

# Robot-to-Camera Calibration: A Generic Approach Using 6D Detections

Christian Nissler\* and Zoltan-Csaba Marton\*

Institute of Robotics and Mechatronics, German Aerospace Center (DLR)

Muenchner Str. 20, D-82234 Wessling, Germany

Email: christian.nissler@dlr.de and zoltan.marton@dlr.de

## I. INTRODUCTION

With nowadays available optical sensors and modern computer vision approaches, robots can explore and interact with their environment in a multitude of situations autonomously. In order to allow for a reliable and safe interaction, especially in the presence of humans, the robot has to have a precise knowledge about its manipulators' poses. This demands a common frame of reference of the camera systems and the manipulators. The classical way of solving this problem is the well-studied hand-eye calibration [1]. For vision controlled manipulation with a robotic arm, knowledge of the camera to tool center point (TCP) transformation is important. This is typically done by classical hand-eye calibration, with which an unknown pose of a camera and its internal parameters (so-called intrinsics) are obtained in parallel. This is conducted by recording many pictures of a precisely known calibration pattern, usually a chessboard style pattern, at different configurations of the robot. The calibration algorithm then finds the optimal solution for both intrinsics of the camera and the sought transformation between camera and robot frame of reference. For good results it is crucial to obtain sample images at different configurations and a maximal range of angles. Even if this is often possible, there are cases where this method would fail because the robot is not able to achieve the desired range of angles. Another problem is that this method yields very poor results if the precision of the robot arm is insufficient. To avoid this problem, usually no hand-eye calibration is done at all in such cases, but Visual Servoing [2], [3] is performed, where the robot is commanded with direct visual feedback. This of course brings a whole range of other problems, like the need for a continuous line of sight of the camera to the scene and high computational complexity [4]. In this study, we want to present a new calibration method, which allows for a precise hand-eye calibration also in these situations. In the next section a new algorithm is introduced, with which it is possible to estimate a mathematical model connecting the kinematics of a robot to the camera frame of reference by sampling a big amount of visual markers and fitting a Bingham distribution. Hereby it is equivalent if the camera is attached to the moving arm, and the pattern is fixed, or the other way around. Additionally, the kinematic chain does not have to be an arm, it can be for



Fig. 1. Justin humanoid robot platform (top left) with a camera mounted at the base (top right), as example of complex robotic systems where traditional hand-eye calibration is difficult to apply.

example also just a mobile base. For our experiments we will focus on a humanoid robot, but the method described in the consecutive sections would be equivalent in other setups as well.

## II. MATERIALS AND METHODS

The basic idea of our approach is to track visible features, in our case AprilTags fiducial markers [5] during a rotational motion of the robot. We obtain for each visible marker a full 6D pose describing the relation of the camera to each marker. The relative rotation of each marker in consecutive camera images is the estimated according to section II-B. The result of one example motion can be seen in Figure 2: In the middle the camera frame is depicted, on the right the markers location (colored dots) and the estimated relative rotation axes can be seen. After applying a Bingham distribution to all of the estimated relative rotation axes, the peak of this Bingham distribution is shown in the origin. After an optimization step the centroid and axis of the rotation are estimated according to section II-C. The result, showing the found centroid and axis of rotation, is shown on the left in Figure 2. This axis corresponds to the physical rotation axis of the robot. If we repeat this rotational motion with a different pose of the robot, we again obtain the new rotation axis. The intersection of those axes correspond to a known point of the robot.

In the case of the humanoid robot platform Justin, we have a camera attached to the robot (Figure 1) and rotate it around its vertical axis, tracking markers in the robot's surroundings at the same time. By repeating this rotational motion for different

This work has partly been supported by the European Commission under contract number H2020-ICT-730014-INFUSE

\* Both authors contributed equally to this work

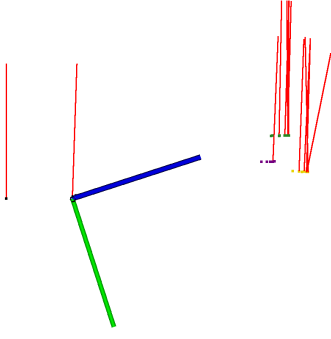


Fig. 2. Different axis estimates as unit vectors (red) shown in the camera's frame, orthogonally projected to the YZ plane. The Z axis of the camera (blue) is the viewing direction, Y points down. Relative rotation axes are visualized for each marker (where consecutive detections were made), whose positions are coded with different colors. The rotation axis estimate based on the fitted Bingham is shown in the origin, while the optimized axis (i.e. the upright direction) is at the estimated center of rotation.

