In Euclidian space parallel lines do not intersect. In the projective space \mathbb{P}^2 from theorem 8 on the other hand one is always able to get an intersection point even for parallel lines. It can now be used to define parallelism.

Theorem 9 (Parallelism of lines) *Two lines l_1 and l_2 in the projective space \mathbb{P}^2 are parallel if their intersection point x is on the line at infinity l_∞ , i.e.*

$$l_\infty x^T = l_\infty (l_1 \times l_2)^T = 0.$$

Proof: Two parallel lines can be represented by the vectors $l_1 = [a, b, c_1]$ and $l_2 = [a, b, c_2]$ because they have the same slope but different intersections with the y -axis. Therefore the first two components of the vector describing a line can be written using the same first two entries for both lines. With theorem 8 the intersection point p of the lines can be computed as $p = (c_1 - c_2)(b, -a, 0)$. The scale $(c_1 - c_2)$ can be ignored due to the projective equality up to scale. According to (A.8) it can be seen that this point is on the line at infinity l_∞ . \square

Affine transformations in \mathbb{P}^2 Affine transformations are given by a 2×2 invertible transformation $A_{2 \times 2}$ from \mathbb{R}^2 into \mathbb{R}^2 and an afterwards applied translation $t_{2 \times 1} \in \mathbb{R}^2$. According to (A.9) this can be written as

$$\mathbf{x}_2 = \underbrace{\begin{bmatrix} A_{2,2} & t_{2 \times 1} \\ 0_{1 \times 2} & 1 \end{bmatrix}}_{T_A} \mathbf{x}_1 \text{ with } \mathbf{x}_1, \mathbf{x}_2 \in \mathbb{P}^2. \quad (\text{A.14})$$

This affine transformation T_A maps points on the line at infinity l_∞ to points on the same line as shown in section A.2.2. Further properties of the affine transformations in \mathbb{P}^2 are

⁴The is nullspace spanned by the eigenvectors that corresponds to the eigenvalue zero.

1. **Preserved parallelism** Two parallel lines are mapped into two parallel lines. This can be seen looking at the intersection point of the parallel lines in projective space \mathbb{P}^2 . Hence with theorem 9 it follows that the intersection point of these parallel lines is on the line at infinity l_∞ . The affine transformation maps the intersection point of the parallel lines to another point that also lies on the line at infinity l_∞ . Consequently the lines still intersect in a point at the line at infinity; therefore with theorem 9 the lines stay parallel.
2. **Invariant ratio of lengths of parallel line segments** The ratio of the lengths of two parallel line segments is constant under affine transformation of the line segments. The affine transformation contains the invertible transformation $A_{2,2}$ and afterwards a translation is applied. Obviously the translation does not change the ratio of the line segments. The transformation $A_{2,2}$ is able to scale the line segments anisotropically⁵ but this scale only depends on the direction of the line segments. Hence for parallel line segments the scale is the same and the ratio of the lengths is invariant against the anisotropic scale.
3. **Invariant ratio of areas** The ratio of two areas is invariant under affine transformations. It can be seen that each area is scaled by the determinant of the regular matrix $A_{2,2}$ from (A.14) [HZ03]. Therefore the ratio of these areas is constant.

The subgroup of the similarity functions that contains an isotropically scaled orthogonal matrix $A_{2,2}$ in (A.14) scales areas isotropically under the transformation. Furthermore the angle between two lines remains constant.

The projective transformations have one major invariant. The class of projective transformations leaves the cross ratio of four line segments of one line invariant. The cross ratio is the ratio of the ratios between the line segments. E.g. for four line segments l_1, l_2, l_3 and l_4 on line l the cross ratio is

$$\frac{\frac{l_1}{l_2}}{\frac{l_3}{l_4}} = \frac{\frac{l_1^p}{l_2^p}}{\frac{l_3^p}{l_4^p}},$$

where l_1^p, l_2^p, l_3^p and l_4^p are the projectively transformed line segments. Each projective transformation from \mathbb{P}^2 into \mathbb{P}^2 is a three by three matrix. According to the scale independency of projective transformations it has 8 degrees of freedom.

In the projective two dimensional space the points and lines are dual entities. If a point \mathbf{x}_1 is transformed by a projective point transformation H then the transformed point is given as $\mathbf{x}_2 = H\mathbf{x}_1$. The dual transformation for a line is developed as follows. The original point \mathbf{x}_1 lies on the line l_1 according to (A.11) the scalar product $l_1^T \mathbf{x}_1$ of the line and the point is zero. Analogous the scalar product $l_2^T \mathbf{x}_2$ is zero for the transformed point and the transformed line. This leads to

$$l_1^T \mathbf{x}_1 = l_2^T \mathbf{x}_2 \Leftrightarrow l_1^T \mathbf{x}_1 = l_2^T H\mathbf{x}_1.$$

Reordering and transposition of the equation leads to

$$l_2 = H^{-T} l_1. \tag{A.15}$$

⁵Anisotropic scaling is a directionally dependent scaling.

Consequently the line is transformed by the inverse transposed projective point transformation H^{-T} . After introducing the basic entities, point and lines, of the projective space \mathbb{P}^2 the next section will introduce conics. Conics are often used for self-calibration and therefore important here.

A.2.4 Conics

In the two dimensional Euclidian space \mathbb{R}^2 conics appear in three different representations: hyperbola, ellipse and parabola. In the Euclidian \mathbb{R}^2 space conics are non-degenerate⁶ curves generated by the intersections of a plane with one or two nappes of a cone. This section will describe how these conics are represented in the projective space. A conic \mathbf{C} is described by a polynomial of degree two in the two coordinates of \mathbb{R}^2

$$ax^2 + bxy + cy^2 + dx + ey + f = 0 \quad \forall [x, y]^T \in \mathbf{C}. \quad (\text{A.16})$$

Hence the conic has five degrees of freedom, namely the five independent ratios $\{a : b : c : d : e : f\}$. Applying the mapping function (A.7) from projective space \mathbb{P}^2 to Euclidian space \mathbb{R}^2 to a point $\mathbf{x} = [x, y, w]^T \in \mathbb{P}^2$ yields the point $\bar{x} = [\frac{x}{w}, \frac{y}{w}]^T \in \mathbb{R}^2$. Using this point in (A.16) leads to

$$ax^2 + bxy + cy^2 + dxw + eyw + fw^2 = 0 \quad \forall [\frac{x}{w}, \frac{y}{w}]^T \in \mathbf{C}. \quad (\text{A.17})$$

It can be written as a matrix equation

$$[x, y, w]C[x, y, w]^T = [x, y, w] \begin{bmatrix} a & \frac{b}{2} & \frac{d}{2} \\ \frac{b}{2} & c & \frac{e}{2} \\ \frac{d}{2} & \frac{e}{2} & f \end{bmatrix} [x, y, w]^T \quad \forall [x, y, w]^T \in C, \quad (\text{A.18})$$

with the symmetric matrix C representing the conic. Please note that in the following it will not be distinguished between the conic \mathbf{C} and the matrix C representing the conic. This conic is called a **point conic** because its defined by (A.18) incorporating points. A conic C that does not have full rank is called degenerate. Degenerate conics represent two lines (rank 2) or a repeated line (rank 1). The matrix C in general has 6 degrees of freedom in contrast to the conic \mathbf{C} in Euclidian space that only has five degrees of freedom. Since the multiplication of the conic C with a non-zero scalar has no effect on the conic itself because equation (A.18) still describes the same conic. Five degrees of freedom remain for a conic in \mathbb{P}^2 .

The theorem 7 about the duality of points and lines gives us the equations for a **dual (line) conic** C^* . Using lines to describe the conic \mathbf{C} in (A.18) leads to the matrix C^* for the dual (line) conic

$$l^T C^* l = 0, \quad (\text{A.19})$$

for lines that are tangent to the conic C . Please note that C and C^* characterize the same conic but C defines the conic by point mappings and C^* defines the conic with the mapping of lines. In order to prove that the dual conic is a conic in \mathbb{R}^2 an additional theorem is needed.

⁶I. e. a hyperbola that does not collapses into two lines crossing at a point, through a family of hyperbola having those lines as common asymptotes, a parabola that is not a line and an ellipse that is not collapsed into a point.

Theorem 10 (Tangent lines to a conic) *The tangent line l to a conic at a point \mathbf{x} on the conic in matrix form C is given as*

$$l = C\mathbf{x}.$$

Proof: First it will be shown that the point \mathbf{x} is on the line $l = C\mathbf{x}$. Using (A.10) the scalar product of the point and the line has to be zero:

$$l^T \mathbf{x} = (C\mathbf{x})^T \mathbf{x} = \mathbf{x}^T C^T \mathbf{x} = 0.$$

Due to the symmetry of the matrix C representing a non-degenerate conic this is equal to (A.18) and therefore valid for all points \mathbf{x} on the conic. Furthermore it has to be proven that the line intersects the conic only in the point \mathbf{x} . The line l can be written as $l = \mathbf{x} + \lambda d$ where d is the direction of the line. Suppose the line intersects the conic in a second point. Then there exists a $\lambda \neq 0$ that fulfills the conic equation (A.18). The conic equation for all points on the line is

$$(\mathbf{x} + \lambda d)^T C (\mathbf{x} + \lambda d) = \underbrace{\mathbf{x} C^T \mathbf{x}}_{=0} + 2\lambda \mathbf{x}^T C d + \lambda^2 d^T C d = 0.$$

This is only valid if the conic is equal to the line l because $\mathbf{x}^T C d$ and $d^T C d$ are only in this case equal to zero otherwise is not. \square

With this result equation (A.19) can be proved. Each line l tangent to the conic C intersects it according to theorem (10) at the point $\mathbf{x} = C^{-1}l$. Equation (A.18) is then equal to

$$0 = \mathbf{x}^T C \mathbf{x} = (C^{-1}l)^T C (C^{-1}l) = l^T C^{-T} \underbrace{C C^{-1}}_{I_{3 \times 3}} l = l^T C^{-1} l.$$

With the symmetric conic matrix $C^* = C^{-T} = C^{-1}$ this is equal to (A.19) for an invertible conic C . Moreover this proof gives an explicit connection between the conic and the dual conic as the adjoint matrix of the conic matrix C .

Computation of the conic The computation of a conic can be done for points or lines as follows. Equation (A.18) provides for each point one constraint on the conic C . As seen before the conic has five degrees of freedom therefore five points are sufficient to uniquely determine the conic. The computation of the conic can be done writing (A.17) as a scalar product of a vector $\mathbf{c} = [a, b, c, d, e, f]^T$ that contains the coefficients of the conic from (A.17) and the vector $\vec{x} = [x^2, xy, y^2, xw, yw, w^2]$ that contains the data of the projective point $\mathbf{x} = [x, y, w]$,

$$\vec{x} \mathbf{c} = 0. \tag{A.20}$$

From a set of five points $\{\mathbf{x}_1, \dots, \mathbf{x}_5\}$ the conic can then be computed through

$$\underbrace{\begin{bmatrix} \vec{x}_1 \\ \vec{x}_2 \\ \vec{x}_3 \\ \vec{x}_4 \\ \vec{x}_5 \end{bmatrix}}_{A_c} \mathbf{c} = 0_{5 \times 1} \tag{A.21}$$

as the nullspace of the 5×6 matrix A_c . Similarly the dual conic can be computed using equation (A.19) and lines instead of points.

In the following the transformation rule for a conic under a projective point transformation will be deduced. The projective point transformation H maps the point \mathbf{x}_1 to a point $\mathbf{x}_2 = H\mathbf{x}_1$. For each point \mathbf{x}_1 on the conic C_1 with equation (A.18) follows $\mathbf{x}_1^T C_1 \mathbf{x}_1 = 0$. The equation $\mathbf{x}_2^T C_2 \mathbf{x}_2 = 0$ holds then for the mapped point \mathbf{x}_2 on the mapped conic C_2 . The equality of both conic equations leads to

$$\mathbf{x}_2^T C_2 \mathbf{x}_2 = \mathbf{x}_1^T C_1 \mathbf{x}_1 \Leftrightarrow \mathbf{x}_1^T H^T C_2 H \mathbf{x}_1 = \mathbf{x}_1^T C_1 \mathbf{x}_1.$$

Accordingly the transformation of a conic is given by

$$C_2 = H^{-T} C_1 H^{-1}. \quad (\text{A.22})$$

The next section will introduce the three dimensional projective space \mathbb{P}^3 . Many of the properties of the projective three dimensional space are straight forward extensions of the properties of the projective two dimensional space transferred to a space with one more dimension.

A.2.5 Three dimensional projective space \mathbb{P}^3

The three dimensional projective space \mathbb{P}^3 is the extension of the three dimensional Euclidian space \mathbb{R}^3 by the plane at infinity π_∞ . The hyperplanes of the \mathbb{P}^3 are two dimensional planes in contrast to lines in the previous section as hyperplanes of \mathbb{P}^2 . The duality in the \mathbb{P}^3 is therefore defined between points and planes. The extension to the third dimension adds some new properties to this space compared to the \mathbb{P}^2 but most of the properties of the \mathbb{P}^2 can be extended to \mathbb{P}^3 .

The hyperplane in \mathbb{P}^3 is a plane as mentioned before. At first the properties of a plane and the representation of the plane in \mathbb{P}^3 will be introduced. For each point $\bar{X} = [X, Y, Z]$ on a plane π in the Euclidian space \mathbb{R}^3 the following equation holds

$$\pi_1 X + \pi_2 Y + \pi_3 Z + \pi_4 = 0. \quad (\text{A.23})$$

Comparable to the line equation (A.11) this contains the three independent ratios ($\{\pi_1 : \pi_2 : \pi_3 : \pi_4\}$) and is scale invariant. Hence the plane has three degrees of freedom in \mathbb{R}^3 . The plane equation for a point $\mathbf{X} = [X, Y, Z, W]^T \in \mathbb{P}^3$ can be deduced from (A.23) by using the transfer function (A.7) from the projective space \mathbb{P}^3 into the Euclidian space \mathbb{R}^3 . Then the point $\bar{X} \in \mathbb{R}^3$ is given as $\bar{X} = [X/W, Y/W, Z/W]^T$. Using this in (A.23) results in the following equation for a plane in projective space

$$\pi_1 X + \pi_2 Y + \pi_3 Z + \pi_4 W = 0 \Leftrightarrow \underbrace{[\pi_1, \pi_2, \pi_3, \pi_4]}_{\pi} \mathbf{X} = 0, \quad (\text{A.24})$$

which holds for all points $\mathbf{X} \in P^3$ on the plane. Reordering yields the Hesse normal form of a plane

$$\underbrace{[\pi_1, \pi_2, \pi_3]}_{n_\pi} \bar{X} + \pi_4 = 0 \quad \forall \mathbf{X} \in \mathbb{P}^3 \quad (\text{A.25})$$

if \mathbf{X} is on the plane. Accordingly n_π is the normal vector of the plane and $d_\pi = \frac{\pi_4}{\|n_\pi\|_2}$ is the distance of the plane to the origin. The above deduction is not possible for ideal points because the fourth coordinate W of an ideal point is zero. The plane at infinity has the following plane vector $\pi_\infty = [0, 0, 0, 1]^T$. One can easily verify that only ideal points can fulfill the plane equation (A.24). Hence the normal vector is given by $n_\infty = [0, 0, 0]^T$ and the distance d_∞ to the origin is infinity as expected.

