LOCALIZING AND VISUALIZING THE DEGREE OF PEOPLE CROWDING WITH AN OMNIDIRECTIONAL CAMERA BY DIFFERENT TIMES

Tomu Muraoka, Satoshi Kubota & Yoshihiro Yasumuro
Kansai University, Japan

ABSTRACT: *The Corona Disaster increased the demand for information on the degree of human crowding, as it was essential to balance avoiding restricting behavior and reducing the risk of crowding. Although there are many technologies for detecting people using monitoring cameras, the number of cameras installed in a wide area is costly, and coverage is limited. In this study, we propose a method to qualitatively visualize the distribution of people by using images captured by a moving omnidirectional camera from the viewpoint of facility management during regular security patrols. Omnidirectional images are used for both 3D modeling of the target space based on SfM (structure from motion) and person detection/tracking by machine learning. The distribution of people is visualized qualitatively by obtaining the positions of the extracted people on the 3D model of the site and mapping them. The parallel software processing of visitor observation and mapping is expected to be highly cost-effective in terms of implementation and operation. On the other hand, although there are time deviations in the mapping depending on the location, the visualization and the updated time show their usefulness in understanding the distribution of congestion.*

KEYWORDS: *COVID-19, people's congestion, omnidirectional camera, SfM (Structure from Motion), machine-learning*

1. INTRODUCTION

1.1 Research background

COVID-19 infection was moved to category five infectious disease in Japan on May 8, 2023. The wearing of masks has been left to the discretion of individuals and businesses, and the condition is coming to an end. During the coronavirus outbreak, infection control measures such as wearing masks, hand sanitizers, and refraining from going to places with a high risk of becoming infected became widespread at the individual level and effectively controlled other infectious diseases. However, with the relaxation of waterfront measures and the increase in the number of foreign tourists, there is a risk that other contagious diseases may be brought into Japan, and the number of older people at high risk of serious illness is expected to increase due to the aging of society. As shown in Fig. 1, the number of influenza cases reported from medical institutions (fixed points) nationwide in 2023 showed an increasing trend, with many weeks exceeding the average number of cases reported over the previous five years (National Institute of Infectious Diseases, 2023).

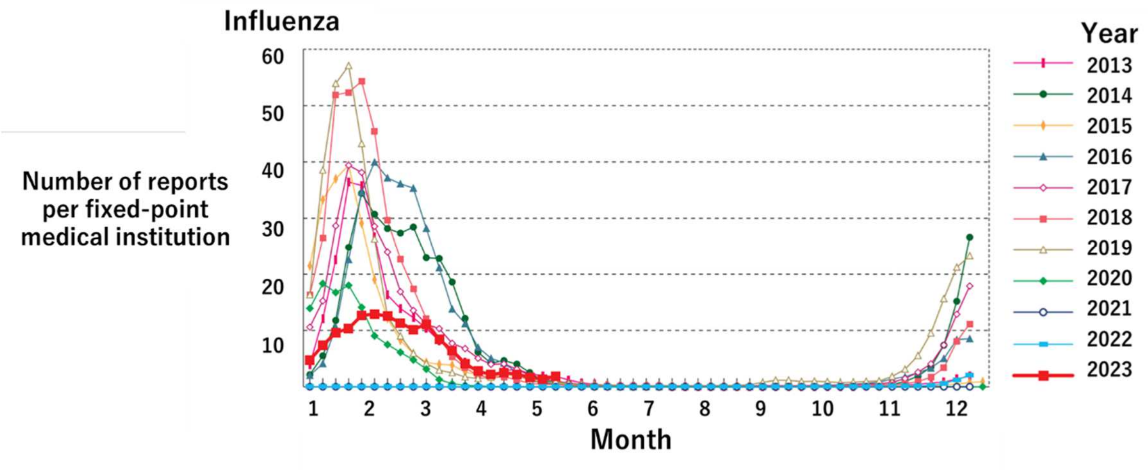

Fig. 1: Number of Influenza Infections
(National Institute of Infectious Diseases, 2023, partially modified and translated)

Referee List (DOI: 10.36253/fup_referee_list)
FUP Best Practice in Scholarly Publishing (DOI 10.36253/fup_best_practice)

Tomu Muraoka, Satoshi Kubota, Yoshihiro Yasumuro, *Localizing and Visualizing the Degree of People Crowding with an Omnidirectional Camera by Different Times,* pp. 657-668, © 2023 Author(s), CC BY NC 4.0, DOI 10.36253/979-12-215-0289-3.65

Therefore, in the after-coronas, it is necessary to continue infection control measures for individuals such as the older generations and people with underlying medical conditions who are at high risk of serious illness.

On the other hand, consumers are restricted in their purchasing behavior by spatial interference and competition among customers in a retail environment with high customer density within a store. As a result, it has been confirmed that consumers' purchasing decisions are negatively affected, resulting in lower satisfaction (Eroglu et al., 2005). Therefore, in terms of large-scale facility management, a system to ensure social distance is still essential in tourist attractions, commercial facilities, and other places where many people usually gather, and there is a high demand for being able to check the level of congestion in an environment where people tend to gather before visiting.

1.2 Previous work

An example of a familiar means for consumers to obtain local congestion information in advance is the "Congestion Radar" published on the web by Yahoo! (Yahoo! JAPAN,2023). Fig. 2 shows an example of the "Congestion Radar" display of congestion in the vicinity of Tokyo Station (Japan). The "Congestion Radar" visualizes the degree of congestion on a heat-map-like color-coded map of Yahoo! Japan by generating statistics of user location information based on the usage of applications provided by Yahoo! As shown in Fig. 3, EXPOCITY (Japan), a large-scale commercial facility, uses existing technology to visualize the traffic volume aggregated by sensors that detect intrusion and passage in a color-coded format, allowing users to view parking lot congestion on the official homepage. These are all abstract and have the disadvantage that it is difficult to confirm the state of each part of a commercial facility and the distribution of individual people, making it difficult to visualize the state of congestion. In addition, many technologies detect and visualize people using monitor cameras inside facilities (Hitachi, Ltd, 2020, for example). However, since it is necessary to install multiple cameras in a wide area to check and compare the field of view individually, the operation of existing monitor cameras is costly in terms of the number of cameras installed, and there are also limitations in terms of the stability and comprehensiveness of the images. As a study to visualize human distribution from monocular wide-field images, the authors have conducted 3D modeling of the target space based on SfM (Structure from Motion) and human detection and tracking processing by machine learning from images captured by a 360 deg camera and mapped them onto a 3D model of the site and demonstrated its effectiveness in principle (Muraoka et al., 2022). However, this method has drawbacks in systemization regarding scalability and continuous operation since the 3D data of each site is regenerated each time it is photographed to update the information on the distribution of people. In addition, because the people's distribution is visualized on the 3D model of the site, it was not easy to grab the surrounding location relationships and thus limits readability for users as a map.

