Preprints.org

Article

Development of Driver State Monitoring System based on Multi-Sensor Data Fusion

Sunhong Park [*] and Jongwon You [*]

Article

Development of Driver State Monitoring System Based on Multi-Sensor Data Fusion

Sunhong Park [1],* and Jongwon You [2],*

[1] Department of Driving Control Technology, Korea Automotive Technology Institute, Chungnam 31214, Republic of Korea; sunhpark@katech.re.kr

[2] Team of Research and Development, FUSIONSOFT, Daegu 41085, Republic of Korea; jwyou@fusionsoft.co.kr

* Correspondence: sunhpark@katech.re.kr (S.H.P.); jwyou@fusionsoft.co.kr (J.W.Y.); Tel.: +82-536097059

Abstract: A stable driver state is essential in the process of manually transition control, which inevitably occurs in Level 3 automated driving situations. To this end, this paper proposed a CNN algorithm-based driver state monitoring system that uses multi-sensor data such as driver's face image, biometric information, and vehicle behavior information as input. This system calculates the probability of drowsiness for each of the four time periods using a convolutional neural network (CNN) based on ToF camera-based eye blinking, ECG information (pulse rate) embedded in the steering wheel, and vehicle information (steering angle data). In order to build a reliable and high-quality training dataset (Ground Truth) for the CNN algorithm, a baseline was established by matching the driver's face image with the electrocardiogram (ECG) and electroencephalogram (EEG) changes in the drowsy and normal states. In a simulation test of the proposed CNN algorithm using more than 20,000 driver image data acquired using a driving simulator, the TNR was 94.8% and the accuracy was 94.2%. Our proposed method is expected to minimize human errors that may occur when switching control by monitoring inappropriate driver state (drowsiness) in real time.

Keywords: Driver state monitoring system; Multi-sensor data fusion; CNN Model; ToF(Time of Flight) camera; Hand on detection; EEG; ECG;

1. Introduction

With the recent shift in mobility, which was used as a means of transportation, automated vehicles are accelerating the paradigm shift to future cars with the latest technologies that can significantly change the way drivers drive, the driving behavior itself, and the user behavior. These automated vehicles are divided into six stages according to the roles of users and automation systems (response to emergency situations while driving, monitoring the surrounding environment, and the subject of control or driving responsibility). According to Society of Automotive Engineers (SAE) [1] criteria, level 3 of automated driving is a conditional driving automation stage that can be entrusted to the system installed in the vehicle to a certain extent, such as perception, judgment, and maneuvering during driving. During automated driving, the driver can take their hands off the wheel, allowing them to drive in a comfortable position. In addition, conditional eye-off, which is the biggest difference from automated driving levels 1 and 2, it is possible to use a smartphone, operate a navigation screen, or watch a movie while driving. However, even when the driving automation system is in operation, the driver must be in a condition to immediately return to driving on behalf of the system at any time if the system requests driver intervention because continuous automated driving is not possible due to vehicle factors such as sensor or system instability or failure and driving environment factors such as sharp curves or road obstacles. Like this, transition from automated driving by the system to manual driving by the driver is called the driver control transition. If a problem occurs in this process, the possibility of a gap or incompleteness occurring in the control subject increases, leading to a traffic accident. According to the automated driving disengagement report of the California Department of Motor Vehicles (DMV) (December 2017 to November 2019), a total of 33 companies reported a total of 106,714 incidents, with driver factors accounting for about 53%, vehicle factors accounting for about 46%, and other external environmental factors accounting for about 1% [2]. The possibility of a traffic accident is relatively low if the driver determines that it is safe and

requests to switch from manual driving to automated driving. However, in the opposite case, the physical and cognitive state of the driver becomes an important variable in the safety of the control transition when the system switches from automated driving to manual driving with an intervention request. In other words, it can be said that it is safe only when the driving control authority is converted in an appropriate state for the driver to drive. If a driver was transferred control from a unstable or inappropriate state, the ability to recognize and respond to changes in surrounding traffic conditions decreases and the ability to cope with situations increases the likelihood of a traffic accident.

In this paper, we proposed a convolutional neural network (CNN) algorithm-based driver state monitoring system that uses multi-sensor data as input, including driver's face image, biometric information, and vehicle behavior information. The probability of drowsiness is calculated for each of the four time periods using a convolutional neural network (CNN) based on ToF camera-based eye blinking, ECG information (pulse) embedded in the steering wheel, and vehicle information (steering angle data). Through this system, it is possible to minimize human errors that may occur when switching control by monitoring the state of an inappropriate driver in real time and warning the driver in case of danger.

The rest of the paper is organized as follows. In Section 2, we investigated and analyzed driver state monitoring systems, which are essential for automated driving control transition, through conventional research. Section 3 describes the multi-sensor fusion data-based CNN algorithm for driver state monitoring system. It describes the main sensor configuration of the system and the proposed CNN algorithm. Section 4 presents the analysis of simulation results, a comparison with existing studies, to demonstrate the validity of the proposed algorithm and discusses improvements and enhancements. Section 5 discusses the conclusions, limitations, and considerations of this paper.

