Person Identification Method Based on Face Recognition and Multiple Object Tracking for Person Following Robot

Taiga Kuroiwa[1,a,*], Xiongwen Jiang[1,b], Takahiro Kawaguchi[1,c], Haohao Zhang[1,d], and Seiji Hashimoto[1,e]

[1]Gunma University, 1-5-1 Tenjin-chou, Kiryu-shi, Gunma 376-0052, Japan

* Corresponding author

[a]<t231d026@gunma-u.ac.jp>, [b]<t212d601@gunma-u.ac.jp>, [c]<kawaguchi@gunma-u.ac.jp>, [d]<h-zhang@gunma-u.ac.jp>, [e]<hashimotos@gunma-u.ac.jp>

Keywords: person following robot, face recognition, multiple object tracking

Abstract. This paper proposes a method for a mobile robot to track and follow a specific person. Traditionally, detecting and identifying the target person necessitates the use of various sensors, including expensive options like LiDAR and stereo cameras. However, proposed method utilizes images from an inexpensive monocular camera for this purpose. By employing multiple object tracking and face recognition, the system can detect and identify the target person effectively. Even when the person's face is not visible, the system can still detect and identify the target by combining these two methods. To compare the performance of the conventional and proposed methods, both were applied to a video containing four scenarios. The conventional method correctly identified the target person in only one scenario, while the proposed method was able to identify the target person in all four scenarios. Furthermore, the effectiveness of the proposed person following method, based on face recognition and multiple object tracking, is evaluated through person following experiments. The proposed method successfully detected and identifies the target to be followed for a longer time under various conditions compared to the conventional method.

1. Introduction

Following a person is among the most crucial capabilities for mobile robots. To achieve effective person following, the initial step involves detecting the target person. Moreover, in scenarios where multiple persons are detected, the system must be able to accurately identify the target individual among them. The method proposed in [1] utilizes two types of sensors: a laser rangefinder and a sonar sensor, for person detection. Meanwhile, in [2-5], the authors propose a method employing a stereo camera for person following. The authors in [6] utilized a laser rangefinder to follow a person. Further advancements include methods for identification based on specific features of the target person [7].

Since monocular cameras are less expensive than LiDAR and stereo cameras, and since it is difficult to identify a person with LiDAR, an algorithm [8] was proposed to detect and identify individuals and track a specified target person, focusing on monocular cameras. However, the approach outlined in [8] encountered challenges in maintaining target identification over extended durations. One contributing factor was the absence of certain identifying features, such as clothing or facial characteristics, within the system's methodology. To address this limitation, face recognition is introduced into the conventional method to enhance target identification performance. Furthermore, we provide an illustrative example of person following using the proposed methodology.

2. Person Following Method

The process flow for person following is illustrated in Fig. 1. Initially, the image captured by a monocular camera undergoes multiple object tracking (MOT) and face recognition. Subsequently, the

target person is identified based on the outcomes of these processes. Following that, a projective transformation is executed to compute the distance and angle between the tracked target and a robot. Eventually, the robot is directed accordingly. This sequence of steps enables the robot to accomplish person following. The method proposed in [8] is defined as the conventional method. Noteworthy enhancements to the conventional method entail the integration of face recognition and the utilization of MOT and face recognition outcomes for target person identification, which constitute the central focus of this paper.

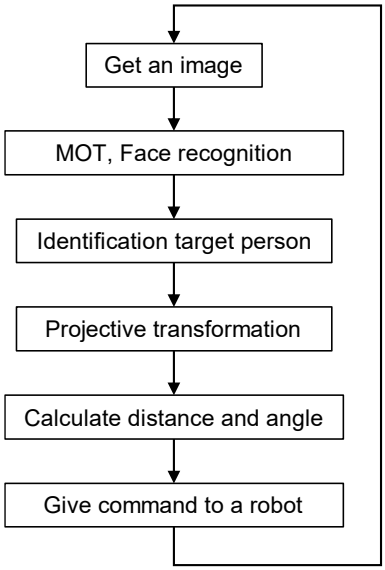


Fig. 1. Process flow for person following.

