Vision and Tactile Pose Identification for Picking a Target without Collision

Saekwang Nam
Dexterous Robotics Group
University of Bristol
Bristol, UK
s.nam@bristol.ac.uk

Loong Yi Lee
Softrobotics Lab.
University of Bristol
Bristol, UK
ly.lee@bristol.ac.uk

Max Yang
Dexterous Robotics Group
University of Bristol
Bristol. UK
max.yang@bristol.ac.uk

Naoki Shitanda
Sch. of Electrical & Electronic Eng.
Kyoto University
Kyoto, Japan
shitanda.naoki.38c@st.kyoto-u.ac.jp

Jonathan Rossiter
Softrobotics Lab.
University of Bristol
Bristol, UK
jonathan.rossiter@bristol.ac.uk

Nathan Lepora
Dexterous Robotics Group
University of Bristol
Bristol, UK
n.lepora@bristol.ac.uk

***Abstract*—Gripper pose is vital in leveraging visuotactile response for successful object grasps. This relationship between pose and tactile feedback may play a large role in successful grasps of objects in clutter, particularly scenarios where deciding the best grasping candidate involves multiple objects and a constrained environment with obstacles. This work presents a bin-picking solution for homogenous objects, using an in-arm camera to seek and approach target objects such that a visuotactile sensor may be safely positioned in a good location. The virtual environment which simulates the system can also be used to test virtual visuo-tactile sensors.**

***Index Terms*—Robot arm, object detection, pose estimation, camera calibration, motion planning**

I. Introduction

Current advancements in visuotactile sensors enable a robot to quickly determine information about a target object through touch. The tactile information generated by contact with an object can serve in reproducing the object's contact shape, hardness, and force exerted to the contact surface. Nevertheless, contact must first be established, and as such, manipulators need to position the gripper or visuotactile sensor in the right location and pose. Picking in cluttered or constrained environments extends this challenge, whereby both vision-guided target selection and subsequent tactile manipulation will need to work together for successful grasps.

This work introduces a system that can autonomously generate a position, posture, and motion path of a robot arm to pick up a single target from a bin containing multiple objects (See Fig. 1(b)). Using YOLOv5, a camera attached near a gripper recognises the target object. Using the depth map captured by the camera, Fast Graspability Evaluation (FGE) algorithm determines the grasping position and gripper orientation (Section II). The calibrated camera coordinate then computes the position and orientation of the target in the system coordinate Section III). Finally, the virtual environment in MoveIt engages in motion planning so that the gripper approaches the object without collisions (Section IV).

II. Choosing a Target Object in Cluttered Environment

A. Object detection

We used the You Only Look Once (YOLO) algorithm to quickly detect and classify objects [1]. The algorithm generates a box containing an object from an RGB image and attach a name on it through classification. In our food-picking system, the pre-trained YOLOv5x model was chosen [2]. Open Images V6-Food [3] and Vietnamese food pictures were used for object detection, while the MAFood121 dataset was used for classification training [4]. The results of our detection of real cookies and fake fried eggs are shown in Fig. 1(a).

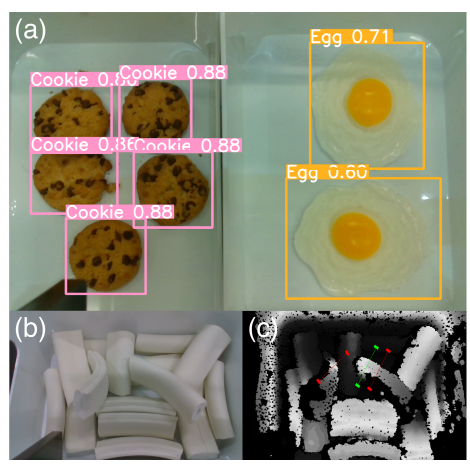

Fig. 1. Examples of object detection and position/pose suggestions. (a) YOLOv5 detects food items given RGB images. (b) Multiple objects are placed in a bin. (c) Given a depth image, Fast Grapability Evaluation algorithm proposes the best position/pose for a gripper (shown in green lines) and the next two candidates (shown in red lines).

