Active Self-calibration of Multi-camera Systems

Marcel Brückner and Joachim Denzler

Chair for Computer Vision
Friedrich Schiller University of Jena
{marcel.brueckner, joachim.denzler}@uni-jena.de

Abstract. We present a method for actively calibrating a multi-camera system consisting of pan-tilt zoom cameras. After a coarse initial calibration, we determine the probability of each relative pose using a probability distribution based on the camera images. The relative poses are optimized by rotating and zooming each camera pair in a way that significantly simplifies the problem of extracting correct point correspondences. In a final step we use active camera control, the optimized relative poses, and their probabilities to calibrate the complete multi-camera system with a minimal number of relative poses. During this process we estimate the translation scales in a camera triangle using only two of the three relative poses and no point correspondences. Quantitative experiments on real data outline the robustness and accuracy of our approach.

1 Introduction

In the recent years multi-camera systems became increasingly important in computer vision. Many applications take advantage of multiple cameras observing a scene. Multi-camera systems become even more powerful if they consist of *active* cameras, i. e. pan-tilt zoom cameras (Fig. 1). For many applications, however, the (active) multi-camera system needs to be calibrated, i. e. the intrinsic and extrinsic parameters of the cameras have to be determined. Intrinsic parameters of a camera can be estimated using a calibration pattern [1] or camera self-calibration methods for a rotating camera [2, 3]. The focus of this paper is on (active) extrinsic calibration which consists of estimating the rotation and translation of each camera relative to some common world coordinate system.

Classical methods for extrinsic multi-camera calibration need a special calibration pattern [1] or user interaction like a moving LED in a dark room [4, 5]. From a practical point of view, however, a pure self-calibration is most appealing. Self-calibration in this context means that no artificial landmarks or user interaction are necessary. The cameras estimate their position only from the images they record. An example for self-calibration of a *static* multi-camera system is the work of Läbe and Förstner [6]. Given several images they extract point correspondences and use these to estimate the relative poses. Another example is the graph based calibration method proposed by Bajramovic and Denzler [7] which considers the uncertainty of the estimated relative pose of each camera pair. However, both methods are designed for static cameras and do not use the benefits of active camera control.

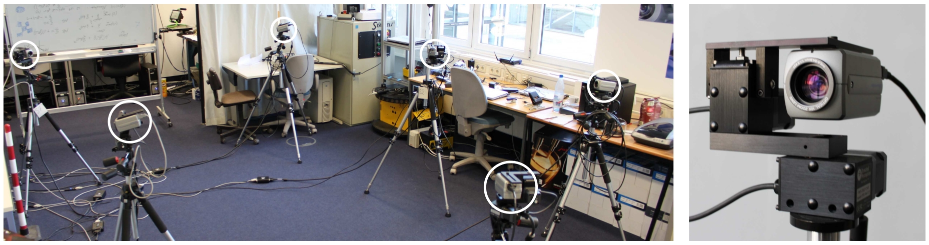

Fig. 1. A multi-camera system (left) consisting of six pan-tilt zoom cameras (white circles). The cameras are mounted near the intersection of the pan and tilt axes (right).

Sinha and Pollefeys [8] suggest a method where each pan-tilt zoom camera builds a high resolution panorama image. These images are used for relative pose estimation. However, these huge images can contain many ambiguities which affect the extraction of correct point correspondences. The calibration method of Chippendale and Tobia [9] defines an observer camera which searches for the continuously moving other cameras. If the observer spots some other camera the relative pose between the two cameras is extracted by detecting the circle shape of the camera lens and tracking some special predefined camera movements. The applicability and accuracy of this method highly depends on the distance between the cameras.

One of the biggest problems in extrinsic camera calibration is extracting *correct* point correspondences between the camera pairs. This problem is called wide baseline stereo and several approaches can be found in the literature [10, 11]. However, if the cameras have very different viewpoints on a scene, projective influences and occlusions complicate or make it impossible to establish correct point correspondences. Active cameras could use rotation and zoom to reduce these projective influences.

