Development of Methods for High-Density Crowd Measurement and Tracking in Railway Station Concourses

Ryosuke Hishida[1], Miho IRYO[2], Rizki ALHAMDANI[2], Kazuki MANO[3] Youhei YAMAGUCHI[3],
Graduate School of Environmental Studies, Nagoya University／Technology Research and Development Department, General Technology Division, Central Japan Railway Company[1] (1545-33 Ohyama, Komaki, 485-0801 Japan),
Graduate School of Environmental Studies, Nagoya University[2] (Furo-cho, Chikusa-ku, Nagoya, 464-8603 Japan),
Kurusugawa Computer Inc.[3] (1-29-23 Sinsakae, Naka-ku, Nagoya, 460-0007 Japan)

Corresponding author: Ryosuke Hishida
Email: ryosuke.hishida@jr-central.co.jp

Abstract

Understanding the characteristics of pedestrian flows in large-scale railway stations is demanding. Not only for designing new stations or renovating existing ones, but also for applying new technologies such as autonomous robots for goods transportation, it is essential to conduct appropriate evaluations based on a fine measurement of pedestrian flow. However, there are only a limited number of cases of large-scale measurements in public spaces, and quantitative verification of tracking accuracy targeting high-density crowds has not been conducted. Furthermore, there are no publicly available 3D point cloud datasets for station environments. In this research, we developed a wide-area, high-density pedestrian flow measurement and tracking system using twenty 3D-LiDAR sensors in a real-world environment to accurately capture and quantify pedestrian flow in crowded large-scale station concourse. By employing an offline person detection model using deep learning and tracking compensation processing using both past and future point cloud data, we confirmed that high-accuracy person detection and tracking with HOTA exceeding 90% is possible in an area of approximately 900m². Using the developed system, we successfully measured walking trajectory data with high accuracy and visualized flow trends using basic indicators and heatmaps derived from the measurement data, thereby evaluating the pedestrian flow characteristics in the station concourse environment.

Keywords: 3D-LiDAR, Point cloud data, Pedestrian flow measurement, People detection and tracking, Deep learning

1 Introduction

Due to the decline in the working-age population in the aging society, there is a need for the use of autonomous robots to support logistics in public spaces such as railway stations. Especially in large-scale terminals, there is an increasing number of cases where commercial facilities are developed by effectively using spaces within the stations. Accordingly, the increasing amount of goods transported within the terminals raises the demand of autonomous robots.

However, coexistence of pedestrians and autonomous robots under high pedestrian demand is a challenging issue. In order to make a path planning of the robots or to calibrate their avoidance behavior, measuring the pedestrian flow in large-scale areas under dense pedestrian flow is required.

The needs of flow measurement are not limited to autonomous robots. When designing new stations or renovating existing ones, it is essential to conduct appropriate evaluations based on the characteristics of pedestrian flow. Practically, many existing studies rely on qualitative and experiential judgments based on the accumulated knowledge of relevant stakeholders, which currently play a central role in decision-making.

Ideally, it is necessary to accurately understand the unique characteristics of pedestrian flow within station spaces—such as the coexistence of passing pedestrians and lingering pedestrians, the diversity of pedestrian purposes, and temporal variations due to factors like day of the week, time of day, and train arrivals—to create pedestrian-friendly spaces. Knowledge of these characteristics is currently limited, and data collection in real-world environments has not been extensively conducted to date. Pedestrian simulation models are often applied these days for the evaluation of new station designs, while their accuracy depends on the quality of the input data. We believe that understanding actual pedestrian flow in extensive and high-density environments is essential for crowd safety management and service optimization at public transportation hubs.

For conducting pedestrian flow surveys within station premises, Wi-Fi packet sensors [1] or image analysis of camera footage [2] are often utilized to track movement patterns in addition to traditional manual flow surveys. However, the Wi-Fi sensors are intended for wider-scale flow estimation due to the limitation of location accuracy

and are not suitable for detailed trajectory analysis. While the latter, being based on two-dimensional image analysis, makes it difficult to identify locations accurately. It also faces privacy concerns that restrict data collection.

Therefore, attention is being focused on the use of LiDAR, which has high anonymity because it only acquires depth information. 2D-LiDAR is inexpensive and easy to handle because it only scans planes horizontally or vertically, and has been used in many studies. For example, Lopes et al. have linked multiple 2D LiDAR sensors in an indoor environment to recognize behavior and cover a wide area [3]. However, the lack of either horizontal or vertical information makes it difficult to separate overlapping objects and detect dense obstructions, thus the risk of false detections is high. On the other hand, using 3D-LiDAR allows for height information to mitigate the impact of obstructions, enabling high-resolution analysis such as wheelchair detection. Na et al. have also demonstrated real-time multi-person tracking using a single 3D-LiDAR and applied it to robot navigation [4]. However, capturing pedestrian trajectories in the whole design area needs multiple sensors, which provides us with many challenges.