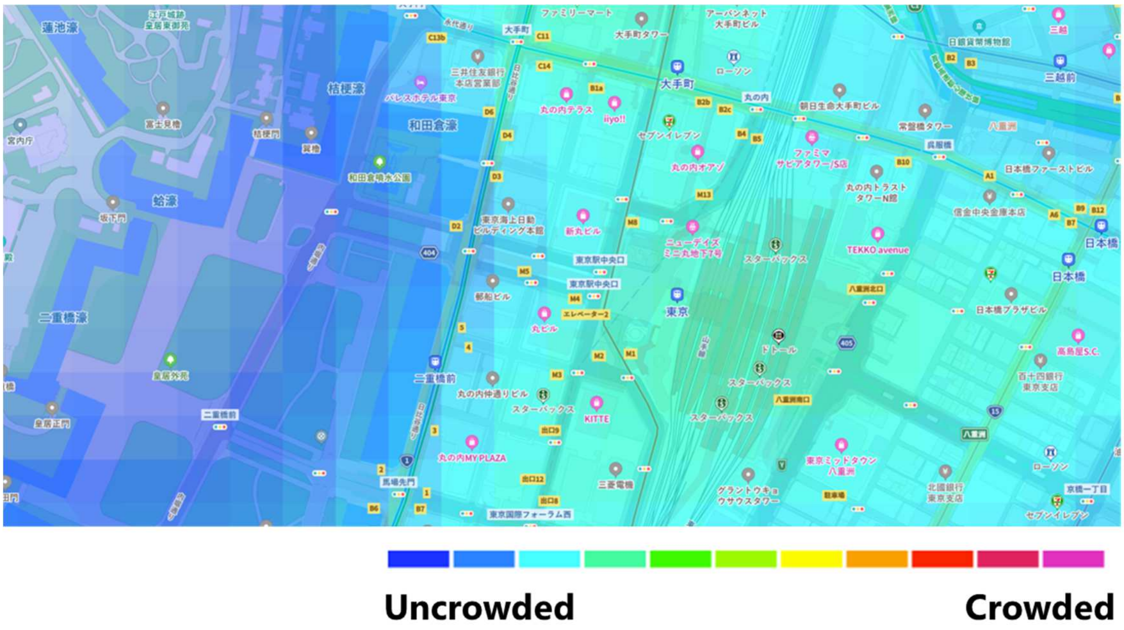

Fig. 2: Example of congestion radar on a map
(Yahoo! JAPAN , 2023, partially modified)

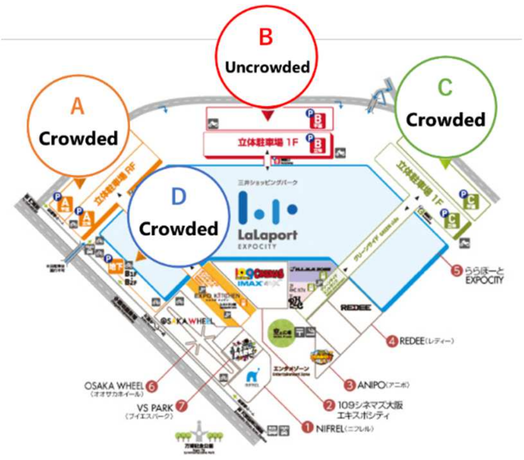

Fig. 3: Example showing parking congestion
(EXPOCITY,2023, partially modified)

2. PROPOSED METHOD

2.1 Outline of the proposed method

To solve the above problem, this study proposes a method of displaying and updating congestion information on a floor map, as shown in Fig. 4, by moving around while taking pictures with an omnidirectional camera carried by security guards or regular base patrol officers in large commercial premises where people move in and out relatively frequently. Fig. 5 shows the processing steps of the proposed method. First, an omnidirectional camera captures images of the surroundings while moving around the target site. Using the photos from the video frames as input for the SfM process, which performs 3D reconstruction of the target scene, and a 3D coordinate system is constructed. As shown in Fig. 6, the captured video is processed for large-scale sites by dividing it into multiple areas. SfM is performed for each area to generate a 3D reconstruction and coordinate system. The video is zenith-corrected so that the vertical axis direction of the image is aligned with the vertical direction of the SfM model. Next, a coordinate transformation equation from the SfM model coordinates to the image coordinates of the floor map image of the facility is calculated by solving a point set matching problem, and the positional relationship between the SfM model and floor map for each patrol position is mapped as shown in Fig. 7. The above process is the preliminary processing performed only at the beginning.

In the person placement process, the same zenith correction processing as above is performed on video frames taken by security guards and others during their patrol duties, and the video data is stored in the cloud service. The video frames are used as input for person detection and tracking. Fig. 8 shows a person detection and tracking process for a video image V taken at a particular location. A specific detector detects multiple targets to be tracked in each video frame, and the same ID is assigned to the same target tracked from frame to frame. In a group of images obtained at regular intervals from video frames (Frame t in Fig. 8), the information on the same person in the frames is integrated to determine whether the object detected as a person in each image is new. If the person is newly detected, the image in which the person is detected is added to the set of input images for SfM. When SfM is executed again, the coordinate system that has already been generated is maintained, and the information of the added image is processed incrementally to calculate the coordinates of the shooting position of the new image. Then, the coordinates of the feet of the new person in the image are obtained, and the person's position in the SfM model is calculated based on the relationship with the camera coordinates. In this way, the distribution of persons can be additionally and successively updated in a consistent coordinate system.

The position of a person on the floor map is calculated by transforming the coordinates from the previously calculated SfM coordinates to the floor map image, and a symbol of the person's size is placed. The symbols on the floor map are then deleted each time data captured by the omnidirectional camera is input, and the map displaying the distribution of people is updated at each time based on the input timestamps. Even over a wide area, the map can be updated piecemeal, corresponding to the location relation on the floor map. SfM is performed on the video frames captured in each area. In this way, the system patrols and displays the positions of persons on the floor map, providing the user visualization of congestion in advance.

