Proceedings of the 2025 SICE Festival with Annual Conference
September 9-12, 2025, Chiang Mai, Thailand

Proposal of a pedestrian recognition method for mobile robots using drones flying overhead and a pedestrian avoidance navigation method based on it

Masafumi Yoneoka[1†], Linghao Ye[1], Daisuke Chugo[1], Satoshi Muramatsu[2], Sho Yokota[3] and, Hiroshi Hashimoto[4]

[1]Graduate School of Science and Technology, Kwansei Gakuin University, Sanda, Hyogo, Japan
(Tel: +81-070-2311-7430; E-mail: {yoneoka, yelinghao, chugo}@chugolab.com)
[2]Department of School of Information Science and Technology, Tokai University, Hiratsuka, Kanagawa, Japan
(Tel: +81-463-58-1211; E-mail: muramatsu@tokai.ac.jp)
[3]Department of Faculty of Science and Engineering, Toyo University, Kawagoe, Saitama, Japan
(Tel: +81-49-239-1420; E-mail: s-yokota@toyo.jp)
[4]Department of Master Program of Innovation for Design and Engineering, Advanced Institute of Industrial Technology, Shinagawa, Tokyo, Japan
(Tel: +81-3-3472-7831; E-mail: hashimoto@aiit.ac.jp)

Abstract: In crowded urban environments, service robots need to guide pedestrians safely while avoiding them. Traditional approaches such as artificial potential field methods enable real-time path planning at low computational cost, but they assume full visibility of pedestrians using on-board sensors such as LiDAR and cameras. In practice, these sensors suffer from occlusion and limited field of view, reducing safety and reliability. To overcome this limitation, we propose a novel pedestrian-aware navigation system that combines a drone-based bird's-eye view with YOLOv8 object detection and artificial potential method. The drone captures overhead images of the environment, from which YOLOv8 detects the positions of pedestrians and mobile robots in real time. The geometric relationship between the drone and the ground is modeled using a pinhole camera model to accurately calculate the coordinates of the pedestrian. These coordinates are transmitted to the robot via ROS, enabling global situational awareness and safe path planning even when on-board sensors fail due to occlusion. Experimental validation, demonstrates that the proposed system accurately detects all pedestrians in the vicinity and enables early and effective collision avoidance. The robot successfully navigates complex scenarios in which pedestrians suddenly appear or do not have direct line-of-sight contact, validating the effectiveness of the proposed drone-assisted navigation system.

Keywords: Drone-based detection, YOLOv8, Pedestrian-aware navigation, Artificial potential field

1. INTRODUCTION

In recent years, demand for service robots has been rapidly increasing. These service robots need to safely avoid pedestrians around them while navigating to their destinations in crowded and complex environments such as sidewalks, parks, and residential areas.

In previous studies, artificial potential methods have been used as a way for mobile robots to avoid pedestrians [1]. This method has low computational costs and enables real-time path planning for mobile robots. Furthermore, a potential method has been proposed that synchronises with surrounding pedestrians in crowded environments and generates paths that follow the overall pedestrian flow [2].

However, these previous studies assume that the robot can detect all pedestrians in the environment. Although the effectiveness of this approach was confirmed in a simulation experiment in [2], in actual environments, robots are often equipped with external sensors, and their installation positions are limited by the size of the robot itself. In many cases, external sensors (such as laser range finders (LRFs) and CCD cameras) are used, but these sensors are these sensors are sensitive to occlusion. Therefore, even if a mobile robot avoids a person in front of it, there is a high probability of colliding with other pedestrians that were not detected due to obstructions.

While methods using fixed cameras on ceilings or high places have been proposed [3], they are limited by a fixed field of view and high installation costs. Although YOLO [4] and potential functions [1] are established techniques, this study's novelty lies in the direct integration of a drone's real-time aerial view with a robot's navigation logic. This approach uniquely overcomes the occlusion and range limitations inherent to ground-based sensors. By combining these elements, this work develops a path planning strategy that allows a mobile robot to safely navigate complex environments with a comprehensive understanding of its surroundings.

This approach utilizes a drone to detect the mobile robot and all surrounding people from a bird's-eye view using the real-time object detection algorithm YOLO [4]. By integrating this approach with the aforementioned potential method, we develop a path planning strategy that enables a mobile robot to navigate a complex environ-

† Masafumi Yoneoka is the presenter of this paper.

ment with many people while maintaining the natural pedestrian flow of movement.

