

# Fixture calibration with guaranteed bounds from a few correspondence-free surface points

Rasmus Laurvig Haugaard, Yitaek Kim and Thorbjørn Mosekjær Iversen

**Abstract**—Calibration of fixtures in robotic work cells is essential but also time consuming and error-prone, and poor calibration can easily lead to wasted debugging time in downstream tasks. Contact-based calibration methods let the user measure points on the fixture’s surface with a tool tip attached to the robot’s end effector. Most such methods require the user to manually annotate correspondences on the CAD model, however, this is error-prone and a cumbersome user experience. We propose a correspondence-free alternative: The user simply measures a few points from the fixture’s surface, and our method provides a tight superset of the poses which could explain the measured points. This naturally detects ambiguities related to symmetry and uninformative points and conveys this uncertainty to the user. Perhaps more importantly, it provides *guaranteed* bounds on the pose. The computation of such bounds is made tractable by the use of a hierarchical grid on SE(3). Our method is evaluated both in simulation and on a real collaborative robot, showing great potential for easier and less error-prone fixture calibration. [sites.google.com/view/ttpose](https://sites.google.com/view/ttpose)

## I. INTRODUCTION

The use of robots in industrial production promises high flexibility, however in practice, this flexibility is realized through reconfiguration of robotic work cells, which in many cases requires expert knowledge. For small and medium enterprises, which often lack in-house robotics experts, the promised flexibility can only be achieved if the reconfiguration process is fast and easily done by non-experts.

An essential part of the configuration process is the calibration of the positions and orientations, also referred to as poses, of the fixtures in the work cell. Calibrating fixture poses can enable transfer of fixture-related skills which have either been programmed or demonstrated in another configuration. Accurate calibration with consumer measuring devices, like rulers, is near-infeasible, and vision based solutions require additional expensive equipment and setup. A promising alternative is contact-based pose estimation, in which the user moves a tool tip, which is attached to the robot’s end effector, to various points on the surface of a fixture. The surface point coordinates are measured in the robot’s frame of reference using the robot’s forward kinematics and the sampled points are then used to infer the fixture’s pose relative to the robot.

Existing contact-based methods rely on correspondences to infer the pose. Most methods require the user to manually

All authors are from SDU Robotics, Maersk Mc-Kinney Møller Institute, University of Southern Denmark. The authors gratefully acknowledge the support from Innovation Fund Denmark through the project MADE Fast. {rlha, yik, thmi}@mmmi.sdu.dk

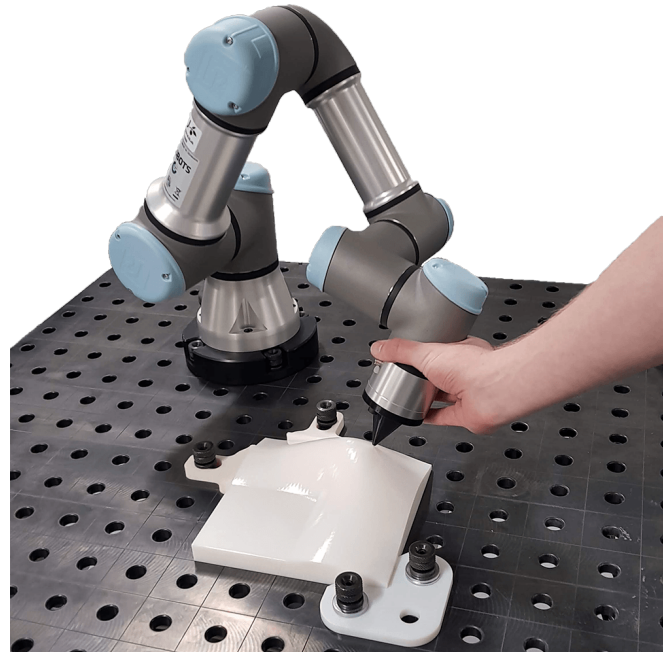


Fig. 1. A user moves a robot to measure a few points on a fixture using a calibrated tool-tip attached to the end-effector of the robot. The tool consist of a 3D-printed mount with a 3 mm steel ball at the tip. The fixture is a 3D-printed MATLAB logo, fixated to the table using additional fixtures.

annotate these correspondences, either in form of point-to-point, point-to-line and/or point-to-plane correspondences. Repeatedly going back and forth between demonstrating points with the robot and annotating correspondences in a computer program is not only a cumbersome user experience; It also requires the user to identify features on the fixture which the user can measure with low uncertainty. This process results in uncertainties that are primarily related to human ability, and it is thus difficult to assume a maximum correspondence error, especially for non-expert users.

In this work, we greatly simplify the user experience by not requiring correspondences. The user is only required to measure a few correspondence-free points on the object surface. The user thus *only* has to make sure, that the measured points are in fact from the fixture’s surface and not, for example, from another object; so it is reasonable to assume this to hold, even for non-expert users. The measurement uncertainties are thus restricted to forward kinematics and tool tip calibration, and a conservative estimate of the maximum error can be determined experimentally.

Given a CAD model and surface measurements with bounded errors, our method finds a tight superset of the set

of poses which could explain the measured surface points. The superset is *guaranteed* to include the true pose, and we can thus infer bounds on positional and rotational error.

