

Flexible State-Aware Planning for Robust Object Placement in Home Tidy-Up with Autonomous Mobile Manipulators*

Natee Buttawong¹, Kosei Isomoto¹, and Hakaru Tamukoh^{1,2}

Abstract—This paper aims to achieve flexible placement onto tray-type storage areas using a mobile manipulator. In a previous study, tray-packing was accomplished by preparing object masks in advance before tidy-up and planning the arrangement by matching these masks with the available tray space. However, such an approach faced difficulties in handling unknown objects and postures for which prior information could not be obtained, and was unable to take the current state of the tray into account, resulting in limited flexibility. The proposed system estimates masks for previously unregistered objects from an RGB-D image, verifies and corrects object posture to a default, builds a tray mask from the current state of the tray, and computes collision-aware, space-efficient placements via a 2D irregular packing algorithm. The proposed system performed a tidy-up on five objects placed in various postures, achieving a success rate of 93%. These results indicate improved flexibility, robustness, and practicality for real-world tidy-up compared to the previous system.

I. INTRODUCTION

Japan is facing a demographic shift with a rapidly aging population and a declining birth rate [1], leading to significant labor shortages in various sectors. Automation and robotics have gained attention as potential solutions to address labor shortages and assist elderly people in the future. Human support robots (HSR) [2] have emerged as a promising solution, and several studies have worked to develop and extend its capabilities [3]–[9]. Fig. 1 shows the HSR components, designed to assist humans in daily activities and can be deployed in various environments, including homes, restaurants, stores, and healthcare facilities. For perception, the robot is equipped with an RGB-D camera mounted on its head and a laser range finder located at its base. Each HSR is equipped with a single robot arm with a two-finger gripper for object manipulation. Its hardware configuration includes an onboard computer. While core tasks are processed locally to enable autonomy, the robot's LAN/Wi-Fi connectivity allows computationally intensive algorithms to be offloaded to an external computer when necessary.

One practical application is the tidy-up task, which involves detecting, grasping, and relocating objects to assigned

*This study is based on results obtained from a project, JPNP16007, commissioned by the New Energy and Industrial Technology Development Organization (NEDO). This work received support from JSPS KAKENHI Grant No. 23K03468 and JST ALCA-Next Grant No. JPMJAN23F3.

¹All authors are with Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu, Kitakyushu 808-0196, Japan. buttawong.natee205, isomoto.kosei778@mail.kyutech.jp

²Hakaru Tamukoh is with Research Center for Neuromorphic AI Hardware, Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu, Kitakyushu 808-0196, Japan tamukoh@brain.kyutech.jp

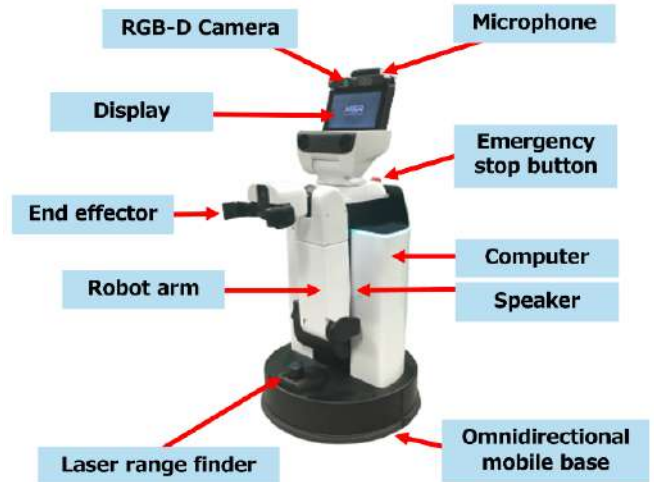


Fig. 1. Components and structure of the human support robot



Fig. 2. Placement by the fixed-point deposition method [11]

storage areas. There is an existing system for the tidy-up task, which won first place in the World Robot Challenge 2020 [10]. However, this existing system faces the problem of tray-type storage deposition. Fig. 2 illustrates the outcome of fixed-point deposition methods that ignore object characteristics or storage conditions, leading to inefficient space utilization, object stacking, and collisions.

