Camera Calibration with known Rotation

Jan-Michael Frahm and Reinhard Koch
Multimedia Information Processing, Institute of Computer Science and Applied Mathematics
Christian-Albrechts-University of Kiel, Herman-Rodewald-Str. 3, 24098 Kiel, Germany

Abstract

We address the problem of using external rotation information with uncalibrated video sequences. The main problem addressed is, what is the benefit of the orientation information for camera calibration? It is shown that in case of a rotating camera the camera calibration problem is linear even in the case that all intrinsic parameters vary. For arbitrarily moving cameras the calibration problem is also linear but underdetermined for the general case of varying all intrinsic parameters. However, if certain constraints are applied to the intrinsic parameters the camera calibration can be computed linearly. It is analyzed which constraints are needed for camera calibration of freely moving cameras. Furthermore we address the problem of aligning the camera data with the rotation sensor data in time. We give an approach to align these data in case of a rotating camera.

1. Introduction

Scene analysis from uncalibrated image sequences is still an active research topic. During the last decade we have seen a lot of progress in camera selfcalibration and 3D-scene reconstruction. All these approaches use image data alone or sometimes additional constraints for camera motion, scene structure respectively camera calibration. Since they have to rely on the available image content they may suffer from degeneracies. These approaches use uncertain data therefore they tend to be sensitive to noise. Fortunately, in many applications we have additional information from other sensors available for example future cars will be equipped with fixed or even rotating or zooming cameras for driver assistence, where at least partial orientation and translation information is available. Another type of application is surveillance with rotating and zooming cameras. These data could be integrated to improve camera calibration.

In this contribution we will discuss the possibilities to use this external orientation information for selfcalibration. We will first review the literature and existing image-based selfcalibration methods in sections 2 and 3. Selfcalibration from image and rotation data will be discussed in detail in section 4. Section 5 will give an approach to align camera data and orientation information in time. Finally we will discuss some experiments and conclude.

2. Previous work

Camera calibration has always been a subject of research in the field of computer vision. The first major work on selfcalibration of a camera by simply observing an unknown scene was presented in [14]. It was proven that selfcalibration was theoretically and practically feasible for a camera moving through an unknown scene with constant but unknown intrinsics. Since that time various methods have been developed.

Methods for the calibration of rotating cameras with unknown but constant intrinsics were first developed in [16]. The approach was extended for rotating cameras with varying intrinsic parameters by [7]. Camera selfcalibration from unknown general motion and constant intrinsics has been discussed in [14, 17, 20]. For varying intrinsics and general camera motion the selfcalibration was proven by [11, 21, 18]. An interesting approach was recently proposed by Rother and Carlsson [9]. They jointly estimate fundamental matrices and homographies from a moving camera that observes the scene and some reference plane simultaneously. The homography induced by the reference plane generates constraints that are similar to a rotation sensor and selfcalibration can be computed linearily. However, some structural constraints on the scene are necessary, while our proposed approach applies the constraints on the camera sensor only. All these approaches for selfcalibration of cameras only use the images of the cameras themselves for the calibration.

Only few approaches exist to combine image analysis and external rotation information for selfcalibration. In [15, 12] the calibration for cameras with constant intrinsics and known rotation was discussed. They use nonlinear optimization to estimate the camera parameters. More often, calibrated cameras are used in conjunction with rotation sensors to stabilize sensor drift [19]. This lack of attention is somewhat surprising since this situation occurs frequently in a variety of applications: cameras mounted in cars for driver assistence, robotic vision heads, surveillance cameras or PTZ-cameras for video conferencing often provide rotation information.

In this paper we will address one of the few cases which have not yet been explored, that of a rotating or generally moving camera with varying intrinsics and known rotation information. We will show that orientation information is helpful for camera calibration. Furthermore it is possible to detect degenerate cases for calibration like pure translation of the camera.

To use the sensor information for camera calibration an alignment of the sensor data with the image data in space and time has to be performed. An approach for the spatial alignment called hand eye calibration was used in [13]. The alignment of two imaging devices without any overlap in the views was developed by [5]. We will modify the technique of [5] to the case of a rotation sensor and a camera to align both devices in time.

3. Selfcalibration from images

In this section we will explain some general notation and summarize previous attempts for selfcalibration from images alone.

3.1. Notation

The projection of scene points onto an image by a camera may be modeled by the equation $m = PM^h$. The image point in projective coordinates is $m = [x, y, w]^T$, where $M^h = [X, Y, Z, 1]^T$ is the world point in homogeneous coordinates and P is the 3×4 camera projection matrix. The matrix P is a rank-3 matrix. If it can be decomposed as $P = K[R^T|-R^T t]$ the P-matrix is called metric, where the rotation matrix R and the translation vector t represent the Euclidian transformation between the camera and the world coordinate system. The intrinsic parameters of the camera are contained in the matrix K which is an upper triangular matrix

$$K = \begin{bmatrix} f & s & c_x \\ 0 & a \cdot f & c_y \\ 0 & 0 & 1 \end{bmatrix}, \tag{1}$$

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 (c_x, c_y) and s is a skew parameter which models the angle between columns and rows of the CCD-sensor.