Our method consists of a search through a hierarchical grid on the pose space,  $SE(3)$ , recursively expanding cells to obtain a finer pose discretization and discarding pose cells which can be guaranteed to not include the true pose. The result is a discrete set of pose cells of which one of them encompasses the true pose. This way, our method naturally detects ambiguities, irrespective of whether they arise from a lack of informative measurement points or object symmetries. This ensures that non-experts are able to easily perform reliable fixture calibrations.

Additionally, we derive the likelihoods of poses within the pose superset, enabling confidence intervals which are tighter than the bounds. The combination of bounds, confidence intervals and pose distributions makes our method relevant both in situations where a single fixture pose is needed with guaranteed or low-uncertainty accuracy, but also for potential future work where the expressive pose distributions open the possibility for downstream tasks to actively account for calibration uncertainties.

## II. RELATED WORK

Fixture calibration can be done in many ways. The two primary approaches are vision-based, where a vision system is installed and calibrated in the cell; and contact-based, where the robot's internal encoders and forward kinematics are used to measure points on the fixture's surface.

Most contact-based methods assume that the user provides correspondences between measured surface points and features on the CAD model. Some of the methods are tailored to specific sets of correspondences. For example, [1] provides a closed-form least squares solution, when exactly three point-to-point correspondences are available, and [2] proposes a calibration method, which also requires exactly three correspondences, but consisting of one point-to-point, one point-to-line and one point-to-plane correspondence.

Other methods represent more general approaches. [3] provides a closed-form least-squares solution for three or more point-to-point correspondences. [4] extends this approach to allow point-to-line and point-to-plane correspondences using an iterative branch and bound algorithm to avoid getting stuck in local minima, and [5], [6] improves on this approach using Lagrangian relaxation.

While the combination of the three correspondence-types do reflect a degree of generality and flexibility, the methods still requires the user to provide accurate correspondences. For objects consisting only of smoothly varying surfaces, none of the three correspondence types are suitable, and while the user might attempt to provide correspondences in such cases, they would be associated with significant and unknown uncertainties. Even for objects, where the three correspondence types are suitable, there may still be a not insignificant amount of uncertainty related to the correspondences. How accurately can a human mark a point-to-point or point-to-line correspondence? Also, for near-

symmetrical objects or objects with other ambiguities, the user might unknowingly provide correspondences which are entirely wrong.

If dense surface samples are available, local 3D features can be utilized to estimate the pose [7]–[9] without requiring the user to annotate correspondences, however, moving a tool-tip across a surface provides a curve - not a surface, so methods which require dense samples are not directly applicable to contact-based pose estimation.

Recent work [10] proposes and validates a framework for automatically finding point correspondences using point pair features [11] on sampled curves on the object surface. The user does thus not have to manually annotate correspondences, and they show, that they are robust to correspondence outliers, however, they do not provide any estimate of the uncertainty of the point estimate they obtain.

To the best of our knowledge, our method is the first fully correspondence-free, contact-based pose estimation method, and the first to provide guaranteed bounds on  $SE(3)$ .

There are also vision-based approaches to fixture calibration. If depth sensors are available, the aforementioned methods using dense surface samples can be used. However, since the obtained point clouds also contain background points, it is not clear how any guarantees can be made for the resulting pose, even with guarantees for the depth errors.

Methods using color images typically either estimate correspondences as part of a pose estimation method [12], [13], use strategies that compute similarities between the observed vision data and a vast set of hypotheses [14]–[16], or directly regress object poses using deep learning [17]–[19].

For the methodologies described above, it is by far most common to provide a single best estimate of the pose, and while state of the art single pose estimation methods [12], [18], [19] achieve impressive results on standard pose estimation benchmarks [20], they are unable to express uncertainties arising from visual ambiguities such as those imposed by object symmetries.

Recently, there has been an increased interest in rotation and pose distribution estimation, defined as the estimation of a probability density function on  $SO(3)$  and  $SE(3)$ , respectively. State of the art methods for both rotation [21]–[23] and full pose [24] distribution estimation provide a function for estimating the unnormalized likelihood of a pose hypothesis, and estimate the relative likelihoods for a large amount of samples on  $SO(3)$  or  $SE(3)$ . This approach has resulted in highly expressive, non-parametric multi-modal distributions on  $SO(3)$  which with a hierarchical grid, proposed in [21], [24], has been expanded to work on  $SE(3)$ . Because of the inherent complexity in natural images, however, to the best of our knowledge, no vision-based pose estimation method has provided guaranteed bounds on pose estimates.

A hierarchical grid on  $SE(3)$  is also at the core of our proposed method, however, we use it in the setting of contact-based pose estimation, where it is reasonable to assume bounds on sampling errors, and we show, that we can propagate these bounds to pose space.

### III. METHODS

Our method is based on hierarchical pose grids on SE(3). For each level in the grid, we can calculate bounds on the maximum point-to-mesh distance which can be caused by the discretization error in the grid. Poses with at least one point-to-mesh distance which cannot be explained by discretization and sampling noise can then be discarded, and only the remaining poses proceed to the next level in the hierarchical grid of higher pose resolution.