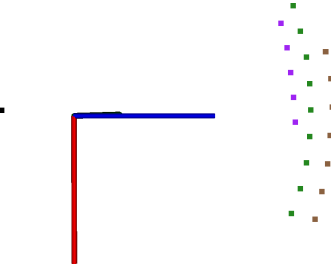


Fig. 3. Top down view of Figure 2, showing the marker detections in the camera's coordinate frame (in different colors per marker), with the estimated rotation center (black point on the left).

tilts of the robot, e.g. slightly tilted to the back or left, we can estimate the intersection point of the rotational axes, which corresponds to the center point of the base.

#### A. Bingham Distribution for Representing Rotations

In our work we are describing rotations using unit quaternions because of their computational advantages compared with other representations and their lack of singularities [6]. The Bingham distribution, named after Christoph Bingham, describes an antipodally symmetric distribution on the n-sphere [7]. Because of this antipodal symmetry, it is very suitable to describe a distribution of 3D rotations in unit quaternion form. The probability density function of a Bingham distribution is:

$$B(x; K, V) := \frac{1}{F(\kappa_1, \kappa_2, \kappa_3)} \exp\left(\sum_{i=1}^3 \kappa_i (v_i^T x)^2\right) \quad (1)$$

where  $\kappa_1, \kappa_2, \kappa_3$  are concentration parameters and  $V = [v_1, v_2, v_3]$  are orthogonal basis vectors.  $F(\kappa_1, \kappa_2, \kappa_3)$  is a normalization term such that the distribution integrates to one on the surface of the hypersphere  $S^3$ .

#### B. Estimation of Rotation Axis

There are several ways of obtaining the axis of rotation based on a set of 6D detections of one or more markers.

The straightforward approach is to treat each seen marker individually. Each marker ID, tracked over time then describes

its own rotational motion relative to the robot frame of reference. By sampling the absolute rotation of each ID at every time step the marker is seen, a Bingham distribution can be fitted to each marker separately. From the least concentrated component of this distribution the axis can be estimated.

Because each marker undergoes the same rotation relative to the robot, the relative rotations between time steps can be used to estimate the axis of rotation. Then one Bingham distribution for all seen markers combined can be fitted to these, and the rotation axis computed from the mode of this distribution.

#### C. Estimation of Rotation Centroid

Once a rotation axis is defined, a point defining the center of rotation is needed. Based on a single rotation it is not possible to yet define a unique point on the line around which the robot rotates, but an arbitrary point on the line can be used, and the final center of rotation will be estimated by intersecting two or more of these different lines the robot rotated around.

Since the detections of a marker will be located at equal distance from this line of rotation, we can define such a point by fitting a set of concentric circles to each marker's detections (see Figure 3). These circles lie on a plane having the rotation axis as normal, assuming the robot rotated around the vertical direction, i.e. perpendicular to the ground. This information helps to reduce the problem to 2D, by projecting the detections along the rotation axis, and fitting the circles in a plane.

As a first step, we can obtain a small correction on the rotation axis estimated based on the Bingham distribution by ensuring that the marker rotations really lie on a plane having this normal. By using the Bingham's estimate as initial guess, the variance of the projection of each marker's detections is minimized jointly, by adjusting the orientation of the rotation axis. This is equivalent to fitting of planes to the marker's detections, but robust to cases where there are only a few noisy detections, and/or they are close to being collinear.

Projecting the detections along the identified axis, a common center of rotation needs to be computed. To obtain a robust initial estimate, we find the best individual circle for each marker using RANSAC, and compute their average by weighting the estimated centers according to inlier size. Then this location is optimized for all markers jointly, using a Levenberg-Marquardt optimizer. The resulting 2D point is transformed back to 3D, marking the position of the current rotation axis.

