Simultaneous Calibration and Mapping

Christian Nissler, Maximilian Durner, Zoltán-Csaba Márton, and Rudolph Triebel

Abstract We present evaluation experiments of a hand-eye calibration and camera-camera calibration method, which is applicable to cases where classical calibration methods fail. As described in our earlier works, the calibration works by performing rotational movements with the robot and estimating the rotational axes by tracking fiducial markers or other static parts of the environment (if the camera is moved with the robot, as in this experiment; otherwise tracking the robot itself or markers on it with a static camera). We extend our earlier work by virtually increasing the field of view of the cameras by using a mapping approach. We compare our results with an extended classical approach for the challenging case of calibrating a compliant humanoid robot having cameras with non-overlapping fields of view. We also show another application of this method: By observing markers attached to a robot's end effector, we can calibrate the markers to each other as well as to the robot's frame of reference.

1 Motivation

Cameras are the most commonly used sensors for robotic perception due to their excellent combination of data density and cost and energy efficiency. Therefore, important robotic perception tasks such as navigation, localization, or object detection are usually addressed by processing images from cameras that are mounted on the robot. This requires an accurate computation of the transformation between the camera coordinate frame and the robot or end effector frame, as well as the pairwise transformation between the cameras. The former is usually denoted *hand-eye calibration*, and it can be done by estimating the unknown transformations from the for-

All authors are with: German Aerospace Center (DLR), Institute of Robotics and Mechatronics, Münchner Str. 20, Oberpfaffenhofen, 82234 Weßling, Germany; e-mail: First.Last@dlr.de
R. Triebel is also with: Dep. Computer Science; Technical University of Munich (TUM), Boltzmannstrasse 3, 85748 Garching, Germany

ward kinematics of the robot [12] or by employing an external tracking system [5]. The latter is known as *camera-to-camera calibration*, and it is usually performed by observing a known calibration pattern with both cameras [13] and minimizing a set of error equations from pairs of corresponding features. This can also be done for other sensors such as 3D laser scanners [1].

However, those classical calibration methods can fail, e.g. if no (precise) kinematic information of the robot is available, or if there is not enough overlap between the cameras and no external tracking is available (e.g. for recalibration during deployment). In this study, we show how calibration can be done in those challenging situations. In particular, we perform rotational movements with the robot and estimate the rotational axes by tracking fiducial markers (AprilTags [10]) in the surrounding. Note that in principle any feature or object detection providing 6D poses would work for our approach. If only 3D positions are given, as in [14], the mapping part can be skipped, as explained below.

This paper builds on our previous work [8, 9] and extends it two ways: First, we address here the limitation of cameras with relatively small opening angles and without any overlap, which can lead to unstable results. In the improved method presented here, we additionally build a *map* of relative transformations between markers in the surrounding to virtually increase the field of view.

Second, we show the applicability of this new approach by calibrating a humanoid robot using this map. Before, only hand-eye calibration could be done in this particular case, but now we also provide an accurate camera-to-camera calibration.

2 Technical Approach

2.1 Mapping

If several markers are seen in a camera image, this additional information can be used. A graph of connected markers is built, refer to figure. 1. Here, each node is a seen marker, and each edge is a vector of relative transformations. By averaging multiple detections in this connected graph, an optimal solution for the relative transformations between makers is obtained. In order to do this, a weight/quality value for each edge of the graph is computed. For the rotational part, a Bingham distribution [3] is fitted to all relative rotations of this marker pair. The concentration term of the Bingham distribution describes the spread of the distribution and can therefore be used as a quality term for the rotation. For the translational part, a Principal Component Analysis (PCA) is performed. The Eigenvalue of the PCA describes the quality of the translational distribution. The translational and rotational quality value are then combined using the SRT distance [11]. The optimal paths are found using the Dijkstra algorithm, resulting in the ”best” connections (depending on the quality values) of all markers to one chosen parent marker.

2.2 Calibration

By obtaining the relative transformations between the markers, a map of the environment is known. This can be used for calibration by transforming all markers to a common coordinate system. The robot is firstly rotated around one axis. During this motion, all visible markers are tracked. By fitting the detections with a Bingham distribution (for the orientation) and a set of concentric circles (for the centroid), an axis of rotation can be obtained. Following this, the robot is tilted around one consecutive axis and the same rotation is repeated. For a more detailed description of the procedure please see [8, 9].

