2026 IEEE/SICE International Symposium on System Integration (SII)
January 11-14, 2026. Cancún, México

Mobile Manipulation System for In-Store Product Stocking, Disposal, and Customer Detection with Minimal Environment Modification*

Ryusei Sawada, Ryota Nakamura, Asahi Miyaji, Kousuke Shimizu, Koki Sunaga, Koji Aikawa, Kyohei Fujita, Ryunosuke Maruta, Shota Nakajima, Mihoko Niitsuma, Member, IEEE

Abstract— **In recent years, competitions such as the World Robot Summit (WRS) have spurred the development of robotic systems aimed at automating operations in convenience stores and other retail outlets. However, many existing methods require extensive modifications to products and shelves, and are highly dependent on the environment, making their deployment in brick-and-mortar stores problematic. In this study, we propose a mobile manipulation system that performs highly accurate product stocking, disposal, and customer detection while minimizing environmental modifications. The proposed system consists of three elements: (1) highly accurate product position recognition using an industrial camera and ArUco markers; (2) robot arm control with variable grasping strategy, coordinate correction, collision prevention, and retry functions; and (3) customer approach detection using 2D LiDAR. Testing in a competition scenario simulating a real-world environment demonstrated improved grasping success rates, detection of discarded products under obstructed visibility, and safety, confirming both practicality and competition performance.**

I. INTRODUCTION

A. Competition Task Overview

This paper introduces Team HARChuo's proposal for the stock and disposal tasks and technical challenge at the World Robot Summit (WRS) [1] Future Convenience Store Challenge (FCSC 2025) [2] held in July 2025.

The stock and disposal task in the FCSC competition consists of two subtasks: the stock task and the stock-and-disposal task. According to the rulebook [3], the system is expected to improve work efficiency and energy conservation in convenience stores, and its performance is evaluated based on speed and accuracy.

In the stock task, the robot places products from a container onto store shelves. The container holds three types of rice balls arranged in varied positions to simulate real environments. In the stock-and-disposal task, the robot removes designated items from the shelf and replaces them with items from the container. Eighteen products (nine types, two each) are placed on the shelves, and one of each type in the container is used for restocking. One item per type is randomly selected for disposal, requiring the robot to identify and remove the target items and then arrange all products face-up, meaning items closer to expiration are placed at the front. Only small product markers are allowed, with a total area limit of 400 mm².

Based on these requirements, the stock task mainly demands product placement and recognition under varied object arrangements, while the stock-and-disposal task additionally requires disposal handling and customer detection during operation.

B. Related research and the position of this proposal

Previous systems developed for the stock task relied on an RGB-D camera and YOLACT-based object detection, but suffered from arm-length limitations, slow path planning, and a limited number of items that could be grasped at once. In prior work on the stock and disposal task [4], a mobile manipulation system using an RGB-D camera, YOLACT, and principal component analysis was proposed, but its accuracy declined due to insufficient training and viewpoint-dependent errors. Other related research has introduced environment-dependent systems, such as motorized shelves or shelves equipped with multiple embedded sensors for product and customer detection [5-10]. While these systems simplify perception, they increase installation cost, complicate recovery from failures, and restrict layout changes, which is the issues common across retail environments.

In contrast, our approach aims to achieve accurate and efficient stock and disposal operations without environmental modification. The proposed system uses sliding shelves instead of motorized mechanisms and integrates ArUco-based position recognition, an industrial camera, and adaptive grasping to improve accuracy and operational flexibility. Customer detection is achieved using 2D LiDAR to ensure safe interaction while the robot is working near customers. This positions our work as a practical, minimally intrusive solution that balances accuracy, robustness, and deployability in real retail settings.

II. METHOD

A. System Overview

1) Hardware

The hardware constraints for the mobile robot and product shelves are defined in the rulebook [3]. The mobile robot and product shelves shown in Fig. 1 were designed to satisfy these constraints. The mobile robots shown in Fig. 1 (a) and Fig. 1 (b) are equipped with a robotic arm, an industrial CMOS camera, a two-finger parallel gripper, and a suction gripper.

