Motion Planning to Retrieve an Object from Random Pile

Shusei Nagato
Graduate School of Engineering Science, Osaka University, 1-3, Machikaneyama, Osaka, Japan
E-mail: nagato@hlab.sys.es.osaka-u.ac.jp

Tomohiro Motoda
Graduate School of Engineering Science, Osaka University

Keisuke Koyama
Graduate School of Engineering Science, Osaka University

Weiwei Wan
Graduate School of Engineering Science, Osaka University

Kensuke Harada
Graduate School of Engineering Science, Osaka University

Abstract

It is challenging to retrieve a target object from a randomly stacked pile by using a robot due to the occlusion of the target object. In this study, we propose a novel retrieval method in which a robot selects the viewpose to observe the occlusion part of the target object using the RGB-D images, and then selects the motion of grasping/dragging to retrieve the object depending on the configuration of the pile. We experimentally confirm that a robot effectively observes a pile with a complex configuration and successfully retrieves a target object.

Key Words: viewpose selection, motion planning, object recognition

1. Introduction

In recent years, robots have been expected to replace human workers in logistics warehouses. However, since many items are usually crammed onto a shelf in a logistics warehouse, it becomes challenging for a robot to pick an item from the shelf. Even if a robot tries to pick an item from the shelf, it may not be able to see the target product since the product is occluded by other items, or even if it can see the target product, its hand may not be able to grasp the product since the products are close together.

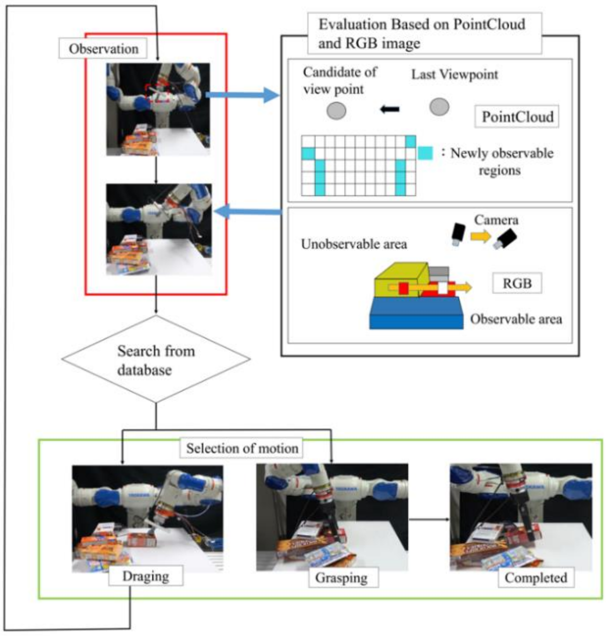


Fig1 Overview of this work

Robotic manipulation of a daily object in clutter has been studied by many researchers [1-6]. Most of the studies have focused on picking the topmost object [1][2], picking an object while avoiding its surrounding objects [3][4], and the push–grasp strategy [5][6]. However, these studies did not consider the problem of directly picking an object on which other objects are placed. If other objects are placed on top of the target object, it becomes difficult for a robot to identify the pose of this object because it is heavily occluded. In addition, even if the pose can be detected, the graspable area of the target object is heavily limited owing to the occlusion. However, if a robot can directly pick the target object from among other objects placed on it, the robot does not need to remove those objects first.

In this study, we propose a motion planning method in which a robot repeatedly captures the image of a pile until it can grab the configuration of objects surrounding the target object (Fig. 1). Once the robot recognizes the configuration, it determines the manipulation strategy for picking the target object. If the robot assesses that the target object is not graspable, it drags the object before grasping it. In our proposed method, the robot first observes the workspace using an RGB-D sensor attached to its arm. At this time, the occlusion of the object is assessed using RGB-D information. When only point cloud information is used, it becomes difficult to reduce the occlusion of a specific target object. When the RGB image information is used, the viewpose can be moved to reduce the occlusion of a specific target object. Based on this result, we have searched for an effective grasping posture from the database to grasp the target object. If no effective grasping posture is found, the target object is dragged in the direction to reduce the occlusion by assessing the occlusion using RGB image features. After the extraction, we observe the object again and repeat the process until the target object is retrieved using the grasping operation. In this way, the target object can be retrieved even when it is difficult to recognize the state of the object or grasp the target object.

