Safety Visualization in Cooperative Operation with Workers and Construction Vehicles Using Temporal LiDAR Data

Kenta Ishizuka*[1], Masafumi Nakagawa[1], Masahiro Miyo[2]

[1]Department of Civil Engineering, Shibaura Institute of Technology

3-7-5, Toyosu, Koto-ku, Tokyo, 135-8548, Japan

Email: ah21014@shibaura-it.ac.jp

[2]Watanabe Engineering Co., Ltd.

7-4-69, Noda-machi, Fukushima City, Fukushima, 960-8055, Japan

***Abstract*:**

In Japan, the declining birthrate and aging population have led to a severe labor shortage. Therefore, the construction industry is trying to improve efficiency using autonomous and ICT construction vehicles with construction information modeling and management. Although an unmanned operation is an ideal scenario, construction works in urban areas have technical issues in poor GNSS environments because of multipath problems, in managing autonomous and ICT construction vehicles. Moreover, construction areas are too small and narrow for autonomous construction vehicles in urban areas. Thus, cooperative operation with workers and ICT-assisted construction vehicles is required for construction work in urban areas. Therefore, the efficiency and safety of cooperative work between construction vehicles and workers must improve. In our previous research, we focused on laser scanning from a construction vehicle to monitor the safety of construction fields. Although we integrated horizontal LiDAR and vertical LiDAR to recognize workers, the vertical field of view was insufficient to recognize workers. In this study, we integrated horizontal LiDAR, vertical LiDAR, and oblique LiDAR to create a wider field of view from a construction vehicle. First, the three types of LiDARs are integrated to obtain composite LiDAR point clouds. Next, the bucket of the construction vehicle and workers are detected from LiDAR point clouds. Then, worker tracking was applied with range image processing and simultaneous localization and mapping processing. Estimated bucket and worker tracking data were used for a safety assessment in the construction field. Through our experiments, we confirmed that our methodology can recognize objects such as a bucket and workers. Moreover, our methodology can visualize safe and dangerous situations using temporal composite LiDAR point clouds. However, technical issues remained such as several types of failures in worker recognition processing. In addition, in safety visualization, technical issues in real-time processing remained such as sudden stop estimation.

Keywords: Point clouds, Laser scanning, Object recognition, Object tracking, Construction vehicle

Introduction

The current construction industry in Japan faces several challenges, including accidents during construction, a declining and aging workforce, and overall project efficiency. To solve these problems, the Japanese construction industry is developing and introducing advanced construction vehicles equipped with ICT technology based on building information modelling (BIM). In addition, construction experiments using remotely operated unmanned construction vehicles are being conducted in large construction spaces. Remote-operated unmanned construction replaces backhoe work in space by improving the working environment with a joystick-type remote control and a monitor covering a wide viewing angle (Kajita et al., 2017). Taisei Corporation tested the camera system at Unzen Fugendake (Kondo et al., 2011). The operator's viewpoint is shared with a mobile remote control room using an onboard camera attached to the driver's seat of the construction equipment. This system assists an operator to evacuate quickly to a safe location in an emergency event, such as a landslide. Kajima Construction has also been conducting demonstration tests in the unmanned construction space for construction vehicles using ICT technology since 2005. In 2018, full-scale unmanned construction was carried out at a dam construction site. The construction vehicles were not operated remotely, but instructions from the control room were sent to several construction vehicles to carry out unmanned construction (Kajima Construction, 2020). Various ICT construction vehicles have also been developed in the past, but several technical issues remain in terms of cooperation with workers. First, the small size of the construction space in urban areas such as Tokyo makes it difficult to apply existing ICT construction equipment. Therefore, construction vehicles are required to collaborate with workers to carry out detailed tasks such as excavation and digging works. Second, GNSS signal receiving is difficult in urban areas such as Tokyo because of multipath effects. Therefore, it is necessary to consider how to operate ICT technology in manned rather than unmanned construction. Previous studies have attempted to visualize safety in the construction space by object recognition and tracking from horizontal scanning data using multilayer LiDAR. However, horizontal scanning has difficulty in detecting and omitting objects because of insufficient scanning resolution and vertical scanning angles, making it difficult to understand the behavior of the bucket. Therefore, in our previous work, we have integrated horizontal and vertical LiDARs mounted on a construction vehicle to recognize objects from point clouds. However, this system was still insufficient to understand the behavior of construction vehicle's buckets. Therefore, in this study, in addition to horizontal and vertical LiDARs, we add oblique LiDAR to spot-observe the working area of the bucket. The objectives of this study are to construct a measurement system combining horizontal, vertical, and oblique LiDARs, to devise a method for integrating point clouds acquired by each LiDAR, to develop an object recognition and tracking method using the integrated results, and to propose a method for evaluating risk based on human tracking data and bucket behavior data.

Methodology

The proposed methodology consists of LiDAR time synchronization, LiDAR point cloud integration, ground surface estimation, backhoe bucket recognition, worker recognition and tracking, and risk estimation, as shown in Figure 1.