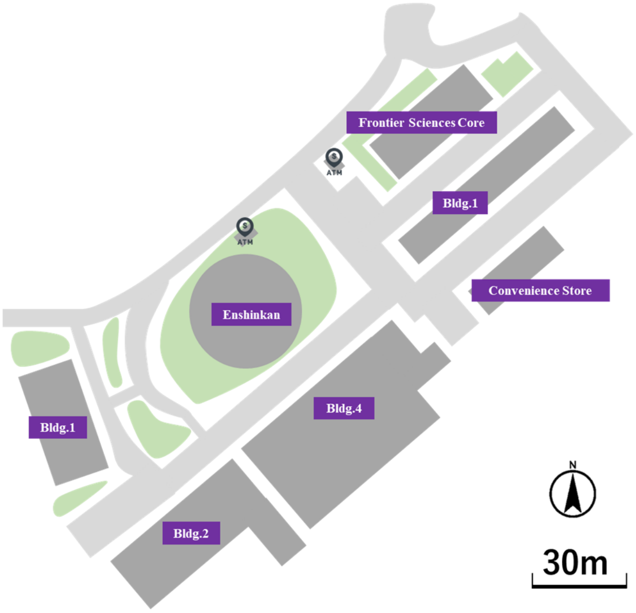

Fig. 4: An example of a floor map (Kansai University in japan)

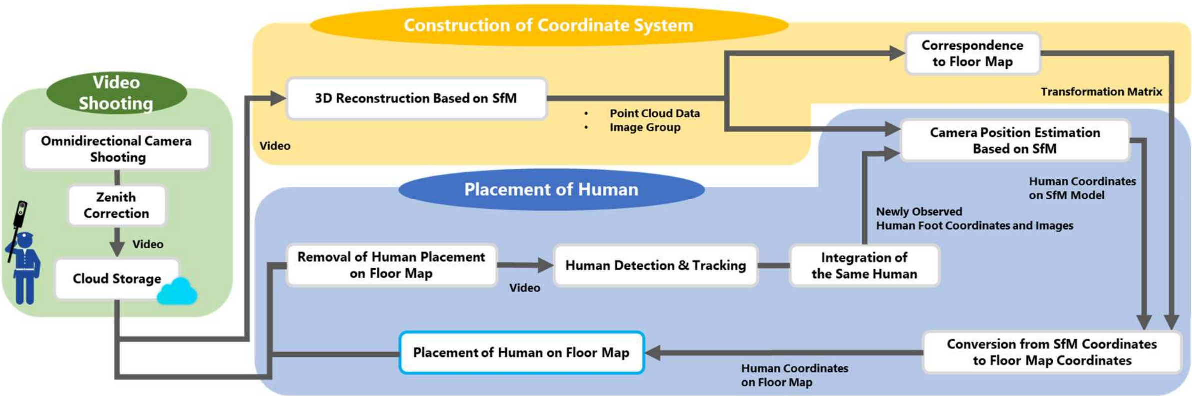
Construction of Coordinate System
Video Shooting
Omnidirectional Camera Shooting
Zenith Correction
Video
Cloud Storage
Video
3D Reconstruction Based on SfM
• Point Cloud Data
• Image Group
Correspondence to Floor Map
Transformation Matrix
Camera Position Estimation Based on SfM
Human Coordinates on SfM Model
Placement of Human
Newly Observed Human Foot Coordinates and Images
Removal of Human Placement on Floor Map
Human Detection & Tracking
Video
Integration of the Same Human
Placement of Human on Floor Map
Human Coordinates on Floor Map
Conversion from SfM Coordinates to Floor Map Coordinates

Fig. 5: Processing procedures by the proposed method

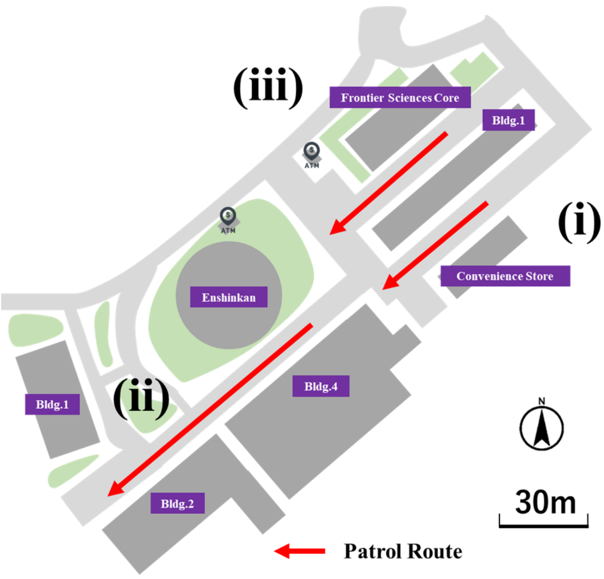

Fig. 6: Example patrol routes at the target site

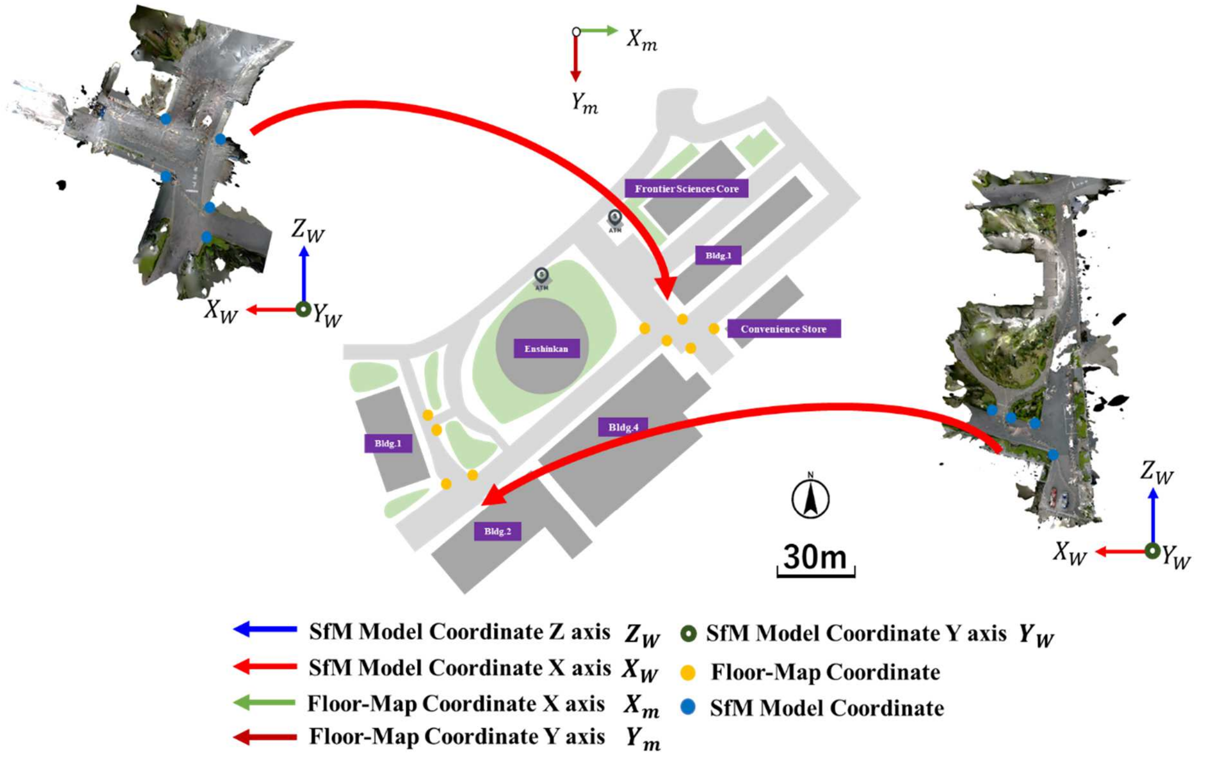
X_m
Y_m
Z_W
X_W
Y_W
Frontier Sciences Core
Bldg.1
Convenience Store
Enshinkan
Bldg.4
Bldg.1
Bldg.2
30m
Z_W
X_W
Y_W
SfM Model Coordinate Z axis Z_W
SfM Model Coordinate X axis X_W
Floor-Map Coordinate X axis X_m
Floor-Map Coordinate Y axis Y_m
SfM Model Coordinate Y axis Y_W
Floor-Map Coordinate
SfM Model Coordinate