#### D. Estimation of Robot's Frame of Reference

For each rotation the robot makes, the line of rotation is estimated as described, and since we know that the relative pose between these lines is a rotation around an axis, we can set out to identify this other axis of rotation. In keeping with the explanations related to the setup on Justin, we will refer to this second rotation as the robot's tilt, but in general this can be any other rotation axis that intersects the first one. Since robotic arms have consecutive rotation axes that intersect by design (and are described by its Denavit-Hartenberg parameters), producing the two rotations is not a problem. Typically the point of intersection is inside the more basic joint. Similarly, in the Justin example this point is defined by construction as the center of base tilting movements.

By performing the rotations around two such tilt axes, we can define the axis of tilting as the cross product between the rotation axes, and the point around which it is performed as the intersection point of the two lines. Given the fixed direction along the line of rotation, this then defines a unique frame of reference, set by the robot by “demonstrating” its rotation and tilt axes. In practice, we accept the two lines to be slightly off, s.t. the point of rotation will be computed as the half-point of the line-to-line distance segment. The estimate can be scored by this distance, and how close the angle between them matches what was executed by the robot.

In the following section we will verify each step of the procedure, and validate if the made assumptions hold.

### III. RESULTS

For our experiments the humanoid robot Justin [8], [9] was used. Three rotational movements around the Z-axis of the robot were performed, whereas the tilt of the robot was different. During the first movement the robot was standing upright, in the second movement the robot was tilted back and left, in the third movement the robot was tilted to the front right. During the rotation images were recorded by a camera attached to the front of the robot. Markers (AprilTags) were placed in the surrounding of the robot and were tracked by the robot, i.e. the 6D pose from the camera to each visible marker was known for every time step.

#### A. Rotation Axis Estimation

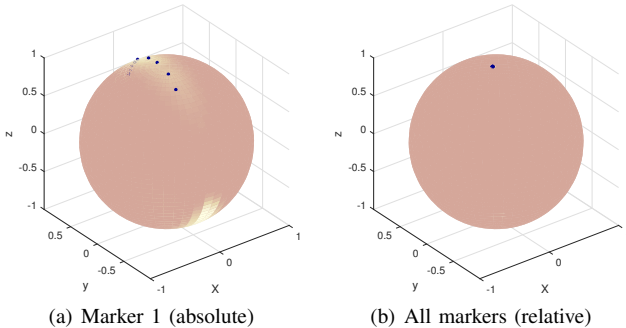


Fig. 4. The Bingham distribution based on different marker rotations depicted as a heat map on the unit hypersphere  $S^2$  of one example rotation. The subfigure *a* shows the resulting Bingham distribution of the absolute rotations of each marker, resulting in a distribution with a smaller concentration around one degree of freedom. Subfigure *b* shows the result based on the relative angles of all the markers combined, leading to a peaked distribution.

As discussed in subsection II-B, we can either fit a Bingham to the detections directly, or the relative transformations between detections (see Figure 4). We tested both approaches and show in Table I how much do the estimated axes vary between markers individually, and the two approaches described in subsection II-B. Since all markers should define the same rotation, the angles between the estimated axes should be minimal. We can see that for the individual markers, taking the detections directly (i.e. “absolute” detections) and estimating the axis of rotation as the least constrained axis in the Bingham, yields more stable results than using relative rotations separately for each marker ID (probably due to the fact that relative

TABLE I  
USING ALL RELATIVE ROTATIONS OF EACH MARKER GIVES A ROTATION AXIS WITH THE SMALLEST ERRORS (IN DEGREES) TO OTHER ESTIMATES

Consistency between markers			Differences per marker		
	<i>abs.</i> vs <i>abs.</i>	<i>rel.</i> vs <i>rel.</i>		<i>abs.</i> vs <i>rel.</i>	<i>abs.</i> vs <i>all</i>
Marker 1-2	1.52	1.05	Marker 1	2.29	<b>0.87</b>
Marker 2-3	2.18	3.40	Marker 2	1.19	<b>0.65</b>
Marker 3-1	0.66	3.15	Marker 3	0.26	<b>1.52</b>
					<b>1.76</b>

rotations for consecutive timestamps exist only for a subset of detections). However, taking the relative rotations has the advantage that in fact all markers estimate the same relative transformation, therefore all of them can be used together to estimate a final rotation. This estimate (denoted “all” in the table) has overall the lowest differences to the other estimation methods, and was used in the following.