* This study was partially supported by Joint Research Project, Institute of Science and Engineering, Chuo University. R. Sawada, R. Nakamura, A. Miyaji, K. Shimizu, K. Sunaga, K. Aikawa, K. Fujita, R. Maruta, S. Nakajima, M. Niitsuma are with Chuo University, Bunkyo-ku, Tokyo, 112-8551. (e-mail:niitsuma@mech.chuo-u.ac.jp)

The two-finger parallel gripper and suction gripper were used selectively for the stock task and the stock and disposal task. For the stock task, image processing was performed on images acquired by the RGB-D camera mounted on the upper part of the mobile robot to control the robotic arm and grippers. For the stock and disposal task, control was performed based on images acquired by the industrial CMOS camera mounted on the robotic arm. Additionally, a 2D LiDAR is mounted on the mobile robot, considering the robotic arm's movable range, for customer detection as described later in Section II.C.1.

The product shelf shown in Fig. 1 (c) features a mechanism where the shelf boards can be pulled out and pushed back via a sliding action. This design anticipates tasks like pulling out shelves for product display in actual convenience stores, aiming to secure the robotic arm's workspace and improve efficiency.

2) Software

Fig. 2 shows the system configuration and overall processing flow for the stock task, and Fig. 3 for the stock and disposal task. Furthermore, to clarify the robot's workflow in the stock and disposal task, Fig. 4 shows a task flowchart illustrating the robot's actions during each phase of task execution, such as recognition, grasping, disposal, and face-up. The specific methods and algorithms for each component to implement these tasks are detailed in Sections II.B and II.C.

B. Stock Task

1) Planning

This section describes the task planning for a robotic arm to perform the task of displaying rice balls. This task plan consists of two stages: static display position planning, which determines the placement of the rice balls, and dynamic grasping and transport planning, which determines actions in real-time in conjunction with object detection.

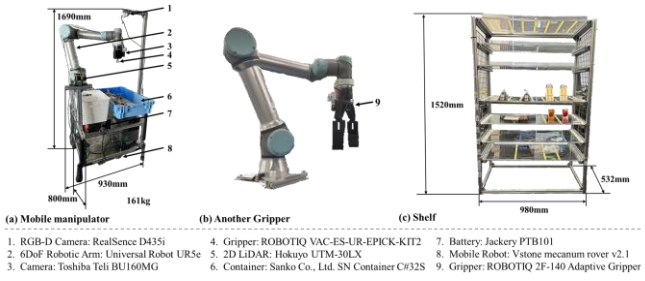

Fig. 1 Hardware components.

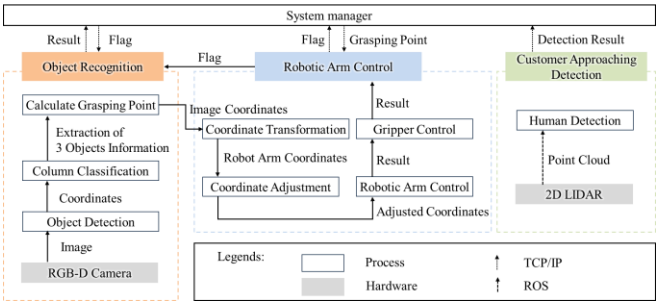

Fig. 2 Stock Task Software System.

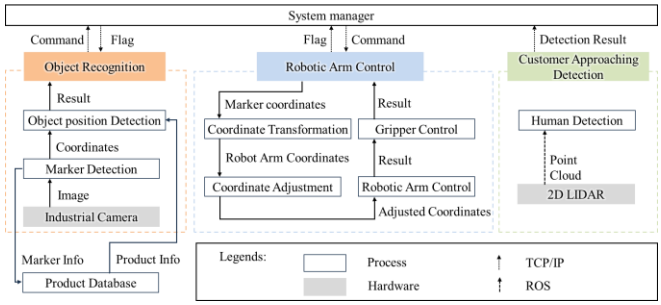

Fig. 3 Stock and Disposal Task Software System.

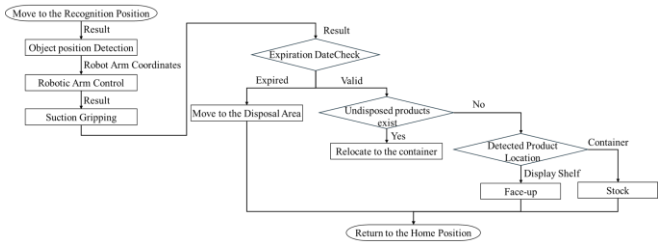

Fig. 4 Robot Workflow in the Stock and Disposal Task.

