IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received October 3, 2019, accepted October 22, 2019, date of publication October 25, 2019, date of current version November 6, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2949707

The Fusion of Electroencephalography and Facial Expression for Continuous Emotion Recognition

DAHUA LI[1], ZHE WANG[1], CHUHAN WANG[1], SHUANG LIU[1], WENHAO CHI[1], ENZENG DONG[1], XIAOLIN SONG[2], QIANG GAO[1], AND YU SONG[1]

[1]Tianjin Key Laboratory for Control Theory and Applications in Complicated Systems, Tianjin University of Technology, Tianjin 300384, China
[2]Engineering Training Center, Tianjin University of Technology, Tianjin 300384, China

Corresponding authors: Qiang Gao (gaoqiang@tjut.edu.cn) and Yu Song (jasonsongrain@hotmail.com)

This work was supported in part by the Fundamental Research on Advanced Technology and Engineering Application Team, Tianjin, China, under Grant 20160524, and in part by the Natural Science Foundation of Tianjin under Grant 18JCYBJC87700.

ABSTRACT Recently, the study of emotion recognition has received increasing attentions by the rapid development of noninvasive sensor technologies, machine learning algorithms and compute capability of computers. Compared with single modal emotion recognition, the multimodal paradigm introduces complementary information for emotion recognition. Hence, in this work, we presented a decision level fusion framework for detecting emotions continuously by fusing the Electroencephalography (EEG) and facial expressions. Three types of movie clips (positive, negative, and neutral) were utilized to elicit specific emotions of subjects, the EEG and facial expression signals were recorded simultaneously. The power spectrum density (PSD) features of EEG were extracted by time-frequency analysis, and then EEG features were selected for regression. For the facial expression, the facial geometric features were calculated by facial landmark localization. Long short-term memory networks (LSTM) were utilized to accomplish the decision level fusion and captured temporal dynamics of emotions. The results have shown that the proposed method achieved outstanding performance for continuous emotion recognition, and it yields 0.625 ± 0.029 of concordance correlation coefficient (CCC). From the results, the fusion of two modalities outperformed EEG and facial expression separately. Furthermore, different numbers of time-steps of LSTM was applied to analyze the temporal dynamic capturing.

INDEX TERMS Continuous emotion recognition, EEG, facial expressions, signal processing, decision level fusion, temporal dynamics.

I. INTRODUCTION

Emotion is a psychophysiological process of perception and cognition to object or situation, and it plays an important role in human-human natural communication. However, the emotion recognition was neglected in the field of human-computer interaction (HCI). Due to the explosion of machine learning in the cognitive science, affective computing has emerged to integrate emotion recognition into HCI. Nowadays, emotion recognition system aims to establish a harmonious HCI by endowing computers with the ability to recognize, understand, express and adapt to human emotions [1]. Hence, it provides potentially applications for emotion recognition in many fields, such as human robot interaction (HRI) [2], safe driving [3], social networking [4] and distance education [5]. These applications manifested the different modalities interaction which provides complementary information to improve the precision and robust of the emotion recognition system.

In order to represent emotions, the discrete models and dimensional models were proposed by psychologists [6]. The discrete emotion models turned emotion recognition into classification problem. Six basic emotions (happiness, anger, sadness, surprise, fear and disgust) can be recognized as prototypes from which other emotions are derived [7]. However, the emotions expressed in communication are complex, and one basic emotion can hardly describe the human feeling under certain situation. In addition, emotions could be mapped in multi-dimensional spaces that could maximize the largest variance of all the possible emotions [8]. The valence-arousal plane is one of the well-known dimensional models of emotion that maps the emotions into a two-dimensional circular space. Valence ranges from negative to positive and arousal is the concentration of physical

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

activity from calm to excited [9]. Based on the valence-arousal plane, the emotions can be represented to the dimensional space in continuous time. The core of continuous emotion recognition is to analyze the time-varying affect phenomenon that are elicited by stimulus [10]. Different materials are utilized to elicit emotions such as affective pictures [11], music videos [12], movie clips [13], and conversation [14]. The characteristics of time-varying visual and audio stimuli in context makes movies seem to be the one of the most effective ways to elicit emotions [15].

In this paper, we focused on continuous movie-induced emotion recognition by the fusion of EEG and facial expression. The goal of this work is to improve the precision of emotion recognition. Three classes of movie clips that represent the negative, neutral, and positive emotions were utilized to elicited emotions, and the EEG signals and facial expressions of the subjects are recorded simultaneously. Firstly, facial expressions response of subjects watching movie clips were annotated continuously by ten annotators. The averaged continuous annotations were considered as the ground truth for the regression of continuous emotions. Next, EEG is preprocessed by 4-47 Hz bandpass filtering and independent component analysis (ICA) based spatial filtering to weaken the influence of eyes gazing and head moving. PSD features of different frequency bands were extracted from EEG by short-time Fourier transform (STFT) and feature selection algorithm t-Distributed Stochastic Neighbor Embedding (t-SNE) was applied to eliminate irrelevant or redundant features to improve the regression performance. For the facial expression, facial geometric features are extracted by the facial landmark localization model. The predictions of EEG and facial expressions were obtained by the Support Vector Regression (SVR) separately. Finally, LSTM was used to accomplish the decision level fusion and temporal dynamics capturing for the continuous emotion recognition. The performance of different numbers of time-steps of LSTM was analyzed for temporal dynamic capturing. The main contributions are as follows. (ş) We adopted an LSTM-based decision level fusion framework for the fusion of EEG and facial expressions based emotion recognition. (şş) Temporal dynamics of emotion were captured and analyzed to improve the precision of continuous emotion recognition.

The structure of the remaining parts of this paper is as follows. The related work of multimodal emotion recognition and temporal dynamics of emotions are given in Section 2. The methods for the fusion of EEG and facial expression based continuous emotion recognition are displayed in Section 3. The experimental results are presented in Section 4. The discussions of the results are shown in Section 5. Finally, the paper is concluded in Section 6.

II. RELATED WORK

A. MULTIMODAL EMOTION RECOGNITION

The performance of multimodal emotion recognition is generally better than single modality [7]. Recent advances in emotion recognition tended to evaluate the affects by multi-modalities. Zheng *et al.* [16] proposed EEG and eye-tracking combined emotion recognition method. The results showed that both feature level fusion and decision level fusion of EEG signals and eye tracking can improve the performance of emotion recognition model. Song *et al.* [17] designed a multi-modal physiological emotion database, including EEG, galvanic skin response (GSR), respiration (RSP) and electrocardiogram (ECG). Physiological signals are sampled for discrete emotion recognition. The attention-long short-term memory (A-LSTM) algorithm was proposed to extract discriminative features to improve classification accuracy. The multimodal emotion databases DEAP [12] and MAHNOB HCI [18] including both physiological signal and facial expression response were widely used in emotion recognition researching. Huang *et al.* [8] performed binary classification of the statues of valence and arousal by combining the EEG and facial expressions. The best accuracies of valence and arousal were achieved 80.30% and 74.23% for the DEAP database. Soleymani *et al.* [19] fused EEG and facial expressions for movie-induced continuous emotion recognition, the results demonstrated that EEG signals could provide the complementary information in presence of facial expressions.

B. THE TEMPORAL DYNAMICS OF EMOTIONS

Many researchers focused on time window based method to model the complex correlations of emotions in the temporal. Liu *et al.* [20] clipped EEG data every 1 second, and use the 2 s window with 50% overlap to construct temporally related EEG data segment. Hence, the current emotional state are predicted by three consecutive EEG data segment. That is potential to improve the performance of online emotion recognition system. Khosrowabadi *et al.* [21] proposed a biologically inspired feedforward neural network for emotion recognition by the EEG signal. This network concludes six layers, and genetic algorithms are applied for selecting the optimum length of EEG data to achieve the short-term memory at the second layer. Compared with various feature extraction algorithms, the proposed network achieved the best accuracies of valence and arousal classification. Nicolaou *et al.* [22] proposed output-associative relevance vector machine (OA-RVM) to learn non-linear input and output associations and temporal dynamics. OA-RVM significantly outperforms both SVR and RVM in classification accuracies by employing a predefined time windows, which could induce the learning of temporal dynamics. Thus, temporal dynamics capturing is potential to improve the robustness and precision of emotion recognition.

