Pose Estimation for Multi-Camera Systems

Jan-Michael Frahm, Kevin Köser and Reinhard Koch

Institute of Computer Science and Applied Mathematics
Hermann-Rodewald-Str. 3,
24098 Kiel, Germany
{jmf,koeser,rk}@mip.informatik.uni-kiel.de

Abstract. We propose an approach for pose estimation based on a multi-camera system with known internal camera parameters. We only assume for the multi-camera system that the cameras of the system have fixed orientations and translations between each other. In contrast to existing approaches for reconstruction from multi-camera systems we introduce a rigid motion estimation for the multi-camera system itself using all information of all cameras simultaneously even in the case of non-overlapping views of the cameras. Furthermore we introduce a technique to estimate the pose parameters of the multi-camera system automatically.

1 Introduction

Robust scene reconstruction and camera pose estimation is still an active research topic. During the last twelve years many algorithms have been developed, initially for scene reconstruction from a freely moving camera with fixed calibration [2] and later even for scene reconstruction from freely moving uncalibrated cameras [3]. All these approaches are using different self-calibration methods, which have been developed in the last decade, to estimate the internal calibration of the camera. This self-calibration can be used to estimate the internal parameters of multi-camera systems (MCS).

However, all these methods still suffer from ill-conditioned pose estimation problems which cause flat minima in translation and rotation error functions [4]. Furthermore the relatively small viewing angle is also a problem which influences the accuracy of the estimation [4]. Due to these problems we introduce a new pose estimation technique which combines the information of several rigidly coupled cameras to avoid the ambiguities which occur in the single camera case. In our novel approach we estimate a rigid body motion for the MCS as a whole. Our technique combines the observations of all cameras to estimate the six degrees of freedom (translation and orientation in 3D-space) for the pose of the MCS. It exploits the fixed rotations and translations between the cameras of the MCS. These fixed rotations and translations are denoted as a *configuration* in the following. We also give a technique to determine these parameters automatically from an image sequence of the MCS.

The paper is organized as follows. At first we discuss the previous work in pose estimation from a single camera or a MCS. Afterwards we introduce our novel

pose estimation approach. In section 4 we provide a technique to automatically estimate the configuration of the MCS. Furthermore in section 5 we show some experimental results to measure the robustness of our approach.

1.1 Notation

In this subsection we introduce some notations. The projection of scene points onto an image by a calibrated camera may be modeled by the equation $x = PX$. The image point in projective coordinates is $x = [x^x, x^y, x^w]^T$, while $X = [X^x, X^y, X^z, X^w]^T$ is the 3D-world point $[\frac{X^x}{X^w}, \frac{X^y}{X^w}, \frac{X^z}{X^w}]^T$ 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 $P = [R^T | -R^T C]$, the P-matrix is called metric, where the rotation matrix R (orientation of the camera) and the translation vector C (position of the camera) represent the Euclidian transformation between the camera coordinate system and the world coordinate system.

2 Previous work

For a single moving camera, Fermüller et. al. discussed in [4] the ambiguities for motion estimation in the three dimensional space. They proved that there were ambiguities in estimation of translation and rotation for one camera for all types of given estimation algorithms. These ambiguities result in flat minima of the cost functions. Baker et. al. introduced in [5] a technique to avoid these ambiguities when using a MCS. For each camera the pose is estimated separately and the ambiguities are calculated before the fusion of the ambiguous subspaces is used to compute a more robust pose of the cameras. In contrast to our approach the technique of [5] does not use one pose estimation for all information from all cameras simultaneously.

There is some work in the area of polydioptric cameras [7] which are in fact MCSs with usually very small translations between the camera centers. In [8] a hierarchy of cameras and their properties for 3D motion estimation is discussed. It can be seen that the pose estimation problem is well-conditioned for an MCS in contrast to the ill-conditioned problem for a single camera.

The calibration of a MCS is proposed in [5]. The line-based calibration approach is used to estimate the internal and external parameters of the MCS. For a MCS with zooming cameras a calibration approach is introduced in [9, 10]. An approach for an auto-calibration of a stereo camera system is given in [1]. Nevertheless, all standard calibration, pose-estimation and structure from motion approaches for stereo camera systems exploit the overlapping views of the cameras, what is in contrast to our pose estimation approach, which does not depend on this.

3 Pose estimation for multi-camera systems

In this section we introduce our novel approach for rigid motion estimation of the MCS. The only assumptions are that we have a MCS with an internal

calibration K_i for each of the cameras and a fixed configuration. That assumption is valid for most of the currently used MCSs because all these systems are mounted on some type of carrier with fixed mount points. The computation of the configuration from the image sequence itself is introduced in section 4. The internal camera calibration K_i can be determined using the techniques of [10, 3]. For convenience we will always talk about K-normalized image coordinates and P-matrices, therefore K_i can be omitted for pose estimation.

