

# Tactile SLAM: Real-time inference of shape and pose from planar pushing

Sudharshan Suresh<sup>1</sup>, Maria Bauza<sup>2</sup>, Kuan-Ting Yu<sup>3</sup>,  
Joshua G. Mangelson<sup>4</sup>, Alberto Rodriguez<sup>2</sup>, and Michael Kaess<sup>1</sup>

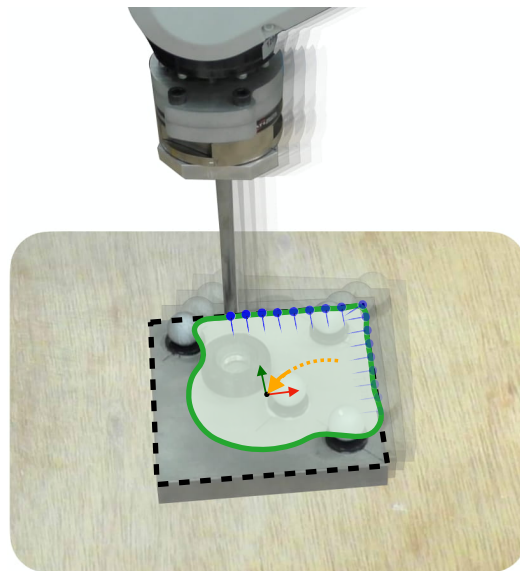
**Abstract**—Tactile perception is central to robot manipulation in unstructured environments. However, it requires contact, and a mature implementation must infer object models while also accounting for the motion induced by the interaction. In this work, we present a method to estimate both object shape and pose in real-time from a stream of tactile measurements. This is applied towards tactile exploration of an unknown object by planar pushing. We consider this as an online SLAM problem with a nonparametric shape representation. Our formulation of tactile inference alternates between Gaussian process implicit surface regression and pose estimation on a factor graph. Through a combination of local Gaussian processes and fixed-lag smoothing, we infer object shape and pose in real-time. We evaluate our system across different objects in both simulated and real-world planar pushing tasks.

## I. INTRODUCTION

For effective interaction, robot manipulators must build and refine their understanding of the world through sensing. This is especially relevant in unstructured settings, where robots have little to no knowledge of object properties, but can physically interact with their surroundings. Even when blindfolded, humans can locate and infer properties of unknown objects through touch [1]. Replicating some fraction of these capabilities will enable contact-rich manipulation in environments such as homes and warehouses. In particular, knowledge of object shape and pose determines the success of generated grasps or nonprehensile actions.

While there have been significant advances in tactile sensing, from single-point sensors to high-resolution tactile arrays, a general technique for the underlying inference still remains an open question [2]. Visual and depth-based tracking have been widely studied [3], but suffer from occlusion due to clutter or self-occlusions with the gripper or robot. We provide a general formulation of pure tactile inference, that could later accommodate additional sensing modalities.

Pure tactile inference is challenging because, unlike vision, touch cannot directly provide global estimates of object model or pose. Instead, it provides detailed, local information



**Fig. 1:** The shape and pose estimation problem in a pusher-slider system. A robot manipulator pushes along a planar object, while recording a stream of tactile measurements. Our method builds a shape contour in real-time as a Gaussian process implicit surface, and optimizes for pose via geometry and physics-based constraints. Figure shows tactile measurements (•), estimated motion (→), estimated shape/pose (○), and ground-truth (---).

that must be fused into a global model. Moreover, touch is intrusive: the act of sensing itself constantly perturbs the object. We consider tactile inference as an analog of the well-studied simultaneous localization and mapping (SLAM) problem in mobile robotics [4]. Errors in object tracking accumulate to affect its predicted shape, and vice versa.

Central to this problem is choosing a shape representation that both faithfully approximates arbitrary geometries, and is amenable to probabilistic updates. This excludes most parametric models such as polygons/polyhedrons [5], superquadrics [6], voxel maps [7], point-clouds [8], and standard meshes [9]. Gaussian process implicit surfaces (GPIS) [10] are one such nonparametric shape representation that satisfies these requirements.

In this paper, we demonstrate online shape and pose estimation for a planar pusher-slider system. We perform tactile exploration of the object via contour following, that generates a stream of contact and force measurements. Our novel schema combines efficient GPIS regression with factor graph optimization over geometric and physics-based constraints. The problem is depicted in Fig. 1: the pusher moves along the object while estimating its shape and pose.

We expand the scope of the batch-SLAM method by Yu et al. [5] with a more meaningful shape representation, and

<sup>1</sup>Sudharshan Suresh and Michael Kaess are with the Robotics Institute, Carnegie Mellon University <suddhu, kaess>@cmu.edu

<sup>2</sup>Maria Bauza and Alberto Rodriguez are with the Mechanical Engineering Department, Massachusetts Institute of Technology <bauza, albertor>@mit.edu

<sup>3</sup>Kuan-Ting Yu is with XYZ Robotics peterkty@gmail.com

<sup>4</sup>Joshua G. Mangelson is with the Electrical and Computer Engineering Department, Brigham Young University joshua\_mangelson@byu.edu

This work was partially supported by the National Science Foundation under award IIS-2008279. We thank Eric Dexheimer, Daniel McGann, Allison Wong, and Paloma Sodhi for their discussions and feedback.

real-time online inference. Our contributions are:

- (1) A formulation of the tactile SLAM problem that alternates between GPIS shape regression and sparse nonlinear incremental pose optimization,
- (2) Efficient implicit surface generation from touch using overlapping local Gaussian processes (GPs),
- (3) Fixed-lag smoothing over contact, geometry, and frictional pushing mechanics to localize objects,
- (4) Results from tactile exploration across different planar object shapes in simulated and real settings.