Recurrent neural network (RNN) is the classic model for time series prediction, and it is widely used in emotion recognition. Weninger *et al.* [23] changed mean squared error (MSE) loss function of RNN to the CCC [24], it improved the performance of continuous emotion recognition obviously. Banda *et al.* [25] proposed Nonlinear Auto Regressive with eXogenous inputs recurrent neural network (NARX-RNN) that speeds up convergence and increases

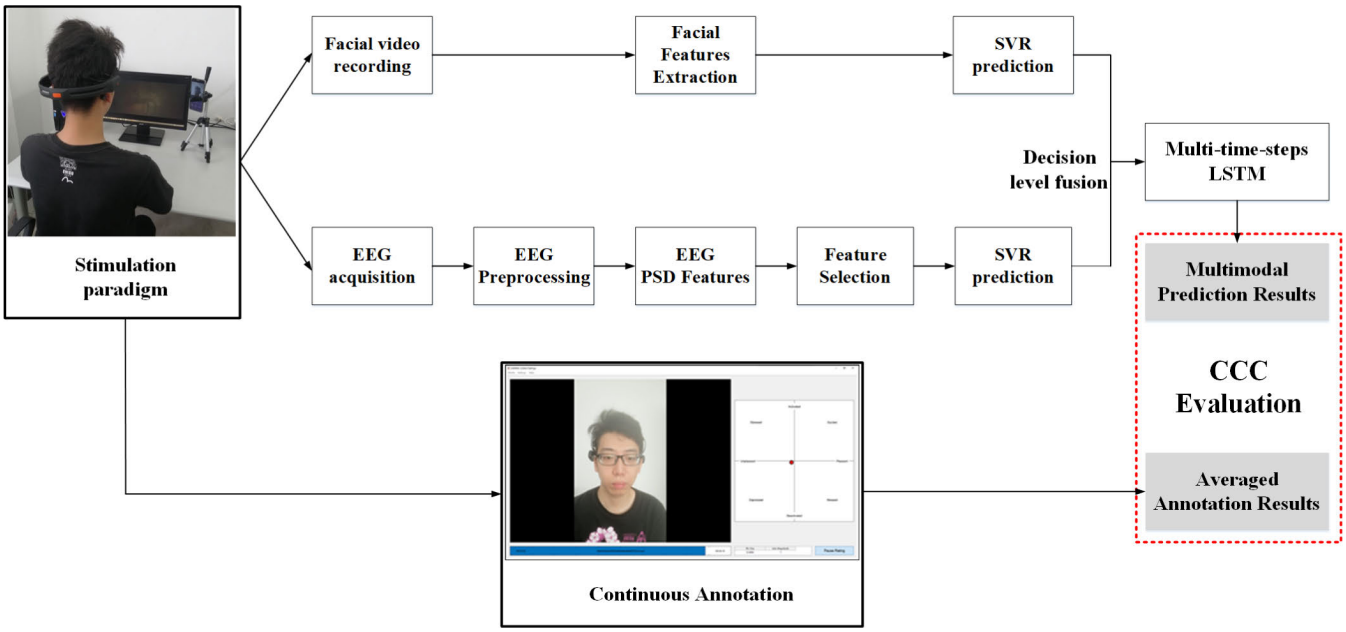


FIGURE 1. An overview of the fusion of EEG and facial expression based continuous emotion recognition framework.

generalization ability. To solve the potential gradient vanishing and explosion problem of RNN, the LSTM was proposed which added gate controlling and memory cells on the basis of RNN to obtain the ability to learn long-range dependencies. Nakisa *et al.* [26] proposed a new framework to automatically optimize LSTM hyperparameters using differential evolution (DE) and that improve the accuracy effectively. Moreover, bidirectional LSTM (BLSTM) was also used to explore the complex relationship between features and labels [27], [28]. However, the current emotional state could not be predicted by future information in practical applications. The LSTM has the advantages in time series analysis and long-range dependencies learning for temporal dynamics capturing in continuous emotion recognition.

III. MATERIALS AND METHODS

A. ARCHITECTURE

In this paper, EEG and facial expressions were fused to constitute the continuous emotion recognition system. Firstly, EEG and facial expressions are recorded by the EEG acquisition device and camera during the experiment separately. Facial expressions response of subjects was annotated continuously, the averaged continuous annotations were considered as the ground truth. Next, PSD features are extracted from EEG, and then the feature selection was also necessary to achieve. For the facial expression, the facial geometric features are calculated by the facial landmark localization model. Finally, the LSTM is utilized to accomplish the decision level fusion and capture the temporal dynamics of emotions. The continuous multimodal emotion recognition framework is described in Figure 1.

B. EXPERIMENT PARADIGM

For inducing spontaneous emotions of the subjects, the movie clips from the SEED dataset [29] are chosen as stimuli in this work. The film clips are divided into three classes which are positive, neutral and negative. Table 1 shows the details of selected film clips.

TABLE 1. The description of SEED dataset film clips.

No.	Label of the film clips	Title of film clips
1	positive	Lost in Thailand
2	positive	Flirting Scholar
3	neutral	World Heritage in China I
4	neutral	World Heritage in China II
5	negative	Aftershock
6	negative	Back to 1942

Ten healthy subjects participated in the experiment, including 5 males and 5 female age from 22 to 26. The process of the experiment is shown in Figure 2. Six film clips were presented to each subject. The length of each clip is about 200 s ($M = 201.33$ s, $SD = 17.18$ s). We asked subjects to keep the body as fixed as possible during the experiment. Before viewing film clips a prompt will appear on the screen, the subject should keep in the relax state for the next 10 s which aims to get the baseline to capture the changing of emotion. EEG and facial expression response are recorded simultaneously in duration of movie clips viewing. The subjects can have a rest between the two sessions for 30 s.

The EEG data was sampled by the Emotiv EPOC headset. Emotiv EPOC is a low-cost dry electrode EEG signal acquisition equipment, including 14 acquisition channels and a 2-axis gyroscope, following the international 10-20 position layout: AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4,

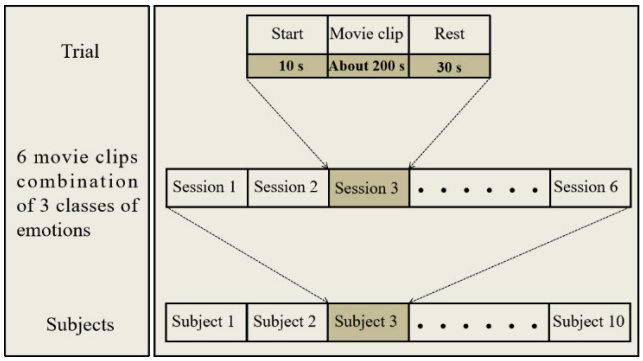


FIGURE 2. The process of the experiment.

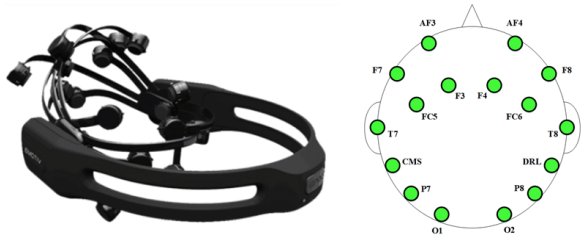


FIGURE 3. The Emotiv EPOC headset and electrodes distribution.

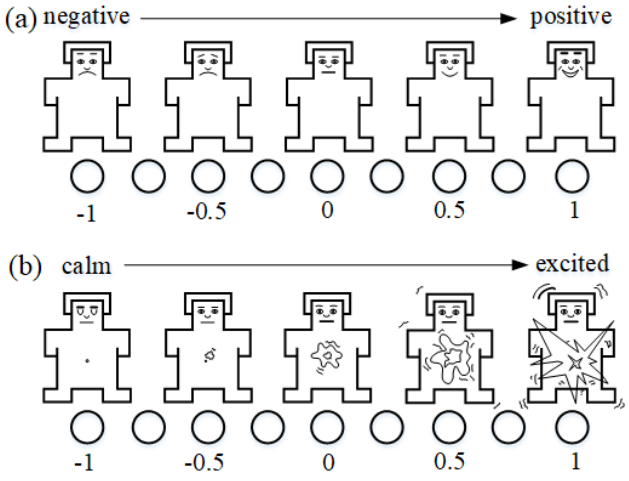


FIGURE 4. The Valence and Arousal of the SAM. (a) Valence and (b) Arousal.

