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

Development of a Product Recognition and Task Planning System for a Shelf-Stocking Robot

Tomohiro Hosokawa[1], Kaname Yamauchi[1], Maria Masuda[1], Masashi Seki[1] and Kazuyoshi Wada[1]

Abstract— **This paper describes the winning robot system in the WRS 2025 Future Convenience Store Challenge (WRS2025 FCSC), an international competition focused on tasks related to managing convenience store merchandise stock. To ensure reliable and efficient task execution, we enhanced the XYZ stage-type display system, which automates stock and disposal tasks. We introduced a two-stage recognition method that combines object detection using YOLO and identification using ArUco markers, enabling the high-precision acquisition of product type, position, and orientation. Furthermore, the work planning software was divided into three components: product management, task planning, and motion planning, improving program readability and debugging efficiency. Evaluation of the developed system at WRS2025 FCSC yielded the following results: in the preliminary round, it achieved a perfect score of 54 points in the Stock Task and 49 points in the Stock and Disposal Task. In the final round, it scored 54 points in the Stock Task and 31 points in the Stock and Disposal Task, securing the overall championship.**

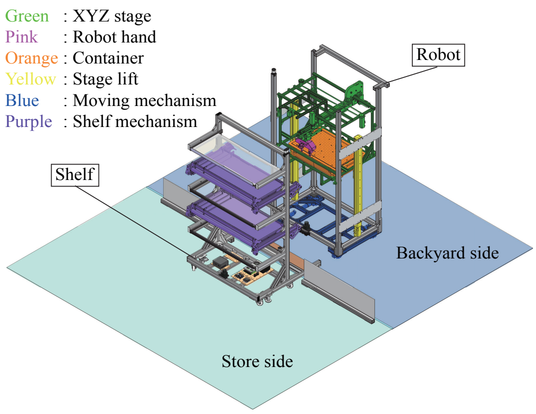


Fig. 1: XYZ stage type display system

I. INTRODUCTION

In Japan, declining birth rates and an aging population have led to a severe labor shortage. This shortage is particularly pronounced in convenience stores, where a small number of staff handle diverse tasks, driving high demand for automation and labor-saving solutions. Among these tasks, product restocking and stock operations are reported to be among the most time-consuming [1]. To address this challenge, our laboratory has developed an XYZ stage-type display system [2]. However, the conventional system had the following two issues: (A) The adoption of ArUco markers for product recognition required large markers for long-distance recognition from above the robot, significantly impacting package design. (B) The software governing work planning was integrated, making information management and debugging cumbersome. This paper proposes a two-stage recognition method that combines YOLO and ArUco markers, along with a work planning component architecture divided by role, to address these issues. It also reports on the results achieved using this system at the World Robot Summit 2025 Future Convenience Store Challenge, where it won first place (Minister of Economy, Trade and Industry Award) and the Stock and Disposal Task Winner [3].

The structure of this paper is as follows. Section II introduces related research, and Section III explains the system architecture. Section IV describes the product recognition method, Section V details the work planning software, and Section VI presents the competition results. Section VII provides a discussion, and Section VIII concludes the study.

*This research was supported by TMU local 5G research support

[1]H. Tomohiro, Y. Kaname, M. Maria, M. Seki, and K. Wada are with the Graduate School of Systems Design, Tokyo Metropolitan University, 6-6 Asahigaoka, Hino, Tokyo, Japan. k_wada@tmu.ac.jp

II. RELATED WORK

In the Future Convenience Store Challenge (FCSC), numerous robotic systems targeting stock and disposal tasks have been developed, with their results reported [4]–[9]. This section focuses on the system's product recognition methods and work planning software architecture, organizing their key features.

A. Product Recognition Methods

Product recognition methods in FCSC can be broadly categorized into marker-based methods [4]–[8] and markerless methods [9]. In marker-based methods, markers such as ArUco markers are affixed to products. Camera recognition of these markers allows easy estimation of the product's ID, position, and orientation. This provides advantages such as the stable acquisition of information necessary for product grasping and the ability to determine whether a product should be discarded, but it also has the disadvantage of impairing the design of the product packaging. Therefore, it is desirable for the markers to be small and inconspicuous. Markerless systems, on the other hand, learn product features in advance using deep learning and directly recognize products using RGB or RGB-D images as input. This approach eliminates the need for additional marking work and offers greater operational flexibility, but it is difficult to determine whether a product should be discarded.