2. ROBOT SYSTEM ARCHITECTURE

2.1. Summary of the Proposed Aerial Monitoring Approach Using Drones

In this system, an autonomous drone is placed above a mobile robot to detect pedestrians and mobile robots on the ground from a bird's-eye view. The position information of detected pedestrians is sent to the mobile robot, allowing the mobile robot to avoid the pedestrians. The configuration of this robot system is shown in Fig. 1.

The developed system utilizes a DJI Air 2S drone, which has a camera with a horizontal field of view of 88 degrees and can capture the ground over a 29m×16m area from a height of 15m. The Air 2S was chosen for its sufficient performance to meet the requirements of our approach. The drone transmits its captured images to a laptop computer wirelessly via HDMI using a DJI controller RC PRO. The laptop runs both YOLO and ROS, where YOLO performs real-time detection of pedestrians and mobile robots in the video feed. Subsequently, the processed detection results are used by ROS to generate a route for the mobile robot.

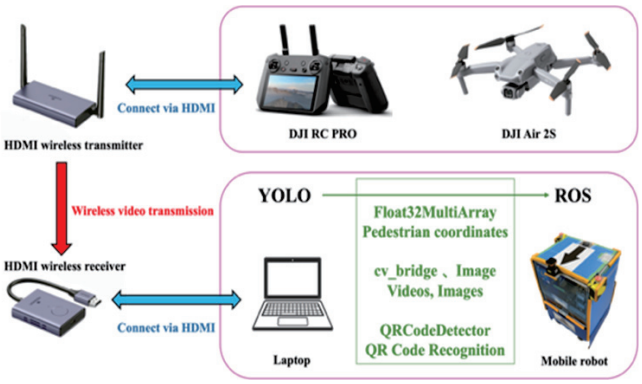


Fig. 1: Robot system configuration .

2.2. Image Detection Techniques

For this system, YOLO is used to recognize and track pedestrians and moving robots. Specifically, YOLOv8 is adopted, the latest version of YOLO, which has a small computational complexity and the highest detection accuracy [5].

YOLOv8 is provided with five types of trained models: n, s, m, l, and x. The YOLOv8s model is used in consideration of the GPU load to install YOLOv8 on a laptop. To recognize a moving robot from a drone image, the system uses the image annotation tool Labelimg to mark pedestrians in the image, as shown in Fig. 2. We generate a position file of the marked pedestrians and train the YOLO model along with the original image. In this preliminary study, 220 photos were prepared.

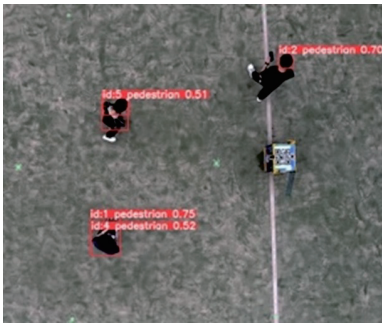


Fig. 2: Pedestrian recognition results.

3. PEDESTRIAN RECOGNITION

3.1. Handling Pedestrian Coordinates

When flying a drone at a high altitude, the size of the robot and pedestrians on the screen changes depending on the camera angle and the drone's flight altitude. Therefore, taking into account their geometric relationship, we derive the coordinates of the pedestrians using Eq. (1) [6].

Fig. 3 represents a pin-hole camera model for an optical camera mounted at a drone in the air, where W and H are width and height of monitoring range, X and Y are width and height of a target object on the ground, D is an altitude of the drone, Θ_x is a viewing angle of the optical camera according to the width orientation, F is a focal length between a camera lens and an image sensor, w and h is width and height of the image sensor, x and y is width and height of the target object captured at the image sensor, m and n is the number of pixels for width and height of the captured image, and x_m and y_m are the number of pixels for width and height of the target object on the capture image.

