https://doi.org/10.2352/ISSN.2470-1173.2020.13.ERVR-380

Development and evaluation of immersive educational system to improve driver's risk prediction ability in traffic accident situation

Hiroto Suto, Department of Imaging Science, Chiba University, Chiba, Japan
Xingguo Zhang, Department of Mechanical Systems Engineering, Tokyo University of Agriculture and Technology, Tokyo, Japan
Xun Shen, Department of Mechanical Systems Engineering, Tokyo University of Agriculture and Technology, Tokyo, Japan
Pongsathorn Raksincharoensak, Department of Mechanical Systems Engineering, Tokyo University of Agriculture and Technology, Tokyo, Japan
Norimichi Tsumura, Graduate School of Engineering, Chiba University, Chiba, Japan

Abstract

Improving drivers' risk prediction ability can reduce the accident risk significantly. The existing accident awareness training systems show poor performance due to the lack of immersive sense. In this research, an immersive educational system is proposed for risk prediction training based on VR technology. The system provides a highly realistic driving experience to driver through 360 degrees video using VR goggle. In the nearly actual driving scene, users are expected to point out every potential dangerous scenario in different cases. Afterwards, the system evaluates users' performances and gives the corresponding explanations to help users improve safety awareness. The results show that the system is more effective than previous systems on improving drivers' risk prediction capability.

1. Introduction

Traffic accidents causes about 1.2 million fatal accidents and more than 50 million injury accidents every year. Traffic accidents are mainly caused by drivers in violation of laws and regulations. However, most drivers were not intended to violate the laws and regulations. Almost 70% of total accidents were due to recognition failures of driver's risk factors which is presented in [1] and shown in Figure 1. If the driver can improve the ability of predicting danger, the reaction time to the danger will be nearly shorten to half as shown in Figure 2. and discussed in [2]. With shorter reaction time, the accidents will be more likely prevented. Thus, it is necessary to improve the driver's risk prediction ability for danger. In addition, the driver habit also affects the risk of traffic accidents. Dangerous driving habit dramatically increase the risk of traffic collision [3]. It is therefore important to train the driver's driving behavior towards safety driving which can reduce the risk of traffic accidents. Obviously, if the driver is able to predict danger in advance, he or she will pay attention to the risk and take safe driving action with more time.

This paper proposes an immersive educational system for risk prediction training, intending to enhance drivers' prediction ability towards danger and their perception of safety driving. The system is constructed based on VR technology, providing a nearly realistic driving experience to driver. Users should point out all potential dangerous scenarios in various situations. Then, the system gives users their ability and personalized instruction for improving their safety awareness.

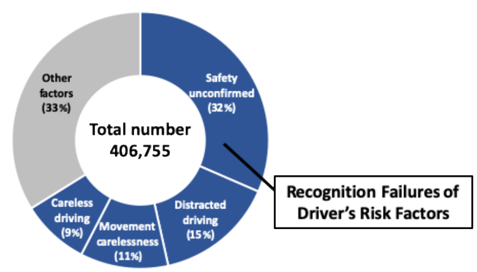


Figure 1. *Number of traffic accidents by driver in violation of laws and regulations*

