An unsupervised eye blink artifact detection method for real-time electroencephalogram processing

2016 Physiol. Meas. 37 401

(http://iopscience.iop.org/0967-3334/37/3/401)

IOP Publishing | Institute of Physics and Engineering in Medicine Physiological Measurement

Physiol. Meas. **37** (2016) 401–417 doi:10.1088/0967-3334/37/3/401

An unsupervised eye blink artifact detection method for real-time electroencephalogram processing

Won-Du Chang, Jeong-Hwan Lim and Chang-Hwan Im

Department of Biomedical Engineering, Hanyang University, Seoul, Korea

E-mail: ich@hanyang.ac.kr

Received 13 August 2015, revised 27 December 2015
Accepted for publication 14 January 2016
Published 19 February 2016

Abstract

Electroencephalogram (EEG) is easily contaminated by unwanted physiological artifacts, among which electrooculogram (EOG) artifacts due to eye blinking are known to be most dominant. The eye blink artifacts are reported to affect theta and alpha rhythms of frontal EEG signals, and hard to be accurately detected in an unsupervised way due to large individual variability. In this study, we propose a new method for detecting eye blink artifacts automatically in real time without using any labeled training data. The proposed method combined our previous method for detecting eye blink artifacts based on digital filters with an automatic thresholding algorithm. The proposed method was evaluated using EEG data acquired from 24 participants. Two conventional algorithms were implemented and their performances were compared with that of the proposed method. The main contributions of this study are (1) confirming that individual thresholding is necessary for artifact detection, (2) proposing a novel algorithm structure to detect blink artifacts in a real-time environment without any *a priori* knowledge, and (3) demonstrating that the length of training data can be minimized through the use of a real-time adaption procedure.

Keywords: electroencephalogram (EEG), electrooculogram (EOG), ocular artifact, eye blink

S Online supplementary data available from stacks.iop.org/PM/37/401/mmedia

(Some figures may appear in colour only in the online journal)

1. Introduction

Scalp electroencephalogram (EEG) is easily contaminated by other physiological signal sources (Hagemann and Naumann 2001, Durka *et al* 2003, Lawhern *et al* 2013) because there are many signal sources near the brain, such as heart, teeth, facial muscles, and eyes (Croft and Barry 2000, Jiang *et al* 2007, Yong *et al* 2008). The signals from other non-brain sources, known as artifacts, are mixed with the brain signal and interfere with the accurate analysis of EEG. For this reason, most EEG studies adopt a preprocessing procedure for obtaining artifact-free EEG signals (Jung *et al* 2000, Hagemann and Naumann 2001, Mognon *et al* 2010).

Eye blinks are one of the most influential sources of artifacts contaminating electroencephalogram (EEG) signals (Hagemann and Naumann 2001). In most EEG paradigms using visual stimuli or visual feedbacks, EEGs are inevitably contaminated by eye blink artifacts because spontaneous blinking is an autonomic process of the human body (Pult *et al* 2013). Although theta band is distorted most severely when an EEG signal is contaminated by eye blink artifacts, eye blink artifacts have also been reported to significantly affect alpha and beta band EEG signals (Hagemann and Naumann 2001). Therefore, cleaning EEG signals contaminated by eye blink artifacts is an important preprocessing procedure to obtain accurate EEG analysis results.

The general approach for the EEG data cleaning is either rejecting the artifact-contaminated time windows (contaminated epochs) or recovering the artifact-free data (Croft and Barry 2000), where both approaches commonly require artifact detection process. Although detecting contaminated epochs is obviously indispensable for the rejection approach (Krishnaveni *et al* 2006, Kook *et al* 2008), the recovery approach also utilizes the artifact detection procedure in order to minimize distortion of source data by applying the recovery procedure only to the detected regions (Krishnaveni *et al* 2006) or applying different correction factors for eye blink artifacts and other muscle movement artifacts (Gratton *et al* 1983). Detection of artifacts is also helpful for independent component analysis (ICA) of EEG signals as the artifact-contaminated independent components need to be detected for the automated artifact removal (Shao *et al* 2009, Mognon *et al* 2010).

There have been many previous studies on the automatic detection of eye blink artifacts, which can be roughly classified into two main categories: feature-based approaches and distance-based approaches. Feature-based approaches determine the presence of artifacts in an epoch using features extracted after certain preprocessing procedures such as band-pass filtering (Hoffmann and Falkenstein 2008, Klein and Skrandies 2013) and component separation (Mognon *et al* 2010). Various features have been introduced, including maximum absolute value (Nolan *et al* 2010), kurtosis (Barbati *et al* 2004, Mognon *et al* 2010), entropy (Barbati *et al* 2004), second-order difference (Klein and Skrandies 2013), and the Teager-Kaiser energy operator (Breuer *et al* 2014).

Distance-based approaches use a template representing the generic shape of the target artifact. After the template is built by an expert (Aarabi *et al* 2009) or generated by averaging sample artifact signals (Kim and McNames 2007), the distance (or similarity, which is the inverse of distance) between the template and a part of the signal is calculated. A variety of methods such as support vector machine (SVM; (Shao *et al* 2009), dynamic time warping (DTW; (Chang and Im 2014), cross-correlation (Li *et al* 2006, Kim and McNames 2007), and correlation between EEGs in different channels (Durka *et al* 2003) have been used to calculate the distance.

In general, both approaches to detect artifacts need to set a certain threshold value to make binary decisions on whether a specific epoch is contaminated by artifacts or not. The simplest way is to use a constant threshold value determined empirically from previous analyses.

However, applying the same threshold value can result in significant detection errors in some individuals' EEGs due to the large individual variance in the shapes and amplitudes of the eye blink artifacts (Galley *et al* 2003). Therefore, the threshold values need to be customized for each individual's EEG. A method that uses the variance in artifact distribution was proposed by (Krishnaveni *et al* 2006); however, their method includes expert intervention for calculating the variance in the artifact distribution. In their method, a training procedure requires an additional dataset, in which all the artifacts should be manually labeled by an expert (Please note that most machine learning algorithms have the same requirements because the most algorithms train their models in supervised ways). Mullen *et al* (2015) suggested a semi-supervised method to determine a threshold from distribution of artifact-free data. We categorized this method as a semi-supervised approach, because this method does not require blink-labeled EEG data but does require labeled artifact-free EEG data to estimate the distribution of normal data. This kind of supervised algorithms can hardly be applied to many end-user applications often requiring real-time EEG processing.

