IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received August 5, 2017, accepted August 29, 2017, date of publication September 11, 2017,
date of current version September 27, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2750701

A Robust Online Saccadic Eye Movement Recognition Method Combining Electrooculography and Video

XIAOJUAN DING, ZHAO LV, CHAO ZHANG, XIANGPING GAO, AND BANGYAN ZHOU
School of Computer Science and Technology, Anhui University, Hefei 230601, China

Corresponding author: Zhao Lv (kjlz@ahu.edu.cn)

This work was supported in part by National Nature Science Foundation of China under Grant 61401002, and in part by Anhui provincial natural science research project of colleges and Universities under Grant KJ2016A043.

ABSTRACT Eye movement is proven to be the most frequent activities of human beings; as a result research on recognition of unit eye movement has become a hotspot in human activity recognition. In this paper, we propose a robust online saccade recognition algorithm, which integrates electrooculography (EOG) and video together. Initially, EOG signals and video data are collected simultaneously from eight saccadic directions. Then online active eye movement segment detection algorithm is developed to detect the effective saccadic signal from ongoing eyeball activities. Furthermore, we extract features from different modalities and explore two fusion strategies [i.e., feature level fusion (FLF) and decision level fusion (DLF)]. In laboratory environment, the average recognition accuracy of FLF and DLF achieves 89.37% and 89.96%, respectively, which reveals that the proposed method can improve the performance of consecutive saccade recognition in comparison with sole modality.

INDEX TERMS Electrooculography, video recording, wavelet packets, wavelet transforms, human computer interaction.

I. INTRODUCTION

The motivation of Human Activity Recognition (HAR) is to perceive user's intention and present by a natural language, which uses comprehensive information such as the kinds of reflection and patterns of behaviors by signal collecting and pattern recognizing [1]. Due to the extraordinary ability of human-computer interaction, HAR system has emerged as a key research area in intelligent surveillance, video retrieval, motion analysis, virtual reality, health management system and so on [2]–[5].

At present, wearable body information sensors and contact-free environmental sensors have been used to achieve HAR. Wearable body information sensors refer to common ambient sensors such as accelerometers, gyroscope and bio-sensors, etc. In general, common environmental sensors, like reed switches or temperature sensors, are restricted only by perceived basic behavior states, for example, in and out of the room or control a device on or off. Acceleration sensors and gyroscopes are primarily concerned with the perception of physical activity and cannot be used in the dominant visual tasks. Therefore, some subtle clues that provide move valuable information for human behavior recognition are ignored. Fortunately, wearable bio-sensors promise to extend the potential applications using HAR algorithms at anytime, anywhere, through a comprehensive analysis of behavior from the user's perspective [6]. Compared with peripheral bio-electrical signals, electrooculography (EOG) has the advantages of low cost, easy operation, non-invasive, less influence, etc.; and it has been becoming a hotspot in HAR implementation [7]–[9]. Equally important, video-based sensors, one of the contact-free sensors, play a key role in HAR. Because it does not contact the user's skin, it becomes a most acceptable approach to perceive user's behaviors.

During the procedure of design and implementation of online HAR system, several solutions have been proposed to address detection and recognition of unit saccadic eye movement. In the aspect of EOG-based methods, Bulling *et al.* developed the continuous wavelet transform-saccade detection (CWT-SD) algorithm [10]. The CWT-SD regards EOGh and EOGv, which are the denoised EOG signal components, as input arguments to computes the continuous

1-D wavelet coefficients, then applies a specific threshold to divide EOGh and EOGv into saccade or non-saccade. Larsson *et al*. proposed a new method combining saccade detection in the acceleration domain with specialized on- and offset criteria to detect saccades in smooth pursuit movements [11]. The performance of the algorithm is evaluated by comparing the results of the algorithm with the existing results based on the velocity detection method. Similarly, in the aspect of video-based methods, Pauly *et al*. suggested a novel method involving Haar based cascade classifier for eye tracking and a combination of HOG features with SVM classifier for eye blink detection in video frames [12]. The acquisition device used in this algorithm is an ordinary network camera; and the algorithm shows good performance under the uncontrolled lighting condition. Picot *et al*. used two energy signals extracted from the video analysis to detect and characterize blinks from the video [13].

