

Virtual-Reality-Based Training System for the Improvement of the Driver's Ability to Predict Dangers in Driving Situations

Hiroto Suto¹, Xingguo Zhang², Xun Shen², Pongsathorn Raksincharoensak² and Norimichi Tsumura³

1. Department of Imaging Science, Chiba University, Chiba 263-8522, Japan

2. Department of Mechanical Systems Engineering, Tokyo University of Agriculture and Technology, Tokyo 183-8538, Japan

3. Graduate School of Engineering, Chiba University, Chiba 263-8522, Japan

Abstract: Improving the danger prediction during driving can significantly reduce the risk of accidents. However, previous danger prediction training systems had not been sufficiently effective owing to the lack of realism. In this study, we propose an immersive training system for danger prediction training using virtual reality (VR) technology. This system provides drivers with a highly realistic training environment with 360° videos viewed with VR goggles. Users can practice various dangerous scenarios in an environment that simulates a real-driving situation. In addition, we introduced a system to select dangerous spots with a controller and implement training schemes on a voluntary basis. This setup enables them to train in a highly interactive state. In addition, we proposed a method to express multiple indices numerically so that users can understand the training effect. We tested the effect of the system on the danger prediction abilities of various users with two experiments by using this approach. These results show that our system was more effective in improving the driver's danger prediction ability than previous systems.

Key words: VR, VR goggle, traffic accident, danger prediction, 360° video, interaction.

1. Introduction

Approximately 1.35 million people die each year worldwide as a result of traffic fatalities [1]. The main causes of road accidents in many countries are attributed to violations of the law. However, the most significant factor related to these accidents is the driver's failure to recognize dangerous factors [2, 3]. In particular, safety and careless movement were responsible for the driver's inability to anticipate and understand dangerous locations. Safety is a factor involved in accidents caused by failure to adequately check the safety of the vehicle even though the vehicle was stopped or slowed down. Conversely, inattention to movement is a factor responsible for accidents to movement caused by the lack of attention to the

movement of vehicles and pedestrians, even if the driver was aware of them.

In addition, the driver's reaction time when the driver was anticipating the danger was approximately half as long compared with the case at which the driver did not anticipate the danger [4, 5]. This suggests that it is important for drivers to anticipate the dangers of driving in advance to reduce the number of accidents. Therefore, to reduce traffic accidents, it is necessary to improve the driver's ability to predict danger.

To achieve this, a danger prediction training system was developed. This system was used to improve the driver's ability to predict the danger spots on the traffic scene, while the user watched the projected video of this scene. However, there are two major problems with previous systems. Most of the previous systems used animated videos [6, 7] or live-action videos displayed on a two-dimensional monitor [8, 9], but they did not yet reproduce the realistic sensation of

Corresponding author: Hiroto Suto, master, research field: computer vision.

*Part of this research is presented in Electronic Imaging 2020.

actual driving. However, most of these systems only allow users to view the traffic scene, and they do not allow users to interact sufficiently. In addition, a driving simulator was introduced for driver training. However, this requires expensive equipment and is not suitable for ordinary people to train easily.

In this study, we propose an immersive training system to enhance the driver's ability to anticipate dangers and to raise their awareness of safe driving. This system is based on virtual reality (VR) technology and provides a more realistic training environment for users. In addition, this system is equipped with a controller that allows users to interactively identify the danger zone. This system requires only a computer, VR goggles, and a controller. Therefore, it can be introduced more easily than a driving simulator.

In addition, we performed a validation experiment on our proposed system. This result shows that the proposed system is more effective in improving the danger prediction ability compared with previous systems.

2. Previous Systems

The danger prediction training system was developed as a method to improve the driver's ability to predict danger. The objective of this system is to improve the driver's ability to anticipate danger in advance by experiencing dangerous or near-dangerous scenes with near-live action videos. At this time, the danger prediction training system for drivers has penetrated that it has actually been made available to the public. Conversely, these systems have two problems in terms of how realistic the videos are and the interactivity with users.