Fig. 7: Mapping examples of SfM model coordinates to the floor map coordinate

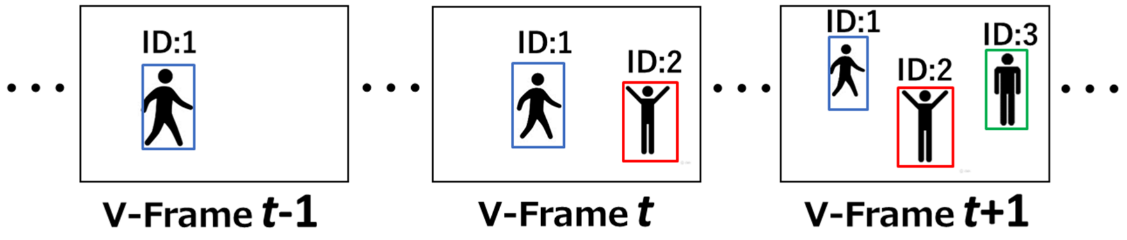

Fig. 8: Human detection and tracking through consecutive frames

2.2 Identification of the human's position

In identifying the standing position of a human on the floor map, the first step is to obtain the pixel positions (a, b) of the feet of the person in the omnidirectional image with width W px and height H px, as shown in Fig. 9. Since the height direction of the omnidirectional image corresponds to the Y_c axis of the camera coordinates through the zenith correction process, as shown in Fig. 10, let θ_1 be the angle between the vertically downward direction of the omnidirectional camera and the line-of-sight direction through the human's feet, and θ_2 be the azimuth angle of the person's feet based on the Z axis of the camera coordinates, $\theta_1 = \pi a/H$ and $\theta_2 = 2\pi b/W$ and The following is a calculation of the azimuth angle between the ground and the origin of the camera coordinates. Next, the height h from the ground to the origin of the camera coordinates is the difference between the Y_w coordinates of each camera and the Y_w coordinates of the point cloud of the ground detected by plane estimation using RANSAC (M. A. Fischler et al.,1981) on the point cloud data obtained by the 3D reconstruction process of the target site. In the X_c-Z_c plane of the camera coordinate system, if the distance from the origin to the person is d, the position of the human (X, Z) projected onto the X_c-Z_c plane is obtained by the following equation.

$$d = h \tan \theta_1 \qquad (1)$$

$$\mathrm{X} = d \sin \theta_2 \qquad (2)$$

$$\mathrm{Z} = d \cos \theta_2 \qquad (3)$$

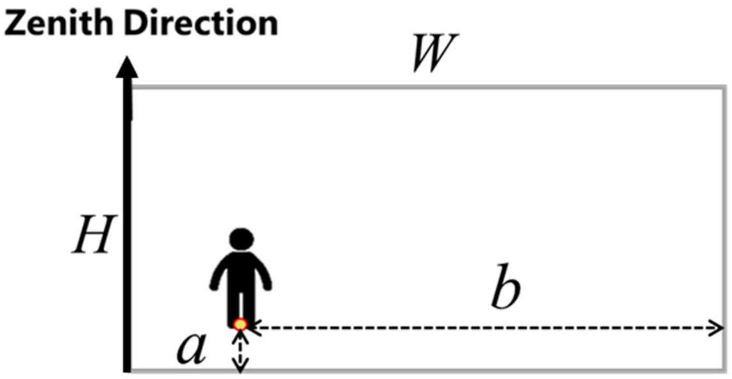

Fig. 9: Position of a person in an omnidirectional image

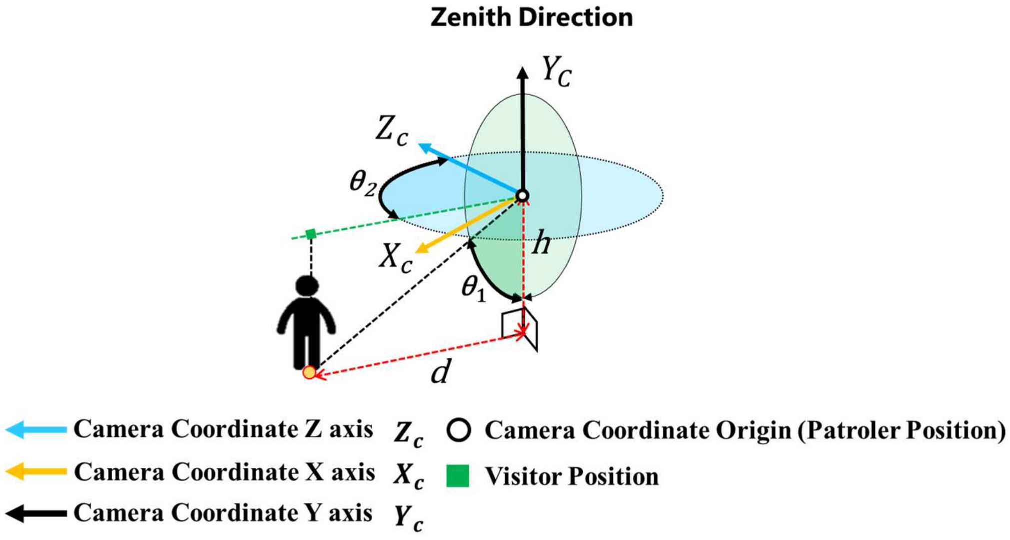
Zenith Direction
Y_C
Z_C
θ_2
X_C
h
θ_1
d
Camera Coordinate Z axis Z_c
Camera Coordinate X axis X_c
Camera Coordinate Y axis Y_c
○ Camera Coordinate Origin (Patroler Position)
🟩 Visitor Position

