Proceedings of the Sixth International Conference on
Railway Technology: Research, Development and Maintenance
Edited by: J. Pombo
Civil-Comp Conferences, Volume 7, Paper 11.6
Civil-Comp Press, Edinburgh, United Kingdom, 2024
ISSN: 2753-3239, doi: 10.4203/ccc.7.11.6

Obstacle Detection using Camera and LiDAR for Train Forward Surveillance

R. Kageyama

Information and Communication Technology Division,
Railway Technical Research Institute
Tokyo, Japan

Abstract

For the further improving the railway safety, we are developing a train forward surveillance method using cameras and sensors. In train forward surveillance, it is important to detect obstacles entering the tracks at a distance. Therefore, we investigated the combination of images obtained from cameras and point cloud data obtained from LiDAR as a suitable senser configuration for the railway environment. In our proposed method, the detection area is set by predicting the area of the rail track from the image, and objects such as people and automobiles are recognized using the deep learning. In addition, by combining detection using point cloud data, we can avoid missing objects even in conditions where recognition from images is difficult, such as at nighttime.

Keywords: train forward surveillance, camera, LiDAR, deep learning, point cloud processing, sensor fusion

1 Introduction

To further improve railway safety, it is important to reduce the risk of contact accidents between trains and obstacles on the tracks. If obstacles on the tracks can be detected by a forward monitoring system that assists the driver, contact accidents can be avoided and damage can be mitigated. In the automobile industry, the development of Advanced Driver Assistance Systems (ADAS) has been progressing, and these systems have been introduced in mass-market vehicles. On the other hand, in the case of railways, the coefficient of friction between the railway wheels and the track is low,

and it takes a greater distance from the detection of an obstacle to the stopping point than for a car traveling at the same speed. Currently in Japan, the requirements for sensor technology necessary to detect obstructions from a distance far enough away to allow trains to slow down or stop have not been established. Therefore, we are developing an obstacle detection method to detect objects at a distance using a camera and a sensor, aiming to establish the requirements for application to train forward surveillance. This paper first describes the results of the sensor configuration based on the current status of collisions between trains and obstructions and the characteristics of the sensors. Next, an overall view of the obstacle detection method is presented. Finally, the result of the verification of the detection performance of the proposed method is shown.

2 Background

2.1 Accidental contact between train and obstacles

In order to understand the current status of contact accidents between trains and obstacles in Japan, we compiled data on past accidents. Specifically, among the accident information stored in the "Railway Safety Database" [1] published by Railway Technical Research Institute, we analysed accidents caused by contact between trains and obstacles and cases that led to significant delays (8,343 cases in total) over the past 20 years.

Figure 1 shows the breakdown of obstacles in the total number of cases. More than half of the cases involved contact with persons entering the railway tracks. Other than people, most cases involve contact with automobiles and motorcycles, and there are also cases of contact with fallen trees, gravel and avalanches.

Figure 2 shows the number of accidents by time of day. In most rail lines, trains operate from 5:00 to midnight the next day, and there is no significant difference in the number of incidents between day and night during this time period.

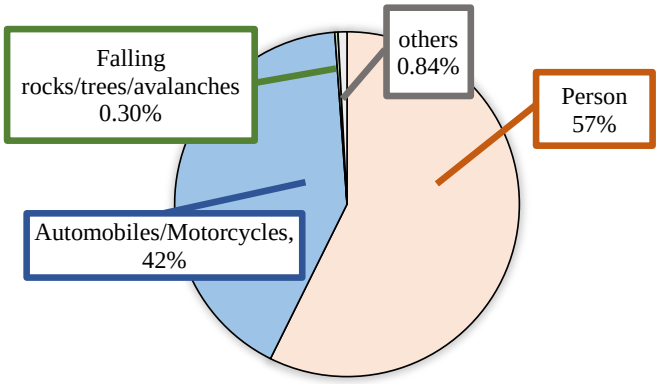


Figure 1: Breakdown of collisions between trains and obstacles by type of obstacles.

