2025 IEEE/SICE International Symposium on System Integration (SII)
January 21-24, 2025, Munich, Germany

Path Planning Method for Mobile Robots in Pedestrian-Friendly Environments Using Drone Bird's-Eye View

Linghao Ye, Daisuke Chugo, Satoshi Muramatsu, Sho Yokota, Hiroshi Hashimoto

Abstract— **The objective of this paper is to develop a robot system that can transmit the video taken by the drone to the computer in real time, use YOLO to recognize the image of pedestrians, and enable the ground robot to plan the path to avoid pedestrians. External sensors, such as laser range finders (LRF) and CCD cameras, are often used on mobile robots. As we all know, these sensors are susceptible to occlusion. Therefore, even if the mobile robot avoids the pedestrian in front of it, it is likely to collide with other pedestrians that cannot be detected due to occlusion. In this study, we used a drone to shoot videos at high altitudes, transferred the videos to a computer, used YOLO on the computer to perform image recognition of pedestrians around the robot, and passed the recognition results to ROS, which enabled it to plan a path to avoid pedestrians around it. As a result, showed that the videos shot at high altitudes could effectively identify pedestrians around the robot after being recognized by YOLO.**

I. Introduction

In recent years, the demand for service robots has been increasing. These service robots must safely avoid surrounding pedestrians while moving toward their goal in crowded and complex environments such as sidewalks, parks, and residential areas.

In previous studies, the artificial potential method has been used as a method for mobile robots to avoid pedestrians [1]. This method requires little computational effort and can realize real-time path planning for mobile robots. In addition, a potential method has been proposed that generates a path in a crowded environment to synchronize with surrounding pedestrians and follow the overall pedestrian flow [2].

However, these previous studies assume that the robot can detect all pedestrians in the environment. The study in [2] confirmed its effectiveness in simulation experiments, but in real environments, external sensors are often mounted on robots, and their installation positions are restricted by the robot's body size. In many cases, external sensors mounted on the robot are used, such as laser range finders (LRF) and CCD cameras, but these sensors are vulnerable to occlusion. Therefore, even if a mobile robot avoids a pedestrian in front of it, there is a high possibility that it will collide with other pedestrians that it could not detect due to occlusion.

On the other hand, methods have been proposed to recognize pedestrians by installing cameras on the ceiling indoors or at high places outdoors [3]. However, these methods cannot detect pedestrians outside the camera range. Therefore, for a mobile robot to travel safely in a city, it is necessary to install a large number of cameras in the environment, which is not realistic in terms of cost.

In this study, we use 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] and combine it with the above-mentioned potential method. Finally, we propose a path planning method that allows a mobile robot to travel to its destination in a complex environment with many people without disrupting the flow of people.

II. Robot system configuration

A. How the mobile robot system works

In this study, 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 drone used is a DJI Air 2S. The Air 2S has a camera with a horizontal field of view of 88 degrees and can capture the ground in a range of 29m x 16m at a height of 15m. The Air 2S was selected because it has sufficient performance for this study. The image captured by the drone is sent to a laptop computer via HDMI wirelessly via a DJI controller RC PRO. Both YOLO and ROS are running on the laptop computer. YOLO detects pedestrians and mobile robots in the video in real time. Furthermore, the detection results are processed, and a route is generated for the mobile robot by ROS.

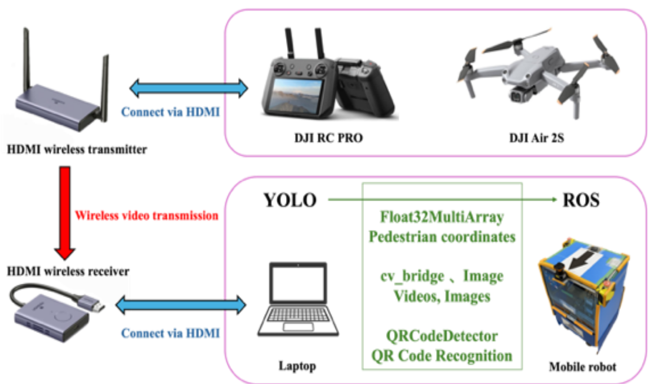


Figure 1. Robot system configuration

Linghao Ye and Daisuke Chugo are with Graduate School of Science and Technology, Kwansei Gakuin University, Sanda, Hyogo 6691330 Japan (corresponding author to provide phone: +81-79-565-7043; fax: +81-79-565-7043; e-mail: {yelinghao, chugo}@chugolab.com).

