An Object Acquisition Based on Human-Robot Cooperation

Kota Ito
Graduate of Engineering, Kyushu Institute of Technology, Sensui-cho Tobata-ku
Kitakyushu, 804-8550, Japan

Masuhiro Nitta, Seiji Ishikawa[†], Joo Kooi Tan
Faculty of Engineering, Kyushu Institute of Technology, [‡]N&N Inc. Sensui-cho1-1 Tobata-ku
Kitakyushu, 804-8550, Japan
E-mail: {ito-kota ,etheltan}@ss10.cntl.kyutech.ac.jp
http://lab.cntl.kyutech.ac.jp/~etheltan/

Abstract

In this paper, we propose a human-robot cooperative system to support shopping refugees. In the system, a robot acquires an object specified by a person in a distant site. The normal vector is calculated from the depth image, and the region is segmented using GBS on an RGB image. The two obtained clues are used to accurately detect the position of the specified object. The effectiveness of the proposed method was verified by experiments.

Keywords: Region Division, Normal Vector, Newton-Raphson Method, GBS

1. Introduction

According to the report by Ministry of Internal Affairs and Communications in 2017, the number of shopping refugees in Japan has been increasing in recent years; as of 2014, the number of shopping refugees in Japan was about 7 million. It also increased by about 1 million people in the six years between 2008 and 2014.[1] The term "shopping refugees" refers to "the people who have difficulty in making daily purchases such as groceries". The number of shopping refugees is on the rise, making it necessary to take countermeasures.

According to a survey by the Ministry of Economy, Trade and Industry (METI) in 2012, the market size of the Japanese robotics industry is expected to expand more than tenfold between 2012 and 2035.[2] Among them, robots in the service field are expected to account for about 51% of the total robot market in Japan by 2035. Looking at the world, the market for household and personal service robots is growing at a high rate of 16.3 million units in 2018, a 59% increase over the previous year, and the service robot market is also expanding year by year, doubling in size.[3] The current robot market is dominated by industrial and manufacturing robots used in limited locations, but service robots that help people in their daily lives are expected to increase in the future. In this paper, we propose an object grasping robot system for assisting shopping refugees among service robots.

In previous researches, various robots have been proposed to take the place of humans for object grasping [4-8]. However, these methods do not deal with a large number of unknown objects, such as methods for grasping objects at specific locations[4], methods limited to specific objects or environments[5,6], or methods that require the robot to learn unknown objects[7,8]. The proposed method is an extension of the method of Sato et al[9].

This paper proposes a method of object grasping by a human-robot cooperative system using an RGB-D camera. The proposed method detects the position and pose of an object specified by a person from two pieces of information: the normal vector calculated from the depth image and the result of region segmentation obtained using Graph Based Segmentation (GBS)[10] for RGB images. In the proposed method, a person instructs a robot to go to a specified site. The person selects an object he/she wants on the images sent from the robot. Then the robot knows which object the person selected and it acquires the object. The proposed system is adaptable to unknown environments and objects.

2. Outline of the Proposed System

In this section, an outline of the proposed human-robot cooperative system is described. In the system, an instruction screen, as shown in **Fig. 1**, is displayed to the user from the image transmitted to the user from a robot in a remote location. The user instructs the robot to move by touching the instruction screen, and the robot moves around the site. Depending on the area touched by the user in the instruction screen, the robot's behavior is determined as follows;

(i) Original image area: Move forward or object selection,
(ii) Green area: Retreat,
(iii) Blue area: Turn left,
(iv) Red area: Turn right,
(v) STOP button: Stop.

When the original image area is touched, it judges whether to move forward or select an object based on the average distance from the surrounding area of the touched pixel. If the selected object is found, the object is separated into regions based on the coordinates touched by the user. The details of region segmentation are described in Section 3.

3. Region segmentation of an object

In this section, we describe a method of region segmentation of user-specified objects. The depth image used in this method is characterized by the fact that the distance of the object boundary is often not obtained. Therefore, if we perform region segmentation using only distance information, the estimated object region will be smaller than the actual object. The proposed method can accurately detect objects even in the area where it is difficult to obtain distance information, such as object boundaries, by using two types of information; normal vector information from depth images and region segmentation results from RGB images.

Fig. 1. Robot motion instruction screen

3.1. *Normal vector*

In the proposed method, the normal vector is first calculated for each pixel in the entire depth image. Then, the object region is expanded by calculating the inner product of the normal vector of the pixel specified by the user and the normal vectors of its surroundings[11]. In this section, we describe a method of estimating the candidate region of an object using normal vectors.

