

TACTFUL: Tactile-Driven Exploration For Object Localization and Identification in Confined Environments

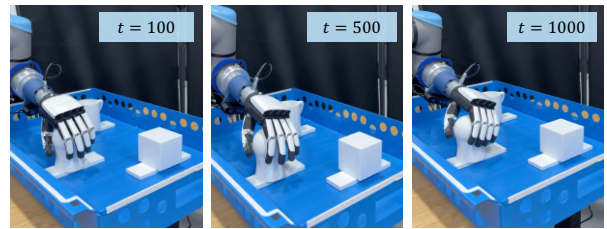
Shivani Kamtikar^{1,2}, Chung Hee Kim^{1,3}, Camilla Tabasso¹, Tye Brady¹, Joshua Migdal¹, Taşkın Padır¹

Abstract—Humans effortlessly locate and identify objects by touch alone, even without vision. In contrast, robotic systems rely heavily on vision and struggle with autonomous tactile exploration and object identification. We present TACTFUL, a vision-free tactile exploration framework that enables a multi-fingered robot to autonomously explore confined workspaces, discover objects through contact, and identify them via tactile reconstruction. Trained entirely on real hardware without simulation, our system learns a single policy that balances global workspace exploration with local surface refinement through a dynamic reward schedule. Our results demonstrate that tactile sensing, when paired with structured learning, can serve as an effective primary modality for object-level reasoning, achieving 77% success with 0.015 m average reconstruction error and outperforming baseline approaches on real-world objects.

I. INTRODUCTION

Humans possess a remarkable ability to locate and identify objects in visually occluded scenarios, through a combination of tactile sensing and prior geometric knowledge, effortlessly distinguishing between a set of keys and a water bottle in a cluttered bag. Tactile sensors offer direct, meaningful feedback about contact interactions and local surface geometry [1]–[3], particularly in occluded or confined settings. Yet, robotic manipulation has traditionally relied heavily on vision, with tactile sensing often used as a secondary or corrective modality [4]–[7]. While prior work has explored tactile-based object mapping and identification [8]–[11], autonomous tactile exploration for object-level reasoning remains underexplored in real-world settings. Moreover, most prior tactile systems have primarily been evaluated in simulation using simplified or binary contact signals, with comparatively limited work demonstrating real-world performance using high-fidelity tactile data [12], [13]. Consider a robot exploring a confined workspace containing multiple objects. When visual perception is available, object localization and identification are relatively straightforward, as camera-based systems can provide dense scene observations. In vision-denied scenarios, robots must rely on local tactile feedback

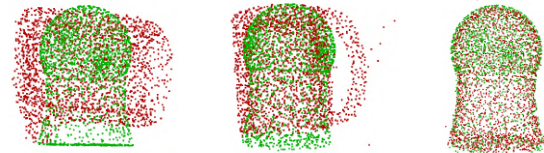
to explore workspaces, localize objects, and reconstruct geometry. This requires simultaneously addressing workspace exploration for object discovery and contact-driven surface reconstruction for identification.



(a) Active Tactile Exploration



(b) Generated Tactile Point Cloud



(c) R_t (red) & R_{GT} (green)

Fig. 1: (a) Active tactile exploration by the agent at various timesteps (no vision), (b) tactile point cloud accumulated at timestep t , and (c) shape-completed reconstruction, R_t (red), overlaid on the ground truth geometry, R_{GT} (green)

We present *TACTFUL*, a real-world, vision-free tactile exploration and identification framework for settings where objects are drawn from a known object library but must be located without vision, such as warehouse bin-picking, medical instrument retrieval, or in manufacturing industries. Our framework makes the following contributions: 1) we develop a learned tactile exploration policy that enables a multi-fingered robot to autonomously *explore* and *localize* objects within confined workspaces, 2) we introduce a dynamic reward schedule that balances global exploration with local contact refinement, leveraging object-level geometric priors to guide 3D reconstruction, and finally, 3) we demonstrate reliable object *identification* based solely on reconstructed tactile geometry.

II. RELATED WORK

Tactile-based localization. Tactile sensing has been explored as a means for object localization when visual information is unreliable or unavailable. Early work demonstrated tactile

*This work was done when Shivani Kamtikar and Chung Hee Kim were interns at Amazon.

¹Amazon Fulfillment Technologies & Robotics, Westborough, MA, USA. {bradytye, ptaskin, jmigdal}@amazon.com

²Siebel School of Computing and Data Science, University of Illinois at Urbana-Champaign, Champaign, IL, USA. skk7@illinois.edu

³Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, USA. chunghek@andrew.cmu.edu

Taşkın Padır holds concurrent appointments as a Professor of Electrical and Computer Engineering at Northeastern University and as an Amazon Scholar. This paper describes work performed at Amazon and is not associated with Northeastern University.