Most relative rotation angles are around 5 degrees, which was the targeted frequency of taking images (see Figure 5), with most deviations being small (60% of the angles is in the  $\pm 0.5$  range). Small deviations can appear due to timing errors of the image grabbing, and detection errors.

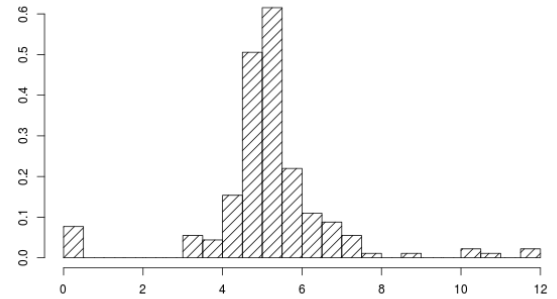


Fig. 5. Histogram of the measured relative rotation angles in degrees

We are filtering out bad detections based on the relative angles, by agglomerative clustering, starting at center and accepting angles with the difference threshold of 0.1 degrees. An example result showing the found rotation axis and centroid for a given marker can be seen in Figure 6.

#### B. Rotation Centroid Optimization

Our concentric circle fitting is different from standard approaches, like the 2D and 3D circle fitting from the Point Cloud Library, in that it fits multiple concentric circles at once, and that it does not optimize the radius explicitly. Rather it is minimizing the pairwise differences of the center-to-point distances within a set of circle inliers, resulting in more residuals, less parameters and better convergence.

The distribution of the errors can be seen in Figure 7, where we compare a version with and without explicit radius optimization. To model the distribution of the magnitudes of the errors, we fitted a Generalized Extreme Value (GEV) distribution to the resulting point-to-circle distances using L-Moments (`lmomco` R package), and found it follows as Weibull

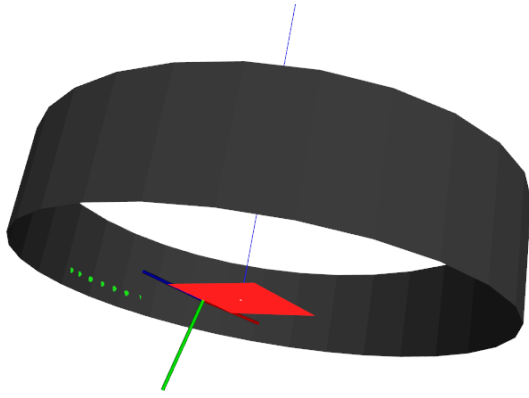


Fig. 6. Example result of found centroids and axes, depicted as 3D cylinder and the plane of rotation of the samples, shown in the centroid (orange), as well as the rotation axis (thin blue line) and found markers in green.

distribution<sup>1</sup>. The expected mode of this distribution is  $5.36E-04$  in the proposed radius-less case, and  $6.20E-04$  with radius optimization, meaning a small but considerable reduction of the most likely errors (86.44% relative reduction of the mode).

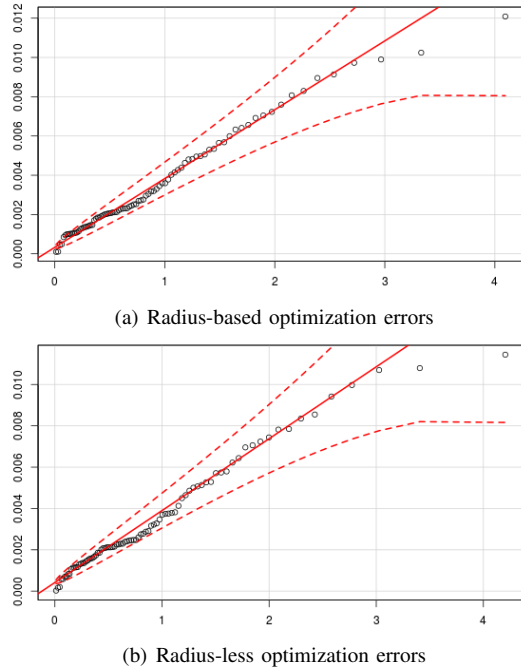


Fig. 7. Q-Q plot of fitted GEV distribution (Weibull) quantiles against the radius-based and the radius-less optimization errors, with a 95% confidence envelope, representing expected value ranges given the distribution<sup>2</sup>

