TOWARD BETTER DRIVING WITH GAZE AWARENESS ENVIRONMENT SUPPORTED BY AREA SEGMENTATION

Taketo Yamada[1], Kenji Matsuura[1], Hironori Takeuchi[1], Akihiro Kashihara[2], Kenichi Yamasaki[3] and Genta Kurita[3]
[1]Tokushima University, Japan
[2]The University of Electro-Communications, Japan
[3]Mitsubishi Precision Co., Ltd., Japan

ABSTRACT

It is important to make car-drivers improve their way of looking for recognizing key objects or areas precisely. This study designs a system following such a motivation that distinguishes several areas in a display with weights of importance. A present proposing function for successful area detection offers drivers an opportunity to compare their gaze with experts. Concrete method for this implementation includes U-Net that is one of major techniques of machine learning combined with grid segmentation.

KEYWORDS

Driving Simulator, U-Net, Grid Segmentation, Eye Tracker

1. INTRODUCTION

It is important for car-drivers to keep both safety way of driving and mature their cognitive ability. More than 40,000 people die and more than one million are injured in traffic accidents each year in European Union member countries. Road fatalities account for only a small number of deaths but are the leading cause of death in developed countries for people under the age of 40 (Plainis et al., 2006). Most automobile accidents are caused by drivers' violation of their duty of safety. In road traffic accidents, 57% are directly caused by human factors and 90% revolved human factors (Green & Senders, 1999). Among other things, findings from the 100-car naturalistic study showed that almost 80% of traffic accidents could be attributed to inattention (Klauer et al., 2006). In fact, it has been argued that young drivers have higher accident rates because of their immature cognitive abilities (Deery, 1999). These facts indicate that drivers may lack the ability to recognize danger objects, routes, places or areas and the awareness of danger avoidance. Thus, it is necessary to approach the internal processing of humans such as cognition and judgment, in order to reduce the number of accidents caused by these factors.

However, it is difficult for learners to judge whether they appropriately recognize and understand the conditions of traffic compared with internal vital state and sensed result of the environment. Even though a driver sees something, sometimes s/he might not consciously recognize it. Therefore, the function that enables a system to determine whether a learner as a driver is aware of the place to be focused and to make an objective judgment to identify a better way for her/his recognition.

Driving a car or a motorbike is available to apply the traditional knowledge in motor skill learning, in which the following processes can be transferred: recognition of external and environmental aspects, correct judgment based on proper recognition (selection of driving behavior), and implement the actual behavior based on prior judgment (operation of a car or a motorcycle). The result of the action is recognized again both externally and intrinsically and reflected succeeding actions during driving.

In the present study, the project is conducted using a driving simulator (DS) with Virtual Reality (VR), rather than learning on the premise of actual car driving. The DS used can change the course with deliberate events according to the driving scenario applied. Our study presents the images of accidents and the driving conditions leading up to them play an important role as human visual input (Yamada et al. 2022).

This study designs an environment to support driving skill improvement using both machine learning and eye tracking technologies. It is assumed that tracked data of eye movements presents high quality of appropriate eye tracking measures. Then, the study analyzes the data using machine learning technique for classifying areas of same semantics and provides a learning support information following the context. However, for the sake of simplicity, the present study starts using only gaze information related to the perception of the outside world, without direct manipulation of the driving object such as accelerating, braking, or steering wheel operation. For example, it is possible to capture how the viewpoint differs from that of skilled drivers and how the viewpoint changes when behavioral change is observed.

2. ANALYSIS BY SCREEN AREA SEGMENTATION

2.1 Analysis of Gaze Coordinates

There are a lot of tools that treat time-series data of eye movements based on (x, y) screen-coordinates. Eye tracking is often performed with either fixed camera or a specific devise of wearing glass during driving. Our environment adopts a DS, in which we can select a fixed camera in front of the screen because the scenery of the outside world can be reflected on it. Actual analysis method for such time-series data evaluates the similarity between two different series, for example, with DTW (Dynamic Time Warping) (e.g. Stana and Philip 2007, Naito et al. 2020) which is available to apply under certain conditions. The advantage of applying DTW is that the analyst only needs to focus on the series data, regardless of object-meanings such as a walker and a car displayed in the background. However, the general viewing angle (e.g., 40 degrees for both left and right side respectively) may not be sufficiently considered from physiological perspective.