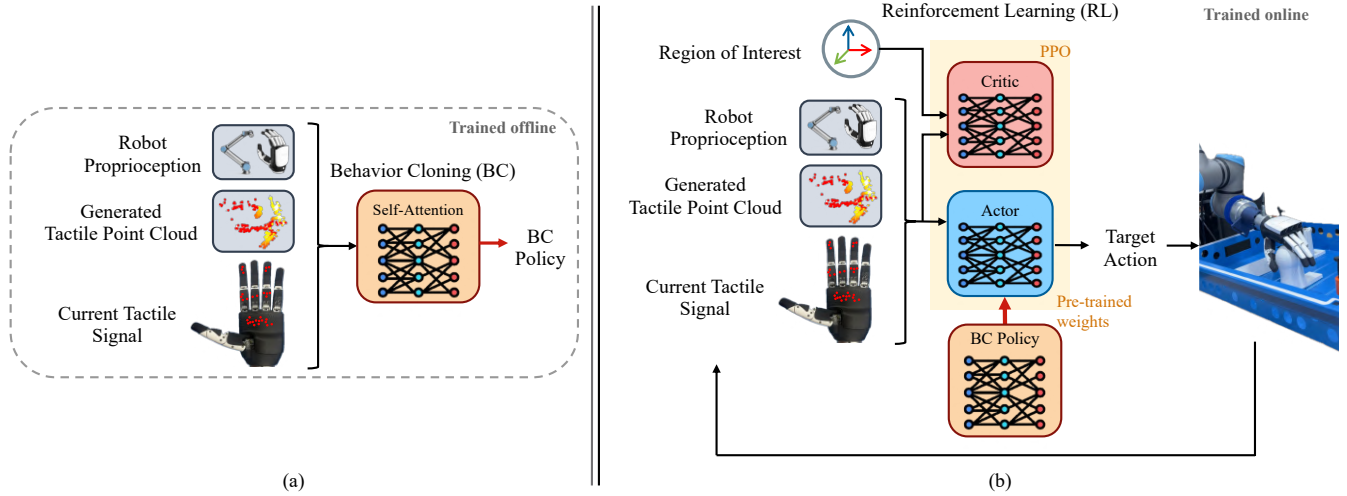


Fig. 2: **Pipeline of TACTFUL:** (a) Offline behavior cloning (BC) model trained using proprioception, tactile signals, and tactile point cloud; used to initialize the actor in the PPO-based reinforcement learning (RL) policy. (b) Online PPO training with the same inputs and an additional region of interest (ROI) for the critic, resulting in an exploration policy for object-guided reconstruction in confined spaces.

SLAM using whisker-inspired sensors [14], while subsequent approaches employed simple tactile probes for active localization [15]–[17]. However, these methods typically rely on low-dimensional contact signals and lack the rich spatial information afforded by multi-fingered tactile hands. Kissoum *et al.* [18] addressed simultaneous localization and reconstruction using particle filters, but their approach is limited to 2D simulated environments with simplified contact dynamics. In contrast, our work leverages high-resolution, real-world tactile sensing to jointly localize and reconstruct objects through learned exploration policies.

Object reconstruction using tactile data. Reconstructing object geometry from touch has gained increasing attention in vision-denied settings, with some works exploring efficient shape exploration using rigid tactile arrays [19] and learning-based methods that fuse contact positions and normals to build 3D object representations [12]. Other approaches focus on reconstructing visually challenging objects, including transparent objects [20]. More recent systems, such as TactoFind [13], demonstrate tactile-only object retrieval in clutter, but depend on hand-crafted exploration heuristics rather than learned policies. However, the majority of tactile reconstruction methods assume that the object has already been localized and focus exclusively on surface exploration. Our work addresses a more general and realistic setting in which the robot must first explore an unknown workspace to discover objects and then perform detailed tactile interaction to reconstruct their geometry, requiring policies that tightly couple global exploration with local shape refinement.

Deep reinforcement learning for tactile perception. Deep reinforcement learning has been widely applied to tactile-driven perception and exploration tasks. Prior work has highlighted the importance of memory for sequential tactile decision-making, using recurrent architectures such as LSTM-augmented A3C [21], [22]. More recent approaches leverage attention mechanisms to generalize tactile perception across manipulation primitives [16]. Similarly, Lee *et*

al. [23] apply reinforcement learning for occluded object retrieval, but focus on reactive behaviors rather than systematic object reconstruction. In the context of exploration, intrinsic motivation strategies such as D-optimality [24] and contact-based exploration bonuses [10] have been proposed to encourage informative interactions. However, these methods typically operate under the assumption that object locations are known and focus on localized surface exploration. In contrast, our approach learns a single tactile-only policy that explores an entirely unknown workspace, localizes objects through contact, and incrementally reconstructs their geometry in the real world using a structured reward design that balances search, interaction, and shape understanding.

III. TACTILE ROBOTIC SETUP

Our system, shown in Fig. 1, consists of a five-fingered Inspire RH56DFTP Series Dexterous hand mounted on a UR10e robotic arm. The robotic hand is equipped with resistive 1062 high-resolution taxels on the fingertips, fingerpads, and the palm area, measuring normal contact forces. The workspace consists of a bin with known bounds having 3D-printed objects placed in arbitrary locations. We do not use any cameras.

IV. METHOD

We address vision-free object localization and identification in confined workspaces containing multiple objects. Given a bin of objects drawn from a known library, the robot is assigned a target object and must locate it using only tactile feedback. During exploration, it may encounter distractor objects and must distinguish the target through accurate tactile reconstruction. This requires jointly exploring the workspace to discover objects via contact and reconstructing surface geometry with sufficient fidelity for reliable identification. TACTFUL learns a single reinforcement learning (RL) policy that enables a multi-fingered robot to autonomously explore and identify objects using tactile sensing alone. The policy

is trained with a dynamic reward schedule that balances global workspace exploration with local, contact-rich surface refinement. At the start of each episode, object locations are unknown, requiring active exploration through continuous end-effector and finger motions. As the robot moves, it records end-effector poses and accumulates tactile contacts, forming a sparse tactile point cloud of the contacted surface. Once sufficient data is collected, the sparse point cloud is passed to a learned shape completion model to produce a dense reconstruction, which is then matched against a known object library for identification.

A. Behavior Cloning Initialization

Learning tactile exploration policies directly through RL is particularly challenging in real-world settings, where random exploration is inefficient and can lead to unsafe contacts [25], [26]. Furthermore, the lack of high-fidelity tactile simulations [27] prevents reliable pre-training in simulation. To overcome these limitations, we initialize the RL policy using Behavior Cloning (BC) [28], [29], allowing it to learn structured exploration behaviors from 5311 tactile-only teleoperated demonstrations, consistent with our vision-free policy. During BC training, the policy is trained to regress the expert action, represented as a concatenation of delta end-effector pose and delta hand configuration, given the current observation state. The BC policy training pipeline is shown in Fig. 2(a). The loss function (L_{BC}) is formulated as a mean squared error (MSE) between predicted and expert action deltas:

$$L_{BC} = \frac{1}{N} \sum_{i=1}^N \|a_i^{\text{pred}} - a_i^{\text{expert}}\|_2^2, \quad (1)$$

where a_i^{pred} and a_i^{expert} denote the predicted and expert actions (detailed in Section IV-B) at sample i , respectively, and N is the number of demonstration samples. The resulting BC policy captures basic contact strategies and local exploration patterns demonstrated by the expert. This pretrained policy is then used to initialize the RL policy (Section IV-B).

