2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
October 14-18, 2024. Abu Dhabi, UAE

Sensor-agnostic Visuo-Tactile Robot Calibration Exploiting Assembly-Precision Model Geometries

Manuel Gomes[1*], Michael Görner[2], Miguel Oliveira[1], Jianwei Zhang[2]

***Abstract*— Visual sensor modalities dominate traditional robot calibration, but when environment contacts are relevant, the tactile modality can provide another natural, accurate, and highly relevant modality. Most existing tactile sensing methods for robot calibration are constrained to specific sensor-object pairs, limiting their applicability. This paper pioneers a general approach to exploit contacts in robot calibration, supporting self-touch throughout the entire system kinematics by generalizing touchable surfaces to any accurately represented mesh surface. The approach supports different contact sensors as long as a simple single-contact interface can be provided. Integrated into the *Atomic Transformation Optimization Method (ATOM)* calibration methodology, our work facilitates seamless integration of both modalities in a single approach. Our results demonstrate comparable performance to single-modality calibration but can trade off accuracy between both modalities, thus increasing overall robustness. Furthermore, we observe that utilizing a touch point at the end of a kinematic chain slightly improves calibration over touching the chain links with an external sensor but find no significant advantage of restricting touch to end-effector contacts when calibrating a dual-arm system with our method.**

I. INTRODUCTION

Robot calibration is a critical aspect of robotics research and application, essential for improving the reaching accuracy for manipulation [1]. Traditionally, efforts have centered around developing calibration methodologies utilizing standard sensors like RGB cameras [2]. However, generic RGB measurements are less accurate when compared to alternative sensing methods such as tactile sensing [2].

Tactile sensing provides direct feedback on local geometries during contact. By leveraging tactile feedback, particularly through self-touch [3] and interaction with known external objects [4], it becomes possible to capture more precise information on the kinematic structure of the robot.

Motivated by the potential of tactile sensing for calibration, this paper introduces a new tactile approach to kinematic and extrinsic robot calibration. Our approach aims to exploit generic tactile self-touch contacts, thus alleviating the need for dedicated and prepared contact areas in the workspace during calibration. Notably, this work supports touching arbitrary links in the involved kinematic chains. We rely on sufficiently accurate mesh-based surface representations of the touched surfaces to achieve this generic applicability.

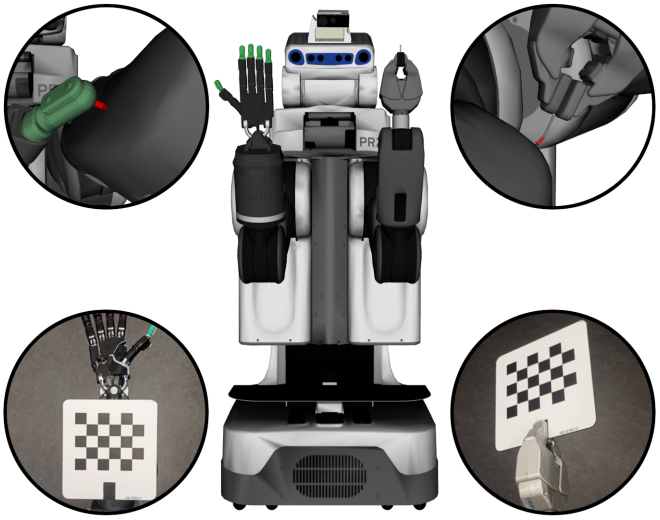


Fig. 1. **(center)** PR2 robot equipped with a Shadow Hand featuring BioTac sensors on its right arm and an Azure Kinect sensor mounted on its head. Our approach exploits observed contacts on self-touch: **(top left)** the right hand interacts with the left forearm, **(top right)** the left gripper's probe touches the right upper arm. Red arrows highlight the positional error between the sensed contact and the touched link. For the complementary visual calibration, chessboard patterns were mounted **(bottom left)** on the right forearm and **(bottom right)** on the left gripper.

We integrate our work into ATOM [5] to achieve multi-modality calibration as it provides modular support for RGB, Light Detection And Ranging (LiDAR), and depth sensors.

To showcase the approach, we focus on the kinematic calibration of an adapted PR2 robot, illustrated in Figure 1. Same as the original PR2, it is a bimanual mobile robot, with the left arm featuring 7 Degrees of Freedom (DOF), a pan/tilt head unit with 2 DOF, and capacitive pressure sensors in its gripper. Significant modifications of the original robot encompass the inclusion of a 19 DOF Shadow Dexterous Hand with BioTac fingertips as the right forearm, reducing the PR2-specific kinematic arm chain to 5 DOF, and the addition of an Azure Kinect camera atop the head.

The primary contributions of this paper are:

- A novel tactile-based calibration approach that exploits high-quality link models and readily supports different forms of contact modalities and sensed surfaces;
- The integration of the approach into the multi-modal calibration methodology ATOM;
- An evaluation of calibration performance considering the different modalities, dataset size, and sensor placement.

*This work was supported by the Foundation for Science and Technology (FCT), in the context of a Ph.D. Scholarship, under Grant 2023.01882.BD and the DFG Transregional Research Centre CML, TRR-169.

* corresponding author manuelgomes@ua.pt
[1] with IEETA-LASI, Department of Mechanical Engineering, University of Aveiro, Portugal.
[2] with Department of Informatics, Universität Hamburg, Germany.

II. RELATED WORK

Extrinsic calibration is a fundamental procedure in robotics that estimates the geometric transformations between sensors within a robotic system [6]. This problem has been extensively studied due to its critical importance for multi-sensor systems, which heavily rely on the assumption of correct transformations. Such systems find applications in diverse domains, including underwater navigation with stereo cameras [7], autonomous vehicles [8], and shared human-robot workspace scenarios [9].