First, in display position planning, the layout of the display shelf is set before the task begins. Specifically, layout information, such as which rice balls and how many to place in which row, the spacing between the rice balls, and the three-dimensional coordinates that serve as the reference point for calculations, are defined in advance. This reference point indicates the center of a specific rice ball, and the placement coordinates of all rice balls are automatically calculated as relative positions from this point. This method improves the system's flexibility and maintainability, as the placement plan can be easily readjusted simply by changing the layout information.

Next, in grasping and transport planning, the system dynamically plans the motion to transport rice balls from a container to their planned positions based on a Greedy Algorithm. First, to maximize operational efficiency, it creates a plan to grasp up to a maximum of three rice balls of the same type at once. Then, based on the results of object detection, it verifies the feasibility of the plan. If grasping is difficult (e.g., the object is not detected or is beyond the gripper's range of motion), it reduces the number by one and reverifies. Through this verification, only feasible plans are reflected in the robot's motion. Furthermore, to enhance the system's robustness in this planning process, a timeout function has been implemented. If planning fails a specified number of times for reasons such as the object being temporarily undetectable due to changes in lighting conditions or the object no longer being in the container, the system cancels the transport of the object and moves on to the next transport task. This prevents a single failure from leading to a system-wide halt and increases the task completion rate.

In this way, by combining static display position planning with dynamic grasping and transport planning using a greedy algorithm and a timeout function, we have achieved efficient display work with a high task completion rate.

2) Object recognition

This system performs object detection using an RGB-D camera fixed to the top of the container. For object detection, YOLACT [11], a Real-time Instance Segmentation method, is employed. YOLACT was selected because it enables real-time object detection and achieves high detection accuracy even when objects overlap. Fig. 5 shows the results of object detection using YOLACT.

After object detection, the system calculates the position information of displayed products based on the detection results. First, it utilizes the condition that products are aligned by type and determines the baseline x-coordinate values for each column through prior calibration. It then clusters the detected object coordinates based on these baseline values to identify columns. Next, it assigns product type labels to each column and aligns them in y-coordinate order. This allows the system to flexibly adapt to changes in the arrangement order of product columns.

Next, the center of gravity and the required grasping width are calculated for cases where one to three objects (n = 1, 2, 3) of each type are arranged. Let n be the number of objects, and let the 3D position coordinates of the i-th object be $\mathbf{p}_i = (x_i, y_i, z_i)$. The overall center of gravity $\mathbf{C}$ of the n objects is determined as the arithmetic mean of their respective coordinate components, using (1).

$$\mathbf{C} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{p}_i = \left(\frac{1}{n}\sum_{i=1}^{n}x_i, \frac{1}{n}\sum_{i=1}^{n}y_i, \frac{1}{n}\sum_{i=1}^{n}z_i\right) \quad (1)$$

The width is calculated as the difference between the upper and lower edges of the extracted region.

The calculated center of gravity position and width information are transmitted via communication according to the conditions requested by the robot control system. Note that gripping by the gripper becomes difficult when the object is in an overturned state. Therefore, the area and width of the object region are calculated, and if either exceeds the threshold, the object is excluded from processing. This reduces the risk of display failure.

3) Robotic arm control

This section describes the control method for the robotic arm in the stock task. TCP/IP communication is used for communication between the robotic arm and the control PC, where an operation script written in URScript is sent from the PC to the robotic arm. URScript is a programming language specifically designed to control Universal Robots' arm robots.

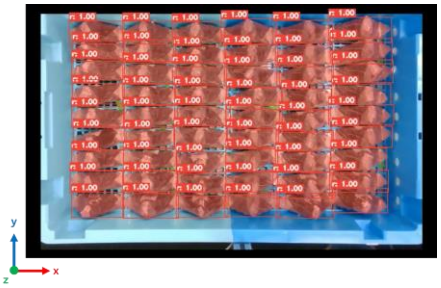

Fig. 5 Results of object detection using YOLACT. (The characters at the top of the bounding box indicate the object label and confidence level.)

The work process is executed in conjunction with the object detection system. First, based on a grasping and transport plan, the object's center coordinates and gripper opening and closing width are received from the object detection system. The coordinate transformation and a two-stage correction, which will be described later, are applied to these coordinates to determine the final target coordinates. Finally, an operation script describing the series of operations to display the rice ball is generated and executed using these corrected coordinates.