Fig. 10: Position of the camera and the person

Fig. 11 shows the relationship between the SfM model's coordinate system and each camera's coordinate system for each video record. The position of the camera coordinate system is the coordinate system calculated by SfM, which is also the position of the patrolman. Since the Y_c axis of the camera coordinate system and the Y_w axis of the SfM model correspond to the zenith direction, the positions of the persons detected at each shooting point can be integrated into the same coordinate system by rotation and translation on the X_w - Z_w plane. Let $R(\varphi)$ be the rotation and (t_x, t_z) be the camera position relative to the world coordinate; the position of a person (X_w , Z_w) in the 3D model can be calculated from the local camera coordinate (X ,Z) by equation (4).

$$\begin{bmatrix} X_w \\ Z_w \end{bmatrix} = R(\varphi) \begin{bmatrix} X \\ Z \end{bmatrix} + \begin{bmatrix} t_x \\ t_z \end{bmatrix} \qquad (4)$$

Next, the coordinate system X_m-Y_m plane of the floor map image of the facility and the X_w-Z_w plane of the SfM model coordinate system can be transformed into the position of the human on the SfM model by rotation φ_w around the Y axis in the SfM model coordinate system and translation shift, as before. Let S be a scaling matrix, R(φ_w) be a rotation matrix, and the origin coordinates (m_x, m_y) of the SfM model coordinate system in the X_m-Z_m plane and the human's position (X_m, Y_m) can be calculated as follows.

$$\begin{bmatrix}X_m\\Y_m\end{bmatrix} = S\cdot R(\varphi_w)\begin{bmatrix}X_w\\Z_{w'}\end{bmatrix}+\begin{bmatrix}m_x\\m_y\end{bmatrix} \qquad (5)$$

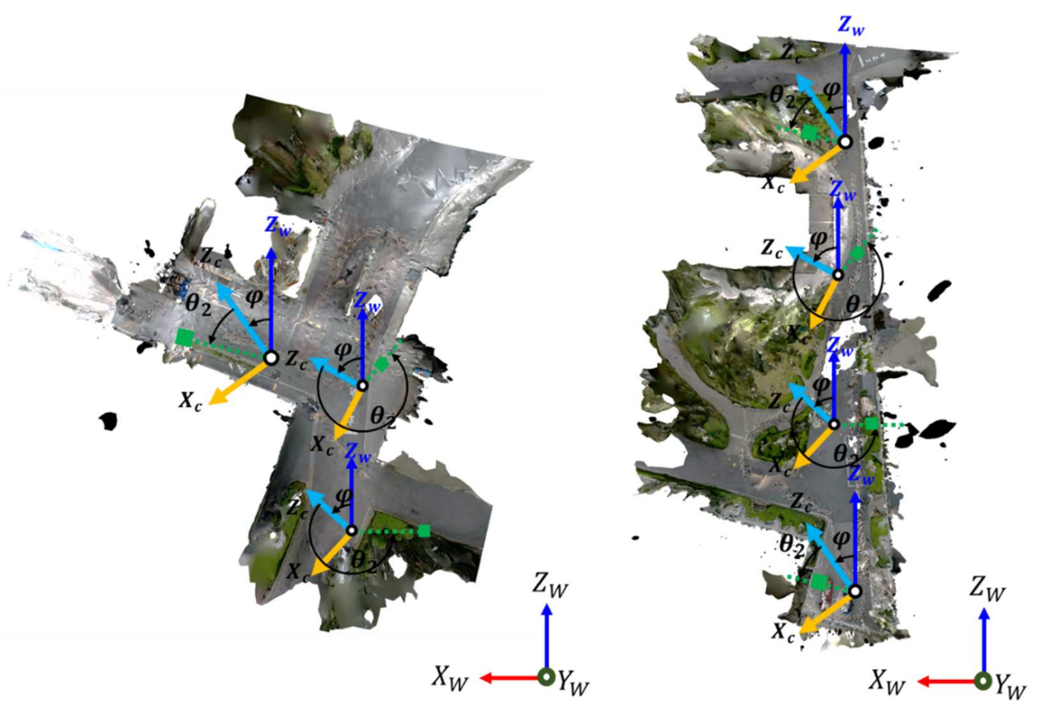

Fig. 11: Relationship between the SfM model coordinate X_w-Y_w-Z_w and the coordinates of the moving camera

3. EXPERIMENT

3.1 Equipment

For the experimental environment, we used a location in front of a convenience store on the Senri-yama campus of Kansai University (area (i) in Fig. 6), assuming a commercial facility crowded with many people. Theta X (RICHO) was used for the omnidirectional camera, with a video resolution of 5760 × 2880, Metashape Professional (Agisoft) for SfM, YOLO-X (Z. Ge et al.,2021) was used for the human detection model, and motpy (motpy - simple multi-object tracking library, 2022) was used for human tracking, as it is easy to combine various object detection models and OpenCV, an image processing library, was used to display the location of the person on the floor map of the target site.

3.2 Construction of coordinate system

As shown in Fig. 12, a 3D reconstruction based on SfM was performed using 91 images taken at the site, and a coordinate system was constructed. The average error was approximately 0.05 m. The number of tie points was approximately 4.7. The number of tie points was about 47,000, and the RMSE (root mean square error) of the re-projection of the feature points estimated from the image set onto the original image was about 3.0 pixels, indicating that the 3D shape is generally accurate. In the mapping between the coordinate system of the SfM model and that of the floor map, first, we visually picked up the corners of buildings and roads in each coordinate system to prepare five pairs of coordinates of (X_w , Z_w) and (X_m, Y_m). Next, the distance from the center of gravity of the five points to each of the five points in each coordinate was calculated, and the average value was calculated. There is a method to calculate the rotation matrix and translation vector in the point set matching problem by SVD (singular value decomposition) (K.S.Arum et al.,1987). In this study, $R(\varphi_w)$ and (m_x, m_y) are calculated by SVD. 5 points on the SfM model are transformed to positions on the floor map by equation (5). The results displayed on the floor map are shown in Fig. 13. The exchanged positions of the five points on the floor map generally corresponded to those of the five points on the SfM model.