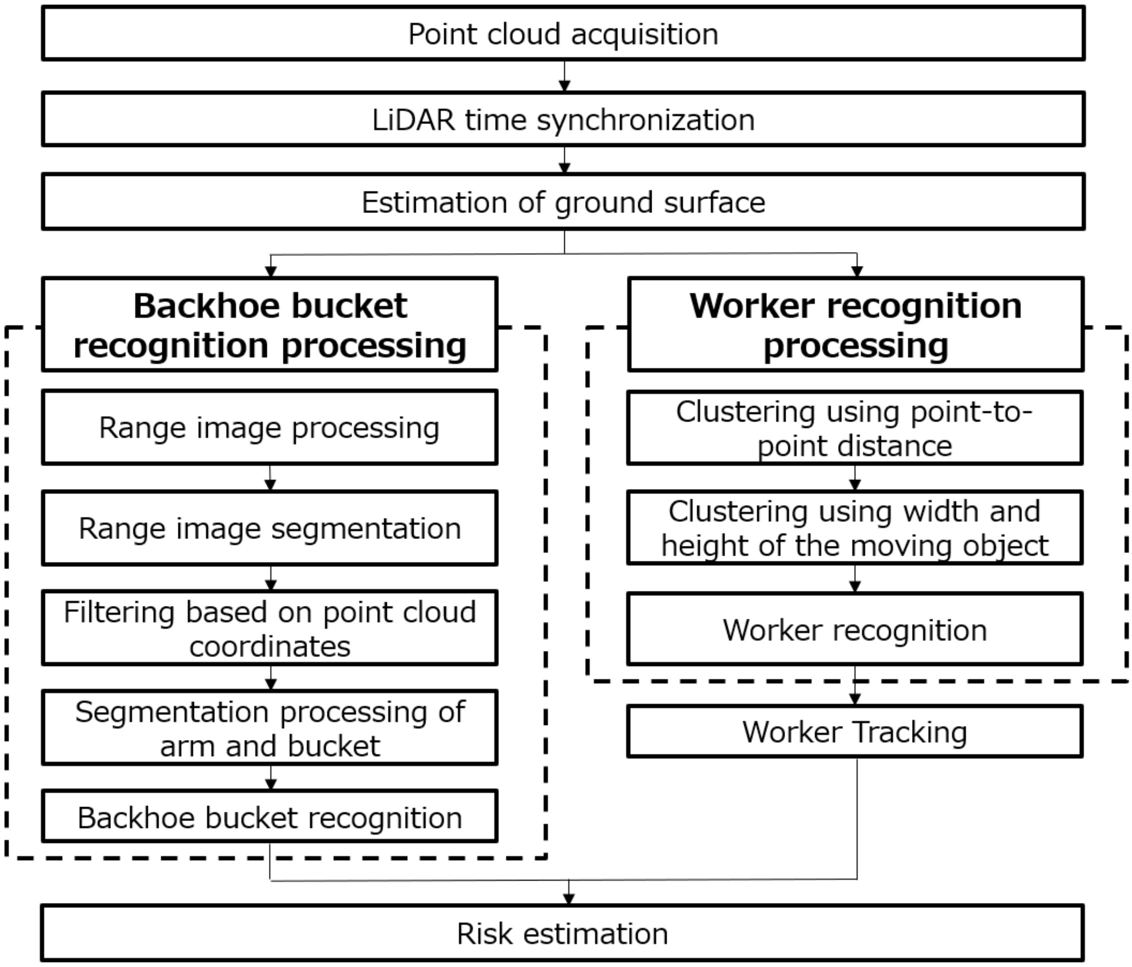


Figure 1: Proposed methodology.

The object recognition processing consists of the following preprocessing steps. First, the horizontal, vertical, and oblique LiDARs are integrated after the time synchronization processing. Next, the ground surfaces are estimated with the RANSAC algorithm. After these steps, labeling is applied to each object on the range image generated from the acquired point clouds. Next, the construction equipment arm and bucket are segmented to extract from point clouds of extracted construction vehicle parts. In parallel, the position and behavior of the operator in the construction space are estimated by a motion (worker) recognition process consisting of clustering point clouds based on Euclidean distance and a clustering process using the width and height of the moving object. By applying the tracking process, the same ID is assigned to the same worker in each scene.

a. Point Cloud Integration:

Integrate the coordinate space (space–time) of point clouds acquired by horizontal, vertical, and oblique LiDARs, as shown in Figure 2. First, the LiDAR system is synchronized using the GPS time recorded during point cloud acquisition. Second, a rotation correction of the point cloud acquired by the vertical LiDAR is performed, as the vertical LiDAR is mounted at a 90° tilt for vertical scanning. Finally, as each LiDAR is located at a small distance from the other, they are integrated using the offset values between the LiDARs.

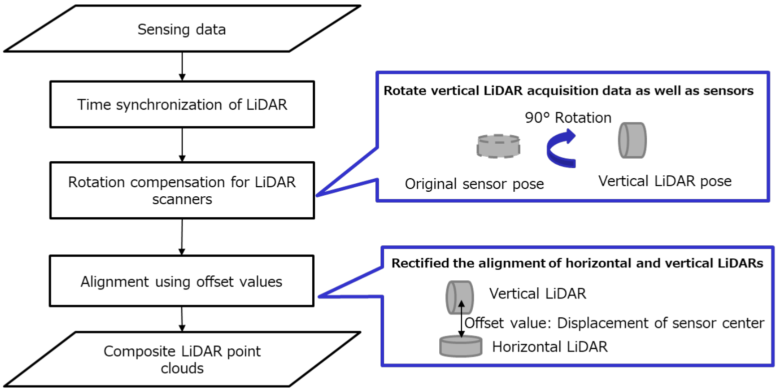


Figure 2: Overview of the Point Cloud Integration.

b. Range Image Processing:

We focus on image-based labeling to improve the performance of point cloud labeling for bucket recognition. First, the range images are generated using LiDAR channel information in the line direction and scanning direction in the column direction as shown in Figure 3. The information in the range image contains the point cloud coordinates, reflection intensity values, and point cloud processing results, which are managed as layers (multiple images). Range image processing consists of distance image generation using the ranging values of each point, distance edge image generation using the difference in ranging values between neighboring pixels, region segmentation image generation of the distance image using the distance edge image, and labeled image generation of the region segmentation results.

