Augmenting Group Performance in Target-Face Recognition via Collaborative Brain-Computer Interfaces for Surveillance Applications

Davide Valeriani, Caterina Cinel, and Riccardo Poli[1]

***Abstract*— This paper presents a hybrid collaborative brain-computer interface (cBCI) to improve group-based recognition of target faces in crowded scenes recorded from surveillance cameras. The cBCI uses a combination of neural features extracted from EEG and response times to estimate the decision confidence of the users. Group decisions are then obtained by weighing individual responses according to these confidence estimates. Results obtained with 10 participants indicate that the proposed cBCI improves decision errors by up to 7% over traditional group decisions based on majority. Moreover, the confidence estimates obtained by the cBCI are more accurate and robust than the confidence reported by the participants after each decision. These results show that cBCIs can be an effective means of human augmentation in realistic scenarios.**

I. INTRODUCTION

A. Motivation

Face recognition is the process of identifying a target person in a static image or a video stream. The possibility of making this process automatic has been studied at great length in computer vision [1]. Years of research in this field have allowed to develop algorithms with very good performance on face images captured in controlled conditions. However, in dynamic environments (e.g., with changes of lighting) [1] or when only a very limited number of training examples of the target face are available [2], the performance of automatic face recognition systems deteriorates significantly. In contrast to machine vision, humans are generally very good at recognising faces, even if they have seen the target person only once or in the presence of different facial expressions or lighting conditions. Our brain has a complex network of regions dedicated to process face information, the fusiform face area being its computational hub [3].

When we see familiar faces, our brain generates some specific event-related potentials (ERPs), for example the N250, a negative potential occurring about 230 ms after the stimulus onset [4]. The generation of such brain-activity patterns in response to faces of interest has made possible the development of brain-computer interfaces (BCIs) that can improve human performance in face recognition [5], [6], [7]. The accuracy and the speed with which we recognise faces can also be further enhanced with collaborative BCIs (cBCIs) which integrate information from multiple brains [8].

We should note that the promising results of the above-mentioned BCIs were obtained with tightly controlled forms of face recognition. Namely, participants saw a sequence of *individual faces* and had to decide which of them were target faces. However, in a real-world environment we usually deal with pictures or video frames of crowded scenes, possibly taken from different viewpoints, where faces can even be partially occluded, making the task much harder for an individual. This is the situation in which automatic face recognition usually struggles and where, we believe, BCIs have the potential to augment human performance.

In previous research [9] we have described a hybrid EEG-based cBCI for improving group decisions in a task where participants had to decide whether or not a polar bear appeared in a picture of a crowded environment shown for 250 ms. The cBCI used a combination of ERPs, response times (RTs) and eye movements to estimate the confidence of each participant in a decision. Group decisions were then obtained by weighing individual responses (recorded via mouse clicks) according to these confidence estimates. The cBCI-assisted groups were significantly more accurate than individuals and equally-sized traditional groups using standard majority.

This study extends the cBCI work described in [9] to the case of face recognition in real-world environments.

B. Contribution

This study makes two major contributions.

Firstly, we explored the possibility of using a cBCI to improve face recognition. To apply our cBCI to a real-world version of this problem, we used a publicly-available dataset of images designed for person identification under real-world surveillance conditions. In our experiment, observers were asked to search for a specific target face in a picture of a crowded environment representing people walking indoor shown for a very limited time. The high perceptual load (due to the presence of multiple potential targets – see Fig. 1), the fast presentation of each image and the absence of features in the image to be used to simplify the task (e.g., colours) made the task very difficult for individual participants.

Secondly, we asked participants to report their degree of confidence after each decision and we compared it with the confidence estimated by the cBCI. Research has shown that humans often tend to overestimate or underestimate their confidence, ending up in reporting high values of confidence when they made the wrong decision [10], [11]. If this happens, the group decisions made using a decision-system based on self-reported confidence are likely to be badly influenced by error-prone group members, leading to poor performance. However, EEG is also unreliable and

* This work was supported by the Defence and Security National PhD programme through DSTL.

[1]The authors are with the Brain Computer Interfaces and Neural Engineering Laboratory, School of Computer Science and Electronic Engineering, University of Essex, CO4 3SQ Colchester, UK. {dvaler,ccinel,rpoli}@essex.ac.uk

noisy, which might also produce confidence estimates of poor quality. By comparing self-reported and cBCI-based confidence, here we aim at assessing which estimate is more reliable for a realistic face-recognition task.