There is a previous system [12] that proposed a shape-aware placement method that considers the shape of the object and the storage state to flexibly place items and maximize space efficiency in home environments. This system achieved higher space utilization, greater flexibility, and reduced risks of collisions and stacking compared to fixed-point deposition on an HSR.

However, the previous system relied on pre-collected ob-

ject mask data and faced difficulties in handling objects and postures for which prior information could not be obtained, and was unable to take the current state of the tray into account, resulting in limited flexibility.

To address these limitations, we develop a state-aware planning system that can work with a wider variety of objects and postures and plans placements from the current tray state. Our proposed system contributions are as follows:

- We propose an object placement system that executes flexible plans based on state recognition, taking into cognizance the objects and their postures prior to grasping, as well as the current state of the tray.
- We confirmed that the proposed system is more suitable for real-world scenarios than the previous system in the experiment.

II. RELATED WORK

A. Bin Packing Problems in Robotic Applications

Bin packing problems involve efficiently arranging items within a limited space to maximize the number or total size of packed items without exceeding capacity. Zhao et al. [13] categorize packing problems by dimensionality: one-dimensional consider in a single axis, efficient for linear materials, two-dimensional (2D) consider two axis (width and height), modeling planar shapes for compact layouts, and three-dimensional (3D) considering 3 axis (width, height, and depth), the most realistic with orientation and stacking.

Some packing algorithms have been applied in the robotics field, and many studies have focused on 3D box packing with static robot arms [14], [15]. However, when these methods are adapted to tray-type storage with a mobile manipulator, the cost becomes high relative to the required outcome. The 3D packing algorithm requires accurate 3D scans of each object, which makes it difficult to add new data of objects to a dataset and consumes significant computing resources and time. By contrast, our target home setting must work with diverse object types and requires real-time performance for smooth operation of tidying up, where the robot needs to move among locations.

A 2D packing algorithm is more appropriate for the application of tray-type storage packing with a mobile manipulator. It represents objects and storage footprints as 2D shapes (rectangles, polygons, or object masks), making it easier to collect object data, enabling faster planning, and avoiding the overhead of full 3D reasoning when stacking is not required. For tray placement on a robot, the expected output from the algorithm is a planar pose of the target point within the storage area, along with the orientation of the object, which a 2D packing algorithm provides naturally.

B. Previous Work and Limitations

In recent years, the field of robotics has seen active research on Vision-Language-Action (VLA) models [16], which aim to understand complex linguistic instructions and generalize diverse tasks in an end-to-end manner. However, adapting these models to new environments requires reconstructing large-scale datasets and performing extensive

fine-tuning, which demands significant computational resources. Moreover, in advanced manipulation tasks such as object placement, robots must consider the positions and orientations of pre-existing objects to achieve precise and space-efficient arrangements. When such complex spatial relationships are handled directly by a VLA model, the task structure becomes increasingly complicated, making real-world application challenging.

As a related approach, Buttawong et al. [12] focused on adding the decision-making capability to the robot to enable flexible object placement and efficient space utilization by considering both object shape and storage information. They employed the 2D packing algorithm to calculate suitable placement points using non-overlapping convolution of pre-collected object and tray masks. To identify the tray location, the robot detects an ArUco marker and uses it as a reference point. By combining the placement plan from the packing algorithm with the marker-based reference point, the robot placed the objects on the tray according to the calculated target positions.

Although the previous work achieved higher space utilization than the fixed-point deposition method, its applicability in home environments was limited by the following factors:

- Required collect object data as a datasets: The approach depended on a predefined dataset, making the system unable to handle objects not included in the collected data.
- Not adaptable any posture of objects that not in datasets: If the posture of an object differs from that in the pre-acquired dataset, the placement location cannot be accurately estimated.
- Does not consider the current state of storage: Instead of considering the current state, the system compared with the mask in the image plan, so it could not account for misalignment after placing the object.
- Required a physical marker for tray localization: The system required a physical marker, such as an ArUco marker, to determine object placement within the storage area. In practical home environments, this reliance could be intrusive and negatively affect comfort and aesthetics.

