Grasp Synthesis of Dishware Using Mean Shape Fitting for a Table Bussing Robot

Jeongho Lee and Dong Hwan Kim

Korea Institute of Science and Technology
5 Hwarang-ro 14-gil, Seongbuk-gu, Seoul 02792, Korea
kape67@kist.re.kr, gregorykim@kist.re.kr

EXTENDED ABSTRACT

1 Introduction

As the robotics industry has developed, robots are frequently encountered in various environments, including coffee barista robots, chicken frying robots, cleaning robots and so on. Recently, robots serving food from the kitchen to the customers' tables are often seen in restaurants. However, the development of table bussing robots, which are designed to clear dishes from the customers' tables and are expected to effectively reduce manpower, is still in the beginning stage. In order to provide stable service, table bussing robots need the ability to robustly detect and grasp a variety of dishes.

Therefore, in this paper, we propose a method for grasp synthesis of dishware using mean shape fitting for a table bussing robot. First we employ off-the-shelf instance segmentation network to detect dishes. Then pose information of the detected dishes is estimated by applying a category-level object pose estimation network with mean shape fitting. Once the object's pose information is obtained, grasp synthesis is performed by analyzing the deformed mean shape with the PCA(principal component analysis) algorithm.

2 Problem description

For robotic grasping, there has been many studies on estimating the pose of an object. Instance-level object pose estimation approaches need whole 3D models for all dishes, which has a limitation that they cannot be applied to unknown dishes. Therefore we apply a category-level object pose estimation approach because it needs a relatively small number of 3D models for each type of dishware. Wen et al. [4] proposed a grasp synthesis algorithm based on category-level pose estimation, which generates grasp information by transferring pre-defined grasp information in the mean shape model. However, it can be failed when the estimated transformation between the detected object and its corresponding mean shape severely stretches or deforms the shape of the object. In order to solve this problem, we propose a grasp synthesis algorithm that computes grasp information by analyzing the transformed mean shape model to fit the detected object data.

3 The applied methods

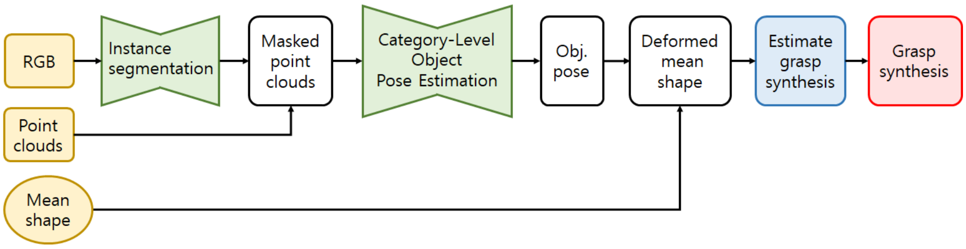
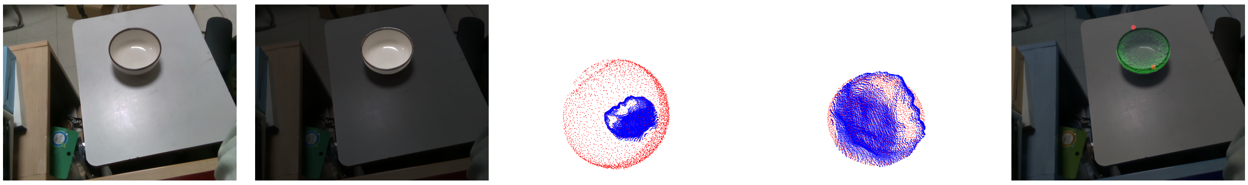


Figure 1: The architecture of the proposed system

In this paper, we propose a method for grasp synthesis of dishware using mean shape fitting, which can be applied to a table bussing robot. Fig.1 shows the overall architecture of our proposed system. First, we employ an off-the-shelf instance segmentation network [5] to detect dishware and their corresponding segmentation masks. For each detected object area, pose information of the object can be efficiently estimated because the background data are removed by the segmentation mask. Therefore, the masked point clouds contain only the 3D data of the object area, and the corresponding mean shape model is transformed to fit them. For mean shape fitting, a canonical representation of objects in the NOCS(normalized object coordinate space)[1] per each category is used, which helps robust estimation of object poses when data are partially missing or occluded.

