2024 IEEE International Conference on Robotics and Automation (ICRA 2024)
May 13-17, 2024. Yokohama, Japan

# TEXterity: Tactile Extrinsic deXterity

*Antonia Bronars[1], *Sangwoon Kim[1], Parag Patre[2] and Alberto Rodriguez[1]
*Equal Contribution, [1]MIT, [2]Magna International Inc.

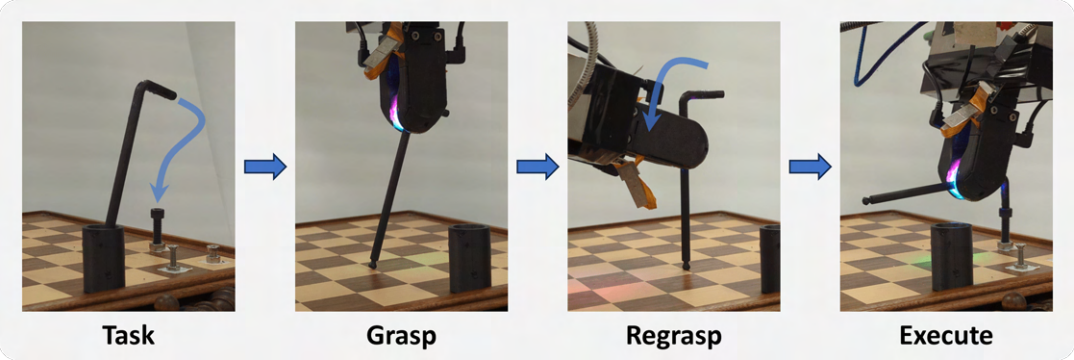


**Fig. 1:** An example task that requires tactile extrinsic dexterity. A proper grasp is essential when using an Allen key to apply sufficient torque while fastening a hex bolt. The proposed method utilizes tactile sensing on the robot's finger to localize and track the grasped object's pose and also regrasp the object in hand by pushing it against the floor - effectively leveraging extrinsic dexterity.

***Abstract*— We introduce a novel approach that combines tactile estimation and control for in-hand object manipulation. By integrating measurements from robot kinematics and an image-based tactile sensor, our framework estimates and tracks object pose while simultaneously generating motion plans in a receding horizon fashion to control the pose of a grasped object. This approach consists of a discrete pose estimator that tracks the most likely sequence of object poses in a coarsely discretized grid, and a continuous pose estimator-controller to refine the pose estimate and accurately manipulate the pose of the grasped object. Our method is tested on diverse objects and configurations, achieving desired manipulation objectives and outperforming single-shot methods in estimation accuracy. The proposed approach holds potential for tasks requiring precise manipulation and limited intrinsic in-hand dexterity under visual occlusion, laying the foundation for closed-loop behavior in applications such as regrasping, insertion, and tool use. Please see this url for videos of real-world demonstrations.**

## I. INTRODUCTION

The ability to manipulate objects within the hand is a long-standing objective in robotics for its potential to increase the workspace, speed, and capability of robotic systems. For example, the ability to change the grasp on an object can improve grasp stability and functionality, or prevent collisions and kinematic singularities. In-hand manipulation is challenging from the perspectives of state estimation, planning, and control: firstly, once the object is enveloped by the grasp, it becomes difficult to perceive with external vision systems; secondly, the hybrid dynamics of contact-rich tasks are difficult to predict [1] and optimize over [2].

Existing work on in-hand manipulation emphasizes the problem of sequencing contact modes, and can be broken down into two prevailing methodologies. One line of work relies on simple object geometries and exact models of contact dynamics to plan using traditional optimization-based approaches [2]–[6], while the other leverages model-free reinforcement learning to learn policies directly that only consider or exploit contact modes implicitly [7]–[12]. Much less consideration has been given to the challenge of precisely controlling such behaviors, despite the fact that prominent tasks like connector insertion or screwing in a small bolt require high precision.

Tactile feedback is a promising modality to enable precise control of in-hand manipulation. Image-based tactile sensing [13]–[15] has gained traction in recent years for its ability to provide high-resolution information directly at the contact interface. Image-based tactile sensors have been used for pose estimation [16], object retrieval [17], and texture recognition [18]. They have also been used to estimate the location of contacts with the environment [19]–[21], to supervise insertion [22], and to guide the manipulation of objects like boxes [23], tools [24], cable [25], and cloth [26].

We study the problem of precisely controlling in-hand sliding regrasps by pushing against an external surface, i.e. extrinsic dexterity [27], supervised only by robot proprioception and tactile sensing. Our framework is compatible with arbitrary, but known, object geometries and succeeds even when the contact parameters are known only approximately.

This work builds upon previous research efforts. First, *Tac2Pose* [16] estimates the relative gripper/object pose using tactile sensing, but lacks control capabilities. Second, *Simultaneous Tactile Estimation and Control of Extrinsic Contact* [28] estimates and controls extrinsic contact states between the object and its environment, but has no understanding of the object's pose and therefore has limited ability to reason over global re-configuration. Our approach combines the strengths of these two frameworks into a single system. As a result, our method estimates the object's pose and its associated contact configurations and simultaneously controls them. By merging these methodologies, we aim to provide a complete solution for precisely controlling general planar in-hand manipulation.

## II. RELATED WORK

**Tactile Estimation and Control.** Image-based tactile sensors are particularly useful for high-accuracy pose estimation, because they provide high-resolution information about the

*This research was supported by MAGNA.