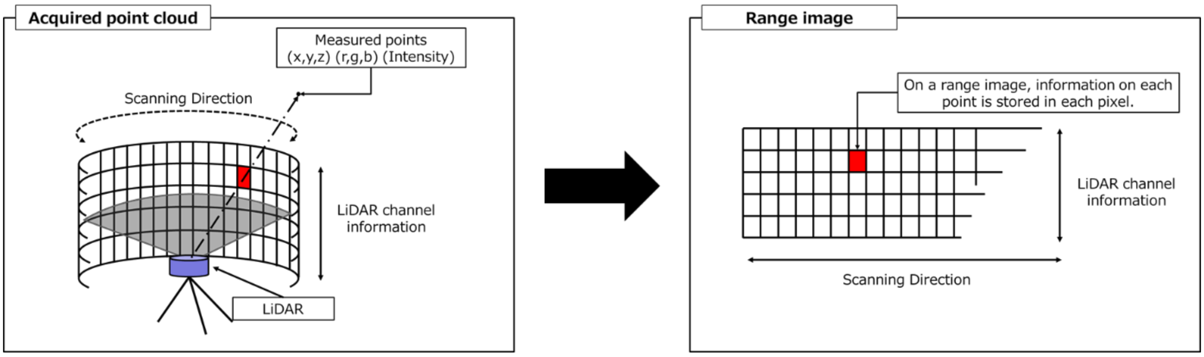


Figure 3: Overview of the range image processing.

c. Labeling of Point Cloud:

On the label image generated by range image processing, labels are also assigned to the point clouds in conjunction with the distance image. However, labels relating to geographic objects are not assigned. Therefore, labels relating to geo-objects are assigned using the relative positional relationship with the LiDAR on the construction vehicle. Here, the objects are classified into three categories: objects under the ground surfaces (e.g., buried pipes and steel sheet piles), objects in front of the construction vehicle (e.g., buckets and arms of the construction vehicle), and other objects (e.g., workers, dump trucks, and buildings). Figure 4 shows a conceptual image of labeling regarding geo-objects.

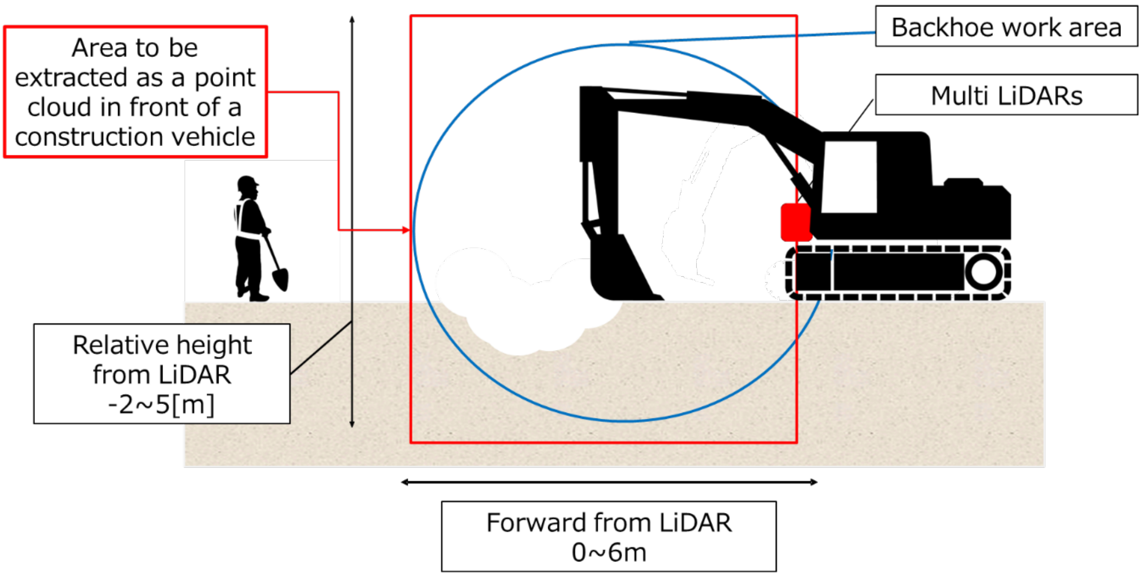


Figure 4: Overview of the labeling of the point cloud.

d. Bucket Recognition Processing

Point clouds of the arm and bucket are connected as a feature. Thus, in the backhoe bucket, extraction from point clouds, the close point clouds in front of the construction vehicle are divided into the arm and bucket. The arm and bucket can be extracted from point clouds based on 3-D object recognition with a prepared CAD model of the arm and bucket. However, detailed shapes and scale parameters should be tuned in advance. Therefore, the point cloud-based bucket model is - used to extract the bucket from point clouds based on point-to-point model fitting. In this study, the iterative nearest point (ICP) algorithm, is applied to bucket extraction from labeled point clouds (Figure 5) by aligning the bucket with the previously prepared point clouds of the bucket.

