AN OCULAR ARTEFACTS CORRECTION METHOD FOR DISCRIMINATIVE EEG ANALYSIS BASED ON LOGISTIC REGRESSION

Xinyang Li, Cuntai Guan*, Kai Keng Ang*, Chuanchu Wang*, Zheng Yang Chin**
Haihong Zhang, Choon Guan Lim† and Tih Shih Lee‡*

* Institute for Infocomm Research, Agency for Science Technology and Research, Singapore
† Department of Child and Adolescent Psychiatry, Institute of Mental Health, Singapore
‡ Duke-National University Singapore Graduate Medical School, Singapore,

ABSTRACT

Electrooculogram (EOG) contamination is a common critical issue in general EEG studies as well as in building high-performance brain computer interfaces (BCI). Existing regression or independent component analysis based artefacts correction methods are usually not applicable when EOG is not available or when there are very few EEG channels. In this paper, we propose a novel ocular artefacts correction method for processing EEG without using dedicated EOG channels. The method constructs estimate of ocular components through artefacts detection in EEG. Then, an optimization based on logistic regression is introduced to remove the components from EEG. Specifically, the optimization ensures that the discriminative information is maintained in the corrected EEG signals. The proposed method is offline evaluated with a large EEG data set containing 68 subjects. Experimental results show that, through the artefacts removal correction by the proposed method, EEG classification accuracy can be improved with statistical significance.

Index Terms— EEG, ocular artefacts correction, brain computer interface

1. INTRODUCTION

The technical field of EEG-based BCIs has seen rapid growth in recent years, with a wide range of promising applications such as motor rehabilitation and cognitive training. One of the major challenges in EEG signal classification/detection in BCI, is EEG's high susceptibility to artefacts contamination by e.g., ocular activities [1]. Therefore, EOG is often recorded along with EEG to capture eye movement information for further artefacts removal.

Eye movement correction procedure (EMCP) based on regression analysis is a commonly used preprocessing method that uses EOG to remove the ocular artefacts from EEG [2,3]. By conducting regression on EEG and EOG, the relation coefficient between EEG channels and EOG channels is calculated to estimate a propagation factor, and subsequently, the EOG portion is subtracted from EEG after scaling by the propagation factor. However, as EOG may also contain brain activity information, such regression based subtraction would cause the loss of relevant EEG signals. Independent component analysis (ICA) proves to be a more preferable method in eliminating the portion of EEG signals that are contaminated by EOG artefacts [3]. ICA approaches the ocular artefacts correction by blind source separation (BSS). Assuming that the observed EEG signal is a mixture of multiple unknown and mutually statistically independent sources, ICA solves the inverse problem and estimates the sources. Then, the source components corresponding to eye movement can be manually identified and removed based on the inverse matrix. The maximum number of independent components estimated by ICA is the number of the EEG and EOG channels. Usually, ICA is applied to data sets recoded from at least 10 EEG channels [4]. In [5], it is found that as few as 35 channels are needed for source estimation in the study of concurrent locomotor and cognitive tasks.

Although the requirement of minimum number of channels may vary in different experiment tasks, the sources separation would be less applicable when only a few EEG channels are available. However, in practical BCI systems, the number of the available channels could be limited for the comfort and the convenience of subjects [6,7]. In such cases, it is difficult to implement EMCP and ICA. To address this problem, here we present a novel ocular artefacts correction method which does not require EOG data and is applicable to single-channel EEG data. Besides the requirements of EEG and EOG, the other major challenge for artefacts removal lies in that the preprocessing for artefacts removal may cause the loss of the class information in EEG signals, and, subsequently, the drop of the classification accuracy. To ensure that the artefacts correction does not degenerate the discriminating power of the data for EEG classification in BCI, our computational model used uses a supervised approach based on a logistic regression.

This paper is organized as follows. In Section 2, the oc-

This project was supported by a grant from NMRC, Singapore.

ular correction based on logistic regression is presented. In Section 3, the validity of the proposed method is verified by experimental study on EEG based attention detection. Concluding remarks are given in Section 4.