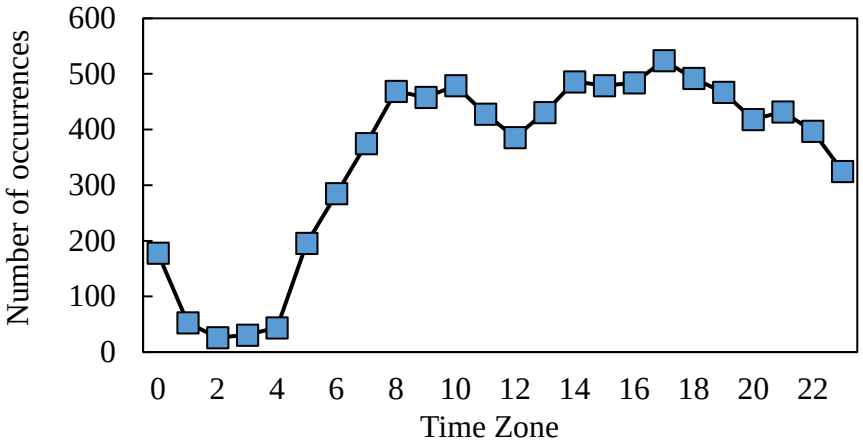


Figure 2: Number of collisions by time of day.

2.2 Characteristics of sensors used for obstacle detection

Table 1 summarizes the characteristics of four typical sensors (visible light camera, LiDAR, millimetre wave radar) that are mainly used in automotive ADAS.

A visible light camera is a sensor that records visible light reflected from an object onto an imaging element. During the daytime, the camera can capture objects with high resolution in the distance, but at night or under other conditions where illumination is not available, it becomes difficult to see the image of the object. In addition, it is difficult for a camera alone to capture the distance to an object or its shape in three dimensions.

LiDAR (Light Detection and Ranging) is a sensor that emits near-infrared light (light with a wavelength of approximately 900~1,000 nm), measures the distance to an object from the time it takes to return to the light receiving surface, and acquires a set of distance information as point cloud data. It can capture the shape of an object in three dimensions, and its performance is independent of illumination. On the other hand, since point cloud data only contains information on the distance to the object and the reflectance of the laser, it is difficult to identify the class of object unless the reflectance is different from that of the surroundings.

Millimetre wave radar is a sensor that uses millimetre wave reflections to make measurements. Depending on the bandwidth, it can detect the presence or absence of objects with high resolution, but it is difficult to determine the detailed shape of objects at a distance.

Sensor	Time zone	Object identification	Resolution	3D measurement
Visible light camera	Day	Identifiable by colour and contour	Order a few centimetres	Impossible due to the principle of the sensor
	Night	Depends on ambient illumination		
LiDAR	Day	Only objects with differences in reflectance can be identified	Order a few centimetres	Possible
	Night			
Millimetre wave radar	Day	Only objects with differences in reflectance can be identified	Orders from a few centimetres to several tens of centimetres	Difficult to measure in detail
	Night			

Table 1: Characteristics of the main sensors

3 Investigation of sensor configuration suitable for obstacle detection

Based on the current status of contact accidents between trains and obstructions presented in Section 2.1, the following two requirements for obstacle detection were defined.

Requirement 1: It is possible to identify that a person, vehicle, or other object has entered the railway tracks.

Requirement 2: It is possible to detect targets during both daytime and nighttime.

Based on the characteristics of the sensors described in Section 2.2, we examined sensor configurations that satisfy the above requirements. To satisfy requirement 1, the sensor must be able to correctly recognize the inside and outside of a railway track and the type of obstacles. Also, to satisfy requirement 2, the sensor must be able to detect obstacles independent of illumination. Requirement 1 can basically be satisfied by using images from a visible light camera and object detection by machine learning. However, as mentioned above, the camera alone cannot clearly capture the image of an object in low-light conditions, so it is necessary to cover the weak point of the camera by using it in combination with other sensors in order to satisfy requirement 2 as well. In this study, we decided to investigate a sensor configuration that combines LiDAR as a sensor that can correctly grasp the shape of an object up to a distance.