B. POMDP Formulation

We formulate the tactile exploration task as a Partially Observable Markov Decision Process (POMDP) defined by the tuple $M = (S, A, T, \mathcal{O}, \Omega, R, \gamma)$, where S is the underlying state space, A is the action space, $T : S \times A \rightarrow S$ is the state transition function, $\mathcal{O}$ is the observation space, $\Omega : S \rightarrow \mathcal{O}$ is the observation function mapping states to observations, $R : S \times A \rightarrow R$ is the reward function, and $\gamma \in [0, 1]$ is the discount factor. The true state includes the complete 3D geometry and pose of all objects in the workspace, information that is not directly observable to the agent. Instead, the agent must infer object locations and shapes through accumulated tactile feedback over time. At each timestep t , the agent receives an observation $o_t = \Omega(s_t) \in \mathcal{O}$ consisting of proprioceptive information (end-effector pose, joint angles, hand configuration) and high-resolution tactile sensor readings from the 1062 taxels distributed across the

robot hand. These observations provide only local, contact-based information about the environment, making the process partially observable. The goal is to learn a policy $\pi : \mathcal{O}^k \rightarrow A$ that maps a history of k recent observations to actions that maximize the expected cumulative discounted reward:

$$\pi^* = \arg \max_{\pi} \mathbb{E}_{\tau \sim \pi} \left[\sum_{t=0}^T \gamma^t r_t \right] \quad (2)$$

where $\tau = (o_0, a_0, r_0, o_1, a_1, r_1, \dots)$ is a trajectory sampled under policy π , and T is the episode horizon. We solve this POMDP using Proximal Policy Optimization (PPO) [30], an on-policy actor-critic reinforcement learning algorithm.

Observation Space ($\mathcal{O}$): At timestep t , the agent observes $o_t \in \mathcal{O}$ consisting of:

- **Tactile readings** $f_t \in \mathbb{R}^{1062}$: Force measurements from each taxel, indicating contact locations and intensities across the hand surface.
- **Tactile Point Cloud** $c_t \in \mathbb{R}^{1088}$: Learned embedding of the accumulated tactile point cloud history obtained through Sonata [31], a self-supervised transformer-based point cloud encoder. This embedding encodes both spatial coverage and geometric structure of the explored surface across the entire episode history, providing the agent with a compressed representation of its exploration progress.
- **End-effector pose** $p_t \in \mathbb{R}^9$: Normalized position and sine-cosine encoded orientations.
- **Joint angles** $j_t \in \mathbb{R}^6$: Normalized arm joint configurations from forward kinematics.
- **Hand configuration** $h_t \in \mathbb{R}^6$: Normalized finger joint angles from forward kinematics.

At each timestep t , a single observation $o_t \in \mathbb{R}^{2171}$ is constructed as:

$$o_t = [p_t, j_t, h_t, f_t, c_t] \in \mathbb{R}^{2171} \quad (3)$$

State Representation (s_t): Since individual observations provide insufficient context for long-horizon tactile exploration (e.g., determining movement direction along a surface requires temporal information), the state is represented as a sequence of $k = 10$ recent observations:

$$s_t = o_t, o_{t-1}, \dots, o_{t-(k-1)} \quad (4)$$

This temporal sequence is processed by a transformer architecture (described in IV-D), which naturally handles the sequential dependencies through its self-attention mechanism, eliminating the need for explicit recurrent connections.

Action Space (A): The action space is defined as:

$$a_t = [\Delta p_t, \Delta h_t] \in \mathbb{R}^{12} \quad (5)$$

where $\Delta p_t \in [-1, 1]^6$ represents incremental changes to the end-effector pose, and $\Delta h_t \in [-1, 1]^6$ represents changes to the finger joint angles. Actions are normalized using z-score standardization to stabilize training.

Partial Observability: A challenge in our setting is that the agent never observes the true object locations, geometries, or the completeness of its exploration. Without vision, the agent must rely entirely on the history of tactile contacts and proprioceptive feedback to build an implicit understanding of the workspace structure, necessitating memory mechanisms (the point cloud embedding c_t and observation history) to enable effective exploration and reconstruction.

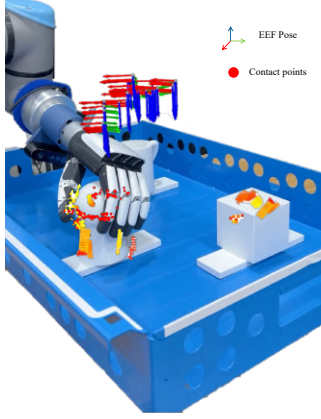


Fig. 3: Example map and the tactile point cloud generated during exploration, with the agent tracking end-effector (EEF) poses and contact regions.

C. Dynamic Reward Design

We design a composite reward function to guide the agent toward effective tactile exploration, shape reconstruction, and diverse surface coverage. The different reward components are explained below.

Contact Reward (r_{contact}): Promotes informative surface exploitation by encouraging the agent to refine contact around already discovered objects through novel, spatially distinct tactile interactions. For each activated taxel, we compute its 3D position in world coordinates by combining the taxel's local position on the hand surface with the current end-effector pose p_t and hand configuration h_t . We also maintain a history of all contact points accumulated throughout each episode. By referencing this history, actions that create new, previously unseen contact points are incentivized rather than redundant or repetitive touches. A new contact point is considered novel if its Euclidean distance to all previously accumulated points exceeds a minimum distance threshold of 5 mm. It is computed as the number of new, previously unseen points in the reconstruction during time step t , normalized by the maximum possible number of points:

$$r_{\text{contact}} = \frac{\text{diff}_t}{N_{\text{max}}} \quad (6)$$

where diff_t denotes the number of new points, and N_{max} is the normalization constant.