The aforementioned methods have been obtained success in detection of unit eye movement signals respectively; however, some limitations to EOG-based or video-based methods are inevitable. For example, under the condition of EOG acquisition, although the normal activities are not affected, the range of activities is limited. Moreover, the user's slight movements will lead to serious interference as well as affect the performance of HAR system. By contrast, video-based method can effectively overcome this problem, but it is often interfered by the luminance, especially in the dark condition, the performance will decrease sharply. Apparently, since EOG-based or video-based method only focuses on a certain specific aspect, it is difficult to build a robust saccade detection method using single modality. Considering the advantages of two different modalities, we propose a saccadic eye movement recognition method combining EOG and video, so as to improve the performance and robustness of online unit saccadic movement simultaneously. In this paper, we first exploit an online Active Eye Movement Segment Detection (AEMD) algorithm for continuous eye movement signals, and then explain how to extract EOG features and video features respectively. Finally, we propose two fusion strategies to take advantage of their complementary information. This paper is organized as follows: Section II introduces the principle of EOG, Section III details the methodology followed in the course of the work. The experiments are shown in Section IV, and Section V concludes this paper.

II. EOG GENERATION AND ACQUISITION

The eyeball can be seen as a bipolar model with a positive pole at the cornea and a negative pole at the retina (see Fig. 1 (a)) and the movement of the eyeballs can generate a potential between cornea and retina [10]. The amplitude of eye signal will change with the movement of eyeball, if draw this potential to the timeline, we can get a curve called Electrooculogram(EOG). Compared with other bioelectrical signals, EOG has a relatively large signal to noise ratio (SNR) and shows signal amplitudes ranging from 5 to 20μV/degree [10]. Therefore, these features make it an ideal signal for eye movement detection. In our work, we use the NeuroScan instrument to collect EOG signals. A pair of electrodes which are located at the above and below of the eyeball are to collect vertical EOG components and a pair of electrodes arranged near the corner of the left or right eyeball are to detect horizontal EOG components. The other two electrodes are the reference electrode (A1) and the ground electrode (GND) that is attached to the mastoid. The distribution of EOG collecting electrodes is shown in Fig.1 (b).

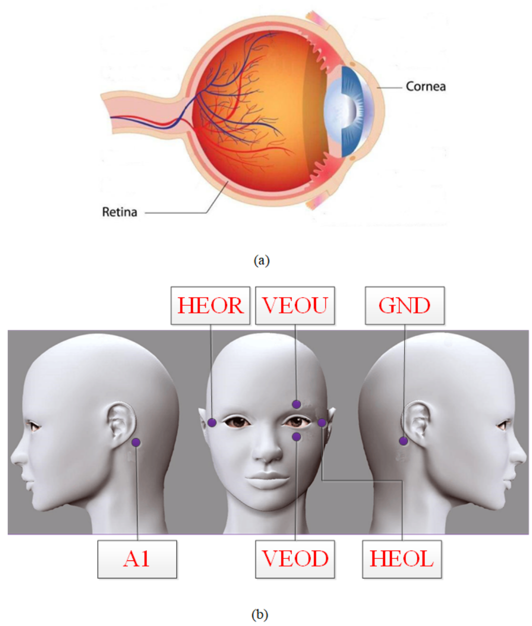


FIGURE 1. Anatomy of the eye ball and setting of EOG collecting electrode. (a) Anatomy of the eye ball. (b) Setting of EOG collecting electrode.

III. METHODS

To start with, we design the experimental paradigm to ensure the reliability and validity of both EOG signals and video data. Then, the raw input signals are preprocessed so as to detect the active eye movement segments and extract features. Finally, the Support Vector Machine (SVM) is utilized to recognize the eye movement according to up, down, left, right, upper left, upper right, lower left and lower right respectively. The overall framework of the algorithm is given in Fig.2.

A. DESIGN OF EXPERIMENTAL PARADIGM

In our work, EOG signals are collected by the Neroscan system with a sampling frequency of 250Hz, and video data are recorded by a high-definition network camera at 30fps. The camera locates in front of the computer screen, the distance between them is about 10 cm. The height of the camera is the same level as the subject's eyeball and 25cm apart from it. In order to ensure the synchronization between EOG and video signals, we use a synchronous acquisition software which developed by our laboratory. Each trial starts with characters of "start" following a short warning tone "beep."