Affine transformations in \mathbb{P}^3 An affine transformation in \mathbb{P}^3 is given by a regular transformation $A_{3 \times 3}$ and a succeeding translation $t_{3 \times 1}$ in \mathbb{R}^3 . According to (A.9) this can be written as

$$\mathbf{X}_2 = \underbrace{\begin{bmatrix} A_{3,3} & t_{3 \times 1} \\ 0_{1 \times 3} & 1 \end{bmatrix}}_{T_{A_{3,3}}} \mathbf{X}_1 \text{ with } \mathbf{X}_1, \mathbf{X}_2 \in \mathbb{P}^3. \quad (\text{A.26})$$

Again the plane at infinity π_∞ is invariant under affine transformations. I.e. points on the plane at infinity stay on the plane at infinity but the points can be moved on the plane.

Projective transformation in \mathbb{P}^3 A projective transformation in the three dimensional projective space \mathbb{P}^3 is given by a regular 4×4 matrix. It can be seen in section 2.7 that a projective transformation generally skews the projective space and changes the position of the plane at infinity π_∞ .

Computation of a plane Each plane in \mathbb{P}^3 has three degrees of freedom as seen earlier, namely the three ratios. Each point provides one constraint to the plane using (A.24). Hence to determine the plane π three points are needed. The plane π can be computed using (A.24) and three points \mathbf{X}_1 , \mathbf{X}_2 and \mathbf{X}_3 from the projected space \mathbb{P}^3 on the plane as nullspace of

$$\underbrace{\begin{bmatrix} \mathbf{X}_1^T \\ \mathbf{X}_2^T \\ \mathbf{X}_3^T \end{bmatrix}}_{A_\pi} \pi = 0_{3 \times 1}. \quad (\text{A.27})$$

The nullspace can be computed explicitly from the minors of the matrix A_π [HZ03].

After introducing points and planes as the basic entities of the three dimensional projective space \mathbb{P}^3 the frequently used quadric will be explained, which provides an attractive concept for the camera self-calibration.

A.2.6 Quadric

In general a quadric and respectively a quadric surface is any 3-dimensional surface represented by a second-order equation in the coordinates of a point $\bar{X} = [X, Y, Z]$ in the space \mathbb{R}^3

$$aX^2 + bY^2 + cZ^2 + dXY + eXZ + fYZ + gX + hY + iZ + k = 0, \quad (\text{A.28})$$

which holds for all points on the quadric \mathbf{Q} . The quadric equation for a point \mathbf{X} from the projective space \mathbb{P}^3 can be deduced using the transformation (A.7) between \mathbb{P}^3 and \mathbb{R}^3 for points that are not on the plane at infinity

$$aX^2 + bY^2 + cZ^2 + dXY + eXZ + fYZ + gWX + hWY + iWZ + kW^2 = 0. \quad (\text{A.29})$$

It can be written in matrix form as

$$\mathbf{X}^T Q \mathbf{X} = \mathbf{X}^T \begin{bmatrix} a & \frac{d}{2} & \frac{e}{2} & \frac{g}{2} \\ \frac{d}{2} & b & \frac{f}{2} & \frac{h}{2} \\ \frac{e}{2} & \frac{f}{2} & c & \frac{i}{2} \\ \frac{g}{2} & \frac{h}{2} & \frac{i}{2} & k \end{bmatrix} \mathbf{X} = 0 \quad (\text{A.30})$$

which is fulfilled for all points $\mathbf{X} \in \mathbb{P}^3$ on the quadric surface. The 4×4 symmetric matrix Q describes the surface of the quadric \mathbf{Q} . Therefore the matrix Q and the quadric \mathbf{Q} will not be distinguished here. Further it can be seen that equation (A.30) can be scaled with $\rho \in \mathbb{R} \setminus \{0\}$ without changing the quadric Q . Accordingly a quadric in \mathbb{P}^3 has nine degrees of freedom. The 10 degrees of the symmetric 4×4 matrix minus one for the scale invariance.

Due to the duality theorem 7 the **dual (plane) quadric** Q^* holds

$$\pi^T Q^* \pi = 0 \text{ with } \pi \in \mathbb{P}^3 \quad (\text{A.31})$$

for all planes π that are tangent to the quadric \mathbf{Q} . This can be proved analogous to the proof for the dual conic in section A.2.4. The dual quadric describes also a quadric in \mathbb{P}^3 , and Q^* is the adjoint matrix of Q [HZ03]. This is simply the inverse for a regular quadric Q ; that means Q has full rank. If Q has not full rank the quadric is called degenerate. The calculation of the projection of the quadric into a camera is much easier for a dual quadric than for a quadric. The intersection of a quadric with a plane leads to a conic in that plane. A quadric Q_1 transforms under the projective point transformation H with $x_2 = Hx_1$ as

$$Q_2 = H^{-T} Q_1 H^{-1} \quad (\text{A.32})$$

where Q_2 is the transformed quadric. This can be deduced similar to the transformation of the conic in section A.2.4. The transformation rule for the dual quadric is

$$Q_2^* = H Q_1^* H^T. \quad (\text{A.33})$$

Computation of a quadric from points The quadric has nine degrees of freedom due to the overall scale ambiguity as explained before. Each point on the quadric surface provides one constraint (A.30) to the quadric Q . Thus nine points are sufficient to determine the quadric uniquely up to scale. The quadric Q can be computed using nine points $\mathbf{X}_1 = [X_1, Y_1, Z_1, W_1]^T, \dots, \mathbf{X}_9 = [X_9, Y_9, Z_9, W_9] \in \mathbb{P}^3$ as the solution of

$$\begin{bmatrix} X_1^2 & Y_1^2 & Z_1^2 & X_1 Y_1 & X_1 Z_1 & Y_1 Z_1 & W_1 X_1 & W_1 Y_1 & W_1 Z_1 & W_1^2 \\ \vdots & \vdots \\ X_9^2 & Y_9^2 & Z_9^2 & X_9 Y_9 & X_9 Z_9 & Y_9 Z_9 & W_9 X_9 & W_9 Y_9 & W_9 Z_9 & W_9^2 \end{bmatrix} Q_{vec} = 0, \quad (\text{A.34})$$

where $Q_{vec} = [a, b, c, d, e, f, g, h, i, k]^T$ is a vector that defines the quadric matrix Q by (A.30).

After introducing the geometric spaces for the imaging process of a camera and the properties of these spaces, the camera projection will be modeled using these spaces. This will linearize the nonlinear projection (A.3) of the 3D scene point into the image point. The next section will introduce the internal camera parameters that model the transformation of coordinates of an image point on the camera image plane into the pixel coordinate system of a digital image.

A.2.7 From world to pixel coordinates

In section A.1.1 the image point on the camera image plane was given as the intersection point of the projection ray and the camera image plane. This explanation neglected some parts of the imaging process to get a digital image for computation. The imaging process for a digital camera uses a digital sensor that is a grid of light sensitive cells. Each cell integrates the photons of the light-rays that hit the cell. Usually each cell represents a pixel of the resulting image. To get the coordinates of an image point the cell on the camera image plane has to be determined to get the pixel coordinates of the image point. Obviously the topology of the cells on the camera image plane is important for the mapping between the coordinates on the camera image plane and the pixel coordinates of the image point. This transformation is an affine transformation in the image plane.

At first the different parts of this transformation will be explained in detail.

- **Pixel scale** The pixel scale depends on the size of a cell of the imaging sensor. It scales the coordinates of the intersection point of the light-ray and the image plane into the pixel coordinate system. The scale factor is the number of pixels per unit of the camera image plane coordinate system. This pixel scale is usually also denoted as focal length f . For the remainder of the thesis it will be not distinguished between focal length and pixel scale.
- **Aspect ratio** Only for quadratic pixels the focal length is the same for both coordinate directions because the focal length depends on the number of pixels per unit in the camera image plane coordinate system. The aspect ratio a is the ratio of the pixel size in x -direction and y -direction. The focal length for the y -coordinate is the product of the focal length f and the aspect ratio a .
- **Principal point** The principal point $[u, v]$ is the position of the intersection point of the camera image plane and the optical axis. It is given in pixel coordinates. The principal point is the shift between the origin of the pixel coordinate system and the image plane coordinate system.
- **Skew** Sometimes the rows and columns of the imaging sensor are not perpendicular to each other. Hence the pixel coordinate system is skewed with respect to the image plane coordinate system. The skew s models this effect and can be computed as the cosine of the angle between rows and columns scaled with the focal length f .

Accordingly the transformation between a homogenous point on the camera image plane in camera image plane coordinates and pixel coordinates is given as the affine transformation K in \mathbb{P}^2

$$K = \begin{bmatrix} f & s & u \\ 0 & a \cdot f & v \\ 0 & 0 & 1 \end{bmatrix}. \quad (\text{A.35})$$

The matrix K is called the calibration matrix of the camera. In the thesis it will not be distinguished between the camera calibration and the camera calibration matrix K . The parameters focal length f , skew s , principal point $[u, v]$ and aspect ratio a are also known as the intrinsic camera parameters because they describe the transformation inside the camera. For digital cameras usually the skew is zero due to the perpendicular rows and columns of cells on the CCD-sensor. For the digitalization of an analog image this is not always fulfilled.

The goal of self-calibration is to recover the calibration matrix K from the images themselves without using special calibration objects. If the camera calibration is known the coordinates of a point on the image plane in the camera coordinate system can be computed. This is done by applying the inverse calibration matrix K^{-1} to the point \mathbf{x}_p in pixel coordinates to get the point $\mathbf{x} = K^{-1}\mathbf{x}_p$ in the camera coordinate system. The point \mathbf{x} is transformed into **normalized coordinates** i.e. it does not depend on the properties of the imaging sensor.

The next section will introduce the camera transformation that models the camera projection of a homogenous 3D point \mathbf{X} onto the camera image plane and the transformation into the image point in the pixel coordinate system.

A.2.8 Projection matrix

For a 3D scene point it was deduced in section A.1.1 that the image point on the camera image plane depends on the cameras position $\bar{C} \in \mathbb{R}^3$ and the cameras orientation R . The affine transformation used for the 3D scene point $\bar{X} \in \mathbb{R}^3$ can be written as

$$\bar{X}_c = R^T(\bar{X} - \bar{C}), \quad (\text{A.36})$$

where $\bar{X}_c \in \mathbb{R}^3$ is the 3D scene point given in the camera coordinate system. Using the corresponding point \mathbf{X} in projective space \mathbb{P}^3 of the scene point equation (A.36) can be written in matrix form as

$$\mathbf{X}_c = \begin{bmatrix} R^T & -R^T\bar{C} \\ 0_{3 \times 1} & 1 \end{bmatrix} \mathbf{X}, \quad (\text{A.37})$$

where $\mathbf{X}_c \in \mathbb{P}^3$ is the homogenous point in the camera coordinate system corresponding to the scene point \mathbf{X} . Then the perspective projection T_{per} can be used to project \mathbf{X}_c onto the camera image plane. Here a projection into a canonical camera can be used because \mathbf{X} is already mapped into the camera coordinate system. The general central projection onto the

image plane at $[0, 0, f]$ is given by

$$T_p = \underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{T_{par}} \underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & \frac{1}{f} & 1 \end{bmatrix}}_{T_{per}}, \quad (\text{A.38})$$

where T_{per} is the perspective transformation and T_{par} is the parallel projection onto the x - y -plane. The perspective transformation skews the projective space so that all projection rays are parallel. It moves the optical center of the camera from the origin of the coordinate system onto the plane at infinity. This can be deduced from the fact that the optical center is the intersection of all projection rays. According to theorem 9 if all rays are parallel, the intersection point has to be on the plane at infinity. The parallel projection T_{par} simply projects from \mathbb{P}^3 into the image plane of the camera. Then T_p applied to the point \mathbf{X}_c in the camera coordinate system yields the point on the camera image plane

$$\mathbf{X}_i = T_p \begin{bmatrix} R^T & -R^T \bar{C} \\ 0_{3 \times 1} & 1 \end{bmatrix} \mathbf{X}. \quad (\text{A.39})$$

where $\mathbf{X}_i \in \mathbb{P}^3$ is the point on the camera image plane. For an image plane at the canonical position $z = 1$ and neglecting of the z -component of the projected point \mathbf{x} equation (A.39) is equal to

$$\mathbf{x}_i = [R^T | -R^T \bar{C}] \mathbf{X} \quad (\text{A.40})$$

where $\mathbf{x}_i \in \mathbb{P}^2$ is a projective point on the camera image plane. Please note the projection in (A.40) is described by the matrix product from (A.39) which is the 3×4 matrix $[R^T | -R^T \bar{C}]$. The transformation of this image point \mathbf{x}_i from image plane coordinates into pixel coordinates is given by the camera calibration from (2.3). This leads to the camera transformation that projects a 3D scene point \mathbf{X} into an image point \mathbf{x} in pixel coordinates on the camera image plane

$$\mathbf{x} = K[R^T | -R^T \bar{C}] \mathbf{X}. \quad (\text{A.41})$$

The matrix $P = K[R^T | -R^T \bar{C}]$ is called the camera projection matrix. It transfers a 3D scene point into the image point for a camera with intrinsic calibration K at position \bar{C} and with orientation R . Due to the equality up to scale in projective spaces the camera projection matrix P is defined only up to scale. The next paragraph will show how the camera projection matrix P can be estimated.