The most critical issue is the computational costs. Among the methods for 3D object detection from point clouds, numerous approaches have been reported, such as VoxelNet [5] and PointPillars [6]. However, these methods involve voxelization and large-scale convolution processing, making them computationally intensive for real-time inference on edge devices. Recent studies have proposed a modular approach for real-time person detection and tracking using LiDAR [7], but it is designed to detect several people in the vicinity using a small computer installed in a robot, not to handle large crowds in crowded places such as stations. On the other hand, IA-SSD [8] is a method that achieves fast and memory-efficient point cloud detection by prioritizing the processing of only the foreground points of the detection targets and reducing the computation and memory usage for background points. IA-SSD has been applied for real-time detection purposes for autonomous vehicles, while the features of saving computation costs would also be helpful for large-scale detection with multiple sensors.

Another issue is the static obstacles. To meet the requirements of structural strength, pillars are densely located in some stations. There are also various facilities, such as shops or work booths, for the diverse needs of station users. This condition makes it difficult to capture the crowd measurement in large-scale public spaces. Furthermore, quantitative tracking accuracy verification targeting high-density crowds in railway stations has not been conducted, and there are no publicly available 3D point cloud datasets for station environments. With the development of such a system and improvements in accuracy, it is expected that the data will eventually be made available to the public.

Therefore, this study aims to construct a high-precision tracking system by installing twenty 3D-LiDAR sensors in a railway station concourse where visibility is reduced due to numerous pillars. In the next section, the method to integrate measurement data from multiple sensors and the quantitative evaluation of its measurement accuracy under high-density crowds are described. Then, the analysis of pedestrian flow characteristics using the observed data is explained, followed by discussions and conclusions.

2 Construction of crowd measurement system

2.1 Target site

Nagoya Station in Japan is served by all Shinkansen (high-speed railway) trains on the Tokaido Shinkansen line, with conventional (non-Shinkansen) trains operating express services to various destinations from the station. Multiple other railway lines also connect at the station. Additionally, the station is surrounded by several large-scale commercial facilities directly accessible from the station, making it the largest terminal station in the Tokai region. The daily average number of passengers at Nagoya Station is approximately 220,000 (Shinkansen and conventional lines combined, as of the 2018 fiscal year [9]), a significant number of passengers.

In this study, the measurement area was defined as approximately 57m × 28m near the central gate of the conventional railway line in the central concourse as shown in Figure 1, which extends east-west and is located at the center of Nagoya Station. This area is bustling with people and features tourist information centers, booth-style shared office (Figure 2) work booths, restaurants, and shops. In this space, the particularly heavy pedestrian flow moves mainly in an east-west direction, toward the Shinkansen ticket gates and large commercial facilities. This is joined by people coming from the conventional railway ticket gates. Twenty 3D-LiDAR sensors were installed on-site as shown by the circles in Figure 1 to establish the measurement environment.

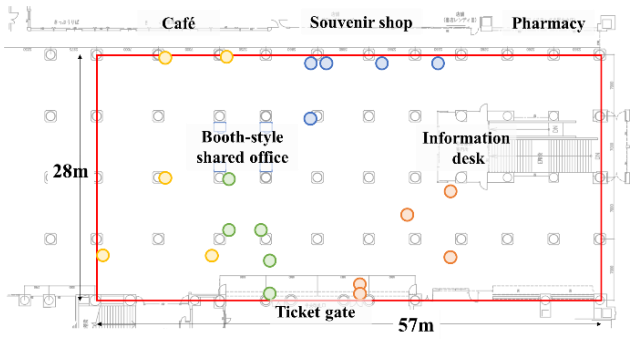

Figure 1. Measurement area and positions of 3D-LiDAR sensors

Figure 2. booth-style shared office installed in the passageway

2.2 System configuration

The on-site equipment consists of four sets, each comprising two small Linux PCs (NVIDIA Jetson Orin NX) processing five LiDAR sensors. This configuration was chosen due to network bandwidth limitations, which made it difficult to connect all equipment to a single network. The specifications of the 3D-LiDAR system adopted in this study are shown in Table 1. The software configuration is as shown in Figure 3. The driver receives point cloud data from the LiDAR sensors, which operate at 10 Hz, processes it through a preprocessing module, and then processes it in the person detection module. The detection results are stored in real time on an SSD within the edge device. The stored data is then integrated offline and the tracking processing is performed. The reason for performing offline integration and tracking processing is that, as mentioned earlier, the network must be divided, making online integration difficult, and offline tracking processing allows for a more accurate method that does not prioritize real-time performance.