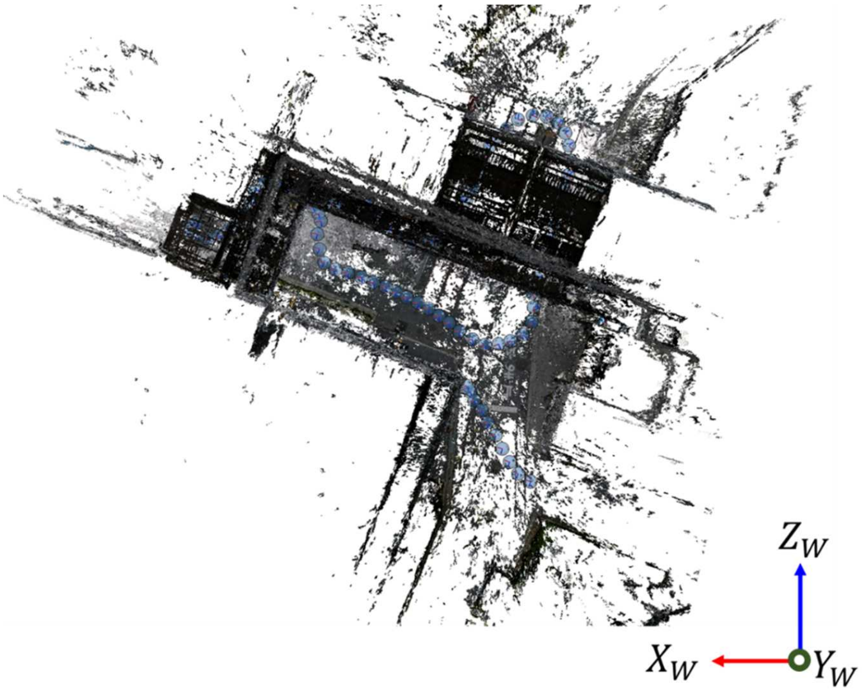

Fig. 12: 3D model and camera shooting position: The blue sphere indicates the shooting position.

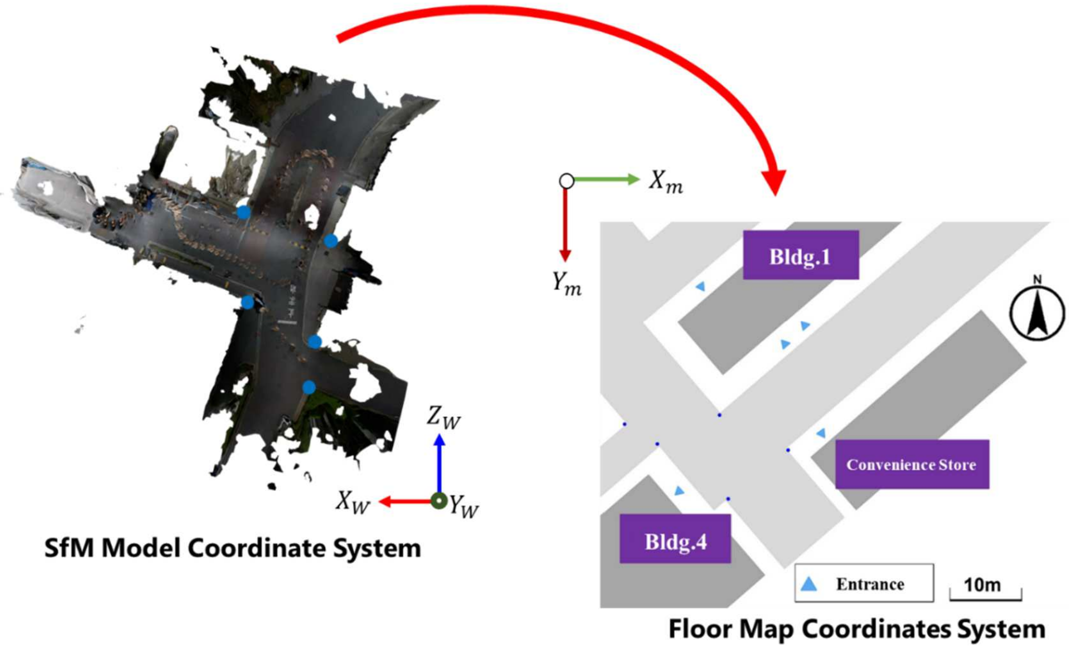

Fig. 13: Correspondence between SfM coordinate system and floor map coordinate system：
blue dots indicate 5 points in each coordinate system

3.3 Integration and display of person location information

YOLO-X was used for performing human detection for obtained image set from each section's video, as shown in Fig. 14. If the detected person has been tracked from a previous image frame by motpy using the video as input, the detected human is integrated into the existing person's information. If the human was not tracked and was newly observed, the observed image was used to estimate the camera position in the coordinate system of the SfM model, and the camera coordinates were calculated. The coordinates of the human's feet and the coordinates of the floor map were used to calculate the position coordinates of the human using Equations (1)-(5) and displayed on the floor map (Fig. 15, 16, and 17). The positions of the people mapped on the floor generally corresponded to the original images, and their positions about the buildings were also confirmed. However, because YOLO-X can detect people even in the case of body parts, and because the images are taken while moving, there are many cases where people in occlusion due to occlusion can be observed and mapped in other locations, making it possible to visualize the distribution of the people quantitatively. In addition, we were able to confirm changes in the degree of crowding at different times, such as around noon (Fig. 15), when the area is crowded at lunchtime, there is extreme crowding near the store entrance, many people move during class breaks (Fig. 16), and the area is empty during class (Figure17). Area (ii) in Fig. 5 was also patrolled around noon on the same day. The distribution of people observed was displayed on the floor map in the same manner as above (Fig. 18). As a result, it was possible to visualize the occurrence and resolution of queues and crowding of people near building entrances and the vicinity of stores, depending on the time of day. In a wide area, it was confirmed that information on the distribution of people could be updated for different areas in parallel by patrolling the area with several people. These functions are practical for visitors and others to know the trend of human distribution.

