Robots Learning Through Physical Interactive Intelligence

Roberto Martín-Martín

Department of Computer Science
University of Texas at Austin, Austin, TX, USA
robertomm@cs.utexas.edu

Artificial Intelligence (AI) has advanced significantly in fields like computer vision (CV) and natural language processing (NLP), fueled by breakthroughs in machine learning. However, these strides have yet to fully extend to robotics, which faces unique challenges in long-horizon decision-making. Predicting and utilizing physical interactions to alter the environment remains complex for artificial agents. Consequently, robots minimize physical interaction to avoid risks in uncontrolled environments, relying heavily on computation for perception, planning, and control.

In contrast, humans demonstrate that autonomy and intelligent behavior arise from exploiting rich interactions with the environment, reducing reliance on precise computation. My research aims to balance interaction and computation in robot learning systems that reason, perceive, and plan within the interaction loop, emphasizing physical interaction as essential in unstructured environments. To that end, my work spans three complementary research directions:

Imitation Learning for Physical Interactive Intelligence: Humans excel at tasks in complex environments like homes and offices. Imitating human behavior is promising for enabling robots to interact with their surroundings. My work has introduced methods for robots to **learn from a single video** demonstration of bimanual tasks by projecting perception into the screw space of relative hand motion (Bahety et al. 2024). I also developed teleoperation interfaces to collect demonstrations tailored to a robot's embodiment, enabling human-in-the-loop learning algorithms (Mandlekar et al. 2020; Liu et al. 2024).

Physical Interactive Intelligence in Navigation: Navigation in human environments inherently involves physical interactions —pushing obstacles, opening doors, or pressing elevator buttons. However, conventional robotic navigation systems are restricted to search for collision-free paths between locations. In my research, I introduced a new AI concept, *Interactive Navigation*, navigation problems that require physical interaction to be solved, and several hierarchical reinforcement learning architectures that allow robots to address them (Li et al. 2020; Xia et al. 2021).

Perceiving Through Physical Interactions: Humans and other biological agents interact with the environment to obtain task-relevant information (Gibson 1979). Differently, robots are "afraid" of contacting the environment, strongly relying on passive modes of information gathering. This restricts their ability to acquire rich models of the environment. I equipped robots with the algorithms to **perceive** their environment **through interaction**, enabling tasks such as object search (Danielczuk et al. 2019) or understanding and manipulating articulated objects like doors, drawers, and scissors (Martín-Martín and Brock 2014, 2017), critical for robots operating in human spaces.

Ever Improving Physically Interacting Robots: Looking ahead, I aim to build on this work by developing **continual learning** algorithms that allow robots to incrementally acquire new skills through imitation and exploration, leveraging increasingly rich representations derived from physical interactions.

References

Bahety, A.; et al. 2024. ScrewMimic: Bimanual Imitation from Human Videos with Screw Space Projection. In *Robotics: Science and Systems (RSS)*. Best Student Paper Finalist.

Danielczuk, M.; et al. 2019. Mechanical search: Multi-step retrieval of a target object occluded by clutter. In *IEEE International Conference on Robotics and Automation (ICRA)*.

Gibson, J. J. 1979. *The ecological approach to visual perception: classic edition*. Psychology press.

Li, C.; et al. 2020. Hrl4in: Hierarchical reinforcement learning for interactive navigation with mobile manipulators. In *Conference on Robot Learning (CoRL)*, 603–616. PMLR.

Liu, H.; et al. 2024. Model-based runtime monitoring with interactive imitation learning. In *IEEE International Conference on Robotics and Automation (ICRA)*, 4154–4161. IEEE.

Mandlekar, A.; et al. 2020. Learning to generalize across long-horizon tasks from human demonstrations. In *Robotics: Science and Systems (RSS)*.

Martín-Martín, R.; and Brock, O. 2014. Online interactive perception of articulated objects with multi-level recursive estimation based on task-specific priors. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2494–2501. IEEE. Best Paper, Best Student Paper and Best Cognitive Paper Finalist.

Martín-Martín, R.; and Brock, O. 2017. Building Kinematic and Dynamic Models of Articulated Objects with Multi-Modal Interactive Perception. In *AAAI Spring Symposium Series*. AAAI Pubs.

Xia, F.; et al. 2021. ReLMoGen: Leveraging motion generation in reinforcement learning for mobile manipulation. In *IEEE International Conference on Robotics and Automation (ICRA)*.