4 Obstacle detection method

4.1 Overview of proposed method

The overview image of the obstacle detection algorithm using cameras and LiDAR is shown in Figure 3. The data obtained from the sensor in front of the train includes objects that should not be detected as obstacles, such as people on the platform, for example. To exclude such objects, a detection area is set around the train tracks at first. Next, targets are detected using both images and point cloud data. The detection results are judged to be inside or outside the detection range, and an alert is issued only when the target is within the detection range. Of these, the detection range setting method is described in Section 4.2, and the object detection method by camera-LiDAR fusion is described in detail in Section 4.3, as the main technical elements.

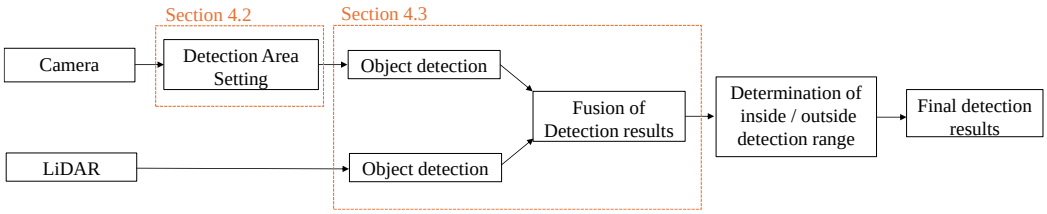


Figure 3: Overview of obstacle detection algorithm

4.2 Detection area setting from images.

The detection area is set toward the direction in which the train is moving in the area around the tracks in the front image. First, an image of the detection area setting is shown in Figure 4. Specifically, the regions of rail track are extracted by PIDNet [2], a deep learning method that predicts regions from an image in pixel units. Next, multiple horizontal lines are set in the image and the intersection points of each horizontal line with the extracted rail region are obtained. The approximate gauge at each horizontal line can be determined by giving the position of the vanishing point of the rail in the image and the width of the rail at the bottom edge of the image as parameters. From this, the points of rails at the intersection of each horizontal line that are close to the length of the gauge are extracted as the left and right rail pair. By connecting pairs belonging to the same rail area, starting from the bottom edge of the image, the detection range can be configured in the direction of moving. In addition to the inside of the rail, a certain amount of space (clearance gauge) must be maintained in the outside direction as an area where the object must not intrude. Since the width of the building limit is fixed according to the width of the rail (for example, the width of the clearance gauge is 3,800 mm on a conventional rail line with a gauge of 1,067 mm in Japan), this information is used to set the detection area by expanding the extracted rail area to the left or right by the amount of the clearance gauge.

An example of the application of the proposed method to an image in the dataset of train forward images (RailSem19[3]) is shown in Figure 5. The pink circles on the rails represent the extracted left and right rail pairs. It can be seen that the detection range can be set along the direction of the line regardless of weather conditions. Furthermore, even at locations where turnouts exist, the detection range can be set smoothly toward the direction where the turnouts are open.

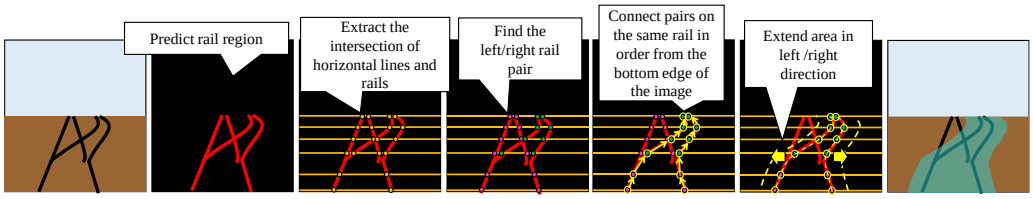


Figure 4: Image of the detection area setting.