In an actual situation for tracking, however, it is necessary to take into account that mixed elements may influence errors with visual processing by humans as well as the processing conditions of the computer, including observation errors. Specifically, humans do not always accurately perceive the smallest unit of (x, y) coordinates corresponding to the screen resolution. It may be, if anything, more natural to perceive the target area or the object as part of a plane. Therefore, we propose a method to divide the entire screen into partial areas.

2.2 Analysis of Gaze Coordinates in a Predefined Area

Gaze information is one of the major keys for solving the traffic accidents (Vicente, F. et al. 2015). A lot of studies introduce machine learning techniques to this domain from technical perspectives (Yoon, H.S. et al. 2019). However, they are still focusing on the technical innovation of precise detection of eyes and therefore we need the applied approach for learning a way of gaze. When drivers learn the better way of looking at several important objects or areas according to the condition of actual road, the driving way including cognition and operation is refined.

In this study, we aim to provide a learning support system that improves the driver's ability to recognize hazardous areas by guiding the driver's gaze through comparison with other gaze data which is obtained from skilled drivers in advance and regarded as a referential model. In this case, it is necessary to develop a system that estimates the degree of hazard discrimination from the gaze data and prompts the driver to pay attention to the hazardous area.

By dividing the entire screen into partial areas of homogeneity based on a latticed pattern, it is possible to determine which area the driver is looking at if the driver's gaze coordinates are available for tracking. However, it is necessary to determine the scale of the unit area occupied and the location of the area that can be considered to correspond to the gaze. Since the DS used in this study introduces VR technology, it is possible to calculate the two-dimensional coordinates of the objects drawn on the DS screen based on the relationship between the position of the object itself and the camera from which it is drawn. However, our system is different from the original DS and the proposed function is independently developed from the DS itself considering its future extension to the real view in real driving.

3. METHODS INTEGRATION AND GAZE EVALUATION

3.1 U-Net

Segmentation is the process of dividing an image into subregions with similar features and meanings. In recent image processing fields, the regions of the target object and background in an image are often segmented and recognized pixel by pixel, and the image cells are classified to simultaneously recognize the region and class of the object in the image. In 2012, a method using feature maps obtained by CNN, which is an abbreviation of convolutional neural network, was proposed (Farabet, C. et al, 2012.), and since then, deep learning has become the mainstream method for image segmentation.

U-Net has a CNN architecture for image segmentation developed by Oraf et al. (2015). The network consists of a contraction path and an expansion path. It is called U-Net because of its U-shaped architecture. In the contracting path, features are mapped by convolution. In U-Net, the same hierarchy of contraction and expansion paths are connected by a mechanism called skip connection, which ensures that the positional relationship of each pixel in the input image is not lost.

As for our actual implementation of U-Net, an original image is converted into 160 × 160 square size at first in terms of a contraction phase. We configured batch size as 4 and 40 epochs. We used *Keras* and *Tensorflow* for the default library of U-Net and we set *Adam* as an optimizer.

This study uses the GTA5 dataset created by Stephan R. et al. (2016). These images are rendered using the open-world video game Grand Theft Auto 5, and are all from the perspective of a car on the streets of a virtual city. The dataset consists of 2,500 original images of 1,914 × 1,052 pixels and 2,500 segmented images, 30% of which are used as test data for comparison and evaluation. Actual number of the test data we used was 750 and the rest was used for train data. The training was performed with a batch size of 4 and an epoch count of 40, and predictions were made for each frame of the DS video.

3.2 Grid Segmentation

In this study, the model gaze data of a skilled driver and the gaze data of a learner are compared for each frame of the DS video using U-Net based area segmentation, and identification ability for danger areas/objects is evaluated by the degree of agreement between the areas segmented with both techniques. However, even if the gaze coordinate in an area of a skilled driver and that of a learner are the same, they may be looking at locations that are quite far apart in terms of coordinates of the same area by U-Net segmentation. In this case, for example, a road spans a wide area from the left end to the right end, and unified area segmentation using simple U-Net would result in a problem where gaze on the extreme left or right are regarded identically. Therefore, in addition to area segmentation by U-Net, the entire image is independently divided into an N × N grid (3 × 3 in this experiment) to subdivide the viewing direction. This process suppresses the problematic patterns and enables highly accurate evaluation.

