A high-speed driver behavior detection deep learning system using the amount of change in contrast between frames

Min Woo Yoo, Jihun Kim and Dae Woong Cha and Woo Sung Son and Donggyu Lee and Dong Seog Han*
School of Electronic and Electrical Engineering
Kyungpook National University
Daegu, Republic of Korea
dshan@knu.ac.kr*

***Abstract*—This paper proposes a deep learning system that detects driver behavior for safe driving. A method of detecting the dangerous behavior of an existing driver uses a method of deep learning object detection that detects a class and a location of an object in an image. However, the deep learning object detection algorithm uses many computational resources, so it cannot be used in vehicle embedded environments with limited computational resources. In the case of an object classification algorithm that classifies a single object in an image, fewer computational resources are used than that of a deep learning object detection algorithm. However, it cannot be applied because various objects in the camera image cannot be classified as a single object. In the paper, We propose an algorithm that infers the driver's behavioral area using the driver's static movement in a vehicle and then applies deep learning objects to the inferred area. The proposed algorithm may be applied to a vehicle embedded environment because the calculation time is faster and more accurate than the deep learning object detection algorithm.**

***Keywords*—deep learning, object detection, classification**

I. INTRODUCTION

The driver monitoring system monitors the driver's condition and alerts the driver to prevent accidents. Recently, deep learning technology has been widely applied to driver monitoring systems. When detecting the driver's face [1] and mobile phone [2] use and smoking [3], a deep learning object detection algorithm is used. Representative deep learning object detection algorithms include SSD [4], YOLO [5], and RCNN [6]. SSD and YOLO are very fast to detect, but RCNN is very slow. A deep learning object classification algorithm is used when classifying the driver's emotions [7] and the driver's gaze [8]. Representative lightweight object classification algorithms include MobileNet [9] and SqueezeNet [10]. The object detection model generally uses more operator resources than the object classification model. In a limited vehicle embedded environment, a deep learning object detection algorithm that uses many computational resources causes system instability. Therefore, it is inappropriate to detect mobile phones and cigarettes using a deep learning object detection model.

The object classification algorithm's existing driver behavior detection algorithm consists of two steps. The first step detects a face area, and the second step is to enter the area of the ear or mouth into the object classification model [11]. However, this method has a problem in that behavior cannot be detected when there are no objects around the mouth and ears. In addition, additional computational resources are consumed because the driver's face must be detected to detect behavior.

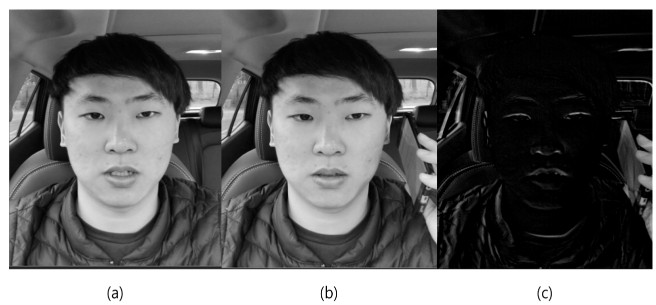

Figure 1. Changes in contrast when a driver uses a mobile phone: (a) Previous frame, (b) current frame, (c) contrast change image

The vehicle interior environment is very different from the general environment. The general environment is very dynamic. On the other hand, the environment observed by the vehicle interior camera is very static. Figure 1 shows the previous and current frames when driving after setting the camera frame to 5 FPS. Figure 1(a) is the previous frame, and figure 1(b) is the current frame using a mobile phone. Figure 1 (c) can be obtained by subtracting Figure 1 (a) and Figure 2(b). Through Figure 1, we confirmed a significant change in the contrast of the behavioral area when the driver acted in a static vehicle environment. On the other hand, there is no movement in the non-behavior area, so the change in contrast is negligible.

