IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received August 27, 2021, accepted September 23, 2021, date of publication October 19, 2021, date of current version November 3, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3121543

Emotion Recognition Using a Glasses-Type Wearable Device via Multi-Channel Facial Responses

JANGHO KWON[1,2], JIHYEON HA[1,3], DA-HYE KIM[1,4], JUN WON CHOI[2], (Member, IEEE), AND LAEHYUN KIM[1,5], (Member, IEEE)

[1]Center for Bionics, Korea Institute of Science and Technology, Seoul 02792, South Korea
[2]Electrical Engineering Department, Hanyang University, Seoul 04763, South Korea
[3]Biomedical Engineering Department, Hanyang University, Seoul 04763, South Korea
[4]Brain and Cognitive Engineering, Korea University, Seoul 02841, South Korea
[5]Department of HY-KIST Bio-Convergence, Hanyang University, Seoul 04763, South Korea

Corresponding authors: Laehyun Kim (laehyunk@kist.re.kr) and Jun Won Choi (junwchoi@hanyang.ac.kr)

This work was supported in part by the Institute for Information and Communications Technology Promotion (IITP) Grant by the Korean Government through the Ministry of Science and ICT (MSIT) (Development of non-invasive integrated Brain–Computer Interface (BCI) Software (SW) platform to control home appliances and external devices by user's thought via Augmented Reality (AR)/Virtual Reality (VR) interface) under Grant 2017-0-00432, and in part by the Korea Institute of Science and Technology (KIST) Institutional Program under Grant 2E31122.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Institutional Review Board of KIST under Application No. 2015-012 and Application No. 2016-013.

ABSTRACT We present a glasses-type wearable device to detect emotions from a human face in an unobtrusive manner. The device is designed to gather multi-channel responses from the user's face naturally and continuously while he/she is wearing it. The multi-channel facial responses consist of local facial images and biosignals including electrodermal activity (EDA) and photoplethysmogram (PPG). We had conducted experiments to determine the optimal positions of EDA sensors on the wearable device because EDA signal quality is very sensitive to the sensing position. In addition to the physiological data, the device can capture the image region representing local facial expressions around the left eye via a built-in camera. In this study, we developed and validated an algorithm to recognize emotions using multi-channel responses obtained from the device. The results show that the emotion recognition algorithm using only local facial images has an accuracy of 76.09% at classifying emotions. Using multi-channel data including EDA and PPG, this accuracy was increased by 8.46% compared to using the local facial expression alone. This glasses-type wearable system measuring multi-channel facial responses in a natural manner is very useful for monitoring a user's emotions in daily life, which has a huge potential for use in the healthcare industry.

INDEX TERMS Wearable device, emotion recognition, affective computing, facial expression, biosignal, physiological responses.

I. INTRODUCTION

Emotion recognition is a technology to predict people's emotional states based on user responses such as verbal or facial expressions [1]; this technology can be applied in various fields, such as health care [2], [3], gaming [4], and education [5], [6]. To aid these applications, the technology should recognize emotions in real-time and naturally while the user is experiencing them. Recently, wearable devices have garnered attention for emotion recognition applications [7].

Most existing wearable devices for emotion recognition have relied on biosignals. The biosignals are electronic signals that indicate physiological responses, such as a pulse or a sweating response. These signals originate from changes in the autonomous nervous system (ANS), which is a control system that regulates human bodily functions. With regard to the ANS, there is a risk of misunderstanding a user's emotional state if the wearable devices relies on

The associate editor coordinating the review of this manuscript and approving it for publication was Arianna Dulizia.

the biosignals alone, because the ANS is affected not only by emotion but also by other factors, including cognitive stress [8], [9] or physical activities [10]. For instance, many people in workplaces experience physical activities and cognitive stress, which affect their biosignals; therefore, using only the bio-signal may not be reliable [11]. In this case, it would be desirable to use additional modalities to obtain more reliable emotional information.

Facial expression can be used as an additional modality here as it provides important cues for emotion recognition. This modality is already used in studies on wearable devices [12]. There are two methods for extracting facial expressions via the wearable devices. The first is the sensor-based method, which measures the movements of facial muscles to reflect emotions [11], [12]. This approach might cause discomfort due to contact with the skin on the facial muscles.

The other method is the camera-based method, which captures facial expressions using a camera [13]. This method has advantages over the sensor-based method because the camera is not attached to the skin. Nevertheless, camera-based methods have been not used frequently in wearable devices because the modules were cumbersome and heavy to wear [14]. However, owing to advancements in technology, the sizes of camera modules have become small enough to wear comfortably.

Recently, camera modules have been used in commercial wearable devices [15]. In particular, they have been primarily used in head-mounted wearable devices to capture the user's perspective, and there have also been emotion recognition studies using the captured pictures. However, these studies used pictures to recognize the emotions of other persons and not those of the users [16]. We assume that it is beneficial to use camera modules for monitoring the user's own emotions using these glasses-type wearable devices.

In addition to situational changes, individual differences between users could also affect the use of wearable devices. We have considered two cases as examples. Some people inherently have less sweat gland activity or less heart rate variability [17], [18]. In this case, Biosignal based device might not properly recognize emotions. On the contrary, some people show very little difference in facial expression. In this case, facial expression based wearable devices would not be suitable for such users. Multi-channel facial wearable devices could widen the customer range by increasing the robustness of these device to personal differences.

In summary, to check the feasibility of our idea, we propose the following two hypotheses.

1. The local, side facial image can be used to monitor a user's emotional state.
2. Combining local facial expressions and biosignal yields a better outcome than using a single channel.

To validate this hypothesis, we developed a glasses-type wearable device that can measures multi-channel facial responses; local facial expressions and biosignals. Using the developed device, we conducted emotion recognition experiments with video clips to elicit emotional responses. The experimental results demonstrated that we were able to classify the four emotions with an accuracy of 76.09%, which is higher than the case where a single channel is used.

TABLE 1. Comparison with related studies. ECG: Electrocardiogram, EEG: Electrocardiogram, and SKT: Skin temperature.

Author	Device	Channel	Number of Subject	Accuracy	Protocol
Ours	Glasses Type	Biosignal (EDA, PPG), Local Facial Expression	20	76.09% (Quadrant)	Induced Emotion
[23]	Glasses Type	Facial Expression	3	92.8% (8 emotions)	Forced expression
[24]	Headset	Biosignal (EEG, ECG)	60	75% (arousal) 70% (valence)	Induced emotion
[25]	Glasses Type	Local Facial Expression	4	76% (5 emotions)	Forced expression
[11]	Glasses Type	Local Facial Expression	1	94% (Frown)	Forced expression
[12]	Headset	Biosignal (EMG)	2	99% (smile) 95% (frown)	Forced expression
[26]	Wristband	Biosignal (EDA, ECG, SKT)	2	75.56% (Quadrant)	Induced emotion
[27]	Headset, Wristband	Biosignal (EEG, PPG)	20	77.68% (Quadrant)	Induced emotion

Our study proposed a new wearable system to recognize user's emotions using multi-channel responses including facial expression. The proposed system provides more accurate, stable, and natural emotion recognition for healthcare applications, for instance monitoring patients with mental disorder [19]–[21].

A preprint of this paper is at https://arxiv.org/abs/1905.05360 [22].

A. RELATED WORKS

Table 1 lists related studies. We mainly compared the studies that used glasses-type wearable devices or conducted experiments to recognize the induced emotional state.

Most studies using glasses-type wearable devices were conducted mainly using forced expression-based recognition experiments, which means that the user deliberately made facial expressions. Such experiments lack natural expressions from the user. The most natural way to obtain the emotional state is to record the user's response while performing some task, such as watching a video. Therefore, we designed an experiment to induce emotional states and captured natural expression of facial expression. Although such experiments have been used in studies on biosignal-based wearable devices [24], [26], and [27], they have not been used facial expression-based wearable devices. Therefore, our study could help researchers studying facial expression-based wearable devices for emotion recognition in situations with high naturalness.