2. OCULAR ARTEFACTS CORRECTION

2.1. Ocular Artefacts Detection

To conduct the ocular artefacts correction without EOG recording, the artefacts need to be extracted using raw EEG data $X_0(t) \in \mathbb{R}^{n_c \times n_t}$, where n_c is the number of channels and n_t is the number of time samples. Given the analysis of the morphology characteristic of the potentials induced by eye movements in [2, 8], the moving average filter is applied to the raw EEG data to obtain the smoothed signal $x_s(t)$ for further potential extraction, as below

$$x_s(t) = \frac{1}{m} \sum_{j=-\frac{m}{2}}^{j=\frac{m}{2}} x_0(t+j) \tag{1}$$

where m is the number of the neighboring points used in the moving average filter, and $x_0(t) \in \mathbb{R}^{n_t}$ is the EEG signal from one arbitrary channel, or in other words, one arbitrary row of $X_0(t)$. Calculate the relative height of the peaks as

$$h(t) = \max(|x(t) - x(t-1)|, |x(t+1) - x(t)|) \tag{2}$$

Define the peak height range parameter h_r as

$$h_r = [h_b, h_u] \tag{3}$$

Then, find the set $\mathcal{P}_t$ containing time indexes of those peaks with height in the range h_r as

$$\mathcal{P}_t = \{t : \frac{m}{2} < t < n_t - \frac{m}{2} \text{ and } h_b < h(t) < h_u\} \tag{4}$$

For each element $t^i \in \mathcal{P}_t$, $i = 1, 2, ..., |\mathcal{P}_t|$, let z_b^i and z_a^i be the nearest zero points before and after t^i. Then, the artefacts signal $x_a(t)$ is obtained as

$$x_a(t) = \begin{cases} x_s(t), & z_b^i < t < z_a^i\ ; \\ 0, & \text{else.} \end{cases} \tag{5}$$

Examples of constructing $x_a(t)$ from $x_s(t)$ with different h_r are illustrated in Figure 1. As shown by subfigures (a) and (b), $x_a(t)$ are zeros except those points belonging to peaks with heights within a certain range. In this way, EEG data that are not contaminated by the ocular artefacts could be kept as intact as possible in the artefacts removal procedure.

Moreover, with different height ranges $h_r^j = [h_b^j, h_u^j]^T, j = 1, 2, ..., n_h$, $x_a^j(t)$ can be extracted correspondingly, where n_h is the number of peak height ranges. Define $X_a(t)$ as the matrix containing all artefacts signals $x_a^j(t)$, as follows

$$X_a(t) = \begin{bmatrix} x_a^1(t) \\ \vdots \\ x_a^{n_h}(t) \end{bmatrix} \tag{6}$$

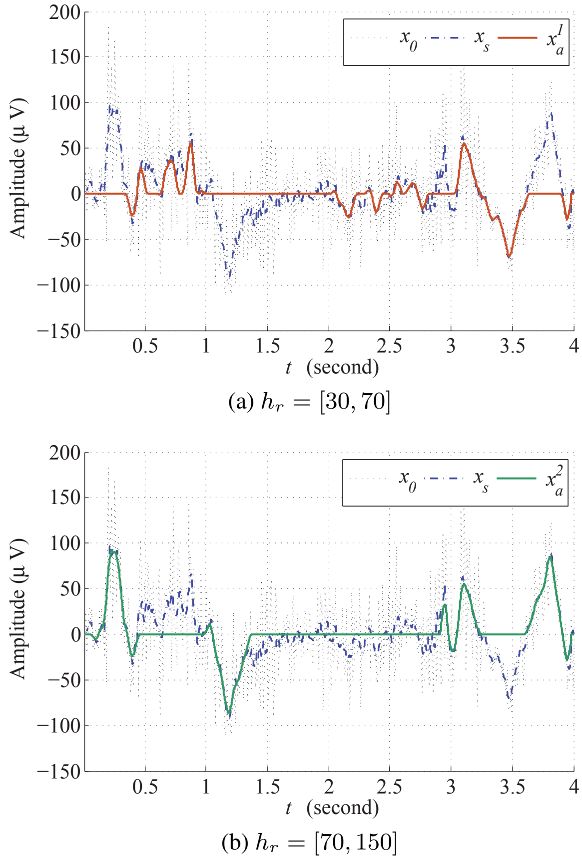


Fig. 1. Examples of artefacts signal construction.

