

# Research of Robust Follower Recognition Function for Airport Guide Robot

Zijun SHA, Ph.D.<sup>1</sup>Yoichi NATORI<sup>1</sup>Takahiro ARIIZUMI<sup>1</sup>

## ABSTRACT

Hospitality is a key factor for the success of guide robot service. This paper defines a new recognition function called follower recognition, and proposes a hybrid approach to realize the follower recognition function. This approach creates a fusion of facial recognition, human body tracking, and person re-identification algorithms and takes advantage of the respective technological characteristics to realize robust follower recognition against issues such as recognition distance change, illumination change, and occlusion from the robot's perspective. The proving test result at Narita International Airport showed that the proposed approach is practical in guide robot service.

## 1. Introduction

With the shrinking of the working population due to a chronically low birthrate and rapidly aging population, an increase in single-person elderly households, and heightened interest in enhancement of living standards as lifestyles have become more diverse and efficient, demand is increasing for service robots that can coexist and cooperate with people. One example of this is social implementation of guide robots that are capable of autonomous movement in environments where people come and go<sup>(1)</sup>. The authors had the goal of realizing a robot that performs useful functions in society while coexisting and cooperating with people, and developed and conducted a proving test of an airport guide robot system equipped with follower recognition functionality that enables a robot to lead a follower to the destination within Narita International Airport (hereafter, Narita Airport). This system provides people with the new experience of being guided to their destination by a robot, and has the goal of showing the potential of new guide services, creating robots that are not demonstrations but rather coexist with people and perform useful functions in the real world<sup>(2)</sup>.

The proposed guide service is described in detail in references<sup>(2), (3)</sup>, and Fig. 1 shows an outline of the utilization flow. As shown in Step 1, the follower performs guide reception such as setting the destination through interaction with a guide service reception robot (reception robot) installed in the airport. At this time, a registration image for follower face identification is acquired by the

reception robot, and the facial feature values are transferred to the management system. Next, the guide robot acquires the destination and the facial feature values from the management system and commences guide operations (Step 2). During guide operations, the guide robot recognizes the follower's position using a camera mounted on the back of its head, as shown in Fig. 2, automatically adjusts its speed to match the movement speed of the follower, and leads the follower while maintaining an appropriate distance, as shown in Step 3. When the destination is reached, the guide service ends and the robot returns to the reception area (Step 4).

In order to provide the follower with the experience of "being guided by a robot," it is thought that simply following a robot that moves monotonically toward the destination is not sufficient, and that the follower needs to feel that the robot is operating according to the follower. In other words, the follower feels "guided" when the robot proceeds with a comfortable sense of distance according

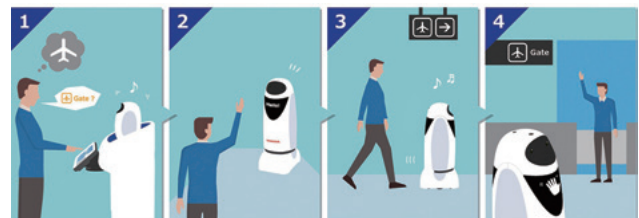


Fig. 1 Utilization flow of guide robot service

<sup>1</sup> Life Creation Center

to the follower's movement pace, as shown in Fig. 2. It is thought that robot operation according to the follower's pace in this manner will lead to enhanced guide service hospitality and increase overall satisfaction with guide service. A key factor for achieving this is that the robot continues to recognize the follower's position in a stable manner so that robot operation does not become unnatural or awkward. This paper proposes an approach for realizing a robust follower recognition function that can be implemented in a guide robot that moves autonomously in the complex indoor environment of an airport, and demonstrates robust recognition through evaluation using image data acquired in an airport. Hereafter, Chapter 2 describes the relationship between conventional approaches and follower recognition, and Chapter 3 describes follower recognition using the proposed approach. Chapter 4 describes the evaluation data, and Chapter 5 examines the evaluation results and the usefulness of the proposed approach.

