

Feature extraction for drowsiness detection using facial skin temperature distribution

by A. Masaki*, K. Nagumo*, K. Oiwa* and A. Nozawa*

* Aoyama Gakuin University, Japan, akio@ee.aoyama.ac.jp

Abstract

Level 3 autonomous driving vehicle, the driver needs to drive himself when the system is in an emergency. Therefore, the system to detect the lowering of arousal degree is desired. Commercially available drowsiness detection technologies can be roughly classified into two methods. These include: 1) Vehicle-based methods; 2) Driver behavior-based methods. These two techniques may only work if drivers are fully drowsy and the risk of an accident increases. So far, there has been proposed the model that detects drowsiness instantaneously on facial skin temperature distribution using convolutional neural network (CNN). However, CNN, a supervised learning algorithm, has difficulty preparing datasets for all drowsiness levels. To solve this problem, we focused on an unsupervised learning algorithm, the variational autoencoder (VAE). In this study, we tried to construct a model to detect drowsiness using VAE by using only the facial skin temperature distribution during arousal. The datasets were prepared by conducting the drowsiness induction and suppression experiment.

1. Introduction

Autonomous driving technology is being actively studied. Level 3 autonomous driving vehicle as defined by Society of Automotive Engineers (SAE) [1], the driver does not need to monitor traffic conditions when the autonomous driving system is normal, but the driver needs to avoid danger himself when the system is in an emergency [2]. Therefore, drivers need to maintain concentration and suppress drowsiness in order to cope with the switching of driving in an emergency, and a system to detect the lowering of arousal degree is desired [3].

Commercially available drowsiness detection technologies can be roughly classified into two methods. These include: 1) Vehicle-based methods; 2) Driver behavior-based methods. Examples of 1) are drowsiness detection from the state where the vehicle deviates from the lane and repeats meandering operation [4][5]. Examples of 2) are drowsiness detection from closed eyes or continuous blinking [6][7]. These two techniques may only work if drivers are fully drowsy and the risk of an accident increases.

In contrast to the preceding argument, the method that uses the physiological information of the driver may be able to detect the sign of drowsiness according to the physiological mechanism [8]. So far, studies have been reported to estimate drowsiness from physiological indicators such as brain wave (EEG) [9], electroencephalogram (EOG) [10], heart rate [11], and hemodynamics [12]. However, the physical and mental load on the driver can be large, since these physiological indicators require a sensor to be directly attached to the driver.

We previously evaluated drowsiness states based on facial skin temperature distribution taken by infrared thermography, which can record remotely. It has been clarified that the facial skin temperature rises with drowsiness, which is one of the autonomic nervous system indices that may be a sign of drowsiness [13]. So far, there has been proposed the model that detects drowsiness instantaneously on facial skin temperature distribution using convolutional neural network (CNN) [14]. Therefore, the features extracted automatically from the facial skin temperature with drowsiness induction by deep learning were observed [15]. This study had two limitations. These include: Limitation 1) It is not clear that features related to drowsiness could be extracted because skin temperature has various fluctuation factors; Limitation 2) Difficult to estimate drowsiness because a dataset for all drowsiness levels must be prepared due to the characteristics of CNN.

Given this background, we worked with two approaches. First of all, in order to solve Challenge 1), the experiment was conducted to induce drowsiness induction and drowsiness suppression to confirm whether the desired feature can be extracted. As a next step, in order to solve Challenge 2), the variational autoencoder (VAE) [16], a deep learning algorithm that has been successful in the field of anomaly detection in image space, was adopted. In this study, we defined "drowsy" as "anomaly" and tried to extract features related to drowsiness using only the facial skin temperature distribution during arousal.

The objective of study is detecting drowsiness using facial skin temperature distribution to prevent traffic accident. In order to extract the drowsiness-related features, we constructed the model that extracts the drowsiness-related features and detects drowsiness by learning only the facial skin temperature during arousal using VAE. In this paper, the task of making the subject repeat drowsiness induction and suppression was examined by experiment. Further, using the data set obtained by the experiment, the feature extraction regarding drowsiness was performed using VAE.



