7th Annual International IEEE EMBS Conference on Neural Engineering
Montpellier, France, 22 - 24 April, 2015

Rapid Face Recognition Based on Single-Trial Event-Related Potential Detection over Multiple Brains*

Lei Jiang, Yun Wang, Bangyu Cai, Yiwen Wang*, Weidong Chen and Xiaoxiang Zheng

Abstract—**The automatic machine face recognition has achieved great performance but is still far from satisfactory in the uncontrolled environments. Human brain has a powerful ability to recognize faces across various conditions, which makes it possible to introduce brain-computer interface (BCI) technology to face recognition. However, the performance of single-participant based BCI suffers from the low signal-to-noise ratio of electroencephalography (EEG) signals, especially detecting the event-related potential (ERP) in single trial. Here, we propose a rapid face recognition approach based on single-trial ERP detection, but the EEG signals are integrated from multiple participants. After the first-layer classifier to detect the ERP of each individual, the support vector machine scores of all individuals are concatenated and inputted to a second-layer classifier by the voting method to obtain the final classification results. The results show that our approach significantly outperforms the single-participant based BCI, and the voting method is superior to the ERP averaging method. In addition, when the area under the receiver operating characteristic curve is required to be greater than 0.9, our approach can recognize target faces about 300 ms ahead of button-press responses when integrating EEG signals from 9 participants. These results indicate that our approach can integrate decisions from multiple individuals to achieve rapid and accurate face recognition.**

I. INTRODUCTION

Face recognition is a hot topic in pattern recognition and machine vision. As research continues, automatic machine face recognition has achieved satisfactory performance on face images captured in well-controlled environments, however, uncontrolled lighting, large pose and facial expression variations, and severe partial occlusions can significantly affect the performance of automatic machine face recognition [1]. In contrast to machine vision, human brain has an extraordinary ability to recognize individual faces. We can recognize faces accurately and effortlessly across large viewpoint changes, under poor lighting conditions, even only partial views are available. And some event-related potential (ERP) components related to face recognition have been found [2, 3]. Among these ERP components, the N250 which is believed to be the earliest ERP component that is involved in face recognition, emerges as early as around 230 ms after stimulus onset [2, 3]. This makes it possible to introduce brain-computer interface (BCI) technology to directly decode ERP components related to face recognition [4], thus bypassing the motion related procedures (*e.g.*, pressing a button to indicate the target faces), therefore implementing an electroencephalography (EEG)-based rapid face recognition system.

However, due to the low signal-to-noise ratio (SNR) of EEG signals, the performance of BCI applications is relatively low. To enhance the SNR of EEG signals, trial-averaging approach has been widely used in BCI systems [4]. In these systems, each stimulus is repeated multiple times to produce enough trials for the subsequent averaging process. But for a real-time BCI system, this conventional trial-averaging approach is practically inappropriate. Under these circumstances, integrating EEG signals from multiple individuals may be a feasible alternative. Combining opinions from multiple brains leading to greater decision accuracy, a phenomenon known as collective wisdom, is ubiquitous in human society [5]. Similarly, integrating single-trial EEG signals from multiple individuals, a technique known as multi-brain computing [5], might obtain better overall BCI performance. For instance, in [6], Peng Yuan *et al.* propose a BCI system to detect visual targets by integrating and identifying multi-participants' visual evoked potentials, and the classification performance of this system is significantly higher than the single-participant based classification.

We are interested in whether robust face recognition can be obtained by the multi-brain computing method. Ten participants are instructed to perform a face recognition task in the rapid serial visual presentation paradigm, and their corresponding single-trial EEG signals are integrated and analyzed to identify which face images are the targets. To integrate the EEG signals collected from multiple participants, we use the voting method where the output of the sub-classifier trained for each participant is concatenated as the input of another classifier to obtain the classification result. We compare the voting method with the ERP averaging method where single-trial EEG signals are averaged across participants and these averaged EEG signals are used for classification. We investigate the feasibility of the multi-brain computing method in improving the speed and accuracy of face recognition.

*This work is supported by grants from National High Technology Research and Development Program of China (No. 2012AA011602), National Basic Research Program of China (No.2013CB329506), Natural Science Foundation of China (No.61473261, 61305146, 31371001, 61233015), Zhejiang provincial international science and technology cooperation program (No.2012C24025), Zhejiang provincial Natural Science Foundation of China (No. LY14F030015), and the Fundamental Research Funds for the Central Universities.