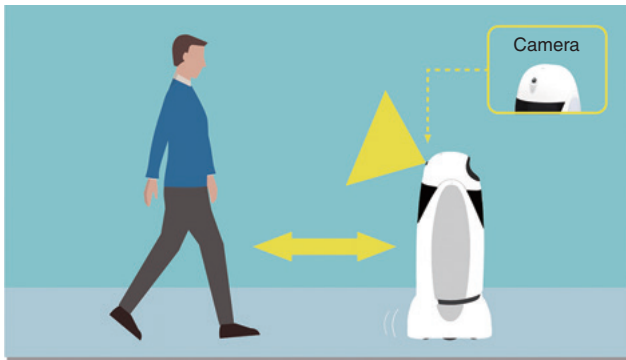


Fig. 2 Adjustment of distance from follower

## 2. Conventional Approaches

The follower recognition function continues to stably recognize the relative position of the person who is the follower from the perspective of the leading guide robot in a dynamic environment of movement. No precedents could be found at the time of writing this paper of technology for a guide robot to maintain an appropriate relative distance by using a camera to recognize the relative position of the follower and adjust its movement speed accordingly.

Although there are few cases of R&D for follower recognition, the person-following<sup>(4)</sup> function whereby a robot follows a person, as shown in Fig. 3, has been mentioned and active R&D has been carried out since the 1990s as a technique for recognizing the dynamic relative position with a person from the robot's perspective in the same manner as follower recognition<sup>(5)-(7)</sup>. The biggest difference from follower recognition is that the person leads in front of the robot. However, there are many common issues with follower recognition in terms of recognizing

people in dynamic and complex environments, and R&D is underway on corresponding measures using a wide variety of sensors and algorithms. Pierre<sup>(8)</sup> performed person recognition and tracking from a computer vision approach using a monocular camera. Chen et al.<sup>(9)</sup> and Wang et al.<sup>(10)</sup> realized person following by combining a computer vision approach using a stereo camera and RGBD images with depth information. Leigh et al.<sup>(11)</sup>, Peng et al.<sup>(12)</sup>, and Germa et al.<sup>(13)</sup> realized tracking by various means such as a laser range finder (LRF), ultrasonic sensor, and radio frequency identifier (RFID).

Person re-identification (Person re-id)<sup>(14)</sup> algorithms identify persons from whole-body and half-body images between cameras and frames, as shown in Fig. 4, and have continued to evolve. A Person re-id algorithm mainly uses a whole-body color image of a person as the input, extracts appearance features such as the shape and color of the person using a deep learning model, and compares



Fig. 3 Person following



Fig. 4 Person re-id

them with the same features acquired from the registered image to check if it is the same person. The authors previously worked to adapt a Person re-id algorithm to robot perspective images<sup>(15)</sup>. Person re-id algorithms have lower personal authentication accuracy compared with face identification, but high robustness against posture and distance changes, and thus they are also thought to be adaptable to person following and follower recognition. However, there is the precondition that advance registration of the whole-body or half-body image of the recognition subject is needed.

It is thought that these conventional approaches related to person following can serve as references for realizing a follower recognition function. However, there are also characteristics that differ from those of person following, such as being able to see the follower's face and the challenge of predicting the follower's behavior (looking away, detour, sudden stop, U-turn, etc.), and thus an approach similar to person following is not necessarily suitable.

A method based on conventional approaches that continually recognize the follower by facial recognition (including face detection, face identification, and face tracking), human body tracking, and the use of special devices (RFID, etc.) and marks (QR code, etc.) can be considered as a way of realizing follower recognition. However, these conventional approaches have various issues, such as those described below.

When using facial recognition, robustness is extremely low with respect to issues specific to robots arising from a low viewpoint and dynamic changes in the surrounding environment due to movement<sup>(16), (17)</sup>. These issues include recognition distance and illumination changes, image blur, cutoff from the camera angle of view, and occlusion by objects or people. For example, the follower's face may be lost suddenly due to face cutoff, such as when the follower turns sideways or as a result of sudden illumination changes associated with indoor movement. In addition, multiple variations of face images can be acquired and registered in advance to increase robustness even slightly, but it is not preferable from the viewpoint of usability to take multiple photos with different postures when acquiring images for face identification with the reception robot.

Human body tracking performs tracking between frames in a time series by comparing relatively monotonous feature values of the person (for example, local features such as color and edges, and relative positions of bounding boxes in the person detection results), in the area around the tracking target. When using this human body tracking, it is necessary to recognize by some means at the start of guide operations that the person in the camera image is the follower. In addition, it is expected that there will be frequent cases during tracking where tracking is unsuccessful due to occlusion or other factors in the same manner as with facial recognition described above. Furthermore, individuals cannot be identified by human body tracking alone, which

makes recovery from unsuccessful tracking a challenge<sup>(18)</sup>.