III. METHODOLOGY

An improved performance mainly comes from making a proposed system more autonomous, perception-driven, and adaptable. Instead of relying on pre-collected data or physical markers, the robot now perceives information from objects and tray conditions during the operation, allowing it to handle diverse object postures and achieve more accurate, flexible, and efficient placement.

A. Overview of Proposed System

Fig. 3 illustrates the overview of a proposed system. The operation begins with the robot moving to the search area and performing object detection to extract essential information, which it passes to subsequent processes. An object mask generation process (B) obtains a mask of each

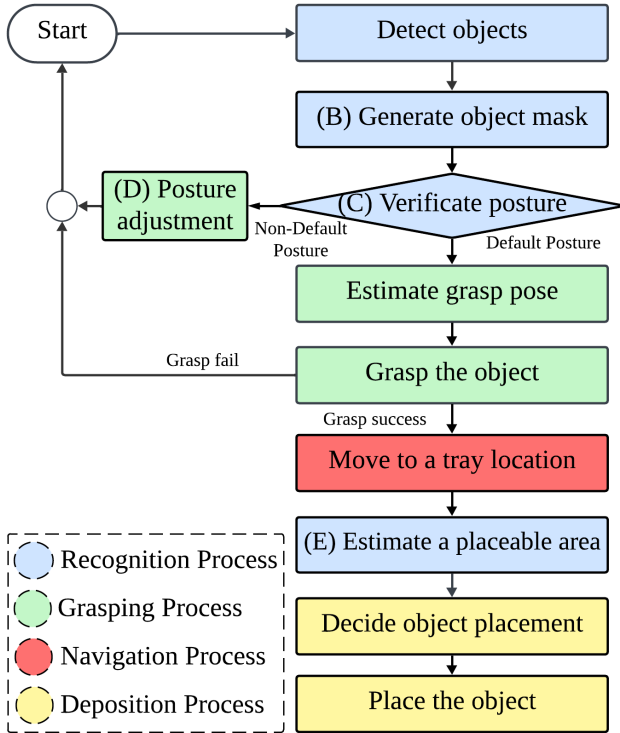
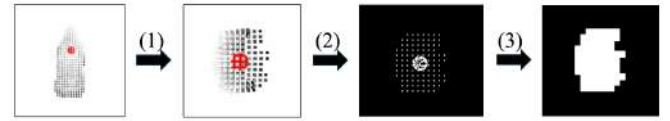


Fig. 3. Overview of the proposed system

object, enabling the system to work with varied items without prior data collection. To address diverse postures, the posture verification process (C) checks that each object is in its default posture. If an object is in a non-default posture, the posture adjustment process (D) grasps and re-places the object to achieve the correct posture. When the object is properly oriented, the system estimates a grasp pose using depth and mask data for stable grasping. After reaching the storage area, the placeable area estimation process (E) detects the corners of the tray and uses them as reference points to enable markerless placement, then generates a dynamic tray mask to assess the current tray state and produces the tray mask. This process can update the state of the tray and can indicate the position of an existing object, resulting in avoiding a collision even if it has an error from the previous placement. We use a two-dimensional irregular packing algorithm based on the non-overlapping convolution method [12], which efficiently handles binary mask inputs of object and tray shapes that we extract from the previous process to determine the optimal placement. Finally, the robot places the object at the computed target point.

B. Dynamic Object Mask Generation Process

Fig. 4 shows the process of obtaining an object mask by extracting the 2D shape from the 3D object shape to use it as an input in the packing algorithm. Since we use an RGB-D camera for object detection, we can obtain both the object information and the corresponding depth data, which together enable 3D reconstruction of each object as input for this process. In step (1), the system aligns the view to a



(1) Set the perspective to a top-down view
 (2) Render the 3D object from above to obtain a 2D mask
 (3) Extract the object's contour and generate a filled binary mask