Estimation of the projection matrix The projection equation for the camera from (A.41) provides two constraints on the camera projection matrix P for a 3D scene point \mathbf{X} and the corresponding image point \mathbf{x} . In (A.41) are three equations contained but only two linear independent due to the two degrees of freedom for a point \mathbf{x} in \mathbb{P}^2 . A counting argument leads to the fact that six correspondences $(\mathbf{X}^i, \mathbf{x}^i)$ of a 3D scene point and an image point are needed

to compute the projection matrix linearly. For a given correspondence $(\mathbf{X}^i, \mathbf{x}^i)$ of a 3D scene point \mathbf{X}^i and an image point \mathbf{x}^i equation (A.41) is only defined for equality up to scale

$$\mathbf{x}^i \cong K[R^T | -R^T C] \mathbf{X}^i. \quad (\text{A.42})$$

Euclidian equality can be achieved by eliminating the projective scale. This can be done by multiplying the right hand side in (A.42) with the last component of the image point \mathbf{x}_i and the left hand side with the third component of the mapped 3D scene point $P\mathbf{X}$. Reordering leads to

$$\begin{aligned} & -x_3^i(P_{2,1}X_1^i + P_{2,2}X_2^i + P_{2,3}X_3^i + P_{2,4}X_4^i) + \\ & \quad x_2^i(P_{3,1}X_1^i + P_{3,2}X_2^i + P_{3,3}X_3^i + P_{3,4}X_4^i) = 0 \\ & \quad x_3^i(P_{1,1}X_1^i + P_{1,2}X_2^i + P_{1,3}X_3^i + P_{1,4}X_4^i) - \\ & \quad x_1^i(P_{3,1}X_1^i + P_{3,2}X_2^i + P_{3,3}X_3^i + P_{3,4}X_4^i) = 0 \\ & -x_2^i(P_{1,1}X_1^i + P_{1,2}X_2^i + P_{1,3}X_3^i + P_{1,4}X_4^i) + \\ & \quad x_2^i(P_{2,1}X_1^i + P_{2,2}X_2^i + P_{2,3}X_3^i + P_{2,4}X_4^i) = 0 \end{aligned}$$

this can be written in matrix form as

$$\underbrace{\begin{bmatrix} 0_{1 \times 4} & -x_3^i(\mathbf{X}^i)^T & x_2^i(\mathbf{X}^i)^T \\ x_3^i(\mathbf{X}^i)^T & 0_{1 \times 4} & -x_1^i(\mathbf{X}^i)^T \\ -x_2^i(\mathbf{X}^i)^T & x_1^i(\mathbf{X}^i)^T & 0_{1 \times 4} \end{bmatrix}}_{A_P} \begin{bmatrix} P_{1,1} \\ \vdots \\ P_{3,4} \end{bmatrix} = 0_{3 \times 1}. \quad (\text{A.43})$$

for each correspondence $(\mathbf{X}^i, \mathbf{x}^i)$. The conditions for six correspondences then give a sufficient number of constraints for the projection matrix which is the nullspace of the matrix A_P in (A.43).

For 3D scene points in general positions (A.43) computes the camera projection matrix. For points on a plane the nullspace dimension is greater than one. Hence the solution for the camera projection matrix is ambiguous. In this case another parameterization of the camera projection matrix together with a nonlinear optimization has to be chosen in order to compute a unique camera projection matrix P . More details about nonlinear estimation methods for the camera projection matrix can be found in [HZ03].

A.2.9 Radial distortion

In the previous sections the imaging process was assumed to be linear. Hence the camera projection could be modeled as a linear process in (A.43). Unfortunately for real lens systems this assumption is violated especially for cheaper lens systems. This nonlinear effect is called radial distortion. In figure A.4 the effect of radial distortion is illustrated. Radial distortion is visible at straight lines that become curves.

The radial distortion is caused by the bending of the light rays in the lens system of the camera. The correction of the radial distortion can be performed on the image plane. An

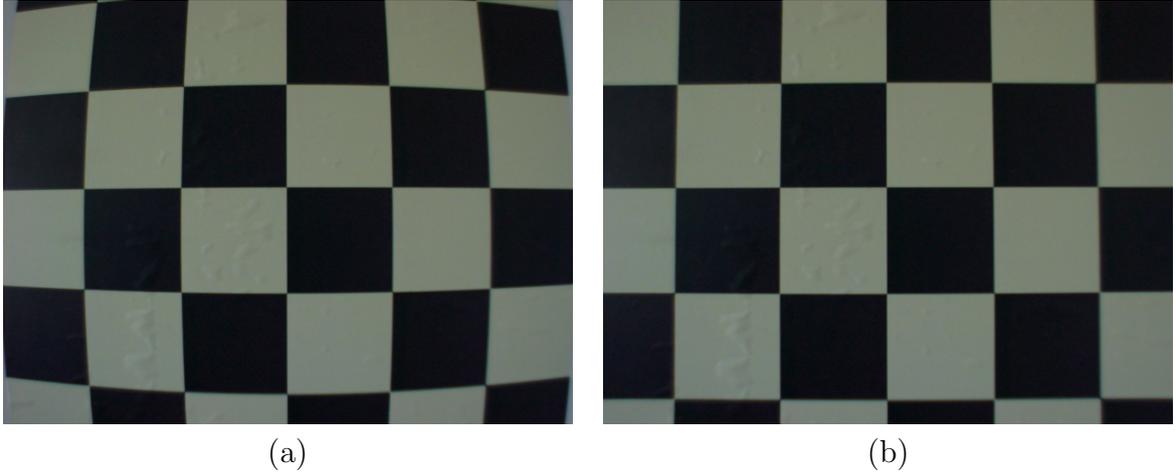


Figure A.4: (a) Camera image of a chess board with radial distortion. Please note the straight lines of the chess board become curves. (b) Image from (a) with corrected radial distortion of the lens.

example is shown in figure A.4. It can be seen that the radial distortion mainly depends on the distance to the principal point. Albeit the real center of the radial distortion $[u_r, v_r]^T$ is only close to the principal point of the camera [Ste93], the principal point is often used as center of the radial distortion $[u_r, v_r]^T = [u, v]^T$. The radial distortion $L(\cdot)$ can be modeled for a distorted point $\bar{x}^d = [x_1^d, x_2^d]^T$ as

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = L(r_d(\bar{x})) \begin{bmatrix} x_1^d \\ x_2^d \end{bmatrix}, \quad (\text{A.44})$$

where $\bar{x} = [x_1, x_2]^T$ is the not distorted image point of the linear model and $r_d(\bar{x})$ is the Euclidian distance of the observed distorted point \bar{x} to the center of the radial distortion $[u_r, v_r]^T$. Since it is not easy to exactly model all effects of the lens system an approximation for the true function is used. The Taylor expansion for the radial distortion $L(\cdot)$ interpreted as a general function

$$L(r_d(\bar{x})) = 1 + \kappa_1 r_d + \kappa_2 r_d^2 + \kappa_3 r_d^3 + \dots$$

is used as approximation to the true radial distortion function. The coefficients of the radial distortion function $L(\cdot)$ and the center $[u_r, v_r]^T$ of the radial distortion are part of the internal camera parameters. Then the radial distortion model (A.44) can be written as

$$x = (1 + \kappa_1 r_d + \kappa_2 r_d^2 + \kappa_3 r_d^3 + \dots) \begin{bmatrix} x_1^d \\ x_2^d \end{bmatrix}. \quad (\text{A.45})$$

In order to compensate the radial distortion this function has to be inverted. In figure A.4 (b) the result of the radial distortion compensation is illustrated. Hence the imaging process becomes linear by use of the radial distortion compensation. Often the Taylor approximation is terminated after the linear term κ_1 . The effect of the radial distortion in dependence on the sign of κ_1 is illustrated in figure A.5.

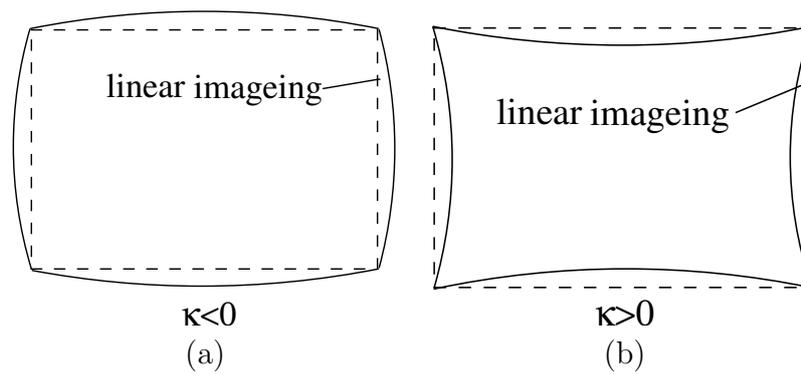


Figure A.5: (a) Effect of radial distortion for κ_1 with negative sign. (b) Effect of radial distortion for κ_1 with positive sign.

This appendix introduced the camera model used to describe the imaging process of a real camera. Furthermore the projective space was introduced. It leads to a linear model for the camera transformation in contrast to a nonlinear camera transformation in Euclidian space.

Appendix B

Self-calibration of a purely rotating camera with known orientation

In this chapter the calibration equations for a purely rotating camera with known orientation will be introduced in detail. First of all the self-calibration equations for a camera with varying intrinsic parameters are shown in section B.1. Due to the fact that digital cameras have a zero skew the self-calibration equations for cameras with zero skew are introduced in section B.2. The additional constraints provided by a camera with constant intrinsic parameters are investigated in section B.3.

B.1 Self-calibration with varying intrinsics and known orientation

In section 4.1 for a purely rotating camera with varying intrinsic parameters the self-calibration problem is given in equation (4.3). The comparison of the entries of the matrices in (4.3)

provides the following nine self-calibration equations

$$\begin{aligned}
0 &= f_j (H_{j,i}^\infty)_{1,1} - \frac{f_i}{\rho_{j,i}} (R_{j,i})_{1,1} - \frac{u_i}{\rho_{j,i}} (R_{j,i})_{3,1} - \frac{s_i}{\rho_{j,i}} (R_{j,i})_{2,1} \\
0 &= a_j f_j (H_{j,i}^\infty)_{1,2} - \frac{f_i}{\rho_{j,i}} (R_{j,i})_{1,2} - \frac{u_i}{\rho_{j,i}} (R_{j,i})_{3,2} - (R_{j,i})_{2,2} \frac{s_i}{\rho_{j,i}} + s_j (H_{j,i}^\infty)_{1,1} \\
- (H_{j,i}^\infty)_{1,3} &= u_j (H_{j,i}^\infty)_{1,1} + v_j (H_{j,i}^\infty)_{1,2} - \frac{f_i}{\rho_{j,i}} (R_{j,i})_{1,3} - \frac{u_i}{\rho_{j,i}} (R_{j,i})_{3,3} - (R_{j,i})_{2,3} \frac{s_i}{\rho_{j,i}} \\
0 &= f_j (H_{j,i}^\infty)_{2,1} - \frac{a_i f_i}{\rho_{j,i}} (R_{j,i})_{2,1} - \frac{v_i}{\rho_{j,i}} (R_{j,i})_{3,1} \\
0 &= a_j f_j (H_{j,i}^\infty)_{2,2} - \frac{a_i f_i}{\rho_{j,i}} (R_{j,i})_{2,2} - \frac{v_i}{\rho_{j,i}} (R_{j,i})_{3,2} + (H_{j,i}^\infty)_{2,1} s_j \\
- (H_{j,i}^\infty)_{2,3} &= u_j (H_{j,i}^\infty)_{2,1} + v_j (H_{j,i}^\infty)_{2,2} - a_i \frac{f_i}{\rho_{j,i}} (R_{j,i})_{2,3} - \frac{v_i}{\rho_{j,i}} (R_{j,i})_{3,3} \\
0 &= f_j (H_{j,i}^\infty)_{3,1} - (R_{j,i})_{3,1} \frac{1}{\rho_{j,i}} \\
0 &= a_j f_j (H_{j,i}^\infty)_{3,2} + (H_{j,i}^\infty)_{3,1} s_j - (R_{j,i})_{3,2} \frac{1}{\rho_{j,i}} \\
- (H_{j,i}^\infty)_{3,3} &= u_j (H_{j,i}^\infty)_{3,1} + v_j (H_{j,i}^\infty)_{3,2} - (R_{j,i})_{3,3} \frac{1}{\rho_{j,i}} \tag{B.1}
\end{aligned}$$

in the entries of the calibration matrix K_j of camera j and the scaled calibration \tilde{K}_i of camera i . The homography $H_{j,i}^\infty$ in (B.1) can be computed from the images themselves as seen in section 2.5.1. The rotation matrix $R_{j,i}$ can be determined using the known orientation of the cameras. Hence a linear equation system for the internal calibration of camera j and the scaled intrinsics of camera i can be deduced from (B.1). To have a more compact notation the entry (m, n) of the homography $H_{j,i}^\infty$ is denoted with $h_{m,n}$ and similar the entry (m, n) of $R_{j,i}$ is denoted as $r_{m,n}$. The equation system deduced from (B.1) is then given by

$$A_H \underbrace{\left[\begin{array}{ccccccccccc} f_j & s_j & u_j & a_j f_j & v_j & \frac{f_i}{\rho_{j,i}} & \frac{s_i}{\rho_{j,i}} & \frac{u_i}{\rho_{j,i}} & \frac{a_i f_i}{\rho_{j,i}} & \frac{v_i}{\rho_{j,i}} & \frac{1}{\rho_{j,i}} \end{array} \right]^T}_{\mathcal{K}} = b_H \tag{B.2}$$

with

$$A_H = \left[\begin{array}{ccccccccccc} h_{1,1} & 0 & 0 & 0 & 0 & -r_{1,1} & -r_{2,1} & -r_{3,1} & 0 & 0 & 0 \\ 0 & h_{1,1} & 0 & h_{1,2} & 0 & -r_{1,2} & -r_{2,2} & -r_{3,2} & 0 & 0 & 0 \\ 0 & 0 & h_{1,1} & 0 & h_{1,2} & -r_{1,3} & -r_{2,3} & -r_{3,3} & 0 & 0 & 0 \\ h_{2,1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -r_{2,1} & -r_{3,1} & 0 \\ 0 & h_{2,1} & 0 & h_{2,2} & 0 & 0 & 0 & 0 & -r_{2,2} & -r_{3,2} & 0 \\ 0 & 0 & h_{2,1} & 0 & h_{2,2} & 0 & 0 & 0 & -r_{2,3} & -r_{3,3} & 0 \\ h_{3,1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -r_{3,1} \\ 0 & h_{3,1} & 0 & h_{3,2} & 0 & 0 & 0 & 0 & 0 & 0 & -r_{3,2} \\ 0 & 0 & h_{3,1} & 0 & h_{3,2} & 0 & 0 & 0 & 0 & 0 & -r_{3,3} \end{array} \right],$$

with the right hand side b_H given by

$$b_H = [0 \ 0 \ -h_{1,3} \ 0 \ 0 \ -h_{2,3} \ 0 \ 0 \ -h_{3,3}]^T.$$

The rank of the system matrix A is at most 9, because the number of rows is 9. In the following it will be proved that the rank for a general rotation is exactly nine. To prove this it is sufficient to find a submatrix of A of size 9×9 that has a non-zero determinant. Here the determinant of the submatrix $A_{1-9,3-11}$ is chosen which is given by

$$\begin{aligned} \det (A_{1-9,3-11}) &= ((r_{1,3}r_{2,2} - r_{1,2}r_{2,3})r_{3,1} + (r_{1,1}r_{2,3} - r_{1,3}r_{2,1})r_{3,2} + (r_{1,2}r_{2,1} - r_{1,1}r_{2,2})r_{3,3}) \\ &\quad h_{3,2}(h_{2,2}h_{3,1} - h_{2,1}h_{3,2})r_{3,1}(r_{2,2}r_{3,1} - r_{2,1}r_{3,2}). \end{aligned} \quad (\text{B.3})$$