Lei Jiang and Weidong Chen are with College of Computer Science and Technology, Zhejiang University, China (e-mail: fishjianglei@gmail.com, chenwd@zju.edu.cn). Yun Wang, Bangyu Cai and Xiaoxiang Zheng are with College of Biomedical Engineering & Instrument Science, Zhejiang University (e-mail: 56wangyun@gmail.com, cbangyu@gmail.com, zxx667@gmail.com). Yiwen Wang and Xiaoxiang Zheng are also with Key Laboratory of Biomedical Engineering of Ministry of Education, Zhejiang University (phone: 86-571-87952339; fax: 86-571-87952865; e-mail: eewangyw@zju.edu.cn).

II. Methods

A. Experiment Design and Data Acquisition

Ten participants (5 males, 5 females) participate in this study. Face stimuli are composed of 1854 different color images of 46 celebrities (half male and half female), each celebrity has 34-56 different exemplars (mean = 40.3, STD = 3.2). To exclude non-facial identity cues of each image, all face images are cropped and scaled to 200 X 200 pixels. The face images have varying illumination, poses and expressions, even are partially occluded.

Participants are seated about 90 cm in front of a computer screen. The whole experiment consists of 14 blocks of target face recognition task. In each block, one celebrity is randomly selected from the 46 candidate celebrities, 10 face images of the selected celebrity serve as targets, which are needed to be recognized by the participants. In addition, 150 face images of the other 45 celebrities serve as nontargets, making a block composed of 160 face images (10 targets, 150 nontargets) in total. In the time course of the experiment, each face image is presented centrally in a random order for 500 ms, followed by a blank inter-stimulus interval (ISI) of 500 ms. Before the start of a block, participants are shown one face image of the target celebrity to familiarize them with the facial features of the target celebrity. Throughout each block, participants are required to maintain central eye fixation, and press left button as quickly and accurately as possible when they recognize the target faces. Breaks are encouraged between blocks to minimize fatigue and eye movements.

The EEG signals are recorded by 60 Ag/AgCl electrodes using Neuroscan Synamps system (1000 Hz sampling rate, 200 Hz low-pass and 50 Hz notch, impedances < 30 kΩ). EEG signals are referenced to the nose, electrode AFz serves as grounding electrode. Blinks are monitored by vertical electrooculogram (EOG) electrodes located above and below the left eye.

B. Data Preprocessing

Button-press responses falling between 300 and 1000 ms post target face stimuli onset are considered correct. We calculate the distribution, mean and standard deviation of the response time of correct button-press responses.

The EEG signals are first band-pass filtered from 0.5 to 30 Hz using a second order Butterworth filter. Then the EEG signals are segmented into epochs from 0 to 800 ms after stimuli onset, with the average of the 200 ms pre-stimulus interval as baseline. For the subsequent classification, EEG epochs are downsampled from 1000 Hz to 100 Hz. Then the EEG epochs are concatenated by channel to create a feature vector for each face stimulus.

C. Classification

For the single-participant based classification, we adopt a Gaussian kernel support vector machine (SVM) implemented by LIBSVM [7]. The trials in the first 8 blocks (160 trials per block) serve as training data, and those in the last 6 blocks serve as test data. The parameters of the Gaussian kernel (kernel width σ^2 and regularization parameter C) are determined by a 3-fold cross validation. In addition, we use the area under the receiver operating characteristic (ROC) curve (AUC) as the performance evaluation criterion [8].

For the multi-brain computing based classification, we adopt two voting methods, named "two-layer SVM" and "SVM + linear discriminant analysis (LDA) [9]".

1) Two-layer SVM: The first-layer SVM classifiers are used for individual classification for the participants. The outputs of the first-layer SVM classifiers from multiple participants, which indicate the probabilities that the single-trial EEG epochs are corresponding to target face stimuli, are concatenated as the input of the second-layer SVM classifier to obtain the final classification results.

2) SVM + LDA: Similar to the "two-layer SVM" method, except that the second-layer classifier is replaced by the LDA classifier.

For comparison, we also use the ERP averaging method where single-trial EEG signals are averaged across multiple participants and these averaged EEG signals serve as inputs to a SVM classifier for target/nontarget face classification. Since the number of participants is a crucial factor for a multi-brain computing based BCI system, we first evaluate the performance of our system with respect to the number of participants. For each number N (from 1 to 10), we randomly sample 1000 groups of N participant(s). For each random group of N participant(s), we integrate the single-trial EEG signals (0 to 800 ms after stimuli onset) across participants using the two voting methods and the ERP averaging method, respectively. To examine the effect of the number of participants on the system performance, we conduct one-way Analysis of Variances (ANOVAs). Additionally, post hoc two-sample t-tests are employed to investigate whether the performance of a multi-brain computing based BCI is significantly better than the single-participant based BCI.