In this paper, we propose an algorithm to detect drivers' dangerous behaviors by using the change in contrast in the behavioral area. The proposed algorithm detects the final behavior through two steps. The first step is to detect the driver's behavior area. The second step is to classify the final driver behavior using a deep learning classification algorithm

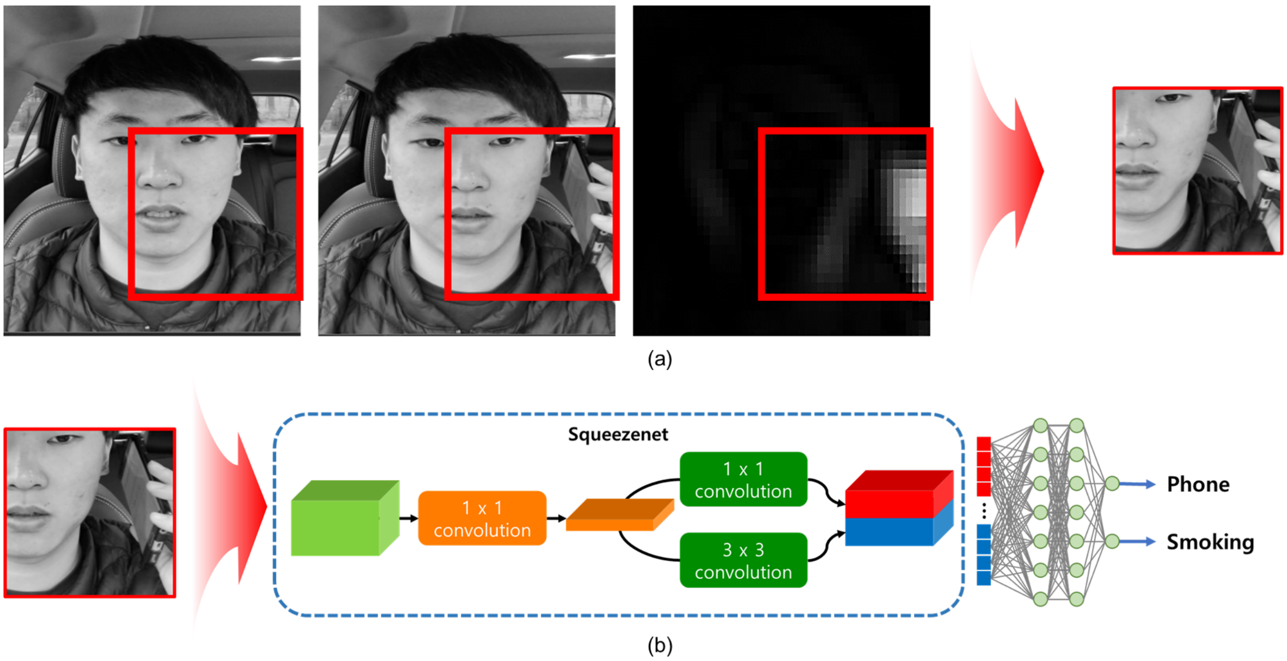

Figure 2. The structure of the driver behavior detection algorithm: (a) inferring the behavior area, (b) classifying the behavior area

in the behavioral area. Through this process, the driver's behavior can be detected using only deep learning object classification without using a deep learning object detection algorithm.

The rest of the paper consists of the following. The proposed driver behavior detection algorithm is described in Section II. The experimental results and analysis are described in Section III, and the conclusions are summarized in Section IV.

II. Behavior detection algorithm

Figure 2 is the structure of the proposed algorithm. The proposed algorithm consists of two steps behavioral area inference and classification. The first step is to compare the contrast between the previous and current frames, as shown in Figure 2(a), to find the active area with the most considerable contrast. The second step classifies the driver behavior by inputting the active region of Figure 2(a) into the deep learning classification algorithm, as shown in Figure 2(b).