B. Generation of grasping position/pose

From a depth map, the FGE algorithm generates multiple candidates for the optimal grasping position and orientation of a tactile sensor [5]. The algorithm calculates grasp-ability in order to determine the top three candidates. Figure 1(c) demonstrates the result of applying FGE to an object-filled bin, while Figure 1(b) shows the RGB image of the same scene. We referred to Xinyi Zhang's code for implementation [6].

III. Calibration for Camera-in-hand System

We mounted a camera to the robot arm's wrist in order to detect the target object and provide the arm with its rotation and translation relative to the robot's origin (H^{ROB}_{OBJ}). The coordinate relationships between the major components used to calculate H^{ROB}_{OBJ} are depicted in Fig. 2. Because H^{EE}_{CAM} is fixed and H^{ROB}_{EE} is provided automatically by the robot arm, H^{ROB}_{OBJ} is easily extracted by only identifying H^{CAM}_{OBJ}. In our research, we utilised the Universal Robots UR10 robot arm and Intel's RealSense camera.

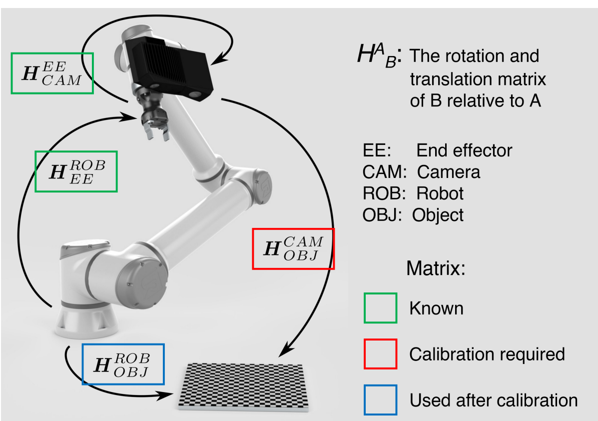

Fig. 2. Brief illustration of the in-hand camera calibration. To precisely predict the pose and location of an object captured from the camera, the H^{CAM}_{OBJ} matrix must be determined. (The primary picture has been captured from [7])

IV. Motion Planning

How to move the robot arm and tactile sensor-attached gripper while avoiding collisions with the local environment is the next challenge for successful object pickup. Particularly, the system with a large arm and gripper will struggle to find a pose to pick up a small object from a tight free-space, as shown in the space of the rack of Fig. 3(a). Therefore, we created a virtual environment in the MoveIt software and instructed it to perform motion planning using the TRAC-IK plugin [8]. This allows us to determine whether the gripper can approach an object without colliding with the environment. The optimal scenario, in the case of Fig. 3, is that no component of the robot system touches the components on the assembly and source tables.

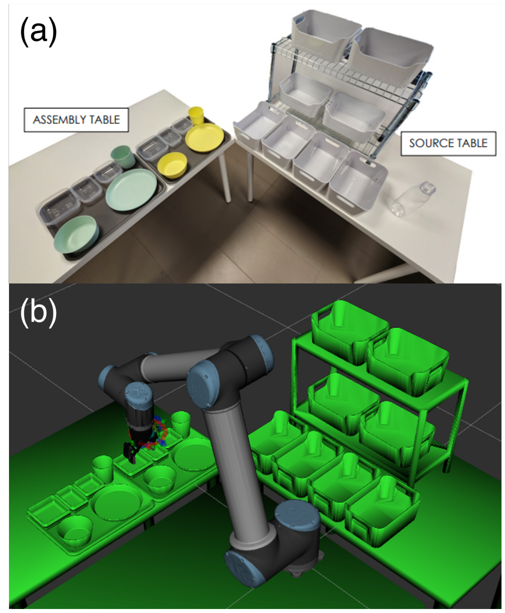

Fig. 3. Our objective is to relocate a target object from the source table to the assembly table. The virtual environment that replicates the actual one assists in designing the motion of the robot arm to avoid collisions with the environment, maximizing the sense-ability of a tactile sensor. (Fig. 3(a) has been captured from [9].)