Extrinsic calibration problems can take various forms [5], including sensor-to-sensor calibration [10], [11], sensor-in-motion calibration [12], [13], and sensor-to-frame calibration [14]. Estimating sufficiently good transformations becomes even more challenging as the number of sensors increases since calibration involves more active degrees of freedom and requires calibrating multiple sensor modalities together. Despite significant research focusing on pairwise configurations of different modalities [15], [16], comprehensive solutions for multi-modal, multi-sensor systems still need to be more thoroughly explored [5].

A novel methodology named ATOM has been proposed to address the limitations of existing calibration methods [5]. ATOM introduces a generalized approach to extrinsic calibration using a sensor-to-pattern paradigm, estimating transformations between the sensor and a calibration pattern. Unlike conventional methods that estimate transformations between sensors, ATOM preserves the complete topological structure of transformations, focusing solely on estimating atomic transformations of interest. This approach offers versatility across various calibration problems. Additionally, ATOM provides a complete calibration framework integrated into the Robot Operating System (ROS), covering all stages of the calibration process, from initial sensor pose definition to data collection and optimization [17]. Notably, ATOM has successfully calibrated diverse systems, including collaborative factory cells [18], hand-eye systems [19], intelligent vehicles [20], and agricultural robots [21], featuring a wide range of sensors such as RGB, LiDAR, and depth sensors.

Pradeep et al. implemented a bundle-adjustment approach to calibrate system parameters of multi-arm multi-sensor robots, specifically the original PR2 [22]. Their approach models Gaussian measurement errors to normalize observed error terms across different RGB and LiDAR sensors, facilitating optimization of geometric transformations and joint parameters.

All of the approaches discussed above focus on different forms of visual sensors for feature detection and thus readily generalize to different robotic systems using these combinations of sensors. On the other hand, robotic systems built for manipulation regularly touch surfaces inside their workspace, supporting various system components in intelligent robots [23]. Approaches can exploit such contacts for calibration as well. Specifically, calibration approaches can model contacts perceived through pressure sensors, force/torque sensors, and various kinds of advanced tactile perception as calibration errors. Tactile information can be used to calibrate the kinematic parameters of the moving robot links [24] and, due to their local perception, do not suffer from resolution changes across the workspace.

One of the simplest ways to use tactile sensors for calibration is to place the end-effector, or a probe attached to it, in contact with a planar surface in its workspace and exploit the corresponding planar constraint [4]. Despite the unknown real position of the contacts, they form a plane, facilitating closed-form calibration methods. An enhanced version of this approach involves embedding contact sensors on the contact plane, such as using a tablet [25], so contact positions can be determined in all three dimensions. Additionally, incorporating probabilistic time-varying and configuration-dependent joint errors, which can be tracked through Cartesian contact events, can improve precision in specific tasks [26].

While computationally easy to model, known planar touch surfaces are not accessible in the general case. However, they suggest viable generalizations to alternative methods of self-touch, a technique that involves the robot using itself as a known contact surface, usually between two end-effectors. A recent strategy for robotic hands, proposed by Tenhumberg et al. [27], moves the tips of the fingers pairwise into contact. The error is measured by computing the distance between the modeled fingertip geometries, as the actual measurement is always zero. Some approaches also incorporate self-visualization to the self-touch [3], [28]. To ground touch events in visual data, they attach fiducial markers to the touching end-effectors of a dual-armed robot, thereby facilitating a visual representation of the contact. Consequently, this method enables the calibration of RGB cameras based on perceived contacts. This method also enables the incorporation of planar constraints and has demonstrated that calibrating the robot through tactile based calibration yields superior outcomes compared to visual based calibration. A limitation of all discussed tactile-based methods lies in their specificity, as they rely solely on predefined tactile sensors and interactions with predetermined robot-object pairs.

III. PROPOSED APPROACH

Our tactile calibration method, i.e., a calibration method that exploits tactile measurements for whole-robot kinematic calibration, is geared to integrate seamlessly with the ATOM calibration framework [5]. This integration allows us to take advantage of ATOM's methodology for visual calibration, which relies on detecting keypoints of calibration patterns to estimate the pose of the camera and pattern across a dataset of measurements.

In contrast, our approach for tactile calibration involves measuring contacts and analyzing the currently assumed robot state to quantify deviations between model and actual link contacts. As we model tactile events in a kinematic model of a robot, we require a sufficiently accurate surface representation of the robot geometry, which determines a lower bound on the achievable calibration accuracy. We assume rigid representations of each robot link, which are typically available through the manufacturer in the form of

(a) (b)

Fig. 2. Tactile sensors used for contact detections: (a) PR2 fingertip pressure sensors; (b) SynTouch BioTac mounted on the fingertips of the Shadow Dexterous Hand.

meshes. We also assume a sufficiently accurate kinematic representation through a Unified Robot Description Format (URDF) description, except for those well-specified system parameters that will be calibrated. Typically, robot descriptions split geometry into high-quality *visual geometry* and low-accuracy but strongly overestimating *collision* geometry for fast online computation. We consistently utilize visual geometries as we rely on high model accuracy and calibration is run offline.

The tactile calibration process detailed below unfolds in four steps: individual contact detection, dataset acquisition & aggregation, and calibration. Contact detection is pivotal to extracting relevant information about contacts from tactile inputs. Data acquisition & aggregation gathers essential data required for calibration at each instance. The calibration step optimizes a clearly defined list of joint parameters and geometric transformations.

