Evaluation and Improvement of Global Pose Estimation with Multiple AprilTags for Industrial Manipulators

Christian Nissler*, Stefan Büttner*, Zoltan-Csaba Marton*, Laura Beckmann* and Ulrike Thomas†

*Institute of Robotics and Mechatronics, German Aerospace Center (DLR), 82234 Wessling, Germany
Email: christian.nissler@dlr.de
†Robotics and Human Machine Interaction Lab, Technical University Chemnitz, 09126 Chemnitz, Germany
Email: ulrike.thomas@etit.tu-chemnitz.de

Abstract— Given the advancing importance for light-weight production materials an increase in automation is crucial. This paper presents a prototypical setup to obtain a precise pose estimation for an industrial manipulator in a realistic production environment. We show the achievable precision using only a standard fiducial marker system (AprilTag) and a state-of-the art camera attached to the robot. The results obtained in a typical working space of a robot cell of about 4.5m x 4.5m are in the range of 15mm to 35mm compared to ground truth provided by a laser tracker. We then show several methods of reducing this error by applying state-of-the-art optimization techniques, which reduce the error significantly to less than 10mm compared to the laser tracker ground truth data and at the same time remove existing outliers.

I. INTRODUCTION

In automobile manufacturing, aerospace construtin and many more fields Carbon-fiber-reinforced polymers (CFRP) structures are becoming a more and more important material, providing good material characteristics at a very low weight [1], [2]. The production with CFRP materials is leading to several challenges. So is the placing of CFRP patches right now a very little automated process, involving much manual work, which is very tedious for the worker, containing for example many steps where work has to be done overhead. There are also possible health risks involved because fine, respirable dust develops often when working with CFRP materials [3]. Another important aspect is that the whole production with CFRP is relatively expensive compared to "traditional" material like e.g. aluminium.

This motivates to intensify the automation of the production with CFRP. Several challenges arise thus, one being the sensitivity of the CFRP patches in their "raw" state, making the gripping, picking up and placing down of them an open problem. Another problem is the needed high absolute precision of the placing of the CFRP pieces. A typical CFRP structure consists of many individual CFRP components of different sizes, which are placed in a sandwich-like pattern in a mold, where the alignment is chosen to optimize the strength of the final structure. The placing of these parts requires a very high precision (typically few mm). The DLR founded the Center of Lightweight Production (ZLP) to investigate

some of the mentioned fields and offering a prototypical assembly line to test automated CFRP manufacturing in a realistic scenario. Because CFRP components are put in molds, an overhead configuration of robots, meaning robots hanging from the ceiling instead of standing on the ground, is desirable. This leads to the problem that the weight of the robot is deflecting its linear axis. Also robots influence each other and themselves by, for example, inducing vibrations. All of this makes the pose estimation of the robot control, based on the robot kinematics, unreliable. One approach to solve this problem and achieve a reliable high-precision pose estimation is by using a laser tracker. In the described environment a laser tracker is available (Leica AT901). This offers a very precise pose estimation by tracking a fiducial point on the robot. This tracking fails however if there are occlusions by, e.g., other robots or tools and if the robot "dives" into a mold. It also is heavily affected by the industrial environment, e.g. dust can prevent a robust tracking.

An open question which we address here is if the described system would benefit from a camera system attached at the robot end effector combined with a fiducial marker system attached at outer structures. Many techniques used in visual tracking, like Shi-Tomasi features and KLT [4], [5], [6], rely on distinctive keypoints in the image data that can be re-detected in subsequent frames. Azad et al. propose a method for tracking (single, rigid) 3D models by instead rendering their edges and computing the overlap between the image edges [7]. Ulrich et al. similarly match 3D CAD models in monocular images [8]. In our case, in an industrial environment with very little texture natural landmarks as mentioned above are very hard to obtain in a robust and precise manner. State of the art image features like SIFT [9] or SURF[10] fail when confronted with objects with very little texture. On the other side, fiducial marker based systems like the ARTag [11] and the AprilTags [12] can be easily attached in the working space of the robot. Those marker systems are widely used in e.g. Augmented Reality [13], mobile robotics [14], even camera calibration [15] and demonstrated to be reliable feature trackers.

In this work we focus on evaluating the performance of the AprilTag system, and improving its performance by taking into account multiple detections, in the same image or over time. This problem bears close resemblance to (and uses techniques from) the 6DoF pose estimation domain, i.e. 3D registration.
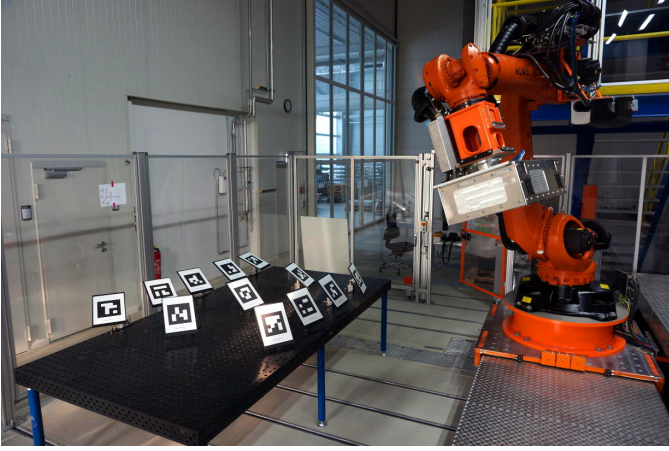
Fig. 1. The experiment setup, showing the used robot with the attached end effector containing the camera on the right and a table with the attached AprilTag markers on the left

There are two main types of registration approaches, one where the initial state is already close to the optimal (e.g. during sequential data acquisition), called local registration, and the other being global registration, where the initial state is arbitrary (and typically not used for finding the solution). Since in robotic applications as considered in this paper, the targeted tags are observed multiple times as the robot moves, local registration techniques could be used to localize the robot more precisely.

