Train Frontal Obstacle Detection Method with Camera-LiDAR Fusion

Ryo KAGEYAMA  Nazomi NAGAMINE  Hiroki MUKOJIMA

Image Analysis Laboratory, Information and Communication Technology Division

Recently, the importance of obstacle detection methods for railway has been increasing. In the field of automobiles, obstacle detection systems with sensors have been introduced on mass-produced vehicles. However, in railway, a practical detection system does not exist because railways require longer detection distances than do automobiles. Therefore, we have developed a train frontal obstacle detection method using a camera and LiDAR. We confirmed that our method detects a person 200 m away, which a camera alone cannot detect, with 45% accuracy at night.

Key words: train front monitoring, deep learning, LiDAR, 3D point cloud processing sensor fusion

1. Introduction

Reducing the number of collisions is important to further improve the safety of railways. In the field of automobiles, Advanced Driver-Assistance Systems (ADAS) using cameras and sensors have been already developed and introduced on mass-produced vehicles [1]. In the case of railways, the coefficient of friction between wheels and rails is low, and braking distances are about three times longer than for cars traveling at the same speed. For this reason, it is difficult to apply ADAS for automobiles directly to railways, and sensors and algorithms that can detect more distant objects are needed. However, so far no system that meets these requirements has been put into practice. Therefore, we developed a method for detecting obstacles in front of trains which could be applied to operational support systems for railways. Our past studies examined detection methods based on machine learning [2] and detecting differences in successive images [3]. In this study, we developed a detection method using deep learning to further improve detection accuracy. In addition, as a countermeasure against the weakness visible light cameras, which is the deterioration of detection performance at night, we developed a detection method with integrated LiDAR.

2. Study of obstacle detection methods for railway

2.1 Target setting for detection performance

2.1.1 Consideration of detection targets

First, we examined the target object. According to the results of a survey of the percentage of railway crossing accidents by object involved in collisions, collisions with pedestrians accounted for the highest percentage of the 208 railway crossing accidents that occurred in FY 2019, which was about half the total number of accidents, or 95 cases [4]. Based on this observation, it was considered that detection of a person and avoidance of a collision would contribute to a significant reduction in the number of accidents. Therefore, in this study, we set people as the main target for detection.

2.1.2 Consideration of detection distance

The detection distance that should be satisfied by the obstacle detection method for railway should be at least 600 m, which is the standard braking distance in an emergency on conventional railway in Japan [5]. However, it is difficult to detect an object as large as a person 600 m away with the performance of sensors available currently. In this study, we investigated distances below 600 m which would still mitigate disaster and set them as the target detection distances. For the main JR lines (about 10,300 km length), we calculated the percentage of line lengths that could be stopped at different braking distances from maximum speed of each line section and vehicle emergency deceleration [6]. Figure 1 shows a graph with the horizontal axis representing the braking distance and the vertical axis representing the percentage of line segments where this stopping distance was possible. It was found that when the braking distance is more than 300 m, stopping is possible on more than 50% of the route. From this, it can be inferred that if the target can be detected 300 m away, disaster mitigation would be effective for at least half of the line segments. Therefore, the target detection distance was set at 300 m in this study.

2.2 Target setting for detection performance

Figure 2 shows the overall picture of the detection method to achieve the target detection distance of 300 m. The proposed method extracts rails from the image of what can be seen from the front of the train, obtained from the camera, to limit the detection area to the direction in which the train is moving and more than 300 m ahead, which is the target detection distance. For the extracted detection area, object detection was performed only from the image information during daytime. On the other hand, during nighttime, when visible light cameras are not good at detecting objects, detection performance is ensured by integrating 3D point clouds obtained from LiDAR.

![Fig. 1 Relationship between braking distance and percentage of route with suitable stopping distance.](image1.png)
3. Over-detection suppression method

3.1 Suppression of over-detection by rail detection

Images from the front of the train contain information on various objects such as tracks, equipment, roads, and buildings, etc., which can cause over-detection. Since this leads to excessive train emergency stops, it should be prevented. Therefore, we developed a method to limit the detection area in the direction of train travel in order to suppress excessive detection in the preliminary stage of inputting the train front image to the detection algorithm.