## II. RELATED WORK

### A. SLAM and object manipulation

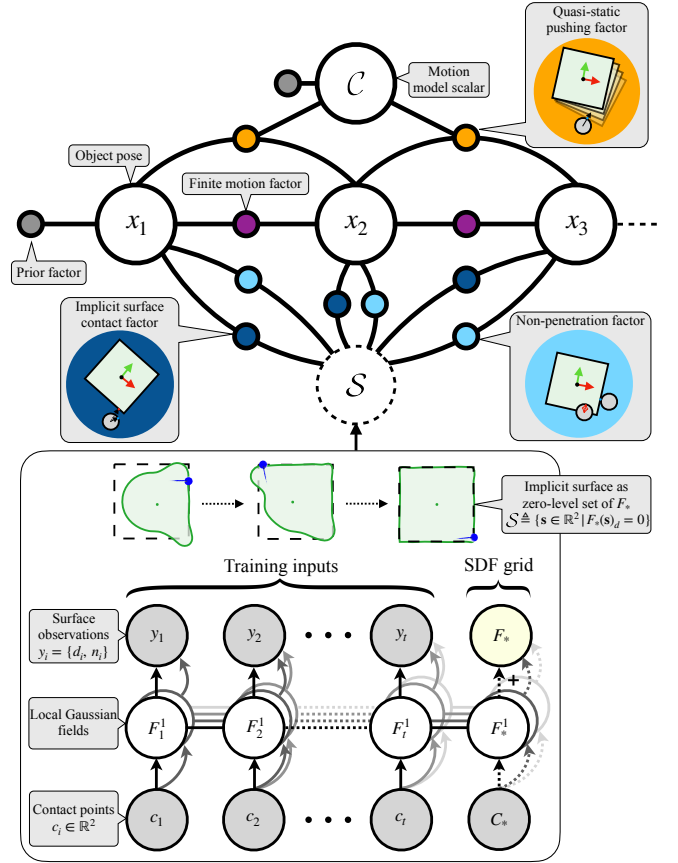
Our work is closely related to that of Yu et al. [5], that recovers shape and pose from tactile exploration of a planar object. The approach uses contact measurements and the well-understood mechanics of planar pushing [11, 12] as constraints in a batch optimization. Naturally, this is expensive and unsuitable for online tactile inference. Moreover, the object shape is represented as ordered control points, to form a piecewise-linear polygonal approximation. Such a representation poorly approximates arbitrary objects, and fails when data-association is incorrect. Moll and Erdmann [13] consider the illustrative case of reconstructing motion and shape of smooth, convex objects between two planar palms. Strub et al. [14] demonstrate the full SLAM problem with a dexterous hand equipped with tactile arrays.

Contemporaneous research considers one of two simplifying assumptions: modeling shape with fixed pose [8, 15, 16, 17, 18], or localizing with known shape [19, 20, 21, 22, 23]. The extended Kalman filter (EKF) has been used in visuo-tactile methods [24, 25], but is prone to linearization errors. At each timestep, it linearizes about a potentially incorrect current estimate, leading to inaccurate results. Smoothing methods [26] are more accurate as they preserve a temporal history of costs, and solve a nonlinear least-squares problem. These frameworks have been used to track known objects with vision and touch [27, 28], and rich tactile sensors [23].

### B. Gaussian process implicit surfaces

A continuous model which can generalize without discretization errors is of interest to global shape perception. Implicit surfaces have long been used for their smoothness and ability to express arbitrary topologies [29]. Using Gaussian processes [30] as their surface potential enables probabilistic fusion of noisy measurements, and reasoning about shape uncertainty. GPIS were formalized by Williams and Fitzgibbon [10], and were later used by Dragiev et al. to learn shape from grasping [15]. It represents objects as a signed distance field (SDF): the signed distance of spatial grid points to the nearest object surface. The SDF and surface uncertainty were subsequently used for active tactile exploration [17, 31, 32]. To our knowledge, no methods use GPIS alongside pose estimation for manipulation tasks.

Online GP regression scales poorly due to the growing cost of matrix inversion [30]. Spatial mapping applications address this by either sparse approximations to the full



**Fig. 2:** The combined formulation between the factor graph (Section V) and GPIS (Section IV). **[top]** The graph illustrates the relationship between the variables to be optimized for (circles) and the factors that act as constraints (colored dots). **[bottom]** Our GPIS builds an implicit surface shape representation that is the zero level-set of GP potential function. Spatial partitioning with local GPs enables efficient regression.

GP [33], or training separate local GPs [34, 35]. Lee et al. [36] propose efficient, incremental updates to the GPIS map through spatial partitioning.

## III. PROBLEM FORMULATION: TACTILE SLAM

We consider a rigid planar object on a frictional surface, moved around by a pusher (see Fig. 1). The interaction is governed by simple contour following for tactile exploration. Given a stream of tactile measurements, we estimate the 2-D shape and object's planar pose in real-time.

**Object pose:** The object pose at the current timestep  $t$  is defined by position and orientation  $\mathbf{x}_t = (x, y, \theta) \in SE(2)$ .

**Object shape:** The estimated object shape is represented as an implicit surface  $S \in \mathbb{R}^2$  in the reference frame of  $\mathbf{x}_t$ .

**Tactile measurements:** At every timestep we observe:

$$\mathbf{z}_t = \left\{ \underbrace{\mathbf{p}_t \in \mathbb{R}^2}_{\text{pusher position}}, \underbrace{\mathbf{f}_t \in \mathbb{R}^2}_{\text{force vector}}, \underbrace{\Theta_t \in \{0, 1\}}_{\text{contact/no-contact}} \right\} \quad (1)$$

Pusher position is obtained from the robot's distal joint state, and force is sensed directly from a force/torque sensor. Contact is detected with a minimum force threshold, and the estimated contact point  $\mathbf{c}_t$  is derived from knowledge

of  $\mathbf{p}_t$ ,  $\mathbf{f}_t$  and probe radius  $r_{\text{probe}}$ . This is consistent with the formulation in [5], with the addition of direct force measurements  $\mathbf{f}_t$ . For simplicity we consider a single pusher, but it can be expanded to multiple pushers or a tactile array.

**Assumptions:** We make a minimal set of assumptions, similar to prior work in planar pushing [5, 27, 28]:

- Quasi-static interactions and limit surface model [12, 37],
- Uniform friction and pressure distribution between bodies,
- Object’s rough scale and initial pose, and
- No out-of-plane effects.

The remainder of the paper is organized as such: Section IV formulates building shape estimate  $\mathcal{S}$  as the implicit surface of a GP, given current pose estimate and measurement stream. Section V describes factor graph inference to obtain pose estimate  $\mathbf{x}_t^*$  with  $\mathcal{S}$  and the measurement stream. These two processes are carried out back-and-forth for online shape and pose hypothesis at each timestep (Fig. 2). Finally, we demonstrate its use in simulated and real experiments (Section VI) and present concluding remarks (Section VII).