Fig. 4. Dynamic object mask generation process

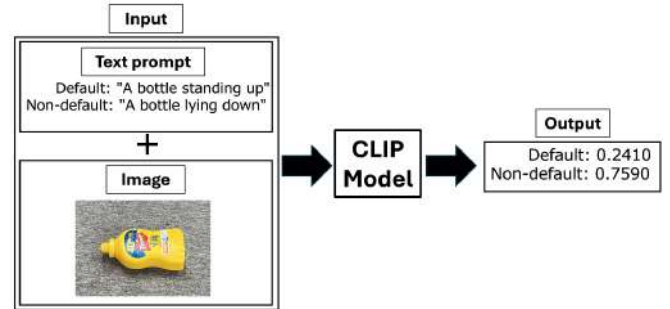


Fig. 5. Verify object posture using CLIP

top-down perspective. Next, step (2) renders the object from above to produce a 2D projection that initially appears as a sparse set of dots. Finally, step (3) converts this dot mask into a complete binary mask by extracting contours and applying morphological closing to fill gaps and recover the full object shape.

C. Object Posture Verification Process

To ensure that objects remain stable after placement and remain easy to grip and immediately usable, the system assigns a default posture to each type of object. Before grasping the object, the posture verification process checks whether the object is in its default posture. Fig. 5 illustrates the method used to indicate the current posture of an object. This process uses a contrastive language-image pretraining (CLIP) model [17] combined with a human-in-the-loop approach for prompt selection. For each object type, a pair of text prompts representing the default posture and non-default postures is selected with human guidance. The CLIP model computes similarity scores between the captured image of the object and each text prompt. Based on these scores, the system determines whether the object is in its default posture.

D. Object Posture Adjustment Process

Fig. 6 illustrates the object posture adjustment process. In the first step, the robot estimates the grasp point of the target object [18] and grasps it. The second step is a base face estimation to indicate which side of the object should be placed on the floor. To estimate the base face, we slice a 3D object shape from the top and bottom along the z-axis (represented by the red and green areas). For each slice, we calculate the difference between the maximum and minimum Z values to indicate flatness and the XY spread to represent the area coverage in the horizontal plane. The side with a

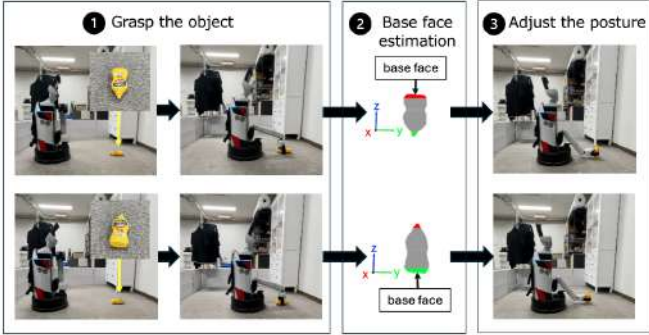


Fig. 6. Object posture adjustment process



Fig. 7. Reference coordinates for tray localization based on tray corners

flatter surface and wider spread serves as the base face of the object. After we defined the base face of the object, we rotated the robot hand joint to turn the base face facing the floor surface and placed the object to adjust the posture as shown in step 3.

E. Placeable Area Estimation Process

This process segments the tray and extracts its corners, which serve as reference points for object placement. When the robot arrives at the deposit location, the NanoSAM [19] is used to perform segmentation and generate a mask of the tray. Fig. 7 shows the tray segmentation and corner detection process. The robot captures an RGB image of the tray and provides it along with a text prompt to the NanoSAM model. The model then returns the result of the tray mask image. With this result, we can continue to use it to extract the top left, top right, bottom left, and bottom right from each corner of the tray. These corners are then used as placement reference points, eliminating the need for a physical marker.

After obtaining the tray segmentation mask and the corner coordinates of the tray, the system can capture the current state of the tray and the positions of existing objects. Fig. 8 illustrates the dynamic tray mask generation process. Using the coordinates of each corner together with the segmentation mask of the tray, the region of the tray is defined. Since the