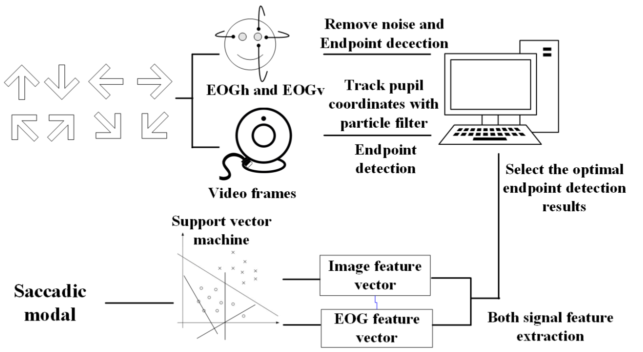


FIGURE 2. Overall frame works of the algorithm and experiment processing.

During second 1 to second 4, a red arrow, pointing either to up, down, left, right, upper left, upper right, lower left and lower right representing one of eight different saccadic tasks, randomly display in the screen. At the same time, participants are instructed to perform the corresponding eye movement. The trial ends with a disappeared arrow and a blank screen at second 4, then the participant has a rest for 2s to relax. Each participant is required to collect 20 groups of data, with each group including 20 saccades. The single experimental paradigm is shown in Fig. 3.

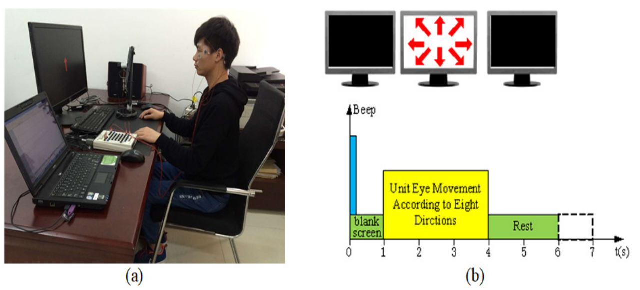


FIGURE 3. The experimental scene and the experimental paradigm. (a) Experimental scene. (b) Experimental paradigm.

A total of 8 subjects (five females and three males whose ages range from 23 to 25) record EOG and video data in a quiet environment. All the participants are graduate students with normal hearing and vision, and right-handed. Before the experiment, they are told that the purpose of the experiment. Since saccades in the four directions (up, down, left, and right) are more common in life, these four directions are scanned 3 times in each group, which is one more than saccade in the oblique directions.

B. DATA PREPROCESSING

It's inevitable that the raw eye movement data will be corrupted by different noise sources, such as the measurement circuitry, participant's slight movement, power interference and so on. In order to preserve the effective eye movement components as well as suppress the artificial noises, a 32-order band-pass digital filter with cut-off frequency of 0.01-10Hz is used in the proposed method [14].

C. ACTIVE EYE MOVEMENT SEGMENT DETECTION (AEMD)

AEMD, a necessary and important step for achieving online function, is mainly used to extract the useful saccadic EOG clips exactly as well as reduce the redundancy information. In this paper, we use a sliding window technology to dynamically estimate energy of saccadic EOG signals. Meanwhile, in order to improve the performance of the AEMD algorithm in noisy environment, we compute and optimize the results of energy based EOG detection and results of particle filter based video detection.

1) EOG DETECTION

To achieve high-speed performance of online processing, we can derive a recursive algorithm to calculate the energy in each sliding window. Suppose the index of the current sliding window is n, the length of this windows is L, the dynamic estimation procedure can be expressed as follows:

$$\mu^{(w)}(n) = \frac{n}{L}[\mu(n) - \mu(n_L)] + \mu(n_L) \tag{1}$$

$$S_2^{(w)}(n) = S_2(n) - S_2(n_L) - \frac{L \times n_L}{n}[\mu^{(w)}(n) - \mu(n_L)]^2 \tag{2}$$

Where $\mu^{(w)}(n)$ and $S_2^{(w)}(n)$ denote the mean and short time energy of the sliding window according to time point n. The recursive results will continuously update sample by sample following the window's slide. In this way, the short-time energy at time point n can be acquired and applied to the AMED algorithm.