F8, AF4, which also has the reference electrode (CMS/DRL) of the ear. Figure 6 shows the distribution of all electrodes of the Emotiv EPOC headset. The internal sampling rate of EPOC is 2048 Hz, which is converted to a 128 Hz output by down-sampling, and data is transmitted to the computer via Bluetooth. In this work, we choose all the 14 acquisition channels. The EEG and facial expressions are recorded simultaneously. We applied the HUAWEI Honor 8x mobile phone (30 FPS, 1920x1080 image size) fixed with a holder as the low-cost facial expression sampling equipment.

In this work, 10 postgraduates majored in psychology was guided to finish the continuous annotation of subjects' facial expression response. We applied DARMA for

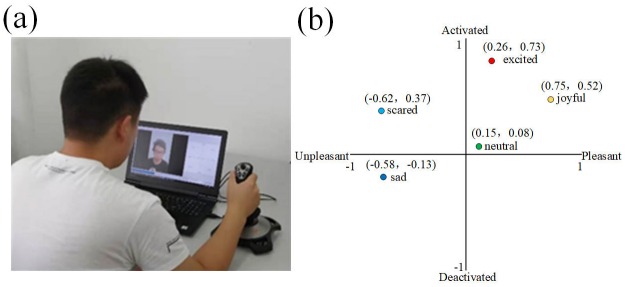


FIGURE 5. (a) The annotation environment and (b) annotation examples.

emotion annotation. DARMA is a media annotation program that collects continuous ratings of valence and arousal while displaying audio and video files and the annotators can complete continuous emotion annotation visually with a joystick [30]. The annotation environment and annotation examples are shown in Figure 5. Here, the values of valence and arousal were defined by the Self-Assessment Manikin (SAM). It is a generally accepted model of emotion evaluation that uses the image of a cartoon manikin to represent the values of different dimensions of emotion are applied as criterion of emotion evaluation [31]. Figure 4 shows the valence and arousal of SAM, the annotators could complete the emotion annotation according to the state of the manikins. Each dimension of the emotion was continuously valued in the range of -1 to 1. Each annotator needs to finish the valence rating for 60 facial videos. The sample rating of the joystick is 20 Hz, and the averaged value of 1 s was as the real output. Considering that the subjects carefully and quietly watch the movie clips during the experiment, it is difficult for the annotators to rate arousal by facial videos of subjects [19]. Therefore, only the valence annotations were used to evaluate the emotional state in this study.

C. EEG PREPROCESSING AND FEATURE EXTRACTION

The raw EEG signal contains a variety of artifacts, especially the Electrooculogram (EOG) and utility frequency. The preprocess of EEG is accomplished by the EEGLAB which is the toolbox for physiological signal analysis on Matlab [32]. The original EEG data are processed by a 4-47 Hz bandpass filter to reduce the utility frequency artifacts. The EOG related to eyes movement which could be rejected by topographic maps. The ICA can decompose the EEG data into 14 (Number of the acquisition channels) independent components. The independent components can be visualized by of EEG field as 2-D circular views which maps the original EEG data to the spatial domain. Ultimately, the EEG signal will be reconstructed for feature extraction.

The EEG signal can be seen as a type of non-stationary random signal that is hard to extract features only in the time domain. However, time-frequency analysis can capture the dynamic change in the time and frequency domain. Hence, we used STFT to decompose signal into small segments by the time window, and each segment is considered to be stationary signal approximately. The EEG data are mapped

to the time-frequency plane by STFT, and the Hamming window is chosen as the short-time analysis window. The PSD features of EEG signals in different bands are correlated with emotions, and theta bands (4-7 Hz), alpha bands (8-12 Hz), beta bands (13-30 Hz) and gamma bands (30-47 Hz) are the general selections [33]. The 56 features (14 channels $\times 4$ frequency bands) was selected for representing the EEG signal. The subtraction of power spectrum between the movie clips viewing durations and neutral durations (10 s before the movie clips viewing) of each subject are calculated. This may present the change of emotions more obviously and minimize individual variability to some extent. Finally, PSD of theta, alpha, beta and gamma bands are calculated separately.

D. THE FEATURE SELECTION OF EEG

Feature selection is a necessary step for EEG analysis. In practical, some of the EEG features are irrelevant or may be interdependent. Feature selection aims to simplify the model and improve generalization performance by reducing irrelevant or redundant features. In this paper, we apply a non-linear dimensionality reduction algorithm for EEG feature selection.

Traditional linear dimensionality reduction algorithm, Principal Components Analysis (PCA) [34] and Partial Least Square (PLS) [35], have been widely used in feature selection of EEG. A major problem of linear dimensionality reduction algorithms is that they preserve as much of the significant structure of the high-dimensional data as possible in the low-dimensional map by maximize the variance, that caused data points are far apart even the points are similar [36]. However, it's very essential to keep similar data points close together in mapping the high-dimensional data to low-dimensional. Hence, in this work we utilized the t-SNE, a non-linear feature selection algorithm, make the features more characteristic to improve the prediction performance. The t-SNE has been demonstrated its advantage in the field of computer vision [36].

Stochastic Neighbor Embedding (SNE) was proposed by Hinton and Roweis [37]. Its main principle is to convert the high-dimensional Euclidean distance between data points to the conditional probability of similarity. Though SNE manifests the outstanding performance in imaging dimensionality reduction, the complexity of cost function and the crowding problem are worth to optimize. t-SNE employed the student's t-distribution as heavy-tailed distribution in low-dimensional to compute the similarity between points which could alleviate the growing problem. The similarities of low-dimensional and high-dimensional can be presented by equation (1) and equation (2):

$$q_{ij} = \frac{\exp\left(-\left\|y_i - y_j\right\|\right)}{\sum_{k \neq l} \exp\left(-\left\|y_k - y_l\right\|\right)} \tag{1}$$

$$p_{ij} = \frac{\exp\left(-\left\|x_i - x_j\right\| / 2\sigma^2\right)}{\sum_{k \neq l} \exp\left(-\left\|x_k - x_l\right\| / 2\sigma^2\right)} \tag{2}$$

where $X = \{x_1, x_2, \ldots, x_i\}$ is the input data which is high-dimensional, $Y = \{y_1, y_2, \ldots, y_i\}$ is the low-dimensional mapping of X. p_{ij} and q_{ij} are the pairwise similarities in the high-dimensional space and low-dimensional space respectively. σ is the variance of the Gaussian that is centered on datapoint x_i. Furthermore, the cost function of t-SNE is the symmetric version of SNE that could simplify the form of the gradient, and it is given by:

$$C = KL\left(P||Q\right) = \sum_i \sum_j p_{ij} log \frac{p_{ij}}{q_{ij}} \tag{3}$$

The cost C which is represented by the Kullback-Leibler divergence between joint probability distribution P (in the high-dimensional) and joint probability Q (in the low-dimensional). And the gradient of t-SNE was calculated to minimize the cost C which is simpler than the gradient of SNE, it can be shown as:

$$\frac{\delta C}{\delta y_i} = 4 \sum_i \left(p_{ij} - q_{ij}\right)\left(y_i - y_j\right) \tag{4}$$

In this work, X refers to the raw EEG features, and Y refers to the mapped EEG features. The feature dimension is settled as 50 by the PCA preprocessing. The parameter of perplexity is 30, and the number of iterations is set to 1000. At last, the mapped dimension is needed to optimize in the range of 15 to 30.

E. FEATURE EXTRACTION OF FACIAL EXPRESSIONS

Facial expression is a simple way to express emotions in daily life, so facial expressions are always applied in multimodal emotion recognition. The facial landmark localization is an effective method to extract facial geometric features. The 68 points facial landmark model was established by Dlib C++ library on the Python platform [38]. In this study, we selected the slope of the brow, the extent of eyes opening, the extent of mouth opening, and slope of a corner of the mouth as the facial features. These features are calculated by coordinates of 29 landmarks that mainly located at the area of eyes and mouth. The detailed description of facial landmarks has been shown in our previous research [39], and examples of facial landmarks are expressed in Figure 6.