Honda Motor Co., Ltd. developed a danger prediction training system [6] that allows all people to use it online. The system was designed to automatically pause the video when a dangerous part appears on the screen, and the user selects the dangerous spot. This system is based on our method. However, the problem is that it is very different from the actual driving environment because it uses animated videos. As an approach to this problem, Suto et al. [8] and Shimazaki

et al. [9] developed systems that projected a live-action video on a tablet and allowed the user to select a dangerous spot. However, these systems still lack a sense of presence compared with the real driving environment because training was conducted while images were projected on a two-dimensional planar monitor. To solve this problem, we developed a danger prediction system with VR technology. It was designed to simulate a 360° video on the VR goggles to allow the user ability to predict danger while sitting in a real driver's seat.

The Japanese Automobile Federation has published a training system that uses live-action video of near-danger situations online [10]. This provides a high-level of realism based on actual, dangerous situations. Conversely, there is a problem in that the system is unidirectional, wherein the user only sees videos. A system that introduces VR goggles has also been published. However, it is a unidirectional system, and we have not been able to modify it to enhance the user's positive attitude. Our system solves this problem by introducing a controller. Incorporation of the controller allows the user to aggressively search for the dangerous spot in a 360° video.

In addition, a driving simulator is mentioned as a system similar to the danger prediction training system. This system is a combination of a monitor that displays images and a space that replicates the driver's seat and steering wheel. It allows the driver to simulate driving while viewing animated or live-action videos. In addition, several studies have been conducted on the application of VR [11, 12]. However, the system requires a driver's seat, steering wheel, and a monitor that can span the field-of-view. Our system requires only a computer, VR goggles, and a controller. It is easier to introduce it compared with the driving simulator.

3. System Overview

3.1 Overview

In this research study, we developed a system to improve both the realism and user interaction of a

danger prediction training system. To achieve this, we propose a system that uses 360° videos, VR goggles, and a controller.

First, an omnidirectional camera was placed in the driver's seat to capture the driving video for use as a dangerous scene video. At this time, we have selected near-dangerous scenes from the videos we had acquired.

Videos were subsequently projected in the VR goggles, wherein video scenes could be paused. Correspondingly, the danger spot was set up so that the user could operate it with the controller. The system was configured to repeat the process until the correct spot was selected. The system was configured to repeat the process until the correct area was selected. When this was achieved, the system displayed a video of the dangerous scene and an explanation for it.

We have also implemented a system to record and display the reaction time and number of times to select dangerous spots as a score to quantify the ability to predict the danger. After training with this system, users can evaluate their scores and understand their abilities to predict danger quantitatively.

In this study, we describe the operation of this system in Section 2, while the structure of this system—including the video recording and projecting method—is described in Section 3. The point conversion for quantifying the reaction time and number of selections is outlined in Section 4.

3.2 System Behavior

Fig. 1 shows the main flow, Fig. 2 shows our system setup and Fig. 3 shows the main mechanism for selecting the dangerous spot in this system. First, we explain the operation to select the dangerous spots.

When the system begins to operate, an explanation of the traffic situation is displayed. By letting the user understand the traffic situation in advance, it makes the driving situation in the video clearer. The purpose of this is to create a situation that simulates actual driving. The video is played at the time the user clicks a button on the controller. A few seconds after the video is played, a dangerous traffic scene appears in the video. The user then moves to the next step pertaining to the selection of a dangerous spot. This was achieved in two steps. The first step was the selection of a dangerous scene from the traffic scene that was played, and the second step was the selection of a dangerous spot from the paused screen. The user clicks the trigger button on the controller shown in Fig. 4 at the time at which the user senses the danger. By doing so, the video can be paused. This action allows the user to check if the traffic scene at the selected time is dangerous. If it is a dangerous scene, the user moves to the step of selection of the dangerous spot in a paused traffic scene. When the user selects the dangerous spot, the controller moves in the same way, the selection can be made by moving a red pointer to the spot that needs to be selected. If the most dangerous part is selected, a video and an explanation of the dangerous scene are displayed. Conversely, if the video is paused in a non-dangerous scene, and a non-dangerous spot is selected, the user is instructed to search for the selection of the dangerous scene from the beginning of the video again. The user must repeat this action until the most dangerous spot is selected, that is, until the correct answer is identified. The user will acquire the ability to predict dangers by learning the existence of dangerous spots in traffic scenes.