The detailed steps are as follows: (1) the filtered EOG data is processed by frame window, and the energy threshold amp is empirically set to a certain value. (2) Calculate the energy value (E) in the current sliding window, compare it with the energy threshold *amp*, if $E > amp$, the corresponding point will be marked as a possible start-point. (3) Backward search from start, the corresponding S_2 value and the energy threshold are compared, here are three cases: A. The sample point is still in EOG signal segment: if S_2 value is greater than *amp*, then the point is still in EOG active segment, activecount (sample points in active segment; initial value is set 0) plus one point; or the point is in nonactive segment, nonactivecount plus one point. If the nonactivecount (sample points in nonactive segment; initial value is set 0) is less than the maxnonactivecount (the maximum allowable sample points in nonactive segment), it is still in EOG signal segment. B. The sample point is in nonactive segment: if the nonactivecount is larger than the maxnonactivecount then determine whether the activecount is greater than the minactivecount (the minimum allowable sample points in active segment); C. If the activecount is greater than the minactivecount: we find the active segment, conversely, the activecount and nonactivecount will be reset to zero. Return to the start position, re-search for the start of the EOG signal segment.

2) VIDEO DETECTION

In the first frame of the video, we empirically initialize a rectangle to select the sclera and pupil, and regard the center coordinates of the rectangular frame as the pupil center. Then the particle filter algorithm is performed to automatically track the eye movement and record the trajectory of the pupil center [16]. Once the movement trajectory is obtained, we will adopt it as a reference coordinates (x_0, y_0), which is key factor in determining the direction of saccade. In addition, baseline shift of the movement trajectory over time is the biggest hurdle in obtaining a high accuracy rate. The least square method is being applied on the whold data to remove baseline shift. Moreover, because of the involuntary blink is unavoidable during the experiment and will affect the performance of endpoint detection. Thus, rely on a suitable threshold (*thbl*: is the multiple of the maximum amplitude of video data) and the duration of the vertical signal (X_t: the time difference between the threshold and two intersection of the video waveform is set to X_t), the vertical signal above the threshold (*thbl*) value is set to y_0. This operation eliminates interference of blink signal. Furthermore, the horizontal and vertical waveforms are smoothed by moving average algorithm. The two sets of waveforms are superposed to obtain the absolute value, and an appropriate threshold is set up according to the average value as the threshold of the endpoint detection. The amplitude and threshold value is compared from the first frame. By applying a threshold (*thw*) on the video signal, we create a vector W with element W_i:

$$W_i = \begin{cases} 1, & V_i > thw \\ 0, & V_i < thw \end{cases} \tag{3}$$

V stands for vertical signal; i represent the subscript of each signal point. This step divides video signal into active ($W = 1$) and nonactive ($W = 0$) segment. The start and end points of the saccade signal are stored in the corresponding array.

3) OPTIMAL AMED RESULTS SELECTION

Considering the difference of results between the above-mentioned methods (especially in the noisy environment), we introduce a comparison algorithm to choose the best one. For example, if the HAR system is interfered by luminance, the video-based AMED does not work well but the EOG-based AMED can exactly detect the end-point; similarly, if the subject slightly sways his body, the performance of the EOG-based AMED will decrease due to micro-movement of bio-electrode, but the video-based AMED can keep the stable detection. We observe that the time-domain waveforms will be severely distorted in the case of noisy environment. As a result, the duration between the start and end point will become very short than normal cases. In addition, the longer duration can carry more eye movement information in the normal case; we therefore compare the duration of endpoint detected signals respectively and choose the longer duration as the optimal results.

D. FEATURE EXTRACTION

1) EOG FEATURE EXTRACTION

Wavelet transform is a powerful transformation analysis method, which can analyze the signal in time and frequency. Unlike wavelet transform, the wavelet packet transform (WPT), for the subsequent decomposition level, not only decomposes the approximation coefficients, but also decomposes the detail coefficients. Therefore, we use the wavelet packet decomposition to extract the EOG feature after preprocessing. This decomposition process generates a wavelet packet tree structure, and the node representation of the tree is associated with different frequency location feature subspace. We chose sym4 as the mother wavelet and set the decomposition level to 3. Then we randomly select an EOG segment and decompose it with wavelet packet transform. As we all know, eye signal is mainly concentrated in the low frequency band. By many comparison experiments, the algorithm finally selects the low frequency coefficients of the third layers as the classification feature and feed directly into the SVM for classification.