3.1.1. *Calculating normal vector*

To calculate the normal vector, we use the position vectors of the three neighboring points of the pixel of interest for all pixels in the depth image to calculate the normal vector as shown in the following equation.

$$\boldsymbol{n} = (\boldsymbol{b} - \boldsymbol{a}) \times (\boldsymbol{c} - \boldsymbol{a}) \quad (1)$$

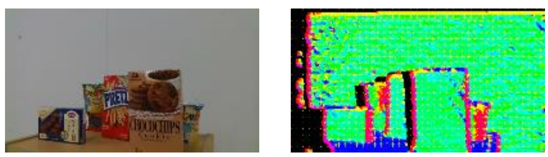
Then, the length of the calculated normal vector is normalized to 1 by the following equation.

$$\hat{\boldsymbol{n}} = \boldsymbol{n}_G / \|\boldsymbol{n}_G\|_2 \quad (2)$$

The result of calculating the normal vectors for all pixels in the depth image is shown in **Fig. 2**.

3.1.2. *Region extension*

Using the normal vectors calculated in 3.1.1, we expand the candidate object region. In the region expansion, an



(a) (b)

Fig. 2. The result of normal vector calculation: (a) An input image, (b) obtained normal vectors.

$n \times n$ pixels region centered at the coordinates obtained from the user is set as the initial object region. Then, for each of the 8-neighbor pixels of the object region, it is determined whether the pixel is an object region or an edge region.

For the judgement, the inner product of the normal of the center pixel and the normal of the pixel which is p pixels apart from the center pixel is calculated in each of the 8 directions. The inner product obtained in each direction is then used to perform thresholding as shown in the following;

$$\delta_{min} = \min\{\delta_N, \delta_{NE}, \cdots, \delta_W, \delta_{NW}\}$$
$$L = \begin{cases} 1 & \delta_{min} > \delta_{th} \\ 0 & otherwise \end{cases} \quad (3)$$

The pixels with L=1 are judged as the pixels in the object region, whereas the pixels with L=0 are judged as those in the edge region.

In the proposed method, the new pixel identified as the pixel in the object region is chosen as the next search point. The region expansion is iterated until all the object regions are surrounded by edge regions.

3.2. *Region correction using color information*

The object candidate regions estimated using normal vectors in Section 3.1 are corrected using the color information of the image. In the proposed method, GBS is performed on RGB images and the obtained segmentation results are used for region correction.

3.2.1. *Graph based segmentation (GBS)*

In the proposed method, GBS is used as a method for segmenting RGB images. GBS is a method of dividing an image into multiple regions by grouping the regions into pixels that have similar pixel values. In the method, GBS is used for the segmentation because it is fast and the parameters for the likelihood of region segmentation can be set.

3.2.2. *Region correction*

The object candidate regions estimated in Section 3.1 are corrected using the segmentation results obtained by GBS described in Section 3.2.1. Since the RGB-D camera used in this method has the characteristic that it is difficult to obtain the distance information of the object boundary, GBS is performed on n pixels around the object candidate region estimated in Section 3.1. By limiting the range of the GBS, the processing time can be reduced and the number of false estimates can also be reduced. Then, for each region created by GBS, we calculate the percentage of the region occupied by the candidate object region estimated in Section 3.1, and if the percentage is greater than a threshold, we estimate the region as the object region.

4. Object Grasping

In the proposed method, the object is grasped based on the estimated position and posture information of the object after the object is divided into regions specified by the user. To estimate the position of an object, an $n \times n$ rectangle is placed at the center of gravity of the detection area, and the average of the 3D positions is calculated. The calculated 3D position is then converted from the camera coordinate system to the robot arm coordinate system to estimate the position of the object.

For object pose estimation, the object pose is obtained from the normal vector of the center of gravity pixel in the detection area. From the estimated position and posture of the object, the joint angle of the robot arm is determined by solving the inverse kinematics problem using the Newton-Raphson method, and the object is grasped.

5. Experiment

Experiments were conducted to show the performance of the proposed method. The camera used in the experiment was an Intel RealSense D415. The camera was placed in front of the robot.

5.1. *Region segmentation experiment*

The first experiment was an object segmentation experiment. In the experiment, we performed region segmentation on the images taken from three different viewpoints for three types of objects with different arrangements and evaluated the accuracy. For objects A and C, the whole shape is visible in the image, but object B is partially hidden by other objects. The accuracy of region segmentation is evaluated using the F value.

As a result, the F value was 90.9% for object A, 77.7% for object B, and 77.5% for object C.

An example of the experimental results is shown in **Fig. 3**. Figure 3(a) is an original image, Fig. 3(b) shows

(a) (b)

(c) (d)

Fig. 3. An experimental result on region segmentation: (a) Input image, (b) normal vectors, (c) after correction, (d) the result.

Fig. 4. Scenes on object grasping experiment

the object candidate region estimated by the normal vector, Fig. 3(c) shows the result of correction using color information, and Fig. 3(d) shows the result on the object region detection. Note that the yellow frame shows the detected object.