3.1 Multiple Object Tracking

Multiple object tracking is a technique used to detect the location and identity of multiple specific objects within a sequence of images or videos. In this study, our focus is solely on detecting individuals, specifically people, and their positions within the imagery are represented by rectangular regions known as bounding boxes. These bounding boxes are defined by the coordinates of two points, denoted as $\boldsymbol{x^{min}}$ and $\boldsymbol{x^{max}}$, as illustrated in Fig. 2. The results of Multiple Object Tracking (MOT) can be expressed mathematically as:

$$\boldsymbol{x}_i^{min} = \left(x_i^{min}, y_i^{min}\right), i \in \{1, 2, \dots, n\} \quad (1)$$

$$\boldsymbol{x}_i^{max} = \left(x_i^{max}, y_i^{max}\right), i \in \{1, 2, \dots, n\} \quad (2)$$

where n represents the number of individuals detected. In MOT, a unique ID is assigned to each detected object.

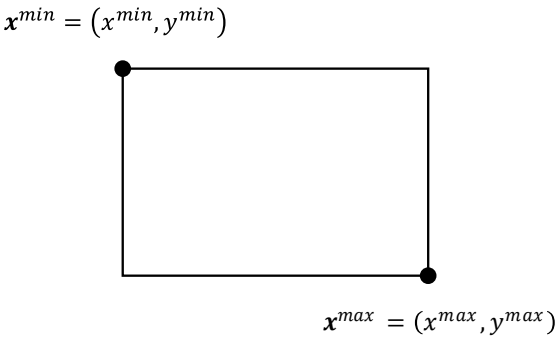


Fig. 2. Bounding box.

3.2 Face Recognition

Face recognition is a technology used to detect a face in an image or video and match it with a known face. The position of the target person's face can be represented using a bounding box, similar to MOT, where it can be denoted as

$$\boldsymbol{x}_t^{min} = \left(x_t^{min}, y_t^{min}\right) \tag{3}$$

$$\boldsymbol{x}_t^{max} = \left(x_t^{max}, y_t^{max}\right) \tag{4}$$

3.3 Person Identification Method

For the results of MOT and face recognition, $\boldsymbol{x}_i^{min}$ and $\boldsymbol{x}_i^{max}$ indicate the bounding box of the target person if there is an object that satisfies all the following conditions:

$$x_{\mathrm{t}}^{min} \geq x_i^{min} \tag{5}$$

$$y_{\mathrm{t}}^{min} \geq y_i^{min} \tag{6}$$

$$x_t^{max} \leq x_i^{max} \tag{7}$$

$$y_t^{max} \leq y_i^{max} \tag{8}$$

If face recognition is not possible due to factors such as the face being obscured or for other reasons, the system resorts to using the ID assigned to each object in MOT for person identification. Specifically, the system stores the ID of the target person, and when face recognition is not feasible, the person is identified by referencing the stored ID of the target.

3. Experiments and Evaluation

The model and hardware to be used for experiments are described as follows. YOLOv8 [9] was employed as the model for MOT due to its rapid processing capabilities, while face recognition [10] served as the model for facial recognition tasks. YOLOv8 [9] is the most recent model in the YOLO series, compatible with MOT and capable of real-time inference. YOLOv8 has several pre-trained models using the common objects in context (COCO) dataset [11], and we used YOLOv8n as it is the fastest performing among the others. The face recognition model utilizes deep learning techniques to extract information from 68 facial landmarks detected on the face. This information is then compared with previously acquired data regarding the target's facial landmarks to facilitate accurate identification of the face.

Fig. 3 displays the appearance of the robot utilized in the experiment, while Table 1 provides details regarding the hardware installed within the robot. Two PCs handle computation tasks, and a Jetson Xavier, equipped with a GPU, is utilized. The camera employed is an Intel RealSense D435i, possessing both depth and RGB capabilities; however, only RGB images are utilized in this study. ROS [12] serves as the middleware. In the current configuration, the proposed method can execute five operations per second.

Table 1. Details regarding hardware installed in robot.

Hardware	Details
PC1	Intel NUC
PC2	NVIDIA Jetson Xavier
Camera	Intel Realsense D435i