Fig. 1 Summary of the main flow of our system.



Fig. 2 Our system setup.

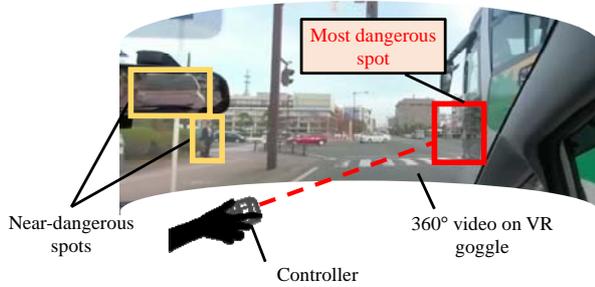


Fig. 3 Schematic explaining the mechanism of our system.



Fig. 4 Controller for operation.

In addition, we recorded the reaction time and the number of times the user selected a dangerous point. This result can be confirmed directly by the user after the training. The details are explained in Section 4.

3.3 Structure of the System

A 360° camera was installed in the driver's seat to capture the 360° videos used in this system. This camera was installed at the same height position as the driver's eyes to be closer to the driver's actual field-of-view to assist watching the video with the VR goggles. The resolution of the video was $3,840 \times 1,920$. In this system, we asked drivers to drive several times on a street in urban or residential areas, and selected dangerous or near-dangerous scenes from the recorded videos.

We built this system with Unity (2018.2.0 Beta 10). To project videos, we created a virtual spherical object and pasted the image inside it, as shown in

Fig. 5. Users can watch videos projected inside the sphere, and 360° videos through VR goggles by installing a virtual camera in the center of the sphere. At this time, we used the Oculus Rift for the VR goggles.

We achieved the selection process of the danger location by the controller by placing the object that incorporated the selection and discrimination process inside the virtual sphere that projected the image shown in Fig. 5. In addition, this object was set to move or stay according to the movement of the video. The user can train the danger prediction ability by selecting the most dangerous spot based on the setting of three selectable spots for each traffic situation.

This most dangerous spot was set based on data relevant to traffic accident causes [3]. Its details are described in Section 4.

3.4 Function Used to Display Results and Point Conversion

This system has a function to display results so that the user can understand the training effect quantitatively. This function enables the user to objectively understand the change in the danger prediction ability by displaying the transition in the training repetition on the VR goggles in the form of a graph, as shown in Fig. 6. In this graph, we can see the transition of the individual as in Fig. 6a and the results compared with the average data from other users as in Fig. 6b. The horizontal axis of the graph shows the order of the traffic situations in which the training was performed, and the vertical axis is the value of the multiple evaluation factors converted as a score. We explain the method of score conversion as follows.

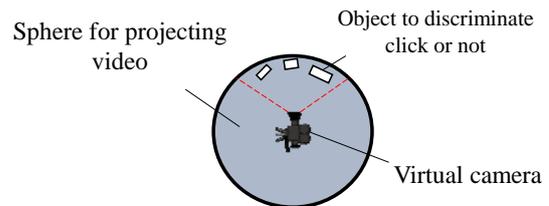


Fig. 5 Virtual environment used for projection.

