Learning Adaptive Representations for Robust Mobile Robot Navigation from Multi-Modal Interactions

PIs: Matthew Walter (TTIC), Thomas Howard (UR)
Students: Siddharth Patki (UR), Ethan Fahnestock (UR), Andrea Daniele (TTIC), Charles Schaff (TTIC), David Yunis (TTIC), Michael Napoli (UR), and Harel Biggie (UR)

Motivation

Robots need models that provide detail relevant to specific tasks and environments.

Minimal but sufficient representations enable efficient mapping, planning and reasoning for mobile manipulation.

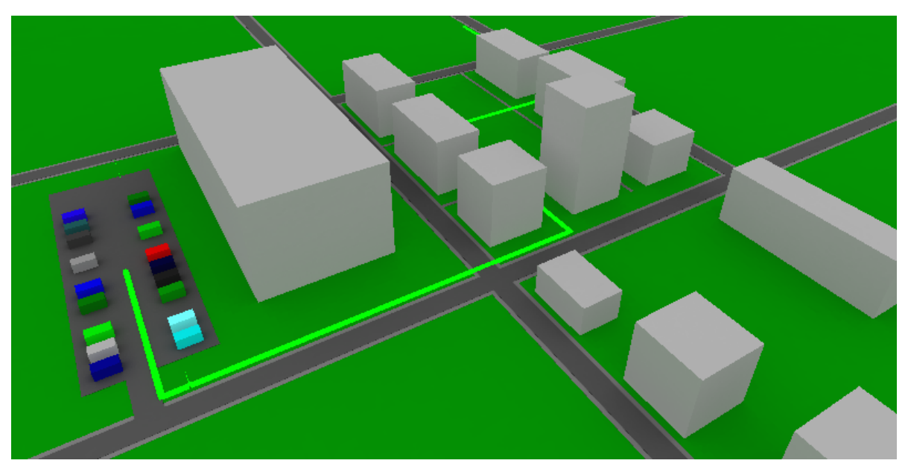
Designing minimal representations is task-specific, may vary over time, and may require input from multiple different sensing modalities.

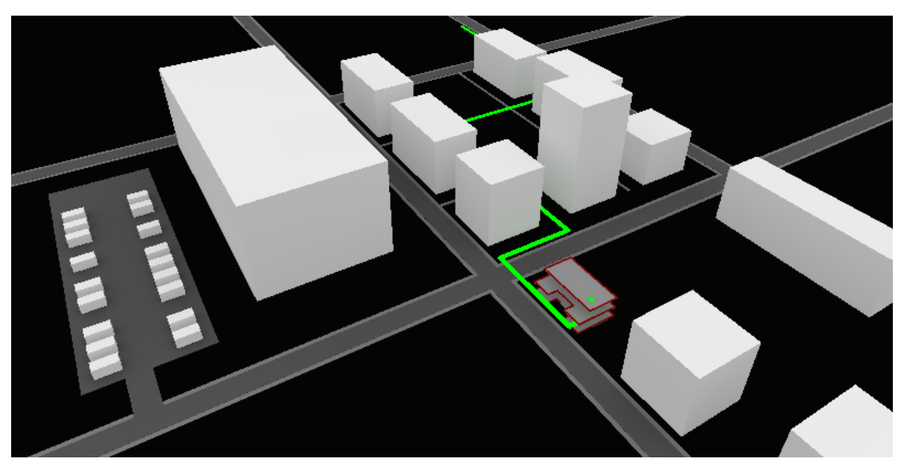
Objectives

Improve scalability and heterogeneity of robot behaviors and environments.

Learn models for adapting the representation of the environments for efficient robot motion planning.

Develop algorithms for extracting succinct representation from multi-modal interactions.





Learned Models for Adaptive Representations

Predictively Adaptive State Lattices

Developed predictive algorithm for optimizing local trajectory libraries to improve relative optimality of mobile robot navigation in complex environments (Fig. 1). Demonstrated effectiveness of learning optimized representations of graph connectivity for improving path search.

