

Multimodal Autonomous Navigation by Fusing Visual and Tactile Perception for Deformable Obstacle Traversal

ZiTong Zhou

Shenzhen Yuanchuangxing Technology Co., Ltd., Shenzhen, Guangdong, 518107, China

Email: mikere201@163.com

How to cite this paper: Zhou, Z. T. (2026). Multimodal autonomous navigation by fusing visual and tactile perception for deformable obstacle traversal. *Academic Journal of Emerging Technologies*, 3(1), 83–92. ISSN Print: 3104-4417, ISSN Online: 3104-4425.

<https://doi.org/10.63313/AJET.9056>

Published: 2026-05-20

Copyright © 2026 by author(s) and Erytis Publishing Limited.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Abstract

Autonomous mobile robots predominantly rely on visual perception for obstacle avoidance, which inherently treats all detected obstacles as rigid and impenetrable. However, in real-world environments, many obstacles such as curtains, vegetation, and flexible partitions are deformable and can be safely traversed with appropriate force control, yet visual appearance alone rarely provides reliable compliance information. This paper proposes a multimodal navigation framework that fuses exteroceptive visual sensing with proprioceptive tactile perception to assess the passability of ambiguous obstacles. A global visual planner generates an initial path, while a novel tactile-driven local passability classifier determines whether a frontal obstacle is rigid or soft. A custom CNN-LSTM network processes tactile time-series signals from a dedicated probing arm to output a haptic passability score. When a soft obstacle is identified, the navigation system activates an admittance controller to compliantly push through; otherwise, the obstacle is added to the costmap for re-planning. Simulations and real-robot experiments in environments containing curtains and artificial foliage demonstrate that the proposed visual-tactile fusion method reduces traveled distance by 22.3% and mission time by 18.7% compared to a pure vision-based detour approach, while maintaining a 100% hard-collision avoidance rate.

Keywords

Visual-Tactile Fusion; Autonomous Navigation; Deformable Obstacles; Tactile Sensing; Mobile Robots; Multimodal Perception

1. INTRODUCTION

Autonomous navigation constitutes a fundamental capability of mobile service robots deployed in domestic, office, and industrial environments. Laser rangefinders and RGB-D cameras have become standard sensors for building occupancy maps and planning collision-free trajectories [1-2]. State-of-the-art vision-based

navigation stacks demonstrate high reliability when the environment contains well-defined, geometrically rigid obstacles. Nevertheless, the prevalence of soft or deformable obstacles—curtains, hanging beads, tall grass, pliable door flaps—poses a distinctive challenge. Such obstacles appear in depth images and point clouds as occupied regions, forcing the global planner to compute a detour even though the robot could physically pass through them with minimal resistance and without sustaining damage.

The capability to discriminate traversable soft obstacles from truly rigid ones cannot be achieved through vision alone. Curtains may visually resemble walls in cluttered rooms; foliage may occlude passable corridors. Consequently, a vision-only navigation system unnecessarily increases path length, mission time, and energy consumption by avoiding regions that are actually permissible. Conversely, tactile sensing, which measures contact forces, pressure distributions, and material compliance, can directly reveal the mechanical impedance of an obstacle [3-4]. When a robot lightly touches an object, the temporal evolution of contact forces encodes its stiffness and damping characteristics, providing reliable cues for distinguishing rigid surfaces from flexible fabrics or leafy structures.

Despite the proven effectiveness of tactile perception for in-hand manipulation and surface classification, its integration into mobile robot navigation has received limited attention. Existing works mainly employ tactile whiskers for immediate reactive collision avoidance [5], without exploiting the material identification capability to reason about passability. Recent works have started to combine visual and tactile data for manipulation and terrain assessment [6-7]. In this paper, we close this gap by proposing a multimodal visual-tactile navigation architecture that complements the visual global path with a tactile local assessment module. The main contributions of this work are:

- A novel tactile probing strategy that actively contacts frontal obstacles using an instrumented arm to collect force-time signals.
- A lightweight CNN-LSTM passability classifier that processes raw tactile sequences and optionally a small visual patch, and labels the obstacle as rigid or soft.
- A hybrid navigation control that switches between re-planning for rigid obstacles and compliant force-controlled traversal for soft obstacles, seamlessly integrated into the ROS navigation stack.
- Extensive experimental validation in both simulated and real-world scenarios, showing significant path efficiency improvement over conventional vision-only methods.