2) VIDEO FEATURE EXTRACTION

Two dimensional wavelet transform decomposes the image into four subimages, as shown in Fig.4, which are the average image and three details of the image. The four subimages combined into a general plan which can be seen in Fig.4 that the dimension is $68 * 88$. In order to achieve the purpose of saccade classification, we use the two dimensional wavelet transform decomposes the image of the active saccadic eye movement segments and convert the general plan into an image feature vector. Then these image feature vector are integrated in a two-dimensional matrix. Connect each row in the two-dimensional matrix to get a feature vector that identifies the saccade state. This feature vector is used as saccade movement representation for the purpose of saccade classification.

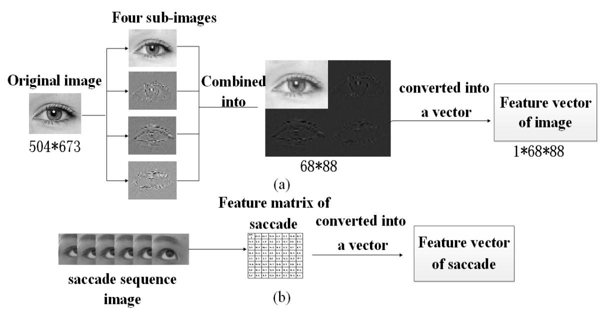


FIGURE 4. **Diagram of the 2-D wavelet decomposition. (a) 2-Dimensional wavelet transform for one frame image. (b) 2-Dimensional wavelet transform for saccade sequence image.**

3) SACCADE RECOGNITION

a: Saccade Recognition Based On Single Modality

According to the characteristics of EOG and video signals, we choose SVM with polynomial kernel and SVM with linear kernel as classifier to recognize them respectively [17].

b: MODALITY FUSION

Signals from different modalities represent different aspects of eye movement and the complementary information from different modalities can be integrated to build a more robust saccadic action recognition model. Within this section, we present two fusion strategies. Let x_{eog} and x_{vedio} as feature vector of EOG and video respectively, they can be described as follows:

c: FEATURE LEVEL FUSION (FLF)

The EOG feature vectors and the video feature vectors are directly combined into a new feature vector, i.e.,

$$x_{FLF} = x_{eog} + x_{vedio} \tag{4}$$

Where x_{FLF} denotes the new FLF feature.

d: DECISION LEVEL FUSION (DLF)

The maximum value method is a kind of decision level fusion method, which has the advantages of no need of training, simple method and so on. It selects a maximum possible class label as the result tag according to the posterior probability. The max rule is thus defined as follows for a given trial x, i.e.,

$$C(x) = \arg\max\{P(w_a|x)\}|_{a=1\ldots k} \tag{5}$$

$$C_{final}(x) = \arg\max\{P_q(w_a|x)\}|_{\substack{a=1\ldots k \\ q=1\ldots Q}} \tag{6}$$

In (5) and (6), C is the primary class decision output. C_{final} is the output of the decision level fusion method based on the max rule. Q is the ensemble of all classifiers that can be selected for fusion. This experiment has a total of k categories. $P(w_a|x)$ is the posterior probability of having class w_a the sample is x according to classifier q. We first calculate the posterior probabilities of the eight categories for the two models. Then, the maximum posterior probability corresponding to the eight categories in EOG and video are compared with each other. Finally, we choose the category with the highest posterior probability as the final decision.

IV. EXPERIMENTS AND RESULTS ANALYSIS

A. ACTIVE EYE MOVEMENT DETECTION

In EOG detection, the initial threshold of energy (*amp*) is 0.5053 and the length of window is 80 sample points. In video detection, the initial threshold (*thw*) is 0.3492. Fig.5 describes waveform of saccade EOG and video signals that are randomly selected, red vertical lines indicate the start-points, while black lines denote the end-points.

Furthermore, Probability of correct point detection (Pc-point) and Probability of false point detection (Pf-point) are computed to evaluate the validity of the proposed AEMD algorithm, they are defined in [15]:

The results of algorithm are listed in Table 1. In AEMD experiments, the average pc-point of the eight-direction saccade are 77.05%, 82.19%, 77.6%, 78.63%, 81.46%, 81.64%, 80.44%, and 81.16% respectively, while, the average pc-point and pf-point are 80.02% and 7.56%. Experimental results reveal that the proposed AEMD algorithm can acquire an