Figure 3 shows a flow chart illustrating how the detection area is limited. In the proposed method, first, the rail position is learned by deep learning using single frames of the video taken from the front of the train and a dataset of mask images with pixel values only around the rail position in the corresponding frame. Next, the rail position is detected using the learned deep learning model. From among the rail detection frames, the detection region is then limited to an area centered on the vanishing point of the rail.

3.2 Over-detection suppression effect

Using the results of rail extraction using this method, the detection area limited around the extracted vanishing point are shown in Fig. 4. It was confirmed that the rails in the direction of travel were correctly extracted regardless of the track surface condition and alignment, and that the detection area was appropriately narrowed down.

The number of frames in which over-detection occurred for 2,500 frames of forward video was also compared. The number of frames in which over-detection occurred before limiting the detection area was 52, while the number of frames in which over-detection occurred after limiting the detection area was 1. This result confirms that this method can reduce the over-detection to less than 2%.

4. Detection methods using camera alone

4.1 Deep-learning based detection method

We developed a method to detect people and other objects using only a camera, on the basis of the detection area limited as described in the previous section. Figure 5 shows an image of the detection method. The proposed method uses an object detection algorithm based on deep learning [7]. This algorithm simultaneously predicts the object-likeness and type of objects in an image, and outputs the position, size, probability of existence, and classification of the object. The detection result is output for each frame, but information from the past 10 frames is used for the detection decision, and a frame is considered “detected” if 3 or more frames are observed in which “the existence probability is 80% or more and the person is correctly classified” out of 10 frames. This reduces the influence of instantaneous noise that occurs when the image of a person is momentarily obscured by windshield wipers, etc. in rainy weather.

4.2 Experiment for performance evaluation of the proposed detection method

4.2.1 Method of experiment

A field experiment was conducted to evaluate the detection method. A camera was installed inside the cab window at the back of a train, as shown in Fig. 6. Then, the camera took pictures as the train moved away from the subject (about 475 m away) again from the starting point near the subject on the track, as shown in Fig. 7. This single run is referred to as one trial run. The subjects were...
people in different postures and clothing as shown in Fig. 8. In order to obtain sufficient resolution (more than 50 pixels) for the person at a distance of 300 m, a USB camera with a pixel count of 4,096 (W) x 2,160 (H) and a lens with a focal length of 50 mm was used.

4.2.2 Method of evaluating detection rate per distance

The detection rate for each distance between the train and the subject was obtained from the captured forward video using the following procedure.

I. The total amount of movement is calculated by integrating over the frames after calculating the amount of movement of feature points between frames in the video.

II. The distance between the train and the subject in each frame is estimated from the correspondence between the total distance traveled by the train per test number (475 m) and the total movement of the feature points.

III. The entire frame is divided into 10-m distance increments from 0 m to 480 m, and the detection decision logic described in Section 4.1 is applied in each section. The detection rate is defined as the percentage of “detection” states.

For example, if the detection process is executed a total of 10 times for images obtained at 100 m to 110 m from the subject, and the “detection” status is obtained 9 times out of 10 times, the detection rate is 90%.

4.2.3 Result of evaluation

The detection algorithm was applied to a total of 56 test videos of a standing person during the daytime, and the detection rate for each distance was evaluated according to the procedure described in Section 4.2.2. Figure 9 shows an example of detection of a standing person facing 400 m away for one trial. It can be seen that the red frame indicating detection is displayed around the person even at a distance that is difficult to see with the naked eye. Figure 10 shows the results of obtaining the average value of all trial numbers (hereinafter referred to as “average detection rate”) and the maximum/minimum value for each trial number for the detection rate at each distance. The average value of the trial number confirms that the detection rate exceeds 98% at a target distance of less than 300 m. Figure 11 shows the results of the evaluation of the detection rate at different distances for different illuminance levels, using a person in a standing position facing forward as the subject and repeatedly photographed in the evening. The figure shows that the distance at which more than 90% of the subjects could be detected became shorter as the illuminance decreased, and this tendency was particularly pronounced at illuminances of less than 10 lx.