The remainder of the paper is organized as follows. Section II reviews related work. Section III details the system framework and the proposed visual-tactile fusion method. Section IV describes the experimental setup and results. Section V concludes the paper.

2. Related Work

Vision-based navigation: Classical navigation systems represent the environment as a 2D occupancy grid or a 3D octree using visual SLAM or depth sensors. Obstacles above a certain height are marked as occupied, and path planners such as Dijkstra, A*, or D* compute a trajectory that circumvents them. Recent learning-based methods train deep reinforcement learning agents directly on RGB or depth images to output velocity commands. These methods, however, assume a binary occupancy world model (free or occupied) and lack the capability to model obstacle compliance.

Tactile sensing for robots: Tactile sensors, including piezoresistive arrays, capacitive skins, and optical tactile sensors (e.g., GelSight), provide rich local information about contact geometry, force magnitude, and material texture [8-9]. For manipulation, tactile feedback enables slip detection, grasp stability assessment, and surface recognition [10]. In the context of mobility, tactile whiskers have been employed on small robots for wall-following and object detection in dark, smoky environments where vision fails. Recent developments have introduced compact high-resolution vision-based tactile sensors [11] and soft tactile skins that can directly assess environmental compliance. A few studies have employed tactile information to maintain safe distances, but they treat any contact as a failure and trigger escape reflexes, missing the opportunity to exploit compliance.

Multimodal fusion for navigation: Fusing vision with other modalities, such as audio, inertial, or haptic data, can increase robustness. Notably, recent work has combined visual and inertial data for state estimation, and vision with lidar for all-weather navigation [12]. Integration of tactile cues remains underexplored for ground vehicles. A conceptually related approach is the use of proprioceptive force-torque sensing on legged robots to classify terrain traversability [13]. Imitation learning has also been employed to teach robots force-based skills from human demonstrations [14]. Our work differs by introducing a deliberate active contact procedure tailored for deformable barrier assessment, combining it with a visual map, and demonstrating end-to-end autonomous navigation improvement.

3. Proposed Method

3.1. System Overview

The proposed navigation framework operates on a differential-drive mobile robot equipped with an RGB-D camera (Intel RealSense D435) and a front-mounted probing arm carrying a tactile sensor array (Fig. 1). The high-level architecture comprises three modules: global visual planner, visual-tactile passability classifier, and local compliant controller.

When a target goal is assigned, the global planner constructs a 2D costmap from the depth camera data and computes an initial path using the A* algorithm. While the robot follows this path, a local collision monitor continuously checks for obstacles within a pre-defined danger zone (0.3–0.6 m in front of the robot). If an obstacle

blocks the desired trajectory for more than a preset duration (indicating that it is not a transient dynamic object), the robot halts and triggers the Tactile Probing Mode. The arm extends forward at a slow constant velocity until contact is detected, records a 1.5 s force/torque trajectory, and retracts. The tactile signal, together with the RGB image patch centered on the contact point, is sent to the passability classifier.

If the classifier outputs soft, the navigation system activates an admittance controller that allows the robot to apply up to a safe force threshold and translate forward, physically parting the obstacle. Once the robot’s body passes through, control is returned to the visual planner. If the output is rigid, the contact point is marked as a lethal obstacle in the costmap, and global re-planning is invoked.