Exploration Reward (r_{explore}): Encourages the agent to explore new, previously unvisited regions in the workspace. It incentivizes the agent to explore more areas in the confined tray. We track the agent's trajectory by keeping a memory of all visited end effector (EEF) poses. We discretize the workspace into a 1cm^3 voxel grid and mark

a voxel as explored when the end-effector center enters it. By incorporating this history, actions that lead to the agent exploring previously unvisited areas in the tray are rewarded prominently. This also helps the agent to prevent getting stuck in local minima by exploring the same region repeatedly. Adding this reward ensures that the agent has a broader understanding of the workspace and, in turn, helps in convergence. r_{explore} is defined as:

$$r_{\text{explore}} = \frac{V_t^{\text{new}}}{V_{\text{max}}}, \quad (7)$$

where V_t^{new} is the number of new voxels explored at time t , and V_{max} is the maximum number of voxels in the grid.

Reconstruction Reward (r_{recon}): Encourages the agent to produce accurate and complete object reconstructions as it explores by penalizing imprecise or incomplete surface representations. Specifically, we use the Chamfer distance, which measures the average closest-point discrepancy between two point sets, to quantify reconstruction quality. A lower Chamfer distance indicates better alignment and higher fidelity to the true object shape. By incorporating this metric, the agent is incentivized to exploit gathered tactile data to refine local geometry and close coverage gaps, leading to more precise object understanding. This reward thus complements r_{contact} and r_{explore} : while the r_{contact} promotes local, high-resolution coverage and r_{explore} encourages global scene coverage, r_{recon} ensures that the overall shape representation remains faithful to the true object geometry. At each timestep t , the accumulated sparse tactile point cloud is passed through a pretrained shape completion network (detailed in IV-E) to generate a dense reconstructed point cloud R_t . This reconstruction is compared against the ground-truth point cloud R_{GT} of the target object using normalized Chamfer distance:

$$r_{\text{recon}} = \frac{\text{Chamfer}(R_t, R_{GT})}{\text{Chamfer}_{\text{max}}}, \quad (8)$$

The total reward at each time step t is given by:

$$r_t = \alpha_t r_{\text{contact}} + \beta_t r_{\text{explore}} - \lambda_t r_{\text{recon}}, \quad (9)$$

where $\alpha_t, \beta_t, \lambda_t \in \mathbb{R}^+$ are the time-varying weighting coefficients.

To balance global workspace exploration early in an episode with local surface refinement later, we employ a time-varying reward schedule in which the weights of the reward components evolve as a function of the timestep t and episode horizon T . Specifically, we use a linear schedule:

$$\alpha(t) = \alpha_{\min} + (\alpha_{\max} - \alpha_{\min}) \cdot \frac{t}{T} \quad (10)$$

$$\beta(t) = \beta_{\max} - (\beta_{\max} - \beta_{\min}) \cdot \frac{t}{T} \quad (11)$$

$$\lambda(t) = \lambda_{\min} + (\lambda_{\max} - \lambda_{\min}) \cdot \frac{t}{T} \quad (12)$$

Here $\beta(t)$ initially prioritizes workspace coverage to facilitate object discovery through incidental contact, while $\alpha(t)$

and $\lambda(t)$ progressively increase to emphasize informative surface interaction and penalize inaccurate reconstructions once contact is established. We set $\alpha_{\min} = 0.1$, $\alpha_{\max} = 0.5$, $\beta_{\min} = 0.1$, $\beta_{\max} = 0.7$, $\lambda_{\min} = 0.2$, and $\lambda_{\max} = 0.6$. Although a single policy is used throughout an episode, this evolving reward landscape induces a behavioral transition from broad exploratory motions to deliberate, contact-rich interactions around discovered objects. Importantly, r_{explore} remains active throughout the episode, ensuring that late object discovery does not prevent successful reconstruction.

D. Training Procedure

The RL policy, shown in Fig. 2(b), is trained entirely using data collected on the hardware setup with no simulation. Both the actor (policy) and critic (value) networks share the same architecture, consisting of two transformer layers with an embedding dimension of 512 and 4 attention heads. An adaptive average pooling layer is applied to condense the sequence of encoded observations into a global representation, allowing the model to capture long-range temporal dependencies and high-dimensional tactile data into a compact embedding. We train for 10 episodes with a maximum horizon of 2000 steps per episode. We use standard PPO hyperparameters: batch size of 16, initial learning rate of 0.001 with linear decay, clipping $\epsilon = 0.2$, discount factor $\gamma = 0.99$ and 0.016 KL threshold to ensure stable policy updates. The observation history length is $k = 10$, and the workspace voxel grid uses 1cm resolution for exploration tracking. To stabilize value estimation and guide learning in early stages, we apply *asymmetric observation* between the policy and value networks [21], [23]. The value network is provided with additional privileged information - the approximate 6-DoF pose (within 0.10 m) of the target object (region of interest as shown in Fig. 2(b)). This is not accessible to the policy network during deployment. Note that when multiple objects are present in the workspace, the critic only observes the target object’s approximate pose, not the poses of distractor objects.

E. Shape Completion

To transform sparse tactile measurements into complete object representations, we employ a point cloud diffusion model [32] trained to perform shape completion from partial observations. The model is pretrained on a diverse object set and learns to infer full object geometry from partial, contact-based point clouds. Given tactile measurements, it predicts a dense reconstruction that can be directly compared against the target object. This model effectively serves as a learned prior over our known object set, allowing us to bridge the gap between sparse tactile observations and complete object geometries.

V. EXPERIMENTS AND RESULTS

A. Task Formulation

We evaluate our framework on three geometrically diverse real-world objects (Fig. 4), selected based on graspability metrics [33]: a *cube* (high graspability), a *deformed cylinder*,

and a *deformed cup* (lower graspability). In each episode, all three objects are placed in a bin with known spatial bounds at randomized positions. The robot is assigned one of the three objects as its target.