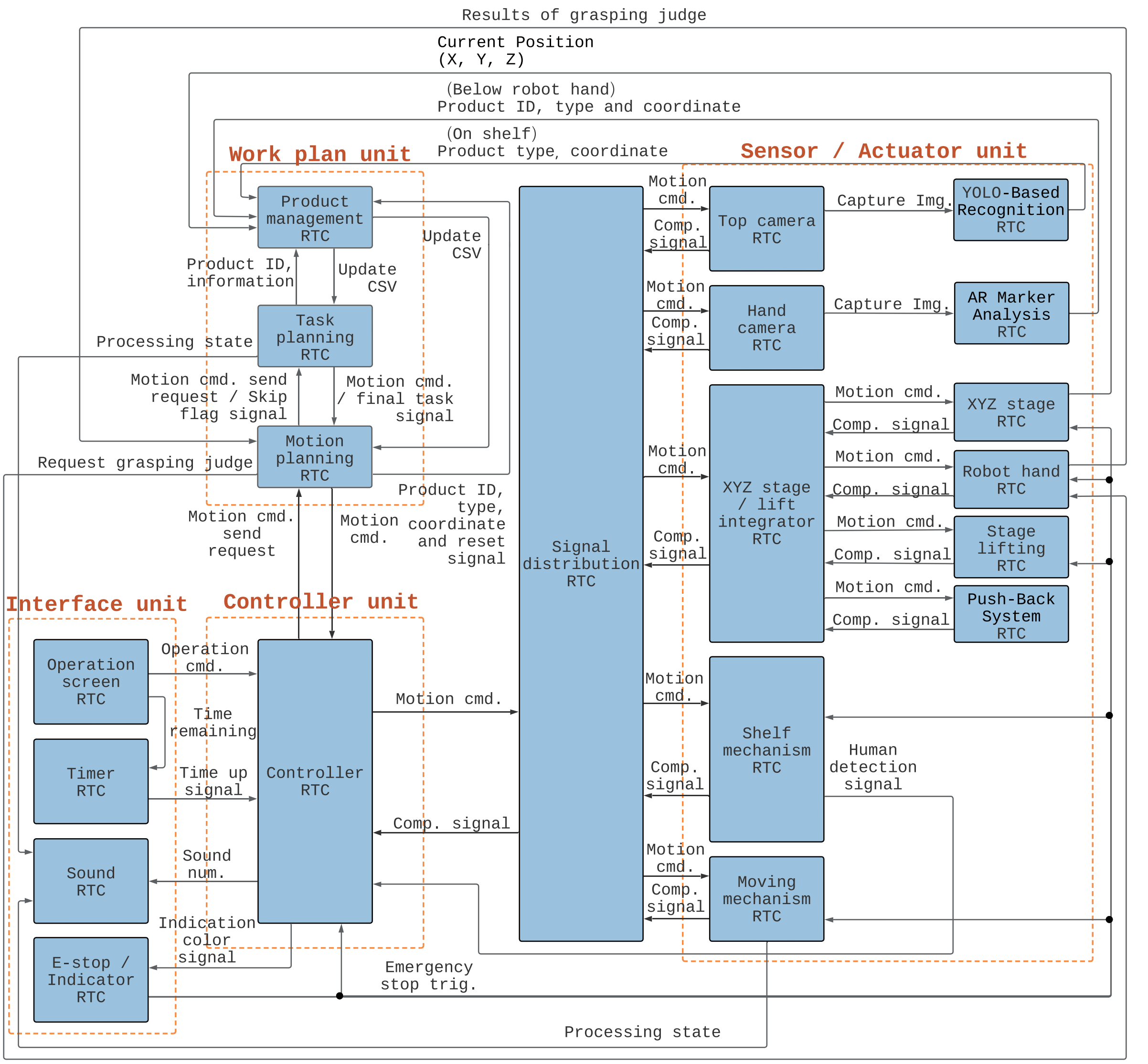


Fig. 2: Software component configuration

B. Work Planning Software Architectures

Sakai et al. introduced a distributed architecture centered on a system manager, which integrates environmental information and perception results to control behavior [4]. Garcia et al. adopted a rule-based strategy based on product characteristics and success rates while constructing a ROS-based integrated architecture [6], [7]. Takubo et al. used the sum of the distance to the grasping target and the distance to the placement target as an evaluation function, minimizing the total travel distance through sequential search [8]. Tomikawa et al. adopt a hierarchical structure with the management system at the apex, under which modules such as human detection, autonomous navigation, intelligent shelf, object pose detection, and manipulation [9].

