9th International IEEE EMBS Conference on Neural Engineering
San Francisco, CA, USA, March 20 - 23 , 2019

Classification of Five Emotions from EEG and Eye Movement Signals: Complementary Representation Properties

Li-Ming Zhao, Rui Li, Wei-Long Zheng and Bao-Liang Lu* *Senior Member, IEEE*

Abstract— **Recently, various multimodal approaches to enhancing the performance of affective models have been developed. In this paper, we investigate the complementary representation properties of EEG and eye movement signals on classification for five human emotions: happy, sad, fear, disgust, and neutral. We compare the performance of single modality and two different modality fusion approaches. The results indicate that EEG is superior to eye movements in classifying happy, sad and disgust emotions, whereas eye movements outperform EEG in recognizing fear and neutral emotions. Compared with eye movements, EEG has the advantage of classifying the five emotions, with the mean accuracies of 69.50% and 59.81%, respectively. Due to the complementary representation properties, the modality fusion with bimodal deep auto-encoder significantly improves the classification accuracy to 79.71%. Furthermore, we study the neural patterns of five emotion states and the recognition performance of different eye movement features. The results reveal that five emotions have distinguishable neural patterns and pupil diameter has a relatively high discrimination ability than the other eye movement features.**

I. INTRODUCTION

Emotions play an important role in how we think and behave. Although we have a rich vocabulary for describing emotions (e.g., joy, love, fear, angry, and so forth), it is difficult to directly quantify and measure the specific emotional state. Ekman *et al.* [1] first proposed several principles in mind and defined the six basic emotions including happiness, fear, disgust, anger, surprise, and sadness. These six emotions are considered as typical emotions in our daily life, and there is a great distinction from each other.

Various studies indicate that different modalities can describe different aspects of emotions. As one of the most popular modalities for emotion recognition, facial expressions have been widely used in affective technologies in the past few decades [2]. Meanwhile, several approaches to recognizing emotions from speech have been reported [3]. However, there is a growing interest in other modalities, such as fMRI, EEG, and eye movements.

Saarimäki *et al.* presented a method for measuring emotions in humans using fMRI on six basic emotions [4]. In their work, a linear neural network classifier without hidden layers was used for classifying emotions and the results suggested basic emotions have a discrete neural basis. Soleymani *et al.* presented a multi-modal approach that fused EEG signals, pupillary response and gaze distance for emotion recognition [5]. In our previous studies, we explored the complementary characteristics of EEG and eye movements for classifying three emotions (happy, sad, and neutral) and four emotions (happy, sad, fear, and neutral) [6][7][8]. However, the complementary representation properties of EEG and eye movements for discriminating the five emotions (happy, sad, fear, disgust, and neutral) are still unclear.

It is difficult to represent emotional states comprehensively and accurately depending on only single modality. Compared with external behavioral information such as facial expressions, voices, and gestures, EEG signals can reflect individual emotional states more objectively and accurately. Meanwhile, eye movement signals can not only provide physiological signals, but also important subconscious activities, which offer contextual clues to emotion recognition. By combining external subconscious behavior from eye movement signals and internal neural patterns from EEG signals, we can build a preferable emotion recognition system.

The aim of this paper is to reveal the complementary representation properties of EEG and eye movement signals for classifying five emotions, including happy, sad, fear, disgust, and neutral. We adopt a linear kernel SVM as the baseline classifier and use two different modality fusion approaches to enhance the classification performance, the feature-level fusion and the bimodal deep auto-encoder (BDAE) [9].

II. METHODS

A. Data Set

The data set used in this paper is the same as [10]. In the experiment, carefully selected film clips are used as the stimuli, which have been explored to have reliability in eliciting emotions [11]. A total of 45 video clips with highly emotional contents are used and edited into 3 segments, each of which consists 15 clips (3 for per emotions). For each segment, 15 clips are placed at random, but for the convenience of subsequent 3-fold cross-validation, the 5 clips in each fold are guaranteed to have different emotion labels. Sixteen healthy subjects (6 males and 10 females) participant in the experiments and each subject is required to perform the experiments for three sessions, at an interval of one week or longer. Both EEG and eye movement signals are collected simultaneously when the subjects are watching the film clips.

This work was supported in part by grants from the National Key Research and Development Program of China (Grant No. 2017YFB1002501), the National Natural Science Foundation of China (Grant No. 61673266), and the Fundamental Research Funds for the Central Universities.