In this paper, we present a method which uses active camera control to calibrate a multi-camera system consisting of pan-tilt zoom cameras. After an initial coarse calibration which uses the common field of view detection of Brückner et al. [12] to reduce ambiguities in the point correspondence detection, the best relative pose for each camera pair is selected based on its probability. Hence we present an image based probability distribution for relative poses. Given the initial poses, each camera pair rotates and zooms in a way that the points of view of the two cameras are very similar. The resulting similar camera images significantly simplify the problem of establishing new point correspondences which are used to reestimate the relative poses. In a final step we use the relative poses and their probabilities to calibrate the complete multi-camera system from a minimal set of relative poses. In order to estimate the scale factors of the relative poses in a camera triangle, we use only two of the three relative poses and we do not need any triple point correspondences. Instead we use active camera control and our image based probability distribution for relative poses. This reduces the number of relative poses needed for the complete calibration and totally avoids outlier point correspondences. The remainder of this paper is organized as follows: in Section 2 we introduce some basics and notation. Our method is described in Section 3. In Section 4 we present and discuss our experiments. Conclusions are given in Section 5.

2 Basics

2.1 Camera Model and Relative Pose between Cameras

A world point $\boldsymbol{X}_w$ is projected to the image point $\boldsymbol{x} \overset{\text{def}}{=} \boldsymbol{K}\boldsymbol{R}_{ptu}\left(\boldsymbol{R}_i\boldsymbol{X}_w + \boldsymbol{t}_i\right)$, where $\boldsymbol{R}_i, \boldsymbol{t}_i$ are the extrinsic camera parameters (rotation and translation), $\boldsymbol{K}$ is the pinhole matrix [13] and $\boldsymbol{R}_{ptu}$ is the rotation of the pan-tilt unit. We assume the pan and tilt axes to be identical to the Y and X axes of the camera coordinate system, respectively. Throughout the paper we use image points which are normalized with respect to the camera and pan-tilt rotation $\tilde{\boldsymbol{x}} \overset{\text{def}}{=} \boldsymbol{R}_{ptu}^{-1}\boldsymbol{K}^{-1}\boldsymbol{x}$. From this point on, when talking about the camera orientation and position we actually mean $\boldsymbol{R}_i, \boldsymbol{t}_i$ with no pan-tilt rotation $\boldsymbol{R}_{ptu} = \boldsymbol{I}$. The relative pose between two cameras i and j is defined as $\boldsymbol{R}_{i,j} \overset{\text{def}}{=} \boldsymbol{R}_j\boldsymbol{R}_i^{-1}$ and $\boldsymbol{t}_{i,j} \overset{\text{def}}{=} \boldsymbol{t}_j - \boldsymbol{R}_j\boldsymbol{R}_i^{-1}\boldsymbol{t}_i$.

2.2 Common Field of View Detection

Common field of view detection consists of deciding which image pairs show a common part of the world. We will briefly describe the probabilistic method of Brückner et al. [12] which gave the best results in their experiments.

Given two camera images, the difference of Gaussian detector [11] is used to detect interest points $\mathcal{C}_i = \{\tilde{\boldsymbol{x}}_1, \ldots, \tilde{\boldsymbol{x}}_n\}$ and $\mathcal{C}_j = \{\tilde{\boldsymbol{x}}'_1, \ldots, \tilde{\boldsymbol{x}}'_{n'}\}$. For each point $\tilde{\boldsymbol{x}}_i$, the SIFT descriptor $\mathbf{des}(\tilde{\boldsymbol{x}}_i)$ is computed [11]. These descriptors are used to construct a conditional correspondence probability distribution for each $\tilde{\boldsymbol{x}}_i$

$$p\left(\tilde{\boldsymbol{x}}'_j \mid \tilde{\boldsymbol{x}}_i\right) \propto \exp\left(-\frac{d_d^{i,j} - d_N(\tilde{\boldsymbol{x}}_i)}{\lambda_d\, d_N(\tilde{\boldsymbol{x}}_i)}\right) , \qquad (1)$$