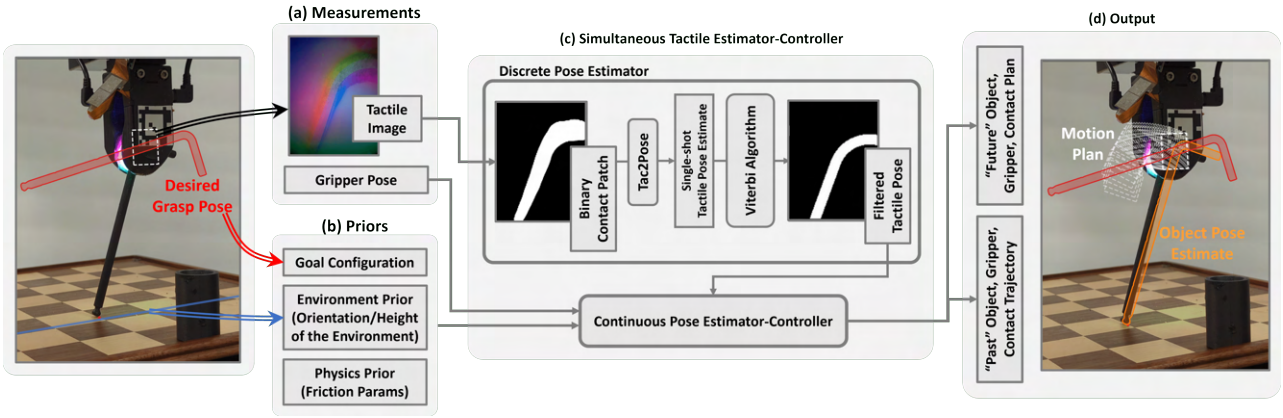


**Fig. 2:** Overview of the Simultaneous Tactile Estimation and Control Framework.

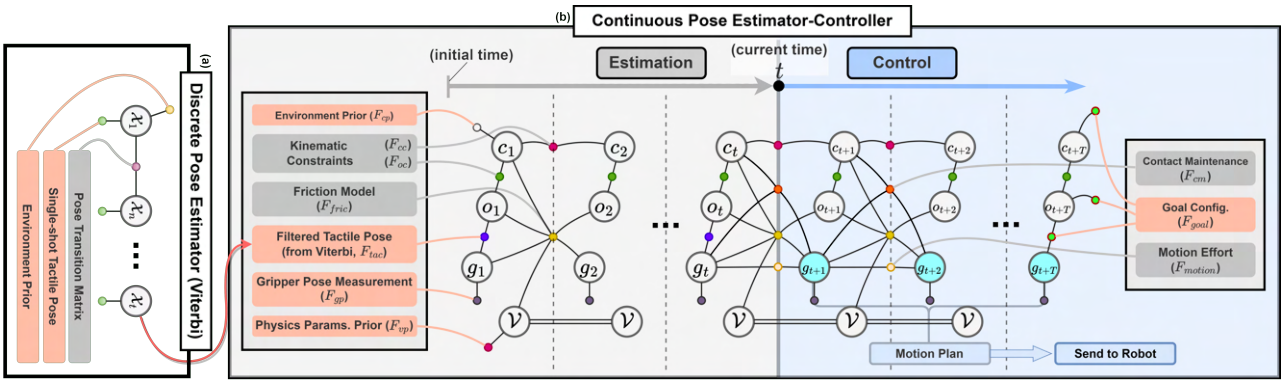


**Fig. 3:** Graph Architecture of the Simultaneous Tactile Estimator-Controller.

object geometry throughout manipulation. They have been successfully used to track object drift from a known initial pose [29], [30], build a tactile map and localize the object within it [31]–[33], and estimate the pose of small parts from a single tactile image [34]. Because touch provides only local information about the object geometry, most tactile images are inherently ambiguous [16]. Some work has combined touch with vision [35]–[38] to resolve such ambiguity. Our approach is most similar to a line of work that estimates distributions over possible object pose from a single tactile image [16], [39], [40], then fuses information over streams of tactile images using particle [39] or histogram [40] filters. [39] tackles the estimation, but not control, problem, assuming that the object is rigidly fixed in place while a human operator slides a tactile sensor along the object surface. Similarly, [40] also assumes the object is fixed in place, while the robot plans and executes a series of grasp and release maneuvers to localize the object. Our work, on the other hand, tackles the more challenging problem of estimating and controlling the pose of an object sliding within the grasp while not rigidly attached to a fixture.

**In-Hand Manipulation.** In-hand manipulation is most commonly achieved with dexterous hands or by leveraging the surrounding environment (extrinsic dexterity [27]). One line of prior work formulates the problem as an optimization over exact models of the hand/object dynamics [2]–[6], [41], but only for simple objects and generally only in simulation [2], [3], or by relying on accurate knowledge of physical parameters to execute plans precisely in open loop [4]. Another line of prior work focuses on modeling the mechanics of contact itself in a way that is useful for planning and control, either analytically [42]–[44] or with neural networks [45], [46].

Some work has avoided the challenges of modeling contact altogether, instead relying on model-free reinforcement learning with vision to directly learn a policy for arbitrary geometries. Some policies have been tested on simulated vision data only [7], [8], while others operate on real images [9]–[12]. They, however, suffer from a lack of precision. As an example, [9] reports 45% success on held out objects, and 81% success on training objects, where success is defined as a reorientation attempt with less than 0.4 rad (22.9°) of error, underscoring the challenge of precise reorientation for arbitrary objects.

There have also been a number of works leveraging tactile sensing for in-hand manipulation. [25], [26], and [47] use image-based tactile sensors to supervise sliding on cables, cloth, and marbles, respectively. [24] detects and corrects for undesired slip during tool manipulation, while [48] learns a policy that trades off between tactile exploration and execution to succeed at insertion tasks. Some works rely on proprioception [49] or pressure sensors [50] to coarsely reorient objects within the hand. State estimation from such sensors is challenging and imprecise, leading to policies that accrue large errors.

We consider the complementary problem of planning and controlling over a known contact mode (in-hand sliding by pushing the object against an external surface), where the object geometry is arbitrary but known. We leverage a simple model of the mechanics of sliding and supervise the behavior with high-resolution tactile sensing, in order to achieve precise in-hand

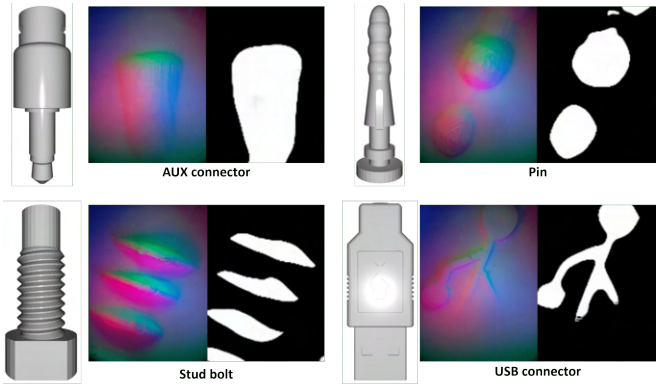


**Fig. 4:** Test objects with example tactile images and contact patch reconstruction.

manipulation. By emphasizing the simultaneous estimation and control for a realistic in-hand manipulation scenario, this work addresses a gap in the existing literature and paves the way for executing precise dexterous manipulation on real systems.