To circumvent the limitation of the supervised approach, some studies proposed unsupervised artifact detection methods, which adopted fully automatic training procedure and thus do not require expert labeling. Mongon *et al* (2010) adjusted the threshold value using the expectation-maximization (EM) algorithm; however, their method was not fit for the real-time processing because of the time-consuming procedure of maximization. Geetha and Geethalakshmi (2012) utilized Otsu's thresholding technique (Otsu 1975) originally proposed for image binarization, and Breuer *et al* (2014) used 80th percentile of the individual data distribution as the threshold value, but both methods were still based on the empirical parameterization.

The present study proposes a new approach for the real-time detection of eye blink artifacts in an unsupervised manner. The proposed approach combines our previous method to detect eye blink artifact from single-channel EEG using the first derivative sum in multiple sliding windows (Chang *et al* 2015) with an automatic thresholding method introduced by (Kim and McNames 2007). Both methods were modified to operate in real-time as they were originally developed for off-line spike signal detection. The performance of the proposed approach is evaluated in both simulated-real-time and real-time environments. In the validation process, we tried to demonstrate the followings: (1) whether the use of individual threshold enhances the accuracy of detecting eye blink artifacts in comparison with the common threshold; (2) how well the real-time adaptation algorithm is working; (3) how we can balance false-positives and false-negatives in real-time thresholding.

2. Methods

2.1. Experimental data

Experiments were conducted in a simulated-real-time environment using an EEG dataset acquired from 24 healthy participants (Chang and Im 2014). In order to simulate the real-time data acquisition, each data sample from the pre-saved EEG dataset was read one by one with a sampling frequency identical to that of real-time recording and then immediately transferred to an artifact detection program, which means that the developed artifact detection program could be directly applied to real-time artifact detection without any modification. The original dataset was recorded using a multi-channel EEG recording system (ActiveTwo AD-box™, BioSemi, Netherlands) at a sampling rate of 2,048 Hz (afterwards the data were down-sampled to 64 Hz for the eye blink detection), while the study participants performed spot-the-difference puzzles (participants were asked to find the differences between two

images shown on the left and right of a monitor). The ground-truth of eye blinking ranges was marked manually by visual inspections of frontal channel EEG (Fp2 in the international 10–20 system) and vertical electrooculogram (EOG). We used the first 290 s EEG data for the analysis. The study participants signed a consent form and received monetary reimbursement for their participation; the study was approved by the institutional review board (IRB) of Hanyang University.

2.2. Artifact detection using the first derivative sum in multiple sliding windows

This section briefly introduces an artifact detection method proposed by the authors (Chang *et al* 2015), which adopted a simple digital filter to accurately extract the artifact ranges. The filter can be written as follows:

$$F(t) = S(t) - S(t-|W|), \tag{1}$$

where $|W|$ is the width of a sliding window and $S(t)$ is the t-th sample of the original signal. The filter was named the '*s*ummation of first *d*erivatives in a sliding *w*indow (SDW)' because the equation (1) was derived from the following equation:

$$F(t) = \sum_{k=t-n+1}^{t} S(k) - S(k-1). \tag{2}$$

As readily recognized from the above equation, the SDW filter emphasizes spikes in a specific wave-width. To emphasize spikes with various wave-widths, SDWs in multiple window sizes should be considered since the durations of eye blink artifacts vary (Verleger 1993, Fukuda and Stern 2005). Hence, an empirical procedure was introduced for selecting a value among SDWs of different window sizes as a representative at time t. This procedure was called MSDW (Multiple-window SDW) as it utilizes multiple sliding windows. For every t, the following selection steps are performed:

(1) Calculate SDWs with different $|W|$'s, considering a typical eye blink range. Let $F_{|W|}$ denote an SDW with a window size of $|W|$.
(2) Choose the maximum $F_{|W|}$ at time t, denoted as MSDW(t), as a representative if it satisfies the conditions below.
    a. The numbers of local minima and maxima are the same within the range of $[t,\ t-|W|]$.
    b. All of the first derivatives from time t to $t-|W|+1$ should be within $S'(t-|W|+1)$ and $S'(t)$, where $S'(t)$ represents the first derivatives at time t.

It is possible that there exist multiple $|W|$'s that maximize $F_{|W|}$. In such a case, the smallest $|W|$ is selected for MSDW(t). When there are no SDWs that satisfy all of the conditions, the minimum window size is used.

Let us denote the window size selected through the above procedure at time t as $|W_{\mathrm{MSDW}(t)}|$. Then, the artifact range is defined as

$$R = \left[T(\mathrm{Max}_{i-j}) - \left|W_{\mathrm{MSDW}(T(\mathrm{Max}_{i-j}))}\right|,\ \ T(\mathrm{Min}_i) \right], \tag{3}$$

where $\mathrm{Max}_{i-j} - \mathrm{Min}_i > \theta$, Min_i and Max_i are the ith local minimum and maximum MSDW values, respectively, and $T(\mathrm{Max}_i)$ and $T(\mathrm{Min}_i)$ are the time positions of the ith local maximum and minimum, respectively. The integer value j is chosen among integer values equal to or larger than zero in order for $\mathrm{Max}_{i-j} - \mathrm{Min}_i$ to be maximized and in order for the time