where λ_d is the inverse scale parameter of the exponential distribution, $d_d^{i,j} = \text{dist}(\mathbf{des}(\tilde{\boldsymbol{x}}_i), \mathbf{des}(\tilde{\boldsymbol{x}}'_j))$ is the Euclidean distance between the descriptors of the points $\tilde{\boldsymbol{x}}_i$ and $\tilde{\boldsymbol{x}}'_j$, and $d_N(\tilde{\boldsymbol{x}}_i) = \min_j(d_d^{i,j})$ denotes the distance of the nearest neighbor of the point $\tilde{\boldsymbol{x}}_i$. Each of the resulting conditional probability distributions $p(\tilde{\boldsymbol{x}}'_j \mid \tilde{\boldsymbol{x}}_i)$ has to be normalized such that $\sum_{\tilde{\boldsymbol{x}}'_j \in \mathcal{C}_j} p(\tilde{\boldsymbol{x}}'_j \mid \tilde{\boldsymbol{x}}_i) = 1$ holds.

The conditional probability distributions are used to calculate the normalized joint entropy which is defined as

$$H(\mathcal{C}_i, \mathcal{C}_j) \overset{\text{def}}{=} -\frac{1}{\eta} \sum_{\tilde{\boldsymbol{x}}_i \in \mathcal{C}_i} \sum_{\tilde{\boldsymbol{x}}'_j \in \mathcal{C}_j} p(\tilde{\boldsymbol{x}}_i) p\left(\tilde{\boldsymbol{x}}'_j \mid \tilde{\boldsymbol{x}}_i\right) \log\left(p(\tilde{\boldsymbol{x}}_i) p\left(\tilde{\boldsymbol{x}}'_j \mid \tilde{\boldsymbol{x}}_i\right)\right) , \qquad (2)$$

where $\eta = \log(nn')$ is the maximum joint entropy and $p(\tilde{\boldsymbol{x}}_i)$ is a uniform distribution if no prior information about the interest points is available. A low joint entropy $H(\mathcal{C}_i, \mathcal{C}_j)$ indicates similar images. For further details the reader is referred to [12].

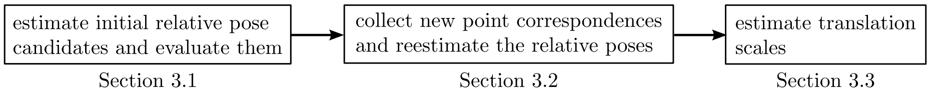
estimate initial relative pose candidates and evaluate them
collect new point correspondences and reestimate the relative poses
estimate translation scales
Section 3.1
Section 3.2
Section 3.3

Fig. 2. The three steps of our multi-camera calibration method. Each step is described in the indicated Section.

3 Active Multi-camera Calibration

We calibrate an active multi-camera system consisting of c pan-tilt zoom cameras. For each camera the intrinsic parameters for different zoom steps are assumed to be known. Our calibration method consists of three steps which are illustrated in Fig. 2: an initial relative pose estimation with an evaluation of the relative poses, an optimization of these relative poses and a final estimation of the translation scale factors. Each step uses active camera control in a different way and to a different extent.

3.1 Initial Relative Pose Estimation and Evaluation

Given the intrinsic parameters, each camera records as many images as necessary to cover its complete environment. Now each camera pair searches for image pairs sharing a common field of view (Section 2.2). This search can be viewed as a prematching of point correspondences which considers the local environment of each interest point. Hence, it decreases the chance of ambiguities disturbing the point matching process. Between each of these image pairs point correspondences are extracted using the difference of Gaussian detector, the SIFT descriptor, the Euclidean distance, and the two nearest neighbors matching with rejection as proposed by Lowe [11].

Based on all extracted point correspondences of a camera pair we estimate the relative pose using the five point algorithm [14]. An important point is that the translation of these relative poses can only be estimated up to an unknown scale factor. For the complete calibration of a multi-camera system consistent scale factors for all translations have to be estimated.