For coordinate transformation, a rigid body transformation matrix estimated from five pairs of corresponding points in the camera coordinate system and the UR coordinate system is used. Specifically, the centroid of the corresponding point clouds is calculated and centered, after which the rotation matrix R is obtained by performing singular value decomposition on their correlation matrix. The translation vector t is calculated by applying the rotation matrix R to the centroid of the camera coordinate system, and finally, the coordinate transformation is performed based on equation (2).

$$P_{ur} = RP_{cam} + t \quad (2)$$

Furthermore, to ensure the accuracy and safety of the operation, a two-stage coordinate correction is implemented. First, as a simple correction, the system corrects for the center point deviation caused by the oblique viewing angle, which results from the distance between the camera's center and the target object. This correction amount is determined through a calibration process that involves grasping rice balls at various locations within the container beforehand and analyzing the relationship between the horizontal distance from the camera center and the measured position error. Next, to avoid collisions of the robotic arm, the system verifies that the final corrected coordinates do not deviate from a predefined safe movable area within the container. If the coordinates are outside this area, they are corrected to the inner boundary of the area so that the entire gripper fits inside the container.

For the trajectory planning of the robotic arm, a semi-static method is adopted to avoid complex real-time path calculations and achieve high-speed, stable operation. This method uses a safe path template that connects three points: the gripping position, the display position, and a predefined waypoint for collision avoidance. The operation script is dynamically generated by embedding the calculated coordinates into this template.

This dual correction and safety function improves the accuracy and safety of the system.

C. Stock and Disposal Task

1) Planning

This method requires markers to be visible, but there are cases where detection may fail due to overlapping products. Therefore, when non-discarded items intended for disposal cannot be found, we added a process to temporarily relocate non-disposal items into the container. This prevents disposal omissions caused by occlusion. Based on the above, we propose a system that comprehensively performs tasks for all products in the order of disposal, face-up, and stock.

2) Object recognition

This paper describes a product recognition method using ArUco markers. Since poses obtained from a single frame are susceptible to noise and other factors, leading to grasping failures, this method implements a robust pose estimation algorithm. First, to improve marker detection rates, Gaussian filtering is applied to the input image for noise reduction, followed by edge enhancement using a sharpening filter [12]. Next, to mitigate the impact of momentary detection errors, all pose information detected over a 3-second period is accumulated as time-series data.

To calculate the most probable pose from the accumulated data, multi-stage outlier removal is performed. First, the marker facing most directly toward the camera is selected as the processing target. Subsequently, two-stage filtering is applied to the time-series data of that marker. First, DBSCAN (Density-Based Spatial Clustering of Applications with Noise) clustering is used to remove positional outliers. DBSCAN is suitable for this task because it does not require pre-specifying the number of clusters and can automatically isolate points in low-density regions as noise. The parameters were set to a maximum distance of 5 mm for forming clusters and a minimum of 2 data points per cluster. The maximum distance value was empirically set to be larger than the measurement error of the products but sufficiently smaller than the distance between adjacent products. Setting the minimum number of data points to 2 also effectively removes isolated, sudden false detections as noise. Second, statistical methods (mean ± 2σ) are used to remove angular outliers. The threshold of 2σ was chosen to exclude outliers caused by measurement errors while preserving valid data. Finally, the 3D pose of the target object is determined by calculating the average of the reliable data set that passes through all these filtering steps. Furthermore, for objects with markers attached to multiple surfaces, pose estimation is performed using the predefined rotation matrix R_{offset} and translation vector t_{offset}. Using the rotation matrix R_A and translation vector t_A of the marker in the camera coordinate system, the object's pose (R_B, t_B) is calculated using (3).

$$R_B = R_A R_{offset} \quad , \quad t_B = t_A + R_A t_{offset} \tag{3}$$

This enables stable detection of the grasping position, as estimation is possible based on other markers even when a specific marker cannot be detected within the camera's field of view.

3) Robotic arm control

This section describes the control method for the robotic arm in the stock and disposal task. The system receives the product name, task type, and position/orientation coordinates from the marker detection system, and applies coordinate transformation and correction to these coordinates. Finally, based on this information, the system generates and executes an operation script detailing the sequence of operations for suction-gripping and transporting the target product. The control system configuration and the robotic arm's trajectory planning adopt the same methods as the stock task.