2. Related Work

In order to safely transfer control in unexpected situations during automated driving, it is necessary to clearly communicate and recognize driving status information, operating status information, emergency situation information, and control transfer request information to the driver. However, first and foremost, it is necessary to transfer control by consideration the driver state. This is because if the driver does not respond within a certain period of time after the automated driving system requests the transfer of control, it should automatically conduct minimal risk maneuver (MRM) from then on. This is closely related to the driver factor, which is one of the before mentioned disabling factors for automated driving. In this regard, techniques for monitoring driver state in real time can be broadly categorized into detecting changes in driver behavior, physiological responses, and vehicle interior environment and behavior [3–5]. Typically, driver behavior is a physical change that occurs naturally in the process of drowsiness, such as head shaking, driver's facial movement, the frequency of eye closure and yawning, pupil focus, and pupil movement. Driver physiological responses include electrocardiogram, oxygen saturation, and EEG, and changes in vehicle cabin environment and behavior include the amount of carbon dioxide in the vehicle, the number and width of steering adjustments, and changes in lateral position due to steering wheel maneuvers [6]. Mbouna et al.[7] continuously monitored the driver's eye condition and head pose in the vehicle and classified them as warning or non-warning events using Support Vector Machine (SVM) technique using features of eye index, pupil activity, and head pose. The head pose provides useful information about inattention when the driver's eyes are not visible. Experiments with people of different races and genders in real-world road driving conditions demonstrated low error and high warning accuracy. S. Maralappanavar et al. [8] used driver gaze estimation to determine driver distraction behaviors such as talking on a cell phone and talking to bystanders while driving, and showed a gaze detection accuracy of about 75%. F. Vicente et. al [9] proposed a camera-based system consisting of facial feature tracking, head pose and gaze estimation, and 3D geometric inference to detect Eyes off the road (EOR). O. Kopuklu et al. [10] proposed an open set recognition approach for driver state monitoring applications, which is a contrastive learning approach to learn metrics to distinguish between normal and abnormal driving. To this end, they constructed a video-based benchmark, the Driver Anomaly Detection (DAD) dataset. It contains training sets for normal and abnormal driving. We experimented with a 3D CNN structure and found that it performs similarly to the ResNet-18 model but requires 11 times fewer parameters and 13 times less computation. M. Sabet et al. [11] used Haar feature algorithm for eye and face detection and SVM for drowsiness classification using closed and open eye

images, and A. George et al. [12] used CNN for gaze direction classification. F. Zhang et al. [13] proposed a CNN-based eye state recognition method using infrared images to detect fatigue by calculating PERCLOS and eye blink frequency. M. Awais et al. [14] proposed a method to improve driver drowsiness detection performance by combining EEG and ECG and using SVM classifier. G. Zhenhai et al. [15] proposed a novel driver drowsiness detection method based on time series analysis of steering wheel angular velocity. For this purpose, steering behavior was analyzed under fatigue over time. Meanwhile, J. G. Gaspara et al. [16] conducted a study to evaluate the effectiveness of warnings against lane departure due to drowsiness. They tested a total of 72 young adult drivers (21-32 years old) and found that multi-stage warnings were more effective than single-stage discrete alerts in reducing drowsy lane departures.

3. Driver State Monitoring System

The driver state monitoring system consists of an input section to collect video, biometric, and vehicle information, a processing section to determine the driver state (drowsiness) based on multi-sensor data, and an output section to provide a danger warning to the driver based on the determined results. The input part consists of a ToF camera for acquiring the driver's face image inside the vehicle, 12 capacitive sensors and 1 capacitive ECG sensor embedded in the steering wheel, and a device for collecting the driver's steering angle data through in vehicle network (IVN). The proposed CNN algorithm based on multi-sensor fusion data determines the drowsiness status and warns the driver with video and audio.

3.1. Configuration Overview

3.1.1. ToF camera

The ToF camera is a Meere company product with a center wavelength of 955nm, VGA-grade (640*480), 30fps, a detection effective range of 0.2~1.8m and is mounted on the top of the steering column to collect images of the driver's face in real time.
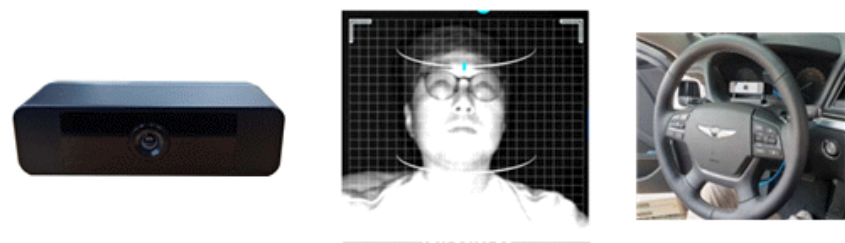

Figure 1. ToF camera Design and Installation.

Table 1. ToF camera specification

Item	Specification
Operation range	0.2 ~ 1.5m
Illuminator	2W-2VCSEL
Depth resolution	VGA(640 x 480)
Field of view	122°(D) x 76°(V) x 100°(H)
Frame rate	30fps
Dimension	60mm(H) x 28mm(V) x 34mm(D)

3.1.2. HoD(Hand on Detection) sensor

The steering wheel is equipped with a capacitive sensor 12ch and a capacitive ECG sensor 1ch(Figure 2 and Table 2), which can detect changes in pulse, a bio-signal, as well as whether the driver makes contact with the steering wheel while driving.
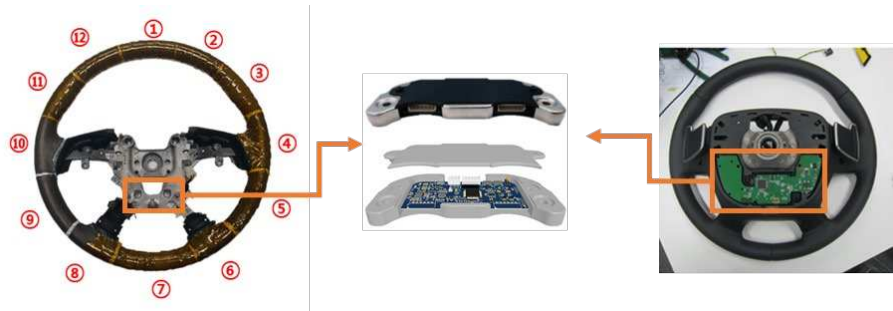
Figure 2. HoD Design and Configuration.

Table 2. HoD specification.

Item		Specification
Sensor type	Hand on detection	12ch(capacitive)
	Bio signal	Electrocardiogram 1ch (Standard Limb Lead II)
Response time		100ms
Interface		CAN (External) / UART, I2C, SPI (Internal)

We measured the recognition rate of steering wheel grip (contact) in various road conditions (straight, Belgian, gravel road, etc.) as shown in Figure 3 by installing HoD sensors in real cars, and the average rate was 97.4%.

