Driver Drowsiness Detection based on Multimodal using Fusion of Visual-feature and Bio-signal

Hyung-Tak Choi*
Chungnam National University
Computer Engineering
Deajeon, Korea
poo4967@gmail.com

Moon-Ki Back
Chungnam National University
Computer Engineering
Deajeon, Korea
zmzment@gmail.com

Kyu-Chul Lee◆
Chungnam National University
Computer Engineering
Deajeon, Korea
kclee@cnu.ac.kr

***Abstract*—It is very dangerous for a driver to fall into a momentary drowsiness. Previous studies to prevent this have classified the driver's condition mainly by using features that appear when drowsiness occurs. At this time, there is insufficient information to judge the driver's condition by using only single data (physiological data, visual data). In this study, we propose a system based on Multimodal Deep Learning that recognizes both visual and physiological changes in drowsiness. Because using different kind of data, heterogeneity problem arise. So in order to eliminate heterogeneity between data, using generative model to representation. Since drowsiness is a change that occurs with time, we use a deep learning network consisting of Long Short-Term Memory (LSTM) to classify the driver's condition**

Keywords—multimodal, deep learning, drowsiness detection, visual feature, physiological feature

I. Introduction

Drowsy driving is a very dangerous accident and many methods are being studied to prevent it. Methods for detecting driver drowsiness can be divided into three categories[]. The first is to use physiological data. This can determine correct drowsiness, but it can be difficult to use in a limited situation of driving. However, at present, basic physiological data can be collected through wearable devices. The second is to grasp the driving pattern and identify the case of lane invasion. This also depends on the characteristics of each driver, so that the versatility can be deteriorated. The third is to utilize the driver's visual data. It is a method widely used in the limited situation of driving by detecting the drowsiness of the driver by observing changes that occur when drowsiness occur when drowsiness. However, visual data determines the driver's condition through the designed algorithm, which makes it difficult to say that it is an accurate drowsiness.

In this study, we propose a system to detect driver's drowsiness by using physiological data as well as visual data. Using all of the driver's physiological changes along with visual changes can be a clear basis for drowsiness decisions. At this time, the heterogeneity of the two data should be removed because different data are used, and various methods are suggested[1]. Actually, Drowsiness is detection by an algorithm designed through shallow classification of two data (image, BPM). But it is difficult to use all the features of both data because it uses a shallow classifier. Therefore, the characteristics of each data are extracted and included in all the properties of the used modality by the fusion method[2].

The driver's condition is one class and deep learning is used to classify it. The field of machine learning show high performance in various fields, and image processing is highly accurate in Computer Vision. So some studies have used Convolutional Network Neural, a deep learning network, to detect driver drowsiness[3]. However, since drowsiness is a change that occurs over time, it is difficult to assume that only the use of images at the moment has taken into account, and it is impossible to be sure that drowsiness is judged only from visual data. Therefore, in this study, we use Long Short-Term Memory which can consider the temporal change to classify the driver's condition. The driver's condition is classified into 3 class because considering visual data as well although it has 6 steps in total based on physiological data.

II. Related Work

In this section, we describe the following: (1) the technology of drowsiness detection by feature(visual, physiological), and (2) fusion of multi-modality.

A. Drowsiness detection

Many researches have been proposed to detect driver's drowsiness. A method based on visual characteristics extracts features that indicate change in drowsiness in the driver's face image. The movement of the eyes, mouth, and head posture [4], [5], [6] are mainly used, and the precise extraction of feature from the driver's image can provide high accuracy. As a method of recognizing the driver's face and extracting the feature mainly use the Viola&Jones algorithm [7] or the face landmark detection. Reference [8] designed a system for using physiological data BPM as well as visual data, and there is a decision tree of each data.

B. Multi-modality

Multimodal methods for using heterogeneous data have also been studied [1], [2], [9], and a method of utilizing all the features of each modality through a preprocessing process that extracts the feature of each data has been proposed. As RBM(Restricted Boltzmann Machine is generative model, has reconstruct data. At this time if probability distribution of reconstruct data has similar probability distribution of input data, generative model understand about input data well. So joint representation by RBM has features about multi-modality

III. Multimodal Deep Learning

* First Author
◆ Corresponding Author.