3 Experiments

Two experiments are presented: a hand-eye calibration and camera-camera of a humanoid robot and a hand-eye calibration of a robotic arm of a mobile platform. In figure 3 the humanoid robot be seen during the experiment, in figure 2 the experiment using the mobile platform is depicted.

3.1 Hand-eye and Camera-Camera Calibration of Humanoid Robot

The goal of the experiment was to calibrate the humanoid robot Rollin'Justin [2],[7]. For this robot, no reliable (forward) kinematics are available due to its compliant joints and it employs cameras with very little to no overlap, which makes a classical calibration challenging. See also figure 3 for a picture of the robot and its cameras, which are attached at the front, left, right and back of the platform of the robot.

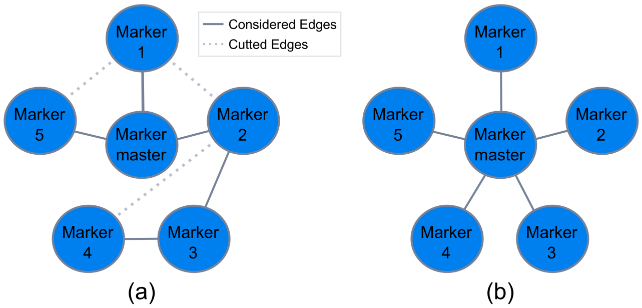

Fig. 1 (a) The initial graph, showing connected marker pairs; (b) the optimized graph, in which the optimal pose between a chosen master marker and all seen markers is computed

In a first step, a map of surrounding markers is built by moving the robot and tracking markers from several angles (see section 2).By optimizing the graph, all tracked markers can be transformed to a common frame of reference. This is then used in the actual calibration procedure: while rotating the robot around one axis, AprilTags in the surrounding are tracked and transformed into a common frame of reference (found before). The robot is tilted around another axis and the same rotation like before is repeated (e.g. rotation 1 is performed leaned back, rotation 2 is performed leaned to the front), at the same time tracking AprilTags again. The goal of the calibration is to find the rotation axis in space, around which the robot's rotation occurred. The estimation of the rotation axis can be split up in two parts: first, the vector around which the rotation occurs is estimated by fitting a Bingham distribution to the relative rotations of the transformed markers. Afterwards, the point around which the rotation occurs is estimated by fitting a circle to all detected marker positions. More details about this procedure can be found in [9]. After finding the rotation axes for several tilts, the intersection of those axes are estimated. Based on this, the transformation of all cameras relative to each other and to the robot's frame can be estimated. In total three different robot orientations are used in the experiment: a position of the robot tilted back relative to its upright axis (called orientation1 in the following), an upright position (orientation2) and leaning to the front (orientation3). In the next section we evaluate the quality of the found camera-camera calibration regarding the different configurations.

3.2 TCP Modeling

In addition to the presented marker graph generation , we are able to replace the master April Tag with an external coordinate frame. Therefore, the corresponding pose of the desired coordinate frame to each April Tag detection is recorded and treated as the master marker. In this experiment the marker graph is modeled with respect to the Tool Center Point (TCP) of a robotic arm which results in a marker graph representing the relative pose of each April Tag with respect to the TCP (see $T_{tcp}^{marker_x}$ in Fig. 2). In other words, by observing at least one of the related AprilTags the TCP pose can be estimated visually, which is an essential part of vision-based robot control. Furthermore, given the relative TCP-to-April Tag poses the Hand-Eye calibration (Camera-to-TCP) can be evaluated by comparing the visual TCP pose estimation and the one obtained by the forward kinematics.

The robotic system used in this experiment, consists of a mobile platform, additionally equipped with a robotic arm as well as a stereo-camera mounted on a pan tilt unit (see Figure 4). To generate the TCP-to-April Tag graph the robotic arm is placed in various (here five) arm joint configurations and rotates the TCP stepwise (in this set-up 60 steps) around its z-axis.