Based on this investigation, it is considered that the top cameras, which can capture the entire shelf, can identify all products on it by using deep learning-based object recognition. This provides useful information for the system to determine an efficient work procedure. On the other hand, the hand cameras can shoot from a close distance to the products. Therefore, it is thought that small ArUco markers can be used, enabling the acquisition of detailed product information without significantly compromising the package design. Furthermore, dividing the work planning software into Task Planning, Motion Planning, and Product Management is expected to contribute to streamlining development and debugging efforts while enhancing software extensibility.

III. SYSTEM OVERVIEW

This paper describes the XYZ stage-type display system currently under development in our laboratory [2]. The system consists of a product stock robot and stock shelves.

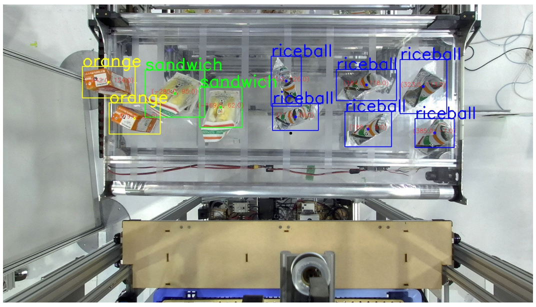


Fig. 3: YOLO recognizing items on the shelf

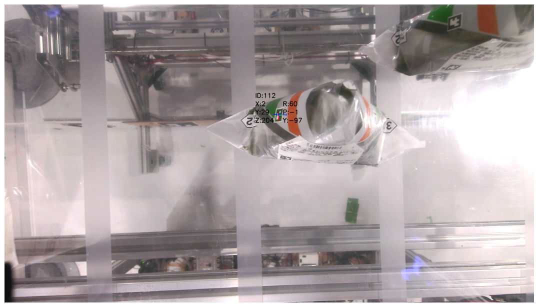
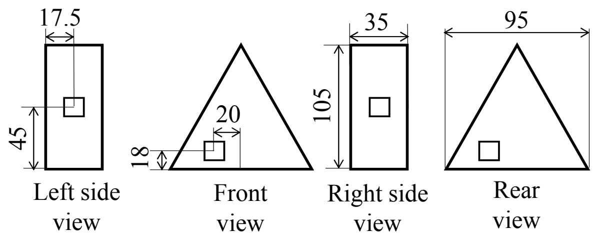
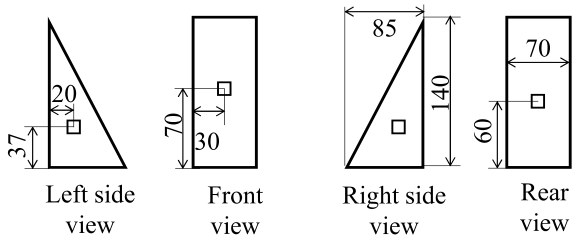


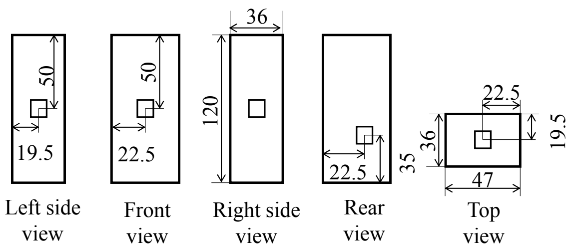
Fig. 4: ArUco markers for product recognition