Episode Termination: An episode terminates under three conditions: (1) *Success*: the agent correctly identifies the target object (shape completion achieves Chamfer distance below 0.02 m for 10 consecutive timesteps), (2) *Failure*: the agent incorrectly identifies a non-target object with high confidence, or (3) the maximum horizon $T = 2000$ steps is reached. Safety violations (workspace boundary exceeded or collision detected) also trigger termination and are counted as failures.

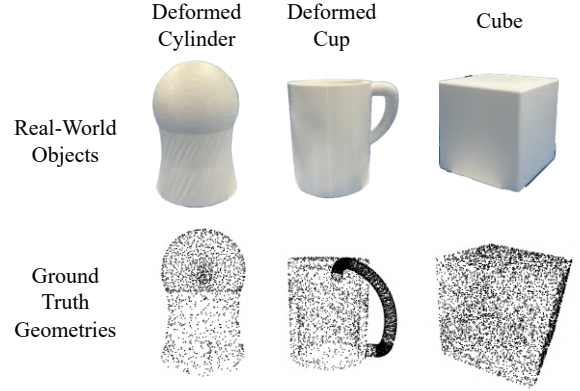


Fig. 4: Real-world objects used for the experiments, along with the ground truth point clouds for each of them.

Metrics: The tests are evaluated based on 2 different success metrics:

- 1) *Success Rate*: This is the success rate over trials. A trial is a success if the agent identifies and locates the target object.
- 2) *Chamfer- L_2* : We measure the Chamfer- L_2 distance between the generated (shape-completed) and ground-truth point clouds. Let $P_{\text{GT}} = \{\mathbf{p}_i\}_{i=1}^N$ be the set of points in the aligned ground truth point cloud, and $P_t = \{\mathbf{q}_j\}_{j=1}^M$ be the set of points in the reconstructed point cloud. The one-way Chamfer distances are defined as:

$$\text{Chamfer}_{\text{GT} \rightarrow t} = \frac{1}{N} \sum_{i=1}^N \min_{\mathbf{q}_j \in P_t} \|\mathbf{p}_i - \mathbf{q}_j\|_2 \quad (13)$$

$$\text{Chamfer}_{t \rightarrow \text{GT}} = \frac{1}{M} \sum_{j=1}^M \min_{\mathbf{p}_i \in P_{\text{GT}}} \|\mathbf{q}_j - \mathbf{p}_i\|_2 \quad (14)$$

The full symmetric Chamfer distance is the sum of both directions:

$$\text{Chamfer}(P_{\text{GT}}, P_t) = \text{Chamfer}_{\text{GT} \rightarrow t} + \text{Chamfer}_{t \rightarrow \text{GT}} \quad (15)$$

Baselines (All results detailed in Section V-B and Table I)- We do not compare against visuotactile baselines (e.g., [4], [6], [7]) as they assume visual observability, which directly contradicts our vision-denied setting. Similarly, purely vision-based methods are inapplicable. We instead compare against tactile-only approaches, which operate under identical sensory constraints.

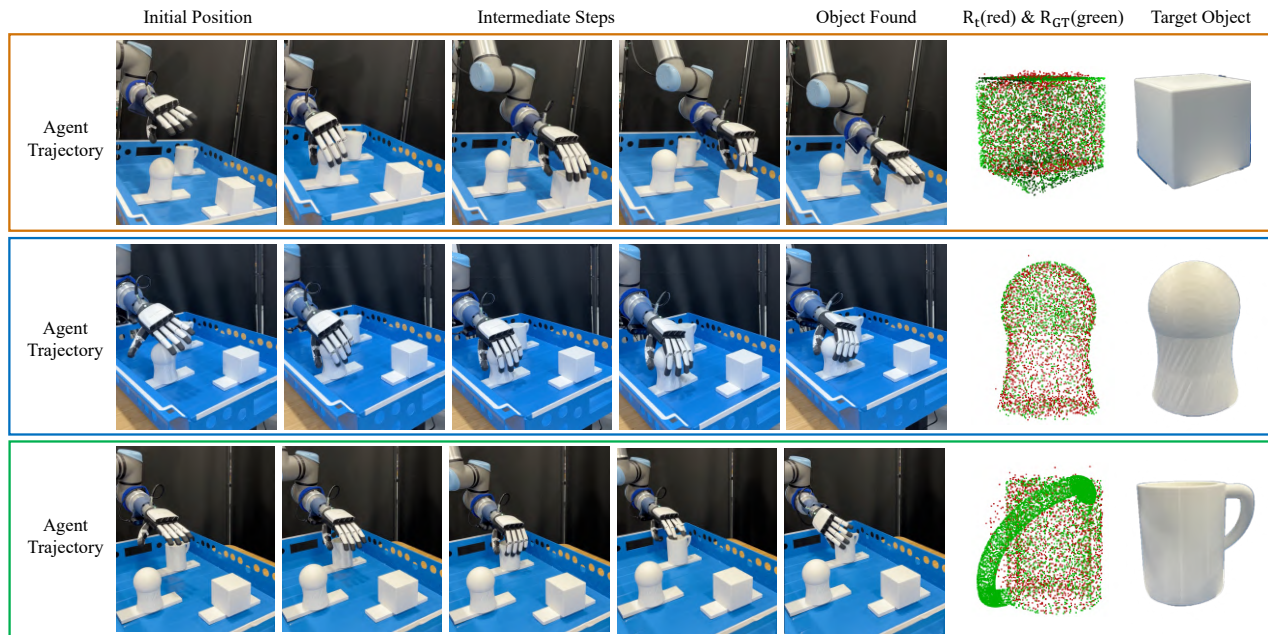


Fig. 5: **Experimental results (one example test run for each of the objects)**: Shows initial view, three intermediate views, and the final view (end of test) of the exploration sequence. The shape-completed reconstruction, R_t (red), overlaid on the ground truth geometry, R_{GT} (green), is also shown. The final column shows the goal object that the robot was tasked to identify. (Also see supplementary video.)

