Attention Evaluation with Eye Tracking Glasses for EEG-based Emotion Recognition

Zhen-Feng Shi, Chang Zhou, Wei-Long Zheng and Bao-Liang Lu* *Senior Member, IEEE*

***Abstract*— Attention of subjects in EEG-based emotion recognition experiments determines the quality of EEG data. Traditionally, self-assessment with questionnaires is used to evaluate the attention degree of subjects in experiments. However, this kind of self-assessment approach is subjective and inaccurate. Low quality EEG data from subjects without attention might influence the experiment evaluation and degrade the performance of affective models. In this paper, we extract scanpaths of subjects while watching emotion clips with eye tracking glasses and propose an attention evaluation method with spacial-temporal scanpath analysis. Based on the assumption that subjects with attention have similar scanpath patterns under the same clips, our approach clusters these similar scanpath patterns and evaluate the attention degree. Experimental results demonstrate that our proposed approach can cluster EEG features under attentive conditions effectively and significantly improve the classification performance. The mean accuracy of emotion recognition based on clustered high quality data is 81.70%, whereas the mean accuracy of using the whole dataset is 68.54%.**

I. INTRODUCTION

In order to provide relevant context information about users affective state, affective brain computer interfaces (aBCIs) require a reliable detection of user's emotion. Among multifarious approaches of emotion recognition, the method based on electroencephalogram (EEG) are more reliable due to its high accuracy and objective evaluation comparing to other methods based on outward manifestation like face expression and body gestures. In studies of emotion recognition based on EEG, using film clips as stimuli to elicit emotions has been proved to be reliable [1] and is widely accepted. Only the data with right elicited emotions, regarded as high quality data, should be used in further analysis. In previous studies [1] and [2], self-assessment with questionnaires was used to evaluate the quality of data, which is rather subjective and inaccurate. Therefore, finding an objective method which can automatically distinguish the high quality data and low quality data is necessary.

With an increase in the availability of eye trackers and a reduction in their costs, eye tracking technology has been extensively employed in various fields including but not limited to advertising [3] and language [4] research. In our previous work [5][6], we proposed multimodal emotion recognition framework combining EEG and eye movement data and demonstrated its effectiveness. We extracted twenty-one features from eye movement, identified the intrinsic patterns of three emotional states and reached an accuracy of 77.8%. Meanwhile, eye tracking data can be used to research for usability analysis and assessment since they can provide a natural and efficient way to observe the behavior of users. In the work of [7], seven descriptors are extracted from eye movement data to evaluate the engagement of subjects. However, very few methods are available for studying the sequential properties of fixations, despite they are fundamentally sequential (one fixation after another). Markov models have been used successfully in certain domains, such as eye movement modeling in reading [8], but the implementation is too complex.

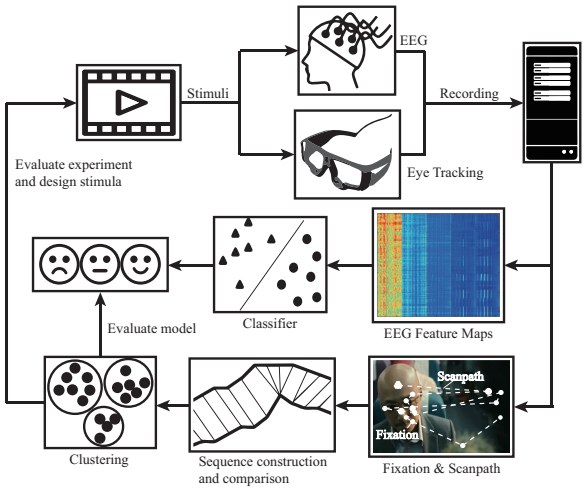


Fig. 1. The proposed framework for evaluating the attention of subjects during experiments by using spacial-temporal scanpath analysis.

In this work, we proposed a noval spacial-temporal scanpath analysis method as illustrated in Figure 1, which automatically classifies the high quality data and low quality data by evaluating the attention of subjects.

