Happy Robot 2026 Team Description Paper

Kosei Demura Aoi Hayashi Takuya Shimada Yuta Okitsu
Kouki Sugimoto Masaya Wada Zinnsei Arai
Yamato Koizumi Natsuki Sasaki Keitatsu Sawanobori
Yoshiki Takahashi Masami Takizawa Yukitaka Tachibana
Sho Taniguchi Masaya Watanabe

February 10, 2026

Abstract. This paper presents the Happy Robot team ' s system for RoboCup@Home 2026, built on our open-hardware platform, Happy Edu, which lowers the entry barrier for new teams through a modular and extensible design. The robot integrates multimodal perception for speech interaction, object understanding, and human-aware recognition. We further introduce key research contributions: a cost-effective 3D LiDAR–based re-identifiable human-following system, a foundation-model-driven planning framework for shelf management, a robot hand capable of four primitive motions for Bento assembly, and a Fluorescent AR Marker–based method for automatic 6 DoF pose annotation and estimation. In addition, we are working on fine-tuning robot foundation models toward RoboCup tasks. These developments advance practical domestic service robotics.

1 Introduction

The Happy Robot team has participated in the RoboCup@Home league of the RoboCup Japan Open since 2012 and has competed in the RoboCup World Competition since 2015. Our results include 9th place in 2015, 8th in 2016, 9th in 2017, and 5th in 2018. Participation in the world competition was suspended from 2020 to 2022 due to the COVID-19 pandemic. In 2023, we took part in the @Home Education Workshop and Challenge at RoboCup 2023 Bordeaux and obtained 2nd place. In 2024, we participated in the @Home Playground at RoboCup 2024 Eindhoven and achieved another 2nd place. These events serve as an educational bridge between RoboCup Junior and RoboCup@Home, providing undergraduate students with meaningful research experience and contributing to the continuous development of the @Home league.

In addition, our team has been actively engaged in the World Robot Summit (WRS) Future Convenience Store Challenge (FCSC), a competition evaluating advanced robotic technologies for realistic convenience-store operations.

The challenge focuses primarily on shelf-stocking and disposal tasks, requiring accurate item recognition, dependable grasping, and precise placement or removal of expired products. In 2024, our team, Happy Robot, earned first place in the FCSC, a competition that featured leading RoboCup@Home teams such as Hibikino-Musashi and Er@sers. This accomplishment highlights our capability to integrate perception, manipulation, and task planning into a highly reliable system. Our robot consistently demonstrated robust recognition and manipulation performance across diverse products, satisfying the stringent operational standards expected in the competition.

The Happy Robot team is jointly organized by the Demura Laboratory in the Department of Robotics and the Yumekobo Projects at the Kanazawa Institute of Technology (KIT). Yumekobo, known as the “factory for dreams and ideas,” is a distinctive educational initiative established in 1993 to foster creativity, hands-on engineering skills, and character development among students. A central activity of Yumekobo is the support of project-based student teams—known as Yumekobo Projects—that aim to cultivate technical proficiency, collaboration, and leadership.

The mission of the Happy Robot team is to develop robots that bring happiness and comfort to people. Our robot, shown in Fig. 1, is intentionally designed with the appearance of a small child and features a bright yellow color scheme to emphasize friendliness and emotional approachability. We believe that domestic service robots should embody designs that are welcoming to both older adults and children. Since 2015, our team has pioneered this design philosophy within the RoboCup@Home league.

The remainder of this paper is organized as follows. Section 2 details the hardware architecture of our robotic platform. Section 3 describes the software architecture. Section 4 presents our research contributions. Section 5 concludes the paper with a discussion of applications and outlines future directions.

2 Open Hardware

Since 2015, our robots have pioneered a distinctive design concept within the RoboCup@Home league, adopting a friendly, approachable, and childlike appearance intended to promote familiarity and emotional acceptance among children, women, and elderly individuals in domestic environments. Building on this design philosophy, we began developing Happy Edu in 2023—a new small, lightweight, and cost-effective open hardware robot, shown in Fig.1, designed to reduce entry barriers for new teams in the @Home league.