## III. METHOD

### A. Problem Formulation

We address the task of manipulating objects in-hand from unknown initial grasps to achieve desired configurations by pushing against the environment. The target configurations encompass a range of potentially simultaneous manipulation objectives:

- Changing the grasp pose (i.e., relative rotation/translation between the gripper and the object)
- Changing the orientation of the object in the world frame (i.e., pivoting against the environment)
- Changing the location of the extrinsic contact point (i.e., sliding against the environment)

A wide variety of regrasping tasks can be specified via a combination of the above objectives.

We make several assumptions to model this problem:

- Grasped objects are rigid with known 3D models.
- The part of the environment that the object interacts with is flat, with a known orientation and height.
- Contact between the grasped objects and the environment occurs at a single point.
- Grasp reorientation is constrained to the plane of the gripper finger surface.

### B. Overview

Fig. 1 illustrates our approach through an example task: using an Allen key to apply sufficient torque while fastening a hex bolt. Adjusting the grasp through in-hand manipulation is necessary to increase the torque arm and prevent the robot from hitting its motion limit during the screwing.

Fig. 2 provides an overview of the framework of our approach. The system gathers measurements from both the robot and the vision-based tactile sensor (Fig.2a). Robot proprioception provides the gripper's pose, while the GelSlim 3.0 sensor [15] provides observations of the contact interface between the gripper finger and the object in the form of an RGB tactile image. The Apriltag attached to the gripper is solely employed for calibration purposes during the quantitative evaluation in Section IV-A and is not utilized as input to the system. The framework also takes as input the desired goal configuration and estimation priors (Fig.2b):

- **Desired Goal Configuration**: A combination of the manipulation objectives discussed in Section III-A.
- **Physics Parameter Priors**: The friction parameters at both the intrinsic contact (gripper/object) and the extrinsic contact (object/environment). These priors do not need to be accurate and are manually specified based on physical intuition.
- **Environment Priors**: The orientation and height of the environment in the world frame.

Utilizing these inputs, our **simultaneous tactile estimator-controller** (Fig.2c) calculates pose estimates for the object, along with a motion plan to achieve the manipulation objectives (Fig.2d). This updated motion plan guides the robot's motion. The framework comprises two main components: **discrete pose estimator** and **continuous pose estimator-controller**, which are described in the next subsections.

### C. Discrete Pose Estimator

The discrete pose estimator computes a probability distribution within a discretized grid of relative gripper/object poses. We first reconstruct a binary mask over the region of contact from raw RGB tactile images using a pixel-to-pixel convolutional neural network (CNN) model as described in [16]. Subsequently, the binary mask is channeled into the Tac2Pose estimator [16], which generates a distribution over possible object poses from a single contact mask.

We then filter the stream of tactile information and fuse it with the environment prior using the Viterbi algorithm [51], yielding a filtered probability distribution of the relative object pose. We discretize the pose space with 5mm of translational resolution and 10°of rotational resolution. The discretized state space consists of 5k-9k poses, depending on the object size. The inference step takes 2-6 seconds per iteration using PGMax [52], yielding a slow and coarse but global object pose signal.

Fig. 3a provides insight into the architecture of the discrete pose estimator, where $\mathcal{X} \in SE(2)$ represents the relative pose between the gripper and the object. At the initial timestep, the environment prior is introduced. Given our prior knowledge of the environment's orientation and height, we can, for each discrete relative object pose within the grid, ascertain which point of the object would be in closest proximity to the environment and compute the corresponding distance. The integration of the environment prior involves the multiplication of a Gaussian function over these distances. In simpler terms, it assigns higher probabilities to the relative poses that are predicted to be closer to the environment.

Subsequently, we incorporate the single-shot tactile pose estimation distribution at every $n^{th}$ step of the continuous pose estimator-controller, where $n$ is approximately five (see Fig. 3a), since the discrete pose estimator runs slower than the continuous pose estimator-controller. Instead of integrating tactile observations at a fixed frequency, we add the next tactile observation as soon as the discrete filter is ready, once the marginalization step to incorporate the previous tactile observation has been completed.

The transition probabilities impose constraints on tactile observations between consecutive time steps in the discrete graph, including:

- *Continuity*: The pose can transition only to neighboring poses on the pose grid to encourage continuity.
- *Persistent Contact*: The height of the closest point to the environment remains consistent across time steps due to the flat nature of the environment. This consistency is enforced through the multiplication of a Gaussian function that factors in the height difference.

Together, they encode the assumption that the object slides continuously within the grasp. This enables the discrete pose estimator to compute and filter the distribution of relative gripper/object poses, taking into account tactile information, robot proprioception, and environmental priors.

### D. Continuous Pose Estimator-Controller

The continuous pose estimator-controller serves a dual purpose: it takes as input the filtered discrete probability distribution of relative gripper/object poses and outputs a continuous pose estimate and an iteratively updated motion plan in a receding horizon fashion. The Incremental Smoothing and Mapping (iSAM) algorithm [53], which is based on the factor graph model [54], [55], serves as the computational backbone of our estimator-controller. We leverage its graph-based flexible formulation to combine estimation and control objectives as part of one single optimization problem.

The factor graph architecture of the continuous pose estimator-controller is illuminated in Fig. 3b. Noteworthy variables include $g_t$, $o_t$, and $c_t$, each frames in $SE(2)$, representing the gripper pose, object pose, and contact position, respectively. The orientation of $c_t$ is fixed and aligned with the normal direction of the environment. Additionally, $\mathcal{V}$ represents the set of physics parameters:

- Translational-to-rotational friction ratio at the grasp: $F_{max}/M_{max}$, where $F_{max}$ and $M_{max}$ are the maximum pure force and torque that it can endure before sliding.
- Friction coefficient at the extrinsic contact between the object and the environment: $\mu_{max}$.

