Proc. IROS-2018 Workshop on
Towards Robots that Exhibit Manipulation Intelligence
Madrid, Spain, Oct. 2018

Sweeping Motions for Finding and Picking up a Specific Object in a Pile of Objects

Jun Miura, Kota Chikaarashi, and Wataru Miyazaki
Department of Computer Science and Engineering
Toyohashi University of Technology

***Abstract*— Sweeping is effective for moving multiple object at a time in such a case where we remove objects around the target object to pick it up or where we remove objects possibly occluding the target to find it. Such a situation often happens when, for example, finding a small toy in a box or in finding a stationary in a drawer. A key to using sweeping effectively is to determine when and where to sweep. This is realized by recognition of the scene and planning of sweeping motions. This paper presents two example scenarios where sweeping is effective, and the developed robotic systems for them.**

***Index Terms*— Sweeping motion, hand motion planning, object placement estimation.**

I. Introduction

Personal service robots have been expected to be widely used in a near future that can support people in their everyday situations. Possible tasks of such robots are: fetching a user-specified object, putting tableware away, and tidying up a room. Object search and manipulation is one of the commonly-used functions for such robotic tasks. Fig. 1 shows typical situations we face in our daily life. Although learning-based grasp planning that can cope with such situations have recently been studied extensively [1], [2], another approach is required when the target object is blocked or occluded by surrounding objects.

In this paper we focus on the use of *sweeping motion*, which is a motion to sweep (move) multiple objects at once and effective for the situations like Fig. 1. This paper deals with two cases of utilizing sweeping motions: one is to remove objects which are obstructing a hand motion for picking up a target object; the other is to remove objects which are (fully or partially) occluding a target object.

Object manipulation other than just pick-and-place ones has been one of the main research areas in manipulation domain. A typical manipulation motion without grasping is *pushing* [3], [4], [5]. A new use of pushing is to physically segment object clusters for visual recognition [6], [7], however only applied to cases where objects are on a flat table.

Fig. 1. Typical cases where sweeping is effective for object search.

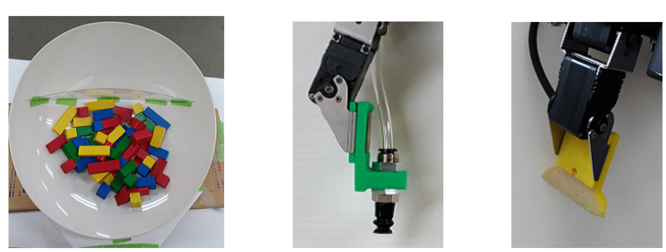
This paper presents our attempts to utilizing sweeping motion in object search in a bowl. Our contributions lie in: introducing sweeping motion for object search, presenting heuristic sweeping motion planing strategies, and evaluating our approach in integrated hand-eye systems. The paper is based on our previous papers [8], [9] and tries to extract a common structure of sensing and manipulation.

II. Tasks and Approaches

Fig. 2 shows two tasks dealt with in this paper. The first task (see Fig. 2(a)) is to find a target object in a bowl with objects. The shape and appearance of the target is given so that an image-based object localization is utilized. Since each object is not small, objects easily tumble in a bowl, and sweeping is used for making a target object graspable. The second task (see Fig. 2(b)) is to find a target object in a clutter of smaller-sized blocks with similar shapes. Since many objects occlude the target, sweeping for removing them at once is effective.

A system to find and pick up a target object is usually composed of a *sensing* part and a *planning and execution* part (see Fig. 3). The sensing part analyzes the scene and

(a) Picking up a known object. (left) Humanoid HIRO and a pile of objects in a bowl. (right) A Target object (front and back side).



(b) Picking up a specific block with suction mechanism. (left) Many objects in a bowl. (center) Suction nozzle. (right) Plate for sweeping.

Fig. 2. Target tasks.

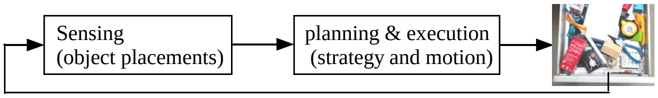


Fig. 3. General system structure.

locates the target object. If it is recognized with a high confidence, a hand motion for picking it up is planned and executed. Additional hand motions could be generated for removing surrounding objects which prevent the target from being grasped. If the target is not recognized due to a partial or complete occlusion by other objects, a hand motion for removing such objects is generated. In both cases, sweeping motions are utilized if necessary.