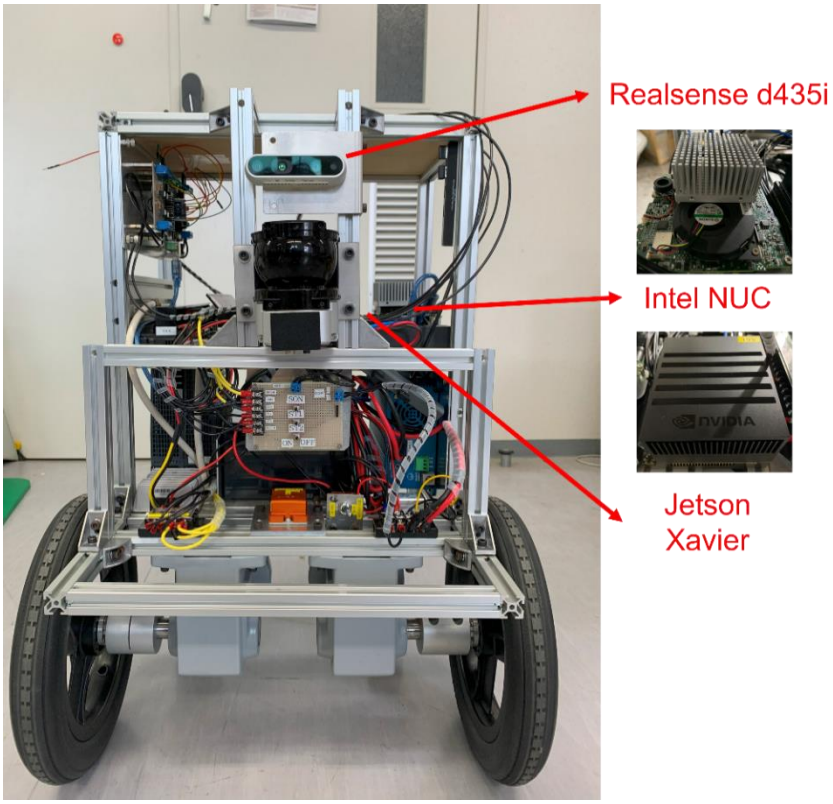


Fig. 3. Robot utilized in experiment.

3.1 Target Person Identification Performance

The target person identification performance of the proposed method was compared with that of the conventional method. Both methods were applied to a video containing two individuals, one of whom was the target person. The video presented four main scenarios: firstly, when the target's entire body and face were visible; secondly, when only the target's face was visible without his entire body; thirdly, when the target's face and body were visible again; and fourthly, when the target's face were visible, but not his entire body. The latter situation pertains to cases where the target's face is shown without the entire body. For these videos, only the person identification capabilities of the conventional and proposed methods were evaluated, with the results presented in Table 2.

In Case 1, both methods successfully identified the target person as the video included both the face and the entire body. In Case 2, the proposed method failed to identify the target due to the absence of the target's entire body in the video. However, the proposed method managed to identify the target through face recognition and utilizing information from one frame earlier. In Case 3, where both the face and entire body were visible again, the conventional method was unable to identify the target person once lost. In Case 4, although the entire body was visible, the face was not. Therefore, face recognition could not be utilized, but the object could still be identified through MOT.