The coordinate transformation is an extension of the method used in the stock task, adapted to accommodate multiple areas and tiers. It estimates the transformation matrix from the marker coordinate system to the robot coordinate system using corresponding points measured in advance for each area. Position transformation uses the same method as the stock task. For orientation transformation, the Euler angles in the camera coordinate system and the rotation vector in the UR coordinate system at each corresponding point are converted into rotation matrices. The relative rotation between them is then averaged and subsequently orthonormalized using singular value decomposition to suppress rotation errors. This achieves the transformation of the 3D position and orientation of objects belonging to any tier and area.

Furthermore, two functions were implemented to enhance operational reliability and safety. One is a collision avoidance function to prevent collisions with shelf boards. It avoids the risk of collision and damage to the robotic arm by forcibly correcting the coordinates if the calculated suction position falls below a preset height threshold based on the number of shelf tiers. The other is a retry function to improve the suction success rate. When a pickup attempt fails, it retries up to three times while changing the pushing depth. If it is still unsuccessful, the system requests the marker detection system to remeasure the marker and retries the same suction operation with the new coordinates. If it ultimately fails, the task for that product is interrupted, and the system moves on to the next operation, thereby increasing the success rate while avoiding delays in the overall process.

As with the stock task, this control flow and the dual safety and reliability functions improve the system's precision and safety.

4) Customer approaching detection

It is dangerous for visitors to approach a mobile robot while it is working. For this reason, it is necessary to notify people in the vicinity that a mobile robot is operating. In this system, point clouds obtained from a single 2D LiDAR are clustered, and people are detected based on the size of the cluster.

Person detection in this system is performed using the following steps. If the distance d between point clouds satisfies (4), they are considered to be in the same cluster. The above process is performed for all point clouds. Then, if the number of point clouds N in each cluster satisfies (5), the cluster is determined to be a person.

$$d \leq 30 \text{ mm} \tag{4}$$

$$20 \leq N \leq 200 \tag{5}$$

If a cluster determined to be a person enters the audio notification area of the mobile robot, which is surrounded by a red frame in Fig. 6, audio information is provided to urge the customer to leave. However, because there is a possibility that the robot arm may be determined to be a person, the range of motion of the robot arm is set outside the range of the person detection process. The range of person detection can be set arbitrarily.

III. EXPERIMENT ON THE STOCK TASK

A. Experiment

A rice ball display experiment was conducted to evaluate the effectiveness of a rice ball display method using a robotic arm. First, a total of 27 rice ball mockups, consisting of three types with nine of each type, were manually placed and

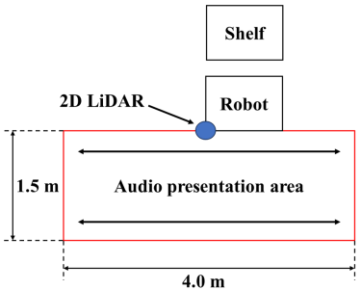

Fig. 6 Experimental environment for evaluating human detection.

arranged at equal intervals in a container. From this setup, under indoor fluorescent lighting, one trial was conducted for each type of rice ball to simultaneously grasp one, two, and three items, respectively. A series of trials to display these 18 rice balls was defined as one set, and the experiment was conducted for a total of five sets.

The evaluation items were success rate and task time. The success rate was calculated as the ratio of the number of successful trials to the total number of trials for each grasping condition. The task time was measured as the time from when the robotic arm started to move to grasp the rice balls until it completed their placement on the display shelf.

B. *Results and Discussion*

As shown in Table I, the stock task achieved a success rate above 87.5% under all conditions. The number of trials for grasping three rice balls was smaller because their combined width exceeded the gripper's maximum opening, requiring the system to divide the operation into two and one. The average work time per trial was about 20 seconds and varied mainly with the physical distance between the grasping and display positions. According to FCSC rules, 54 rice balls must be displayed within 10 minutes; assuming three items per grasp, the estimated total work time is about 6 minutes. These results indicate that the proposed method is both effective and time-efficient.

The main cause of failures was position calculation error. This likely resulted from imaging conditions, including slight camera tilt affecting correction accuracy and deviations in the detected center point when rice balls near the container edge were captured from an oblique angle. As the number of rice balls decreased, previously hidden sides became visible, altering their appearance and reducing recognition accuracy.

Two improvements may further increase success rates: installing a camera at the robot arm's end to capture images from directly above, and adding grasp success/failure detection using a force sensor. These enhancements would enable recovery actions and improve overall robustness.