3.1 Relation between world and multi-camera system

The general structure from motion approach uses an arbitrary coordinate system $\mathcal{C}_{world}$ to describe the camera position by the rotation R of the camera, the position C of the camera center and the reconstructed scene. Normally the coordinate system $\mathcal{C}_{world}$ is equivalent with the coordinate system of the first camera. In this case the projection matrix of camera i with orientation R_i and translation C_i is given by

$$P_i = \left[R_i^T | - R_i^T C_i\right]. \tag{1}$$

For a multi camera-system we use two coordinate systems during the pose estimation. The absolute coordinate system $\mathcal{C}_{world}$ is used to describe the positions of 3D-points and the pose of the MCS in the world. The second coordinate system used, $\mathcal{C}_{rig}$, is the relative coordinate system of the MCS describing the relations between the cameras (configuration). It has its origin at C_v and it is rotated by R_v and scaled isotropically by λ_v with respect to $\mathcal{C}_{world}$.

Now we discuss the transformations between the different cameras of the MCS and the transformation into the world coordinate system $\mathcal{C}_{world}$. Without loss of generality we assume all the translations ΔC_i and rotations ΔR_i of the cameras are given in the coordinate system $\mathcal{C}_{rig}$. Then with (1) the camera projection matrix of each camera in $\mathcal{C}_{rig}$ is given by

$$P_i^{\mathcal{C}_{rig}} = \left[\Delta R_i^T | - \Delta R_i^T \Delta C_i\right]. \tag{2}$$

The position C_i of camera i and the orientation R_i in $\mathcal{C}_{world}$ is given by

$$C_i = C_v + \frac{1}{\lambda_v} R_v \Delta C_i, \qquad R_i = R_v \Delta R_i, \tag{3}$$

where translation C_v, orientation R_v and scale λ_v are the above described relations between the MCS coordinate system $\mathcal{C}_{rig}$ and the world coordinate system $\mathcal{C}_{world}$. Then the projection matrix of the camera i in $\mathcal{C}_{world}$ is given by

$$P_i = \left[\Delta R_i^T R_v^T | - \Delta R_i^T R_v^T (C_v + \frac{1}{\lambda_v} R_v \Delta C_i)\right]. \tag{4}$$

With (3) we are able to describe each camera's position in dependence of the position and orientation of camera i in the coordinate system of the multi-camera system $\mathcal{C}_{rig}$ and the pose of the MCS in the world $\mathcal{C}_{world}$. Furthermore with (4) we have the transformation of world points X into the image plane of camera i in dependence of the position and orientation of the MCS and the configuration of the MCS.

3.2 Virtual camera as a representation of a multi-camera system

We now introduce a *virtual camera* as a representation of the MCS, which is used to determine the position of the MCS in $\mathcal{C}_{world}$ independent of its configuration.

The virtual camera v which represents our MCS is at the origin of the coordinate system $\mathcal{C}_{rig}$ and is not rotated within this system. It follows immediately that it has position C_v and orientation R_v in $\mathcal{C}_{world}$ because it is rotated and translated in the same manner as the MCS. With (1) the projection matrix P_v of the virtual camera v is

$$P_v = \left[R_v^T | - R_v^T C_v\right], \tag{5}$$

where rotation R_v and position C_v are the above given rotation and position of the MCS. From (4) and (5) it follows that the projection matrix P_i of camera i depends on the virtual camera's projection matrix P_v :

$$P_i = \Delta R_i^T \left(P_v + \left[0_{3x3} | - \frac{1}{\lambda_v} \Delta C_i\right]\right). \tag{6}$$

3.3 Pose estimation of the virtual camera

Now we propose a pose estimation technique for the virtual camera using the observations of all cameras simultaneously. The image point x_i in camera i of a given 3D-point X is given as $x_i \cong P_i X$, where $x_i \in \mathbb{P}^2$, $X \in \mathbb{P}^3$ and $\cong$ is the equality up to scale. With equation (6) the image point x_i depends on the virtual camera's pose by

$$x_i \cong P_i X = \Delta R_i^T \left(P_v + \left[0_{3x3} | - \frac{1}{\lambda_v} \Delta C_i\right]\right) X, \tag{7}$$

For a MCS with known configuration, namely camera translations ΔC_i, camera orientations ΔR_i and scale λ_v, this can be used to estimate the virtual camera's position C_v and orientation R_v in dependence of the image point x_i in camera i as a projection of 3D-point X.