In order to increase the robustness against outliers we embed the five point algorithm into a RANSAC scheme [15]. As distance measure we use the closest distance between two viewing rays

$$d_e^{i,j}\left(\tilde{\boldsymbol{x}}_i,\tilde{\boldsymbol{x}}'_j\right) \overset{\text{def}}{=} \min_{\lambda_i,\lambda_j}\left\|\left(\lambda_i \boldsymbol{R}_{i,j}\tilde{\boldsymbol{x}}_i + \frac{\boldsymbol{t}_{i,j}}{\|\boldsymbol{t}_{i,j}\|_2}\right) - \lambda_j\tilde{\boldsymbol{x}}'_j\right\|_2 \quad \text{with} \quad \lambda_i,\lambda_j > 0\ . \qquad (3)$$

Since we normalize the translation to unit length, it is possible to define the inlier threshold relative to the camera distance. The scale factors λ_i and λ_j need to be positive which affects the direction of the viewing rays and is similar to the constraint of 3D points to lie in front of both cameras.

Instead of selecting a single best pose, we select the m_p best poses based on the number of inliers. Since most of these poses are quite similar, we additionally constrain the selection to take only relative poses that satisfy a minimum rotation difference θ_R and translation difference θ_t to the already selected relative poses.

Now, each camera pair i, j performs the following procedure for each of its m_p relative pose candidates. First, the two cameras are rotated in a way that they look into the same direction and their optical axes are aligned (or a setup as close as possible to this). Camera i has to look in the direction $-\boldsymbol{R}_{i,j}\boldsymbol{t}_{i,j}$ and camera j looks at $-\boldsymbol{t}_{i,j}$. From each of the resulting camera images interest points are extracted. Now, the cameras repeat the first step, but in the opposite direction. The result of this procedure is a set of interest points $\mathcal{C}_i$ and $\mathcal{C}_j$ for each of the two cameras i and j. Given these interest point sets we want to evaluate the relative pose candidate. Therefore we calculate the probability

$$p\left(\boldsymbol{R}_{i,j}, \boldsymbol{t}_{i,j}\right) \propto \sum_{\tilde{\boldsymbol{x}}_i \in \mathcal{C}_i} \sum_{\tilde{\boldsymbol{x}}'_j \in \mathcal{C}_j} p\left(\boldsymbol{R}_{i,j}, \boldsymbol{t}_{i,j} \mid \tilde{\boldsymbol{x}}'_j, \tilde{\boldsymbol{x}}_i\right) p\left(\tilde{\boldsymbol{x}}'_j \mid \tilde{\boldsymbol{x}}_i\right) p\left(\tilde{\boldsymbol{x}}_i\right) \ , \tag{4}$$

where $p\left(\boldsymbol{R}_{i,j}, \boldsymbol{t}_{i,j} \mid \tilde{\boldsymbol{x}}'_j, \tilde{\boldsymbol{x}}_i\right) \overset{\text{def}}{=} \exp\left(-d_e^{i,j}\left(\tilde{\boldsymbol{x}}_i, \tilde{\boldsymbol{x}}'_j\right)/\lambda_e\right)$ is an exponential distribution using the distance measure of (3) and the inverse scale parameter λ_e, $p\left(\tilde{\boldsymbol{x}}'_j \mid \tilde{\boldsymbol{x}}_i\right)$ is the conditional correspondence probability of (1) and $p\left(\tilde{\boldsymbol{x}}_i\right)$ is a uniform distribution if no prior information about the interest points is available. We note that this probability distribution can also be viewed as an image similarity measure which is based on image and geometric information. For each camera pair the relative pose candidate with the highest probability is selected.

3.2 Actively Optimizing the Relative Poses

Given the initial relative poses $\boldsymbol{R}_{i,j}, \boldsymbol{t}_{i,j}$ we optimize these poses by steering each camera pair in a way that it can easily establish new point correspondences.

As mentioned in Section 1, the biggest problem in finding correct point correspondences are projective influences. These influences depend on the relation between camera distance and scene distance and the difference in the viewing directions between the cameras. To reduce these influences we first rotate the two cameras in a way that their optical axes are aligned as described in Section 3.1. Additionally we search for the zoom step z of the backmost camera i with the highest image similarity by

$$\operatorname*{argmin}_z H\left(\mathcal{C}_i\left(z\right), \mathcal{C}_j\right) \ , \tag{5}$$

where $\mathcal{C}_i\left(z\right)$ is the interest point set of camera i at zoom step z and $H\left(\mathcal{C}_i, \mathcal{C}_j\right)$ is the normalized joint entropy (2). Again, this procedure is repeated for the opposite direction and yields in an interest point set for each camera. Similar to the initial calibration we extract point correspondences and use these to estimate the relative pose. Since we expect the descriptors of two corresponding points to be very similar due to the high similarity of the camera images, we choose a stricter rejection threshold for the two nearest neighbors matching than in the initial calibration. The estimated relative poses are evaluated as described in Section 3.1. For each camera pair the reestimated relative pose will only be used if it has a higher probability than the initial relative pose.