We are also interested in the potential time-savings in face recognition through analyzing EEG signals integrated from multiple participants, compared to the button-press responses. The face recognition is speeded up if the faces can be accurately recognized before the button-press responses.

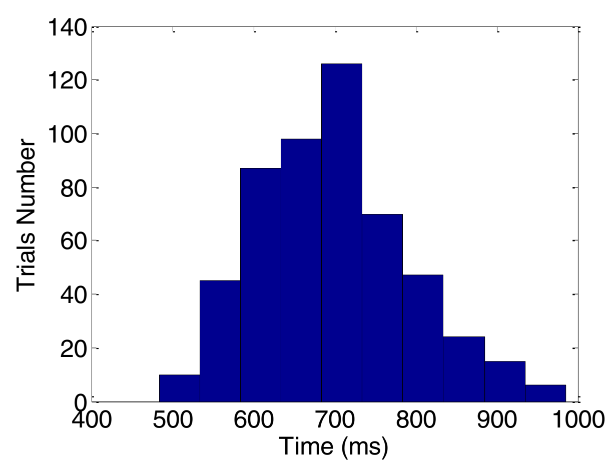


Fig. 1. The distribution of the button-press response time (RT) over the participants.

Therefore, we evaluate the system performance with different time windows used for classification. Time windows with a length of L ms, which start from the face stimuli onset, are used for classification, L ranges from 20 ms to 800 ms with a step size of 20 ms (*i.e.*, time windows are 0-20 ms, 0-40 ms, …, 0-800 ms).

III. Results

A. Behavioral Results

Fig. 1 shows the distribution of the button-press response time (RT) over the participants. The RT is 698 ± 94 ms, and no RT is shorter than 480 ms in each trial.

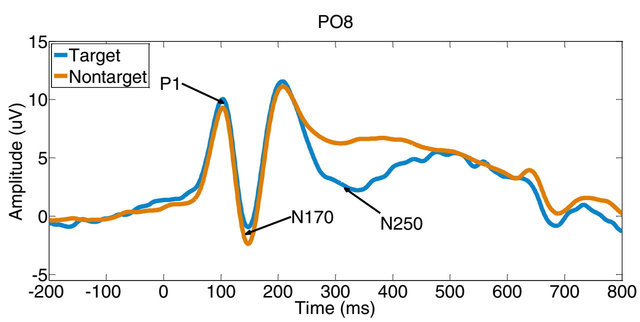
B. Grand-average ERPs

In Fig. 2, the grand-average ERPs are visualized. Within 210 ms after face stimuli onset, the ERPs (P1 and N170) elicited by target and nontarget face stimuli are highly similar. While, compared to nontarget face stimuli, the target face stimuli also elicit robust N250 (230-400 ms), P3 (400-600 ms) and N4 (600-760 ms). Therefore, it could be inferred that the ERP components occur within 210 ms after face stimuli onset have no contributions to the classification, while the subsequent N250, P3 and N4 would play important roles in the classification.