5.2. *Object grasping*

The second experiment was object grasping. A user gave instruction to the robot to move and acquire the object specified by the user at two different locations. In the object grasping, the user controlled the robot from a remote location. The experiments were conducted five times for a total of ten object grasps, and the accuracy was evaluated.

As a result of the experiment, the success rate of object grasping was 80% in five experiments. The experimental scene is shown in **Fig. 4**.

6. Conclusion

In this paper, we proposed a method of object acquisition by a human-robot cooperative system using an RGB-D camera. The effectiveness of the proposed method was verified by experiments. In the object region segmentation experiment, the *F* value was 82.0%, whereas, in the object grasping experiment, the success rate of object acquisition was 80%. In the future, we aim at developing a system that can handle complex environments with obstacles, such as an actual store.

References

1. Ministry of Internal Affairs and Communications: "Fact-finding survey on measures for shopping refugees", pp.2-4, 2017. (In Japanese)
2. Ministry of Economy, Trade and Industry: "Market Trends in the Robotics Industry", pp.17-19, 2012. (in Japanese)
3. Japan Machinery Federation: "FY2019 Research Report on the Promotion of Robotics Industry and Technology", pp.18-19, 2020.
4. H. Masuta, H. Lim, T. Motoyama, K. Koyanagi, T. Oshima: "Invariant perception for grasping an unknown object using 3D depth sensor", *2015 IEEE Symposium Series on Computational Intelligence*, pp.122-124, 2015.
5. S. Takaki, J. K. Tan, S. Ishikawa: "A human care system by a mobile robot employing cooperative objects recognition", *Proceedings of the 2017 IEEE Region 10 Conference*, pp.1148-1152, 2017.
6. J. Stückler, R. Steffens, D. Holz, S. Behnke: "Efficient 3D object perception and grasp planning for mobile manipulation in domestic environments" *Robotics and Autonomous Systems*, Vol.61, No.10, pp.1106-1115, 2013.
7. K. Suzuki, Y. Yokota, Y. Kanazawa, T. Takebayashi: "Online self-supervised learning for object picking: detecting optimum grasping position using a metric learning approach", *2020 IEEE/SICE International Symposium on System Integration (SII)*, pp.1-8, 2020.
8. E. Maiettini, G. Pasquale, L. Rosasco, L. Natale: "Interactive data collection for deep learning object detectors on humanoid robots", *Proceedings of the 2017 IEEE-RAS 17th International Conference on Humanoid Robotics (Humanoids)*, pp.862-868, 2017.
9. Joo Kooi Tan, Akitoshi Sato: "Human-robot Cooperation Based on Visual Communication", *International Journal of Innovative Computing, Information and Control*, Vol.16, No2, pp.543-554, 2020.
10. P. F. Felzenszwalb, D. P. Huttenlocher: "Efficient graph-based segmentation", *International Journal of Computer Vision*, Vol.59, No.2, pp.167-181, 2004.
11. Akitoshi Sato, Joo Kooi Tan, Yuta Ono: "Development of a Human-Robot Cooperative System Based on Visual Information, *Proceedings of SPIE International Workshop on Advanced Image Technology (IWAIT2019)*, 6pages, 2019.

Authors Introduction

Mr. Kota Ito

He received his B.E. degree in Department of Control Engineering in 2020 from the Faculty of Engineering, Kyushu Institute of Technology in Japan. He is acquiring the M.E. in the same University. His research interests are image processing, object detection, robot vision and robotics.

Dr. Masuhiro Nitta

He received his PhD from Nara Institute of Science and Technology. He is now an assistant professor in Faculty of Engineering at Kyushu Institute of Technology in Japan .

Emeritus Prof. Seiji Ishikawa

He graduated from Tokyo University and was awarded BE, ME and PhD there. He is now Emeritus Professor of Kyushu Institute of Technology. He was Visiting Researcher of the University of Sheffield, UK, and Visiting Professor of Utrecht University, NL. His research interests include visual sensing & 3-D shape/motion recovery. He was awarded The Best Paper Award in 2008, 2013 and 2015 from Biomedical & Fuzzy System Association, Japan. He is a member of IEEE.

Prof. Dr. Joo Kooi Tan

She is currently with Department of Mechanical and Control Engineering, Kyushu Institute of Technology, as Professor Her current research interests include ego-motion, three-dimensional shape and motion recovery, human detection and its motion analysis from videos. She was awarded SICE Kyushu Branch Young Author's in 1999, the AROB Young Author's Award in 2004, Young Author's Award from IPSJ of Kyushu Branch in 2004 and BMFSA Best Paper Award in 2008, 2010, 2013 and 2015. She is a member of IEEE, The Information Processing Society, The Institute of Electronics, Information and Communication Engineers, and The Biomedical Fuzzy Systems Association of Japan.