We developed systems for dealing with these two tasks with a combination of sensing, planning, and execution. The following two sections explain the system in some detail for the first and the second task, respectively.

III. Sweeping for Grasping

A. *Object recognition and pose estimation*

The task is to pick up a specified target object in a pile of objects in a bowl (see Fig. 2(a)). All objects have box-shape and the size and the appearance (i.e., images of six surfaces) of each target object are given in advance. A set of SIFT features [10] are also extracted and recorded in advance, which are used for recognizing and estimating the pose of a target object. Fig. 4 shows results of recognition and pose estimation. Observation is done by a camera attached to a hand (called *hand camera*).

Data for other (unknown) objects are not given, although their shapes are known to be boxes. We use depth data from Kinect to extract their rectangular surfaces. The steps for extraction are: depth data acquisition, background (i.e., bowl) subtraction, and a plane fitting in 3D followed by a rectangular fitting in 2D. Fig. 5 shows an example of surface extraction. Observation is done by a Kinect on the head.

B. *Hand motion planning*

Hand motion planning is to determine a hand trajectory with predefined target poses (e.g., grasp point), given the

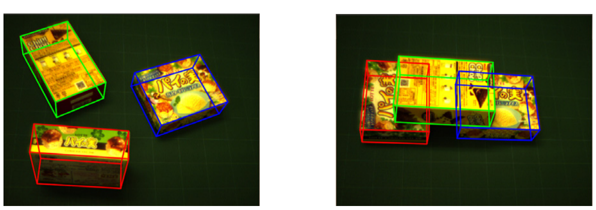


Fig. 4. SIFT-based recognition and pose estimation examples.

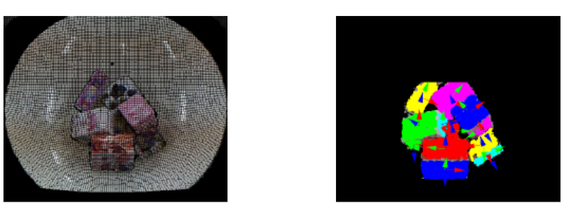


Fig. 5. Depth data and extracted surfaces with normals.

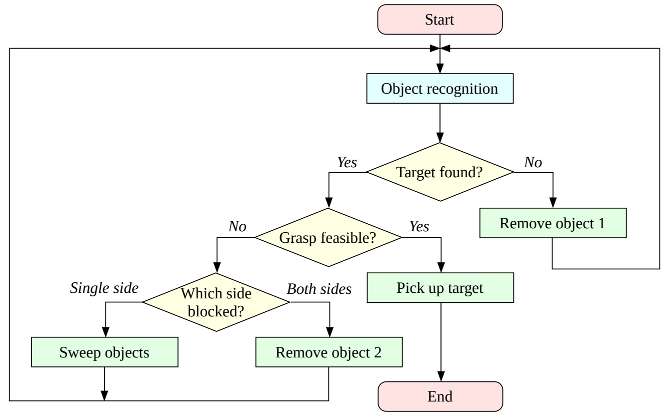


Fig. 6. Flow of object search. Green boxes indicate planning and execution of sweeping or pickup hand motions. The blue box indicates the object recognition operation which is carried out every time a sweeping or a pickup motion is executed.

pose of the target object, those of the others, and the robot configuration. We consider the following three hand motions:

Pickup motion for picking up an object. One of the two hands is used.

Sweeping motion for moving and holding multiple objects by a hand so that the other hand can pick up a target object.

Recognition motion for making a hand camera face a target surface for recognition.

Fig. 6 shows the outline of operations. The robot first recognizes the current situation (i.e., how objects are piled up) and chooses an appropriate action. The recognition is carried out every time a sweeping or a pickup motion is executed because such a motion may alter the poses of other objects. We divide possible situations into the following four cases, where hand motions are planned accordingly (see green boxes in Fig. 6).

case 1: target object is not found:

This is the case where the target object is occluded by others. Considering the case where the target object is partially visible (but not sufficiently visible for SIFT-based recognition), we search the extracted surfaces for the one which has the most similar color with the target, and try to remove objects above this target candidate. Among such objects, the robot chooses the object at the highest position as the starting one, and checks all objects in the ascending order of the distance to the starting one to determine the object to pick up.