IV. TEMPORAL DYNAMICS OF EMOTIONS

The continuous emotion recognition could be considered as a type of the time series prediction, and LSTM has the capability of time series analysis. Thus, the LSTM was utilized to accomplish the decision level fusion of EEG and facial expression and capture the temporal dynamics of emotions.

A. LONG SHORT-TERM MEMORY NETWORKS AND HYPERPARAMETERS SELECTION

The LSTM was implemented by Keras [40], a high-level neural networks API, and the Tensorflow was chosen as Keras backend. The model of LSTMs was running on NVIDIA GeForce GTX 950M with Compute Unified Device Architecture (CUDA) acceleration. The weights of LSTM are

FIGURE 6. The facial landmark localization of different subjects.

TABLE 2. The average performance of different LSTM schemes.

NU_1	One layer	Two layers (1/2 of first layer units)		Two layers (1/4 of first layer units)	
10	0.548±0.044	[10,5]	0.561±0.042	[10,3]	0.559±0.041
20	0.551±0.043	[20,10]	0.563±0.042	[20,7]	0.565±0.040
30	0.554±0.043	[30,15]	0.566±0.039	[30,8]	**0.567±0.041**
40	0.546±0.044	[40,20]	0.555±0.040	[40,10]	0.558±0.039

TABLE 3. The hyper-parameters of LSTM.

Hyper-parameters	L	NU_1	NU_2	LF	GD	LR	MS	C
Selected results	2	30	8	MAE	Adam	0.001	40	0.1

randomly initialized according to the normal distribution to improve the efficient of gradient descent.

The hyper-parameters of LSTM are settled as follows, the number of hidden layers L ∈ {1,2}, the number of hidden units in the first hidden layer NU_1 ∈ {10, 20, 30, 40}. If the LSTM has two hidden layers, the units in second hidden layer (NU_2) will be the half or a quarter of the first layer. The loss function LF∈ {Mean Squared Error (MAE), Root Mean Squared Error (MSE)}. The gradient descent optimizer GD∈ {SGD, Momentum, Adam}. The learning rate LR∈{0.0005, 0.001}. The mini-batch size MS∈{10, 20, 30, 40}. To avoid overfitting, we add the dropout in the hidden layer and the coefficient C∈{0.1, 0.2}. The results were evaluated in a 10-fold cross validation which is subject-dependent. The best performance parameters on the test set are selected as the final hyper-parameters. The selected hyper-parameters are listed in Table 3 and we gave the performance by different hidden layers and hidden units in the Table 2. In this paper, CCC was used to evaluate the performance.

B. TEMPORAL DYNAMICS OF EMOTIONS AND DECISION LEVEL FUSION

The minimum time window of EEG is settled as 1 s, 128 sample points in total. According to the EEG, the images of facial expression are picked out in 1 s video. The raw EEG data should be preprocessed by bandpass filtering and ICA spatial filtering, and then PSD features are extracted by STFT. Meanwhile, the facial expression images are processed by the facial landmark model, the facial features are calculated by the coordinates of significant landmarks.

EEG and facial expressions can get the current prediction of valence by inputting the processed features to SVR respectively. What's more, inherent temporal dependencies exist in emotion recognition research [22], so the temporal dynamic of emotion should be considered in continuous emotion recognition. In the paper, the decision level fusion of EEG and facial expressions was implemented by LSTM which could capture temporal dynamics, and Figure 7 has shown the decision level fusion framework. For example, the current valence prediction not only correlated with the current time-step predictions but also the previous time steps. And each time-step of LSTM is fed with 1 s affective data, so the number of the time-steps can be selected as 1, 2, 3 and 4 that correspond to decision level fusion by different time length.

V. RESULTS

A. CONTINUOUS AFFECTIVE ANNOTATION

The ground truth of each trail was the averaged annotations by 10 annotators. The intraclass correlation coefficients (ICC) are calculated to measure the consistency of the valence annotations (M = 0.730, SD = 0.083), and 51.67% of annotations of the movie clips are over 0.75. The appropriate movie clips selection and visual annotation by DRAMA potentially improved the consistency of annotations that made the ground truth more reliable. Figure 8 shows an example of negative clip session detection result.

B. THE RESULTS OF FEATURE SELECTION

Figure 9 shows precision of EEG-based emotion recognition by two feature selection methods (PCA and t-SNE) with different dimensions. The results indicated that the precision of EEG-based emotion recognition was promoted as the decreasing of feature dimensions. The PCA and t-SNE shows the effectiveness in EEG-based emotion recognition. Moreover, t-SNE achieved more significantly improvement than PCA. The best performances achieved by t-SNE and PCA were 0.534 ± 0.028 and 0.464 ± 0.032 separately when the dimension of the mapped feature is 15.

C. THE ANALYSIS OF DIFFERENT TIME-STEPS

We applied different numbers of time-steps (1, 2, 3, 4) of LSTM to accomplish the decision level fusion and temporal dynamic capturing. The results are summarized in Figure 10. Overall, all classes of movie clips achieved the highest CCC at three time-steps. And the averaged CCC at one time-step is 0.567±0.041, two time-steps is 0.615±0.037, three time-steps is 0.625±0.029 and four time-steps is 0.614±0.027. The averaged performances were not significantly different

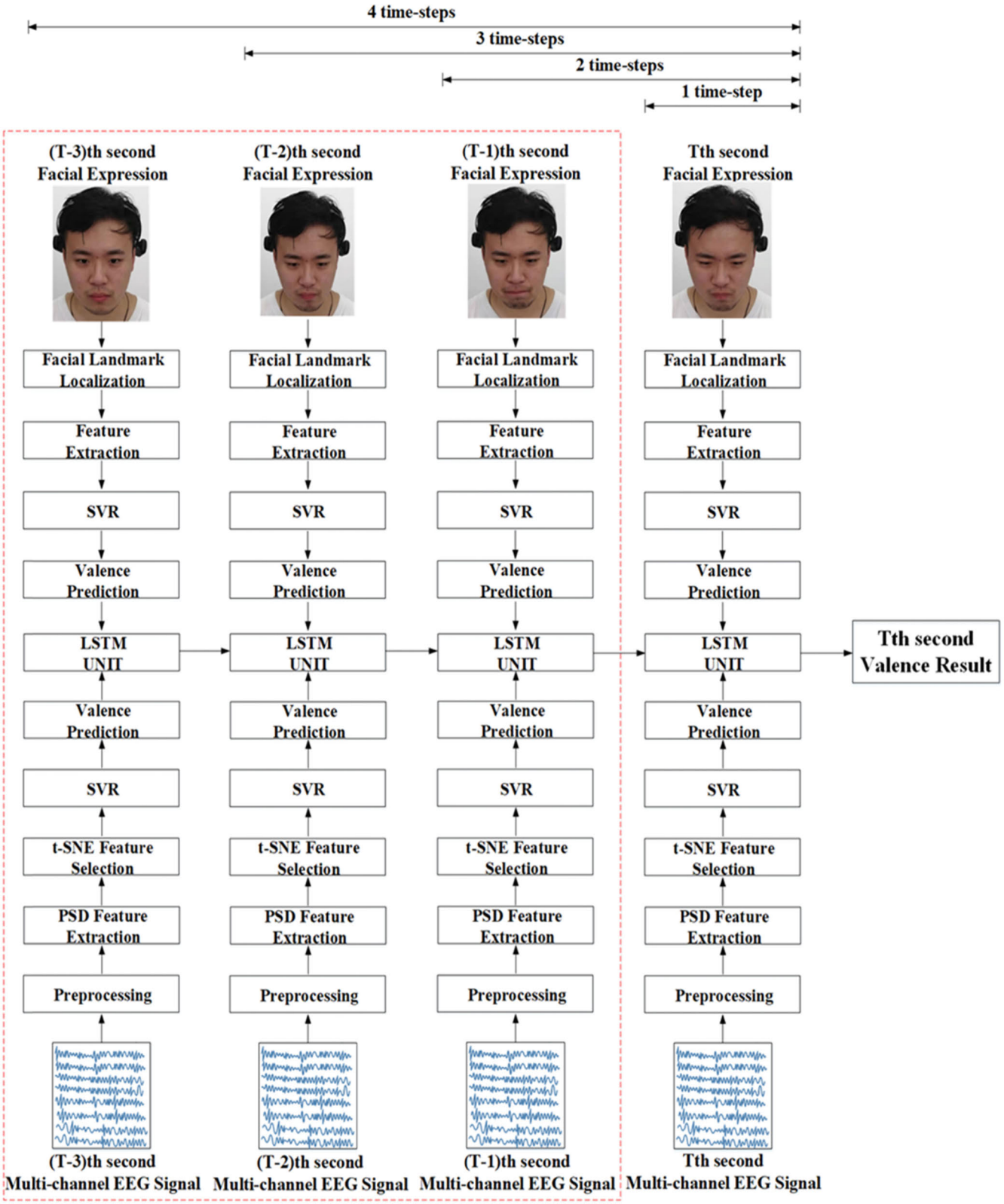


FIGURE 7. The LSTM decision fusion with different time-steps.