2. Experiment

2.1. Experimental system

The preliminary study was conducted in order to carry out an experiment in which drowsiness induction and drowsiness suppression were periodically evoked. Preliminary experiments are described here. The measured physiological indices were the following: facial skin temperature (FST), electroencephalograms (EEG) as brainwaves. The bioinstrumentation system consisted of an infrared thermography device (A-600 Series, FLIR) and wireless biological measuring equipment (Polymate Mini AP108, TEAC Co.). The infrared thermography device was set at a distance 100 cm from the face. Thermal images were created at 1-s sampling intervals. The size of each thermal image was 640 × 480 pixels, and the temperature resolution was less than 0.1 °C. The infrared emissivity of the skin was $\epsilon = 0.98$. The wireless biological measuring equipment recorded the EEG with a sampling frequency of 500 Hz. To evaluate relative an alpha waves, EEG was recorded by a referential electrode derivation method. The EEG electrode was fixed at parietal (Pz) positions according to the international 10-20 system. The right ear lobe (A2) was used as a reference. Before data acquisition, the contact impedance between the EEG electrodes and scalp was calibrated to be less than 10 k Ω .

Facial expressions (for example, blink cycles and lips movements) represent characteristic changes in accordance with awareness level [17]. In this study, an infrared camera recording the subject's face was placed 60 cm in front of the subject to assess facial expression.

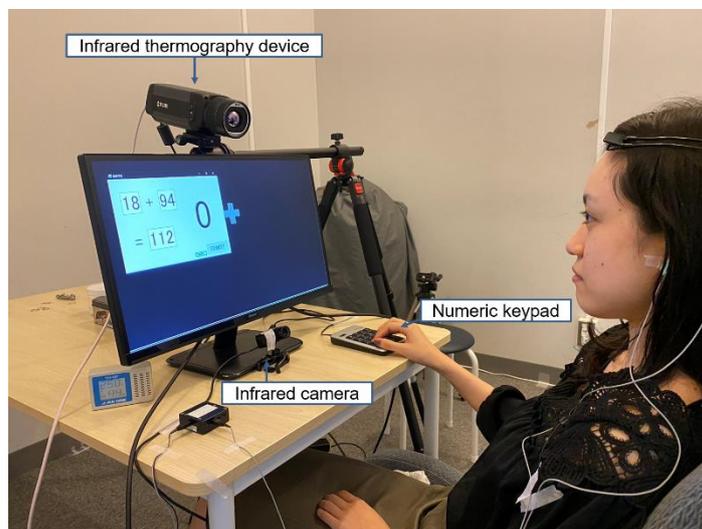


Fig. 1. Experiment image

2.2. Procedure and conditions

The subject was a man (age: 31 years). He was fully informed about the experiment and objectives of our study before participation and signed a consent form. The experiments were carried out during the day to control for potential circadian rhythm effects. The experiments began after the subject had been in the room for at least 30 minutes so that they could become acclimated to the experimental room temperature.

The experiment consisted of four 3-min resting state segments, "Rest 1", "Rest 2", "Rest 3", "Rest 4", and a 15-min induced drowsiness segment, "Drowsiness induction", and a 5-min suppressed drowsiness segment, "Drowsiness suppression", and a 30-sec break time segment, "Break". In each "Rest" segment, subject was instructed to sit in a rested state. Furthermore, the "Rest" segment consisted of three 1-min resting state segments, "First opened eyes(FOE)", "Second closed eyes(SCE)", "Last opened eyes(LOE)". In the "Drowsiness induction", subject was instructed to gaze at the video for, shown in Fig. 2, with only eye movements in order to induce drowsiness. The image for inducing drowsiness was displayed on an LCD monitor placed in front of the subject. The circular moving target moved to draw a circular orbit. One cycle was set to two seconds. In the "Break", subject stretched to reduce her fatigue. In the "Drowsiness suppression", to begin with, the amazing sound was played for about 20 sec. And then, for about 5 minutes, arithmetic questions were presented on LCD monitor, while displaying each for 4 seconds, shown in Fig. 2. The subject input an answer using a numeric keypad. The experiment was conducted with the lights off to enhance the effect of inducing drowsiness.