For the proposed self-calibration approach a rotation about the x -axis and the y -axis was assumed (general rotation). This leads to a rotation matrix $R_{i,j}$ with the structure

$$R_{j,i} = \begin{bmatrix} \cos(\alpha_y) & 0 & \sin(\alpha_y) \\ \sin(\alpha_x) \sin(\alpha_y) & \cos(\alpha_x) & -\cos(\alpha_y) \sin(\alpha_x) \\ \cos(\alpha_x) \sin(\alpha_y) & \sin(\alpha_x) & \cos(\alpha_x) \cos(\alpha_y) \end{bmatrix}, \quad (\text{B.4})$$

where α_x is the rotation angle of the rotation about the x -axis and α_y is the rotation angle of the rotation about the y -axis. Using this structure of the rotation matrix (B.3) is reduced to

$$\det(A_{1-9,3-11}) = h_{3,2}(h_{2,1}h_{3,2} - h_{2,2}h_{3,1}) \cos(\alpha_x) \sin^2(\alpha_y). \quad (\text{B.5})$$

The terms $\cos(\alpha_x)$ and $\sin^2(\alpha_y)$ are not zero for the assumed general rotation about the x -axis and the y -axis. Exploiting the structure of the infinite homography (2.16) and applying the rotation $R_{j,i}$ from (B.4) leads to a determinant of

$$\det(A_{1-9,3-11}) = \frac{a_2 f_2}{a_1^2 f_1^4} \cos(\alpha_x) \sin^3(\alpha_y) (f_1 \sin(\alpha_x) + s_1 \cos(\alpha_x) \sin(\alpha_y)). \quad (\text{B.6})$$

Hence the determinant (B.6) is different from zero for a general rotation $R_{j,i}$. This leads to the fact that the matrix A in (B.2) has rank 9 because it exists a submatrix of A that has rank 9. Accordingly A has nine linear independent rows. \square

The angle of rotation about the y -axis influences the determinant in (B.6) with the third power of the sinus. As a consequence the determinant is close to zero for very small rotations in y direction which introduces numerical instabilities for the estimation process. I.e. the rotation about the y -axis should be ensured to be not too small for a stable estimation, whereas the rotation about the x -axis is not critical for the numerical stability of the estimation process.

B.2 Self-calibration equations for cameras with zero skew

To reach a linear calibration from a single pair of cameras in section 4.1.1 it was proposed to use the zero skew constraint for CCD cameras. In that case the linear equations provided by

(4.3) for the calibration become

$$\begin{aligned}
0 &= f_j (H_{j,i}^\infty)_{1,1} - \frac{f_i}{\rho_{j,i}} (R_{j,i})_{1,1} - \frac{u_i}{\rho_{j,i}} (R_{j,i})_{3,1} \\
0 &= a_j f_j (H_{j,i}^\infty)_{1,2} - \frac{f_i}{\rho_{j,i}} (R_{j,i})_{1,2} - \frac{u_i}{\rho_{j,i}} (R_{j,i})_{3,2} \\
-(H_{j,i}^\infty)_{1,3} &= u_j (H_{j,i}^\infty)_{1,1} + v_j (H_{j,i}^\infty)_{1,2} - \frac{f_i}{\rho_{j,i}} (R_{j,i})_{1,3} - \frac{u_i}{\rho_{j,i}} (R_{j,i})_{3,3} \\
0 &= f_j (H_{j,i}^\infty)_{2,1} - \frac{a_i f_i}{\rho_{j,i}} (R_{j,i})_{2,1} - \frac{v_i}{\rho_{j,i}} (R_{j,i})_{3,1} \\
0 &= a_j f_j (H_{j,i}^\infty)_{2,2} - \frac{a_i f_i}{\rho_{j,i}} (R_{j,i})_{2,2} - \frac{v_i}{\rho_{j,i}} (R_{j,i})_{3,2} \\
-(H_{j,i}^\infty)_{2,3} &= u_j (H_{j,i}^\infty)_{2,1} + v_j (H_{j,i}^\infty)_{2,2} - \frac{a_i f_i}{\rho_{j,i}} (R_{j,i})_{2,3} - \frac{v_i}{\rho_{j,i}} (R_{j,i})_{3,3} \\
0 &= f_j (H_{j,i}^\infty)_{3,1} - (R_{j,i})_{3,1} \frac{1}{\rho_{j,i}} \\
0 &= a_j f_j (H_{j,i}^\infty)_{3,2} - (R_{j,i})_{3,2} \frac{1}{\rho_{j,i}} \\
-(H_{j,i}^\infty)_{3,3} &= u_j (H_{j,i}^\infty)_{3,1} + v_j (H_{j,i}^\infty)_{3,2} - (R_{j,i})_{3,3} \frac{1}{\rho_{j,i}}.
\end{aligned} \tag{B.7}$$

The linear equation system for cameras that have zero skew is given by

$$A_S \underbrace{\begin{bmatrix} f_j & u_j & a_j f_j & v_j & \frac{f_i}{\rho_{j,i}} & \frac{u_i}{\rho_{j,i}} & \frac{a_i f_i}{\rho_{j,i}} & \frac{v_i}{\rho_{j,i}} & \frac{1}{\rho_{j,i}} \end{bmatrix}^T}_{\mathcal{K}} = b_S \tag{B.8}$$

with the system matrix A_S

$$A_S = \begin{bmatrix} h_{1,1} & 0 & 0 & 0 & -r_{1,1} & -r_{3,1} & 0 & 0 & 0 \\ 0 & 0 & h_{1,2} & 0 & -r_{1,2} & -r_{3,2} & 0 & 0 & 0 \\ 0 & h_{1,1} & 0 & h_{1,2} & -r_{1,3} & -r_{3,3} & 0 & 0 & 0 \\ h_{2,1} & 0 & 0 & 0 & 0 & 0 & -r_{2,1} & -r_{3,1} & 0 \\ 0 & 0 & h_{2,2} & 0 & 0 & 0 & -r_{2,2} & -r_{3,2} & 0 \\ 0 & h_{2,1} & 0 & h_{2,2} & 0 & 0 & -r_{2,3} & -r_{3,3} & 0 \\ h_{3,1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -r_{3,1} \\ 0 & 0 & h_{3,2} & 0 & 0 & 0 & 0 & 0 & -r_{3,2} \\ 0 & h_{3,1} & 0 & h_{3,2} & 0 & 0 & 0 & 0 & -r_{3,3} \end{bmatrix},$$

with the right hand side b_S given by

$$b_H = [0 \ 0 \ -h_{1,3} \ 0 \ 0 \ -h_{2,3} \ 0 \ 0 \ -h_{3,3}]^T.$$

It is clear that the system matrix A_S has at most rank nine because it is a 9×9 matrix. In order to prove that it has exactly rank nine it is sufficient to show that the determinant is not zero. As in the previous section for the calibration a general rotation like in (B.4) is assumed.

Exploiting the structure of the infinite homography (2.16) leads to the determinant $\det(A_S)$ of the system matrix A_S

$$\det(A_S) = \frac{a_2 f_2}{a_1^2 f_1^4} \cos(\alpha_x) \sin^2(\alpha_x) \sin^2(\alpha_y) (u_2 \cos(\alpha_x) \sin(\alpha_y) - f_2 \cos(\alpha_y)), \quad (\text{B.9})$$

which is not equal to zero for a general rotation between the cameras. \square

Again it can be seen that for small rotations about the y -axis or the x -axis the determinant is close to zero which implies that the estimation process becomes unstable due to numerical problems.

B.3 Self-calibration equations for a purely rotating camera with constant intrinsics

This section introduces the self-calibration equations for two cameras with constant intrinsic parameters. Then the self-calibration equation (4.3) simplifies to equation (4.8) as explained in section B.3. The self-calibration equations provided by (4.8) are:

$$\begin{aligned} 0 &= f \left((R_{j,i})_{1,1} - (H_{j,i})_{1,1} \right) + s (R_{j,i})_{2,1} + u (R_{j,i})_{3,1} \\ 0 &= f (R_{j,i})_{1,2} + s \left((R_{j,i})_{2,2} - (H_{j,i})_{1,1} \right) + u (R_{j,i})_{3,2} - af (H_{j,i})_{1,2} \\ (H_{j,i})_{1,3} &= f (R_{j,i})_{1,3} + s (R_{j,i})_{2,3} + u \left((R_{j,i})_{3,3} - (H_{j,i})_{1,1} \right) - v (H_{j,i})_{1,2} \\ 0 &= -f (H_{j,i})_{2,1} + af (R_{j,i})_{2,1} + v (R_{j,i})_{3,1} \\ 0 &= -s (H_{j,i})_{2,1} + af \left((R_{j,i})_{2,2} - (H_{j,i})_{2,2} \right) v (R_{j,i})_{3,2} \\ (H_{j,i})_{2,3} &= -u (H_{j,i})_{2,1} + af (R_{j,i})_{2,3} + v \left((R_{j,i})_{3,3} - (H_{j,i})_{2,2} \right) \\ -(R_{j,i})_{3,1} &= -f (H_{j,i})_{3,1} \\ -(R_{j,i})_{3,2} &= -s (H_{j,i})_{3,1} - af (H_{j,i})_{3,2} \\ (H_{j,i})_{3,3} - (R_{j,i})_{3,3} &= -u (H_{j,i})_{3,1} - v (H_{j,i})_{3,2}. \end{aligned} \quad (\text{B.10})$$

These equations can be written as a linear equation system for the unknown calibration parameters $\mathcal{K}_c = [f, s, u, af, v]^T$

$$\underbrace{\begin{bmatrix} r_{1,1} - h_{1,1} & r_{2,1} & r_{3,1} & 0 & 0 \\ r_{1,2} & r_{2,2} - h_{1,1} & r_{3,2} & -h_{1,2} & 0 \\ r_{1,3} & r_{2,3} & r_{3,3} - h_{1,1} & 0 & -h_{1,2} \\ -h_{2,1} & 0 & 0 & r_{2,1} & r_{3,1} \\ 0 & -h_{2,1} & 0 & r_{2,2} - h_{2,2} & r_{3,2} \\ 0 & 0 & -h_{2,1} & r_{2,3} & r_{3,3} - h_{2,2} \\ -h_{3,1} & 0 & 0 & 0 & 0 \\ 0 & -h_{3,1} & 0 & -h_{3,2} & 0 \\ 0 & 0 & -h_{3,1} & 0 & -h_{3,2} \end{bmatrix}}_{A_c} \mathcal{K}_c = \begin{bmatrix} 0 \\ 0 \\ h_{1,3} \\ 0 \\ 0 \\ h_{2,3} \\ -r_{3,1} \\ -r_{3,2} \\ h_{3,3} - r_{3,3} \end{bmatrix} \quad (\text{B.11})$$

In order to prove that the linear equation system (B.11) has five linear independent equations it is sufficient to show that the system matrix has five linear independent rows. It is clear that the matrix A_c has at most rank five because it has five columns. To show that A_c has exactly rank five it is sufficient to show that a submatrix of size 5×5 of A_c exists that has a determinant different from zero. Here the determinant of $(A_c)_{\{1,4,7,8,9\},\cdot}$ is calculated which contains the first, fourth, seventh, eighth, and ninth row of the system matrix A_c . Using the structure of a general rotation (B.4) and the structure of the infinite homography (2.16) for constant camera calibration leads to the determinant

$$\det((A_c)_{\{1,4,7,8,9\},\cdot}) = \frac{\cos^3(\alpha_x) \sin(\alpha_x) \sin^4(\alpha_y) (f \sin \alpha_x + s \cos(\alpha_x) \sin(\alpha_x))}{af^4}, \quad (\text{B.12})$$

which is not zero for a general rotation. Hence the system matrix A_c has rank five. \square

Further it can be seen in (B.12) that the determinant is close to zero for small rotations about the x -axis and y -axis.

Appendix C

Self-calibration of a freely moving camera with known orientation

This chapter investigates the properties of the self-calibration equations for a freely moving camera in detail.