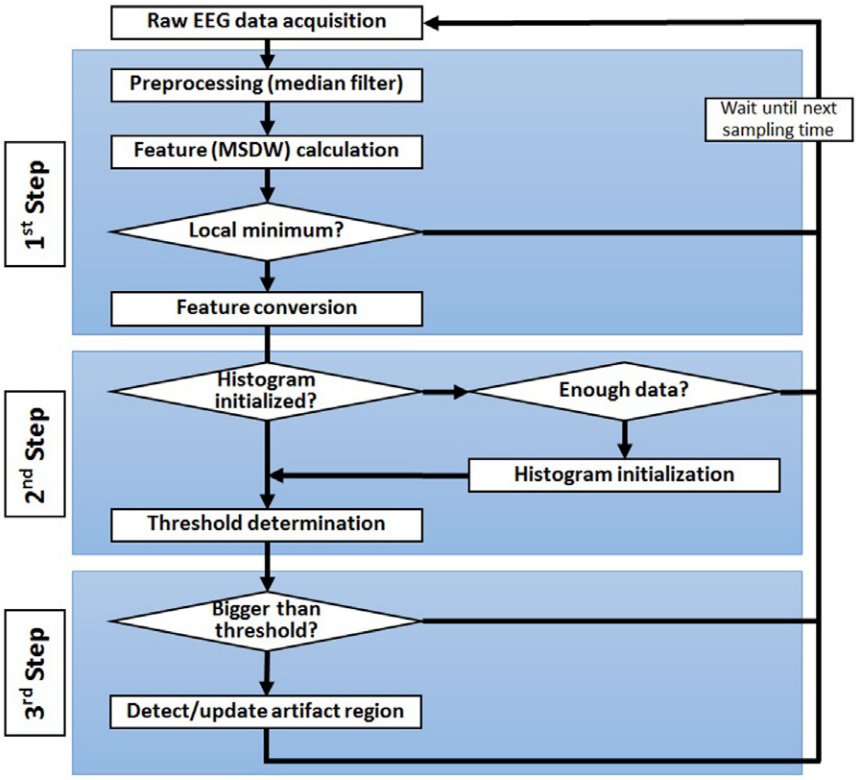

Figure 1. The overall architecture of the proposed system.

difference between Max_{i-j} and Min_i not to exceed the maximum window size of MSDW. We assume that the first local maximum precedes the first local minimum in an MSDW series: any local minimum before the first maximum is ignored when determining the order.

2.3. Proposed method for unsupervised artifact detection

This article proposes a novel method to detect eye blink artifacts in an unsupervised way. The proposed method is composed of three main steps after EEG data acquisition. Figure 1 describes the overall architecture of the algorithm. The first main step is to calculate a feature for assessing the presence of artifacts. The MSDW method described in the previous section was utilized for this step because this method has several advantages over the conventional eye blink artifact detection methods in terms of usability in real-time applications. First, it does not require any expert actions such as template construction except determining the individual threshold values. Second, it can detect artifact ranges as well as the presence of the artifact, enhancing the overall performance of artifact detection. Third, it is computationally light and can be applied to single frontal-channel EEG data without the need for EOG reference. Therefore, it is particularly useful in brain-computer interface applications using wearable EEG devices with a few frontal EEG channels.

As is illustrated in figure 1, an MSDW value is calculated using newly acquired data after preprocessing the data with a median filter (a data buffer with the same size as the median

filter is necessary for this purpose). The original procedure for artifact detection is separated into parts (the 1st step and the 3rd step) in order to insert an automatic thresholding procedure (the 2nd step). The MSDW value is converted for automatic thresholding procedure because the original MSDW value does not directly represent dissimilarity or similarity. Hence, we converted the MSDW values into new values using the difference between local minima and their adjacent maxima. If the latest MSDW value is a local minimum, the largest value among the recent maxima was chosen to calculate the difference value. The converted MSDW feature was defined as

$$f(t) = \mathrm{Max}_{i-j} - \mathrm{Min}_i, \tag{4}$$

where all of the notations are the same as in (3).

The second main step is to automatically determine the threshold from the given feature (converted MSDW feature) series. Since storing all the feature series might be inadequate for a long recording or might require unnecessarily large storage size, a density histogram accumulated over time was utilized in order to investigate the feature population. The histogram could be initialized with some given data and readily updated by increasing a value at a specific bin. To calculate the histogram, the width of the histogram bin was calculated using the following equation:

$$\mathrm{hist\_width} = (M - m)/(n\mathrm{Bin} - 2), \tag{5}$$

where M is an expected maximum value of the entire distribution, m is the minimum value of the distribution, and nBin is the number of bins. In this experiment, the number of bins was fixed to 20, and six-standard deviation of the acquired data was assumed to M. Based on this histogram, the threshold could be automatically determined with conventional thresholding algorithms. We implemented three conventional algorithms and modified one of them considering our conditions. The details of the thresholding algorithms will be introduced in the next sections.

The last main step is to detect and update the artifact region using the determined threshold. When the current MSDW value is a local minimum at time t and the dissimilarity feature, $f(t)$, is larger than the threshold, the region of the artifact is determined as

$$R = \left[T(\mathrm{Max}) - \left|W_{\mathrm{Max}}\right|, \quad t \right], \tag{6}$$

where Max is the local maximum to calculate $f(t)$, $T(\mathrm{Max})$ is its time points, and $\left|W_{\mathrm{Max}}\right|$ is the window size of MSDW for calculating Max. Essentially, the determined region is readily overlapped because the detection process is conducted for every local minimum. To avoid multiple detection of the same artifact, overlapped ranges were merged into a single range.

2.4. Automatic thresholding using a fixed portion of data

Breuer *et al* (2014) introduced a method using a percentage of an individual data (feature) distribution as the threshold. This can be calculated by finding m satisfying the following equation:

$$\int_0^m F(k) \cdot \mathrm{d}k = P \cdot N, \tag{7}$$

where $F(k)$ is a density function, P is the portion of normal data, and N is the total number of data points in the distribution.

2.5. Automatic thresholding by analyzing the data distribution of offline detections

This section briefly describes an algorithm proposed by (Kim and McNames 2007), hereafter referred to as KM. The algorithm was proposed to detect spikes in extracellular neural recordings in offline environment. KM utilized the kernel density function (probability density function: PDF) of local maxima to detect a spike. The PDF distribution was separated into two modes, primary and secondary modes, where the secondary mode was assumed to contain the spikes. The procedure of the automatic thresholding algorithm is as follows:

(1) Calculate the series of similarity (using cross-correlation in the article) and calculate PDF using the local maxima of the series. $F(k)$ denotes the density at a specific similarity k.
(2) Assume that the largest mode in PDF represents normal signal and the rightmost mode represents the spikes. If there exist only a single mode, the threshold is determined according to the predefined signal-to-noise ratio (SNR).
(3) Find local maxima and minima in the PDF. All abscissas of the local minima between the two modes are regarded as candidates of the threshold. The abscissas of the ith maximum and minimum after the normal mode are denoted as M_i and m_i, respectively.
(4) The largest candidate among those satisfying the following two conditions is chosen as the threshold.

   a. $\tau_{\min} \leqslant F(m_i) \leqslant \tau_{\max}$, where $\tau_{\min}$ and $\tau_{\max}$ are the minimum and maximum thresholds known in advance, respectively.
   b. $F(m_i) < v \cdot \min\{F(M_i),\ F(M_{i+1})\}$, where v denotes a user-specified parameter to eliminate the shallow valleys of the minima.