Besides the images, which are recorded by the stereo-camera directed towards the gripper, we also log the TCP poses obtained by the forward kinematic for each

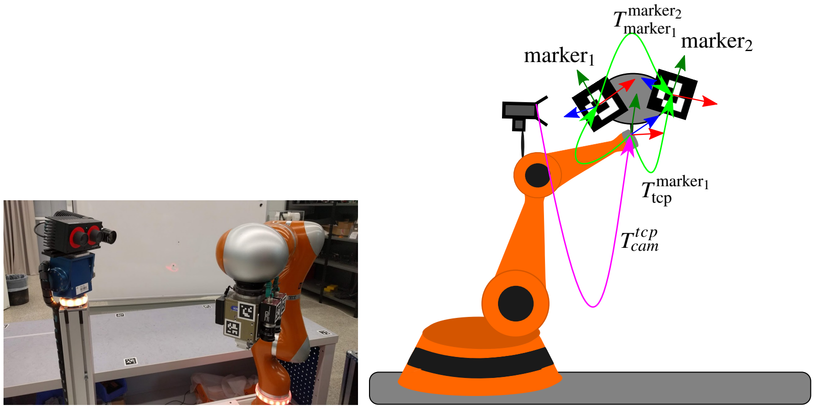
$T^{\text{marker}_2}_{\text{marker}_1}$
marker$_1$
marker$_2$
$T^{\text{marker}_1}_{\text{tcp}}$
T^{tcp}_{cam}

Fig. 2 (left) a picture of the used robotic arm and camera attached to a mobile platform, (right) an overview of the transformations, where the transformation colored in pink is assumed static and provided to the algorithm, and the transformations colored in green are estimated by the algorithm

step. This results in 60 images of the robot's end effector and corresponding TCP-logs for each of the five arm joint configurations.

The gathered data is then divided by its joint configurations into five data splits. In a 5-fold cross-validation manner the marker graph is modeled on four data splits and evaluated on the remaining one.

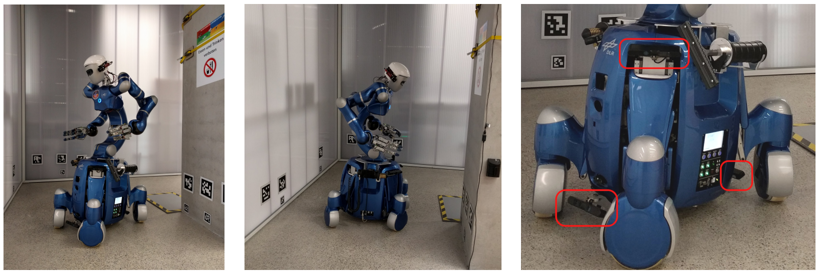

Fig. 3 The robot used in the experiment during the rotational motion around its up axis, tracking AprilTags in the surrounding(placed at the walls). The image on the right shows a close-up of the robot with its used cameras (marked red)

4 Results

In order to evaluate the quality of the resulting relative marker transformations described in section 2, ground truth was recorded by measuring a test object equipped with several AprilTags, by using a measurement arm (FaroArm Quantum M). The differences to the result of our mapping yield a translational error of $0.44mm \pm 0.36mm$ and a rotational error of $0.65^\circ \pm 0.34^\circ$ (mean ± STD based on pairwise tag detections in 15 frames).

The result of this obtained map is applied to the tracked markers by transforming them to a common frame of reference, the result of this can be seen in Fig. 4. In Fig. 4 (left) the individual markers seen by a camera are shown, whereas the localization of the common frame of reference yields much larger segments (Fig 4 right).

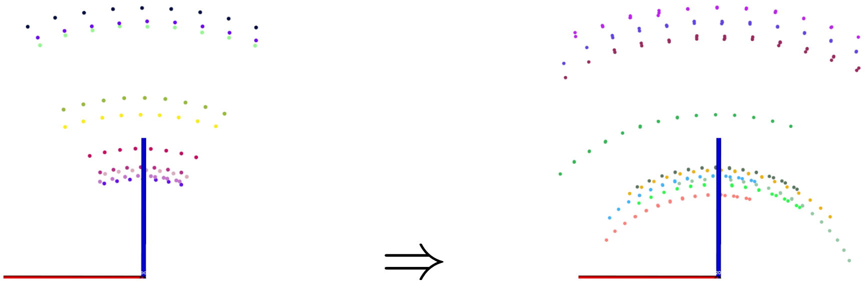

Fig. 4 (left) The tracked AprilTags during a rotation seen by one camera, the colors representing connected components of the graph, i.e. markers seen together; (right) By changing the detections to the mapped common frame of reference the field of view gets virtually increased, facilitating a better circle fit

By estimating the axis of rotation based on the transformed markers instead of the individual markers, a stable estimation of the rotation axis can be obtained. The result of this can be seen in figure 4 (left), which shows two rotation axes as seen by two different cameras. Those two rotation axes are describing the same rotation, which can be seen by overlaying one on another (Fig.4 (right)): they match perfectly.