The framework closely resembles that of earlier work [28], where further explanation of the iSAM implementation can be found. We encourage readers to refer to this seminal work for a description of the factors that are directly borrowed from [28] ($F_{cc}, F_{oc}, F_{gp}, F_{motion}$).

The continuous estimator-controller comprises two main sections: the left segment, spanning from the initial time to the current moment $t$, is dedicated to the **estimation** of the object's pose. The right segment, covering the time from $t$ to the control horizon $t+T$, is responsible for devising a motion plan to **control** the system and achieving the manipulation objectives. In the following sections, we define each new factor. The arguments of each factor definition are the variables, priors, and observations that the factor depends on. The right-hand side specifies the quantity we are trying to optimize.

**Estimation.** Within the estimation segment, the factor graph takes filtered pose estimations from the discrete pose estimator:

$$F_{tac}(g_i, o_i; \mathcal{X}_{i,\text{MAP}}) = \mathcal{X}_{i,\text{MAP}}^{-1}(g_i^{-1} o_i), \tag{1}$$

where $\mathcal{X}_{i,\text{MAP}}$ denotes the filtered maximum a posteriori (MAP) discrete relative pose, and $(g_i^{-1} o_i)$ denotes the continuous estimate of the gripper/object relative pose. Given the higher operating speed of the continuous pose estimator-controller (0.1∼0.2 seconds per iteration) compared to the discrete pose estimator (2∼6 seconds per iteration), the discrete pose estimation factor is integrated when an update is available every few steps within the continuous estimator-controller.

Similar to the discrete pose estimator, the environment (contact) prior is established during the initial time step:

$$F_{cp}(c_1; c^*) = c^{*-1} c_1, \tag{2}$$

where $c^* \in SE(2)$ contains the prior information about the environment's orientation and its height.

Additionally, physics priors are imposed:

$$F_{vp}(\mathcal{V}; \mathcal{V}^*) = \mathcal{V} - \mathcal{V}^*. \tag{3}$$

where $\mathcal{V}^*$ is the prior for the physics parameters.

Furthermore, we impose a friction model based on the limit-surface model [43], [56] as a transition model to capture the dynamics of sliding ($F_{fric}$). This model provides a relation between the kinetic friction wrench and the direction of sliding at the grasp. In essence, it serves as a guide for predicting how the object will slide in response to a given gripper motion and extrinsic contact location. The correlation is formally represented as follows:

$$[\omega, v_x, v_y] \propto [\frac{M}{M_{max}^2}, \frac{F_x}{F_{max}^2}, \frac{F_y}{F_{max}^2}]. \tag{4}$$

Here, $[\omega, v_x, v_y]$ denotes the relative object twist in the gripper's frame, i.e. sliding direction, while $[M, F_x, F_y]$ signifies the friction wrench at the grasp. To fully capture the friction dynamics, additional kinematic and mechanical constraints at the extrinsic contact are also considered. These constraints are formulated as follows:

$$M\hat{z} - \vec{l}_{gc} \times \vec{F} = 0, \tag{5}$$

$$v_{c,N}(g_{i-1}, o_{i-1}, c_{i-1}, g_i, o_i) = 0, \tag{6}$$

$$v_{c,T}(g_{i-1}, o_{i-1}, c_{i-1}, g_i, o_i) = 0 \tag{7}$$

$$\perp (F_T = -\mu_{max} F_N \text{ OR } F_T = \mu_{max} F_N), \tag{8}$$

In these equations, $\vec{l}_{gc}$ is the vector from the gripper to the contact point, and $v_{c,N}$ and $v_{c,T}$ represent the local velocities of the object at the point of contact in the directions that are normal and tangential to the environment, respectively. $F_N$ and $F_T$ denote the normal and tangential components of the force. Eq. 5 specifies that no net torque should be present at the point of extrinsic contact since we are assuming point contact. Eq. 6 dictates that the normal component of the local velocity at the point of extrinsic contact must be zero as long as contact is maintained. Eq. 7 and Eq. 8 work complementarily to stipulate that the tangential component of the local velocity at the contact point must be zero (Eq. 7), except in cases where the contact is sliding. In such instances, the contact force must lie on the boundary of the friction cone (Eq. 8). By combining Eq. 4∼8, we establish a fully determined forward model for the contact and object poses, which allows the object pose at step $i$ to be expressed as a function of its previous poses, the current gripper pose, and the physics parameters:

$$o_i^* = f(g_{i-1}, o_{i-1}, c_{i-1}, g_i, \mathcal{V}) \tag{9}$$

This relationship can thus be encapsulated as a friction factor:

$$F_{fric}(g_{i-1}, o_{i-1}, c_{i-1}, g_i, o_i, \mathcal{V}) = o_i^{*-1} o_i. \tag{10}$$

Together, the estimation component formulates a smooth object pose trajectory that takes into account tactile measurements, robot kinematics, and physics model.

**Control.** The control segment incorporates multiple auxiliary factors to facilitate the specification of regrasping objectives. First, the desired goal configuration is imposed at the end