FIGURE 1. **Proposed hardware. (1) EDA sensors that measure the skin conductance from the user's nose and mastoid. (2) PPG sensor with an ear clip that measures the pulse of the user's earlobe. (3) Built-in camera that measures the user's facial expressions around the eyes.**

Furthermore, our experiment was conducted with several subjects. Because emotional response patterns are highly dependent on individual differences [14], a large number of subjects is important to generalize the results. In this aspect, our study used a much larger number of subjects than other studies on wearable devices, except for large-scale research [24] and [27]. This number of subjects guarantees the reliability of our results.

We published our collected data online. The dataset is available at https://neurocomputinglab.wixsite.com/neulab/products.

II. METHODS

A. HARDWARE IMPLEMENTATION

The multi-channel wearable device for emotion recognition was designed to extract facial expressions and biosignal. To easily acquire measurements, the device was designed in the form of glasses-type wearable, similar to Google Glass [15] or the prototype sunglasses for emotion recognition presented at CES 2017 [28]. The Internet protocol (IP) camera module was attached to the left side of the device to capture the local facial expressions around the left eye (from the eyebrows to the cheeks). To measure electrodermal activity in response to emotional state change, electrodermal activity (EDA) sensors were attached to the surface of the skin in contact with the nose and mastoid. To perform plethysmography (PPG), ear-clip-type SpO_2 sensors were attached

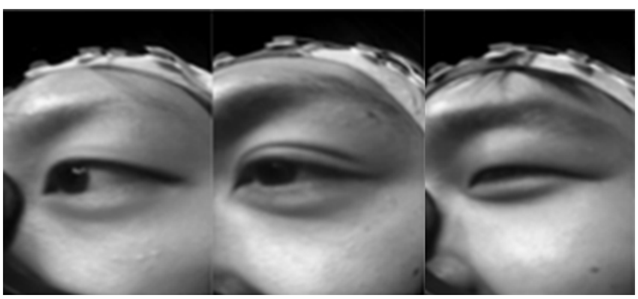


FIGURE 2. **Samples of local facial expressions obtained using the built-in camera in the wearable device.**

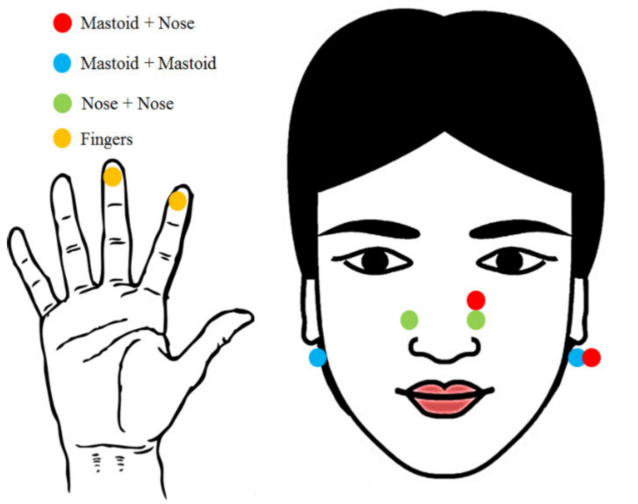


FIGURE 3. **Candidate positions for EDA sensor placement. The orange dots indicate the fingers used to measure the reference signal.**

to the earlobes, which has been used frequently in previous studies [29], [30].

We used wireless communication for transferring acquired data. Two different wireless communication protocols were used for each modality to prevent wireless interference. The facial expression images were transferred via WIFI, and the biosignals were transferred via Bluetooth protocol. All the attached devices were powered using a custom rechargeable Li-polymer battery of capacity 3000 mAh; the devices could be continuously operated for about 2 h. The device was designed to swap out the battery when the battery runs out.

B. DETERMINING EDA SENSOR LOCATION ON THE GLASSES

Fig. 2 shows the facial expression samples acquired from the built-in camera. The parts of facial expression were captured on side of the face.

The position of the EDA sensors was determined more carefully than those of the other sensors. The sweat glands, which are the source of the EDA signal, are distributed with different densities in the body [31]. Therefore, it is necessary to find the best location to contact the EDA sensors. Two facial parts were selected considering natural contact with the device and little change by the facial expression by analyzing the skin movement during the facial expression using the 3D camera [32]; nose and mastoid. Then, we prepared three candidate positions by a combination of these two parts. The first position is on both sides of the nose. This position can be contacted by the nose-supporting part of the glasses. The second is on both sides of the mastoid, which is just behind the ears. This position can be contacted by the earpiece of the glasses. The last position is a combination of the ear and mastoid. The EDA signal measurement was analyzed to compare these positions in the experiments. The experiments were carried out with 10 subjects. The subjects sat on a chair 90 cm away from a 19-inch LCD monitor. EDA sensors were placed in the three candidate positions and on the left index and middle fingers, and the subjects were asked to watch video clips for approximately 90 minutes. The EDA signals were measured simultaneously from all the positions while the subject watched the movie clips. All the procedures in the experiment were certified by the Institutional Review Board of KIST-2015-012.

C. INDUCED EMOTION RECOGNITION

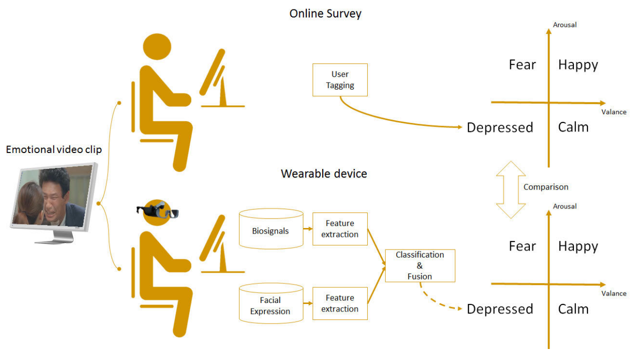


FIGURE 4. **Overall flow diagram of the induced emotion recognition experiment. We first labeled emotional video clips by online survey (upper part) and acquired emotional biosignal and facial expression response via the proposed wearable device while the user watched the labeled emotional video (lower part).**

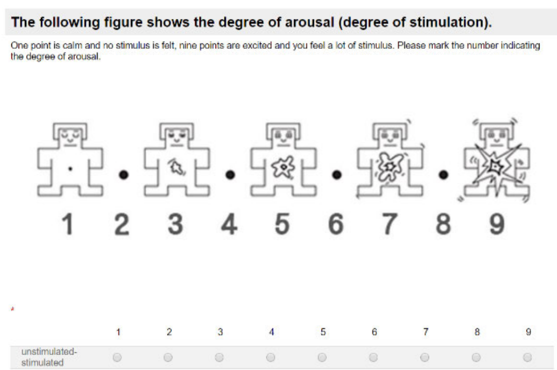


FIGURE 5. **Questionnaire form to evaluate the arousal score resulting from emotional videos in the online survey.**

D. VIDEO STIMULUS SELECTION BY USER TAGGING

We prepared a video-clip-based stimulus to induce emotion in the subjects. We intended to induce the four-dimensional emotional state [33]. The target emotional states include high arousal with high valence (HAHV), high arousal with low valence (HALV), low arousal with low valence (LALV), and low arousal with high valence (LAHV), which correspond to each quadrant of the arousal–valence plane [34].

Two-minute movie clips were used as stimuli to induce emotions. All clips were extracted from Korean movies because the language was important for inducing the HAHV state in the pilot study. The movie clips were carefully selected to induce only one emotion to avoid mixing with other emotions. Ten movie clips were selected for each emotion category. Next, we recruited sixty subjects for the online survey, and each subject ranked the degree of emotion of the clips to determine whether the stimulus induced strong emotions. During the survey, each clip was watched and each subject scored arousal and valence between 1 and 9. Figs. 5 and 6 are the questionnaires after watching each emotional video. For subjects unfamiliar with the concept of arousal and valence, the SAM (Self-Assessment Manikin) pictures were provided. After the survey, the clips were sorted according to the distance from the origin, and two clips were excluded for each quadrant plane, which were close to the origin. As a result, a total of 32 video clips were selected.