Figure 3. Road test conditions with HoD sensor.

3.2. Method

The overall framework of the proposed driver state monitoring system is shown in Figure 4. The process of determining the probability of driver drowsiness by applying a CNN algorithm based on multi-sensor fusion data from data collection for artificial intelligence learning is described.
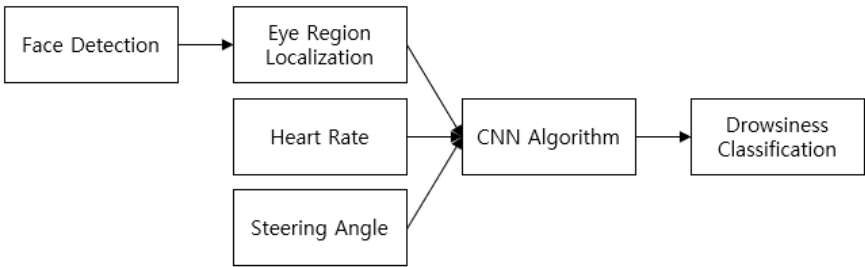

Figure 4. Framework of Driver state monitoring system.

3.2.1. Dataset for AI training

In order to obtain reliable and high quality ground truth for CNN training, we measured EEG data related to the central nervous system and ECG data related to the autonomic nervous system (sympathetic and parasympathetic) in a total of 28 people (14 men and 14 women, with an equal ratio of people in 20s~50s). The collected driver biometric data was then matched with the drowsiness video data and correlated. EEG electrode placement was based on the international 10-20 electrode system. As shown in Figure 5, Fp1, Fp2, F7, F3, Fz, F4, F8, T3, C3, Cz, C4, T4, T5, P3, Pz, P4, T6, O1, O2 A total of 19 EEG locations were selected to measure EEG signals while driving using a driving simulator, with reference electrode A1 attached behind the right earlobe and ground electrode attached to the forehead. The EEG electrodes were placed according to the standard 12-guidance method, and two electrodes were attached to RA(Right Arm) and LA(Left Arm) to measure the Lead I signal.
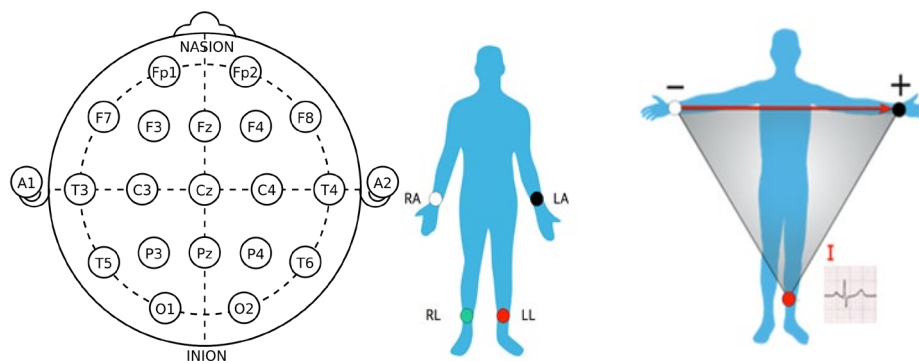

Figure 5. Bio-signal sensing position

The acquired EEG time series data were filtered to the frequency band of interest using BioScan (BioBrain Inc.), a specialized analysis program for bio-signal data, and Fourier transforms were performed to create EEG rhythm data for each frequency domain. Kubios Premium, a heart rate variability (HRV) analysis software, was used to compare the results of preprocessing and analyzing electrocardiogram data. Table 3 shows the results of calculating the spectrum reflecting the quantitative amount of EEG rhythm during the EEG analysis process. The EEG analysis indicators used here are AT, AA, AB, AG, AFA, ASA, ALB, AMB, AHB, RT, RA, RB, RG, RFA, RSA, RLB, RMB, RHB, RST, RMT, RSMT, RAHB, which are defined by wavelength[17], were analyzed in batches using an analysis program, and the abbreviations, full names, and frequency ranges of the indicators calculated through the analysis are shown in Table 4.

Table 3. Results of EEG measurement between drowsiness and normal condition.

Index	Statistic	p-value	Drowsy phase		Normal phase	
			Mean	S.D.	Mean	S.D.
P3_AT	1.74559	0.114837	0.0368098	0.0253582	0.0190881	0.01581871
C3_AT	0.760058	0.466663	0.0323774	0.0182424	0.0261164	0.0241279
F3_AT	2.30342	0.0467372	0.0505931	0.0312916	0.0236024	0.0189182
Fz_AT	-0.702137	0.500348	0.032248	0.0202582	0.0507222	0.0866852
F4_AT	-0.287468	0.780263	0.051055	0.0235523	0.0563553	0.0565932
C4_AT	0.34488	0.738107	0.0326393	0.0209903	0.0306183	0.030243
P4_AT	2.18261	0.0569315	0.0452105	0.0342071	0.0237076	0.0190626
Cz_AT	1.03146	0.329246	0.0729789	0.141693	0.0272307	0.0297778
Fp1_AT	0.687474	0.509111	0.0358831	0.0254056	0.0306054	0.0277164
Fp2_AT	0.660054	0.525749	0.0400698	0.0309069	0.0351219	0.0309297
T3_AT	1.7714	0.110263	0.0285004	0.024878	0.0130038	0.0104697
T5_AT	1.37797	0.201503	0.0571663	0.0864881	0.0169613	0.0163606