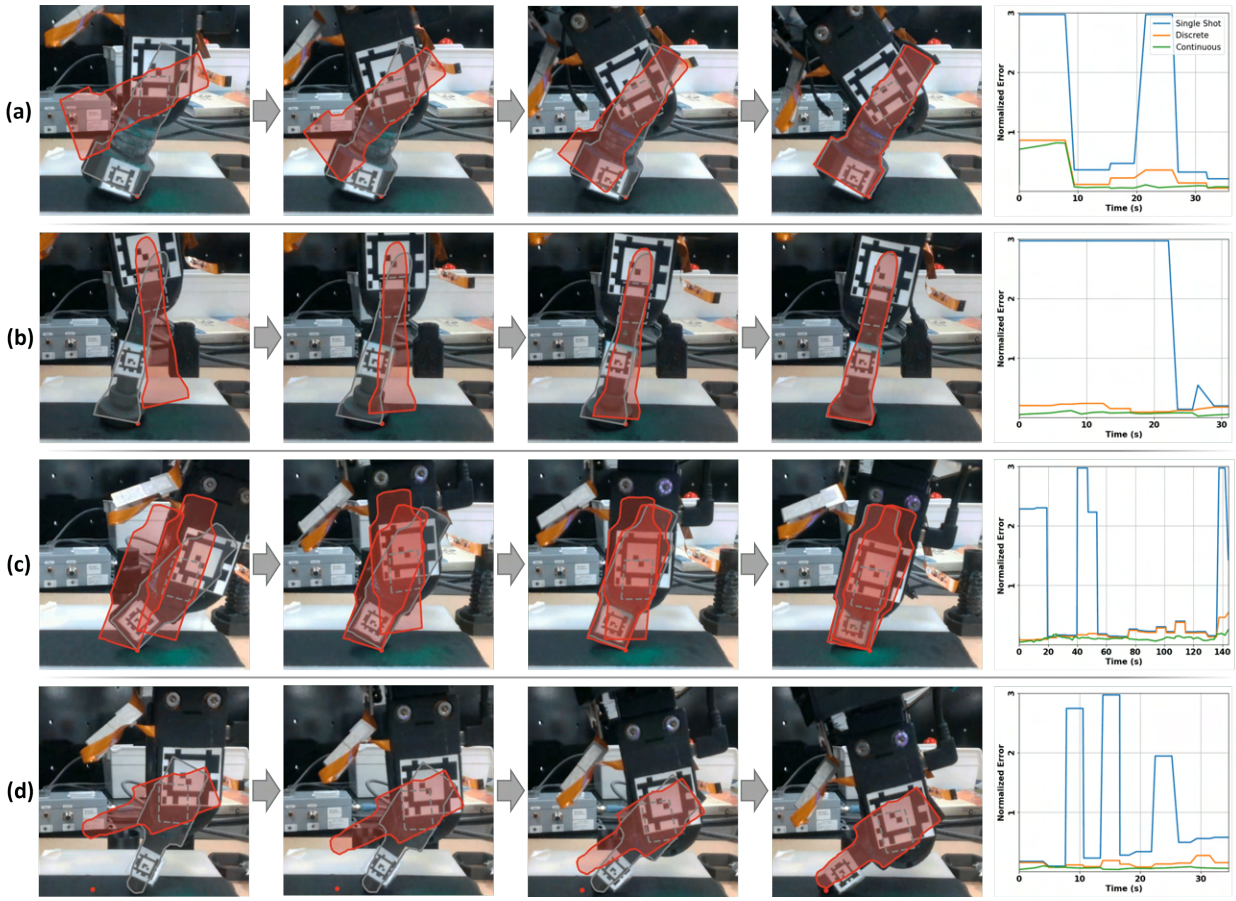


**Fig. 5:** Demonstrations of four types of goal configurations: (a) Relative Orientation + Stationary Extrinsic Contact, (b) Relative Orientation/Translation + Stationary Extrinsic Contact, (c) Relative Orientation + Global Orientation + Stationary Extrinsic Contact, and (d) Relative Orientation + Sliding Extrinsic Contact. The right column depicts normalized estimation accuracy for the proposed method and ablation models.

**TABLE I.** Estimation Performance Comparison

| | AUX | Pin | Stud | USB | Overall |
|---|---|---|---|---|---|
| Single Shot (Tac2Pose [16]) | 1.02 (5.0 mm / 52.8 deg) | 2.47 (10.3 mm / 131.6 deg) | 1.45 (13.6 mm / 67.3 deg) | 1.14 (6.85 mm / 57.0 deg) | **1.52** (8.3 mm / 78.0 deg) |
| Discrete Est. (this paper) | 0.17 (2.8 mm / 6.8 deg) | 0.22 (4.12 mm / 8.7 deg) | 0.20 (4.6 mm / 5.9 deg) | 0.20 (4.8 mm / 5.6 deg) | **0.20** (3.9 mm / 6.9 deg) |
| Continuous Est. (this paper) | 0.09 (1.7 mm / 3.2 deg) | 0.10 (1.9 mm / 3.9 deg) | 0.08 (2.2 mm / 2.1 deg) | 0.11 (2.5 mm / 3.4 deg) | **0.10** (2.0 mm / 3.3 deg) |

of the control horizon ($F_{goal}$). This comprises three distinct sub-factors, corresponding to the three manipulation objectives described in Section III-A, which can be turned on or off, depending on the desired configuration:

1) $F_{\text{goal,go}}$ regulates the desired gripper/object relative pose at $o_{t+T}$ and $g_{t+T}$.
2) $F_{\text{goal,o}}$ enforces the object's orientation within the world frame at $o_{t+T}$.
3) $F_{\text{goal,c}}$ dictates the desired contact point at $c_{t+T}$, thereby facilitating controlled sliding interactions with the environment.

These sub-factors are mathematically expressed as follows:

$$F_{goal,go}(g_{t+T}, o_{t+T}) = p^{g\ -1}_{o,goal}(g^{-1}_{t+T} o_{t+T}), \quad (11)$$

$$F_{goal,o}(o_{t+T}) = o^{-1}_{goal} o_{t+T}, \quad (12)$$

$$F_{goal,c}(c_{t+T}) = c^{-1}_{goal} c_{t+T}. \quad (13)$$

Here, $p^{g}_{o,goal}$ signifies the target relative gripper/object pose, $o_{goal}$ represents the desired object orientation in the world frame, and $c_{goal}$ is the intended contact point.

Additionally, the $F_{motion}$ factor minimizes the gripper motion across consecutive time steps, encouraging motion smoothness. Concurrently, a contact maintenance factor, $F_{cm}$, serves as a soft constraint to direct the gripper's motion in a way that prevents it from losing contact with the environment:

$$F_{cm}(g_{i-1}, c_{i-1}, g_i; \epsilon_i) = max(0, \zeta_i(g_{i-1}, c_{i-1}, g_i) + \epsilon_i), \quad (14)$$

$\zeta_i$ represents the normal component of the virtual local displacement from step $i-1$ to $i$ at the contact point. The term $\epsilon_i$ is a small positive scalar, encouraging $\zeta_i$ to be negative, thus fostering a motion that pushes the object against the environment.

Taken together, these factors cohesively formulate a motion plan, which is then communicated to the robot. The robot continues to follow the interpolated trajectory of this motion plan until it receives the next update.