In the RoboCup Japan Open, the Education League, introduced in 2015 as an entry-level league for the @HomeLeague, was renamed the Bridge Competition in 2025 to further support the development of the @Home ecosystem. However, the number of new entrant teams has remained limited. A major contributing factor is the dependence on the TurtleBot2 (Kobuki), a previously affordable platform that is no longer in production, making it increasingly difficult to obtain. Consequently, suitable robot platforms for beginner teams have be-

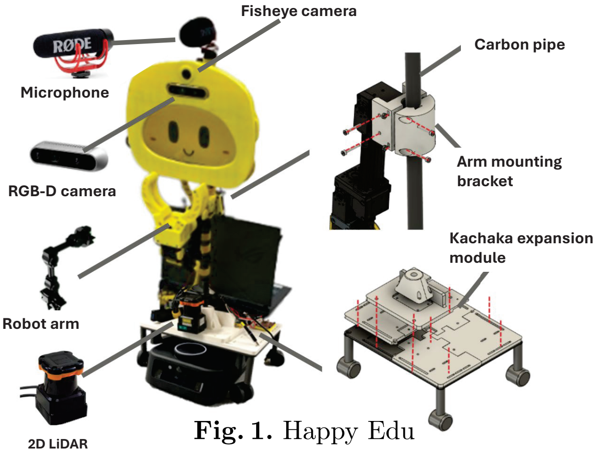


Fig. 1. Happy Edu

come scarce. To address this issue, we developed Happy Edu, an open-hardware robot based on the Kachaka platform, as a practical and accessible successor to TurtleBot2-based systems.

Happy Edu consists of the Kachaka mobile base and three primary modules: the torso, arm, and head. Technical details are provided in the Annex. To equip the robot with the functions required for competition, we fabricated mounting fixtures for components such as cameras and robotic arms using an FDM 3D printer. These fixtures are clamped to the carbon-pipe torso using four screws, resulting in a lightweight, modular structure that facilitates rapid on-site assembly and maintenance.

The original Kachaka docking base offered mounting holes only at its ends, limiting hardware expandability. To overcome this constraint, we designed a resin plate with embedded nuts, enabling flexible placement of devices including the onboard PC, robotic arm, and sensor modules.

All expansion components are provided as open-source CAD data and can be manufactured using a 3D printer or outsourced fabrication services [1]. The CAD data can be freely modified—for example, to add custom mounting holes—allowing new @Home teams to rapidly build or adapt their robots while significantly reducing development time.

3 Software

3.1 Speech Recognition and Speech Synthesis

For speech recognition, we use Whisper, a Transformer-based ASR model developed by OpenAI. Trained on large-scale multilingual datasets, Whisper offers high robustness to noise, speaker variation, and reverberation, making it suitable for domestic service-robot environments in RoboCup@Home. Its accurate timestamp alignment and reliable command transcription improve downstream task execution.

For speech synthesis, we employ Mimic3, an open-source TTS engine based on the VITS architecture. VITS integrates a variational autoencoder, normalizing flows, and adversarial learning in an end-to-end framework, enabling natural-sounding waveform generation without intermediate spectrograms. Mimic3 provides clear prosody, high intelligibility, and low-latency inference, allowing smooth and responsive human–robot interaction on embedded hardware.

Together, Whisper and Mimic3 form a robust speech interface that supports reliable command understanding and expressive verbal responses in real-world home environments.

3.2 Recognition

For natural language understanding, we employ RoBERTa, an enhanced variant of BERT fine-tuned for sentiment and intention analysis, enabling reliable interpretation of user utterances. For visual scene understanding, a Generative Image-to-Text Transformer combines a CLIP encoder with a Transformer decoder to produce semantic captions, supporting contextual reasoning and scene explanation.

For human-aware perception, MTCNN is used for face detection, followed by DeepFace for estimating emotion, age, and gender, which supports personalized and socially appropriate interaction. For object perception, GroundingDINO provides open-vocabulary detection based on natural-language prompts, allowing recognition of unseen objects without additional training.

These four models collectively form a complementary perception pipeline for the RoboCup@Home league: RoBERTa interprets speech intent, the Image-to-Text model supplies semantic scene descriptions, MTCNN/DeepFace handle human-centered recognition, and GroundingDINO enables flexible object detection. Together, they support tasks such as locating requested objects, understanding user context, and generating appropriate robot responses in domestic environments.

4 Research Contribution

4.1 Re-identifiable Human-following System Using 3D LiDAR

We have been developing a re-identifiable human-following system using 3D LiDAR. The system consists of two main components: human detection and re-identification. For detection, we employ PointPillars, whose performance is enhanced by training with a custom dataset to mitigate domain gaps. For re-identification, we use ReID3D, which distinguishes individuals based on physical attributes such as height and clothing, as well as gait features, enabling robust re-identification even after occlusions.