$X_a(t)$ in (6) can be regarded as the pseudo-EOG signal. As mentioned above, using $X_a^j(t)$ for artefacts correction is more advantageous than using the real EOG signal as it is zero at most of the time points, which would cause less information loss in the artefacts removal procedure. Besides, in conventional EMCP, one propagation factor is estimated for one EEG-EOG pair. Different components $x_a^j(t)$ can be regarded as artefacts corresponding to different eye movements. By estimating different filtering parameters for them, it is more flexible to maintain the discriminative information in EEG signals after the artefacts correction. In the following section, we will introduce the discriminative learning of the filtering model for artefacts removal.

2.2. Ocular Artefacts Removal

Let $x_c(t)$ be the signal after the artefacts correction, as follows

$$x_c(t) = x_0(t) - \theta X_a(t) \tag{7}$$

where $\theta \in \mathbb{R}^{n_h}$ is the filtering parameter, which is similar to the propagation factor in the conventional EMCP, representing the portion of the artefacts in EEG to be subtracted. To ensure that the artefacts correction could benefit the classification in BCI, θ is learnt in a discriminative manner, which is different from estimating the regressive coefficients used in EMCP. As the power of the EEG signal is one of the most common features in BCI, the power of $x_c(t)$ plus a bias parameter b is used to construct the latent variable z [9–11], as follows

$$\begin{aligned} z &= x_c x_c^T + b \\ &= \theta^T R_a \theta - 2\theta^T r_a + r_0 + b \end{aligned} \tag{8}$$

where $R_a \in \mathbb{R}^{n_h \times n_h}$ is the covariance matrix of $X_a(t)$, i.e.,

$$R_a = X_a X_a^T \tag{9}$$

$r_a \in \mathbb{R}^{n_h}$ is the vector containing the correlation between $x_a(t)$ and $x_0(t)$, i.e.,

$$r_a = X_a x_0^T \tag{10}$$

and r_0 is the variance of $x_0(t)$. Thus, the function used to predict labels based on the logistic function is

$$h(z) = \frac{1}{1+e^{-z}} \tag{11}$$

Therefore, the final objective function is

$$\begin{aligned} J(\theta, b) = \frac{1}{n_j} \sum_{j=1}^{m} [-y^i \log(h^i(z)) - \\ (1-y^i)\log(1-h^i(z))] \end{aligned} \tag{12}$$

where y^i and $h^i(z)$ are the true label and the predicted label for trial i, respectively, and n_j is the number of trials for model training. The filtering parameter $\hat{\theta}$ could be obtained by minimizing $J(\theta, b)$.

By applying the ocular artefacts detection and removal for each EEG channel $x_0^k(t)$, we could obtain the corrected signal for each channel, as follows

$$x_c^k(t) = x_0(t) - (\hat{\theta}^k)^T X_a^k(t) \tag{13}$$

where $\hat{\theta}^k$ and $X_a^k(t)$ are the filtering parameter and artefacts signal for channel k. Let $X_c(t)$ be the multi-channel data after the artefacts correction, i.e.,

$$X_c(t) = \begin{bmatrix} x_c^1(t) \\ \vdots \\ x_c^{n_c}(t) \end{bmatrix} \tag{14}$$

Then, a proper feature extraction method could be applied to $X_c(t)$ followed by the classification.

3. EXPERIMENTAL STUDY

3.1. Experiment Setup and Data Processing

EEG data were recorded using a Neurosky dry EEG headband with one bipolar channel, which was positioned at the frontal site Fp1. The sampling rate was 256 Hz. 68 subjects participated the experiment, and for each subject, 3 sessions of Color Stroop test were recorded [12]. The second and third sessions were recorded 20-30 weeks after the first session was recorded. In each session, there were 40 Stoop trials, during which the subject was assumed to concentrate on the test. Each Stroop trial was followed by a rest period when the subject could relax. Each Stroop trial lasted around 10s while the rest period between 2 Stroop trials was around 15s.