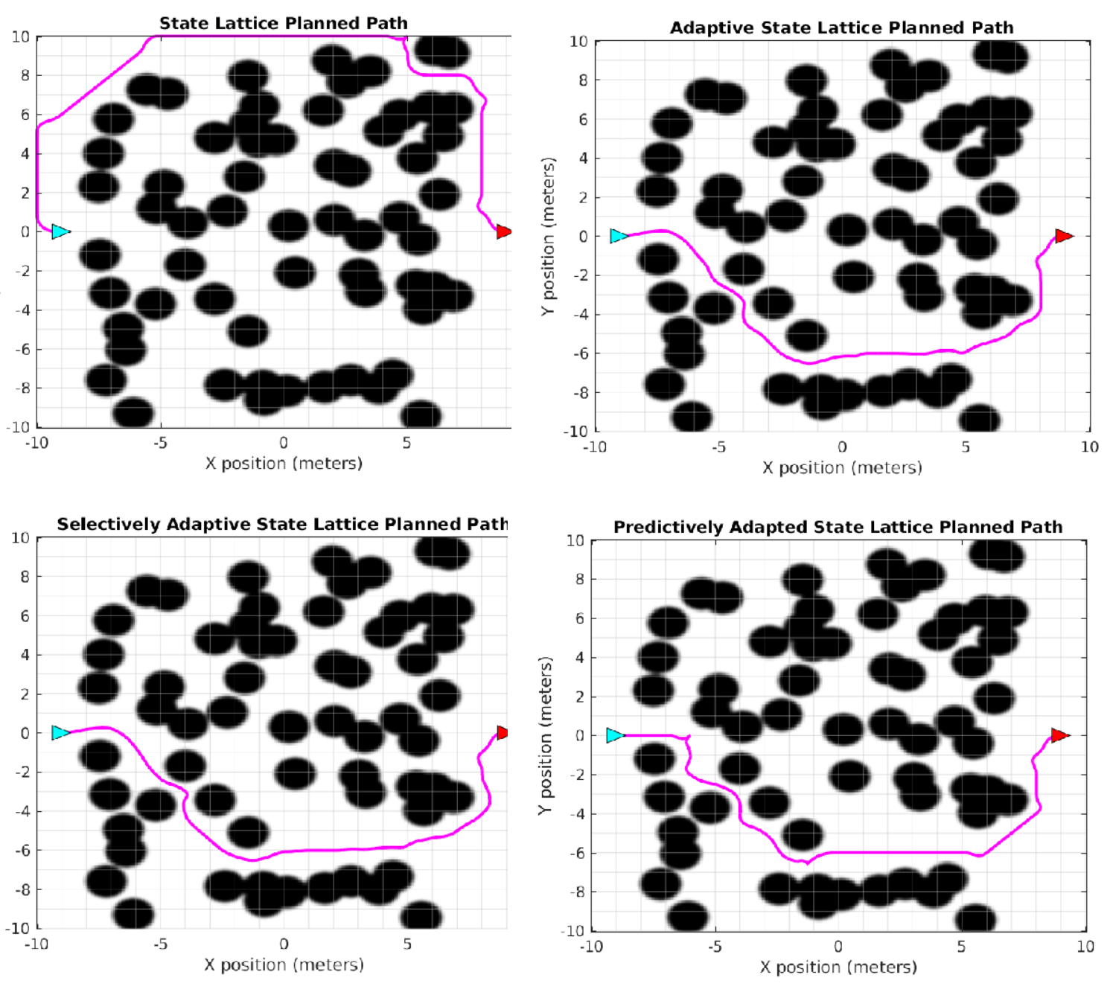


Fig.1. Local mobile robot motion planning with two-dimensional and three-dimensional environments [1]

Articulated Models from Vision & Language

Adapted DCG [2] to infer affordances between objects in a joint model that extracts rich representations from language and vision (Fig. 3).