V. Preliminary Tests and Results

The system was set up for the manipulation challenge at the IEEE International Conference for Soft Robotics (Robosoft) 2023 at Singapore. The environment was as such set up as shown in Fig. 3, with the challenge being pick and place of food items from bins on the source table to specific containers on the source table. Food items include green beans, broccoli, sausages, meatballs, fried eggs, cookies, carrot slices, and a bottle of orange juice. With the motion planning running in the background, the robot was able to navigate to defined image acquisition positions, utilize the detection and grasp pose algorithms, and servo into target objects in the bins. One of the findings from these tests is that the motion planning via the virtual environment would need many physical tolerance modifications such that the end effector can be place in more varied grasp poses safely. Different motion planning control algorithms may also help improve robustness of collision detection. The system also had a limitation in its object acquisition approach where tactile sensing can close the loop to approach objects safely and effectively, as the current method relies on defined primitives to pick objects up after approaching the target grasp pose.

VI. Conclusion and Discussions

This study presents a system in which a gripper-equipped robot arm autonomously identifies the best target object among a collection of objects and determines the optimal location and orientation for successful picking without causing collisions with the surrounding environment. We had the camera calibration to figure out the target's position and orientation relative to the robot's origin using a camera. In addition, the captured RGB images and depth maps are used for object detection and optimal gripper pose estimation. The robot system is able to avoid collisions thanks to the virtual environment and the TRAC-IK kinematics solver.

The next step is integration of visuotactile sensing and tuning the system to take advantage of the sensor morphology in making contact with objects. The virtual environment setup allows possible integration with virtual tactile sensing and training as shown in [10]. In consideration of the fact that the system is still in development, several changes must also be made to improve it. First, the integrated system should be evaluated further to identify unforeseen issues. In addition, the gripper's design would limit its performance. We must conduct comparative performance tests with a number of major grippers. We believe a combination of visual pose estimation and visuotactile sensing can solve cluttered bin picking.

ACKNOWLEDGMENT

The authors would like to thank the technician team at Bristol Robotics Laboratory for their help with hardware development. This research has been supported by the Made Smarter Innovation - Research Centre for Smart, Collaborative Industrial Robotics (Grant number: EP/V00784X/1) and the UKRI Trustworthy Autonomous Systems Node in Functionality (Grant number: EP/V026518/1). We thank Ocado technology for providing UR10 robot arms for this research.

REFERENCES

[1] G. Jocher, "YOLOv5 by Ultralytics," 5 2020. [Online]. Available: https://github.com/ultralytics/yolov5

[2] H.-L. Nguyen and M.-K. Pham, "Meal analysis with Theseus," 7 2022. [Online]. Available: https://github.com/lannguyen0910/food-recognition-meal-analysis-with-theseus-

[3] "Open images dataset," accessed: 2023-03-13. [Online]. Available: https://storage.googleapis.com/openimages/web/index.html

[4] E. Aguilar, M. Bolaños, and P. Radeva, "Regularized uncertainty-based multi-task learning model for food analysis," *Journal of Visual Communication and Image Representation*, vol. 60, pp. 360–370, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1047320319301002

[5] 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 *2014 IEEE International Conference on Robotics and Automation (ICRA)*, 2014, pp. 1997–2004.

[6] "bin-picking-robot," accessed: 2023-03-13. [Online]. Available: https://github.com/xinyiz0931/bin-picking-robot/blob/master/bpbot/binpicking.py

[7] "Hand-eye calibration sollution," accessed: 2023-03-13. [Online]. Available: https://support.zivid.com/en/latest/academy/applications/hand-eye/hand-eye-calibration-solution.html

[8] "Trac-ik kinematics solver," accessed: 2023-03-13. [Online]. Available: https://ros-planning.github.io/moveit_tutorials/doc/trac_ik/trac_ik_tutorial.html

[9] "Robosoft competition 2023 scenarios and rules," accessed: 2023-03-13. [Online]. Available: https://softroboticsconference.org/wp-content/uploads/2023/01/RoboSoft-Competition-2023_v6.pdf

[10] Y. Lin, J. Lloyd, A. Church, and N. F. Lepora, "Tactile gym 2.0: Sim-to-real deep reinforcement learning for comparing low-cost high-resolution robot touch," *IEEE Robotics and Automation Letters*, vol. 7, no. 4, pp. 10 754–10 761, 2022.