2. Related Works

Changjoo et al.[7] proposed a path generation method to retrieve a target object in a cluttered environment by once relocating objects out of the workspace where they introduced a graph expressing the collision possibility among objects to find a path. Sang et al. [8] used this method to retrieve the target object by determining the order of the objects to be relocated. However, in these studies, the target object cannot be retrieved when the objects cannot be relocated.

As for the research on planning a viewpose of vision sensor, Harada et al. [9] proposed a method to determine the sensor pose that maximizes the visibility of piled objects stored in a box. In addition, Motoda et al. [10] used this method to maximize the visibility [9] as an indicator to understand the pile's state. Although the viewpose selections used in these studies can reduce the occlusion included in the image of the pile, they are not always efficient because they are not optimal for observing the target object.

On the other hand, this research proposes a novel approach where a robot selects the grasping and dragging motions to select the target object to retrieve the target object from the pile depending on the situation of the pile. In addition, viewpose selection to reduce the occlusion of the target object is performed using both an RGB image and point cloud.

3. Proposed Method

This study proposes a motion planning method for a robot to grasp and retrieve a single target object which shape is known in advance. We use a dual-arm robot where one arm is equipped with a sensor to acquire RGB-D images from an arbitrary viewpose and a finger to drag an object, and the other arm is equipped with a two-fingered gripper to grasp the target object.

Fig. 1 shows the outline of our proposed method. First, we capture the RGB-D image of the target object by selecting a viewpose that can focus on its occluded part. Then, according to the observation results, a robot selects either the grasping or the dragging operation. If the dragging operation is selected, then the above process is repeated until the robot retrieves the target object by selecting the grasping motion. In the following sections, we describe the details of this method. We first explain the criteria used to select the action according to the state of the pile. Next, we explain the method used to recognize the state of the target object using the RGB images. Finally, we explain the selection of viewpose for better observation of the target object.

3.1. *Manipulation selection based on state*

Fig. 2 shows a sequence of events that leads the target object to be extracted. First, the occlusion of the target object is detected based on the feature matching of the RGB images of the pile. Next, we select a viewpose with which the robot can observe the occlusion existing in the pile's RGB image [9]. Subsequently, we search for an effective grasping pose from the grasping pose database that was created in advance of executing the planner where the grasping pose database stores the pose of the hand relative to the object to stably grasp the object. If a grasping pose is successfully found, the target object is grasped using the identified grasping pose. If the target

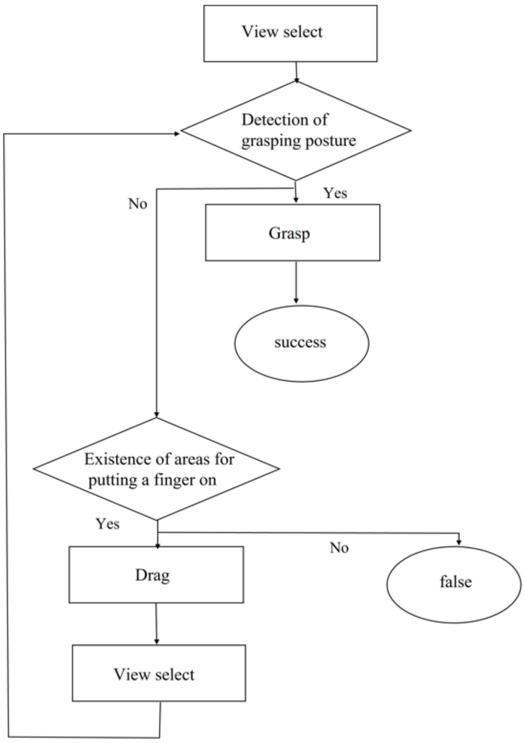

View select
Detection of grasping posture
Yes
No
Grasp
success
Existence of areas for putting a finger on
Yes
No
Drag
false
View select