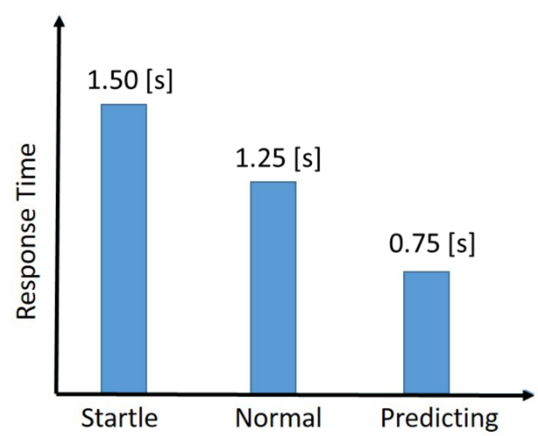


Figure 2. *Reaction time from notice of danger to action to avoid accidents*

2. Objective

Currently, a risk prediction system for drivers has already been developed, and the system is also provided to the general public. Honda Motor Co., Ltd. has developed "Risk Prediction Training System" [5] and published it on the Internet. In this method, it is trained by watching traffic scene images displayed on the computer screen as animation. This is a mechanism in which the video is automatically stopped when a dangerous spot exists, and user selects a dangerous spot from three candidate spots on the stopped dangerous scene. When an appropriate spot is selected, animated dangerous scene is played, and user can predict dangerous scene.

Suto *et al.* have also developed “Driver educational system using a tablet PC” [6]. This is a training system while viewing live-action camera videos. User can recognize pedestrians who appear on the videos and touch the dangerous point. This system can be trained in more dangerous situations.

However, these systems have a problem that realism is low during actual driving. Since the system developed by Honda Motor Co., Ltd. uses animation videos, the situation is different from the actual driving, and the realism is low. The system developed by Suto *et al*. is used by live-action videos, but the video is also displayed on a plane monitor, which makes it less realism than in actual driving.

Another problem is that user's interaction is low. The system developed by Honda Motor Co., Ltd. has a mechanism that automatically stops the video when a dangerous spot is appeared. For this reason, user usually don't need to actively find a dangerous spot. Therefore, we propose a new system that aims to improve both realism of the system and user's interaction.

3. The proposed system

3.1 Features representation

In this paper, we propose a system that improves both the realism of the system and user's interaction by using 360 degrees video and a controller for operation. This system is used by 360 degrees video instead of an animated video or a video displayed on a plane monitor and can create realism during actual driving. It is possible to watch the video in a state close to actual driving because this system moves the content displayed in VR goggles with the rotation of the head. Also, it is possible to train in a wide variety of danger areas because user can watch a wide range of videos including side mirrors and the lateral direction by moving the neck.

In addition, this system implements a function to operate using a controller, so it can interact with the system instead of only watching the training videos. By pausing the video in a dangerous scene, it is possible to select a dangerous scene from the video that is being played or click to select a dangerous spot. This will give users more interaction to simulate a risk prediction ability needed in actual driving.

3.2 Operation procedure

There are two main steps in the operation of this system: a selection of dangerous spots and confirmation of dangerous scenes. The main mechanism of this system is shown in Figure 3. First, we explain the operation to select a dangerous spot. At first, when the video is played, an explanation about the traffic situation is shown. This is to allow user to predict the danger in a situation close to actual driving by letting user grasp the situation of the traffic scene in advance. The video is played at a time that user clicks a button. A short time after the video is played, a dangerous spot appears in the video. Next, the operation moves to the selection of dangerous spots. In this system, the dangerous spots are selected in two steps. The first stage is an operation to select a dangerous scene from the traffic scene being replayed, and the second stage is an operation to select a dangerous spot from the paused screen. User can pause the video by pressing the trigger button of the controller at the timing when the dangerous spot is supposed to exist. At this time, user can see whether or not the temporarily stopped traffic scene is dangerous. If there is no danger, the system moves to search for the dangerous scene from the beginning again. If it is a dangerous scene, the operation moves to newt step of selecting the dangerous spot from the temporarily stopped scene. To click a dangerous spot, move the controller shown in Figure 4. and align the pointer that appears on the video with the most dangerous spot for user. In this system, three selectable dangerous points are set in one scene, one target with highest danger level, and the other two are the target with lower danger level. If you select the most dangerous spot, a scene that is close to danger or reproduction of the dangerous scene will be played, and at the same time, an explanation will be displayed in the driver's seat. The video is played again from the beginning. Up to here is the flow of movement for one traffic scene. After training in one traffic scene, it moves to the next traffic scene video.