2.6. Automatic thresholding for real-time artifact detection

In this study, we modified the KM's automatic thresholding algorithm in order for the algorithm to be used in real-time environments. Unlike the offline environment, a real-time system has little *a priori* knowledge on the signal characteristic, and a data distribution at a certain time may not accurately reflect the distribution of whole data. The modification was conducted to remove all the factors related to *a priori* knowledge and to allow robust operation with only a small amount of data. We changed the following parts in the KM's original algorithm: First, a histogram was used instead of the PDF because the calculation of a histogram is simpler, efficient, and faster than that of the PDF. Second, a previously determined threshold was used instead of calculating a new threshold when there is only a single mode in the distribution. Because it was empirically known that the optimal threshold does not change dramatically, it was better to use the previous threshold than to determine a new threshold based on the assumed SNR. Third, the smallest candidate was chosen instead of the largest. We found that the smallest candidate separates the signal and artifacts better, and consequently a novel condition was introduced to identify the shallow valleys. These changes eliminated the need for *a priori* knowledge on the artifact range in the determination of the threshold. The newly-defined condition and the modified procedure for real-time automatic thresholding are given below (hereafter the proposed algorithm will be referred to as 'real-time KM'):

(1) Accumulate sufficient length of data for initialization. Calculate the histogram using MSDWs. $F(k)$ denotes the density at a specific dissimilarity k.
(2) If there is only a single mode, the threshold will not be changed until a secondary mode is observed.
(3) Find local maxima and minima in the histogram. The largest maximum represents the normal mode. All of the local minima after the normal mode become candidates for the

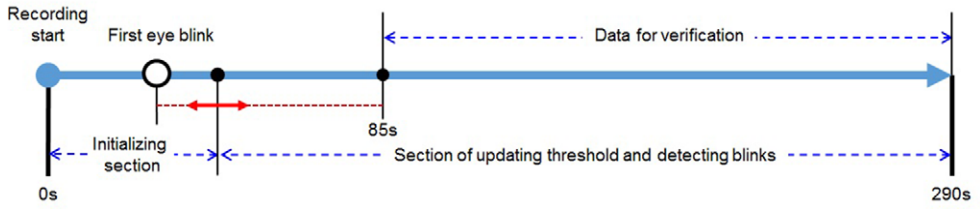

Figure 2. The experimental paradigm for the evaluation of methods. Data in the initializing phase was used to construct the initial histogram, and the histogram was updated in the latter phase. The initializing phase must include at least one eye blink artifact, and a different initializing time was tested in the simulations.

threshold. The abscissas of the ith maximum and minimum after the normal mode are denoted as M_i and m_i, respectively.

(4) The smallest candidate among those satisfying the condition below is chosen as the threshold.

   a. $F(M_i) < v \cdot F(M_0)$, where v denotes a user-specified parameter to eliminate the shallow minima.

After a threshold is chosen based on the feature distribution, an adjustment process for balancing false-positives and false-negatives may be necessary. A new optional parameter (α) is developed for this purpose. The histogram is divided into two groups, normal and blink, based on the originally-determined threshold (θ^{old}), which is then adjusted to include the α-standard deviation of the blink distribution:

$$\theta^{\text{new}} = \theta^{\text{old}} + \alpha \cdot \sigma. \tag{8}$$

For more comprehensive understanding of our procedure, please download the library source codes freely available at http://cone.hanyang.ac.kr/BioEST/Eng/EyeblinkMasterOnline.zip.

3. Results and discussion

A simulated real-time environment was designed for evaluation as illustrated in figure 2. The data were read from pre-recorded EEG data at every sampling time. The data were divided into two sub-sections. The data in the first section were used to construct an initial histogram (let us call this the initializing section), and the histogram was updated using the data in the succeeding section. For the successful construction of the initial histogram, the initializing section needed to include at least one eye blink artifact. Right after the initialization of the histogram, the real-time eye blink detection process started. The threshold was determined using the methods described in section 2, and the results were evaluated by counting the number of correctly detected (true positives) and falsely detected (false positives) events. The true and false positives were counted only in the range from 85 to 290 s in order to sufficiently separate the initializing section and the evaluation section. Note that the length of the initializing section varied in the simulation study. In this study, we regarded that an artifact was correctly detected if the peak point of the artifact was included in the detected artifact range.

For the evaluation of the automatic thresholding methods described in section 2, three different assessment criteria of accuracy were utilized: precision, recall, and F1 score. Precision reflects false detection, recall represents correct detection, and the F1 score is the harmonic mean of precision and recall. The equations for the accuracy measures are as follows:

Table 1. Accuracy of detecting a blinking event: M1, M2, and M3 denote different thresholding algorithms: Common threshold, fixed portion, and the proposed algorithm (real-time KM), respectively. Unit: %.