A. *infer the behavior area*

The process of inferring the behavioral area consists of four steps. In the first step, the previous frame and the current frame are downsampled to a size of 48×36. The second step is to create an active image, a contrast change image, as shown in Figure 3. Figures 3(a) and 3(b) are the results of applying a 5×5 smoothing filter to the frame. The active image in Figure 3(c) is generated by the difference in contrast values between Figure 3(a) and Figure 3(b), to which smoothing is applied. The third step is to downsample the active image to a size of 8×6, as shown in Figure 4. The brightest area in Figure 4 is the area where the driver acted. If the average of the total contrast values of the active image exceeds 60, the light change is excessive. The typical environment where there are many light changes is the tunnel's entrance and exit. In such an environment, it is impossible to infer the behavior area. Therefore, only when the average of the total contrast values of the active image is less than 50 does it move on to the area inference step. The final step is to store each sliding value by sliding window size of 3×3 in the downsampled active image. An area with the most considerable value in the sliding window becomes an active area. When the sum of pixels in the extracted active area exceeds 300, it is input to the following behavioral area classification process.

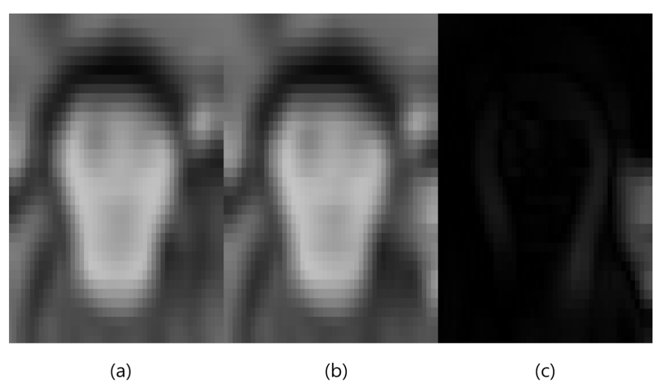

Figure 3. Generating an active image through a difference between a previous frame and a current frame: (a) previous frame, (b) current frame, (c) active image

In the process of inference of behavioral areas, downsampling uses area interpolation. Two benefits can be obtained by downsampling by area interpolation. First, as the image becomes smaller, the following operation's calculation time can be reduced. Second, to find the behavioral area, the change in some areas is more important than the change in a specific pixel.

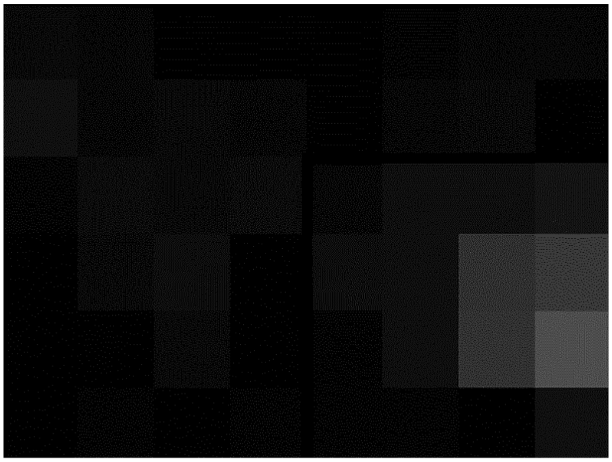

Figure 4. The result of downsampling the active image to 8×6

B. Classification of behavioral areas

To classify behavior, this paper used Squeezenet and a deep learning classification algorithm [10]. Existing deep learning classification models focus on increasing accuracy. However, SqueezeNet is a small classification network with a small number of parameters. Models with few parameters are suitable for hardware with limited computational resources, such as vehicle embedded environments. Learning data were collected to learn the behavior classification model. Learning data consisting of smoking and cigarettes were collected 1,000 sheets each. 800 data were used for learning, and the remaining 200 were used for algorithmic performance tests. Data amplification used rotation, flip, and movement. Rotational amplification was applied at $\pm 20^\circ$ considering the rotation of the driver's face. Flip amplification was applied only to the left and right. Movement applied 20% of the image size. Figure 5 is the result of classifying the behavioral area through the classification algorithm. https://ko.overleaf.com/project/61a713fa1d603bc1333327dc