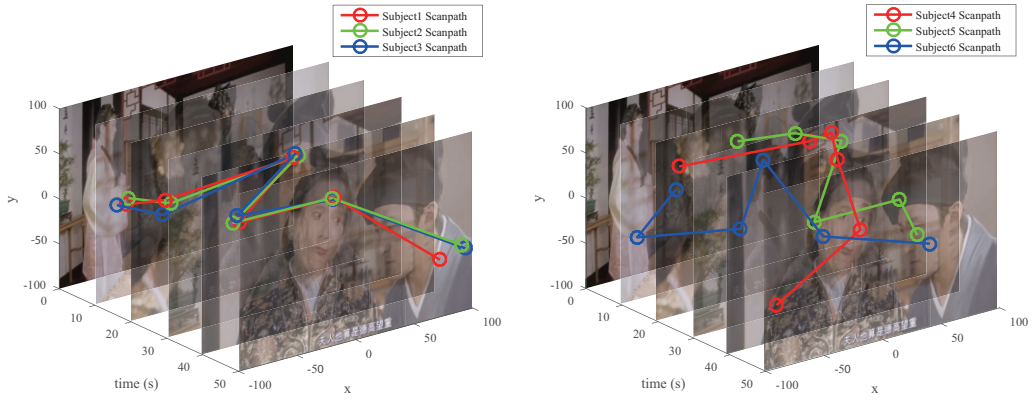
II. EXPERIMENT SETUP

Using film clips as stimuli to elicit subject's emotions has been proved to be reliable by previous studies. Fifteen clips, 5 for positive, 5 for negative and 5 for neutral, are selected and edited properly to form 15 sessions. Each

This work was supported in part by the grants from the National Natural Science Foundation of China (Grants No. 61272248 and No. 61673266), the National Basic Research Program of China (Grant No.2013CB329401), and the Major Basic Research Program of Shanghai Science and Technology Committee (15JC1400103).

Zhen-Feng Shi, Chang Zhou, 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)



(a) Scanpaths with attention

(b) Scanpaths without attention

Fig. 2. Scanpaths for subjects watching film clips with attention and without attention

session consists of a 5 s hint for starting and 60 s rest after each clip for subjects to recover from elicited emotions. Experiments are held in a soundproof room and the light condition is strictly controlled by the indoor illumination system. All the subjects are instructed to sit comfortably, facing a large screen and move as least as possible. In these common conditions, Two separate experiments were held.

A. Experiment 1

In this experiment, only eye tracking data are collected. We use the SensoMotoric Instruments (SMI) eye tracking glasses (ETG) to accurately track the eye movement. Ten reliable subjects (all males, aged between 20 and 23), with normal or corrected-to-normal vision and normal hearing participated in this experiment. They are informed about the experiment and instructions before starting. In particular, 5 of them are required to watch each clips with high attention and elicit their own corresponding emotions. The other 5, on contrary, are required to watch each clips without attention.

B. Experiment 2

In this experiment, both eye tracking data and EEG data are collected simultaneously. We use the same equipment to collect eye tracking data. EEG was recorded using an ESI NeuroScan System at a sampling rate of 1 KHz from 62-channel electrode cap. The impedance of each electrode is less than 5 $K\Omega$. Twenty-six subjects, 15 males and 11 females, aged between 20 and 26, with self-reported normal or corrected-to-normal vision and normal hearing, participated in this experiment. All of them are supposed to watch the clips with attention and elicit their own corresponding emotions.

III. Proposed Algorithm

Eye movements can be regarded as a spacial-temporal sequence. For an identical static picture, the area of interests (AOIs) of different subjects remain similar. For an identical video clip, it can be regarded as a sequence of static pictures playing raw. Therefore, when subjects are watching the same clip with high attention, the sequence of their scanpath should be alike. Figure 2(a) illustrates the part of the scanpath of 3 different subjects watching film clips with attention. Figure 2(b) illustrates the part of the scanpath of 3 different subjects watching film clips without attention.