TABLE I. DISPLAY METHOD EVALUATION EXPERIMENT RESULTS.

Number of grasps	Number of grasp attempts	Number of successes	Number of failures	Success rate [%]
3	14	13	1	92.0
2	16	14	2	87.5
1	16	14	2	87.5

IV. EXPERIMENT AND DISCUSSION ON THE STOCK AND DISPOSAL TASK

A. *Experiment*

1) *Object recognition and Robotic arm control*

To validate the object detection method, we conducted a static accuracy experiment using ArUco markers under indoor fluorescent lighting. The marker was placed approximately 600 mm from the camera at an angle of about 40°, and five measurements were taken for three marker types: (i) 4x4, 0.6 cm, (ii) 6x6, 0.6 cm, and (iii) 6x6, 0.8 cm.

Using the selected marker type, grasping tests were conducted on six products: rice ball, sandwich, packaged juice, stick salad, coleslaw, and lunch box. An industrial CMOS camera with a resolution of 1,440×1,080 and a focal length of 8 mm was used. The frame rate was approximately 27 frames per second. A suction gripper equipped with a 30 mm diameter silicone suction pad was used for grasping, with a suction pressure of approximately 10.1 kPa. The experimental results are shown in Tables II and III.

2) *Customer approaching detection*

To evaluate the person detection method, we conducted an experiment in the environment shown in Fig. 6. People passed each other 10 times over approximately 50 seconds, yielding 478 frames at 0.1-s intervals. Two conditions were tested: a two-person case and a four-person case. Detection accuracy was defined as the match rate between the detected number of people and the actual number. The results are summarized in Table IV.

B. *Results and Discussion*

1) *Object recognition and Robotic arm control*

As presented in Table II, condition (iii) exhibited errors approximately three times larger than the others; therefore, conditions (i) and (ii) were adopted. Although their average errors were similar, (i) produced false detections at positions without markers, while (ii) provided more stable recognition with fewer false positives. Using the (ii)-type markers in the grasping experiments, Table III shows that lunch boxes, coleslaw, and stick salad achieved high success rates, whereas rice balls and packaged drinks performed poorly.

These results indicate that grasping performance strongly depends on the material and shape of each object. Items with rigid plastic containers provided good adhesion to the suction pad, allowing stable vacuum retention during transport. In contrast, rice balls wrapped in flexible film easily deform, and packaged drinks with many creases cannot form a sealed surface, leading to unstable suction.

This shows that a single suction-based method is insufficient for handling diverse products. Flexible objects such as rice balls may require alternative gripping methods, such as clamping. To address this challenge, we propose an adaptive handling strategy that integrates ArUco marker recognition for position/orientation estimation with object recognition for identifying product types and characteristics. Object recognition restricts the marker search region, improving processing speed and reducing false detections. This enables faster and more reliable extraction of the marker

TABLE II. MARKER ACCURACY VERIFICATION RESULTS.

	1st error [mm]	2nd error [mm]	3rd error [mm]	4th error [mm]	5th error [mm]	average [mm]
(i)	11.85	20.38	27.09	11.67	19.96	18.19
(ii)	27.04	20.02	13.86	10.84	12.05	16.76
(iii)	43.43	43.24	105.6	49.23	54.16	59.12

TABLE III. GRASPING VERIFICATION EXPERIMENT RESULTS.

Product type	Number of trials [times]	Number of successes [times]	success rate [%]
rice ball	10	0	0
sandwich	10	6	60
juice	10	4	40
stick salad	10	10	100
coleslaw	10	9	90
lunch box	10	8	80

TABLE IV. RESULTS OF THE PERSON DETECTION METHOD EVALUATION EXPERIMENT.

(a) Accuracy verification by two people

Number of people	Detection rate [%]
0	0
1	7.53
2	92.3

(b) Accuracy verification by four people

Number of people	Detection rate [%]
0	0
1	0.628
2	9.68
3	45.1
4	45.3

corresponding to the target item, increasing positioning accuracy in subsequent motion planning.

Furthermore, based on the recognized object type, the system dynamically selects the appropriate manipulation method. For deformable items such as rice balls and packaged drinks, it switches from suction to a parallel gripper; for rigid items such as lunch boxes, it retains suction gripping. This allows autonomous selection of an optimal manipulation strategy according to object characteristics.