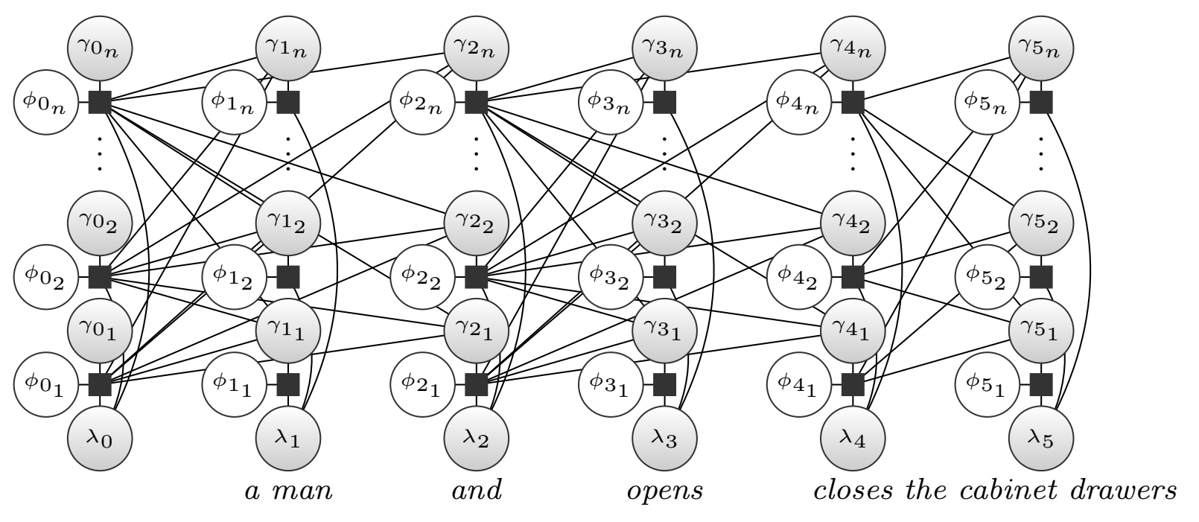


Fig.2. Distributed Correspondence Graph for inferring affordance relationships between pairwise objects

Model improves accuracy over vision-only based models [3] and demonstrates ability to extract affordance and object information from multi-modal interactions in a manipulation-based domain.

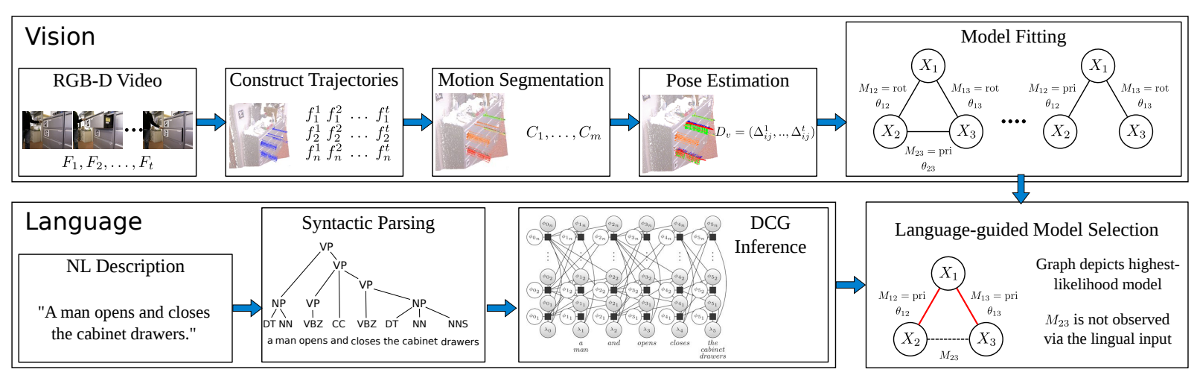


Fig.3. Fusing natural language descriptions with learned models from visual perception

Language-guided Adaptive Perception (AP)

Information in the utterance can be leveraged to infer task specific configurations of the perception pipeline leading to task adaptive world model inferences. DCG [2] is adapted to learn salient visual detectors for robot manipulation tasks in a table-top scenario [4] (Fig. 4). Experimental results show that inferring task adaptive representations (Fig. 4) improves run-time of both perception and symbol grounding.