A. Contact Detection

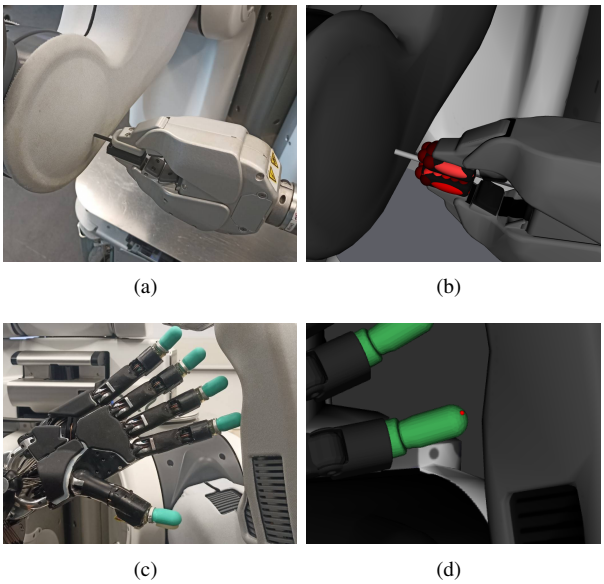
The mechanisms and sensor response for contact detection vary strongly between sensors. Some sensors, such as the TacTip [29], can deform significantly upon contact, leading to a high-contact surface with enough measured information to classify patches of surface geometry. Capacitive sensor arrays on the other hand, provide a discrete number of taxels that indicate contact forces along the surface normal. Force-Torque sensors provide a single wrench measurement indicating contact in any connected link.

In the following, we assume single-point contacts $\mathbf{C}$ and provide a minimal definition for them, which can be extracted from different sensors:

$$\mathbf{C} \equiv \langle \mathbf{t}, \mathbf{r}, \mathbf{x} \rangle, \tag{1}$$

where $\mathbf{t}$ represents the time of contact, $\mathbf{r}$ represents the geometric frame of reference of the tactile probe, i.e., the sensor responsible for detecting the contact, and $\mathbf{x}$ represents the Cartesian position of contact in $\mathbf{r}$.

This work focuses on contact detection using two tactile sensors in our setup: the PR2 fingertip pressure sensors [30] and SynTouch BioTac sensors [31] on the Shadow Hand, both of which can be seen in Figure 2.



(a) (b) (c) (d)

Fig. 3. Usage of tactile probes and data visualization in RViz: (a) 3D printed probe in PR2 Gripper used to sense contact with another part of the robot; (b) Data visualization of the contact. Each red blob represents a pressure sensor. The brighter the color of the blob, the larger the pressure sensed; (c) Shadow Hand's BioTac sensors used to sense contact with another part of the robot; (d) Data visualization of the contact. The red spot on the surface of the BioTac sensor represents a contact detected at that particular point.

1) PR2 Fingertip Pressure Sensors: Each PR2 gripper finger features 22 pressure-sensitive sensors (i.e., *taxels*). Their layout provides valuable information on normal forces especially for object grasping as most taxels are located on the insights of the parallel fingers. However, as the fingers cannot fully spread, this restricts their use for self-touch experiments where the touched geometry exceeds the maximum opening width. To address this limitation, we utilize a 3D-printed probe grasped by the gripper, as shown in Figure 3(a). The probe establishes a single point of contact at the tip that transmits forces to the pressure sensors, effectively reducing the contact position $\mathbf{x}$ to a single point and simplifying the approach.

Contact detection is achieved by numerically differentiating the pressure signal from all taxel signals. When an empirically determined threshold is exceeded, a contact is detected at the rising edge of the signal. Figure 3(b) visualizes such a contact with illustrated taxel response.

2) SynTouch BioTac Sensor: Contact detection and single-point localization are well-established capabilities of the BioTac sensor [32]. Before first contact, we normalize the electrode readings and aggregate them into a single signal. Contact is inferred when this signal exceeds a low threshold, which defines the contact time $\mathbf{t}$. During contact, the signal-weighted Cartesian average of the electrode values is computed, resulting in a 3D point within the BioTac sensor. This point is projected onto the sensor's surface to determine the contact point $\mathbf{x}$. Note that although the design of the biotac slightly deforms its rubber skin on contact, this deformation remains sub-millimeter and is thus below our expected final

accuracy for the robot setup. The Shadow Hand, with the sensors mounted, and a visualization of an estimated contact point, is present in Figure 3(c) and Figure 3(d).

3) Alternatives: While we use the two aforementioned sensors in our experiments, and various other tactile sensors could also be used, there are two more noteworthy alternatives. First, the probe-tool approach, which reduces $\mathbf{x}$ to a single point, transfers to any other force-measuring sensor, including gravity-compensated force-torque sensors. However, we did not utilize the ATI mini45 sensor mounted above the left gripper of our PR2 as the mounted setup did not prove sensitive enough for light contacts in early trials. Second, the same approach can be used with no tactile sensor, relying on online human labeling when a human operator guides the robot into contact. While such an approach will slightly reduce precision, this even enables using the framework without additional sensors.

B. Data Acquisition

Autonomously maneuvering a robotic arm into controlled self-contacts with inaccurate calibration would require additional sensors for local servoing and further assumptions. To focus on the core calibration method, we instead forego automatic data acquisition and rely on kinesthetic demonstrations by a human operator. The operator guides each sensor probe (i.e., either end-effector) repeatedly into contact with different body parts to accumulate contact events. As this procedure comes at the cost of a human bias in sampling self-touch configurations, an efficient autonomous (or semi-autonomous) sampling of self-touch for calibration constitutes a relevant area for future work.

C. Aggregating Collections

When detecting a contact, the system extracts self-contained samples that capture all data required to compute an error term for that contact during the optimization. We call each such sample a *collection*. For tactile calibration, a collection encompasses the contact as defined in Equation 1 as well as the rigid link that was touched, which we refer to as the *sensed link*.

Given that the uncalibrated robot state will not reliably register collisions between parts during physical contact, we devise a heuristic method to identify the sensed link. We numerically approximate the Cartesian velocity vector of the contact point based on the kinematic measurements in the last 100 ms. We project the contact point along this vector and consider the first mesh this ray penetrates as the sensed link.

D. Calibration

Our integration into ATOM is particularly evident in the optimization process based on ATOM's sensor-to-pattern paradigm. ATOM needs a pattern and feature detections on said pattern to function. When using cameras, commonly supported patterns include chessboards and ChArUco boards, with the set of detections comprising the 2D coordinates of detected pattern corners.