One of the main uses of local registration is to improve alignments obtained with other methods and the large field of Simultaneous Localization And Mapping (SLAM). An overview and evaluation principles are presented in [16]. For our application data correspondence is not a problem, and there is no need to create a map, so one needs to only deal with measurement and ego-motion uncertainties. This can be efficiently dealt with using Sequential Monte Carlo methods [17], [18], [19]. Therefore, we are using a particle filter to fuse the results obtained at individual timestamps, which is a widely used method for localization tasks, with various techniques available for improving their accuracy [20], [21], [22].

An important part of global registration is how to find and filter correspondences between the source and target model. As discussed in [23], several methods focus on finding consistent/common transformations among many hypotheses derived from point, surflet or local reference frame correspondences. Two common approaches are voting, or density-based estimations [24], [25], [26], and ones based on random sampling consensus (RANSAC) [27], like for example [28], [29], [23]. In the case of AprilTags, the problem then becomes to find the most likely robot/camera pose at each timestamp, given the estimates based on the detections of multiple tags. To find a maximal consensus set, i.e. rejecting misdetections, we employ RANSAC and compare it to a least-squares solution. The idea of RANSAC is, instead of using the full set of data points and trying to find a model minimizing the global error, to use only

a minimal subset of datapoints, iteratively estimating a model with those points and evaluating this model with the residual data points. This approach has the advantage to least-squares methods that it also works with very noisy data and in presence of outliers. However, it is computationally more complex and - being a iterative approach - it's not possible to guarantee a runtime (but a maximal runtime). While the proposed solution based on the particle filter integrates information over time, the RANSAC one optimizes over multiple detections in the same image.

In summary, this paper deals with the question what precision a pose estimation with a camera attached to the end effector of a robot can achieve in a realistic setting. In order to show this, experiments with a prototypical end effector were conducted and compared to ground truth obtained by the laser tracker mentioned before. We present methods to improve the precision in case of a single tag being detected/tracked, and for the case where multiple ones are visible in the camera image.

This work can be considered an extension of a previous preliminary result presented at ETFA 2015 [30].

II. Experiment Description

The experimental setup consists of a KUKA KR 210 robot with an attached end effector and a test rig, consisting of a table with 12 AprilTags mounted on top (see figure 1). This rig is supposed to simulate a tooling, e.g. a mold.

The end effector is designed and constructed as a prototype, containing a AVT GigE camera and an embedded PC. A fanless embedded computer is used, which is industry-grade hardware suitable for environments found in production lines, providing an i7 quad core processor, 8GB of RAM and a SSD drive in a rugged, fanless chassis. This setup was constructed with regard to the special environment. Because very fine dust can exist during handling of CFRP materials, a fully enclosed space for the camera and the embedded computer is desirable. This is an even bigger problem considering that CFRP material can be electroconductive and is therefore potentially dangerous to the entire electronic equipment. For cooling and preventing dust from getting inside, pressured air is constantly maintaining a positive pressure against the surrounding. It is an integrated system which only needs power and pressured air connections. All the calculations for the pose estimation and marker detection run on the embedded computer, thus making a usage on any robot possible.

In order to simulate typical process steps, three different motion profiles are executed for these experiments: A horizontal and a vertical arc around the table and another linear motion towards the markers, as shown in figure 2.

Figure 3 shows the principal setup of the experiment, depicting the robot with the attached image processing end effector including the camera, the laser tracker, and the simulated mold with the attached AprilTags. For transformations the following notation is used: $\mathbf{T}_a^b$, meaning a (passive) transformation from reference frame a to reference frame b, therefore: $\mathbf{T}_a^c = \mathbf{T}_b^c \mathbf{T}_a^b$. The used transformations are thus:
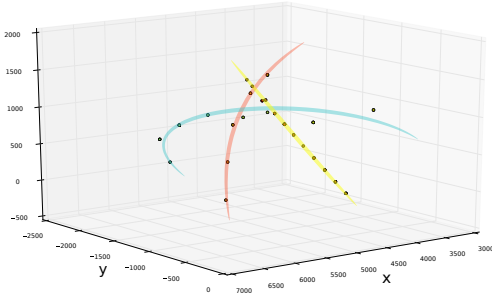
Fig. 2. Concept of the three camera motion profiles.

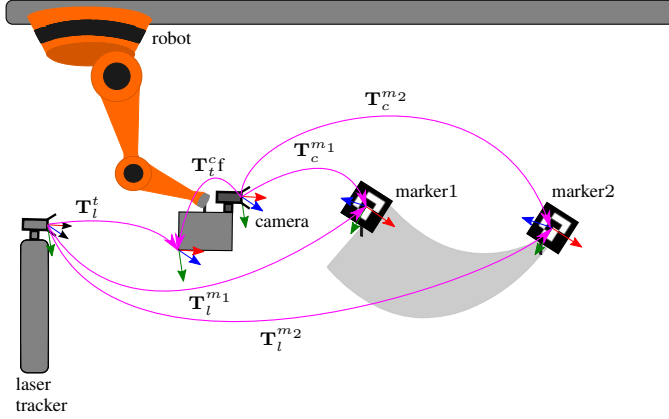| yellow | linear movement towards tag |
| red | arc in vertical direction |
| blue | arc in horizontal direction |

Fig. 3. Overview of the transformations

- $\mathbf{T}_l^t$ the transformation from the laser tracker to the tracked ficudcial point on the robot end effector,
- $\mathbf{T}_t^c$ the transformation from the the tracked fiducial point to the camera,
- $\mathbf{T}_c^{m_i}$ the transformation from the camera to the observed AprilTag marker, where i is the id of that marker,
- $\mathbf{T}_l^{m_i}$ the transformation from the laser tracker to the AprilTag marker, where i is the corresponding id.