O1_AT	2.42004	0.0386078	0.043547	0.026796	0.0217965	0.0180512
O2_AT	2.79673	0.0208236	0.0667467	0.0497118	0.0241085	0.0182638
F7_AT	1.27465	0.234358	0.060407	0.0782538	0.0254135	0.0231573
F8_AT	0.328574	0.749994	0.048504	0.0340096	0.0422901	0.0489318
T6_AT	2.17032	0.0580827	0.059463	0.0533442	0.0221804	0.019894
T4_AT	1.6686	0.129541	0.0414121	0.0332289	0.0256395	0.021453
...						
T4_RFA	4.0154	0.00303903	0.125282	0.0307242	0.097252	0.0280358
P3_RSA	2.53436	0.0320056	0.140471	0.0314127	0.109417	0.0296315
C3_RSA	1.99097	0.0776747	0.112568	0.0364409	0.0934568	0.0343815
F3_RSA	2.1419	0.0608331	0.120932	0.0405498	0.102234	0.0357021
Fz_RSA	0.622018	0.549361	0.130187	0.0339955	0.121796	0.0434039
F4_RSA	1.77442	0.109738	0.135762	0.0455147	0.110789	0.0371417
C4_RSA	2.08924	0.0662654	0.132276	0.0392044	0.108867	0.0265308
P4_RSA	2.23954	0.0518819	0.151325	0.031486	0.123405	0.0323593
Cz_RSA	2.24294	0.051595	0.141195	0.0263102	0.12026	0.0283306
Fp1_RSA	2.22445	0.0531766	0.119558	0.0583884	0.0933145	0.0460626
Fp2_RSA	4.14723	0.00249463	0.110722	0.0506854	0.0842736	0.0451414
T3_RSA	3.42305	0.00759127	0.100524	0.0361446	0.0715178	0.0307846
T5_RSA	2.2196	0.053599	0.149101	0.0527068	0.10974	0.0412117
O1_RSA	1.63455	0.136574	0.154235	0.0556591	0.124454	0.0397242
O2_RSA	1.91773	0.087374	0.158262	0.0411623	0.124784	0.0376523
F7_RSA	1.4229	0.18849	0.117814	0.0362028	0.0980929	0.0478469
F8_RSA	2.65399	0.0263036	0.131613	0.0441668	0.0988031	0.0258764
T6_RSA	1.59353	0.145505	0.144259	0.0331236	0.121232	0.0398572
T4_RSA	4.46852	0.00155799	0.108711	0.0364635	0.0769052	0.0247091
...						

Table 4. Definition of EEG indices and frequency range

No.	Abbreviation	Full Terminology	Frequency Range (Hz)
1	AT	Absolute Theta Power Spectrum	4-8
2	AA	Absolute Alpha Power Spectrum	8-13
3	AB	Absolute Beta Power Spectrum	13-30
4	AG	Absolute Gamma Power Spectrum	30-50
5	AFA	Absolute Fast Alpha Power Spectrum	11-13
6	ASA	Absolute Slow Alpha Power Spectrum	8-11
7	ALB	Absolute Low Beta Power Spectrum	12-15
8	AMB	Absolute Mid Beta Power Spectrum	15-20
9	AHB	Absolute High Beta Power Spectrum	20-30
10	RT	Relative Theta Power Spectrum	4-8/4-50
11	RA	Relative Alpha Power Spectrum	8-13/4-50
12	RB	Relative Beta Power Spectrum	13-30/4-50
13	RG	Relative Gamma Power Spectrum	30-50/4-50
14	RFA	Relative Fast Alpha Power Spectrum	1-13/4-50
15	RSA	Relative Slow Alpha Power Spectrum	8-11/4-50
16	RLB	Relative Low Beta Power Spectrum	12-15/4-50
17	RMB	Relative Mid Beta Power Spectrum	15-20/4-50
18	RHB	Relative High Beta Power Spectrum	20-30/4-50
19	RST	Ratio of SMR to Theta	12-15/4-8

20	RMT	Ratio of Mid Beta to Theta	15-20/4-8
21	RSMT	Ratio of (SMR~Mid Beta) to Theta	12-20/4-8
22	RAHB	Ratio of Alpha to High Beta	8-13/20-30
23	SEF50	Spectral Edge Frequency 50% = Median Frequency	4-50
24	SEF90	Spectral Edge Frequency 90%	4-50
25	ASEF	Spectral Edge Frequency 50% of Alpha Spectrum Band	8-13

Result of analysis the indicator data of AT, AA, AB, AG, AFA, ASA, ALB, AMB, AHB, RT, RA, RB, RG, RFA, RSA, RLB, RMB, RHB, RST, RMT, RSMT, RAHB, the changes in RSA data were more significant. Among them, the data measured in Fp2_RSA showed features that could distinguish between drowsy and normal states, and this value (above 0.120) was used as the standard. Based on this, the trend of drowsy conditions could be confirmed as a result of matching and comparing with the driver's image in the same section. Therefore, ground truth for four sections (P1~P4) within a one-minute time window was obtained by linking video and biometric data, as shown in Figure 6. Then, the probabilistic driver drowsiness of P1~P4 was calculated into three intervals [0, 0.5, 1] to form the training data set, and the simulation test and verification data sets were also formed in the same proportion.
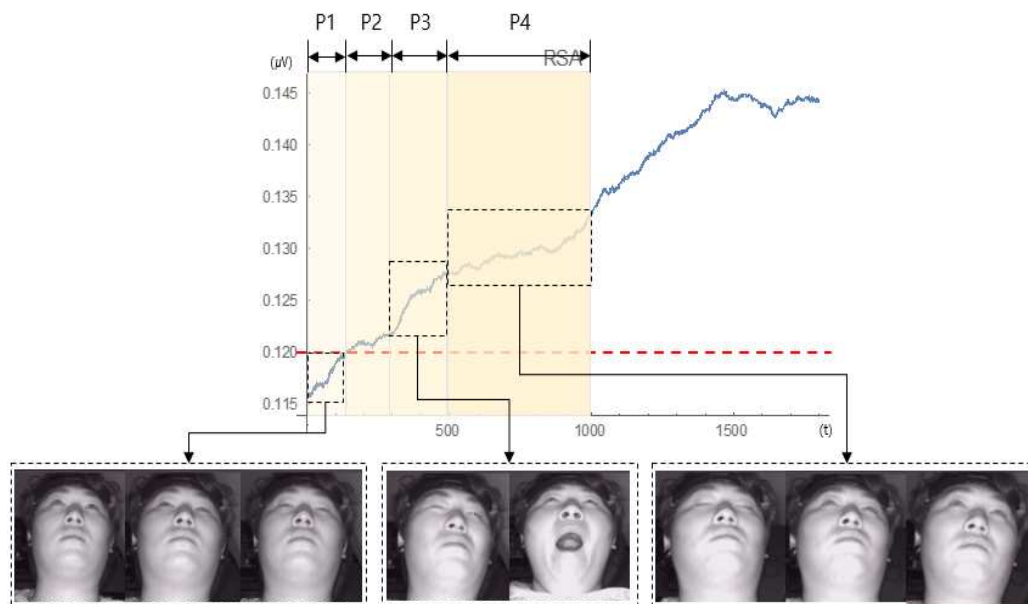