TABLE I: **Quantitative results on exploration and object identification**. Each row shows results when that object is the designated target, with all three objects present in the workspace. The table shows the average success rate over 12 trials for every object is shown along with the Chamfer distance (m) (Mean $\pm$ STD) for all methods. Additionally, the initial distance from the target (Init. Δ (m)) is also given.

#	Method	Deformed Cylinder			Cube			Deformed Cup		
		Success $\uparrow$	Chamfer- L_2 $\downarrow$	Init. Δ	Success $\uparrow$	Chamfer- L_2 $\downarrow$	Init. Δ	Success $\uparrow$	Chamfer- L_2 $\downarrow$	Init. Δ
1	Heuristic	0.416	0.050 $\pm$ 0.023	0.230 $\pm$ 0.029	0.5	0.052 $\pm$ 0.020	0.318 $\pm$ 0.004	0.416	0.130 $\pm$ 0.100	0.334 $\pm$ 0.009
2	BC only	0.416	0.044 $\pm$ 0.015	0.246 $\pm$ 0.062	0.583	0.063 $\pm$ 0.007	0.253 $\pm$ 0.063	0.583	0.120 $\pm$ 0.100	0.331 $\pm$ 0.071
3	RL w/o BC	0.750	0.020 $\pm$ 0.001	0.201 $\pm$ 0.041	0.750	0.023 $\pm$ 0.003	0.256 $\pm$ 0.048	0.583	0.031 $\pm$ 0.015	0.321 $\pm$ 0.061
4	Ours	0.833	0.010$\pm$0.004	0.214$\pm$0.051	0.833	0.012$\pm$0.009	0.255$\pm$0.059	0.667	0.021$\pm$0.011	0.322$\pm$0.068

1) *Heuristic Exploration Policy (Heuristic)*: For the heuristic baseline, we implemented a policy similar to TactoFind [13], with the addition of more hand configurations like full grasp, partial grasp, and wrist movements. The policy first locates the objects within the voxelized bin to map them, then systematically explores them to obtain fine-grained tactile data for identification.

2) *Behavior Cloning (BC only)*: We use only our BC policy without the addition of the RL policy and rewards.

3) *Reinforcement Learning without Behavior Cloning (RL w/o BC)*: We train the RL policy without a BC initialization.

Effect of Various Reward Functions: We conduct an ablation study to evaluate the contribution of each reward component in our PPO training framework. We train the PPO model with only 2 reward functions at a time and conduct experiments on all three objects using this model. Results detailed in V-B and Table II.

Effect of Shape Completion Model: We ablate the shape completion model by removing it and comparing the raw sparse point cloud to the ground truth, evaluating its impact on success rate and Chamfer distance. Results detailed in V-B and Table III.

B. Experimental Results

Result 1: Our method enables effective tactile-driven exploration for object localization and identification.

Despite the challenging setting where three objects are present simultaneously and the agent must locate a specific target while avoiding false identifications of distractors, our method consistently outperforms both *BC Only* and *Heuristic* baselines in terms of reconstruction accuracy, consistency, and object identification success. As shown in Table I, we achieve higher success rates and lower Chamfer- L_2 distances across all objects. Fig. 5 illustrates representative test runs, while Fig. 6(a) shows a steady decrease in reconstruction error over training episodes.

The *Heuristic* baseline, which follows predefined exploration patterns without learning from contact feedback, exhibits high variance in contact coverage and reconstruction quality. Although it occasionally succeeds through incidental dense contact, it lacks consistency across trials and object geometries. In contrast, our method achieves significantly lower variance in Chamfer distance, indicating more reliable and repeatable behavior (Fig. 6(b)).

Compared to the *BC Only* policy, which is constrained by offline expert demonstrations, our policy improves performance through on-policy adaptation guided by structured

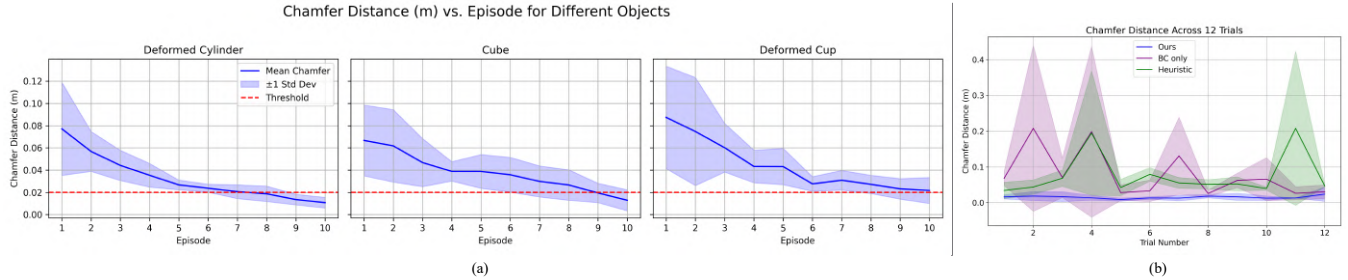


Fig. 6: **(a) Chamfer distance vs Episode of our method across three objects:** Each curve shows the mean Chamfer distance over 10 episodes, with shaded regions representing ± 1 standard deviation. The decreasing trend shows improved reconstruction accuracy over training. ; **(b) Chamfer distance comparison for various policies:** Plot shows the mean Chamfer distance over 12 trials for each method averaged across three test objects, with shaded regions indicating standard deviation. The plot shows that our method achieves significantly lower standard deviation in Chamfer distance, indicating more reliable and consistent behavior)

reward signals. BC initialization gives RL a structured prior over safe contact strategies, avoiding costly exploration from scratch on real hardware. This behaviour also explains why our policy performs better than the *RL w/o BC* baseline, as it allows the policy to focus on refining exploration and reconstruction behaviours rather than discovering viable contact strategies from scratch, which is time-consuming in real-world setups.