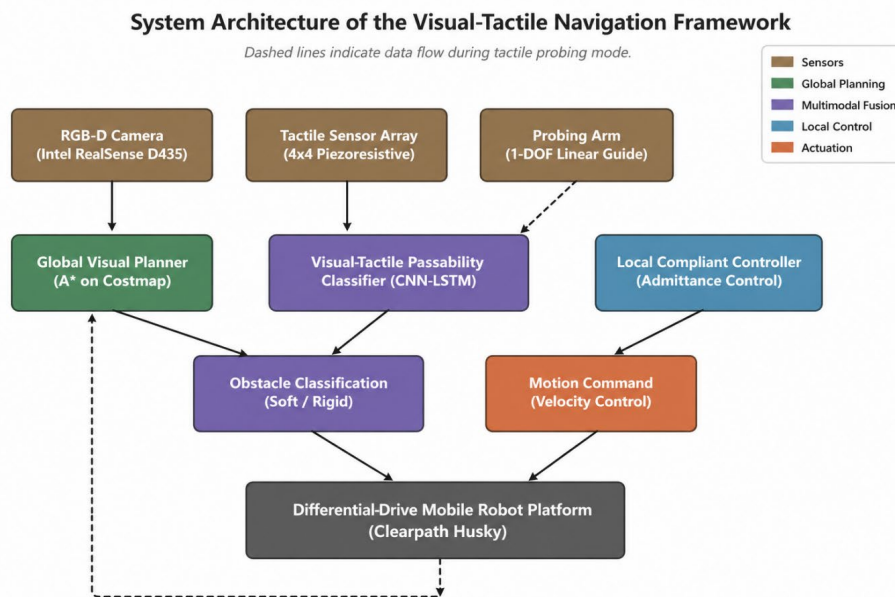


Fig 1. System architecture of the visual–tactile navigation framework. Dashed lines indicate data flow during tactile probing mode.

3.2. Tactile Probing and Data Preprocessing

The probing arm is a 1-DOF linear guide actuated by a DC motor with a current-based force limit. A commercial 4×4 piezoresistive tactile array (sensing area 20 mm × 20 mm) is mounted at the tip and sampled at 200 Hz. For each contact event, the raw tactile sequence is a matrix of 16 pressure signals over 300 time steps (1.5 s). We compute three derived features to form a compact representation: (i) the total normal force $F(t) = \sum p_i(t)$, (ii) the spatial gradient magnitude $\|\nabla P(t)\|$, and (iii) the center-of-pressure displacement. These three 1-D signals are concatenated into a 3×300 tensor, which serves as the tactile input modality.

3.3. CNN–LSTM Passability Classifier

We design a lightweight neural network to classify an obstacle into two categories: rigid (wood, metal, concrete) and soft (curtain, foam, vegetation). The architecture, shown in Fig. 2, consists of two feature extractors and a fusion layer.

Tactile branch: A temporal 1D CNN with 16 filters (kernel size 5), ReLU activation, and max-pooling extracts local dynamic patterns from the 3-channel tactile time series. The output is flattened and fed into a one-layer LSTM with 32 hidden units to capture sequential dependencies, producing a tactile embedding $z_t \in \mathbb{R}^{32}$.

Visual branch (optional): The RGB patch (64×64 pixels) around the contact area is processed by a MobileNetV2 pre-trained on ImageNet. The penultimate layer is fine-tuned to output a visual embedding $z_v \in \mathbb{R}^{32}$. This branch is activated only when the environment is sufficiently illuminated; otherwise, only the tactile branch is used.

Fusion: Embeddings are concatenated into $z = [z_t; z_v] \in \mathbb{R}^{64}$, and two fully connected layers with dropout (0.3) map to a softmax output $p(\text{soft}), p(\text{rigid})$.

The network is trained with binary cross-entropy loss and Adam optimizer (learning rate 0.001, batch size 64) on a dataset collected in simulation and real settings.