In our implementation, we use the Livox Mid-360, a cost-effective 3D LiDAR sensor. Because ReID3D was originally trained on datasets captured with

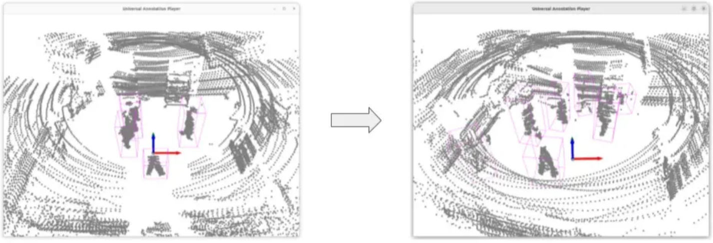


Fig. 2. Data arugumentation for fine-tuning PointPillars

Table 1. Performance comparison of pedestrian AP [%]

Method	Metric	Pedestrian AP
PointPillars	3D-BBox	47.9446
	BEV	54.3456
Proposed	3D-BBox	46.5419
	BEV	62.4932

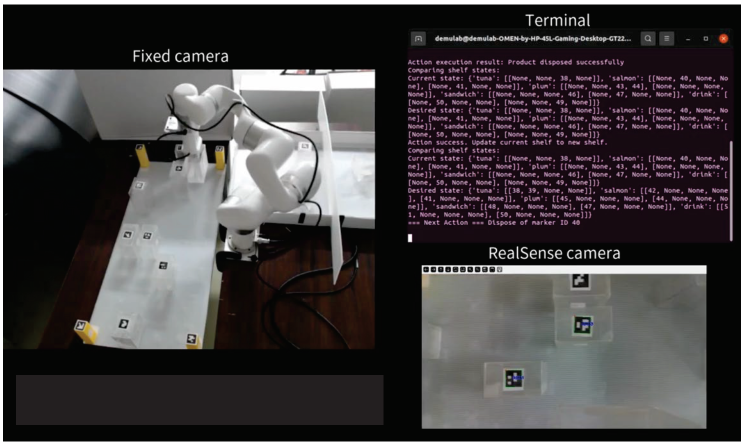
Velodyne LiDAR, a noticeable domain gap arises when applying it directly to Livox data. To address this, we perform fine-tuning of ReID3D using Livox-based data. Specifically, as shown in Fig. 2, we generated augmented training samples by adding human point clouds to diversify the data distribution. As a result, and as summarized in Table 1, this data augmentation strategy enabled us to obtain significantly improved AP scores, although these results are still preliminary as the system remains under development.

This study represents a significant contribution to the @Home league, where traditional systems have relied primarily on 2D LiDAR. By enabling robust human following with 3D LiDAR, and by demonstrating that high-level person re-identification can be achieved using an affordable sensor, our approach lowers the hardware barrier and advances the reliability of human–robot interaction in real domestic environments.

4.2 Foundation-Model-Based Planning System for Shelf Display and Disposal

This study proposes an automated planning system for robotic shelf display and disposal tasks in convenience stores to address labor shortages and the limitations of rule-based approaches in dynamic retail environments. The system combines perception modules, foundation-model-based reasoning, and a closed-loop execution architecture to achieve adaptive and reliable task performance (Fig. 3) [3].

For environment perception, an arm-mounted RGB camera observes the shelf, and ArUco markers provide product identifiers. Homography estimation reconstructs the shelf plane regardless of camera pose, enabling grid discretization. Each detected product is mapped to the nearest cell to form a structured



```
Action execution result: Product disposed successfully
Comparing shelf states:
Current state: {'tuna': [[None, None, 38, None]], 'salmon': [[None, 40, None, No
ne], [None, 41, None, None]], 'plum': [[None, None, 43, 44], [None, None, None,
None]], 'sandwich': [[None, None, None, 46], [None, 47, None, None]], 'drink': [
[None, 50, None, None], [None, None, 49, None]]}
Desired state: {'tuna': [[None, None, 38, None]], 'salmon': [[None, 40, None, No
ne], [None, 41, None, None]], 'plum': [[None, None, 43, 44], [None, None, None,
None]], 'sandwich': [[None, None, None, 46], [None, 47, None, None]], 'drink': [
[None, 50, None, None], [None, None, 49, None]]}
Action success. Update current shelf to new shelf.
Comparing shelf states:
Current state: {'tuna': [[None, None, 38, None]], 'salmon': [[None, 40, None, No
ne], [None, 41, None, None]], 'plum': [[None, None, 43, 44], [None, None, None,
None]], 'sandwich': [[None, None, None, 46], [None, 47, None, None]], 'drink': [
[None, 50, None, None], [None, None, 49, None]]}
Desired state: {'tuna': [[38, 39, None, None]], 'salmon': [[42, None, None, None
], [41, None, None, None]], 'plum': [[45, None, None, None], [44, None, None, No
ne]], 'sandwich': [[48, None, None, None], [47, None, None, None]], 'drink': [[5
1, None, None, None], [50, None, None, None]]}
=== Next Action === Dispose of marker ID 40
```

Fig. 3. Execution results of the proposed system

array representation of the shelf state, which is embedded into a dynamically generated prompt for the foundation model.

