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 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, nonlinear 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:

\[
\mathcal{F}: \{x_i, y_i\}_{i=1}^{t} \rightarrow \{d_i, n_{x_i}, n_{y_i}\}_{i=1}^{t}
\]

where \(d\) represents signed-distance from object surface:

\[
\begin{align*}
    d &= 0, \quad \text{on surface} \\
    d &> 0, \quad \text{outside object} \\
    d &< 0, \quad \text{inside object}
\end{align*}
\]

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.

\[
\tilde{F}_* = k_* \left( K + \sigma^2 I \right)^{-1} Y
\]

where \(k_*\), \(K\), and \(\sigma\) 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 \(S\) is the zero-level set contour of \(\mathcal{F}\), such that:

\[
S = \{ (x, y) \mid \mathcal{F}_d(x, y) = 0 \}
\]

While more observations, \(S\) further delineates the true object shape (Fig. 1). In our experiments, 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 \(x_{i..t}\). 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 time-steps, 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.
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 to get the best online shape and pose hypothesis at each timestep (Fig. 2). We note that without global pose constraints (e.g. vision), these estimates will drift over a large number of timesteps.

III. TACTILE EXPLORATION EXPERIMENTS

Qualitative results for the three shapes (Fig. 3a) show the evolution of the shape contour along with planar pose tracking. We also compute root mean squared error (RMSE) of pose (Table I), and modified Hausdorff distance ($d_{MH}$) [11] of the transformed contour with respect to the ground truth (Fig. 4). $d_{MH}$ measures mutual proximity, and neatly captures both shape and pose error over time:

$$d_{MH}(t) = \max \left\{ \frac{1}{|S_t^X|} \sum_{s \in S_t^X} d(s, G^X_t), \frac{1}{|G^X_t|} \sum_{g \in G^X_t} d(g, S_t^X) \right\}$$

where $d(a, B) = \min_{b \in B} \|a - b\|$. At timestep $t$, $G^X_t$ is the ground truth model transformed by the true pose, $S_t^X$ is the online GPIS contour transformed by the estimated pose.

IV. DISCUSSION AND FUTURE WORK

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) employing alternate pusher configurations, (iii) efficient GPIS updates for real-time inference, and its tighter integration into the factor graph, (iv) real-world testing, (v) incorporating vision, and richer tactile data [12].

REFERENCES