Tactile calibration, on the other hand, diverges from this concept, as it does not require a traditional calibration pattern. Instead, the robot can interact with arbitrary objects or parts, assuming a predefined mesh representation of the object or part. Nevertheless, we can adapt tactile calibration to this formulation with a minimal change of perspective. Here, the probe and sensed link function analogously to the sensor and pattern, respectively, constituting the sensing and sensed components of the contact. Notably, the contact point seamlessly aligns with the concept of detection, representing the precise Cartesian location where contact is detected.

The original ATOM optimization, adapted to include joint offset calibration, is expressed as follows:

$$\underset{\{\hat{\mathcal{T}}\},\{\hat{\mathcal{J}}\}}{\operatorname{argmin}} \sum_{s\in\mathcal{S}} \sum_{c\in\mathcal{C}} \sum_{d\in\mathcal{D}} e\Big({}^{s}\mathbf{T}_{c}^{p}, d_{[c,s]}, \{\lambda_s\}\Big). \tag{2}$$

In this context, $e(\cdot)$ represents an error function, instantiated per modality. The error function is computed for each sensor $s \in \mathcal{S}$, considering each collection $c \in \mathcal{C}$. It is dependent on $d_{[c,s]}$, which denotes a detection of the pattern from the set of detections $\mathcal{D}$, and on λ_s, which denotes a set of additional sensor parameters required to compute the error, such as the intrinsic parameters for camera sensors. It is also dependent on ${}^{s}\mathbf{T}_{c}^{p}$, which denotes the geometric transformation from the sensor s to the singular pattern p, as computed via forward kinematics involving the optimization parameters. In ATOM, this transformation is decomposed into a set of atomic transformations $\mathcal{T}$, and a set of joints $\mathcal{J}$. From these atomic transformations, one must define which set of them are to be estimated, $\hat{\mathcal{T}}$. Along with $\hat{\mathcal{T}}$, one must also define which set of joint parameters $\hat{\mathcal{J}}$ are to be estimated.

In our interpretation for tactile calibration, we need to support many patterns, i.e., all sensed links across data acquisition. To accommodate, we extend the equation to include a set of patterns $\mathcal{P}$:

$$\underset{\{\hat{\mathcal{T}}\},\{\hat{\mathcal{J}}\}}{\operatorname{argmin}} \sum_{p\in\mathcal{P}} \sum_{s\in\mathcal{S}} \sum_{c\in\mathcal{C}} \sum_{d\in\mathcal{D}} e\Big({}^{s}\mathbf{T}_{c}^{p}, d_{[c,s,p]}, \{\lambda_s\}\Big), \tag{3}$$

where p now denotes a pattern within the set of patterns $\mathcal{P}$. As the different modalities can detect multiple different patterns in the same collection, we aggregate over all patterns instead of specifically extracting the sensed link of each collection and sensor. $\hat{\mathcal{T}}$ and $\hat{\mathcal{J}}$ are predefined by the user but will only be estimated for touch collections if they are in the kinematic chain of the aggregate transformation ${}^{s}\mathbf{T}_{c}^{p}$.

The error function $e(\cdot)$ must be defined for each sensor modality. The RGB error function has been previously defined by [5] and compares the detected corners of a pattern in an image with their expected positions projected in the same image. This error function can be formulated as

$$e_{[p,c,s,d]} = \Big|\Big| x_{[c,s,d]} - \mathcal{P}\big({}^{s}\mathbf{T}_{c}^{p} \cdot \mathbf{x}_d, k_s, u_s\big) \Big|\Big|_F^2, \tag{4}$$

where $\mathbf{x}_d$ represents the homogeneous three-dimensional coordinates of the pattern corner corresponding to detection d, k_s and u_s represent the intrinsic and distortion parameter

vectors of the sensor s, $\mathcal{P}$ is the camera projection function, $x_{[c,s,d]}$ is the two-dimensional image coordinate of detection d, at collection c, for sensor s, and $||\cdot||_F^2$ represents the Frobenius norm.

We base the tactile error function on the distance between the perceived contact point $\mathbf{x}$ and the sensed link. Ideally, this distance should be zero after successful calibration, indicating contact in the model. To formalize, let $\mathbf{x}_{[c,s,d]}$ denote the 3D coordinates of the contact point w.r.t. the probe frame $\mathbf{r}$. The set $\mathcal{Q}$ encompasses the surface of the mesh representation of the sensed link, w.r.t. its coordinate frame. The calculation of the distance between the probe and the sensed link involves determining the shortest distance between any point in $\mathcal{Q}$ and the contact point $\mathbf{x}_{[c,s,d]}$ after transforming it into the coordinate frame of the sensed link. This error function is formulated as follows:

$$e_{[p,c,s,d]} = \min_{q \in \mathcal{Q}} \left(\left|\left| q - ({}^{s}\mathbf{T}_{c}^{p})^{-1} \cdot \mathbf{x}_{[c,s,d]} \right|\right|_F^2 \right), \qquad (5)$$

where q is a single 3D point on the link surface $\mathcal{Q}$. This error function differs from other tactile calibration methods, as it does not rely on a previously defined sensed link or point, but rather on the closest point on the sensed link to the contact point.

IV. RESULTS

This section demonstrates the approach for the kinematic calibration of our modified PR2 robot based on a comprehensive dataset. We investigate calibration accuracy across modalities, assess the effect of dataset sizes on the results, and perform a series of ablation studies to analyze the influence of touch and sensor locations.

Our experiments aim at a kinematic and extrinsic calibration of the PR2 robot and the Azure Camera system. This calibration includes the estimation of the joint offsets of all arm joints (12 parameters[1]) and all head joints (2 parameters[2]). Additionally, it includes estimating the transformation between the Azure Kinect camera and the head of the PR2 (6 parameters), the transformations between the arms and their respective chessboard (12 parameters), totalling 32 parameters. The non-linear relationship between the motors and joint states, a factor beyond the scope of this study, precluded the calculation of joint offsets for the Shadow Hand.