Fig.2 Flowchart of the proposed method

object is successfully retrieved, the task is completed. On the other hand, if a grasping pose cannot be found, we check whether or not there is enough area on the target object to place a robot's finger. If there is no enough area to place a finger, the task fails; however, if a finger can be placed on the target object, the robot performs the dragging operation. After the dragging operation, the occlusion of the target object is detected again based on the feature matching. If the occlusion area is still larger than the threshold, the observation is performed using the same aforementioned method. Otherwise, the target object is observed with the current viewpose. Based on the observation results, we repeat the sequence of operations until the robot successfully extracts the target object.

3.2. *State recognition of target object*

To extract the target object from the pile, we need to know the occlusion of the target object. As shown in Fig. 3, we acquire the RGB image of the pile using the current viewpose and extract the local features of the image by

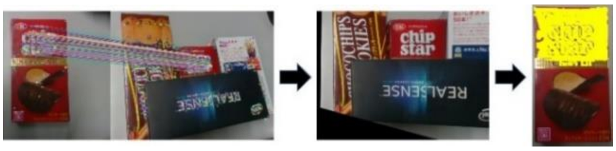


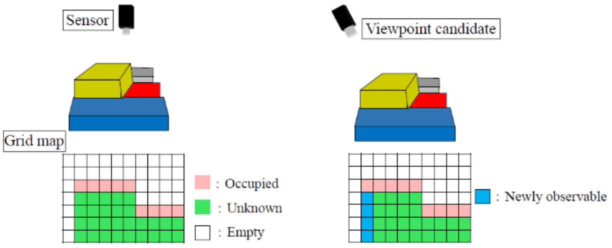
Fig. 3 Flow of state recognition

using A-KAZE [11], and the image of the target object that we have prepared in advance. Next, feature matching is performed between two images, and the corresponding points included in the two images are obtained. In this process, random sample consensus (RANSAC) [12] is used to eliminate matching with outliers. Next, we perform the homographic transformation to align the target object in the original and its transformed images. The image in the center of Fig. 3 shows the result of the homography transformation. Subsequently, a match is detected based on the average of RGB values among nine pixels, including the pixels of interest and neighboring pixels. This is shown in the right figure of Fig. 3, where the area which is not drawn in yellow is occluded.

3.3. *Viewpose*

In this study, we use both RGB and point clouds where RGB is used to detect the target object while point cloud is used to identify occlusions in the pile’s image. This information is used to reduce the occlusions of the target object caused by other objects and move the sensor to the position where the target object is less occluded. The viewpose is selected from multiple candidates explained in the following.

We now explain how to construct candidates of viewposes and how to select one from multiple candidates. The view direction is determined such that the camera faces the center of the table. We assume a regular polyhedron at the center of the table and a set of lines passing through both the center of each face and the center of the polyhedron. We assume that the vision sensor is placed along this line with facing the center of the table. Here, the position of the vision sensor is selected to minimize the occlusion of the target object. First, the workspace is divided into three-dimensional grids [13], and the occupancy grid map [9] is created for each grid based on the point cloud information acquired with the current viewpose. The next position of the vision sensor is selected such that the largest number of



(a) (b)

Fig.4 Next viewpose exploration by points clouds: (a)State of grids by last observations, and (b) Evaluation method in viewpose candidates

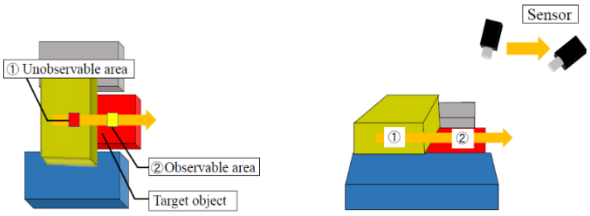
unobserved grids can be observed as shown in Fig. 4.