Virtual-Reality-Based Training System for the Improvement of the Driver's Ability to Predict Dangers in Driving Situations

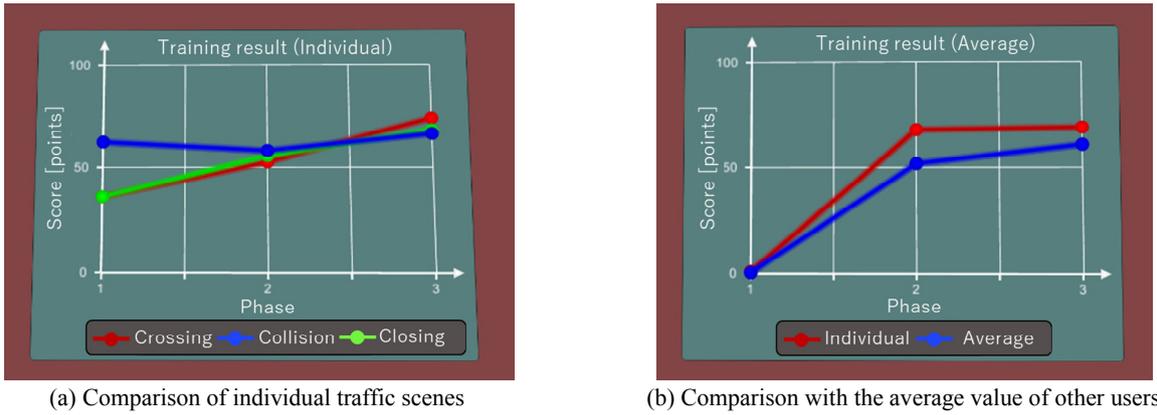


Fig. 6 Training result graphs (example).

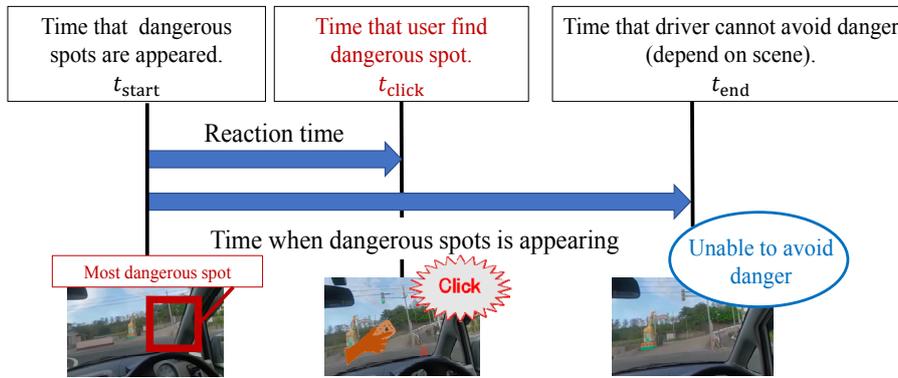


Fig. 7 Danger reaction time.

In this system, we assessed four items: reaction time for a dangerous scene, search time for a dangerous spot, the number of erroneous selections of the dangerous scene, and the number of erroneous selections of the dangerous spot, as the necessary factors for quantitative evaluation of the training effect. These factors are based on the evaluation factors used by Suto et al. [8]. Each element should be set up to represent the maximum number of n points.

As shown in Fig. 7, the reaction time for a dangerous spot is the time from the time the dangerous scene is displayed on the screen until the instant the user notices the dangerous scene and pauses the video, as shown in Fig. 7. However, this measurement time can not be directly converted to several points because the time required for the danger point to appear on the screen differs for each traffic situation. Therefore, it is necessary to modify the system so that the reaction time can be treated in the same way even in traffic

scenes where the time of appearance of the dangerous spot is different. As shown in Fig. 7, if the time for the appearance of the dangerous spot on the video is t_{start} [s], the time for the danger to become unavoidable is t_{end} [s], and the time for the user to notice the dangerous scene and pause the video is t_{click} [s]. The reaction time to the danger converted as a number (P_1) is then calculated as follows:

$$P_1 = \left(1 - \frac{t_{click} - t_{start}}{t_{end} - t_{start}}\right) \times n \quad (1)$$

The search time of the dangerous spot is the measured time from the time that the user notices the danger scene and pauses the video to the time that the user selects the dangerous spot. Let the time that the user selects the dangerous spot be t_{search} [s], and the maximum time required to search for the dangerous spot be t_{max} [s]. The search time for the dangerous spot as a number (P_2) is then calculated as follows:

Table 1 Point convention table listing the number of times that the dangerous spot was selected incorrectly.