## IV. Experiments and Results

We conducted a series of experiments on four distinct 3D-printed objects (illustrated in Fig. 4) to validate the efficacy of our algorithm. The experiments were designed to:

1) Evaluate the algorithm's performance across a variety of target configurations.
2) Assess the algorithm's applicability to specific real-world tasks, such as object insertion.

### A. Performance Across Various Goal Configurations

We assessed our algorithm's performance using a total of 18 diverse goal configurations. Our framework allows for specifying goals relative to the gripper (regrasping) and relative to the world frame (reorienting), facilitating different downstream tasks. For example, regrasping can improve grasp stability, enable tactile exploration, and establish a grasp optimized for both force execution and the avoidance of collisions or kinematic singularities in downstream tasks. On the other hand, reorienting the object can enable mating with target objects in the environment or prevent collisions with obstacles. The configurations we evaluate fall into four distinct categories:

- Relative Orientation + Stationary Extrinsic Contact
- Relative Orientation/Translation + Stationary Extrinsic Contact
- Relative Orientation + Global Orientation + Stationary Extrinsic Contact
- Relative Orientation + Sliding Extrinsic Contact

Examples of these four goal configuration types are illustrated in Fig. 5, along with corresponding plots showcasing estimation accuracy. The red silhouettes that move along with the gripper represent the desired relative pose between the gripper and the object. The grey silhouettes depict object poses as measured by Apriltags, which we use as the ground truth object pose. The red dots mark the desired extrinsic contact location. In Fig. 5c, the other red silhouette signifies the desired object orientation in the global frame. The time series plots in the right column depict the normalized estimation error of the proposed model ('Continuous'), alongside two ablation models: 1) single-shot Tac2Pose estimation ('Single Shot'), 2) discrete filtered estimation from Viterbi decoding ('Discrete'). These results attest to the algorithm's adeptness in attaining desired goal configurations while showing better estimation performance compared to the ablation models.

A summary of each algorithm's average estimation performance is presented in Table I. The values denote the normalized estimation error, computed as follows:

$$\epsilon_{\text{norm}} = ||(\epsilon_{\text{rot}}, \epsilon_{\text{trn}}/(l_{\text{obj}}/2)||_1 \tag{15}$$

Here, $||\cdot||_1$ signifies the L1-norm, $\epsilon_{\text{rot}}$ indicates rotation error in radians, $\epsilon_{\text{trn}}$ denotes translation error, and $l_{\text{obj}}$ represents the object's length. This analysis reveals a marked reduction in overall normalized error, progressing from 1.52 to 0.20 when transitioning from single-shot estimation to the discrete pose estimator's filtered estimation. The single-shot estimator suffers due to ambiguity in individual tactile images, as explored thoroughly in [16]. For most grasps of the objects we experiment with, a single tactile imprint is not sufficient to uniquely localize the object. The discrete pose estimator is able to reduce ambiguity by fusing information over a sequence of tactile images, obtained by traversing the object surface and therefore exposing the estimator to a more complete view of the object geometry. The discrete pose estimator also consumes information about the ground height and orientation, providing an additional constraint on the object pose. In this way, the tasks of tactile object pose estimation and in-hand manipulation are synergistic: tactile object pose estimation supervises in-hand manipulation, while in-hand manipulation allows object pose to be unambiguously estimated from tactile images. Furthermore, a subsequent improvement is observed, decreasing from 0.20 to 0.10 when employing the continuous pose estimator-controller. This improvement comes from refining the estimation accuracy beyond the resolution of the discrete grid.

### B. Real-World Application: Insertion Task

To validate our algorithm's practical utility, we applied it to a specific downstream task — object insertion with small clearance (1∼0.5 mm). For these experiments, we sampled random goal configurations from the first category (adjusting relative orientation) described in Section IV-A. Following this, we aimed to insert the grasped object into holes with 1 mm and 0.5 mm total clearance in diameter.

Table II summarizes the outcomes of these insertion attempts. The AUX connector, which features a tapered profile at the tip, had a success rate exceeding 90%. On the other hand, the success rate dropped considerably for objects with untapered profiles, especially when the clearance was narrowed from 1 mm to 0.5 mm. The varying performance is consistent with our expectations, given that the algorithm's average pose estimation accuracy is approximately 2.0 mm as quantified in Section IV-A.

These findings indicate that our algorithm is useful in tasks that necessitate regrasping and reorienting objects to fulfill downstream objectives by meeting the goal configuration. However, for applications requiring sub-millimeter accuracy, the algorithm's performance would benefit from integration with a compliant controlled insertion policy (e.g., [19], [22], [57], [58]).

**TABLE II.** Insertion Experiment Results (Success/Attempt)

| Clearance | AUX | Pin | Stud | USB |
|---|---|---|---|---|
| 1 mm | 10 / 10 | 6 / 10 | 7 / 10 | 7 / 10 |
| 0.5 mm | 9 / 10 | 3 / 10 | 5 / 10 | 6 / 10 |

## V. Conclusions

This paper introduces a novel simultaneous tactile estimator-controller tailored for in-hand object manipulation. The framework harnesses extrinsic dexterity to regrasp a grasped object while simultaneously estimating object poses. We demonstrate the framework for the particular case of precisely controlling sliding regrasps by pushing against an external surface. This innovation holds particular promise in scenarios necessitating object or grasp reorientation for tasks like insertion or tool use, particularly in cases where the precise visual perception of the object's global pose is difficult due to occlusions. In future research, we aim to explore methodologies for autonomously determining optimal target orientations for task execution, rather than relying on manual specification.

## References

[1] M. Bauza, F. R. Hogan, and A. Rodriguez, "A data-efficient approach to precise and controlled pushing," in *Conference on Robot Learning*. PMLR, 2018, pp. 336–345.

[2] I. Mordatch, Z. Popović, and E. Todorov, "Contact-invariant optimization for hand manipulation," in *Proceedings of the ACM SIGGRAPH/Eurographics symposium on computer animation*, 2012, pp. 137–144.

[3] B. Sundaralingam and T. Hermans, "Geometric in-hand regrasp planning: Alternating optimization of finger gaits and in-grasp manipulation," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 231–238.

