Grasping Pattern Estimation Based on Co-occurrence of Object and Hand Shape

Takuya Kawakami*, Tadashi Matsuo*, Nobutaka Shimada*
*Ritsumeikan University, Shiga, Japan

***Abstract*— There is a close relation between the function of the object and the posture and shape of the human hand grasping it. In this research, we describe how to grasp an object quantitatively by using descriptive parameter space of grasping method (interaction descriptor space) that was acquired by an unsupervised Shift Invariant Auto-encoder. Furthermore, it is assumed that the object is a combination of shape parts having a specific way of grasping, and learns the relationship between the parts and grasping method (grasping pattern). In the experiment using the model obtained by learning, it was confirmed whether the interaction descriptor space can quantitatively describe the similarity or dissimilarity of the grasping pattern. We also estimated the grasping pattern for each part by inputting only unknown objects.**

I. Introduction

Objects to be grasped by human have various functions. Also, when grasping an object, human change the hand shape according to the function of the object[1]. In this paper, we estimate the grasping pattern of objects — focused on how a human approach to objects by using machine learning. Kitahashi, et al. 'S report[2] is a work that tackles the task of estimating the function of objects from visual information, but when a human uses an object, the object is limited to an object accompanying the movement. In this paper, in order to recognize the function by the shape of the object and the hand when grasping the object, an object not accompanied by movement at the time of use is also to be recognized. The final goal is to estimate how to grasp from the shape even if the robot does not register an object in advance when it recognizes the object, and the robot can grasp the object properly.

II. Outline of the Proposed Method

Figure 1 shows the procedure for creating a learning model that recalls how to grasp. In this paper, we describe how to grasp in 30 dimensional descriptor (Interaction descriptor), and the combination of grasping method and object shape is called Grasping pattern. Ultimately, we aim to create a model that can recall the interaction descriptor corresponding to the object when inputting an unknown object image into the learned model.

First, learn the space which the interaction descriptors is mapped from the grasping images (three-channel image consisting of depth image, hand mask image, object mask image.) using Auto-encoder. To create a interaction descriptor, use the Encoder part of the learned Auto-encoder, and use the Decoder part when restoring the grasping image from the interaction descriptor. Next, with the interaction descriptor that is the learning result of the Auto-encoder as a teacher, let only the object learn the relation between the image and interaction descriptor using CNN. Next, we will learn the relationship between only object image and interaction descriptor (learning result of Auto-encoder) using CNN. In this method, in order to cope with fine positional deviation of the object region in the input image, an auto-encoder which is invariant to the displacement of the image called Shift Invariant Auto-encoder[3] is used. Also, since objects are grasped differently for each part, 32 px x 32 px patch images of each channel are cut out from the grasping images of 64 px x 64 px x 3 channels and used for learning.

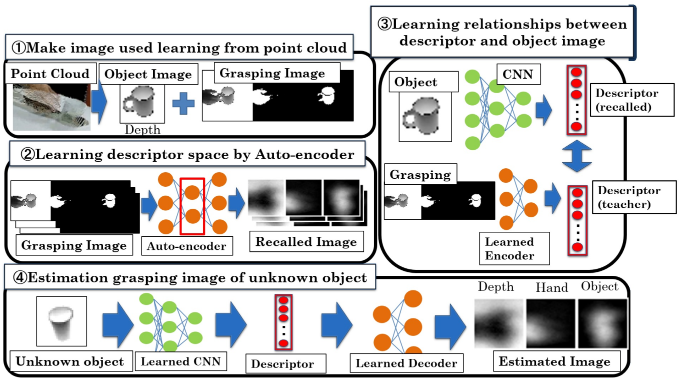

Fig. 1. Outline of the proposed method

III. Experiment on Grasping Pattern Recall

As shown in Fig. 2, the object used is four kinds of objects in four categories: a mug, (no handle) cup, ball, and spray(total of 16 objects). Approximately 100 grasping images were prepared for each object, and using it for learning estimation model.

A. Learning result of Interaction descriptor space by Auto-encoder

First of all, in order to confirm whether the interaction descriptor obtained by learning of Shift Invariant Auto-encoder are divided on the interaction descriptor space for each category, the first principal component, the second principal component of the interaction descriptor space The distribution of the components is shown. From FIG. 3, it can be seen that the same object is plotted at a position which is substantially close even in the interaction descriptor space. Even if they are other objects, if they are in the same category, they are gathered at

Fig. 2. 16 objects used for learning

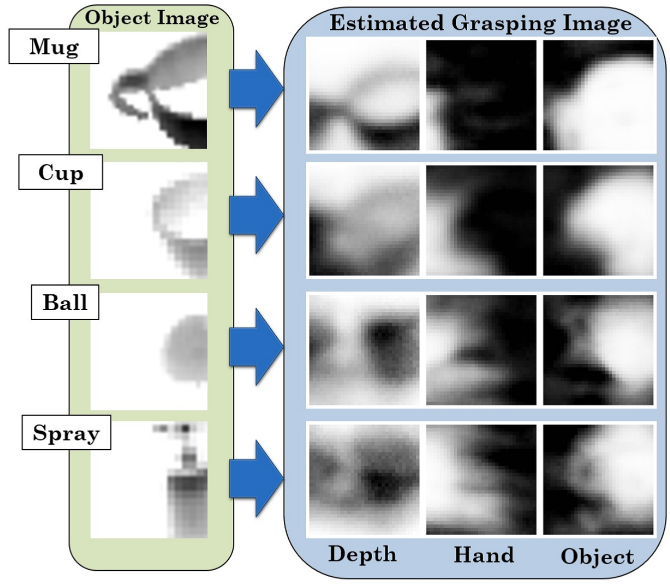

Fig. 4. A grasping image (test image) recalled from the object image

positions close to each other on the first and second principal component spaces.