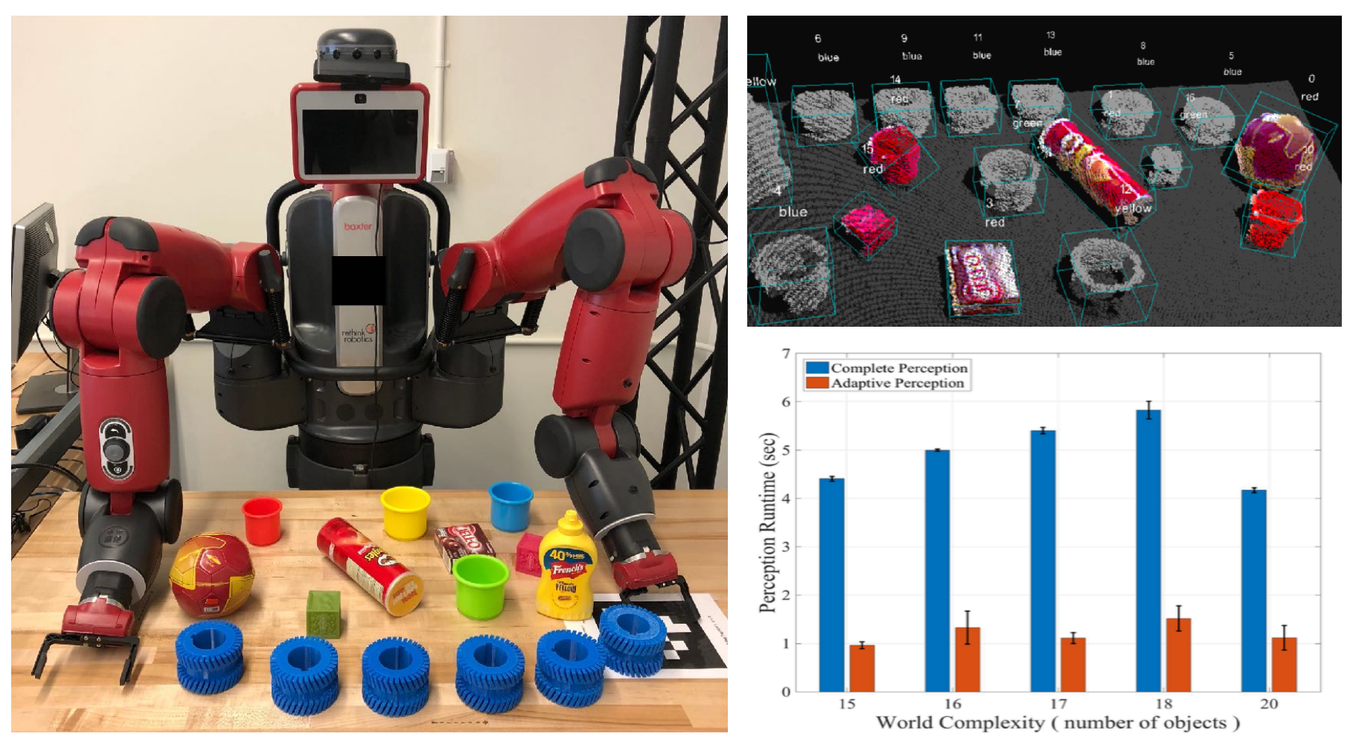


Fig.4. On the left is the physical environment for adaptive perception pipeline experiment. On the right is a representation extracted for "pick up the nearest red object" (top), while the graph (bottom) shows the gain in perception run-time for both adaptive and exhaustive modalities.

Compact Representations via Observations Filtering and AP

Environment information encoded within the instructions can be exploited to further narrow the set of task-relevant observations. To achieve this, three probabilistic graphical models trained from a corpus of annotated instructions infer salient scene semantics, perceptual classifiers, and grounded symbols from language [5].

