Horizontal Multi-Surface Random Sample Consensus for Robust Customizable Shelf Perception

Yu-Tang Peng[1], Justin Huang[2], and Maya Cakmak[1]

***Abstract*— Many robot tasks across domains from warehouses to homes, such as object fetching, reconfiguration, and stowing, involve interactions with shelves. In this paper, we present a variant of the Random Sample Consensus (RANSAC) algorithm, with a sampling strategy that exploits the regular structure of shelves to robustly detect them. We examined the efficacy and usability of our algorithm by comparing it to a state-of-the-art surface segmentation algorithm implemented in PCL both with general parameters and with shelf-specific customized parameters. The results indicate that our algorithm has better precision and much higher recall in both settings.**

I. Introduction

Many tasks we can imagine robots doing in environments from homes to warehouses involve fetching, stowing, or reconfiguring objects on horizontal surfaces, including shelves. While there has been a long line of research on detection of tabletops, counter tops, or other horizontal open work spaces and segmentation of unknown objects on them [1], [2], [3], many of the developed algorithms explicitly assume that there is just one surface and fail to work for detecting shelves. Recent events like the Amazon Picking Challenge[1] have challenged researchers to work in and around shelves. One algorithm that was developed by Trevor et al. to address scenes with multiple surfaces is the Organized Multi Plane Segmentation (OMPS) [4], which was implemented and deployed in the widely-used Point Cloud Library (PCL) [5]. This algorithm can be used to detect multiple planar surfaces of any orientation. The generality comes at the cost of not taking advantage of the known structure of specific surface compositions, including shelves which consist of multiple *horizontal* planes, often at equal distance intervals. Other approaches that allow shelf perception either require complete 3D mesh models of the shelf, modeling surfaces as lines [6] or involve forming dense maps of the scene that might include shelves without segmentation [7], [8]. In this paper, we propose Horizontal Multi-Surface RANSAC (HMS-RANSAC), a variant of the popular Random Sample Consensus algorithm [9] that can detect multiple horizontal surfaces. Examples of outputs from our algorithm can be found in Fig. 1.

[1]Paul G. Allen School of Computer Science and Engineering, University of Washington, Seattle, WA 98195. {pengy25,mcakmak}@cs.washington.edu

[2]Jet Propulsion Laboratory, Pasadena, CA 91109. Work done while affiliated with the University of Washington. justin.huang@jpl.nasa.gov

[1]https://amazonpickingchallenge.org/

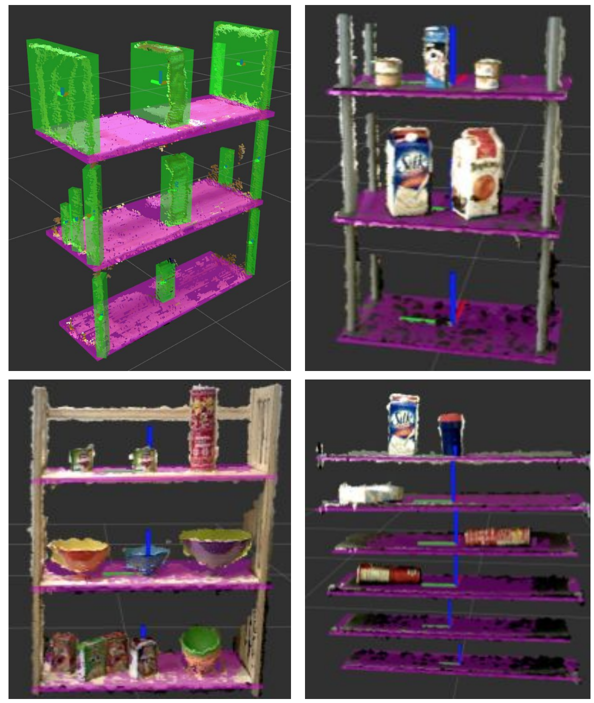

Fig. 1. Examples of surface segmentations using our algorithm HMS-RANSAC. Detected surfaces are visualized with purple boxes. Shelf segmentation can help segment objects on the surfaces for manipulation tasks, as shown with green boxes in the top left.

II. Proposed Method

We assume that we are only searching for horizontal surfaces in a given point cloud, within a certain angular threshold. To locate surfaces, we iteratively sample a single point from the point cloud and hypothesize the existence of a horizontal plane passing through the sampled point. We then count the number of *inliers*, points that are close to the plane. Our algorithm has a parameter, *inlier distance*, that specifies how close a point must be to a plane to be considered an inlier. This parameter can be used to govern the "thickness" of the output surfaces. If the hypothesized plane actually coincides with a surface in the point cloud, it will have many more inliers than other randomly chosen planes. Our algorithm outputs all surfaces that have at least a certain number of inliers. If two output surfaces are within a certain distance of each other, we assume that they are duplicates and pick the surface with more inliers.

We also refine each output surface by running RANSAC to fit a plane to its inliers. If the refined plane is still horizontal according to the angular tolerance parameters, we replace the

TABLE I
PERFORMANCE COMPARISON BETWEEN HMS-RANSAC AND OMPS.

Metrics	HMS-RANSAC	HMS-RANSAC (tuned)	OMPS	OMPS (tuned)
Precision	98.78%	99.25%	50.00%	97.92%
Recall	91.60%	99.25%	14.15%	44.34%
Time spent	948.80 ms	489.20 ms	194.04 ms	228.37 ms

candidate surface with the refined plane.