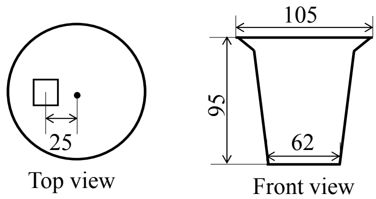
(a) Rice ball



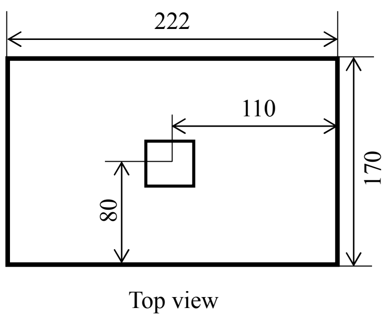
(b) Sandwich



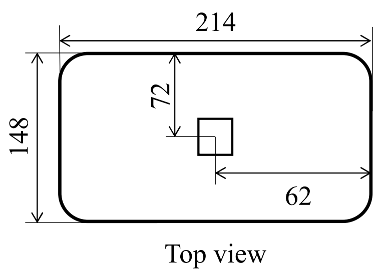
(c) Juice



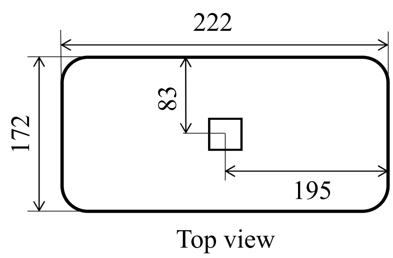
(d) Stick salad



(e) Coleslaw



(f) Lunch box 1



(g) Lunch box 2

Fig. 5: ArUco marker attached position [mm]

A. Hardware Configuration

The robot is a Cartesian coordinate robot equipped with an XYZ stage capable of vertical movement, mounted on a mobile cart. Its end effector features a specialized robotic hand [10] designed to grasp diverse products handled in convenience stores from directly above. All motions are generated in Cartesian coordinates, simplifying motion planning and allowing the robot's range of motion to be easily visualized from its mechanism. To recognize products, the robot is equipped with two cameras. The top camera, mounted above the XYZ stage, captures the entire shelf surface to detect product types and approximate coordinates. Meanwhile, the hand camera attached to the robotic hand captures the ArUco marker affixed to the product, enabling high-precision recognition of the target product's ID, position, and orientation. The XYZ stage can accommodate the actual trays used in convenience stores. The stock shelves are designed specifically for this system and can be automatically deployed via actuators embedded in each shelf board. Furthermore, the shelf boards incorporate a belt conveyor mechanism, enabling automatic face-up orientation of products [11]. The robot recognizes ArUco markers installed at the bottom of the stock shelves to correct its positional relationship with them.

B. *Software Configuration*

The robot is controlled using RT middleware [12]. Software modules called RT component (RTC) are created for each component, and the entire system is constructed by combining multiple RTCs. The system consists of four elements: (A) Controller Unit, (B) Sensor/Actuator Control Unit, (C) Work Planning Unit, and (D) Interface Unit. The controller RTC in the controller unit provides overall control of the entire system. Fig. 2 shows the component structure. The work planning unit is divided into three functions: task planning, motion planning, and recognized product management. These functions collaborate to execute tasks.

IV. PRODUCT RECOGNITION SYSTEM

The recognition function of this system consists of two components: object detection using YOLO and identification using ArUco markers. Product recognition is performed in two stages: overall recognition and detailed recognition. First, the top camera mounted on the robot captures an image of the entire shelf (Fig. 3), and YOLO recognition estimates the position and type of products. Next, based on the position information obtained from YOLO recognition, the hand camera is moved above the target product to capture an image. ArUco marker recognition is then applied to this captured image (Fig. 4) to obtain detailed product information.

1) YOLO-Based Recognition RTC: This RTC uses images captured by the top camera as input, detects the areas and types of products within the images, and transmits the results. Product detection employs YOLOv8n, utilizing a dedicated model pre-trained on target product images to perform estimations. The initial model utilized over 6,800 images taken under varying conditions, including individual shots on desks and shots on densely packed shelves. Furthermore, over 200 images were added shots taken in the venue environment. Training was conducted with epoch 20–100, batch size 16, and image size 1920. The probability that all products can be correctly detected is 85%.

2) ArUco Marker Recognition RTC: This RTC detects ArUco markers present in images captured by a hand camera. It outputs detailed product information, including type, expiration date, location, and orientation, as detection results. Since product shapes vary, the positions and number of attached markers also differ per product. Fig. 5 shows the marker attachment positions for each product. Per competition regulations, the total area of markers affixed to each product is capped at $400\,\mathrm{mm}^2$. Therefore, the actual size of each marker is inherently determined by the number of markers attached. Since the products handled by this system differ in the number and size of markers applied, recognition involves identifying them by combining a unique three-digit number with size information. By pre-assigning numbers and sizes to each marker, the system enables simultaneous operation of markers of different sizes.