II. METHODOLOGY

A. Participants

We gathered data from 10 healthy participants (37.8 $\pm$ 4.8 years old, 7 females, all right-handed) with normal or corrected-to-normal vision and no reported history of epilepsy. All volunteers signed a consent form before taking part in the experiment and were paid GBP 16 plus an additional amount up to GBP 4 depending on their performance in the experiment, in order to further motivate them. This research received UK Ministry of Defence and University of Essex ethical approval in July 2014.

B. Stimuli and Procedure

Participants were presented with 6 blocks of 48 trials, for a total of 288 trials. Each block started with the presentation of a display showing the cropped face of the target person assigned to that block. Observers were asked to memorise the face and press the left mouse button when they felt ready to start the presentation of the 48 trials.

As shown in Figure 1, each trial started with the presentation of a fixation cross for 1000 ms. This time allowed participants to prepare for the next stimulus and EEG signals to return to baseline after the previous trial. Then an image of a crowded scene was presented for 300 ms in fullscreen, subtending approximately 14.4 degrees horizontally and 11.0 degrees vertically. After that, a screen showing (again) the image of the cropped target face associated to that block was shown as a reminder to the user. During the presentation of this display, the user had to decide, as quickly as possible, whether or not the target person was present in the scene, by clicking the left or the right mouse buttons, respectively. The mouse was controlled with the preferred hand and RTs were recorded. After indicating their decision, the participants were also asked to indicate the degree of confidence in that decision (0–100%) using the mouse wheel (i.e., scrolling up/down to increase/decrease the confidence by 10%) within a time window of 4 s.

The images used as stimuli have been gathered from the sequences P2E_S5 and P2L_S5 of the ChokePoint video dataset [12], which was designed for person identification under real-world surveillance conditions. The two sequences include 29 people (6 female) walking indoor and passing through two different portals. Three cameras were positioned at the top-left (L), top-center (C) and top-right (R) of each portal, respectively, so that every scene was represented in three pictures taken concurrently from different viewpoints and contained between 2 and 11 people. Since in video sequences consecutive frames contain similar information, we randomly sampled the 700+ images in each sequence to select 48 scenes represented by one image for each viewpoint and shuffled the selected scenes. This procedure allowed to

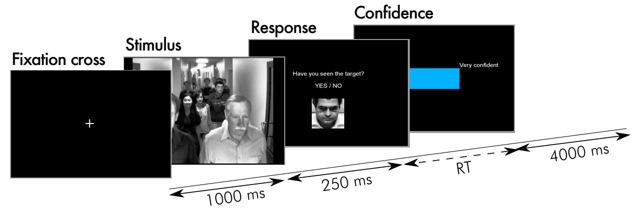

Fig. 1: Sequence of stimuli presented in a trial.

reduce the possibility that participants used previous knowledge to make decisions. Each image has been converted to greyscale and its histogram has been equalised. In each sequence, a different person has been chosen as "target". The images have then been labelled as "target" or "non-target" depending on the presence or absence of the target person, respectively. For each sequence, a total of 36 images (12 per viewpoint) were labelled as "target" and 108 (36 per viewpoint) as "non-target". So target frequency was 25%.

In each block, the images corresponding to a particular sequence (1 or 2) and a particular viewpoint (L, C or R) were used. All possible combinations of sequences and viewpoints have been tested, namely (1, L), (1, C), (1, R), (2, L), (2, C) and (2, R). Each stimulus was used exactly once. The images of each block were presented in the same order for each participant to be able to simulate (offline) concurrent group decisions. However, the blocks were presented in random order for each participant to reduce the impact of learning on the average performance associated to each block.

Participants were comfortably seated at about 80 cm from a LCD screen. The experimental session started with briefing and preparation of the volunteers. Then, two training sessions of 10 trials each were undertaken by the users to familiarise with the task. Preparation and practice took approximately 45 minutes, while the experiment took roughly 25 minutes.