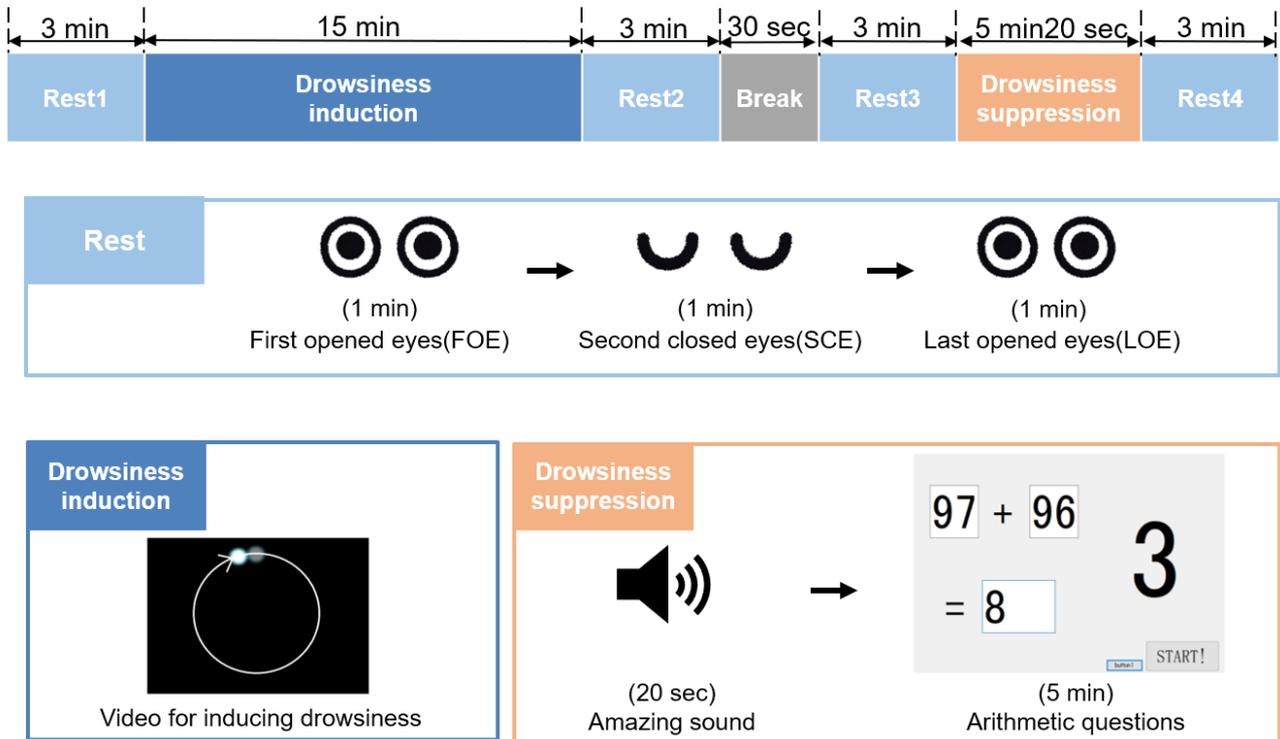


Fig. 2. Experimental protocol

3. Definition of the drowsiness level

We adopted the drowsiness level to objectively assess drowsiness. Thus, the facial skin temperature distribution was labeled by drowsiness level. The drowsiness level was assessed objectively every 20 seconds by the assessment criteria of New Energy and Industrial Technology Development (NEDO) in Japan [18], shown in Table 1, based on the facial expressions during “Drowsiness induction”. Three experimental collaborators (i.e., no subjects) made the assessment.

Table 1. Drowsiness rating criteria based on facial expression by the NEDO procedure^[18]

Drowsiness level	State	Example of facial expression and motions
Level 1	Awake	<ul style="list-style-type: none"> · Eye movements are quick and frequent · Blink cycles are stable at approximately two per two seconds · Body motions are active
Level 2	Slightly drowsy	<ul style="list-style-type: none"> · Lips are parted · Motions of eye movements are slow
Level 3	Drowsy	<ul style="list-style-type: none"> · Blinks are slow and frequent · Re-positions body on seat · Touches hand to face
Level 4	Very drowsy	<ul style="list-style-type: none"> · Blinks assumed to occur consciously · Shakes head · Frequently yawns
Level 5	Extremely drowsy	<ul style="list-style-type: none"> · Eyelids closing · Leans head back and forth