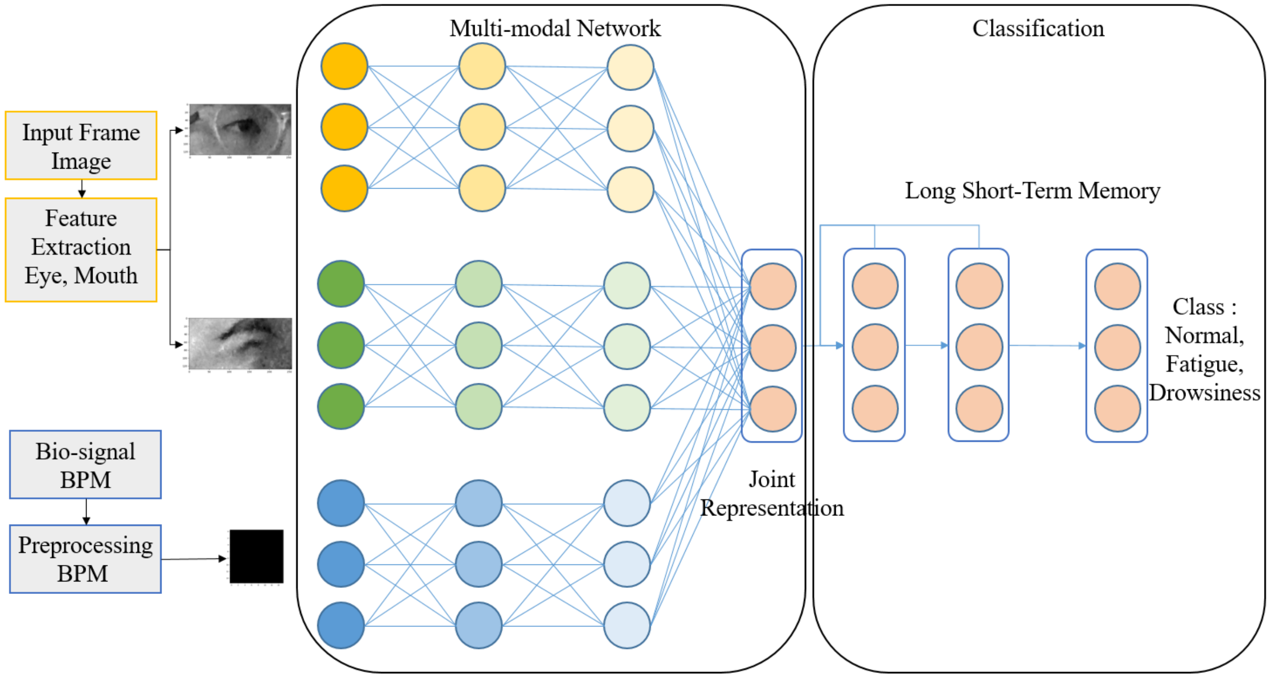

Fig. 1. Drowsiness detection system

The features used in the proposed system have basic preprocessing tasks. Also, it is divided into two stages after input into the system as shown in Fig. 1. The first joint representation the input feature data as one data by generative model. The second, classifies the driver's condition using representation data.

A. *Preprocessing*

We used the features of the driver's eyes and mouth. Keeping eye closure duration after continuous yawning it is more accurate drowsiness rather than simply keeping eye closure. In addition, we used the bio-signals (heart rate, EEG etc.) collected from wearable devices and a total of three modalities are used shown in Fig. 2. First we need to collect visual feature from driver's face image. We used MTCNN(Multi-task Cascaded Convolutional Networks) to extract feature[10]. MTCNN have three stage convolution networks to predict face and landmark location. We need to extract feature(eye, mouth) from driver's face and MTCNN is good at study.

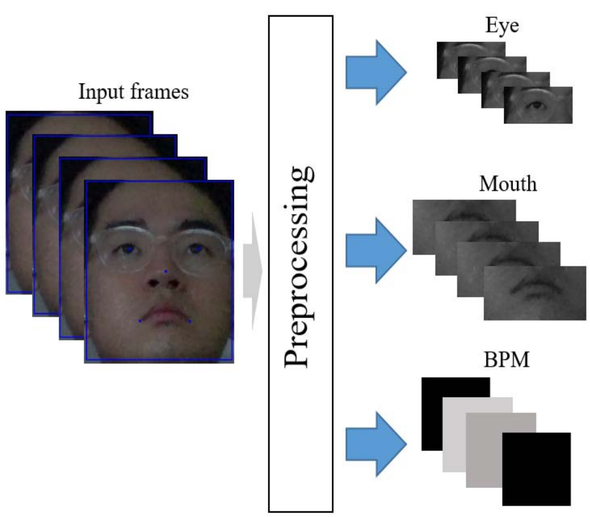

Fig. 2. Extract feature

B. *Maintaining the Integrity of the Specifications*

Three modalities(eye, mouth, BPM) are characterized have different data size and format each other. In addition driver's face frame have simply one value of bio-signals so we have a preprocessing. this study before each modalities train, to make image include bio-signals feature and each pixel have a value of BPM. After to representation remove heterogeneity three modalities used RBM(Restricted Boltzmann Machine) If input data and reconstruct data have similar probability distribution, the hidden layer that constitutes the RBM can represent and understand for input data.

C. *Maintaining the Integrity of the Specifications*

Representation data have three state of drowsiness, fatigue, normal. And used LSTM to classify driver state to consider temporal change. LSTM consider input data and pre input data and it is an appropriate method to detect drowsiness with temporal change.

IV. Experiment

A. *Dataset*

Experiments are carried out by 5 volunteers, aged between 20 to 30 years old. The volunteers consist of both male and female drivers have driver's license. We built a virtual driving environment and used the game Euro Truck. Because to collect data for driver drowsiness very dangerous. So the equipment shown in Fig. 3 was constructed to collect drowsiness in a similar environment. Table 1 describe the

Fig. 3. Experiment environment

composition of the database collected in virtual environment Configured data is progressing. Table2 describe the composition of the training data and test data. Suggested system compare with LSTM used single modality for accuracy of drowsiness detection.