For cameras rotating about their optical center the translation vector t is the null vector. Therefore the projection of scene points is equal to $m = AM$, where A is a 3×3 matrix and $M = [X, Y, Z]$ denotes the point in 3D space. The matrix A is also a rank-3 matrix which may be decomposed as $A = KR$ for metric P-matrices. The mapping of image points from camera j to camera i over the plane at infinity π_∞ is given by the homography $H^\infty_{j,i} = A_i A_j^{-1}$ which is

$$H^\infty_{j,i} = K_i R_i^T R_j K_j^{-1}, \tag{2}$$

where R_i resp. R_j is the rotation of camera i resp. camera j. We summarize these rotation matrices to $R_{j,i}$ which represents the rotation between camera j and i.

3.2. Rotating camera with known calibration

In case of known calibration matrices K_i, K_j and known rotation $R_{j,i}$ between image j and i we can fully predict the camera homographies $H^\infty_{j,i}$ by

$$H^\infty_{j,i} = K_i R_{j,i} K_j^{-1}. \tag{3}$$

The image i can be rotationally compensated w.r.t. image j and we can exploit any estimated image homography to compensate for sensor and calibration errors. Applications to this case exist for fast and robust realtime tracking if known feature markers are used [19].

We can extend this approach for varying calibration if only the first calibration K_0 of camera 0 is known. Rewriting (3) it is also possible to compute the varying calibrations K_i from the known rotations $R_{0,i}$ and estimated homographies $H^\infty_{0,i}$ by

$$K_i = \rho_{j,i} \tilde{H}^\infty_{0,i} K_0 R_{0,i}^T. \tag{4}$$

where $\tilde{H}^\infty_{0,i}$ is the homography $H_{0,i}$ of (3) scaled by $\frac{1}{\rho_{j,i}}$ which is estimated from the images themselves.

3.3. Selfcalibration of a rotating camera

Often all calibration matrices K_i are unknown or only some contraints on K_i's are given. In this case it is not possible to use (3) directly to compute the camera transformations $H^\infty_{j,i}$ and they have to be computed from the images.

There are many techniques to estimate the camera homographies $\tilde{H}^\infty_{j,i}$ from the given images[6, 7, 8]. We summarize the technique from [7] where the internal and external camera calibration can be computed from images even in the case of varying internal parameters. In [7] the infinite homography constraint (IHC)

$$K_i K_i^T = \rho^2_{j,i} \tilde{H}_{j,i} K_j K_j^T \tilde{H}^T_{j,i} \tag{5}$$

is used to get a set of linear equations by using some constraints like zero camera skew. The solution of this equation set is the dual of the image of the absolute conic (DIAC) $\omega^* = KK^T$. The calibration K_j is the Cholesky decomposition of the DIAC ω^*_j. The other calibrations K_i are computed as Cholesky decompositions of ω^*_i by using (5) to compute ω^*_i. This involves a nonlinear optimization with an IHC-based error function.

IHC calibration optimizes only the algebraic error in dependency of the estimated homographies $\tilde{H}^{\infty}_{j,i}$. In a refinement step the error is minimized statistically by a maximum likelihood (ML) or a maximum a posterori (MAP) optimization. Considering that the estimated feature positions for homography computation are disturbed by Gaussian noise and to avoid a fit to noise of poorly constrained parameters like the principal point a MAP estimation is used. In [7] the MAP estimation is given by

$$\text{MAP} = \text{argmin}_{K_i,R_i,M_l} \sum_{i=1}^{\#cameras} \sum_{l=1}^{\#points} \|m_{i,l} - K_i R_i M_l\|^2 +$$

$$\sum_{i=1}^{\#cameras} (c^i - c^{pri})^T \left[\begin{array}{cc} \sigma_x^2 & 0 \\ 0 & \sigma_y^2 \end{array} \right] (c^i - c^{pri}) \qquad (6)$$

with a given distribution $[\sigma_x, \sigma_y]$ of the principal point c and a prior c^{pri} of the principal point. $\hat{\mathbf{M}}_l$ is the projection ray of the l-th image point m_l.

In case of noise the approach of [7] sometimes fails to compute the calibration K with Cholesky decomposition of ω^* because the estimated ω^* may not be positive definit. Another problem is that if U is the number of unknown intrinsics in the first frame, and V is the number of intrinsics which may vary in subsequent images, the following counting argument for the unknown intrinsics of the cameras must hold:

$$U + V(n-1) \leq (n-1) \cdot 5 \qquad (7)$$

where n is the number of cameras. This leads to the limitation that not all intrinsics are allowed to vary in this approach.