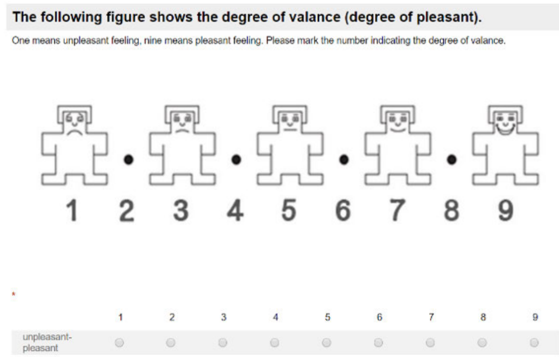


FIGURE 6. **Questionnaire form to evaluate the valence score resulting from emotional videos in online survey.**

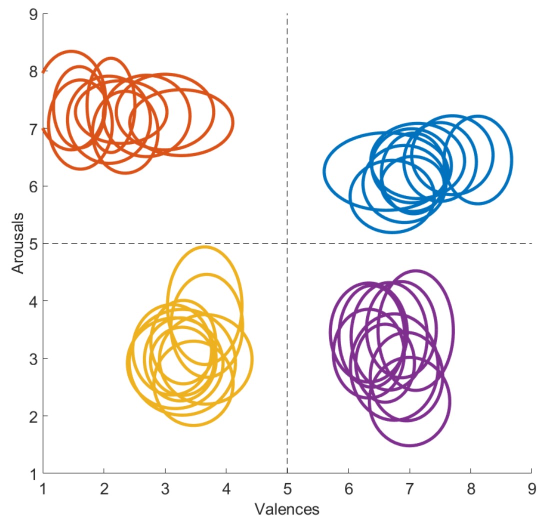


FIGURE 7. **Distribution of SAM scores on arousal–valence plane based on survey results.**

Fig. 7 indicates the distribution of the SAM (Self-Assessment Manikin) scores obtained from the online survey. Each circle represents the video clip. The position of the circle indicates the average SAM score of the clip, and the width and height of the circle represent the standard deviation of the valence and arousal scores of the clip, respectively. The distance from the center (5, 5) was measured. The average distance for the HAHV videos was 2.52 and the standard deviation was 0.48. The average distances for the HALV, LALV, and LAHV videos were 3.56 (standard deviation: 0.51), 2.48 (standard deviation: 0.37), and 2.56 (standard deviation: 0.42).

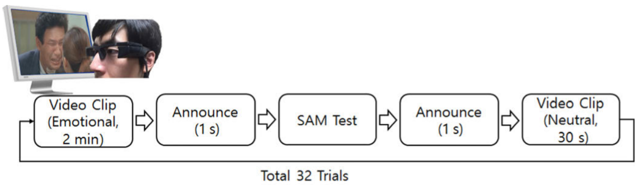


FIGURE 8. Flowchart of emotion-inducing experiment.

E. EXPERIMENT TO ACQUIRE INDUCED EMOTIONAL RESPONSE

Emotion-inducing experiments were conducted using the selected stimuli. The experiment was organized into 32 trials. In each trial, the stimulus clips were shown for 2 min and the neutral clips for 30 s. Neutral clips were shown to neutralize the emotional state between trials. The emotions induced during the trials were counterbalanced to avoid label bias. E-Prime 2.0 [35] was used to present the stimulus.

Experiments were conducted using stimulus clips. A total of 24 subjects participated in the experiment. Data from the four subjects were excluded due to technical error during data acquisition. A 1.7 m $\times$ 1.9 m $\times$ 3.0 m shield room was used for the experiment. The room contained a 19-inch monitor and a two-way speaker. The distance between the subject's head and the monitor was 1 m. To acquire data, the subject wore the wearable device and the physiological signals and facial expressions were recorded while the subject watched the clip. Biosignal data was recorded at 180 Hz and transferred via Bluetooth. The facial expressions were captured at 5 Hz and transferred over Wi-Fi. The data acquisition procedure was manually programmed in MATLAB 2017a. Furthermore, to compare the performance of emotion recognition with the reference, we attached the EDA and PPG sensors to the subject's left hand. A Biosemi Active Two (Biosemi Inc.) was used to acquire the EDA and PPG signals on the subject's finger. The sampling rate of the reference EDA and PPG signals was 512 Hz.

The subjects were in the age group of 21 to 35 years, the average age was 26.7. The experimental protocol was carefully explained to the subjects upon arrival. The experiment consisted of 32 trials. In each trial, a 30-s-long neutral video clip was shown to neutralize the former emotional state of the subject. Then, 2-min-long emotional video clips were shown followed by a 2-s-long black screen. Subjects were requested to rest for 5 min after completing 16 trials to minimize the effect of stress on the physiological signal. The total trial was 1 h and 32 min long, but the entire process, including preparation, took ~2 h. All procedures in the experiment were certified by the Institutional Review Board of KIST-2016-013.

F. DATA PROCESSING

The acquired raw data were processed in a traditional feature-based machine learning manner. Emotion-related features were extracted from each raw data channel. The features were extracted from a 120-s-long biosignal per trial. The observation length for feature extraction was 100 s, and the features were extracted by moving the observed region in increments of 5 s. Therefore, three feature vectors were acquired from each trial and, in total, 160 feature vectors were acquired for each subject. Two types of features were extracted: statistical features and domain-related features.

The statistical features were extracted from the raw signal regardless of the acquired channel. The commonly used statistical features extracted from the raw signal are: (1) average value, (2) standard deviation, (3) mean of absolute value of the first difference, (4) mean of absolute value of the second difference, (5) ratio of mean of absolute value of the first difference and standard deviation, and (6) ratio of mean of absolute value of the second difference and standard deviation.

Table 2 shows the domain-related features that were processed differently for each channel. For the PPG signal, the average acceleration [14] of the pulse and heart rate variability (HRV)-related features was extracted. The HRV features are extracted mainly from the peak-to-peak (PP) interval and power spectral density (PSD) of the PP interval. Fig. 9 shows the PPG signal with PP interval and the corresponding PSD. In the EDA signal, most features were

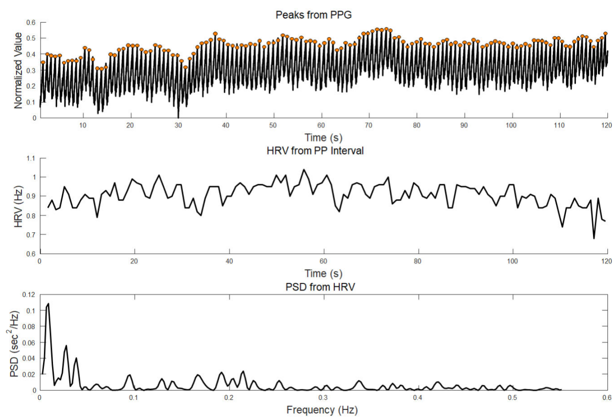


FIGURE 9. Extraction of HRV-related features from PPG signal. the orange dots indicates the detected peak of PPG signal for extract PP(Peak to peak) interval, the orange dots presented upper figure indicate peaks in each pulse to extract PP Interval.