The use of a special device or mark places the extra burden on the follower of carrying or wearing the device or mark during guidance, which leads to reduced hospitality and usability. Furthermore, RFID is able to grasp that the target is nearby, but not the relative position of the target.

In this manner, with regard to a follower recognition function that uses separate algorithms based on conventional approaches, issues remain for realizing a guide robot with high hospitality.

### 3. Proposed Approach

This paper proposes a follower recognition approach that creates a fusion of Face/Body tracking and Person re-id algorithms in order to address the issues with conventional approaches described in Chapter 2 and realize a highly robust follower recognition function. The proposed approach took the two-step method of creating a base function by combining facial recognition and human body tracking for RGB monocular camera images, and then combining Person re-id with that base function. Use of only a monocular camera instead of combining various sensors was decided on with a view toward guide robot safety (impact reduction in the event of collision or tip-over) as well as payload and cost aspects, etc.

#### 3.1. Follower Recognition by Fusion of Face/Body Tracking

This section proposes an approach for fusing the features of facial recognition and human body tracking. Figure 5 shows an outline of the Face/Body tracking processing. Face/Body tracking detects the face and body

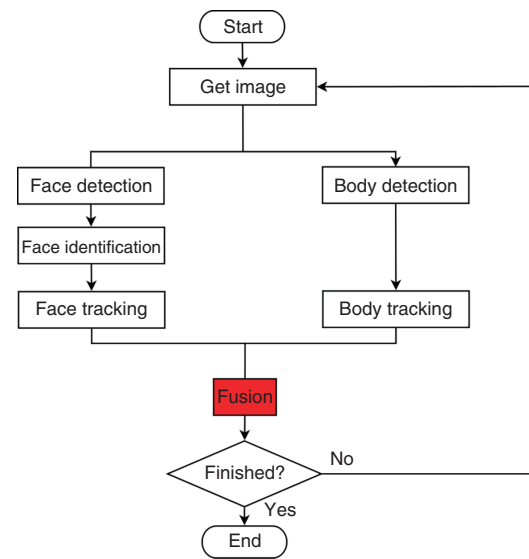


Fig. 5 Flow chart of Face/Body tracking based follower recognition

from the image acquired by the camera, and assigns a face-tracking ID to the detected face area and a body-tracking ID to the body. For the face area, the degree of difference from the registered facial features of the follower is calculated and an ID score is obtained. When the ID score exceeds the threshold, the face is determined to be that of the registered follower (face identification successful) and the face-tracking ID corresponding to the follower is updated. Tracking is performed separately for the face and the person, and the same tracking ID is retained for each when tracking is successful.

The fusion processing shown in Fig. 5 creates a fusion of the facial recognition and human body tracking results. Figure 6 shows the specific fusion method. The face-tracking ID (face\_id) and body-tracking ID (body\_id) for the same person are linked using the relative positional relationship between the bounding boxes of the detected/tracked face and person, and the face- and body-tracking IDs corresponding to the follower are updated. Figure 7 shows an example of linking judgment by the relative positions of the face and body bounding boxes corresponding to the red part in Fig. 6. By bundling and retaining the face- and body-tracking IDs corresponding to the follower in this manner, follower recognition can be expected to continue through mutual complementation even if one of the recognition approaches is unsuccessful. That is

to say, even if facial recognition is unsuccessful, follower recognition can continue as long as tracking of the linked body-tracking ID is successful. Conversely, even if the body cannot be detected/tracked, follower recognition can be performed as long as the face can be detected and facial recognition is successful.

In a complex environment such as an airport, however, facial recognition and human body tracking may be unsuccessful at the same time, and further enhancement of robustness is necessary to achieve stable follower recognition.