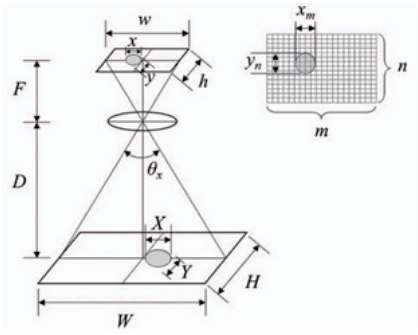


Fig. 3: Optical camera mounted at a drone [6].

$$
\begin{aligned}
x_m &= f\left(\frac{m}{w}x\right) = f\left(\frac{m}{w}\cdot\frac{wX}{W}\right) \\
&= f\left(\frac{mX}{2D\tan\left(\frac{\theta_x}{2}\right)}\right) = f\left(\frac{-mX}{2D\left(\frac{w}{2F}\right)}\right) \\
&= f\left(\frac{mXF}{wD}\right) \qquad (1)
\end{aligned}
$$

W Width of the monitoring range
H Height of the monitoring range
X Width of a target object on the ground
Y Height of a target object on the ground
D Altitude of the drone
θ_x Viewing angle of the optical camera according to the width orientation
F Focal length between a camera lens and an image sensor
w Width of the target object captured at the image sensor
h Height of the target object captured at the image sensor
x Width of the target object captured at the image sensor
y Height of the target object captured at the image sensor
m Number of pixels for width of the captured image
n Number of pixels for height of the captured image
x_m Number of pixels for width of the target object on the captured image
y_n Number of pixels for height of the target object on the captured image

As shown in Fig. 3, the drone captures riverside areas in vertical view and the size of the target object projected on the aerial image is related to lots of parameters. Based on the pinhole camera model, the width size of the target object in pixels can be represented by the related parameters as follows.

3.2. Preliminary Experiment

To evaluate the detection accuracy of the proposed system, three preliminary experiments were conducted on the Sanda Campus of Kwansei Gakuin University. The front of the robot was defined as the positive Y-axis direction.

3.2.1 Experimentation Setup

In the three experiments, a pedestrian either approached the robot from 5m to 1m, moved away from 1m to 5m, or stood still at 4m in front of the robot. Each experiment was repeated three times, with the robot remaining stationary and only the pedestrian in motion.

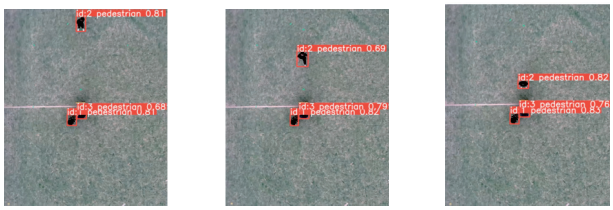
A drone was automatically maintained at an altitude of 15m directly above the robot, with an altimeter accuracy of 0.1m. The robot was equipped with a 2D LiDAR sensor (LRF) to measure the actual distance to the pedestrian. Simultaneously, the drone recorded video footage, which was later used for pedestrian detection using YOLO.

The purpose of the experiments was to verify whether YOLO could detect the pedestrian's position and transmit the coordinates to ROS for visualization. Additionally, the distance estimated by YOLO was compared with the true distance measured by the LRF, and the error was calculated.

3.2.2 Experimental results

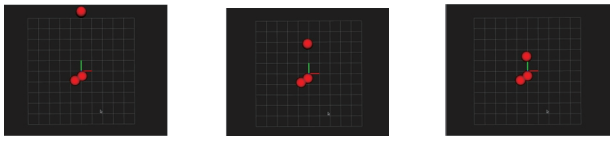
Figs. 4, 5 and 6 show a typical result of Experiment 1, including the real scene, YOLO detection, and Rviz visualization. Similar procedures were applied in the other two experiments, confirming that pedestrian positions could be detected and visualized in ROS.

Fig. 4: Preliminaly Experimental Setup.



(a) approaching (b) estrangement (c) stillness

Fig. 5: YOLO recognition resolts.



(a) approaching (b) estrangement (c) stillness

Fig. 6: Recognition result of YOLO which is imported into Rviz.

4. ROBOT NAVIGATION

4.1. Coordinate construction using AprilTag

In this work, a coordinate system for the robot is constructed using AprilTag, a vision-based binary marker developed by a research team at the University of Michigan, which is capable of highly accurate pose estimation and target tracking. It is effective for robot navigation and path planning because of its higher detection accuracy, faster processing speed, and suitability for real-time processing compared to QR codes. It is also less susceptible to changes in brightness and partial shielding, enabling stable detection even in outdoor environments.