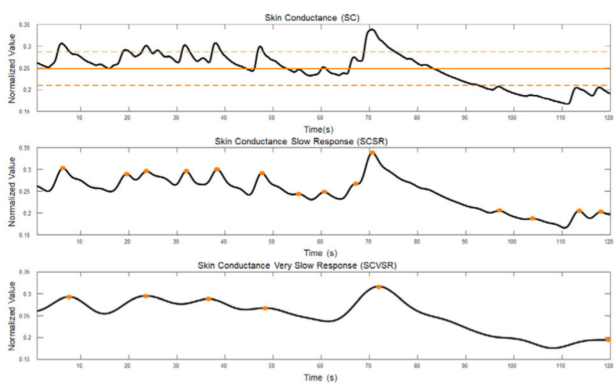


FIGURE 10. Extraction of SCR-related features from EDA signal. the orange line and dashed line indicates the mean and standard deviation of skin conductance signal. the orange dots indicates the local maximum of the signals to extract SCR-related features.

related to the skin conductance response (SCR) [36]. Two low-pass filters were used to acquire the SCRs in different time resolutions. First, the raw signal was passed through a low-pass filter with 0.2 Hz cut-off frequency, which was called the skin conductance slow response (SCSR). Second, a low-pass filter with a cut-off frequency of 0.08 Hz was applied, and the signal was called the very slow response of the skin conductance (SCVSR). The SCR-related features were extracted from each preprocessed signal. we extracted the number of the SCR in the SCSR and the SCVSR, and the average amplitude of the SCR in the SCSR and the SCVSR. and we also extract the recovery time of the SCR in the SCSR.

After extracting the biosignal features, a feature-selection method was applied to filter the features that did not vary between emotional states. The selection algorithm used was ReliefF [37].

TABLE 2. Types of domain-related features.

Channels	Features
PPG	(1) Average acceleration of pulse, (2) Number of interval differences of successive NN intervals greater than 50ms (NN50), (3) The root mean square of the successive differences (RMSSD), (4) Standard deviation of the NN interval (SDNN), (5) Standard deviation of differences between adjacent NN intervals (SDSD), (6) Power in very low frequency (VLF) range (≤ 0.04 Hz), (7) Power in low frequency(LF) range (0.04-0.15Hz), (8) Power in high frequency(HF) range (0.15Hz – 0.4Hz)
EDA	(1) Ratio of recovery time to signal length in SCSR, (2) Number of SCRs in SCSR, (3) Number of SCRs in SCVSR, (4) Average amplitude of SCRs in SCSR, (5) Average amplitude of SCRs in SCVSR

To process the response to facial expression, we used the Fisherface method [38]. We used this method because this method does not need facial landmarks that cannot be acquired in our local facial expression. Labels were determined according to the clip the subject watched, regardless of whether the subject actually expressed emotion. First, principal component analysis (PCA) was applied to reduce the dimensionality of the captured images. The principal components and the eigenvalues were extracted from the training dataset and sorted based on the magnitude of the eigenvalues. We rejected the eigenvectors with the two largest eigenvalues because the components that have large eigenvalues typically describe illumination changes [38], not expression changes. Additionally, we also rejected components with small eigenvalues to remove noise. The smallest eigenvectors that summation of its eigenvalues accounts for 15% of the total summation were rejected.

After obtaining the refined eigenvectors W_{PCA}, The pixel values of the captured images were mapped on to the selected components. The projected values were analyzed using linear discriminant analysis (LDA) to extract more discriminant features based on the emotional state of the user. A weight matrix W was obtained from the LDA, which satisfies the following expression [38]:

$$W_{LDA} = argmax_W \frac{|W^T W_{PCA}^T S_B W_{PCA} W|}{|W^T W_{PCA}^T S_W W_{PCA} W|}$$

The extracted feature vectors were used to train the emotion recognition model. We used the support vector machine with RBF kernel as the emotion recognition model. The size of the SVM kernel was 0.5 for facial expression and 0.2 for other modalities.

To recognize the arousal and valence states, we used the binary RBF SVM model. To recognize quadrant emotional states, we implemented a two-level classification model because combining binary classifiers in multiple levels could be a more optimal solution to recognize quadrant emotional states that can be decomposed into arousal and valence states [39]. Therefore, we first recognized the user's arousal and valence states separately and then combined each state to estimate the user's quadrant emotional state.

G. FEATURE FUSION AND CLASSIFICATION

Our wearable device can acquire the user's emotional response from biosignals and facial expressions simultaneously. To recognize the emotional state of these two individual sources of information, we fused acquired biosignal and facial expression features. The most critical challenge in combining these features was that the sampling rate in each channel was different. The built-in camera module can capture 500 sequences of images while extracting a biosignal feature vector within a 100-s-long biosignal. Therefore, we attempted to extract a representative expression image from the sequential images. However, simple averaging is not effective because there are so few meaningful expressions in the sequence of expression images as the emotional expressions occur in a short period of time. In particular, the initial sequences contain few emotional expressions. This is because we designed the stimulus video clip to gradually induce the user's emotional state. Therefore, we designed the algorithm to extract meaningful expression data from the facial images sequence with respect to the biosignal features. Algorithm 1 describes the extraction of meaningful expression data. We first divided the user eigenface scores into two groups using the k-means clustering algorithm. Then, we computed the group membership of the sequence in an initial 30 seconds to find the group with emotional expressions. Finally, the representative fisherface was obtained by averaging the emotional expression group. The obtained fisherface score was used to match facial expression features.

Finally, we validated the feasibility of our wearable system by comparing the performance of emotion recognition. There are four validations to measure the performance. The first was the accuracy using reference biosignal. These signals include PPG, the GSR signal from the user's left hand, which was acquired by the Biosemi ActiveTwo. The second was the validation using wearable biosignal and facial expression independently, the last was the accuracy using the feature-fused modality.

We validated each channel in a leave-one-trial-out manner. In a preliminary study [40], we evaluated the data in 10-fold cross-validation with random sampling. This data split has the problem that the training data was extracted and the test

Algorithm 1 Pseudo Code for Emotion Recognition

Input:
S_{Facial}: sequence of facial expression images
S_{EDA}, S_{PPG}: sequence of facial EDA biosignal, facial PPG biosignal
$T_{initial}$: initial time to start emotion recognition (second)
T_{final}: final time to stop emotion recognition (second)
Output: Emotion recognition results using the wearable system at each time
Assign theinitial variable: offset $= 0$, $T_{final} = 120$sec
While $T_{window\_end} \leq T_{final}$:
  // Set the window to extract the matched features
  $T_{window\_end} = T_{initial} + \text{offset} \cdot (5\text{sec})$
  $T_{window\_start} = T_{window\_end} - 100\text{sec}$
  P_{EDA}, P_{PPG}: sequence of facial EDA biosignal and PPG biosignal within time series $[T_{window\_start}, T_{window\_end}]$
  P_{Facial}: sequence of facial expression images within time series $[T_{window\_start}, T_{window\_end}]$
  // Obtain matched biosignal features
  Extracted EDA features F_{EDA} and PPG features F_{PPG} from P_{EDA}, P_{PPG}
  $F_{Biosignal}$: Concatenation of F_{EDA} and F_{PPG}
  // Extract emotional eigenface expression
  N: number of the eignface within time range $[T_{window\_start}, T_{window\_end}]$
  P: number of data within early 30 second time range $[T_{window\_start}, T_{window\_start}+30\text{sec}]$
  $S_{PCA} = \{s_1, s_2, \ldots, s_N\}$: Eigenface scores by projecting P_{Facial} onto W_{PCA}
  $L_{PCA} = \{l_1, l_2, \ldots, l_N\}$, $l_i \in \{1,2\}$: Label of eigenface scores in two unsupervised groups using k-means clustering
  $\text{count\_g1} = \sum_{i=1}^{i=P} (l_i == 1)$
  $\text{count\_g2} = \sum_{i=1}^{i=P} (l_i == 2)$
  if count_g1 $\geq$ count_g2
    EmotionalExp $= 2$
  else
    EmotionalExp $= 1$
  $F_{PCA} = \frac{\sum_1^N (l_i == EmotionalExp) \cdot S_i}{\sum_1^N (l_i == EmotionalExp)}$
  // Obtain matched facial feature vector
  F_{LDA}: Fisherface scores by projecting F_{PCA} onto W_{LDA}
  // Obtain matched fusion features
  F_{Fusion}: Concatenation of $F_{Biosignal}$ and F_{LDA}
  // Obtain and display emotion recognition results
  ER_Result = ClassificationModel(F_{Fusion})
  Display or Validate the ER_Result
  offset $+= 1$
  repeat

data was close in time. This temporal similarity could lead to a high correlation between the test and training datasets. Therefore, to prevent this temporal similarity problem, we set the data in one trial as a test dataset and let other datasets for training and repeat this validation for all trials.