Qualitatively, the learned policy produces contact trajectories that target geometrically informative and previously unexplored regions, enabling the shape completion model to generate more accurate reconstructions (Fig. 5). This leads to earlier and more confident episode termination based on reconstruction quality. Despite allowing independent control of all six finger joints, the policy exhibits coordinated grasp-like motions interleaved with sliding and grazing behaviors (see supplementary video). Such coordination emerges naturally from BC initialization and reward shaping and yields richer tactile information per action, particularly on curved or convex surfaces.

TABLE II: **Ablation study on the effect of different reward components.** All three rewards are essential; removing any leads to reduced exploration, contact quality, or reconstruction accuracy.

Reward Components	Success $\uparrow$	Chamfer- L_2 (m) $\downarrow$
$r_{\text{recon}} + r_{\text{contact}}$	0.33	0.045 ± 0.025
$r_{\text{recon}} + r_{\text{explore}}$	0.41	0.054 ± 0.032
$r_{\text{explore}} + r_{\text{contact}}$	0.50	0.060 ± 0.041
$r_{\text{recon}} + r_{\text{contact}} + r_{\text{explore}}$	0.77	0.015 ± 0.010

Result 2: All three rewards are necessary for effective exploration and object reconstruction. Our results, Table II, indicate that all three rewards are necessary to effectively guide the agent through the stages of workspace exploration, object interaction, and shape reconstruction. Removing the r_{explore} causes the agent to remain near its initial configuration, leading to failures unless an object is encountered early. Excluding r_{contact} results in poor surface coverage despite successful object discovery, while omitting r_{recon} yields partial reconstructions with higher geometric error. Only the full reward formulation consistently achieves high success rates and low reconstruction error, confirming the necessity of jointly balancing exploration, contact quality, and reconstruction accuracy.

TABLE III: **Ablation study on the role of shape completion.** Shape completion improves object identification success and reconstruction accuracy, reducing ambiguity from sparse tactile inputs.

Method	Success $\uparrow$	Chamfer- L_2 (m) $\downarrow$
No shape completion	0.34	0.038 ± 0.050
With shape completion	0.77	0.015 ± 0.010

Result 3: Shape completion model is essential for robust object identification. As shown in Table III, removing the shape completion model significantly reduces object identification success, even when the Chamfer distance appears low. Sparse tactile point clouds can be geometrically ambiguous and may incidentally match multiple objects, leading to premature episode termination without correct identification. The learned shape completion model provides a strong prior that resolves this ambiguity by producing dense, object-specific reconstructions, resulting in substantially improved success rates and reconstruction accuracy.

VI. LIMITATIONS

While our framework demonstrates effective tactile exploration and object localization in real-world settings, several limitations warrant discussion. Our method assumes operation from a known library of three object models. While our experiments demonstrate the method’s core mechanisms, broader validation on larger object sets (e.g., YCB benchmark objects) and more complex geometries would strengthen generalization claims. Real-world data collection remains costly, limiting our experimental scope. Furthermore, our experiments focus on static objects, as estimating object displacement or pose changes without vision is challenging. Extending the framework to dynamic environments can be addressed by integrating visuo-tactile sensing to infer object displacement.

VII. CONCLUSION

In this work, we presented *TACTFUL*, a tactile-based framework for autonomous exploration and object identification in environments without visual perception. Our policy trained in the real-world enables a multi-fingered robot to *explore* an unknown workspace, *localize* objects through contact, and *reconstruct* their 3D geometry without any

visual input. Our results demonstrate that our method outperforms baselines in tactile exploration and object identification through reconstruction in real-world settings. Quantitatively, we observed an overall success rate of 77% with an average Chamfer- L_2 loss of 0.015 m. Qualitatively, we observe that the agent learns to explore the workspace effectively to perform purposeful contact interactions that lead to accurate reconstructions. These results highlight the potential of tactile sensing not only as a reactive fallback for vision but as a primary modality for perception in unstructured, occluded, or confined environments. Future work will extend our system to dynamic scenes and improved generalization.

