Robotic Grasping of Common Objects: Focusing on Edge Detection for Improved Handling

Tomoya Shiba
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan

Hakaru Tamukoh
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan

Email: shiba.tomoya627@mail.kyutech.jp, tamukoh@brain.kyutech.ac.jp
https://www.brain.kyutech.ac.jp/~tamukoh/

Abstract

Grasping objects like plates and cups poses unique challenges for robots because of their irregular shapes and the difficulty of finding reliable grasp points. Traditional approaches often attempt to grasp the object at its center, but this strategy tends to fail for items like plates or cups, whose shapes deviate from simple forms like cubes or spheres. To address this issue, we propose a new method that utilizes AI-powered image analysis to identify the best edges for grasping. Through experiments conducted with a home service robot and a set of YCB objects, we evaluated the effectiveness of our approach compared to conventional methods. The results revealed a significant improvement in the success rate, particularly for objects with prominent edges, such as cups.

Keywords: Object recognition, Dataset, Service robot, Mobile manipulator, RoboCup@Home,

1. Introduction

Autonomous mobile robots have become more prevalent in a variety of industries in recent years, from supermarkets to restaurants [1],[2],[3]. These robots have primarily been employed thus far for transportation-related jobs that need their ability to navigate. Nevertheless, we may significantly improve their functionality by giving them robotic arms, so transforming them into mobile manipulators [4].

The capacity to grip objects is a crucial characteristic that sets mobile manipulators apart from simple transport robots. Using items with predetermined placements and orientations or placing marks where objects should be held are the easiest ways to guarantee a solid grasp. However, because service robots are designed to operate in human-centered settings, they must be able to recognize appropriate grasp points on their own without assistance from others.

Several approaches that integrate depth information with RGB-based object recognition have been put forth to address this problem. For simple geometric shapes like cuboidal items (like boxes) and cylindrical objects (like beverage cans), these methods enable robots to estimate object size and identify grab locations [5]. Additionally, more sophisticated methods have been created by utilizing neural networks that have been trained on specific datasets, such as depth pictures and 3D point clouds [6]. These techniques do have a major disadvantage, though, in that they necessitate a great deal of dataset preparation and 3D annotation, which makes large-scale data collecting extremely time-consuming and impracticable in many situations.

By using an effective grasp point estimate framework that just uses RGB images and does not require 3D point clouds or depth information, our suggested approach, which is shown in (Fig. 1), seeks to get around these restrictions. We can train reliable models while drastically cutting down on the overhead related to conventional data gathering techniques by employing a simulator-driven dataset generation approach.

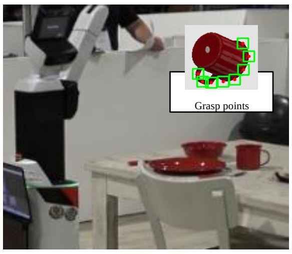

Fig. 1. Select Grasping Points

In this study, we provide a dataset generating method in this study that makes use of a simulator. The approach makes it possible to efficiently create large-scale training datasets by randomly placing different items within a virtual environment and automatically annotating grasp point information. Additionally, we want to show that grasp point estimation may be done without the usage of 3D point cloud data by using object recognition based just on RGB images.

We experiment using YOLOv8 [7] for object recognition in order to test the efficacy of the suggested method and determine how it affects grip accuracy in practical situations. YCB objects (Fig.2) [8], which are frequently used to benchmark grasping tasks, are used for the evaluation.

Fig. 2. YCB Object dataset (kitchen item) [8]

2. Related works

2.1. *Object recognition*

Previous studies have explored methods for home service robots to detect the locations of people and objects and respond accordingly. Research has been conducted on object recognition techniques for tasks such as tidying up and detecting people in home environments. Approaches utilizing YolactEdge [9] and point cloud data [10] have been proposed for estimating object positions and identifying available space.

2.2. *Simulator-Based Dataset Generation*

Studies on simulator-based dataset generation have demonstrated that training solely with simulated data can produce models capable of performing effectively in real-world robotic applications. These findings suggest that properly designed simulation environments can serve as a viable alternative to costly and labor-intensive real-world data collection and annotation [11].

Fig. 3. Dataset Generator

3. Proposed Method

3.1. *Overview*

In this study, we propose an automatic dataset generation and detection system for identifying graspable regions of objects using RGB images. The dataset generation system captures images from multiple angles and automatically annotates them using 3D models of target objects and a PyBullet-based physics simulator [12]. For the detection system, we adopt YOLOv8, which offers high-speed object detection while maintaining a relatively lightweight architecture.

3.2. *Grasp-Point Definition*

Since our method utilizes an object detection model, we define the center of the bounding box as the grasp point. As a result, grasp points are recorded as bounding boxes in the RGB images captured by the simulator, as illustrated in (Fig. 4).