### 3.2. Follower Recognition by Fusion of Face/Body Tracking and Person re-id

The proposed approach uses Person re-id to supplement the abovementioned Face/Body tracking fusion approach and further enhance the robustness of follower recognition. Compared with face identification, Person re-id can identify a person from a whole-body or half-body image even when occlusion occurs or in a long-distance (low-resolution) image, and has a high likelihood of being able to recognize the follower even in scenes where recognition by the Face/Body tracking fusion approach is a challenge. However, the identification performance is not as good as that of facial recognition under conditions ideal for recognition, such as illumination and face orientation. In addition, as there are many feature values to handle and the processing load is high, Person re-id is slower than facial recognition. Furthermore, a registered whole-body or half-body image of the follower is needed, but the camera of the reception robot mainly takes a photo of the follower's face during guide service reception. For these reasons, the proposed approach mainly uses the results of successful recognition by high-speed Face/Body tracking based on highly reliable face identification, and uses Person re-id in a supplementary manner.

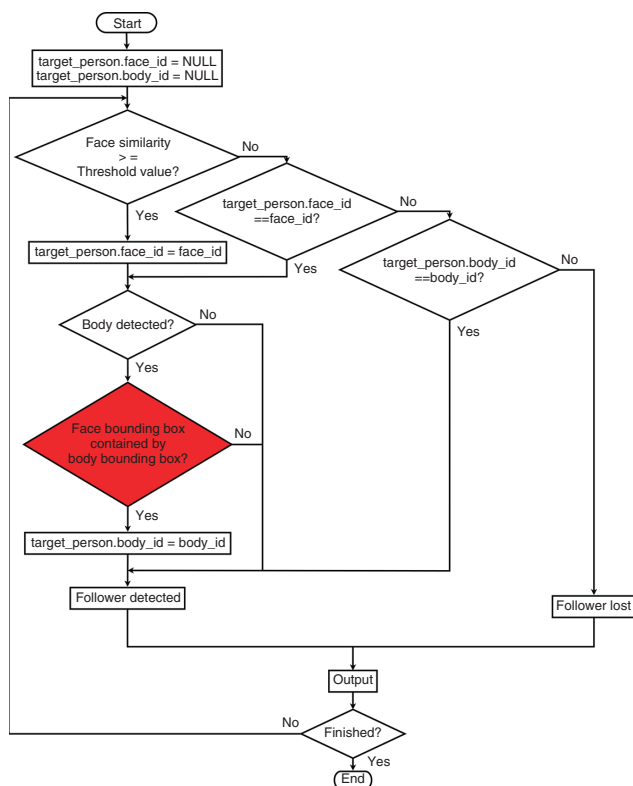


Fig. 6 Flow chart of Face/Body tracking fusion

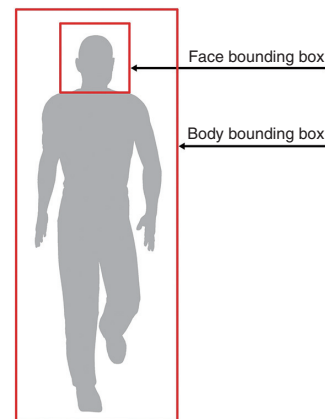


Fig. 7 Link of Face/Body bounding boxes

Figure 8 shows an outline of the processing. The part enclosed by the red frame in the figure corresponds to the Face/Body tracking approach shown in Fig. 5. The Face/Body tracking and Person re-id processing proceed separately. The system first confirms the follower recognition results of the Face/Body tracking approach. While follower recognition by Face/Body tracking is successful, a whole-body registration image for Person re-id is acquired at regular intervals and added to update the

database. This enables retention of the most recent follower whole-body image, which serves to increase the robustness of recognition even with respect to changes in the illumination environment or changes in appearance, such as taking off a jacket or changing the way a bag is held. When follower recognition using the Face/Body tracking approach is unsuccessful, the Person re-id processing results are used as the final follower recognition results. This fusion of the Face/Body tracking and Person re-id approaches can be expected to realize an effective follower recognition function that compensates for the disadvantages of each approach.

#### 4. Evaluation Data

The follower recognition function was verified by actually operating a guide robot in the departures area of Narita Airport Terminal 1, as shown in Fig. 9<sup>(19)</sup>. In order to quantitatively and comprehensively evaluate the performance of the approach proposed in Chapter 3 in various scenes, images covering sufficient distance as a guide route and including various environmental changes along the way were used as evaluation data.

