Semantic Robot Programming:
High-level Programming of Robots for Manipulation and Mobility in Human Environments

Odest Chadwicke Jenkins
Laboratory for Perception RObotics and Grounded REasoning SystemS
Computer Science and Engineering Division / Robotics Institute
University of Michigan

So, where is my robot?

As a roboticist, this is the ever-present question posed to me by nearly everyone. People seem to be especially keen on having their snacks, mail, or medicines delivered by robot. And, if such robots were available, they could serve an important role in the ***caretaking of older adults***, to both improve cost and quality of life for an essential sector of our economy.

Figure 1: The Michigan Progress Fetch Robot

My response is that such robots are both closer and further from becoming a reality than people expect. We now have highly capable robot platforms, such as the Michigan Progress Fetch Robot (Figure 1), physically capable of mobility and dexterous manipulation in human environments. Further, we now have computing resources and inference models to make such robots taskable [1] and autonomously capable of performing a wide variety of tasks from the instructions of human users, *but assuming* full observation of the robot's environment. Thus, there is one critical assumption that stands in our way: perception in highly uncertain and unstructured environments.

In other words, ***robots can't work if they cannot see***, and right now robots can see very little.

Our work aims to change this fact and usher in the next generation of autonomous robots capable of functioning in common unstructured human environments. Think about the clutter that forms in a typical kid's bedroom (or graduate student office), and the uncertainty this creates with objects that are occluded, piled, or strewn about. Robust perception is the primary challenge for robotics to realize the levels of capability, reliability, and usability necessary to reasonably meet the expectations of people in these scenarios. Towards this end, we propose the **Semantic Robot Programming (SRP)** [2] to enable high-level programming of robots at the convergence of research in semantic mapping for perception and robot programming by human demonstration.

Semantic Robot Programming =
Semantic Mapping + High-level Programming

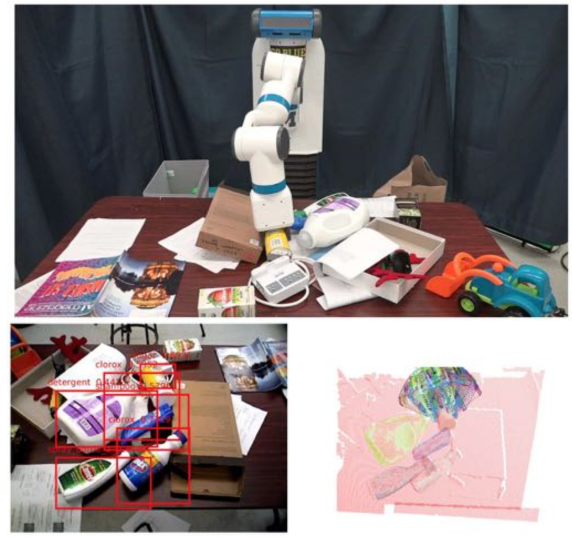

Figure 2: The Michigan Progress Fetch Robot sorting a cluttered set of objects on a table into cleaning (right bin) and non-cleaning categories (left bin). Our methods perform perception by using (bottom left) RGB object recognition to inform (bottom right) sequential pose estimation from 3D point cloud observations and the feasible grasp poses on the manipulated object.

Our work will advance the state of the art in semantic mapping, as scene perception that combines the flexibility of convolutional neural networks with the robustness of generative probabilistic inference [3,4] (Figure 2, video link), assuming known object geometries. This work will enable SRP as a new form of high-level robot programming, where robots autonomously reason to achieve user goals declared spatially as desired world states. We will explore the use of SRP to perform autonomous delivery of snack items in our building, and eventually meal preparation and cleaning tasks for older adults in nursing homes.

Robots must understand where objects are in its environment and how to use such objects dexterously. This is the essence of a robot's semantic map. A semantic map would provide the basic computational symbols to program robots (and their environments) with high-level expressions. In this paradigm, robots are abstracted and become simply the instruction executors for carrying out actions on objects, or using objects as tools. Robot programs will be written or demonstrated as goal states declaring satisfying scene graphs in the world, which will transcend robot-specific configuration spaces (the robotics analogy of assembler). With such abstractions, we can realize many of the usability, portability, and interoperability benefits of high-level languages for the general-use programming of robots. For example, we might not even care which robot carries out a task. We only care that the task is done, similar to how a compute server can run on a variety of operating (or distributed) systems.

Robotics and funding agencies mistakenly assume that computer vision will solve the perception problem. As a result, modern robotics uses a myriad of technical crutches to enable robots to perceive, including: monocolor backgrounds (e.g., green screens), augmented reality tags, highly salient objects (e.g., the ubiquitous "coke can on a table"), and isolated objects on flat surfaces. Such constraints have limited robots to one-off research demonstrations/videos in controlled environments. Our work towards SRP is poised to push past these crutches and deliver viable robot perception, as the missing ingredient in the autonomous control loop for real robots.

Budget - This project is expected have an 18-month scope at $209K to support one graduate student, one faculty month effort, GPU-enabled workstations and cloud compute time, authorship fees, professional and project travel, and miscellaneous supplies.

References

[1] J.E. Laird, K. Gluck, J. Anderson, K. Forbus, O.C. Jenkins, et al. Interactive Task Learning. *IEEE Intelligent Systems*, 32(4): 6-21, 2017.
[2] Z. Zeng, Z. Zhou, Z. Sui, O.C. Jenkins. Scene-level Programming by Demonstration. *arXiv preprint* arXiv:1704.01189
[3] Z. Sui, Z. Zhou, Z. Zeng, O.C. Jenkins. SUM: Sequential Scene Understanding and Manipulation. In *IEEE International Conference on Intelligent Robots and Systems* (IROS), 2017.
[4] Z. Sui, L. Xiang, O.C. Jenkins, and K. Desingh. Goal-directed robot manipulation through axiomatic scene estimation*.* *The International Journal of Robotics Research*, 36(1):86–104, 2017.