at two, three and four time-steps, but that obviously outperformed the performance one time-step. Moreover, the previous time-step information is beneficial to the prediction of current time-step. Thus, the LSTM with multi-time-steps could capture the temporal dynamic of emotions effectively.

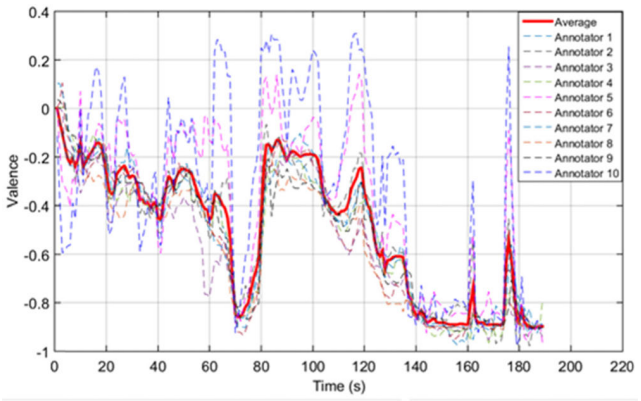


FIGURE 8. The annotation results were annotated by ten annotators *An annotation example of the subject's response to the Aftershock (a negative movie clip).*

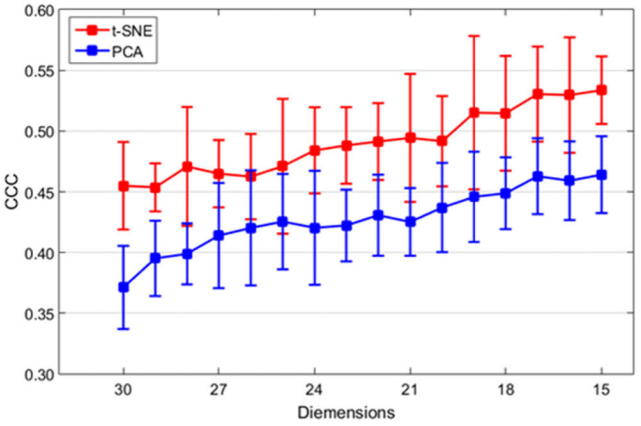


FIGURE 9. The performances of mapped features by different feature selection algorithm.

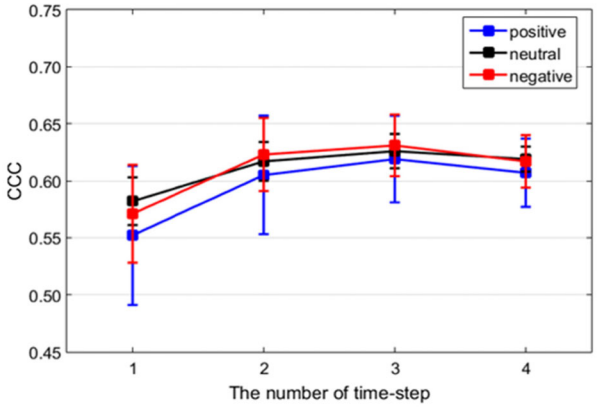


FIGURE 10. The decision level fusion performance by LSTM with different time-steps.

D. THE RESULTS OF CONTINUOUS EMOTION RECOGNITON

The results of continuous emotion recognition are shown in Table 4. The fusion of two modalities achieved better results than EEG and facial expressions separately. For the results of single modality, the facial expressions outperformed EEG. This might be the annotations are based on the facial expressions video and facial expressions, and facial expressions are easier to express the emotion. What's more, the negative movie clips yield better results than the other two classes. The war and disaster related plots of the negative movie clips conveyed pain feelings to subjects that would elicit sadness intuitively. For the neutral movie clips, a better result was achieved by facial expressions. Subjects did not show the significant fluctuations in mood, so facial features and annotations were more reliable. Three examples of continuous emotion detection are given in Figure 11.

TABLE 4. The average CCC evaluations of different modalities and three classes of movie clips.

Emotion States	EEG	Facial Expressions	Decision Fusion
Positive	0.524±0.038	0.559±0.027	0.619±0.032
Neutral	0.531±0.016	**0.577±0.013**	0.626±0.015
Negative	**0.546±0.027**	0.569±0.025	**0.631±0.027**

VI. DISSCUSSION

In this work, we aimed to improve the precision of continuous emotion recognition by the fusion of EEG and facial expression. From Table 4, the multi-modal emotion recognition yields better results than a single modality, suggesting multi-modalities provides comprehensive and complementary emotional information.

However, the performance of positive movie clips did not achieved a desirable results (as shown in Table 4). Though we guided the subject to maintain stability during the experiment period, the subjects watched the video for a longer period would blink and move the head inadvertently, especially in some exciting or ridiculous plots. The features extracted by the facial landmark localization are very sensitive to landmark coordinates, and the artifacts would also influence EEG. In response to this situation, we applied ICA spatial filtering in EEG preprocessing to partially eliminate the effects of blinking and head moving. Although facial landmark localization has defects in the above cases, the extracted features are explainable and it is convenient to implement. The extent of mouth opening and slope of mouth corner are the most relevant features. Besides, negative movie clips conveyed the pain of disaster and war has triggered subjects' sadness successfully. In future work, we will attempt different types of negative movie clips as stimulis.

Measurements recorded over various parts of the brain including the amygdala potentially enable observation of the emotions felt [41]. The amygdala plays an important role of emotions, the measurement via EEG could obtain emotions correlated information from amygdala. Therefore, EEG is significantly related to valence. In order to analyze the relationship between EEG and valence, the correlations between PSD features with different electrodes and valence labels are calculated. The top 5 valence related electrodes are given in Table 5. And statistical significance with all the p-values are smaller than 0.05. T7 and T8 which distributed in left and right temporal accordingly have shown the strong

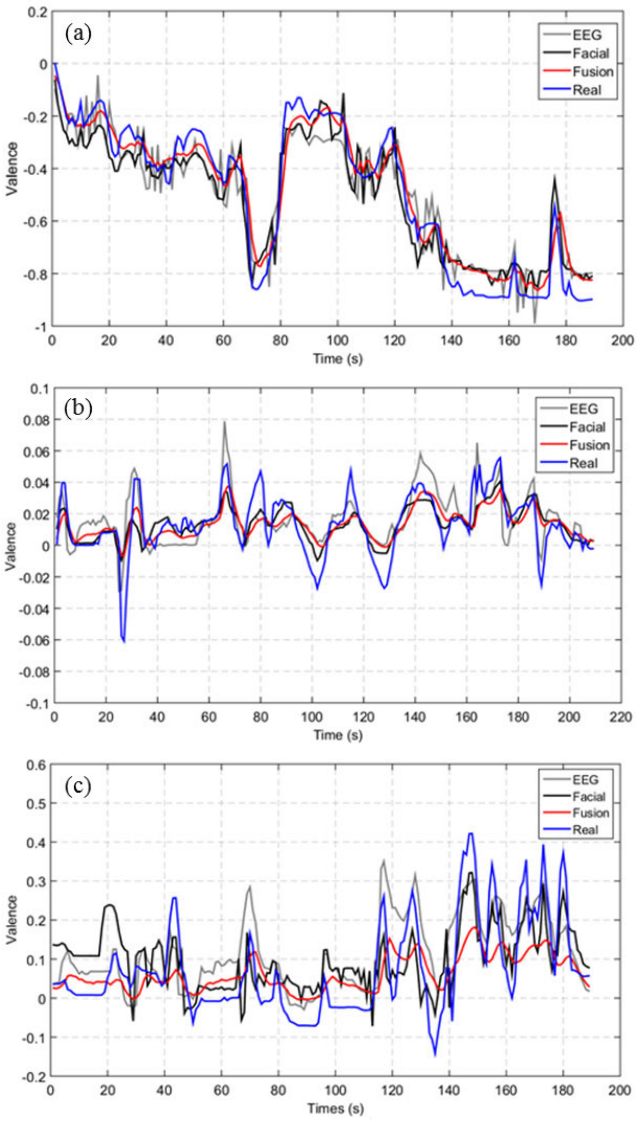


FIGURE 11. Examples of continuous emotion recognition results. (a) is the negative emotion state example which elicited by Aftershock. (b) is the neutral emotion state example which elicited by World Heritage in China II. (c) is the positive emotion state which elicited by Lost in Thailand.