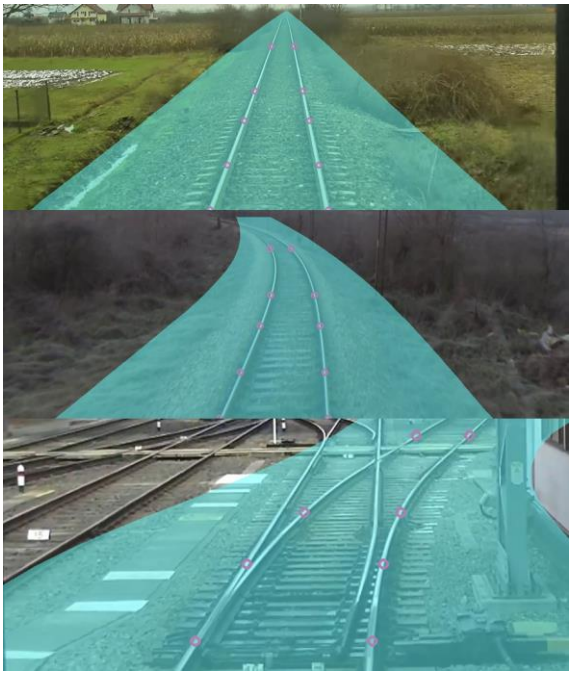


Figure 5: The example of detection area setting.

4.3 Obstacle detection by camera-LiDAR fusion

After setting the detection area, objects are detected from each of the camera image data and LiDAR point cloud data, and the results are fused. In order to combine the detection results of both sensors, the coordinate system of each sensor must be aligned through prior calibration. Let (u, v) be the coordinate system with the center of the camera image as the origin and (X_L, Y_L, Z_L) be the coordinate system with the light-

receiving surface of LiDAR as the origin, the relationship between the two is expressed by the following equation,

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \mathbf{K}[\mathbf{R}, \mathbf{T}] \begin{pmatrix} X_L \\ Y_L \\ Z_L \\ 1 \end{pmatrix} \quad (1)$$

where **K** is the intrinsic parameters of the camera and **R**, **T** are the extrinsic parameter (rotation matrix and translation vector) between the camera and LiDAR coordinate system. Once the intrinsic parameters are obtained in advance, extrinsic parameters can be estimated by the PnP algorithm, which minimizes the distance between the image and the corresponding points in the point cloud, given multiple combinations of points. Once the correct extrinsic parameters are obtained, the point cloud data can be projected onto the image, as shown in Figure 6. In Figure 6, the projected points are coloured according to distance (red for near areas and blue for far areas), indicating that red points are projected on the track surface in front of the sensor and on grass and trees along the track.

Figure 6: The result of projection of point cloud into image.

On the assumption that the sensors are properly calibrated with each other, detection results from point clouds can be projected into detection results from images to achieve highly robust detection. An image of the integration of detection results is shown in Figure 7. From the image, the position, size, and type of the object are simultaneously predicted by using YOLOX [4], a deep learning method used in object detection. In the output layer of the deep learning model, an array (feature map) containing the following information is obtained for each grid that divides the image evenly.

(1) Probability of objectness p_{obj} $(0 \leq p_{obj} \leq 1)$

(2) Class of object

(3) probability of determining class p_{class} $(0 \leq p_{class} \leq 1)$

(4) Coordinates and size of objects in the image

From the point cloud data, the area corresponding to the rail track is removed from the image, and the process of extracting three-dimensional objects above the rail track is applied. To remove the track area, RANSAC [5], an algorithm that estimates the plane from the 3D point cloud, is applied to the point cloud data projected onto the detection area in Section 4.2. Since the extracted point cloud data includes noise caused by ambient light, etc., DBSCAN [6], a density-based clustering process that extracts regions of high density of point clouds as a single chunk, is applied.