If only the point cloud is used, the effect of viewpose selection minimizing the occlusion of a specific object is limited[9]. Therefore, we select a viewpose making the occlusion of only the target object be visible by using RGB information. First, we determine the occlusion of the target object using the current viewpose. In this way, the center of the observable and unobservable part of the target object can be identified. Let the center of the observable and unobservable parts of the target object be O and A, respectively. In addition, let the ith horizontal position of the viewpose candidate be $B_i = (b_{xi}, b_{yi})$ (i = 1,2,⋯,n). In this case, it is expected that the unobserved part can be observed by moving the sensor following the vector from the unobserved part to the observed part with imposing the limitation on the viewpose candidates:

$$distance(O, A, (b_{xi}, b_{yi})) < l \tag{1}$$

$$\left| arccos \frac{\overrightarrow{OA} \cdot \overrightarrow{OB_i}}{|\overrightarrow{OA}||\overrightarrow{OB_i}|} \right| < \theta \tag{2}$$

where the $distance(O, A, (b_{xi}, b_{yi}))$ denotes the Euclidean distance between the line including both O and A and the ith position of the viewpose candidate. Equation (1) limits the viewpose candidates by the distance between the line connecting the two unobserved points and the observed point, whereas Equation (2) limits the direction of movement by the arrow shown in Fig. 5(a) where an example of movement is shown in Fig. 5(b). The overall flow of the shifting position of the vision sensor is explained in the following. The initial viewpose is selected randomly from viewpose candidates. Feature matching by A-KAZE is performed from this viewpose. If the number of matches is small, it is assumed that the target object has not been found, and feature matching is performed again by using a viewpose with less occlusion based on the point cloud information. We assume that the target object is found when the number of matches exceeds a specific threshold, and the target object is directly observed. From the method shown above, the occlusion of the target object is checked, and the visible area can be obtained. When the observable area is larger than 80%, the viewpose selection is terminated, and action selection is performed. When more than 80% of the target object cannot be observed, the viewpose is selected based on the evaluation value in Equations (1) and (2).



(a) (b)

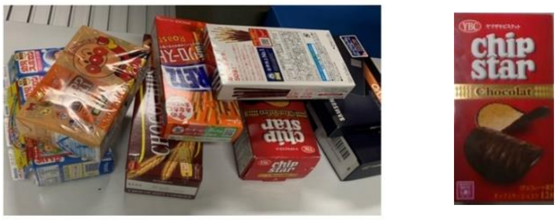
Fig.5 Exploration of next camera position by RGB image: (a)Determining the direction of the following viewpose, and (b) example of viewpose selection

4. Experiments

In this section, we describe the environment used in the experiments. We used a Motoman-SDA5F manufactured by Yaskawa Electric Corporation. For observation, we attached the Intel Realsense SR305 at the tip of the left hand. The left hand was used for the dragging operation, and the right hand was equipped with Robotiq 2F-140 Adaptive Grippers for the grasping operation. A workspace was set in front of the robot, and various objects were placed around the target object whose size and appearance were known.

4.1. *Results of the target object extraction*

Figs. 7(a)-(f) show the results of the proposed method In Fig. 7(a), the observation is performed by randomly selecting a viewpose from multiple candidates. The observed target object is shown in Fig. 7(b), where occlusion is evaluated by the RGB image. Since less than 80% of the target object was observed, the point cloud is