3.4. Selfcalibration from general motion

Finally we discuss the problem of selfcalibration from freely moving cameras. For freely moving cameras and general 3D-scenes, the relation between two consecutive frames is described by the fundamental matrix [4] if the camera is translated between these frames. The fundamental matrix $F_{j,i}$ maps points from camera j to lines in camera i. Furthermore the fundamental matrix can be decomposed into a homography $H^{\pi}_{j,i}$ which maps over the plane π and an epipole e

$$F_{j,i} = [e]_x H^{\pi}_{j,i}, \qquad (8)$$

where $[\cdot]_x$ is the cross product matrix. The epipole is contained in the null space of $F_{i,j}$: $F_{i,j} \cdot e = 0$.

The fundamental matrix is independent from any projective skew. This means that the projection matrices P_j and P_i lead to the same fundamental matrix $F_{j,i}$ as the projectively skewed projection matrices $\tilde{P}_j$ and $\tilde{P}_i$ [4]. This property poses a problem when calibrating from projection matrices. Most techniques for calibration of translating and rotating cameras first estimate the projective skewed camera matrices $\tilde{P}_i$ and the inversely skewed positions $\tilde{\mathbf{M}}_k$ of the scene points from the image data with a Structure-from-Motion approach. The estimated projection matrices $\tilde{P}_i$ and the reconstructed scene points may be projectively skewed by a projective transformation $H_{4\times4}$. Then they estimate skewed projection matrices $\tilde{P}_i = PH_{4\times4}$ and inversely skewed scene points $\tilde{\mathbf{M}} = H^{-1}_{4\times4}\mathbf{M}$. For uncalibrated cameras one cannot avoid this skew and selfcalibration for the general case is concerned mainly with estimating the projective skew matrix $H_{4\times4}$ via the DIAC [4] or the absolute quadric [11, 18].

4. Selfcalibration with known rotation

In this section we will develop novel techniques to use available external orientation information for camera selfcalibration. We will address both cases of purely rotating and arbitrarily moving cameras. It is assumed that the alignment in time between orientation data and camera data is given. In section 5 we will discuss a technique to align these data.

4.1. Rotating cameras

We can exploit given rotational information to overcome the limitations on the number of varying intrinsics and the problems caused from noise during computation of ω^* in [7]. Equation (2) can be rewritten as

$$K_i R_{j,i} - \rho_{j,i} \tilde{H}^{\infty}_{j,i} K_j = 0_{3x3} \qquad (9)$$

where the rotation $R_{j,i}$ is known from the orientation sensor and the homography $\tilde{H}^{\infty}_{j,i}$ can be estimated from the images themselves. Furthermore, for known scale $\rho_{j,i}$ (9) is linear in the components of K_i and K_j. It provides nine linear independent contraints on the intrinsics of the cameras

Normally the scale $\rho_{j,i}$ is unkown then (9) can be written as

$$0_{3x3} = \tilde{K}_i R_{j,i} - H^{\infty}_{j,i} K_j \text{ with } \tilde{K}_i = \frac{1}{\rho_{j,i}} K_i, \qquad (10)$$

which is also linear in the intrinsics of the camera j and linear in the elements of $\tilde{K}_i$. Note that due to the unknown scale we now have six unknowns in $\tilde{K}_i$. K_j is unchanged, therefore we know that the matrix element $K_j(3,3)$=1. Eq. (9) provides nine linearily independent equations for each camera pair for the five intrinsics contained in K_j and the five intrinsics of K_i plus the scale $\rho_{j,i}$ contained in $\tilde{K}_i$. If there are no constraints available for the intrinsics, (10) has no unique solution for a single camera pair. With two constraints for the intrinsics or the scale $\rho_{j,i}$ the solution is unique. Alternatively, if we consider a camera triplet

(i, j, k) with estimated homographies $\tilde{H}^{\infty}_{j,i}$ and $\tilde{H}^{\infty}_{j,k}$ (10) provides

$$\begin{aligned} \tilde{K}_i R_{j,i} - H^{\infty}_{j,i} K_j &= 0_{3x3}, \\ \tilde{K}_k R_{j,k} - H^{\infty}_{j,k} K_j &= 0_{3x3}, \end{aligned} \quad (11)$$

with 17 unknowns and up to 9 independent equations for each camera pair. Therefore, for each camera triplet the solution for the intrinsics and scales is unique and can be solved even for fully varying parameters. The use of a reference image j in (11) is no limitation because the homography $H_{j,k}$ can be computed as $H_{j,k} = H_{j,i} H_{i,k}$ if it is only possible to compute pairwise homographies.