Now we deduce a formulation for the estimation of the virtual camera's position C_v and orientation R_v given the translations ΔC_i, orientations ΔR_i, and scale λ_v of the cameras of the MCS. From (7) we get

$$\underbrace{\Delta R_i x_i}_{\tilde{x}_i} \cong \underbrace{P_v X - \frac{X_w}{\lambda_v} \Delta C_i}_{\hat{x}},$$

where $X = [X^x, X^y, X^z, X^w]^T \in \mathbb{P}^3$ is the 3D-point in the 3D projective space. Using the same affine space for $\tilde{x}_i$ and $\hat{x}$ leads to the following linear equations

$$\begin{aligned} &X^x \tilde{x}_i^x (P_v)_{3,1} + X^y \tilde{x}_i^x (P_v)_{3,2} + X^z \tilde{x}_i^x (P_v)_{3,3} + X^w \tilde{x}_i^x (P_v)_{3,4} \\ &- \left(X^x \tilde{x}_i^w (P_v)_{1,1} + X^y \tilde{x}_i^w (P_v)_{1,2} + X^z \tilde{x}_i^w (P_v)_{1,3} + X^w \tilde{x}_i^w (P_v)_{1,4}\right) \end{aligned}$$

$$= (\Delta\tilde{C}_i)_3 X^w \tilde{x}_i^x - (\Delta\tilde{C}_i)_1 X^w \tilde{x}_i^w, \tag{8}$$
$$X^x \tilde{x}_i^y (P_v)_{3,1} + X^y \tilde{x}_i^y (P_v)_{3,2} + X^z \tilde{x}_i^y (P_v)_{3,3} + X^w \tilde{x}_i^y (P_v)_{3,4}$$
$$- (X^x \tilde{x}_i^w (P_v)_{2,1} + X^y \tilde{x}_i^w (P_v)_{2,2} + X^z \tilde{x}_i^w (P_v)_{2,3} + X^w \tilde{x}_i^w (P_v)_{2,4})$$
$$= (\Delta\tilde{C}_i)_3 X^w \tilde{x}_i^y - (\Delta\tilde{C}_i)_2 X^w \tilde{x}_i^w \tag{9}$$

in the entries of P_v with $\tilde{x}_i = [\tilde{x}_i^x, \tilde{x}_i^y, \tilde{x}_i^w]^T$ and $\Delta\tilde{C}_i = \frac{1}{\lambda_v}\Delta C_i$.

Note that the above equations are a generalization of the case of a single camera which can be found in [1] and analogous methods to those given in [1] can be used to estimate P_v from these equations and to finally extract the unknown orientation R_v and the unknown position C_v. The extension for the MCS is that the rotation compensated image points $\tilde{x}_i$ are used and terms for the translation ΔC_i of camera i in the multi-camera coordinate system $\mathcal{C}_{rig}$ are added. In the case of pose estimation for a single camera using our approach it is assumed without loss of generality that the coordinate system $\mathcal{C}_{rig}$ is equivalent to the camera's coordinate system. Then ΔC_i vanishes and the rotation ΔR_i is the identity. In this case (8) and (9) are the standard (homogeneous) pose estimation equations from [1].

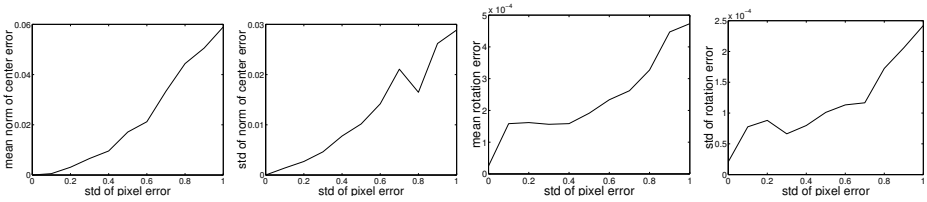
4 Calibration of the multi-camera system

In the previous section we always assumed that we know the orientation ΔR_i and translation ΔC_i of each camera in the coordinate system $\mathcal{C}_{rig}$ and the scale λ_v between $\mathcal{C}_{rig}$ and $\mathcal{C}_{world}$. In this section we present a technique to estimate these parameters from the image sequence of a MCS with overlapping views. However, note that the simultaneous pose estimation of the MCS itself does not depend on overlapping views, once the configuration is known.