Also, because the object in the spray category is a mask image that is significantly different from other objects, it is plotted on the space in Fig. 3 at a position distant from the samples in other categories.

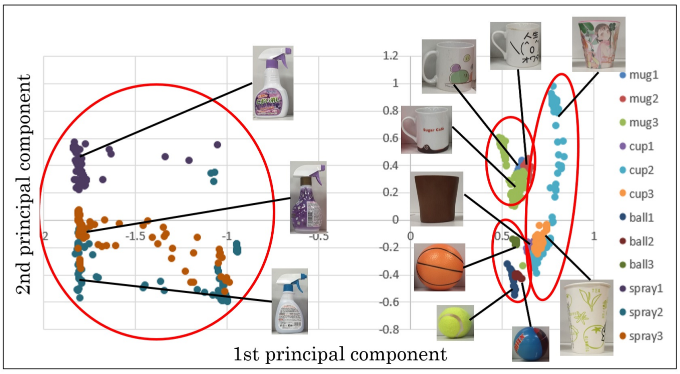

Fig. 3. Distribution of estimated interaction descriptor

B. Recalling result of grasping image by CNN

Next, a restored image created using Decoder from the interaction descriptor recalled in Fig. 4 is shown.

In the mug, a depth image of the edge part is recalled differently from the position of the edge of the object image, but a hand shape image such as a hand shape grasping part of the handle was recalled. The cup is reminiscent of a hand shape like wrapping the part of the torso, and the ball is generally good because it is reminiscent of a hand shape that covers the whole. Regarding spray, although it was well recalled to the extent that the shape of the finger can be recognized, the object mask was significantly different from the object shape of the input.

C. Changing how to grasp by mug handles including or hiding

Finally, in order to confirm whether the grasping method changes for each part even for the same object, two different positions from one object image are cut out as a patch image and recalled. The results are shown in Fig. 5. This time I recalled each grasping image from the patch image including the handle of the mug and the patch image hiding the handle. Regarding patch images with handles, a hand shape that grasps the handle was recalled, and for the patch image hiding the handle, a way to grasp the bottom portion was recalled. Also, comparing the patch image hiding the handle of the mug with a patch image of a similar part of the cup, almost similar hand shapes were recalled. From this result, it was confirmed that the learning model recognizes the handle, which is an important part of the mug, and recalls an appropriate hand shape for the part.

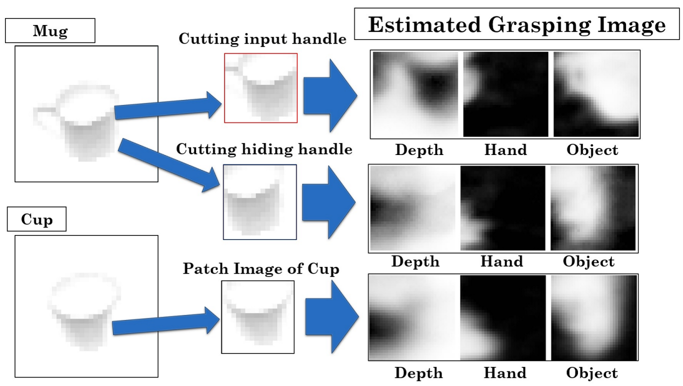

Fig. 5. Difference in remembrance result depending on presence / absence of handle of mug

IV. Conclusion

In this paper, based on the characteristic that the function of an object and the posture and shape of the hand of a human grasping it are closely related, the description of the object function from the hand shape at the time of grasping the object, We propose a method to recall appropriate hand grasping shape from an object by using the Interaction descriptor. Quantitatively describe how to grasp each part of an object part by interaction descriptor acquired by an unsupervised Shift Invariant Auto-encoder from the depth image at the time of grasping the object and the mask image of the hand and object (grasping image). In addition, we created a model that recalls the corresponding grasping hand shape from the object image, and recalled the hand shape corresponding to the part of the object. In the future we will aim to create a probability map of the hand region of the whole object by pasting the hand shapes recalled from patch images of the object.

References

[1] N. Kamakura, *Shape of hand and Hand motion*. Ishiyaku Publishers, 1989.

[2] Tadahiro Kitahashi et.al，“Cooperative Recognition of Human Movements and Objects and Its Modeling”，Information Processing Society of Japan. CVIM，pp.109-116，2005

[3] T.Matsuo et.al，“ Evaluation Function for Shift Invariant Auto-encoder ”，MPR2016，P1-9，2016