Figure 3. A mechanism of our system

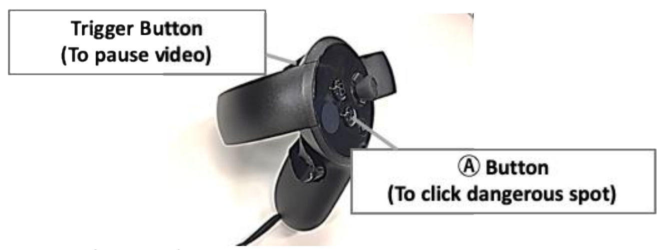


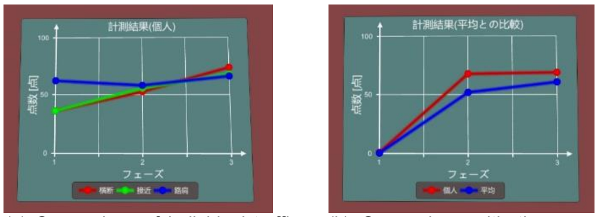
Figure 4. Controller for operation

3.3 The method of video presentation

This system is prepared three types and numbers of traffic scene videos and allows user to experience a total of nine videos. A group containing one type of traffic scene is considered as one phase, and this is prepared for three phases. The videos are all different. All videos will be played randomly. This will avoid the influence of getting used to it on the experimental results. By adopting such a presentation method, it is possible to prevent user's training effect from changing depending on the pattern in which the video is presented and to see the training effect for each traffic scene.

3.4 A function to display results and point conversion

This system implements a result display function for the purpose of allowing user to grasp the training effect themselves. This is because the change when training is repeated, as shown in Figure 5, the experiment results are shown in the VR goggles as a graph. (A) is a graph that allows you to see changes in individual traffic scenes, and shows changes in the risk prediction ability for each type of dangerous scene. (B) is a graph that shows the average value for each phase of all dangerous scenes, and shows the change by comparing the average value of individuals with other users. It is possible to grasp whether the risk prediction ability is sufficiently useful compared with the general public. The horizontal axis of this graph is the traffic scene presentation number, and the vertical axis shows multiple evaluation factors as points. It is a numerical value. Next, this point conversion method is explained.



(a) Comparison of individual traffic scenes *(b) Comparison with the average value of other users*

Figure 5. Output interface of training results

The reaction time for danger measures the time from when a dangerous spot appears on the screen until user clicks the button in a scene that seems to be dangerous and pauses the video as shown in Figure 6. However, since the time at which the danger appears on the screen varies from scene to scene, this measured time cannot be converted directly into points. For example in figure 6, the timing at which the dangerous spot appears and user discovers the danger and pauses the video are the same timing, but the timing at which the danger cannot be avoided is different. This time, although the reaction time is equal, the timing at which the danger becomes unavoidable is different, so the measured reaction time cannot be treated as equal. So, it is necessary to make a calculation to make the reaction time comparable even in traffic scenes with different times of appearance. If the point of reaction time expressed as a comparable value by converting points is P_1 points, the time at which the location appears on the video and the danger can be predicted is t_{start} [s], the time when the danger is unavoidable is t_{end} [s], and the time until user finds the dangerous spot and pauses the video is t_{click} [s] as shown in Figure 6. and the full score is n points, the conversion of the reaction time for the danger is calculated as shown in the equation below.

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

As a result, it is possible to compare even when the time of occurrence of danger is different, and to handle the measurement result as a numerical value. In this system, 25 points are set as the perfect score for the reaction time, so 25 is substituted for n.

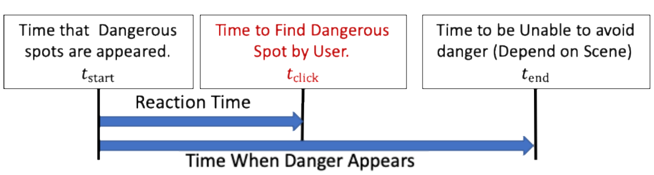


Figure 6. Risk predictable time

The searching time for a dangerous spot is a time measured from the moment when the video is paused and move to the operation of selecting a dangerous spot until the dangerous spot is clicked as shown in Figure 6.