#### IV. SHAPE ESTIMATION WITH IMPLICIT SURFACES

##### A. Gaussian process implicit surface

A Gaussian process learns a continuous, nonlinear function from sparse, noisy datapoints [30]. Surface measurements are in the form of contact points  $\mathbf{c}_i$  and normals  $\mathbf{n}_i$ , transformed to an object-centric reference frame. Given  $N$  datapoints, the GP learns a mapping  $X \mapsto Y$  between them:

$$F : \underbrace{\{c_{ix}, c_{iy}\}_{i=1 \dots N}}_{X \in \mathbb{R}^2} \mapsto \underbrace{\{d_i, n_{ix}, n_{iy}\}_{i=1 \dots N}}_{Y \in \mathbb{R}^3} \quad (2)$$

$$\begin{array}{l} \text{where } d \text{ represents} \\ \text{signed-distance from} \\ \text{object surface} \end{array} \quad \begin{cases} d = 0, & \text{on surface} \\ d < 0, & \text{inside object} \\ d > 0, & \text{outside object} \end{cases} \quad (3)$$

The posterior distribution at a test point  $\mathbf{c}_*$  is shown to be  $F_* \sim \mathcal{GP}(\bar{F}_*, \sigma_*^2)$ , with output mean and variance [30]:

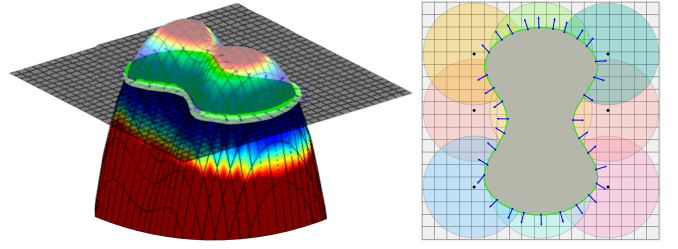
$$\begin{aligned} \bar{F}_* &= k_*^T (K + \sigma_{\text{noise}}^2 I)^{-1} Y \\ \sigma_*^2 &= k_{**} - k_*^T (K + \sigma_{\text{noise}}^2 I)^{-1} k_* \end{aligned} \quad (4)$$

where  $K \in \mathbb{R}^{N \times N}$ ,  $k_* \in \mathbb{R}^{N \times 1}$  and  $k_{**} \in \mathbb{R}$  are the train-train, train-test, and test-test kernels respectively. We use a thin-plate kernel [10], with hyperparameter tuned for scene dimensions. The noise in output space is defined by a zero-mean Gaussian with variance  $\sigma_{\text{noise}}^2$ . While contact points condition the GP on zero SDF observations, contact normals provide function gradient observations [15]. Thus, we can jointly model both SDF and surface direction for objects.

We sample the posterior over an  $M$  element spatial grid of test points  $\mathbf{C}_*$ , to get SDF  $F_{*d}$ . The estimated implicit surface  $\mathcal{S}$  is then the zero-level set contour:

$$\mathcal{S} \triangleq \{\mathbf{s} \in \mathbb{R}^2 \mid F_*(\mathbf{s})_d = 0\} \quad (5)$$

The zero-level set  $\mathcal{S}$  is obtained through a contouring subroutine on  $F_{*d}$ .  $\mathcal{S}$  is initialized with a circular prior and updated with every new measurement  $\{\mathbf{c}_i, \mathbf{n}_i, d = 0\}$ . In Fig. 3 we



**Fig. 3:** [left] Gaussian process potential function and its implicit surface (green) for noisy contact measurements on the butter shape [38]. The colormap shows spatial grid uncertainty. [right] The overlapping local GPs  $F^1 \dots F^L$ . Each GP is responsible for training and test points within its radius, and the overlapping regions ensure continuity in the shape contour.

reconstruct the butter shape [38] with a sequence of noisy contact measurements.

##### B. Efficient online GPIS

In a naïve GP implementation, the computational cost restricts online shape perception. Equation 4 requires an  $N \times N$  matrix inversion that is  $O(N^3)$ , and spatial grid testing that is  $O(MN^2)$ . We use local GPs and a subset of training data approximation for efficient online regression:

**Local GP regression:** We adopt a spatial partitioning approach similar to [34, 35, 36]. The scene is divided into  $L$  independent local GPs  $F^1 \dots F^L$ , each with a radius  $r$  and origin (Fig. 3). Each  $F^i$  claims the training and test points that fall within  $r$  of its origin. The GPs effectively govern smaller domains ( $N_{F^i} \ll N$  and  $M_{F^i} \ll M$ ), and not the entirety of the spatial grid. At every timestep: (i) a subset of local GPs are updated, (ii) only the relevant test points are resampled. Kim et al. [34] demonstrate the need for overlapping domains to avoid contour discontinuity at the boundaries. Thus, we increase  $r$ , and in overlapping regions, the predicted estimates are averaged among GPs.

**Subset of data:** Before adding a training point  $\mathbf{c}_i$  to the active set, we ensure that the output variance  $\sigma_i^2$  is greater than a pre-defined threshold  $\sigma_{\text{thresh}}^2$ . This promotes sparsity in our model by excluding potentially redundant information.

**Implementation:** Rather than direct matrix inversions (Equation 4), it is more efficient to use the Cholesky factor of the kernel matrix [30]. In our online setting, we directly update the Cholesky factor  $\mathcal{LL}^T = (K + \sigma_{\text{noise}}^2 I)$  with new information. We multi-thread the update/test steps, and perform these at a lower frequency than the graph optimization. We set  $L = 25$ , grid sampling resolution 5 mm, and circular prior of radius 40 mm for our experiments.