Satoshi Muramatsu is with School of Information Science and Technology, Tokai University, Hiratsuka, Kanagawa 2591292 Japan. (e-mail: muramatsu@tokai.ac.jp).

Sho Yokota is with Faculty of Science and Engineering, Toyo University, Kawagoe, Saitama 3508585 Japan. (e-mail: s-yokota@toyo.jp).

Hiroshi Hashimoto is with Master Program of Innovation for Design and Engineering, Advanced Institute of Industrial Technology, Shinagawa, Tokyo 1400011 Japan. (e-mail: hashimoto@aiit.ac.jp).

The YOLO used in this study is run by Python. A dedicated ROS node is created to send the coordinates of pedestrians to ROS. The detection results are sent to a ROS node using Python, and the coordinates stored by the mobile robot in the ROS node are read to plan the robot's path.

B. Image Recognition Methods

In this study, YOLO is used to recognize and track pedestrians and moving robots. In this study, we adopt YOLOv8, 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. In this study, we use the YOLOv8s model in consideration of the GPU load to install YOLOv8 on a laptop. In this study, to recognize a moving robot from a drone image, we use 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. 3 shows the results of pedestrian recognition using the trained YOLO model.

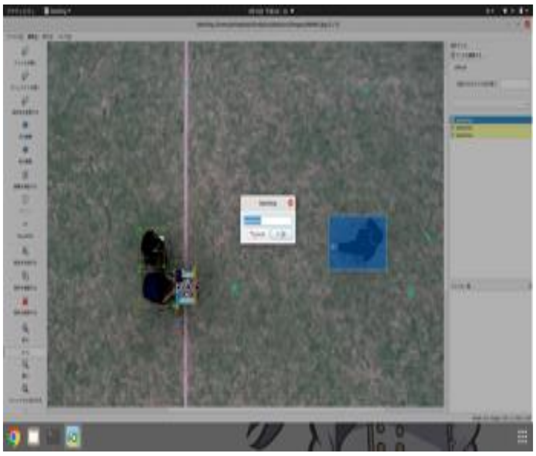


Figure 2. Mark pedestrians with Labelimg

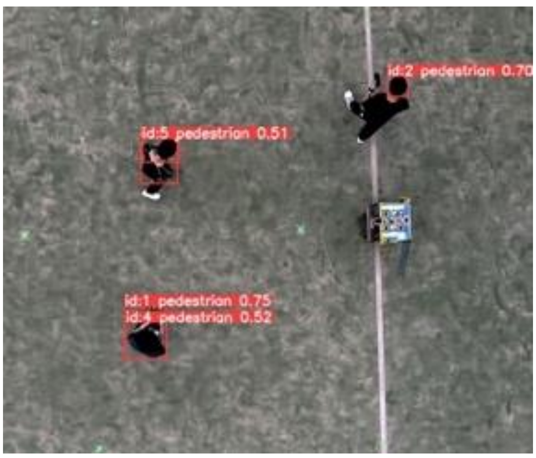


Figure 3. Pedestrian recognition results

III. Processing 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 (1) [6].

Fig. 4 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, θx 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_n are the number of pixels for width and height of the target object on the capture image.

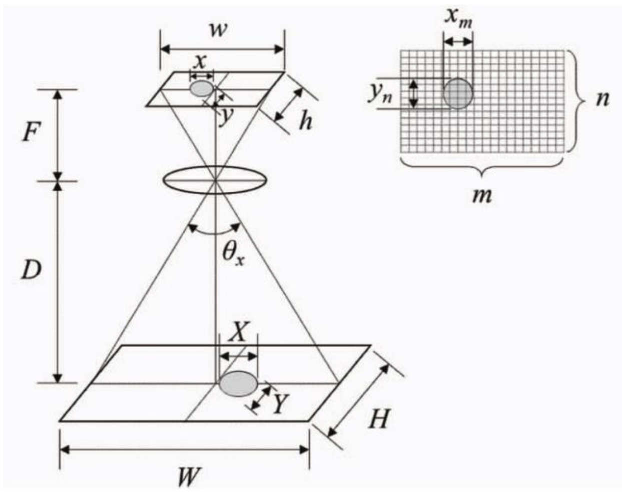


Figure 4. Optical camera mounted at a drone [6]

As shown in Fig. 4, 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:

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

where $f\{\}$ is the function of floor operation. The height size of the target object in pixels can be represented by the same way of the width size representation.