3.3 Gaze Evaluation

When evaluating gaze using the combination of U-Net and lattice domain segmentation described in the previous sections, we focus on the vertical and horizontal distance traveled to the opponent's gaze based on the position of a learner's or an expert's gaze and add the evaluation points (EP) as shown in Table 1. where N is the number of grid segmentations, and L is the cost of reaching the line of sight from the referential point. For example, if the cell in upper left corner of Figure 1 is the reference and the lower right corner is the same area in U-Net, the EP sets 0, but for adjacent cells, the EP sets 0.5.

Table 1. Evaluation weight points at the time of gaze detection

Grid Segmentation	Segmentation by U-Net	Evaluation Point
Match	Match	1
Match	Not match	0.75
Not match	Match	1-(L/(N-1))
Not match	Not match	0

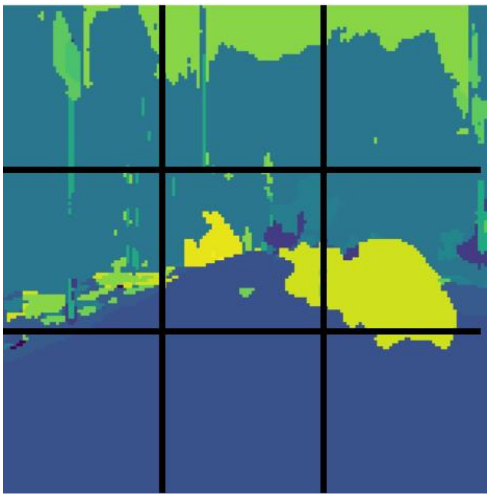

Figure 1. Combined method of area segmentation using U-Net and grid

3.4 Feedback to Learners

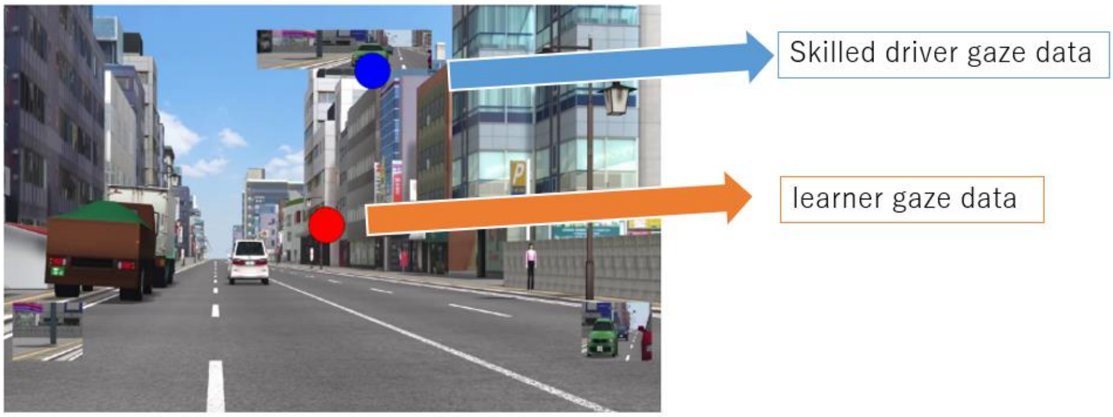

Figure 2. Screen dump; feedback circles on the screen of driving simulator

In this study, after acquiring the gaze information of the learner, the gaze coordinates of the learner and those of the skilled driver acquired beforehand are simultaneously displayed on the screen as a circular area together with the DS video of a background layer, as shown in Figure 2, to support learning how to set the viewpoint on the screen at each frame. In this process, a gradation of colors is applied to the drawing according to the evaluation points in the previous section to make it easier to understand visually which points are different each other. The fundamental idea for gaze improvement is that they can identify how and where the instructor or the expert looks on in case learners are aware of the model view. They can get such information with two circles with independent colors. The color is set based on the previous discussion based on the analysis shown in Figure 1. At that time, the gaze coordinate data of the skilled driver is always drawn in blue. On the other hand, the gaze data of the learner is also drawn in blue, the same color as the gaze coordinate data of the skilled driver when the evaluation score is 1. The color is set as green when the score is 0.75, and yellow when the score is 0.5. When the score is 0, the color is set as red. The difference indicates the gap between them and therefore the learner should learn where and how s/he should look on with such colored circles as feedback.