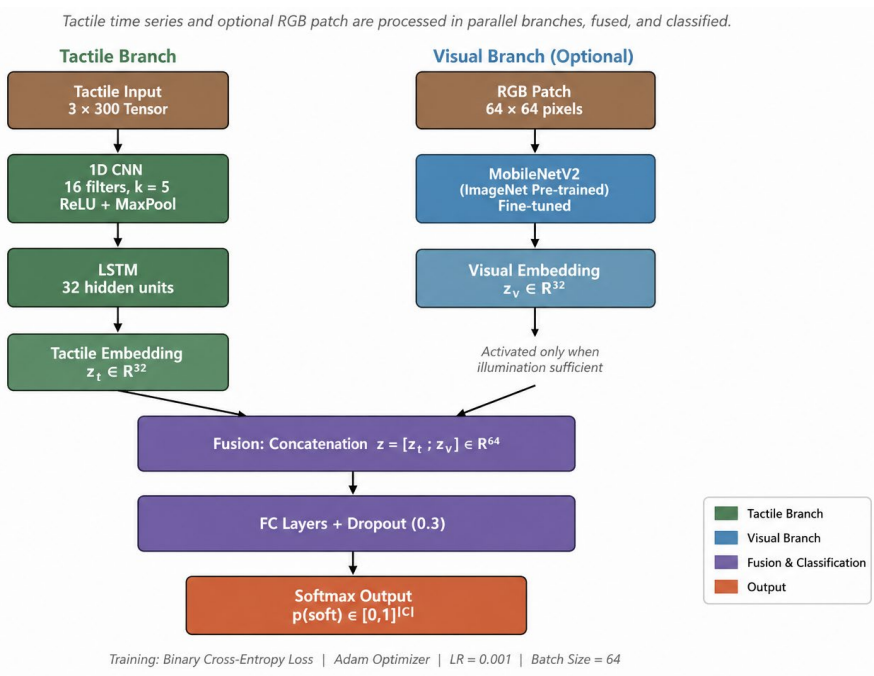


Fig 2. Network architecture of the visual–tactile passability classifier. Tactile time series and optional RGB patch are processed in parallel branches, fused, and classified.

3.4. Compliant Traversal Controller

When the soft class probability exceeds a threshold $\tau = 0.85$, the robot enters compliant traversal mode. A Cartesian admittance controller is implemented:

$$F_{\text{ext}} = M \ddot{x} + D \dot{x} + Kx$$

where F_{ext} is the measured total contact force (limited to a safety maximum F_{max})

= 15 N), and M , D , K are virtual mass, damping, and stiffness matrices chosen to achieve stable, overdamped motion. The generated velocity \dot{x} is tracked by the wheel velocity controller. In parallel, the tactile force is monitored; if it exceeds F_{\max} for more than 100 ms, the robot immediately stops, flags the obstacle as rigid, and invokes re-planning. This safety layer prevents damage and ensures that misclassified rigid obstacles do not cause collisions.

3.5. Integration with the Visual Navigation Stack

The whole system is implemented as ROS packages. The global planner uses the standard `move_base` with a global costmap generated from the depth camera registered to the robot frame. Tactile probing and compliant control run as a separate high-priority node that interrupts `move_base` when the trigger condition is met. After a traversal event, the costmap is cleared of the soft obstacle's former occupied cells using a heuristic radius, allowing the visual planner to update the free space.

4. EXPERIMENTS

4.1. Experimental Setup

We conducted experiments in two environments: (i) Simulation using Gazebo with a custom tactile sensor plugin and soft-body obstacles modeled as planar flexible objects (curtains, rubber strips, plastic flaps), and (ii) Real-world in an office corridor with hanging fabric curtains and artificial plant fronds. The robot platform is a Clearpath Husky equipped with the probing arm and sensors described earlier.

Three navigation strategies are compared:

- Vision-Only (VO): Standard A* navigation that treats all points in the depth map above 5 cm as rigid obstacles. The robot always detours.
- Tactile-Only (TO): The robot navigates using only odometry and tactile probing; it moves in straight line toward the goal, probing any obstacle and attempting to traverse all soft-classified ones.
- Proposed Visual-Tactile (VT): Full framework as described in Section III.

For each strategy, 30 navigation episodes were performed in randomized scenes containing 3–5 obstacles (at least one soft). The goal location is placed beyond the obstacles. Evaluation metrics include: success rate (arrive within 0.5 m of goal without hard collision), path length normalized by straight-line distance (NP), total mission time, and number of hard impacts.