V. WORK PLANNING SYSTEM

The work planning function of this system consists of three components: Product Management RTC, Task Planning RTC, and Motion Planning RTC. Below, we describe the role and functions of each component. The connection points are as shown in Fig. 2.

A. *Product Management RTC*

This component creates and updates databases required for task planning and motion planning based on image data acquired by the top camera and hand camera. The input data ports primarily receive the current coordinates of the hand, information about the product captured by the hand camera, information about the product captured by the top camera, and information about the product currently being processed. The output data ports primarily send signals indicating database updates.

This component creates and updates two types of databases. The first is the product management database provided to the task planning component, which registers data for all products detected on shelves by the top camera and for new products whose information is known in advance. Each time the hand camera recognizes a marker, it registers whether the product is for disposal or restock, and updates the current coordinates during each operation. The second is the grasping database provided to the motion planning component, which registers the position, orientation, and type of products detected when images are captured by the hand camera.

B. *Task Planning RTC*

This component determines the order of target products based on the product management database and sends work commands to the motion planning component. The input data ports primarily receive database update signals, task request signals from the Motion Planning RTC, or task skip signals. The output data ports primarily transmit task instructions and information about the currently processed item.

The task instruction includes the target product type and the target coordinates for the hand. At task start, the hand is moved to the target coordinates above the product. After grasping is complete, the target coordinates for product placement are sent. The task order is determined by sorting the database to process taller products first, reducing collision risk.

State transitions are as shown in Fig. 6. Tasks proceed in complete cycles per shelf level, executed in the following steps: (1) After initialization completes, the system enters a standby state. (2) Capture images with the top camera, update the product management database, then transition to disposal tasks. (3) After discarding expired items, transition to restocking operations, rotating non-discarded items to face

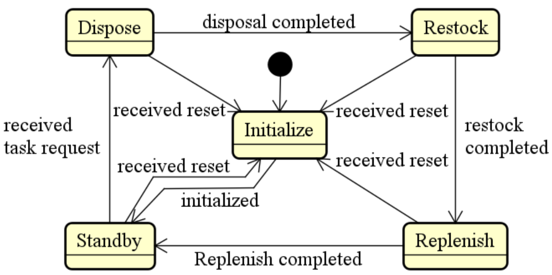


Fig. 6: State transition diagram of the task planning RTC

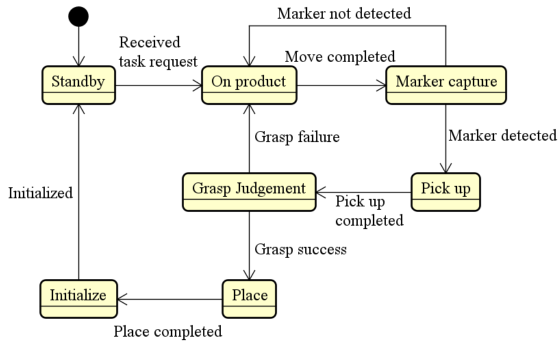


Fig. 7: State transition diagram of the motion planning RTC

TABLE I: Results of the Stock Task

Trial	Salmon Onigiri	Tuna Onigiri	Plum Onigiri	Time [mm'ss"]	Retries	Total [pt]
1	18	18	18	3'47"	0	54
2	18	18	18	3'44"	0	54
3	18	18	18	3'21"	0	54
4	18	18	18	6'55"	1	54

TABLE II: Results of the Stock and Disposal Task

Trial	Face-up (existing)	Disposal collection	Mis-dispose (penalty)	New Item Stocking	All-items Bonus	Infrastructure Bonus	Customer Detection Bonus	Time	Retries	Total
1	6	15	-3	1	0	0	5	9'54"	0	24
2	0	0	0	0	0	0	0	10'00"	1	0
3	6	27	0	1	10	0	5	9'57"	0	49
4	15	9	0	2	0	0	5	9'55"	0	31

forward. (4) After rotating all items, transition to replenishment operations, stocking new products. (5) After completing all operations, return to the waiting state. Repeat this process for the other shelf.