For planning, the system uses OpenAI ' s o1-mini foundation model, which receives static task rules and the updated shelf state. It generates high-level actions—such as removing expired items and placing new products—that are parsed into executable robot commands. After each step, the robot re-observes the shelf and updates the prompt, forming an adaptive, feedback-driven loop.

Simulated and real-robot experiments show that the system produces near-optimal action sequences, maintains robustness under varying success rates, and can perform complete shelf display and disposal operations. Remaining challenges are mainly due to perception errors. Overall, the results demonstrate that foundation-model-driven planning offers a flexible and scalable alternative to conventional rule-based methods for real convenience-store environments.

4.3 Development of a robot hand with four primitive motions for efficient assembly of various ingredients in Bento boxes

In recent years, numerous robotic hands have been developed for food-handling applications. Nevertheless, arranging a wide variety of ingredients within a Bento lunch box remains a significant challenge due to the limited dexterity and substantial finger thickness of conventional grippers. Robot hands designed for general-purpose food manipulation often lack the precision and versatility required for the detailed assembly tasks involved in Bento preparation.

To address this limitation, we developed a novel robotic hand that incorporates an additional linear-motion degree of freedom into a standard two-jaw gripper as shown in Fig. 4[2]. This mechanism enables the execution of four fundamental manipulation primitives—grasping, releasing, pulling, and sliding off—which we identify as essential for the effective arrangement of Bento ingredients.

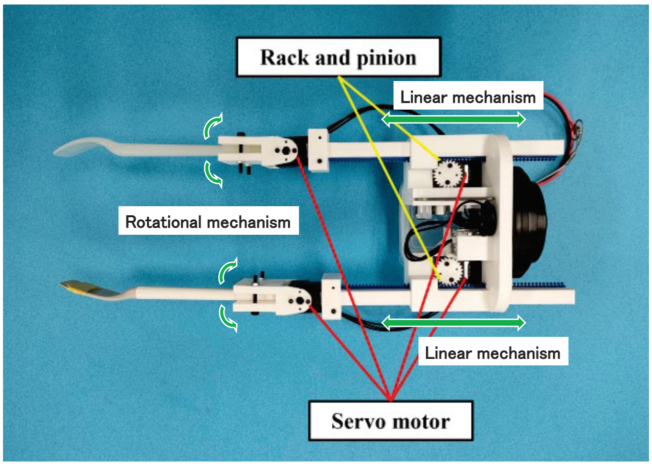


Fig. 4. Developed robotic hand for Bento box assembly

We conducted a series of systematic experiments using a single-arm robotic platform equipped with a vision system based on the Segment Anything Model. The experimental results demonstrate that the proposed hand successfully performs all four primitive motions, thereby overcoming the operational constraints of conventional grippers. Notably, the pulling motion enabled reliable manipulation of thin food items that are difficult to handle through grasping alone. Furthermore, the proposed hand successfully arranged all ingredients in the Bento box used in our study, indicating its strong potential for practical deployment in Bento assembly tasks.

4.4 The Fluorescent AR Marker

We developed a 6DoF pose estimation network using a dataset automatically annotated with our Fluorescent Augmented Reality (AR) Marker [4], as shown in Fig. 5. The marker consists of a transparent film coated with fluorescent paint that becomes visible under ultraviolet (UV) illumination while remaining invisible under normal lighting. By alternating visible and UV light, the system automatically generates RGB images, segmentation masks, and ground-truth 6DoF pose information, enabling efficient creation of large-scale datasets that are difficult to obtain with conventional methods.

Using this dataset, we trained a 6DoF pose estimation network that learns object poses with the marker but can estimate them without the marker during inference. The average positional error is 7.2 mm, comparable to traditional AR marker–based systems. This method significantly reduces the cost and effort of collecting high-quality pose data and provides a practical solution for manipulation tasks. Future work will extend the approach to transparent and reflective objects, which remain challenging for existing 6DoF pose estimation techniques.

