Integrating State Estimation and Perception for Picking

Caris Moses[1] and Jane Shi[2]

I. INTRODUCTION

Going from an RGB-D image of a cluttered scene to successfully picking all objects in the scene is a challenging task in robotics. One approach is to learn an end-to-end model which maps directly from images to grasp success [1]. Another approach is to use a segmentation algorithm, then select grasp points from proposed object segments [2]. We take the latter approach in this work.

Segmentation-based methods aim to generalize to novel objects with enough training data, but if the goal of picking is to be 100% robust, then relying on perfect segmentation is unrealistic. The way in which objects interact in clutter is very complex, making segmentation challenging. Objects that look similar or that form a plane together can often be segmented together, and objects with distinct subparts can be oversegmented.

In this work we aim to predict errors in segmentation by incorporating object-level information, specifically object dimensions. We apply this method to packages, which are cuboidal objects. While this requires some prior information on the novel objects, dimensions can be relatively easy to calculate (eg. from pointcloud data).

II. METHOD

Segmentation-based picking pipelines work by first segmenting a scene, then searching for pick points within the segmented objects. Each pick point is then scored and ranked. This is shown in green in Figure 1. What we propose is a method in which potential segmentation errors influence pick point ranking, shown in blue in Figure 1. To achieve this, an RGB-D image of the scene is captured and run through a segmentation network. We represent the estimated dimensions of detected packages with particle distributions, and initialize the particles uniformly at random.

For each segment, we fit a 3D bounding box which serves as an observation on the package dimensions. We update each distribution with the observation using a Gaussian observation model. Then, for each segment, we calculate the maximum likelihoods that the observation belonged to a true package. We take the maximum of the maximum likelihood values for each true package dimensions, and that serves as the maximum likelihood that this segment corresponds to a true package. Finally, we use a threshold
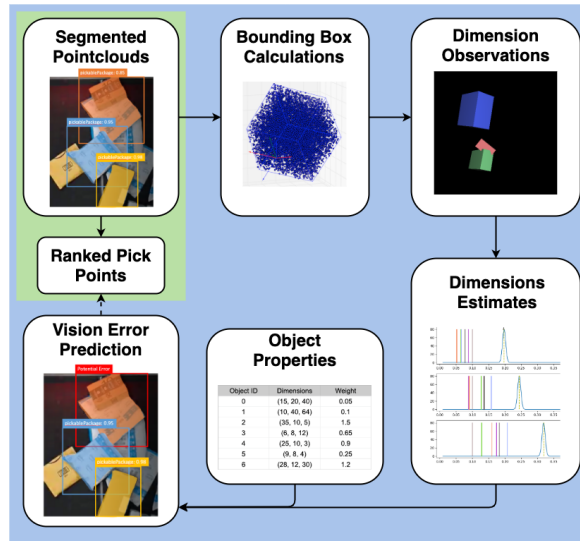

Fig. 1. Our proposed pipeline for pick point ranking.

likelihood value to predict if any segments are erroneous (eg. oversegmented a package, segmented multiple packages, etc.). For future work, these predictions would then influence pick point ranking.

III. PRELIMINARY RESULTS

Our preliminary results were gathered on scenes consisting of 6-10 packages from a set of 12 total packages. A total of 105 package segments were detected. We found a maximum likelihood threshold using receiver operator characteristic (ROC) curves, and selected the threshold which resulted in the highest classification accuracy. With this threshold we were able to achieve a True Negative Rate of 89% and a True Positive Rate of 87%.

This work was a promising first look into incorporating object-level information in a segmentation-based picking pipeline. If we hope to achieve a fully reliable and robust picking pipeline, then we need to aim for systems which are able to reason about not just a scene or an image segment, but object-level information.

[1]Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA, USA carism@mit.edu

[2]Amazon Robotics, North Reading, MA, USA janeshi@amazon.com

REFERENCES

[1] S. Levine, P. Pastor, A. Krizhevsky, J. Ibarz, and D. Quillen, "Learning hand-eye coordination for robotic grasping with deep learning and large-scale data collection," *The International Journal of Robotics Research*, vol. 37, no. 4-5, pp. 421–436, 2018.

[2] D. Morrison, A. W. Tow, M. Mctaggart, R. Smith, N. Kelly-Boxall, S. Wade-Mccue, J. Erskine, R. Grinover, A. Gurman, T. Hunn *et al.*, "Cartman: The low-cost cartesian manipulator that won the amazon robotics challenge," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 7757–7764.