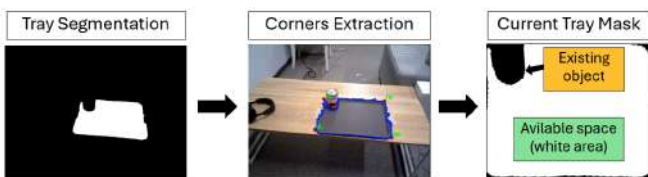


Fig. 8. Dynamic tray mask generation process

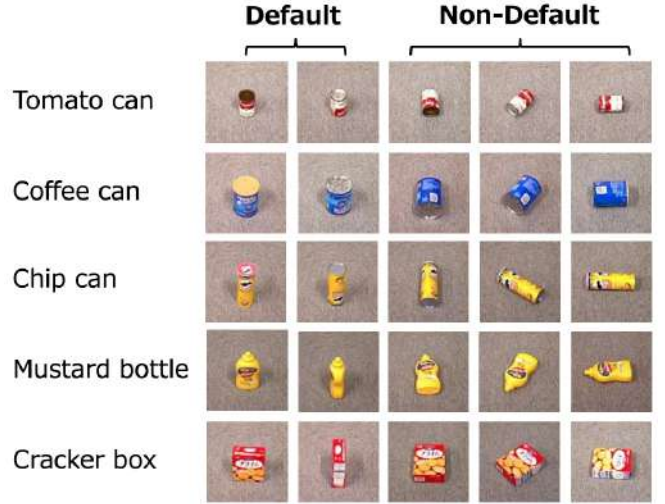


Fig. 9. Images of objects with two default and three non-default postures

actual dimensions of the tray are known, the segmentation mask of the tray from the perspective of the robot can be projected into a top-view perspective and scaled to match the ratio of the object mask. The resulting tray mask represents the current state of the tray, where black areas indicate existing objects, and white areas indicate available space for the next object. This updated tray mask is then provided to the packing algorithm, along with the object mask, to calculate a suitable placement position for the target object.

IV. EXPERIMENT

A. Experiment I: Evaluation of Posture Verification Using CLIP

This experiment evaluates the performance of CLIP in verifying the posture of objects. We select five objects from the Yale-CMU-Berkeley (YCB) dataset [20]. Each object is tested in five different postures: two default postures and three non-default postures. In total, we prepare 25 images for evaluation, as shown in Fig. 9.

Table I presents a categorization of objects into four groups: can, bottle, box, and others. For each category, we designed a pair of contrastive prompts describing both the default and non-default postures, where the default posture refers to a stable placement in which the object remains easy to grasp and immediately usable for humans. We then applied these prompts to classify each prepared image using CLIP and recorded the prediction results.

Table II presents the result of this experiment. The CLIP model correctly predicted the posture of 22 of 25 objects, achieving an accuracy of 88%. These results demonstrated that CLIP could predict object posture with high accuracy.

B. Experiment II: Performance Comparison of the Tidy-Up Task Using the Previous and the Proposed System

We conducted experiments using the HSR in a domestic indoor environment configured to simulate a typical home scenario. Fig. 10 shows the experimental environment, which

TABLE I
A SET OF A PAIR OF CONTRASTIVE PROMPTS FOR EACH TYPE OF OBJECT

Object types	Text prompt represents the default posture	Text prompt represents the non-default posture
Can	a can standing upright on its bottom on the floor	a can rolling on the floor, not standing
Bottle	a bottle standing upright on its bottom on the floor	a bottle lying on its side on the floor
Box	a box standing upright on its bottom on the floor	a box lying flat on the floor
Others	an object placed correctly with its base on the floor	an object lying on its side on the floor

TABLE II
THE RESULTS OF PREDICTION OF OBJECT POSTURE USING CLIP

Object	The number of correct prediction of default posture (/2)	The number of correct prediction of non-default posture (/3)	Incorrect prediction (/5)
Tomato can	2	2	1
Coffee can	2	3	0
Chip can	2	2	1
Mustard bottle	2	3	0
Cracker box	1	2	1
Average accuracy	88%		

TABLE III
COMPARISON OF PROPOSED AND PREVIOUS SYSTEMS ACROSS TRIALS USING STRICT SUCCESS CRITERIA

System	Trial	Successful placements (/5)	Fell out (/5)	Notes
Previous	1	0	1	3 objects placed on tray but in non-default posture, unstable; 1 fell; 1 skipped.
	2	0	2	2 objects placed on tray but in non-default posture, unstable; 2 fell; 1 skipped.
	3	0	0	4 objects placed on tray but all in non-default posture, unstable; 1 skipped.
Proposed	1	5	0	All placed correctly in default posture, stable.
	2	5	0	All placed correctly in default posture, stable.
	3	4	1	One object fell due to gripper opening during placement.

consisted of two main areas: a search area, where the objects were initially located, and a storage area containing a tray for deposition. The tray used throughout the experiments measured 37×29 cm. We used the same object set as in Experiment I, where three objects were initially placed in the default posture, and two objects were in the non-default posture.

