Perception for General-purpose Robot Manipulation

Karthik Desingh

Department of Computer Science & Engineering, Minnesota Robotics Institute (MnRI)
University of Minnesota - Twin Cities
kdesingh@umn.edu

To autonomously perform tasks, a robot should continually perceive the state of its environment, reason with the task at hand, plan and execute appropriate actions. In this pipeline, perception is largely unsolved and one of the more challenging problems. Common indoor environments typically pose two main problems: 1) inherent occlusions leading to unreliable observations of objects, and 2) the presence and involvement of a wide range of objects with varying physical and visual attributes (i.e., rigid, articulated, deformable, granular, transparent, etc., see Fig. 1). Thus, we need algorithms that can **accommodate perceptual uncertainty in the state estimation** and **generalize to a wide range of objects**. Probabilistic inference methods have been highly suitable for modeling perceptual uncertainty, and data-driven approaches using deep learning techniques have shown promising advancements toward generalization. Perception for manipulation is a more intricate setting requiring the best from both worlds. My research aims to develop robot perception algorithms that can generalize over objects and tasks while accommodating perceptual uncertainty to support robust task execution in the real world. In this presentation, I will briefly highlight my research in these two research threads.

Accommodation of perceptual uncertainty in state estimation: Starting from a collection of rigid objects on a table to highly articulated objects with rigid body parts such as cabinets, my research has contributed to various Generative Probabilistic Inference (GPI) techniques for state estimation (Desingh et al. 2019; Sui et al. 2017). The most impactful contribution in this line of work is "Pull Message Passing for Nonparametric Belief Propagation (PMPNBP) (Desingh et al. 2019)," where the problem of indoor scene estimation is formulated as a Markov Random Field (MRF). PMPNBP caters to the high-dimensional and multimodal demands of cluttered indoor scenes while being computationally feasible. With the computational advantage of PMPNBP, we show the first demonstration of practically applying a belief propagation method for articulated object pose estimation, with full pose prediction for each object part.

Generalization to a wide range of objects: Real-world objects are too diverse, and assigning handcrafted explicit object representations such as 6DoF poses is non-trivial. We

Figure 1: To understand the perceptual demands of a general-purpose robot manipulator, let us consider an example task of making an omelet. Various categories of objects are involved in this single task; *eggs, butter, salt, pepper, omelet* (brittle, semi-solid items, granular substances), *fridge, dishwasher, cabinet* (articulated objects), *pan, fork, bowl* (rigid objects), *counter, stove* (place/object locations), and *wipes or cloth* (deformable objects). A robot must perceive and interact with each of these categories of objects to accomplish this seemingly simple task.

explore ways to learn object-centric representations from large amounts of simulated task execution data. "Spatial Object-centric Representation Network (SORNet)" (Yuan et al. 2021), a preliminary work in this direction, showed promising generalization of learned object representations to unseen object instances. SORNet representation captures continuous spatial relationships while being trained purely on logical relationships.

References

Desingh, K.; Lu, S.; Opipari, A.; and Jenkins, O. C. 2019. Efficient nonparametric belief propagation for pose estimation and manipulation of articulated objects. *Science Robotics*, 4(30).

Sui, Z.; Xiang, L.; Jenkins, O. C.; and Desingh, K. 2017. Goal-directed robot manipulation through axiomatic scene estimation. *The International Journal of Robotics Research*, 36(1): 86–104.

Yuan, W.; Paxton, C.; Desingh, K.; and Fox, D. 2021. SORNet: Spatial Object-Centric Representations for Sequential Manipulation. In *5th Annual Conference on Robot Learning*.