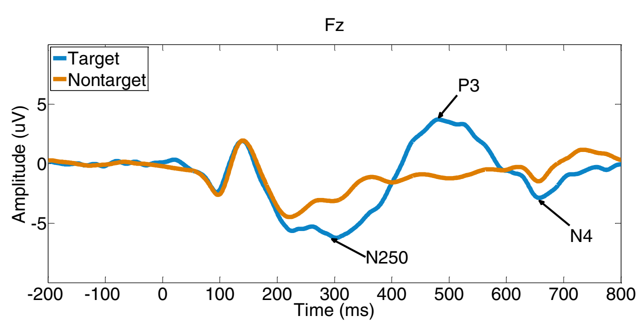
C. Classification Performance

Fig. 3 shows the classification performance for the two voting methods and the ERP averaging method as a function of the number of participants. The EEG signals from 0 to 800 ms after stimuli onset are used for feature extraction. The performance of the single-participant based face recognition, in term of the AUC, is 0.859 ± 0.0564. As the number of participants increased, the average classification performance for the two voting methods and the ERP averaging method is improved along with the standard deviation decreased. One-way ANOVAs show that the "number of participants" has significant effect on the classification performance (ERP averaging: $F(9,9990) = 2057.48$, $p = 0$; Two-layer SVM: $F(9,9990) = 2542.57$, $p = 0$; SVM + LDA: $F(9,9990) = 3162.43$, $p = 0$). Post hoc two-sample t-tests (10 samples for the single-participant based classification versus 1000 samples for the multi-brain computing based classification) reveal that two participants are enough for the multi-brain computing based classification to obtain significantly better performance than the single-participant based classification when using the two voting methods and the ERP averaging method (one-tailed t-tests; ERP averaging: $t(1008) = -2.6038$, $p = 0.0047$; Two-layer SVM: $t(1008) = -2.4093$, $p = 0.0081$; SVM + LDA: $t(1008) = -4.7455$, $p = 1.19e-006$). When integrating EEG signals from all the 10 participants, the two voting methods and the ERP averaging method all achieve high performance (AUC = 0.984 for the ERP averaging method, 0.996 for the "two-layer SVM" method, 0.997 for the "SVM + LDA" method).

In addition, as we can see from Fig. 3, the two voting methods show similar performance, and always obtain greater performance improvements than the ERP averaging method as EEG signals from multiple participants are integrated. The reason why the ERP averaging method results in the least performance improvements may be due to the individual differences in ERPs. As reported in [10], the amplitudes and latencies of the ERPs show large variations among individuals. As a result, although ERP averaging reduces the background noise in EEG signals, it might also filter out much useful information. For example, when two adjacent ERP components have opposite polarities, due to the ERP latency differences among individuals, ERP signals might be cancelled out when averaging across multiple participants. In contrast to the ERP averaging method, since the EEG signals are independent between participants, it is reasonable to assume that our voting methods do not lose useful information for classification [11]. Hence, the two voting methods are superior to the ERP averaging method in our system.



(a)



(b)

Fig. 2. The grand-average ERPs in electrode PO8 and electrode Fz.

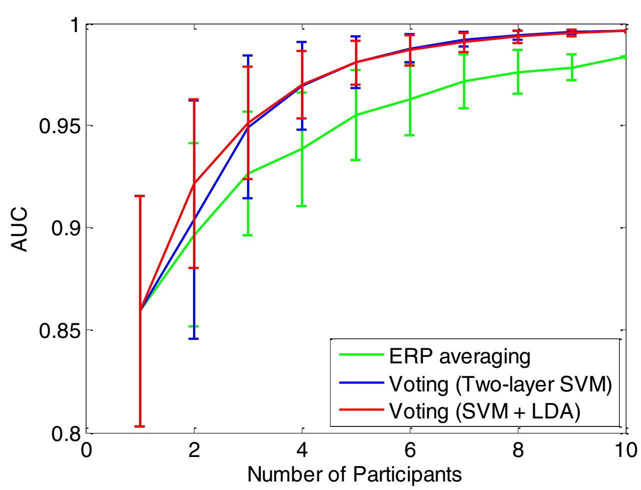


Fig. 3. The classification performance for the two voting methods and the ERP averaging method as a function of the number of participants.

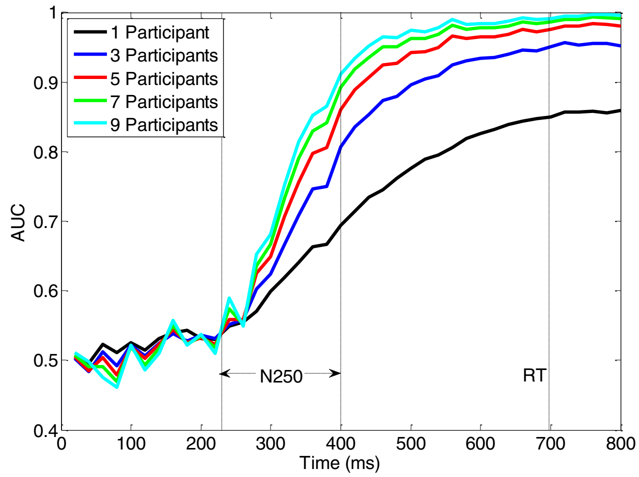


Fig. 4. The classification performance for the "SVM + LDA" method as a function of the length of the time windows used for feature extraction. The time window of the N250 component and the mean button-press response time (RT) are marked.