#### V. POSE ESTIMATION WITH FACTOR GRAPHS

##### A. Factor graph formulation

The *maximum a posteriori* (MAP) estimation problem gives variables that maximally agree with the sensor measurements. This is commonly depicted as a factor graph: a bipartite graph with variables to be optimized for and factors that act as constraints (Fig. 2). The augmented state comprises object poses and a motion model scalar  $\mathcal{C}$ :

$$\mathcal{X}_t = \{\mathbf{x}_1, \dots, \mathbf{x}_t; \mathcal{C}\} \quad (6)$$



Prior work in pushing empirically validates measurement noise to be well-approximated by a Gaussian distribution [27]. With Gaussian noise models, MAP estimation reduces to a nonlinear least-squares problem [39]. Our MAP solution (given best-estimate shape  $\mathcal{S}$ ) is:

$$\mathcal{X}_t^* = \underset{\mathcal{X}_t}{\operatorname{argmin}} \sum_{i=1}^t \left( \underbrace{\|Q(\mathbf{x}_{i-1}, \mathbf{x}_i, \mathbf{z}_{i-1}, \mathcal{C})\|_{\Sigma_Q}^2}_{\text{QS pushing factor}} + \underbrace{\|I(\mathbf{x}_i, \mathbf{z}_i, \mathcal{S})\|_{\Sigma_I}^2}_{\text{IS contact factor}} \right) + \underbrace{\|P(\mathbf{x}_i, \mathbf{z}_i, \mathcal{S})\|_{\Sigma_P}^2}_{\text{Non-penetration factor}} + \underbrace{\|F(\mathbf{x}_{i-1}, \mathbf{x}_i)\|_{\Sigma_F}^2}_{\text{Finite motion factor}} + \underbrace{\|\mathbf{p}_0\|_{\Sigma_0}^2 + \|\mathbf{c}_0\|_{\Sigma_c}^2}_{\text{Priors}} \quad (7)$$

This is graphically represented in Fig. 2, and the cost functions are described in Section V-B. Given covariance matrix  $\Sigma$ ,  $\|v\|_{\Sigma}^2 = v^T \Sigma^{-1} v$  is the Mahalanobis distance of  $v$ . The noise terms for covariances  $\{\Sigma_Q, \dots, \Sigma_c\}$  are empirically selected. The online estimation is performed using incremental smoothing and mapping (iSAM2) [26]. Rather than re-calculating the entire system every timestep, iSAM2 updates the previous matrix factorization with new measurements. In addition, we use a fixed-lag smoother to bound optimization time over the exploration [39]. Fixed-lag smoothing maintains a fixed temporal window of states  $\mathcal{X}_w$ , while efficiently marginalizing out preceding states (100 timesteps in our experiments). Note that this is different from simply culling old states and discarding information.

### B. Cost functions

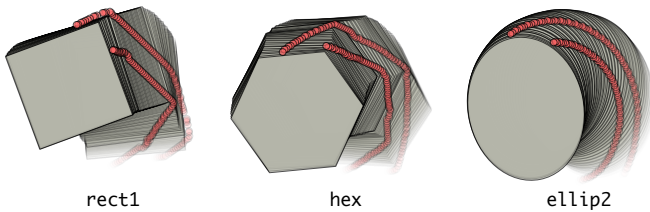
• **QS pushing factor:** The quasi-static model uses a limit surface (LS) model to map between pusher force  $\mathbf{f}_t$  and object motion [12]. Specifically, Lynch et al. [11] develop an analytical model using an ellipsoid LS approximation [37]. The factor ensures object pose transitions obey the quasi-static motion model, with an error term:

$$Q(\mathbf{x}_{t-1}, \mathbf{x}_t, \mathbf{z}_{t-1}, \mathcal{C}) = \left[ \frac{v_x}{\omega} - \mathcal{C}^2 \frac{f_{t-1,x}}{\tau}, \frac{v_y}{\omega} - \mathcal{C}^2 \frac{f_{t-1,y}}{\tau} \right] \quad (8)$$

- $(v_x, v_y, \omega)$  is the object's velocity between  $\mathbf{x}_{t-1}$  and  $\mathbf{x}_t$ ,
- $\tau$  is the applied moment w.r.t. pose center of  $\mathbf{x}_{t-1}$ ,
- $\mathcal{C} = \tau_{\max}/f_{\max}$  is an object-specific scalar ratio dependent on pressure distribution.

For a more rigorous treatment, we refer the reader to [5]. We weakly initialize  $\mathcal{C}$  with our known circular shape prior, and incorporate it into the optimization.

• **Implicit surface contact factor:** Given  $\Theta_t = 1$  (contact), we encourage the measured contact point  $\mathbf{c}_t$  to lie on



**Fig. 4:** Snippets of the tactile exploration data collected in the PyBullet simulator. We use a two-finger pusher to perform contour following, and collect the tactile measurements and ground-truth poses.

the manifold of the object. We define  $\Phi = \Phi(\mathbf{x}_t, \mathbf{z}_t, \mathcal{S})$  that computes the closest point on  $\mathcal{S}$  w.r.t. the pusher via projection. The error term is defined as:

$$I(\mathbf{x}_t, \mathbf{z}_t, \mathcal{S}) = [\Phi - \mathbf{c}_t, \|\Phi - \mathbf{p}_t\| - r_{\text{probe}}] \quad (9)$$

This ensures that in the presence of noise, the contact points lie on the surface and the normals are physically valid [5].

• **Non-penetration factor:** While we assume persistent contact, this cannot be assured when there is more than one pusher. When  $\Theta_t = 0$  (no contact) we enforce an intersection penalty (as used in [28]) on the pusher-slider system. We define  $\Psi = \Psi(\mathbf{x}_t, \mathbf{z}_t, \mathcal{S})$  to estimate the pusher point furthest inside the implicit surface, if intersecting. The error is:

$$P(\mathbf{x}_t, \mathbf{z}_t, \mathcal{S}) = \begin{cases} \|\Psi - \Phi\|, & \text{when intersecting} \\ 0, & \text{when not intersecting} \end{cases} \quad (10)$$

• **Finite motion factor:** Given persistent contact, we weakly bias the object towards constant motion in  $SE(2)$ . The magnitude is empirically chosen from the planar pushing experiments. We observe this both smooths the trajectories, and prevents an indeterminate optimization.

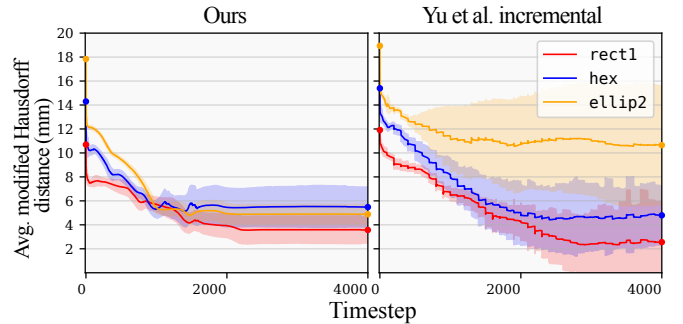
• **Priors:** The prior  $\mathbf{p}_0$  anchors the optimization to the initial pose.  $\mathcal{C}$  is initialized with  $\mathbf{c}_0$  using the circular shape prior.