4. Analysis method

4.1. Alpha wave

The frequency component from 8 Hz to 13 Hz of the EEG is called an alpha wave, which is prominently expressed at rest, eye closure and arousal. The alpha wave decreases with eye opening and during sleep. EEG data time series are subjected to fast Fourier transform at 512 points per second to calculate the power spectra $P(t, f)$, where t is time in the experiment. Alpha power spectra $P_\alpha(t)$ is an average value per second represented as

$$P_\alpha(t) = \frac{1}{N} \int_{f_1}^{f_2} p(t, f) df \quad (t = 1, 2, \dots, 2075) \quad (1)$$

where f_1 and f_2 are 8 and 13, respectively, which are the limits of the alpha frequency band, and N is the number of samples in the band. The averaged alpha power spectra in each drowsiness level based on the objective drowsiness $P_{\alpha,i}$ are computed in the following equation:

$$P_{\alpha,i} = \frac{1}{N'} \sum P_{\alpha,i}(t) / P^*_\alpha \quad (i = 1, 2, \dots, 4) \quad (2)$$

where N' is the number of samples in each resting state segments and P^*_α is the average of $P_\alpha(t)$ at resting state segments.

4.2 . Variational Autoencoder

Variational Autoencoder(VAE) is a generative model based on deep learning. Fig.3 shows overview of the VAE algorithm. X, \tilde{X}, NN represents input, output, and a neural network, respectively. The VAE network is divided into an encoder section and a decoder section. Given observation $X = \{\vec{x}_1, \vec{x}_2, \dots, \vec{x}_N\}$, VAE identify probability distribution $p(\vec{x}_* | X)$ that produce unobserved value (\vec{x}_*) . VAE is designed on the assumption that latent variables that serve as explanatory variables are normally distributed. The Encoder performs dimensional compression of X , and it calculates the mean vector $(\vec{\mu}_\phi(x))$ and variance $(\Sigma_\phi(x))$, which are parameters of the normally distribution. The blue part in Fig.3 indicates that sampling is performed from the standard normal distribution. The points are then sampled from the latent space distribution (\vec{z}) . In the decoder, the model likelihood parameters $(\vec{\eta}_\theta(\vec{z}))$ is calculated, and the reconstruction error can be computed. Finally, the reconstruction error is backpropagated through the network. Since the reconstruction error to be optimized includes a regularization term that brings the mean $(\vec{\mu}_\phi(x))$ to 0 and the variance $(\Sigma_\phi(x))$ close to the unit matrix, the distribution of the latent variable \vec{z} has a shape close to a standard normal distribution. There is a tendency to regularize the organization of the latent space by bringing the distribution returned by the encoder closer to the standard normal distribution. For this reason, VAE can avoid overfitting and achieve a high recall compared to a normal autoencoder.

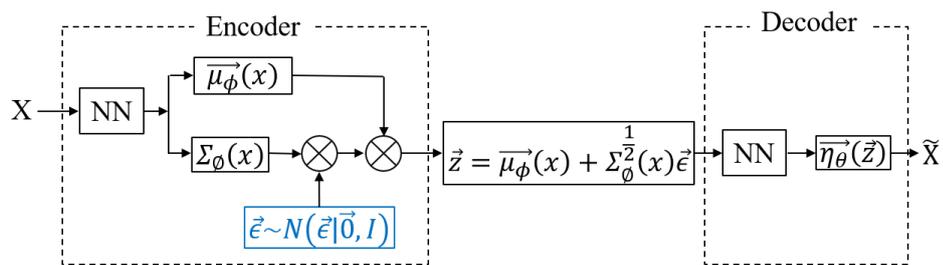


Fig. 3. Overview of VAE

4.3. Drowsiness detection using VAE

Variational We explain the concept of drowsiness detection in facial skin temperature using VAE. Fig.4 shows the conceptual diagram. In this study, we defined thermal face images in arousal state as Arousal, in drowsy state as Drowsy. Only Arousal is used for drowsiness detection using VAE. Firstly, by learning a large amount of Arousal, the drowsiness detection model was constructed. Secondly, testing data (Arousal and Drowsy) were input to the drowsiness detection model. If the pattern was similar to Arousal, the testing data was decided Arousal; otherwise the testing data was decided Drowsy.