To increase the number of trials, a 4s window with a window-shift is applied to segment EEG data recorded during Stoop trials, which yields data of the attention class. The same segmentation is also applied to EEG recorded during rest periods, which yields data of the idle class. In this way, the final data set is balanced between the two classes, i.e., attention class and idle class. Moreover, the first half of the original Stroop trials with rest period were truncated into training trials and the second half test trials. For each subject, the number of total trials was around 240.

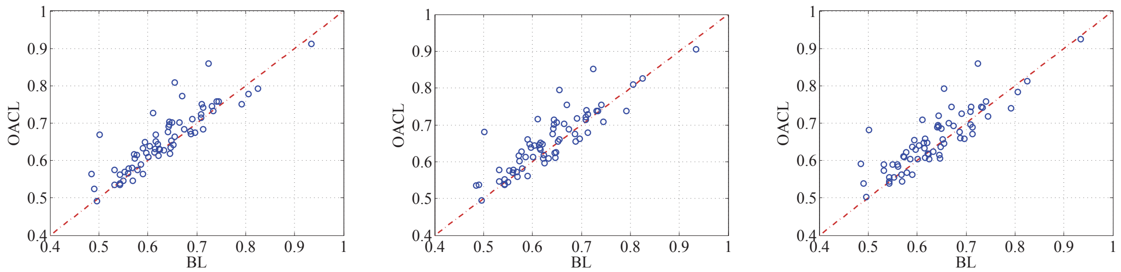
In this study, raw EEG data are smoothed with $m = 10$ in (1), and 2 spike height ranges are used, i.e., $n_h = 2$, yielding $X_a(t) \in \mathbb{R}^{2\times n_t}$. h_r^1 and h_r^2 are set manually by examining the morphology of spikes in the training set. After the ocular artefacts correction, a filterbank containing 9 frequency bands (2-6Hz, 6-10Hz, ..., 34-38Hz) is applied to calculate the bandpower of $X_c(t)$, and, subsequently, mutual information is applied to select 4 from the 9 bandpower features [13]. Selected features are classified into the attention class or the idle class by the linear discriminant analysis (LDA) classifier.

3.2. Results

Table 1 summarizes the classification results of the proposed ocular artefacts correction method based on logistic regression (OACL) compared with the baseline method for which no artefacts correction preprocessing is applied. As shown by Table 1, for all three sessions the proposed method improves both the median and average classification accuracies, the significance of which is validated by Wilcoxon signed-rank test with ***p***-values all below 0.05. Since OACL yields significant improvements in all three parameter settings, the proposed method is robust against h_r to a certain extend. Figure 2 shows the comparison of the classification results in a more observable way, where each dot represents an average result of one subject across three sessions. As x-axis represents the accuracies with baseline (BL) and y-axis represents OACL, it is shown that the proposed method could achieve improvements for most of the subjects with more dots above the line $y = x$.

Table 1. Classification results (%)

h_r^1, h_r^2	Baseline			OACL								
				[30, 70], [70, 150]			[30, 80], [80, 150]			[30, 90], [90, 150]		
Session	1	2	3	1	2	3	1	2	3	1	2	3
mean	64.95	63.88	61.65	67.10	65.66	64.08	66.82	65.48	64.12	66.69	65.61	64.22
median	62.23	61.53	60.00	66.40	64.38	62.39	64.74	64.59	61.72	64.01	64.88	61.57
std	12.80	11.40	11.69	11.85	11.33	11.64	11.81	10.99	11.58	11.79	11.18	11.01
***p*-value**	-	-	-	**0.041**	**0.009**	**0.001**	**0.037**	**0.029**	**0.001**	**0.043**	**0.035**	**0.001**



(a) (b) (c)

Fig. 2. Classification accuracy comparison. Subfigures (a), (b), (c) correspond to OACL with $h_r^1 = [30, 70], h_r^2 = [70, 150]$, $h_r^1 = [30, 80], h_r^2 = [80, 150]$ and $h_r^1 = [30, 90], h_r^2 = [90, 150]$, respectively.