Based on the above characteristics, this study uses AprilTag to construct a coordinate system for the robot. Within this framework, a camera coordinate system is constructed with the center of the drone's camera as the origin, and an AprilTag is attached on the robot to define the robot coordinate system. The coordinates of the pedestrian are extracted within the camera coordinate system and the corresponding coordinate transformation is performed. The relationship between the robot coordinate system and the camera coordinate system is obtained by Eq. (2) and (3), and R_{tag} and T_{tag} can be directly obtained from the AprilTag detector.

$$P_{\text{tag}} = R_{\text{tag}}^{-1} \cdot (P_{\text{cam}} - T_{\text{tag}}) \tag{2}$$

$$\begin{bmatrix} X_{\text{tag}} \\ Y_{\text{tag}} \\ Z_{\text{tag}} \end{bmatrix} = R_{\text{tag}}^{-1} \left(\begin{bmatrix} X_{\text{cam}} \\ Y_{\text{cam}} \\ Z_{\text{cam}} \end{bmatrix} - \begin{bmatrix} T_x \\ T_y \\ T_z \end{bmatrix} \right) \tag{3}$$

P_{tag}	Point in the AprilTag coordinate system
P_{cam}	Point in the camera coordinate system
T_{tag}	Translation vector of the AprilTag in the camera coordinate system
R_{tag}	Rotation matrix of the AprilTag in the camera coordinate system
R_{tag}^{-1}	Inverse of R_{tag} (orthogonal matrix, so $R_{\text{tag}}^{-1} = R_{\text{tag}}^{\text{T}}$)

4.2. Processing of pedestrian location data

The pedestrian's position relative to the robot is found via coordinate transformation [7]. While Eq. (4) presents the general 3D case, this system uses the 2D planar version in Eq. (5) for practical navigation.

$$\begin{bmatrix} X_{\text{human}} \\ Y_{\text{human}} \\ Z_{\text{human}} \end{bmatrix} = R_{\text{tag}}^{-1} \left(\begin{bmatrix} X_{\text{img}} \\ Y_{\text{img}} \\ Z_{\text{img}} \end{bmatrix} - \begin{bmatrix} X_{\text{tag}} \\ Y_{\text{tag}} \\ Z_{\text{tag}} \end{bmatrix} \right) \cdot \delta \tag{4}$$

$$\begin{bmatrix} X_{\text{human}} \\ Y_{\text{human}} \end{bmatrix} = R_{\text{tag}}^{-1} \left(\begin{bmatrix} X_{\text{img}} \\ Y_{\text{img}} \end{bmatrix} - \begin{bmatrix} X_{\text{tag}} \\ Y_{\text{tag}} \end{bmatrix} \right) \cdot \delta \tag{5}$$

P_{human}	Pedestrian position in the robot coordinate system
P_{img}	Detected pedestrian position in the camera coordinate system
P_{tag}	AprilTag position in the camera coordinate system
R_{tag}^{-1}	Inverse of the rotation matrix from robot to camera coordinate system
δ	Pixel scale: real-world distance per pixel

4.3. Robot Navigation via Potential Functions

The robot's path is planned using the potential method [1]. The potential method creates a potential function that attracts the robot to the target point and a virtual potential field that has an exclusion effect on obstacles such as pedestrians, and combines the two to determine the robot's direction of travel. This method is a suitable algorithm for robot guidance because it generates a path based on the surrounding information, which is computationally inexpensive, and the robot's target path can be changed in real time.The exclusion potential function due to obstacles such as pedestrians and the attraction potential function toward the target point are given by the following Eqs. (6) and (7). Adding them together, we can calculate the potential field at the robot's current position Eqs. (8) and (9). The potential field is considered as the coordinates (x,y) of the robot and (x_1,y_1), (x_2,y_2), ..., (x_m,y_m) of the walkers [8]. Thus, the exclusion potential function for m pedestrians is expressed as follows Eq. (7) The closer the distance to the target coordinates (x_d,y_d), the stronger the effect.

$$P_d(x, y) = -\frac{1}{\sqrt{(x - x_d)^2 + (y - y_d)^2}} \tag{6}$$

$$P_i(x, y) = \sum_{i=1}^{m} \frac{1}{\sqrt{(x - x_i)^2 + (y - y_i)^2}} \tag{7}$$

$$P(x, y) = w_d P_d + \sum_{i=1}^{m} w_i P_i \tag{8}$$

$$-\nabla P(x, y) = -\left[\frac{\partial P}{\partial x}, \frac{\partial P}{\partial y} \right] \tag{9}$$

5. EXPERIMENTS

5.1. Prototype mobile robot and drone specifications

Figs 7 and 8 show the drone and the Prototype Mobile Robot used in this study, respectively. The Prototype Mobile Robot has two driving wheels and one caster. The April Tag is installed to construct the coordinate system of the Prototype Mobile Robot. In addition, 2DLRF, which is used for comparison experiments, is installed. The drone is a DJI Air 2S, which has a camera with a horizontal field of view of 88 degrees and can capture images of the ground in an area of 19 x 11 meters at a height of 10 meters.