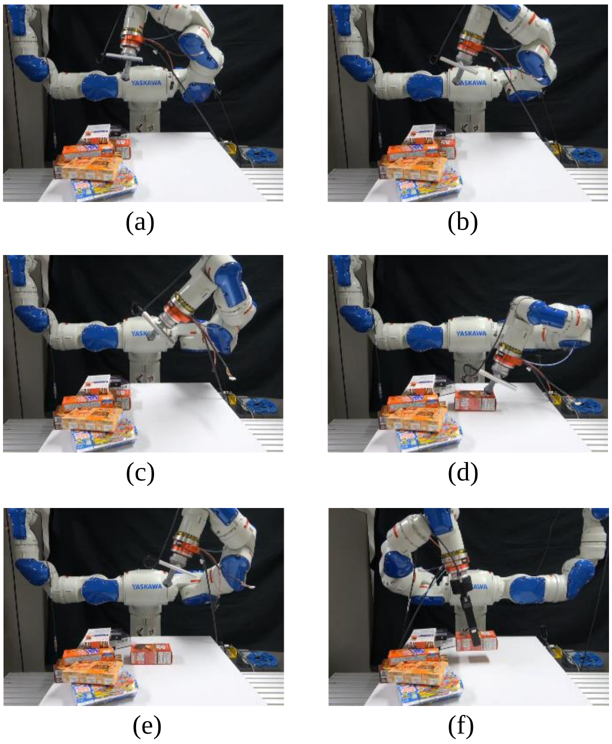
(a)Situation (b) Target object

Fig.6 Experimental environment

acquired, and the occlusion in the workspace is calculated. In Fig. 7(c), the target object is observed from the perspective of reducing the occlusion. There is no grasping posture in the initial state; therefore, the dragging operation is performed from the observation result of the current viewpose. In the dragging motion, the boundary between the observable and non-observable points is determined based on the result of the occlusion assessment; the finger is placed at the center of the observable point, and the dragging motion is performed in the direction perpendicular to the boundary. Fig. 7(d) shows the actual dragging operation. Because the position of the target object changes depending on the dragging, the viewpose candidates are calculated again. In Fig. 7(e), the occlusion of the target object is assessed by observing it again, and the observable area is more than 80%. Therefore, there is no need to reduce the occlusion area, and the viewpose selection is terminated. The grasping motion is performed as shown in Fig. 7(f). Because the grasping posture exists, the retrieval of the target object is completed.

4.2. *Consideration*

As shown in the results, we succeeded in grasping and extracting the target object from a pile using the proposed method. In this study, we confirmed that we could observe the target object using RGB-D information to position the sensor to observe the part of the target object that could not be observed using the previous viewpose. The experiment showed that it is possible to retrieve the target object after dragging it. However, the overall success rate was approximately 20%. This was because the estimation of the target object’s occlusion was sometimes incorrect. Because only RGB information was used to estimate the occlusion of a target object, incorrect results were obtained when an object with an RGB value close to that of the target object was an obstacle or when the RGB value of the target object differed from that of the original image because of the influence of light condition. In the future, we improve the accuracy of the occlusion assessment of the target object using not only RGB images but also depth information.



(a) (b) (c) (d) (e) (f)

Fig.7 Flow of action: (a) Observation from initial viewpose, (b) Observation from above, (c) Observation of the occlusion area, (d) Drag, (e) Observation from initial viewpose, (f)Grasp

5. Conclusion

In this study, we proposed a method for selecting a viewpose to reduce the occlusion of a target object using RGB-D information of the pile. According to the observation results, we confirmed that the target object could be retrieved by combining the dragging and grasping motions where the oject could not be retrieved only by using the grasping motion.

As for future work, we plan to improve the accuracy of occlusion assessment by using RGB image by simuntaneously using the point cloud information. In addition, we estimate the posture of the target object, assess obstacles based on the observed information, and verify whether the target object can be retrieved even in an unknown environment. Furthermore, we plan to determine the direction and amount of dragging by considering the situation of objects piled up around the target object.