This experiment aimed to evaluate whether the proposed system effectively addressed the limitations identified in the previous system. Therefore, this section compares the performance of the proposed system with the previous work through three evaluation trials. Performance was evaluated using two criteria: (1) successful placement of objects in their default posture without collision, and (2) occurrence of objects falling from the tray. The results are presented in Fig. 11 and summarized in Table III.

The proposed system demonstrated higher results than the previous system. Since the previous system relied on pre-collected object masks for planning, and placement became inaccurate when the actual object shape or posture did not match the stored mask. This mismatch caused differences between the planned and actual placements.

In contrast, the proposed system placed all five objects in their default posture without falling in trials 1 and 2. In trial 3, the robot placed four objects correctly, but one object fell when the gripper opened unexpectedly while placing another item. In total, the proposed system achieved 14 correct placements out of 15, corresponding to a 93% success rate.

These results clearly demonstrated that the proposed system significantly improved placement reliability and stability by overcoming the limitations of the previous system. In particular, it addressed errors in object posture, collisions during placement, reliance on physical markers for tray localization, and failures caused by unregistered object. The complete tidy-up process required approximately one minute per object for both systems. However, the proposed system needed an additional 30 seconds when adjusting from a non-default posture.

V. DISCUSSION

In the first experiment, the accuracy of the CLIP model depended heavily on the quality and clarity of the text prompts. By dividing the objects into four types, we provided

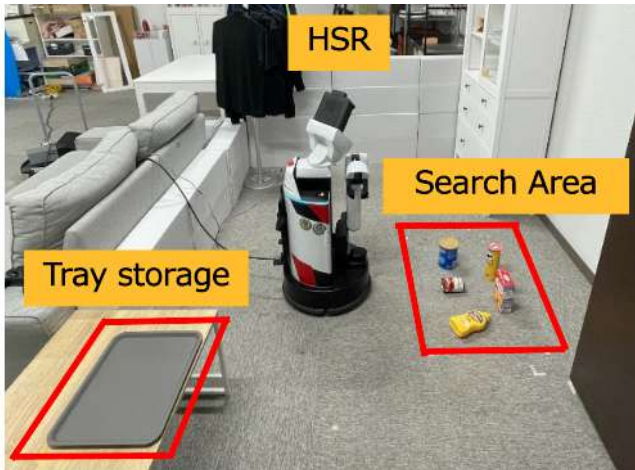


Fig. 10. Experimental environment



Fig. 11. Placement results from the proposed and the previous system

a more specific set of text prompts to the model, which made the prompts clearer and achieved higher prediction accuracy.

The proposed system in the second experiment applied the dynamic object mask generation, enabling the system to handle a broader range of objects without being constrained to a pre-collected set. In contrast, the previous system required a large pre-collected dataset to address object diversity.

The integration of posture verification and adjustment allows the robot to handle objects presented in different postures. Although two objects in the search area were initially in non-default postures, the system adjusted and placed them correctly in their default postures in the final arrangement, while the previous system would suffer mask–pose mismatch relative to the pre-collected masks.

Furthermore, the proposed system employs dynamic tray mask estimation, enabling the tray status to be updated before each calculation of the suitable placement point by the packing algorithm. This dynamic feedback significantly reduced the risk of collisions with previously placed objects, whereas the result of the experiment in the previous system still had collisions with shifted existing objects.

The results of this experiment demonstrate that the proposed system effectively addresses the limitations identified in the previous implementation, improving both flexibility and robustness in real-world operation.