For the visual modality, we recorded images through the Azure Kinect camera, with a calibration pattern strategically positioned on both the robot's left gripper and right forearm, as depicted in Figure 1. Visual calibration also requires estimating each pattern pose w.r.t. the robot link on which it is positioned. For the tactile modality, self-touch actions were performed with both contact sensors as described in subsection III-A.

The collected dataset is detailed in Table I.

[1]left shoulder pan, left shoulder lift, left upper arm roll, left elbow flex, left forearm roll, left wrist flex, left wrist roll, right shoulder pan, right shoulder lift, right upper arm roll, right elbow flex, right forearm roll

[2]head tilt and head pan

TABLE I
DATASET COMPOSITION

Collection Type	Count	Pattern/Sensor Used
RGB Collections	30	Pattern on right forearm
	23	Pattern on left gripper
Contact Collections	117	BioTac on index finger
	127	PR2 capacitive pressure sensors
Total	**297**	

An obvious complication of our evaluation is the need to consider separate evaluation functions for the different modalities. As we do not have an external ground truth, we utilize the error terms in Equation 4 and Equation 5 for evaluation, but note that either induces a bias towards calibration of their respective modality.

A. Modal Analysis

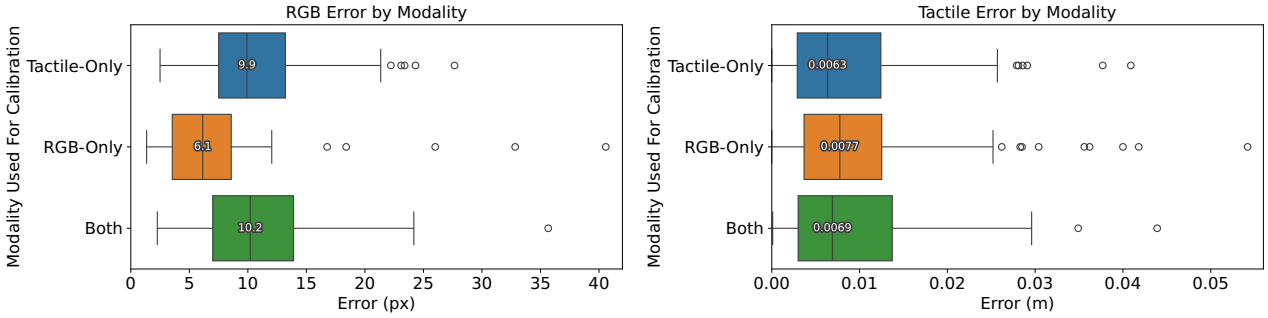
First, we investigate the calibration results for the whole dataset across modalities. We compare three conditions, calibrating either RGB data only (RGB-Only), tactile data only (Tactile-Only), or both modalities simultaneously (Both).

By its nature, tactile only calibration cannot calibrate the optical frame of the RGB camera. However, we still aim to estimate calibration accuracy in the visual error. To achieve a lower bound on such a measure, we follow up all Tactile-Only calibration runs with a traditional Hand-Eye calibration step to estimate only the camera and pattern transforms given the optimized joint offsets. This adjustment aims to achieve an ideal camera transformation, so the evaluation highlights only the errors in kinematic calibration. However, this process also introduces bias as the whole evaluation set is involved in the latter calibration, and we can only consider the result a lower bound.

For a thorough analysis of the entire dataset, we employ leave-one-out cross-validation. The results are depicted in Figure 4. It is relevant to point out that there is no generic translation between both error units, as the Cartesian distance represented by pixel errors depends on the pattern distance and the error orientation. Still, both metrics represent calibration errors in the same order of magnitude.

Specifically, Figure 4(a) reports the RGB errors across all conditions. As expected, RGB-Only yields superior performance as calibrating with only the evaluated modality biases the evaluation. Still, both other conditions yield comparable results. As the proposed Tactile-Only evaluation here only serves as a lower bound estimate, we conclude that utilizing both modalities for calibration trades off the precision of the modalities while accepting a slightly increased error spread. At the same time, the number of distribution outliers reduces significantly, hinting at an increased robustness of the result.

In Figure 4(b), the same conditions are analyzed w.r.t. the observed tactile errors, and we find essentially the same results with inverted modalities. Interestingly, the median discrepancy between both single-modality conditions is very weak, while RGB-Only still shows a larger spread of outliers.



(a) (b)

Fig. 4. Box plots illustrating (a) RGB errors and (b) tactile errors when calibrating across different modalities. The value depicted in each box corresponds to the median.

This indicates the expected good performance of purely visual calibration but also implies that visual calibration overfits part of the evaluated joint space. The combined condition, again, slightly increases the error spread but is more resistant to outliers. The error magnitude for all conditions is in the order of 1cm, which is generally high for robot manipulation. However, in the PR2 this can be partly explained by the non-linear cumulative error of the encoder measurements for all joints. These encoders are placed on the motors and not on the joints, having gearing between them which induce deadband. Another reason for this error would be the 3D model itself, not providing sub millimeter accuracy.

The reported visual errors can be approximately compared to those presented in [22], which depict median RGB errors post-calibration between the PR2 head cameras on the order of 7 pixels. Although our median error seems slightly higher, our experiments employ a camera resolution of 2048x1536, while their approach uses a resolution of 752x480. Thus, errors can be amplified by a factor of up to 3.2 between the experiments.

B. Dataset Size Analysis

Next, we investigated the impact of dataset size on the calibration process of a robot utilizing our methodology. To assess this impact, we employed stratified shuffle split cross-validation. This approach involves partitioning the dataset into calibration and evaluation subsets in a stratified manner. The stratification criteria were based on the sensors used in data collection: the Azure Camera, Left Gripper Probe, and Right Hand Index Finger.