4. EXPERIMENTAL USE AND THE RESULTS

4.1 Experimental Use

The project conducted an experimental use of this system. The purpose of this experiment was to evaluate the effectiveness of the proposed method for supporting gaze learning and its implemented system. First, five people who drive at least once every two days tried the system as skilled drivers, and their gaze data were obtained through the system. Though we have to provide a model data for reference to subjects, there are several ways for this purpose. Among several options, the model data to be used in the experiment was determined by voting within the skilled drivers at this time. Then 15 men and women ranged in their 20s to 50s whose driving frequency were less than once every two days joined voluntarily as learners and randomly divided into Group A, Group B, and Group C. Group A was the group that used the system (proposed system) in which both the learner's and the skilled person's gaze coordinates were displayed together in different colors according to the evaluation points, and Group B was the group that used the system that displays only the gaze coordinates of the proficient driver. Group C was the group that used the system that displayed only the gaze coordinates of the learner.

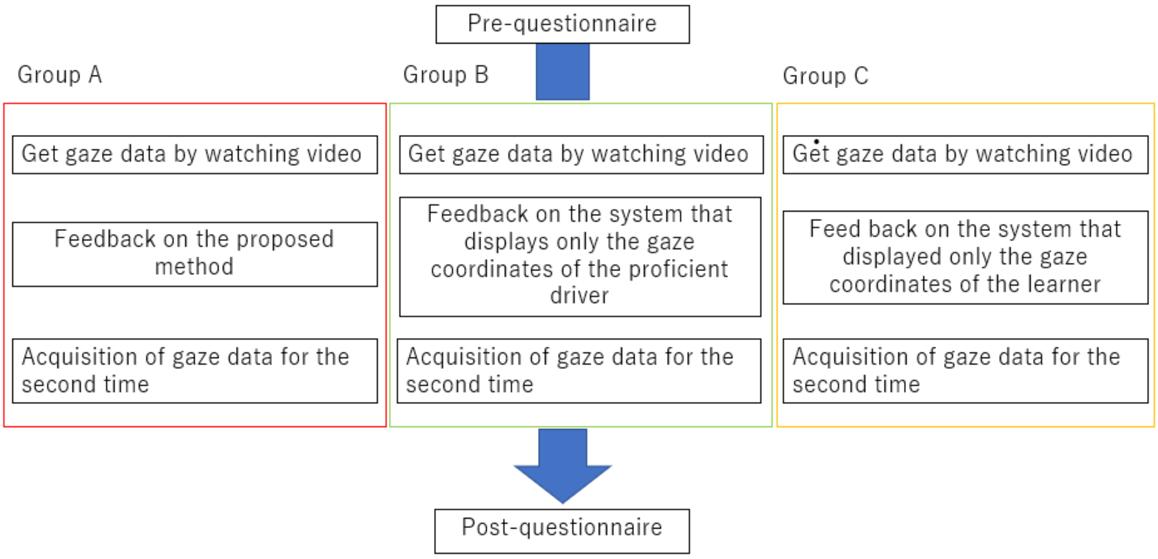
Pre-questionnaire
Group A
Group B
Group C
Get gaze data by watching video
Feedback on the proposed method
Acquisition of gaze data for the second time
Get gaze data by watching video
Feedback on the system that displays only the gaze coordinates of the proficient driver
Acquisition of gaze data for the second time
Get gaze data by watching video
Feed back on the system that displayed only the gaze coordinates of the learner
Acquisition of gaze data for the second time
Post-questionnaire

Figure 3. Flow of the experiment

Experiments on learners were conducted according to the flow shown in Figure 3. First of all, all the learners were asked to answer the questionnaire for the preliminary survey. The contents of the preliminary questionnaire asked about the mileage in the last year, accident history, violation history and so forth. Then, they watched the image of each group in drive simulator and we acquired gaze data. After that, the system gave feedback according to each group property. After the feedback, each group members watched the same video again and acquired the gaze data. The difference in cumulative evaluation scores between the groups was then investigated. In order to investigate changes in learners' behaviors, each group was given a questionnaire with a 4-point scale after the feedback given at each trial on how much they contributed to the learning of their viewpoints.

