GelSLAM: A Real-time, High-Fidelity, and Robust 3D Tactile SLAM System

Anonymous

Abstract— **Accurate object pose and shape perception are essential for precise manipulation. Compared to vision-based methods, tactile sensing is robust to occlusion, making it well-suited for in-hand tasks. We present GelSLAM, a real-time tactile-only 3D SLAM system that estimates object pose over long horizons and reconstructs shapes with high fidelity. By leveraging surface normals and curvatures instead of point clouds, it achieves low-drift tracking and submillimeter reconstruction accuracy, even for low-texture objects. GelSLAM extends tactile sensing beyond local contact to enable global, long-horizon spatial perception, and we believe it will serve as a foundation for many precise manipulation tasks involving interaction with objects in hand.**

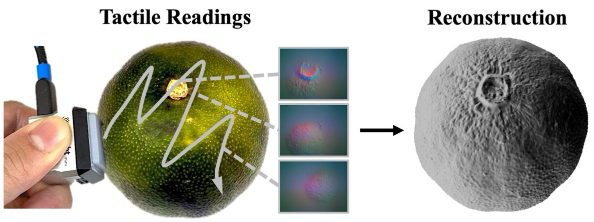

Fig. 1: GelSLAM enables robust, high-fidelity object-level 3D reconstruction and real-time, accurate long-horizon object tracking using only tactile sensing.

I. INTRODUCTION

Tactile sensing plays a critical role in how humans perceive and interact with the world, enabling inference of object properties such as shape and pose [1]. In robotics, tactile sensing is particularly valuable for contact-rich manipulation tasks, where high precision is required, but vision often suffers from occlusion or poor lighting. Vision-based tactile sensors such as GelSight [2] provide high spatial resolution for capturing local surface geometry, making them well-suited for geometric perception. However, enabling robots to achieve real-time, long-horizon object tracking and high-fidelity 3D reconstruction using tactile sensing alone remains a significant challenge.

In this work, we present GelSLAM, a real-time tactile SLAM system that enables long-horizon 6DoF object tracking and high-fidelity 3D reconstruction using only tactile input. The key challenge lies in the inherently local nature of tactile sensing: each contact provides only a small surface patch, making it difficult to establish global spatial relationships. Existing tactile SLAM attempts remain limited in scale, robustness, or geometric fidelity [3], [4], [5]. GelSLAM addresses this by introducing a differential representation of tactile data, from which robust tracking, loop closure, and reconstruction modules are built within a unified system. This allows robust drift correction, relocalization after contact loss, and accurate reconstruction of diverse objects, including low-texture and large objects.

II. METHOD

A. System Overview

GelSLAM takes a stream of tactile images from a GelSight sensor and estimates the object-relative sensor trajectory while reconstructing a global 3D shape. The system consists of three modules running in parallel: (1) a tracking module that estimates relative motion between nearby frames and selects keyframes, (2) a loop closure module that detects revisited contact regions and performs pose graph optimization to correct drift, and (3) a reconstruction module that fuses local surface patches into a global model. Together, these components enable real-time operation with long-horizon consistency using only tactile sensing, without relying on vision, object models, or proprioception.

B. Key Idea: Differential Representation for Tactile SLAM

The key insight of GelSLAM is to represent tactile observations using surface normals and curvature, rather than point clouds or height maps. Due to the small range of surface deformation in tactile contact, point clouds exhibit limited geometric variation and are poorly suited for registration. In contrast, differential representations capture rich local structure: normal maps encode first-order surface geometry, while curvature maps provide rotation-invariant features for robust SIFT [6] matching. By operating directly on these representations, GelSLAM achieves reliable frame-to-frame alignment, robust loop detection, and accurate fusion of local geometry into a globally consistent reconstruction.

III. EXPERIMENTS AND RESULTS

We evaluate GelSLAM on two tasks using only tactile sensing: long-horizon object 6DoF pose tracking and object 3D reconstruction.