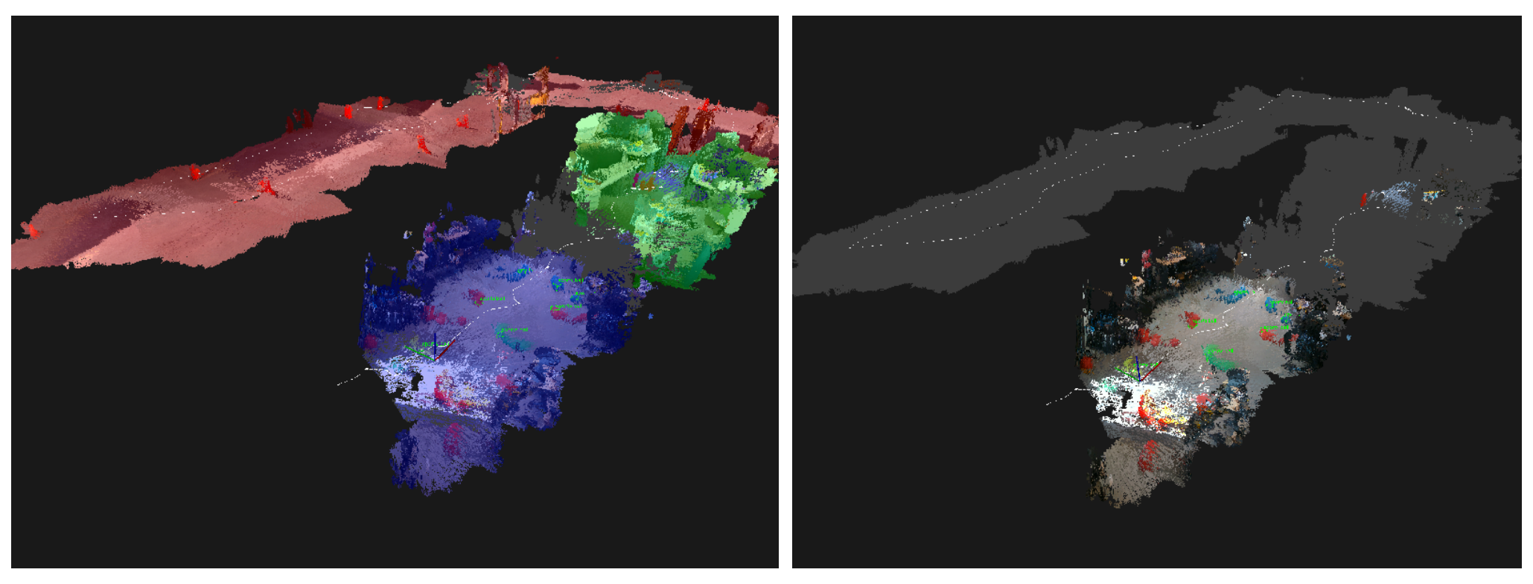


Fig.5. On the left are the inferred scene semantics, On the right is the compact representation extracted for "drive to the nearest ball in the lab".

