Belief-Guided Interactive Perception for Manipulation in Clutter under Noisy Proprioception

Author Names Omitted for Anonymous Review.

***Abstract*—Robotic manipulation often operates under perception that is partially observable and noisy. This limitation is especially pronounced for low-cost robot arms, which may lack tactile skins, force–torque sensors, or well-calibrated cameras. In this project, we study whether such arms can still acquire useful information about the environment through interactive contact exploration, using only noisy proprioceptive signals. Through exploratory motions and interaction with contacts, the robot identifies free space, movable objects, and fixed collisions. These interaction outcomes are used to maintain and update a belief over the workspace, which is then exploited for tasks such as goal reaching in cluttered environments without prior knowledge of the clutter arrangement. Our current pipeline uses rule-based contact abstractions and manually designed policies, and is tested on goal-reaching tasks in tabletop clutter and constrained cabinet settings. Preliminary results suggest that even low-cost arms with noisy proprioception can use contact to update a task-relevant map, enabling manipulation in cluttered environments.**

I. Introduction

Robot manipulation is moving beyond structured environments towards homes and cluttered workspaces [2, 1]. Low-cost robotic arms have recently emerged as an attractive path toward manipulation in the wild, reducing hardware barriers and improving reproducibility [14]. However, these platforms remain limited in both sensing and control capabilities. They often exhibit lower precision, noisier actuation, less reliable vision, and limited access to fine-grained tactile or force sensing [14, 10, 5]. These limitations become more pronounced in unstructured scenes, where occlusion, clutter, and unpredictable contact further complicate manipulation.

Humans can complete manipulation tasks even when sensing is incomplete. When reaching for a shutoff valve inside a cluttered cabinet (Fig. 1A), human relies on contact feedback to inspect the region, avoid collisions and reach the targets. Although human arms provide much richer sensing and control, the underlying principle is useful for robotics: uncertain contact can provide critical information for action.

We ask whether a low-cost robot arm (USD 4.4k) can select actions from proprioception alone. Although such signals are insufficient for precise contact localization, object identification, or full scene reconstruction, our preliminary results suggest that they can still support action-relevant inference. We therefore formulate the problem as belief-guided contact exploration. Through active interaction, the robot incrementally builds a belief map of the workspace, enabling goal reaching in cluttered environments under partial observability.

Prior work has shown that contact can serve as an informative signal for manipulation rather than merely a source of disturbance. Whole-arm tactile systems have used distributed contact sensing to navigate cluttered environments and reach occluded targets [7]. Another line of contact-driven methods exploit interaction to infer occupancy, reduce pose uncertainty, or retrieve objects under partial observability [13, 12, 15]. Active perception for grasping handles heavy clutter and difficult optical properties with camera motion [3, 9]. Sensorless and proprioception-only methods recover contact events or memories without tactile skins [10, 6, 11, 5]. Our setting is less observable: sparse, noisy proprioceptive events must produce only the belief variables needed for action selection.

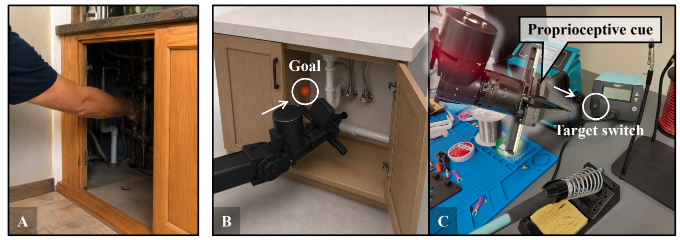

Fig. 1. Motivation and target settings. (A) Human exploratory touch in a dark cabinet. (B) Simulated cabinet reach to a goal. (C) Hardware reach to an occluded switch on a cluttered workbench.

We make three preliminary contributions: (1) we formulate manipulation without exteroceptive sensing on low-cost arms as belief construction from weak proprioceptive interaction signals; (2) we define an actionable contact belief over free, blocked, uncertain, and pushable regions; and (3) we build an end-to-end interaction-perception framework that closes the loop between proprioceptive contact sensing, task-level belief modeling, and contact-aware control, enabling a low-cost arm to manipulate in clutter without vision.