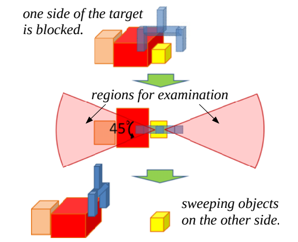
case 2: target object is found and can be picked up:

In this case, the robot picks up the target object and the task is completed.

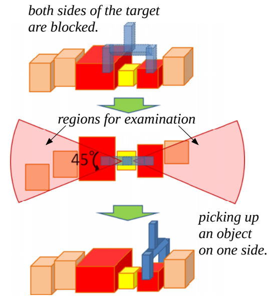
case 3: target object is found but cannot be picked up:

This case is further divided into the following two sub-cases depending on how objects are placed around the target:

case 3a: one side of the target object is open: in this case, if the objects on the other side are all removed, the robot can pick up the target object. So the robot does a sweeping motion for moving them away from the target. Fig. 7(a) shows this sub-case.

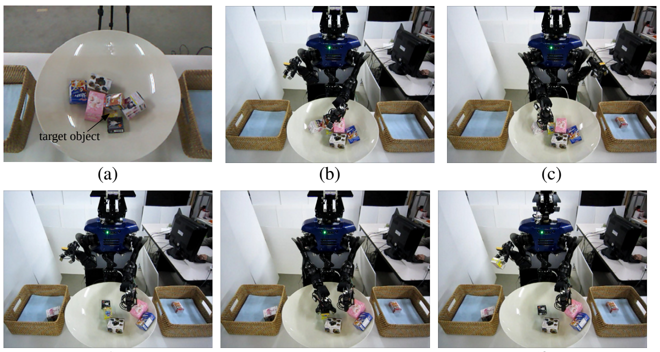


(a) One side of the target is blocked; a sweeping motion is chosen.



(b) Both sides are blocked; a pickup motion is chosen.

Fig. 7. Choice of picking or sweeping motions. Regions to examine on either side of the target are set as circular sectors with a central angle of 45 [deg] as shown in the figures.



(a) (b) (c)

(d) (e) (f)

Fig. 8. Sweeping and keeping objects to grasp the target object: (a) initial state; (b) remove one object; (c) remove a blocking object; (d) sweep and hold two objects; (e) grasp the target; (f) pick up the target.

case 3b: both sides of the target object are blocked: in this case, the robot picks up an object from one of the regions where the number of objects is smaller. The reason of this choice of the region is that the objects on the other side (i.e., those in the region where the number of objects is larger) are expected to be swept away later, and that sweeping as many objects as possible will reduce the total time for completing the task. Fig. 7(b) shows this sub-case.

Fig. 8 shows the initial state of objects and hand motion sequences for picking up a target object.

IV. Sweeping for Finding

A. Probabilistic object placement estimation

In a pile of many objects, due to frequent occlusions, a target object may not be visible or may be only partially visible. Since ignoring partially-visible objects in a very cluttered scene as shown in Fig. 2(b) is inefficient, we infer object identities also for such partially occluded objects.

In this task, objects are all wooden blocks. Each object is characterized by its shape (a set of surface descriptions) and color. Fig. 9 shows all of twelve surface models and seven object models. We developed a method of probabilistically estimating object identities using surface-object relationships in the models and detected surfaces [9]. Fig. 10 shows examples of object identification and placements. Objects 0 and 1 have one or two candidate models. Objects 2 and 3 have much more candidates because only one surface is detected.

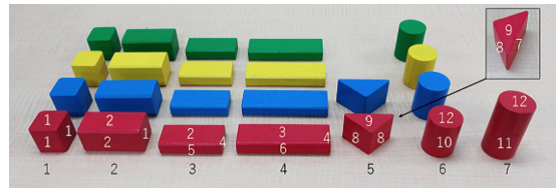
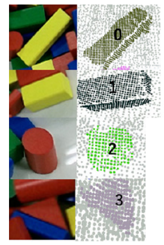


Fig. 9. Object models and surface models.



Detected object	Candidate model objects (probability)
0	3(1.000)
1	2(0.649),5(0.351)
2	1(0.346),2(0.116),3(0.146),6(0.196),7(0.196)
3	1(0.125),2(0.151),3(0.055),4(0.166),5(0.498)

Fig. 10. Object identification results.