C. Data Acquisition and Feature Extraction

A BioSemi ActiveTwo EEG system has been used to acquire neural data from 64 electrode locations according to the international 10-20 system. Each channel was referenced to the average of the electrodes placed on each earlobe, sampled at 2048 Hz and band-pass filtered between 0.15 and 40 Hz. Artefacts caused by ocular movements were removed by using a standard subtraction algorithm based on correlations. For each trial, stimulus- and response-locked epochs lasting 1900 ms were extracted from the EEG data. The former started 200 ms before the stimulus onset, while the latter started 1200 ms before the participants response. The data of each epoch were then low-pass filtered (cut-off frequency 14 Hz) and downsampled to 32 Hz. The first and last 200 ms of each epoch were then trimmed. Therefore, each trial was represented by 48 samples for each channel (i.e., a total of 3,072 values). Each epoch was then associated to the class "correct" (confident) or "incorrect" (not confident) depending on the correctness of the decision made by the participant in the corresponding trial.

Stimulus- and response-locked epochs have then been transformed using Local Temporal Correlation Common

Spatial Pattern (LTCCSP) [13], which, in previous research [14], we found yielded better performance than traditional CSP for EEG data. Since LTCCSP is a supervised method, we used 10-fold cross-validation to split the dataset in training and test sets, computed the LTCCSP matrix on the training set and used that matrix to transform the data in the test set. The variances of the first and the last columns of the transformed epochs have been used as neural features. Therefore, we have used four LTCCSPs in total, two representing stimulus-locked epochs and two representing the response-locked ones.

The feature vector of each trial was then completed by adding the RT, which is known to correlate with the decision confidence of the participant [15]. RTs were measured by timestamping the click of an ordinary USB mouse.

D. Making Group Decisions

Least Angle Regression (LARS) [16] was used to predict the decision confidence from the neural features and RTs. For each participant p, LARS coefficients were identified from the data in the training set and then used to predict the decision confidence $c_{p,i}$ in the unseen trials i of the test set. Group decisions were then made as follows:

$$d_{group,i} = \sum_{p=1}^{m} (w_{p,i} \cdot d_{p,i}),$$

where m is the group size, $d_{p,i}$ is the decision of group's member p in trial i, and $w_{p,i} = \exp(-2.5 - c_{p,i})$ is the corresponding weight.

To compare the cBCI performance with traditional methods, we also obtained group decisions by using: (a) the majority rule, where $w_{p,i} = 1$ for all i and p, and (b) a weighted majority rule where $w_{p,i}$ was the confidence indicated by participant p after the i-th decision.

III. Results

A. Individual Performance

Figure 2 shows the error rates of the participants in the experiment. Individual performance was quite varied as we did not perform any selection on the participants involved in the experiment. On average, observers made approximately one wrong decision out of four, with some participants (e.g., 8 and 9) performing close to random.

B. Group Performance

We considered all $\binom{10}{n}$ groups of size $n = 2, \ldots, 10$ that could be assembled with our 10 participants. For each group, we computed the errors made by the group using: (1) the standard majority rule, (2) a weighted majority where the confidence values reported by the participants after each decision were used as weights, and (3) a cBCI based on neural features and RTs. We then used the one-tailed Wilcoxon signed-rank test to compare the error distributions within each group size for different methods.

Figure 3 shows the error rates of groups of increasing size when making decisions with different methods. Results indicate that, for group of sizes 2–9, the confidence-based

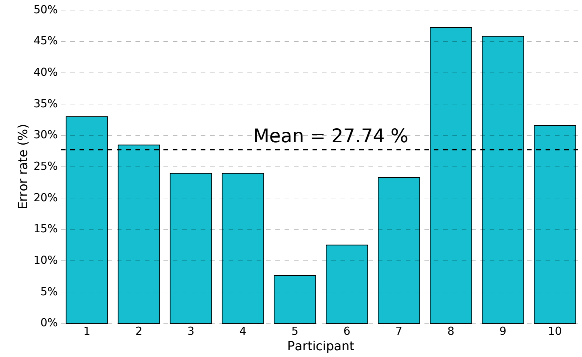

Fig. 2: Error rates for each participant across the experiment.