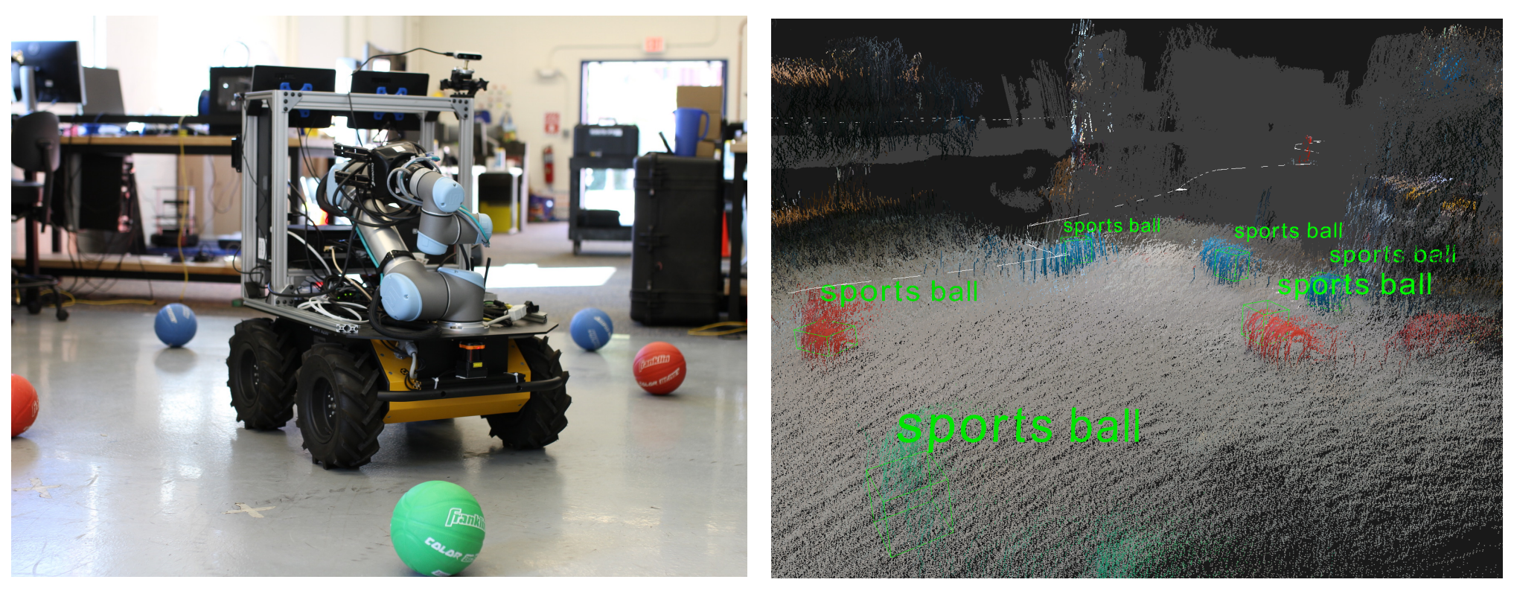


Fig. 6. The robot on the left is instructed to "navigate to the nearest ball in the lab". The image on the right shows a compact world model inferred to facilitate grounding of the instruction.

Language-guided Semantic Mapping and Mobile Manipulation in Partially Observable Environments

Our framework learns to exploit environment and task-related information implicit in a given utterance to ground instructions in partially observed environments. Traditional approaches to language grounding involve reasoning over a highly detailed model of the environment that is assumed to be known at the time of utterance and is computationally expensive to maintain. Our approach [6] learns to reason over a distribution of compact maps that model only task-relevant objects by adapting perception based on the instruction as in [4,5] but also exploits spatial relationships expressed in the utterance to inform the sampled poses of unobserved objects.

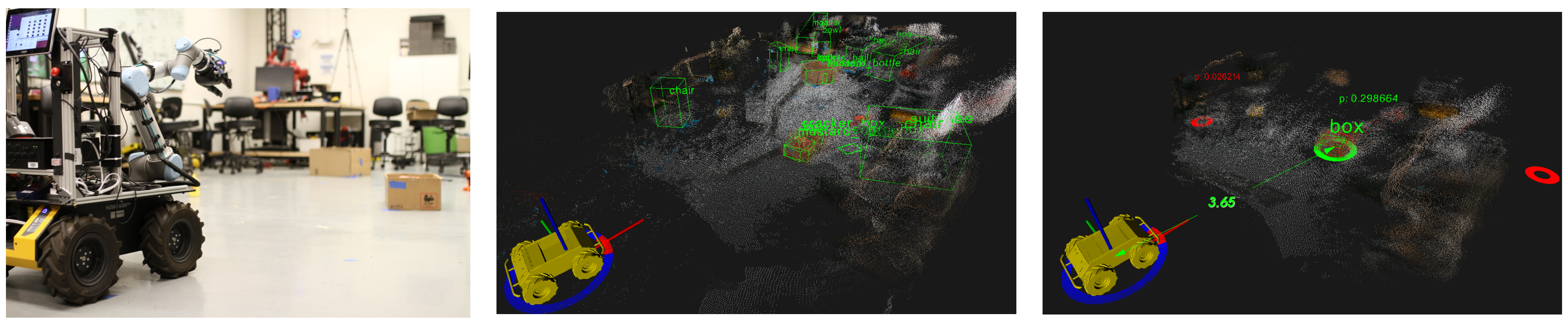


Fig.7. On the left is a robot instructed to "retrieve the ball from the box" in a priori unknown environment. Middle image shows a highly detailed world model that is unavailable for unknown environments and is computationally expensive to maintain. On the right is the inferred distribution of compact maps generated by adapting perception to selectively model task relevant objects. The circles show the distribution of possible locations of the boxes containing a ball.