#### A. Problem Formalization

Let  $P = \{p_1, p_2, \dots, p_n\}$  be a set of surface points relative to a robot's base frame, obtained by probing the surface of a fixture with a calibrated tool-tip attached to the end-effector of the robot. Due to various uncertainties in the forward kinematics from the base of the robot to the tool-tip, the obtained points are not exactly on the surface of the fixture. Let  $p_i^*$  be the point on the fixture's surface which is closest to  $p_i$ :  $p_i^* = \arg \min_{p \in \mathcal{S}} \|p - p_i\|$ , where  $\mathcal{S}$  denotes the fixture surface. The point-to-mesh error is then defined as  $e_i = \|p_i - p_i^*\|$ . Given a bound on the point-to-mesh error,  $e_i \leq b \forall i$ , there exists a unique set of poses which satisfy the point-to-mesh constraints:

$$\mathcal{X} = \{x \in \text{SE}(3) \mid D(\mathcal{S}, x, p_i) \leq b \quad \forall i\}, \quad (1)$$

where  $D(\mathcal{S}, x, p_i)$  is the point-to-mesh distance from the point  $p_i$  to the fixture surface  $\mathcal{S}$  given the pose  $x$ .

We aim to find a tight superset  $\hat{\mathcal{X}} \supset \mathcal{X}$ , to be able to guarantee bounds on the true pose,  $x^\dagger = (R^\dagger, t^\dagger)$ .

#### B. Initial Positional Bound

First, an initial bound on the fixture's position is found. Without loss of generality, we define the fixture's origin to be at the center of the smallest sphere enclosing the fixture's mesh. We find this sphere from the fixture's 3D-model using the method described in [25]. The radius of the sphere will be referred to as the fixture radius,  $r$ .

It follows from the point-to-mesh constraints, that the distance from the sampled points to the fixture position,  $t^\dagger$ , is bounded by the sum of the fixture radius,  $r$ , and the maximum sample error,  $b_s$ .

Intuitively, the set of fixture positions satisfying these constraints is the intersection of spheres with radius  $r + b_s$  centered around each sampled point,  $p_i$ . Formally, the set is defined by

$$\mathcal{T} = \{t \in \mathbb{R}^3 \mid \|t - p_i\| \leq r + b_s \quad \forall i\}. \quad (2)$$

Finding the Axis-Aligned Bounding Box (AABB) of  $\mathcal{T}$  can be formalized as six Second-Order Cone Programs (SOCP):

$$\begin{aligned} \min_t \quad & c^\top t \\ \text{subject to} \quad & \|t - p_i\| \leq r + b_s \quad \forall i, \end{aligned} \quad (3)$$

with two programs per axis with varying  $c$ . Specifically, the AABB of  $\mathcal{T}$  is given by the AABB of the solutions to  $c \in \{\hat{i}, -\hat{i}, \hat{j}, -\hat{j}, \hat{k}, -\hat{k}\}$ , where  $\hat{i}, \hat{j}, \hat{k}$  are the orthonormal axis vectors. We use the SOCP solver in CVXOPT [26]. The

AABB provides an initial bound on the position, and the positional part of the hierarchical pose grid is initialized such that it encompasses the AABB.

#### C. Discretization error bound on the $\mathbb{R}^3$ grid

The positional part of the hierarchical SE(3) grid is an octree where each node represents a cube in a regular grid. For each recursion, each cube is divided into eight equal cubes. We let each cube be represented by its center point, and the largest distance from the center to any point in the cube is the distance to a cube corner,  $\sqrt{3}l/2$ , where  $l$  is the side length of the cube. Since  $l$  is halved at each recursion, we can write the bounds of the positional discretization error as  $b_p = \sqrt{3}l_0/2^{\tau+1}$ , where  $l_0$  is the side length of cubes at recursion zero, and  $\tau$  is the recursion level.

#### D. Discretization error bound on the SO(3) grid

For the rotational part of the grid, we use the HealPix grid [27] extended to SO(3) by [28]. Specifically, HealPix defines a grid on the two-sphere, which describes two of the three rotational degrees of freedom; azimuth and elevation. [28] proposes to take the cartesian product of the HealPix grid and a one-dimensional grid, describing the last rotational degree of freedom: Tilt.

HealPix provides the largest angular distance from the center of a rotation cell to any rotation in the cell, however, the angular distance incurred by tilt must also be taken into account. Let  $\theta$  be the largest angular distance from HealPix, and let  $\phi$  be maximum tilt distance, which is simply half of the tilt resolution. The maximum angle between a rotation cell center and any rotation in the cell is then given by

$$\gamma = \max_{\alpha \in [0, \theta], \beta \in [0, \phi]} \cos^{-1} \left[ \frac{\text{tr}[R_x(\alpha)R_z(\beta)] - 1}{2} \right] \quad (4)$$

$$= \max_{\alpha \in [0, \theta], \beta \in [0, \phi]} \cos^{-1} \left[ \frac{C_\beta + C_\alpha C_\beta + C_\alpha - 1}{2} \right] \quad (5)$$

$$= \cos^{-1} \left[ \frac{C_{\beta^*} + C_{\alpha^*} C_{\beta^*} + C_{\alpha^*} - 1}{2} \right], \quad (6)$$

where  $C_\bullet = \cos \bullet$ ,  $\bullet^* = \min(\bullet, \pi)$ , and

$$R_x(\alpha)R_z(\beta) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & C_\alpha & -S_\alpha \\ 0 & S_\alpha & C_\alpha \end{bmatrix} \begin{bmatrix} C_\beta & -S_\beta & 0 \\ S_\beta & C_\beta & 0 \\ 0 & 0 & 1 \end{bmatrix}. \quad (7)$$

Note that while we use  $R_x$  here, it holds for rotations around any vector in the xy-plane. Eq. 6 provides a bound on the angular discretization error.