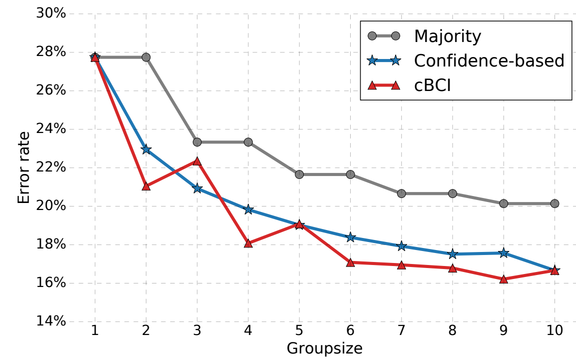

Fig. 3: Error rates obtained by groups of different sizes when using: the standard majority, the self-reported-confidence-based weighted majority, and the cBCI.

methods are significantly superior to the standard majority rule (Wilcoxon $p < 0.003$). Moreover, cBCI-assisted groups achieve significantly better performance than groups making decisions using the self-reported confidence for group sizes 2, 4, 6, 7, 8, 9 ($p < 0.005$). The decision of reported-confidence-based groups were significantly better than the cBCI ones only for group size 3 ($p < 10^{-4}$) and were on par for groups of size 5 ($p = 0.180$). On the one hand, for even-sized groups the cBCI provides a large reduction in error rates due to its ability of breaking ties. On the other hand, ties do not occur in odd-sized groups and, so, to improve performance the cBCI has to allow a minority of users to decide on behalf of the group. For small odd-sized groups this task is quite hard considering the distribution of cBCI weights for the two classes (see next Section). This is why the improvement in performance in these cases is smaller.

C. Confidence Estimates

Figures 4 and 5 compare the distributions of confidence values estimated by participants and the cBCI, respectively, for "correct" and "incorrect" trials. The p-values of the Kruskal-Wallis test comparing the two distributions of each confidence estimate are also reported.

The results show that both confidence estimates have statistically significantly different distributions between the two sets of trials. This indicates that they are both predictors of correctness. However, it should be noted that there is

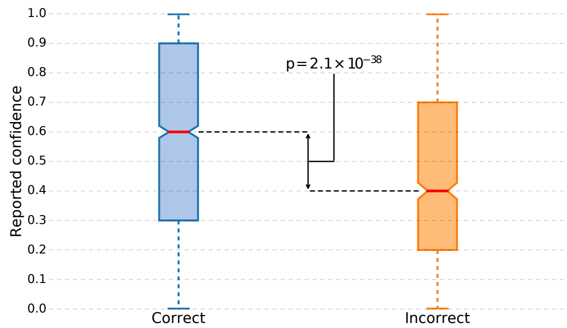

Fig. 4: Distributions of the confidence values indicated by participants after each response for correct and incorrect decisions. The corresponding Kruskal-Wallis p-values comparing such distributions are also reported.

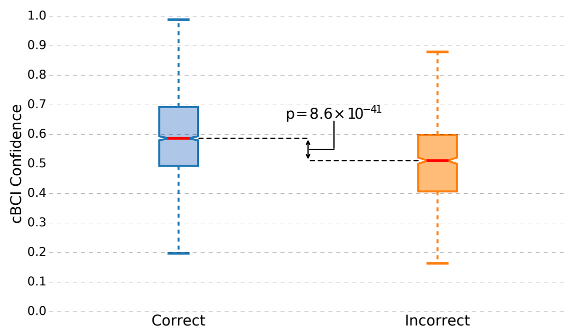

Fig. 5: Distributions of the confidence weights (divided by 34 for plotting purposes) estimated by the cBCI for correct and incorrect decisions. The corresponding Kruskal-Wallis p-values comparing such distributions are also reported.

significant overlap between the box-plots in Figure 4. Also, there is a marked asymmetry in the "incorrect" distribution, caused by the presence of trials with an inverse relation between confidence and correctness (i.e., the overconfident behaviour [10]). Conversely, we see more separation between the distributions in Figure 5, and no significant asymmetry, suggesting that the cBCI's estimates of decision confidence are more robust and accurate predictors of correctness than the self-reported confidence. This, in turn, explains the improvement in group performance achieved by the cBCI and documented in the previous section.

IV. CONCLUSIONS

This paper has presented a hybrid cBCI which uses neural signals and RTs to estimate the decision confidence of isolated observers undertaking an extremely difficult target-face recognition task with real-world stimuli. The cBCI's confidence estimates were used to weigh individual responses and obtain group decisions. Such confidence estimates effectively were obtained by tapping in the unconscious mind of users.

Results show that the cBCI's estimates are better predictors of correctness than the confidence values reported by participants after each decision. Moreover, results with either form of confidence estimation were superior to traditional majority-based group decisions.