## VI. EXPERIMENTAL EVALUATION

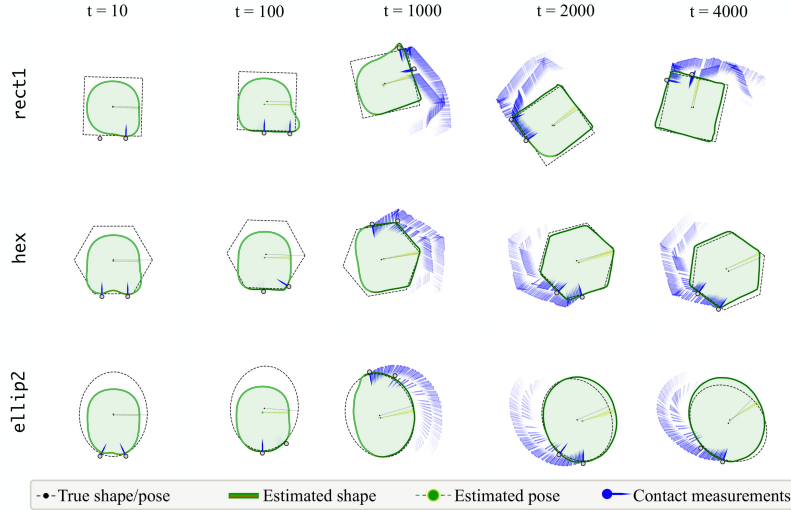
We demonstrate the framework in both simulated (Section VI-A) and real-world planar pushing tasks (Section VI-B).

**Evaluation metrics:** For pose error, we evaluate the root mean squared error (RMSE) in translation and rotation w.r.t. the true object pose. For shape, we use the modified Hausdorff distance (MHD) [40] w.r.t. the true object model. The Hausdorff distance is a measure of similarity between arbitrary shapes that has been used to benchmark GPIS mapping [35], and the MHD is an improved metric more robust to outliers.

**Baseline:** We compare against the work of Yu et al. [5], which recovers shape and pose as a batch optimization



**Fig. 5:** Average MHD w.r.t. the ground-truth model for the 50 logs and for each of the 3 objects. Comparing ours [left] to Yu et al. incremental [right], we see less variance across all shapes, and much lower error in ellip2. This shows that while the piecewise-linear representation is suited to polygonal objects, it understandably fails to generalize to more arbitrary shapes. The GPIS faithfully approximates both classes. Moreover, the errors in data association lead to large variance among trials in rect1 and hex.



**Fig. 6:** Estimated shape and pose of representative simulation trials, with timesteps  $t$ . We compare these against the ground-truth, and overlay the stream of tactile measurements.

(Section II-A). For fairness, we implement an online version as our baseline, that we refer to as *Yu et al. incremental*. We use  $N_s = 25$  shape nodes in the optimization to better represent non-polygonal objects.

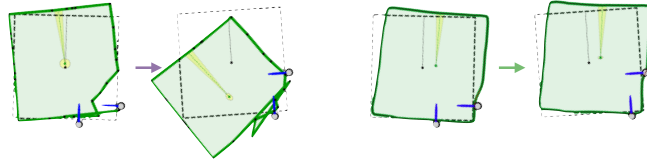
**Compute:** We use the GTSAM library with iSAM2 [26] for incremental factor graph optimization. The experiments were carried out on an Intel Core i7-7820HQ CPU, 32GB RAM without GPU parallelization.

#### A. Simulation experiments

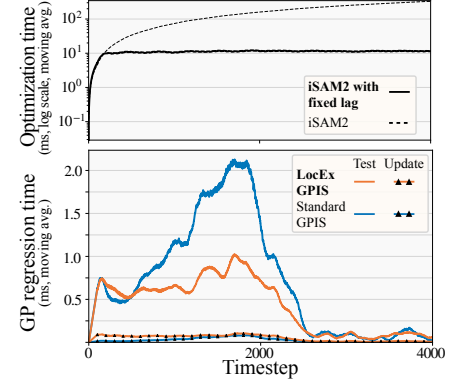
**Setup:** The simulation experiments are conducted in PyBullet [41] (Fig. 4). We use a two-finger pusher ( $r_{\text{probe}} = 6.25$  mm) to perform tactile exploration at 60 mm/s. Contour following is based on previous position and sensed normal reaction. The coefficients of friction of the object-pusher and object-surface are both 0.25. Zero-mean Gaussian noise with std. dev. [0.1 mm, 0.01 N] are added to  $[\mathbf{c}_t, \mathbf{f}_t]$ .

We run 50 trials of 4000 timesteps each, on three shape models [38]: (i) *rect1* (90 mm side), (ii) *hex* (60.5 mm circumradius), and (iii) *ellip2* (130.9 mm maj. axis). While our framework can infer arbitrary geometries, the contour following schema is best suited for simpler planar objects. The object’s initial pose  $\mathbf{x}_0$  is randomly perturbed over the range of  $\pm(2$  mm, 2 mm,  $5^\circ)$ .

**Results:** We highlight the qualitative results of a few representative trials in Fig. 6. We observe the evolution of object shape from the initial circular prior to the final shape, with pose estimates that match well with ground-truth.



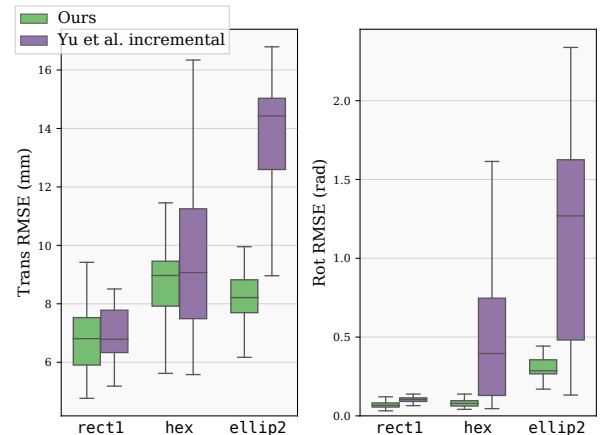
**Fig. 8:** [left] An example of data association failure in the baseline parametric representation [5]. Without discriminating features, committing to vertex associations affects the entire optimization. [right] The GP does not require associations, and the kernel smooths out outlier measurements.