W width of monitoring range
H height of 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
θx 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 the number of pixels for width of the captured image
n the number of pixels for height of the captured image
x_m the number of pixels for width of the target object on the capture image
y_n the number of pixels for height of the target object on the capture image

IV. Preliminary Experiment

To verify the detection accuracy of the proposed system, three preliminary experiments were conducted on the grounds of the Sanda Campus of Kwansei Gakuin University, as shown in Fig. 5. The front of the robot is defined as the positive direction of the Y axis.

A. Experimental procedure

In experiment 1, the robot approached from 5m in front of the robot to 1m (Fig. 6). In experiment 2, the robot moved away from 1m to 5m in front of the robot (Fig. 7). In experiment 3, the robot moved to a position 4m in front of the robot (Fig. 8). Each experiment was conducted three times.

The drone was automatically maintained at a height of 15m directly above the robot. The accuracy of the altimeter built into the drone is 0.1m. The robot is equipped with a 2D radar sensor (LRF) to detect pedestrians and measure the true distance between the pedestrian and the robot. While measuring the true distance, the drone also took a video, which was later used to calculate YOLO's measurement values. In all three experiments, the robot remained stationary while the pedestrian in front was in motion. The robot is located at the position of the QR code in the picture.

The purpose of this experiment was to test whether the position of pedestrians can be detected by YOLO and the position coordinates of pedestrians can be transmitted to ROS. And calculate the error between the calculated distance detected by YOLO and the actual distance detected by the 2D radar sensor (LRF) during the three experiments.

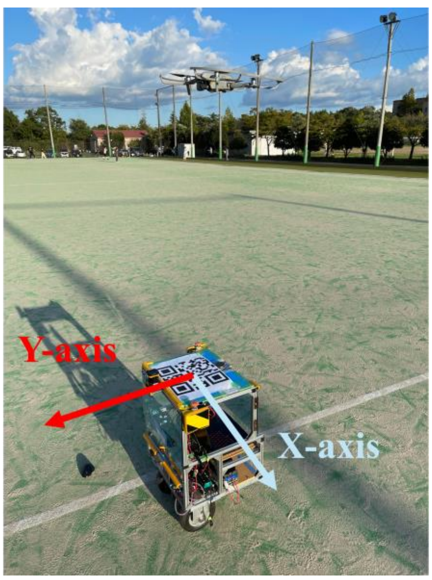


Figure 5. Experimental scenario

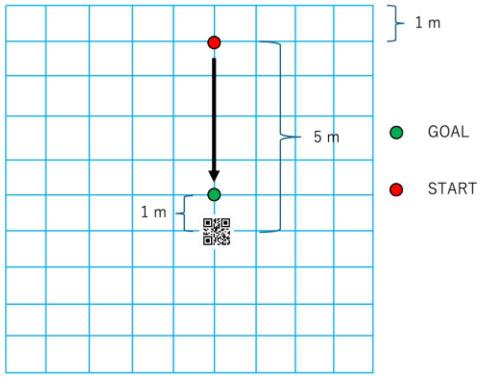


Figure 6. Experiment 1

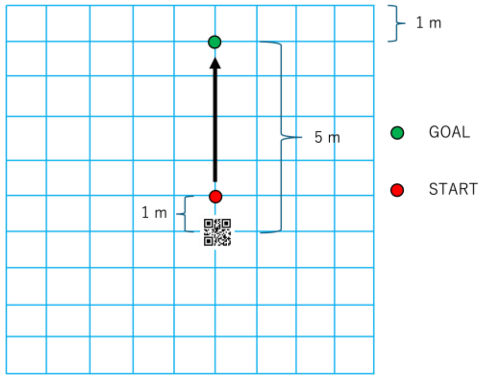


Figure 7. Experiment 2

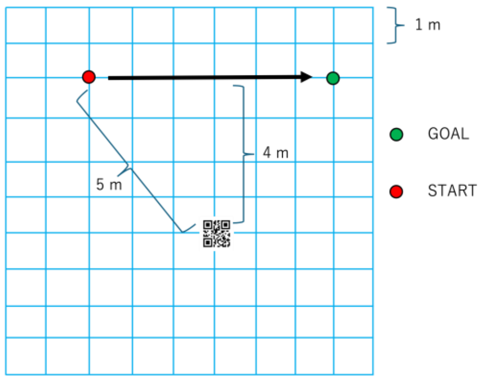


Figure 8. Experiment 3

B. YOLO recognizes and passes it to ROS

The following pictures show the situation during the experiment. Fig. 9 shows the situation during Experiment 1, Fig. 10 shows the result of Experiment 1 after YOLO recognition, and Fig. 11 shows the scene where the recognition result of YOLO is imported into Rviz.