For generating grasp synthesis, We define the grasp information as two direction vectors and a pair of points similar to [2]: GAD(Gripper Approaching Direction), GCD(Gripper Closing Direction), and GCP(Gripper Contact Points). For simplicity, we constrain the GAD to the top-down direction, which corresponds to the height direction of objects. As the GAD is determined in a top-down direction, the GCD can be obtained using two axes orthogonal to the GAD, which corresponds to the width and length direction of objects. However, when the deformation is processed, as the deformed mean shape can be distorted independently of the object's coordinates, the width and length axes should be obtained. In order to obtain two axes, the deformed mean shape data is projected to width-length plane. Then PCA is used to get width and length axes. The axis of shorter distance along to width or length have become the GCD. Finally, the GCP is determined by combining these pieces of information: the largest height value, the mean value of the longer distance along the two axes, and the minimum and maximum values of the shorter distance along the axes.

4 Experimental results



(a) raw RGB (b) detected dishware (c) before fitting (d) after fitting (e) result of grasp synthesis

Figure 2: Examples of process for our proposed method: (a) raw RGB data from a camera (b) the mask of detected dish using instance segmentation (c) the example of masked point clouds(blue dots) and mean shape data(red dots) before mean shape fitting (d) the result of mean shape fitting (e) the result of grasp synthesis(Red Dots) and pose(green dots) of the object

We experimented with dishware on tables in the real world. In the Fig2, examples are listed in the order of process of our proposed method. When RGB data are extracted from a camera, the 2d-based mask by using instance segmentation. For this, we adopt mask2former[5] which is the SOTA on segmentation tasks. To extract point clouds according to object pixel for estimating object pose, the point clouds are removed correspond to background pixels of 2d-based mask. Before category-level object pose estimation[3] with mean shape fitting, the difference of the pose is quite big between masked point cloud and mean shape shown Fig.2c. However, after mean shape fitting, mean shape can be fit masked point clouds by obtaining pose of the object. By using the pose of the object, we can generate grasp synthesis. The green dots represent point clouds of mean shape and red dots are the GCPs in Fig.2e

5 Conclusion

We proposed grasp synthesis of dishware using mean shape fitting for a table bussing robot. We employed off-the-shelf instance segmentation network to detect dishes. Then pose information of the detected dishes was estimated by applying a category-level object pose estimation network with mean shape fitting. Once the object's pose information is obtained, grasp synthesis was performed by analyzing the deformed mean shape with the PCA algorithm. In the experiments, it showed that the proposed methods could generate grasp synthesis in real world.

Acknowledgments

This work was supported by the Technology Innovation Program and Industrial Strategic Technology Development Program (20018256, Development of service robot technologies for cleaning a table)

References

[1] Wang, He, et al. "Normalized object coordinate space for category-level 6d object pose and size estimation." in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019.

[2] C. Nam et al., "A Software Architecture for Service Robots Manipulating Objects in Human Environments," IEEE Access, vol. 8, pp. 117900-117920, 2020, doi: 10.1109/ACCESS.2020.3003991.

[3] Akizuki et al., "ASM-Net: Category-level pose and shape estimation using parametric deformation." in Proceedings of the British Machine Vision Conference. 2021.

[4] Wen, Hongtao, et al. "TransGrasp: Grasp Pose Estimation of a Category of Objects by Transferring Grasps from Only One Labeled Instance." Computer Vision–ECCV 2022: 17th European Conference, 2022.

[5] Cheng, Bowen, et al. "Masked-attention mask transformer for universal image segmentation." in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022.