Specifically, a camera was installed at a height of 1 m and an elevation angle of 25° on a cart simulating an actual guide robot. The image data used for evaluation consisted of approximately 9 minutes (5541 frames) acquired at a resolution of 2056 × 1542 and a shooting speed of 10 fps while moving from the Central Building to the edge of the No. 4 Satellite of Terminal 1 in a manner simulating actual guide operations. The guide robot did not adjust its speed by follower recognition while acquiring the test data. In addition, the test subject follower followed the robot while maintaining a distance simulating the condition of being guided by the robot, but performed various actions along the way, such as stopping in front of shops in the airport

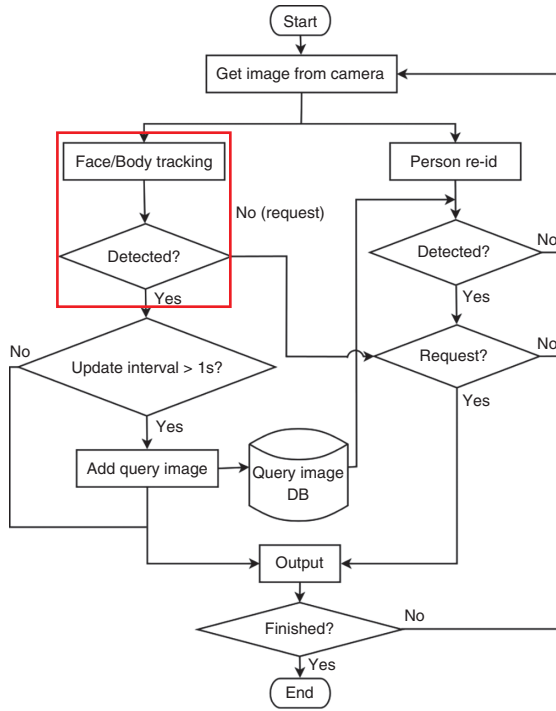


Fig. 8 Fusion of Face/Body tracking and Person re-id



Fig. 9 Evaluation test path on 3rd floor of Narita Airport Terminal 1

and looking to the left and right. Figure 10 shows examples of the test data, and the corresponding shooting locations are indicated in Fig. 9. Various situations were recorded in which robust recognition functions are needed due to dynamic changes in the environment, such as illumination changes due to factors such as backlight and lighting changes, image blur due to guide robot movement, face orientation and distance changes and cutoff due to changes in follower behavior, taking off and putting on a mask, and occlusion due to intersection with people and objects, etc. The red mark in the figures indicates the test subject follower.

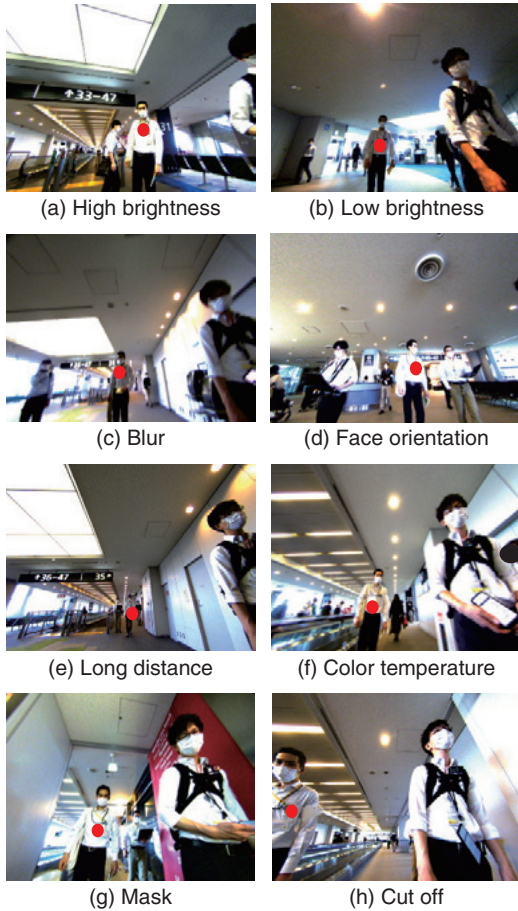


Fig. 10 Scenes that call for robust recognition

## 5. Evaluation Results and Examination

Figure 11 shows result images of the follower recognition processing, and the corresponding shooting locations are indicated in Fig. 9. Figure 11(a) shows the case when the Face/Body tracking approach succeeded in both facial recognition and human body tracking, and the respective bounding boxes (orange frames in this