The transformations $\mathbf{T}_t^c$ and $\mathbf{T}_l^{m_i}$ are static, which means they have to be determined once before the experiment, in the calibration step, see II-A. The transformation $\mathbf{T}_c^{m_i}$ is obtained from the AprilTag algorithm explained in section II-B and the transformation $\mathbf{T}_l^t$ is the ground truth obtained by the laser tracker. The latter two transformations are not static and are measured throughout the experiment. The frame of reference of the laser tracker is a stationary coordinate system and is therefore (without loss of generality) considered as the world coordinate system.

A. Calibration Process

For handling of CFRP material, e.g. putting components in a mold or handling material with two robots simultaneously, an absolute positioning, meaning not only relative to the robot is needed. In order to be able to achieve a positioning to a determined world coordinate system, the size and the location

of the AprilTag markers have to be measured. The size is determined by a high-precision microscope, whereas the location and rotation is measured by a laser tracker (Leica AT901), which allows for a 3D measurement precision of $0.5\mu m$ and an angular accuracy of $0.14\ arc\ sec$ [31]. Therefore, it is possible to know precisely the location and orientation of the markers and their respective sizes. For every marker, its corresponding marker based coordinate system, located at the middle point of the marker, was then calculated.

The intrinsic parameters of the camera (focal length, principal point and distortion parameters) and the transformation $\mathbf{T}_c^t$ from the camera frame to the fiducial point tracked by the laser tracker were established by taking pictures of a precisely know chessboard pattern at different poses of the robot tracked by the laser tracker. This is a classical *hand-eye calibration* problem [32], [33] and was solved with the DLR CalDe and CalLab [34] camera calibration software.

B. Image Acquisition and Processing

The tag detection software is a C++ port of the original AprilTag algorithm presented in [12] allowing full feature localization (position and orientation) relative to the camera from a single image. We chose the AprilTag system because it improves upon previous ones, incorporating a fast and robust line detection system, a stronger digital coding system and greater robustness to occlusions, warping and lens distortions than for example the ARTag [11] system. In order to compare the estimated poses, the output of the tag detection system (translation and rotations of the AprilTag relative to the camera frame) and the corresponding marker id are saved in every step along with the ground truth transformation from the laser tracker to the robot end effector.

Because we want to evaluate the absolute error between our camera based pose estimation and the ground truth we estimate:

$$\mathbf{T}_l^c = \left(\mathbf{T}_c^{m_i}\right)^{-1}\mathbf{T}_l^{m_i} \tag{1}$$

which is the transformation from the laser tracker to the camera frame based on the pose estimation of the camera. Note that we assume the error in the transformation from laser tracker to be neglectable. This is because we measured the AprilTags like explained in section II-A very precisely.

$$\widetilde{\mathbf{T}}_l^c = \mathbf{T}_t^c\left(\mathbf{T}_t^l\right)^{-1} \tag{2}$$

is the transformation from the laser tracker to the camera based on measurements. Note that the transformation T_t^c is obtained in the hand-eye calibration phase. Therefore we can't assume to have "true" ground truth here. By comparing the transformations we can however make a statement about the overall error in the transformation chain. Because we want to evaluate the absolute error between our camera based pose estimation and the assumed ground truth we compare 1 and 2:

$$\mathbf{E} = \mathbf{T}_l^c\left(\widetilde{\mathbf{T}}_l^c\right)^{-1} \tag{3}$$

which is our error matrix. If we split this error matrix E in its translational vector t_E and its rotational part described by the axis-angle representation $(\mathbf{v}, \alpha)$ we can evaluate the translational error and angular error:

$$e_t = \|\mathbf{t}_E\|_2 \qquad e_r = |\alpha| \qquad (4)$$

C. Single AprilTags

The three motion profiles mentioned before are performed several times and in certain intervals measurements are obtained, meaning the captured image frames, depicting the marker locations, are saved and the location of the laser-tracked point at the end effector of the robot are measured. Note that our system also would make dynamic measurements possible, but because a synchronization of the laser tracker and our system is not yet possible at the moment, static measurements are done.

To show the benefits of integrating multiple measurements of the same tag over time, the individual detections are used in a standard particle filter [17], [19]. The measurement and motion models assume zero-mean Gaussian errors, with the variances determined based on approximate accuracies of the robot and the AprilTag detection method. More specifically, for representing the rotational part of the motion and measurement model the Bingham distribution is used [35]. The rotational and translational parts are sampled and weighted independently, with the final weight being computed as the product of the translational and rotational probabilities.

For estimating the most likely 6DoF pose, the expected value of the distribution of the particles are used. The rotational part requires again special care, as averaging rotations is not straightforward. However, the normalized arithmetic mean was shown to be a good approximation of the optimal solution [36], and is therefore used in our system. We use 100,000 particles in our experiments, of which 1% are uniformly sampled in each time step to avoid depletion.