III. RESULTS

A. THE RESULT OF EDA SENSOR LOCATION EXPERIMENT

Before the emotion recognition experiment, we experimented with EDA correlation between facial parts and reference locations. The results of the EDA measurement experiments show that the highest correlation with fingers was shown when the sensors were placed near the nose and mastoid. The average correlation of the mastoid and nose is 0.6757, which is higher than those in other locations, and the standard deviation between subjects is 0.1663, which is lower than those in other locations. This result explains why the EDA electrodes are placed on the mastoid and nose contact parts of the glasses.

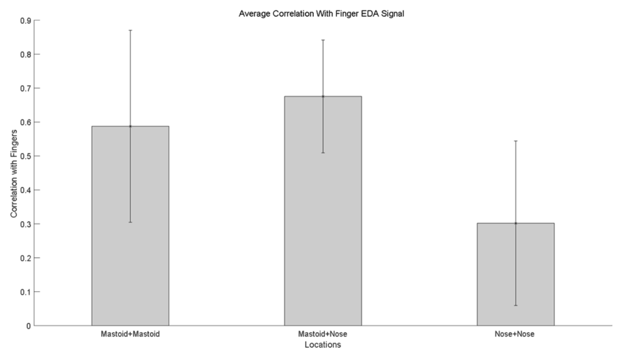


FIGURE 11. Correlation coefficients between fingers and candidate facial sensor placements.

B. THE RESULT OF THE INDUCED EMOTION RECOGNITION EXPERIMENT

Table 3 presents the average accuracies for each channel and the fusion method. Although the recognition performance in the fusion method is not always better than that of other methods, its average score was better than those of those other wearable modalities.

Fig. 15 shows samples of the acquired fisherface. In the acquired facial expression, it appears that the eyebrow and upper cheek regions are highly activated compared to other regions. The fisherface for the valence state shows higher activation at the laughter line near the mouth compared to the fisherface for the arousal state.

Table 5 shows a comparison of the average emotion recognition rates between male participants ($n = 10$) and female participants ($n = 10$). The results indicate that the female participants show better emotion recognition rates for facial expressions, which is consistent with the results in [41], which implies that women use facial expressions more frequently than men. It is observed that the system can recognize emotional states more easily from frequent facial expressions. On the contrary, male participants show better emotion recognition rates for facial biosignal than female participants.

IV. DISCUSSION

Existing wearable devices for emotion recognition may not yield good and stable results due to dynamic situations in

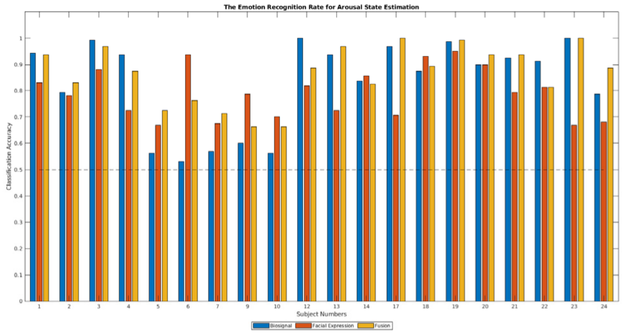


FIGURE 12. The emotion recognition rates for estimating arousal state for each subject. The dashed black line indicates the probability of random chance.

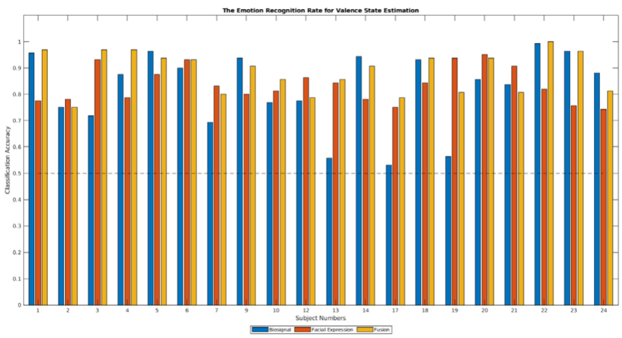


FIGURE 13. The emotion recognition rates for valence state estimation for each subject. The dashed black line indicates the probability of random chance.

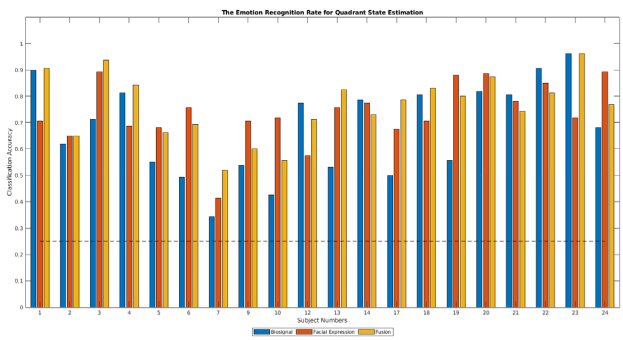


FIGURE 14. Emotion recognition rates for quadrant state estimation for each subject. The dashed black line indicates random chance probability.

TABLE 3. Average accuracies in emotion recognition for each modality. The number in parentheses indicates the number of target emotional states.

	Arousal (2)	Valence (2)	Quadrant (4)
Biosignal	0.8313	0.8197	0.6763
Facial Expression	0.7916	0.8359	0.6678
Feature fusion	0.8641	0.8844	0.7609
Reference Biosignal	0.9049	0.8664	0.7842

real life and individual difference of emotional response. To solve this problem, this study proposed a new wearable system to improve emotion recognition performance by using physiological signals and facial image simultaneously.

TABLE 4. Confusion matrices for estimation of the quadrant state. For improved clarity, the names of each quadrant emotion have been replaced with similar basic emotions; Funny replaced HAHV, Disgust replaced HALV, Depressed replaced LALV, and Calm replaced LAHV.

	Wearable			
	Funny	Disgust	Depressed	Calm
Funny	448(14%)	99(3.09%)	56(1.75%)	197(6.16%)
Disgust	103(3.22%)	484(15.13%)	209(6.53%)	4(0.13%)
Depressed	0(0%)	36(1.13%)	622(19.44%)	142(4.44%)
Calm	17(0.5%)	21(0.6%)	152(4.75%)	610(19.06%)
	Facial Expression			
	Funny	Disgust	Depressed	Calm
Funny	549(17.16%)	58(1.81%)	16(0.5%)	177(5.53%)
Disgust	143(4.47%)	495(15.47%)	124(3.88%)	38(1.19%)
Depressed	42(1.31%)	116(3.63%)	532(16.63%)	110(3.44%)
Calm	121(3.78%)	33(1.03%)	85(2.6%)	561(17.53%)
	Feature Fusion			
	Funny	Disgust	Depressed	Calm
Funny	575(17.97%)	104(3.25%)	16(0.5%)	145(4.53%)
Disgust	58(1.81%)	550(17.19%)	145(4.53%)	7(0.22%)
Depressed	12(0.38%)	62(1.94%)	688(21.50%)	78(2.44%)
Calm	43(1.34%)	5(0.16%)	90(2.81%)	622(19.44%)

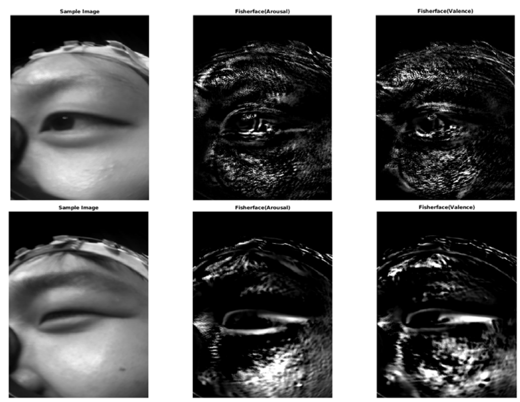


FIGURE 15. Samples of the acquired fisherface for the arousal state estimation(center) and valence state estimation(right).