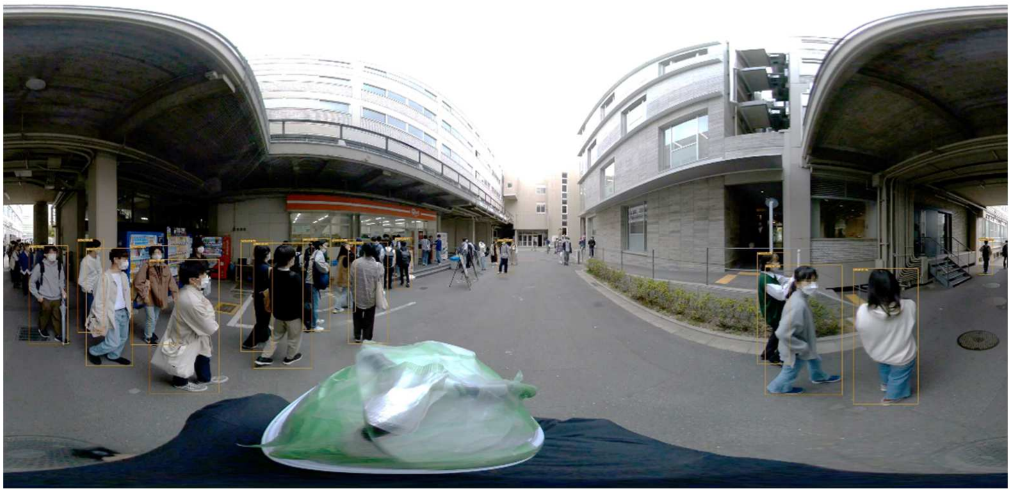

Fig. 14: Human detection using YOLOX for an omnidirectional image

4. CONCLUSION

In this study, we proposed a method for mapping and updating the distribution of people on a floor map in a fragmented manner, even over a wide area, simply by walking around the site and taking pictures with an omnidirectional camera and confirmed the method's effectiveness through experiments. The proposed method cannot perform synchronized observations because both the observer and the visitor are moving. However, it requires far fewer cameras and is equivalent to observing from many viewpoints because of the moving point observation. Therefore, parallel software processing of visitor observation and mapping will be highly cost-effective in terms of implementation and operation. Although the visitor mapping has time deviations depending on the location of the observation, visualization of the updated time together with the map will help understand the distribution of congestion. In future work, the authors plan to investigate a method of real-time mapping by online processing using the live streaming function of the camera and to construct a system that allows users to view maps showing the distribution of people on the Web.

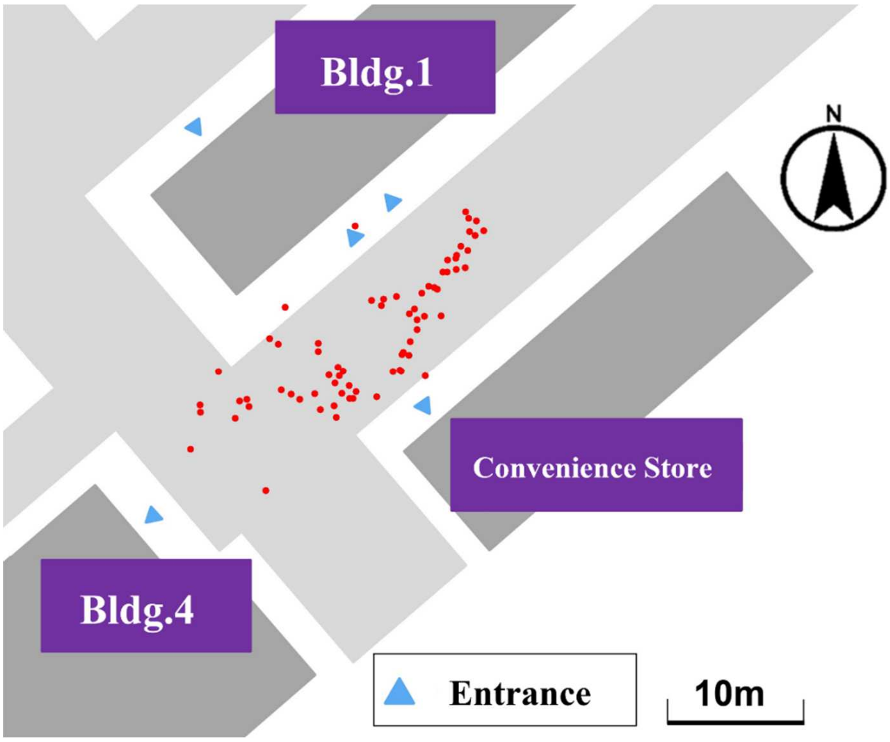

Fig. 15: Placement of people on the area (i) map: 12:00 PM: Red dots indicate the visitor's location.

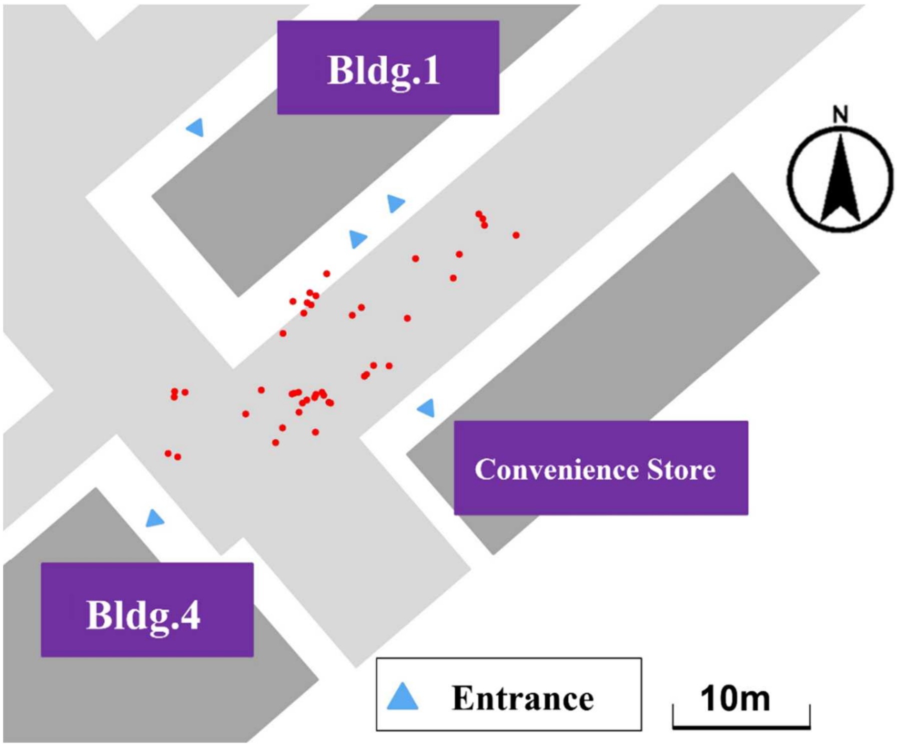

Fig. 16: Placement of people on the area (i) map: 2:30 PM: Red dots indicate the visitor's location.