D. Multiple AprilTags

If more than one marker is visible, a common pose estimation based on these several markers may be more beneficial than one based on individual markers. For this the outer corners of all visible markers in a camera image are obtained from the AprilTag detection software. Because the location and orientation of the markers were measured in the calibration step (II-A), these corner points can be calculated in the world frame of reference. With that, the pose $\mathbf{T}$ of the camera in a world coordinate system is sought. This can be solved in a Least-Squares approach:

$$\mathbf{T}_l^c = \arg_{\mathbf{T}} \min \sum_i \left\| \begin{bmatrix} u_i \\ v_i \\ w_i \end{bmatrix} - \mathbf{AT} \begin{bmatrix} x_i \\ y_i \\ z_i \\ 1 \end{bmatrix} \right\|_2^2 \qquad (5)$$

where $\mathbf{A}$ is the camera matrix which maps from camera space to pixel coordinates and which was obtained by the camera calibration (depicting the camera intrinsics). $\begin{bmatrix} x_i & y_i & z_i & 1 \end{bmatrix}^t$

is the vector of the object coordinates (of the i-th corner point), which are in our case the known marker corners in the laser tracker frame of reference, and $\begin{bmatrix} u_i & v_i & w_i \end{bmatrix}^t$ are the corresponding marker corners detected in the image. Note that this is a 2D-homogenous vector and w_i is just a scale parameter. This is a classical Pose-from-n-Points (PnP) algorithm and can be solved iteratively by a Levenberg-Marquardt optimization.

To address for example misdetections or bad measurements, that can affect the least squares solution, RANSAC was used. In contrast to the Least-Squares Approach (Equation 5) only the minimal amount of data to estimate a pose (i.e. 4 marker corners) are used to obtain one possible solution. Then the distance of every data point to this estimated model (i.e. the reprojection error) is calculated. If the distance is under a certain threshold this point is considered as an inlier. These steps are then repeated until a maximum number of iterations is reached and the resulting model is the one with the most inliers over all these iterations. In each step, the maximum number of iterations can be obtained from the best number of inliers found so far using a probabilistic formula [27], ensuring that the probability of finding a solution with more inliers is below a user-given threshold (typically $\ll 1\%$).

III. EXPERIMENTAL RESULTS

A. Single AprilTags

Figure 4 shows the mean, minimal, and maximal translational and rotational error (4) for the motions described in section II, namely a horizontal semicircle motion in figure 4a, a vertical semicircle in figure 4a and a linear motion approaching the markers in 4c. The average error over all measurements is in all cases around $15mm$. Even if this sounds rather imprecise compared to the robot precision, one has to keep in mind the large working space of the robot: in case of the linear motion we start the motion from a distance of about $4.5m$.

One interesting issue is the behaviour in the linear motion case: the error decrease as the robot comes closer to the markers, as expected. However, the errors increase again after reaching a minimum of about $5mm$. This can be explained by the chosen configuration of the camera's optics. It was chosen to be in focus in almost the complete workspace distance, however, if the camera gets too close the image becomes too blurry reducing the performance of the april tag detection.

The detections are then used in a particle filter, that report in each step the expected value of the relative transformation between the tag and the camera. As marker number 8 is visible most often, we report the results for that in figure 5. The different trajectories are treated again separately, but the detections from the three runs are merged to better illustrate the behavior of the filter over time. The estimated 6DoF pose's error to the ground truth provided by the laser tracker is in general more reliable than the raw detections, being able to filter out the more inaccurate detections' effects. The translational errors are below $10mm$, and rotational ones below 0.5 degrees (or even lower). Additionally, the MAP estimate are also computed for the last step, which differs
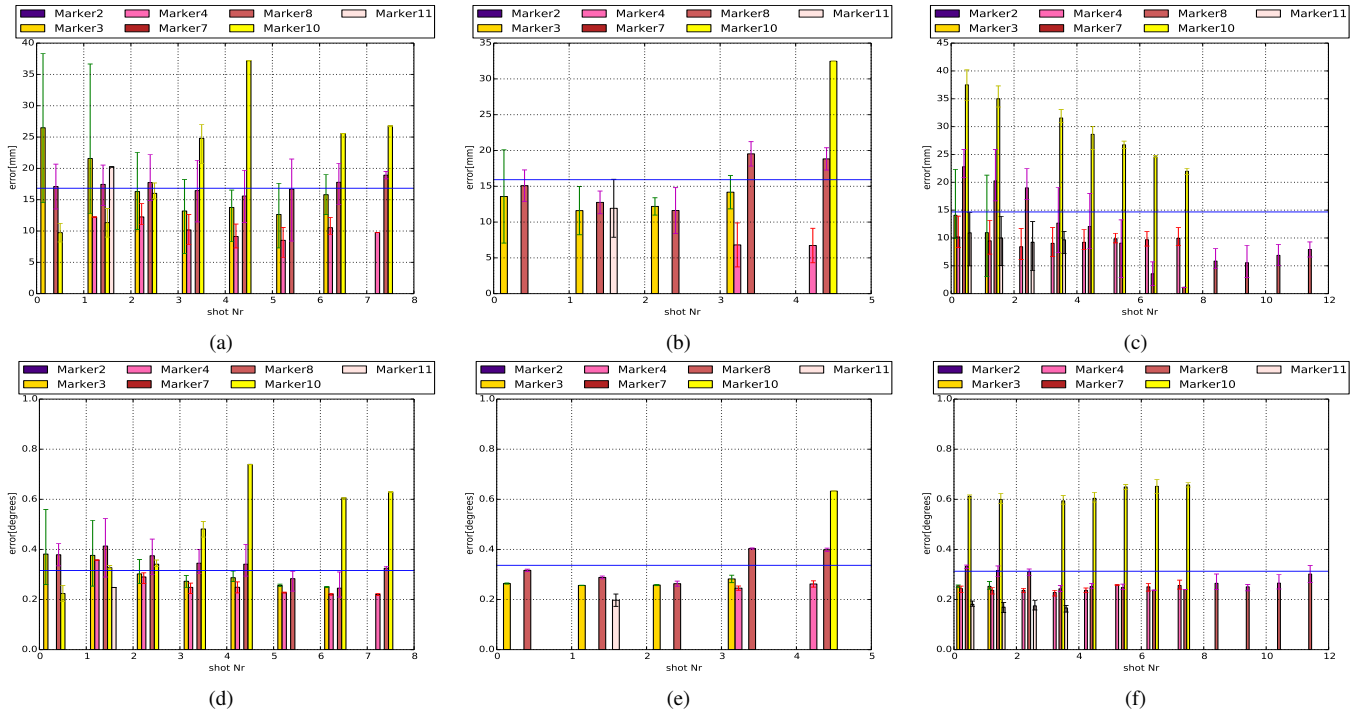
Fig. 4. Mean of the translational (top row) and the rotational error (bottom row) of the pose estimation based on individual AprilTag detections in mm/degrees, for