### C. Validation of the Calibration

We performed the two rotation s.t. once the robot was tilting to the front right, then back to the left, and rotating with the

<sup>1</sup>The Weibull distribution is a special case of the GEV distribution, being a very versatile function, it is often used in a wide range of applications to model sizes/magnitudes of different kinds

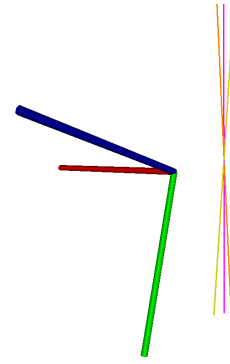


Fig. 8. The location of the three lines estimated as the centers of the three rotations, shown in orange, yellow and magenta, relative to the camera axes

wheels in a place. The line-to-line distance between the two estimated axes in the camera's frame was:  $1.738mm$ , meaning that the assumed center of rotation had to be moved off these lines by half of this distance, i.e. less than  $1mm$ . The angle of  $6.279^\circ$  was measured between the lines, which was  $0.09^\circ$  off our estimate of the relative tilt.

For validation, we estimated a 3rd axis using the same procedure, and it also crosses very close the same point:  $1.566mm$  away from the first line, and  $0.177mm$  from the second. The three axes are visualized in Figure 8.

As an independent evaluation was performed by exploiting the previously unused depth data, by detecting the normal directions of vertical planes in the environment (walls, furniture, etc), resulting in a mean angle of  $90.246$  degrees to our estimated vertical robot position (with STD of  $0.575$  degrees).

### IV. FUTURE WORK

In future work we plan to enhance the precision by adding more markers and by the use of pre-learned marker transformations. An interesting line of research is to use natural landmarks / keypoint descriptors instead of fiducial markers.

**Acknowledgment:** We would like to thank Daniel Leidner and Laura Beckmann for their help with the experiments.

### REFERENCES

- [1] K. H. Strobl and G. Hirzinger, "Optimal hand-eye calibration," in *Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on*. IEEE, 2006, pp. 4647–4653.
- [2] B. Espiau, F. Chaumette, and P. Rives, "A new approach to visual servoing in robotics," *Robotics and Automation, IEEE Transactions on*, vol. 8, no. 3, pp. 313–326, 1992.
- [3] P. I. Corke, "Visual control of robot manipulators-a review," *Visual servoing*, vol. 7, pp. 1–31, 1993.
- [4] F. Chaumette, "Potential problems of stability and convergence in image-based and position-based visual servoing," in *The confluence of vision and control*. Springer, 1998, pp. 66–78.
- [5] E. Olson, "Apriltag: A robust and flexible visual fiducial system," in *2011 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2011, pp. 3400–3407.
- [6] J. Diebel, "Representing attitude: Euler angles, unit quaternions, and rotation vectors," *Matrix*, vol. 58, no. 15-16, pp. 1–35, 2006.
- [7] C. Bingham, "An antipodally symmetric distribution on the sphere," *The Annals of Statistics*, pp. 1201–1225, 1974.
- [8] B. Bäumel, F. Schmidt, T. Wimböck, O. Birbach, A. Dietrich, M. Fuchs, W. Friedl, U. Frese, C. Borst, M. Grebenstein *et al.*, "Catching flying balls and preparing coffee: Humanoid rollin' justin performs dynamic and sensitive tasks," in *2011 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2011, pp. 3443–3444.

- [9] D. Leidner, C. Borst, A. Dietrich, and A. Albu-Schaeffer, “Classifying compliant manipulation tasks for automated planning in robotics,” in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2015, pp. 1769–1776.

<sup>2</sup>In both plots the points around 0 error sightly leave the envelope downwards, meaning we observed a bit fewer errors of 0 than predicted by the distribution