C.1 Freely moving camera with known orientation and varying parameters

For a freely moving camera the Fundamental matrix $F_{j,i}$ describes the mapping of a point of camera j to a line in camera i . In section 4.2 from the Fundamental matrix the calibration equation (4.26) was deduced. Comparing the matrix coefficients provides nine calibration equations

$$\begin{aligned} 0 &= -F_{3,1}f_j - e_2r_{1,1}\frac{1}{\rho_{j,i}}f_i + \frac{1}{\rho_{j,i}}r_{2,1}(e_1a_if_i - e_2s_i) + \frac{1}{\rho_{j,i}}r_{3,1}(-e_2u_i + e_1v_i) \\ 0 &= -a_jf_jF_{3,2} - e_2r_{1,2}\frac{1}{\rho_{j,i}}f_i + \frac{1}{\rho_{j,i}}r_{2,2}(e_1a_if_i - e_2s_i) + \frac{1}{\rho_{j,i}}r_{3,2}(-e_2u_i + e_1v_i) - F_{3,1}s_j \\ F_{3,3} &= -e_2r_{1,3}\frac{1}{\rho_{j,i}}f_i + \frac{1}{\rho_{j,i}}r_{2,3}(e_1a_if_i - e_2s_i) + \frac{1}{\rho_{j,i}}r_{3,3}(-e_2u_i + e_1v_i) - F_{3,1}u_j - F_{3,2}v_j \\ 0 &= -F_{2,1}f_j + e_3r_{1,1}\frac{1}{\rho_{j,i}}f_i + e_3r_{2,1}\frac{1}{\rho_{j,i}}s_i + \frac{1}{\rho_{j,i}}r_{3,1}(e_3u_i - e_1) \\ 0 &= -a_jf_jF_{2,2} + e_3r_{1,2}\frac{1}{\rho_{j,i}}f_i + e_3r_{2,2}\frac{1}{\rho_{j,i}}s_i - F_{2,1}s_j + \frac{1}{\rho_{j,i}}r_{3,2}(e_3u_i - e_1) \\ F_{2,3} &= e_3r_{1,3}\frac{1}{\rho_{j,i}}f_i + e_3r_{2,3}\frac{1}{\rho_{j,i}}s_i - F_{2,1}u_j - F_{2,2}v_j + \frac{1}{\rho_{j,i}}r_{3,3}(e_3u_i - e_1) \\ 0 &= -F_{1,1}f_j - e_3r_{2,1}\frac{1}{\rho_{j,i}}a_if_i + \frac{1}{\rho_{j,i}}r_{3,1}(-e_3v_i + e_2) \\ 0 &= -a_jf_jF_{1,2} - e_3r_{2,2}\frac{1}{\rho_{j,i}}a_if_i - F_{1,1}s_j + \frac{1}{\rho_{j,i}}r_{3,2}(-e_3v_i + e_2) \\ F_{1,3} &= -e_3r_{2,3}\frac{1}{\rho_{j,i}}a_if_i - F_{1,1}u_j - F_{1,2}v_j + \frac{1}{\rho_{j,i}}r_{3,3}(-e_3v_i + e_2) \end{aligned} \tag{C.1}$$

of a freely moving and rotating camera with known orientation. The equations (C.1) are linear in the scaled intrinsics of camera j and the intrinsics of camera i . Due to the rank two of the Fundamental matrix at most six of the equations (C.1) are linear independent. The equations can be written as a linear equation system

$$A_F \underbrace{\left[\begin{array}{ccccccccccc} f_j & s_j & u_j & a_j f_j & v_j & \frac{f_i}{\rho_{j,i}} & \frac{s_i}{\rho_{j,i}} & \frac{u_i}{\rho_{j,i}} & \frac{a_i f_i}{\rho_{j,i}} & \frac{v_i}{\rho_{j,i}} & \frac{1}{\rho_{j,i}} \end{array} \right]^T}_{\mathcal{K}} = b_F \quad (\text{C.2})$$

with the system matrix A_F for the self-calibration of a freely moving camera

$$A_F = \begin{bmatrix} -F_{3,1} & 0 & 0 & 0 & 0 & -e_2 r_{1,1} & -e_2 r_{2,1} & -e_2 r_{3,1} & e_1 r_{2,1} & e_1 r_{3,1} & 0 \\ 0 & -F_{3,1} & 0 & -F_{3,2} & 0 & -e_2 r_{1,2} & -e_2 r_{2,2} & -e_2 r_{3,2} & e_1 r_{2,2} & e_1 r_{3,2} & 0 \\ 0 & 0 & -F_{3,1} & 0 & -F_{3,2} & -e_2 r_{1,3} & -e_2 r_{2,3} & -e_2 r_{3,3} & e_1 r_{2,3} & e_1 r_{3,3} & 0 \\ -F_{2,1} & 0 & 0 & 0 & 0 & e_3 r_{1,1} & e_3 r_{2,1} & e_3 r_{3,1} & 0 & 0 & -e_1 r_{3,1} \\ 0 & -F_{2,1} & 0 & -F_{2,2} & 0 & e_3 r_{1,2} & e_3 r_{2,2} & e_3 r_{3,2} & 0 & 0 & -e_1 r_{3,2} \\ 0 & 0 & -F_{2,1} & 0 & -F_{2,2} & e_3 r_{1,3} & e_3 r_{2,3} & e_3 r_{3,3} & 0 & 0 & -e_1 r_{3,3} \\ -F_{1,1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -e_3 r_{2,1} & -e_3 r_{3,1} & e_2 r_{3,1} \\ 0 & -F_{1,1} & 0 & -F_{1,2} & 0 & 0 & 0 & 0 & -e_3 r_{2,2} & -e_3 r_{3,2} & e_2 r_{3,2} \\ 0 & 0 & -F_{1,1} & 0 & -F_{1,2} & 0 & 0 & 0 & -e_3 r_{2,3} & -e_3 r_{3,3} & e_2 r_{3,3} \end{bmatrix}.$$

The right hand side b_F of the linear equation system (C.2) is

$$b_F = [0 \ 0 \ F_{3,3} \ 0 \ 0 \ F_{2,3} \ 0 \ 0 \ F_{1,3}]^T.$$

For a general rotation from (B.4) it will be proved in the following that the matrix A in (C.2) has rank six.

Proof: The upper bound of at most six linear independent equations is given from the rank two constraint of the Fundamental matrix because one row of the Fundamental matrix is a linear combination of the other two rows. Hence three of the nine calibration equations can be described as linear combinations of the other six equations. It will shown now that the columns corresponding to the focal length, skew, and aspect ratio respectively the scaled ones of A_F in (C.2) are linear independent. Due to the fact that column rank is equal to the row rank this is equivalent to show that six linear independent rows exist which represent our calibration equations. This leads to the general result that the matrix has exactly rank six.

The columns of A_F from (C.2) that correspond to to the focal length, skew, and aspect ratio respectively the scaled focal length, aspect ratio and skew lead to a system matrix \tilde{A}_F

$$\tilde{A}_F = \begin{bmatrix} -F_{3,1} & 0 & 0 & -e_2 r_{1,1} & -e_2 r_{2,1} & e_1 r_{2,1} \\ 0 & -F_{3,1} & -F_{3,2} & -e_2 r_{1,2} & -e_2 r_{2,2} & e_1 r_{2,2} \\ 0 & 0 & 0 & -e_2 r_{1,3} & -e_2 r_{2,3} & e_1 r_{2,3} \\ -F_{2,1} & 0 & 0 & e_3 r_{1,1} & e_3 r_{2,1} & 0 \\ 0 & -F_{2,1} & -F_{2,2} & e_3 r_{1,2} & e_3 r_{2,2} & 0 \\ 0 & 0 & 0 & e_3 r_{1,3} & e_3 r_{2,3} & 0 \\ -F_{1,1} & 0 & 0 & 0 & 0 & -e_3 r_{2,1} \\ 0 & -F_{1,1} & -F_{1,2} & 0 & 0 & -e_3 r_{2,2} \\ 0 & 0 & 0 & 0 & 0 & -e_3 r_{2,3} \end{bmatrix}.$$

Gaussian elimination of \tilde{A}_F leads to a matrix $\tilde{A}_F^{Gaussian}$

$$\tilde{A}_F^{Gaussian} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

in echelon form¹ which has rank six and therefore \tilde{A}_F has also rank six. \square

The above result shows that for a freely moving camera there is no unique solution for the general self-calibration problem. From this proof it follows directly that the proposed constraints in section 4.2.2 lead to unique solutions of the self-calibration. Since all sets of constraints lead to a calibration that uses linear independent columns.

¹An $m \times n$ matrix is in echelon form if

- All rows that consist entirely of zeros are grouped together at the bottom of the matrix;
- In every non-zero row, the first entry (left to right) is 1;
- If the $(i + 1)$ st row contains non-zero entries, then the first non-zero entry is in a column to the right of the first non-zero entry in the i th row.

Appendix D

Sensitivity of the self-calibration with known rotation

This appendix will discuss the sensitivity of the self-calibration to errors in the measured rotation. The presented analysis will be done for a single camera pair with varying focal length. All other intrinsic parameters are fixed and known. At first the influence of errors in the rotation measurements for a purely rotating camera will be analyzed. In section D.2 the influence for a freely moving camera will be discussed.

D.1 Sensitivity of the self-calibration of a purely rotating camera

The sensitivity of the achieved self-calibration for the focal lengths was the subject of section 4.1.6. This section will give the detailed solution for the estimated focal lengths \hat{f}_j and \hat{f}_i of the equation system (4.23)

$$A_H[\hat{f}_j, \hat{f}_i] = b_H,$$

with the system matrix A_H similar to equation (B.2) where the columns for the known parameters skew, principal point and aspect ratio of both cameras were substituted. The resulting linear equation system was solved analytically afterwards.

This analytical solution is of high complexity. In order to get a helpful form of the resulting solution here the case for a rotation about the x -axis only and for a rotation about the y -axis only will be presented separately. Since the aspect ratio is known both rotations provide a solution of the focal lengths \hat{f}_j and \hat{f}_i . The solution for \hat{f}_j for a rotation about the x -axis is given by

$$\hat{f}_j = -\frac{4 \sin(\phi) \sin(\phi + \Delta\phi) ((\cos(\Delta\phi) + \cos(2\phi + \Delta\phi) + 2) f_i^2 + 4) f_j}{(s(\phi, \Delta\phi) + 8 \cos(2\phi + \Delta\phi) + \cos(2(2\phi + \Delta\phi)) - 14) f_i^2 - 16 \sin^2(\phi)} \quad (\text{D.1})$$

where $s(\phi, \Delta\phi)$ is

$$s(\phi, \Delta\phi) = -6 \cos(2\phi) + 8 \cos(\Delta\phi) + \cos(2\Delta\phi) + 2 \cos(2(\phi + \Delta\phi)). \quad (\text{D.2})$$

For a pure rotation about the y -axis the following solution for \hat{f}_j can be computed from (4.23)

$$\hat{f}_j = -\frac{4 \sin(\Phi) \sin(\Phi + \Delta\Phi) ((\cos(\Delta\Phi) + \cos(2\Phi + \Delta\Phi) + 2) f_i^2 + 4) f_j}{(s(\Phi, \Delta\Phi) + 8 \cos(2\Phi + \Delta\Phi) + \cos(2(2\Phi + \Delta\Phi)) - 14) f_i^2 - 16 \sin^2(\Phi)}, \quad (\text{D.3})$$

with $s(\Phi, \Delta\Phi)$ from (D.2). These solutions (D.1) and (D.3) show that the estimated focal length is disturbed with terms that depend on the cosine and the sine of the angular disturbances $\Delta\phi$ respectively $\Delta\Phi$. The resulting relative error of the estimated focal length \hat{f}_j is plotted for a fixed absolute rotation with $\phi = 10^\circ$ and $\Phi = 10^\circ$ in dependence on the angular disturbance $\Delta\phi$ in figure D.1 (a). Together with the observation that the error is nearly an odd function for small disturbances of less than four degrees with a present absolute rotation of ten degrees. For smaller absolute rotations the range of linearity is also smaller. The nearly odd function delivers the result that the mean of multiple estimations of the self-calibration is not effected by angular errors with distributions of zero mean.

Furthermore from (D.2) and (D.3) it can be seen that the error increases with increasing values of the angular disturbances $\Delta\phi$ and $\Delta\Phi$. The solution for the focal length \hat{f}_i of the second camera leads to comparable properties which are not shown in detail here.

After this analysis of the self-calibration error of purely rotating cameras for a disturbed measured rotation the next section will investigate the disturbance of the self-calibration for freely moving cameras.

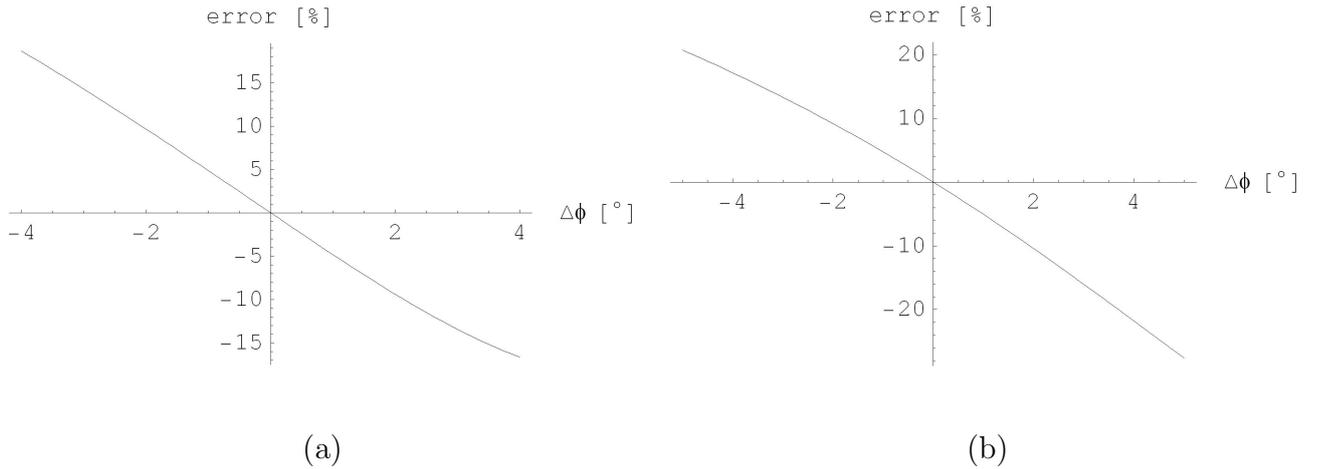


Figure D.1: (a) Relative error for a purely rotating camera of the estimated focal length \hat{f}_j in dependence on the disturbance of the rotation $\Delta\phi$ in the rotation about the x -axis. (b) Relative error for a freely rotating camera of the estimated focal length \hat{f}_j in dependence on the disturbance of the rotation $\Delta\phi$ in the rotation about the x -axis.

D.2 Sensitivity of the self-calibration of a freely moving camera

This section will show the results of the analysis done in section 4.2.4. The focal lengths \hat{f}_j and \hat{f}_i were computed as the solution of the linear equation system

$$A_F \begin{bmatrix} \hat{f}_j \\ \hat{f}_i \end{bmatrix} = b_F, \quad (\text{D.4})$$

where the system matrix A_F and the right hand side b_F have been deduced from equation (C.2) with substituted known parameters skew, principal point and aspect ratio of both cameras. Please note, that due to the known aspect ratio the focal length here can be estimated from a rotation about the x -axis as well as from a rotation about the y -axis.