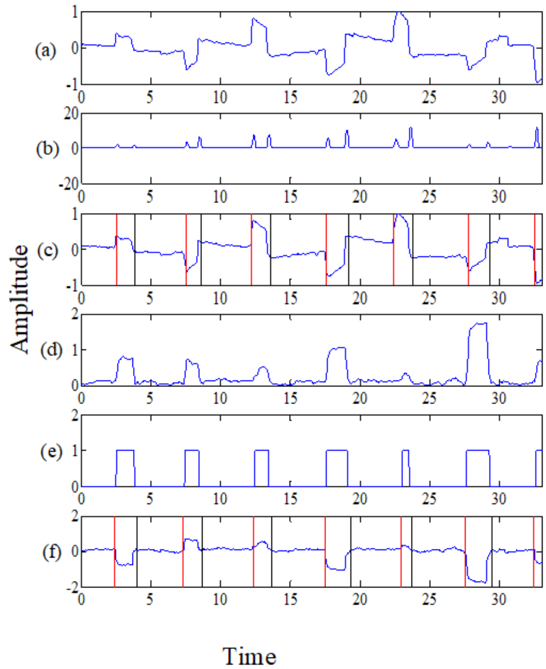


FIGURE 5. Eye movement signal endpoint detection. (a) EOG signal. (b) Short time energy. (c) Endpoint detection result. (d) Original video data. (e) Vector W. (f) Video endpoint detection result.

TABLE 1. Performance of the endpoint detection algorithm.

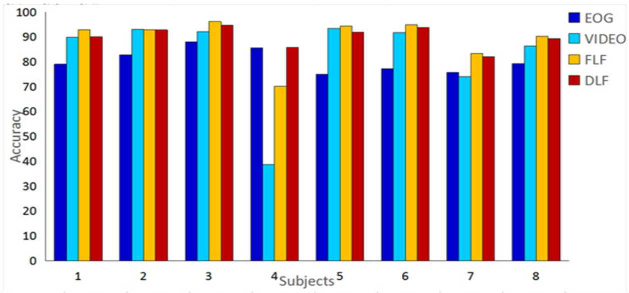
Type	Pc-point(%)	Type	Pc-point(%)
Up	77.05	Upper left	81.46
Down	82.19	Upper right	81.64
Left	77.60	Lower right	80.44
Right	78.63	Lower left	81.16
Average pc-point		80.02	
Average pf-point		7.56	

ideal accurate ratio in continuous activity eye movement segment detection.

B. SACCADE RECOGNITION

Recognition results are compared with data labels for acquiring the classification accuracy ratio. If they are the same, the classification is proven correct, otherwise, it is wrong. For improving the validity of recognition results, the penalty coefficient is set to one and 3-fold cross-validation method is adopted. It works as follows: all samples are sorted at random and divided into 3 parts, then one part is taken out as test database and the rest two parts are used to train the classifier; repeat this procedure 10 times for each participant's data.

The Fig.6 (a) shows the accuracies for different unimodalities and multi-modalities using SVM as classifier. As we can see from Fig.6 (a), we achieved 80.33% and 82.41% average accuracies from EOG and video data, respectively. The feature level fusion method give an average accuracy of 89.37%. For decision level fusion, we use maximum value method to combine two models trained with EOG and video data, which achieved 89.96% average accuracy. The results show that, the models with modality fusion perform is better than



	1	2	3	4	5	6	7	8	Average
• EOG	79	82.9	88.1	85.66	74.95	77.16	75.66	79.24	80.33
• VIDEC	89.8	93.1	92.1	38.75	93.42	91.66	74.07	86.41	82.41
• FLF	92.8	92.9	96.2	70.16	94.34	94.99	83.33	90.24	89.37
• DLF	90.1	92.8	94.7	85.83	91.94	93.83	82.13	88.33	89.96

(a)

	↑	↓	←	→	↖	↗	↘	↙
↑	0.94	0	0	0	0	0.02	0	0.04
↓	0.01	0.8	0	0.12	0	0.04	0.03	0
←	0.02	0	0.88	0	0.04	0	0	0.06
→	0.01	0.08	0	0.86	0	0.04	0.01	0
↖	0.01	0	0.35	0	0.6	0	0	0.04
↗	0.04	0.11	0	0.07	0	0.78	0	0
↘	0.02	0.08	0.01	0.09	0	0.01	0.79	0
↙	0.22	0	0.09	0	0	0	0	0.69

