Interactive singulation of objects from a pile

Lillian Chang[*†], Joshua R. Smith[†‡], Dieter Fox[†]
[†] Computer Science & Engineering, University of Washington, Seattle, USA
[‡] Electrical Engineering, University of Washington, Seattle, USA
[*] Intel Corporation, University of Washington Intel Science and Technology Center, Seattle, USA

I. INTRODUCTION

An ongoing challenge in robotics is the interaction with clutter in unstructured environments. In particular, objects may be placed close together or overlapped such as in a pile of toys or groceries. The composition of a pile may be one of many classes depending on the scale and properties of what is considered the unit item. A common example is a pile of primarily rigid objects, such as a set of game pieces or a stack of books. In other cases, the unit items may be articulated or deformable objects, such as pile of dishrags or a jumble of rope. The pile itself may even be considered more as a single deformable material when the unit pieces have small granularity, such as a heap of chopped vegetables or sugar. Robust exploration and interaction with piles will be necessary for a service robot that encounters clutter in household environments. The interaction strategies may range from the singulation of individual units to directly classifying and manipulating a material pile as one entity.

In this work, we focus on the first example of a pile composed of primarily rigid objects. Cluttered placement within a pile is a challenge for both perception and manipulation: object recognition may be poor due to occlusion from surrounding objects, or a grasp attempt may fail when mistakenly applied to a target that is actually multiple objects. Our goal is that the robot be able to identify and separate individual objects from a novel pile of unmodeled items. This requires the integration of perception of the scene state with a manipulation plan for interacting with the pile. This singulation process is intended as an initial exploration stage that would precede and facilitate future actions such as object recognition or grasping, where spatial separation improves performance.

Our video demonstrates an example arranging task requiring interaction with multiple objects in a pile. In this example, the PR2 counts the number of objects in a pile through exploration of the object by pushing or perturbation actions. If the observed motion of a pile is consistent with a single rigid motion hypothesis, additional moves are attempted to test whether there is continued evidence of singulation. Grasp attempts are executed after sufficient evidence, and the item is arranged in the goal area. Otherwise, the pile is re-perturbed to introduce new motion segments which then become candidates for individual separated items.

This material is based upon work supported by the National Science Foundation under Grant # 1019343 to the Computing Research Association for the CIFellows Project.

II. RELATED WORK

Interaction with piles of objects has been examined previously for the bin-picking application. Several methods for workpiece acquisition reframe the problem as the identification of potential grasping or hold sites based on the local shape of objects, without explicitly segmenting an individual workpiece from the pile. For example, Kelley et al. [1] describes vision algorithms for identifying smooth surfaces suitable for vaccuum grippers or hold sites for parallel jaw grippers (see also [2]). More recently, this approach of matching or filtering surface features for a particular gripper geometry was also achieved using 3D range data and the PR2 grippers for a cashier checkout operation [3]. Other investigations of bin-picking have focused on machine vision techniques for pattern recognition of a known workpiece shape [4, 5] or segmentation of the topmost piece in a pile [6]. Our approach does not assume prior modeling of the objects and can also be used to interact with objects that are not graspable by the robot end effector.

Related work has also investigated the integration of robot actions with object perception or modeling. Object segmentation from video sequences of robot pushing actions has been demonstrated for rigid planar objects by Kenney et al. [7] and for articulated objects by Katz and Brock [8, 9]. These works focused on the segmentation of objects from a video of a pre-planned robot motion that results in clear motion cues of the target object. Our work extends these approaches to interactive perception by examining how the push or perturbations can be planned to accumulate a sequence of motion cues, which individually may not provide sufficient evidence for singulation.

Another area of related work is next-best view planning for interactive object modeling, investigated by e.g. Krainin et al. [10]. These methods focus on a single object either already grasped in-hand by a robot or fixed in the world relative to a moving camera. The aim of these methods is to create a complete model of the object geometry, and these techniques would complement our results to create an object model once it is separated from clutter.

Recent work by Dogar and Srinivasa [11] presents push-grasping as a framework for dealing with clutter in the environment. The robot motion plans depend on previous knowledge of object models for the target and the cluttering

Fig. 1. Multiple objects may appear as a single segment after clustering 3D data due to lack of spatial separation.

objects. Our method could be used as an initial exploration step for identifying individual movable items among the clutter.

III. Method overview and initial results

Our demonstration shows how a robot can learn what portion of a cluttered pile is an individual rigid segment and then separate it spatially. The objects and pile configurations are not known a priori. We assume that the objects have some visual or geometric features that can be used to identify candidate correspondences between two states of the pile.

We use image and spatial data from a Kinect depth-camera (Microsoft Corporation) mounted on the PR2 robot (Willow Garage). For each pile state between actions, we capture the high-resolution 2D image and the corresponding 3D point cloud.

Piles are identified from the cloud data by first fitting and removing the plane of the support surface and then spatially clustering the remaining cloud points. As shown in Fig. 1, spatial clustering is not necessarily sufficient to segment individual objects in clutter.