Adaptive Perception with Hierarchical Symbolic Representations for Mobile Manipulators

Work on adaptive perception has primarily focused on generating minimal world representations for efficient symbol grounding. Planning and control algorithms put additional constraints on minimal representations as relations between objects or their affordances may need to be modeled to select and execute suitable actions. To begin exploring this space, we introduced a novel symbolic representation that selectively represents single-layer hierarchies between object detectors in adaptive perception pipelines [4,5,6] for mobile manipulation [7].

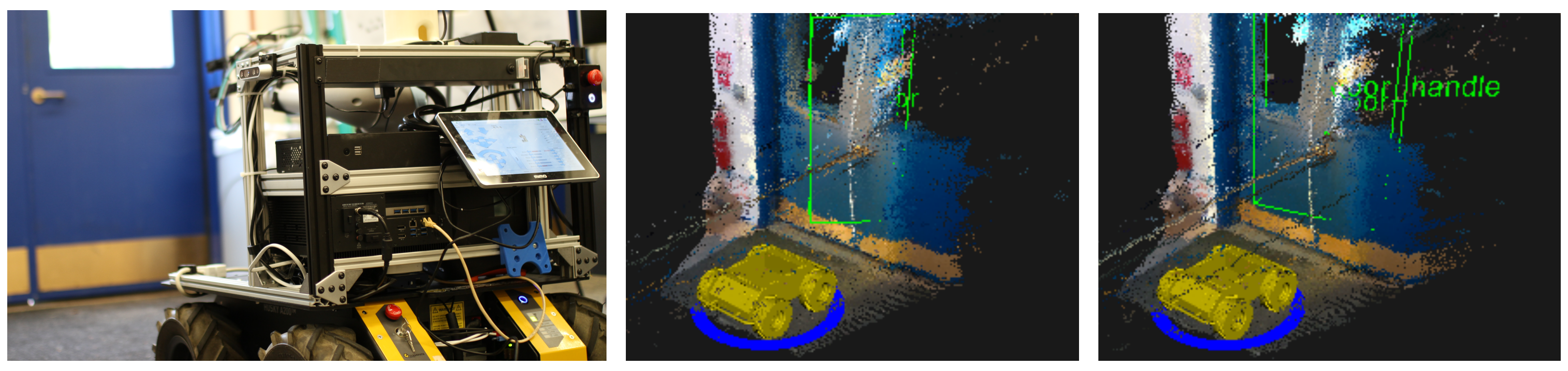


Fig. 8. Experiments in the environment on the left involved commands "go to the door" and "open the door". We demonstrated that using the proposed symbol set, adaptive perception could selectively represent hierarchies between object detectors by learning to condition the inference on the verb in the sentence. In these experiments the "open the door" task inferred door and door handle detectors.

[1] M. Napoli, H. Biggie, and T.M. Howard, "Learning Models for Predictive Adaptation in State Lattices". In: Proceedings of the 11th Conference on Field and Service Robotics. Sep. 2017.

[2] T.M. Howard, S. Tellex, and N. Roy, "A Natural Language Planner Interface for Mobile Manipulators," in Proceedings of the 2014 International Conference on Robotics and Automation, 2014.

[3] A. Daniele, T.M. Howard, and M. Walter, "A Multiview Approach to Learning Articulated Motion Models". In: Proceedings of the International Symposium on Robotics Research. Dec. 2017.

[4] S. Patki and T. M. Howard, "Language-guided adaptive perception for efficient grounded communication with robotic manipulators in cluttered environments," In: Proceedings of the 19th Annual Meeting of the Special Interest Group on Discourse and Dialogue, 2018.

[5] S. Patki, A. Daniele, M. Walter, and T. M. Howard " Inferring Compact Representations for Efficient Natural Language Understanding of Robot Instructions". In: Proceedings of the 2019 International Conference on Robotics and Automation, 2019

[6] S. Patki, E. Fahnestock, T.M. Howard, and M. Walter, "Language-guided Semantic Mapping and Mobile Manipulation in Partially Observable Environments," In *Conference on Robot Learning*. Oct. 2019

[7] E. Fahnestock, S. Patki, and T.M. Howard, "Language-guided Adaptive Perception with Hierarchical Symbolic Representations for Mobile Manipulators," In *6th AAAI Fall Symposium Series on Artificial Intelligence for Human-Robot Interaction*. Nov. 2019