Table 1. Specifications of the adopted sensor

Specifications	Livox HAP(TX)	Robosense M1P
Detection distance [m]	150	180
Horizontal FOV [°]	120	120
Vertical FOV [°]	25	25
Horizontal angular resolution [°]	0.23	0.2
Vertical angular resolution [°]	0.18	0.2
Point density [point/s]	452,000	750,000
Frame rate [Hz]	10	10

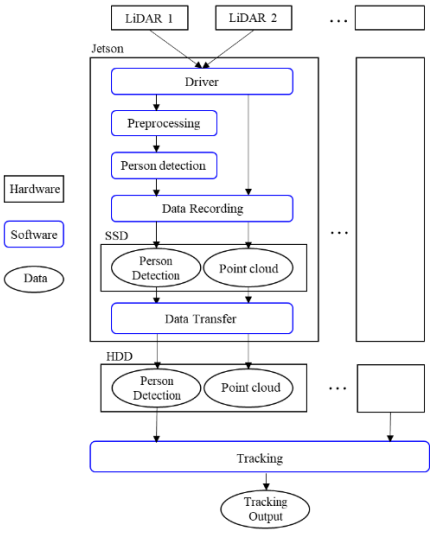

Figure 3. Software configuration

Figure 4 shows a single scene obtained by integrating point clouds measured by 20 LiDAR units within the station concourse. As in the figure, the pillars are evenly spaced, which makes visibility poor. This visibility deterioration and the crowd in the concourse space cause occlusions in detection. We carefully selected the sensor installation locations after thorough on-site surveys to avoid occlusion as much as possible. Note that for the integration of point cloud data from each LiDAR, we first created a SLAM map of the installation environment, performed a rough alignment using the point cloud visualization tool CloudCompare, and then used the ICP method to perform detailed alignment by searching for the nearest points between the point cloud data of the 3D shapes and converging them to minimize evaluation errors.

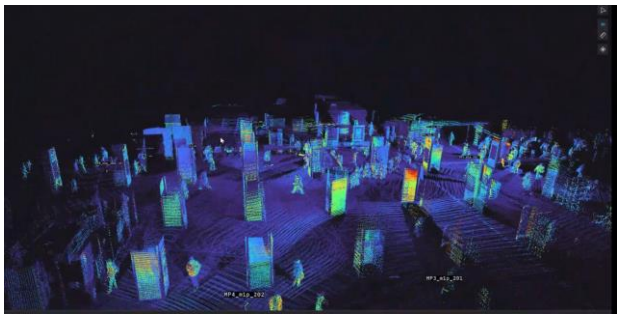

Figure 4. Actual measured point cloud

2.3 Person detection using deep learning

Person detection is performed by annotating point cloud data measured in advance at the site and then constructing a model based on that annotation. The annotation work uses a segmentation method to identify each person in the point clouds and was performed using the annotation tool AnnoFab (Figure 5) provided by Kurusugawa Computer Inc. The annotation was performed for 10 seconds per scene (20 frames at 0.5-second intervals) for a total of 312 frames.

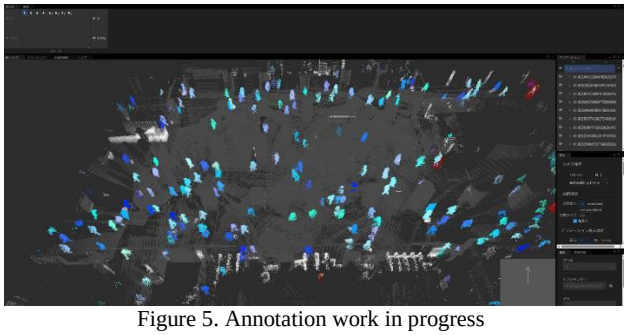

Figure 5. Annotation work in progress

For person detection, we adopted a machine learning method called IA-SSD [10]. IA-SSD is a method that excels in fast processing speed and efficient memory usage. In this study, we separated the input point cloud into foreground point clouds and background point clouds, and input only the foreground point clouds into IA-SSD.

The background point clouds represent static environmental —information. We manually defined background areas on a pre-constructed SLAM map, and the corresponding point clouds were removed. This provides the following advantages.

1. Reduction of false detections: By removing background point clouds, false detections of background areas such as walls and ground are prevented.
2. Improved model generalization: By suppressing context differences caused by the background, the same detection model can be applied more easily in different environments, making it easier to reuse models across multiple LiDAR sensors (20 sensors in this study).

In the conventional IA-SSD, learning-based downsampling selects points close to the center of the bounding box. However, in this study's setup, which uses only foreground point clouds as input, learning-based sampling was found to be ineffective. Therefore, we replaced the downsampling module with a geometry-based method and successfully suppressed the decrease in recall.

Furthermore, to accelerate inference, we optimized the pre-trained IA-SSD model using TensorRT, NVIDIA's deep learning inference optimization SDK. TensorRT enables significant improvements in inference speed on GPUs through techniques such as model structure optimization and low-precision conversion (FP16/INT8).

2.4 Tracking algorithm

In this system, as described in Section 2.2, offline tracking processing is performed on data obtained from real-time person detection. The reason for performing tracking processing offline is not only to reduce the load on edge-side processing but also to improve tracking accuracy by utilizing future trajectory information, taking into account the impact of person detection errors caused by occlusions that inevitably occur to some extent. When assigning tracking IDs, as shown in Figure 6, we first determine the trajectory of a person from adjacent time points, then perform processing to connect and interpolate the disconnected segments of the trajectory before and after the gap.