Table 1 Results

|                  | Total (Frame) | True Positive (Frame) | False Positive (Frame) | False Negative (Frame) | Success (%) |
|------------------|---------------|-----------------------|------------------------|------------------------|-------------|
| Face/Body (1)    | 5541          | 4540                  | 0                      | 1001                   | 81.9        |
| Person re-id (2) | 1001          | 835                   | 32                     | 134                    | 83.4        |
| (1) + (2)        | 5541          | 5375                  | 32                     | 134                    | 97.0        |

case) could be linked. Figure 11(b) shows the case when facial recognition was unsuccessful due to the impact of a mask, but the linked human body tracking was successful. Figure 11(c) shows the case when both facial recognition and human body tracking were unsuccessful, but the follower could be recognized by Person re-id. Figure 11(d) shows the case when all approaches were unsuccessful, resulting in unsuccessful follower recognition.

Table 1 shows the results of evaluating the proposed approach using the data described in Chapter 4. Evaluation was carried out by visually confirming whether the bounding box in the processing results surrounded the follower as shown in Fig. 11. True Positive (TP) in Table 1 is the number of frames where the system was judged to have correctly recognized the follower. False Positive (FP) is the number of frames where a person other than the follower was recognized as the follower. False Negative (FN) is the number of frames where the follower was present but could not be recognized. Success is the follower recognition success rate. In other words, out of a

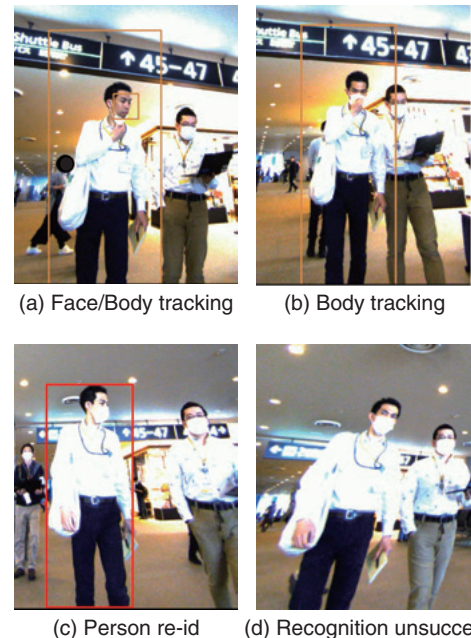


Fig. 11 Result images of proposed follower recognition

total 5541 frames of test data images, recognition by only the Face/Body tracking fusion approach was successful in 4540 frames, with 0 FP frames and 1001 FN frames, for a recognition success rate of 81.9%. Of the 1001 frames where Face/Body tracking was unsuccessful, the Person re-id approach was able to correctly recognize the follower in 835 frames, with 32 FP frames and 134 FN frames, for a Person re-id approach recognition success rate of 83.4%. Based on the above, the final recognition success rate of the proposed approach that created a fusion of Face/Body tracking and Person re-id was 97.0%.

Figure 12 plots the time series data of the distance from the robot to the follower estimated based on the follower recognition results for the test data used for evaluation. This graph shows the results of the Face/Body tracking approach alone in orange and the results of Person re-id alone in blue. The distance to the follower was estimated by the size of the bounding box that detected the follower. Frames with the estimated distance plotted as lost are scenes in which follower recognition was unsuccessful. Of these, locations where lost is plotted in orange but the estimated distance is plotted in blue for the same frame are scenes where the Face/Body tracking approach was unsuccessful but the follower could be correctly recognized by Person re-id. The red frame in Fig. 12 is an example of this, and shows that recognition by the Face/Body tracking approach was unsuccessful because the distance to the follower increased, but recognition by Person re-id was successful. In this manner, stable follower recognition can be realized by creating a fusion of the Face/Body tracking and Person re-id approaches. Recognition by the Face/Body tracking approach was unsuccessful in each of the scenes that call for robustness shown in Fig. 10, but use of Person re-id led to recognition success, which shows that the desired effect was achieved.

