Embodied Tactile Intelligence: From Adaptive Manipulation to Knowledge Accumulation

Zihang Zhao
Institute for Artificial Intelligence, Peking University
Email: zhaozihang@stu.pku.edu.cn

To operate robustly in unstructured environments, robots must evolve beyond static, pre-programmed behaviors and acquire a grounded physical understanding of environmental dynamics [1, 10], defined here as the laws governing how object states evolve in response to actions [6]. While vision provides the necessary global context, it can be limited in resolving local contact properties, due to occlusion and visual ambiguity—systems with substantially different dynamics may have identical visual appearance [7, 25, 23].

My research posits that tactile sensing is the fundamental component of the state observer required for accurate system identification in open worlds. I aim to develop a framework that treats physical interaction and sensory as a rich information source of environmental dynamics rather than a mere actuation result. By leveraging high-resolution touch to monitor action effects, robots can refine dynamic models in real-time, allowing for the immediate adjustment of controllers optimized for the current interaction scenario.

My PhD work systematically addresses this challenge through a bottom-up methodology, starting with the development of **high-resolution full-hand tactile embodiment** to enable high-fidelity distributed perception [22]. Building on the hardware foundation, I derived a tactile controller capable of regulating contact mechanics to robustly manipulate articulated objects **without relying on object kinematics priors** [23]. This progression culminated in the integration of global visual context with proactive tactile control [24, 3], enhancing both algorithmic efficiency and pipeline automation. My ongoing and future work aims to close the loop on this architecture: by leveraging these robust tactile interactions, I propose to accumulate structural information that self-supervises high-level dynamics prediction systems, mirroring the developmental trajectory where embodied sensory calibrates cognition [11, 20].

A. High-Resolution Full-Hand Tactile Embodiment

To rigorously validate the role of touch, I first established a hardware foundation capable of capturing high-fidelity contact dynamics. Conventional robotic hands suffer from a trade-off: they either prioritize mechanical dexterity at the expense of sensing [4, 17, 13] or rely on sparse, fingertip-restricted sensors [5, 19, 16, 9, 14]. In response, I developed a biomimetic system embedding high-resolution tactile sensing **(0.1 mm spatial resolution) across 70 % of its inner surface area** [22], as shown in Figure 1(a).

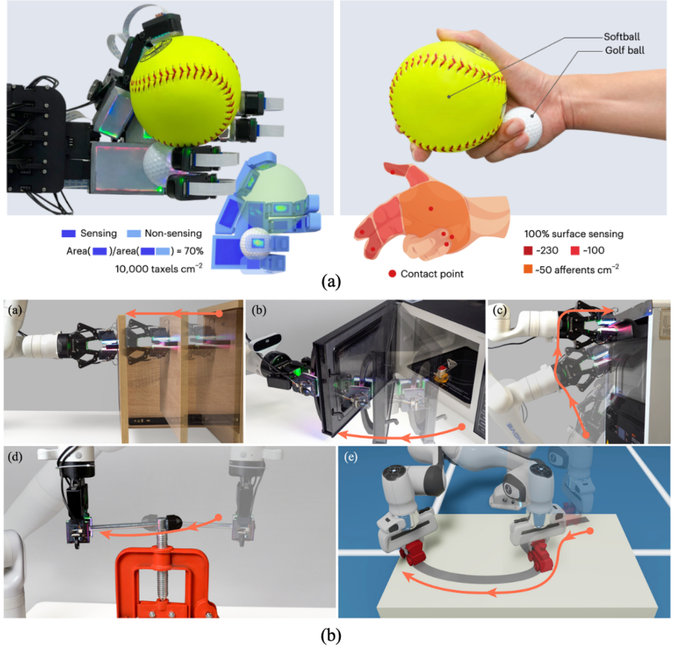

Fig. 1. **Enhancing robotic interaction capabilities with tactile embodiment.** (a) Full-hand high-resolution tactile sensing for adaptive **multi-object grasping**. (b) Tactile-informed manipulation of articulated objects without requiring object kinematics priors, **extending beyond pure translation and rotation to include helical and randomly-generated articulations**.