For the case of constant but unknown intrinsics (9) is equal to

$$K R_{i,j} - \rho_{j,i} \tilde{H}^{\infty}_{j,i} K = 0_{3x3}. \quad (12)$$

In this case the solution for the intrinsics is unique for a single camera pair. The scale $\rho_{i,j}$ of $\tilde{H}_{j,i}$ can be computed directly from the homography $\tilde{H}_{j,i}$ itself because it is a conjugated rotation matrix[1], therefore the eigenvalues of $\tilde{H}^{\infty}_{j,i}$ are [10, 1]

$$\text{eigval}(\tilde{H}^{\infty}_{j,i}) = \rho_{i,j}[1, \cos(\phi) + i\sin(\phi), \cos(\phi) - i\sin(\phi)] \quad (13)$$

where ϕ is the rotation angle about the axis given by the eigenvector corresponding to the eigenvalue 1. The scale $\rho_{i,j}$ is the unknown scale of the homography. With (13) we are also able to decide whether the camera calibration is constant or the camera has varying intrinsics.

This novel linear technique to compute the intrinsics of the cameras substitutes the linear and the nonlinear estimation steps of the algorithm from [7]. To avoid error propagation caused by algebraic error during estimation of $\tilde{H}^{\infty}_{j,i}$ and error of orientation sensor we can use the MAP optimization from [3] to get exact calibration and to stabilize the orientation sensor.

Evaluation for rotating cameras: To measure the noise robustness of the calibration we test the approach with synthetic data. The center of the rotating camera is at the origin of the coordinate system. The camera rotates about x-axis and y-axis with up to six degrees and observes a uniformly distributed scene in front of the camera. The scene points are uniformly distributed in a cube and projected into the images of size 512x512. The location of the projected points is disturbed by uniformly distributed noise with maximum of 2 pixel. The known camera orientation is also disturbed by uniformly distributed angular noise of up to 2 degrees per axis. We varied both pixel and rotational noise.

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

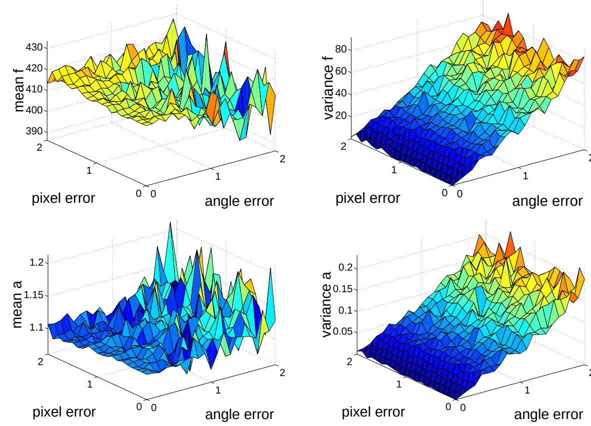


Figure 1: Noise robustness measurements. Top left: mean of estimated focal length f, top right: variance of estimated focal length f, bottom left: mean estimated aspect ratio a, bottom right: variance of estimated aspect ratio a.

The homographies $\tilde{H}^{\infty}_{j,i}$ are estimated from point correspondences by least squares estimation using feature point normalisation as in [2]. The measurements for the first camera with focal length $f = 415$ and aspect ratio $a = 1.1$ are shown in figure 1. The measured errors and variances for the other images are similar to this results.

It can be seen from figure 1 that the estimated calibration is rather stable and the variance is below 10% if the pixel noise is less than one pixel and the orientation data are noisy by angular errors of less than one degree. The influence of the orientation noise is much larger since the absolute rotation angle between the cameras is in the range of the noise (6 degree camera rotation with up to 2 degree noise). Since the error of orientation sensors like the InertiaCube[2] from InterSense is normally in the range below one degree, we can rely on the rotation information. The homography estimation can also be estimated with an error of less than 1 pixel for the features' positions in most situations. This shows that the proposed calibration with (11) is robust for most applications.

4.2. Calibration for freely moving cameras

We will investigate how to combine rotational information and the Fundamental matrix $F_{j,i}$ in the general motion case as introduced in section 3.4. The Fundamental matrix is not affected by projective skew, therefore we will use $F_{j,i}$ in the following to calibrate the cameras.

In (8) the homography $H^{\pi}_{j,i}$ is element of the three parameter family [4]

$$H^{\pi}_{j,i} = H^{\infty}_{j,i} - e v^T,$$

where H^{∞} is the homography which maps over the plane at infinity. Without loss of generality we assume that $v =$

$[0, 0, 0]^T$. With (2) and (8) we get

$$F_{j,i} = [e]_x K_i R_{j,i} K_j^{-1}. \tag{14}$$

This is equivalent to

$$[e]_x K_i R_{j,i} - F_{j,i} K_j = 0_{3\times 3}, \tag{15}$$

which is linear in the intrinsics of camera i and camera j. Please note the relationship to Eq. (9). One can see that (15) is an extension of (9) which contains the unknown camera translation t in the epipole. Equation (15) provides six linear independent equations for the intrinsics of the cameras. So we need five image pairs to compute the calibration of the first camera in case of fully varying intrinsics.