Figure 6. Correlation between EEG and driver image during drowsiness

3.2.2. Facial and Eye Feature Extraction

To extract the region of interest (RoI) of both eyes of the driver, an image of size 24 x 24 is extracted through a large four-step process as shown in Figure 7. First, the Multi-task Cascaded Convolutional Networks (MTCNN) Face Detection model is used to extract the face area of interest from the image data of the ToF camera. Second, we initialize 68 facial feature points set as Mean in the Shape Model, including eyes, nose, mouth, chin, and eyebrows in the recognized face area. Third, we extract the Response Map for 68 feature points using the trained Local Model (CLM) model and finally extract 68 feature points through RLMS (recursive least mean square) optimization. The CLM (Constrained Localization Model) model consists of a shape model by constructing a matrix of feature points of a representative face through principal component (PCA) analysis for the face shape information calculated through PDM (point distribution model), and a patch expert model by extracting feature points for major landmarks and matching them to the correct location. The probability of each landmark is then calculated using the learned parameters. From the 68 extracted points, we calculate the rotation angle for 12 eyelid points to align them horizontally with the center position of the right and left eyes. Finally, the face image region width is reduced to 24%, and the right and left eye images

adjusted to 24x24 pixels in size are extracted using the Affine transform to obtain the right and left eye images located at each center.
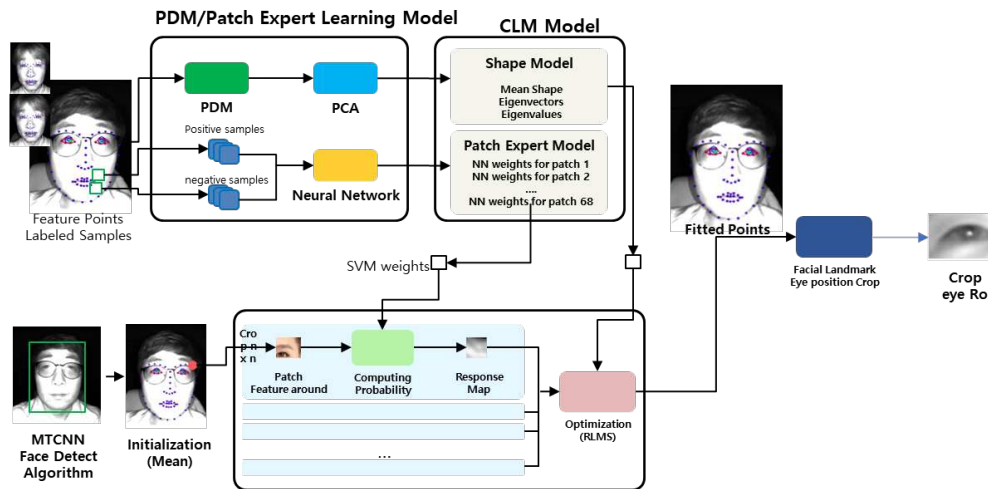

Figure 7. Facial and Eye Feature Extraction process

3.2.3. Calculation of eye aspect ratio(EAR)

The eye aspect ratio (EAR) of both eyes is obtained by a CNN algorithm with images of the driver's eyes as input. As shown in Figure 8, the image of the eye passes through eight 3x3 convolutional layers (stride 1, padding 2, depth 32, 32, 64, 64, 128, 128, 256, and 256 respectively, where ReLU and batch normalization are connected). There are three 2x2 Max pooling layers between the convolution layers and one for every two convolution layers. Next, after conducting Global Max Pooling, connect the Fully Connected Layer to derive the EAR. This way through the network calculates a real number type of EAR. Then, the steering angle data is normalized from pulse data and in-vehicle signals via the ECG sensor embedded in the steering wheel.
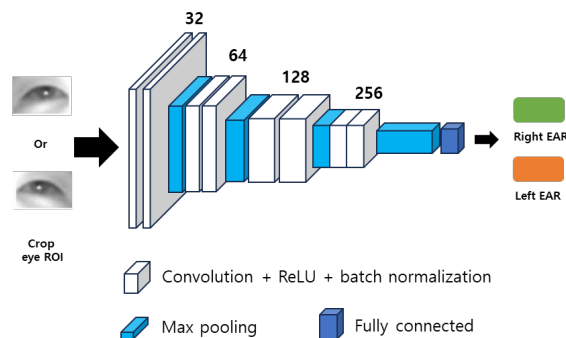
Figure 8. CNN algorithm with EAR.