III. EXPERIMENT RESULTS AND ANALYSIS

To analyze the performance of the behavior position detector, 100 pairs of general state frames were collected. We also collected 100 pairs of data on driver behavior. We confirmed that the behavioral location detector has an accuracy of 90%. For the performance analysis of the behavior classifier, 200 behavior data were collected. We confirmed that the behavior classifier has an accuracy of 95%. Finally, We confirmed that our final overall system has an accuracy of 85%.

To compare the operation speed, the SSD detector and the proposed algorithm were compared [4]. Computer equipment and environment are CPU i7-8086, RAM 16GB, Windows, Keras. The operation speed of SSD is 8FPS, and the proposed algorithm is 25FPS.

Figure 5. The result of detecting a driver using a mobile phone

The proposed algorithm was more accurate in the environment where there was no change in the interior contrast caused by lighting and sunlight. However, the accuracy was low in an environment where there was much change in vehicle interior contrast due to lighting and sunlight. The cause of the performance degradation is the behavioral position detector. The behavior region detector uses the contrast between the previous frame and the current frame.// If there is a difference in contrast due to lighting and sunlight, not the difference in contrast due to behavior, the proposed algorithm incorrectly detects the location.

IV. Conclusion

In this paper, we propose a behavior detection algorithm that reduces the computation time by 70% compared to the deep learning object detection model by using the static movement of the driver. The proposed algorithm consists of two steps. The first step is finding the behavior position in the previous and current frames. The second step is to input the behavioral location into the classification network to detect the final driver's behavior. The accuracy of the proposed behavior detector is 85%.

Acknowledgment

This research was supported by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2021-2020-0-01808) supervised by the IITP (Institute of Information Communications Technology Planning Evaluation); and the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2021R1A6A1A03043144).

References

[1] X. Sun, P. Wu, and S. C. Hoi, "Face detection using deep learning: An improved faster rcnn approach," *Neurocomputing*, vol. 299, pp. 42–50, 2018.

[2] T. Hoang Ngan Le, Y. Zheng, C. Zhu, K. Luu, and M. Savvides, "Multiple scale faster-rcnn approach to driver's cell-phone usage and hands on steering wheel detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2016, pp. 46–53.

[3] T.-C. Chien, C.-C. Lin, and C.-P. Fan, "Deep learning based driver smoking behavior detection for driving safety," *Journal of Image and Graphics*, vol. 8, no. 1, 2020.

[4] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "Ssd: Single shot multibox detector," in *European conference on computer vision*. Springer, 2016, pp. 21–37.

[5] J. Redmon and A. Farhadi, "Yolo9000: better, faster, stronger," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 7263–7271.

[6] R. Girshick, "Fast r-cnn," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1440–1448.

[7] S. A. Hussain and A. S. A. Al Balushi, "A real time face emotion classification and recognition using deep learning model," in *Journal of Physics: Conference Series*, vol. 1432, no. 1. IOP Publishing, 2020, p. 012087.

[8] R. A. Naqvi, M. Arsalan, G. Batchuluun, H. S. Yoon, and K. R. Park, "Deep learning-based gaze detection system for automobile drivers using a nir camera sensor," *Sensors*, vol. 18, no. 2, p. 456, 2018.

[9] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 4510–4520.

[10] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, "Squeezenet: Alexnet-level accuracy with 50x fewer parameters and< 0.5 mb model size," *arXiv preprint arXiv:1602.07360*, 2016.

[11] K. Seshadri, F. Juefei-Xu, D. K. Pal, M. Savvides, and C. P. Thor, "Driver cell phone usage detection on strategic highway research program (shrp2) face view videos," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2015, pp. 35–43.