5. Detection method by camera-LiDAR fusion

5.1 Selection of a sensor used with the camera

The results of Section 4.2.3 suggest that in low-light conditions, in addition to cameras, it is necessary to combine other sensors that can be applied regardless of illumination. In automotive driver assistance systems, sensor fusion technology, which integrates information from multiple sensors to provide robust detection against environmental changes, has been adopted [8]. Millimeter-wave radar and LiDAR (Light Detection And Ranging) are rep-
Millimeter wave radar emits radio waves in the millimeter wave band and uses the reflected wave information for object detection. LiDAR emits lasers in various directions and acquires reflections from objects as a point cloud. Like millimeter-wave radar, LiDAR can be used even at night, and depending on the model, high-density point cloud data can be obtained.

Sensors used together with cameras at night in a railway environment should be able to measure far distances regardless of brightness and have sufficient spatial resolution to detect objects. We decided to investigate the integration of a camera and LiDAR as a sensor to satisfy these conditions. In this study, we selected a sensor that can measure objects at distances of at least 300 m, regardless of the reflectance of the object and which had an angular resolution of 0.03° or less between adjacent lasers so that at least two lasers could hit a front-facing, standing person at 300 m away in the shoulder-width direction. The main characteristics of the selected LiDAR are shown in Table 1, and the point cloud of the test line in the Railway Research Institute acquired by LiDAR is shown in Fig. 12. Results shown in Fig. 12 confirm that a high-density 3D point cloud of the railway environment can be acquired.

### 5.2 Detection method by camera-LiDAR fusion

We developed a detection method by integrating a camera and selected LiDAR. An image of the detection method is shown in Fig. 13. In this method, the 3D point cloud data measured by LiDAR is superimposed on the deep learning detection method described in Section 4.1. In the process of detecting objects from images by deep learning, predicted values of object-likeness and object classification are output for each small region (grid) that delimits the image. When the image of the object is clear, the area around the grid with high object-likeness is converted into the position and size of the object and output as the final detection result along with the classification of the object in this area. However, in low-light conditions, such as at night, the object-likeness value is reduced, and the object cannot be detected. Therefore, after the camera and LiDAR are aligned in advance, the point cloud data is projected onto this image to supplement the object-likeness information. The projected point cloud is preprocessed by removing ground points using the Random Sample Consensus (RANSAC) algorithm [9] and by using Density-Based Spatial Clustering of Applications with Noise (DBSCAN) [10]. The algorithm removes the point clouds corresponding to the surrounding buildings, and only the candidate objects for detection are extracted. The degree of occupancy of the point cloud in each grid in the image is then determined. Occupancy is defined as the ratio of the pixel on which the point cloud is projected to the surrounding pixels in the small area. For each grid, the degree of occupancy is compared with the object likeness, and if the degree of occupancy is greater than the value of object-likeness, the value of object-likeness is replaced by a degree of occupancy. This allows detection of an object even when the object is obscured in the image, as long as a point cloud data of a certain density is observed. Finally, the object detection results for each frame are output in the same format as in Section 4.1. Therefore, the same detection decision logic as in Section 4.1 can be applied to this method.

### Table 1 Specification of selected LiDAR

| Detection distance | 320 m at 10% Reflectivity |
|-------------------|---------------------------|
| Angular resolution | 500 m at 50% Reflectivity |
| Distance resolution | 15.7 cm (at 300 m) |
5.3 Experiment for performance evaluation of the proposed detection method

5.3.1 Method of experiment

Evaluations of the method’s performance described in the previous section were conducted in the evening and at night on a straight section (approximately 200 m in length) of the test line within the Railway Technical Research Institute. For the test, a device was made that imitated the head of a train. As shown in Fig. 14, the device consisted of a hand-pushed bogie and an aluminum frame, and a camera and LiDAR were installed at a height equivalent to that of the driver’s cab. As shown in Fig. 15, an LED light was installed at a position equivalent to the front light of a train and was always on during the night. All subjects were standing people facing the front.

When the illumination was low before and after sunset, the light was turned off and the dolly was positioned at 200 m from the subject, and the images were repeatedly taken for 10 seconds every minute. During the night, the dolly was positioned between 50 m and 200 m from the subject in 25 m increments for 10 seconds, and the images were repeatedly taken. The series of data taken from 50 m to 200 m was considered to be one trial number. The detection rate was calculated as the ratio of the number of “detection” events in the same way as when the detection judgment described in Section 4.1 was applied to the 10-second shooting data.