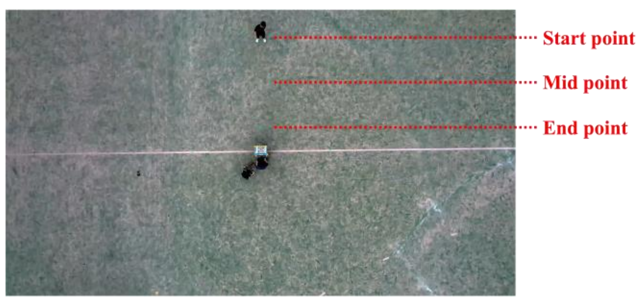
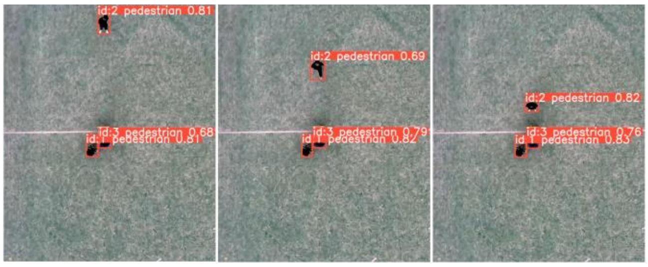
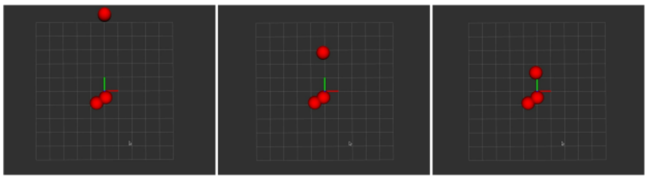


Figure 9. Situation during Experiment 1



(a) Start (b) Mid (c) End

Figure 10. Result of Experiment 1 after YOLO recognition



(a) Start (b) Mid (c) End

Figure 11. Experiment 1 recognition result of YOLO is imported into Rviz

Fig. 12 shows the situation in progress of Experiment 2, Fig. 13 shows the result of Experiment 2 after YOLO recognition, and Fig. 14 shows the scene after YOLO recognition result is imported into Rviz.

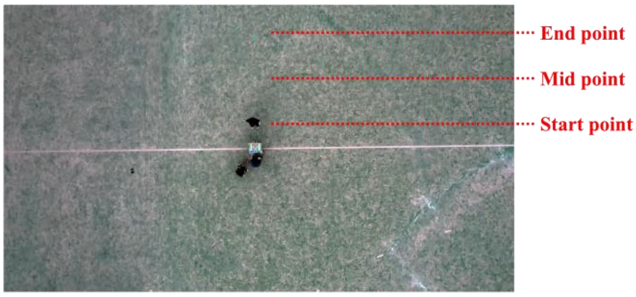
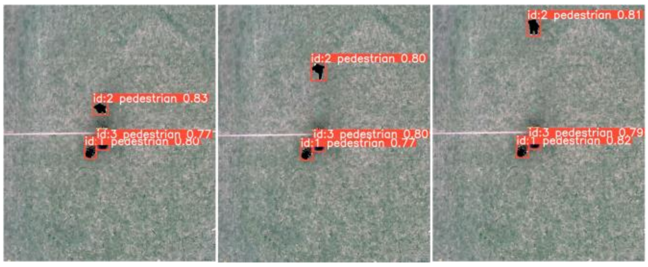
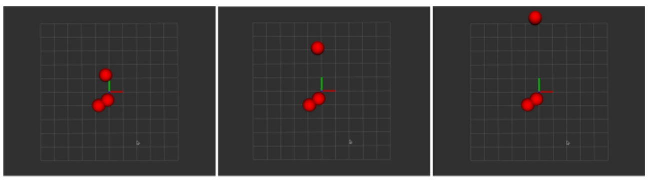


Figure 12. Situation during Experiment 2



(a) Start (b) Mid (c) End

Figure 13. Result of Experiment 2 after YOLO recognition



(a) Start (b) Mid (c) End

Figure 14. Experiment 2 recognition result of YOLO is imported into Rviz

Fig. 15 shows the situation in progress of Experiment 3, Fig. 16 shows the result of Experiment 3 after YOLO recognition, and Fig. 17 shows the scene after YOLO recognition result is imported into Rviz.