As shown in the Table II, although the posture verification model achieved an 88% success rate, the classification errors

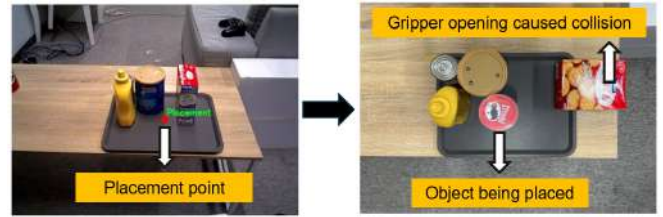


Fig. 12. Cracker box fall due to gripper opening during placement.

in the remaining cases demonstrate a need for a robust mechanism to handle prediction uncertainty.

VI. CONCLUSIONS AND FUTURE WORK

This paper addressed the practical problem of placing diverse objects into a tray both space-efficiently and safely during tidy-up with a mobile manipulator. Previous work [9] required pre-collected data, was not adaptable to any posture of objects that are not in the datasets, did not refer to the current state of the tray, and still required a physical marker for the tray localization. This requirement leads to limitations of object and posture variety, collision with the existing objects, and could be intrusive and negatively affect comfort and aesthetics in the home environment.

To address these limitations, we developed a flexible state-aware planning placement system and evaluated it experimentally. The first experiment demonstrated that CLIP was effective for posture verification. The results of the second experiment showed that the proposed system performed better in several key aspects. The dynamic mask generation process enables the system to complete operation without the pre-collected data, while the posture verification and adjustment process enables the system to work with different object postures. By updating the tray state in real time, the system can avoid collisions with already placed items. Importantly, these improvements were achieved without relying on pre-collected data or physical markers.

Together, these findings highlight that the proposed system is more flexible, more reliable, and better suited for real-world applications, addressing the critical limitations of the earlier system and enabling more robust performance in domestic tidy-up tasks.

Future work will focus on two main improvements. First, the packing algorithm will be updated to include the gripper opening range when planning object placements. This will help create placement positions that match the robot's physical limits and reduce the chance the gripper hits nearby objects during placement.

Second, an error recovery system will be added to fix small errors between the planned and actual placement positions. This will allow the robot to detect and correct errors during operation, creating a closed-loop system that is more reliable and better suited for real-world use.

By addressing these areas, the proposed approach can further advance the practicality of mobile manipulators for autonomous tidy-up tasks in everyday environments.

Moreover, the proposed system can automate the generation of datasets that are currently collected manually through human teleoperation. By performing manipulation tasks autonomously and recording object states, actions, and outcomes, the robot can create large-scale, high-quality datasets without human intervention. These datasets can then be used to train or fine-tune Vision-Language-Action models, enabling them to learn high-level reasoning from real robot experience more efficiently and at a larger scale than manual collection allows.