TABLE 5. Average accuracies for emotion recognition for each gender group. The number in parentheses indicates the number of target emotional states.

	Arousal (2)		Valence (2)		Quadrant (4)	
	Male	Female	Male	Female	Male	Female
Biosignal	0.8438	0.8188	0.8219	0.8175	0.6881	0.6644
Facial	0.7738	0.8094	0.8188	0.8531	0.6450	0.6906
Fusion	0.8613	0.8669	0.8506	0.9181	0.7331	0.7888
Reference Biosignal	0.8875	0.9206	0.8694	0.8637	0.7681	0.7988

The multi-modal wearable device has a strength that can be applied to various situations in real life. In addition to improving recognition accuracy, using two modalities has an advantage, for example; when in a situation the facial expression did not reveal effectively, such as a huge illumination change, the biosignal could compensate the facial expression. In contrast, facial expression could compensate for the biosignal under some condition that the biosignal could not work effectively, such as a condition that the user has cognitive stress. Additionally, the results in Table 5 indicate that multi-modal approach could be robust to gender bias of emotion recognition. Table 5 shows that the biosignal-based emotion recognition rates were higher in male users, while the facial expression-based emotion recognition rates were higher in female users. The proposed wearable device could be a reasonable approach to compensate for gender-based bias, because it uses both modalities, which could expand the range of selective use of advantageous features according to gender.

We designed the glasses-typed wearable device to get the advantage of a multi-channel wearable device. To find an optimal facial location to measure the biosignal, we conducted EDA signal acquisition experiments by comparing the correlation of the EDA signal with fingers and facial locations. The results indicate that it is best to place EDA sensors on the nose and mastoid rather than on the noses and mastoids only. The higher correlation could be obtained if we place the EDA sensors in other locations, such as the forehead [31]. However, other locations, except the nose and mastoid, could result in a noisy EDA signal due to skin movement by facial expression, which was experimented in a preliminary study [32]. In this study, we found the best locations for EDA sensors in reasonable locations that can be used in facial wearable devices.

According to the results of the EDA correlation experiment, the wearable device was designed to measure the sweat response of the mastoid and nasal skin. The built-in camera acquired local facial expression on the side of the user's face.

The initial motivation of our study was the compensation effect between each channel. Unfortunately, this compensation effect was not fully covered in the experiment because we did not control the change in the user's situation, such as giving a stress [8], we will experiment to prove that the multi-modal wearable device has robust emotion recognition performance in various situations. Although the experiment did not cover the advantages of the fusion modality in situational changes, it validated the benefits of the fusion modality with respect to personal differences. We believe that the experimental results shown in Figs. 11–13 agree with the main hypothesis of our study. As shown in the figures, the use of a single modality is not always suitable for all subjects. For example, subject 19 shows poor recognition performance using biosignal, but subject 24 shows poor recognition performance using facial expression, but the other channel compensated for these poor emotional responses. Table 1 shows the overall accuracies. although these compensatory effects of the fusion modality are not shown in all subjects, these effects are clear in overall average performance. The fusion of two modalities shows better performance than any of the single modalities. The classification performance increased by 8.46% compared to the method using biosignal alone. Additionally, as we observed during the experiment, the individual differences were found. Some

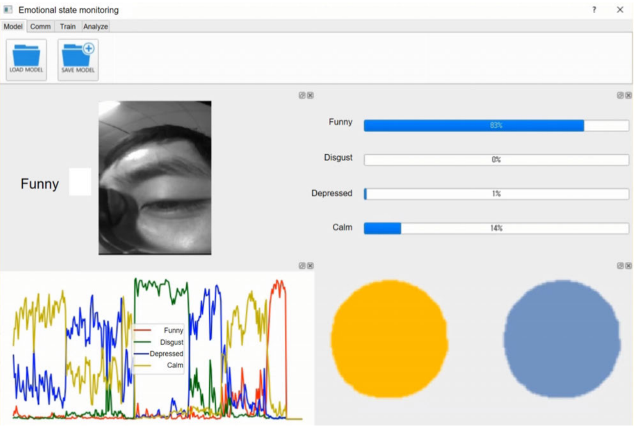


FIGURE 16. An example application of the proposed emotional recognition system. Each emotional state is logged in real-time (bottom of the figure), and current emotional state is determined by most dominant emotional state (left side of the figure).

subjects remained mostly expressionless and some subjects rarely presented EDA responses during the emotion-inducing experiment.

The confusion matrix in Table 4 shows that using a combination of biosignals and facial expressions reduces the confusion of the negative emotional states (i.e., disgust and depressed). For example, confusion cases that predict the disgust state as the funny state or cases that predict the depressed state as the calm state are clearly less in number when a fusion of modalities is used compared to single modalities, while the other number of confusion cases are between the number of two single modalities. This may indicate that the fusion of two facial modalities has a synergistic effect in detecting negative emotional states. Therefore, the fusion of two channels could lead to a more accurate monitoring of negative emotional state for healthcare, such as monitoring depressed state [19].

Fig. 16 shows an example application of our emotion recognition system for mental health care. In this application, the ratio of quadrant emotional states is logged in order to monitor subtle and complicated emotional change in the real world. The user can observe daily emotional state patterns using the logged emotional state and get more detailed assistance from mental health care.

Compared with other studies on emotion recognition based on facial expressions, the system proposed here shows a lower performance on emotion recognition [23], [42], [43]. However, most comparative studies on wearable devices classified data created by intentional facial expressions rather than data created by eliciting actual emotions. Therefore, we suggest that the lower accuracy is due to differences in the experimental paradigms. Compared to other studies recognized the induced emotional state, our study shows higher emotion recognition rates for the classification of the valence state and arousal state [24].

Compared with existing multi-modal emotion recognition studies [33], [36], the existing studies measured biosignals or facial expressions at a standard measurement location. For example, the sweat response uses the EDA response of the

TABLE 6. The source and description of video stimulus used in the emotion-inducing experiment. The descriptions presented in the table are all understandable in each clip without the prior experience watching the source.

	Source name	Type of source	Description	Target emotion
1	Beatles code	Talk show	A Comedian host teasing a guest in a talk show	HAHV (Funny)
2	Infinite challenge	Entertainment program	A broadcast writer is dancing awkwardly	HAHV (Funny)
3	Running man	Entertainment program	Two entertainers are having a funny conversation	HAHV (Funny)
4	Radio star	Talk show	Two entertainers prepare some comedy and show in	HAHV (Funny)
5	Non-summit	Talk show	Two entertainers prepare some comedy and show in	HAHV (Funny)
6	Gag concert	Comedy show	A couple have a secret dating at workplace	HAHV (Funny)
7	Radio star	Talk show	A guest telling a funny story	HAHV (Funny)
8	Radio star	Talk show	A non-native guest telling a story about misheard	HAHV (Funny)
9	Shadows in the Palace	Movie (horror)	A maid teacher torture her student with needle	HALV (Disgust)
10	Bedevilled	Movie (horror)	The abused wife brutally murders her husband	HALV (Disgust)
11	Inside Men	Movie (criminal)	A gangster is being cut his arm in exchange for betrayal	HALV (Disgust)
12	Blind	Movie (thriller)	A woman runs away from murderer chasing her	HALV (Fear)
13	New World	Movie (criminal)	A gangster is fighting other gangsters with a knife in an elevator	HALV (Disgust)
14	The Divine Move	Movie (criminal)	A Gangster boss violently revenge against the traitor	HALV (Fear)
15	I Saw the Devil	Movie (thriller)	A man is hitting murder's head with bat	HALV (Disgust)
16	Old boy	Movie (thriller)	A man is cutting his tongue as an apologies	HALV (Disgust)
17	The Princess' Man	TV series (drama)	Last farewell with the princess and her husband ahead of the death penalty	LALV (Sad)

TABLE 6. *(Continued.)* The source and description of video stimulus used in the emotion-inducing experiment. The descriptions presented in the table are all understandable in each clip without the prior experience watching the source.