	Precision			Recall			F1 Score		
	M1	M2	M3	M1	M2	M3	M1	M2	M3
1	90.74	100.00	97.96	98.00	86.00	96.00	94.23	92.47	96.97
2	100.00	100.00	100.00	97.89	50.53	95.79	98.94	67.13	97.85
3	85.71	85.71	86.36	85.71	85.71	90.48	85.71	85.71	88.37
4	94.12	50.00	94.12	100.00	100.00	100.00	96.97	66.67	96.97
5	100.00	100.00	97.50	64.44	62.22	86.67	78.38	76.71	91.76
6	88.89	61.54	72.73	100.00	100.00	100.00	94.12	76.19	84.21
7	98.51	100.00	98.48	100.00	69.70	98.48	99.25	82.14	98.48
8	100.00	100.00	100.00	100.00	61.84	100.00	100.00	76.42	100.00
9	91.67	100.00	94.83	98.21	91.07	98.21	94.83	95.33	96.49
10	85.71	89.66	87.88	100.00	86.67	96.67	92.31	88.14	92.06
11	65.22	53.57	83.33	100.00	100.00	100.00	78.95	69.77	90.91
12	100.00	100.00	100.00	94.59	83.78	94.59	97.22	91.18	97.22
13	88.00	84.62	90.48	91.67	91.67	79.17	89.80	88.00	84.44
14	95.24	95.24	95.24	100.00	100.00	100.00	97.56	97.56	97.56
15	78.26	58.06	100.00	100.00	100.00	100.00	87.80	73.47	100.00
16	98.11	100.00	96.30	98.11	71.70	98.11	98.11	83.52	97.20
17	100.00	93.10	93.75	16.98	50.94	84.91	29.03	65.85	89.11
18	86.96	90.00	95.00	100.00	90.00	95.00	93.02	90.00	95.00
19	100.00	100.00	100.00	88.00	70.00	86.00	93.62	82.35	92.47
20	93.75	80.00	100.00	93.75	100.00	87.50	93.75	88.89	93.33
21	97.96	100.00	100.00	96.00	46.00	94.00	96.97	63.01	96.91
22	92.54	100.00	98.21	92.54	41.79	82.09	92.54	58.95	89.43
23	90.91	47.62	95.24	100.00	100.00	100.00	95.24	64.52	97.56
24	96.77	100.00	100.00	100.00	66.67	93.33	98.36	80.00	96.55
Avg	92.46	87.05	94.89	92.33	79.43	94.04	90.70	79.33	94.20
Stdev	8.27	18.33	6.66	17.84	19.46	6.39	14.34	11.19	4.55
Min	65.22	47.62	72.73	16.98	41.79	79.17	29.03	58.95	84.21
Max	100.00	100.00	100.00	100.00	100.00	100.00	100.00	97.56	100.00

$$\begin{aligned} \text{Precision} &= |\text{TP}|/(|\text{TP}|+|\text{FP}|) \\ \text{Recall} &= |\text{TP}|/(|\text{TP}|+|\text{FN}|) \\ \text{F1 score} &= 2 \cdot \text{Precision} \cdot \text{Recall}/(\text{Precision} + \text{Recall}), \end{aligned} \tag{9}$$

where $|\text{TP}|$, $|\text{FP}|$, and $|\text{FN}|$ denote the numbers of true positives, false positives, and false negatives, respectively.

Table 1 compares the accuracies of the proposed methods implemented with three different thresholding algorithms: (1) common threshold, (2) threshold at a fixed percentile, and (3) real-time KM. The values of the common threshold and the percentiles were chosen experimentally to yield the highest accuracy (140 for the common threshold; 96th percentile for the fixed percentile threshold). All procedures except for the threshold determination were controlled to be the same for each method, i.e. the same algorithm for histogram initializing and updating was employed. The initialization time was fixed to 10 s from the end of the first eye blink, and v for the real-time KM was fixed to 0.1 empirically. The results of the common threshold method (M1) show the necessity for individual thresholding. Although

the mean detection accuracies were acceptable (92.46%, 92.33%, and 90.70% for precision, recall, and F1 score, respectively), relatively low accuracies were observed for some subjects. The F1 scores of three subjects were lower than 80% due to the imbalance between precision and recall. Furthermore, F1 score of one subject (#17) was lower than 30%. Due to the large individual variability, the standard deviation of the method M1 was greater than 10%. The accuracies of a fixed percentile method (M2) were rather lower than those of the common threshold, with an average F1 score of 79.33%. This is because the number of eye blinks varied from 24 to 141 in our dataset (the mean and standard deviation were 59.8 and 32.43, respectively). The accuracies of the real-time KM showed the best performance in all three assessment criteria (94.89%, 94.04%, and 94.20% for precision, recall, and F1 score, respectively). In comparison with the common thresholding algorithm (M1), the improvement of the performance of real-time KM mainly resulted from the improvement of the accuracies in some subjects' data that showed low accuracy in M1 (#5, 11, and 17). Accordingly, the mean F1 score increased by 3.50 %, but the standard deviation dramatically decreased from 14.34 to 4.55.

Tables 2 and 3 show the results of evaluating the initialization and update procedures of the histogram. Table 2 shows the changes in the accuracy of the proposed method with respect to the initialization time. Time zero (0) in the table indicates the point when the first eye blink signal finishes, since the initialization must include at least one eye blink signal. As is clearly shown, the mean F1 score was greater than 90% even when the first five-second signal is included. The standard deviation of the score decreased dramatically when the initialization time exceeded 10 s, after which the standard deviation kept stable. Table 3 shows the changes in the accuracy with respect to the initialization time when histogram updating process was not included in the procedure (the same as table 2 except that there was no histogram updating). In this case, once a threshold was determined with the initial histogram, the threshold was not changed at all during whole processes. In comparison with the results in table 2, the accuracy reported in table 3 generally was relatively lowered and became unstable. These results demonstrate the importance of histogram updating in our proposed method.

Figure 3 shows the change of FP and FN balance with respect to the new parameter (α) introduced in (4). The default precision and recall (when $\alpha = 0$) were 94.89 and 94.13%, respectively, and they traded off as α varied. The changes in the two accuracies were gradual when α was between -1 and 1, whereas recall decreased dramatically when α exceeded 1. These results show that the α value needs to be set to a value between -1 and 1 to achieve acceptable FP and FN ratios in eye blink artifact detection.

The speed of the proposed method was measured for all subjects using the default parameters used in table 1. The mean processing time was 1.5 ms with a standard deviation of 0.79 when the method was coded with MATLAB and run on a Windows 7 platform on an Intel® i5 processor with 8 GB RAM. No significant difference among subjects was found. Since the artifact detection procedure was performed in every 15.63 ms (64 data samples per second), the speed of the proposed method was fast enough to be used for real-time applications.

After the simulated real-time experiments, we performed a fully real-time online experiment with a subject (male, 25 years old). For this online experiment, the proposed method was implemented with C/C++ and tested with a mobile EEG device (Enobio 8, Neuroelectrics, Spain) with a single prefrontal EEG channel (See figure 4). A movie clip demonstration is available in the online version of this article (please find the Supplementary Movie file attached to this article) (stacks.iop.org/PM/37/401/mmedia). Ten eye blinks were observed for 65 s, and all the blinks were detected successfully without any false detection.

Table 2. Changes in the accuracy of the proposed method with respect to the initialization time (mean ± standard deviation, unit: %).