REFERENCES

- [1] HirePlanner, “Japan labor shortage scheduled to double by the year 2025,” [Online]. Available: <https://home.hireplanner.com/blog-example>. Accessed: Jun. 16, 2025.
- [2] T. Yamamoto, K. Terada, A. Ochiai, F. Saito, Y. Asahara, and K. Murase, “Development of human support robot as the research platform of a domestic mobile manipulator,” *ROBOMECH J.*, vol. 6, p. 4, 2019.
- [3] K. Isomoto, Y. Yano, Y. Tanaka, and H. Tamukoh, “Robust trash can lid opening system,” in *Proc. 2023 Int. Workshop Smart Info-Media Systems in Asia (SISA)*, 2023.
- [4] Y. Yano, K. Isomoto, T. Ono, and H. Tamukoh, “Autonomous waiter robot system for recognizing customers, taking orders, and serving food,” in *Proc. 26th RoboCup Int. Symp.*, 2023.
- [5] T. Shiba, T. Ono, and H. Tamukoh, “Object search and empty space detection system for home service robot,” in *Proc. 2023 Int. Conf. Artificial Life and Robotics (ICAROB2023)*, 2023.
- [6] D. Kanaoka, Y. Tanaka, and H. Tamukoh, “Applying center loss to multidimensional feature space in deep neural networks for open-set recognition,” in *Proc. 17th Int. Conf. Computer Vision Theory and Applications (VISAPP2022)*, vol. 5, pp. 359–365, 2022.
- [7] T. Shiba and H. Tamukoh, “Robotic Grasping of Common Objects: Focusing on Edge Detection for Improved Handling,” in *Proc. 2025 Int. Conf. Artificial Life and Robotics (ICAROB 2025)*, Vol. 30, pp. 487–490, February 13, 2025.
- [8] H. Miura, R. Youh, and Y. Yoshimoto, “Development of a Collaborative System Between A Drone and A Home Service Robot for Enhanced Operational Efficiency,” in *Proc. 2025 Int. Conf. Artificial Life and Robotics (ICAROB2025)*, Vol. 30, pp. 504–508, February 13, 2025.
- [9] R. Terashima, Y. Yano, K. Arimura, and H. Tamukoh, “Proposal of a Grasp Verification Method Utilizing Background Subtraction and Depth Information,” in *Proc. 2025 Int. Conf. Artificial Life and Robotics (ICAROB2025)*, Vol. 30, pp. 491–494, February 13, 2025.
- [10] T. Ono, D. Kanaoka, T. Shiba, S. Tokuno, Y. Yano, A. Mizutani, I. Matsumoto, H. Amano, and H. Tamukoh, “Solution of world robot challenge 2020 partner robot challenge (real space),” *Adv. Robotics*, pp. 870–889, 2022.
- [11] Hibikino-Musashi atHome, “RoboCup JapanOpen @Home Domestic Standard Platform League,” [Online Video]. Available: <https://www.youtube.com/watch?v=AsQeRL0.YC8>. Accessed: Aug. 12, 2025.
- [12] N. Buttawong, K. Isomoto, K. Yamao, N. Fuengfusin, and H. Tamukoh, “Optimizing object placement for human support robots using a two-dimensional irregular packing algorithm for efficient tray storage,” in *Proc. 2025 Int. Conf. on Artificial Life and Robotics (ICAROB2025)*, Oita, Japan, Feb. 13–16, 2025, paper OS17-1. doi: 10.5954/ICAROB.2025.OS17-1.
- [13] B. Guo, Y. Zhang, J. Hu, J. Li, F. Wu, Q. Peng, and Q. Zhang, “Two-dimensional irregular packing problems: A review,” *Frontiers in Mechanical Engineering*, vol. 8, art. 966691, Aug. 2022. doi: 10.3389/fmech.2022.966691.
- [14] F. Wang and K. Hauser, “Stable bin packing of non-convex 3D objects with a robot manipulator,” in *Proc. 2019 IEEE Int. Conf. Robot. Autom. (ICRA)*, Montreal, Canada, pp. 8698–8704, 2019.
- [15] F. Wang and K. Hauser, “Robot packing with known items and nondeterministic arrival order,” *IEEE Trans. Autom. Sci. Eng.*, vol. 18, no. 4, pp. 1901–1915, 2021.
- [16] M. Reuss, “State of VLA Research at ICLR 2026,” *Technical Blog Post*, Oct. 2025. [Online]. Available: https://mbreuss.github.io/blog_post_iclr_26_vla.html. Accessed: Sep. 10, 2025.
- [17] OpenAI, “Contrastive Language-Image Pre-Training (CLIP),” [Online]. Available: <https://github.com/openai/CLIP>. Accessed: Aug. 15, 2025.
- [18] T. Ono and H. Tamukoh, “A fast pick-and-place method for home service robots using 3D point clouds,” in *Proc. 2020 Int. Conf. Artificial Life and Robotics (ICAROB2020)*, Beppu, Japan, pp. OS20–3, 2020.
- [19] X. Sun, J. Liu, H. T. Shen, X. Zhu, and P. Hu, “On efficient variants of segment anything model: A survey,” *arXiv preprint arXiv:2410.04960*, 2024.
- [20] B. Calli, A. Walsman, A. Singh, S. Srinivasa, P. Abbeel, and A. M. Dollar, “Benchmarking in manipulation research: Using the Yale–CMU–Berkeley object and model set,” *IEEE Robot. Autom. Mag.*, vol. 22, no. 3, pp. 36–52, 2015.