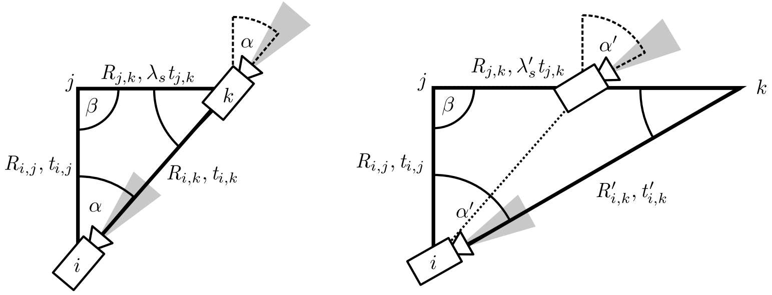

Fig. 3. A camera triangle (i, j, k). Cameras i and k rotate with angle α around the plane normal of the camera triangle. The scale λ_s depends on the angle α. There will be only one angle α where the optical axes of both cameras are aligned (left). At this point the triangle is correctly scaled. In all other cases the cameras will not look into the same direction and the scaling between the relative poses is incorrect (right).

3.3 Estimation of the Translation Scale Factors

At this point of our calibration we have for each camera pair i, j a relative pose $\boldsymbol{R}_{i,j}, \boldsymbol{t}_{i,j}$ and a probability of this pose $p\left(\boldsymbol{R}_{i,j}, \boldsymbol{t}_{i,j}\right)$. We do not know the correct scale factor of each translation. Scaling a relative pose always means scaling the translation. The final calibration can only be estimated up to one common scale factor [13]. In order to estimate the scale factors in a camera triangle, traditional methods use either all three relative poses in the triangle [6, 7] or they try to establish point correspondences between all three camera images [13]. Our proposed method uses only two of the three relative poses and does not need any point correspondences at all. Instead we use active camera control and the probability distribution of (4). This reduces the number of required relative poses and totally avoids the chance of outlier point correspondences.

The final calibration is represented by a relative pose graph where each vertex represents a camera and each edge represents the relative pose between two cameras. Two vertices i and j are *simple connected* if there exists a path between them and they are called *triangle connected* if there exists a path of triangles between them [7]. The important difference is that only triangle connected subgraphs have a consistent scaling. In the beginning this graph has no edges. The following procedure is repeated until the graph is triangle connected.

We search for the camera triangle (i, j, k) which has the highest product of the probabilities of two of its relative poses $p\left(\boldsymbol{R}_{i,j}, \boldsymbol{t}_{i,j}\right) p\left(\boldsymbol{R}_{j,k}, \boldsymbol{t}_{j,k}\right)$ and no edge between the two vertices i and k. We now simultaneously estimate the third relative pose $\boldsymbol{R}_{i,k}, \boldsymbol{t}_{i,k}$ and all translation scale factors of the triangle. This is done by rotating camera i and k simultaneously around the plane normal of the camera triangle. In the beginning both cameras look into the direction defined by the translation $\boldsymbol{t}_{i,j}$. Now, we search for the rotation angle α that

$$\max_{\alpha} p\left(\boldsymbol{R}_{i,k}, \boldsymbol{t}_{i,k}(\alpha)\right) \text{ with } \boldsymbol{R}_{i,k} \overset{\text{def}}{=} \boldsymbol{R}_{j,k} \boldsymbol{R}_{i,j} \text{ and } \boldsymbol{t}_{i,k}(\alpha) \overset{\text{def}}{=} \boldsymbol{R}_{j,k} \boldsymbol{t}_{i,j} + \lambda_s \boldsymbol{t}_{j,k} \ , \tag{6}$$