The fundamental matrices $\tilde{F}_{j,i}$ that have to be estimated from the images are scaled by an arbitrary scale $\rho_{j,i}$

$$\tilde{F}_{j,i} = \rho_{j,i} F_{j,i}. \tag{16}$$

For these estimated fundamental matrices $\tilde{F}_{j,i}$ (15) is

$$\begin{aligned} 0_{3\times 3} &= [e]_x K_i R_{j,i} - \tilde{F}_{j,i} K_j \\ &= [e]_x \tilde{K}_i R_{j,i} - F_{j,i} K_j \end{aligned} \tag{17}$$

which is also linear in the intrinsics of camera j and the scaled intrinsics of camera i in conjunction with the scale $\frac{1}{\rho_{j,i}}$. The matrices $[e]_x \tilde{K}_i R_{j,i}$ and $F_{j,i} K_j$ have rank 2, for this reason we only have to use two rows of (17) for computation and it provides only six linearily independent equations for the scale and the intrinsics of the cameras. It follows from the counting argument that the solution is never unique if no constraints for the scales $\frac{1}{\rho_{j,i}}$ or the intrinsics of the cameras are available.

Now we will discuss the most important constraints to get a unique solution to the calibration problem. We can constrain the K_i's by different parameters settings:

- known principal point: The solution for the focal length, aspect ratio and skew is unique if we use two image pairs.
- known skew and principal point: We can estimate the focal length and aspect ratio directly from a single fundamental matrix and the rotation.
- known skew, known aspect ratio and principal point: The solution for the focal length is unique for one image pair. Note that this case is also linear in the case of unknown rotation [4].

These constraints can be applied for efficient selfcalibration in case of general camera trajectory.

Furthermore the known rotation can be used to detect critical motion sequences for the solution of the selfcalibration problem. Critical motion sequences mean that it is not possible to fully determine the projective skewing homography $H_{4\times 4}$ and therefore the camera can't be calibrated completely. Pure translation of the camera can be detected from zero rotation about all axes. In this case the reconstruction is only affine. Another critical motion is planar translation of the camera and rotation about an axis perpendicular to that plane. This critical motion can be detected by measuring the orthogonality of the eigenvector corresponding to the real eigenvalue one of the rotation matrix (Eq. (13) in section 4) and the camera motion plane.

Evaluation for freely moving cameras: To measure the noise robustness of the proposed calibration for arbitrarily moving cameras we use synthetic data with known noise and ground truth information. Six cameras are positioned on a sphere looking inside and observing randomly distributed points inside the sphere. The 3D points are projected to the cameras and the corresponding image points are disturbed with pixel noise of up to 2 pixel. The images are 512x512 pixel. These projections are normalized [2] and we calculate the fundamental matrices $\tilde{F}_{j,i}$ for the image pairs by least squares estimation. The computed fundamental matrices $\tilde{F}_{j,i}$ are used for the robustness measurements. The known orientation of the cameras is also disturbed by angular noise of up to 2 degrees. The results for the case of known principal point (c_x, c_y) and known skew s are shown in figure 2 for the first camera with focal length $f = 415$ and aspect ratio $a = 1.1$. The errors and variances for the other images are very similar to these measurements.

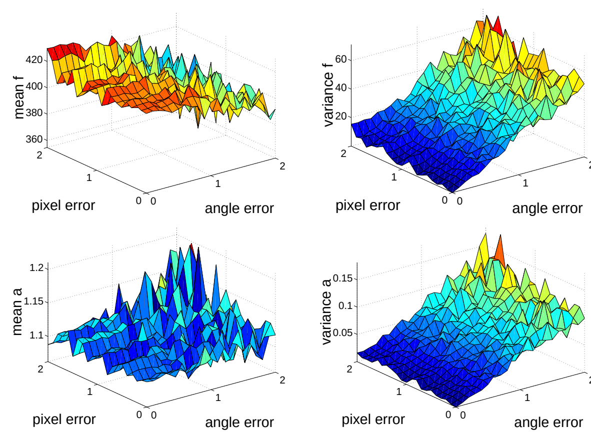


Figure 2: Noise robustness measurements. Top left: mean of estimated focal length f, top right: variance of estimated focal length f, bottom left: mean estimated aspect ratio a, bottom right: variance of estimated aspect ratio a.

It can be seen that for orientation noise of up to 1 degree and pixel noise of up to 1 pixel the calibration is rather stable. The noise sensitivity for this calibration is very similar to the rotational case, but one can see a slightly larger influence of pixel noise for F-estimation. For larger noise levels the quality of the estimated parameters is not as good as in the purely rotational case but they may still be used as good starting values for full nonlinear selfcalibration.