18	Closer To Heaven	Movie (drama)	A conversation between a patient who is about to die and his girlfriend	LALV (Sad)
19	EBS Documentary Prime	Documentary	Patients who are about to die and their families at the hospital	LALV (Sad)
20	Misaeng: Incomplete Life	TV series (drama)	A middle-aged man is being scolded by his boss and frustrated	LALV (Depressed)
21	Misaeng: Incomplete Life	TV series (drama)	A working mother is frustrating at a situation where she cannot take good care of her child.	LALV (Depressed)
22	Misaeng: Incomplete Life	TV series (drama)	A middle-aged man talks to a co-worker who is hospitalized for overwork	LALV (Depressed)
23	Misaeng: Incomplete Life	TV series (drama)	A middle-aged man talking to a subordinate who fell out of an internship	LALV (Depressed)
24	Signal	TV series (criminal)	A man is crying at the cinema	LALV (Sad)
25	Lump sugar	Movie (drama)	A girl is playing with a horse in the meadow	LAHV (Calm)
26	Scandal Makers	Movie (comedy)	A boy is fell in love with a girl in the kindergarten	LAHV (Calm)
27	Ode to My Father	Movie (drama)	A man and a woman are dating	LAHV (Calm)
28	You Are My Sunshine	Movie (drama)	A man and a woman are having a conversation lying down under the cherry blossom	LAHV (Calm)
29	One Fine Spring day	Movie (drama)	A man and a woman recording the sound of nature in the bamboo forest	LAHV (Calm)
30	Welcome to Dongmakgol	Movie (drama)	Villagers are sharing food in the fields after harvest	LAHV (Calm)
31	The Way Home	Movie (drama)	A boy having a conversation with his grandmother in her house	LAHV (Calm)
32	The Classic	Movie (drama)	A man and an woman is running in the rain, their head covered with man's clothes	LAHV (Calm)

finger, and the facial expression uses a frontal view of the full face, and EEG response from sensors attached precise locations on scalp under the hair. Our study, on the other hand, measured the user response in areas confined to the face, because the proposed devices measured user's facial responses without modification the shape of glasses as much as possible, such as facial expressions from the side of the eye and sweat responses of the nasal skin. Therefore, the modalities used in our experiment differ from those considered in existing multi-modal emotion recognition studies. Our study could provide support to multi-modal emotion recognition researchers and facial wearable device developers to explore more applicable modalities in the real world.

A limitation of our study is that there was no scale to measure the amount of emotion induced by stimulation. Although we carefully collected the emotional state to induce exactly one emotional state, which might increase the ambiguity of the labeled data. We recommend that a dominance scale [34] should be adopted the experiment, which will be helpful to develop more accurate emotion recognition models in future studies. The results of our study may be of value to other researchers studying multi-modal emotion recognition in wearable devices.

V. CONCLUSION

In this study, to increase the reliability of emotion recognition, we propose a glasses-type wearable device that measures local facial expressions in addition to biosignal. The facial expressions were acquired in an unobtrusive manner using a camera, and the location of the biosensors on the wearable device was determined by signal measurement experiments. Experiments with video clips were conducted to evaluate the performance of the device. Our results show that the glasses-type wearable device can be used to accurately estimate a user's emotional state.

The proposed method can expand the range of healthcare applications using wearable devices, such as monitoring physical and mental health states. In the future, we hope to develop a wearable device capable of monitoring user emotions with respect to various tasks in daily life.

APPENDIX

See Table 6.

ACKNOWLEDGMENT

The authors would like to thank all subjects for their participation in the data acquisitions. They would also like to thank editors and anonymous reviewers who have given insightful comments.

An earlier version of this paper was presented at https://arxiv.org/abs/1905.05360 [22]. The only citable version of the paper is this one.

REFERENCES

[1] M. Brackett, "Emotional intelligence," *Character Lab Playbook*, vol. 21, p. 91, Aug. 2019.

[2] M. Fahim, M. Idris, R. Ali, C. Nugent, B. Kang, E.-N. Huh, and S. Lee, "ATHENA: A personalized platform to promote an active lifestyle and wellbeing based on physical, mental and social health primitives," *Sensors*, vol. 14, no. 5, pp. 9313–9329, May 2014.

[3] C. D. Katsis, G. Ganiatsas, and D. I. Fotiadis, "An integrated telemedicine platform for the assessment of affective physiological states," *Diagnostic Pathol.*, vol. 1, no. 1, p. 16, Dec. 2006.

[4] E. Reynolds, "Nevermind: Creating an entertaining biofeedback-enhanced game experience to train users in stress management," in *Proc. ACM SIGGRAPH Posters (SIGGRAPH)*, 2013, p. 1.

[5] C.-H. Wu, Y.-M. Huang, and J.-P. Hwang, "Review of affective computing in education/learning: Trends and challenges," *Brit. J. Educ. Technol.*, vol. 47, no. 6, pp. 1304–1323, Nov. 2016.

[6] R. Mehmood and H. Lee, "Towards building a computer aided education system for special students using wearable sensor technologies," *Sensors*, vol. 17, no. 2, p. 317, Feb. 2017.

[7] K. Wac and C. Tsiourti, "Ambulatory assessment of affect: Survey of sensor systems for monitoring of autonomic nervous systems activation in emotion," *IEEE Trans. Affect. Comput.*, vol. 5, no. 3, pp. 251–272, Jul. 2014.

[8] C. Setz, B. Arnrich, J. Schumm, R. L. Marca, G. Tröster, and U. Ehlert, "Discriminating stress from cognitive load using a wearable EDA device," *IEEE Trans. Inf. Technol. Biomed.*, vol. 14, no. 2, pp. 410–417, Mar. 2010.

[9] N. Nourbakhsh, Y. Wang, F. Chen, and R. A. Calvo, "Using galvanic skin response for cognitive load measurement in arithmetic and reading tasks," in *Proc. 24th Austral. Comput.-Hum. Interact. Conf. (OzCHI)*, 2012, pp. 420–423.

[10] R. A. Calvo and S. D'Mello, "Affect detection: An interdisciplinary review of models, methods, and their applications," *IEEE Trans. Affective Comput.*, vol. 1, no. 1, pp. 18–37, Jan. 2010.

[11] J. Scheirer, R. Fernandez, and R. W. Picard, "Expression glasses: A wearable device for facial expression recognition," in *Proc. CHI Extended Abstr. Hum. Factors Comput. Syst. (CHI)*, 1999.

[12] A. Gruebler and K. Suzuki, "Design of a wearable device for reading positive expressions from facial EMG signals," *IEEE Trans. Affect. Comput.*, vol. 5, no. 3, pp. 227–237, Jul. 2014.

[13] A. Teeters, R. El Kaliouby, and R. Picard, "Self-Cam: Feedback from what would be your social partner," in *Proc. ACM SIGGRAPH Res. Posters (SIGGRAPH)*, 2006, p. 138.

[14] R. W. Picard, E. Vyzas, and J. Healey, "Toward machine emotional intelligence: Analysis of affective physiological state," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 10, pp. 1175–1191, Oct. 2001.

[15] T. Starner, "Project glass: An extension of the self," *IEEE Pervas. Comput.*, vol. 12, no. 2, pp. 14–16, Apr. 2013.

[16] P. Washington, C. Voss, N. Haber, S. Tanaka, J. Daniels, C. Feinstein, T. Winograd, and D. Wall, "A wearable social interaction aid for children with autism," in *Proc. CHI Conf. Extended Abstr. Hum. Factors Comput. Syst.*, May 2016, pp. 2348–2354.

[17] W. Prokasy, *Electrodermal Activity in Psychological Research*. Amsterdam, The Netherlands: Elsevier, 2012.

[18] H. Kobayashi, "Inter- and intra-individual variations of heart rate variability in Japanese males," *J. Physiolog. Anthropol.*, vol. 26, no. 2, pp. 173–177, 2007.

[19] T. Roh, S. Hong, and H.-J. Yoo, "Wearable depression monitoring system with heart-rate variability," in *Proc. 36th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2014, pp. 562–565.