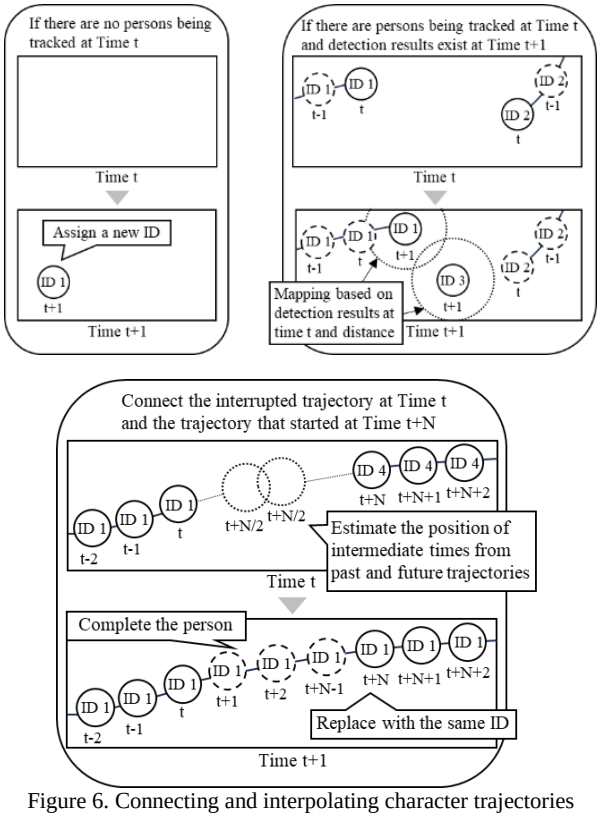

Figure 6. Connecting and interpolating character trajectories

Specifically, the system remembers the time and position of N frames (set to 5 frames = 0.5 seconds) before and after the trajectory is interrupted. It then assumes that the trajectory moves at a constant speed in a straight line during the N frames and the interrupted portion, and estimates the trajectory of the interrupted portion. At the intermediate time t between the start and end of the interruption, the system calculates the estimated positions of the trajectories before and after the interruption. If the distance between the estimated positions is within a certain limit (0.5 meters), the trajectory is connected as belonging to the same person.

2.5 Accuracy Indicator

In this study, we adopt the HOTA [11] (High Order Tracking Accuracy) metric proposed by Luiten et al. HOTA is a metric that calculates the score as follows, where objects for which the IoU (Intersection over Union) between the predicted bounding box and the ground truth bounding box is greater than or equal to the threshold α in each frame are considered True Positive (TP). A False Negative (FN) is a ground truth boundary box that does not match any predicted boxes. A False Positive (FP) is a predicted box that does not march any ground truths.

HOTA_α, HOTA with given α, is decomposed to detection accuracy score $\mathrm{DetA}(\alpha)$ and association accuracy score $\mathrm{AssA}(\alpha)$ as below.

$$\mathrm{HOTA}_\alpha = \sqrt{\mathrm{DetA}(\alpha) \times \mathrm{AssA}(\alpha)} \quad (1)$$

$$\mathrm{DetA}(\alpha) = \frac{|\mathrm{TP}|}{|\mathrm{TP}| + |\mathrm{FN}| + |\mathrm{FP}|} \quad (2)$$

$$\mathrm{AssA}(\alpha) = \frac{1}{|\mathrm{TP}|} \sum_{c \in \mathrm{TP}} \frac{|\mathrm{TPA}(\alpha)|}{|\mathrm{TPA(c)}| + |\mathrm{FNA(c)}| + |\mathrm{FPA(c)}|} \quad (3)$$

Note that |TP|, |FP| and |FN| represent the numbers of TP, FP and FN in a certain frame, respectively. True Positive Associations, TPA(c), is defined using a given $c \in$ TP. The set of TPA is the set of TPs in a different frame from the object frame, in which the tracking IDs in both ground truth and predicted trajectories are the same as *c* in Figure 7. This indicates the trajectories in the other frames are continuously traced with the same tracking ID. FNA (False Negative Association) is a set of ground truth boxes which have the same tracking ID as *c* but the IDs of predicted ones do not match. FPA (False Positive Association) is a set of predicted boxes which have the same tracking ID as *c* but the IDs of ground truth do not match.

In the original paper by Luiten et al., considering the localization errors, HOTA_α is evaluated in the whole valid range of parameter α from 0 to 1 by taking the integral. The final HOTA value is calculated as below.