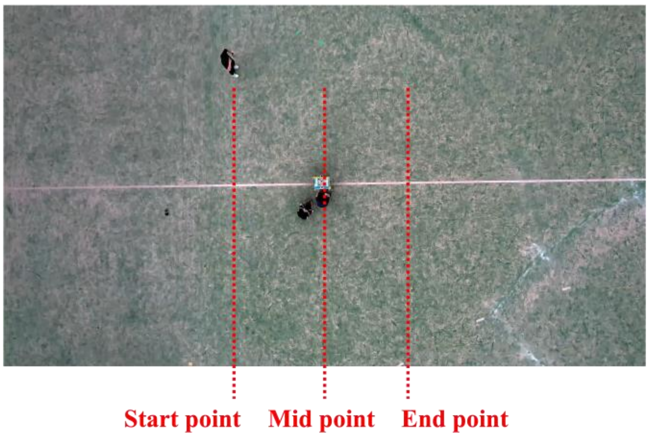
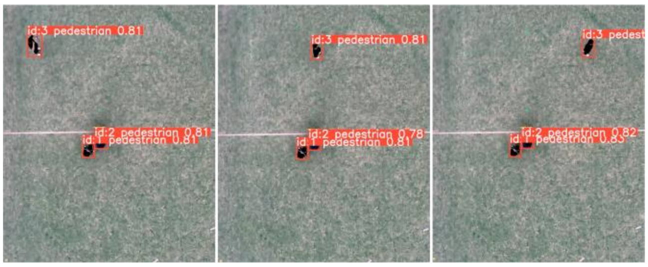
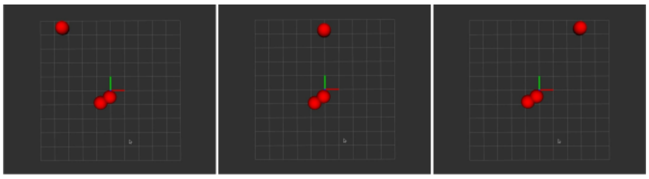


Figure 15. Situation during Experiment 3



(a) Start (b) Mid (c) End

Figure 16. Result of Experiment 3 after YOLO recognition

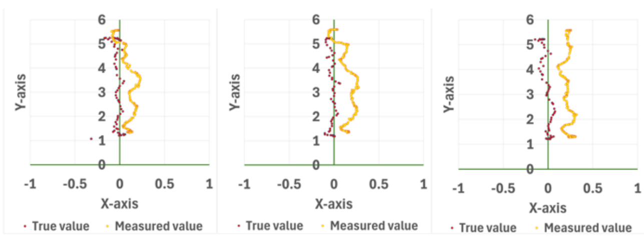


(a) Start (b) Mid (c) End

Figure 17. Experiment 3 recognition result of YOLO is imported into Rviz

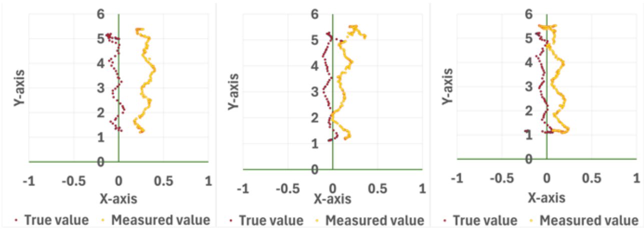
V. EXPERIMENTS DISCUSSION

The results of the three experiments are shown below. The results of Experiment 1 when the pedestrian moves from 5m away from the robot to 1m in front of the robot are shown in Fig. 18; the results of Experiment 2 when the pedestrian moves from 1m away from the robot to 5m in front of the robot are shown in Fig. 19; the results of Experiment 3 when the pedestrian moves to 4m in front of the robot are shown in Fig. 20. The error results of the three experiments are shown in Fig. 21.