II. From Weak Contact Cues to Actionable Belief

We study manipulation in cluttered workspaces without exteroceptive sensing. Let $\mathcal{W}$ be discretized into local regions $\mathcal{R} = \{r_i\}_{i=1}^{N}$. The robot must move from configuration q_0 to a target region $\mathcal{G} \subset \mathcal{W}$, while each region has an unobserved task-level contact state $z_i \in \mathcal{L} = \{F, B, P\}$ for free space, immovable blockage, or pushable contact. The robot receives only weak proprioceptive observations y_t from tracking error, motor current, stall events, and motion outcomes. These cues are useful but do not identify object shape, contact force, or an exact contact point. Thus z_i is an action-relevant contact class, not a semantic or geometric scene label.

The useful state is therefore smaller than a reconstructed scene. We maintain a contact belief

$$b_t^i(\ell) = \Pr(z_i = \ell \mid h_t), \quad h_t = (q_{0:t}, a_{0:t-1}, y_{1:t}), \tag{1}$$

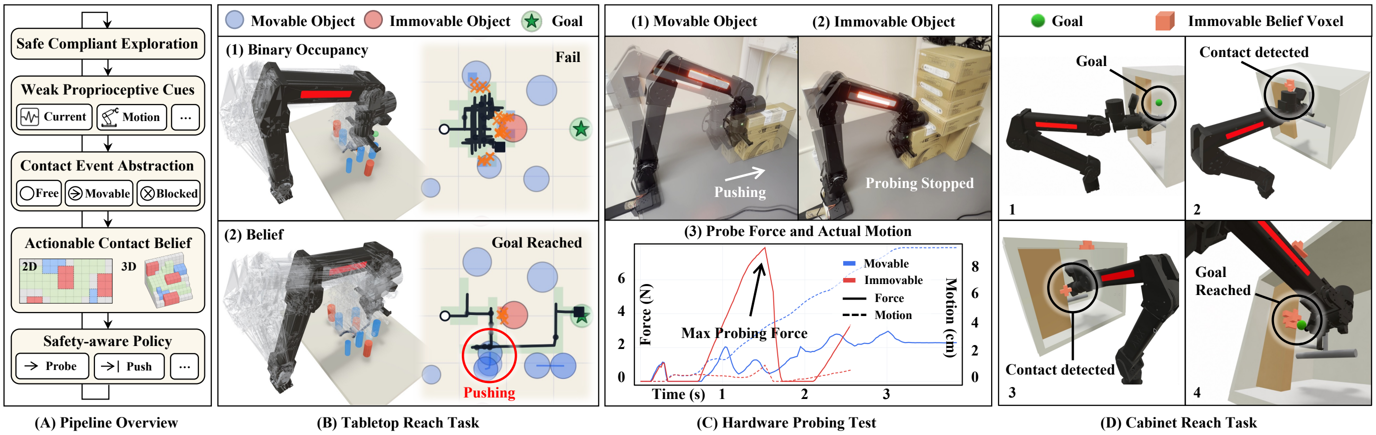

Fig. 2. Pipeline and preliminary experiments. (A) Safe compliant exploration converts weak proprioceptive cues into contact events, updates an actionable contact belief, and passes it to a policy designed for safe contact. (B) In tabletop reaching, binary occupancy treats contact as blockage and fails, while the belief distinguishes pushable from blocked regions and reaches the goal. (C) Hardware force and motion traces separate movable from immovable contacts under a bounded probing force. (D) In cabinet reaching, detected contacts update immovable belief voxels for retreat, replanning, and goal reaching.

where $\ell \in \mathcal{L}$ and $a_t \in \{\text{reach}, \text{probe}, \text{retreat}, \text{push}\}$. For action selection, the belief is converted into an actionable label

$$\hat{\ell}_t^i = \begin{cases} U, & \max_{\ell \in \mathcal{L}} b_t^i(\ell) < \tau, \\ \arg\max_{\ell \in \mathcal{L}} b_t^i(\ell), & \text{otherwise}, \end{cases} \tag{2}$$

where U denotes insufficient evidence rather than another physical state. Free regions enable reaching, blocked regions call for avoidance or retreat, uncertain regions call for probing, and pushable regions may justify a gentle push or sweep. Unlike planning with known object arrangements or detailed pushing dynamics [4, 8], the planner must first build this task-relevant state from weak contact evidence.