$$\mathrm{HOTA} = \int_0^1 HOTA_\alpha d\alpha \quad (4)$$

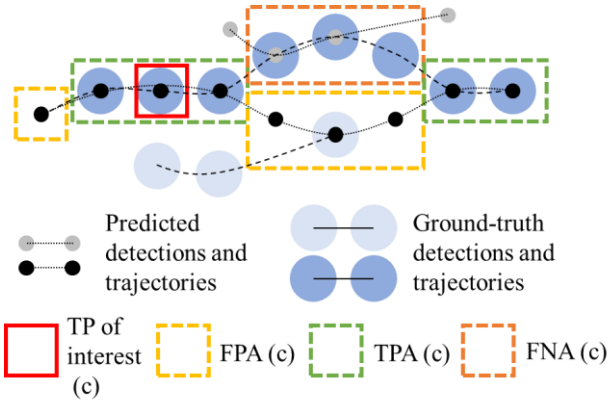

Figure 7. Conceptual diagram of HOTA

Since the evaluation environment of this study deals only with people, we introduced a few simplified treatments from the original definition. Firstly, a distance-based method instead of IoU as a matching condition to enable more intuitive and appropriate evaluation. The MatchScore is calculated as follows for the Euclidean distance *d* between the detection center coordinates $\boldsymbol{c_{det}}$ and the true center coordinates $\boldsymbol{c_{gt}}$.

$$d = \left\| \boldsymbol{c_{det}} - \boldsymbol{c_{gt}} \right\|_2 \quad (5)$$

$$\mathrm{MatchScore} = \max\left(0, 1 - \frac{d}{0.5}\right) \quad (6)$$

Considering the average shoulder width of a person, this MatchScore is evaluated as TP or TPA when it exceeds the threshold α=0.5m, and $\mathrm{HOTA}_{\alpha=0.5}$ is used as the evaluation index. DetA and AssA represent detection performance and tracking performance, respectively, and HOTA is their geometric mean, with higher values indicating higher performance in each category. Since this study focuses solely on humans, we use the HOTA score (HOTAα) when α is fixed at 0.5, and the value before integration with respect to α is used as the metric. This setting enables evaluation optimized for humans.

2.6 Evaluation results of person detection and tracking

The evaluation dataset consists of point cloud data and person detection and tracking annotation data, using scenes not used for training. The results were calculated separately for the entire area and for Area 1 (approximately 29 m × 28 m as in Figure 8), where all pillars could be measured by two or more LiDAR sensors, and Area 2, where the number of LiDAR sensors was limited due to construction constraints, resulting in measurements from only one side.

Table 2 shows the evaluation results. DetRe and DetPr are the recall and precision of the detection at each frame, and AssRe and AssPr are the recall and precision of the association (tracking), respectively. The results in both areas showed significant differences in the metrics. In the verification environment of the Nagoya Station concourse, which is a crowded area with many pillars and poor visibility, tracking interruptions frequently occur due to occlusion even when using a single LiDAR from one direction. However, by configuring two or more LiDAR sensors to reduce occlusion, we confirmed that association accuracy significantly improves. Although the accuracy cannot be directly compared with other case studies because we fixed α as 0.5, the HOTA value over 90% in an area of approximately 900 m² is considerably high for the purpose of crowd analysis.

In Area 2, the detection accuracy is relatively lower than the association accuracy. This also implies that the lack of detection due to the occlusion with single sensor is not avoidable. Meanwhile, the better value of association accuracy indicates that the interpolation method enables to keep tracking the pedestrians once they are detected.

The precisions for both detection and association are higher than the recalls. The reason for the higher precision rate is because there are no other moving obstacles than pedestrians in the study area. As static obstacles are removed in the process, the detected boundary boxes correspond to pedestrians. Meanwhile, the recalls may occur due to the occlusions.

Table 2. Quantitative evaluation results by area

Detection area	HOTA	DetA	AssA	DetRe	DetPr	AssRe	AssPr
Whole area	75.25%	70.75%	80.04%	79.04%	87.10%	83.31%	92.63%
Area 1	90.57%	90.70%	90.45%	95.12%	95.12%	92.09%	97.68%
Area 2	66.54%	63.79%	69.41%	72.69%	83.90%	74.38%	87.61%

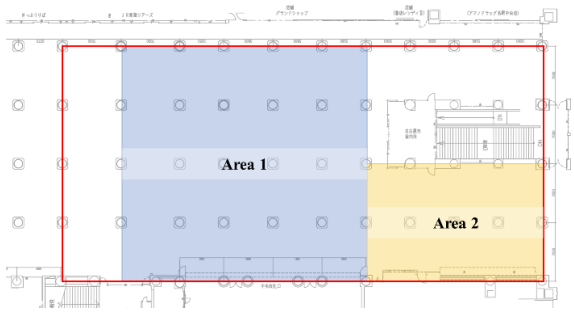

Figure 8. Detection area classification

As explained in Section 2.4, this method uses future trajectory information for interpolation. Table 3 shows a comparison of the accuracy with and without using the future information. In overall, the accuracy of each indicator significantly improved with the future trajectory information.

Table 3. Quantitative evaluation results with and without future trajectory data

Algorism	HOTA	DetA	AssA	DetRe	DetPr	Assre	AssPre
Without	45.44%	38.88%	53.10%	48.70%	65.87%	58.22%	80.56%
With	75.25%	70.75%	80.04%	79.04%	87.10%	87.10%	92.63%

3 Analysis of pedestrian flow

3.1 Analysis policy and method

Using the proposed system, we demonstrated basic analysis of pedestrian flow characteristics. The tracking results are generated at a frequency of approximately 10 Hz, depending on the processing status at the time of point cloud acquisition, and are output as JSON format data containing multiple entries in list format including the items in Table 4.