3.2.4. Proposed CNN algorithm based on multi-sensor fusion data

For real-time driver status judgement, we proposed a system to extract binary drowsiness probabilities for four time series using a convolutional neural network (CNN) based on data on the degree of driver eye closure, pulse, and steering angle for four-time intervals (5s, 15s, 30s, and 60s). Drowsiness induced in short time intervals shows early responsiveness but is a low-accuracy drowsiness estimate, while drowsiness induced in long time intervals is accurate but relatively less responsive than the results estimated in short time intervals. Accuracy and responsiveness were improved with the CNN algorithm using sequential multi-sensor-based fusion data for one minute. Figure 9 shows the process of using multi-sensor fusion data for one minute to construct the proposed system to make four drowsy inferences.
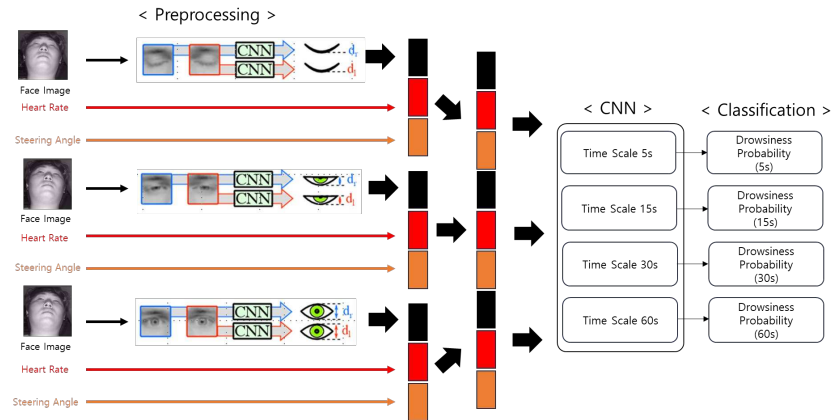
Figure 9. Overview of CNN Model based on multi-sensor fusion data

First, a one-minute face image acquired from the ToF camera is used to extract the RoI regions of the face and eyes using an open-source algorithm, and the EAR (distance between eyelids) is calculated using a CNN algorithm. At the same time, pulse from the ECG sensor embedded in the steering wheel and steering angle data from the in-vehicle network (IVN) are acquired and vectorized to be concatenated in time order for input to the time series CNN algorithm. The concatenated input vectors in four different time section to infer the probability of drowsiness through a convolutional neural network (CNN). Here, the input is a minute-long EAR sequence (30 frames per second, 1800×2 values) associated with both eyes, as well as pulse and steering angle (30x2) data. As shown in Figure 10, the output produces four estimates of the probability of drowsiness, one for each of the different time interval.
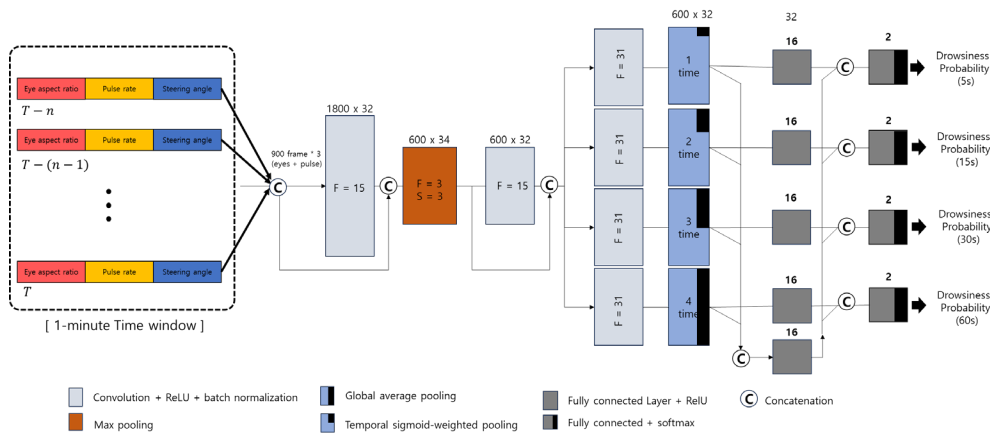

Figure 10. Proposed CNN algorithm based on multi-sensor fusion data.

The input sequence of EAR, biometric, and vehicle information is processed by two temporal convolutional layers. These two layers have a depth of 32 each, with a Receptive Field of 15, stride of 1, and padding of 7. Activation ReLU and batch normalization are connected. These two convolutional layers are densely connected so that the CNN layer outputs and previous inputs are concatenated to produce an output sequence with dimensions of 34 and 66. The output sequence (depth 66) is then passed to four branches, each of which performs a task to infer the probability of drowsiness. Each brunch is configured as follows. First a temporal convolutional layer (depth 32, receptive field k is 31, stride s is 1, padding is 15, ReLU and batch normalization are connected, and there is no dropout connection). Second a different global pooling layer for each brunch is connected. Thirdly, the first fully connected layer (depth 16, followed by ReLU), which is connected to the last fully connected layer (depth 2), which is followed by the softmax function. This configuration allows us to estimate the probability of drowsiness in four different time intervals. In fact, since the ground truth drowsiness signal changes rapidly in short time windows, the drowsiness estimation should be based on short time windows, mainly to react to the rapid change of the EAR closing the eyes. Therefore,

the global pooling of the first 3 brunches is designed to focus on the length of recent fluctuations (for each time scale of 5, 15, and 30 seconds) through a "Time sigmoid weighted pooling" layer. This layer is calculated by the following equation.

$$\alpha(n_0) \ = \ \sum_{n=1}^{600} \frac{\sigma(\frac{3}{2}(n-600-10n_0))}{\sum_{k=1}^{600}\sigma(\frac{3}{2}(k-600-10n_0))} v_n \tag{1}$$

where α is the activation output feature vector, v_n is the feature vector of the nth position in the input sequence, $\sigma(x)$ is represented by $(1 + e^{-x})^{(-1)}$ as a sigmoid function, and n_0 is the time (in seconds) we are interested in. We chose the sigmoid function to allow the temporal weight to decrease quickly and smoothly at n_0. And global pooling in brunch (time scale of 60 seconds) is global average pooling. We also add multiple time-branch contexts to each brunch. the output of the global pooling layer of each brunch is concatenated into one and processed through a Fully Connected layer (with depth 16 and ReLU connections), which then goes back to each brunch and connects with the output of the corresponding first Fully Connected layer. This adds dependencies between branches, which can be robust to drowsiness estimation in low time intervals.