For each target object, grasp points are predefined in its 3D model by identifying key regions suitable for grasping. Specifically, areas such as the rim or handle of a cup are selected based on the 3D mesh structure to ensure the model learns practical grasping locations. During the annotation process, a 2D bounding box is placed around each predefined grasp point, covering 10–30% of the object's total size, allowing the detection model to effectively learn and generalize graspable areas.

Fig. 4. Annotated Grasp Points (Mug / Bowl / Cup)

3.3. *Dataset Generation*

To generate training data, we capture images of objects within a 3D physics simulator using PyBullet , as illustrated in (Fig. 5). Inspired by previous research [11], we apply domain randomization techniques to minimize the gap between simulated and real-world images. Specifically, we vary camera angles, adjust lighting conditions, and change background colors randomly to improve the model's generalization and robustness. These variations ensure that the dataset includes a diverse range of scenarios, helping the trained model adapt better to real-world conditions.

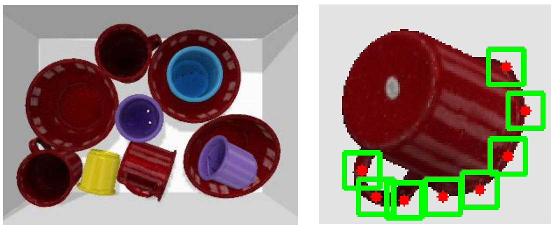

Fig. 5. Simulator-Based Dataset

4. Experiments and Results

4.1. *Recognition experiments*

We conducted an experiment using Toyota's HSR [13][14][15], a mobile manipulator, to verify whether it could correctly select grasp points. The effectiveness of the proposed method was evaluated by performing object grasping.

For the object grasping dataset, we used the YCB dataset. In this experiment, we selected bowls, mugs, and cups as target objects for grasp point selection. The training dataset consisted of 10,000 images, with each image containing at least two overlapping objects. The model was trained exclusively on simulator-generated images. The validation dataset consisted of images captured by the robot's head-mounted camera, showing objects placed on the floor. The results are shown in (Table. 1).

Table. 1. Performance

Object	BBox mAP	Grasp-Point MPJPE[mm]
Mug	75.3 %	28.1 [mm]
Bowl	72.3 %	32.3 [mm]
Cup	81.8 %	12.7 [mm]

4.2. *Grasping experiments*

Additionally, we conducted a comparison with a conventional grasping system. The training settings were identical to those in Experiment 4.1. The conventional system followed the method proposed by Ono et al. [11], which attempted to grasp the center of an object. The robot attempted to grasp three types of objects—bowls, mugs, and cups—placed on the floor in front of it. The objects were positioned randomly in different orientations, and the robot performed 20 grasping attempts for each object. A grasp was considered successful if the robot lifted the object from the floor. The results are presented in (Table. 2).

Table. 2. Grasping Performance Evaluation

Object	**Proposed Method**	Conventional Method	Success Rate Diff
Mug	**18/20 (90%)**	17/20 (85%)	+5%
Bowl	**12/20 (60%)**	5/20 (25%)	+35%
Cup	16/20 (80%)	16/20 (80%)	0%

5. Discussion

To evaluate the grasp point estimation AI, we analyzed both bounding box mean Average Precision (BBox mAP) and Mean Per Joint Position Error (MPJPE), as summarized in (Table. 1). The model performed well in detecting graspable regions across different objects, with BBox mAP values ranging from 72.3% to 81.8%. However, the accuracy of grasp point estimation varied depending on the object shape. The cup had the lowest MPJPE (12.7 mm), indicating high precision, while the bowl had the highest error (32.3 mm), suggesting that grasp points were more difficult to estimate. This difference likely stems from the object's geometry—cups tend to have more well-defined grasp points, whereas bowls, with their curved surfaces, pose a greater challenge.

To further assess how well the model's grasp point predictions translated into actual grasping success, we conducted experiments using a mobile manipulator. The results, shown in (Table. 2), indicate that the proposed method improved grasping success rates for the mug (+5%) and bowl (+35%) compared to the conventional approach. Notably, the success rate for cups remained unchanged at 80%, suggesting that the conventional method was already effective for this particular object type. The large performance gap observed for the bowl highlights the advantage of using AI-based grasp point estimation for objects where a precise grasp is harder to determine.

Overall, these findings suggest that the proposed AI model enhances grasping accuracy, especially for objects with irregular shapes. However, the higher MPJPE for the bowl indicates a potential limitation in predicting grasp points for objects with smooth, less-defined surfaces. Further refinements, such as incorporating additional geometric features or optimizing the dataset to include more varied object poses, may help improve the model's performance in these cases.

6. Conclusion

In this work, we proposed a grasp point estimate method that does not require 3D point clouds or depth information, and instead uses only RGB images. Our methodology maintains good gripping accuracy while drastically reducing the workload associated with manual data collecting and annotation by utilizing a simulator-based dataset generation mechanism.