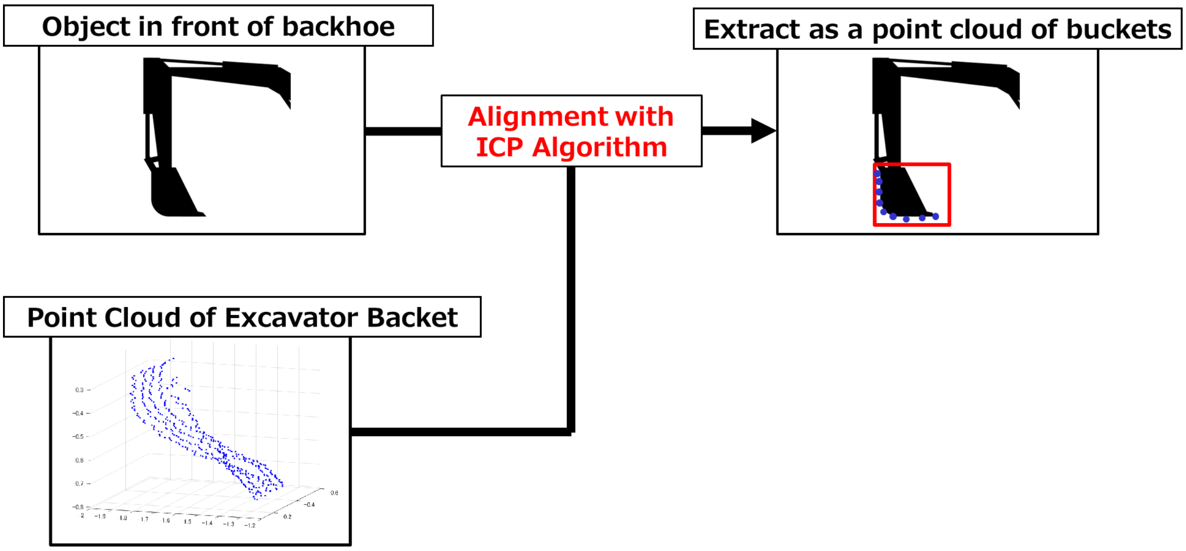


Figure 5: Extraction of a bucket from point clouds.

e. Worker Recognition Processing:

For worker recognition, points are divided into clusters by setting the minimum Euclidean distance between points in different clusters and the minimum number of points required to be recognized as a cluster (Figure 6). Next, based on the knowledge of the height and width of a person, threshold values for the width ($\varDelta W$) and height ($\varDelta H$) of each cluster are set. Then, only clusters that satisfy the threshold values are extracted as workers (Figure 6).

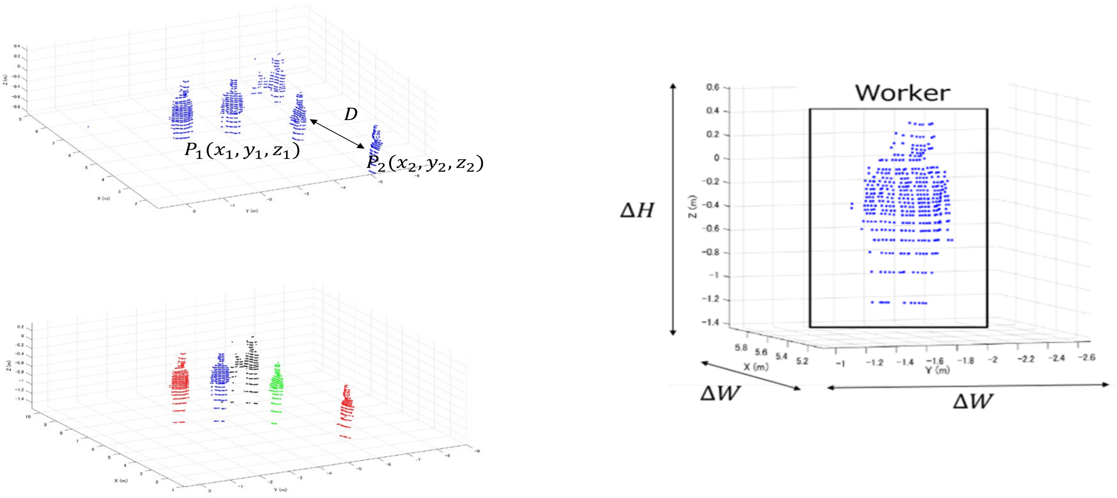


Figure 6: Conceptual image of clustering using Euclidean distance.

(Upper left: acquired point cloud, Lower left: clustering points using Euclidean distance, Right: clustering points using cluster width and height).

f. Worker Tracking Processing:

The worker recognition process is sequential for each scene. Therefore, the tracking process is used to establish the continuity of the worker's time series. The method is to calculate the center of gravity of an object and track its center of gravity. The worker tracking process compares the previous and current scenes of the center of gravity. The worker tracking process also performs the correspondence by distance. However, the time-series point clouds acquired by LiDAR from a turning moving construction vehicle contain rapid horizontal rotation. If the search range of the moving object is set accordingly, there is a high possibility of erroneous correspondence. Therefore, in the tracking process in this study, the simultaneous localization and mapping (SLAM) process is applied to reconstruct the time-series point cloud with the amount of horizontal rotation corrected. However, only the amount of horizontal rotation is corrected in this study, as the amount of translational movement because of the caterpillar movement of construction vehicles in the construction space is only a few centimeters. The time-series point cloud with horizontal rotation correction is used to track the workers, and the search range of the workers in the tracking process is set according to the range within which the moving object may move during one scene. Figure 7 describes the process.