Finally, the results of clustering are projected onto the image and fusion processing is executed. Since it can be said that an object almost certainly exists at the location where the cluster exists, the value of the feature map corresponding to the location where the centre of the cluster is projected is referenced, and the value of objectness p_{obj} is set to 1. For the class of object, if the class determination probability p_{class} determined from the image at the projected location exceeds a certain threshold value, the value predicted from the image is adopted; otherwise, the "Unknown" is considered. In the image in Figure 7, the value of objectness at the centre of the person detected in the image is 0.7 in the image alone, but by projecting the cluster centre coordinates of the point cloud, the value is updated to 1, indicating a higher probability of objectness.

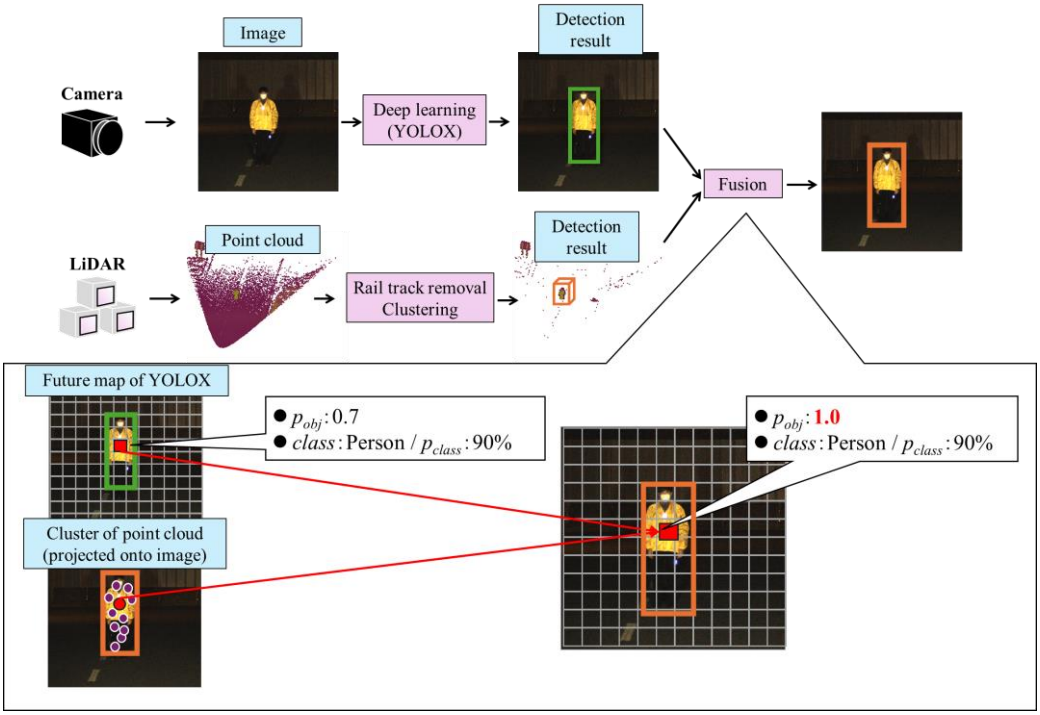


Figure 7: The image of object detection from images and point cloud.

5 Performance evaluation of the proposed method

5.1 Summary of the evaluation experiment

To evaluate the detection performance of the detection algorithm at each distance, a stationary test was conducted on a test road where a straight section of 600 m was available. To simulate the situation where the sensors are installed in front of an actual vehicle, a sensor unit (a camera and nine LiDARs) and two headlights were installed on the back of a truck, as shown in Figure 8. In the sensor unit, nine LiDARs were lined up vertically and horizontally with a spacing of approximately 1 cm. The distance between the headlights was 1.7 m. During the experiment, the data was repeatedly recorded and moved in a fixed position for 10 seconds from a point 50 m away from the subject to a point 600 m away from the subject. As subjects, we targeted people dressed differently, as shown in Figure 9.