Viewed this way, the problem separates into belief update and belief-conditioned control:

$$b_{t+1} = f_\theta(b_t, q_t, a_t, y_{t+1}), \qquad a_t \sim \pi_\psi(q_t, b_t, \mathcal{G}), \tag{3}$$

where f_θ models how contact evidence changes the map and π_ψ models how the map changes action choice, with the goal of reaching $\mathcal{G}$ while limiting unsafe contact. The current controller follows the loop in Fig. 2A. Safe, compliant exploration generates weak cues; simple contact abstractions update the belief; and a manually designed policy selects from probe, reach, retreat, and gentle push. This makes the prototype an evaluation scaffold: the belief should be judged by whether it changes action choices under the same weak observations, not by how completely it reconstructs the scene.

III. Prototype Suite and Planned Evaluation

Fig. 2 summarizes the current prototype suite and the planned evaluation path. The tabletop reach task in Fig. 2B tests whether a belief built from contact changes decisions beyond binary occupancy. A free/occupied map can mark blockage, but it has no state for whether contact should be avoided or used. In mixed movable and immovable clutter, this causes the robot to stop short of the goal. The actionable belief records contact outcomes as free, blocked, uncertain, or pushable regions. This lets the policy probe around hard contacts and apply a bounded push when movable clutter blocks an otherwise useful path.

The hardware experiment in Fig. 2C checks whether the low-cost arm provides enough proprioceptive separation to support those labels. We run bounded probes against movable and immovable objects with a compliant arm controller, estimating force from motor current and tracking error. The arm yields under contact up to a force limit: movable contacts allow continued motion, while immovable contacts quickly reach the limit and stop. These force-motion traces calibrate the contact labels used in hardware reach settings such as the cluttered switch task in Fig. 1C.

The cabinet reach task in Fig. 2D stresses the same interface under constrained 3D contact. Contacts may occur on the gripper, wrist, or arm body. The system only needs a coarse update for action selection. When a contact is detected, it adds immovable belief voxels near the inferred region; the robot then retreats and replans toward the goal. This tests whether contact memory reduces repeated collisions with the same surface while still allowing progress.

Together, these tasks test the same interface between weak contact evidence and action choice. We will compare stopping on contact, random probing, a free/blocked map, the full actionable belief, and an oracle map across sparse clutter, mixed movable/immovable clutter, and constrained cabinet reaching. The ablations separate belief construction from belief use: false blocked or pushable labels expose update errors, while success, probes, repeated blocked contacts, and completion time evaluate control.

IV. Discussion

The next step is to make the full belief-to-action loop learnable from weak proprioceptive evidence. This includes calibrated belief updates from proprioceptive streams and policies that act on this belief under explicit safety and information-gain tradeoffs. This keeps the representation task-level: it need not reconstruct clutter geometry, only the contact state needed for the next safe action.

References

[1] Jeannette Bohg, Karol Hausman, Bharath Sankaran, Oliver Brock, Danica Kragic, Stefan Schaal, and Gaurav S. Sukhatme. Interactive Perception: Leveraging Action in Perception and Perception in Action. *IEEE Transactions on Robotics*, 33(6):1273–1291, 2017. doi: 10.1109/TRO.2017.2721939. URL https://doi.org/10.1109/TRO.2017.2721939.

[2] Sachin Chitta, E. Gil Jones, Matei Ciocarlie, and Kaijen Hsiao. Mobile Manipulation in Unstructured Environments: Perception, Planning, and Execution. *IEEE Robotics & Automation Magazine*, 19(2):58–71, 2012. doi: 10.1109/MRA.2012.2191995. URL https://doi.org/10.1109/MRA.2012.2191995.

[3] Yixiang Dai, Siang Chen, Kaiqin Yang, Dingchang Hu, Pengwei Xie, Guosheng Li, Yuan Shen, and Guijin Wang. Active-Perceptive Language-Oriented Grasp Policy for Heavily Cluttered Scenes. *IEEE Robotics and Automation Letters*, 10(11):11094–11101, 2025. doi: 10.1109/LRA.2025.3604750. URL https://doi.org/10.1109/LRA.2025.3604750.