If the search time score expressed as a numerical value that can be compared is P_2 points, the maximum search time is t_{max} [s], the time when the most dangerous spot is clicked by user is t_{search} [s], and the full score is if n points is used, the search time is converted to a point as shown in the following formula.

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

The number of mistakes is count of times user mistakes until user selects the most dangerous spot. This point conversion is based on the point conversion table shown in Table. 1. The system assigns 25 points to the perfect score, so there is no mistake is 25 points, one mistake is 15 points and two times mistake is 5 points. That is set to be lower 10 points at a time.

Table. 1 Point convention table for the number of times the uncorrected spot is mistaken

Number of mistakes	Score point
0	25
1	15
2	5

The number of repeating is count of times one scene video is repeated when a dangerous location cannot be found or when an attempt is made to pause the video in a scene without danger. If the number of repetitions for point conversion is P_3 points, the number of repeating is R times, and the number of full points is n points, the point conversion of the number of times the traffic scene is repeated is in the below.

$$P_3 = n - R \tag{3}$$

It is possible to show the training effect in the form of numerical values and graphs by obtaining the total value of these four factors obtained by scoring. In terms of reaction and searching time, it is necessary to record the value only when target with highest danger level is clicked. Therefore, when user tries to stop the video in a scene where there is no dangerous part, or clicks parts with a low degree of danger, or cannot find a dangerous place, the results of preservation is not performed. Through the above methods, we can conduct a numerical evaluation of the driver's safety awareness.

4. Experiment

The verification is performed by a quantitative evaluation based on the result of quantification by point conversion. Quantitative evaluation can evaluate the individual training effect for each traffic scene.

In the experiment, three different types of traffic scenes are presented three at a time. For each traffic scene, three virtual objects are placed to click on a dangerous spot. There are four evaluation elements for the experiment: the reaction time for danger, the searching time for a dangerous spot, the number of mistakes for spot selection, and the number of repeating. There are three types of traffic scenes to be used: “Scenes where a pedestrian cross”, "Scenes that the car is too close to the front of car” and "Scenes where the car is parked on roadside". “Scenes where a pedestrian cross” is the most frequent accident caused by contact between the vehicle and others when the safety is not confirmed, and movement is carelessness. Since rear-end collisions are the most common cause of mutual accidents [1], the “Scenes that the car is too close to the front of car” is used. And " Scenes where the car is parked on roadside" is used because it is a place often used in conventional systems. In addition, video is played randomly.

Before conducting the experiment, we explained how to operate the system using different images to subjects, and then practiced with the images of traffic scenes that were not used in the experiment. When the most dangerous part is clicked, an explanation of a dangerous spot is displayed, and the subject is instructed to look carefully, including the explanation. The subjects here are 10 male people in their 20s who have a driver's license.

5. Results

Figure 7. shows the result of the experiment, which shows the average value of 10 subjects for each scene. The data are arranged along the horizontal axis in the order in which they were presented. The vertical axis indicates the score point converted from evaluation effects. From this result, the score point of evaluation effects was increased by training with this system in two of the three types of traffic scenes.

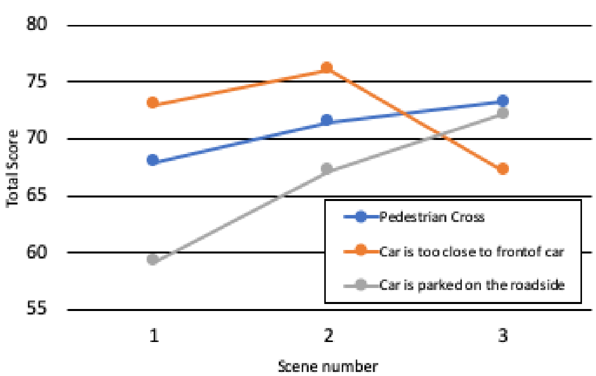


Figure 7. Transition of the averaged score of the measurement results for the three factors

6. Examination and Summary

In this research, we developed a risk prediction educational system for drivers using 360 degrees video, VR goggles, and an input controller for the purpose of enhancing the realism and user’s interaction. In addition, we confirmed the training effect of this system by experiment.