By seamlessly embedding 17 custom compact vision-based sensors [19] into a layout that preserves the hand's full capacity of grasping, this work demonstrated that dense tactile feedback is a necessity, not an add-on, for adaptive behavior. To leverage this capability, I developed a generative algorithm that synthesizes human-like hand configurations, enabling the system to perform stable, adaptive grasping of multiple objects simultaneously, a task where vision alone frequently fails due to occlusion [1]. This work confirmed that comprehensive tactile embodiment unlocks manipulation capabilities that vision or proprioception cannot support, laying the hardware groundwork for my subsequent research on dynamic interaction.

B. Prior-Free Manipulation via Tactile Regulation

Building on this hardware foundation, I shifted focus to the more rigorous challenge of articulated object manipulation. While free-object grasping tolerates 6-DoF errors, articulated objects impose strict kinematic constraints, confining motion to specific, often unknown, manifolds. Traditionally, robots rely on prior kinematic knowledge to interact with these objects efficiently. This knowledge is typically embedded either explicitly—using kinematic models for control and planning [2, 12]—or implicitly within learned policy models [18]. However, these strategies often prove brittle when facing the novel scenarios inevitably encountered in open-world settings.

I reformulated this challenge by prioritizing local interaction and proposed a framework for tactile-informed, prior-free manipulation [23]. Rather than relying on a pre-computed kinematic model, the robot uses real-time tactile feedback to regulate contact stability. By treating contact deviation as the divergence between the right interaction direction and the current one, the controller dynamically adjusts the end-effector to minimize the deviation, complying with the object's natural trajectory to complete the manipulation. Triggered by a rough interaction direction, this 'act-and-regulate' paradigm eliminates the need for prior kinematic knowledge, **enabling robust manipulation of prismatic, revolute, and even complex helical mechanisms** (such as a vise), as shown in Figure 1(b). Leveraging rigorous initialization methods [21], I demonstrate significant improvements over traditional approaches in both real-world experiments and large-scale simulations.

C. Efficiency and Visuo-Tactile Synergy

While purely reactive tactile regulation guarantees stability, its exclusive reliance on instantaneous error feedback inherently bottlenecks execution speed. To bridge this gap between robust safety and operational efficiency, I developed an efficient version [24]. This controller transitions from passive adjustment to proactive control by leveraging short-horizon historical tactile data to regress local kinematic properties—specifically the instantaneous velocity twist—on the fly. By explicitly modeling the object's motion constraints, the controller anticipates the trajectory rather than waiting for contact errors to manifest. This predictive capability accelerates manipulation speeds by over 10×—**reducing task times from minutes to seconds**—while maintaining a $100\,\%$ success rate across complex prismatic, revolute, and helical joints.

However, even a high-speed tactile controller is limited without effective initialization. Recognizing that vision is indispensable for the global context, I train a vision-based neural network [3] to resolve the cold-start problem. This framework establishes a complementary synergy: Vision provides a coarse initialization, proposing interaction directions, while the tactile controller handles the local precision required for actuation. This hierarchical structure, global visual guide coupled with local tactile verification, enables the system to achieve **fully autonomous manipulation in unstructured real-world environments**.

D. Ongoing Work: Self-Supervised Learning from Interaction

My current research tries to redefine the relationship between low-level control and high-level reasoning, viewing embodiment not merely as an execution platform, but as an active mechanism for grounding artificial intelligence in the physical world. While large-scale foundation models possess vast semantic knowledge, they typically remain frozen and prone to hallucinating physical dynamics. The robust interaction capabilities developed in my previous work now serve as a critical, real-time data generator to bridge this gap.

I am currently exploring how this accumulated interaction data can steer these static models toward physical reality. In this paradigm, the robot enters a new environment and cautiously probes unknown objects using tactile regulation. These interactions reveal 'tactile residuals'—the high-dimensional discrepancies between the model's visual predictions (imagination) and the actual sensed contact dynamics (reality).

By treating these residuals as an error signal, the system minimizes the divergence, effectively updating the model's latent physical parameters to resolve instance-specific constraints. This process mimics the neuroplasticity of human learning: a continuous dialogue where distinct phases of cautious exploration transform into efficient, prior-based interaction, ensuring that internal representations remain tightly coupled with an evolving physical world.