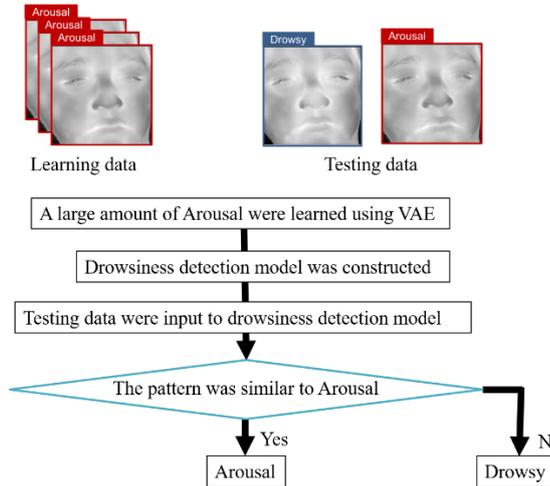


Fig. 3. Conceptual diagram of drowsiness detection using VAE

Drowsiness detection using VAE requires a large amount of normal data. In this study, we verify whether the proposed algorithm is effective in drowsiness detection in facial skin temperature. Drowsy is thermal face images in which the skin temperature in drowsiness level 5 using the NEDO methods. Arousal were thermal face images obtained when the subject was in "Drowsiness suppression" section. Image sizes are 640×480 pixels. The number of Arousal were 120. The thermal face image was normalized such that the maximum was 1 and the minimum was 0 because the difference in the temperature value in the skin area is small. In this study, 640×480 pixels of the thermal image were resized to 210×210 pixels and therefore, the facial skin area remained. When performing VAE learning, a part of the thermal image is randomly cut out at 8×8 size pixels, and this patch is used as learning data as shown Fig.4. The local thermal image of the skin area was expanded to 10,000 sheets.

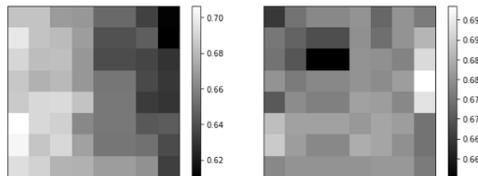


Fig. 4. Sample learning data

Table 2. Construction of encoder

Layers	Function	Filters	Size	Stride
Conv1	-	16	2×2	2
BatchNorm1	ReLU	-	-	-
Conv2	-	32	2×2	2
BatchNorm2	ReLU	-	-	-
FC	-	-	-	-

Table 3. Construction of decoder

Layers	Function	Filters	Size	Stride
FC	-	-	-	-
BatchNorm1	ReLU	-	-	-
ConvTrans1	-	32	2 × 2	2
BatchNorm2	ReLU	-	-	-
ConvTrans2	-	16	2 × 2	2
BatchNorm3	ReLU	-	-	-
ConvTrans3	-	1	4 × 4	-
ConvTrans4	-	1	4 × 4	-
BatchNorm4	ReLU	-	-	-
BatchNorm5	ReLU	-	-	-

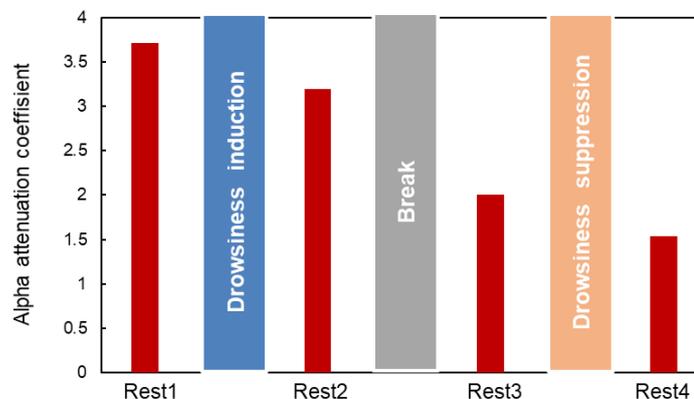
In this study, convolutional layers were placed before the FC layer of encoder in order to extract the features of the skin temperature pattern of the skin blood vessels. Along with that, deconvolutional layers were placed after the decoder. The construction of the encoder is depicted in Table 2. The structure of the VAE encoder consisted of two convolutional layers and one fully connected layer. In this table, Conv, BatchNorm, and FC indicate the convolutional, batch normalization, and fully connected layers, respectively. The mean vector and variance were output from FC. The structure of the decoder is paired with the structure of the encoder, and the structure is opposite to that of the encoder. The construction of the decoder is depicted in Table 3. ConvTrans indicates a transpose convolution. The gradient descent method was used for VAE parameter learning, and the optimization algorithm at that time was Adam. The number of dimensions of the latent variable was 6. The number of epochs was 20 and the batch size was 128. When testing, the spatial unregularized anomaly score was calculated with reference to [19] and used as an index for detecting anomaly data and detecting anomaly points inside thermal image. The equation for the spatial unregularized anomaly score is shown below.