Table 2 and Figure 3 show the mean difference between the pre- and post-feedback scores of weighted gaze agreement for each group of learners in the experiment, and the box plots with medians and quartiles, respectively. The higher scores indicate the improvement which is the influence of the system contribution. Figure 4 shows the results of a questionnaire in which participants rated on a 4-point scale the extent to which the feedback provided in each trial contributed to the learning of the learner's point of view after each feedback session. The results of the questionnaire after the feedback showed that about 67% of the responses was "Contributed" and about 27% was "Contributed a little" in group A. In group B, 20% of the responses was "contributed" and 67% was "contributed a little". As for the same question, in Group C, about 33% of the responses was "contributed" and about 53% was "contributed a little". It means that positive effects were found dominantly at Group A, Group B and Group C just in order.

Table 2. Mean difference in gaze scores before and after feedback

learners	Group A	learners	Group B	learners	Group C
A1	0.0787	B1	-0.0113	C1	0.0961
A2	0.0471	B2	0.0106	C2	-0.0239
A3	0.1312	B3	0.1869	C3	- 0.0445
A4	0.1587	B4	0.1558	C4	0.0188
A5	0.0625	B5	0.0047	C5	-0.0107
average	0.0956	average	0.0693	average	0.0072

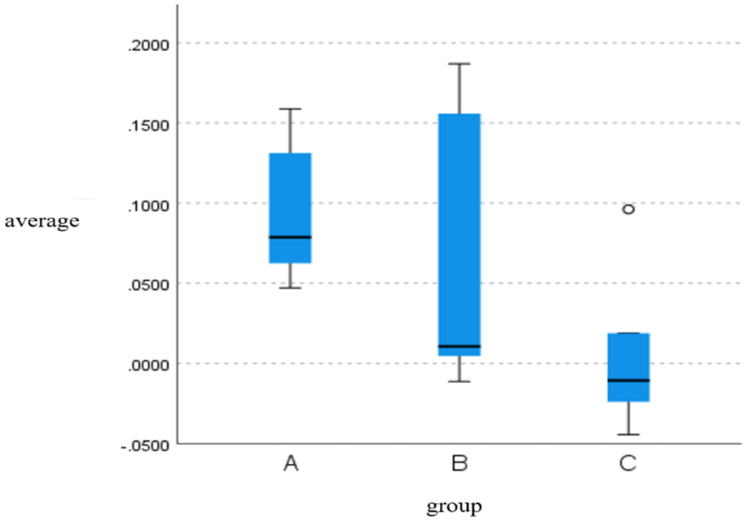

Figure 4. Resulting graph in a box-plot format

4.2 Discussion

The results of the experiment showed that Group A had the highest improvement scores on the average though the highest improvement among all subjects appears in Group B (see Table 2). The median values in cumulative evaluation score were different between the groups, which was seen in Figure 4. The box plot also indicates that the degree of data distribution. Interesting discussion could arise between Group A of comparative middle variation and B of large variation. Our belief is that subjects with our system were influenced by a sound comparison including their own gaze tracking data to some extent while the improvements on subjects using model trajectory without their own gaze had diverse influence. In particular, one of subjects in group B indicated that the behavioral change had been worse through the feedback. In addition, improvements by two subjects, B2 and B5, were almost zero which means no actual improvements found while the other two indicated highly improved.

Next, as the number of subjects were small but when we compare group A with group C, we can find large difference between them in several viewpoints. Therefore, we think the feedback through the gaze-detection with area-segmentation brings positive effects.

If a way of driving operation conforms to the model, the learner gets good feedback but otherwise not. In addition, the results of the questionnaire requested after the feedback showed that Group A had the highest ratio about "contributed" with 67% of the learners. It means that they got positive impression from subjective viewpoint.