A. Long-Horizon Tracking

We evaluate 6DoF tracking accuracy on a dataset of 18 objects with 126 episodes. We compare GelSLAM against ICP [7], the most common point cloud–based tracking method, and NormalFlow [8], the state-of-the-art method for tactile-based object tracking. As shown in Table I, ICP performs poorly due to the limited geometric variation of tactile point clouds, while NormalFlow achieves accurate short-horizon tracking but accumulates drift over time. In

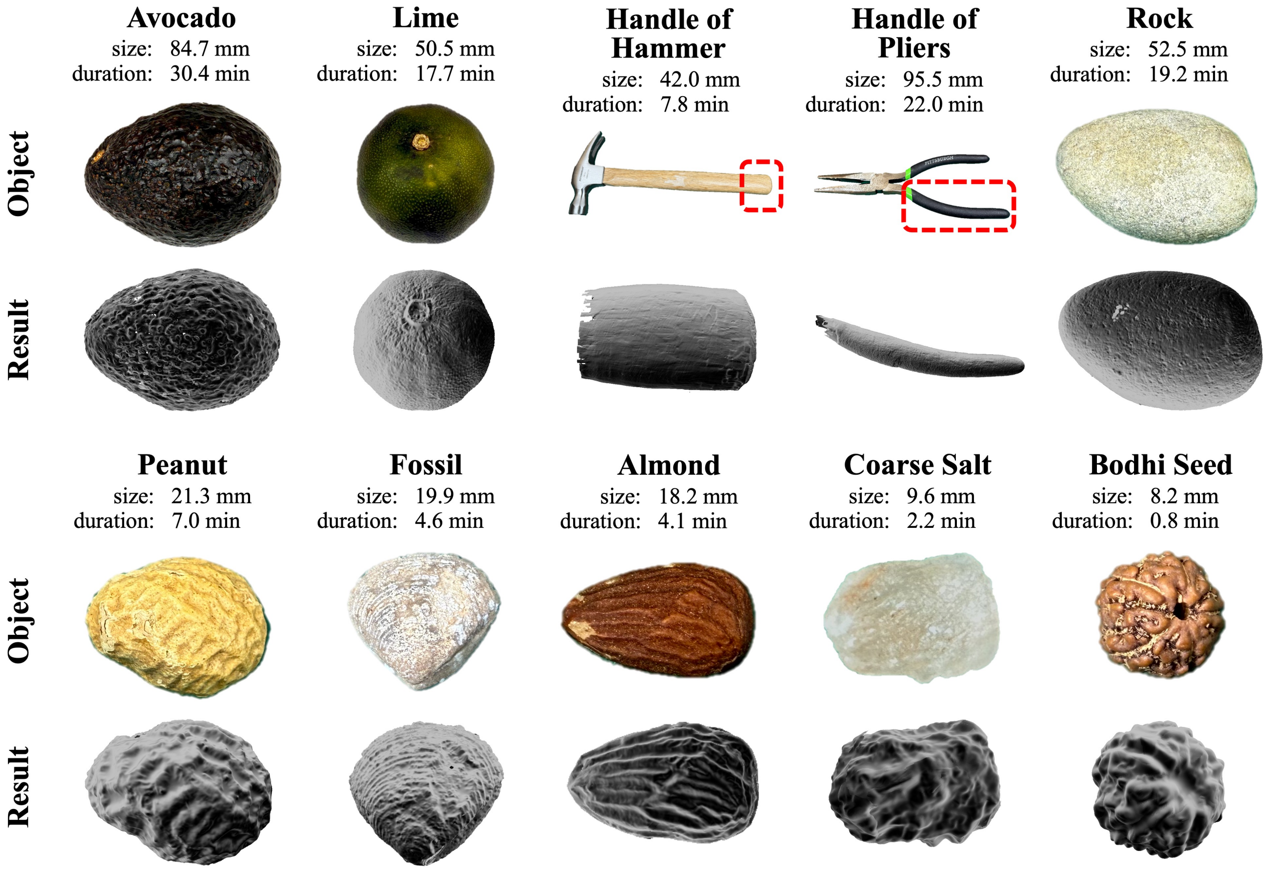

Fig. 2: GelSLAM reconstruction results, objects ordered roughly from largest to smallest. Object size and tactile video duration are shown. GelSLAM successfully reconstructs detailed global 3D models using only local tactile patches.

contrast, GelSLAM significantly reduces both rotation and translation errors by correcting drift through loop closure, achieving consistently low error over long horizons.