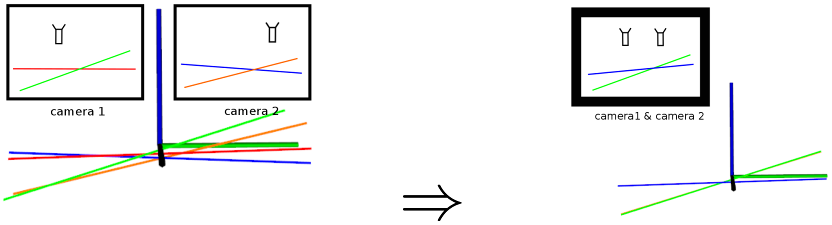

Fig. 5 (left) Two rotation axes seen by two cameras, by overlaying one onto the other we can find the relative transformations between the cameras (right)

4.1 Calibration Evaluation

In table 1 we are evaluating the quality of the found rotation axis in several ways: Firstly, we estimate both the line-to-line distance (d) for each found axis pair and the estimated angle (α) between the two rotation axes. Secondly, we calculate the area spanned by the estimated rotation axes intersection. The are of this triangle, for which each edge consists of an intersection point of an axis pair gives an estimate of how good three axes intersect. The result can be seen in table 3 for the case without the marker modeling part described before for all combinations of orientations.

Table 2 shows the corresponding results after marker modeling. It can be seen that the line-to-line distances are reduced and the area of the spanned triangle reduced drastically. Note that a small line-to-line distance d doesn't necessarily mean a better axis estimation, because (non-parallel) lines in space always produce a point with a minimal distance. However, the opposite is true: if the distance d is very high, the found axes intersection is also of poor quality. The spanned area of the triangle is a good estimate of how well the axis intersect. By repeating the experiment for different orientations (more than 3) this would allow to find the optimal combination.

Table 1 Intersection error of estimated axes (d) of each camera and estimated tilt (α) of the robot without marker modeling

	orientation1-3		orientation1-2		orientation2-3		all orientations
	$d(mm)$	$\alpha(^\circ)$	$d(mm)$	$\alpha(^\circ)$	$d(mm)$	$\alpha(^\circ)$	$A(m^2)$
front	10.09	8.07	9.54	4.40	2.89	3.67	0.0050
left	113	6.84	5.94	4.29	76.52	3.07	0.0267
right	34.77	8.50	5.94	4.29	27.17	4.22	0.0011
back	6.16	8.03	6.26	4.26	1.12	3.77	0.0117

Table 2 Intersection error of estimated axes (d) of each camera and estimated tilt (α) of the robot after marker modeling

	orientation1-3		orientation1-2		orientation2-3		all orientations
	$d(mm)$	$\alpha(^\circ)$	$d(mm)$	$\alpha(^\circ)$	$d(mm)$	$\alpha(^\circ)$	$A(m^2)$
front	2.08	8.00	0.47	4.13	3.17	3.83	0.0008
left	13.06	7.62	12.18	4.06	0.85	3.56	0.0007
right	7.18	8.14	13.50	4.20	5.75	3.94	0.0002
back	4.76	8.15	1.99	4.24	2.42	3.91	0.0024

In table 3 we compare our results with calibration results performed by an extension of classical hand-eye calibration and camera-camera calibration, using the upper cameras at the head of Justin to estimate the transformation between non-overlapping cameras.

Without the marker modeling the results produce relatively large rotational and translational errors, the camera pairs left to right were even not possible to calibrate at all. Table 4 shows the results after marker modeling, which reduces the rotational and translational error in all cases. It can also be seen that by averaging over all found calibrations, a calibration can be found which deviates from the classical solution by a few cm and degrees.