(a), (d) : horizontal semicircle
(b), (e) : vertical semicircle
(c), (f) : linear motion.
Note that the error bars depict the minimal and maximal error and the blue line is the total average over all measurements.

from the expected value if the distribution represented by the particles is skewed and/or is not unimodal. The fact that for the measurements there is usually a nearby particle (so that there is no depletion) is shown by the fact that the particle having the highest data likelihood has similar translational and rotational errors than the detection.

B. Multiple AprilTags

1) Least Squares: Figure 6 shows the mean, minimal, and maximal translational and rotational error after Least Squares Optimization for the motions described in section II.

The Least Squares approach lowers the error in most cases. Note that the missing bars to the right in figure 6c are because the camera is that close to the mold, that only one marker is visible. Then, the Least Square error is the same as the result of the several AprilTag result in figure 4c.

2) RANSAC: Figure 7 shows the mean, minimal, and maximal translational and rotational error after RANSAC Optimization for the motions described in section II. We also report the inlier ratios, which is the number of data points RANSAC treats as valid data points divided by the total amount of data points for each measurement step and a total average in table I.

The RANSAC approach lowers the error vastly — the total average of all measurements goes down to only 8 to $10mm$. It is interesting to see that outliers, which are probably bad

detections of a marker, are rejected, for example marker 10 compared to figure 7c. Another interesting fact is that the error seems to be quite static, also during the approach to the markers in the third figure to the left. This means that the error is the same when the camera is far away from the scene, but can see a lot of markers and the case when the camera is close to the scene and sees only one marker (filling more of the camera image however). The residual error could be explained by static error in the transformation chain, for example errors induced by the hand-eye calibration to the transformation from the fiducial point of the laser tracker to the camera (T_t^c).

IV. CONCLUSIONS

We presented an evaluation of the widely used AprilTag detection algorithm based on a well calibrated system that allows for high-accuracy ground truth estimates. The resulting dataset will be released online to facilitate the development of other improvements.

Our approach focused on improvements in two areas, first for single AprilTags, and second on estimates based on multiple AprilTags. Our experiments show clear improvements over the off-the-shelf methods. This warrants further analysis of the detection and tracking methods in order to improve on the accuracy. Obtaining reliable, very precise pose estimates would benefit not only our application, but many others, especially in industrial, controlled environments.