Method	$\theta_x(^\circ)$	$\theta_y(^\circ)$	$\theta_z(^\circ)$	x(mm)	y(mm)	z(mm)
ICP	13.0	13.9	23.9	8.14	9.94	6.73
NormalFlow	7.11	7.53	7.63	1.13	1.21	0.92
GelSLAM	**4.06**	**4.38**	**3.57**	**1.00**	**0.96**	**0.72**

TABLE I: Average 6DoF tracking MAE.

B. *Object 3D reconstruction*

The 3D reconstruction results using GelSLAM on 10 objects are shown in Fig. 2. Tactile videos are recorded under unconstrained interaction with frequent contact breaks and re-establishment. Fig. 3 shows detailed pose graphs for four representative objects, demonstrating robust detection of a large number of loop closures across many keyframes. GelSLAM produces high-fidelity reconstructions across all cases, capturing fine geometric details that are difficult to recover with vision.

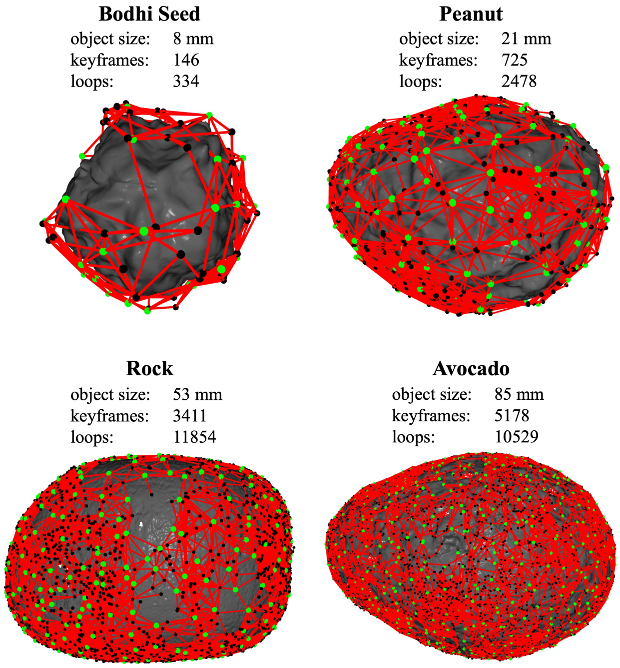

Fig. 3: Pose graphs generated by GelSLAM. Black nodes: keyframes; red edges: pairwise pose constraints.

REFERENCES

[1] S. J. Lederman and R. L. Klatzky, “Extracting object properties through haptic exploration,” *Acta Psychologica*, vol. 84, no. 1, pp. 29–40, 1993, tactile Pattern Recognition.

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

[3] 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)*, 2022, pp. 2164–2170.

[4] S. Suresh, Z. Si, J. G. Mangelson, W. Yuan, and M. Kaess, “Shapemap 3-d: Efficient shape mapping through dense touch and vision,” in *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 2022, pp. 7073–7080.

[5] 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.

[6] D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” *Int. J. Comput. Vision*, vol. 60, no. 2, p. 91–110, Nov. 2004.

[7] Y. Chen and G. Medioni, “Object modeling by registration of multiple range images,” in *Proceedings. 1991 IEEE International Conference on Robotics and Automation*, 1991, pp. 2724–2729 vol.3.

[8] H.-J. Huang, M. Kaess, and W. Yuan, “Normalflow: Fast, robust, and accurate contact-based object 6dof pose tracking with vision-based tactile sensors,” *IEEE Robotics and Automation Letters*, pp. 1–8, 2024.