Note however that in that case the ”classical” calibration cannot be considered ground truth, but also introduces errors by relying on the kinematics of the head cameras.

Table 3 Results of camera calibration compared to classical calibration, without marker modeling, showing both translational (e_{trans}) and rotational errors (e_{rot}) for all camera combinations (rows) and robot orientations (columns), as well for a solution based on averaging all found calibrations (*mean*)

	orientation1-3		orientation1-2		orientation2-3		mean	
	e_{trans}	e_{rot}	e_{trans}	e_{rot}	e_{trans}	e_{rot}	e_{trans}	e_{rot}
left - right	—	—	—	—	—	—	—	—
back - left	1.69	27.03	0.55	25.21	1.06	25.7	1.19	11.14
back - right	0.35	7.87	0.18	5.96	0.27	6.73	0.31	6.73
back - front	0.22	2.84	1.21	2.84	0.58	2.76	0.68	2.8
front - left	1.81	11.51	0.65	9.17	0.5	14.41	0.53	11.69
front - right	0.46	11.7	1.08	10.83	0.33	11.01	0.32	11.11

Table 4 Results of camera calibration compared to classical calibration, after marker modeling, showing both translational (e_{trans}) and rotational errors (e_{rot}) for all camera combinations (rows) and robot orientations (columns), as well for a solution based on averaging all found calibrations (*mean*)

	orientation1-3		orientation1-2		orientation2-3		mean	
	e_{trans}	e_{rot}	e_{trans}	e_{rot}	e_{trans}	e_{rot}	e_{trans}	e_{rot}
left - right	0.03	9.36	0.08	9.85	0.09	9.45	0.04	9.45
back - left	0.24	2.36	0.05	2.85	0.21	1.88	0.16	2.35
back - right	0.25	5.48	0.1	5.33	0.15	5.95	0.16	5.51
back - front	0.17	1.46	0.3	1.5	0.36	1.47	0.08	1.46
front - left	0.08	3.06	0.32	3.59	0.16	2.64	0.1	3.04
front - right	0.09	9.4	0.39	9.37	0.21	9.6	0.1	9.43

4.2 TCP Modeling

In figure 6 the translational errors(right) and rotational errors (left) are shown for five different configurations of the robot. Generally the errors are in the range of less than 1*cm* and 2°, which is consistent to the ground truth estimates shown before. However, there are configurations of the robot which lead to a slightly higher error, like e.g. fold 4 in figure 6. This can be explained by the relatively imprecise angle encoders and therefore poor kinematics of the robot. However, if the TCP is visible in the camera image, this can be accounted for by employing the found transformations of our method.

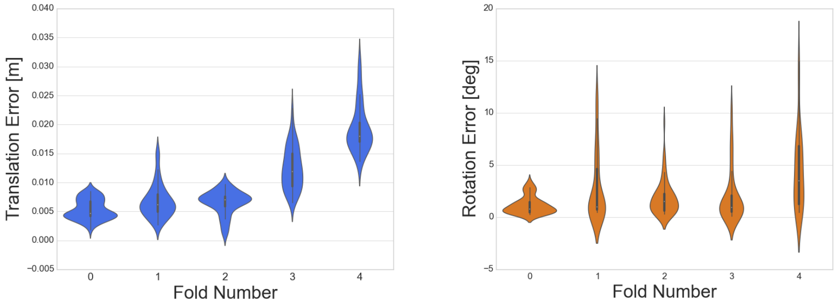

Fig. 6 Pose estimation error per fold. The average over all folds results in a translation error of 0.0089 ± 0.0052 meters and a rotation error of 2.24 ± 2.64 degrees

5 Discussion

Based on the results of the experiments, we have shown two things:

First, we built a map of markers in the environment by tracking markers during movement of the robot. By building a graph and optimizing it, we obtain the relative transformations (rotations and translations) between all seen markers.

This map can be used to calibrate the robot: another point of view is to regard this map as one large calibration object. In contrast to classical calibration methods the idea here is not to estimate the transformations based on the forward kinematics of the robot, but by estimating rotation axes around which the robot rotates based on tracked markers in the surrounding.