We systematically varied the percentage of the dataset allocated for training from 5% to 90%, ensuring 10% of the dataset reserved for evaluation purposes. For each percentage of the dataset reserved for training, there were ten repetitions of sampling and calibration. The procedure to determine the camera position in Tactile-Only was already outlined in subsection IV-A.

The results are depicted in Figure 5. Figure 5(a) illustrates the average RGB error. Tactile-Only calibration results in minimal RGB error, even in low collection usage, due to the discussed lower bound analysis which uses the complete RGB collection for the extrinsic camera calibration. In contrast, utilizing a small subset of RGB images introduces a strong bias towards this subset, leading to larger errors in Both and RGB-Only. As more collections are used for calibration, error rapidly decreases, followed by a plateau after approximately 70% of the dataset is used, with RGB-Only achieving superior results. However, the Tactile-Only error slightly increases as more collections are employed for calibration, which can be attributed to the joint parameters adapting more to the tactile information, compensating the overfit on the image-based calibration.

To analyze the robustness of the whole evaluated joint space, we additionally investigate maximum RGB errors in Figure 5(c). We observe a continuous decrease of maximum outliers in Both and RGB-Only as dataset size increases, indicating a better generalization as more data on the evaluated modality becomes available.

Moving to Figure 5(b), the average tactile error demonstrates a decreasing trend across all modalities as more of the dataset is utilized, plateauing at around 60%. RGB-Only exhibits the highest error, which is reasonable given its lack of tactile information for calibration, while the Tactile-Only and Both modalities achieve comparable results.

Finally, Figure 5(d) shows maximum tactile errors similar to Figure 5(c), with RGB-Only plateauing at approximately 30% with a considerably larger error than both other conditions. Tactile-Only and Both continue to decrease in error until the end of the graph, indicating that the dataset size may still be below optimal for calibration.

C. Tactile Ablation Studies

Lastly, we investigate the effect of tactile sensor placement and the choice of touched links through dataset ablation. The results of restricting the set of calibration contacts to either tactile sensor and restricting calibration contacts exclusively to end-effector interactions can be seen in Table II.

Initially, the effectiveness of calibration using only the left gripper's probe was evaluated through 5-Folds cross-validation, with 20% of the dataset reserved for evaluation. The error for contacts detected on the right arm was higher

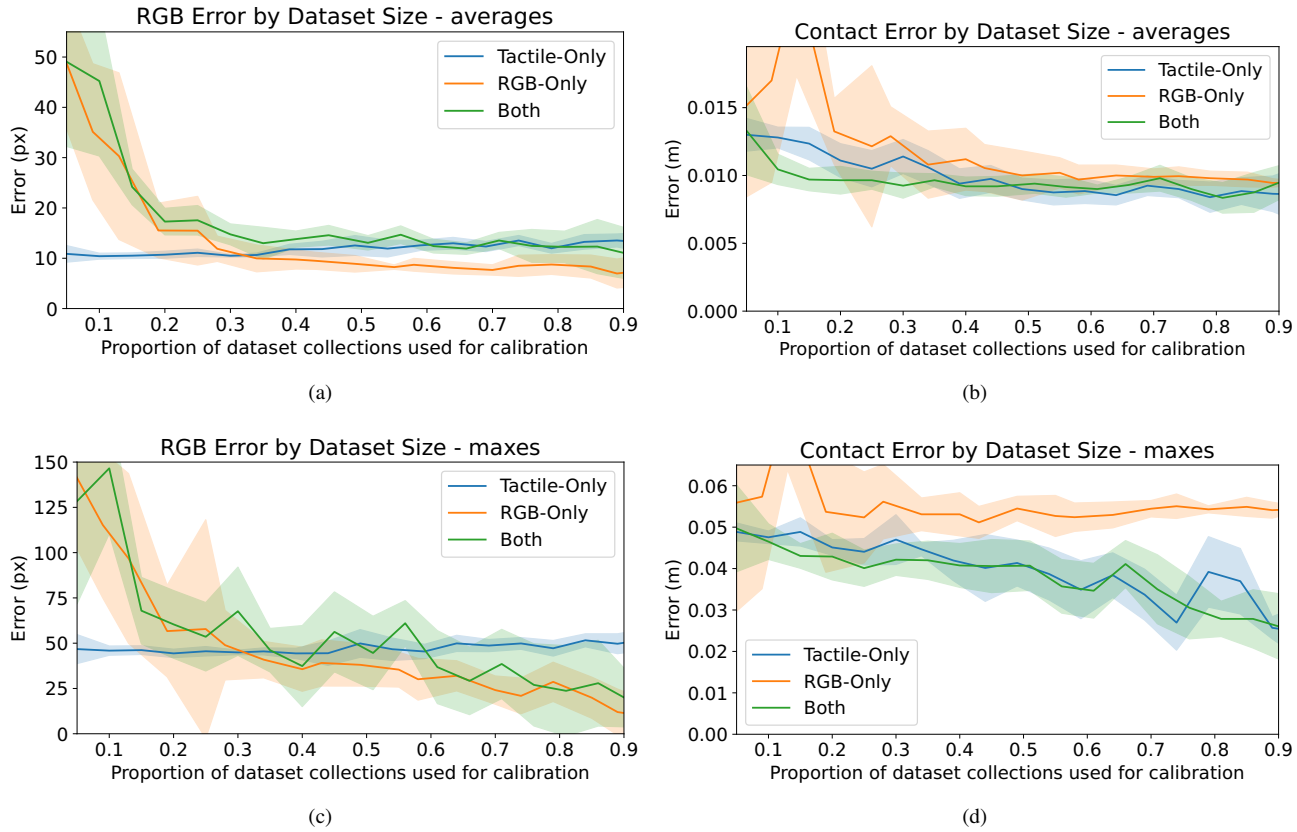


Fig. 5. Line plots illustrating (a) average RGB errors, (b) maximum RGB errors, (c) average tactile errors, and (d) maximum tactile errors when calibrating across different modalities. The results show the average of 10 runs, with the standard deviation presented in an error bar.