Table 4. Tracking output format

Contents	Details
Timestamp	Data source timestamp
position	Object position information [m]
size	Object size [m]
rotation	Rotation of each axis of an object [rad]
speed	Vector representing the speed of an object [m/s]
tracking_id	tracking ID
attributes	Object attribute information
score	Person confidence score (0-1) output by the person detection model

The concourse space was divided into 37 cells based on the pillar positions shown in Figure 9. Among the areas where high accuracy in person detection and tracking was confirmed in the previous chapter, density and walking speed were calculated for Cells 17, 21, 25, and 29 in front of the ticket gates, and Cells 18, 19, 20, and 21 crossing the concourse, and then compared and evaluated by time of day. The colors of the cells used for comparison in Figure 9 correspond to the line colors in the graphs presented in the next chapter.

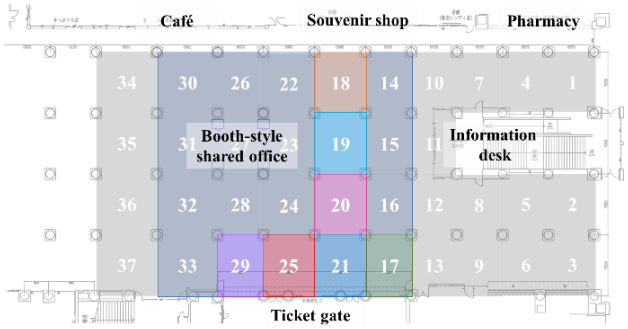

Figure 9. Analysis cell classification

3.2 Quantification of pedestrian flows by basic indicators

Figure 10 shows the pedestrian density ((1-minute moving average) transitions over a 15-minute period of March 1, 2025, 6:00 p.m. Pedestrian density at a frame is calculated as the number of pedestrians existing in the subject cell divided by the area of the cell. For the ticket gates (Cells 17, 21, 25, and 29), it can be seen that density increases and decreases repeatedly at certain intervals. This is thought to be due to trains arriving at the times indicated in red in the figure, causing a temporary increase in density on the ticket gate exit side. On the other hand, the area around the work booth (Cell 19) keeps showing low-density values, indicating that there was little traffic. It is known that it is difficult to form a traffic flow through this area due to the large obstacle of the information desk, even though the east-west flow is predominant. The measurement results also confirm this hypothesis. All results are generally as expected, and the characteristics of the concourse environment are reflected in the measurement results.

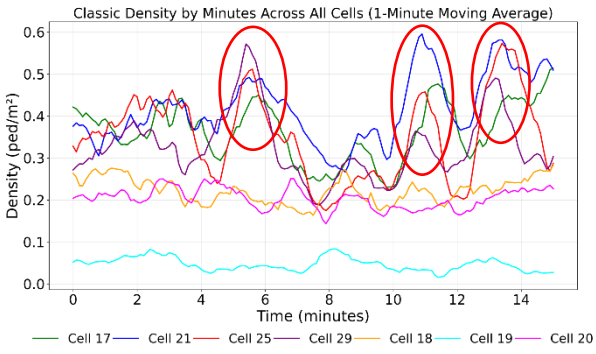

Figure 10. Pedestrian density trend (1-minute moving average)

Figure 11 shows the walking speed trends (moving average values per minute during the same time period. Individual walking speeds were calculated by dividing the difference between the position coordinates of each frame during tracking by the observation time interval, and then calculating the average speed of pedestrians within the cell. Overall, walking speeds were low, with large fluctuations in Cell 19, which had few people.

Upon reviewing the pedestrian flow conditions from the original measurement point clouds, it was observed that multiple stationary individuals were detected around pillars. As they were not excluded from the average walking speed calculations, the resulting values were underestimated.

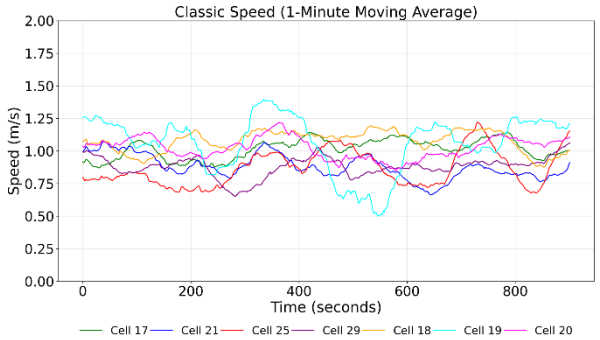

Figure 11. 1-minute average speed trends

Figure 12 shows the distribution of walking speeds in Cells 19 and 21. The speeds in Cell 19, which were unstable in Figure 10, were found to have a clear distinction between walking and stationary individuals. In the lower pedestrian density, some people may feel ease to stay there, and others are able to walk in free-flow speeds within this cell. Conversely, Cell 21, located near the ticket gate, exhibited speed reductions due to congestion, and it was difficult to distinguish between walking and stationary behavior due to temporary stops and other walking actions. These results are consistent with the characteristics of the environment we typically observe in this area.