Initialization time (s)	0	5	10	20	30	40	50	60
Precision	82.18 ± 33.49	94.07 ± 14.74	96.05 ± 4.93	94.89 ± 6.66	95.02 ± 6.35	95.82 ± 4.57	95.44 ± 5 .63	95.89 ± 4.74
Recall	91.13 ± 14.05	93.76 ± 6.52	91.79 ± 9.47	94.04 ± 6.39	94.30 ± 6.15	93.97 ± 6.94	93.97 ± 8.00	93.52 ± 7.18
F1 score	79.50 ± 30.75	92.85 ± 11.38	93.54 ± 5.63	94.20 ± 4.55	94.44 ± 4.45	94.69 ± 4.23	94.50 ± 5.77	94.49 ± 4.34

Table 3. Changes in the accuracy with respect to the initialization time when histogram updating process was not included in the procedure (mean ± standard deviation, unit: %).

Initialization time (s)	0	5	10	20	30	40	50	60
Precision	42.89 ± 43.27	75.56 ± 35.60	81.36 ± 30.52	93.33 ± 9.79	85.25 ± 24.17	86.14 ± 22.61	95.01 ± 6.06	92.07 ± 13.21
Recall	96.71 ± 8.25	81.77 ± 33.02	91.31 ± 14.20	93.72 ± 8.60	95.40 ± 8.63	96.06 ± 3.99	93.37 ± 12.43	95.56 ± 4.70
F1 score	45.87 ± 41.03	72.26 ± 32.15	80.26 ± 26.35	92.95 ± 7.12	87.13 ± 20.56	88.41 ± 18.99	93.48 ± 8.74	93.05 ± 8.71

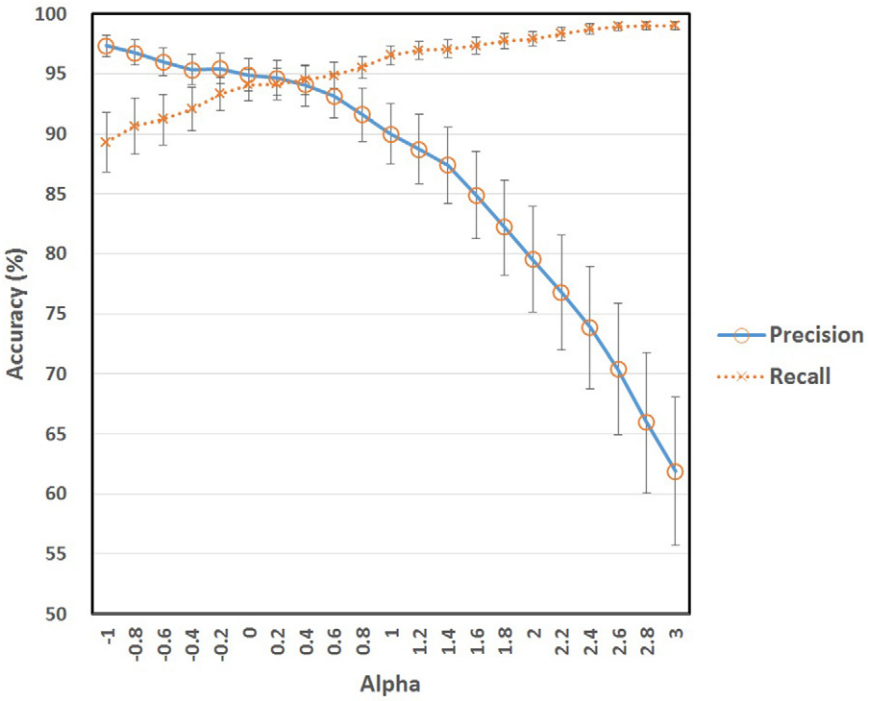

Figure 3. Changes in accuracies (precision and recall) according to alpha value, which shows the trade-off between false-positive and false-negative errors.

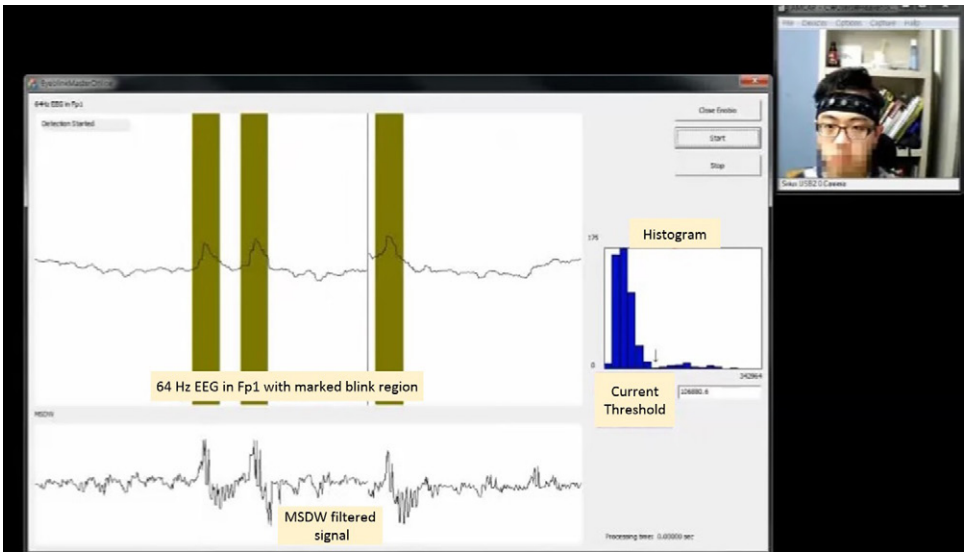

Figure 4. A screenshot of a real-time experiment. The proposed algorithm was tested with a mobile EEG device with a single prefrontal EEG channel. The full movie file can be found as a supplementary material of this manuscript.

The proposed method had a few optional parameters. The influences of initialization time and the balancing factor, alpha, on the overall performance of the proposed method were shown in the results section, where it was demonstrated that the accuracy became stable even with a short initialization time as short as 10 ~ 20 s and the FP and FN trade-off remained acceptable within a certain range of the balancing factor (from -1 to 1). Table 4 shows the influence of another parameter (v) on the accuracy of the proposed method. The parameter v was adopted to avoid misrecognition of a part of the normal (signal) distribution as an eye blink. As shown in the table 4, the influence of the parameter v was quite smaller than expected. The accuracies increased as v increased, but became stable when v was equal to or greater than 0.05 (5%). The proportion of the blinking signal relative to the normal signal was about 4.17% (since the best proportion in the results of the fixed proportion method was 96%); the proposed method exhibited stability even with twice the original proportion.

