Research Statement

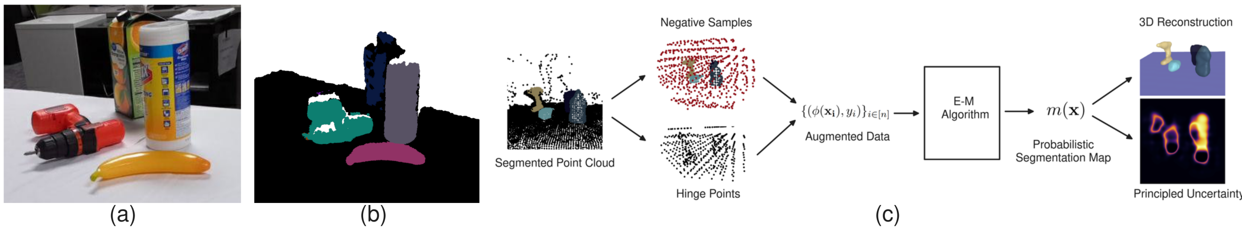

Figure 1: (a) A tabletop scene; (b) segmented partial point cloud; and (c) overview of V-PRISM method

Introduction: A lot of time and effort goes into caretaking for the young, elderly, or those with disabilities. Intelligent robots have the potential to do a lot of the work that goes into such caretaking. These robots could help cook, clean, dress, as well as other necessary activities. Because these activities would be performed in close proximity to humans, the algorithms employed *need* to be reliable and safe. It would be reckless and infeasible to deploy a robotic system that works great 90% of the time, but 10% of the time creates a hazardous situation for the humans involved. Further complicating things is the fact that these robots will also likely have to operate in novel environments and situations. In order to prevent negative outcomes, we need robotic algorithms and systems that are both *robust* and have *uncertainty-awareness*.

Uncertainty during robotic tasks can come from a variety of places. One of the biggest sources of uncertainty is perception. Robots use sensors such as cameras and tactile devices to *perceive* the outside world. These sensors are often noisy (aleatoric uncertainty) and only contain partial information about the relevant outside world (epistemic uncertainty). Modern machine learning approaches to robotic perception often uses types of neural networks that do not produce accurate uncertainty estimates. Probabilistic methods can both provide the needed uncertainty in many of these cases. Consider the task of building a 3D map of cluttered tabletop scene. Such a complete representation is required for many motion planning algorithms, but even a depth camera would only be able to recover partial information about the scene geometry because of various forms of occlusion as shown in Fig. 1a and Fig. 1b. The observation from the sensor could also be very noisy. A non-probabilistic reconstruction method such as PointSDF [1] cannot capture the necessary uncertainty about the scene.

In an effort to probabilistically approach the 3D scene mapping problem, my advisors and I introduced the V-PRISM [2] method (Volumetric, Probabilistic, and Robust Instance Segmentation Maps). Recently, the paper for our work was accepted into the 2024 IEEE/RSJ International Conference on Intelligent Robots and Systems, with me as the first author. The proposed method leveraged the hinge point map representation introduced in [3] that has been utilized in mapping literature. The method first took a segmented depth image of a multi-object scene and constructed a corresponding probabilistic multi-class map. Such depth images can be obtained by using a depth camera such as the RealSense camera used in the paper. One of the main contributions of the work was a novel expectation maximization (EM) algorithm for Bayesian softmax regression, which was performed with respect to the hinge point representation. We showed that this probabilistic approach provided principled and informative uncertainty measurements. Another contribution was a negative sampling technique leveraging tabletop and sensor constraints that removed the need for a training dataset. We showed through both quantitative and qualitative experiments that our method was more robust to novel objects than a learned approach. An overview of the method can be seen in Fig. 1c.