5 Alignment of orientation sensor and rotating camera

We have to address the problem of aligning orientation sensor data and the camera data in time. When grabbing an image or reading sensor data, the computer system adds unknown latencies to the data. Measurements have shown that typical latencies for digital image acquisition are in the range of 50 to 150 ms. Sometimes this latency information is available from time stamps of sensor and camera data, otherwise we have to estimate the timeshift t_s between the camera data and the sensor from the data itself.

In this section we show that for a rotating camera it is possible to align the orientation sensor data to the camera data by using the estimated homographies $\tilde{H}_{j,i}$ and the rotation $R_{j,i}$. We will describe two different alignment approaches to compute t_s depending on the camera type. The first approach assumes a camera with fixed internal parameters. In the second approach we will address the case that the intrinsics of the camera vary.

5.1 Rotation alignment with fixed internal parameters

For a rotating camera with fixed internal parameters we don't have to calibrate the internal camera parameters because the homography between two views is a conjugated rotation matrix for constant K (see section 3.1). For this case Caspi and Irani [5] developed a similarity measure that exploits the eigenvalue structure of the homographies to align two spatially coupled camera sequences 1 and 2. They sort the eigenvalues of the homographies in descending order (eigenvalue vector) and compare them by

$$\text{sim}_{\text{eig}}(\tilde{H}^1_{j,i}, \tilde{H}^2_{j,i}) = \frac{\text{eig}(\tilde{H}^1_{j,i})\text{eig}(\tilde{H}^2_{j,i})}{\|\text{eig}(\tilde{H}^1_{j,i})\|\|\text{eig}(\tilde{H}^2_{j,i})\|}, \qquad (18)$$

where $\|\cdot\|$ is a vectornorm. It measures the parallelity of the eigenvalue vectors of the homographies. For real valued eigenvalues, Eq. (18) provides the cosine of the angle between the two vectors. We have adapted their approach such that we exchange the first homography $\tilde{H}^1_{j,i}$ with the rotation matrix $R_{j,i}$ of the orientation sensor. The second homography can then be replaced by the infinite homography $\tilde{H}^\infty_{j,i}$. In this case we cannot sort the eigenvalues because the eigenvalues of a rotation matrix R are complex and a permutation of (13) and have absolute value 1. So it is not possible to sort the eigenvalues contained in the eigenvalue vector. Furthermore the eigenvalue vector is a function of Φ_i

$$E(\Phi_i) = \rho_{j,i}[1, \cos\Phi_i + i\sin\Phi_i, \cos\Phi_i - i\sin\Phi_i]^T. \qquad (19)$$

This known eigenvalue structure leads to a simpler and more robust matching criterion

$$\text{sim}_{\cos} = \left|\cos\Phi_i - \frac{\cos\Phi_j}{\rho_{j,i}}\right| \qquad (20)$$

The alignment in time between the estimated homograpy sequence and the rotation data can now be performed with this criterion. We now search for the time shift t_s which minimizes

$$t_s = \text{argmin}_{shift}\left\{\text{sim}_{\cos}(\tilde{H}^\infty_{j,i}, R_{t+shift})\right\}. \qquad (21)$$

5.2 Rotating camera with varying internal parameters

Our calibration approach normally has to deal with a camera with varying internal parameters. In this case it is not possible to match eigenvalues of the rotation matrix $R_{j,i}$ and the homography $\tilde{H}^\infty_{j,i}$ because the conjugation assumption is not valid for varying K_j. Therefore we have to calibrate the camera before the alignment between camera and sensor in time. We can use image-based selfcalibration like the approach of [7] but we only need short sequences for this task. In contrast to a camera with fixed but unknown calibration we have the calibration information and we know the accurate rotation. Therefore we don't have to deal with scales for the homographies. For these reasons (20) is equal to

$$\text{sim}_{\text{rot}}(R_i, R_j) = |\cos\Phi_i - \cos\Phi_j|. \qquad (22)$$

In the framework of calibrated cameras we are able to use the information about the rotation axis, too. The difference Δ_{axis} of the normalized rotation axis is

$$\Delta_{i,j,\text{axis}} = \frac{\text{eigvec}(R_i)}{\|\text{eigvec}(R_i)\|} - \frac{\text{eigvec}(R_j)}{\|\text{eigvec}(R_j)\|}. \qquad (23)$$

In order to use (22) and (23) simultaneously as a similarity measure we need to scale each measure to unit range $[0,1]$. Therefore the combined similarity measure is

$$\text{sim}_{\text{rot,axis}} = \frac{\Delta_{i,j,axis}}{3} + \frac{\text{sim}_{\text{rot}}(R_i, R_j)}{2}. \qquad (24)$$

This criterion can also be used with cameras with fixed but unknown calibration.