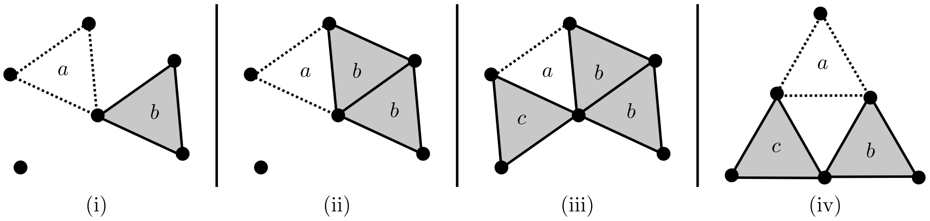

Fig. 4. Situations that can occur when inserting a camera triangle into the relative pose graph. The inserted triangle a has doted lines. Existing triangles are gray and share the same letter if they are in a triangle connected subgraph.

where we assume $\|\boldsymbol{t}_{i,j}\|_2 = \|\boldsymbol{t}_{j,k}\|_2$, the scale factor $\lambda_s \overset{\text{def}}{=} \sin(\alpha) / \sin(\Pi - \alpha - \beta)$ arises from the law of sines and $\beta \overset{\text{def}}{=} \arccos\left(\left(\boldsymbol{t}_{i,j}^{\mathrm{T}} \boldsymbol{t}_{k,j}\right) / \left(\|\boldsymbol{t}_{i,j}\|_2 \|\boldsymbol{t}_{k,j}\|_2\right)\right)$ is the angle between the translation vectors $\boldsymbol{t}_{i,j}$ and $\boldsymbol{t}_{k,j}$. The probability $p\left(\boldsymbol{R}_{i,k}, \boldsymbol{t}_{i,k}(\alpha)\right)$ is the probability of (4). There will be only one rotation angle α where the two cameras i and k look exactly into the same direction. For a clearer understanding the described relations are visualized in Fig. 3. The procedure is repeated in the opposite direction which results in estimating the inverse relative pose $\boldsymbol{R}_{k,i}, \boldsymbol{t}_{k,i}$. Again, we decide between these two poses based on their probability.

If the relative poses and scales of a camera triangle are known, it is inserted into the graph. We distinguish four different situations when inserting a new relative pose triangle into the relative pose graph. These four situations are illustrated in Fig. 4. The first situation is the trivial case of inserting a single triangle without conflicting edges. In the second case the inserted triangle shares a common edge with a triangle connected subgraph. This situation requires a rescaling of the triangle. The scale factor is defined by the relation between the translation lengths of the two common edges (the translation direction of these is identical). The third case creates a triangle connection between two prior simple connected parts of the graph. This requires rescaling the triangle and one of the two graph parts. The relative pose triangle in the fourth case cannot be inserted because it is impossible to correctly rescale the participating subgraphs. After inserting a camera triangle into the graph we need to check if two edges of some camera triangle are in the same triangle connected sub graph. In this case the relative pose of the third edge results from the poses of these two edges.

We note that several single camera triangles can be inserted before some of them build a triangle connected subgraph which reduces error propagation.

4 Experiments and Results

4.1 Experimental Setup

In our experiments we use a multi-camera system consisting of six Sony DFW-VL500 cameras with a resolution of 640 × 480 pixels. Each camera is mounted on a Directed Perception PTU-46-17.5 pan-tilt unit. We use a slightly modified version of this pan-tilt unit which allows to mount the camera quite close to the intersection of the pan and tilt axes (Fig. 1, right). We test our method on a total of 30 calibrations with 5 different setups of the multi-camera system. An example setup can be found in Fig. 1 (left). In order to generate ground truth we use the calibration software of the University Kiel [16] which uses a pattern based calibration method [1] with non-linear refinement. The intrinsic parameters of five zoom steps of each camera are estimated using the self-calibration method of [3]. The radial distortion of the images is corrected using the two parameter radial distortion model of [17]. As explained in Section 3.3, we can only calibrate up to a common scale factor. In order to compare our calibration with the ground truth, we scale our calibration result by the median of the factors $\|\boldsymbol{t}_{i,j}^{\mathrm{GT}}\|_2/\|\boldsymbol{t}_{i,j}\|_2$ of all camera pairs i, j, where $\boldsymbol{t}_{i,j}^{\mathrm{GT}}$ is the ground truth translation.