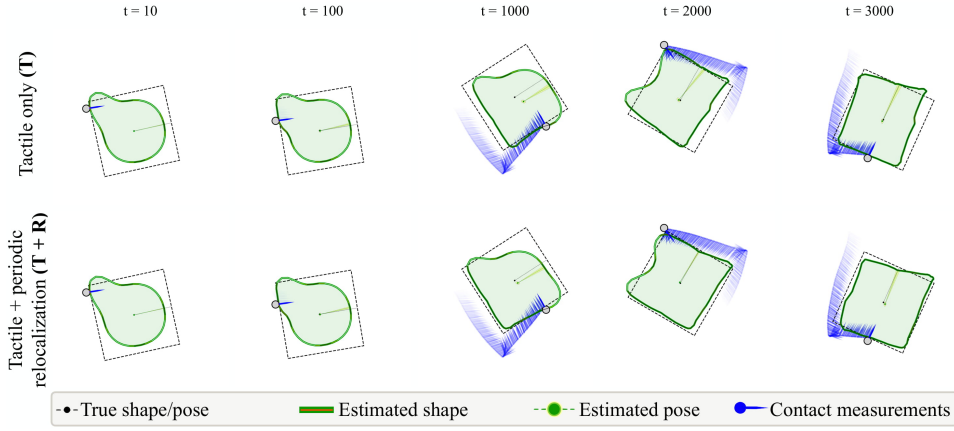
**Fig. 7:** Moving average of execution time of the processes, over all  $3 \times 50$  logs. [top] While the complexity of full iSAM2 grows linearly with time, fixed-lag smoothing maintains a bounded optimization window. [bottom] As a result of spatial partitioning, local GPIS regression has a lower query time compared to the standard implementation.

Fig. 5 shows the decreasing MHD shape error over the 50 trials. The uncertainty threshold of the GPIS  $\sigma_{\text{thresh}}^2$  (Section IV-B) prevents shape updates over repeated exploration, and hence the curve flattens out. This trades-off accuracy for speed, but we find little perceivable difference in the final models. The baseline has larger error for *ellip2*, a shape which their formulation cannot easily represent. Moreover, the uncertainty of the shape estimates are high, due to data association failures. Our representation has no explicit correspondences, and the kernel smooths out outlier measurements. An example of these effects are seen in Fig. 8. A similar trend is seen in pose RMSE across all trials (Fig. 9). The baseline shows comparable performance only with *rect1*, as the shape control points can easily represent it.

Finally, we quantify the computational impact of both the local GPIS regression and incremental fixed-lag optimizer (Fig. 7). Fixed-lag smoothing keeps the optimization time bounded, and prevents linear complexity rise. Spatial partitioning keeps online query time low, and  $\sigma_{\text{thresh}}^2$  results in less frequent updates over time for both methods. The combination of these two give us an average compute time of about 10 ms or 100 Hz. For reference, at 60 mm/s, that



**Fig. 9:** Translation and rotation RMSE box-plots across the 50 simulation trials, for each of the objects. *Yu et al. incremental* performs comparably for *rect1*, but has higher variance and error for the more complex shapes.



**Fig. 10:** Representative results from the real-world estimation task. We compare our tactile only (**T**) result to one aided by periodic relocalization (**T + R**). We add 10 such events in the trial, and the reduced pose drift improves shape prediction.

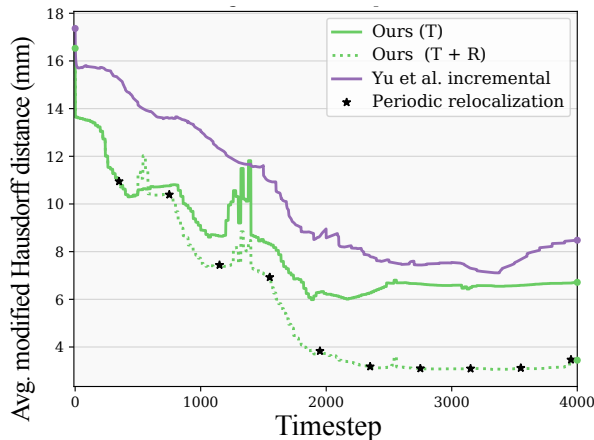
equates to an end-effector travel distance of 0.6 mm per computation. The maximum time taken by a re-linearization step is 55 ms (3.3 mm travel).

### B. Real-world tactile exploration

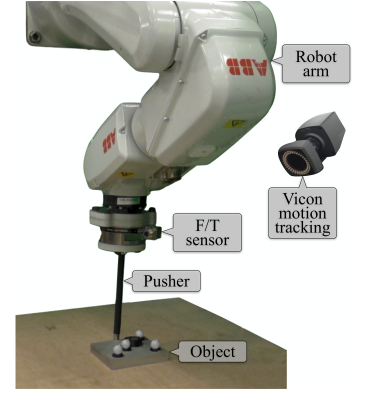
**Setup:** We carry out an identical tactile exploration task with the pusher-slider setup in Fig. 11. An ABB IRB 120 industrial robotic arm circumnavigates a square object (98 mm side) at the rate of 20 mm/s. We perform the experiments on a plywood surface, with object-pusher and object-surface coefficient of friction both  $\approx 0.25$ . We use a single cylindrical rod with a rigidly attached ATI F/T sensor that measures reaction force. Contact is detected with a force threshold, set conservatively to reduce the effect of measurement noise. Ground-truth is collected with Vicon, tracking reflective markers on the object.

We collect three trials of 4000 timesteps each, with tactile measurements and ground-truth. In this case, we do not record force magnitude, but only contact normals. We instead map the pusher velocity to forces via the motion cone, and reasoning about sticking and slipping [42].

**Results:** The top row (**T**) of Fig. 10 shows the evolution



**Fig. 12:** Average MHD w.r.t. the ground-truth model for the real-world experiments, compared against the baseline. With just a few relocalization events, we can achieve far lower shape error.



**Fig. 11:** Our data collection setup: The ABB IRB 120, with F/T sensing, pushes the block on a plywood surface. The Vicon system tracks the object as ground-truth in our evaluation.

**TABLE I:** RMSE for real-world tactile exploration. Apart from *Yu et al. incremental*, we also compare to a method aided by periodic relocalization.