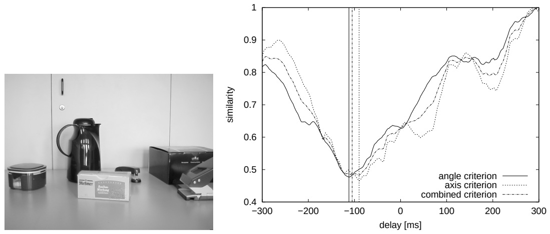


Figure 3: left:Images from the sequence to test alignment. right: Alignment of orientation sensor data with camera data using the different criterions angle criterion (22), axis criterion (23) and the combined criterion (24).

Alignment experiments: We compared the three similarity measures (22), (23) and (24) w.r.t. robustness. It can be seen that all criteria improve with sequence length. The best results were achieved with the combined criterion and the angle criterion, while the axis criterion showed more sensitivity to noise. A real sequence of 57 images was taken and aligned with an InterSense InterTrax2 consumer rotation sensor. This sequence is taken by a PAL camera mounted on a tripod which rotates about Pan-axis and Tilt-axis. A frame of the sequence is shown in figure 3. The orientation sensor provides new orientation data for every 8ms. The different criteria (22), (23) and (24) were used to estimate the time shift t_s. The similarity curves for the different criteria are plotted in figure 3 with minima in the range from -90ms to -112ms.

6. Experiments

6.1. Calibration of rotating camera

We tested the calibration techniques for rotating cameras with a sequence taken by a consumer pan-tilt-zoom camera as used in video conferencing (Sony DV-31). The camera is panning and zooming during the sequence. A frame of the sequence is shown in figure 4. The camera rotation is taken from the camera control commands, which means that we used the angles which are sent to the camera. Therefore the rotation error depends on the positioning accuracy of the pan-tilt head which is in the range of below 0.5 degrees for each axis. As reference for the zoom we interactively estimated the focal length for the different zoom positions from a given calibration object beforehand as ground truth. The focal length of the camera varied between 875-1232 (in pixel). We also compensated the zoom-dependent radial distortion beforehand. This can be done for the different discrete zooming steps of the camera with the knowledge of the zoom step but without knowledge of the correct zoom in pixel.

The sequence was processed by tracking feature points with a KLT-tracker [22]. From these tracks we calculated the homographies for the sequence with RANSAC and

Figure 4: left: Image from the real scene, right: Image from the synthetic scene

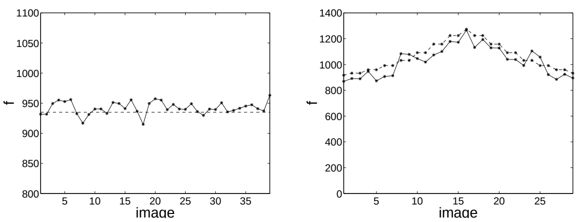


Figure 5: Calibration results for constant and varying focal length

least-squares-estimation over the inliers. The reprojection error gave a mean pixel error of 0.8 pixel. Calibration estimates for the focal length were computed from triples of images.

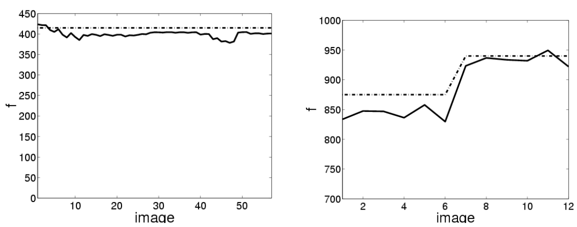
Figure 5 shows results for focal length estimation. The dashed line gives the true values, the solid line the estimated values. The left chart shows the estimated focal length (in pixel) for constant focal length f_{true} =940 pixel, the right chart contains a zooming camera. The average relative estimation error is around 3% for fixed zoom and 7% for changing zoom.

6.2. Calibration from general motion

We tested the calibration of a moving and rotating camera by using images rendered from a photorealistic image renderer. A tilting and panning camera is moving sideways in front of a VRML-scene created from realistic 3D models of buildings (see figure 4). The focal length of the camera was fixed to 415 (in pixel). We tracked features over the image sequence with the KLT-tracker [22]. From these features we estimated Fundamental matrices for different image pairs. The rotation is the given rotation of the ground truth data. The linear estimated focal length has a mean relative error of 4% w.r.t. the true focal length (see figure 6 (a)).

We also tested the calibration technique for a real, moving, and rotating camera. The test sequence was taken by the above mentioned pan-tilt-zoom camera. For the first test sequence the camera only rotates and moves during the sequence (see figure 4). As reference for the zoom we use the manual calibration of the different discrete zoom steps of the camera. The focal length of the camera for the first Sequence is 875 (in pixel). We also compensated the zoom-dependent radial distortion beforehand. The resulting relative mean error is about 5% for the linear calibration.