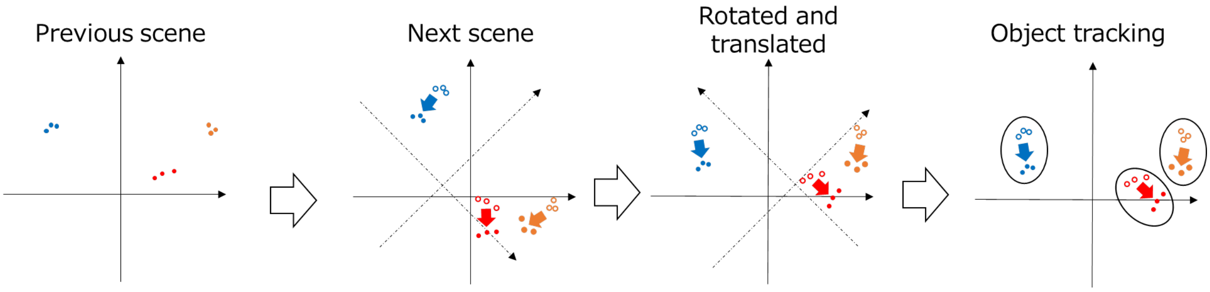


Figure 7: Worker tracking using SLAM.

g. Risk Estimation:

The results of worker tracking and bucket recognition are used to estimate the degree of danger or construction risk assessment.

The proposed risk assessment is expressed by the following equation.

$$Risk = aP + bV + c\theta + dB \ (Equation 1)$$

Here,

P: Relative distance from the bucket to workers estimated from point clouds

V: Speed of worker's movement estimated using temporal point clouds

θ: Relative vector angle between the worker and bucket

B: Variables related to the behavior of the bucket

θ is calculated by finding the angle between the vector from the current scene worker's center of gravity to the center of gravity of the bucket and the inverse vector of the vector from the current scene worker's center of gravity to the previous scene worker's center of gravity. To provide some examples, if θ = 0°, the worker is moving away from the bucket and is, therefore, the safest; if θ = 180°, the worker is coming towards the bucket and is therefore most dangerous, and so on. If the value exceeds 180°, then 360° minus θ is used to ensure accurate results.

B is determined by whether the bucket is moving or not and whether it is turning or not.

Experiment

A simulated construction space (Figure 8) was prepared to represent actual construction activities such as excavation, piping, and backfilling by a backhoe and workers. Horizontal LiDAR (VLP-32C, Velodyne), vertical LiDAR (VLP-16, Velodyne), and oblique LiDAR (Horizon, Livox) were installed in front of the backhoe operator's seat (Figure 8) to acquire temporal point clouds.

Figure 8: Simulated construction space and mounted LiDARs.

Results and Discussion

Figure 9 shows the oblique LiDAR acquisition point cloud results, and Figure 10 shows the object recognition results from the combined LiDAR.

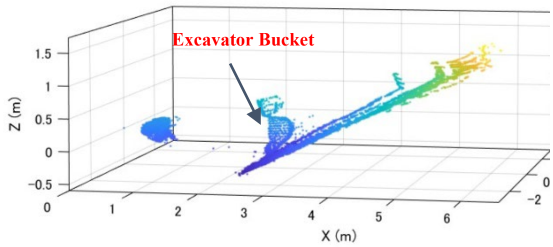


Figure 9: Acquired point clouds (oblique LiDAR).

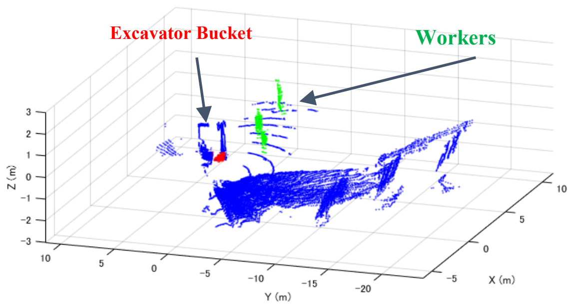


Figure 10: Object recognition results.

Figure 10 shows that the point cloud integration of horizontal, vertical, and oblique LiDARs can cover occluded areas of acquired point clouds. On the ground objects around the construction vehicle, we confirmed that the point cloud density increased from 11.88 [points/m^3] to 42.48 [points/m^3] by vertical and oblique LiDARs. However, we identified the misalignment of the point clouds, as shown in Figure 11. Possible causes include, first, lack of precise registration with positioning by offset values. Second, vibration was caused by the engine of the construction equipment. We also considered the possibility that the scanner was misaligned, but when we checked the processing results during the excavation and turning scenes, no noticeable changes were observed. Therefore, we believe that engine vibration, not scanner misalignment, was the cause. Possible solutions to these problems include registration methods for point clouds and registration methods using planar information of the ground surface. In this study, we attempted to solve this problem using the ICP algorithm. However, it was confirmed that misalignment still occurred. The reason for this is that the search for corresponding points does not work well because of the different scanner scan directions and the occlusion effect caused by the inclusion of the backhoe arm and bucket point cloud in the vertical LiDAR point cloud.

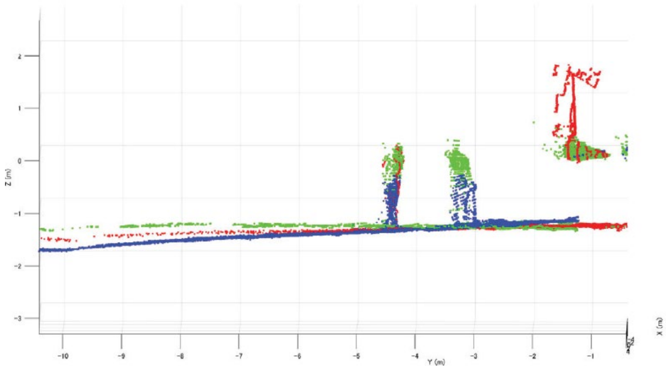


Figure 11: Misalignment of point clouds after alignment.