Fig. 7: Side view of the Air 2.

Fig. 8: Prototype Mobile Robot.

5.2. Experimental setup

This experiment aims to verify whether the Prototype Mobile Robot successfully detects all surrounding pedestrians and plans its route using a drone. The driving environment for this evaluation was the painted surface behind Building V of the University (Fig. 9). The experiment involved four participants, with the scene captured by a drone. It simulates a scenario where a pedestrian suddenly appears while the Prototype Mobile Robot is in motion, a situation where occlusion prevents detection by the LRF. However, our system does not experience occlusion in such cases, allowing us to verify that the robot

can safely avoid pedestrians. Four experimenters are positioned in two rows on the right side of the screen (Fig. 10(a)), with two in the front row and two in the back, all facing left. The prototype mobile robot moves from the left side to the right at a speed of 0.6 m/s, while the experimenters move in the opposite direction at 0.8 m/s. After approximately five seconds, the experimenter in the front row executes an evasive maneuver, shifting to the right in the direction of motion. Meanwhile, the experimenter in the back row, who is unaware of the Prototype Mobile Robot, moves slightly to the left to avoid the front-row experimenter before continuing forward (Fig. 10(b)).

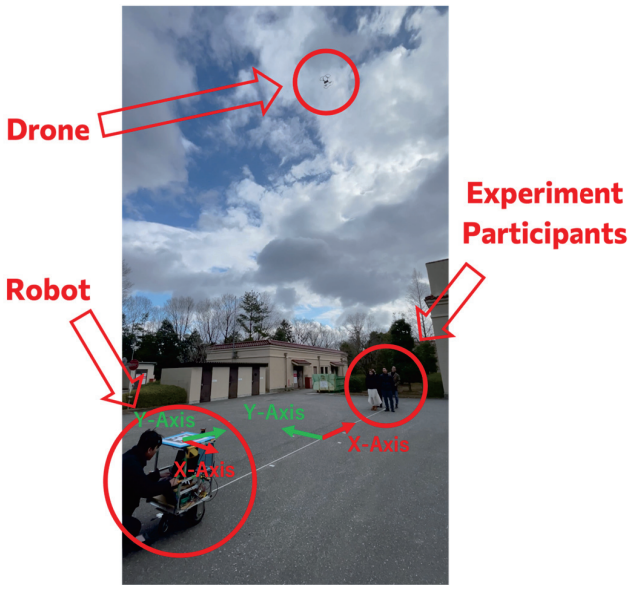
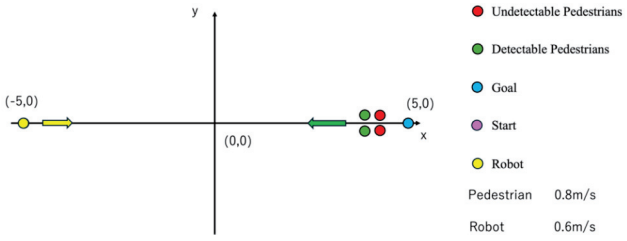
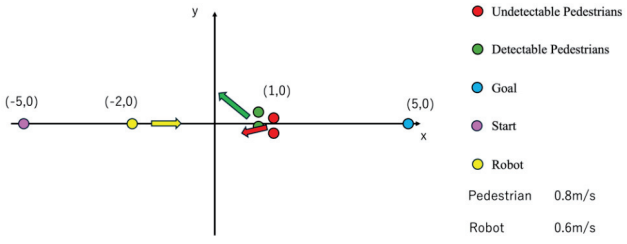


Fig. 9: Scene of the Experiment in Progress.



(a) Example Image 1 of Experiment 3.



(b) Example Image 2 of Experiment 3.

Fig. 10: Comparison of pedestrian detection (Fig.10(a) and Fig.10(b)).