C. Motion Planning RTC

This component plans hardware operations based on instructions from the task planning component and sends commands to the controller component for each actuator. The input data ports primarily receive task commands, task request signals from the Controller RTC, database update signals, and grasping determination results. The output data ports primarily transmit task request signals, grasping determination request signals, and operation commands.

The state transitions are as shown in Fig. 7. The work cycle for a single product is as follows: (1) After initialization completes, it enters a standby state. (2) Upon receiving a work command, it transitions to `onProduct` and moves the hand above the target product. (3) Transitions to `handArUco` to capture images with the hand camera. If marker recognition fails, adjusts position and returns to `onProduct`. (4) Upon successful recognition, transitions to `pickUp` to perform grasping. (5) Performs a grasp judgment in `graspJudge`; if it fails, returns to `onProduct`. (6) If successful, transitions to `place` and places the product at the specified location. (7) After placement, performs initialization processing and returns to the standby state. This cycle is repeated for each work command.

VI. Evaluation in WRS FCSC2025

FCSC2025 was held from July 13 to 19, 2025, at the EXPO Messe “WASSE” venue of the Osaka-Kansai Expo, with eight teams participating [3]. During the event, two types of tasks were conducted: the Stock Task and the Stock and Disposal Task. Rankings were determined by the total points earned from each task [13].

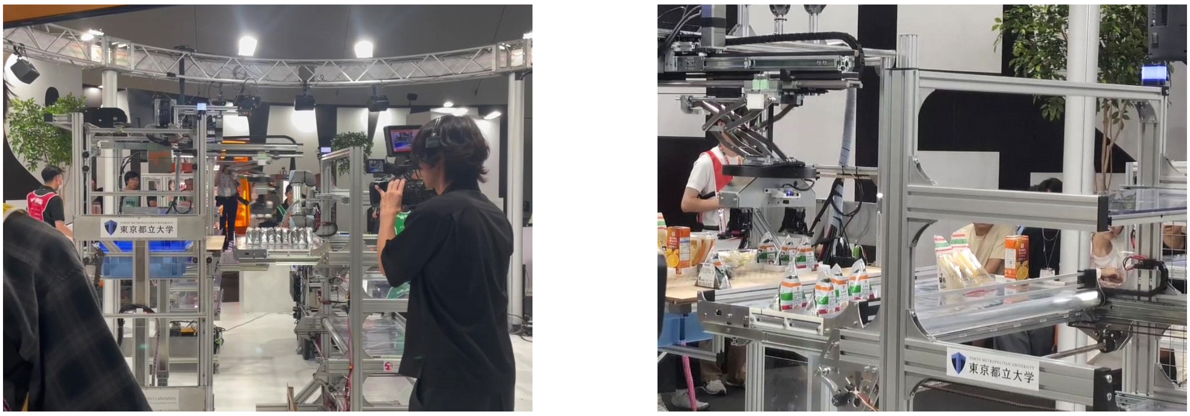
In the Stock Task, 54 rice balls placed in a tray must be accurately positioned on designated spots on the stock shelf. In the Stock and Disposal Task, for products on the shelf, tasks include disposing of expired items, aligning remaining products (face-up), and displaying new products.

During the preliminary round, each task was performed three times, with the top four teams by total score advancing to the finals. Our team's results for the Stock Task and Stock and Disposal Task are shown in Table I and Table II. In the preliminary round, we scored a perfect 54 points in the Stock Task and 49 points in the Stock and Disposal Task, totaling 103 points to rank 1st out of all 8 teams.

In the finals, we achieved a perfect score of 54 points in the Stock Task and 31 points in the Stock and Disposal task, securing the overall championship.

VII. Discussion

Of the four attempts at the competition, the second attempt at the Stock and Disposal Task scored zero points due to hardware failure that could not be resolved within the



(a) Stock Task

(b) Stock and disposal Task

Fig. 8: Demonstration of the developed robot in the competition

competition time. The other three attempts yielded varying scores. Notably, the first attempt, which scored 24 points, involved numerous grasping failures. The causes of these failures are discussed below. The most frequent failure involved the gripper hand interfering with products surrounding the target item. A particularly common scenario involved two bento boxes stacked, with one positioned diagonally. The correct procedure would be to first either grasp the diagonally stacked bento box or reorient it to eliminate the stack. However, attempts to grasp the lower bento box often resulted in interference with the stacked bento box above. For items other than bento boxes, when products are densely packed, the likelihood of interference with non-target items increases. To address this issue, it is considered necessary to recognize the position and orientation of surrounding items around the target object, determine the hand's posture to avoid interference, and implement a flexible task planning mechanism that switches to a different target object if no non-interfering hand posture is found.

