Incremental shape and pose estimation from planar pushing using contact implicit surfaces

Sudharshan Suresh, Joshua G. Mangelson, and Michael Kaess

I. Introduction

Robots need accurate, online estimates of shape and pose while manipulating unknown objects. While vision and depth-based tracking has been well studied [3], they are affected by self-occlusion, cluttered workspaces, and poor visibility. Interestingly—even when blindfolded—humans can infer object properties from local tactile information.

Online tactile inference is hard due to the intrusive nature of touch sensing, and initially unknown object models. Recently, Yu et al. [1] formulated this as a batch-SLAM problem, relying on planar frictional pushing mechanics. However, the method is not built for online tracking, and uses a piecewise-linear discrete shape representation. Incremental, graph-based approaches were later considered, but assume known object model and incorporate vision [2, 4].

A Gaussian process implicit surface (GPIS) shape representation can fuse uncertain measurements in a probabilistic fashion, and is non-parametric—unlike [1]. Dragiev et al. [5] use a GPIS for tactile exploration of fixed pose 3-D objects. To our knowledge, no methods use this representation with online pose estimation for manipulation tasks.

In this work, we combine a GPIS shape representation with sparse nonlinear incremental optimization to localize and infer the shape of planar objects. We demonstrate results with simulated tactile exploration of different objects.

II. Combining Implicit Surface with Factor Graph

A. Gaussian process implicit surface for contacts

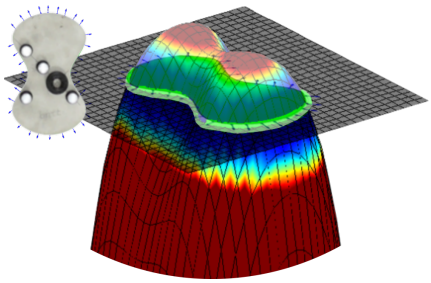

Fig. 1: Gaussian process and its implicit surface (green) for noisy contact measurements on the `butter` shape (taken from [2]). The measurements are added sequentially; the colormap of the GP indicates uncertainty (video).

A **Gaussian process regressor** learns a continuous, non-linear function from sparse, noisy datapoints [6]. In planar pushing, we get surface measurements as global contact locations and normals (x, y, n_x, n_y). The GP learns a mapping:

$$\mathcal{G} : \underbrace{\{x_i, y_i\}^{i=1\cdots t}}_{X} \mapsto \underbrace{\{d_i, n_{xi}, n_{y_i}\}^{i=1\cdots t}}_{Y} \tag{1}$$

where d represents signed-distance from object surface

$$\begin{cases} d = 0, & \text{on surface} \\ d < 0, & \text{inside object} \\ d > 0, & \text{outside object} \end{cases} \tag{2}$$

We use a thin-plate kernel function—as in [7]—which is isotropic and simple to tune. The GP gives an output mean function as its MAP estimate (Eq. 3), and a variance.

$$\bar{\mathcal{G}}_* = k_* \left(K + \sigma^2 I\right)^{-1} Y \tag{3}$$

where k_*, K, and σ represent our test-train kernel, train-train kernel, and noise on output respectively (refer [6]). While contact points provide the GP zero-value observations, contact normals provide gradient observations. The **implicit surface** $\mathcal{S}$ is the zero-level set contour of $\mathcal{G}$, such that:

$$\mathcal{S} \triangleq \{(x, y) \mid \mathcal{G}_d(x,\ y) = 0\} \tag{4}$$

With every timestep, more observations are added, and $\mathcal{S}$ further delineates the true object shape (Fig. 1). In our experiments, all shapes are initialized with a circular prior, an underestimate of the true shape.