5.3. Experimental results

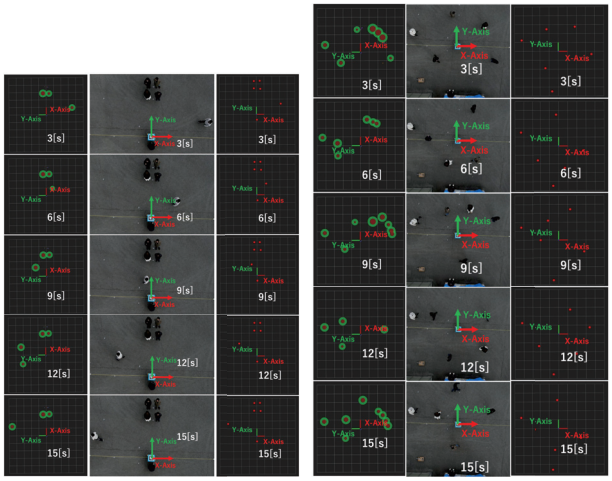
When LRF is used, the robot can only detect the pedestrian in front of it.The robot was unable to avoid a pedestrian who suddenly appeared behind it, and came to a sudden stop just before colliding with the pedestrian. However, our system detected four pedestrians from the beginning. Furthermore, the robot began to avoid the pedestrians when the pedestrians in front of it made their avoidance movements. Finally, the robot avoided all four pedestrians and arrived at the destination safely.

With the robot stationary, four experimenters stood in two rows, two in the front and two in the rear. LRF failed to detect the two experimenters in the back due to occlusion, whereas our system successfully detected all four from the beginning.

The results are shown in Fig 11(a). Next, while the robot remained stationary, four experimenters walked clockwise around it. LRF could not detect pedestrians behind the robot, but our system successfully identified them. The results are shown in Fig 11(b). In the case of using LRF, the robot could only detect pedestrians in front and failed to avoid those who suddenly appeared behind it, stopping abruptly to prevent a collision. In contrast, our system detected all four pedestrians from the start.

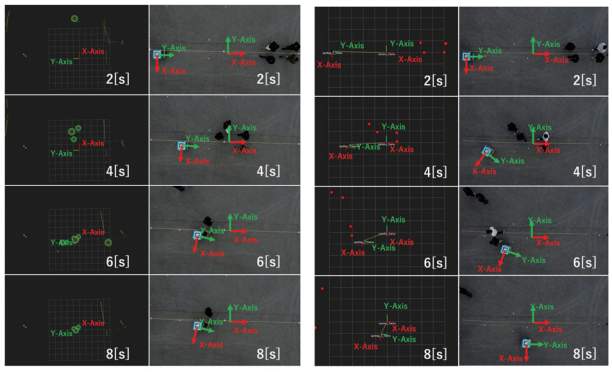
Additionally, the robot initiated avoidance maneuvers as soon as the pedestrians in front began their own movements. Ultimately, the robot successfully avoided all four pedestrians and reached its goal safely. The results are shown in Fig 12.

Fig. 13(a) shows the results of LRF in the previous study. Fig. 13(b) shows the results of the proposed system. Our system successfully detected all pedestrians, eliminating the occlusion seen with the LRF. Quantitatively, the robot initiated avoidance maneuvers 2.1 seconds earlier on average than the LRF-based method. Across all experimental runs, a 100% collision avoidance success rate was achieved, while maintaining a minimum safe distance of 1.2m from pedestrians. This verifies the effectiveness of the proposed method in generating a reliably safe route.



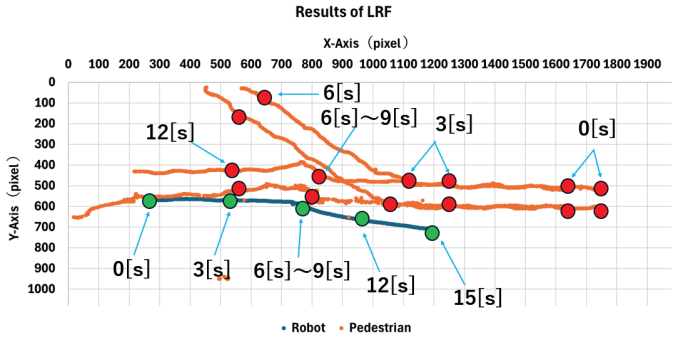
(a) LRF fails to react to pedestrians appearing behind the robot, whereas (b) our system detects all four pedestrians and enables safe navigation

Fig. 11: Comparison of pedestrian avoidance.