Fig. 13 The image of obstacle detection method by camera-LiDAR fusion.

Fig. 14 Equipment used in experiment.

Fig. 15 Scene of experiment at night.

5.3.2 Result of experiment

The relationship between the distance from the subject and the detection rate under illumination of 0.2 to 0.3 lx, which is equivalent to an illuminance with a full moon, for the camera-only method and the integrated camera and LiDAR method respectively. Here, the average values of 20 trials are shown. While the detection rate of the camera alone was 0% at over 75 m from the subject, the detection rate of the proposed method was more than 90% up to a distance of 150 m from the subject and about 45% at 200 m, so that improvement of the detection rate was confirmed.

6. Conclusion

To improve safety and reduce transport disruptions by preventing train collisions, we developed a train frontal obstacle detection method with a camera and LiDAR. A detection area limitation method based on rail extraction was developed to suppress over-detection. We confirmed that the detection area can be set centered on the vanishing point of the rail and that over-detection can be suppressed to less than 2% by limiting the detection area.

We also developed a method to detect obstacles in front of a train using only a camera by applying deep learning object detection. Using the developed method, we confirmed that the detection rate was 98% or higher at a distance of 300 m from the target during the daytime. On the other hand, we confirmed that the detection rate decreases as illuminance falls from evening to night. To complement the method’s nighttime performance, we selected and applied LiDAR. Furthermore, we developed a detection method that integrates the camera and the LiDAR. We have confirmed that the detection of a person in the evening is independent of illumination degradation, and that the camera cannot detect a person at a distance of over 50 m at night, whereas the integrated detection method with LiDAR was able to detect 45% of the human targets at a distance of 200 m. In the future, we will evaluate the performance of the system in a straight section of 300 m or more and the performance when the system is actually installed on a vehicle. The final goal is to establish a technology that can detect objects 600 m away.
Fig. 16  Detection rate of person 200 m ahead at low illumination.

Fig. 17  Detection rate of person by distance at night.

References

[1] Hiraoka, T., “A Driving Support System for Encouraging Drivers to Drive Safely,” *Measurement and Control*, Vol. 51, pp. 742-747, 2012 (in Japanese).

[2] Niwa, J., et al., “The Detection of the Person on the Track by Front Surveillance from the Train based on Single-lens Camera Image,” *The papers of Technical Meeting on TER*, TER-17-020, 2017 (in Japanese).

[3] Nakasone, R., et al., “Frontal Obstacle Detection using Background Subtraction and Frame Registration,” *RTRI Report*, Vol. 31, No. 3, pp. 11-16, 2017 (in Japanese).

[4] Cabinet Office, “Trends in railway traffic accidents,” *Traffic Safety White Paper*, pp. 156-158, 2020 (in Japanese).

[5] MLIT Railway Bureau, *Explanation: Technical standards for railway (8th Edition)*, Japan Train Operation Association, 2020 (in Japanese).

[6] Nagamine, N., et al., “People Detection Method using Deep Learning by Monocular Train Frontal View Camera,” *The papers of Technical Meeting on TER*, TER-19-025, 2019 (in Japanese).

[7] Bochkovskiy, A., et al., “YOLOv4: Optimal Speed and Accuracy of Object Detection,” arXiv: 2004.10934, 2020.

[8] Technical Information Institute, *Development of sensor fusion technology and application examples (1st Edition)*, 2019 (in Japanese).

[9] Fischler, M.A., et al., “Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography,” *Commun. of the ACM*, Vol. 24, No. 6, pp. 381-395, 1981.

[10] Ester, M., et al., “A density-based algorithm for discovering clusters in large spatial databases with noise,” *Proc. KDD 1996*, pp. 226-231, 1996.

Authors

**Ryo KAGEYAMA**
Researcher, Image Analysis Laboratory, Information and Communication Technology Division
Research Areas: Computer Vision, Image Processing, 3D Point Cloud Processing

**Hiroki MUKOJIMA**
Researcher, Image Analysis Laboratory, Information and Communication Technology Division
Research Areas: Computer Vision, Image Processing

**Nozomi NAGAMINE**, Ph. D.
Senior Chief Researcher, Head of Image Analysis Laboratory, Information and Communication Technology Division
Research Areas: Computer Vision, Image Processing, Signalling Systems