correlations in the gamma and beta bands. In the theta and alpha bands, the electrode O2, O1, and P8 which distributed in the occipital region are the most relevant. Similarly, Koelstra et al. found that valence led to an increase of power of the occipital region in the low-frequency bands and right temporal increase of power in the high-frequency bands [12].

According to Figure 9, t-SNE has the obvious advantage over PCA in EEG feature selection. It may be because of removing the redundant features can simplify the model to improve the precision of the model. It perhaps that linear feature selection algorithm (such as PCA) cannot map nonlinear EEG features to low dimensions very well. What's more, PCA pursues the maximization of sample variance that causes the feature points mapped to low dimensions are far apart. However, it's very critical to keep similar feature points together. In contrast to this, the core of t-SNE is to convert

TABLE 5. Top 5 valence related electrodes.

Electrode	Frequency bands	Average correlation coefficient
T7	Gamma	0.16
	Beta	0.13
T8	Gamma	0.15
P8	Alpha	0.13
O2	Theta	0.10
O1	Theta	0.09

the high-dimensional Euclidean distance between data points into conditional similarity probability. The conditional similarity probability is calculated between the low-dimensional data points, and then construct the loss function through KL divergence to compare the similarity of the probability distributions between the low-dimensional and high-dimensional. The reliability of the low-dimensional mapped feature is improved by the gradient descent to the minimization loss function. Therefore, t-SNE mapped nonlinear EEG feature into low dimension more reliably.

The LSTM captured temporal dynamics of emotions and achieved ideally decision level fusion results. According to Figure 10, LSTM has shown the ability to capture the temporal dynamics and obtained the best result 0.625 ± 0.029 at three time-steps. And decision level fusion obviously outperformed single modality except for the one time-step. It might be when the time-step is one, the LSTM would be played like the BP neural network. Hence, it may hardly capture the temporal dynamic appropriately. Moreover, the period of 1-4 s is needed to discriminate an emotional state by EEG signal [42]. The reason is that EEG is assumed to remain stationary during short intervals [20]. Therefore, we set the different number of time-steps from 1 to 4 of LSTM to improve precision of continuous emotion recognition by temporal dynamic caputruring based on previous research.

For comparing our work with the related works, the fusion of EEG and facial expression based emotion recognition works are listed in Table 6. Here we analyzed these works from the methods level. Decision level fusion outperformed feature level decision in the majority. That is one of the reasons we selected the multi-steps LSTM-based decision level fusion, another is that LSTM could capture the temporal dynamics of emotions and it could improve the performance. For EEG feature extraction, most works extracted temporal-frequency features, especially the PSD features of theta, alpha, beta, and gamma bands. More than that, we also noted the EOG artifacts in the long-time video-induced experiment, and the EOG was removed by ICA. Furthermore, feature selection is the essential step of EEG signal processing and the half of the works considered it [43]–[48]. Compared with variance maximized linear feature reduction/selection algorithm, manifold learning algorithm t-SNE might more suitable for nonlinear and nonstationary EEG signals. For the feature extraction of facial expression, CNNs and facial landmark localization are the primary methods. CNNs and improved CNNs have shown the strong ability in facial expression recognition [49], [50]. On the other hand, facial landmark localization is easier to implement, and it

TABLE 6. The fusion of EEG and facial expression based emotion recognition works.

Ref.	EEG features and feature extraction	EEG feature selection	Facial features and feature extraction	Facial feature selection	Fusion scheme	Classification or Continuous detection	Multimodal Results
[19]	Theta, alpha, beta, and gamma PSD features; STFT	—	facial geometric features; SDM based face tracker	—	LSTM based decision level fusion	Valence continuous detection;	RMSE:0.044±0.026 $\overline{\rho}$:0.45±0.35
[43]	Theta, slow alpha (8-10Hz), alpha, beta, and gamma PSD features and symmetric-pairs features	RFE	3 types facial AU related features; AU detection system is trained by MMI dataset for facial expression recognition.	RFE	W-REG based decision level fusion (classification); W-REG^{T} based feature level fusion (Regression);	Binary classification of valence, arousal and control; Valence, arousal and control continuous detection	Classification(F1-score): Valence:0.709 Arousal:0.718 Control:0.670 Regression(R^2-score): Valence:0.231 Arousal:0.276 Control:0.270
[44]	Wavelet features; SWT	—	facial geometric features; OpenPose tool	—	RFC based feature level fusion	Nine emotion states classification	Average classification accuracy: 97.70%
[8]	Theta, slow alpha, alpha, beta, and gamma PSD features T7-T8, Fp1-Fp2, and CP1-CP2 symmetric-pairs features; DWT	SVM-RFE	Viola & Jones face detector (face position dection); Multi-task CNN (feature extraction and arousal-valence classification tasks)	—	Adaboost based decision level fusion	Binary classification of valence and arousal; Valence and arousal are both rated in the range from 1 to 9, the high class rating is (6-9) and the low class is (1-5).	Valence classification accuracy: 80.00 ± 12.40% Arousal classification accuracy: 74.17 ± 14.04%
[45]	Theta, alpha, beta, and gamma PSD features; STFT	—	facial geometric features; Openface application	—	LSTM based feature level fusion	Valence Continuous detection	Subject-dependent RMSE: 0.041 Subject-independent RMSE: 0.046
[46]	Delta, theta, alpha1(8-10Hz), alpha2(11-13Hz), beta1(14-20Hz), beta2(21-30Hz), gamma1(31-40Hz) gamma2(41-50Hz) PSD features; STFT	—	Harr-like features are used to detect position of face, and resize the face part as 48 by 48; Adaboost	PCA	Production rules based decision level fusion	Four emotion states classification (happiness, neutral, sadness, and fear)	Average classification accuracy: 82.75 ± 5.19%
[47]	EEG features are calculated by Hjorth parameters for theta, alpha, beta and gamma bands	mRMR	Haar-like features' cascade of weak classifiers for face position detection; CNN for expression recognition	PCA	The average score of Multi-Sample-Multi-Source based decision level fusion	Valence and arousal classification, and each of them has quantized into three levels	Valence levels at 3, 5, and 7 are 77.5%, 63.2%, and 34.5% separately. Arousal levels at 3, 5, and 7 are 75.2%, 57.2%, and 25.5% separately.
[48]	1-32Hz decomposed into 63 subbands, SP and SPD features are calculated; Welch's averaged periodogram method	KNN	EPF features; STLMBP	KLPP	LWF based decision level fusion	Valence and arousal classification, and each of them has low, medium, and high levels	Valence classification accuracy: 66.28% Arousal classification accuracy: 63.22%
Our work	Theta, alpha, beta, and gamma PSD features; ICA, STFT	t-SNE	facial geometric features; Dlib facial landmark model	—	Multi-steps LSTM based decision level fusion	Valence continuous detection	CCC overall: 0.625±0.029 Positive: 0.619±0.038 Neutral: 0.626±0.015 Negative: 0.631±0.027

One work might propose several fusion methods, we only listed the method corresponding to the optimal result.
Supervised Descent Method (SDM), Root Mean Square Error (RMSE), Pearson Correlation Coefficient (ρ), Recursive Feature Elimination (RFE), Action Unit (AU), Regression-Estimated Weights Fusion (W-REG), W-REG with the training performance (W-REGT), Stationary Wavelet Transform (SWT), Random Forest Classifier (RFC), Discrete Wavelet Transform (DWT), Convolutional Neural Network (CNN), adaptive boosting (Adaboost), Support Vector Machine (SVM), Minimum Redundancy and Maximum Relevance (mRMR), Spectral Power (SP), Spectral Power Difference (SPD), k-Nearest Neighbor (kNN), Expression-Percentage Feature (EPF), Spatio-Temporal Local Monogenic Binary Pattern (STLMBP), Kernel Local Preserving Projection (KLPP), learned weights fusion (LWF)

extracts facial geometric features of emotion related regions (eyes, eye brows, and mouth). What's more, CNNs and facial landmark localization are both unnecessary to complete feature selection.