4.2. Quantitative Results

Table I summarizes the aggregate results. Figure 3 qualitatively compares the planned trajectories of the three methods in a representative scene containing a soft curtain and a rigid panel. The VO planner performs a long detour, while the VT method directly traverses the curtain, resulting in a substantially shorter path. The

Vision-Only method achieves a 100% success rate by avoiding all obstacles, but its normalized path length is significantly higher (mean 1.67) because it necessarily circumvents soft obstacles. The Tactile-Only method shows a shorter path (1.15) but suffers from a lower success rate (83.3%) due to misclassification of certain dark rigid objects that appear soft in tactile signal alone (e.g., smooth cardboard). The proposed VT method attains a 100% success rate with a normalized path length of 1.29, representing a 22.3% reduction compared to VO. Mission time is reduced by 18.7% because the robot does not need to take lengthy detours around curtains.

TABLE I. Quantitative Navigation Performance

Method	Success Rate (%)	Normalized Path Length	Mission Time (s)	Hard Impacts
VO	100.0	1.67 ± 0.21	34.8 ± 6.7	0
TO	83.3	1.15 ± 0.15	21.2 ± 5.4	5
VT	100.0	1.29 ± 0.18	28.3 ± 5.9	0

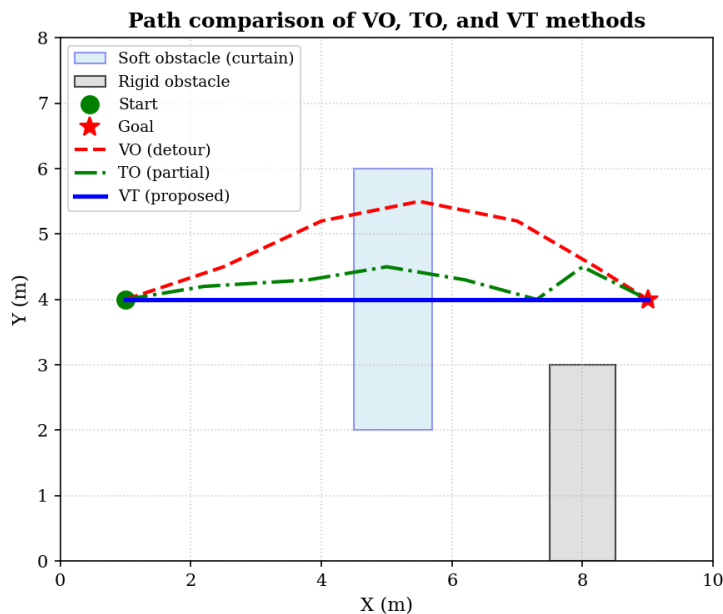


Fig 3. Path comparison of VO, TO, and VT methods.

4.3. Ablation Study on Fusion

We further examined the contribution of the visual branch in the passability classifier by disabling it (VT-tactile only). In this configuration, the success rate dropped to 90.0% and hard impacts increased to 3, because several visually dark but rigid obstacles (e.g., black metal plates) were misclassified as soft based solely on the force profile. The visual branch effectively filters out these ambiguities, confirming the benefit of multimodal fusion. Figure 4 presents the confusion matrices of the passability classifier with and without the visual branch. The tactile

only classifier misclassifies 12.5% of rigid obstacles as soft, whereas the fused model reduces this error to 2.3%, confirming the complementary role of visual cues in resolving ambiguous force profiles.