REFERENCES

- [1] W. M. Bergmann Tiest and A. M. Kappers, "The influence of visual and haptic material information on early grasping force," *Royal Society open science*, vol. 6, no. 3, p. 181563, 2019.
- [2] X. Wei, B. Wang, Z. Wu, and Z. L. Wang, "An open-environment tactile sensing system: toward simple and efficient material identification," *Advanced Materials*, vol. 34, no. 29, p. 2203073, 2022.
- [3] H. Li, J. Akl, S. Sridhar, T. Brady, and T. Padir, "Vita-zero: Zero-shot visuotactile object 6d pose estimation," 2025. [Online]. Available: <https://www.amazon.science/publications/vita-zero-zero-shot-visuotactile-object-6d-pose-estimation>
- [4] J. Hansen, F. Hogan, D. Rivkin, D. Meger, M. Jenkin, and G. Dudek, "Visuotactile-rl: Learning multimodal manipulation policies with deep reinforcement learning," in *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 2022, pp. 8298–8304.
- [5] B. Huang, Y. Wang, X. Yang, Y. Luo, and Y. Li, "3d-vitac: Learning fine-grained manipulation with visuo-tactile sensing," *arXiv preprint arXiv:2410.24091*, 2024.
- [6] M. Bauza, A. Bronars, Y. Hou, I. Taylor, N. Chavan-Dafle, and A. Rodriguez, "Simple, a visuotactile method learned in simulation to precisely pick, localize, regrasp, and place objects," *Science Robotics*, vol. 9, no. 91, p. eadi8808, 2024.
- [7] K. Li, P. Li, T. Liu, Y. Li, and S. Huang, "Maniptrans: Efficient dexterous bimanual manipulation transfer via residual learning," in *Proceedings of the Computer Vision and Pattern Recognition Conference*, 2025.
- [8] M. Bauza, O. Canal, and A. Rodriguez, "Tactile mapping and localization from high-resolution tactile imprints," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 3811–3817.
- [9] Q. Li, O. Kroemer, Z. Su, F. F. Veiga, M. Kaboli, and H. J. Ritter, "A review of tactile information: Perception and action through touch," *IEEE Transactions on Robotics*, vol. 36, no. 6, pp. 1619–1634, 2020.
- [10] A.-H. Shahidzadeh, S. J. Yoo, P. Mantripragada, C. D. Singh, C. Fermüller, and Y. Aloimonos, "Actexplore: Active tactile exploration on unknown objects," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2024, pp. 3411–3418.
- [11] X. Mao, G. Giudici, C. Coppola, K. Althoefer, I. Farkhatdinov, Z. Li, and L. Jamone, "Dexskills: Skill segmentation using haptic data for learning autonomous long-horizon robotic manipulation tasks," in *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2024, pp. 5104–5111.
- [12] J. Xu, H. Lin, S. Song, and M. Ciocarlie, "Tandem3d: Active tactile exploration for 3d object recognition," *arXiv preprint arXiv:2209.08772*, 2022.
- [13] S. Pai, T. Chen, M. Tippur, E. Adelson, A. Gupta, and P. Agrawal, "Tactofind: A tactile only system for object retrieval," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*, 2023, pp. 8025–8032.
- [14] C. Fox, M. Evans, M. Pearson, and T. Prescott, "Tactile slam with a biomimetic whiskered robot," in *2012 IEEE International Conference on Robotics and Automation*. IEEE, 2012, pp. 4925–4930.
- [15] S. Dragiev, M. Toussaint, and M. Gienger, "Uncertainty aware grasping and tactile exploration," in *2013 IEEE International conference on robotics and automation*. IEEE, 2013, pp. 113–119.
- [16] T. Schneider, C. de Farias, R. Calandra, L. Chen, and J. Peters, "Apple: Toward general active perception via reinforcement learning," in *The Fourteenth International Conference on Learning Representations*.
- [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 *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2016, pp. 4925–4930.
- [18] G. KISSOUM and V. PERDEREAU, "Simultaneous tactile localization and reconstruction of an object during robotic manipulation," in *2021 20th International Conference on Advanced Robotics (ICAR)*, 2021, pp. 948–954.
- [19] S. Fleer, A. Moringen, R. L. Klatzky, and H. Ritter, "Learning efficient haptic shape exploration with a rigid tactile sensor array," *PloS one*, vol. 15, no. 1, p. e0226880, 2020.
- [20] P. K. Murali, B. Porr, and M. Kaboli, "Touch if it's transparent! actor: Active tactile-based category-level transparent object reconstruction," in *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2023, pp. 10 792–10 799.
- [21] P. Mirowski, R. Pascanu, F. Viola, H. Soyer, A. J. Ballard, A. Banino, M. Denil, R. Goroshin, L. Sifre, K. Kavukcuoglu *et al.*, "Learning to navigate in complex environments," *arXiv preprint arXiv:1611.03673*, 2016.
- [22] D. Ramani, "A short survey on memory based reinforcement learning (2019)," *arXiv preprint arXiv:1904.06736*, 1904.
- [23] K.-W. Lee, Y. Qin, X. Wang, and S.-C. Lim, "Dextouch: Learning to seek and manipulate objects with tactile dexterity," *IEEE Robotics and Automation Letters*, vol. 9, no. 12, pp. 10 772–10 779, 2024.
- [24] J. A. Placed and J. A. Castellanos, "A deep reinforcement learning approach for active slam," *Applied Sciences*, vol. 10, no. 23, p. 8386, 2020.
- [25] H. Zhu, A. Gupta, A. Rajeswaran, S. Levine, and V. Kumar, "Dexterous manipulation with deep reinforcement learning: Efficient, general, and low-cost," in *2019 International Conference on Robotics and Automation (ICRA)*, 2019, pp. 3651–3657.
- [26] V. G. Goecks, G. M. Gremillion, V. J. Lawhern, J. Valasek, and N. R. Waytowich, "Integrating behavior cloning and reinforcement learning for improved performance in dense and sparse reward environments," in *Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems*, ser. AAMAS '20. Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems, 2020, p. 465–473.
- [27] J. Xu, S. Kim, T. Chen, A. R. Garcia, P. Agrawal, W. Matusik, and S. Sueda, "Efficient tactile simulation with differentiability for robotic manipulation," in *Conference on Robot Learning*. PMLR, 2023, pp. 1488–1498.
- [28] B. Fang, S. Jia, D. Guo, M. Xu, S. Wen, and F. Sun, "Survey of imitation learning for robotic manipulation," *International Journal of Intelligent Robotics and Applications*, vol. 3, no. 4, pp. 362–369, 2019.
- [29] J. Choi, H. Kim, Y. Son, C.-W. Park, and J. H. Park, "Robotic behavioral cloning through task building," in *2020 International Conference on Information and Communication Technology Convergence (ICTC)*, 2020, pp. 1279–1281.
- [30] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," *arXiv preprint arXiv:1707.06347*, 2017.
- [31] X. Wu, D. DeTone, D. Frost, T. Shen, C. Xie, N. Yang, J. Engel, R. Newcombe, H. Zhao, and J. Straub, "Sonata: Self-supervised learning of reliable point representations," in *Proceedings of the Computer Vision and Pattern Recognition Conference*, 2025, pp. 22 193–22 204.
- [32] I. Romanelis, V. Fotis, A. Kalogeras, C. Alexakos, K. Moustakas, and A. Munteanu, "Efficient and scalable point cloud generation with sparse point-voxel diffusion models," *arXiv preprint arXiv:2408.06145*, 2024.
- [33] D. Wang, D. Tseng, P. Li, Y. Jiang, M. Guo, M. Danielczuk, J. Mahler, J. Ichnowski, and K. Goldberg, "Adversarial grasp objects," in *2019 IEEE 15th International Conference on Automation Science and Engineering (CASE)*. IEEE, 2019, pp. 241–248.