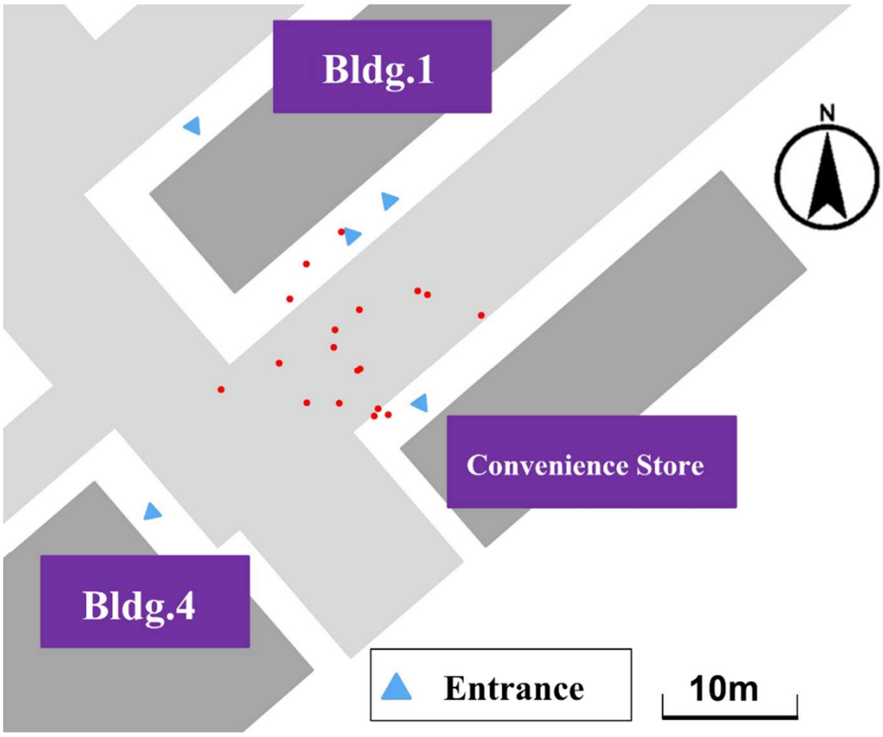

Fig. 17: Placement of people on the area (i) map: 4:00 PM:
Red dots indicate the visitor's location.

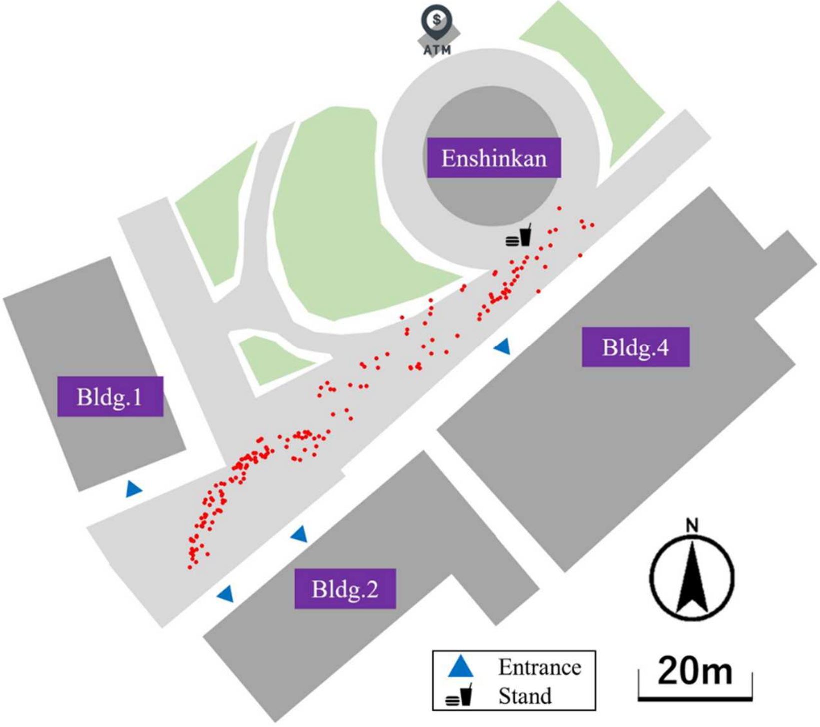

Fig. 18: Placement of people on the area (ii) map: 12:00 PM :
Red dots indicate the visitor's location.

REFERENCES

National Institute of Infectious Diseases (2023), Ministry of Health, Labor and Welfare, iDWR Infectious Diseases Weekly Report, <https://www.niid.go.jp/niid/images/idsc/idwr/IDWR2023/idwr2023-20.pdf>, (Viewed 2023. 6.5).

Yahoo! JAPAN (2023), Yahoo! JAPAN Map, Congestion Radar, <https://map.yahoo.co.jp/congestion>, (Viewed 2023.8.1).

EXPOCITY (2023), Current parking lot congestion, < https://www.expocity-mf.com/expo/parking/>, (Viewed 2023.9.21)

Hitachi, Ltd (2020), Hitachi technology demonstration at Tokyo Dome for infection-prevention measures: Visualization of congestion inside the stadium, <https://social-innovation.hitachi/en/topics/tokyo-dome/> (Viewed 2023.8.3).

T. Muraoka., S. Kubota, and Y. Yasumuro, (2022) Localizing and Mapping of People's Distribution with an Omnidirectional Camera, *Proceedings of the 22nd International Conference on Construction Application of Virtual Reality (CONVR2022)*, pp. 134-142.

S. A. Eroglu, K. Machleit, and T. F. Barr (2005), Perceived Retail Crowding and Shopping Satisfaction: The Role of Shopping Values, *Journal of Business Research*, Vol 58 (8), pp. 1146-1153.

M. A. Fischler, R. C. Bolles (1981). Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Communications of the ACM*, Vol 24, pp 381-395.

Z. Ge, S. Liu, F. Wang, Z. Li, J. Sun (2021), YOLO X: Exceeding YOLO Series in 2021, *The Conference on Computer Vision and Pattern Recognition (CVPR2021).*

motpy - simple multi object tracking library, <https://github.com/wmuron/motpy> (Viewed 2022.10.26)

K. S. Arum, T. Shuang and S. D. Blostein (1987), Least-Squares Fitting of Two 3-D Point Sets, *IEEE Trans. On Pattern Analysis and Machine Intelligence*, Vol.9, No.5, pp. 698-700.