A. Algorithm for Attention Evaluation

1) Encoding a Sequence: Scanpath can be represented by a sequence of fixations. Since each fixation are indicated with 5 parameters, which are fixation position x and y, start time, duration and end time, we encode the scanpath to make it easy to be compared. To begin with, we divide the view of eye tracking glasses into 30 * 40 regions and distribute each region with a pair of unique numbers. We encoded each fixation position x and y with the pair of numbers corresponding to that region and initial a sequence with the order of fixation start time. To take the fixation duration into account, we introduced temporal binning into the sequence by repeating the pair of numbers corresponding to the position in a way that is proportional to the fixation duration. Normal effective fixation duration is longer than 50 ms. Therefore, we ignore the fixation whose duration is less than 50 ms and for those longer than 50 ms, we repeat and insert the pair of numbers once for each 50 ms into the sequence right behind the original one. In this way, the sequence encoded from the fixation data incorporates spacial location, sequential information, and temporal duration.

2) Comparing Sequences: Scanpath sequences encoded as above should share high similarity between data collected from subjects with high attention. However, scanpath sequences tend to be of different length even for identical clips. We apply FastDTW, an approximation algorithm of Dynamic Time Warping in linear time and space developed by S. Salvador and Chan [9]. Dynamic Time Warping (DTW) is an algorithm for measuring similarity between two temporal sequences which may vary in speed and gives a distance-like quantity. It calculates an optimal match between two given sequences with certain restrictions. The sequences are 'warped' non-linearly in the time dimension to determine a measure of their similarity independent of certain non-linear variations in the time dimension. The optimal warp path is the warp path with minimum distance. The distance of a

warp path W defines as

$$Dist(W) = \sum_{k=1}^{k=K} Dist(w_{ki}, w_{kj}),$$

where $Dist(W)$ is the distance of warp path W, and $Dist(w_{ki}, w_{kj})$ is the distance between the two data point indexes in the k^{th} element of the warp path. DTW typically uses dynamic programming, which requires $O(N^2)$ in general and is time consuming on large dataset. Hence, we choose to use a fast method called FastDTW, which adopts a multilevel approach, recursively projects a warp path from a coarser resolution to the current resolution, and refines it.

Based on FastDTW, we calculate the distance for each pair of scanpath sequences from the same clips, to form a distance matrix. For scanpath sequences encoded from subjects with attention, the distance should be relatively smaller comparing to those from subjects without attention.

3) Clustering: The purpose of clusering algorithm is to distinguish subjects with attention and subjects without attention. Based on the distance matrix calculated by the previous steps, we perform a density-based clustering algorithm to fulfil our task. The clustering algorithm should meet the following requirements:

- It should be density-based and allow using precalculated distance matrix as the input;
- It should be able to detect outliers, have a notion of noise, and be robust to outliers;
- It should work fine when the distances in matrix may not hold the triangle inequality.

DBSCAN [10] algorithm is a classic density-based clustering algorithm and has mature implementation. By performing this algorithm, the subjects are classified into outliers and non-outliers, which are considered as subjects without attention and subjects with attention, respectively.

Algorithm 1: Overall Algorithm for Attention Evaluation

input : Raw data F, contains sequence of fixation position, fixation duration, fixation start time for X samples
output: R $= (1, X)$, cluster label for each sample

```
for i ← 1 to X do
    S [i] ← EncodeSequence(F [i]);
Matrix ← zeros(X, X);
for i ← 1 to X do
    for j ← i to X do
        Matrix [i, j] ← FastDTW(S [i], S [j]);
        Matrix [j, i] ← Matrix [i, j];
R ← DBSCAN(Matrix);
return R;
```

B. Algorithm for Emotion Recognition

1) Signal Preprocessing: Due to the contamination of electromyography (EMG) signals from facial expressions and electrooculogram (EOG) signals from eye movements in EEG data [11], a bandpass filter between 1 Hz and 75 Hz is applied to discard the noise and artifacts. Then, a downsampling of 200 Hz is further employed on the 62-channel EEG signal. Features of EEG are extracted from continuous non-overlapping 1 s time windows of each clip.