Li-Ming Zhao, Rui Li, Wei-Long Zheng and Bao-Liang Lu are with the Center for Brain-Like Computing and Machine Intelligence, Department of Computer Science and Engineering, the Key Laboratory of Shanghai Education Commission for Intelligent Interaction and Cognitive Engineering and the Brain Science and Technology Research Center, Shanghai Jiao Tong University, 800 Dong Chuan Road, Shanghai 200240, China.

* Corresponding author (bllu@sjtu.edu.cn)

B. Preprocessing

A key characteristic of EEG signals is that they consist of a mixture of an unknown number of brain and non-brain contributions, which renders the recognition and analysis of brain-related EEG activity difficult. In this paper, we first use a band pass filter (1-50 Hz) to eliminate low-frequency noise and high-frequency noise. Then we use principal component analysis (PCA) [12] in Curry 7 to remove the artifacts.

For eye movement data, we extract pupil diameters, fixation details, saccade details, blink details, and event statistics using BeGaze analyzing software of SMI. It is well known that pupil diameter is sensitive to light [13], and different subjects have the similar patterns of pupil diameter change if they watch the same video. Therefore, we use PCA to remove light reflex to obtain the emotional information in the pupil diameter [5].

C. Feature Extraction

Considering the effectiveness of differential entropy (DE) in EEG-based emotion recognition [14], we choose DE as the EEG feature. The DE features are extracted in five frequency bands: δ (1-3 Hz), θ (4-7 Hz), α (8-13 Hz), β (14-30 Hz), and γ (31-50 Hz). A 256-points Short-time Fourier transform (STFT) with 4 s non-overlapping Hanning window is used to calculate the DE features of each channel on these bands. Since 62 channels of EEG signals are collected, we have 310 dimensional features for each sample. A linear dynamic system approach is used to eliminate the rapid changes of the DE features [8], which makes the features more reliable.

For eye movement feature, we extract 33 eye movement features. The details of the features extracted from eye movement data can be found in our previous work [7]. The number of EEG (eye movements) samples of five emotions from each subject is 602, 952, 746, 612, and 734 of happy, sad, fear, disgust, and neutral, respectively.

D. Classification Methods

We adopt a linear SVM as the classifier and use three-fold cross-validation to evaluate the classification performance. For each subject, we split the 15 film clips of each session into the first 5 segments, the middle 5 segments, and the last 5 segments. After that, we stitch together the data from three sessions. Each of three SVMs are trained on permutations of two out of three parts, and the trained models are evaluated on the remaining third. We define the space of parameter C, where 20 values are evenly distributed between 2^{-10} and 2^{10}, from which we choose the best C. All the performance reported in this paper is the average accuracy across all experiments. We perform experiments on two kinds of emotion recognition tasks. 1) For single modality, we use SVM to classify five emotions on EEG features and eye movement features, respectively. 2) For multiple modalities, we apply SVM to the new features generated by different modality fusion strategies.

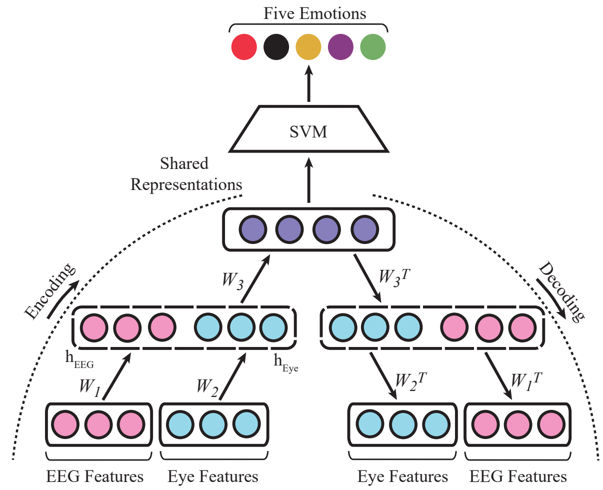


Fig. 1. Deep autoencoder model adopted in this paper. The structure under the black dotted arc shows the learning process of BDAE, including the encoding network and the decoding network. The shared representations are denoted by the purple circles and used as the inputs of SVM to get the final results.

E. Fusion Strategies

We adopt two different strategies to fuse the EEG features and eye movement features. At feature level fusion, we directly concatenate the eye movement feature vector and the EEG feature vector into a larger feature vector. For the other fusion strategy, we adopt a bimodal deep auto-encoder (BDAE) to extract the high level representation of both EEG and eye movements, as shown in Fig. 1. At encoding stage, we use two Restricted Boltzmann Machines (RBMs) for EEG and eye movement features, respectively. Then two hidden layers (h_{EEG}, h_{Eye}) are concatenated directly as the input of an upper auto-encoder. The decoding network is the symmetric structure of the encoding network, trying to reconstruct the original EEG and eye movement inputs. Finally, the new shared representations are input into the linear SVM to get the recognition results.