$$L_{VAE}(x) = \sum_{i=1}^{N_x} \frac{1}{2} \times \frac{(\mu_{x_i} - x_i)^2}{\sigma^2 x_i} \quad (3)$$

5. Result and discussion

5.1. Alpha wave

Fig. 5. shows the alpha attention coefficient (AAC) for each rest segment. Since AAC decreased from Rest1 to Rest2, a drowsiness inducing effect in Drowsiness induction section was confirmed. In contrast to the preceding argument, Since the AAC decreased in also Drowsiness suppression section, the effect of drowsiness suppression could not be confirmed. Improvement of experimental protocol is required in the future.

**Fig. 5.** Alpha attention coefficient

5.2. Unregularized anomaly score

Fig. 6. shows the Unregularized anomaly score for any sample. This score is a logarithmic scale. Blue areas indicate abnormal points. The bluer the color, the higher is the degree of feature. There is little difference in Unregularized abnormal scores between normal and abnormal images. This is probably caused by insufficient learning due to a lack of normal data. In the future, we will review the experimental protocol and increase the data of the arousal state.

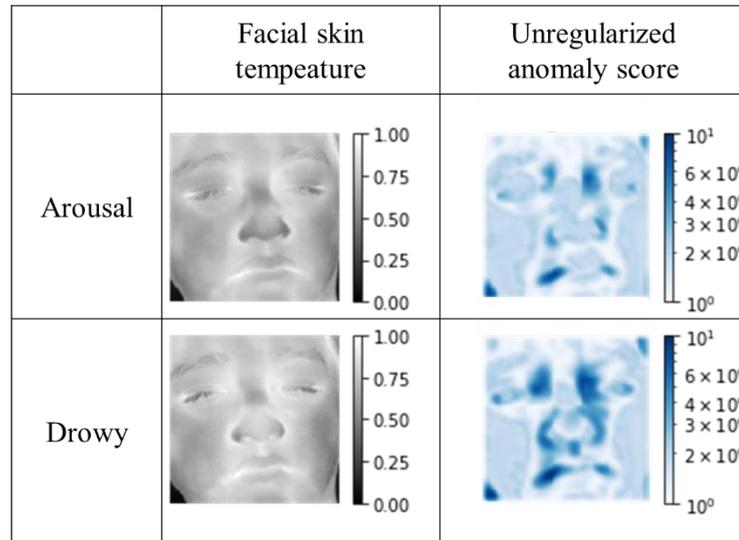


Fig. 6. Unregularized anomaly score

6. Conclusion

The objective of study is detecting drowsiness using facial skin temperature distribution to prevent traffic accident. In order to extract the drowsiness-related features, we constructed the model that extracts the drowsiness-related features and detects drowsiness by learning only the facial skin temperature during arousal using VAE. In this paper, the task of making the subject repeat drowsiness induction and suppression was examined by experiment. Further, using the data set obtained by the experiment, the feature extraction regarding drowsiness was performed using VAE. However, The drowsiness suppression effect was not confirmed in the drowsiness suppression section composed of amazing sound and arithmetic questions. Therefore, in the future, we will review the experimental protocol again and try to extract features related to drowsiness from facial skin temperature.