III. EVALUATION OF THE METHOD

To evaluate HMS-RANSAC, we compared it to the Point Cloud Library's implementation of Organized Multi Plane Segmentation (OMPS). We created a dataset with 26 point clouds of shelf and tabletop scenes and ran both our algorithm and OMPS on each scene. We ran two types of evaluation: one in which each algorithm had a generic set of parameters that was used for all scenes, and one in which we individually tuned the parameters for both algorithms on each scene. The first evaluation represents cases where the algorithm will be applied to many different scenes, while the second evaluation tells us how well each algorithm can perform in known, structured scenes.

We chose HMS-RANSAC generic parameters by optimizing HMS-RANSAC on another smaller dataset of representative scenes, with later adjustment. The generic parameters for HMS-RANSAC were: a minimum iteration of 300, angular threshold of 10 degrees, surface point-distance threshold of 1 cm, and minimum inliers per surface of 8000. The generic parameters for OMPS were from its implementation in PCL: angular threshold of 3 degrees, distance threshold of 2 cm, minimum inliers per surface of 1000. OMPS also requires normal vectors to be computed for the input point cloud. Although [4] used a different algorithm for normal computation [10], we found in early experiments that using PCL's NormalEstimationOMP algorithm improved recall.

We ran both algorithms 10 times on each scene and visually inspected the results. We measured the algorithms' precision, recall, and time spent per scene.

IV. RESULTS AND DISCUSSION

In our evaluation, we found that HMS-RANSAC achieved higher precision and recall than OMPS did in our dataset (Table I). This held true after we hand-tuned OMPS for each individual scene. While we were able to improve OMPS's precision to be comparable with our algorithm's precision, its recall remained low, just 44.34% on average. It may be the case that our dataset (examples shown in Figure 1) differs from what the creators of OMPS intended.

However, our algorithm ran more slowly compared to OMPS per scene. One difference between HMS-RANSAC and OMPS is that OMPS requires computing normal vectors for the input point cloud. The time reported in Table I excludes the normal computation time for OMPS, since it is reasonable to expect that developers might compute normal vectors for other purposes. Depending on the choice of normal computation algorithm, HMS-RANSAC may be faster than OMPS. For example, using PCL's NormalEstimation algorithm, the total average time to run OMPS is 1429.9 ms. Speed was not a primary consideration in the development of HMS-RANSAC, so its speed could also be improved with further development work.

In this work, we demonstrated that the proposed algorithm, with fixed set of parameters, can detect surfaces with high precision and recall in a variety of scenes. Our evaluation also shows that its performance can achieve near-perfect levels if the algorithm is tuned for each specific scene it encounters. This means one can use the general parameters as a reference to effectively optimize the performance of HMS-RANSAC based on the task.

In terms of future work, we want to optimize the speed of HMS-RANSAC. We also want to develop tools that help users of HMS-RANSAC optimize parameters for a variety of scenes. Our implementation is available as an open-source ROS library, and it has been used to segment surfaces and objects in shelf scenes for robot manipulation tasks. Documentation on its usage and links to its source code can be found at `https://wiki.ros.org/surface_perception`.

ACKNOWLEDGMENT

This work was supported by the National Science Foundation IIS-1525251 "NRI: Rich Task Perception for Programming by Demonstration."

REFERENCES

[1] R. B. Rusu, "Semantic 3D object maps for everyday manipulation in human living environments," Ph.D. dissertation, Computer Science department, Technical University of Munich, Germany, October 2009.

[2] D. Holz, S. Holzer, R. B. Rusu, and S. Behnke, "Real-time plane segmentation using rgb-d cameras," in *Robot Soccer World Cup*. Springer, 2011, pp. 306–317.

[3] A. Richtsfeld, T. Mörwald, J. Prankl, M. Zillich, and M. Vincze, "Segmentation of unknown objects in indoor environments," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2012, pp. 4791–4796.

[4] A. J. Trevor, S. Gedikli, R. B. Rusu, and H. I. Christensen, "Efficient organized point cloud segmentation with connected components," *Semantic Perception Mapping and Exploration (SPME)*, 2013.

[5] R. B. Rusu and S. Cousins, "3D is here: Point cloud library (PCL)," in *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, 2011.

[6] R. Li, Z. Huang, E. Kurniawan, and C. K. Ho, "AuRoSS: An autonomous robotic shelf scanning system," in *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, 2015, pp. 6100–6105.

[7] P. Henry, M. Krainin, E. Herbst, X. Ren, and D. Fox, "Rgb-d mapping: Using kinect-style depth cameras for dense 3d modeling of indoor environments," *The International Journal of Robotics Research*, vol. 31, no. 5, pp. 647–663, 2012.

[8] J. Sturm, N. Engelhard, F. Endres, W. Burgard, and D. Cremers, "A benchmark for the evaluation of rgb-d slam systems," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2012, pp. 573–580.

[9] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[10] S. Holzer, R. B. Rusu, M. Dixon, S. Gedikli, and N. Navab, "Adaptive neighborhood selection for real-time surface normal estimation from organized point cloud data using integral images," in *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2012, pp. 2684–2689.