4. Simulation results and discussion

Apart from building a dataset for AI training, men and women under the same conditions (high-medium-low lighting conditions and driver wearing masks, hats, sunglasses, etc. We conducted a simulation test of the proposed CNN algorithm using video data (more than 20,000 images) taken when 28 people drove a public road in the simulator. The test results showed TNR of 71.9%, 88.2%, 90.5%, and 94.8%, TPR of 58.8%, 72.1%, 75.8%, and 76.1%, and Accuracy of 70.4%, 84.3%, 90.2%, and 94.2% for four different time scales(t1-t4), respectively. As a result of complementing each other's features in four different time intervals and comprehensively considering video signals, biometric signals, and vehicle signals, it can be seen that the performance is more than equivalent to the existing driver drowsiness judgement algorithm as shown in the table 5.

Table 5. Comparison of performance measurement with other methods

Conventional Method	Performance measurement	Result
X. Wang et al.[18]	Average recall (3 classes)	56.04%
P. Ebrahim et al.[19]	Average recall (3 classes)	75.9%; 65.2%
E. Vural et al.[20]	Area under the ROC curve (AUC)	0.98
Y. Liang et al.[21]	TNR; TPR; AUC (subject-specific, not -generic)	98%; 67%; 0.92
C. François et al.[22]	TNR; TPR	80%; 72%
I. Garcia et al.[23]	TNR; TPR	95.8%; 85.5%
C. H. Weng et al.[24]	F1 score; Accuracy	85.39%; 84.82%
T. H. Shih et al.[25]	F1 score; Accuracy	82.82%; 82.61%
X. P. Huynh et al.[26]	F1 score; Accuracy	87.97%; 87.46%
Proposed Method	TNR; TPR; Accuracy	t1 : 71.9%; 58.8%; 70.4% t2 : 88.2%; 72.1%; 84.3% t3 : 90.5%; 75.8%; 90.2% t4 : 94.8%; 76.1%; 94.2%

As shown in Table.5 above, it can be said that the characteristics of drowsiness are more expressed as the time window increases. This is because the EAR data-based characteristics are averaged by a long-time window, and the noise is minimized because the proposed algorithm uses the results of previous steps in the final judgement. However, in order to improve responsiveness, good results should be obtained in a short window. Therefore, a trade-off between responsiveness and accuracy is required, and it is necessary to derive the optimal window time. It is also expected that the data from 4 levels can be utilized in driver feedback strategies for real-time response. In addition, there are factors that unexpectedly interfere with detecting the ROI of the driver's eyes (e.g., shielding from hands on the steering wheel and the effects of physical conditions such as driver height), leading

to lower accuracy. Several methods to counteract this could include increasing the camera IR resolution, changing the camera mounting location, adding camera or exception handling.

5. Conclusions

In this paper, we proposed a method to calculate the probability of drowsiness according to four different time intervals based on CNN with multi-sensor data inputs such as video, biometric, and vehicle data to develop a driver state monitoring system. Previous studies have used single-sensor-based driver drowsiness measurement methods, including driver behavior, physiological responses, and changes in vehicle interior environment and behavior. In this study, we derived a CNN algorithm that characterizes the degree of blinking to estimate the driver's drowsiness by using multi-sensor fusion data and used the accumulated changes in one minute to estimate drowsiness. To build a reliable dataset (ground truth) for training the CNN algorithm, we matched the changes in the driver's face image based on the electrocardiogram (ECG) and electroencephalogram (EEG) changes in the drowsy and normal states and established the baseline. This can be objectified a reliable by acquiring and comparing bio-signals as a way to compensate for the uncertainty of the ground truth of drowsiness images in previous studies. In addition, it showed better performance than the preceding drowsiness state determination system, which is judged by other single sensors through multiple sensor data fusion. The results of driver drowsiness judgement derived from the four time criteria suggest that there is a trade-off between fast response and high accuracy, requiring a strategic choice of which performance to prioritize. Of course, the response to the recognition performance degradation problem caused by limiting factors in acquiring the driver's face image should be studied continuous in the future. The results of this study can be applied to driver state monitoring systems that are required when switching control of automated vehicles and are expected to be highly useful in the future.

References

1. ISO/SAE PAS 22736:2021, "Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles", Available online: https://www.iso.org/standard/73766.html(accessed on 29-09-2023)
2. H.N. Yun; S.L. Kim; J.W. Lee; J.H. Yang. Analysis of Cause of Disengagement Based on U.S. California DMV Autonomous Driving Disengagement Report. Transaction of the Korean Society of Automotive Engineers 2018, 26, 464-475. https://doi.org/ 10.7467/KSAE.2018.26.4.464.
3. R. Tian; K. Ruan; L. Li; J. Le; J. Greenberg; S. Barbat. Standardized evaluation of camera-based driver state monitoring systems. IEEE/CAA Journal of Automatica Sinica 2019, 6, 716-732. https://doi.org/10.1109/JAS.2019.1911483.
4. Li, G.; Chung, W.-Y. Detection of Driver Drowsiness Using Wavelet Analysis of Heart Rate Variability and a Support Vector Machine Classifier. Sensors 2013, 13, 16494-16511. https://doi.org/10.3390/s131216494.

5. M. Ramzan; H. U. Khan; S. M. Awan; A. Ismail; M. Ilyas; A. Mahmood. A Survey on State-of-the-Art Drowsiness Detection Techniques. IEEE Access 2019, 7, 61904-61919. https://doi.org/10.1109/ACCESS.2019.2914373.

6. Doudou M.; Bouabdallah A.; Berge-Cherfaoui V. Driver drowsiness measurement technologies: current research, market solutions, and challenges. International Journal of Intelligent Transportation Systems Research 2020, 18, 297–319. https://doi.org/10.1007/s13177-019-00199-w.