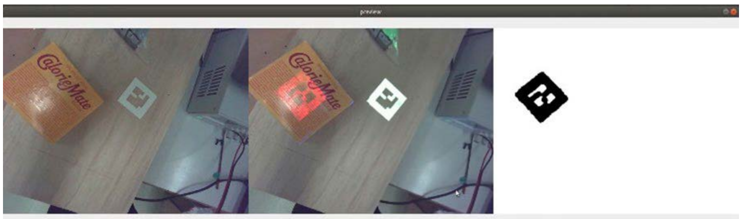


Fig. 5. Fluorescent AR Marker: The left image is captured without UV illumination, the center image with UV illumination, and the right image represents the difference between them

5 Conclusion

This paper presented the Happy Robot team's key developments for RoboCup@Home 2026. Our open-hardware platform, Happy Edu, provides an accessible solution for new teams, while our integrated perception system supports robust multimodal understanding. Research contributions include:

- a robust and cost-efficient 3D LiDAR–based human-following system
- a foundation-model-driven planning method for shelf operations
- a robot hand with four primitive motions for Bento assembly
- an automatic 6DoF annotation and estimation method using a Fluorescent AR Marker

Together, these contributions demonstrate practical effectiveness and potential for further advancement. Future work will focus on improving perception accuracy, manipulation ability, and performance on transparent or reflective objects. In addition, we are working on fine-tuning the robot foundation model $\pi_{0.5}$ using LoRA on the HSR robot for RoboCup tasks.

References

1. Happy Edu CAD data, `https://github.com/demulab/happy_edu_cad_data.git`, (accessed February 10, 2026).
2. A. Hasegawa, K. Demura: Development of a robot hand with four primitive motions for efficient assembly of various ingredients in Bento boxes, Advanced Robotics, 39(8) 441-456, 2025.
3. Y. Ishiyama, K. Demura: Shelf Display and Disposal by a Robotic Arm (in Japanese), Proceedings of the 42nd Annual Conference of the Robotics Society of Japan (RSJ), 3H4-01, 2024.
4. S. Okano, T. Makino, K. Demura: Fluorescent Texture: Proposal of a 2-3D Automatic Annotation Method for Deep Learning (in Japanese), Journal of the Robotics Society of Japan, vol.40, no.1, pp.71-82, 2022.

Annex

Happy Edu Description

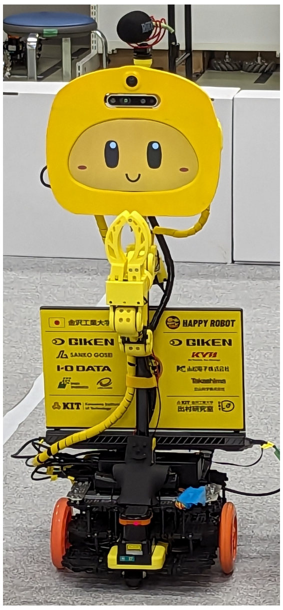


Fig. 6. Happy Edu

- Happy Edu is an open-hardware robot developed by the Demura Laboratory, as shown in Fig.6.
- **Hardware descripton**
  - Height: 0.9 [m], Width: 0.4 [m], Length: 0.4 [m], Weight 18:[kg]
  - Base: Kachaka, Differential pair. Max velocity is 0.8 [m/s].
  - Manipulators: Current, an arm is 4 DoF and a hand is a DoF. Payload is 0.5 [kg]. This manipulator is not sufficient for RoboCup@Home task, so 6 DoF arm is devlopping.
  - Torsos: A carbon pipe. A motorized cane will be used for the lifting mechanism
  - Heads: Equipped with an LCD display, an RGB-D camera, and a fisheye camera, and includes a tilt mechanism.
  - LiDAR: Current Hokuyo UTM 30 LX. 3D LiDAR, Livox Mid-360, will be used.
  - RGB-D camera: RealSense D435i
  - Fisheye camera: ELP 180 [°] USB camera
  - Microphones:RODE VideoMic Go
  - Display: 10.1 inch IPS LCD display

  - Computer: Dell Alienware R16 (GPU: Nvidia RTX 4090 Mobile)
- **Software description**
  - Automated speech recognition: Whisper
  - TTS: Mimic3
  - Manipulation: ROS2 MoveIT2
  - Natural Language Processing: ChatGPT API
  - Navigation, localization, and mapping: ROS2 Nav2
  - Object recognition: Detic
  - Object segmentation: SAM
  - People recognition: Grounding Dino
  - People tracking: Self-developing based on Reid3D and Pointpillars
  - Pose/Gesture recognition: Media Pipe