The analytical solution of the linear equation system (D.4) is of high complexity. In order to obtain a helpful presentation here the case for a rotation about the x -axis and the solution for a rotation about the y -axis will be given separately. The epipole e is given as $e = [e_1, e_2, e_3]^T$. The solution for the focal length \hat{f}_j with a rotation about the x -axis is given by

$$\hat{f}_j = \frac{s_\phi (-e_2(F)_{1,2} + e_1(F)_{2,2} - \omega_\phi)}{D_F} \quad (\text{D.5})$$

with the following substitutions for D_F , ω_ϕ , Ω_ϕ , s_ϕ and c_ϕ

$$\begin{aligned} D_F &= (F)_{1,1}^2 + (F)_{1,2}^2 + (F)_{2,1}^2 + (F)_{2,2}^2 + (F)_{3,1}^2 + (F)_{3,2}^2 \\ \omega_\phi &= \frac{(e_3 (c_\phi(F)_{1,2} - (F)_{2,1}) + e_2(F)_{3,1} - c_\phi e_1(F)_{3,2}) \Omega_\phi}{(e_1^2 + e_2^2 + 2e_3^2) D_F - (e_3 (c_\phi(F)_{1,2} - (F)_{2,1}) + e_2(F)_{3,1} - c_\phi e_1(F)_{3,2})^2} \\ \Omega_\phi &= (e_2(F)_{1,2} - e_1(F)_{2,2}) (e_3 (c_\phi(F)_{1,2} - (F)_{2,1}) + e_2(F)_{3,1} - c_\phi e_1(F)_{3,2}) \\ &\quad - D_F (e_3(F)_{1,3} - e_1(F)_{3,3}) \\ s_\phi &= \sin(\phi + \Delta\phi) \\ c_\phi &= \cos(\phi + \Delta\phi), \end{aligned}$$

where $(F)_{r,c}$ is the element of $F_{j,i}$ in row r and column c . The solution for \hat{f}_j for a rotation about the y -axis is given by

$$\hat{f}_j = \frac{s_\Phi (e_2(F)_{1,1} - e_1(F)_{2,1} - \omega_\Phi)}{D_F} \quad (\text{D.6})$$

where D_F is defined as above and ω_Φ , Ω_Φ , s_Φ and c_Φ are

$$\begin{aligned} \omega_\Phi &= \frac{(e_3 ((F)_{1,2} - c_\Phi(F)_{2,1}) + c_\Phi e_2(F)_{3,1} - e_1(F)_{3,2}) \Omega_\Phi}{(e_1^2 + e_2^2 + 2e_3^2) D_F - (e_3 ((F)_{1,2} - c_\Phi(F)_{2,1}) + c_\Phi e_2(F)_{3,1} - e_1(F)_{3,2})^2} \\ \Omega_\Phi &= (e_2(F)_{1,1} - e_1(F)_{2,1}) (e_3 ((F)_{1,2} - c_\Phi(F)_{2,1}) + c_\Phi e_2(F)_{3,1} - e_1(F)_{3,2}) \\ &\quad - D_F (e_3(F)_{2,3} - e_2(F)_{3,3}) \\ s_\Phi &= \sin(\Phi + \Delta\Phi) \\ c_\Phi &= \cos(\Phi + \Delta\Phi). \end{aligned}$$

From (D.5) and (D.6) it can be seen that the estimated focal length \hat{f}_j is influenced by the cosine of the angular error as well as by the sine of the angular error. Furthermore it can be seen that the sign of the error can be positive as well as negative. The error function is nearly an odd function for disturbances $\Delta\phi$ that are small (approximately less than four degrees with ten degrees absolute rotation). This implies that the mean value of the estimated focal length is not disturbed by the angular error in x -direction and y -direction as long as it has a zero mean. The relative error of the estimated focal length is illustrated in figure D.1 (b) for an absolute rotation of $\phi = 10^\circ$ and $\Phi = 10^\circ$ in dependence on the angular disturbance $\Delta\phi$. A plot of the analytical solution for errors in x -direction and y -direction in dependence on the absolute rotation is shown in figure 4.18 in section 4.2.4.

The camera motion influences the epipole as explained in section 2.5.2 and from (D.5) and (D.6) it follows that it also influences the error of the focal lengths. At first a scale in the epipole changes the scale of the error. This can be avoided by normalizing the epipole to Euclidian norm one. The major influence comes from components of the epipole that are zero which eliminate terms in the estimation equations (D.5) and (D.6). Investigating the error shows that the error in case of zero epipole components also only depends on the rotation angles ϕ respectively Φ and the angular disturbances $\Delta\phi$ respectively $\Delta\Phi$. The results for the focal length of the second image \hat{f}_i are comparable to the results for \hat{f}_j and will be therefore not presented here.

Appendix E

Critical rotations for self-calibration with known orientation

This appendix provides the details for the analysis of the critical rotations for the proposed self-calibration approaches with known camera orientation. The detailed deduction for a rotating camera will be shown here. For a freely moving camera the analysis can be done similarly. Hence for freely moving cameras the results will be discussed shortly after the detailed deduction for the purely rotating camera. These results for freely moving cameras are valid for all translations because for each epipole even with zero components six linear independent equations are provided by (C.2). The analysis will be done for the equations provided by a single camera pair. Due to the block structure of the resulting equation system for three or more cameras each additional camera will provide the same number of linear independent equations as well as the same types of constraints.

For the deductions in the following sections the structure of the homography $H_{j,i}^\infty$ respectively the structure of the Fundamental matrix $F_{j,i}$ is used. In order to take the structure into account the true camera calibrations K_j and K_i

$$K_j = \begin{bmatrix} f_j & s_j & u_j \\ 0 & a_j \cdot f_j & v_j \\ 0 & 0 & 1 \end{bmatrix}, \quad K_i = \begin{bmatrix} f_i & s_i & u_i \\ 0 & a_i \cdot f_i & v_i \\ 0 & 0 & 1 \end{bmatrix}$$

are used together with the rotation matrix to formulate the structure of each element. This is needed for a detailed analysis of the effects of each intrinsic parameter. These explicit forms of the matrix elements in the true camera intrinsics are used in the equation systems (B.2) respectively (C.2). To distinguish between true and estimated parameters the estimated parameters are denoted with a tilde in the following sections. For simplicity and without loss of generality the scale of the homography and the Fundamental matrix is assumed to be one.

E.1 No rotation of the camera

To investigate the effect of no rotation for self-calibration using (4.3) the structure (2.16) of the infinite homography will be used. Using an identity matrix as rotation matrix leads to the

homography $H_{j,i}^\infty$

$$H_{j,i}^\infty = K_i K_j^{-1}. \quad (\text{E.1})$$

This homography is applied together with the identity rotation to the linear equation system (B.2) for the self-calibration. Using Gaussian elimination [GL89] gives an echelon form of the equation system

$$\underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & -\frac{f_j}{f_i} & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & -\frac{f_j}{f_i} & 0 & \frac{f_j s_i - f_i s_j}{a_i f_i^2} & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & -\frac{f_j}{f_i} & 0 & \frac{f_j s_i - f_i s_j}{a_i f_i^2} & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & -\frac{a_j f_j}{a_i f_i} & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & -\frac{a_j f_j}{a_i f_i} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}}_{A_I} \mathcal{K} = b_I \quad (\text{E.2})$$

with the right hand side

$$b_I = \left[0 \quad 0 \quad \frac{a_i f_i f_i u_j - a_i f_i f_j u_i - f_i s_j v_i + f_j s_i v_i}{a_i f_i^2} \quad 0 \quad \frac{a_i f_i v_j - a_j f_j v_i}{a_i f_i} \quad 1 \right]^T$$

and the vector \mathcal{K} of the unknown intrinsic parameters

$$\mathcal{K} = \left[\tilde{f}_j \quad \tilde{s}_j \quad \tilde{u}_j \quad \tilde{a}_j \tilde{f}_j \quad \tilde{v}_j \quad \tilde{\rho}_{j,i} \tilde{f}_i \quad \tilde{\rho}_{j,i} \tilde{s}_i \quad \tilde{\rho}_{j,i} \tilde{u}_i \quad \tilde{\rho}_{j,i} \tilde{a}_i \tilde{f}_i \quad \tilde{\rho}_{j,i} \tilde{v}_i \quad \tilde{\rho}_{j,i} \right]^T$$

It can be seen that the resulting equation system provides six linear independent equations since (E.2) has six linear independent columns. For varying intrinsics of the camera and non-zero skew the principal point in y can be estimated. From the equations (E.2) it can be seen that the focal lengths, the aspect ratio, the skews and the principal point in x of the cameras can't be estimated uniquely since the columns corresponding to these parameters are linear dependent. Furthermore it can be seen that the linear dependent columns provide constraints on the ratios of the focal lengths, the aspect ratios, the skews and the principal points in x . The constraint on the principal point in y is more of theoretical interest since the poorly conditioned estimation of the principal point makes it numerically difficult to compute it here and most cameras have skew zero. The six provided equations lead to a unique solution of the principal point in y from a single camera pair.

For the self-calibration of a freely moving camera using (4.26) with a known camera orientation and the Fundamental matrix $F_{j,i}$ is

$$F_{j,i} = [e_i]_\times K_i K_j^{-1} \quad (\text{E.3})$$

leads to five linear independent constraints on the principal point in y . The constraints are comparable to the previously described constraints for a rotating camera. The number of provided constraints is sufficient to calibrate the remaining principal points in y of the cameras. The next section will discuss the equations for a rotation about the x axis.

E.2 Rotation about the x -axis

The next rotation that was identified as critical is the rotation about the x -axis of the coordinate system. The rotation matrix for this type of rotation is given by

$$R_x(\phi) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \phi & -\sin \phi \\ 0 & \sin \phi & \cos \phi \end{bmatrix}. \quad (\text{E.4})$$

Using the structure (2.16) of the infinite homography together with the rotation $R_x(\phi)$ in the linear equation system (B.2) for the self-calibration provides the following equations in echelon form after Gaussian elimination

$$A_x \mathcal{K} = b_x \quad (\text{E.5})$$

where the system matrix A_x is given by

$$\underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & -\frac{f_j}{f_i} & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & -\frac{f_j \cos(\phi)}{f_i} & -\frac{f_j \sin(\phi)}{f_i} & 0 & 0 & \frac{-f_i s_j + f_j s_i \cos(\phi) + f_j u_i \sin(\phi)}{f_i} \\ 0 & 0 & 1 & 0 & 0 & 0 & \frac{f_j \sin(\phi)}{f_i} & -\frac{f_j \cos(\phi)}{f_i} & 0 & 0 & \frac{\cot(\phi)(-f_i s_j + f_j s_i \cos(\phi) + f_j u_i \sin(\phi))}{f_i} \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & -a_j f_j \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & -a_j f_j \cot(\phi) \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -\frac{1}{\sin(\phi)}(a_i f_i \cos(\phi) + v_i \sin(\phi)) \end{bmatrix}}_{A_x}$$

and the right hand side B_x becomes

$$b_x = [0 \quad 0 \quad b_3^x \quad 0 \quad b_5^x \quad a_i f_i \cos(\phi) \quad -a_i f_i \cot(\phi)]$$

with

$$b_3^x = \frac{1}{\sin(\phi)} \frac{(f_j s_i - f_i s_j \cos(\phi) + f_i u_j \sin(\phi))}{f_i}$$

$$b_5^x = \frac{1}{\sin(\phi)} (v_j \sin(\phi) - a_j f_j).$$

From these equations which exploit the special structure of the homography and the rotation from (E.4). It can be seen that the focal lengths of the cameras become ambiguous due to the linear dependency of the corresponding columns of A_x in (E.5). This leads to a constraint only on the ratio of the focal lengths of the cameras.

All other intrinsic parameters can still be estimated since the corresponding columns are still linear independent. At a first glance it seems to be surprising that the principal point in x still can be estimated. This can be explained by the motions of the features under rotation about the x -axis. In this case the features are moving not on a column of the image, they move on a sinus curve in the image. This slight side motion provides information about the principal point in x direction and also the estimated skew \tilde{s} . It is also clear from the linear independent

columns corresponding to the skew that it can be estimated from a pure rotation about the x -axis. Moreover it can be seen that it is possible to compute the product $\tilde{a}\tilde{f}$ of aspect ratio \tilde{a} and focal length \tilde{f} .

It follows from (E.5) that the equation system provides seven linear independent equations for estimating the product af of aspect ratio \tilde{a} and focal length \tilde{f} as well as for the principal point $[\tilde{u}, \tilde{v}]$ and the skew \tilde{s} .

An analogous deduction for a freely moving camera using (4.26) and the rotation $R_x(\phi)$ provides the same type of constraints. Five of these equations are linear independent. From these five equations the product $\tilde{a}\tilde{f}$ of focal length \tilde{f} and aspect ratio \tilde{a} as well as the skew and principal point can't be computed uniquely at all. In order to get a unique solution of the self-calibration additional constraints on the intrinsic parameters are needed. The next section will investigate the equations for a pure rotation about the y -axis.

E.3 Rotation about the y -axis

The pure rotation about the y -axis often occurs for surveillance cameras that use only a panning motion to extend the field of view to the left and right. This section will show which of the intrinsic parameters of the camera can be estimated under this camera motion. For a purely panning camera the rotation is described by

$$R_y(\Phi) = \begin{bmatrix} \cos \Phi & 0 & \sin \Phi \\ 0 & 1 & 0 \\ -\sin \Phi & 0 & \cos \Phi \end{bmatrix}. \quad (\text{E.6})$$

This rotation $R_y(\Phi)$ was applied to (B.2), afterwards the homography resulting from (2.16) and the special rotation matrix from (E.6) is applied to the linear equations in (B.2). The resulting equation system was reduced to echelon form using Gaussian elimination this leads to

$$A_y \mathcal{K} = b_y \quad (\text{E.7})$$

where the system matrix A_y is given by

$$\underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -f_j \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & -\frac{s_j}{a_i f_i} & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & f_j \cot(\Phi) & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & -\frac{a_j f_j}{a_i f_i} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & -\frac{s_i}{a_i f_i} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & \frac{1}{\sin(\Phi)}(f_i \cos(\Phi) - u_i \sin(\Phi)) & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & -v_i \end{bmatrix}}_y$$

with the right hand side B_y

$$b_y = \begin{bmatrix} 0 & 0 & \frac{1}{\sin(\Phi)}(f_j \cos(\Phi) + u_j \sin(\Phi)) & 0 & v_j & f_i & 0 & f_i \cot(\Phi) & 0 \end{bmatrix}.$$

The resulting equation system still provides nine linear independent constraints on the intrinsic parameters of the cameras as in the case of a general rotation. This asymmetry to a rotation about the x -axis results from the modeling of the skew in the camera calibration matrix which is biased towards the x -direction. It can be seen from (E.7) that the principal points and the focal lengths can be determined uniquely from the resulting equations since the corresponding columns are linear independent. Furthermore it follows that the focal length in y -direction $\tilde{a}\tilde{f}$ and the skew can't be determined. The equations (E.7) only provide constraints about the ratio of the aspect ratios and the ratio of the skews. These ambiguities correspond to the ambiguous estimation of the focal length given a rotation about the x -axis. The principal point and the focal lengths of the cameras can be estimated uniquely using the equations delivered by (E.7).