Number of mistakes	Score points
0	25
1	15
2	5

$$P_2 = \left(1 - \frac{t_{\text{search}}}{t_{\text{max}}}\right) \times n \quad (2)$$

The number of times that the user made a mistake in the selection of the dangerous scene is the number of times that the user could not identify the dangerous scene or made a mistake. This system was designed to make the user repeat the training when the user made a mistake in the selection of a scene. Therefore, we measured the number of repetitions. Let R be the number of times that the training was repeated R . The number of times the training was selected erroneously as a dangerous scene as a number (P_3) was calculated as follows:

$$P_3 = n - R \quad (3)$$

The number of times that the dangerous spot was selected incorrectly was the number of times that the user selected other spots before the selection of the most dangerous spot. The point conversion of this number of times was based on Table 1. In this system, the full score was set to 25 points. Thus, if the selection was never erroneous, we set the score to 25 points. If the selection was erroneous once, the score was 15 points, and if it was erroneous twice, the score was 5 points.

The training effect can be expressed numerically and presented in a graph format based on the calculation of the total value of the aforementioned four elements obtained by the point conversion process. In addition, we set the reaction time and the search time for the dangerous spot only when the point with the most dangerous spot was selected (as a correct response), and the measurement results were updated when the user selected the erroneous dangerous scenes or spots.

4. Experiment

We verified the training effect of our system based on experiments. We performed two types of

evaluations: one was quantitative and was based on the conduct of repeated training experiments with our system, and the other was subjective achieved by training the users with our system and a previous system that used two-dimensional (2D) videos.

4.1 Experiment 1: Quantitative Evaluation

The purpose of this experiment was to test whether the risk prediction ability of our system was changed by its continuous use. The experiment was conducted in nine traffic situations with three different traffic scenarios, each with three different videos. The experiment was divided in three stages, and in each stage, three different traffic scenes were reproduced once. The order in which the videos were played was randomly set, but the conditions for playing three different traffic scenes, one for each stage, were fixed. To evaluate these quantitatively, we calculated four factors described in Chapter 3, which are the reaction time for a dangerous scene, search time for a dangerous spot, the number of times of the erroneous selection of the dangerous scene, and the number of times of the erroneous selection of the dangerous spot, and evaluated the total score of each factor. In this experiment, the maximum number of points n for each element was 25, and the total value of all the points was set to 100. In addition, the maximum time needed to select the dangerous spot was set to 30 s. This time, we used the following three types of traffic situations based on data from the National Police Agency [3]. These situations are also shown in Fig. 8.

- “Pedestrian crosses the road” (the most common cause of collisions between people and vehicles);
- “Collisions with vehicles passing in front or behind” (the second most common cause of collisions between people and vehicles);
- “Driver’s car is very close to the car in front” (the most common cause of collisions between vehicles).

The experiment was conducted with nine male volunteers who had obtained a driver’s license. They have a range of 21 to 29 years old and the average age

of them was 23.3 ± 2.3 years. We explained the operation of the system to the subjects before the experiment and asked them to perform the selection operation afterward to familiarize with the system. This traffic scene was not used in the experiment. And this experiment was conducted after being reviewed by Ethical Committee of our university.

4.2 Experiment 2: Subjective Evaluation

In this experiment, we asked subjects to use our system with 360° videos in conjunction with a previous system that used 2D videos to investigate the differences in the training effects with and without the use of 360° videos. After Experiment 1, we asked subjects to use a system without 360° videos. Afterward, subjects were asked to answer a questionnaire. These were then evaluated. Similar to

Experiment 1, three different situations with nine traffic scenes were used in Experiment 2, but videos of the traffic scenes were different from those in Experiment 1. The videos used in this experiment were captured from a dashboard camera during driving. The 2D videos in this experiment were viewed with VR goggles.

There were two types of questionnaires: one was a multiple-choice questionnaire, and the other is an open-ended questionnaire based on the rating scale method. In total, there were 29 questions. In addition, there will be two types of multiple-choice questions: one that compares our system with 360° videos to a previous system without 360° videos based on a 4-point scale, and the other that only uses our system on a 5-point scale. Nine volunteers participated in Experiment 1.