E. Future Research: General-Purpose Embodied Intelligence

My long-term research agenda focuses on scaling up tactile-driven robotics to achieve more general-purpose autonomy. I propose a three-pillar framework to bridge the gap between raw sensory signals and high-level reasoning:

Phase 1: High-Fidelity Embodiment: To support this vision, the physical vessel must match the complexity of the data required. I plan to develop a next-generation dexterous hand that combines the full degrees of freedom of human-level systems (*e.g.*, the Shadow Hand [13]) with a continuous tactile skin. Unlike current setups that suffer from blind spots (Figure 1(a)), this comprehensive coverage will ensure that no contact dynamic—whether on the palm, fingertip, or side of the finger—is lost during complex in-hand manipulation, providing a dense, unbroken stream of physical ground truth.

Phase 2: Data-Driven Policy Initialization: With the hardware in place, I would like to address the data scarcity problem by capturing human tactile-sensory-motor data. Using high-frequency tactile gloves [15, 8], I aim to record fine-grained manipulation data that captures not just motion, but the subtle force and contact. By mapping these human tactile signals to human motions, I will initialize robot policies with human-like sensorimotor primitives. This process provides the robot with a robust set of initial reflexes, allowing it to start learning from a competent baseline rather than from scratch.

Phase 3: Unsupervised Knowledge Distillation: Finally, I will investigate methods to distill this raw embodied experience into structured, generalizable knowledge. Moving beyond black-box policies, I aim to use physical interaction to construct explicit world models that encode dynamics. By continuously updating these internal models against real-world tactile feedback, the robots are expected to autonomously refine their understanding of the environment. This ultimately enables the emergence of an adaptable, autonomous agent capable of reasoning about novel objects and dynamics with a proficiency that rivals human physical intelligence.

REFERENCES

[1] Aude Billard and Danica Kragic. Trends and challenges in robot manipulation. *Science*, 364(6446):eaat8414, 2019.

[2] Sachin Chitta, Benjamin Cohen, and Maxim Likhachev. Planning for autonomous door opening with a mobile manipulator. In *IEEE International Conference on Robotics and Automation (ICRA)*, 2010.

[3] Leiyao Cui, Zihang Zhao, Sirui Xie, Wenhuan Zhang, Zhi Han, and Yixin Zhu. Vi-tacman: Articulated object manipulation via vision and touch. In *IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2026.

[4] Raphael Deimel and Oliver Brock. A novel type of compliant and underactuated robotic hand for dexterous grasping. *International Journal of Robotics Research (IJRR)*, 35(1-3):161–185, 2016.

[5] Jeremy A Fishel and Gerald E Loeb. Sensing tactile microvibrations with the biotac—comparison with human sensitivity. In *International Conference on Biomedical Robotics and Biomechatronics (BioRob)*, 2012.

[6] Bruce A Francis and Walter Murray Wonham. The internal model principle of control theory. *Automatica*, 12(5):457–465, 1976.

[7] Roland S Johansson and J Randall Flanagan. Coding and use of tactile signals from the fingertips in object manipulation tasks. *Nature Reviews Neuroscience*, 10 (5):345–359, 2009.

[8] Haoyang Luo, Zihang Zhao, Leiyao Cui, Saiyao Zhang, Liu Yang, Zhi Han, Xiyuan Tang, and Yixin Zhu. T-800: An 800 hz data glove for precise hand gesture tracking. *arXiv preprint arXiv:2603.26403*, 2026.

[9] Yiyue Luo, Yunzhu Li, Pratyusha Sharma, Wan Shou, Kui Wu, Michael Foshey, Beichen Li, Tomás Palacios, Antonio Torralba, and Wojciech Matusik. Learning human–environment interactions using conformal tactile textiles. *Nature Electronics*, 4(3):193–201, 2021.

[10] Aslan Miriyev and Mirko Kovač. Skills for physical artificial intelligence. *Nature Machine Intelligence*, 2 (11):658–660, 2020.

[11] Rajesh PN Rao and Dana H Ballard. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature neuroscience*, 2(1):79–87, 1999.