Three types of traffic scenes are used to conduct experiments for verification of the training effect of this system. The scores of evaluation parameters increase after training in the “Scenes where a pedestrian cross” and the “Scenes where the car is parked on roadside”. However, the scores were partially reduced in the “Scenes that the car is too close to the front of car”. Probably, the reason might be that the subject’s concentration feller than other scenes because this scene might be not able to recognize as more danger compares to other scenes.

Since no verification experiment with the conventional method was conducted this time, it is necessary to confirm whether the presence and interaction in this system have improved compared to the conventional method, such as a subjective evaluation experiment using a questionnaire. It is also necessary to increase the types of traffic scenes and the number of scenes and to review the conversion method of evaluation parameters into points.

References

[1] National Police Agency Transportation Bureau , “Status of traffic accidents during 2018”, National Police Agency Transportation Bureau statistics, 2019.

[2] “Risk prediction and accident avoidance training Importance of risk prediction”,(http://www.jaf.or.jp/ecosafety/safety/kyt/important .htm).

[3] Hanae Kosuga, “Relationship between near-miss types and normal driving behavior Driving awareness −Human error analysis for safety education measures consideration−“, Occupational Safety and Health Research, 2018, Vol.11, No.1, p.25-37.

[4] Frank P. Mckenna, Jeffrey Crick, “Developments in Hazard Perception”, Transport Research Laboratory, 1997.

[5] Honda Motor Co., Ltd. “Risk prediction tracing (KYT)”,

[6] Takumi Suto, Kazuki Saruta, Yuki Terata, Guogue Chen, “Development and evaluation of driver education system using tablet PC”, Information Science and Technology Forum Proceedings, 2014, vol..13, no.3, p. 49-54.

Author Biography

Hiroto Suto received a bachelor’s degree from Akita Prefectural University. Currently, He belongs to the Graduate School of Integrated Science and Engineering at Chiba University. The current research field is a material appearance and object texture reproduction using computers.

Xingguo Zhang received the B.E. degree in computer science from the Dalian Polytechnic University of China in 2009, and the M.E. and Ph.D. degrees in Systems Science and Technology from the Akita Prefectural University of Japan in 2012 and 2015. Currently, he is an Assistant Professor with the Department of Mechanical Systems Engineering in Tokyo University of Agriculture and Technology, Japan. His research interests include computer vision, pattern recognition, and machine learning.

Xun Shen received the B.S. degree in Electrical Engineering from Wuhan University, Wuhan, China in 2012, and the Ph.D. degree in Mechanical Engineering from Sophia University, Tokyo, Japan, in 2018. He had been a Post-Doctoral Research Associate with the Department of Engineering and Applied Sciences at Sophia University from April, 2018. Since June, 2019, he has been an Assistant Professor in the Department of Mechanical Systems Engineering, Tokyo University of Agriculture and Technology.

Pongsathorn Raksincharoensak received the B.S. degree in engineering from the Chulalongkorn University, Bangkok, Thailand, in 1999, and the M.S. and the Doctor's degrees in engineering from the Tokyo University of Agriculture and Technology, Tokyo, Japan, in 2002 and 2005, respectively.

He is currently a Professor with the Department of Mechanical Systems Engineering in Tokyo University of Agriculture and Technology. His research interests include vehicle dynamics and motion control of electric vehicle.

Norimichi Tsumura is an associate professor in department of information and image science at Chiba University. He received a PhD from Osaka University and was also visiting associate professor at Dept. of electrical and computer Eng. in University of Rochester. His current research field is mainly color and appearance engineering, affective computing and deep learning.