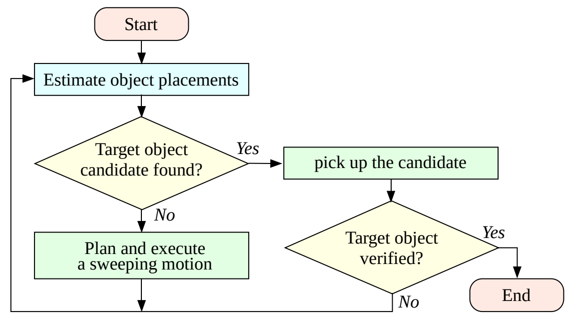


Fig. 11. Flow of target object finding process.

B. Hand motion planning

Fig. 11 shows the process of finding a target object. The observation step determines if a target object is found and a hand trajectory is determined for either picking up the target object or sweeping with hoping that the target object becomes visible.

1) Hand motion for verification: Among target object candidates with probabilities, if the highest one is above a threshold, the target is considered found and will be picked up by a suction mechanism. This pickup motion includes the verification step, in which the robot observes that object from two different viewpoints to verify that it is certainly the target. Fig. 12 shows a verification scene.

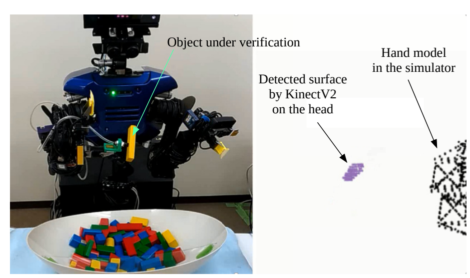
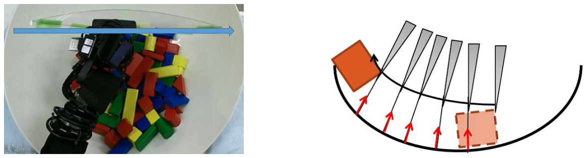
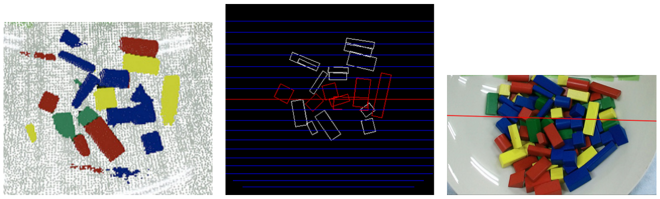


Fig. 12. Verification of the lifted object.



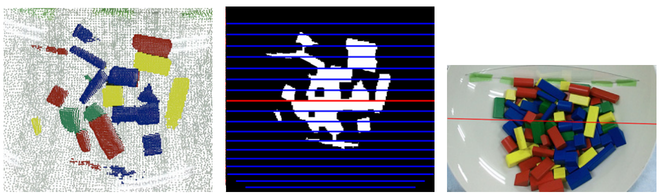
(a) Sideways motion. (b) Align to the bottom.

Fig. 13. Sweep motion model.



(a) Objects above the sweeping motion. (b) Objects to be removed by the motion. (c) The selected motion by Criterion 1.

Fig. 14. Sweeping motion planning by criterion 1.



(a) Objects above the sweeping motion. (b) 3D points to be removed by the motion. (c) The selected motion by Criterion 2.

Fig. 15. Sweeping motion planning by criterion 2.

2) Planning a sweep motion: The variety of sweeping motions is huge and it is difficult to consider all possible motions. Considering the specific shape of the container (i.e., sphere), we limit the sweeping direction to the sideways and align to the bottom of the container (see Fig. 13).

We consider the following two criteria to determine the sweeping motion:

Criterion 1: chooses the sweeping motion which maximizes the currently-occluded volume to be visible by removing objects on the sweeping motion. To do this, the 3D region occluded by each object is calculated as the one between the bottom of the object and the container. Fig. 14 shows an example sweeping trajectory selected by this criterion.

Criterion 2: chooses the sweeping motion which maximizes the point cloud volume to be removed by the sweeping motion. calculate. Fig. 15 shows an example sweeping trajectory selected by this criterion.