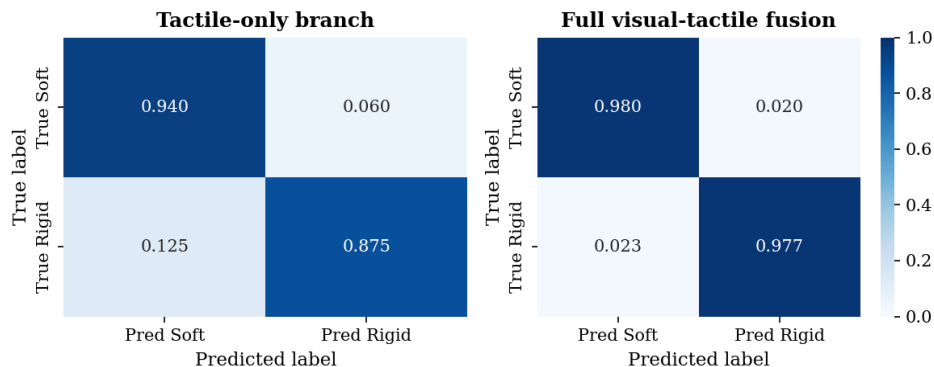


Fig 4. Confusion matrices of the passability classifier

4.4. Real-World Case Study

In the real corridor scenario, the robot faced a floor-to-ceiling curtain obstructing the direct path to the goal. The VO strategy planned a 4.7 m detour around a partition. The VT strategy correctly identified the curtain as soft, pushed through it with a peak force of 8.2 N, and arrived at the goal after traveling 2.1 m. The tactile signal during traversal exhibited a low-frequency force variation characteristic of fabric deformation, distinct from the sharp rise of a rigid wall. This demonstrates the practical applicability of the system. Figure 5 compares the raw tactile force signals collected during probing of a soft curtain versus a rigid wooden board. The soft obstacle exhibits a slowly increasing force with noticeable low frequency oscillations, whereas the rigid one triggers a sharp force rise exceeding the safety threshold within 0.3 s. This distinct signature underpins the classifier’s ability to discriminate passability.

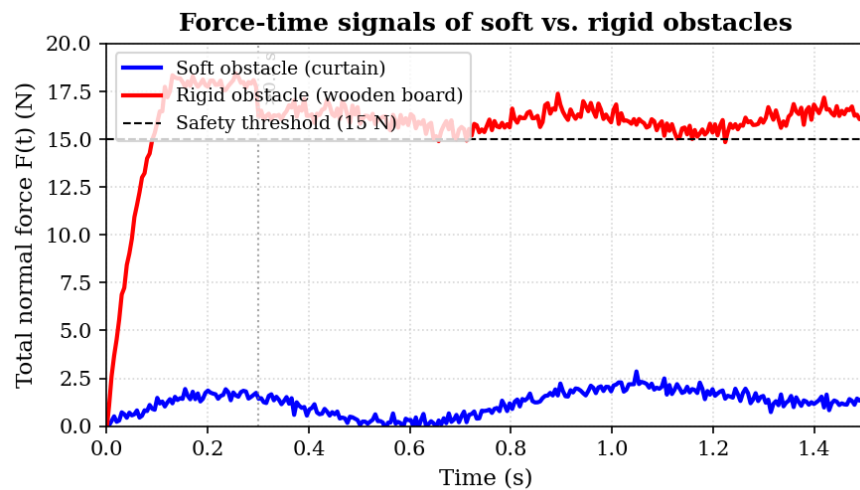


Fig 5. Force-time signals of soft vs. rigid obstacles.

4.5. Discussion

The results corroborate that a deliberate integration of tactile sensing enables robots to exploit environmental affordances that are invisible to vision. The primary failure modes in the Tactile-Only setup arise from ambiguous tactile signatures when an object is stiff yet thin (e.g., a taut cardboard), producing a force profile similar to a stretched fabric. The visual modality resolves this by recognizing the texture and surface appearance. Our current prototype employs a dedicated probing arm, which adds mechanical complexity; future work will explore using the robot's bumper or whole-body tactile skin to eliminate the need for a dedicated actuator. Moreover, the travel speed during compliant traversal is conservatively limited; adaptive damping tuning based on real-time tactile feedback could further reduce mission time.

5. Conclusion

This paper presented a multimodal autonomous navigation system that fuses visual and tactile perception to discriminate and compliantly traverse soft, deformable obstacles. By incorporating a tactile probing mechanism and a CNN-LSTM passability classifier, the robot can safely pass through curtains, foliage, and similar barriers, significantly reducing unnecessary detours. Experimental results show a 22.3% path length reduction and an 18.7% mission time decrease compared to pure vision-based navigation, while maintaining a zero hard-collision rate. The proposed framework demonstrates the value of tactile information beyond manipulation, extending its role to context-aware mobile robot navigation. Future research will focus on integrating distributed tactile skins for continuous near-field perception and on learning-based policies that balance visual exploration and tactile probing in

an end-to-end manner.