VII. CONCLUSION

In this paper, we have presented a framework for the fusion of EEG and facial expression based continuous emotion recognition which achieved significantly better results than the single modality. The LSTM was applied for decision level fusion and it captured the temporal dynamic of emotions that improve the performance significantly. From the experimental results, we found that three time-steps of LSTM yielded the best CCC (0.625 ± 0.029). For the EEG modality, the ICA spatial filtering was utilized to weaken the eye gazing and head moving artifacts. The t-SNE feature selection was not only eliminated redundant features but also got the characterized features. Though facial expressions based emotion recognition outperformed EEG overall, it didn't achieve desirable results. The reason might be that ridiculous and exciting plots and fatigue would caused the eye gazing and head moving, and facial expression features are sensitive to the coordinates of the landmarks. To this situation, more suitable facial expression recognition algorithms are valuable to explore in future work. Besides, we also attempted a wearable and low-cost equipment, Emotiv Epoc headset, for EEG sampling. The application of this low-cost equipment may be essential to the further online experiment and apply EEG-based emotion recognition to daily life. In the future, we will keep on researching the emotion recognition system based on the fusion of EEG and facial expression by appending machine learning algorithms and HCI.

ACKNOWLEDGMENT

(D. H. Li and Z. Wang are co-first authors.)

REFERENCES

[1] R. W. Picard, *Affective Computing*. Cambridge, MA, USA: MIT Press, 1997, pp. 247–248.
[2] F. Cid, L. J. Manso, and P. Núnez, "A novel multimodal emotion recognition approach for affective human robot interaction," in *Proc. FinE*, Hamburg, Germany, 2015, pp. 1–9.
[3] J. Serrano, L. L. Di Stasi, A. Megías, and A. Catena, "Affective-sound effects on driving behaviour," *Transport*, vol. 29, no. 1, pp. 100–106, Mar. 2014.
[4] C. Wu, Y. Zhang, J. Jia, and W. Zhu, "Mobile contextual recommender system for online social media," *IEEE Trans. Mobile Comput.*, vol. 16, no. 12, pp. 3403–3416, Dec. 2017.
[5] C.-H. Yang, "Fuzzy fusion for attending and responding assessment system of affective teaching goals in distance learning," *Expert Syst. Appl.*, vol. 39, no. 3, pp. 2501–2508, Feb. 2012.
[6] K. R. Scherer, "What are emotions? And how can they be measured?" *Social Sci. Inf.*, vol. 44, no. 4, pp. 695–729, Dec. 2005.
[7] S. K. D'Mello and J. Kory, "A review and meta-analysis of multimodal affect detection systems," *ACM Comput. Surv.*, vol. 47, no. 3, Apr. 2015, Art. no. 43.
[8] Y. Huang, J. Yang, S. Liu, and J. Pan, "Combining facial expressions and electroencephalography to enhance emotion recognition," *Future Internet*, vol. 11, no. 5, p. 105, May 2019.
[9] A. Mehrabian and J. A. Russell, *An Approach to Environmental Psychology*. Cambridge, MA, USA: MIT Press, 1974.
[10] H. Gunes and B. Schuller, "Categorical and dimensional affect analysis in continuous input: Current trends and future directions," *Image Vis. Comput.*, vol. 31, no. 2, pp. 120–136, Feb. 2013.
[11] M. M. Bradley and P. J. Lang, "The international affective picture system (IAPS) in the study of emotion and attention," in *Handbook of Emotion Elicitation and Assessment*, J. A. Coan and J. J. B. Allen, Eds. Oxford, U.K.: Oxford Univ. Press, 2007, pp. 19–24. [Online]. Available: https://global.oup.com/
[12] 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, Jan. 2012.
[13] H. Huang, Q. Xie, J. Pan, Y. He, Z. Wen, R. Yu, and Y. Li, "An EEG-based brain computer interface for emotion recognition and its application in patients with Disorder of Consciousness," *IEEE Trans. Affective Comput.*, to be published. doi: 10.1109/TAFFC.2019.2901456.
[14] C. Busso, M. Bulut, C.-C. Lee, A. Kazemzadeh, E. Mower, S. Kim, J. N. Chang, S. Lee, and S. S. Narayanan, "IEMOCAP: Interactive emotional dyadic motion capture database," *Lang. Resour. Eval.*, vol. 42, no. 4, pp. 335–359, Dec. 2008.
[15] X. W. Wang, D. Nie, and B. L. Lu, "Emotional state classification from EEG data using machine learning approach," *Neurocomputing*, vol. 129, no. 10, pp. 94–106, Apr. 2014.
[16] W.-L. Zheng, B.-N. Dong, and B.-L. Lu, "Multimodal emotion recognition using EEG and eye tracking data," in *Proc. 36th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Chicago, IL, USA, Aug. 2014, pp. 5040–5043.
[17] T. Song, W. Zheng, C. Lu, Y. Zong, X. Zhang, and Z. Cui, "MPED: A multi-modal physiological emotion database for discrete emotion recognition," *IEEE Access*, vol. 7, pp. 12177–12191, 2019.
[18] M. Soleymani, J. Lichtenauer, T. Pun, and M. Pantic, "A multimodal database for affect recognition and implicit tagging," *IEEE Trans. Affective Comput.*, vol. 3, no. 1, pp. 42–55, Jan./Mar. 2012.
[19] M. Soleymani, S. Asghari-Esfeden, Y. Fu, and M. Pantic, "Analysis of EEG signals and facial expressions for continuous emotion detection," *IEEE Trans. Affective Comput.*, vol. 7, no. 1, pp. 17–28, Jan./Mar. 2016.
[20] Y.-J. Liu, M. Yu, G. Zhao, J. Song, Y. Ge, and Y. Shi, "Real-time movie-induced discrete emotion recognition from EEG signals," *IEEE Trans. Affective Comput.*, vol. 9, no. 4, pp. 550–562, Oct. 2018.
[21] R. Khosrowabadi, C. Quek, K. K. Ang, and A. Wahab, "ERNN: A biologically inspired feedforward neural network to discriminate emotion from EEG signal," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 3, pp. 609–620, Mar. 2014.
[22] M. A. Nicolaou, H. Gunes, and M. Pantic, "Output-associative RVM regression for dimensional and continuous emotion prediction," *Image Vis. Comput.*, vol. 30, no. 3, pp. 186–196, Mar. 2012.
[23] F. Weninger, F. Ringeval, E. Marchi, and B. Schuller, "Discriminatively trained recurrent neural networks for continuous dimensional emotion recognition from audio," in *Proc. Int. Joint Conf. Artif. Intell.*, New York, NY, USA, Jul. 2016, pp. 2196–2202.
[24] F. Ringeval, B. Schuller, M. Valstar, S. Jaiswal, E. Marchi, D. Lalanne, R. Cowie, and M. Pantic, "AV+EC 2015: The first affect recognition challenge bridging across audio, video, and physiological data," in *Proc. Int. Workshop Audio/Vis. Emotion Challenge*, New York, NY, USA, Oct. 2015, pp. 3–8.
[25] N. Banda, A. Engelbrecht, and P. Robinson, "Continuous emotion recognition using a particle swarm optimized NARX neural network," in *Proc. Int. Conf. Affect. Comput. Intell. Interact.*, Xi'an, China, Sep. 2015, pp. 380–386.
[26] 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.
[27] M. A. Nicolaou, H. Gunes, and M. Pantic, "Continuous prediction of spontaneous affect from multiple cues and modalities in valence-arousal Space," *IEEE Trans. Affective Comput.*, vol. 2, no. 2, pp. 92–105, Apr./Jun. 2011.
[28] M. Wöllmer, M. Kaiser, F. Eyben, B. Schuller, and G. Rigoll, "LSTM-modeling of continuous emotions in an audiovisual affect recognition framework," *Image Vis. Comput.*, vol. 31, no. 2, pp. 153–163, Feb. 2013.
[29] W.-L. Zheng and B.-L. Lu, "Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks," *IEEE Trans. Auton. Mental Develop.*, vol. 7, no. 3, pp. 162–175, Sep. 2015.
[30] J. M. Girard and A. G. C. Wright, "DARMA: Software for dual axis rating and media annotation," *Behav. Res. Methods*, vol. 50, no. 3, pp. 902–909, Jun. 2017.
[31] M. M. Bradley and P. J. Lang, "Measuring emotion: The self-assessment manikin and the semantic differential," *J. Behav. Therapy Experim. Psychiatry*, vol. 25, no. 1, pp. 49–59, Mar. 1994.