[4] Mehmet R. Dogar and Siddhartha S. Srinivasa. A Planning Framework for Non-Prehensile Manipulation under Clutter and Uncertainty. *Autonomous Robots*, 33(3):217–236, 2012. doi: 10.1007/s10514-012-9306-z. URL https://doi.org/10.1007/s10514-012-9306-z.

[5] Wanjia Fu, Hongyu Li, Ivy X. He, Stefanie Tellex, and Srinath Sridhar. UniTac: Whole-Robot Touch Sensing Without Tactile Sensors. *arXiv preprint arXiv:2507.07980*, 2025. doi: 10.48550/arXiv.2507.07980. URL https://arxiv.org/abs/2507.07980.

[6] Maged Iskandar, Alin Albu-Schäffer, and Alexander Dietrich. Intrinsic sense of touch for intuitive physical human-robot interaction. *Science Robotics*, 9(93): eadn4008, 2024.

[7] Advait Jain, Marc D. Killpack, Aaron Edsinger, and Charles C. Kemp. Manipulation in Clutter with Whole-Arm Tactile Sensing. *The International Journal of Robotics Research*, 32(4):458–482, 2013. doi: 10.1177/0278364912471865. URL https://doi.org/10.1177/0278364912471865.

[8] Julius Jankowski, Lara Brudermüller, Nick Hawes, and Sylvain Calinon. Robust Pushing: Exploiting Quasi-Static Belief Dynamics and Contact-Informed Optimization. *The International Journal of Robotics Research*, 44(12):1959–1980, 2025. doi: 10.1177/02783649251318046. URL https://doi.org/10.1177/02783649251318046.

[9] Ginga Kennis and Shogo Arai. HEAPGrasp: Hand-Eye Active Perception to Grasp Objects with Diverse Optical Properties. *IEEE Robotics and Automation Letters*, 11(3):3206–3213, 2026. doi: 10.1109/LRA.2026.3653331. URL https://doi.org/10.1109/LRA.2026.3653331.

[10] Jacky Liang and Oliver Kroemer. Contact Localization for Robot Arms in Motion without Torque Sensing. In *IEEE International Conference on Robotics and Automation*, 2021. doi: 10.48550/arXiv.2011.03142. URL https://arxiv.org/abs/2011.03142.

[11] Federico Pizarro Bejarano, Bryson Jones, Daniel Pastor Moreno, Joseph Bowkett, Paul G. Backes, and Angela P. Schoellig. ProDapt: Proprioceptive Adaptation using Long-Term Memory Diffusion. In *IEEE International Conference on Robotics and Automation*, pages 15588–15594, 2025. URL https://arxiv.org/abs/2503.00193.

[12] Muhammad Suhail Saleem, Rishi Veerapaneni, and Maxim Likhachev. A POMDP-Based Hierarchical Planning Framework for Manipulation under Pose Uncertainty. *arXiv preprint arXiv:2409.18775*, 2024. doi: 10.48550/arXiv.2409.18775. URL https://arxiv.org/abs/2409.18775.

[13] Muhammad Suhail Saleem, Lai Yuan, and Maxim Likhachev. A Contact-Driven Framework for Manipulating in the Blind. *arXiv preprint arXiv:2510.20177*, 2025. doi: 10.48550/arXiv.2510.20177. URL https://arxiv.org/abs/2510.20177.

[14] Tony Z. Zhao, Vikash Kumar, Sergey Levine, and Chelsea Finn. Learning Fine-Grained Bimanual Manipulation with Low-Cost Hardware. In *Proceedings of Robotics: Science and Systems*, Daegu, Republic of Korea, 2023. doi: 10.15607/RSS.2023.XIX.016. URL https://doi.org/10.15607/RSS.2023.XIX.016.

[15] Sheng Zhong, Nima Fazeli, and Dmitry Berenson. Soft Tracking Using Contacts for Cluttered Objects to Perform Blind Object Retrieval. *IEEE Robotics and Automation Letters*, 2022. doi: 10.48550/arXiv.2201.10434. URL https://arxiv.org/abs/2201.10434.