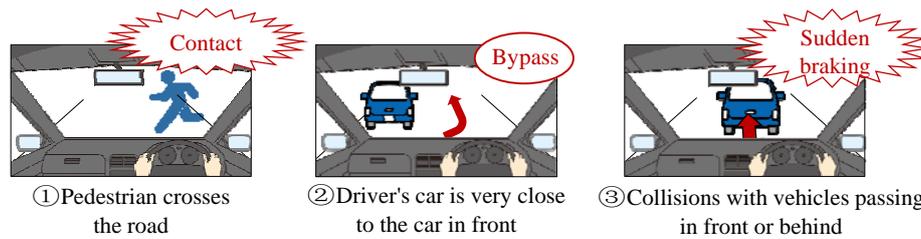


Fig. 8 Traffic scenes using the experiment.

Table 2 Percentage of subjects who chose our system.

Questions	Percentage of people who chose systems (%)			
	360° video	Either, prefer to 360° video	Either, prefer to 2D video	2D video
Which one was more likely to notice the danger?	33.3	11.1	22.2	33.3
Which one was easier to watch videos?	11.1	22.2	55.6	11.1
Which one felt closer to actual driving?	100.0	0	0	0
Which system was easier to use?	22.2	22.2	33.3	22.2
Which system more actively searched dangerous spots?	77.8	22.2	0	0
Which system was more focused during the training?	66.7	22.2	11.1	0.0
Which were the more memorable dangerous scenes?	44.4	55.6	0	0

Table 3 Percentage of subjects who agreed to the questions.

Questions	Percentage who agreed (%)				
	Strongly agree	Agree	Agree a little	Disagree a little	Disagree
Do you want to use our system regularly at home?	0	11.1	33.3	33.3	22.2
Do you think that our system will improve your ability to predict dangerous spots?	22.2	55.6	22.2	0	0
Do you feel that your awareness of predicting dangers has been increased after using our system?	11.1	77.8	11.1	0	0
Do you feel a sense of presence when you use our system?	55.6	33.3	11.1	0	0

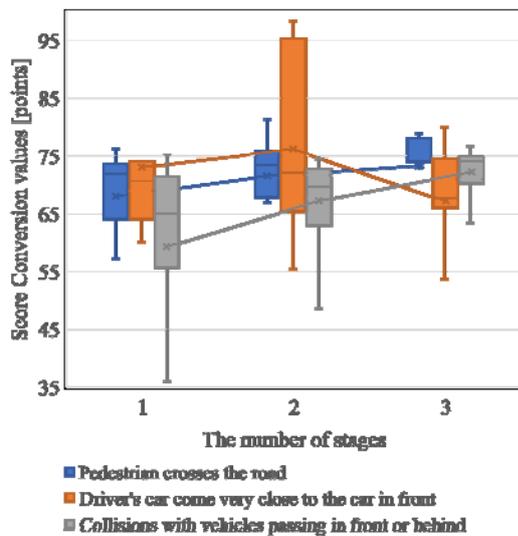


Fig. 9 Average and median values of danger prediction ability score with our system.

5. Results

5.1 Experiment 1

The results of the experiments are shown in Fig. 9. It is a box plot that includes the average and median values of the nine subject scores for each traffic situation. The horizontal axis shows the number of stages of the presented traffic levels, and the vertical axis shows the score conversion values of the evaluation elements. These average and median scores are increased by repeating trains, in two situations: “pedestrian crosses the road” and “collisions with vehicles passing in front or behind”.

5.2 Experiment 2

We present the results of a questionnaire comparing our system with 360° videos and the previous system (which did not use 360° videos) in Table 2. Some of the results of the questionnaire on our system are shown in Table 3. In more than half of the cases, our system with 360° videos was more realistic than the previous system.