In the bucket recognition, some discrepancies were observed, as shown in Figure 12. One of the causes is the influence of noise. The ICP algorithm's search for corresponding points was affected by the noises around the bucket. One solution is a noise reduction of point clouds to improve the accuracy of object recognition. However, the noise reduction has the disadvantage of increasing processing time. The displacement of the bucket was as small as 10 cm; thus, it is not considered necessary to take immediate action. When dividing the arm and bucket from the backhoe, the model point cloud of the bucket was extracted from the measured point clouds. However, visual inspection was required to set the initial position of the bucket. Therefore, a future challenge is the automatic calculation of the initial position to make this method a method that can be handled by anyone.

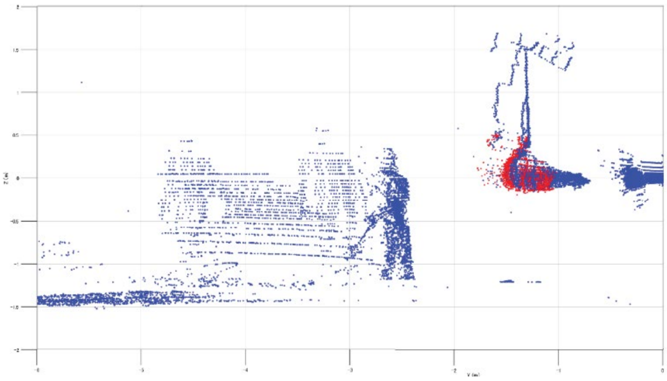


Figure 12: Gaps in recognition and measurement results.

In the worker recognition, it was observed that worker recognition omissions and misrecognition occurred, as shown in Figure 13. The first reason for the recognition omissions was that the worker on the left side of Figure 13 had a long object. This issue caused them to exceed the width threshold in the worker clustering. Second, the workers on the right side of Figure 13 were recognized as one object because of their proximity to each other, causing the width threshold to be exceeded in the same way as before. As a solution, more detailed recognition of workers is necessary. Specifically, worker recognition using color information from cameras with knowledge of body shapes would be effective. One cause of misrecognition is that the point cloud of plants became the same height and width as the workers. The solution is first, accurate recognition of the worker. The second is accurate recognition that they are not workers. Methods include recognition using color information from the camera, and omitting objects that are not in contact with the ground from the workers because the plants are in the air, as shown in Figure 13.

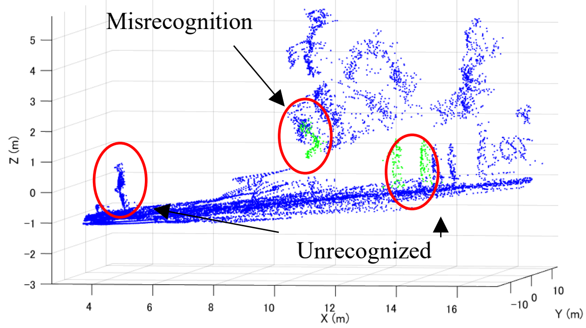


Figure 13: Worker recognition result.

In the worker tracking, it was confirmed in this study that SLAM can accurately output trajectories even when horizontal rotation occurs. However, in our method, when a worker moves out of the LiDAR field of view, the tracking process for the worker who has moved out is stopped at that point. When the worker is recognized again, the tracking of the worker starts again from that scene. Thus, when a worker who was once omitted from recognition appears on the point cloud again, a different ID is assigned. When this problem occurs, even if a risk assessment is performed, the person with the dangerous behavior cannot be identified, and the research lacks development potential. Therefore, it is necessary to understand the behavior of workers in occlusion. The simplest method would be to estimate the position of the worker during occlusion, if the worker will move in the same way in the next scene. However, as discussed later in this discussion, this method cannot handle sudden movements or turns by the worker, and the accuracy of the risk assessment is significantly reduced. The next method to mention is the application of the Kalman filter. The Kalman filter is an algorithm that estimates the state of the system using the state equation calculated based on the input values and observed values of the system. This process is performed sequentially to estimate the state of the system. The Kalman filter is used in various fields, and because it is also used in point clouds, there is a possibility that it can be used in this study.

The current risk assessment and the risk estimation after 5 seconds are shown in Figure 14. The left side of Figure 14 shows the current risk assessment and 5-second risk prediction for each worker at time t, and the right side shows the same risk assessment at time t+5. However, it was found that the risk estimation after 5 seconds was not accurate because the 5-second risk prediction of worker 1 at time t and the current risk evaluation of worker 1 at time t+5 did not match. The reason for this is that the worker behavior prediction assumes that the worker will continue to move in the same manner. Therefore, a more accurate prediction of the worker's behavior is needed. A method that predicts future movements by applying a skeletal model of the worker's point cloud to image data and using deep learning is effective. Research has already been conducted on the skeletal model of a point cloud of a worker, and it can be fully utilized (Horiuchi et al., 2017). In addition, the risk value calculated by the risk evaluation formula in this study was determined to be dangerous when it exceeds 80, based on observation of on-site movements. However, this criterion is not accurate and needs to be improved. Possible improvement methods include the use of statistical methods and simulation of the construction space to investigate appropriate risk value thresholds.

Risk estimation results at time t. Risk estimation results at time t + 5 [s].