7. Mbouna; R. Oyini; S.G. Kong; M. G. Chun. Visual analysis of eye state and head pose for driver alertness monitoring. IEEE Transactions on Intelligent Transportation Systems 2013, 14, 1462–1469. https://doi.org/10.1109/TITS.2013.2262098.

8. S. Maralappanavar; R. Behera; U. Mudenagudi. Driver's distraction detection based on gaze estimation. in Proceedings of IEEE International Conference on Advances in Computing, Communications and Informatics, Jaipur, India, 2016. https://doi.org/ 10.1109/ICACCI.2016.7732431.

9. F. Vicente; Z. Huang; X. Xiong; F. De la Torre; W. Zhang; D. Levi. Driver gaze tracking and eyes off the road detection system. IEEE Transactions on Intelligent Transportation Systems 2015, 16, 2014–2027. https://doi.org/10.1109/TITS.2015.2396031.

10. O. Kopuklu; J. Zheng; H. Xu; G. Rigoll. Driver Anomaly Detection: A Dataset and Contrastive Learning Approach. IEEE Winter Conference on Applications of Computer Vision 2021, 91-100. 2021. https://doi.org/10.48550/arXiv.2009.14660.

11. M. Sabet; R. A. Zoroofi; K. Sadeghniiat-Haghighi; M. Sabbaghian. A new system for driver drowsiness and distraction detection. 20th Iranian Conference on Electrical Engineering, Tehran, Iran, 2012, 1247-1251. https://doi.org/10.1109/IranianCEE.2012.6292547.

12. A. George; A. Routray. Real-time eye gaze direction classification using convolutional neural network. 2016 International Conference on Signal Processing and Communications, Bangalore, India, 2016, 1-5. https://doi.org/10.1109/SPCOM.2016.7746701.

13. F. Zhang; J. Su; L. Geng; Z. Xiao. Driver Fatigue Detection Based on Eye State Recognition. 2017 International Conference on Machine Vision and Information Technology, Singapore, 2017, 105-110. https://doi.org/10.1109/CMVIT.2017.25.

14. M. Awais; N. Badruddin; M. Drieberg. A Hybrid Approach to Detect Driver Drowsiness Utilizing Physiological Signals to Improve System Performance and Wearability. Sensors 2017, 17, 1-16. https://doi.org/10.3390/s17091991.

15. G. Zhenhai; L. DinhDat; H. Hongyu; Y. Ziwen; W. Xinyu. Driver Drowsiness Detection Based on Time Series Analysis of Steering Wheel Angular Velocity. 9th International Conference on Measuring Technology and Mechatronics Automation 2017, Changsha, China, 99-101. https://doi.org/10.1109/ICMTMA.2017.0031.

16. J.G. Gaspar; T. L. Brown; C. W. Schwarz; J. D. Lee; J. Kang; J. S. Higgins. Evaluating driver drowsiness countermeasures. Traffic Injury Prevntion 2017, 18, S58-S63. https://doi.org/10.1080/15389588.2017.1303140.

17. K. S. Kim. The abbreviated word of EEG analysis, index and its definition. BioBrain Inc. 2017, Available online : http://biobraininc.blogspot.kr/.

18. X. Wang; C. Xu. Driver drowsiness detection based on non-intrusive metrics considering individual specifics. Accident Analysis & Prevention 2016, 95, 350-357. https://doi.org/10.1016/j.aap.2015.09.002.

19. P. Ebrahim; A. Abdellaoui; W. Stolzmann; B. Yang. Eyelid-based driver state classification under simulated and real driving conditions. IEEE International Conference on Systems, Man, and Cybernetics 2014, San Diego, CA, USA, 3190-3196. https://doi.org/10.1109/SMC.2014.6974419.

20. E. Vural; M. Cetin; A. Ercil; G. Littlewort; M. Bartlett; J. Movellan. Machine learning systems for detecting driver drowsiness. In-vehicle corpus and signal processing for driver behavior 2009. 97-110.

21. Y. Liang; W. J. Horrey; M. E. Howard; M. L. Lee; C. Anderson; M. S. Shreeve; C. S. O'Brien; C. A. Czeisler. Prediction of drowsiness events in night shift workers during morning driving. Accident Analysis & Prevention 2019, 126, 105-114. https://doi.org/10.1016/j.aap.2017.11.004.

22. C. François; T. Hoyoux; T. Langohr; J. Wertz; J. G. Verly. Tests of a New Drowsiness Characterization and Monitoring System Based on Ocular Parameters. International Journal of Environmental Research and Public Health 2016, 13, 174. https://doi.org/10.3390/ijerph13020174.

23. I. Garcia; S. Bronte; L. M. Bergasa; J. Almazan; J. Yebes. Vision-based drowsiness detector for real driving conditions. 2012 IEEE Intelligent Vehicles Symposium, Alcalá de Henares, Spain, 618-623. 978-1-4673-2118-1.

24. C. H. Weng; Y. H. Lai; S. H. Lai. Driver drowsiness detection via a hierarchical temporal deep belief network. Computer Vision–ACCV 2016 Workshops: ACCV 2016 International Workshops, Taipei, Taiwan, November 20-24, 2016, 117-133. https://doi.org/10.1007/978-3-319-54526-4_9.

25. T. H. Shih; C. T. Hsu. MSTN: Multistage spatial-temporal network for driver drowsiness detection. Computer Vision–ACCV 2016 Workshops: ACCV 2016 International Workshops, Taipei, Taiwan, November 20-24, 2016, 146-153, Lecture Notes in Computer Science, 10118, Springer. https://doi.org/10.1007/978-3-319-54526-4_11.

26.   X. P. Huynh; S. M. Park; Y. G. Kim. Detection of driver drowsiness using 3D deep neural network and semi-supervised gradient boosting machine. Computer Vision–ACCV 2016 Workshops: ACCV 2016 International Workshops, Taipei, Taiwan, November 20-24, 2016, 134-145, Lecture Notes in Computer Science, 10118. Springer. https://doi.org/10.1007/978-3-319-54526-4_10.