TABLE II

TACTILE ERRORS IN ABLATION STUDIES OF TACTILE CALIBRATION.

Evaluating touches on	Calibrating with	Mean error (m)
Left Arm	Left Gripper Probe	0.0089
	Right Hand BioTac	0.0112
	Both	0.0082
Right Arm	Left Gripper Probe	0.0114
	Right Hand BioTac	0.0099
	Both	0.0099
End-effector links	Both / EE contacts only	0.0116
Other links		0.0122

than those on the left. Conversely, calibrating using the right hand's BioTac sensor resulted in higher errors for contacts detected on the left arm than those on the right arm.

These findings suggest improved calibration of the whole arm kinematic chain when the tactile probe is attached to the arm itself rather than mounted on another kinematic chain, that touches links from the arm. This can be attributed to the arm's kinematic chain being fully excited in each touch, as the transformation from the measured contact position to the sensed link always connects through it. In contrast, each external touch of the calibrated kinematic chain is only anchored to the 2-dimensional mesh surface, effectively providing less information on the touched link's position.

Another pertinent analysis concerns exclusive contact with end-effectors. Calibrating solely with the 33 collections involving end-effector contact and evaluating against the remaining data yielded an error of 0.0122 m. This outcome is comparable to the results in Figure 5(b), where the nearest data point of Tactile-Only corresponds to calibration with 36 collections with an average error of 0.0123 m. Thus, focusing solely on end-effector contact does not confer a discernible advantage over touching various links of the robot's body for overall robot calibration.

V. CONCLUSION

We introduced a new method for calibrating robots using tactile and visual feedback within one framework. When compared to other discussed approaches, we extend the concept of self-touch throughout the robot's entire kinematic chain, not requiring predefined contact points or specific sensors. Therefore, the approach offers versatility and robustness across different sensors and robot parts, contingent upon the availability of precise geometric models. It is also integrated in a multi-modal calibration framework, which allows for the simultaneous calibration of a wider range of sensors.

Our results indicate that this multi-modal approach yields comparable performance to single-modal analysis but improves robustness and generalization. The improvement is not solely due to a larger dataset, as we can see data stabilizing in the dataset size analysis, but due to high measurement accuracy of the sensor and the possibility of

collecting samples outside the field of view of the camera. We also found that for our particular bimanual robot setup, we achieved the best results from around 170 contact collections and 35 RGB collections. However, maximum calibration errors will further decrease with increasing data points.

In a bimanual system, better calibration results are achieved when the touch sensor is placed on the arm requiring calibration. Additionally, touching other robot parts besides the end-effector yields similar results for whole-system calibration, enhancing the generality of the calibration process.

Future research might explore methods for enhanced data acquisition by incorporating kinesthetic demonstration data to estimate joint parameter uncertainty and identify optimal touchpoints for calibration. Furthermore, data-driven approaches can be explored to estimate the optimal number of data points needed for calibration. This estimation should consider a balanced sampling of contact poses across the space and the balanced excitation of kinematic chains for both tactile and visual modalities. Incorporation of further modalities, especially LiDAR and depth sensing, could improve calibration accuracy and reliability.

REFERENCES

[1] Z. Li, S. Li, and X. Luo, "An overview of calibration technology of industrial robots," *IEEE/CAA Journal of Automatica Sinica*, vol. 8, no. 1, pp. 23–36, Jan. 2021.

[2] Chen-Gang, Li-Tong, Chu-Ming, J.-Q. Xuan, and S.-H. Xu, "Review on kinematics calibration technology of serial robots," *International Journal of Precision Engineering and Manufacturing*, vol. 15, no. 8, pp. 1759–1774, Aug. 1, 2014.

[3] K. Stepanova, J. Rozlivek, F. Puciow, P. Krsek, T. Pajdla, and M. Hoffmann, "Automatic self-contained calibration of an industrial dual-arm robot with cameras using self-contact, planar constraints, and self-observation," *Robotics and Computer-Integrated Manufacturing*, vol. 73, p. 102 250, Feb. 1, 2022.

[4] X.-L. Zhong, J. M. Lewis, and F. L.N.-Nagy, "Autonomous robot calibration using a trigger probe," *Robotics and Autonomous Systems*, vol. 18, no. 4, pp. 395–410, Oct. 1, 1996.

[5] M. Oliveira, E. Pedrosa, A. P. de Aguiar, *et al.*, "ATOM: A general calibration framework for multi-modal, multi-sensor systems," *Expert Systems with Applications*, vol. 207, p. 118 000, Nov. 2022.

[6] K. Kodagoda, A. Alempijevic, J. Underwood, S. Kumar, and G. Dissanayake, "Sensor registration and calibration using moving targets," in *2006 9th Int. Conf. on Control, Automation, Robotics and Vision*, 2006, pp. 1–6.

[7] S. Yan, Z. Wu, J. Wang, M. Tan, and J. Yu, "Marine Autonomous Navigation for Biomimetic Underwater Robots Based on Deep Stereo Attention Network," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Sep. 2021, pp. 8418–8423.

[8] K. Zhu, W. Chen, W. Zhang, R. Song, and Y. Li, "Autonomous Robot Navigation Based on Multi-Camera Perception," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Oct. 2020, pp. 5879–5885.

[9] T. Linder, N. Vaskevicius, R. Schirmer, and K. O. Arras, "Cross-Modal Analysis of Human Detection for Robotics: An Industrial Case Study," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Sep. 2021, pp. 971–978.