shot Nr.	0	1	2	3	4	5	6	7	8	9	10	11	total
horizontal semicircle	89	74	69	70	78	75	72	77					76
vertical semicircle	91	78	96	67	82								83
linear motion	82	73	100	67	89	79	75	67	69	78	100	100	79
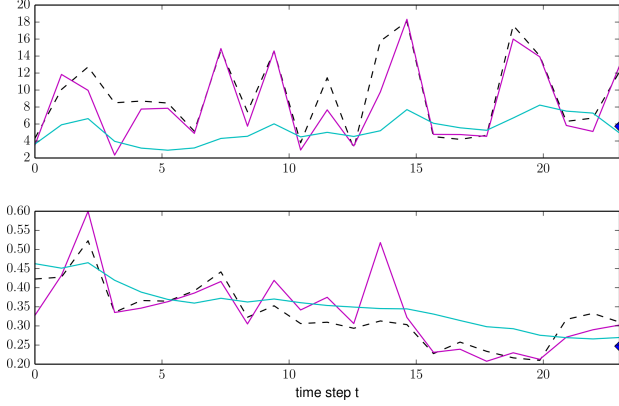
(a) horizontal semicircle
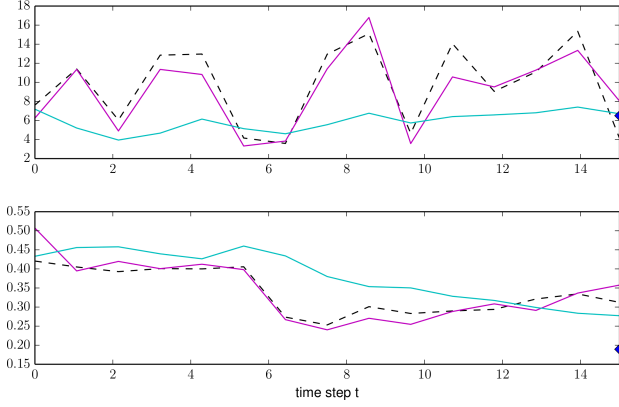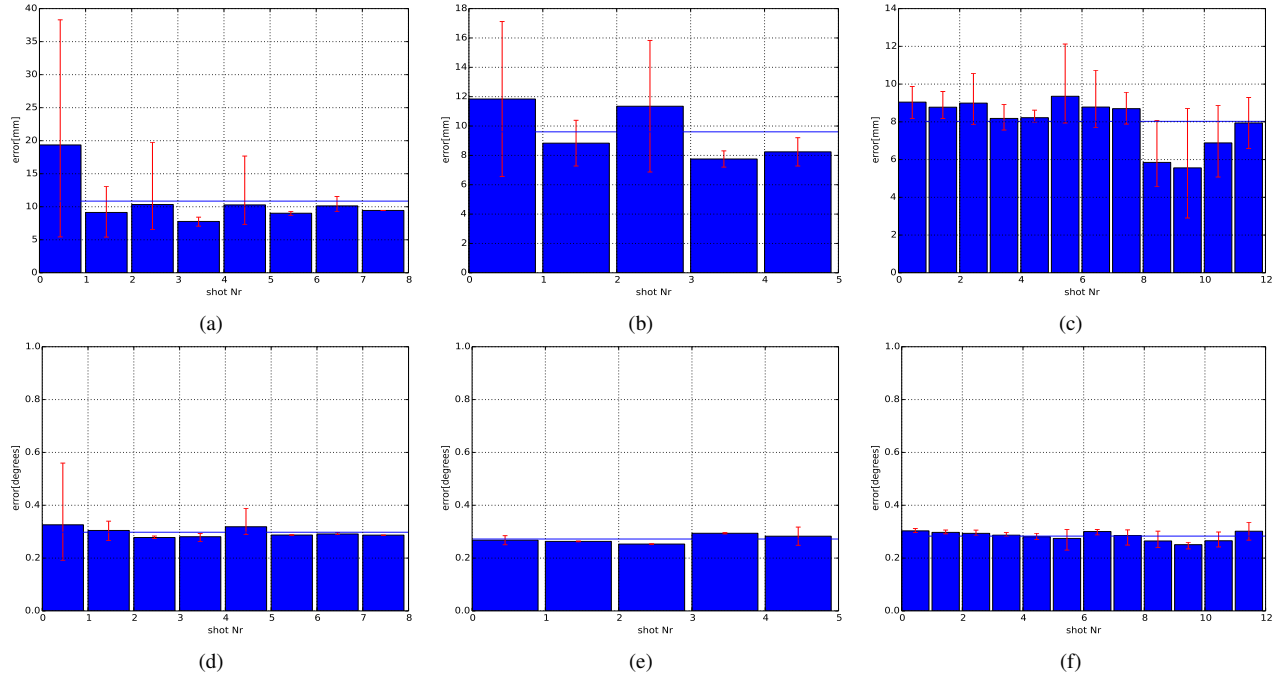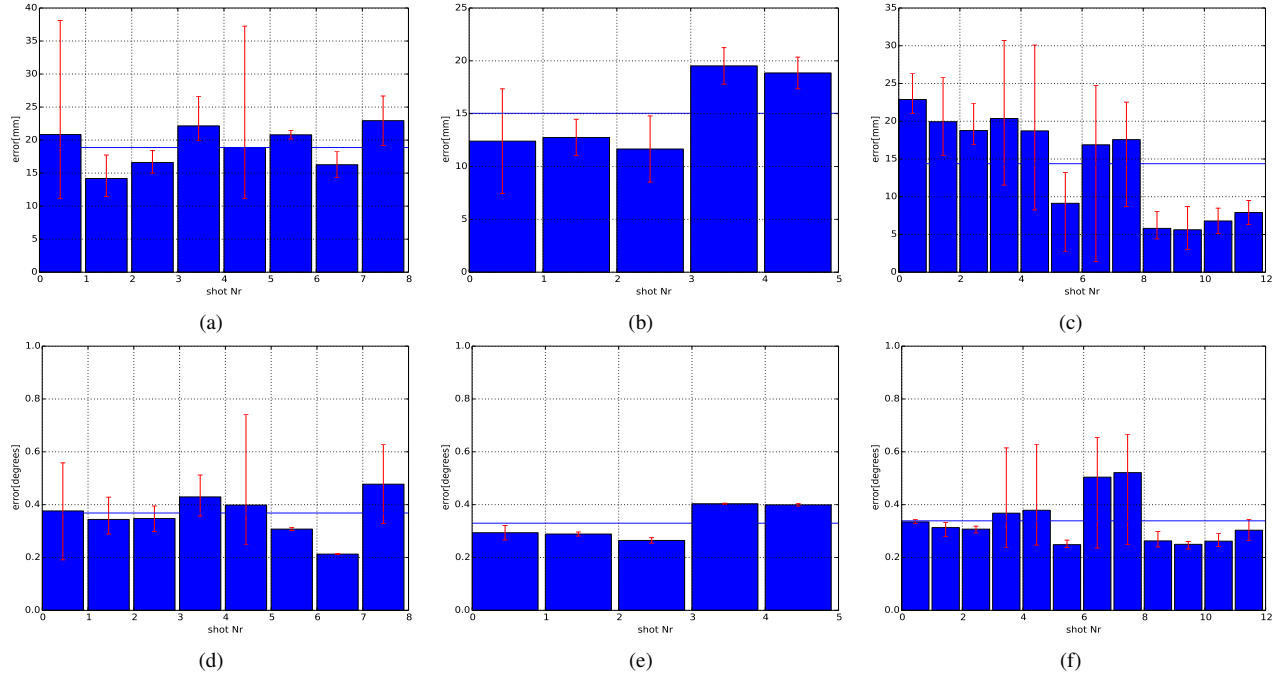