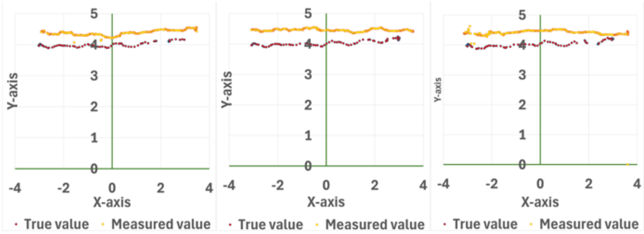
(a) 5m to 1m ver.1 (b) 5m to 1m ver.2 (c) 5m to 1m ver.3

Figure 18. Results of Experiment 1



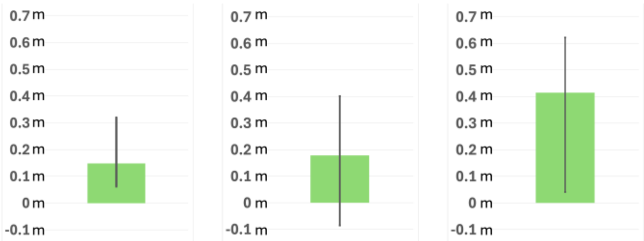
(a) 1m to 5m ver.1 (b) 1m to 5m ver.2 (c) 1m to 5m ver.3

Figure 19. Results of Experiment 2



(a) Passes 4m in front ver.1 (b) Passes 4m in front ver.2 (c) Passes 4m in front ver.3

Figure 20. Results of Experiment 3



(a) 5m to 1m (b) 1m to 5m (c) Passes 4m in front

Figure 21. Error in recognition results

Fig. 18, Fig. 19 and Fig. 20 show the walking trajectories of pedestrians captured during the experiment. The red trajectory is the accurate value detected by the 2D radar sensor. The yellow trajectory is the value calculated by the YOLO detection. According to the results of the preliminary experiment, it can be seen that there is a certain error between the value calculated by the YOLO detection and the accurate value detected by the 2D radar sensor. Fig. 21(a) shows that the maximum error of Experiment 1 is about 0.3m. Fig. 21(b) shows that the maximum error of Experiment 2 is about 0.4m. The difference between the two is not much. The maximum error occurs at the place where Fig. 21(c) passes 4m ago, which is about 0.6m.

According to the results, the farther the distance, the larger the error tends to be. This is most likely due to distortion at the edge of the camera. From now on, we will consider how to solve this problem.

VI. Conclusion

In this preliminary experiment, we successfully used YOLO to identify the video from the drone and sent the identification results to ROS. Then in ROS, we extracted the coordinates from the identification results and calculated the relative position of the pedestrian and the robot through coordinate transformation. We also displayed the relative position of the robot and the pedestrian through Rviz. Next, we will continue to import the obstacle avoidance algorithm on the robot so that the robot can avoid the surrounding pedestrians in real time. It is possible to achieve an unobstructed robot obstacle avoidance system through drones.

References

[1] Nozomi Nishino, Ryo Tsugita, Daisuke Chugo, Sho Yokota, Satoshi Muramatsu, Hiroshi Hashimoto, "Robot navigation according to the characteristics of pedestrian flow" IECON 2016-42nd Annual Conference of the IEEE Industrial Electronics Society, 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, 2017, pp. 8521-8526, doi: 10.1109/IECON.2017.8217496.

[3] A. Jinguji, Y. Sada and H. Nakahara, "Real-Time Multi-Pedestrian Detection in Surveillance Camera using FPGA," 2019 29th International Conference on Field Programmable Logic and Applications (FPL), Barcelona, Spain, 2019, pp. 424-425, doi:10.1109/FPL.2019.00078.

[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, 2016, pp. 779-788, doi: 10.1109/CVPR.2016.91.

[5] Wang, G.; Chen, Y.; An, P.; Hong, H.; Hu, J.; Huang, T. UAV-YOLOv8: A Small-Object-Detection Model Based on Improved YOLOv8 for UAV Aerial Photography Scenarios. *Sensors* 2023, *23*, 7190.

[6] S. Kim, W. Lee, J. Yim, Y. -S. Park and Y. -T. Lee, "Human monitoring system using drones for riverside area," 2017 International Conference on Information and Communication Technology Convergence (ICTC), Jeju, Korea (South), 2017, pp. 1054-1057, doi:10.1109/ICTC.2017.8190852.