Recall the bound on the true fixture center, found in section III-B. Let  $\hat{t}$  denote the AABB center and let  $b_t$  denote half of the AABB diagonal, which is the maximum distance from  $\hat{t}$  to the true translation,  $t^\dagger$ . It follows that the distance between the true fixture center,  $t^\dagger$ , and a given sample,  $p_i$  is bounded:  $\|t^\dagger - p_i\| \leq \|\hat{t} - p_i\| + b_t$ . Also, the distance between a point on the unit circle before and after a rotation of  $\gamma$  radians around the origin is  $\sqrt{2 - 2 \cos \gamma}$ . The sample-wise bounds on the point-to-mesh errors related to rotation discretization are thus given by

$$b_t^i = (\|\hat{t} - p_i\| + b_t) \sqrt{2 - 2 \cos \gamma}. \quad (8)$$

### E. Distance constraints

Given the bounds for the point-to-mesh errors related to discretization errors for the positional,  $b_p$ , and rotational,  $b_r$ , part of the SE(3) grid, as well as sampling,  $b_s$ , and numerical precision,  $b_\epsilon$ , we can provide total bounds on the point-to-mesh errors, as required in Eq. 1:

$$b^i = b_p + b_r^i + b_s + b_\epsilon. \quad (9)$$

Given a pose hypothesis, the sampled surface points are transformed into the fixture frame, and the nearest distance to the mesh can be computed in sub-linear time using a tree to represent bounding volumes of mesh triangles. We use Open3D's [29] implementation, built on Intel Embree [30]. If a pose cell has at least one point-to-mesh distance larger than the bound, the whole pose cell can be rejected.

### F. Expansion Strategy

The pose superset  $\hat{\mathcal{X}}$  is found by recursively expanding the SE(3) grid and rejecting cells using the criteria in Eq. 1. The lowest recursion of the grid consists of the cartesian product of the positional grid at recursion 0 (a single cube containing the AABB) and the rotational grid at recursion 0 (the 72 rotations corresponding to the lowest recursion of the SO(3) HealPix grid).

Our expansion strategy utilizes that the rotational and positional parts of the grid do not have to be expanded simultaneously. Instead, only the part of the grid which incurs the largest point-to-mesh error bounds are expanded: The rotational grid is expanded if  $b_r < b_p$  and vice versa.

In normal cases, this expansion continues recursively until the sampling bound,  $b_s$ , becomes the dominating term in Eq. 9. At this point very few of the expanded cells can be rejected causing the number of cells to increase exponentially.

### G. SE(3) Point Estimate with Guaranteed Bounds

The found superset of poses which explain the measured points,  $\hat{\mathcal{X}} \supset \mathcal{X}$ , describe complex bounds on the true pose. In case of multi-modal ambiguities, e.g. from a symmetric fixture, this ambiguity can be conveyed to the user, as will be discussed in section IV-A. However, when  $\mathcal{X}$  is uni-modal, it makes sense to provide a point estimate with guaranteed error bounds, to provide information which is more easily interpretable than the superset itself.

The positional point estimate is given by the center of the smallest sphere which encloses all possible positions in  $\hat{\mathcal{X}}$ . The positional error bound is given by the radius of the same sphere. For rotation, we approximate the center of rotations by the center of the smallest enclosing sphere of quaternion representations of the rotations in  $\hat{\mathcal{X}}$ , and then calculate the maximum angular distance from the estimated center to all rotations in  $\mathcal{X}$ .

### H. Estimation of Pose Likelihoods

The previous sections involved the computation of guaranteed bounds on the pose. When only the point-to-mesh distance bounds are used, the resulting pose bounds do not necessarily improve significantly with more samples. In

consequence, large parts of the superset,  $\hat{\mathcal{X}}$ , while possible, may be highly unlikely given the surface measurements. For this reason, we also consider a probability distribution over the superset of poses, with the aim of determining confidence intervals which are tighter than the bounds.

The following derives the relationship between the surface point measurements,  $P = \{p_1, p_2, \dots, p_n\}$ , and the pose distribution  $p(x|P)$ . The relationship between the surface measurements and the pose can be expressed using Bayes theorem:  $p(x|P) = p(P|x)p(x)p(P)^{-1}$ . Assuming no prior knowledge of the pose, both  $p(x)$  and  $p(P)$  are constants. It is also reasonable to assume that the surface point measurements are conditionally independent given the pose,  $x$ , which leads to the expression:

$$p(x|P) \propto \prod_{p_i \in P} p(p_i|x). \quad (10)$$

Using the law of total probability, the distribution corresponding to a single measurement can be factorized as

$$p(p_i|x) = \int_{s_i \in \mathcal{S}} p(p_i|x, s_i)p(s_i|x)ds_i, \quad (11)$$

where  $\mathcal{S}$  is the fixture surface and  $s_i$  is the unknown true surface point for which  $p_i$  was measured.