Method	Trans. RMSE (mm)	Rot. RMSE (rad)
<b>Ours (T)</b>	$10.60 \pm 2.74$	$0.09 \pm 0.02$
Ours (T + R)	$4.60 \pm 1.00$	$0.09 \pm 0.01$
Yu et al. incremental	$12.75 \pm 4.01$	$0.17 \pm 0.03$

of shape and pose over the runtime. When compared to the simulation results (Fig. 6), we notice aberrations in the global shape. We can attribute these to: (i) lack of a second pusher, which enables better localization and stable pushes, and (ii) motion model uncertainty in real-world pushing.

The bottom row (**T + R**) of Fig. 10 describes an additional scenario, where we demonstrate better results when we can periodically relocalize the object. This is a proxy for combining noisy global estimates from vision in a difficult perception task with large occlusions. To illustrate this, we crudely simulate 10 such events over each trial using global Vicon measurements with Gaussian noise.

Fig. 12 plots the evolution of shape error over the three trials. Decrease in shape error is correlated to relocalization events, highlighting the importance of reducing pose drift. Finally, Table I shows the RMSE of our experiments.

## VII. CONCLUSION

We formulate a method for estimating shape and pose of a planar object from a stream of tactile measurements. The GPIS reconstructs the object shape, while geometry and physics-based constraints optimize for pose. By alternating between these steps, we show real-time tactile SLAM in both simulated and real-world settings. This method can potentially accommodate tactile arrays and vision, and be extended beyond planar pushing.

In the future, we wish to build on this framework for online SLAM with dense sensors, like the GelSight [43] or GelSlim [44], to reconstruct complex 3-D objects. Multi-hypothesis inference methods [45] can relax our assumption of known initial pose, and learned shape priors [46] can better initialize unknown objects. Knowledge about posterior uncertainty can guide active exploration [31], and perform uncertainty-reducing actions [47].