We have shown that both criteria can generate almost equally-good sweeping motions but Criterion 2 is better in terms of the simplicity of calculation [9].

C. Experiments

Fig. 16 shows the process of finding a target object (object 4, yellow, indicated in the figures) by executing three sweeping motions and one pickup motion.

We also compared the proposed method with a simple method which removes the object at the top of the pile one after another when the target is not visible. The proposed method was much better in various object placement cases. See [9] for the details of the comparison.

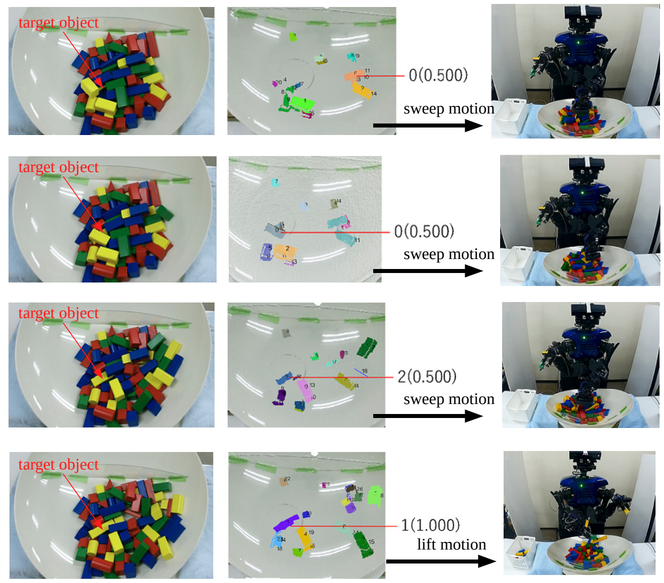


Fig. 16. Example steps of successfully picking up a target object. At each step, description A(B) indicates that object A in the current scene has the highest probability B of being the target object. The target object is heavily occluded at steps 1 and 2, while fairly visible at steps 3 and 4.

V. Summary

This paper have described a new type of manipulation task, finding and picking up a specific object in a pile of objects. To achieve this task efficiently, we adopt sweeping motions, which are effective for moving multiple objects at a time in order to quickly reach the target object. We have shown two examples of sweeping motions based on visual recognition of object placements [1]. The presented methods for hand motion planning are heuristic and developing a more general method is future work.

References

[1] A. ten Pas, M. Gualtieri, K. Saenko, and R. Platt. Grasp Pose Detection in Point Clouds. *Int. J. of Robotics Research*, Vol. 36, No. 13-14, pp. 1455–1473, 2017.

[2] 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. *Int. J. of Robotics Research*, Vol. 37, No. 4-5, pp. 421–436, 2018.

[3] M.T. Mason. *Mechanics of Robotic Manipulation*. The MIT Press, 2001.

[4] J.D. Bernheisel and K.Y. Lynch. Stable Transport of Assemblies by Pushing. *IEEE Trans. on Robotics*, Vol. 22, No. 4, pp. 740–750, 2006.

[5] K. Harada, J. Nishiyama, Y. Murakami, and M. Kaneko. Pushing Manipulation for Multiple Objects. *Trans. of ASME, J. of Dynamic Systems, Measurement and Control*, Vol. 128, No. 2, pp. 422–427, 2006.

[6] M. Gupta and G.S. Sukhatme. Using Manipulation Primitives for Brick Sorting in Clutter. In *Proc. ICRA-2012*, pp. 3883–3889, 2012.

[7] D. Katz, M. Kazemi, J.A. Bagnell, and A. Stentz. Clearing a Pile of Unknown Objects using Interactive Perception. In *Proc. ICRA-2013*, pp. 154–161, 2013.

[8] K. Chikaarashi and J. Miura. Generating Grasping Motion including Sweeping for a Dual-Arm Robot. In *2014 Proceedings of JSME Robotics and Mechatronics Conf.*, 2014. (in Japanese).

[9] W. Miyazaki and J. Miura. Object Placement Estimation with Occlusions and Planning of Robotic Handling Strategies. In *Proc. of AIM-2017*, pp. 602–607, 2017.

[10] D.G. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. *Int. J. of Computer Vision*, Vol. 60, No. 2, pp. 91–110, 2004.

[1] See https://youtu.be/0avHVMR2L4g for the video of experiments.