For the common field of view detection we use $\lambda_d = 0.5$ and $m_r = 71$ as suggested by Brückner et al. [12]. Each camera records 20 images in order to cover its complete environment. For each camera pair we use the $m_i = 20$ image pairs with the highest image similarity. From each of these image pairs 25 point correspondences are extracted using a nearest neighbor rejection threshold of 0.8 as suggested by Lowe [11]. This results in a maximum of $m_c = 500$ point correspondences for each camera pair. We use 50000 RANSAC iterations and an inlier threshold of 0.005 for the initial relative pose estimation. For each camera pair we save the $m_p = 5$ best poses according to the number of inliers and a minimum rotation and translation difference of $\theta_R = \theta_t = 2°$. We set the inverse scale parameter λ_e of the exponential distribution in (4) to 0.005. The choice of this parameter is not that critically as additional experiments show. For the matching during the optimization process we use a stricter nearest neighbor rejection threshold of 0.6.

4.2 Results

We present our calibration results in Fig. 5 using box plots (the box depicts the 0.25 and 0.75 quantiles, the line in the middle is the median and crosses are outliers, for further details please refer to [18]). In the upper row we show the rotation errors in degree. The bottom row displays the translation errors in degree or millimeters depending on the calibration step. We plot the errors of the relative poses for the initial calibration (initial, Section 3.1), after the evaluation and optimization step (opt., Section 3.2) and of the absolute camera poses for the final calibration (final, Section 3.3). We also distinguish whether we used the five zoom steps (zoom) or not (no zoom). For comparison we also present results of the passive uncertainty based calibration method of Bajramovic and Denzler [7] (passive). We manually rotate the cameras to ensure that they share a common field of view for this passive method.

The results show that each step refines the calibration and outliers are rejected. We achieve a final median rotation error of 0.9 degree and a median translation error of about 68 millimeters for the method using zoom. In the case of no zoom the results are slightly worse. In comparison to the passive approach we reach a similar rotation and a much lower translation error.

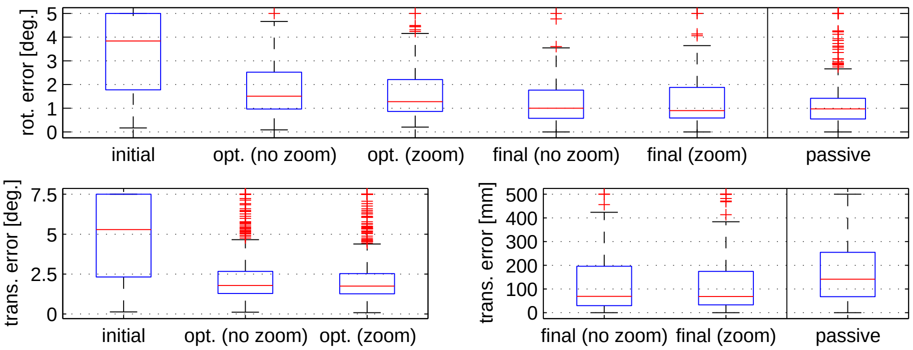

Fig. 5. Top: the rotation error during the different calibration steps in degree. Bottom: the translation error in degree or millimeters depending on the calibration step. For comparison we also present the results of the passive calibration approach of [7] (passive). The results of the initial calibration and some outliers are truncated.

Since we assume the pan and tilt axes to be identical to the Y and X axes of the camera, we also investigated the rotation error between the pan-tilt unit and the camera. We note that a (small) rotation between the camera and the pan-tilt unit has a higher impact on normalized point coordinates extracted from zoomed images. In order to rate the magnitude of this rotation we estimate it with the hand-eye calibration method of Tsai and Lenz [19]. The mean rotation between pan-tilt unit and camera in our experiments is 0.995°.

We also investigate the repeatability of the camera zoom by switching between the zoom steps and calibrating the intrinsic parameters several times. The calculated coefficients of variation for the intrinsic parameters lie in a magnitude of 10^{-3} which indicates good repeatability.