B. Factor graph formulation

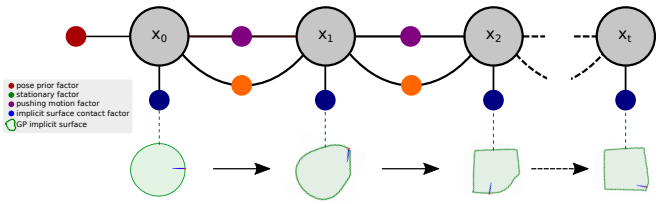

Fig. 2: Combined formulation of the factor graph with GPIS contour updates, optimizing for poses $x_{i\cdots t}$. Initialized with a circular prior, the GP learns a better shape representation, which better informs contact factors.

Online nonlinear least-square optimization is more accurate than filtering methods as it preserves the history of cost functions over timesteps, to compute the optimal solution. Incremental smoothing and mapping (iSAM) [8] performs efficient updates by taking advantage of sparsity.

Given noisy time-series measurements, we formulate our problem as a factor graph (Fig. 2). This is a bipartite graph with variable nodes that we optimize for, and measurement factor that constrain them. In Fig. 2, the nodes are object poses, while the factors depend on frictional pushing mechanics and implicit-surface contact constraints:

The authors are with the Robotics Institute, Carnegie Mellon University, Pittsburgh, PA 15213, USA {suddhu, jmangels, kaess}@cmu.edu. We thank Kuan-Ting Yu and Alberto Rodriguez for access to their codebase [1, 2]

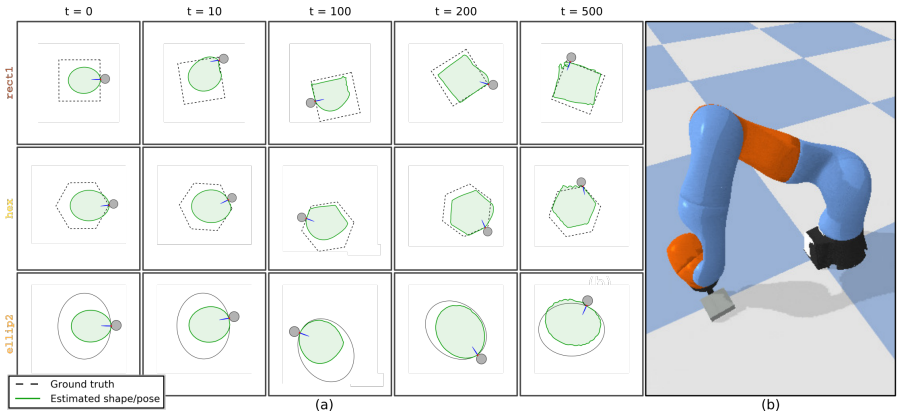

TABLE I: pose RMSE for shapes

Shape	Trans. (mm)	Rot. (rad)
rect1	12.3	0.11
hex	12.2	0.07
ellip2	11.6	0.32

Fig. 3: **(a)** Results from incremental pose estimation, along with the GPIS contours at each timestep. The object shape is delineated over time, while being transformed to the global frame by the object's estimated pose. We notice some aberrations, and consistent underestimates of the object shape—which will be tackled in future work. **(b)** The Pybullet simulator featuring a KUKA arm manipulating the `rect1` object model (video).

Pushing motion factor: Constrains consecutive poses based on the quasi-static frictional pushing. Similar to [1], we use an ellipsoidal limit surface approximation [9], and assume known uniform friction and pressure distribution.

Implicit-surface contact factor: Imposes the contact measurement to lie on *current estimate* of the GPIS contour, and the contact direction to be towards the object.

Stationary factor: Enforces infinitesimal motion between consecutive timesteps.

Pose prior factor: Anchors the optimization to starting pose (assumed known).

At each iteration, the GPIS is estimated using new contact data (Section II-A), and the backend gives us the *maximum a posteriori* (MAP) pose estimate. These two processes are carried out back-and-forth in an online fashion, so as to get the best shape and pose hypothesis at each timestep (Fig. 2).