(i) EOG

	↑	↓	←	→	↖	↗	↘	↙
↑	0.87	0.03	0.03	0.01	0.02	0.01	0.01	0.02
↓	0.05	0.89	0	0	0.02	0	0.02	0.02
←	0.02	0.03	0.86	0.02	0.03	0	0	0.04
→	0.02	0	0.02	0.86	0.04	0.02	0.04	0
↖	0.02	0.02	0.16	0.04	0.7	0.05	0.01	0
↗	0.04	0	0.06	0.12	0.01	0.75	0.02	0
↘	0.03	0.03	0.03	0.8	0.04	0.04	0.72	0.03
↙	0.03	0.07	0.04	0	0	0	0.05	0.81

(ii) VIDEO

	↑	↓	←	→	↖	↗	↘	↙
↑	0.95	0	0.02	0	0.01	0.02	0	0
↓	0.02	0.95	0	0	0	0.01	0.02	0
←	0.02	0	0.89	0	0.05	0	0	0.04
→	0	0	0	0.92	0.01	0.03	0.04	0
↖	0.01	0	0.18	0.01	0.76	0	0	0.04
↗	0.04	0	0	0.06	0	0.9	0	0
↘	0	0.11	0.02	0.12	0	0.01	0.73	0.01
↙	0	0	0.03	0	0	0	0	0.97

(iii) FLF

	↑	↓	←	→	↖	↗	↘	↙
↑	0.98	0	0	0	0	0.02	0	0
↓	0	0.93	0	0.02	0.01	0.01	0.03	0
←	0.01	0.01	0.92	0	0.04	0	0	0.02
→	0	0.02	0	0.92	0.02	0.03	0.01	0
↖	0.01	0	0.16	0	0.8	0	0	0.03
↗	0.06	0.02	0	0.03	0	0.89	0	0
↘	0.01	0.07	0.01	0.04	0	0	0.87	0
↙	0.09	0	0.03	0	0	0	0.01	0.87

(iv) DLF

(b)

FIGURE 6. **Experimental results expressed in different forms. (a) The comparison of the performances of unimodalities and multi-modal. (b) Confusion matrices of single modality and different modality fusion strategies.**

that of unimodality because modality fusion can combine complementary information in each single modality.

For the No.4 subject, the recognition accuracy obtained by video-based unimodality method is relatively poor (38.75%) due to the surrounding luminance interference. Whereas, EOG-based method, which is a bio-electrical signal, is less disturbed by light noise, presents a stable performance (85.66%). Since FLF and DLF strategies can be complementary to improve the performance of unimodality by combining EOG and video, they achieve 70.16% and 85.83% average accuracy, which is much better than video modality. Likewise, for the No.7 subject, the unimodality method gets a lower recognition rate (75.66% and 74.07%) than fusion modalities (83.33% and 82.13% respectively) when the number of blinks is more than normal trials. The experimental results reveal that the proposed combination feature parameters perform a higher robustness than unimodality in noisy environment.

On closer inspection of the accuracy under two fusion modalities, it turns out that the overall performance of FLF method (92.11%) is better than DLF (90.55%) except for luminance interference condition. The likely reason is that FLF directly concatenate two feature vectors of EOG and video to train the detection model, which use more complementary information than DLF.

The confusion matrices of each unimodality and multi-modalities are shown in Fig.6 (b), which presents details of the merits and drawbacks of each modality. Each row of the confusion matrices stands for the target class and each column stands for the predicted class. The elements (i, j) in the matrix denote the percentage of samples in class i that is classified as class j.

From Fig. 6 (b) we can find that the largest between-class substitution error between upper-left and left in all modalities, i.e., 35% and 16% in case of EOG and video respectively that means most of upper-left saccades are falsely returned left saccades. By contrast, the largest error is 18% and 16% according to FLF and DLF. Compared with EOG modality, the proposed fusion strategies can effectively decrease confusion; similarly, compared with video modality, they can keep a lower confusion rate steady as well as classify up, down, left and right saccade with higher accuracy. Therefore, the proposed fusion modalities can be complementary to improve the results for saccadic eye movement recognition.