Suppose we are given n cameras in the MCS and grab images at time t_0. After a motion of the MCS (time t_1), we capture the next image of each camera. We now have $2n$ frames with overlapping views, for which a standard structure-from-motion approach (for example as described in [6]) for single cameras can be applied to obtain their positions and orientations.

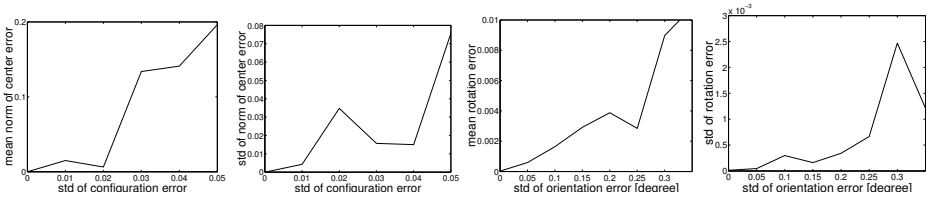
For each of the two groups of n cameras (the MCS at t_0 and t_1) the virtual camera is set to the first camera of the system. Then the rigid transformations for the other cameras are computed and averaged, which yields an initial approximate configuration of the system. In order to obtain a mean rotation we use the axis-angle representation, where axes and angles are averaged arithmetically with respect to their symmetries. If $\mathcal{C}_{world}$ is defined to be the coordinate system of the estimated single cameras, it follows immediately that λ_v has to be set to 1 since $\mathcal{C}_{rig}$ already has the correct scale.

To improve precision the estimate of the configuration is iteratively refined: For each new pose of the system the pose of each single camera is revised with respect to the points seen by that camera. Afterwards the configuration of the refined cameras is computed and averaged with the previously estimated configurations. Since the combined camera system pose estimation is somewhat sensitive to noise in the configuration parameters, this is more robust.



(a) (b) (c) (d)

Fig. 1. Dependency of the standard deviation of the feature position noise in pixel (a) the mean of the norm of camera center error, (b) the standard deviation of the latter error, (c) the absolute value of the angular error of the cameras orientation, (d) the standard deviation of the latter error.



(a) (b) (c) (d)

Fig. 2. Dependency of the standard deviation of the noise in the MCS configuration (a) the mean of the norm of the camera center error, (b) the standard deviation of the norm camera center error, (c) the absolute value of the angular error of the cameras orientation, (d) the standard deviation of the angular error of the cameras orientation.

5 Experiments

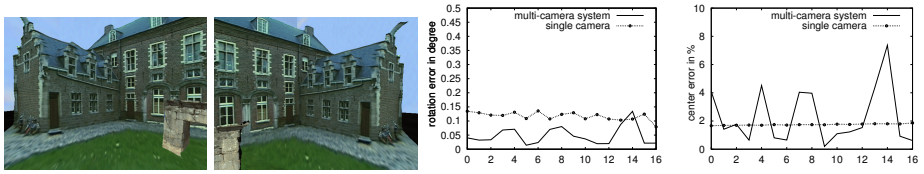
In this section the introduced estimation techniques for the pose of a MCS are evaluated. First we measure the robustness of the technique with synthetic data. Afterwards we use image sequences generated by a simulator and compare our results with the given ground truth data. Finally we also present experimental results for a real image sequence.

To measure the noise robustness of our novel pose estimation technique we use synthetic data. The MCS is placed in front of a scene consisting of 3D-points with given 2D image points in the cameras of the MCS. At first we disturb the 2D correspondences with zero-mean Gaussian noise for each image. Afterwards we use our approach to estimate the pose of the virtual camera, with a least squares solution based on all observed image points. The norm of the position error and the angle error of the estimated orientation can be seen in figure (1). It can be seen that the proposed pose estimation is robust with respect to the pixel location error of up to 1 pixel noise.

In a second test we disturb the configuration of the MCS with a zero-mean Gaussian translation error (with sigma of up to 5% of the camera's original displacement) and a Gaussian rotation error of up to 0.35 degrees in each axis. It can be seen that the proposed pose estimation technique is robust against these

disturbances but the configuration errors cause higher errors in the estimated pose than the noise in the feature positions does.