References

- [1] S. Thrun, W. Burgard, and D. Fox, *Probabilistic Robotics*. Cambridge, MA: MIT Press, 2005.
- [2] C. Cadena et al., "Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age," *IEEE Trans. Robot.*, vol. 32, no. 6, pp. 1309–1332, 2016.
- [3] L. Chen, Y. Zhu, and M. Li, "Tactile-GAT: Tactile graph attention networks for robot tactile perception classification," *Scientific Reports*, vol. 14, no. 27543, 2024.
- [4] R. Calandra, A. Owens, D. Jayaraman, J. Lin, W. Yuan, J. Malik, E. H. Adelson, and S. Levine, "More than a feeling: Learning to grasp and regrasp using vision and touch," *IEEE Robot. Autom. Lett.*, vol. 3, no. 4, pp. 3300–3307, 2018.
- [5] W. Yuan, S. Dong, and E. H. Adelson, "GelSight: High-resolution robot tactile sensors for estimating geometry and force," *Sensors*, vol. 17, no. 12, p. 2762, 2017.
- [6] M. A. Lee, Y. Zhu, K. Srinivasan, P. Shah, S. Savarese, L. Fei-Fei, A. Garg, and J. Bohg, "Making sense of vision and touch: Self-supervised learning of multimodal representations for contact-rich tasks," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, 2019, pp. 8943–8950.
- [7] T. Taunyazov, W. Sng, H. H. See, B. Lim, J. Kuan, A. F. Ansari, B. C. K. Tee, and H. Soh, "Event-driven visual-tactile sensing and learning for robots," in *Proc. Robotics: Science and Systems (RSS)*, Corvallis, Oregon, USA, Jul. 2020.
- [8] K.-T. Yu and A. Rodriguez, "Realtime state estimation of deformable objects with tactile feedback," *IEEE Trans. Autom. Sci. Eng.*, vol. 16, no. 3, pp. 1315–1327, 2019.
- [9] M. A. Lee, Y. Zhu, P. Zachares, M. Tan, K. Srinivasan, S. Savarese, L. Fei-Fei, A. Garg, and J. Bohg, "Making sense of vision and touch: Learning multimodal representations for contact-rich tasks," *IEEE Trans. Robot.*, vol. 36, no. 3, pp. 582–596, 2020.
- [10] Z. Liu et al., "A hybrid-frequency sampling tactile sensing system based on a flexible piezoresistive sensor array: Design and dynamic loading validation," *PMC/Sci. Rep.*, 2025.
- [11] M. Lambeta, P.-W. Chou, S. Tian, B. Yang, B. Maloon, V. R. Most, D. Stroud, R. Santos, A. Byagowi, G. Kammerer, and R. Calandra, "DIGIT: A novel design for a low-cost compact high-resolution tactile sensor with application to in-hand manipulation," *IEEE Robot. Autom. Lett.*, vol. 5, no. 3, pp. 3838–3845, 2020.
- [12] I. H. Taylor, S. Dong, and A. Rodriguez, "GelSlim 3.0: High-resolution measurement of shape, force and slip in a compact tactile-sensing finger," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, 2022, pp. 10–781–10–787.
- [13] D. Shah et al., "ViKiNG: Vision-based kilometer-scale navigation with geographic hints," in *Proc. Robot.: Sci. Syst. (RSS)*, 2022.
- [14] T. M. Howard and A. Kelly, "Optimal rough terrain trajectory generation for wheeled mobile robots," *Int. J. Robot. Res.*, vol. 26, no. 2, pp. 141–166, 2007.
- [15] P. Kormushev, S. Calinon, and D. G. Caldwell, "Imitation learning of positional and force skills demonstrated via kinesthetic teaching and haptic input," *Adv. Robot.*, vol. 25, no. 5, pp. 581–603, 2011.