III. EXPERIMENTAL RESULTS

Table I shows the performance of each single modality and the two modality fusion approaches. For emotion recognition using only EEG, we obtain an average accuracy of 69.50% (Std.=10.28%), which is nearly ten percent higher than that using only eye movements (Mean=59.81%, Std.=8.77%). BDAE achieves the best performance with the average accuracy and standard deviation of 79.71% and 4.76%, respectively. The feature fusion strategy is ranked between BDAE

TABLE I
AVERAGE ACCURACY AND STANDARD DEVIATIONS (%) OF EACH SINGLE MODALITY AND THE TWO MODALITY FUSION APPROACHES

Method	Eye	EEG	Feature Fusion	BDAE
Mean	59.81	69.50	73.65	**79.70**
Std.	8.77	10.28	8.90	**4.76**

and single EEG, with the values of 73.65% and 8.90%. Although feature fusion can also improve the accuracy of emotion recognition, the differences of accuracies suggest that BDAE can learn the high-level shared representations from the EEG and eye movement features.

We analyze the confusion matrices of EEG and eye movements to investigate the complementary characteristics. Fig. 2 present the confusion matrices of each single modality and the two modality fusion approaches. Fig. 3 illustrates a confused graph of EEG and eye movements for a more intuitive comparison between these two modalities. We observe that EEG is superior to eye movements in classifying happy, sad and disgust emotions, with the mean accuracies of 56%, 78%, and 66%, respectively, whereas eye movements have more discriminative power than EEG in recognizing fear (80% versus 77%) and neutral (74% versus 71%) emotions. Disgust emotion has the worst performance on both modalities, especially for eye movements, with an accuracy of only 33%. Despite this, the modality fusion increases the recognition accuracy of disgust to 58%. The experimental results indicate that EEG and eye movement have complementary contribution to emotion recognition.

For the performance of multimodal fusion approaches shown in Fig. 2, these two fusion methods can significantly improve the classification performance in sad, fear, and neutral emotions, especially for BDAE with improvements in accuracies of 43%, 7%, and 18%, respectively, compared with a single modality. However, these two modality fusion approaches have no advantage in classifying happy emotions compared to the single EEG modality. These phenomena are very similar to the conclusions of previous work in classifying three and four emotions [7][8]. Moreover, the classification of disgust emotion still remains the lowest.

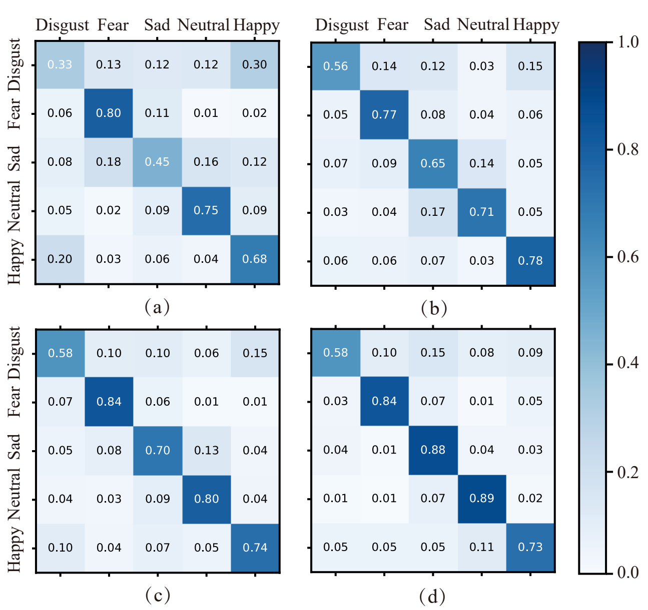


Fig. 2. The confusion matrices of each single modality and the two modality fusion approaches: (a) Eye movements. (b) EEG. (c) Feature Fusion. (d) BDAE.