If there is no prior knowledge of what the true surface point is,  $p(s_i|x)$  can be treated as a uniform distribution over the fixture surface, and consequently as constant in the integral of Eq. 11. If we assume the measured points to be normally distributed,  $p_i = s_i + \epsilon_i$ , with  $\epsilon_i \sim \mathcal{N}(0, \Sigma)$ , it follows that  $p_i$  and  $x$  are conditionally independent given  $s_i$ :  $p(p_i|x, s_i) = p(p_i|s_i) = \mathcal{N}(p_i; s_i, \Sigma)$ . Substituting this into Eq. 11 yields

$$p(p_i|x) = \int_{s_i \in \mathcal{S}} \mathcal{N}(p_i; s_i, \Sigma)ds_i. \quad (12)$$

The integral in Eq. 12 is expensive to compute. Fortunately, the measuring error of a robot tend to be much smaller than the size of the fixture which is to be calibrated. In this work, the measuring error is assumed to be isotropic, i.e.  $\Sigma = \sigma^2 I$ , so for most surface points  $\mathcal{N}(p_i; s_i, \Sigma) \approx 0$ . Since the major contribution to the value of the integral will be for the surface points closest to  $p_i$ , we choose to approximate the integral by the normal distribution corresponding to the closest point,  $p_i^*$ , given the pose,  $x$ . This leads to the following approximation of the likelihood:

$$p(p_i|x) \approx \mathcal{N}(p_i; p_i^*, \sigma^2 I) \propto \exp\left(-\frac{\|p_i - p_i^*\|^2}{2\sigma^2}\right). \quad (13)$$

Since  $p_i^*$  can be found efficiently (see section III-C), substituting Eq. 13 into Eq. 10 provides a relative pose likelihood function which is efficient to compute:

$$p(x|P) \propto \prod_{p_i \in P} \exp\left(-\frac{\|p_i - p_i^*\|^2}{2\sigma^2}\right). \quad (14)$$

### I. Confidence Intervals

We obtain a discrete pose distribution by stratified Monte Carlo samples, sampling  $k$  poses uniformly within each cell of  $\hat{\mathcal{X}}$ . Specifically, the relative probabilities are estimated using Eq. 14 and normalized over the samples. The result is a discrete set of poses,  $x_i \in \tilde{\mathcal{X}}$ , with associated, estimated probabilities,  $p(x_i|P)$ .

The availability of such a distribution motivates a new point estimate of the pose, namely the expected pose. We compute the expected position and rotation as

$$\hat{t} = \sum_{x_i \in \tilde{\mathcal{X}}} p(x_i|P)t_i \quad \text{and} \quad \hat{q} = \sum_{x_i \in \tilde{\mathcal{X}}} p(x_i|P)q_i, \quad (15)$$

respectively, where  $t_i$  is the translation of pose  $x_i$ ,  $q_i$  is a quaternion representation of the rotation in  $x_i$ , and all  $q_i$ 's are mapped to the same quaternion hemisphere. This would cause problems with very large rotational ambiguities such as in case of symmetries, however, then the confidence intervals are very large anyway. Normalization of  $\hat{q}$  is left out for clarity. Confidence intervals are then found as the error for which the desired amount of probability mass is within that error from the expected pose.

## IV. EXPERIMENTS

We evaluate our method both in simulation and on the robot shown in Figure 1. To evaluate the ability to represent potential ambiguities, we run experiments on objects of the SYMSOL [21] dataset with varying degrees of symmetry. We also evaluate bounds and confidence intervals in simulation, where the true pose is known. On the real system, one expert as well as three non-experts, to demonstrate ease of use, measure points on the MATLAB logo shown in Figure 1.

All experiments are run on a laptop with an Intel i7-8650U 1.90GHz CPU and 16GB RAM. We allow up to  $10^7$  cells when expanding the pose grid, and each experiment takes between a few seconds and two minutes, depending on geometry and how informative the sampled points are.

### A. Simulation

To obtain simulation results, we simulate the manual surface sampling procedure. First, we sample a large number of uniform samples from the object surface. Second, we reduce the samples down to a small number, using farthest point sampling. In our experiments, we reduce 1.000 uniform samples to 10 surface samples. Finally, based on the sampling error bound,  $b_s$ , we add noise,  $e$ , to the samples. Specifically,  $e$  is sampled from a normal distribution, which is truncated at  $\|e\| = b_s$ , and with covariance  $(0.3b_s)^2 I$ . For the simulation experiments, we normalize the meshes to have an enclosing sphere with a diameter of 25 cm to mimic typical fixture sizes. Examples of simulated, sampled points are shown in Figure 3.

To present our method's ability to provide multi-modal supersets,  $\hat{\mathcal{X}}$ , we apply our method on objects from the SYMSOL [21] dataset with varying degrees of symmetry. The rotational part of the supersets is shown in Figure 3.

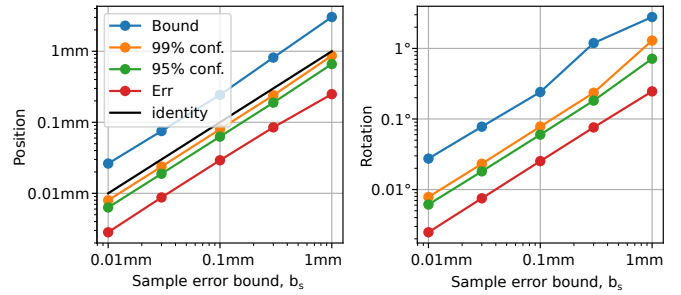


Fig. 2. Pose uncertainty for simulations on the MATLAB logo with varying sample error bounds. The true error for the expected pose (Eq. 15) is shown in red and is within the confidence intervals and bounds. The pose uncertainties decrease proportionally with the sample error bound.