2) Feature Extraction: Differential entropy (DE) features are employed for feature extraction. According to the work of [12], DE features have superior performance than power spectral density (PSD) features and they are equivalent to the logarithm of PSD in a certain frequency band with a fixed length EEG sequence. Based on short-term Fourier transform, DE features can be calculated from five frequency bands (δ: 1-3 Hz, θ: 4-7 Hz, α: 8-13 Hz, β: 14-30 Hz, and γ: 31-50 Hz). The total dimension of extracted features is 310.

3) Feature Smoothing: The features extracted as above tend to have strong fluctuations. Because that emotion changes tend to be slow, we apply the linear dynamic system smoothing [13] to filter out uncorrelated features caused by other brain activities.

4) Classification: We use the first nine sessions, including 3 positive, 3 negative and 3 neutral clips, as the training set. The rest six sessions, including 2 positive, 2 negative and 2 neutral clips are used as the testing set. Linear SVM is employed to train a generic classifier for three emotional states.

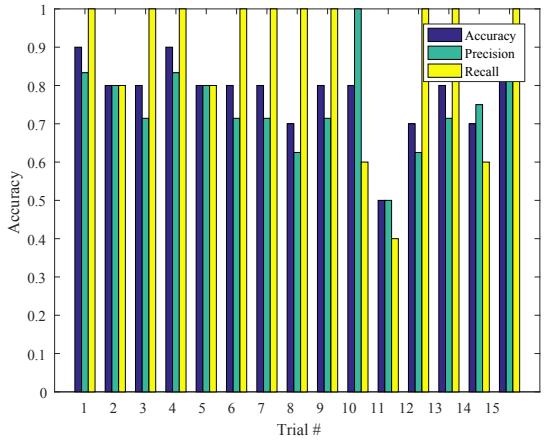


Fig. 3. Attention evaluation for ten subjects on seperate trials

IV. Experiment Result

A. Result of Experiment 1

Ten subjects participated this experiment. Five of them are asked to watch clips with attention and labeled as positive while the other five are asked to watch without attention, labeled as negative. Subjects are numbered from A1 to A10. Subjects A1 to A5, all labeled positive, are clustered together. Subjects from A6 to A10, except subject A7, all labeled negative, are correctly detected as outliers. A possible reason is that subject A7 stared at the subtitle for the whole experiment, which happens to be a popular scanning area for subjects with attention.

Attention evaluation on different types of clips may vary on its performance. Figure 3 shows the accuracy, precision, and recall of attention evaluation on separate trials (clips). The trial 8 and trial 11 have obviously worth performance, because these corresponding clips are from scenery films, and may not share common AoIs even for subjects with attention.

B. Result of Experiment 2

Twenty-six subjects participated in this experiment. All of them are required to watch film clips with attention to elicit their corresponding emotion. Both EEG and eye tracking data have been collected simultaneously from the subjects. Emotion recognition accuracy and attention evaluation are calculated by applying the proposed algorithms.

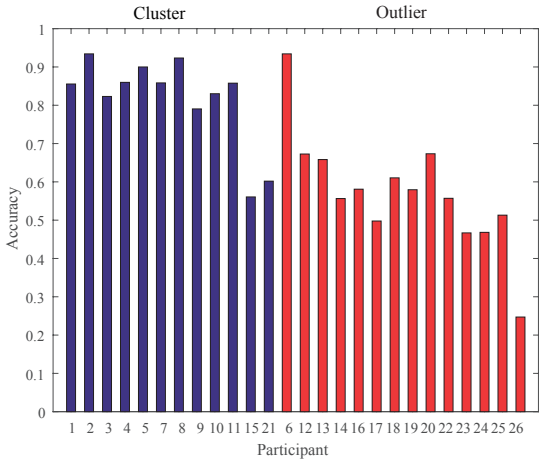


Fig. 4. Emotion recognition accuracy and attention evaluation for 26 subjects

Figure 4 shows the result for 26 subjects. The subjects marked by red at right side are outliers detected based on eye tracking data, which have relative lower emotion recognition accuracy than the subjects marked by blue at the left side. Subject 6 is an abnormity for attention evaluation. This is because subject 6 did watch the clips with attention and elicit his own emotion successfully. However, his eye tracking data was seriously contaminated because he squinted a lot, which made the eye tracking glasses be unable to track his fixation.