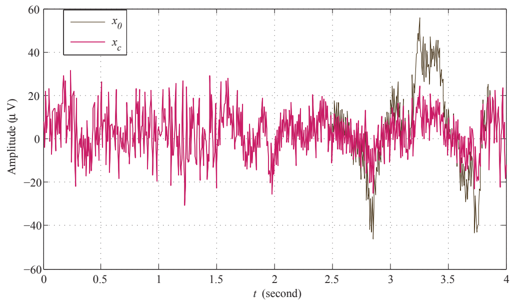


Fig. 3. Raw EEG signal and EEG signal corrected by OACL.

In Figure 3, one example of the comparison between EEG signal and EEG signal corrected by OACL is illustrated, where it can be found that the fluctuation around 3.5 second have been modulated.

3.3. Discussion

In conventional EMCP, the propagation factor represents the portion of the EOG in the EEG to be subtracted, which means that the propagation factor is always a positive value between 0 and 1. Representing a similar portion parameter, $\hat{\theta}$ in (7) is expected to be between 0 and 1 too, although such a constraint is not implemented in the objective function. Surprisingly, $\hat{\theta}$ is negative for some subjects upon the optimization, which means that the amplitudes of certain peaks have been increased.

To find out the effects of the negative $\hat{\theta}$, we have added the regularization in the objective function (12), as follows

$$J_r(\theta, b) = (1 - \lambda)J(\theta, b) + \lambda(\tfrac{1}{2n_h} \sum(\theta_j^2 - \theta_j)) \quad (15)$$

We have tested $\lambda = [0.005, 0.01, 0.1]$, which yield most of $\hat{\theta}$ between 0 and 1, and, however, less improvement in classification accuracies.

For further investigation, the numbers of peaks in the two height ranges h_r^1 and h_r^2 are calculated for each class. The distributions of the number of subjects over the number of the peaks for the two classes have been compared, with the histograms illustrated in Figure 4, where y-axis represents the number of subjects and x-axis represents the number of peaks with height in the range h_r^1 or h_r^2. By comparing the attention class (class A) and the idle class (class I), we can find that there are much more subjects with more peaks for class I than class A, and differences between the two classes are proved to be significant in all 6 cases by Wilcoxon signed-rank test. Therefore, it is possible that differences of the ocular movements between the two classes are so significant that the proposed method captures such differences rather than the mental state differences between the two classes.

4. CONCLUSIONS

This study investigates the ocular artefacts correction for B-CI experiments without EOG or muti-channel EEG, where EMCP and ICA are usually not applicable. In the proposed

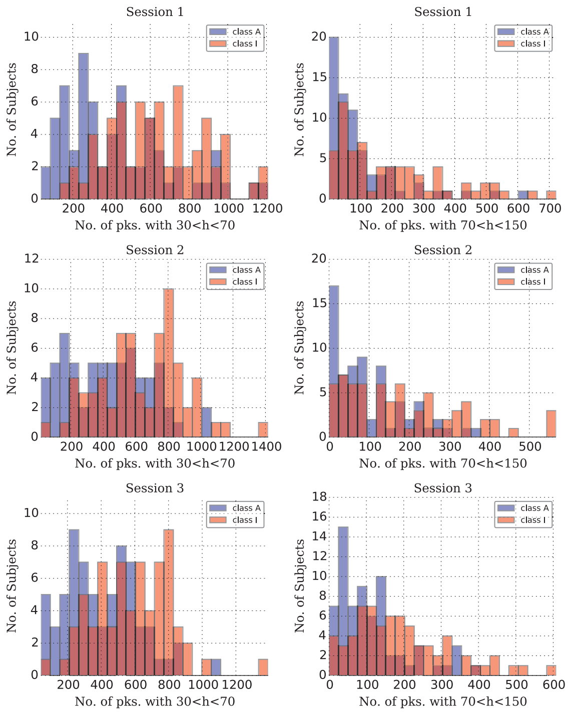


Fig. 4. Comparison of peak number histograms between the two classes.