2) Customer approaching detection

The experiment showed that the system never misdetected "0 people" when someone was within the voice-prompt area, and the detection accuracy for a single person entering the range was 100%, confirming correct operation. The system also detected the number of people when two or more individuals passed each other, although accuracy decreased as the number of people increased (Table Ⅳ). A likely cause is occlusion: when people pass each other, the person closest to the 2D LiDAR blocks those farther away. Increasing the number of LiDAR sensors and improving robustness to occlusion should enable more accurate counting and support congestion estimation.

V. CONCLUSION

This paper presented our approach and technical challenges for the stock and the stock-and-disposal tasks in the World Robot Summit. We developed a mobile manipulation system integrating ArUco-based pose recognition, industrial-camera sensing, variable grasping strategies, coordinate correction, safety functions, and 2D LiDAR-based customer detection. All components were verified in an environment simulating an actual store. Future work includes improving grasping performance for flexible and uneven objects and enhancing customer detection by reducing occlusion using additional LiDAR sensing.

REFERENCES

[1] "World Robot Summit" https://worldrobotsummit.org (accessed Aug. 8, 2025).

[2] "FUTURE CONVENIENCE STORE CONTEST" https://f-csc.org/wrs-fcsc-2025/ (accessed Aug. 8, 2025).

[3] "WRS Future Convenience Store Challenge 2025 Stock and Disposal Task Rulebook" https://fcsc.org/wpcontent/uploads/2025/06/FCSC2025_stock_disposal_task_v006_JPN.pdf (accessed Aug. 8, 2025).

[4] R. Kai et al, "Development of an Environmentally Independent Mobile Manipulation System for Product Disposal in Retail Stores," in Proc. 2024 IEEE/SICE International Symposium on System Integration (SII), Tokyo, Japan, 2024, pp.1060–1065, doi:10.1109/SII56671.2024.10417581.

[5] "Future Convenience Store Challenge DAY2 (September 11,2021)" https://www.youtube.com/live/J3p6wuFzX40?feature=share (accessed Sep. 27, 2025)

[6] "Future Convenience Store Challenge DAY3 (September 11,2021)" https://www.youtube.com/live/aEeoPNXMIEE?feature=share (accessed Sep. 27, 2025)

[7] Gustavo A. Garcia Ricardez, Pedro M. Uriguen Eljuri, Yuta Kamemura, Shiori Yokota, Noriyuki Kugou, Yoshitatsu Asama, Ziyu Wang, Hikaru Kumamoto, Kotaro Yoshimoto, Wai Y. Chan, Tomoki Nagatani, Pattaraporn Tulathum, Bunyapon Usawalertkamol, Lotfi El Hafi, Hiroki Ikeuchi, Masaki Yamamoto, Jun Takamatsu, Tadahiro Taniguchi & Tsukasa Ogasawara (2022) Autonomous service robot for human-aware restock, straightening and disposal tasks in retail automation, Advanced Robotics, 36:17-18, 936-950.

[8] Masashi Seki, Kazuyoshi Wada, Yosuke Kitajima, Masato Hashimoto & Tetsuo Tomizawa (2022) Development of XYZ stage-type display robot system for stock and disposal tasks in convenience stores, Advanced Robotics, 36:23, 1252-1272.

[9] Ryusei Tomikawa, Yusuke Ibuki, Kazufumi Kobayashi, Kazuhisa Matsumoto, Hisanori Suito, Yuma Takemura, Mayu Suzuki, Tsuyoshi Tasaki & Kenichi Ohara (2022) Development of display and disposal work system for convenience stores using dual-arm robot, Advanced Robotics, 36:23, 1273-1290.

[10] Kosuke Mizutani, Nobuhiro Nagasawa & Kosei Demura (2022) Development of automatic opening and closing shelf for convenience stores in cooperation with collaborative robot, Advanced Robotics, 36:23, 1241-1251.

[11] Daniel Bolya, Chong Zhou, Fanyi Xiao, Yong Jae Lee,"Yolact: Real-time instance segmentation," IEEE/CVF International Conference on Computer Vision, pp.9157–9166, 2019

[12] T. Kitsukawa, M. Takahashi, A. Moro, Y. Harada, H. Nishikawa, A. Hamatani, K. Umeda, and M. Noguchi, "A proposal of AR marker recognition system attached to a cart in a factory," in *Proc. SSII2020 The 26th Symposium on Sensing via Image Information*, Yokohama, Japan, 2020, pp. IS2-19.