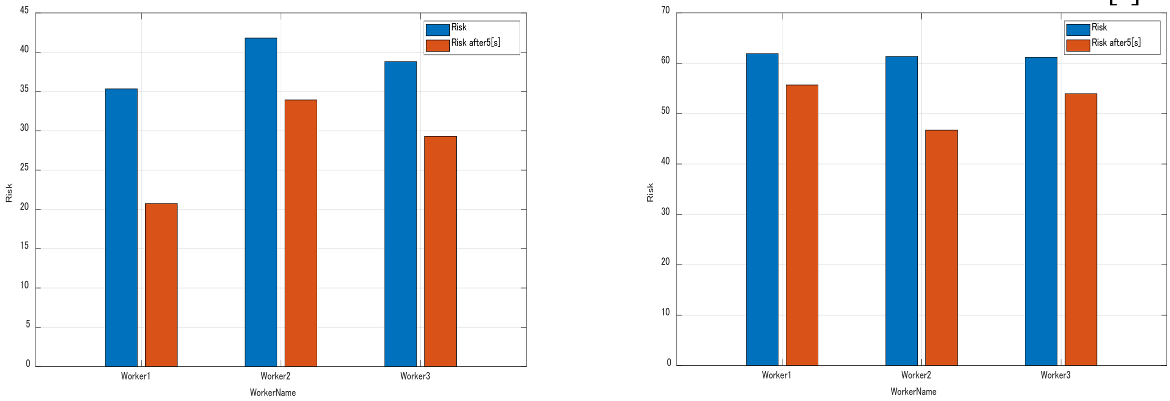


Figure 14: Worker recognition result.

In the processing time, Table 1 confirms that the real-time performance is low. The overall processing in the proposed method took 0.56 [s] per scene, and the output of the risk estimation results was about 2 Hz. The point cloud used in this study was acquired at 10 Hz, which is below the sampling rate of LiDAR. Therefore, point cloud processing needs to be accelerated. In particular, the processes that require higher speed are bucket recognition, worker recognition, and SLAM processing. Because these three processes require 0.53 [s] per scene, real-time performance can be ensured by reducing these processing times. One specific way to shorten the processing time is to adjust the number of input point clouds according to the processing. For example, in the worker recognition process, because many workers exist on the point cloud acquired by horizontal LiDAR, the point clouds of vertical and oblique LiDARs, which are less effective, may be down-sampled. In the SLAM process, it is also effective to use only the point cloud of the channel with the smallest vertical angle, making it easier to acquire the point cloud of the 3-D object, out of the 32 channels, instead of using the entire point cloud of the horizontal LiDAR. Using these methods, we verified the reduction in processing time because of the choice of input point cloud; for the verification of SLAM processing, we prepared a composite LiDAR point cloud, a horizontal LiDAR point cloud, and a point cloud with the number of horizontal LiDAR channels limited to four. Table 2 shows the SLAM processing times for each of these input point clouds. Figure 15 also shows the results for horizontal rotation for each input point cloud.

Table 1: Processing time.

Processing Details	Processing Time [s/scene]
Point Cloud Integration	0.01
Bucket Recognition Processing	0.14
Worker Recognition Processing	0.23
SLAM Processing	0.16
Worker Tracking Processing	0.01
Risk Estimation	0.01
Total Processing Time	0.56

Table 2: Processing time for different input point clouds in SLAM processing.

Input point cloud	Combined LiDAR	Horizontal LiDAR	Horizontal LiDAR (limited)
Processing Time [s/scene]	0.16	0.09	0.02

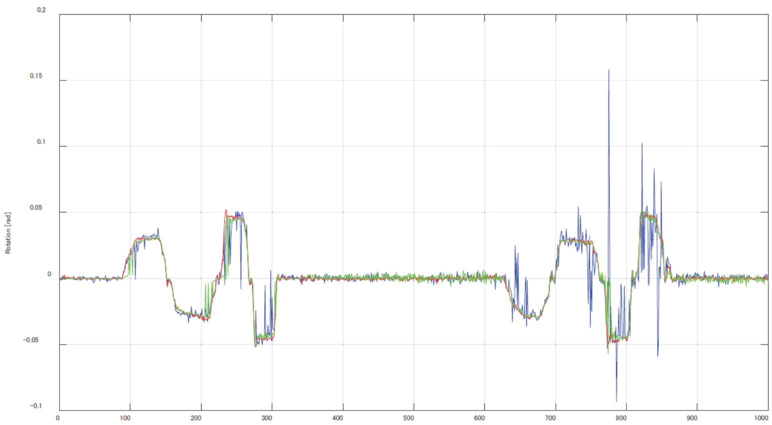


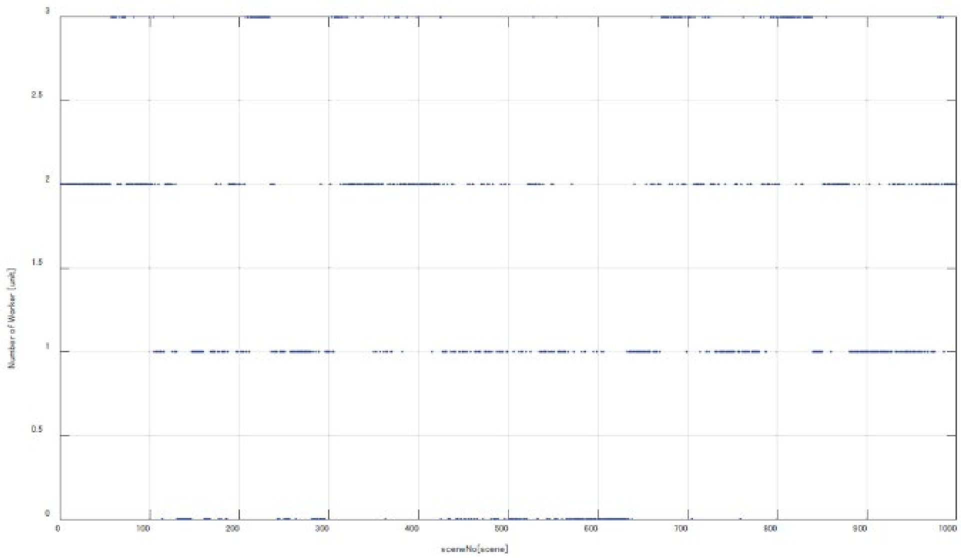
Figure 15: Horizontal rotation estimation results by input point cloud.