REFERENCES

- [1] SAE International., "Taxonomy and Definitions for Terms Related to OnRoad Motor Vehicle Automated Driving Systems". J3016-2014, 2014.
- [2] Kobayashi. M, "Legal issues of autonomous vehicles", Journal of Information Processing and Management, Vol. 60, No. 4, pp. 240-250 ,2017. (in Japanese)
- [3] Murase. T, Oiwa. K, Nozawa. A., "Quantitative evaluation of attention to the driving environment during automated driving at Levels 3 using ERP: evaluation of psychophysiological state of driver". IEEJ Transactions on Electronics, Information and Systems, Vol.138, No.9, pp.1148-1153, 2018.
- [4] Saito. Y, Makoto. I, Inagaki. T., "Driver assistance system with a dual control scheme: Effectiveness of identifying driver drowsiness and preventing lane departure accidents". IEEE Transactions on Human-Machine Systems 46.5, pp.660-671, 2016.
- [5] Eskandarian. A, Mortazavi. A., "Evaluation of a smart algorithm for commercial vehicle driver drowsiness detection". 2007 IEEE Intelligent Vehicles Symposium. IEEE, pp. 553-559, 2007.
- [6] Grace. R, Byrne. V.E, Bierman. D.M, Legrand, J. M, Gricourt. D, Davis. B.K, ... & Carnahan, B., "A drowsy driver detection system for heavy vehicles". In 17th DASC. AIAA/IEEE/SAE. Digital Avionics Systems Conference. Proceedings, Cat. No. 98CH36267, Vol. 2, pp. 136-1-136/8, IEEE, 1998.
- [7] Danisman. T, Bilasco. I.M, Djeraba. C, Ihaddadene. N., "Drowsy driver detection system using eye blink patterns". In 2010 International Conference on Machine and Web Intelligence, pp. 230-233, IEEE, 2010.
- [8] Chowdhury. A, Shankaran. R, Kavakli. M, Haque. M.M., "Sensor applications and physiological features in drivers' drowsiness detection: A review". IEEE Sensors Journal 18.8, pp. 3055-3067, 2018.
- [9] Lin. C.T, Wu. R.C, Liang. S.F, Chao. W.H, Chen. Y.J, Jung T.P., "EEG-based drowsiness estimation for safety driving using independent component analysis". IEEE Transactions on Circuits and Systems I: Regular Papers 52, No.12, pp. 2726-2738, 2005.
- [10] Schleicher. R, Niels. G, Susanne. B, Lars. G., "Blinks and saccades as indicators of fatigue in sleepiness warnings: looking tired? ". Ergonomics 51, No.7, pp. 982-1010, 2008.
- [11] Shinar. Z, Akselrod. S, Dagan. Y, Baharav. A., "Autonomic changes during wake–sleep transition: A heart rate variability based approach." Autonomic Neuroscience 130, No.1-2, pp.17-27, 2006.
- [12] Masaki. A, Nagumo. K, Oiwa. K, Nozawa. A., "Drowsiness Estimation Model Based on Hemodynamics" IEEJ Transactions on Electronics, Information and Systems 140.3, pp. 409-410, 2020.
- [13] Bando. S, Oiwa. K, Nozawa. A., "Evaluation of Dynamics of Forehead Skin Temperature under Induced Drowsiness". IEEJ Transactions on Electrical and Electronic Engineering, Vol.12, Iss.S1, pp.S104-S109, 2017.
- [14] Lawrence. S, Glies. C.L, Tsoi. A.C, Back. A.D., "Face recognition: A convolutional neural-network approach". IEEE transactions on neural networks 8.1, pp.98-113, 1997.
- [15] Adachi. H, Oiwa. K, Nozawa. A., "Drowsiness Level Modeling Based on Facial Skin Temperature Distribution Using a Convolutional Neural Network". IEEJ Transactions on Electrical and Electronic Engineering (TEEE C), Vol.14, No.6, pp.870-876, 2019.
- [16] Jinwon. A, Cho. S., "Variational Autoencoder based Anomaly Detection using Reconstruction Probability". Special Lecture on IE 2.1 , pp.1-18, 2015.
- [17] Zilberg. E, Xu. Z.M, Burton. D, Karrar .M, Lal. S., "Methodology and initial analysis results for development of non-invasive and hybrid driver drowsiness detection systems". The 2nd International Conference on Wireless Broadband and Ultra Wideband Communications (AusWireless 2007), pp.1-6, 2007.
- [18] Research Institute of Human Engineering for Quality Life., "Manual for measuring human senses (in Japanese)" ed. 1, par. 2, p. 146 ,1999.
- [19] Tachibana. R, Matsubara. T, Uehara. K., "Anomaly Manufacturing Product Detection using Unregularized Anomaly Score on Deep Generative Models". Information Processing Society of Japan, 2A1032A103, 2018. (in Japanese)