[32] A. Delorme and S. Makeig, "EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," *J. Neurosci. Methods*, vol. 134, no. 1, pp. 9–21, Mar. 2004.
[33] R. J. Davidson, "Affective neuroscience and psychophysiology: Toward a synthesis," *Psychophysiology*, vol. 40, no. 5, pp. 655–665, Aug. 2003.
[34] Y. Song and F. Sepulveda, "A novel technique for selecting EMG-contaminated EEG channels in self-paced brain–computer interface task onset," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 7, pp. 1353–1362, Jul. 2018.
[35] D. K. Yanti, M. Z. Yusoff, and V. S. Asirvadam, "Single-trial visual evoked potential extraction using partial least-squares-based approach," *IEEE J. Biomed. Health Informat.*, vol. 20, no. 1, pp. 82–90, Jan. 2016.
[36] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, Nov. 2008.
[37] G. E. Hinton and S. T. Roweis, "Stochastic neighbor embedding," in *Proc. Neural Inf. Process. Syst.*, 2002, pp. 857–864.
[38] D. E. King, "Dlib-ml: A machine learning toolkit," *J. Mach. Learn. Res.*, vol. 10, pp. 1755–1758, Jan. 2009.
[39] D. Li, Z. Wang, Q. Gao, C. Wang, X. Yu, and Y. Song, "Facial expression recognition based on Electroencephalogram and facial landmark localization," *Technol. Health Care*, vol. 27, no. 4, pp. 373–387, Jul. 2019.
[40] F. Chollet. *Keras*. Accessed: 2015. [Online]. Available: http://www.keras.io
[41] T. Pun, T. I. Alecu, G. Chanel, J. Kronegg, and S. Voloshynovskiy, "Brain-computer interaction research at the computer vision and multimedia laboratory," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 14, no. 2, pp. 210–213, Jun. 2016.
[42] M. Esslen, R. D. Pascual-Marqui, D. Hell, K. Kochi, and D. Lehmann, "Brain areas and time course of emotional processing," *Neuroimage*, vol. 21, no. 4, pp. 1189–1203, Apr. 2004.
[43] S. Koelstra and I. Patras, "Fusion of facial expressions and EEG for implicit affective tagging," *Image Vis. Comput.*, vol. 31, no. 2, pp. 164–174, Feb. 2013.
[44] V. Chaparro, A. Gomez, A. Salgado, O. L. Quintero, N. Lopez, and L. F. Villa, "Emotion recognition from EEG and facial expressions: A multimodal approach," in *Proc. Annu. Int. Conf., IEEE Eng. Med. Biol. Soc.*, Honolulu, HI, USA, Jul. 2018, pp. 530–533.
[45] D. Şen and M. Sert, "Continuous valence prediction using recurrent neural networks with facial expressions and EEG signals," in *Proc. Signal Process. Commun. Appl. Conf. (SIU)*, Izmir, Turkey, May 2018, pp. 1–4.
[46] Y. Huang, J. Yang, P. Liao, and J. Pan, "Fusion of facial expressions and EEG for multimodal emotion recognition," *Comput. Intel. Neurosc.* , vol. 2017, Sep. 2017, Art. no. 2107451. doi: 10.1155/2017/2107451.
[47] S. Sokolov, Y. Velchev, S. Radeva, and D. Radev, "Human emotion estimation from EEG and face using statistical features and SVM," in *Proc. Int. Conf. Comput. Sci. Inf. Technol.*, Dubai, UAE, 2017, pp. 37–47.
[48] X. Huang, J. Kortelainen, G. Zhao, X. Li, T. Seppänen, M. Pietikäinen, and A. Moilanen, "Multi-modal emotion analysis from facial expressions and electroencephalogram," *Comput. Vis. Image Understand.*, vol. 147, pp. 114–124, Jun. 2016.
[49] Y. Li, J. Zeng, S. Shan, and X. Chen, "Occlusion aware facial expression recognition using CNN with attention mechanism," *IEEE Trans. Image Process.*, vol. 28, no. 5, pp. 2439–2450, May 2019.
[50] B. Yang, J. Cao, R. Ni, and Y. Zhang, "Facial expression recognition using weighted mixture deep neural network based on double-channel facial images," *IEEE Access*, vol. 6, pp. 4630–4640, 2017.

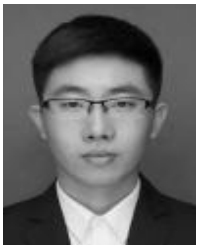
DAHUA LI received the M.S. degree in mechanical manufacturing and automation from Guangxi University, China, in 2004.

Since 2004, he has been an Associate Professor with the School of Electrical and Electronic Engineering, Tianjin University of Technology, where he is currently the Director with the Electrical and Electronic Center. His current research interests include machine vision, 3D inspection systems, and binocular ranging.

ZHE WANG received the B.S. degree in automation from the Tianjin University of Technology, China, in 2017, where he is currently pursuing the M.S. degree with the School of Electrical and Electronic Engineering. His current research interests include affective computing, signal processing, and machine learning.

CHUHAN WANG received the B.S. degree in automation from the Tianjin University of Technology, China, in 2018, where she is currently pursuing the M.S. degree with the School of Electrical and Electronic Engineering. Her current research interests include affective computing and brain–computer interface.

SHUANG LIU received the B.S. degree in electrical automation from the Tianjin University of Technology, China, in 2017, where he is currently pursuing the M.S. degree with the School of Electrical and Electronic Engineering. His current research interests include haptic control and brain–computer interface.



WENHAO CHI received the B.S. degree in automation from the Tianjin University of Technology, China, in 2016, where he is currently pursuing the M.S. degree with the School of Electrical and Electronic Engineering. His current research interests include deep learning and adaptive dynamic programming.

ENZENG DONG received the B.S. and M.S. degrees from the Army Officer Academy of PLA, China, in 2000 and 2003, respectively, and the Ph.D. degree from Nankai University, China, in 2006.

Since 2001, he has been a Full Professor with the School of Electrical and Electronic Engineering, Tianjin University of Technology. His current research interests include chaos systems, brain–computer interface, and image processing.

XIAOLIN SONG received the M.S. degree from Dalian Maritime University, China, in 2007.

Since 2007, she has been a Lecturer with the Engineering Training Center, Tianjin University of Technology, China. Her current research interest includes educational psychology.

QIANG GAO received the M.S. degree in control theory and control engineering from Tianjin University, China, in 1996.

Since 1996, he has been a Full Professor with the School of Electrical Engineering, Tianjin University of Technology, where he is currently the Director with the Engineering Training Center. His current research interests include intelligence controlling, brain–computer interface, and fault diagnosis.

YU SONG received the Ph.D. degree in intelligent mechanical systems engineering from Kagawa University, Japan, in 2018.

He is currently a Lecturer with the School of Electrical and Electronic Engineering, Tianjin University of Technology. His current research interests include haptic control, human robot interaction, teleoperation systems, brain–computer interface, and artificial intelligence.