5. CONCLUSION

In this study, we designed a support environment that captures and analyzes the driver's gaze during driving using a driving simulator in order to make the driver aware of the areas s/he should pay attention to, and to provide visual feedback. The current proposal focuses just on the simple and solid style on detecting gaze and analysis thereof. In addition, we assume the technical proposal is based on the VR environment. However, we can apply it to the real driving with further improvement on machine learning techniqeus.

Future issues to be addressed implied as follows.

- The environment needs to draw on a larger screen because the current screen size used (12 inches) is small and the entire image can be seen, which is because of the limitation of the eye tracking device, including parts of the image that are not normally visible unless the driver gazes at them when driving.
- We need to improve the accuracy of region segmentation by U-Net. The model data is created from the actual drive simulator videos for training U-Net but we should provide more precise one.
- We have to consider the flexibility in viewpoint evaluation based on congruency. With the current system, grid area is congruent, and the U-Net area is incongruent, the agreement points are not uniformly set to a constant score of 0.75, but are variable based on the area ratio and other factors.
- We can arrange the focus not only on viewpoint information, but also on auditory and other non-visual perceptions.

ACKNOWLEDGEMENT

We would like to thank all members joined at project meetings for their fruitful discussions and Mitsubishi Precision Co., Ltd. for its support.

REFERENCES

Deery, H.A., 1999. Hazard and risk perception among young novice drivers. *Journal of Safety Research,* Vol.30, No. 4, pp. 225-236.

Farabet, C., Couprie, C., Najman, L., & LeCun, Y. (2013). Learning hierarchical features for scene labeling. *IEEE transactions on pattern analysis and machine intelligence*, Vol. 35, No. 8, pp. 1915—1929.

Green, M., & Senders, J. (1999). Human error in road accidents. http://www.visualexpert.com/Resources/roadaccidents.html. (accessed 2022 May 19)

Klauer, S.G., Dingus, T.A., Neale, V.L., Sudweeks, J.D., & Ramsey, D.J. (2006). The Impact of Driver Inattention on Near-Crash/Crash Risk: An Analysis Using the 100-Car Naturalistic Driving Study Data. *National Highway Traffic Safety Administration,* DOT HS 810 594, Washington DC.

Naito, H., Matsuura, K., & Yano, S. (2020). Learning Support for Tactics Identification Skills in Team Sports by Gaze Awarenesss. *Proceedings of IIAI-AAI2020,* Taiwan, pp. 209-212.

Plainis, S., Murray, I.J., & Pallikaris, I.G. (2006). Road traffic casualties: understanding the night-time death toll. *Injury Prevention*, Vol. 12, No.2, pp. 125—138.

Richter, S.R., Vineet, V., Roth, S., & Koltun, V. (2016). Playing for Data: Ground Truth from Computer, *Games* , LNCS 9906, pp.102—118.

Ronneberger, O., Fischer. P., & Brox. T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. avab. N. et al. (Eds.) *MICCAI Part III*, LNCS 9351, pp.234—241.

Stana, S., & Philip, C. (2007). Toward accurate dynamic time warping in linear time and space. *Intelligent Data Analysis*, Vol. 11, No. 5, pp. 561-580.

Vicente, F., Huang, Z., Xiong, X., De la Torre, F., Zhang, W. & Levi, D. (2015). Driver Gaze Tracking and Eyes Off the Road Detection System, *IEEE Trans. On Intelligent Transportation Systems*, Vol.16, No.4, pp.2014-2027, 2015.

Yamada, T., Matsuura, K., Takeuchi, H., Kashihara, A., Yamasaki, K., & Kurita, G. (2022). Learning driving with improving gaze through area division methods, *2021 Shikoku student workshop proceedings of JSiSE*, Japan, pp. 215—216. (In Japanese).

Yoon, H.S., Baek, N.R., Truong, N.Q. & Park, K.R. (2019). Driver Gaze Detection Based on Deep Residual Networks Using the Combined Single Image of Dual Near-Infrared Cameras, *IEEE Access*, Vol.7, pp. 93448-93461, 2019.