We previously [9] compared our method to state-of-the art, classical hand-eye and camera-camera calibration methods. In this study we improved on previous limitations of our method by utilizing a parallel running mapping algorithm. We experimentally show the application of this to the camera-camera and hand-eye calibration of a humanoid robot, which could not be calibrated in a stable manner with our method before. The robot does not offer forward kinematic information and its cameras have no overlapping field of view, which hinders the application of classical methods. We show the comparison of our method to a method based on external tracking using the robot's head cameras.

We also show another application of this method: By observing markers attached to a robot's TCP, we are able to both estimate the transformations in between the markers and the transformations between the robot's TCP frame to this markers, thereby finding a model describing the TCP.

One limitation of our method is the dependency on the angle of view of seen markers. This can make the estimation of the centroid of rotation unstable, resulting in relatively large translational errors compared to classical calibration methods. We showed in this study that we can mitigate this effect by building a map of the environment. However, this is a direction we want to pursue further. In the future we want to implement methods from the SLAM community like GTSAM [4] and error redistribution (e.g. g2o [6]), enhancing the quality of the rotation centroid estimation and reducing the translational error.

References

1. H. Andreasson, R. Triebel, and A. Lilienthal. Vision-based interpolation of 3d laser scans. In *Proc. International Conference on Autonomous Robots and Agents (ICARA)*, 2006.
2. Berthold Bäuml, Florian Schmidt, Thomas Wimböck, Oliver Birbach, Alexander Dietrich, Matthias Fuchs, Werner Friedl, Udo Frese, Christoph Borst, Markus Grebenstein, et al. Catching flying balls and preparing coffee: Humanoid rollin' justin performs dynamic and sensitive tasks. In *International Conference on Robotics and Automation (ICRA)*. IEEE, 2011.
3. Christopher Bingham. An antipodally symmetric distribution on the sphere. *The Annals of Statistics*, pages 1201–1225, 1974.
4. Frank Dellaert. Factor graphs and GTSAM: A hands-on introduction. Technical report, Georgia Institute of Technology, 2012.
5. Nadia B Figueroa, Florian Schmidt, Haider Ali, and Nikolaos Mavridis. Joint origin identification of articulated robots with marker-based multi-camera optical tracking systems. *Robotics and Autonomous Systems*, 61(6):580–592, 2013.
6. Rainer Kümmerle, Giorgio Grisetti, Hauke Strasdat, Kurt Konolige, and Wolfram Burgard. g 2 o: A general framework for graph optimization. In *International Conference on Robotics and Automation (ICRA)*. IEEE, 2011.
7. Daniel Leidner, Christoph Borst, Alexander Dietrich, and Alin Albu-Schaeffer. Classifying compliant manipulation tasks for automated planning in robotics. In *International Conference on Intelligent Robots and Systems (IROS)*, 2015.
8. Christian Nissler and Zoltan-Csaba Marton. Robot-to-Camera Calibration: A Generic Approach Using 6D Detections. In *1st Intern. Conf. on Robot Computing (IRC)*. IEEE, 2017.
9. Christian Nissler, Zoltan-Csaba Marton, Hannes Kisner, Ulrike Thomas, and Rudolph Triebel. A Method for Hand-Eye and Camera-to-Camera Calibration for Limited Fields of View. In *International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2017.
10. Edwin Olson. Apriltag: A robust and flexible visual fiducial system. In *International Conference on Robotics and Automation (ICRA)*. IEEE, 2011.
11. M. T. Pham, O. J. Woodford, F. Perbet, A. Maki, B. Stenger, and R. Cipolla. A new distance for scale-invariant 3d shape recognition and registration. In *2011 International Conference on Computer Vision*, pages 145–152, Nov 2011.

12. Klaus H Strobl and Gerd Hirzinger. Optimal hand-eye calibration. In *International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2006.
13. Klaus H Strobl and Gerd Hirzinger. More accurate camera and hand-eye calibrations with unknown grid pattern dimensions. In *International Conference on Robotics and Automation (ICRA)*, pages 1398–1405. IEEE, 2008.
14. Alex Teichman, Stephen Miller, and Sebastian Thrun. Unsupervised intrinsic calibration of depth sensors via SLAM. In *Proceedings of Robotics: Science and Systems*, Berlin, Germany, June 2013.