6. Examinations

6.1 Experiment 1

Our system can be used to train people to predict danger because the number of points scored in two of

the three traffic situations increased. It was thought that subjects can learn the dangerous spot efficiently by selecting the dangerous spot, and by reading the explanation. The speed and accuracy of detection of the dangerous spots is likely to be improved by repeating this training behavior.

Conversely, the number of points decreased when the “Driver’s car came very close to the car in front of it.” This was mainly owing to the problems in the traffic scene videos used. It was difficult to capture traffic scenes in which subjects approached a vehicle in front of them at normal driving conditions, and it was difficult to use videos in the cases that the subject felt dangerous. Therefore, it was difficult for the subjects to predict the danger in this situation, and they could not practice on predicting dangers well. To create more dangerous videos, it is necessary to recreate dangerous scenes using stuntmen, and to collect 360° videos of actual, dangerous scenes.

6.2 Experiment 2

In more than half of the questions about the effectiveness of our system, more than 80% of subjects felt that our system was more realistic and independent compared with the previous system. This suggests that the use of 360° videos can train the ability to predict danger more effectively than the previous system.

In addition, many subjects pointed out the features of our system, such as low resolution, difficulty in seeing owing to the VR goggles, ambiguity of dangerous spots, and VR sickness. The problems of low resolution and viewing difficulties could be solved by installing VR goggles and a 360° camera that can capture high-resolution videos. The problem of ambiguity should be improved by capturing the video for dangerous situations. The problems related with VR sickness can be mitigated by reducing the use of right-left-turn situations and by prompting participants to remove their VR goggles once during the training. By making these improvements, we believe that we can create a system that is easier for users to use.

7. Conclusion

In this research, we developed a danger prediction training system for drivers with 360° videos, VR goggles, and a controller, to enhance the realism and user interaction training to predict dangerous spots. In addition, we verified the effectiveness of the system based on experiments. The results of the experiments showed that the ability to predict danger was improved by repeated training using our system in many traffic situations. In addition, the results showed that the presence and the user's positive attitude were better than those of a previous system.

In the future, it will be necessary to improve this system to make the danger prediction more realistic and less ambiguous by capturing video footage with stuntmen, and by collecting actual scenes of danger.

Acknowledgement

This work was supported by JSPS KAKENHI Grant Numbers 18K13237.

References

- [1] World Health Organization. 2019. *Global Status Report on Road Safety 2018*.
- [2] National Highway Traffic Safety Administration. 2019. *Traffic Safety Facts*.
- [3] National Police Agency. 2020. *Traffic Accident Statics Annual Report*.
- [4] Johansson, G., and Rumeur, K. 1971. "Driver's Brake Reaction Times." *Human Factors* 13 (1): 23-7.
- [5] Green, M. 2000. "How Long Does It Take to Stop? Methological Analysis of Driver Prediction-Brake Times." *Transportaion of Human Factors* 2 (3): 195-216.
- [6] Honda Motor Co., Ltd. 2020. *Honda Driving Safety Promotion Activities 2019*.
- [7] Felez, J., Carlos, V., and Marti'nez, M. L. 1998. "Virtual Reality Applied to Traffic Accident Analysis." *Computer Networks and ISDN Systems* 30: 20-1.
- [8] Suto, T., Saruta, K., Terata, T., and Guoyue, C. 2014 "Development and Evaluation of Driver Education System Using Tablet PC." *Information and Communication Engineers and Information Processing Society of Japan* 13 (3): 49-54.
- [9] Shimazaki, K., Mishina, M., Nakamura, A., Takahashi, A., and Ishida, T. 2012. "Driver Education Using a Tablet Device and Movies of Accidents Recorded by Drive Recorders." *Traffic Phycology* 28 (1): 35-43.
- [10] Japan Automobile Federation. 2019. *About JAF 2019*.
- [11] Mourant, R. R., and Schulthesis, M. T. 2001. *A HMD-Based Virtual Reality Driving Simulator*.
- [12] Bayarri, S., Fernandez, M., and Perez, M. 1996. "Virtual Reality for Driving Simulation." *Communications of the ACM* 39 (5): 72-6.