TABLE I. TABLE TYPE STYLES

Video	20 (hours)		
Extract feature	eye	mouth	Bpm
Number of features	51,363 (frames)	51,363	51,363

B. *Result*

The studies related to this study mainly use single data. Although reference [8] use multi-modality (PPG, eye), lacks one feature type (mouth). So it also demonstrates the benefits of multi-modality by comparing it with the results using only single data.

The results of classify driver's condition with only BPM feature show highest accuracy among the features. However, detection of drowsiness using a single image features has overall low accuracy. Actually, it is very difficult to grasp the driver's condition with only feature image (mouth, eye). Based on the multimodal, it shows higher accuracy than the other experiment when detecting driver's drowsiness as shown in Fig. 4 graph.

TABLE II. TABLE TYPE STYLES

Worked	**Accuracy**
LSTM (eye)	79.1 (%)
LSTM (mouth)	67.2 (%)
LSTM (bpm)	85.4 (%)
LSTM (multi-modality)	90.5 (%)

V. CONCLUSION

Detecting accurate drowsiness in restrict drive situation is difficult study. To detect drowsiness used temporal change but it is difficult to look a definite criterion. For this reason, we used physiological data together. To discriminate(distinguish) the driver's condition used deep learning, Experimental results show that the accuracy is higher than when using a single modality, as shown Fig. 4. because used more the features. We will recruit additional data now, and if we learn more and more data, we will be

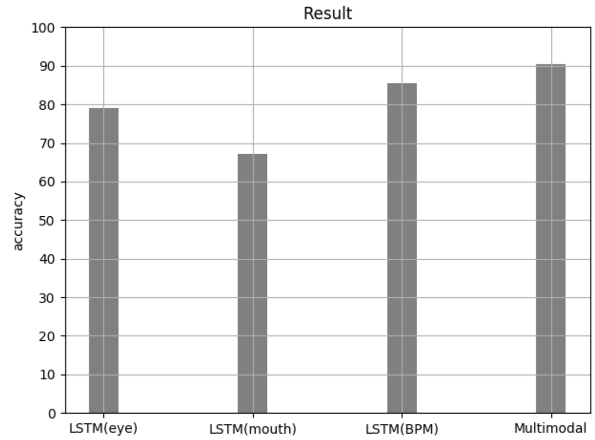

Fig. 4. Experiment result

able to show high accuracy. In addition we will be designed additional layers (i.e. Batch Normalization, CNN) on current system.

ACKNOWLEDGMENT

This work was supported by a grant (18CTAP-C132686-02) from Technology Advancement Research Program (TARP) funded by Ministry of Land, Infrastructure and Transport of Korean government.

REFERENCES

[1] Ngiam, J., Khosla, A., Kim, M., Nam, J., Lee, H., & Ng, A. Y. (2011). Multimodal deep learning. In *Proceedings of the 28th international conference on machine learning (ICML-11)* (pp. 689-696).

[2] Radu, V., Tong, C., Bhattacharya, S., Lane, N. D., Mascolo, C., Marina, M. K., & Kawsar, F. (2018). Multimodal deep learning for activity and context recognition. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *1*(4), 157.

[3] Dwivedi, Kartik, Kumar Biswaranjan, and Amit Sethi. "Drowsy driver detection using representation learning." *2014 IEEE International Advance Computing Conference (IACC)*. 2014.

[4] Jo, J., Lee, S. J., Park, K. R., Kim, I. J., & Kim, J. (2014). Detecting driver drowsiness using feature-level fusion and user-specific classification. *Expert Systems with Applications*, *41*(4), 1139-1152.

[5] Teyeb, I., Jemai, O., Zaied, M., & Amar, C. B. (2014, July). A novel approach for drowsy driver detection using head posture estimation and eyes recognition system based on wavelet network. In *Information, Intelligence, Systems and Applications, IISA 2014, The 5th International Conference on* (pp. 379-384). IEEE.

[6] Saradadevi, Mandalapu, and Preeti Bajaj. "Driver fatigue detection using mouth and yawning analysis." *International journal of Computer science and network security* 8.6 (2008): 183-188.

[7] Viola, P., & Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on* (Vol. 1, pp. I-I). IEEE.

[8] Lee, B-G., S-J. Jung, and W-Y. Chung. "Real-time physiological and vision monitoring of vehicle driver for non-intrusive drowsiness detection." *IET communications* 5.17 (2011): 2461-2469.

[9] Srivastava, Nitish, and Ruslan R. Salakhutdinov. "Multimodal learning with deep boltzmann machines." *Advances in neural information processing systems* (2012): 2222-2230.

[10] Zhang, K., Zhang, Z., Li, Z., & Qiao, Y. (2016). Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Processing Letters*, *23*(10), 1499-1503.