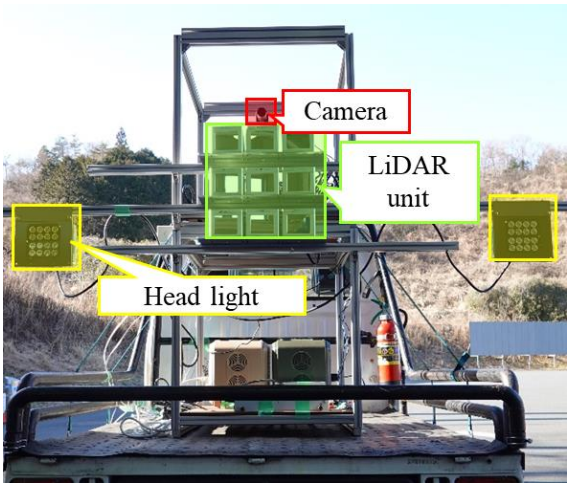


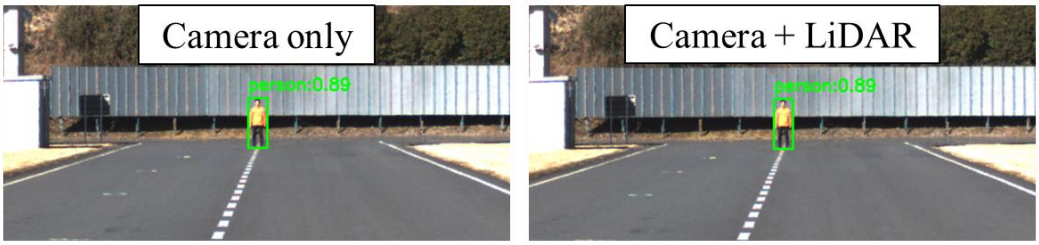
Figure 8: Equipment configuration at the time of the experiment.

Figure 9: Examples of subjects.

5.2 The result of the experiment

For the data acquired at each location, the detection rate was calculated for each distance as the percentage of the data that correctly predicted the presence of an object where it should be located. We used the result of annotating the image in advance as the correct value of the object's position and defined “detection” as when the percentage of overlap (IoU) between the correct position and the detected position exceeds a certain value.

Figure 10 shows examples of detection results for a person 300 m away, both during the (a) daytime and at (b) nighttime. The left side shows the result of using only the camera and the right side shows the result of combining the camera and LiDAR. Although detection using only image information is difficult at night due to the reduced illumination, the detection became possible by combining information from point cloud data.



(a) Daytime

(b) Nighttime

Figure 10: Detection result for a person 300m away.

To investigate the relationship between the density of laser points emitted by LiDARs and detection performance, point cloud data for $N(N = 1, 2, \dots, 9)$ LiDARs were processed to obtain detection rates. The relationship between the number of LiDAR and the detection rate at each distance is shown in Figure 11. The figure shows a correlation between the number of LiDARs used for detection (density of laser points) and the detection rate at each location. When detection processing was performed using all the LiDARs in the sensor unit, it was found that a person could be detected up to 400 m away with a detection rate of more than 90%. Furthermore, based on this relationship, the detection rate was estimated assuming a further increase in the number of LiDAR, and it was predicted that a person could be detected up to 500 m away when the number of LiDAR was increased to 14 units. Table 2 shows the relationship between the number of LiDAR used in the experiment and the number

(density) of points irradiated per m^2 per 0.1 second at 500 m away. From this, it was found that, for example, to correctly detect a person at 500 m away, the sensor unit should be configured so that the density of laser points per 0.1 second is 25 points / m^2.Currently, a larger number of LiDARs results in a larger sensor unit, but if the density of laser points obtained from a LiDAR increases with further improvements in sensor technology in the future, it should be possible to achieve detection with the same accuracy with a more compact sensor configuration.