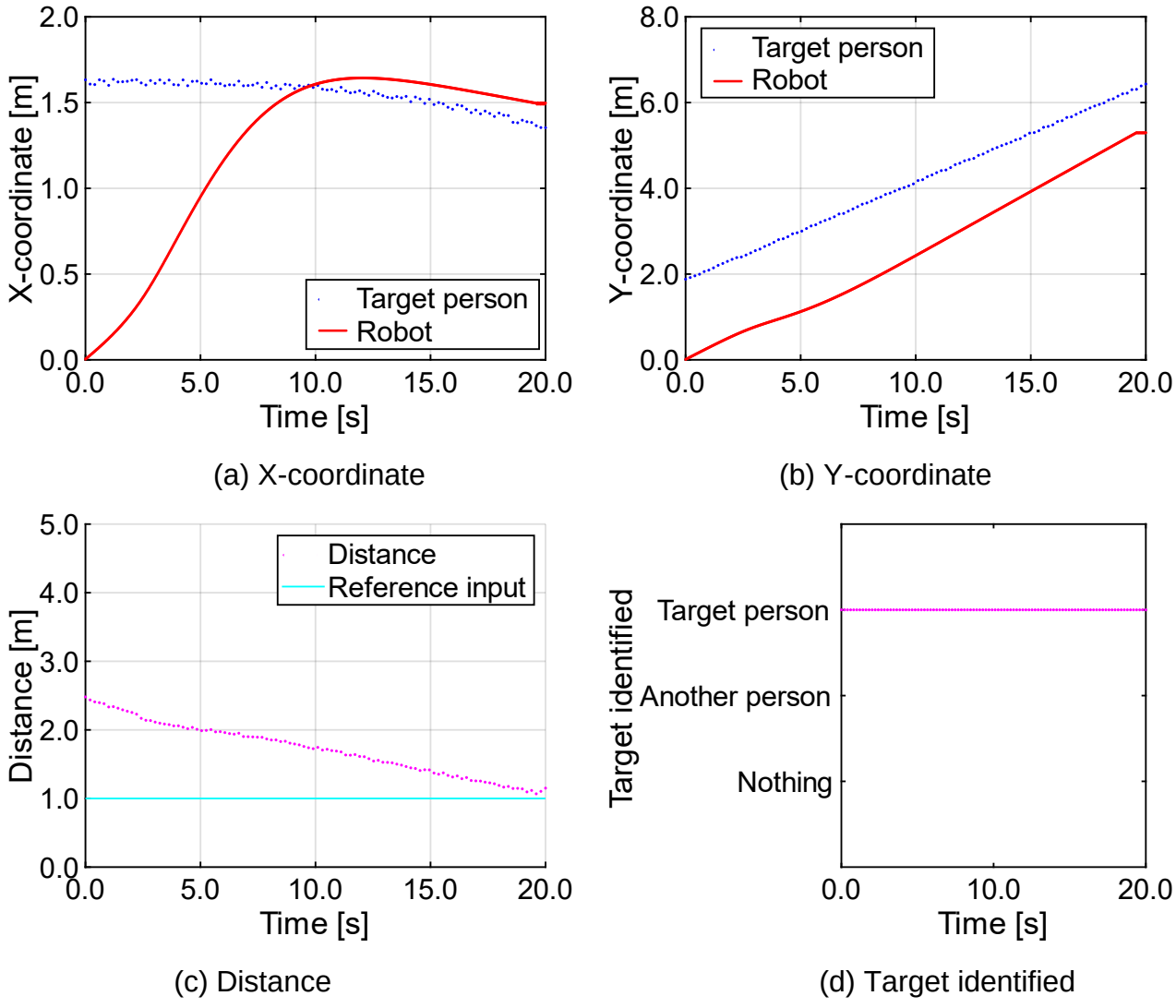
Table 2. Results of target identification.

Case	Face	Whole body	Conventional method	Proposed method
1	○	○	Identified	Identified
2	○	×	Not identified	Identified
3	○	○	Not identified	Identified
4	×	○	Not identified	Identified

(○: Captured, ×: Not captured)

3.2 Experimental Results

An experimental demonstration of person tracking using the proposed method is presented. The robot successfully tracked the target in a scenario involving the detection of two persons. The results are depicted in Figure 4. Panels (a) and (b) display the time response of the x- and y-coordinates of both the robot and the subject. Panel (c) illustrates the time response of the distance between the target person and the robot, while panel (d) presents the temporal variation of the identification status of the target being followed. The robot's coordinate data were captured using visual odometry, solely for recording purposes and not for person tracking. Thanks to visual odometry, the robot's coordinates exhibit smoothness. The target person was detected and identified through the proposed method, with its position inferred from the distance and angle relative to the robot, as well as the visual odometry results. When comparing the coordinates of the robot with those of the target, the target's coordinates appear less smooth. This issue is inherent, as the position of the target is calculated based on the MOT results (i.e., the bounding box). Even when a stationary person is detected through MOT, slight fluctuations occur in the position of the target person due to the minor movements of the bounding box with each detection. The trajectories of movement for the two persons and the robot are depicted in Fig. 5. In this scenario, one person is in motion while the other remains stationary. The distance between the target person and the robot was maintained at 1 meter, leading the robot to halt at this distance from the target.



(a) X-coordinate (b) Y-coordinate

(c) Distance (d) Target identified

Fig. 4. Time responses of experiments.

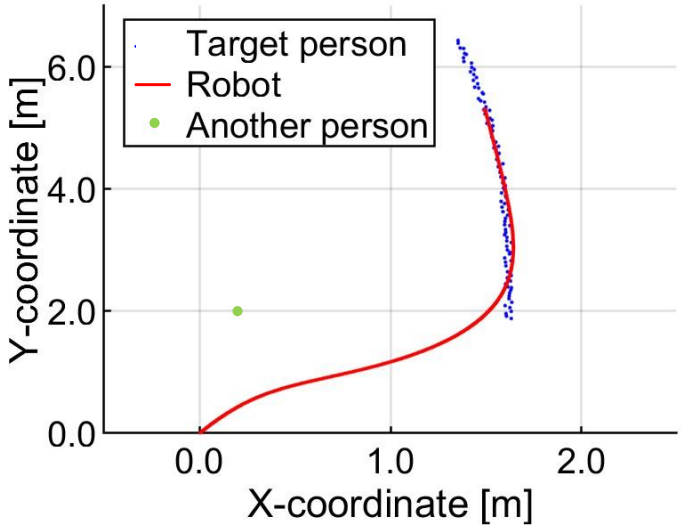


Fig. 5. Movement trajectories of each person and robot.