[20] M. Taj-Eldin, C. Ryan, B. O'Flynn, and P. Galvin, "A review of wearable solutions for physiological and emotional monitoring for use by people with autism spectrum disorder and their caregivers," *Sensors*, vol. 18, no. 12, p. 4271, Dec. 2018.

[21] G. Valenza, M. Nardelli, A. Lanata, C. Gentili, G. Bertschy, R. Paradiso, and E. P. Scilingo, "Wearable monitoring for mood recognition in bipolar disorder based on history-dependent long-term heart rate variability analysis," *IEEE J. Biomed. Health Informat.*, vol. 18, no. 5, pp. 1625–1635, Sep. 2014.

[22] J. Kwon and L. Kim, "Emotion recognition using a glasses-type wearable device via multi-channel facial responses," 2019, *arXiv:1905.05360*. [Online]. Available: http://arxiv.org/abs/1905.05360

[23] K. Masai, Y. Sugiura, M. Ogata, K. Kunze, M. Inami, and M. Sugimoto, "Facial expression recognition in daily life by embedded photo reflective sensors on smart eyewear," in *Proc. 21st Int. Conf. Intell. User Interface*, Mar. 2016, pp. 317–326.

[24] J. Marín-Morales, J. L. Higuera-Trujillo, A. Greco, J. Guixeres, C. Llinares, E. P. Scilingo, M. Alcañiz, and G. Valenza, "Affective computing in virtual reality: Emotion recognition from brain and heartbeat dynamics using wearable sensors," *Sci. Rep.*, vol. 8, no. 1, p. 13657, Dec. 2018.

[25] S. Hickson, N. Dufour, A. Sud, V. Kwatra, and I. Essa, "Eyemotion: Classifying facial expressions in VR using eye-tracking cameras," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2019, pp. 1626–1635.

[26] B. Zhao, Z. Wang, Z. Yu, and B. Guo, "EmotionSense: Emotion recognition based on wearable wristband," in *Proc. IEEE SmartWorld, Ubiquitous Intell. Comput., Adv. Trusted Comput., Scalable Comput. Commun., Cloud Big Data Comput., Internet People Smart City Innov. (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*, Oct. 2018, pp. 346–355.

[27] B. Nakisa, M. Rastgoo, A. Rakotonirainy, F. Maire, and V. Chandran, "Long short term memory hyperparameter optimization for a neural network based emotion recognition framework," *IEEE Access*, vol. 6, pp. 49325–49338, 2018.

[28] B. Andy, *Smith Lowdown Focus: Our First Take*. Portland, OR, USA: Digital Trends, 2017.

[29] K. Shin, Y. Kim, S. Bae, K. Park, and S. Kim, "A novel headset with a transmissive PPG sensor for heart rate measurement," in *Proc. 13th Int. Conf. Biomed. Eng.*, 2009, pp. 519–522.

[30] M.-Z. Poh, N. C. Swenson, and R. W. Picard, "Motion-tolerant magnetic earring sensor and wireless earpiece for wearable photoplethysmography," *IEEE Trans. Inf. Technol. Biomed.*, vol. 14, no. 3, pp. 786–794, May 2010.

[31] M. van Dooren, J. J. G.-J. de Vries, and J. H. Janssen, "Emotional sweating across the body: Comparing 16 different skin conductance measurement locations," *Physiol. Behav.*, vol. 106, no. 2, pp. 298–304, May 2012.

[32] B. Kim, D. Ki, C. Lee, S. Kim, L. Kim, and G. H. Kwon, "Functional design on glasses type wearable device using CAD/CAM," in *Proc. Soc. CAD/CAM Conf.*, 2015, pp. 827–829.

[33] S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, "DEAP: A database for emotion analysis; Using physiological signals," *IEEE Trans. Affective Comput.*, vol. 3, no. 1, pp. 18–31, Oct./Mar. 2012.

[34] J. A. Russell, "A circumplex model of affect," *J. Pers. Soc. Psychol.*, vol. 39, no. 6, p. 1161, 1980.

[35] *E-Prime 2.0*, I. Psychology Software Tools, Sharpsburg, PA, USA, 2012.

[36] M. Soleymani, J. Lichtenauer, T. Pun, and M. Pantic, "A multimodal database for affect recognition and implicit tagging," *IEEE Trans. Affect. Comput.*, vol. 3, no. 1, pp. 42–55, Jan. 2012.

[37] M. Robnik-Šikonja and I. Kononenko, "An adaptation of Relief for attribute estimation in regression," in *Proc. 40th Int. Conf. Mach. Learn., (ICML)*, 1997, pp. 296–304.

[38] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 711–720, Jul. 1997.

[39] J. Kim and E. André, "Emotion recognition based on physiological changes in music listening," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 12, pp. 2067–2083, Dec. 2008.

[40] J. Kwon, D.-H. Kim, W. Park, and L. Kim, "A wearable device for emotional recognition using facial expression and physiological response," in *Proc. 38th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug. 2016, pp. 5765–5768.

[41] D. McDuff, E. Kodra, R. E. Kaliouby, and M. LaFrance, "A large-scale analysis of sex differences in facial expressions," *PLoS ONE*, vol. 12, no. 4, Apr. 2017, Art. no. e0173942.

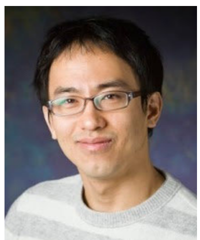
[42] Y. Liu, Y. Li, X. Ma, and R. Song, "Facial expression recognition with fusion features extracted from salient facial areas," *Sensors*, vol. 17, no. 4, p. 712, Mar. 2017.

[43] D. Mehta, M. Siddiqui, and A. Javaid, "Facial emotion recognition: A survey and real-world user experiences in mixed reality," *Sensors*, vol. 18, no. 2, p. 416, Feb. 2018.

JANGHO KWON received the B.S. degree from the University of Seoul, Seoul, South Korea, in 2015, and the M.S. degree from the University of Science and Technology, Daejeon, South Korea, in 2017. He is currently pursuing the Ph.D. degree with the Electrical Engineering Department, Hanyang University, Seoul. He has been a Student Research Scientist at the Center for Bionics, Korea Institute of Science and Technology, Seoul, since 2018.

JIHYEON HA received the B.S. degree from Hanyang University, Seoul, South Korea, in 2016, where he is currently pursuing the Ph.D. degree with the Biomedical Engineering Department. He has been a Student Research Scientist at the Center for Bionics, Korea Institute of Science and Technology, Seoul, since 2016.

DA-HYE KIM Ph.D. candidate of the Department of Brain and Cognitive Engineering, Korea University, Seoul, South Korea. Until 2012, she had been a Researcher with the National Rehabilitation Center. Since 2013, she has been performing research with the Center for Bionics, Korea Institute of Science and Technology, Seoul. Her research interests include neural networks, BCI, and rehabilitation.



JUN WON CHOI (Member, IEEE) received the B.S. and M.S. degrees from the Electrical Engineering Department, Seoul National University, and the Ph.D. degree in electrical computer engineering from the University of Illinois at Urbana–Champaign. In 2010, he joined Qualcomm, San Diego, USA, where he participated in research on advanced signal processing technology for next-generation wireless systems. Since 2013, he has been with the Electrical Engineering Department, Hanyang University, as a Faculty Member. His research interests include signal processing, machine learning, intelligent vehicles, and wireless communications. He has served as an Associate Editor for the IEEE Transactions on Vehicular Technology and the IEEE Transactions on Intelligent Transportation Systems.

LAEHYUN KIM (Member, IEEE) received the Ph.D. degree in computer science from the University of Southern California, Los Angeles, CA, USA. He is currently a Principal Research Scientist with the Center for Bionics, Korea Institute of Science Technology, Seoul, South Korea, and a Professor with the Department of HY-KIST Bio-Convergence, Hanyang University, Seoul. His research interests include haptics, medical robotics, rehabilitation, and BCI.