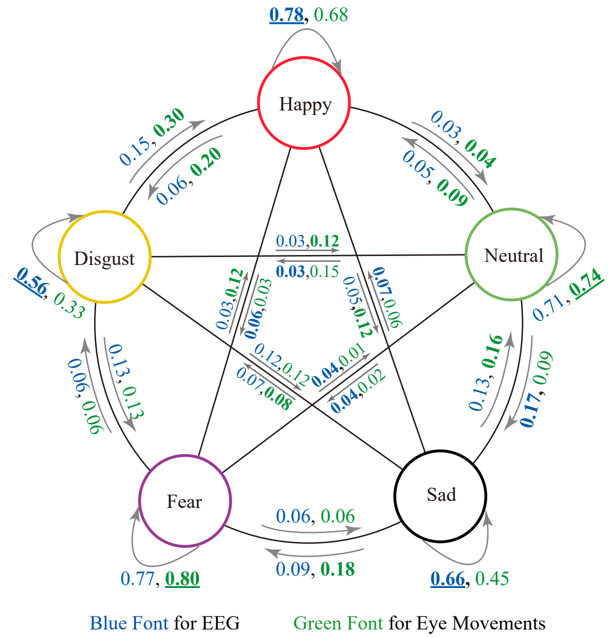


Fig. 3. Confusion graph of EEG and eye movements. The arrow indicates the direction of state transition. The numbers denote the percent corresponding with the number in Fig 2. The underlined and bold digits indicate the higher values.

In order to further explore the characteristics of EEG and eye movements on the five emotions, we study the EEG neural patterns for the five emotions and the classification performances of different eye movement features.

Fig. 4 illustrates the average energy distribution for happy, sad, disgust, fear, and neutral emotions in the gamma band, as we observe the most distinguishable neural patterns compared with other bands. For happy emotion, the lateral temporal areas activate more than the other emotions, while the energy of the prefrontal area is lower than negative emotions (sad, disgust, and fear). As shown in Fig. 2 (b), the sad and neutral emotions are most likely to be confused. Corresponding to Fig. 4 the neural patterns of neutral emotion are similar to sad emotion, which both have less activation in the temporal areas. As for disgust and fear emotions,

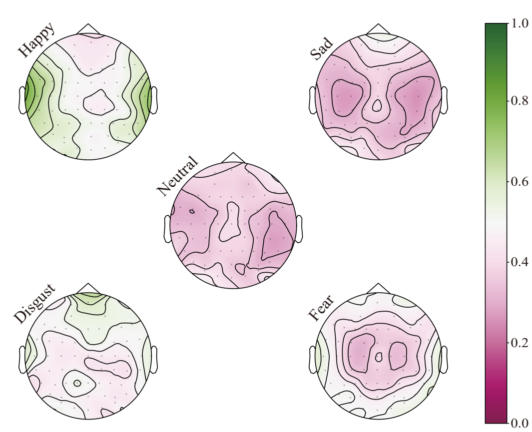
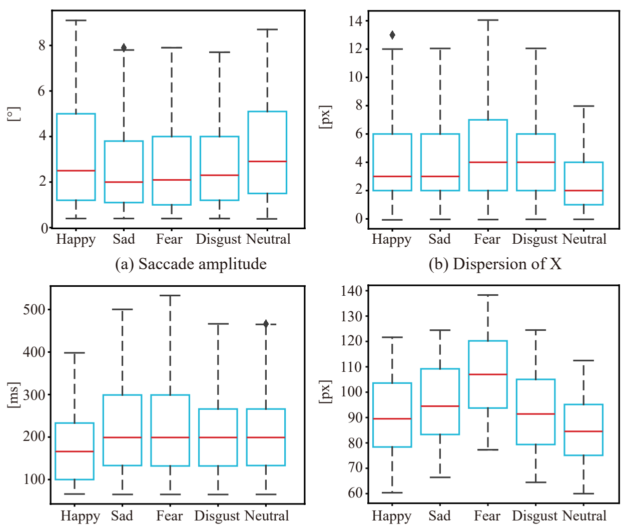


Fig. 4. Topographic maps of the five emotions in the gamma band.

the activated areas are the prefrontal and lateral temporal areas, while fear emotion has lower energy in the parietal area. Furthermore, it should be noted that the activated brain regions in our results are similar with the finding of existing work using fMRI [4].

TABLE II
CLASSIFICATION ACCURACY (%) OF DIFFERENT EYE MOVEMENT FEATURES.

	DP	Saccade	Fixation	Blink	Event	PD
Happy	30.56	40.88	66.54	33.45	67.19	**71.00**
Sad	39.77	43.67	45.27	37.58	**46.07**	34.47
Fear	52.40	35.02	34.73	32.05	58.60	**81.65**
Disgust	**32.23**	24.67	17.71	30.25	27.25	21.79
Neutral	71.68	54.60	40.78	38.40	67.61	**73.57**
All	45.99	40.45	41.10	34.70	53.30	**55.90**



(a) Saccade amplitude
(b) Dispersion of X
(c) Fixation duration
(d) Pupil diameter of X

Fig. 5. Box plots of four eye movement features. The red lines indicate the median.