As Table 2 shows, the processing time was reduced by 50%–90%. However, as shown in Figure 15, the estimation results varied. The large outliers occurred in point clouds with a limited number of horizontal LiDAR channels, and the errors in the horizontal LiDAR itself are small. Therefore, considering the stability and processing time of the SLAM process, we believe that it is appropriate to use only the input point cloud of the horizontal LiDAR.

In the verification of the worker recognition process, we used two types of input point clouds: composite LiDAR point clouds and horizontal LiDAR point clouds. Table 3 shows the processing time of the worker recognition process for each input point cloud. Figure 16 also shows the results of the worker recognition achieved for each input point cloud.

Table 3: Processing time for different input point clouds in worker recognition processing.

Input point cloud	**Combined LiDAR**	**Horizontal LiDAR**
Processing Time [s/scene]	0.23	0.09



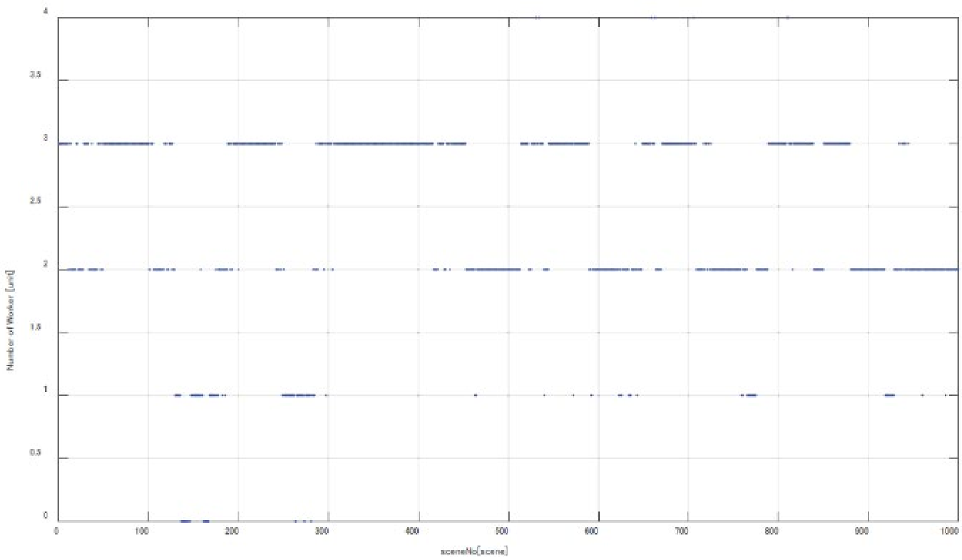


Figure 16: The number of recognized workers from input point clouds.
(Top: combined LiDAR, Bottom: horizontal LiDAR).

As shown in Table 3, the processing time was reduced by 70%. However, as shown in Figure 16, the number of workers makes a difference. Because the composite LiDAR point cloud is more correct when checked, the horizontal LiDAR point cloud alone causes a lot of misrecognition and recognition omissions as mentioned earlier. Therefore, it is appropriate to use the composite LiDAR point cloud in the worker recognition process. If the processing time still must be reduced, other approaches must be considered.

Conclusion and Recommendation

In this study, a method for integrating point clouds acquired from construction vehicles by horizontal LiDAR, vertical LiDAR, and oblique LiDAR, and a method for recognizing backhoe buckets and recognizing and tracking workers using the integration results was developed, as well as a risk assessment method using these results. Through the experiments, we confirmed that the combination of the three LiDAR systems can acquire point clouds of bucket behavior and the bucket’s work area, which is difficult for horizontal scanning LiDAR. Moreover, the combination of the three LiDAR systems can recognize and extract the bucket and workers by object recognition. We also confirmed that risk estimation is possible by using the designed risk evaluation equation. In our future works, it will be necessary to improve the accuracy of risk assessment by accurately identifying workers and more precisely estimating construction vehicles and workers’ behaviors and to improve the processing speed of object recognition and SLAM to archive real-time processing.

References

Kajita, H., Ito, S., Hashimoto, T., 2017. Efforts to Improve Productivity in Unmanned Construction Technology. Civil Engineering Materials, 59 (1), pp. 30-35.

Kondo, T., Aoki, H., Miyazaki, H., 2011. Current Status and Future of Unmanned Constru ction in the Construction Industry. Taisei Corporation Technical Center Report, 44 (19), p p. 1-7

Horiuchi, Y., Makino, Y., Shinoda, H., 2017. Computational Foresight: Forecasting Huma n Body Motion in Real-time for Reducing Delays in Interactive System, ISS '17: Proceedi ngs of the 2017 ACM International Conference on Interactive Surfaces and Spaces, pp. 31 2–317,