The same type of constraints can be deduced for a freely moving camera with rotation $R_y(\Phi)$. Then the self-calibration equations (4.26) still provide six linear independent constraints for the camera parameters. This means that the self-calibration of focal length \tilde{f} and principal point is now possible from six cameras. The only parameters that can't be determined are then the aspect ratio \tilde{a} and the skew of the involved cameras. The next section will discuss the last critical rotation of a camera the rotation about the z -axis.

E.4 Rotation about the z -axis

Similar to the previous cases of rotations about one coordinate axis at a time the special structure of the rotation matrix of a camera that rotates about the z -axis

$$R_z(\Theta) = \begin{bmatrix} \cos \Theta & -\sin \Theta & 0 \\ \sin \Theta & \cos \Theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (\text{E.8})$$

can be used to deduce the structure of the corresponding homography. The rotation and the homography are applied to the linear self-calibration equations (B.2). This provides again a system

$$A_z \mathcal{K} = b_z$$

in echelon form after a Gaussian elimination.

$$A_z = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\frac{f_j}{a_i f_i} & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & (A_z)_{2,7} & 0 & (A_z)_{2,9} & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & (A_z)_{3,8} & 0 & (A_z)_{3,10} & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & -\frac{a_j f_j \tan(\Theta)}{f_i} & 0 & (A_z)_{4,9} & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & (A_z)_{5,8} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & -\tan(\Theta) & 0 & \frac{1}{\cos(\Theta)}(s_i \sin(\Theta) - f_i \cos(\Theta)) & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{a_i f_i}{a_i f_i} & 0 & 1 \end{bmatrix} \quad (\text{E.9})$$

with

$$\begin{aligned}
(A_z)_{2,7} &= \frac{\frac{1}{\cos(\Theta)}(-f_j \cos(\Theta) - s_j \sin(\Theta))}{f_i} \\
(A_z)_{2,9} &= \frac{\frac{1}{\cos(\Theta)}(-f_i s_j \cos(\Theta) + f_j s_i \cos(\Theta) + s_j s_i \sin(\Theta))}{a_i f_i f_i} \\
(A_z)_{3,8} &= -\frac{\frac{1}{\cos^2(\Theta)}(f_j \cos(\Theta) + s_j \sin(\Theta))}{f_i} \\
(A_z)_{3,10} &= \frac{\frac{1}{\cos^2(\Theta)}(-f_i s_j \cos(\Theta) + f_j s_i \cos(\Theta) + s_j s_i \sin(\Theta))}{a_i f_i f_i} \\
(A_z)_{4,9} &= -\frac{\frac{a_j f_j}{\cos(\Theta)}(f_i \cos(\Theta) - s_i \sin(\Theta))}{a_i f_i f_i} \\
(A_z)_{5,8} &= -\frac{a_j f_j \sin(\Theta)}{\cos^2(\Theta) f_i}
\end{aligned}$$

$$b_z = \begin{bmatrix} 0 \\ 0 \\ \frac{\frac{1}{\cos^2(\Theta)}(a_i f_i^2 (u_j + 1) \cos(2z) u_j - 2 f_i (a_i f_j u_i \cos(\Theta) - s_j v_i \cos(\Theta) - a_i s_j u_i \sin(\Theta)) + 2 v_i (f_j s_i \cos(\Theta) + s_j s_i \sin(\Theta))}{2 a_i f_i f_i} \\ 0 \\ -\frac{\frac{a_j f_j}{\cos^2(\Theta)}(f_i \cos(\Theta) - s_i \sin(\Theta))}{a_i f_i f_i} \\ 0 \\ 1 \end{bmatrix}.$$

The resulting equation system has seven linear independent equations that provide a unique solution for the principal point and the skew of the camera. The focal length and the aspect ratio of the camera can't be determined linearly since the corresponding columns are linear dependent. The equations only deliver constraints on the ratio of the focal lengths and the ratio of the focal length in y -direction $\tilde{a}\tilde{f}$. After an automatic detection of this type of critical rotation the provided set of equations is sufficient to compute a unique self-calibration of the remaining parameters principal point and skew.

For a freely moving camera the equations (4.26) deliver six linear independent constraints on the principal point of the camera and the skew. Furthermore also constraints on the ration of the focal lengths of the cameras and on the ratio of the focal length in y -direction $\tilde{a}\tilde{f}$ are submitted. Hence the remaining parameters can be estimated uniquely from a camera triplet.

Bibliography

- [AHP93] A. Azerbayejani, B. Horowitz, and A. Pentland. Recursive estimation of structure from motion using relative orientation constraints. In *Proceedings International Conference of Computer Vision and Pattern Recognition*, pages 294–299, 1993.
- [Ale98] S.M. Alessandrini. A motivational example for the numerical solution of the algebraic eigenvalue problem. *SIAM Review*, 40(4), December 1998.
- [AS01] S. Avidan and A. Shashua. Threading Fundamental Matrices. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 23(1):73–77, 2001.
- [BAHH92] J. R. Bergen, P. Anandan, K.J. Hanna, and R. Hingorani. Hierarchical Model-Based Motion Estimation. In *Proceedings European Conference Computer Vision*, pages 237–252, 1992.
- [Bas93] Anup Basu. Active Calibration: Alternative, Strategy and Analysis. In *Proceedings IEEE Conference on Computer Vision and Pattern Recognition*, pages 495–500, 1993.
- [BM04] S. Baker and I. Matthews. Lucas-Kanade 20 Years On: A Unifying Framework. *International Journal of Computer Vision*, 56(3):221–255, March 2004.
- [Bou98] S. Bougnoux. From projective to Euclidean space under any practical situation, a criticism of self-calibration. In *Proceedings International Conference on Computer Vision*, pages 790–796, 1998.
- [BR96] M.J. Black and A. Rangarajan. On the Unification of Line Processes, Outlier Rejection, and RobustStatistics with Applications in Early Vision. *International Journal of Computer Vision*, 19(1):57–91, 1996.
- [BR97] Anup Basu and K. Ravi. Active camera calibration using pan, tilt and roll. *IEEE Transactions on Systems, Man, and Cybernetics*, 27(3):559–566, June 1997.
- [BTZ96] P. A. Beardsley, P. H. S. Torr, and A. Zisserman. 3D Model Acquisition from Extended Image Sequences. In *Proceedings 4th European Conference on Computer Vision*, Lecture Notes Computer Science, pages 683–695, 1996.
- [BZ87] Andrew Blake and Andrew Zisserman. *Visual Reconstruction*. The MIT Press, Cambridge, Massachusetts, 1987.

- [CI02] Yaron Caspi and Michal Irani. Aligning Non-overlapping Sequences. *International Journal of Computer Vision (IJCV)*, 48(1):39–52, 2002.
- [dAHR98] L. de Agapito, E. Hayman, and I. Reid. Self-calibration of a rotating camera with varying intrinsic parameters. In *Proceedings 9th British Machine Vision Conference, Southampton*, pages 105–114, 1998.
- [dAHR01] L. de Agapito, E. Hayman, and I. Reid. Self-Calibration of Rotating and Zooming Cameras. *International Journal of Computer Vision*, 45(2):107–127, November 2001.
- [Dan99] Kostas Daniilidis. Hand-eye calibration using dual quaternions. *The International Journal of Robotics Research*, 18:286–298, 1999.
- [dARH99] L. de Agapito, I. Reid, and E. Hayman. Linear calibration of a rotating and zooming camera. In *Proceedings IEEE Conference on Computer Vision and Pattern Recognition*, pages 15–21, 1999.
- [DB93] F. Du and M. Brady. Self-calibration of the intrinsic parameters of cameras for active vision systems. In *Proceedings Computer Vision and Pattern Recognition*, 1993.
- [Fau92] Oliver Faugeras. What can be seen in three dimensions with an uncalibrated stereo rig. In *Proceedings European Conference on Computer Vision*, pages 563–578, 1992.
- [Fau93] Oliver Faugeras. *Three-Dimensional Computer Vision: A Geometric Viewpoint*. MIT Press, Cambridge, MA, 1993.
- [FB81] M. A. Fishler and R. C. Bolles. Random sampling Consensus: A paradigm for model fitting with application to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, June 1981.
- [FH92] Oliver D. Faugeras and M. Herbert. The representation, recognition and locating of 3-D objects. *International Journal of Robotics Research*, 5(3):27–52, 1992.
- [FK03a] Jan-Michael Frahm and Reinhard Koch. Camera Calibration and 3D Scene Reconstruction from image sequence and rotation sensor data. In *Proceedings Vision, Modeling, and Visualization*, pages 79–86, 2003.
- [FK03b] Jan-Michael Frahm and Reinhard Koch. Camera Calibration with known Rotation. In *Proceedings IEEE International Conference Computer Vision*, volume 2, pages 1418–1425, 2003.
- [FK03c] Jan-Michael Frahm and Reinhard Koch. Robust Camera Calibration from Images and Rotation Data. In *Proceedings DAGM 2003*, pages 249–256. Springer, 2003.

- [FLM92] O. Faugeras, Q.T. Luong, and S.J. Maybank. Camera self-calibration: Theory and experiments. *European Conference of Computer Vision*, pages 321–334, 1992.
- [Fra83] Wm. Randolph Franklin. Efficient Iterated Rotation of an Object. *IEEE Transactions on Computing*, 11(32):1064–1067, 1983.
- [GL89] G.H. Golub and C.F. Van Loan. *Matrix computations*. The John Hopkins University Press, Baltimore, MD, second edition, 1989.
- [Gol80] H. Goldstein. *Classical Mechanics*. Addison-Wesley Series in Physics. Addison Wesley, second edition, 1980.
- [GWF03] Stefan K. Gehrig, S. Wagner, and Uwe Franke. System architecture for an intersection assistant fusing image, map, and GPS information. In *Proceedings IEEE Intelligent Vehicles Symposium*, 2003.
- [HÅ96] A. Heyden and K. Åström. Euclidean Reconstruction from Constant Intrinsic Parameters. In *Proceedings 13th International Conference Pattern Recognition, Vienna, Austria*, pages 339–343, 1996.
- [HÅ97] A. Heyden and K. Åström. Euclidean reconstruction from image sequences with varying and unknown focal length and principal point. In *Proceedings IEEE Conference Computer Vision and Pattern Recognition*, pages 438–443, 1997.
- [HÅ98] A. Heyden and K. Åström. Minimal conditions on intrinsic parameters for euclidean reconstruction. In *Proceedings Asian Conference on Computer Vision*, 1998.
- [HÅ99] A. Heyden and K. Åström. Flexible calibration: Minimal cases for auto-calibration. In *Proceedings IEEE International Conference Computer Vision*, pages 350–355, 1999.
- [Har92] Richard Hartley. Estimation of relative camera positions from uncalibrated cameras. In *Proceedings European Conference on Computer Vision*, pages 579–587, 1992.
- [Har94a] Richard Hartley. Euclidean reconstruction from uncalibrated views. In J.L. Mundy, A. Zisserman, and D. Forsyth, editors, *Applications of Invariance in Computer Vision*, volume 825, pages 237–256. Springer-Verlag, 1994.
- [Har94b] Richard Hartley. Self-calibration from multiple views with a rotating camera. In *Proceedings European Conference Computer Vision*, volume 1, pages 471–478, 1994.
- [Har95] Richard Hartley. In defence of the 8-Point-Algorithm. In *Proceedings International Conference Computer Vision*. IEEE Press, Cambridge, MA, USA, 1995.

- [Har97a] Richard Hartley. Kruppas equations derived from the Fundamental matrix. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(2):580–593, October 1997.
- [Har97b] Richard Hartley. Self-calibration of stationary cameras. *International Journal of Computer Vision*, 22(1):5–23, February 1997.
- [Har98] Richard I. Hartley. Chirality. *International Journal of Computer Vision*, 26(1):41–61, 1998.
- [HD95] Radu Horaud and Fadi Dornaika. Hand-eye Calibration. *The International Journal of Robotics Research*, 14(3):195–210, June 1995.
- [HHdAR99] Richard Hartley, E. Hayman, L. de Agapito, and Ian Reid. Camera calibration and the search for infinity. In *Proceedings 7th International Conference of Computer Vision, Keerkyra, Greece*, pages 510–517, 1999.
- [Hil84] E. C. Hildreth. *The measurement of visual motion*. Cambridge, MA: MIT Press, 1984.
- [Hil87] E. C. Hildreth. The analysis of visual motion: from computational theory to neural mechanisms. *Annual Review of Neuro-science*, 10, 477-533, 1987.
- [HS88] C. Harris and M. Stephens. A combined corner and edge detector. In *Proceedings Alvey Vision Conference, Univ. Manchester*, pages 147–151, 1988.
- [HT99] J. Hornegger and C. Tomasi. Representation issues in the ML estimation of camera motion. In *Proceedings International Conference Computer Vision*, pages 640–647, 1999.
- [HZ03] Richard Hartley and Andrew Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, second edition, 2003.
- [ITUa] ITU-R Recommendation BT.1358: Studio parameters of 625 and 525 line progressive scan television systems.
- [ITUb] ITU-R Recommendation BT.601-5: Studio encoding parameters of digital television for standard 4:3 and wide-screen 16:9 aspect ratios.
- [Kah99] F. Kahl. Critical motions and ambiguous Euclidian reconstructions in auto-calibration. In *Proceedings International Conference Computer Vision*, pages 469–475, 1999.
- [Kan00] Sing Bing Kang. Radial distortion snakes. In *IAPR Workshop on Machine Vision Applications (MVA2000), Tokyo, Japan*, pages 603–606, November 2000.
- [Kru13] E. Kruppa. Zur Ermittlung eines Objektes aus zwei Perspektiven mit innerer Orientierung. *Sitz.-Ber. Akad. Wiss., Wien, math. naturw. Kl., Abt. IIa.*, 122:1939-1948, 1913.