TABLE I
AVERAGE EMOTION RECOGNITION ACCURACY. THE P-VALUE BETWEEN SUBJECTS WITH ATTENTION AND SUBJECTS WITHOUT ATTENTION IS 0.0001. FOR ALL SUBJECTS AND SUBJECTS WITH ATTENTION, THE P-VALUE IS 0.0290. FOR ALL SUBJECTS AND SUBJECTS WITHOUT ATTENTION, THE P-VALUE IS 0.0560.

	All subjects	Subjects with attention	Subjects without attention
Average accuracy	68.54	**81.70**	57.26
Standard deviation	18.29	**11.80**	15.13

Table I shows the average emotion recognition accuracy for three different datasets. We consider subjects whose eye tracking data is clustered together as the subjects with attention. Those whose eye tracking data is detected as outliers are considered as subjects without attention. Average emotion recognition accuracy for subjects with attention reaches 81.70%, comparing to 68.54% for all subjects and 57.26% for subjects without attention. The *p*-values for *All Subjects* and *Subjects with Attention* is less than 0.05 and for *Subjects with Attention* and *Subjects without Attention* is less than 0.01, respectively.

V. CONCLUSION

In this paper, we have proposed an attention evaluation method for automatically classifying high quality data and low quality data by using spacial-temporal scanpath analysis. The experimental results have demonstrated effectiveness of our proposed method and the attractive features for improving emotion recognition accuracy. In the future, we will carry out experiments with more number of subjects in different ages and investigate the gender difference in attention evaluation.

REFERENCES

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

[2] 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 Transactions on Affective Computing*, vol. 3, no. 1, pp. 18–31, 2012.

[3] L. Maughan, S. Gutnikov, and R. Stevens, "Like more, look more. look more, like more: The evidence from eye-tracking," *Journal of Brand Management*, vol. 14, no. 4, pp. 335–342, 2007.

[4] K. Rayner, "Eye movements and attention in reading, scene perception, and visual search," *The quarterly journal of experimental psychology*, vol. 62, no. 8, pp. 1457–1506, 2009.

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

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

[7] P. K. Podder, M. Paul, T. Debnath, and M. Murshed, "An analysis of human engagement behaviour using descriptors from human feedback, eye tracking, and saliency modelling," in *International Conference on Digital Image Computing: Techniques and Applications*. IEEE, 2015, pp. 1–8.

[8] R. Engbert and R. Kliegl, "Mathematical models of eye movements in reading: A possible role for autonomous saccades," *Biological Cybernetics*, vol. 85, no. 2, pp. 77–87, 2001.

[9] S. Salvador and P. Chan, "Toward accurate dynamic time warping in linear time and space," *Intelligent Data Analysis*, vol. 11, no. 5, pp. 561–580, 2007.

[10] M. Ester, H.-P. Kriegel, J. Sander, X. Xu, *et al.*, "A density-based algorithm for discovering clusters in large spatial databases with noise." in *the 2nd International Conference on Knowledge Discovery and Data Mining*, vol. 96, no. 34, 1996, pp. 226–231.

[11] M. Fatourechi, A. Bashashati, R. K. Ward, and G. E. Birch, "EMG and EOG artifacts in brain computer interface systems: A survey," *Clinical Neurophysiology*, vol. 118, no. 3, pp. 480–494, 2007.

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

[13] L.-C. Shi and B.-L. Lu, "Off-line and on-line vigilance estimation based on linear dynamical system and manifold learning," in *Annual International Conference of the IEEE Engineering in Medicine and Biology*. IEEE, 2010, pp. 6587–6590.