These findings suggest that cBCIs may soon be ready for deployment in real-world critical tasks such as face recognition for security and surveillance.

It should be noted that advances in computer vision have allowed the development of systems that can automatically detect and recognise target faces in pictures with acceptable performance without human intervention. In the future we plan to compare the encouraging results obtained in this study with the performance of cutting-edge computer-vision algorithms for automatic face recognition. We also intend to see whether a form of group augmentation based on a hybrid between computer vision technology and cBCI could achieve even better results.

REFERENCES

[1] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face Recognition: A Literature Survey," *ACM Computing Survey*, vol. 35, no. 4, pp. 399–458, 2003.

[2] X. Tan, S. Chen, Z. H. Zhou, and F. Zhang, "Face recognition from a single image per person: A survey," *Pattern Recognition*, vol. 39, no. 9, pp. 1725–1745, 2006.

[3] A. S. Ghuman, N. M. Brunet, Y. Li, R. O. Konecky, J. A. Pyles, S. A. Walls, V. Destefino, W. Wang, and R. M. Richardson, "Dynamic encoding of face information in the human fusiform gyrus," *Nature Communications*, vol. 5, no. 5672, pp. 1–22, 2014.

[4] J. W. Tanaka, T. Curran, A. L. Porterfield, and D. Collins, "Activation of preexisting and acquired face representations: the N250 event-related potential as an index of face familiarity." *Journal of Cognitive Neuroscience*, vol. 18, no. 9, pp. 1488–1497, 2006.

[5] B. Cai, S. Xiao, L. Jiang, Y. Wang, and X. Zheng, "A rapid face recognition BCI system using single-trial ERP," in *6th Annual International IEEE EMBS Conference on Neural Engineering*, 2013, pp. 89–92.

[6] A. Lage-Castellanos, J. I. Nieto, I. Quiñones, and E. Martínez-Montes, "A zero-training algorithm for EEG single-trial classification applied to a face recognition ERP experiment," in *32nd Annual International Conference of the IEEE EMBS*, 2010, pp. 4209–4212.

[7] Y. Zhang, Q. Zhao, J. Jing, X. Wang, and A. Cichocki, "A novel BCI based on ERP components sensitive to configural processing of human faces," *Journal of Neural Engineering*, vol. 9, no. 2, p. 026018, 2012.

[8] L. Jiang, Y. Wang, B. Cai, Y. Wang, W. Chen, and X. Zheng, "Rapid face recognition based on single-trial event-related potential detection over multiple brains," in *7th Annual International IEEE EMBS Conference on Neural Engineering*, 2015, pp. 106–109.

[9] D. Valeriani, R. Poli, and C. Cinel, "A Collaborative Brain-Computer Interface for Improving Group Detection of Visual Targets in Complex Natural Environments," in *7th International IEEE EMBS Neural Engineering Conference*, 2015, pp. 25–28.

[10] J. Navajas, B. Bahrami, and P. E. Latham, "Post-decisional accounts of biases in confidence," *Current Opinion in Behavioral Sciences*, vol. 11, pp. 55–60, 2016.

[11] P. W. Paese and M. A. Feuer, "Decisions, actions, and the appropriateness of confidence in knowledge." *Journal of Behavioral Decision Making*, vol. 4, no. 1, pp. 1–16, 1991.

[12] Y. Wong, S. Chen, S. Mau, C. Sanderson, B. C. Lovell, P. O. Box, and S. Lucia, "Patch-based Probabilistic Image Quality Assessment for Face Selection and Improved Video-based Face Recognition," in *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2011, pp. 74–81.

[13] R. Zhang, P. Xu, T. Liu, Y. Zhang, L. Guo, P. Li, and D. Yao, "Local Temporal Correlation Common Spatial Patterns for Single Trial EEG Classification during Motor Imagery," *Computational and Mathematical Methods in Medicine*, vol. 2013, pp. 1–7, 2013.

[14] D. Valeriani, R. Poli, and C. Cinel, "Enhancement of Group Perception via a Collaborative Brain-Computer Interface," *IEEE Transactions on Biomedical Engineering*, 2016.

[15] R. D. Luce, *Response Times: Their Role in Inferring Elementary Mental Organization*. Oxford University Press, 1986, vol. 8.

[16] B. Efron and T. Hastie, "Least angle regression," *The Annals of Statistics*, vol. 32, no. 2, pp. 407–499, 2004.