- [KTÅ00] F. Kahl, B. Triggs, and K. Åström. Critical motions for auto calibration when some intrinsic parameters can vary. *Journal of Mathematical Imaging and Vision*, 13(2):131–146, 2000.
- [Kui99] J.B. Kuipers. *Quaternions and Rotation Sequences*. Princeton University Press, 1999.
- [KWT87] M. Kass, A. Wittkin, and Terzopoulos. Snakes: Active contour models. In *Proceedings International Conference Computer Vision*, pages 259–268. IEEE, June 1987.
- [LF97a] Q.-T. Luong and O. Faugeras. Self-calibration of a moving camera from point correspondences and fundamental matrices. *International Journal of Computer Vision*, 22(3):261–289, 1997.
- [LF97b] Q.-T. Luong and O. D. Faugeras. Self-Calibration of a Moving Camera from Point Correspondences and Fundamental Matrices. *International Journal Computer Vision*, 22(3):261–289, 1997.
- [LH81] H. Longuet-Higgins. A computer algorithm for reconstructing a scene from two projections. *Nature*, 293:133–135, 1981.
- [LK81] Bruce D. Lucas and Takeo Kanade. An Iterative Image Registration Technique with an Application to Stereo Vision. In *Proceedings International Joint Conference on Artificial Intelligence*, 1981.
- [Luo92] Q.T. Luong. *Matrice Fondamentale et Calibraton Visulle sur l’environnement*. PhD thesis, Universite de Paris-Sud, Centre D’Orsay, 1992.
- [MF92] S.J. Maybank and Oliver Faugeras. A theory of self-calibration of a moving camera. *International Journal of Computer Vision*, 8(2):123–151, 1992.
- [MM01] F. Mohanna and F. Mokhtarian. Performance Evaluation of Corner Detection Algorithms under Similarity and Affine Transforms. In *Proceedings British Machine Vision Conference*, 2001.
- [Moo98] T. Moons. A Guided Tour Through Multiview Relations. In *SMILE Workshop (post-ECCV’98), Lecture Notes in Computer Science 1506*, pages 304–346. Springer-Verlag, 1998.
- [NF02] L. Naimark and E. Foxlin. Circular Data Matrix Fiducial System and Robust Image Processing for a Wearable Vision-Inertial Self-Tracker. In *Proceedings International Symposium on Mixed and Augmented Reality*, 2002.
- [Nis04] D. Nistér. Untwisting a projective reconstruction. *International Journal of Computer Vision*, 60(2):165–183, 2004.

- [Pea83] C. E. Pearson. *Handbook of Applied Mathematics*. Van Nostrand Reinhold Company, 1983.
- [PFTV88] W.H. Press, B.P. Flannery, S.A. Teukolsky, and W.T. Vetterling. *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press, 1988.
- [PG97a] Marc Pollefeys and Luc Van Gool. A stratified approach to metric self-calibration. In *Proceedings Conference on Computer Vision and Pattern Recognition*, pages 407–412. IEEE Computer Society Press, 1997.
- [PG97b] Marc Pollefeys and Luc Van Gool. Self-Calibration from the Absolute Conic on the Plane at Infinity. In *Proceedings 7th International Conference on Computer Analysis of Images and Patterns*, pages 175–182. Springer-Verlag, 1997.
- [PGO96] Marc Pollefeys, Luc Van Gool, and A. Oosterlinck. The Modulus Constraint: A New Constraint for Self-Calibration. In *Proceedings 13th International Conference on Pattern Recognition*, pages 349–353. IEEE Computer Soc. Press, 1996.
- [PGO97] Marc Pollefeys, Luc Van Gool, and A. Oosterlinck. *Geometry and topology of submanifolds, VIII*. World Scientific, Singapore, New Jersey, London, Hong Kong, 1997.
- [PGP96] Marc Pollefeys, Luc Van Gool, and M. Proesmans. Euclidian 3D Reconstruction from Image Sequences with Variable Focal Length. In *Proceedings European Conference Computer Vision*, volume 1064 of *Lecture Notes Computer Science*, pages 31–42. Springer-Verlag, 1996.
- [PGV⁺04] Marc Pollefeys, Luc Van Gool, Maarten Vergauwen, Frank Verbiest, Kurt Cornelis, Jan Tops, and Reinhard Koch. Visual modeling with a hand-held camera. *International Journal of Computer Vision*, 59(3):207–232, September 2004.
- [PKG98] Marc Pollefeys, Reinhard Koch, and Luc Van Gool. Self-Calibration and Metric Reconstruction in Spite of Varying and Unknown Internal Camera Parameters. In *Proceedings IEEE International Conference Computer Vision*, pages 90–95, 1998.
- [PKG99] Marc Pollefeys, Reinhard Koch, and Luc Van Gool. Self-Calibration and Metric Reconstruction In spite of Varying and Unknown Intrinsic Camera Parameters. *International Journal of Computer Vision*, 32(32):7–25, August 1999.
- [Pol99] Marc Pollefeys. *Self-calibration and metric 3D reconstruction from uncalibrated image sequences*. PhD thesis, ESAT-PSI, K.U.Leuven, 1999.
- [RC02] Carsten Rother and Stefan Carlsson. Linear Multi View reconstruction and Camera Recovery using a Reference Plane. *International Journal of Computer Vision*, 49(2/3):117–141, 2002.
- [SA90] M. Spetsakis and J. Aloimonos. Structure from motion using line correspondences. *International Journal Computer Vision*, 4(3):171–183, 1990.

- [SH98] Y. Seo and K. Hong. Auto-calibration of a rotating camera and zooming camera. In *IAPR workshop on Machine Vision Applications, Chimba, Japan*, pages 17–19, 1998.
- [SH99] Y. Seo and K. Hong. About the self-calibration of a rotating camera and zooming camera: Theory and practice. In *Proceedings 7th International Conference of Computer Vision, Keerkyra, Greece*, pages 183–189, 1999.
- [SHK98] Harpreet S. Sawhney, Steve Hsu, and R. Kumar. Robust video mosaicing through topology inference and local to global alignment. *Lecture Notes in Computer Science*, 1407:103–119, 1998.
- [Sho85] K. Shoemake. Animating Rotation with Quaternion Curves. In *Proceedings 12th annual conference on Computer graphics and interactive techniques*, pages 245–254, San Francisco, USA, July 1985.
- [Shu93] M.D. Shuster. A Survey of Attitude Representations. *The Journal of the Astronautical Sciences*, 41(4):439–517, Oct.-Dec. 1993.
- [SK52] J.G. Semple and G.T. Kneebone. *Algebraic Projective Geometry*. Oxford University Press, 1952.
- [SK93] Richard Szeliski and Sing Bing Kang. Recovering 3D shape and motion from image streams using non-linear least-squares. Technical Report 93, DEC, 1993.
- [SK99] Harpreet S. Sawhney and Rakesh Kumar. True Multi-Image Alignment and Its Application to Mosaicing and Lens Distortion Correction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(3):235–243, 1999.
- [SM95] Cornelia Schmid and Roger Mohr. Matching by local invariants. Technical report, The French National Institute for research in computer science and control, 1995.
- [Smi92] S.M. Smith. A new class of corner finder. In *Proceedings 3rd British Machine Vision Conference*, pages 139–148, 1992.
- [Smi96] S.M. Smith. Flexible filter neighborhood designation. In *Proceedings 13th International Conference on Pattern Recognition*, volume 1, pages 206–212, 1996.
- [SN01] J. Schmidt and H. Niemann. Using Quaternions for Parameterizing 3-D Rotations in Unconstrained Nonlinear Optimization. In T. Ertl, B. Girod, G. Greiner, H. Niemann, and H.-P. Seidel, editors, *Proceedings Vision, Modeling, and Visualization 2001*, pages 399–406, Stuttgart, Germany, November 2001. AKA/IOS Press, Berlin, Amsterdam.
- [SR97] H. Shum and R.Szeliski. Panoramic Image Mosaics. Technical Report MSR-TR-97-23, Microsoft Research, 1997.

- [Ste93] G. P. Stein. Internal camera calibration using rotation and geometric shapes. Master's thesis, Massachusetts Institute of Technology, Artificial Intelligence Laboratory, 1993.
- [Ste95] G. Stein. Accurate internal camera calibration using rotation, with analysis of sources of error. In *Proceedings International Conference of Computer Vision*, 1995.
- [Ste96] G. P. Stein. Lens distortion calibration using point correspondences. A.I. Memo 1595, MIT, November 1996.
- [Stu11] P. Stumpf. Über die Abhängigkeit der visuellen Bewegungsrichtung und negativen Nachbildes von den Reizvorgängen auf der Netzhaut. *Zeitschrift für Psychologie*, 59:321–330, 1911.
- [Stu64] J. Stuelpnagel. On the parameterization of the three-dimensional rotation group. *SIAM Review*, 6(4):422–430, Oct. 1964.
- [Stu96] Peter Sturm. Critical Motion Sequences for Monocular Self-Calibration and Uncalibrated Euclidean Reconstruction. Technical report, INRIA Rhône-Alpes, 1996.
- [Stu97a] Peter Sturm. Critical motion sequences and conjugacy of ambiguous Euclidean reconstructions. In *Proceedings SCIA, Lappeenranta, Finland*, pages 439–446, 1997.
- [Stu97b] Peter Sturm. *Vision 3D non calibration la reconstruction projective et étude des mouvements critiques pour l'auto-calibrage*. PhD thesis, INPG, France, 1997.
- [Stu99] Peter Sturm. Critical motion sequences for the self-calibration of cameras and stereo systems with variable focal length. In *Proceedings British Machine Vision Conference*, 1999.
- [Stu02] Peter Sturm. Critical Motion Sequences for the Self-Calibration of Cameras and Stereo Systems with Variable Focal Length. *Image and Vision Computing*, 20(5–6):415–426, March 2002.
- [TH84] R. Tsai and T. Huang. Uniqueness and Estimation of Three-Dimensional Motion Parameters of Rigid Objects with Curved Surfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:13–27, January 1984.
- [TK92] C. Tomasi and L. Kanade. Shape and motion from image streams under orthography: A factorization approach. *International Journal Computer Vision*, 9(2):137–154, 1992.
- [TL89] R. Y. Tsai and R. K. Lenz. A new technique for fully autonomous and efficient 3D robotics hand/eye calibration. *IEEE Transactions on Robotics and Automation*, 5(3):345–358, June 1989.

- [TM97] Phil H. S. Torr and D. W. Murray. The Development and Comparison of Robust Methods for Estimating the Fundamental Matrix. *International Journal of Computer Vision*, 24(3):271–300, September–October 1997.
- [TM00] B. Tordoff and D. Murray. Violating Rotating Camera Geometry: The Effect of Radial Distortion on Self-Calibration. In *Proceedings International Conference Pattern Recognition*, pages 423–427, 2000.
- [TMHF00] Bill Triggs, Philip McLauchlan, Richard Hartley, and Andrew Fitzgibbon. Bundle Adjustment – A Modern Synthesis. In W. Triggs, A. Zisserman, and R. Szeliski, editors, *Vision Algorithms: Theory and Practice*, LNCS, pages 298–375. Springer Verlag, 2000.
- [Tor97] Phil H. S. Torr. An Assessment of Information Criteria for Motion Model Selection. In *Proceedings Conference on Computer Vision and Pattern Recognition*, page 47. IEEE Computer Society, 1997.
- [Tor98] Phil H. S. Torr. Geometric motion segmentation and model selection. *Philosophical Transactions of the Royal Society*, pages 1321–1340, 1998.
- [Tri97] Bill Triggs. Auto-calibration and the Absolute Quadric. In *Proceedings Conference on Computer Vision and Pattern Recognition*, page 609, 1997.
- [TZ98] Phil H.S. Torr and Andrew Zisserman. Robust computation and parameterization of multiple view relations. In *Proceedings IEEE Sixth International Conference on Computer Vision*, pages 727–732, 1998.
- [VL95] Thierry Viéville and Diane Lingrand. Using Singular Displacements for Uncalibrated Monocular Visual Systems. Technical Report RR-2678, INRIA Sophia-Antipolis, October 1995.
- [Wil65] J. H. Wilkinson. *The algebraic eigenvalue problem*. Clarendon Press, Oxford, 1965.
- [WKSX02] L. Wang, Sing Bing Kang, H.-Y. Shum, and G. Xu. Error analysis of pure rotation-based self-calibration. Technical report, Microsoft Research, 2002.
- [WKSX04] L. Wang, Sing Bing Kang, H.-Y. Shum, and G. Xu. Error analysis of pure rotation-based self-calibration. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 26(2):275–280, February 2004.
- [WW92] A. Watt and M. Watt. *Advanced Animation and Rendering techniques*. Addison-Wesley, 1992.
- [Zel96] C. Zeller. *Calibration projective, affine et Euclidienne en vision par ordinateur et application a la perception tridimensionnelle*. PhD thesis, Ecole Polytechnique, France, 1996.

- [ZF96] C. Zeller and Oliver Faugeras. Camera self-calibration from video sequences: the Kruppa equations revisited. Technical report, INRIA, February 1996.
- [Zha99] Z. Zhang. Flexible Camera Calibration By Viewing a Plane From Unknown Orientations. In *Proceedings International Conference on Computer Vision*, pages 666–673, Corfu, Greece, September 1999.
- [Zha00] Z. Zhang. A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(22):1330–1334, 2000.
- [ZLA98] Andrew Zisserman, David Liebowitz, and M. Armstrong. Resolving ambiguities in auto calibration. *Phil. Transaction Royal Society London*, 356:1193–1211, 1998.
- [Zom91] A. Zomotor. *Fahrwerktechnik: Fahrverhalten*. Vogel Buchverlag Würzburg, second edition, 1991.
- [ZWS01] A. Zomet, Lior Wolf, and A. Shashua. Omni-rig: Linear Self-calibration of a Rig with Varying Internal and External Parameters. In *Proceedings International Conference on Computer Vision, Vancouver*, 2001.