[10] D. Tu, B. Wang, H. Cui, Y. Liu, and S. Shen, "Multi-Camera-LiDAR Auto-Calibration by Joint Structure-from-Motion," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Oct. 2022, pp. 2242–2249.

[11] B. Ponton, M. Ferri, L. König, and M. Bartels, "Efficient Extrinsic Calibration of Multi-Sensor 3D LiDAR Systems for Autonomous Vehicles using Static Objects Information," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Oct. 2022, pp. 6285–6292.

[12] S. Agarwal, N. Snavely, S. M. Seitz, and R. Szeliski, "Bundle adjustment in the large," in *Computer Vision – ECCV 2010*, Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 29–42.

[13] C. Boucher, J. C. Noyer, and M. Benjelloun, "3D structure and motion recovery in a multisensor framework," *Information Fusion*, vol. 2, no. 4, pp. 271–285, Dec. 2001.

[14] B. Lourenço, T. Madeira, P. Dias, V. Santos, and M. Oliveira, "2D lidar to kinematic chain calibration using planar features of indoor scenes," *Industrial Robot: the Int. journal of robotics research and application*, vol. 47, no. 5, pp. 647–655, Jan. 1, 2020.

[15] G. Chen, G. Cui, Z. Jin, F. Wu, and X. Chen, "Accurate Intrinsic and Extrinsic Calibration of RGB-D Cameras With GP-Based Depth Correction," *IEEE Sensors Journal*, vol. 19, no. 7, pp. 2685–2694, Apr. 2019.

[16] V. Q. Dinh, T. P. Nguyen, and J. W. Jeon, "Rectification Using Different Types of Cameras Attached to a Vehicle," *IEEE Transactions on Image Processing*, vol. 28, no. 2, pp. 815–826, Feb. 2019.

[17] M. Gomes, M. Oliveira, and V. Santos, "ATOM Calibration Framework: Interaction and Visualization Functionalities," *Sensors*, vol. 23, no. 2, p. 936, 2 Jan. 2023.

[18] D. Rato, M. Oliveira, V. Santos, M. Gomes, and A. Sappa, "A sensor-to-pattern calibration framework for multi-modal industrial collaborative cells," *Journal of Manufacturing Systems*, vol. 64, pp. 497–507, Jul. 1, 2022.

[19] E. Pedrosa, M. Oliveira, N. Lau, and V. Santos, "A General Approach to Hand–Eye Calibration Through the Optimization of Atomic Transformations," *IEEE Transactions on Robotics*, vol. 37, no. 5, pp. 1619–1633, Oct. 2021.

[20] M. Oliveira, A. Castro, T. Madeira, P. Dias, and V. Santos, "A general approach to the extrinsic calibration of intelligent vehicles using ROS," in *Robot 2019: Fourth Iberian Robotics Conf.*, Cham: Springer Int. Publishing, 2020, pp. 203–215.

[21] A. Aguiar, M. Oliveira, E. Pedrosa, and F. Santos, "A Camera to LiDAR calibration approach through the optimization of atomic transformations," *Expert Systems with Applications*, vol. 176, p. 114 894, Aug. 15, 2021.

[22] V. Pradeep, K. Konolige, and E. Berger, "Calibrating a multi-arm multi-sensor robot: A bundle adjustment approach," in *Experimental Robotics: The 12th Int. Symposium on Experimental Robotics*, Berlin, Heidelberg: Springer Berlin Heidelberg, 2014, pp. 211–225.

[23] Q. Li, O. Kroemer, Z. Su, F. F. Veiga, M. Kaboli, and H. J. Ritter, "A Review of Tactile Information: Perception and Action Through Touch," *IEEE Transactions on Robotics*, vol. 36, no. 6, pp. 1619–1634, Dec. 2020.

[24] B. Mooring, M. Driels, and Z. Roth, *Fundamentals of Manipulator Calibration*. USA: John Wiley & Sons, Inc., 1991, 329 pp.

[25] Y. Cai, H. Gu, C. Li, and H. Liu, "Easy industrial robot cell coordinates calibration with touch panel," *ROBOTICS AND COMPUTER-INTEGRATED MANUFACTURING*, vol. 50, pp. 276–285, Apr. 2018.

[26] L. Meyer, K. H. Strobl, and R. Triebel, "The Probabilistic Robot Kinematics Model and its Application to Sensor Fusion," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Oct. 2022, pp. 3263–3270.

[27] J. Tenhumberg, L. Sievers, and B. Bäuml, "Self-Contained and Automatic Calibration of a Multi-Fingered Hand Using Only Pairwise Contact Measurements," in *2023 IEEE-RAS 22nd International Conference on Humanoid Robots (Humanoids)*, Dec. 2023, pp. 1–8.

[28] K. Stepanova, T. Pajdla, and M. Hoffmann, "Robot Self-Calibration Using Multiple Kinematic Chains—A Simulation Study on the iCub Humanoid Robot," *IEEE Robotics and Automation Letters*, vol. 4, no. 2, pp. 1900–1907, Apr. 2019.

[29] N. F. Lepora, "Soft Biomimetic Optical Tactile Sensing With the TacTip: A Review," *IEEE Sensors Journal*, vol. 21, no. 19, pp. 21 131–21 143, Oct. 2021.

[30] S. Chitta, J. Sturm, M. Piccoli, and W. Burgard, "Tactile sensing for mobile manipulation," *Robotics, IEEE Transactions on*, vol. 27, pp. 558–568, Jul. 2011.

[31] Y. S. Narang, B. Sundaralingam, K. Van Wyk, A. Mousavian, and D. Fox, "Interpreting and predicting tactile signals for the SynTouch BioTac," *The International Journal of Robotics Research*, vol. 40, no. 12-14, pp. 1467–1487, Dec. 1, 2021.

[32] C.-H. Lin, J. A. Fishel, and G. Loeb, "Estimating Point of Contact , Force and Torque in a Biomimetic Tactile Sensor with Deformable Skin," 2013.