To further confirm the general applicability of the proposed method, we applied the proposed method to other EEG dataset recorded under different environment, without changing the parameters used for the previous validation study. The additional dataset was recorded from 20 participants using NeuroScan SynAmps2 amplifier (Compumedics USA, El Paso, TX, USA), while the participants were performing visual oddball tasks, of which the detailed paradigm used for the EEG recording can be found in a recent study (Kim *et al* 2015). EEG signals with the length of 120 s were used for the experiment, when 20 s data from the first blink were used for the initialization of the histogram and the data from 60 to 120 s were used for evaluating the artifact detection accuracy. The ground-truth of eye blinking ranges was marked manually by visual inspection of vertical EOG. Please note that any parameter of the proposed method was not changed in this experiment, whereas the threshold for the common threshold method was adjusted in order to yield the best detection accuracy for the new dataset. For the statistical evaluation, all the data from 44 participants (24 original + 20 additional experiments) were divided into two groups based on the F1 score of the common threshold method, which were a group with low F1 (below averaged F1) and a group with high F1 (above averaged F1). Because a group with high F1 has a very little room for improvement, we hypothesized that the detection accuracy would be enhanced for the 'low-F1 group' and unvaried for the 'high-F1 group' by the use of the proposed method. The detection accuracies are summarized in table 5. In the 'low-F1 group', mean F1 score of the proposed method was 9.25 and 10.73 percent point higher than those of the common threshold and fixed portion methods, respectively, when both differences turned out to be significant ($p = 0.0479$ for the common threshold, $p = 0.0085$ for the fixed portion method according to Wilcoxon signed-rank test). In the 'high-F1 group', there was no significant difference in F1 scores between the proposed method and the common threshold method, while the fixed portion method showed significantly lower F1 score than the other two methods (Bonferroni-corrected $p < 0.001$ for both methods according to paired t-test).

In spite of the significant improvement of the detection accuracies in the low-F1 group, the accuracies are still far from the human-level accuracy in blink detection. The main reasons for the lowered accuracy were irregular shapes of blinking artifacts as illustrated in (Chang *et al* 2015) and extremely uncommon eye blinking patterns. Since the proposed method utilizes a distribution of eye blinks to set a threshold, a threshold may not be correctly adjusted when the distribution is in an extreme condition (e.g. no blinking for 10 to 20 s, or one or more blinks in every second). We expect that this issue could be addressed by modifying the algorithms for the construction and update of histograms, which we would like to study further in the future.

Table 4. Influence of the parameter v on the accuracies of the proposed method (mean $\pm$ standard deviation, unit: %).

v	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1
Precision	71.17 ± 42.31	91.96 ± 19.06	95.68 ± 4.79	95.80 ± 4.69	95.76 ± 4.71	95.78 ± 4.71	95.78 ± 4.71	94.89 ± 6.66	94.89 ± 6.66	94.89 ± 6.66
Recall	65.84 ± 39.10	75.34 ± 31.86	80.86 ± 28.71	89.12 ± 19.66	93.31 ± 7.27	94.04 ± 6.39	94.04 ± 6.39	94.04 ± 6.39	94.04 ± 6.39	94.04 ± 6.39
F1 score	49.52 ± 41.50	77.52 ± 29.50	83.54 ± 25.15	90.48 ± 17.27	94.31 ± 4.44	94.73 ± 4.05	94.73 ± 4.05	94.20 ± 4.55	94.20 ± 4.55	94.20 ± 4.55

Table 5. The detection accuracies for two groups divided based on the F1 score of common threshold method. M1, M2, and M3 denote different thresholding algorithms: Common threshold, fixed portion, and the proposed algorithm, respectively (mean ± standard deviation, unit: %).

	Low F1 group ($N = 14$)			High F1 group ($N = 30$)		
	M1	M2	M3	M1	M2	M3
Precision	81.72 ± 17.87	78.34 ± 18.96	92.38 ± 9.57	96.14 ± 4.44	89.70 ± 20.44	96.96 ± 5.71
Recall	82.95 ± 22.05	82.16 ± 18.34	84.29 ± 15.11	98.57 ± 2.86	97.52 ± 4.32	96.84 ± 4.56
F1 score	78.19 ± 16.68	76.71 ± 10.65	87.44 ± 11.12	97.26 ± 2.66	84.42 ± 16.57	96.73 ± 3.65

4. Conclusion

In this study, we introduced a novel method for automatic detection of eye blink artifacts, which does not require any labeled training data. The novel method was based on our previous method for eye blink artifact detection called MSDW (Chang *et al* 2015) together with an automatic thresholding algorithm, both of which had been designed for offline data processing. In order to implement a real-time eye blink detection system, the architecture of the original MSDW was disassembled and recomposed to be readily combined with the automatic thresholding algorithms. Three different automatic thresholding algorithms were combined with MSDW and their performances were compared with each other. Among the three automatic thresholding algorithms including 'common threshold,' 'fixed proportional threshold,' and 'real-time KM', the proposed 'real-time KM' method outperformed the other two methods.

The contributions of this study are as follows: (1) This study confirmed that individual thresholding is necessary for artifact detection. Our simulated real-time analysis showed that individually customized threshold resulted in higher accuracy in detecting eye blink artifacts than common threshold methods, which implies that individual differences in eye blink amplitude have significant influence upon the detection accuracy. (2) This study proposed a novel algorithm structure to detect eye blink artifacts in a real-time environment without any need for *a priori* knowledge on the signal properties. The results of the present study showed satisfactory detection accuracies in comparison with the conventional thresholding methods, and demonstrated a successful trade-off between false-positives and false–negatives. (3) The length of training data could be minimized by using a real-time adaptation procedure. Our experimental results showed that the minimum training time needed to stabilize detection accuracy could be reduced by the use of adaptation procedure (initialization time of 20 s for the proposed method; that of 50 when no adaptation procedure was applied).

It would be interesting for the proposed method to be tested with other patterns in EEG data such as spikes, sharp waves, or a combination of these because these EEG waveforms have similar shapes with eye blink signals. It is expected that the proposed method can be utilized for recognizing these patterns by changing some optional parameters.