Pushes to perturb the pile are planned based on the target pile location and the surrounding piles or obstacles. The pushes are straight line motion paths of the robot end-effector that approach the centroid of the pile point cloud in a direction corresponding to the most clearance around the pile boundary. The purpose of the push is to perturb the pile enough to provide a state change and does not require an object model or the surface properties for predicting the resulting motion.

The pile state before and after a push perturbation are evaluated to determine whether a single rigid motion is sufficient to describe the observed change (Fig. 2). One challenge is that, even when a pile consists of only a single object, the candidate feature correspondences between image frames may not always be correct. Thus multiple rigid motions drawn from the image point correspondences are evaluated against the full 3D point cloud to determine a candidate explanation for the pile state change.

In the video, repeated successful matches of a candidate motions results in a grasp attempt to place a singulated item in the goal area on another table. Uncertain matches result in new exploration pushes to uncover new candidates for individual items. In this manner, the identification and separation of an item is based on the accumulation of successful matches and can recover from failures to fit a candidate motion. Table I shows results from initial experiments (not shown in the video) testing the perceptual model for matching motions of different everyday objects. We found that even for a separated single item, noisy data and changing views of the objects required several actions and observations before accumulating strong evidence of singulation. Work in progress is developing methods to address some of these challenges.

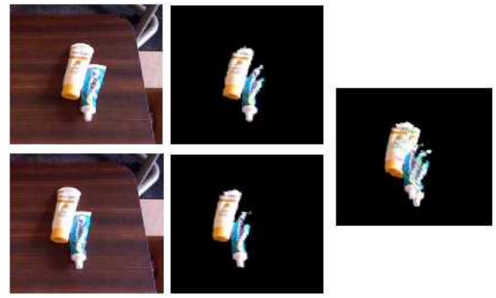

Fig. 2. A pile of two objects shifts shape after a push perturbation. The segmented pile before the push (top) is compared to the state after the push (bottom). The two frames are evaluated to determine if there exists a single rigid transform that explains the motion (right). Rarely do the points in the two frames match completely due to noisy edges of the pile segmentation or changes in the partial object view due to the perturbation.

TABLE I
In the example counting demonstration, piles are pushed until an individual object reaches the goal area and counted.

Number of objects	Number of push perturbations
3 alphabet blocks	39
3: snack box, pastry package, hummus tub	21
3: toothpaste, sunscreen, ribbon	29
3: toothpaste, sunscreen, ribbon	78

References

[1] R. B. Kelley, H. A. S. Martins, J. R. Birk, and J.-D. Dessimoz, “Three vision algorithms for acquiring workpieces from bins,” vol. 71, no. 7, pp. 803–820, 1983.

[2] J.-D. Dessimoz, J. R. Birk, R. B. Kelley, H. A. S. Martins, and C. Lin, “Matched filters for bin picking,” no. 6, pp. 686–697, 1984.

[3] E. Klingbeil, D. Drao, B. Carpenter, V. Ganapathi, O. Khatib, and A. Y. Ng., “Grasping with application to an autonomous checkout robot,” in *Proceedings IEEE Int. Conf. Robotics and Automation ICRA 2011*, 2011.

[4] L. Jacobson and H. Wechsler, “Invariant image representation: A path toward solving the bin-picking problem,” in *Proc. IEEE Int. Conf. Robotics and Automation*, vol. 1, 1984, pp. 190–199.

[5] B. K. P. Horn and K. Ikeuchi, “The mechanical manipulation of randomly oriented parts,” *Scientific American*, vol. 251, no. 2, pp. 100–111, 1984.

[6] H. Yang and A. Kak, “Determination of the identity, position and orientation of the topmost object in a pile: Some further experiments,” in *Proc. IEEE Int. Conf. Robotics and Automation*, vol. 3, 1986, pp. 293–298.

[7] J. Kenney, T. Buckley, and O. Brock, “Interactive segmentation for manipulation in unstructured environments,” in *Proc. IEEE Int. Conf. Robotics and Automation ICRA ’09*, 2009, pp. 1377–1382.

[8] D. Katz and O. Brock, “Manipulating articulated objects with interactive perception,” in *Proc. IEEE Int. Conf. Robotics and Automation ICRA 2008*, 2008, pp. 272–277.

[9] ——, “Interactive segmentation of articulated objects in 3d,” in *Workshop on Mobile Manipulation: Integrating Perception and Manipulation at the IEEE International Conference on Robotics and Automation 2011*, 2011.

[10] M. Krainin, P. Henry, X. Ren, and D. Fox, “Manipulator and object tracking for in-hand 3d object modeling,” *International Journal of Robotics Research*, 2011.

[11] M. Dogar and S. Srinivasa, “Push-grasping with dexterous hands: Mechanics and a method,” in *Proceedings of 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2010)*, October 2010.