Research Plan: My plan for research during my PhD is to build off the momentum of V-PRISM and work towards improving probabilistic methods in robotic perception and decision-making. Concretely, I have identified three main aims of my future research: (1) Perform Bayesian scene reconstruction by leveraging object-level priors; (2) Use uncertainty measurements provided by V-PRISM to perform active learning with tactile or visual sensors; (3) Leverage a dynamics model in tandem with probabilistic reconstructions to both refine reconstructions and propagate uncertainty through time.

Aim 1: One of the limitations of V-PRISM is that while uncertainty is captured, the backside of the objects are not *reconstructed* according to how one would expect everyday objects to look. This is because V-PRISM uses an *uninformative prior* over the map. In this proposed work, we would perform Bayesian inference while using a more informative prior constructed offline from widely available mesh datasets. We already have preliminary results of such a method being able to reconstruct simple, single-object scenes. We hope to extend this to be both robust to out of distribution objects and extend to multi-object scenes. This would allow methods for downstream tasks to be more robust and performant. In order to validate these claims, experiments similar to V-PRISM should be performed. An experiment also showing improvement in a downstream task such as grasping in clutter with our method would be very helpful to the argument.

Aim 2: Active learning has been studied in the context of robotics. The idea is to have the robot take specific actions with the goal of lowering uncertainty about the state of the outside world. Because V-PRISM, and potentially the method from Aim 1 provide principled uncertainty, we could leverage this to perform active learning. An example could be doing next best view guided movements with a wrist-mounted camera. One exciting direction is to leverage tactile sensors. Active learning with tactile sensors has been proposed before to determine individual new points of contact [4], but not for occupancy maps and trajectories. For this project, we would develop a method that continuously updates the map in a feedback loop by detecting whether the tactile sensor is in contact with the object. Our method would determine actions for the robot that move the tactile sensor towards the most uncertain areas of the scene. Experimentally, we would need to show that this method is both able to *refine* the original map and that the final map is *accurate*.

Aim 3: Studying dynamics of robotic systems is a large focus of robotics research. With regard to robotic manipulation, multiple different ways of modeling dynamics have been proposed. For example, analytical differerentiable simulators as well as learned object dynamic models such as [5] have been introduced. We can leverage these models to both inform reconstruction and propagate uncertainty from our reconstruction methods through time. Doing this can help enable robots to take actions during tasks that reduce uncertainty of outcomes. Experiments to verify this aim would focus on showing reconstruction improvement and uncertainty reduction through simple manipulation tasks like pushing objects in both real world and simulation.

Intellectual Merits: The methods addressing the proposed aims would have clear impacts to robotic manipulation as well as broader scientific applications. Methods for Aim 1 could develop new theory and ideas applicable to Bayesian inference at large, similar to the new EM algorithm in V-PRISM. Aim 2 could provide a novel way to quantify reducing uncertainty when information gain is intractable. Aim 3 would also provide relevant insights into leveraging dynamic systems beyond just manipulation.

Broader Impact: Being able to create accurate reconstructions with principled uncertainty (Aim 1), incorporating uncertainty into robotic movements (Aim 2), and being able to reason about how uncertainty propagates over time (Aim 3) can help ensure robotic systems are robust and uncertainty-aware in critical manipulation tasks. Robust, uncertainty-aware robotic systems can allow easier caretaking for those who cannot care for themselves. Specifically, uncertainty-awareness and robustness can ensure that these methods are *safe* and *reliable* when running in proximity to humans.

References

[1] M. Van der Merwe et al., *International conference on robotics and automation*, 2020, pp. 11516–11522.

[2] H. Wright et al., "V-PRISM: Probabilistic mapping of unknown tabletop scenes," *International Conference on Intelligent Robots and Systems (IROS)*, 2024.

[3] F. Ramos et al., *The International Journal of Robotics Research*, vol. 35, no. 14, pp. 1717–1730, 2016.

[4] Z. Yi et al., *International conference on intelligent robots and systems (IROS)*, 2016, pp. 4925–4930.

[5] D. Driess et al., *Conference on robot learning*, 2023, pp. 1755–1768.