Acknowledgments

This work was supported in part by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (No. 2014R1A2A1A11051796) and in part by the Basic Science Research Program through the National Research Foundation of Korea (NRF), funded by the Ministry of Education (NRF-2014R1A1A2A16052334). The authors declare that there is no conflict of interests regarding the publication of this paper.

References

Aarabi A, Kazemi K, Grebe R, Moghaddam H A and Wallois F 2009 Detection of EEG transients in neonates and older children using a system based on dynamic time-warping template matching and spatial dipole clustering *Neuroimage* **48** 50–62

Barbati G, Porcaro C, Zappasodi F, Rossini P M and Tecchio F 2004 Optimization of an independent component analysis approach for artifact identification and removal in magnetoencephalographic signals *Clin. Neurophysiol.* **115** 1220–32

Breuer L, Dammers J, Roberts T P L and Shah N J 2014 Ocular and cardiac artifact rejection for real-time analysis in MEG *J. Neurosci. Methods* **233** 105–14

Chang W-D, Cha H-S, Kim K and Im C-H 2015 Detection of eye blink artifacts from single prefrontal channel electroencephalogram *Comput. Methods Programs Biomed.* **124** 19–30

Chang W-D and Im C-H 2014 Enhanced template matching using dynamic positional warping for identification of specific patterns in electroencephalogram *J. Appl. Math.* **2014** 528071

Croft R J and Barry R J 2000 Removal of ocular artifact from the EEG: a review *Clin. Neurophysiol.* **30** 5–19

Durka P, Klekowicz H, Blinowska K, Szelenberger W and Niemcewicz S 2003 A simple system for detection of EEG artifacts in polysomnographic recordings *IEEE Trans. Biomed. Eng.* **50** 526–8

Fukuda K, Stern J A, Brown T B and Russo M B 2005 Cognition, blinks, eye-movements, and pupillary movements during performance of a running memory task *Aviat. Space Environ. Med.* **76** C75–85 (PMID: 16018333)

Galley N, Schleicher R and Galley L 2003 Blink parameter as indicators of driver's sleepiness—possibilities and limitations *Vis. Vehicles X* ed A Galevol vol 82 (Amsterdam: Elsevier Science Publishers B. V)

Geetha G and Geethalakshmi S N 2012 Artifact removal from EEG using spatially constrained independent component analysis and wavelet denoising with Otsu's thresholding technique *Procedia Eng.* **30** 1064–71

Gratton G, Coles M G and Donchin E 1983 A new method for off-line removal of ocular artifact. *Electroencephalogr. Clin. Neurophysiol.* **55** 468–84

Hagemann D and Naumann E 2001 The effects of ocular artifacts on (lateralized) broadband power in the EEG *Clin. Neurophysiol.* **112** 215–31

Hoffmann S and Falkenstein M 2008 The correction of eye blink artefacts in the EEG: a comparison of two prominent methods *PLoS One* **3** e3004

Jiang J-A, Chao C-F, Chiu M-J, Lee R-G, Tseng C-L and Lin R 2007 An automatic analysis method for detecting and eliminating ECG artifacts in EEG. *Comput. Biol. Med.* **37** 1660–71

Jung T-P, Makeig S, Westerfield M, Townsend J, Courchesne E and Sejnowski T J 2000 Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects *Clin. Neurophysiol.* **111** 1745–58

Kim S and McNames J 2007 Automatic spike detection based on adaptive template matching for extracellular neural recordings *J. Neurosci. Methods* **165** 165–74

Kim D-W, Shim M, Song M J, Im C-H and Lee S-H 2015 Early visual processing deficits in patients with schizophrenia during spatial frequency-dependent facial affect processing *Schizophr. Res.* **161** 314–21 Online: http://linkinghub.elsevier.com/retrieve/pii/S0920996414007300

Klein A and Skrandies W 2013 A reliable statistical method to detect eyeblink-artefacts from electroencephalogram data only *Brain topogr.* **26** 558–68

Kook H, Gupta L, Kota S, Molfese D and Lyytinen H 2008 An offline/real-time artifact rejection strategy to improve the classification of multi-channel evoked potentials *Pattern Recognit.* **41** 1985–96 Online: http://linkinghub.elsevier.com/retrieve/pii/S0031320307004074

Krishnaveni V, Jayaraman S, Anitha L and Ramadoss K 2006 Removal of ocular artifacts from EEG using adaptive thresholding of wavelet coefficients *J. Neural Eng.* **3** 338–46 Online: www.ncbi.nlm.nih.gov/pubmed/17124338

Lawhern V, Hairston W and Robbins K 2013 DETECT: A MATLAB toolbox for event detection and identification in time series, with applications to artifact detection in EEG signals *PLoS One* **8** e62944

Li Y, Ma Z, Lu W and Li Y 2006 Automatic removal of the eye blink artifact from EEG using an ICA-based template matching approach *Physiol. Meas.* **27** 425–36

Mognon A, Jovicich J, Bruzzone L and Buiatti M 2010 ADJUST: an automatic EEG artifact detector based on the joint use of spatial and temporal features *Psychophysiology* **48** 229–40

Mullen T, Kothe C, Chi M, Ojeda A, Kerth T, Makeig S, Jung T-P and Cauwenberghs G 2015 Real-time neuroimaging and cognitive monitoring using wearable dry EEG *IEEE Trans. Biomed. Eng.* **62** 2553–67 Online: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=7274673

Nolan H, Whelan R and Reilly R B 2010 FASTER: Fully automated statistical thresholding for EEG artifact rejection *J. Neurosci. Methods* **192** 152–62

Otsu N 1975 A threshold selection method from gray-level histograms *IEEE Trans. Syst. Man Cybern.* **9** 62–6

Pult H, Riede-Pult B H and Murphy P J 2013 A new perspective on spontaneous blinks *Ophthalmol.* **120** 1086–91

Shao S-Y, Shen K-Q, Ong C J, Wilder-Smith E P V and Li X-P 2009 Automatic EEG artifact removal: a weighted support vector machine approach with error correction *IEEE Trans. Biomed. Eng.* **56** 336–44

Verleger R 1993 Valid identification of blink artefacts: are they larger than 50 microV in EEG records? *Electroencephalogr. Clin. Neurophysiol.* **87** 354–63

Yong X, Ward R K, Birch G E and Society N S 2008 Facial EMG contamination of EEG signals : characteristics and effects of spatial filtering *ISCCSP 2008* (Malta) pp 729–34