[4] Y. Hou, Z. Jia, and M. T. Mason, "Fast planning for 3d any-pose-reorienting using pivoting," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 1631–1638.

[5] J. Shi, J. Z. Woodruff, P. B. Umbanhowar, and K. M. Lynch, "Dynamic in-hand sliding manipulation," *IEEE Transactions on Robotics*, vol. 33, no. 4, pp. 778–795, 2017.

[6] B. Sundaralingam and T. Hermans, "Relaxed-rigidity constraints: kinematic trajectory optimization and collision avoidance for in-grasp manipulation," *Autonomous Robots*, vol. 43, pp. 469–483, 2019.

[7] T. Chen, J. Xu, and P. Agrawal, "A system for general in-hand object re-orientation," in *Conference on Robot Learning*. PMLR, 2022, pp. 297–307.

[8] A. Rajeswaran, V. Kumar, A. Gupta, G. Vezzani, J. Schulman, E. Todorov, and S. Levine, "Learning complex dexterous manipulation with deep reinforcement learning and demonstrations," *arXiv preprint arXiv:1709.10087*, 2017.

[9] T. Chen, M. Tippur, S. Wu, V. Kumar, E. Adelson, and P. Agrawal, "Visual dexterity: In-hand dexterous manipulation from depth," in *Icml workshop on new frontiers in learning, control, and dynamical systems*, 2023.

[10] A. Handa, A. Allshire, V. Makoviychuk, A. Petrenko, R. Singh, J. Liu, D. Makoviichuk, K. Van Wyk, A. Zhurkevich, B. Sundaralingam *et al.*, "Dextreme: Transfer of agile in-hand manipulation from simulation to reality," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2023, pp. 5977–5984.

[11] O. M. Andrychowicz, B. Baker, M. Chociej, R. Jozefowicz, B. McGrew, J. Pachocki, A. Petron, M. Plappert, G. Powell, A. Ray *et al.*, "Learning dexterous in-hand manipulation," *The International Journal of Robotics Research*, vol. 39, no. 1, pp. 3–20, 2020.

[12] W. Huang, I. Mordatch, P. Abbeel, and D. Pathak, "Generalization in dexterous manipulation via geometry-aware multi-task learning," *arXiv preprint arXiv:2111.03062*, 2021.

[13] W. Yuan, S. Dong, and E. H. Adelson, "Gelsight: High-resolution robot tactile sensors for estimating geometry and force," *Sensors*, vol. 17, no. 12, p. 2762, 2017.

[14] M. Lambeta, P.-W. Chou, S. Tian, B. Yang, B. Maloon, V. R. Most, D. Stroud, R. Santos, A. Byagowi, G. Kammerer *et al.*, "Digit: A novel design for a low-cost compact high-resolution tactile sensor with application to in-hand manipulation," *IEEE Robotics and Automation Letters*, vol. 5, no. 3, pp. 3838–3845, 2020.

[15] I. H. Taylor, S. Dong, and A. Rodriguez, "Gelslim 3.0: High-resolution measurement of shape, force and slip in a compact tactile-sensing finger," in *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 2022, pp. 10 781–10 787.

[16] M. Bauza, A. Bronars, and A. Rodriguez, "Tac2pose: Tactile object pose estimation from the first touch," *The International Journal of Robotics Research*, vol. 42, no. 13, pp. 1185–1209, 2023.

[17] S. Pai, T. Chen, M. Tippur, E. Adelson, A. Gupta, and P. Agrawal, "Tactofind: A tactile only system for object retrieval," *arXiv preprint arXiv:2303.13482*, 2023.

[18] S. Luo, W. Yuan, E. Adelson, A. G. Cohn, and R. Fuentes, "Vitac: Feature sharing between vision and tactile sensing for cloth texture recognition," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 2722–2727.

[19] S. Kim and A. Rodriguez, "Active extrinsic contact sensing: Application to general peg-in-hole insertion," in *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 2022, pp. 10 241–10 247.

[20] D. Ma, S. Dong, and A. Rodriguez, "Extrinsic contact sensing with relative-motion tracking from distributed tactile measurements," in *2021 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2021, pp. 11 262–11 268.

[21] C. Higuera, S. Dong, B. Boots, and M. Mukadam, "Neural contact fields: Tracking extrinsic contact with tactile sensing," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2023, pp. 12 576–12 582.

[22] S. Dong, D. K. Jha, D. Romeres, S. Kim, D. Nikovski, and A. Rodriguez, "Tactile-rl for insertion: Generalization to objects of unknown geometry," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 6437–6443.

[23] F. R. Hogan, J. Ballester, S. Dong, and A. Rodriguez, "Tactile dexterity: Manipulation primitives with tactile feedback," in *2020 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2020, pp. 8863–8869.

[24] Y. Shirai, D. K. Jha, A. U. Raghunathan, and D. Hong, "Tactile tool manipulation," in *2023 International Conference on Robotics and Automation (ICRA)*. IEEE, 2023.

[25] Y. She, S. Wang, S. Dong, N. Sunil, A. Rodriguez, and E. Adelson, "Cable manipulation with a tactile-reactive gripper," *The International Journal of Robotics Research*, vol. 40, no. 12-14, pp. 1385–1401, 2021.

[26] N. Sunil, S. Wang, Y. She, E. Adelson, and A. Rodriguez, "Visuotactile affordances for cloth manipulation with local control," in *Conference on Robot Learning*. PMLR, 2023, pp. 1596–1606.

[27] N. C. Dafle, A. Rodriguez, R. Paolini, B. Tang, S. S. Srinivasa, M. Erdmann, M. T. Mason, I. Lundberg, H. Staab, and T. Fuhlbrigge, "Extrinsic dexterity: In-hand manipulation with external forces," in *2014 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2014, pp. 1578–1585.

[28] S. Kim, D. K. Jha, D. Romeres, P. Patre, and A. Rodriguez, "Simultaneous tactile estimation and control of extrinsic contact," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*, 2023, pp. 12 563–12 569.

[29] P. Sodhi, M. Kaess, M. Mukadam, and S. Anderson, "Learning tactile models for factor graph-based estimation," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 13 686–13 692.