Our second test sequence also used the above video conferencing camera. During the sequence the camera is panning, tilting, zooming and moving. A frame of the sequence is shown in figure 4. The focal length in pixel varied between 875 and 940. We also compensated the radial distortion beforehand. The calibration results are shown in figure 6 (b). The resulting error is about 2.5% for the linear calibration.



(a) (b)

Figure 6: (a) calibration for synthetic sequence, (b): estimated focal length for varying intrinsics of real sequence. (dash-dotted: ground truth, solid: estimated values)

7. Conclusions

In this contribution we have proposed a novel selfcalibration approach that exploits knowledge of external rotation information in conjunction with image-based estimation of homographies and Fundamental matrices. The joint calibration from Fundamental matrices and rotation data guarantees metric projection matrices and avoids the problem of projective skew.

Rotation information can be found "for free" in a wide variety of applications and has proven valuable in this context. We have investigated the constraints that can be used to stabilize calibration and have evaluated the robustness of the approach with controlled and synthetic data. First real measurements have shown the viability of the method. We will continue to integrate the approach in a complete structure-from-motion system. We expect that rotation information will not only lead to better calibration but also to faster and more reliable image feature tracking since we can compensate the rotation component. We expect this to be a major step towards uncalibrated realtime tracking in unstructured outdoor environments.

References

[1] R. Franklin, "Efficient Rotation of an Object", IEEE Transactions on Computing, 1983.

[2] R. Hartley, " In defence of the 8-Point-Algorithm", *Proceedings ICCV95, IEEE Press, Cambridge, MA, USA, 1995.*

[3] J.-M. Frahm and R. Koch, "Robust Camera Calibration from Images and Rotation Data" *In Proc. DAGM 2003, Germany*

[4] R. Hartley and A. Zisserman, "Multiple View Geometry in Computer Vision" *Cambridge university press, 2000*

[5] Yaron Caspi and Michal Irani, "Alignment of Non-Overlapping Sequences" *ICCV 2001*

[6] H. Shum and R.Szeliski, "Panoramic Image Mosaics" *Microsoft Research*, Technical Report MSR-TR-97-23, 1997.

[7] L. de Agapito, E. Hayman, and I. Reid, "Self-Calibration of Rotating and Zooming Cameras" *IJCV, Vol. 45(2) Nov 2001*

[8] H. Sawhney, et. al., "Robust Video Mosaicing through Topology Inference and Local to Global Alignment" *ECCV*, 1998.

[9] C. Rother and S. Carlsson,"Linear multi view reconstruction and camera recovery using a reference plane", IJCV 49(2/3).

[10] C. E. Pearson, *Handbook of Applied Mathematics*, S.898, Second Edition, Van Nostrand Reinhold Company, 1983.

[11] B. Triggs,"Autocalibration and the Absolute Quadric", Proceedings Conference on Computer Vision and Pattern Recognition, pp. 609-614, Puerto Rico, USA, June 1997.

[12] F. Du and M. Brady, "Self-calibration of the intrinsic parameters of cameras for active vision systems" *CVPR*, 1993.

[13] R.Y.Tsai and R. K. Lenz, "A new technique for full autonomous and efficient 3d hand/eye calibration" *IEEE Journal of Robotics and Automation*, 5(3):345-358, June 1989 .

[14] O. D. Faugeras and M. Herbert, "The representation, recognition and locating of 3-D objects," *Intl. J. of Robotics Research*, 1992.

[15] G. Stein, "Accurate internal camera calibration using rotation, with analysis of sources of error," *ICCV*, 1995.

[16] R. I. Hartley, "Self-calibration from multiple views with a rotating camera" *ECCV*, 1994.

[17] S.J. Maybank and O. Faugeras, "A therory of self-calibration of a moving camera," *Int. J. of Computer Vision*, 1992.

[18] M. Pollefeys, R. Koch and L. Van Gool, "Selfcalibration and metric reconstruction in spite of varying and unknown internal camera parameters," *ICCV*, 1998.

[19] L. Naimark, E. Foxlin, "Circular Data Matrix Fiducial System and Robust Image Processing for a Wearable Vision-Inertial Self-Tracker". ISMAR'02, Germany.

[20] A. Heyden and K. Aström, "Euclidian Reconstruction from constant intrinsic parameters," *Intl. Conf. PR*, 1996.

[21] A. Heyden and K. Aström, "Euclidian Reconstruction from image sequences with varying and unkwon focal length and principal point," *CVPR*, 1997.

[22] Bruce D. Lucas and Takeo Kanade, " An Iterative Image Registration Technique with an Application to Stereo Vision," *Int. Joint Conference on Artificial Intelligence*, 1981.