4. Conclusion

This paper has proposed a method for a person following robot to identify and track individuals using MOT and face recognition. The focus of the paper is on enhancements compared to conventional methods. Performance evaluation of the proposed method has been conducted against the conventional approach. It has been demonstrated that the proposed method exhibits prolonged person identification capabilities and the ability to re-identify individuals once lost. Furthermore, an illustrative example of person following utilizing the proposed method has been presented, where two individuals were successfully identified and followed by the robot. Looking ahead, future research endeavors aim at improving response speed and implementing modern control techniques. In addition, monocular cameras only detected people in a limited area in front of them. Therefore, an omnidirectional camera will be used to expand the detectable range. In this case, it will be necessary to convert the images obtained by the omnidirectional camera into a form suitable for object detection algorithms such as YOLO. Furthermore, the system currently only follows people and is planned to be integrated with obstacle avoidance.

References

[1] S. Jorge, M. Raul, C. Enric, R. Sergio and P. Javier, “Multi-Sensor Person Following in Low-Visibility Scenarios”, ***Sensors***, Vol. 10, No. 12, pp. 10953-10966, 2010, https://doi.org/10.3390/s101210953.

[2] J. Satake and J. Miura, “Person Following of a Mobile Robot using Stereo Vision”, ***Journal of the Robotics Society of Japan***, Vol. 28, No. 9, pp. 1091-1099, 2010, https://doi.org/10.7210/jrsj.28.1091.

[3] J. Satake and J. Miura, “Multi-Person Tracking for a Mobile Robot using Overlapping Silhouette Templates”, ***Journal of Information Processing Society of Japan***, Vol. 2010, No. 3, pp. 1-8, 2010.

[4] L. Jun, L. Ye, Z. Guyue, Z. Peiru and C. Qiu, “Detecting and Tracking People in Real Time with RGB-D Camera”, ***Pattern Recognition Letters***, Vol. 53, pp. 16-23, 2015, https://doi.org/10.1016/j.patrec.2014.09.013.

[5] A. Eirale, M. Martini, M. Chiaberge, "Human-Centered Navigation and Person-Following with Omnidirectional Robot for Indoor Assistance and Monitoring", ***Robotics***, Vol. 11, No. 108, pp. 1-17, 2022, https://doi.org/10.3390/robotics11050108.

[6] S. Okusako, S. Sakane, "Human Tracking with a Mobile Robot using a Laser Range-Finder", ***Journal of the Robotics Society of Japan***, Vol. 24, No. 5, pp. 605-613, 2006, https://doi.org/10.7210/jrsj.24.605.

[7] M. Chiba, J. Satake and J. Miura, "Improvement of a SIFT-based people identification for a people following robot use of a distance-dependent appearance model", ***Proc. of the Japan Joint Automatic Control Conference***, Vol. 54, pp. 76-81, 2011, https://doi.org/10.11511/jacc.54.0.27.0.

[8] T. Kuroiwa, N. Kan, T. Kawaguchi, S. Hashimoto, T. Isogai, H. Soga, H. Zhang, "Object Tracking Algorithm for Person-Following Autonomous Robots", ***Proc. of The 7th International Conference on Technology and Social Science***, IPS04-06, pp. 1-8, 2023.

[9] Ultralytics, "Ultralytics YOLOv8 Documents", https://docs.ultralytics.com/.

[10] A. Geitgey, "face_recognition", https://github.com/ageitgey/face_recognition.

[11] T. Lin, M. maire, S. Belongie, L. Bourdev, R. Girshick, J. Hays, P. Perona, D. Ramanan, C. Zitnick, P. Dollar, "Microsoft COCO: Common Objects in Context", https://cocodataset.org/#home.

[12] Open Source Robotics Foundation, Inc., "ROS-Robot Operating System", https://www.ros.org/.