III. Tactile Exploration Experiments

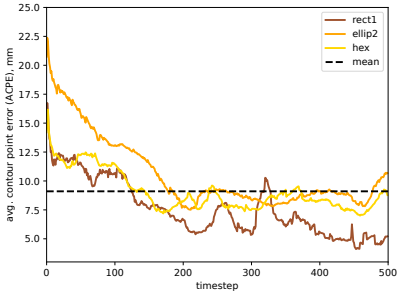

Fig. 4: Decreasing ACPE for shapes due to improving estimates. Dims: (*rect1* 90mm side, *hex* 60.5mm circumradius, *ellip2* 130.9mm maj. axis)

Our method is evaluated on simulation experiments in PyBullet [10]. The incremental graph optimization is performed using GTSAM with iSAM2 [8]. We employ a tactile exploration strategy—a 10mm pusher circumnavigates the objects while pushing them on a frictional plane of $\mu_s = 0.25$ (Fig. 3b). We run the simulations for 500 timesteps, for three shapes models from [2]—`rect1, hex, ellip2`. Noise of $\sigma = 0.1$mm is added to pusher postions and contact points.

Qualitative results for the three shapes (Fig. 3a) show the evolution of the shape contour along with planar pose tracking. The quantitative metrics we compute are root mean squared error (**RMSE**) of pose (Table I) and average contour point error (**ACPE**) (Fig. 4). Our metric, ACPE, intuitively describes both shape and pose error over time:

$$\text{ACPE}(t) = \frac{1}{M_t} \sum_{i=1}^{M_t} \mathcal{D}(\mathcal{S}_{t_i}^{X_t}, \mathcal{O}^{X_t^\dagger}) \tag{5}$$

where, at timestep t, $\mathcal{O}^{X_t^\dagger}$ is the true shape transformed by the true pose, $\mathcal{S}_{t_i}^{X_t}$ is a contour point from the current GPIS transformed by the estimated pose. We average $\mathcal{D}$, the minimum point-to-shape distance, over M_t contour points.

IV. Discussions

This work fuses GPIS with factors graphs for incremental shape and pose estimation from tactile feedback. Future work includes **(i)** modeling contact as lower-dimensional manifold constraints, **(ii)** relaxing the pressure distribution assumption, **(iii)** efficient GPIS computation for real-time inference, and tighter integration into the graph framework, **(iv)** real-world experiments, and dealing with richer contact data.

References

[1] K.-T. Yu, J. Leonard, and A. Rodriguez, "Shape and pose recovery from planar pushing," in *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, pp. 1208–1215, IEEE, 2015.

[2] K.-T. Yu and A. Rodriguez, "Realtime state estimation with tactile and visual sensing - application to planar manipulation," in *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, pp. 7778–7785, IEEE, 2018.

[3] A. Zeng, K.-T. Yu, S. Song, D. Suo, E. Walker, A. Rodriguez, and J. Xiao, "Multi-view self-supervised deep learning for 6D pose estimation in the Amazon picking challenge," in *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, pp. 1386–1383, IEEE, 2017.

[4] 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 *2019 International Conference on Robotics and Automation (ICRA)*, pp. 3165–3171, IEEE, 2019.

[5] S. Dragiev, M. Toussaint, and M. Gienger, "Uncertainty aware grasping and tactile exploration," in *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, pp. 113–119, IEEE, 2013.

[6] C. E. Rasmussen, "Gaussian processes in machine learning," in *Summer School on Machine Learning*, pp. 63–71, Springer, 2003.

[7] M. Li, K. Hang, D. Kragic, and A. Billard, "Dexterous grasping under shape uncertainty," *Robotics and Autonomous Systems*, vol. 75, pp. 352–364, 2016.

[8] 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*, vol. 31, no. 2, pp. 216–235, 2012.

[9] K. M. Lynch, H. Maekawa, and K. Tanie, "Manipulation and active sensing by pushing using tactile feedback.," in *IROS*, vol. 1, 1992.

[10] E. Coumans, "Bullet physics engine," *Open Source Software: http://bulletphysics. org*, vol. 1, no. 3, p. 84, 2010.