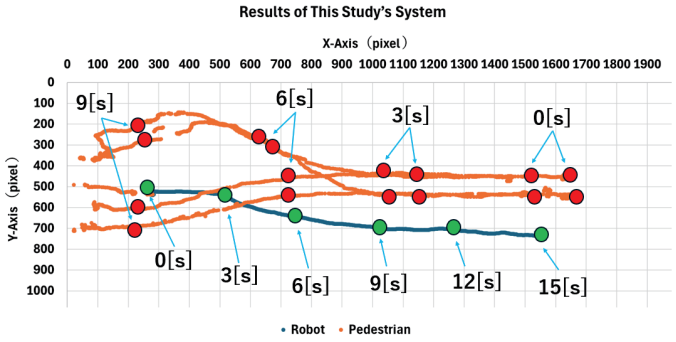


(a) LRF fails to react to pedestrians appearing behind the robot, whereas (b) our system detects all four pedestrians and enables safe navigation

Fig. 12: Comparison of pedestrian avoidance.



(a) Previous method



(b) Our proposed method

Fig. 13: Pedestrians Trajectory

6. CONCLUSION

This paper proposes a navigation method for autonomous mobile robots that takes into account pedestrians, who cannot be detected by conventional sensors due to occlusion, by detecting the autonomous mobile robot and surrounding people from a bird's-eye view using a drone. A camera mounted on a drone captures images of the ground, and YOLOv8 is used to detect pedestrians on the ground and the robot. The relative positions of the robot and the pedestrian are then calculated in the robot coordinate system constructed by April Tag.Using the calculated pedestrian coordinates, we propose a robot navigation method that safely avoids pedestrians invisible from the robot's viewpoint by using the artificial potential method. We implemented the proposed method on a robot and confirmed its effectiveness through experiments, which were built upon public frameworks like ROS and YOLOv8 to ensure reproducibility.

REFERENCES

[1] N. Nishino, R. Tsugita, D. Chugo, S. Yokota, S. Muramatsu and H. Hashimoto, "Robot navigation according to the characteristics of pedestrian flow", *IECON 2016-42nd Annual Conference of the IEEE Industrial Electronics Society*, Vol.6, 2017.8217496, pp.5947-5952, 2016.

[2] N. Nishino, R. Tsugita, D. Chugo, S. Yokota, S. Muramatsu and H. Hashimoto, "Robot navigation according to the characteristics of pedestrian flow", *IECON 2017 - 43rd Annual Conference of the IEEE Industrial Electronics Society, Beijing, China*, Vol.6, pp. 8521-8526, 2017.

[3] A. Jinguji, Y. Sada and H. Nakahara, "Real-Time Multi-Pedestrian Detection in Surveillance Camera using FPGA", *International Conference on Field Programmable Logic and Applications (FPL), Barcelona, Spain*, Vol.2, pp.424-425, 2019.

[4] J. Redmon, S. Divvala, R. Girshick and A. Farhadi,"You Only Look Once: Unified, Real-Time Object Detection", *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA*, Vol.10, pp.779-788, 2016.

[5] G. Wang, Y. Chen, P. An, H. Hong, J. Hu and T. Huang, "UAV-YOLOv8: ASmall-Object-Detection Model Based on Improved YOLOv8for UAVAerial Photography Scenarios", *IECON 2017 - 43rd Annual Conference of the IEEE Industrial Electronics Society, Beijing, China*, Vol.27, pp.1-27, 2023.

[6] S. Kim, W. Lee, J. Yim, Y. Park and Y. Lee, "Human monitoring system using drones for riverside area", *2017 International Conference on Information and Communication Technology Convergence (ICTC), Jeju, Korea (South)*, Vol.4, pp.1054-1057, 2017.

[7] R. Tsugita, N. Nishino, D. Chugo, S. Muramatsu, S. Yokota and H. hashimoto, "Pedestrian detection and tracking of a mobile robot with multiple 2D laser range scanners", *Proceedings of 9th International Conference on Human System Interaction*, Vol.6, pp.412-417, 2016.

[8] A. Toyoshima, N. Nishino, D. Chugo, S. Muramatsu, S. Yokota and H. Hashimoto, "Autonomous Mobile Robot Navigation: Consideration of the Pedestrian's Dynamic Personal Space", *Proceedings of the 27th International Symposium on Industrial Electronics*, Vol.6, pp.1094-1099, 2018.