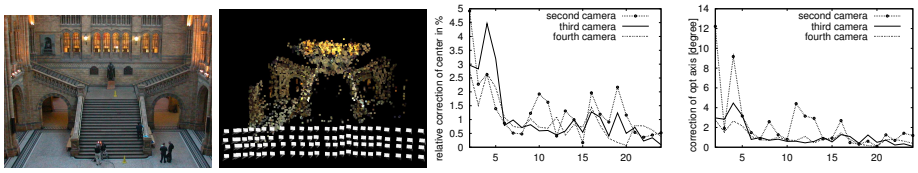
In order to measure the pose estimation errors of the proposed approach in a structure-from-motion framework, we use a sequence of rendered images (see fig. 3) with ground truth pose data. In this sequence a MCS with two fixed cameras with non-overlapping views is moved and rotated in front of a synthetic scene. We implemented a pose estimation algorithm with the following steps: Given a set of Harris corners and corresponding 3d points in an initial image 1.) in each image a Harris corner detector is used to get feature positions, 2.) from the corners a set of correspondences is estimated using normalized cross correlation and epipolar geometry, 3.) using these correspondences (and the referring 3d points) the pose is estimated with RANSAC using eq. (8) and (9), 4.) afterwards a nonlinear optimization is used to finally determine the pose of the MCS. The measured position and orientation errors are shown and compared to a single camera pose estimation in fig. 3. It can be seen that using the MCS pose estimation the rotation is estimated with a smaller error than in the single camera case, but the translation estimatates for a single camera is slightly better for this data.



(a) (b) (c) (d)

Fig. 3. (a),(b): non-overlapping simulator images of MCS (c),(d): error of relative translation and rotation since previous estimate w.r.t. to ground truth (17 image pairs) for standard structure from motion and MCS structure from motion.

Now we show that the pose estimation also works well on real images. The images used have been taken at the National History Museum in London using a MCS with four cameras mounted on a pole. The configuration has been computed from the image data as described in the previous section. Using standard single-camera structure-from-motion approaches, the pose estimation breaks down in front of the stairs. Due to the missing horizontal structure at the stairs there are nearly no good features. However, incorporating all cameras in our approach makes the pose estimation robust exactly in those situations, where some of the cameras can still see some features. Using our approach to compute the MCS configuration the initial estimates for the centers are refined by about five to eight percent in $\mathcal{C}_{rig}$ compared to the finally stable values. After about the seventh pose estimate the center change rate reaches one percent. It is interesting that although the parameters for the second camera are not estimated very well, the system does work robustly as a whole.



(a) (b) (c) (d)

Fig. 4. Museum scene: (a) overview image, (b) reconstructed scene points and cameras, (c) relative corrections of centers in $\mathcal{C}_{rig}$, (d) incremental optical axes rotations. The sequence starts in front of the arc to the left, moves parallel to some wide stairs and finishes in front of the other arc to the right. 25 times 4 images have been taken.

6 Conclusions

We introduced a novel approach for pose estimation of a multi-camera system even in the case of non-overlapping views of the cameras. Furthermore we introduced a technique to estimate all parameters of the system directly from the image sequence itself. The new approach was tested under noisy conditions and it has been seen that it is robust. Finally we have shown results for real and synthetic image sequences.

Acknowledgement This work has been partially funded by the European Union (Project MATRIS, IST-002013).

References

1. R. Hartley and A. Zisserman, "Multiple View Geometry in Computer Vision" *Cambridge university press, Cambrige, 2000*
2. S.J. Maybank and O. Faugeras, "A therory of self-calibration of a moving camera," *International Journal of Computer Vision*, 1992.
3. M. Pollefeys, R. Koch and L. Van Gool, "Selfcalibration and metric reconstruction in spite of varying and unknown internal camera parameters", *ICCV*, 1998.
4. Cornelia Fermüller and Yiannis Aloimonos "Observability of 3d motion" *International Journal of Computer Vision*, 37(1):43-62, June 2000
5. P. Baker, C. Fermüller, Y. Aloimonos and R. Pless, "A Spherical Eye from Multiple Cameras (Makes Better Models of the World)" *CVPR'01*, Volume 1, 2001
6. P. A. Beardsley, A. Zisserman and D. W. Murray "Sequential Updating of Projective and Affine Structure from Motion" *IJCV*, Volume 23 , Issue 3, 1997
7. Jan Neumann, Cornelia Fermüller, and Yiannis Aloimonos "Polydioptric Camera Design and 3D Motion Estimation" *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Volume 2, pages 294-301, 2003
8. J. Neumann, C. Fermüller, and Y. Aloimonos "Eye Design in the Plenoptic Space of Light Rays" *9th IEEE Int. Conference on Computer Vision*, 2003.
9. Jan-M. Frahm and Reinhard Koch, "Camera Calibration with Known Rotation" *Ninth IEEE International Conference on Computer Vision*, Vol. 2, October 2003.
10. A. Zomet et al., "Omni-rig: Linear Self-recalibration of a Rig with Varying Internal and External Parameters", *8th Int. Conf. on Computer Vision*, 2001