References

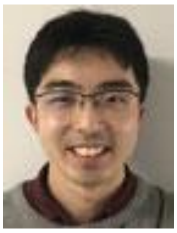
1. H. Zhang, X. Lan, X. Zhou, Z. Tian, Y. Zhang, and N. Zheng, "Visual Manipulation Relationship Network for Autonomous Robotics," in Proc. of 2018 IEEE-RAS 18th Int. Conf. on Humanoid Robots (Humanoids), 2018, pp. 118--125.
2. Y. Domae, H. Okuda, Y. Taguchi, K. Sumi, and T. Hirai, “Fast graspability evaluation on single depth maps for bin picking with general grippers,” in Proc. of IEEE Int. Conf. on Robotics and Automation (ICRA), 2014, pp. 1997--2004.
3. D. Berenson, S. S. Srinivasa, D. Ferguson, A. Collet, and J. J. Kuffner, "Manipulation planning with Workspace Goal Regions," in Proc. of IEEE Int. Conf. on Robotics and Automation (ICRA), 2009, pp. 618-624
4. D. Berenson, S. S. Srinivasa, D. Ferguson, and J. J. Kuffner, "Manipulation planning on constraint manifolds," in Proc. of IEEE Int. Conf. on Robotics and Automation (ICRA), 2009, pp. 625-632
5. L. Berscheid, P. Meißner and T. Kröger, "Robot Learning of Shifting Objects for Grasping in Cluttered Environments," in Proc. of IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS), 2019, pp. 612-618.
6. E. Huang, Z. Jia and M. T. Mason, “Large-Scale Multi-Object Rearrangement,” in Proc. of Int. Conf. on Robotics and Automation (ICRA), 2019, pp. 211-218.
7. C. Nam, J. Lee, S. Hun Cheong, B. Y. Cho, and C. Kim, “Fast and resilient manipulation planning for target retrieval in clutter,” in Proc. of IEEE Int. Conf. on Robotics and Automation (ICRA), 2020, pp. 3777--3783.
8. S. Hun Cheong, B. Y. Cho, J. Lee, C. Kim, and C. Nam, “ Where to relocate?: Object rearrangement inside cluttered and confined environments for robotic manipulation,” in Proc. of IEEE Int. Conf. on Robotics and Automation (ICRA), 2020, pp. 7791--7797.
9. K. Harada, W. Wan, T. Tsuji, K. Kikuchi, K. Nagata, and H. Onda. Experiments on learning based industrial bin-picking with iterative visual recognition,” Industrial Robot: the international journal of robotics research and application, Vol. 45, No. 4, 2018, pp. 446--457.
10. T. Motoda, W. Wan, and K. Harada. Probabilistic action/observation planning for playing yamakuzushi,” in Proc. of IEEE/SICE International Symposium on System Integrations (SII 2020), p. 150--155,2020.
11. P. Alcantarilla and T. Solutions, “Fast explicit diffusion for accelerated features in nonlinear scale spaces,” in British Machine Vision Conference (BMVC), 2013.
12. R. Raguram, O. Chum, M. Pollefeys, J. Matas, and J. Frahm. Usac: A universal framework for random sample consensus. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 35, No. 8, 2013.
13. K. Nagata, T. Miyasaka, D. N. Nenchev, N. Yamanobe, K. Maruyama, S. Kawabata, and Y. Kawai, “Picking up an indicated object in a complex environment,” in Proc. of IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, 2010, pp. 2109--2116

Authors Introduction

Mr. Shusei Nagato

He received the BS degree from Engineering Science Department of Osaka University in 2021. From 2021, he has been a master student in Graduate School of Engineering Science, Osaka University.

Mr. Tomohiro Motoda



He received the MS degree from Graduate School of Engineering Science, Osaka University in 2020. From 2020, he has been a Ph.D. student in Graduate School of Engineering Science, Osaka University.

Prof. Keisuke Koyama

He received the Ph.D. degree from Graduate School of Informatics and Engineering, The University of Electro-Communications in 2017. During 2017-2019, he worked as a Project Assistant Professor, Graduate School of Information Science and Technology, The University of Tokyo. From 2019, he has been working as an Assistant Professor, Graduate School of Engineering Science, Osaka University.

Prof. Weiwei Wan

He received the Ph.D. degree from Graduate School of Information Science and Technology, The University of Tokyo in 2013. During 2013-2015, he was a postdoc researcher at CMU. During 2015-2017, he was a research scientist at AIST. From 2017 he has been an associate professor at Graduate School of Engineering Science, Osaka University.

Prof. Kensuke Harada

1. He received Ph.D. degrees in Mechanical Engineering from Kyoto University in 1997. During 1997-202, he worked as a research associate in Hiroshima University. During 2002-2016, he worked in AIST. From 2016, he has been working as a Professor at Graduate School of Engineering Science, Osaka University.