(b) vertical semicircle
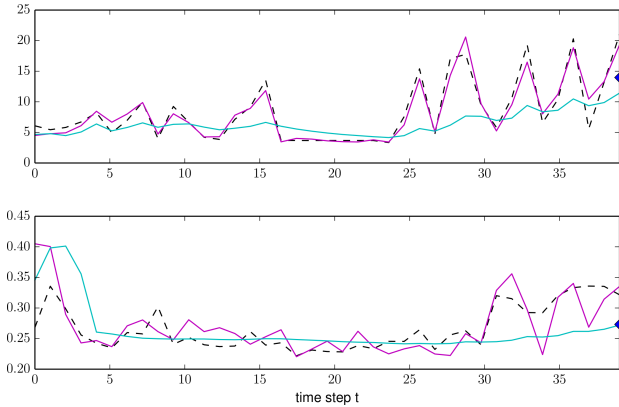
(c) linear motion

Fig. 5. Translational errors in mm (top part) and rotational errors in degrees (bottom part) of the raw detection (dashed black lines), of the estimated relative transformation (blue lines), and of the particle with the highest data likelihood (purple line) in each time step (and the MAP estimate for the last one), separately for the three types of motion.

In our case, we can guarantee a precise hand-eye calibration because of the usage of a laser-tracker in the calibration stage. In cases where this is not possible, and the information resulting from the robot's inverse kinematics is not sufficiently precise, visual servoing is usually performed [37]. This however is computationally complex, requires a line of sight continuously during the operation and can cause some stability problems [38].

We plan to extend this work in several directions: One interesting experiment would be to obtain dynamic measurements instead of the static measurements. This allows to make statements about the accuracy which are closer to the reality in industrial environments. A further line of research is to enhance the precision and robustness (to e.g. camera occlusions) by fusing information from additional sensors (e.g. acceleration sensors) and/or the robot kinematics. Another improvement would be to utilize SLAM based methods in the marker calibration step, so instead of manually measuring every marker, a SLAM based approach could be used to automatically estimate the markers' positions. We wanted to focus on very precise pose estimations and therefore used the measurements obtained by the laser tracker. Finally, the two presented approaches, integration over time and over multiple tags, could be fused, s.t. the particle filter uses the RANSAC solution as a fused measurement, and the reprojection error as weighting the particles.

Fig. 6. Mean of the translational (top row) and the rotational error (bottom row) for the pose estimation based on Least Squares of all visible markers in mm/degrees, for
(a), (d) : horizontal semicircle
(b), (e) : vertical semicircle
(c), (f) : linear motion.
Note that the error bars depict the minimal and maximal error and the blue line is the total average over all measurements.

Fig. 7. Mean of the translation (top row) and the rotational error (bottom row) of the pose estimation based on RANSAC of all visible markers in mm/degree, for
(a), (d) : horizontal semicircle
(b), (e) : vertical semicircle
(c), (f) : linear motion.
Note that the error bars depict the minimal and maximal error and the blue line is the total average over all measurements.

REFERENCES

[1] J. Corum, "Basic properties of reference crossply carbon-fiber composite," Oak Ridge National Lab., TN (US), Tech. Rep., 2001.

[2] R. G. Boeman and N. L. Johnson, "Development of a cost competitive, composite intensive, body-in-white," SAE Technical Paper, Tech. Rep., 2002.

[3] P.-J. Tsai, H.-Y. Shieh, W.-J. Lee, and S.-O. Lai, "Health-risk assessment for workers exposed to polycyclic aromatic hydrocarbons (pahs) in a carbon black manufacturing industry," *Science of the total environment*, vol. 278, no. 1, pp. 137–150, 2001.

[4] C. Tomasi and T. Kanade, "Detection and tracking of point features," Carnegie Mellon University Technical Report CMU-CS-91-132, Tech. Rep., 1991.

[5] J. Shi and C. Tomasi, "Good features to track," in *1994 IEEE Conference on Computer Vision and Pattern Recognition (CVPR'94)*, 1994, pp. 593 – 600.

[6] J. Y. Bouguet, "Pyramidal Implementation of the Lucas Kanade Feature Tracker: Description of the algorithm," Jean-Yves Bouguet, 2002.

[7] P. Azad, D. Munch, T. Asfour, and R. Dillmann, "6-DOF Model-based Tracking of Arbitrarily Shaped 3D Objects," in *ICRA*, 2011.

[8] M. Ulrich, C. Wiedemann, and C. Steger, "Cad-based recognition of 3d objects in monocular images," in *International Conference on Robotics and Automation*, 2009, pp. 1191–1198.

[9] D. G. Lowe, "Object recognition from local scale-invariant features," in *Computer vision, 1999. The proceedings of the seventh IEEE international conference on*, vol. 2. Ieee, 1999, pp. 1150–1157.