When the superset is uni-modal, the bounds and confidence intervals which can be obtained, become of interest. To this end, we simulate experiments on the MATLAB logo, shown in Figure 1, with varying sample error bounds. The results are shown in Figure 2. The positional errors of the expected poses are lower than the sample error bounds and well below both the confidence intervals and pose bounds. We observe similar results for other fixtures.

### B. Real

For the experiments on the real system, we first propose a tool to measure points and provide methods for calibrating the tool tip as well as for estimating the sample error bound. Then we run experiments with one expert and three non-experts to demonstrate ease of use of the proposed method.

To make sure there is contact between tool tip and surface at the time of point sample collection, samples are collected by positioning the tool tip close to a surface, and letting the robot tap the surface with a force-based controller.

1) *Tool design*: We design a low-cost point measuring tool which can be seen in Figure 1 and 4. The tool consists of an 3D-printed cone and an off-the-shelf 3 mm steel ball. The cone is attached to the robot's end effector, and an indent at the tip of the cone allows the steel ball to be glued firmly onto the cone. This way, we get the flexibility of 3D-printing, for easy incorporation in tool designs, and the tolerances and wear-resistance of a steel ball.

2) *Tip calibration*: We first obtain a coarse calibration by approximately rotating the tip around a fixed point with respect to the robot's base frame and record robot configurations to solve the set of linear equations with least squares:

$$\{T_{\text{base}}^{\text{tcp}}\}_i p_{\text{tcp}}^{\text{tip}} = p_{\text{base}}^{\text{tip}} \quad \forall i, \quad (16)$$

where  $p_{\text{tcp}}^{\text{tip}}$  and  $p_{\text{base}}^{\text{tip}}$  are constant but unknown. A coarse table normal calibration can then be obtained by fitting a plane to three non-colinear table points.

From the coarse tip- and table calibrations, the robot can automatically gather a larger dataset. To get an accurate estimate of the table's normal, a set of points on the table is collected with a fixed tcp orientation. Specifically, we subtract the centroid from the table points, apply SVD and

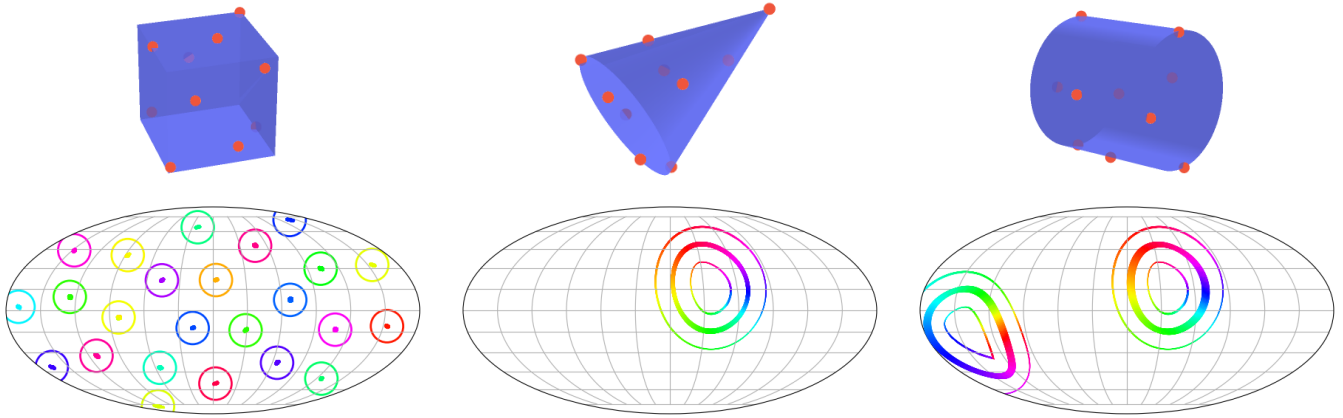


Fig. 3. Qualitative pose estimation results of three different objects from the SYMSOL [21] dataset. From left to right: A cube, a cone and a cylinder. Top: The objects are rendered semi-transparently in blue together with ten simulated surface samples with noise,  $b_s = 0.3\text{mm}$ , in red. From the CAD model and noisy surface samples, our method provides a set of poses which are *guaranteed* to include the true pose. Bottom: We show the rotation part of the pose set with two rotational dimensions shown by the position on a sphere, shown in a Mollweide plot, and the last dimension visualized by color. The true rotation up to symmetry is shown by circles for discrete symmetries (cube) and "stroked" circles for continuous symmetries (cone and cylinder). Our method naturally captures all the modes without any prior knowledge about symmetries, and this ambiguity can be conveyed to a user.