Calibrating a multi-camera system consisting of six cameras takes about 70 minutes in the current (serial) implementation. However, since many steps could be parallelized the runtime could be improved significantly.

5 Conclusions

We presented a method which uses active camera control for calibrating a multi-camera system consisting of pan-tilt zoom cameras. In order to evaluate a relative pose we introduced a probabilistic measure (4) which incorporates image and geometric information. Relative poses were optimized by rotating each camera pair in a way that simplifies the problem of extracting correct point correspondences. The final calibration process was based on these relative poses and their probabilities. The scale factors in each camera triangle were estimated using our probabilistic measure and active camera control. This allowed to reduce the number of necessary relative poses. Our experiments demonstrated the robustness and high accuracy of our approach. We achieved a median rotation error of 0.9° and a median translation error of 68 mm (Fig. 5). In our future work we hope to improve our calibration by considering the hand-eye calibration.

Acknowledgements This research was partially funded by Carl-Zeiss-Stiftung. The authors thank Ingo Schiller for the support on his calibration software, and Bernhard Klumbies and Andreas Langer for modifying the pan-tilt units.

References

1. Zhang, Z.: Flexible Camera Calibration by Viewing a Plane from Unknown Orientations. In: Proceedings of the ICCV. (1999) 666–673
2. Hartley, R.: Self-calibration from multiple views with a rotating camera. In: Proceedings of the ECCV. Volume 800. (1994) 471–478
3. Bajramovic, F., Denzler, J.: Self-calibration with Partially Known Rotations. In: Proceedings of the DAGM. (2007) 1–10
4. Chen, X., Davis, J., Slusallek, P.: Wide area camera calibration using virtual calibration objects. In: Proceedings of the CVPR. Volume 2. (2000) 520–527
5. Svoboda, T., Hug, H., Van Gool, L.: ViRoom–Low Cost Synchronized Multicamera System and Its Self-calibration. In: Proceedings of the DAGM. (2002) 515–522
6. Läbe, T., Förstner, W.: Automatic relative orientation of images. In: Proceedings of the 5th Turkish-German Joint Geodetic Days. (2006)
7. Bajramovic, F., Denzler, J.: Global uncertainty-based selection of relative poses for multi camera calibration. In: Proceedings of the BMVC. Volume 2. (2008) 745–754
8. Sinha, S., Pollefeys, M.: Towards calibrating a pan-tilt-zoom cameras network. In: Proceedings of the IEEE Workshop on Omnidirectional Vision. (2004)
9. Chippendale, P., Tobia, F.: Collective calibration of active camera groups. In: IEEE Conf. on Advanced Video and Signal Based Surveillance. (2005) 456–461
10. Matas, J., Chum, O., Urban, M., Pajdla, T.: Robust wide baseline stereo from maximally stable extremal regions. In: Proceedings of the BMVC. (2002) 384–393
11. Lowe, D.: Distinctive Image Features from Scale-Invariant Keypoints. IJCV **60**(2) (2004) 91–110
12. Brückner, M., Bajramovic, F., Denzler, J.: Geometric and probabilistic image dissimilarity measures for common field of view detection. In: Proceedings of the CVPR. (2009) 2052–2057
13. Hartley, R., Zisserman, A.: Multiple View Geometry in Computer Vision. Cambridge University Press (2003)
14. Nistér, D.: An efficient solution to the five-point relative pose problem. PAMI **26** (2004) 756–770
15. Fischler, M.A., Bolles, R.C.: Random Sample Consensus: a Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. Communications of the ACM **24**(6) (1981) 381–395
16. Schiller, I.: MIP - MultiCameraCalibration http://mip.informatik.uni-kiel.de/tiki-index.php?page=Calibration, last visited on 22-04-2010.
17. Heikkila, J., Silvén, O.: A Four-step Camera Calibration Procedure with Implicit Image Correction. In: Proceedings of the CVPR. (1997) 1106–1112
18. McGill, R., Tukey, J., Larsen, W.A.: Variations of Boxplots. The American Statistician **32** (1978) 12–16
19. Tsai, R.Y., Lenz, R.K.: A new technique for fully autonomous and efficient 3d robotics hand/eye calibration. IEEE Transactions on Robotics and Automation **5**(3) (1989) 345–357