[10] H. Bay, T. Tuytelaars, and L. Van Gool, "Surf: Speeded up robust features," in *Computer vision–ECCV 2006*. Springer, 2006, pp. 404–417.

[11] M. Fiala, "Artag, a fiducial marker system using digital techniques," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 2. IEEE, 2005, pp. 590–596.

[12] 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.

[13] C. Feng and V. R. Kamat, "Augmented reality markers as spatial indices for indoor mobile aecfm applications," in *Proceedings of 12th international conference on construction applications of virtual reality (CONVR 2012)*, 2012, pp. 235–24.

[14] S. Dong, A. H. Behzadan, F. Chen, and V. R. Kamat, "Collaborative visualization of engineering processes using tabletop augmented reality," *Advances in Engineering Software*, vol. 55, pp. 45–55, 2013.

[15] A. Richardson, J. Strom, and E. Olson, "Aprilcal: Assisted and repeatable camera calibration," in *Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on*. IEEE, 2013, pp. 1814–1821.

[16] J. Sturm, W. Burgard, and D. Cremers, "Evaluating Egomotion and Structure-from-Motion Approaches Using the TUM RGB-D Benchmark," in *IEEE International Conference on Robotics and Automation (ICRA)*, Oct. 2012.

[17] S. Thrun, W. Burgard, and D. Fox, *Probabilistic Robotics (Intelligent Robotics and Autonomous Agents)*. The MIT Press, 2005.

[18] C. Andrieu, N. de Freitas, A. Doucet, and M. Jordan, "An Introduction to MCMC for Machine Learning," pp. 5–43, 2003.

[19] O. Cappé, S. J. Godsill, and E. Moulines, "An Overview of Existing Methods and Recent Advances in Sequential Monte Carlo," *Proceedings of the IEEE*, vol. 95, no. 5, pp. 899–924, May 2007.

[20] A. Doucet, S. Godsill, and C. Andrieu, "On Sequential Monte Carlo Sampling Methods for Bayesian Filtering," *Statistics and Computing*, vol. 10, no. 3, pp. 197–208, 2000.

[21] D. Fox, "Kld-sampling: Adaptive particle filters," in *Advances in Neural Information Processing Systems 14*. MIT Press, 2001.

[22] N. Kwak, I. K. Kim, H. C. Lee, and B. H. Lee, "Analysis of Resampling Process for the Particle Depletion Problem in FastSLAM," *Robot and Human interactive Communication, 2007. RO-MAN 2007. The 16th IEEE International Symposium on*, pp. 200–205, Aug. 2007.

[23] A. Aldoma, Z.-C. Marton, F. Tombari, W. Wohlkinger, C. Potthast, B. Zeisl, R. B. Rusu, S. Gedikli, and M. Vincze, "Tutorial: Point cloud library: Three-dimensional object recognition and 6 dof pose estimation," *Robotics & Automation Magazine, IEEE*, vol. 19, no. 3, pp. 80–91, 2012.

[24] U. Hillenbrand, "Consistent parameter clustering: Definition and analysis," *Pattern Recognition Letters*, vol. 28, pp. 1112–1122, Jul. 2007.

[25] B. Drost, M. Ulrich, N. Navab, and S. Ilic, "Model globally, match locally: Efficient and robust 3d object recognition," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2010, pp. 998–1005.

[26] F. Tombari and L. D. Stefano, "Hough voting for 3d object recognition under occlusion and clutter," *IPSJ Transactions on Computer Vision and Applications*, vol. 4, pp. 20–29, 2012.

[27] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[28] C. Chen, Y. Hung, and J. Cheng, "RANSAC-based DARCES: A New Approach to Fast Automatic Registration of Partially Overlapping Range Images," *IEEE Transactions PAMI*, vol. 21, pp. 1229–1234, 1999.

[29] R. B. Rusu, N. Blodow, and M. Beetz, "Fast Point Feature Histograms (FPFH) for 3D Registration," in *IEEE International Conference on Robotics and Automation (ICRA)*, May 2009, pp. 3212–3217.

[30] G. Braun, C. Nissler, and F. Krebs, "Development of a vision-based 6d pose estimation end effector for industrial manipulators in lightweight production environments," in *Emerging Technologies & Factory Automation (ETFA), 2015 IEEE 20th Conference on*. IEEE, 2015, pp. 1–6.

[31] Leica Geosystems, "PCMM System Specification," Tech. Rep., 2013.

[32] 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.

[33] K. Strobl and G. Hirzinger, "More accurate camera and hand-eye calibrations with unknown grid pattern dimensions," in *IEEE International Conference on Robotics and Automation, 2008. ICRA 2008.*, May 2008, pp. 1398–1405.

[34] K. H. Strobl, W. Sepp, S. Fuchs, C. Paredes, and K. Arbter. (2010, July) DLR CalDe and DLR CalLab. Institute of Robotics and Mechatronics, German Aerospace Center (DLR). Oberpfaffenhofen, Germany. [Online]. Available: http://www.robotic.dlr.de/callab/

[35] C. Bingham, "An Antipodally Symmetric Distribution on the Sphere," *Ann. Statist.*, vol. 2, no. 6, pp. 1201–1225, 11 1974. [Online]. Available: http://dx.doi.org/10.1214/aos/1176342874

[36] I. Sharf, A. Wolf, and M. Rubin, "Arithmetic and geometric solutions for average rigid-body rotation," *Mechanism and Machine Theory*, vol. 45, no. 9, pp. 1239 – 1251, 2010.

[37] 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.

[38] 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.