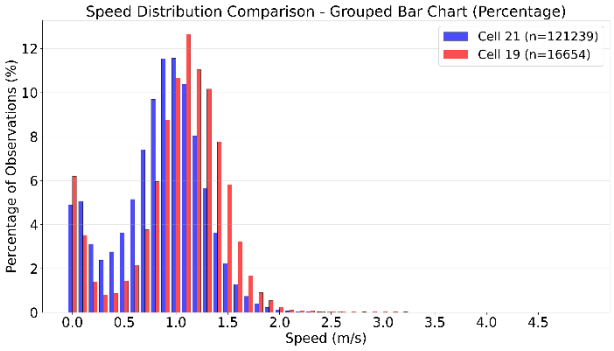

Figure 12. 1-minute average speed trends

3.3 Visualization of flow trends using heat maps

Next, we visualized the average density of the entire measurement area during a 15-minute period of March 1, 2025, 6:00 p.m. using a heat map (as in Figure 13). The grid units used in heat map generation are 0.1 m × 0.1 m. The heat map shows that the passageway near the ticket gates tends to be crowded and that people tend to linger in several places, such as in front of the information desk and around pillars.

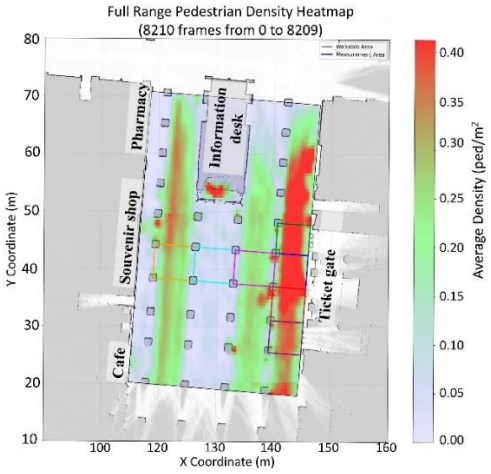

Figure 13. Density heat map (rotate 90° counterclockwise)

We also visualized the trends in walking directions during the same time period using a heat map, as shown in Figure 14. The graph shows that during this time period, people naturally tended to walk on the left side of each passageway between the pillars on the ticket gate side. On the other hand, on the store side, pedestrians heading east tended to walk in the center of the passageway, while those heading west tended to walk on both sides.

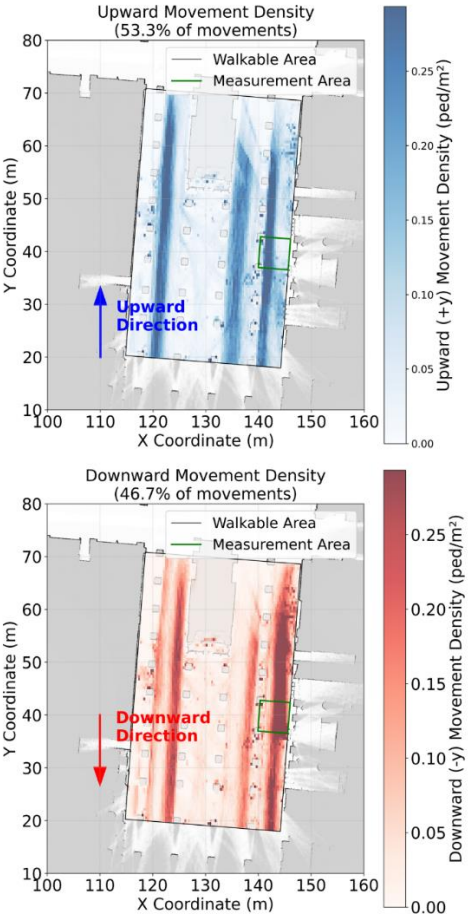

Figure 14. Walking direction heat map
(Top: Westward flow, Bottom: Eastward flow)

4 Discussions and conclusion

In this study, we constructed a wide-area, high-density pedestrian flow measurement and tracking system using twenty 3D-LiDAR sensors in a real environment to accurately grasp and quantify pedestrian flow in a crowded large-scale station concourse. By employing a person detection model using deep learning and tracking compensation processing using future walking trajectories, we confirmed that high-accuracy person detection and tracking with HOTA exceeding 90% is possible in an area of approximately 900 m².

The system successfully measured large-scale, accurate pedestrian trajectory data in real-world environments. Using basic indicators and heatmaps derived from the measurement data, we visualized flow trends and evaluated pedestrian flow characteristics in station concourse environments.

In order to further improve the accuracy of people flow measurement, there are areas where person detection errors are likely to occur, while there are also areas where the number of sensors is considered excessive. Therefore, there is room for optimizing the number of sensors and their installation angles. Additionally, in the construction of a person detection model using machine learning, the amount of data used for training was limited, so further annotation to increase the training data and parameter adjustments are still possible. As a result, significant improvements in accuracy are expected in the future. In this study, tracking was performed using a handcrafted method, but if machine learning can be used to infer walking trajectories from point clouds, it should be possible to address occlusion in crowded environments and complex behaviors. This approach has the potential to contribute not only to improving tracking accuracy but also to reducing costs by reducing the number of sensors and increasing freedom in sensor placement.