[30] P. Sodhi, M. Kaess, M. Mukadanr, and S. Anderson, "Patchgraph: In-hand tactile tracking with learned surface normals," in *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 2022, pp. 2164–2170.

[31] J. Zhao, M. Bauza, and E. H. Adelson, "Fingerslam: Closed-loop unknown object localization and reconstruction from visuo-tactile feedback," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*, 2023, pp. 8033–8039.

[32] S. Suresh, M. Bauza, K.-T. Yu, J. G. Mangelson, A. Rodriguez, and M. Kaess, "Tactile slam: Real-time inference of shape and pose from planar pushing," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 11 322–11 328.

[33] M. Bauza, O. Canal, and A. Rodriguez, "Tactile mapping and localization from high-resolution tactile imprints," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 3811–3817.

[34] R. Li, R. Platt, W. Yuan, A. Ten Pas, N. Roscup, M. A. Srinivasan, and E. Adelson, "Localization and manipulation of small parts using gelsight tactile sensing," in *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2014, pp. 3988–3993.

[35] M. Bauza, A. Bronars, Y. Hou, I. Taylor, N. Chavan-Dafle, and A. Rodriguez, "simPLE: a visuotactile method learned in simulation to precisely pick, localize, regrasp, and place objects," *arXiv preprint arXiv:2307.13133*, 2023.

[36] S. Dikhale, K. Patel, D. Dhingra, I. Naramura, A. Hayashi, S. Iba, and N. Jamali, "Visuotactile 6d pose estimation of an in-hand object using vision and tactile sensor data," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 2148–2155, 2022.

[37] G. Izatt, G. Mirano, E. Adelson, and R. Tedrake, "Tracking objects with point clouds from vision and touch," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2017, pp. 4000–4007.

[38] T. Anzai and K. Takahashi, "Deep gated multi-modal learning: In-hand object pose changes estimation using tactile and image data," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 9361–9368.

[39] S. Suresh, Z. Si, S. Anderson, M. Kaess, and M. Mukadam, "Midastouch: Monte-carlo inference over distributions across sliding touch," in *Conference on Robot Learning*. PMLR, 2023, pp. 319–331.

[40] T. Kelestemur, R. Platt, and T. Padir, "Tactile pose estimation and policy learning for unknown object manipulation," *International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 2022.

[41] Y. Hou, Z. Jia, and M. Mason, "Manipulation with shared grasping," in *Robotics: Science and Systems (RSS)*, 2020.

[42] J. Shi, H. Weng, and K. M. Lynch, "In-hand sliding regrasp with spring-sliding compliance and external constraints," *IEEE Access*, vol. 8, pp. 88 729–88 744, 2020.

[43] N. Chavan-Dafle, R. Holladay, and A. Rodriguez, "In-hand manipulation via motion cones," in *Robotics: Science and Systems (RSS)*, 2018.

[44] N. Chavan-Dafle and A. Rodriguez, "Prehensile pushing: In-hand manipulation with push-primitives," in *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2015, pp. 6215–6222.

[45] A. Nagabandi, K. Konolige, S. Levine, and V. Kumar, "Deep dynamics models for learning dexterous manipulation," in *Conference on Robot Learning*. PMLR, 2020, pp. 1101–1112.

[46] V. Kumar, E. Todorov, and S. Levine, "Optimal control with learned local models: Application to dexterous manipulation," in *2016 IEEE*

*International Conference on Robotics and Automation (ICRA)*. IEEE, 2016, pp. 378–383.

[47] S. Tian, F. Ebert, D. Jayaraman, M. Mudigonda, C. Finn, R. Calandra, and S. Levine, "Manipulation by feel: Touch-based control with deep predictive models," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 818–824.

[48] M. Lepert, C. Pan, S. Yuan, R. Antonova, and J. Bohg, "In-hand manipulation of unknown objects with tactile sensing for insertion," in *Embracing Contacts-Workshop at ICRA 2023*, 2023.

[49] J. Pitz, L. Röstel, L. Sievers, and B. Bäuml, "Dextrous tactile in-hand manipulation using a modular reinforcement learning architecture," *arXiv preprint arXiv:2303.04705*, 2023.

[50] H. Van Hoof, T. Hermans, G. Neumann, and J. Peters, "Learning robot in-hand manipulation with tactile features," in *2015 IEEE-RAS 15th International Conference on Humanoid Robots (Humanoids)*. IEEE, 2015, pp. 121–127.

[51] G. D. Forney, "The viterbi algorithm," *Proceedings of the IEEE*, vol. 61, no. 3, pp. 268–278, 1973.

[52] G. Zhou, N. Kumar, A. Dedieu, M. Lázaro-Gredilla, S. Kushagra, and D. George, "Pgmax: Factor graphs for discrete probabilistic graphical models and loopy belief propagation in jax," *arXiv preprint arXiv:2202.04110*, 2022.

[53] M. Kaess, H. Johannsson, R. Roberts, V. Ila, J. J. Leonard, and F. Dellaert, "isam2: Incremental smoothing and mapping using the bayes tree," *The International Journal of Robotics Research*, vol. 31, no. 2, pp. 216–235, 2012.

[54] F. Dellaert, "Factor graphs and gtsam: A hands-on introduction," *Georgia Institute of Technology, Tech. Rep*, vol. 2, p. 4, 2012.

[55] F. Dellaert, M. Kaess *et al.*, "Factor graphs for robot perception," *Foundations and Trends® in Robotics*, vol. 6, no. 1-2, pp. 1–139, 2017.

[56] S. Goyal, "Planar sliding of a rigid body with dry friction: limit surfaces and dynamics of motion," Ph.D. dissertation, Cornell University Ithaca, NY, 1989.

[57] T. Inoue, G. De Magistris, A. Munawar, T. Yokoya, and R. Tachibana, "Deep reinforcement learning for high precision assembly tasks," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2017, pp. 819–825.

[58] T. Z. Zhao, J. Luo, O. Sushkov, R. Pevceviciute, N. Heess, J. Scholz, S. Schaal, and S. Levine, "Offline meta-reinforcement learning for industrial insertion," in *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 2022, pp. 6386–6393.