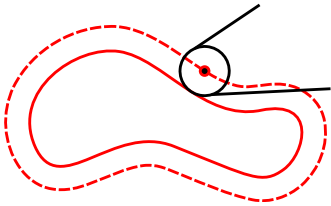


Fig. 4. The tool-tip with the steel ball is shown in black. The fixture is shown in red, and the Minkowski sum of the steel ball and the fixture is illustrated by a dashed, red line. The center of the steel ball lies on the Minkowski sum surface independently of the angle of sampling.

use the vector corresponding to the least singular value to obtain  $\hat{n}_{\text{base}}$ , which is independent of the coarse calibration.

Then, to obtain a better calibration of the tip, we obtain a new dataset with varying orientations, providing the following equations, also solved with least squares:

$$\hat{n}_{\text{base}} \{T_{\text{base}}^{\text{tcp}}\}_i p_{\text{tcp}}^{\text{tip}} = k \quad \forall i, \quad (17)$$

where  $p_{\text{tcp}}^{\text{tip}}$  and  $k$  are constant and unknown.

This calibration provides the center of the steel ball, and it turns out that this is desirable. When sampling points on the fixture surface, it is not clear which point on the steel ball is in contact with the surface, however, it is clear, that the center of the steel ball lies on the Minkowski sum [31] of the fixture and the steel ball. See Figure 4. Since the steel ball and fixture geometries are known, we can compute this Minkowski sum. We use the CGAL implementation [32]. We set the sample error bound above the largest residual observed in Eq. 17:  $b_s = 1.0\text{ mm}$ .

3) *User Study*: We gave three students, which were unfamiliar with our method, a 1-minute explanation of the task and asked them to sample 15 surface points on the setup shown in Figure 1. The resulting positional bounds were 3.5, 2.7 and 3.1 mm, and the rotational bounds were 2.7°, 1.6° and 3.9°. The 99% confidence intervals were 0.80, 0.81 and 0.83 mm for position and 0.73°, 0.50° and 0.99° for rotation.

The expected poses were all within 0.87 mm and 0.74° of each other. The expert obtained bounds of 1.6 mm and 1.2°, and 99% confidence intervals of 0.39 mm and 0.26°.

## V. LIMITATIONS AND FUTURE WORK

One limitation of our method is the assumption of guaranteed bounds on the sample error. The sample error can be caused by everything from flex in the robot or fixture to bias in the robot's internal encoder - and setting the sample error bound too conservatively will lead to large pose bounds.

A current limitation is that the user may not know what points to sample to reduce the bounds. The user would obtain this intuition from trial and error, but identifying points which are informative, and showing them to the user, to build the intuition faster, would be an interesting future work.

This work considered calibration of a single fixture. However, the superset could be computed for several fixtures given the same samples, and since an empty pose set would indicate mesh-samples incompatibility, our method could be used for classification as well as calibration.

Finally, our method is applicable for any sensor where guaranteed bounds on surface measurements can be provided, potentially making our method relevant for use with vision sensors in certain cases.

## VI. CONCLUSION

This paper presented a novel, correspondence-free method for contact-based fixture pose calibration. The user simply has to measure a few points on the surface of the fixture, avoiding manual and error-prone correspondence annotations. The output of the method is a tight superset of the poses which could explain the measured points. The superset provides *guaranteed* bounds on the true pose and naturally represents ambiguities caused by symmetries and uninformative points. Our method was evaluated both in simulation and on a real, collaborative robot, showing great potential for easier and less error-prone fixture calibration.