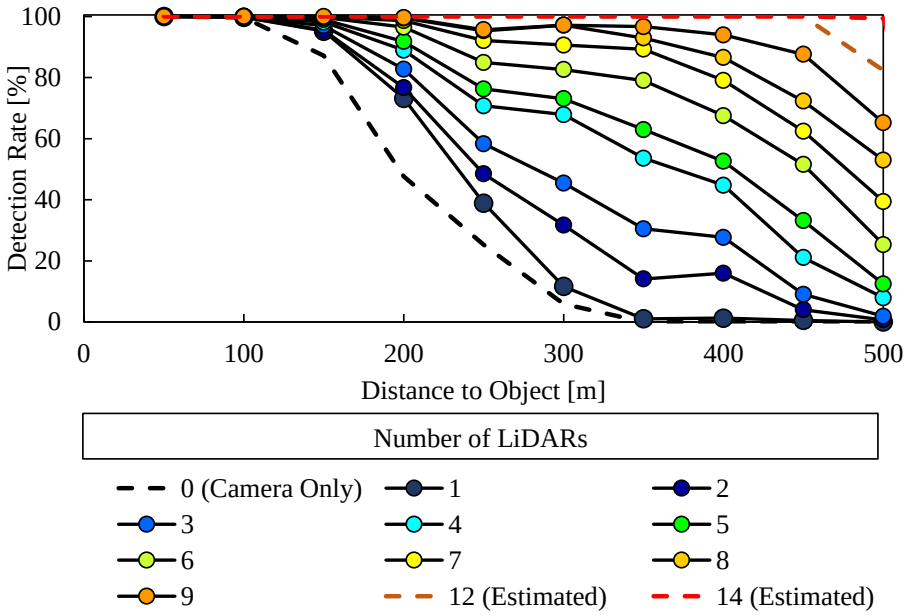


Figure 11: The relationship between LiDAR number and detection rate.

Number of LiDAR	1	3	6	9	14
Number of laser point (/0.1s, $1m^2$)	1.8	5.3	11	16	25

Table 2: The relationship between the number of LiDAR and number of laser point.

6 Conclusion

This paper describes efforts to develop an obstacle detection method using cameras and sensors for train front surveillance. We investigated accident cases and sensor characteristics in Japan, and based on both perspectives, showed that cameras and LiDAR are suitable sensor configurations for detection. We developed a method for setting the detection area based on the rail track predicted from the image, and it was confirmed that the detection area can be set correctly along the direction of moving. We also developed a detection method that combines the results of object detection

from images with the results of clustering on point cloud, and confirmed that it can correctly detect people both day and night in the performance evaluation in stationary conditions. Furthermore, we confirmed the trend that the detection rate improves as the number of sensors comprising the sensor unit is increased, indicating the possibility of further improvement in detection performance with an increase in LiDAR laser density. In the future, we plan to conduct a similar detection performance evaluation under conditions similar to the actual railway environment.

References

[1] Railway Technology Promotion Centre in Railway Technical Research Institute, "Railway Safety Database".

[2] J. Xu, Z. Xiong, S. P. Bhattacharyya, "PIDNet: A real-time Semantic Segmentation Network Inspired by PID Controllers", arXiv: 2206.02066, 2022.

[3] O. Zendel, M. Murschitz, M. Zeilinger, D. Steininger, S. Abbasi, C. Beleznai, "RailSem19: A Dataset for Semantic Rail Scene Understanding", Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2019.

[4] Z. Ge, S. Liu, F. Wang, Z. Li, J. Sun, "YOLOX: Exceeding YOLO Series in 2021", arXiv: 1808.00897, 2021.

[5] M. A. Fischler, R. C. Bolles, "Random sample consensus: paradigm for model fitting with applications to image analysis and automated cartography", Communication of the ACM, 24, 381-395, 1981.

[6] M. Ester, H. P. Kriegel, J. Sander, X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise", KDD-96: Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, 96, 226-231, 1996.