As mentioned in the methods section, we are also interested in the potential time-savings in face recognition through analyzing EEG signals integrated from multiple participants, compared to the button-press responses. Fig. 4 shows the classification performance for the "SVM + LDA" method as a function of the length of the time windows used for feature extraction. The classification performance for 1, 3, 5, 7, and 9 participants is shown. When using EEG signals prior to the emergence of the N250 for classification, the performance remains close to chance level (AUC = 0.5), while the performance increases significantly starting with the EEG signals in the time window of the N250 involved in the classification (Fig. 4). These results are consistent with the findings in ERP analysis (Fig. 2) that the ERPs elicited by target and nontarget faces start to show differences at the time window of the N250. More importantly, from the Fig. 4, we also observe that the potential time-savings in face recognition are influenced by two factors, *i.e.*, the number of participants and the desired performance. When it is required to achieve an AUC of 0.85, face recognition decisions could be made within 400 ms, which is about 300 ms ahead of actual button-press responses, if integrating EEG signals from 5 or more participants. When an AUC of 0.9 is required, it needs to integrate EEG signals from 9 or more participants to make face recognition finished within 400 ms. Considering the P3 occurs beyond 400 ms after face stimulus onset, our results mean that, under certain accuracy requirements, if the EEG signals from enough participants are integrated, then the N250 component in the integrated EEG signals can provide enough discriminative information, the P3 and the subsequent components are not needed to be involved in the classification.

IV. Conclusions and Discussion

In this paper, we propose a face recognition approach based on decoding single-trial EEG signals integrated from multiple individuals. The experiment results demonstrate that our approach significantly outperforms the single-participant based BCI. Two kinds of methods (voting, ERP averaging) for integrating EEG signals are compared. The results show that the voting methods are superior to the ERP averaging method, which may be due to the individual differences in the ERPs. Furthermore, we demonstrate that under certain accuracy requirements, if the EEG signals from enough participants are integrated, our approach can accurately make face recognition decisions 300 ms ahead of actual button-press responses. In these circumstances, the P3 and the subsequent components are not needed to be involved in the classification, the N250 component in the integrated EEG signals can provide enough discriminative information. In the future work, more sophisticated algorithms should be used to improve the performance of single-participant based BCI, thus decreasing the number of participants needed for a feasible multi-brain computing based BCI system. In addition, some more efficient methods should be developed to better integrate the EEG signals from multiple participants.

Acknowledgment

The authors thank Mr. Yimin Shen from Department of BME of Zhejiang University for his assistance and supports in experiments.

References

[1] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *Acm Computing Surveys (CSUR),* vol. 35, pp. 399-458, 2003.

[2] A. Gosling and M. Eimer, "An event-related brain potential study of explicit face recognition," *Neuropsychologia,* vol. 49, pp. 2736-2745, 2011.

[3] X. Zheng, C. J. Mondloch, and S. J. Segalowitz, "The timing of individual face recognition in the brain," *Neuropsychologia,* vol. 50, pp. 1451-1461, 2012.

[4] A. Vallabhaneni, T. Wang, and B. He, "Brain-Computer Interface," in *Neural Engineering*, B. He, Ed., ed: Springer US, 2005, pp. 85-121.

[5] M. P. Eckstein, K. Das, B. T. Pham, M. F. Peterson, C. K. Abbey, J. L. Sy, *et al.*, "Neural decoding of collective wisdom with multi-brain computing," *NeuroImage,* vol. 59, pp. 94-108, 2012.

[6] Y. Peng, W. Yijun, W. Wei, X. Honglai, G. Xiaorong, and G. Shangkai, "Study on an online collaborative BCI to accelerate response to visual targets," in *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*, San Diego, 2012, pp. 1736-1739.

[7] C.-C. Chang and C.-J. Lin, "LIBSVM: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology (TIST),* vol. 2, pp. 1-27, 2011.

[8] T. Fawcett, "ROC graphs: Notes and practical considerations for researchers," *Machine Learning,* vol. 31, pp. 1-38, 2004.

[9] B. Blankertz, S. Lemm, M. Treder, S. Haufe, and K.-R. Müller, "Single-trial analysis and classification of ERP components—a tutorial," *NeuroImage,* vol. 56, pp. 814-825, 2011.

[10] S. J. Luck, *An introduction to the event-related potential technique*: MIT press, 2014.

[11] T. G. Dietterich, "Ensemble methods in machine learning," in *Multiple classifier systems*. vol. 1857, ed: Springer, 2000, pp. 1-15.