According to the results in Table 1, however, there were 166 frames or approximately 3% of the total scenes where recognition was not achieved even by Person re-id, with the result that follower recognition was unsuccessful. Events that may occur when follower recognition is unsuccessful are (1) the FP case where the follower is not present but is recognized as present, and the robot continues to guide a person other than the follower,

and (2) the FN case where the follower is present but is recognized as not present, resulting in a lost judgment even though the follower is present. In both cases, the longer the unsuccessful condition continues, the greater the impact on guide service. In this evaluation, the longest period during which Case (1) continued was 1.5 s. The top speed of the guide robot in this guide system is 1.11 m/s, and thus in this case the robot runs freely for around 1.7 m, but this is not far from the previous recognition successful position, and is thought to be a distance that enables adequate adjustment after re-recognition. The longest period during which Case (2) continued was 3 s. This is a scene where the follower was walking nearly side by side with the robot and was nearly cut off from the angle of view, such as shown in Fig. 10(h). It is explained to the follower beforehand at the start of this guide service that the follower should, in principle, follow, i.e. be behind, the robot. However, this kind of side-by-side positioning and large deviation from the angle of view are also assumed, and the guide robot enters deceleration mode when a lost judgment is made. Deceleration makes it easier for the follower to notice the unusual behavior of the robot and provide more opportunities to return to being behind the robot, which can be expected to increase the likelihood of re-recognition. The robot returns to normal speed after re-recognition, and thus FN of a few seconds is not considered to pose a hindrance. From the above, it can be said that the effective follower recognition approach proposed in this paper, which creates a fusion of Face/Body tracking and Person re-id to take advantage of the strengths of each to complement their respective disadvantages, achieves sufficient recognition performance for guidance in airports with a drastically changing environment. As such, it is considered that prospects have been achieved for practical use.

The factors resulting in unsuccessful follower recognition are analyzed below. The biggest factor for lack of success was confirmed to be an unsuccessful person detection process in the stage before Person re-id processing. This occurred in 96 frames out of the 166 recognition unsuccessful frames, and accounted for nearly 60% of unsuccessful follower recognition. The remaining 40% of unsuccessful recognition cases was

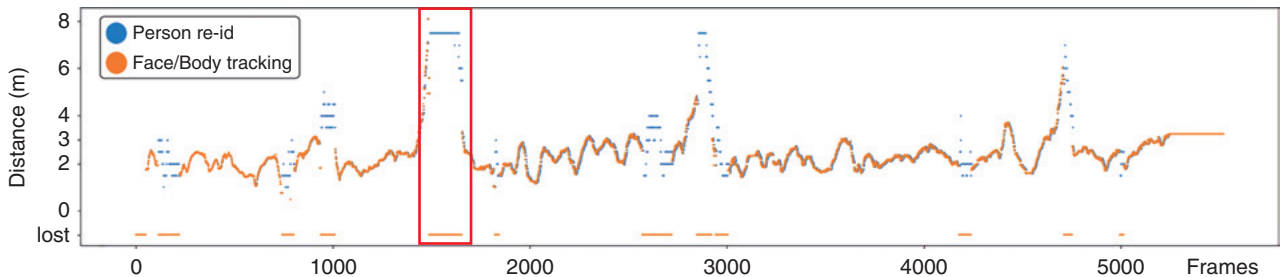


Fig. 12 Estimated distance

misrecognition by Person re-id. Figure 13 shows an example of misrecognition (FP), but the misrecognized person and the follower have similar appearances that even a human may mistake at first glance. Looking ahead to practical use, further enhancement of accuracy is thought to be an issue for recognizing these types of scenes where follower recognition was unsuccessful. Potential means for addressing this include selection of algorithms for each recognition processing (facial recognition, person detection, human body tracking, Person re-id) in this approach that are better suited to specific environments, careful examination and expansion of the amount of learning data, and addition of enhancements such as mutual complementation using various sensors, etc.



Fig. 13 Person re-id False Positive

## 6. Conclusion

This paper proposed a robust follower recognition approach that uses a monocular RGB camera mounted on a guide robot in an airport guide robot system. In the proposed approach, the existing recognition algorithms of the Face/Body tracking and Person re-id approaches were effectively fused to enhance robustness against various issues arising frequently in a dynamic airport environment and that reduce the follower recognition rate, such as occlusion due to crowds of people, changes in illumination due to movement, and changes in recognition distance and face orientation due to follower movement. It was confirmed that high robustness was achieved, with recognition successful for 97% of the evaluation data acquired in an airport.