Results from experiments showed that our suggested approach, especially for structured objects like cups and mugs, was able to accurately identify graspable zones. Our method increased the grip success rate for difficult objects with less distinct grasp points, such as bowls, when compared to traditional grasping strategies. Performance research also showed that the accuracy of grasp point estimate differed based on the shape of the object, suggesting a possible drawback when working with objects that had smooth or uneven surfaces.

Future work will focus on further enhancing grasping performance by integrating additional geometric features and refining dataset diversity. Additionally, real-world deployment and testing will be conducted to validate the model's adaptability beyond simulated environments. We believe that our approach contributes to advancing robotic grasping in home service applications and provides a scalable solution for training grasping models efficiently.

7. Acknowledgements

This research is based on results from a JPNP16007 project commissioned by the New Energy and Industrial Technology Development Organization (NEDO). This research received support from JSPS KAKENHI Grant Number 23H03468 and 23K18495, as well as from JST ALCA-Next Grant Number JPMJAN23F3.

References

1. T. Yamamoto, K. Terada, A. Ochiai, F. Saito, Y. Asahara, and K. Murase, ROBOMECH Journal, 2019.
2. L. Iocchi, D. Holz, J. Ruiz-del-Solar, K. Sugiura, T. van der Zant, Artificial Intelligence, pp. 258-281, 2015.
3. H. Okada, T. Inamura, and K. Wada, Advanced Robotics, 2019.
4. Ghodsian, Nooshin, et al. "Mobile manipulators in Industry 4.0: A review of developments for industrial applications." *Sensors* 23.19 (2023): 8026.
5. O. Khatib,"Mobile manipulators: Expanding the frontiers of robot applications." *Field and Service Robotics*. London: Springer London, 1998.
6. Fang, Hao-Shu, et al. "Graspnet-1billion: A large-scale benchmark for general object grasping." *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2020.
7. Ultralytics “Yolov8 docs”, (2024 12/12 accessed)
8. B. Calli, et al. "The ycb object and model set: Towards common benchmarks for manipulation research." *2015 international conference on advanced robotics (ICAR)*. IEEE, 2015.
9. H. Liu, Soto, R. A. R., Xiao, F., & Lee, Y. J. (2021, May). Yolactedge: Real-time instance segmentation on the edge. In 2021 IEEE International Conference on Robotics and Automation (ICRA) (pp. 9579-9585). IEEE.
10. T. Shiba, T. Ono, and H. Tamukoh, Object Search and Empty Space Detection System for Home Service Robot, The 2023 International Conference on Artificial Life and Robotics (ICAROB2023) OS17-2, Online, February 9-12(10), 2023.
11. T. Ono. "Establishing Fundamental Technologies for Home Service Robots: Data Generation for Sim2Real and Its Application to Pick-and-Place" (Kateiyo Sabisu Robotto no tame no Kiban Gijutsu no Kakuritsu: Sim2Real wo Jitsugen suru Data Seisei to Pick-and-Place he no Oyo). Doctoral Dissertation, 2023. (Japanese)
12. “Bullet Real-Time Physics Simulation”, (2024 12/12 accessed)
13. Yamamoto, T., Terada, K., Ochiai, A., Saito, F., Asahara, Y., & Murase, K. (2019). Development of human support robot as the research platform of a domestic mobile manipulator. ROBOMECH journal, 6(1), 1-15.
14. Hori S, Yutaro I, Kiyama Y, et al. Hibikino-Musashi@Home 2017 team description paper. Preprint. 2017. Avail-able from: arXiv:1711.05457
15. T. Shiba, A. Mizutani, Y. Yano, et al,. “Hibikino-Musashi@Home 2023 Team Description Paper,” arXiv:2310.12650, 2023.

Authors Introduction

Mr. Tomoya Shiba

Tomoya Shiba received his B.Eng. degree from the National Institute of Technology, Kagoshima College, Japan, in 2021, and his M.Eng. from Kyushu Institute of Technology, Japan, in 2023. He is currently a Ph.D. student at the Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology. His research interests include image processing, motion planning, and domestic service robots. He has been actively demonstrating and validating his technologies through the RoboCup@Home competition and is currently conducting research on automatic dataset generation.

Dr. Hakaru Tamukoh

Hakaru Tamukoh received the B.Eng. degree from Miyazaki University, Japan, in 2001, and the M.Eng. and Ph.D. degrees from the Kyushu Institute of Technology, Japan, in 2003 and 2006, respectively. He was a Postdoctoral Research Fellow at the Kyushu Institute of Technology, from April 2006 to September 2007. He was an Assistant Professor with the Tokyo University of Agriculture and Technology, from October 2007 to January 2013. He is currently a Professor with the Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology. His research interests include digital hardware design, neural networks, and home service robots.