VIII. CONCLUSIONS

This study improved the XYZ stage-type display system for automatically displaying and disposing of convenience store merchandise. For product recognition, a two-stage recognition method combining object detection using YOLO and identification using ArUco markers was introduced. Furthermore, the work planning program was divided into three components: product management, task planning, and motion planning to enhance readability and debugging efficiency. This system competed in the WRS2025 Future Convenience Store Challenge, held at the Osaka-Kansai Expo, and won first place, as well as the Stock and Disposal Task Winner. Going forward, we will analyze cases where tasks were incomplete during the competition. By improving recognition accuracy and the flexibility of the task planning, we aim to achieve more stable task execution.

ACKNOWLEDGMENT

This research was supported by TMU local 5G research support.

REFERENCES

[1] K. Miwa, Study of Operation and Staff Scheduling in Retail Store, *The Nagoya Gakuin University review. Social sciences*, vol. 51, No. 4, pp. 143–158, 2015. (In Japanese)

[2] M. Seki, K. Wada, Y. Kitajima, M. Hashimoto, and T. Tomizawa, Development of XYZ stage-type display robot system for stock and disposal tasks in convenience stores, *Adv Robot.*, vol. 36, no. 23, pp. 1252-1272, 2022.

[3] WRS Future Convenience Store Challenge 2025. https://worldrobotsummit.org/en/wrs2025/fcsc/ [Accessed August 26, 2025]

[4] R. Sakai, S. Katsumata, T. Miki, T. Yano, W. Wei, Y. Okadome, et. al., A mobile dual-arm manipulation robot system for stocking and disposing of items in a convenience store by using universal vacuum grippers for grasping items, *Adv Robot.*, vol. 34, no. 3-4, pp. 219-234, 2020.

[5] J. Tanaka, D. Yamamoto, H. Ogawa, H. Ohtsu, K. Kamata and K. Nara, Portable compact suction pad unit for parallel grippers, *Adv Robot.*, vol. 34, no. 3-4, pp. 202-218, 2020.

[6] G. A. Garcia Ricardez, S. Okada, N. Koganti, A. Yasuda, P. M. Uriguen Eljuri, T. Sano, et. al., Restock and straightening system for retail automation using compliant and mobile manipulation, *Adv Robot.*, vol. 34, no. 3-4, pp. 235-249, 2020.

[7] G. A. Garcia Ricardez, P. M. Uriguen Eljuri, Y. Kamemura, S. Yokota, N. Kugou, Y. Asama, et. al., Autonomous service robot for human-aware restock, straightening and disposal tasks in retail automation, *Adv Robot.*, vol. 36, no. 17-18, pp. 936-950, 2022.

[8] T. Takubo, T. Nakamura, R. Sugiyama and A. Ueno, Multifunctional Shelf and Magnetic Marker for Stock and Disposal Tasks in Convenience Stores, *Journal of Robotics and Mechatronics*, vol. 35, no. 1, pp. 18-29, 2023.

[9] R. Tomikawa, Development of Display and Disposal Work System for Convenience Stores Using Dual-Arm Robot, *Adv Robot.*, Vol. 00, No. 00, June 2021, 121.

[10] M. Seki, K. Wada, T. Teraguchi and T. Tomizawa, A robot hand with free-rotating grips for switching three different grasping modes, *Adv Robot.*, vol. 36, no. 17-18, pp. 920-935, 2022.

[11] M. Seki, et al., Development of Automated Display Shelf with Conveyor Belt Type Product Face-Up Mechanism, 2024 IEEE/SICE International Symposium on System Integration (SII), Date of Conference: 08-11 January 2024

[12] OpenRTM-aist. https://www.openrtm.org/openrtm/en [Accessed August 26, 2025]

[13] WRS Future Convenience Store Challenge 2025 Rulebook. https://worldrobotsummit.org/en/wrs2025/fcsc/FCSC2025_stock_disposal_task_v002_ENG.pdf [Accessed: August 26, 2025].