Regarding pedestrian flow analysis, this paper is limited to a comparison using representative indicators for a specific time period, but we plan to standardize station pedestrian flow characteristics by focusing on the movement trajectories of individual pedestrians and examining the flow characteristics when pedestrian flows from multiple directions intersect, using data from other days of the week, times of day, and abnormal situations such as disasters.

Additionally, the point cloud data collected by the system developed in this study is considered sufficient to identify certain pedestrian attributes (e.g., children, passengers with large luggage, wheelchair users, and stroller users). By incorporating attributes into the point cloud data, we aim to deepen pedestrian flow analysis and utilization, enabling more detailed identification of pedestrian flow trends in specific cases and verification of the effectiveness of measures.

As mentioned at the beginning, we also plan to analyze human-robot interactions using this system. With the additional annotation procedure for the robot detection, we will evaluate the impact of introducing the robots on crowds to develop operation methods of the autonomous robots.

5 Acknowledgments

This research was supported by Central Japan Railway Company.

6 Abbreviations

LiDAR: Light Detection and Ranging
SSD: Solid State Drive
FOV: Field of View
SLAM: Simultaneous Localization and Mapping
ICP: Iterative Closest Point
SDK: Software Development Kit
GPU: Graphics Processing Unit

7 Statements and Declarations

Declarations Funding: This research was supported by Central Japan Railway Company.

Competing Interests: The authors declare no conflicts of interest associated with this manuscript.

8 References

1. Fukuzaki, Y., Mochizuki, M., Murao, K. and Nishio, N. “A pedestrian flow analysis system using Wi-Fi packet sensors to a real environment”, UbiComp '14 Adjunct: Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication, 2014
2. Gawande, U., Hajari, K. and Golhar, Y., “Pedestrian Detection and Tracking in Video Surveillance System: Issues”, Comprehensive Review, and Challenges, Recent Trends in Computational Intelligence, IntechOpen Publisher, 2020
3. Bouazizi, M., Mora, A.L., Feghoul, K. and Ohtsuki, T., “Activity Detection in Indoor Environments Using Multiple 2D Lidars”, Sensors 2024, 24(2), 626
4. Na, K.I. and Park, B., “Real-time 3D multi-pedestrian detection and tracking using 3D LiDAR point cloud for mobile robot”, ETRI Journal Volume45 ,2023
5. Zhou, Y. and Tuzel, O., “VoxelNet: End-to-End Learning for Point Cloud Based 3D Object Detection”, IEEE Conference on CVPR, 2018
6. Lang, A. H., Vora, S., Caesar, H., Zhou, L. and Yang, J., “PointPillars: Fast Encoders for Object Detection from Point Clouds”, CVPR 2019
7. Gómez, J., Aycard, O. and Baber, J., “Efficient Detection and Tracking of Human Using 3D LiDAR Sensor”, Sensors, 2023
8. Zhang, Y., Hu, Q., Xu, G. Ma, Y., Wan, J., and Guo, Y., “Not All Points Are Equal: Learning Highly Efficient Point-based Detectors for 3D LiDAR Point Clouds”, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2022

9. JR TOKAI AGENCY CO.,LTD., “JR TOKAI MEDIA GUIDE 2025”
10. Ouyang, J., and Liu, X. and Chen, H., “Hierarchical Adaptive Voxel-guided Sampling for Real-time Applications in Large-scale Point Clouds”, arXiv preprint arXiv:2305.14306, 2023
11. Luiten, J., Osep, A., Dendorfer, P., Torr, P., Geiger, A., Taixe, L. L. and Leibe, B., “HOTA: A Higher Order Metric for Evaluating Multi-object Tracking”, International Journal of Computer Vision (2021) 129:548–578

Ryosuke Hishida received M.Eng. degree from Nagoya University in 2016. He is a research engineer of Central Japan Railway Company, focusing on robotics and research public space traffic in doctoral program for working professionals in Nagoya University.

Miho IRYO received Dr. Eng. from the University of Tokyo in 2007. She is an associate professor at the Graduate School of Environmental Studies, Nagoya University. Her major is traffic flow and safety analysis for road design and traffic operation.

M Rizki Alhamdani graduated with a Bachelor's degree in Civil Engineering from Universitas Islam Indonesia. Research interests focus on traffic flow and pedestrian flow.

Kazuki Mano received M.Eng. degree in Computer Science and Engineering from Toyohashi University of Technology. He is a researcher at Kurusugawa Computer Inc., focusing on computer vision, 3D point clouds and machine learning.

Youhei Yamaguchi received M.Eng. degree from Nagoya Institute of Technology in 2003. He is the Co-founder and Chief Technology Officer of Kurusugawa Computer Inc. His research interests include computational theory, machine learning, and computer vision