[12] John Schulman, Yan Duan, Jonathan Ho, Alex Lee, Ibrahim Awwal, Henry Bradlow, Jia Pan, Sachin Patil, Ken Goldberg, and Pieter Abbeel. Motion planning with sequential convex optimization and convex collision checking. *International Journal of Robotics Research (IJRR)*, 33(9):1251–1270, 2014.

[13] ShadowRobot. Shadow dexterous hand, 2026. URL https://www.shadowrobot.com/dexterous-hand-series/. Accessed: 2026.

[14] Huanbo Sun, Katherine J Kuchenbecker, and Georg Martius. A soft thumb-sized vision-based sensor with accurate all-round force perception. *Nature Machine Intelligence*, 4(2):135–145, 2022.

[15] Subramanian Sundaram, Petr Kellnhofer, Yunzhu Li, Jun-Yan Zhu, Antonio Torralba, and Wojciech Matusik. Learning the signatures of the human grasp using a scalable tactile glove. *Nature*, 569(7758):698–702, 2019.

[16] Benjamin Ward-Cherrier, Nicholas Pestell, Luke Cramphorn, Benjamin Winstone, Maria Elena Giannaccini, Jonathan Rossiter, and Nathan F Lepora. The tactip family: Soft optical tactile sensors with 3d-printed biomimetic morphologies. *Soft robotics*, 5(2):216–227, 2018.

[17] Cai-Hua Xiong, Wen-Rui Chen, Bai-Yang Sun, Ming-Jin Liu, Shi-Gang Yue, and Wen-Bin Chen. Design and implementation of an anthropomorphic hand for replicating human grasping functions. *IEEE Transactions on Robotics (T-RO)*, 32(3):652–671, 2016.

[18] Zhenjia Xu, Zhanpeng He, and Shuran Song. Universal manipulation policy network for articulated objects. *IEEE Robotics and Automation Letters (RA-L)*, 7(2): 2447–2454, 2022.

[19] Wenzhen Yuan, Siyuan Dong, and Edward H Adelson. Gelsight: High-resolution robot tactile sensors for estimating geometry and force. *Sensors*, 17(12):2762, 2017.

[20] Robert J Zatorre, R Douglas Fields, and Heidi Johansen-Berg. Plasticity in gray and white: neuroimaging changes in brain structure during learning. *Nature neuroscience*, 15(4):528–536, 2012.

[21] Zihang Zhao, Leiyao Cui, Sirui Xie, Saiyao Zhang, Zhi Han, Lecheng Ruan, and Yixin Zhu. B*: Efficient and optimal base placement for fixed-base manipulators. *IEEE Robotics and Automation Letters (RA-L)*, 2025.

[22] Zihang Zhao, Wanlin Li, Yuyang Li, Tengyu Liu, Boren Li, Meng Wang, Kai Du, Hangxin Liu, Yixin Zhu, Qining Wang, et al. Embedding high-resolution touch across robotic hands enables adaptive human-like grasping. *Nature Machine Intelligence*, 7(6):889–900, 2025.

[23] Zihang Zhao, Yuyang Li, Wanlin Li, Zhenghao Qi, Lecheng Ruan, Yixin Zhu, and Kaspar Althoefer. Tacman: Tactile-informed prior-free manipulation of articulated objects. *IEEE Transactions on Robotics (T-RO)*, 41: 538–557, 2025.

[24] Zihang Zhao, Zhenghao Qi, Yuyang Li, Leiyao Cui, Zhi Han, Lecheng Ruan, and Yixin Zhu. Tacman-turbo: Proactive tactile control for robust and efficient articulated object manipulation. *IEEE Transactions on Automation Science and Engineering (T-ASE)*, 23:8296–8317, 2026.

[25] Yixin Zhu, Tao Gao, Lifeng Fan, Siyuan Huang, Mark Edmonds, Hangxin Liu, Feng Gao, Chi Zhang, Siyuan Qi, Ying Nian Wu, et al. Dark, beyond deep: A paradigm shift to cognitive ai with humanlike common sense. *Engineering*, 6(3):310–345, 2020.