## REFERENCES

- [1] R. L. Klatzky, S. J. Lederman, and V. A. Metzger, "Identifying objects by touch: An expert system," *Perception & Psychophysics*, vol. 37, no. 4, pp. 299–302, 1985.
- [2] S. Luo, J. Bimbo, R. Dahiya, and H. Liu, "Robotic tactile perception of object properties: A review," *Mechatronics*, vol. 48, pp. 54–67, 2017.
- [3] T. Schmidt, R. Newcombe, and D. Fox, "DART: dense articulated real-time tracking with consumer depth cameras," *Autonomous Robots (AURO)*, vol. 39, no. 3, pp. 239–258, 2015.
- [4] H. Durrant-Whyte and T. Bailey, "Simultaneous localization and mapping: part I," *IEEE robotics & automation magazine*, vol. 13, no. 2, pp. 99–110, 2006.
- [5] K.-T. Yu, J. Leonard, and A. Rodriguez, "Shape and pose recovery from planar pushing," in *Proc. IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, pp. 1208–1215, IEEE, 2015.
- [6] F. Solina and R. Bajcsy, "Recovery of parametric models from range images: The case for superquadrics with global deformations," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 12, no. 2, pp. 131–147, 1990.
- [7] K. Duncan, S. Sarkar, R. Alqasemi, and R. Dubey, "Multi-scale superquadric fitting for efficient shape and pose recovery of unknown objects," in *Proc. IEEE Intl. Conf. on Robotics and Automation (ICRA)*, pp. 4238–4243, IEEE, 2013.
- [8] M. Meier, M. Schopfer, R. Haschke, and H. Ritter, "A probabilistic approach to tactile shape reconstruction," *IEEE Trans. on Robotics (TRO)*, vol. 27, no. 3, pp. 630–635, 2011.
- [9] J. Varley, C. DeChant, A. Richardson, J. Ruales, and P. Allen, "Shape completion enabled robotic grasping," in *Proc. IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, pp. 2442–2447, IEEE, 2017.
- [10] O. Williams and A. Fitzgibbon, "Gaussian process implicit surfaces," *Gaussian Proc. in Practice*, pp. 1–4, 2007.
- [11] K. M. Lynch, H. Maekawa, and K. Tanie, "Manipulation and active sensing by pushing using tactile feedback," in *Proc. IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, vol. 1, 1992.
- [12] S. Goyal, A. Ruina, and J. Papadopoulos, "Planar sliding with dry friction part 1. limit surface and moment function," *Wear*, vol. 143, no. 2, pp. 307–330, 1991.
- [13] M. Moll and M. A. Erdmann, "Reconstructing the shape and motion of unknown objects with active tactile sensors," in *Algorithmic Foundations of Robotics V*, pp. 293–309, Springer, 2004.
- [14] C. Strub, F. Wörgötter, H. Ritter, and Y. Sandamirskaya, "Correcting pose estimates during tactile exploration of object shape: a neuro-robotic study," in *4th International Conference on Development and Learning and on Epigenetic Robotics*, pp. 26–33, IEEE, 2014.
- [15] S. Dragiev, M. Toussaint, and M. Gienger, "Gaussian process implicit surfaces for shape estimation and grasping," in *Proc. IEEE Intl. Conf. on Robotics and Automation (ICRA)*, pp. 2845–2850, IEEE, 2011.
- [16] U. Martinez-Hernandez, G. Metta, T. J. Dodd, T. J. Prescott, L. Natale, and N. F. Lepora, "Active contour following to explore object shape with robot touch," in *2013 World Haptics Conference (WHC)*, pp. 341–346, IEEE, 2013.
- [17] Z. Yi, R. Calandra, F. Veiga, H. van Hoof, T. Hermans, Y. Zhang, and J. Peters, "Active tactile object exploration with Gaussian processes," in *Proc. IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, pp. 4925–4930, IEEE, 2016.
- [18] D. Driess, D. Hennes, and M. Toussaint, "Active multi-contact continuous tactile exploration with Gaussian process differential entropy," in *Proc. IEEE Intl. Conf. on Robotics and Automation (ICRA)*, pp. 7844–7850, IEEE, 2019.
- [19] A. Petrovskaya and O. Khatib, "Global localization of objects via touch," *IEEE Trans. on Robotics (TRO)*, vol. 27, no. 3, pp. 569–585, 2011.
- [20] L. Zhang, S. Lyu, and J. Trinkle, "A dynamic Bayesian approach to real-time estimation and filtering in grasp acquisition," in *Proc. IEEE Intl. Conf. on Robotics and Automation (ICRA)*, pp. 85–92, IEEE, 2013.
- [21] M. C. Koval, N. S. Pollard, and S. S. Srinivasa, "Pose estimation for planar contact manipulation with manifold particle filters," *Intl. J. of Robotics Research (IJRR)*, vol. 34, no. 7, pp. 922–945, 2015.
- [22] M. Bauza, O. Canal, and A. Rodriguez, "Tactile mapping and localization from high-resolution tactile imprints," in *Proc. IEEE Intl. Conf. on Robotics and Automation (ICRA)*, pp. 3811–3817, IEEE, 2019.
- [23] P. Sodhi, M. Kaess, M. Mukadam, and S. Anderson, "Learning tactile models for factor graph-based state estimation," in *Proc. IEEE Intl. Conf. on Robotics and Automation (ICRA)*, May 2021.
- [24] P. Hebert, N. Hudson, J. Ma, and J. Burdick, "Fusion of stereo vision, force-torque, and joint sensors for estimation of in-hand object location," in *Proc. IEEE Intl. Conf. on Robotics and Automation (ICRA)*, pp. 5935–5941, IEEE, 2011.
- [25] G. Izatt, G. Mirano, E. Adelson, and R. Tedrake, "Tracking objects with point clouds from vision and touch," in *Proc. IEEE Intl. Conf. on Robotics and Automation (ICRA)*, pp. 4000–4007, IEEE, 2017.
- [26] M. Kaess, H. Johannsson, R. Roberts, V. Ila, J. Leonard, and F. Dellaert, "iSAM2: Incremental smoothing and mapping using the Bayes tree," *Intl. J. of Robotics Research (IJRR)*, vol. 31, no. 2, pp. 216–235, 2012.
- [27] K.-T. Yu and A. Rodriguez, "Realtime state estimation with tactile and visual sensing - application to planar manipulation," in *Proc. IEEE Intl. Conf. on Robotics and Automation (ICRA)*, pp. 7778–7785, IEEE, 2018.
- [28] A. S. Lambert, M. Mukadam, B. Sundaralingam, N. Ratliff, B. Boots, and D. Fox, "Joint inference of kinematic and force trajectories with visuo-tactile sensing," in *Proc. IEEE Intl. Conf. on Robotics and Automation (ICRA)*, pp. 3165–3171, IEEE, 2019.
- [29] J. F. Blinn, "A generalization of algebraic surface drawing," *ACM Transactions on Graphics (TOG)*, vol. 1, no. 3, pp. 235–256, 1982.
- [30] C. E. Rasmussen, "Gaussian processes in machine learning," in *Summer School on Machine Learning*, pp. 63–71, Springer, 2003.
- [31] S. Dragiev, M. Toussaint, and M. Gienger, "Uncertainty aware grasping and tactile exploration," in *Proc. IEEE Intl. Conf. on Robotics and Automation (ICRA)*, pp. 113–119, IEEE, 2013.
- [32] M. Li, K. Hang, D. Kragic, and A. Billard, "Dexterous grasping under shape uncertainty," *J. of Robotics and Autonomous Systems (RAS)*, vol. 75, pp. 352–364, 2016.
- [33] E. Snelson and Z. Ghahramani, "Sparse Gaussian processes using pseudo-inputs," *Advances in Neural Information Processing Systems*, vol. 18, pp. 1259–1266, 2006.
- [34] S. Kim and J. Kim, "Continuous occupancy maps using overlapping local Gaussian processes," in *Proc. IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, pp. 4709–4714, IEEE, 2013.
- [35] J. A. Stork and T. Stoyanov, "Ensemble of sparse Gaussian process experts for implicit surface mapping with streaming data," in *Proc. IEEE Intl. Conf. on Robotics and Automation (ICRA)*, pp. 10758–10764, IEEE, 2020.
- [36] B. Lee, C. Zhang, Z. Huang, and D. D. Lee, "Online continuous mapping using Gaussian process implicit surfaces," in *Proc. IEEE Intl. Conf. on Robotics and Automation (ICRA)*, pp. 6884–6890, IEEE, 2019.
- [37] S. H. Lee and M. Cutkosky, "Fixture planning with friction," *J. of Manufacturing Science and Engineering*, vol. 113, no. 3, 1991.
- [38] K.-T. Yu, M. Bauza, N. Fazeli, and A. Rodriguez, "More than a million ways to be pushed: a high-fidelity experimental dataset of planar pushing," in *Proc. IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, pp. 30–37, IEEE, 2016.
- [39] F. Dellaert and M. Kaess, "Factor graphs for robot perception," *Foundations and Trends in Robotics*, vol. 6, no. 1-2, pp. 1–139, 2017.
- [40] M.-P. Dubuisson and A. K. Jain, "A modified Hausdorff distance for object matching," in *Proc. Intl. Conf. on Pattern Recognition (ICPR)*, vol. 1, pp. 566–568, IEEE, 1994.
- [41] E. Coumans, "Bullet physics engine," *Open Source Software: http://bulletphysics.org*, vol. 1, no. 3, p. 84, 2010.
- [42] M. T. Mason, "Mechanics and planning of manipulator pushing operations," *Intl. J. of Robotics Research (IJRR)*, vol. 5, no. 3, pp. 53–71, 1986.
- [43] 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.
- [44] E. Donlon, S. Dong, M. Liu, J. Li, E. Adelson, and A. Rodriguez, "GelSlim: A high-resolution, compact, robust, and calibrated tactile-sensing finger," in *Proc. IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, pp. 1927–1934, IEEE, 2018.
- [45] M. Hsiao and M. Kaess, "MH-iSAM2: Multi-hypothesis iSAM using Bayes tree and hypo-tree," in *Proc. IEEE Intl. Conf. on Robotics and Automation (ICRA)*, pp. 1274–1280, IEEE, 2019.
- [46] S. Wang, J. Wu, X. Sun, W. Yuan, W. T. Freeman, J. B. Tenenbaum, and E. H. Adelson, "3D shape perception from monocular vision, touch, and shape priors," in *Proc. IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, pp. 1606–1613, IEEE, 2018.
- [47] M. R. Dogar and S. S. Srinivasa, "A planning framework for non-prehensile manipulation under clutter and uncertainty," *Autonomous Robots (AURO)*, vol. 33, no. 3, pp. 217–236, 2012.