## REFERENCES

- [1] E. Thompson, "An exact linear solution of the problem of absolute orientation," *Photogrammetria*, vol. 15, pp. 163–179, 1958.
- [2] W. Zhang, X. Ma, L. Cui, and Q. Chen, "3 points calibration method of part coordinates for arc welding robot," in *International Conference on Intelligent Robotics and Applications*, 10 2008, pp. 216–224.
- [3] B. K. Horn, H. M. Hilden, and S. Negahdaripour, "Closed-form solution of absolute orientation using orthonormal matrices," *JOSA A*, vol. 5, no. 7, pp. 1127–1135, 1988.
- [4] C. Olsson, F. Kahl, and M. Oskarsson, "The registration problem revisited: Optimal solutions from points, lines and planes," in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, vol. 1. IEEE, 2006, pp. 1206–1213.
- [5] C. Olsson and A. Eriksson, "Solving quadratically constrained geometrical problems using lagrangian duality," in *2008 19th International Conference on Pattern Recognition*, 2008, pp. 1–5.
- [6] J. Briaies and J. Gonzalez-Jimenez, "Convex global 3D registration with lagrangian duality," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 5612–5621.
- [7] C. S. Chua and R. Jarvis, "Point signatures: A new representation for 3d object recognition," *International Journal of Computer Vision*, vol. 25, pp. 63–85, 1997.
- [8] A. E. Johnson and M. Hebert, "Using spin images for efficient object recognition in cluttered 3d scenes," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 21, no. 5, pp. 433–449, 1999.
- [9] A. G. Buch, L. Kiforenko, and D. Kraft, "Rotational subgroup voting and pose clustering for robust 3d object recognition," in *2017 IEEE international conference on computer vision (ICCV)*. IEEE, 2017, pp. 4137–4145.
- [10] Y. Kim, A. Kramberger, A. G. Buch, and C. Sloth, "Contact-based pose estimation of workpieces for robotic setups," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*, 2023, pp. 12 324–12 330.
- [11] B. Drost, M. Ulrich, N. Navab, and S. Ilic, "Model globally, match locally: Efficient and robust 3D object recognition," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 07 2010, pp. 998–1005.
- [12] R. L. Haugaard and A. G. Buch, "Surfemb: Dense and continuous correspondence distributions for object pose estimation with learnt surface embeddings," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 6749–6758.
- [13] R. L. Haugaard and T. M. Iversen, "Multi-view object pose estimation from correspondence distributions and epipolar geometry," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2023, pp. 1786–1792.
- [14] S. Hinterstoisser, V. Lepetit, S. Ilic, S. Holzer, G. Bradski, K. Konolige, and N. Navab, "Model based training, detection and pose estimation of texture-less 3d objects in heavily cluttered scenes," in *Asian conference on computer vision (ACCV)*. Springer, 2012, pp. 548–562.
- [15] M. Ulrich, C. Wiedemann, and C. Steger, "Combining scale-space and similarity-based aspect graphs for fast 3d object recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 34, no. 10, pp. 1902–1914, 2012.
- [16] F. Hagelskjær, T. R. Savarimuthu, N. Krüger, and A. G. Buch, "Using spatial constraints for fast set-up of precise pose estimation in an industrial setting," in *2019 IEEE 15th International Conference on Automation Science and Engineering (CASE)*. IEEE, 2019, pp. 1308–1314.
- [17] Y. Xiang, T. Schmidt, V. Narayanan, and D. Fox, "Posecnn: A convolutional neural network for 6d object pose estimation in cluttered scenes," 2018.
- [18] Y. Labbé, J. Carpentier, M. Aubry, and J. Sivic, "Cosypose: Consistent multi-view multi-object 6d pose estimation," in *European Conference on Computer Vision*. Springer, 2020, pp. 574–591.
- [19] G. Wang, F. Manhardt, F. Tombari, and X. Ji, "GDR-Net: Geometry-guided direct regression network for monocular 6d object pose estimation," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2021, pp. 16 611–16 621.
- [20] T. Hodaň, M. Sundermeyer, B. Drost, Y. Labbé, E. Brachmann, F. Michel, C. Rother, and J. Matas, "Bop challenge 2020 on 6d object localization," in *European Conference on Computer Vision*. Springer, 2020, pp. 577–594.
- [21] K. A. Murphy, C. Esteves, V. Jampani, S. Ramalingam, and A. Makadia, "Implicit-pdf: Non-parametric representation of probability distributions on the rotation manifold," in *International Conference on Machine Learning*. PMLR, 2021, pp. 7882–7893.
- [22] T. M. Iversen, R. L. Haugaard, and A. G. Buch, "Ki-pode: Keypoint-based implicit pose distribution estimation of rigid objects," in *Proceedings of the British Machine Vision Conference (BMVC) 2022*, 2022.
- [23] T. Höfer, B. Kiefer, M. Messmer, and A. Zell, "Hyperposepdf-hypernetworks predicting the probability distribution on so (3)," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2023, pp. 2369–2379.
- [24] R. L. Haugaard, F. Hagelskjær, and T. M. Iversen, "Spyrpose: Se (3) pyramids for object pose distribution estimation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 2082–2091.
- [25] K. Fischer, B. Gärtner, and M. Kutz, "Fast smallest-enclosing-ball computation in high dimensions," in *Algorithms-ESA 2003: 11th Annual European Symposium, Budapest, Hungary, September 16-19, 2003. Proceedings 11*. Springer, 2003, pp. 630–641.
- [26] M. Andersen, J. Dahl, and L. Vandenbergh, "Cvxopt: Convex optimization," *Astrophysics Source Code Library*, pp. ascl-2008, 2020.
- [27] K. M. Górski, E. Hivon, A. J. Banday, B. D. Wandelt, F. K. Hansen, M. Reinecke, and M. Bartelmann, "Healpix: A framework for high-resolution discretization and fast analysis of data distributed on the sphere," *The Astrophysical Journal*, vol. 622, no. 2, p. 759, apr 2005. [Online]. Available: <https://dx.doi.org/10.1086/427976>
- [28] A. Yershova, S. Jain, S. M. Lavalle, and J. C. Mitchell, "Generating uniform incremental grids on so (3) using the hopf fibration," *The International journal of robotics research*, vol. 29, no. 7, pp. 801–812, 2010.
- [29] Q.-Y. Zhou, J. Park, and V. Koltun, "Open3D: A modern library for 3D data processing," *arXiv:1801.09847*, 2018.
- [30] I. Wald, S. Woop, C. Benthin, G. S. Johnson, and M. Ernst, "Embree: a kernel framework for efficient cpu ray tracing," *ACM Transactions on Graphics (TOG)*, vol. 33, no. 4, pp. 1–8, 2014.
- [31] H. Hadwiger, "Minkowskische addition und subtraktion beliebiger punktmengen und die theoreme von erhard schmidt," *Mathematische Zeitschrift*, vol. 53, no. 3, pp. 210–218, 1950.
- [32] P. Hachenberger, "3D minkowski sum of polyhedra," in *CGAL User and Reference Manual*, 5.6 ed. CGAL Editorial Board, 2023. [Online]. Available: <https://doc.cgal.org/5.6/Manual/packages.html#PkgMinkowskiSum3>