method, multiple artefacts signals corresponding to different eye movements are constructed using EEG data. In this way, different filtering parameters could be estimated for different artefacts signals, which is more flexible than estimating the propagation factor for EOG. Moreover, in the proposed method, the filtering parameters are optimized using logistic regression to ensure that the artefacts correction benefits the classification. The proposed method is evaluated using a two-class single-channel EEG data set containing 68 subjects, and the significance of the improvements is validated by statistical tests. In our future work, we will develop an method to automatically optimize the parameters and investigate the extraction of metal states that are independent of ocular movements to have a better understanding of the attention condition.

5. REFERENCES

[1] R. Sameni and C. Gouy-Pailler, "An iterative subspace denoising algorithm for removing electroencephalogram ocular artifacts," *Journal of Neuroscience Methods*, vol. 225, no. 0, pp. 97 – 105, 2014.

[2] G. Gratton, M. G. Coles, and E. Donchin, "A new method for off-line removal of ocular artifact," *Electroencephalogr Clin Neurophysiol*, vol. 55, pp. 468484, 1983.

[3] A. Hoffmann and M. Falkenstein, "The correction of eye blink artefacts in the EEG: A comparison of two prominent methods," *PLoS ONE*, vol. 3, no. 8, pp. e3004, 2008.

[4] T.-P. Jung, S. Makeig, C. Humphries, T. W. Lee, M. J. McKeown, V. Iragui, and T. J. Sejnowski, "Removing electroencephalographic artifacts by blind source separation," *Psychophysiology*, vol. 37, no. 2, pp. 163–178, 2000.

[5] T. M. Lau, J. T. Gwin, and D. P. Ferris, "How many electrodes are really needed for EEG-based mobile brain imaging?," *Journal of Behavioral and Brain Science*, vol. 2, no. 3, pp. 387–393, 2012.

[6] C. G. Lim, T. S. Lee, C. Guan, D. S. Sheng Fung, Y. B. Cheung, S. S. Teng, H. Zhang, and K. R. Krishnan, "Effectiveness of a brain-computer interface based programme for the treatment of adhd: a pilot study," *Psychological Bulletin*, vol. 43, no. 1, pp. 73–82, 2010.

[7] C. G. Lim, T. S. Lee, C. Guan, D. S. Sheng Fung, Y. Zhao, S. S. Teng, H. Zhang, and K. R. Krishnan, "A brain-computer interface based attention training program for treating attention deficit hyperactivity disorder," *PLoS ONE*, vol. 7, no. 10, pp. e46692, 2012.

[8] A. S. Keren, S. Yuval-Greenberg, and L. Y. Deouell, "Saccadic spike potentials in gamma-band eeg: Characterization, detection and suppression," *NeuroImage*, vol. 49, no. 3, pp. 2248 – 2263, 2010.

[9] H. Zhang, Z. Y. Chin, K. K. Ang, C. Guan, and C. Wang, "Optimum spatio-spectral filtering network for brain-computer interface," *IEEE Transactions on Neural Networks*, vol. 22, no. 1, pp. 52–63, Jan 2011.

[10] H. Zhang, H. Yang, and C. Guan, "Bayesian learning for spatial filtering in an EEG-based brain-computer interface," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 24, no. 7, pp. 1049–1060, 2013.

[11] B. Hamadicharef, H. Zhang, C. Guan, C. Wang, K. S. Phua, K. P. Tee, and K. K. Ang, "Learning eeg-based spectral-spatial patterns for attention level measurement," *2009 IEEE International Symposium on Circuits and Systems (ISCAS2009)*, pp. 1465–1468, May 2009.

[12] C. M. MacLeod, "Half a century of research on the stroop effect: an integrative review," *Psychological Bulletin*, vol. 109, pp. 163–203, 1991.

[13] K. K. Ang, Z. Y. Chin, H. Zhang, and C. Guan, "Mutual information-based selection of optimal spatial-temporal patterns for single-trial EEG-based BCIs," *Pattern Recognition*, vol. 45, no. 6, pp. 2137–2144, 2012.