## References

- (1) Velentza, A.-M., Heinke, D., Wyatt, J.: Human Interaction and Improving Knowledge through Collaborative Tour Guide Robots, 2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN), p. 1-7, (2019)
- (2) Saito, Y., Koshiishi, T.: Development and Styling Design of Airport Guide Robot System, Honda R&D Technical Review, Vol. 33, No. 1, p. 1-9
- (3) Onishi, J., Masuda, A., Obayashi, C., Kawasaki, Y., Takebe, Y.: Development of Reception Robot for Airport Guidance, Honda R&D Technical Review, Vol. 33, No. 1, p. 10-18
- (4) <https://talorongilad.com/research-laboratories/follower-me-design-for-person-following-robots/>, 2021/01/19
- (5) Azarbayejani, A., Pentland, A.: Real-Time Self-Calibrating Stereo Person Tracking Using 3-D Shape Estimation from Blob Features, Proceedings of 13th International Conference on Pattern Recognition, Vol. 3, p. 627-632, (1996)
- (6) Wren, C. R., Azarbayejani, A., Darrell, T., Pentland, A. P.: Pfnder: Real-Time Tracking of the Human Body, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 19, No. 7, p. 780-785, (1997)
- (7) Darrell, T., Gordon, G., Harville, M., Woodfill, J.: Integrated Person Tracking using Stereo, Color, and Pattern Detection, Proceedings. 1998 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), p. 601-608, (1998)
- (8) Pierre, J. M.: End-to-End Deep Learning for Robotic Following, Proceedings of the 2018 2nd International Conference on Mechatronics Systems and Control Engineering, p. 77-85, (2018)
- (9) Chen, B. X., Sahdev, R., Tsotsos, J. K.: Person Following Robot Using Selected Online Ada-Boosting with Stereo Camera, 2017 14th Conference on Computer and Robot Vision (CRV), p. 48-55, (2017)
- (10) Wang, M., Su, D., Shi, L., Liu, Y., Miro, J. V.: Real-time 3D human tracking for mobile robots with multisensors, 2017 IEEE International Conference on Robotics and Automation (ICRA), p. 5081-5087, (2017)
- (11) Leigh, A., Pineau, J., Olmedo, N., Zhang, H.: Person Tracking and Following with 2D Laser Scanners, 2015 IEEE International Conference on Robotics and Automation (ICRA), p. 726-733, (2015)
- (12) Peng, W., Wang, J., Chen, W.: Tracking Control of Human-Following Robot with Sonar Sensors, International Conference on Intelligent Autonomous Systems, p. 301-313, (2017)
- (13) Germa, T., Lerasle, F., Ouadah, N., Cadenat, V.: Vision and RFID Data Fusion for Tracking People in Crowds by a Mobile Robot, Computer Vision and Image Understanding, p. 641-651, (2010)
- (14) Gou, M., Karanam, S., Liu, W., Camps, O., Radke, R. J.: DukeMTMC4ReID: A Large-Scale Multi-Camera Person Re-Identification Dataset (Open Access version), 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops, p. 10-19, (2017)
- (15) Sha, Z., Zeng, Z., Wang, Z., Natori, Y., Taniguchi, Y., Satoh, S.: Progressive Domain Adaptation for Robot Vision Person Re-identification, MM '20: Proceedings

- of the 28th ACM International Conference on Multimedia, p. 4488-4490, (2020)
- (16) Chen, J., Chen, J., Wang, Z., Liang, C., Lin, C.-W.: Identity-Aware Face Super-Resolution for Low-Resolution Face Recognition, IEEE Signal Processing Letters, Vol. 27, p. 645-649, (2020)
- (17) Yuan, J., Chen, H., Sun, F., Huang, Y.: Multisensor information fusion for people tracking with a mobile robot: A particle filtering approach, IEEE Trans. Instrumentation and Measurement, Vol. 64, No. 9, p. 2427-2442, (2015)
- (18) Ruan, W., Liu, W., Bao, Q., Chen, J., Cheng, Y., Mei, T.: POINet: Pose-Guided Ovonic Insight Network for Multi-Person Pose Tracking, Proceedings of the 27th ACM International Conference on Multimedia, p. 284-292, (2019)
- (19) <https://www.narita-airport.jp/en/map?terminal=1&map=3>, 2021/01/12

■ Author ■



Zijun SHA



Yoichi NATORI



Takahiro ARIIZUMI