For eye movement data, we further study the recognition performance of different eye movement features, as shown in Table II, where DP denotes dispersion, PD denotes pupil diameter. The values in the first five rows represent the performance of the corresponding emotion under different eye movement features, while the values in the All row denote the average accuracies of the classification for all five emotions. These results indicate that pupil diameter has a relatively higher discrimination ability than the other eye movement features and achieve a comparable result (55.90%) with using total features (59.81%). Fig. 5 shows the box plots of the four features. Just as the pupil diameter contributes the most to the classification, the box plot of pupil diameter is also the most distinguishable.

From Fig. 5 (d), we can conclude that neutral emotion has the smallest pupil diameter among the five emotions, which is consistent with the conclusion of our previous work on the three emotions [6]. On the contrary, fear has the greatest pupil response. Meanwhile, there is no significant difference between sad and disgust emotions. These phenomena explain why eye movements achieve better classification performance in fear and neutral emotions.

IV. CONCLUSIONS

This paper has shown that EEG and eye movements are complementary to emotion recognition in the classification of the five emotional states. The approach by combining EEG and eye movements can considerably enhance the performance of emotion recognition systems in comparison with each single modality. The best accuracy is 79.71% achieved by BDAE, which is much higher than those using single EEG and eye movements (59.81% and 69.50%, respectively). Further exploration shows that the topographic maps indicate the distinguishable characteristics of the neural patterns of the five different emotions, and the pupil diameter contributes the most to the classification. In the future, we plan to use more physiological signals in emotion recognition and explore the complementarity among them.

REFERENCES

[1] P. Ekman and W. V. Friesen, "Constants across cultures in the face and emotion." *Journal of Personality and Social Psychology*, vol. 17, no. 2, p. 124, 1971.

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

[3] T. L. Nwe, S. W. Foo, and L. C. De Silva, "Speech emotion recognition using hidden markov models," *Speech Communication*, vol. 41, no. 4, pp. 603–623, 2003.

[4] H. Saarimäki, A. Gotsopoulos, I. P. Jääskeläinen, J. Lampinen, P. Vuilleumier, R. Hari, M. Sams, and L. Nummenmaa, "Discrete neural signatures of basic emotions," *Cerebral Cortex*, vol. 26, no. 6, pp. 2563–2573, 2015.

[5] M. Soleymani, M. Pantic, and T. Pun, "Multimodal emotion recognition in response to videos," *IEEE Transactions on Affective Computing*, vol. 3, no. 2, pp. 211–223, 2012.

[6] W.-L. Zheng, B.-N. Dong, and B.-L. Lu, "Multimodal emotion recognition using EEG and eye tracking data," in *36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2014, pp. 5040–5043.

[7] Y.-F. Lu, W.-L. Zheng, B. Li, and B.-L. Lu, "Combining eye movements and EEG to enhance emotion recognition." in *International Joint Conference on Artificial Intelligence*, vol. 15, 2015, pp. 1170–1176.

[8] W.-L. Zheng, W. Liu, Y.-F. Lu, B.-L. Lu, and A. Cichocki, "Emotionmeter: A multimodal framework for recognizing human emotions," *IEEE Transactions on Cybernetics*, DOI: 10.1109/TCYB.2018.2797176.

[9] W. Liu, W.-L. Zheng, and B.-L. Lu, "Emotion recognition using multimodal deep learning," in *International Conference on Neural Information Processing*. Springer, 2016, pp. 521–529.

[10] T.-H. Li, W. Liu, W.-L. Zheng, and B.-L. Lu, "Classification of five emotions from EEG and eye movement signals: Discrimination ability and stability over time," *to appear in Proc. of 9th International IEEE EMBS Conference on Neural Engineering*.

[11] A. Schaefer, F. Nils, X. Sanchez, and P. Philippot, "Assessing the effectiveness of a large datebase of emotion-eliciting films: A new tool for emotion researchers," *Cognition and Emotion*, vol. 24, no. 7, pp. 1153–1172, 2010.

[12] O. G. Lins, T. W. Picton, P. Berg, and M. Scherg, "Ocular artifacts in recording EEGs and event-related potentials ii: Source dipoles and source components," *Brain Topography*, vol. 6, no. 1, pp. 65–78, 1993.

[13] M. M. Bradley, L. Miccoli, M. A. Escrig, and P. J. Lang, "The pupil as a measure of emotional arousal and autonomic activation," *Psychophysiology*, vol. 45, no. 4, pp. 602–607, 2008.

[14] R.-N. Duan, J.-Y. Zhu, and B.-L. Lu, "Differential entropy feature for EEG-based emotion classification," in *6th International IEEE/EMBS Conference on Neural Engineering*. IEEE, 2013, pp. 81–84.