V. CONCLUSION

A robust online saccadic eye movement recognition method combining electrooculography and video has been proposed in this paper. We first designed a saccade experiment and collected EOG signals as well as video data in terms of eight saccadic directions. Then we detected the active eye movement segments and extracted different features including wavelet packet coefficient and wavelet coefficient for EOG signals and video data, respectively. The experimental results over seven subjects indicated that FLF strategies achieve the best average classification accuracies of 92.11%, which is nearly 12.54% and 3.46% higher than EOG and video strategy, respectively.

The investigation of HAR system based on EOG would be promising tool to improve the quality of our life, by providing a feasible human computer interaction. In this research, we are mainly concentrated on online recognition of basic saccadic eye movement. In which, the AEMD algorithm plays an important role because it is associated with accurate rate of saccadic actions, especially in noisy environment. Therefore, our future work should prioritize research to develop a more robust AEMD algorithm to process continuous eye movement and apply it to improve the performance of HAR system.

REFERENCES

[1] J. K. Aggarwal and M. S. Ryoo, "Human activity analysis: A review," *ACM Comput. Serv.*, vol. 21, no. 3, pp. 16:1–16:43, 2011.

[2] A. Sathyanarayana *et al.*, "Robust automated human activity recognition and its application to sleep research," in *Proc. IEEE Int. Conf. Data Mining Workshops*, Dec. 2016, pp. 495–502.

[3] U. Fareed, "Smartphone sensor fusion based activity recognition system for elderly healthcare," in *Proc. Workshop Pervasive Wireless Healthcare*, Jun. 2015, pp. 29–34.

[4] H. Wang, O. Dan, J. Verbeek, and C. Schmid, "A robust and efficient video representation for action recognition," *Int. J. Comput. Vis.*, vol. 119, no. 3, pp. 219–238, 2016.

[5] H. Rajabi and M. Nahvi, "An intelligent video surveillance system for fall and anesthesia detection for elderly and patients," in *Proc. Int. Conf. Pattern Recognit. Image Anal.*, Mar. 2015, pp. 1–5.

[6] E. M. Tapia, S. S. Intille, and K. Larson, "Activity recognition in the home using simple and ubiquitous sensors," in *Proc. Int. Conf. Pervasive Comput.*, 2004, pp. 158–175.
[7] Z. Lv, X.-P. Wu, M. Li, and D. Zhang, "A novel eye movement detection algorithm for EOG driven human computer interface," *Pattern Recognit. Lett.*, vol. 31, no. 9, pp. 1041–1047, 2010.
[8] M. S. Hossian, K. Huda, S. M. S. Rahman, and M. Ahmad, "Implementation of an EOG based security system by analyzing eye movement patterns," in *Proc. Int. Conf. Adv. Electr. Eng.*, Dec. 2015, pp. 149–152.
[9] C.-C. Postelnicu, F. Girbacia, and D. Talaba, "EOG-based visual navigation interface development," *Expert Syst. Appl.*, vol. 39, no. 12, pp. 10857–10866, 2012.
[10] A. Bulling, J. A. Ward, H. Gellersen, and G. Troster, "Eye movement analysis for activity recognition using electrooculography," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 4, pp. 741–753, Apr. 2011.
[11] L. Larsson, M. Nystrom, and M. Stridh, "Detection of saccades and postsaccadic oscillations in the presence of smooth pursuit," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 9, pp. 2484–2493, Sep. 2013.
[12] L. Pauly and D. Sankar, "A novel method for eye tracking and blink detection in video frames," in *Proc. IEEE Int. Conf. Comput. Graph., Vis. Inf. Secur.*, Nov. 2015, pp. 252–257.
[13] A. Chaudhuri, A. Dasgupta, and A. Routray, "Video & EOG based investigation of pure saccades in human subjects," in *Proc. 4th Int. Conf. Intell. Human Comput. Interact.*, 2012, pp. 1–6.
[14] R. OuYang, Z. Lv, and X. Wu, "An algorithm for reading activity recognition based on electrooculogram," in *Proc. Int. Conf. Inf., Commun. Signal Process.*, Dec. 2015, pp. 1–5.
[15] Y. Luo, W. U. Xiaopei, L. Zhao, K. Peng, and Y. Gui, "A recursive calculating algorithm for higher-order cumulants over sliding window and its application in speech endpoint detection," *Chin. J. Acoust.*, vol. 2015, no. 4, pp. 436–449, 2015.
[16] F. Zhou, W. Chen, and H. Fang, "Robust eye tracking and location method based on Particle filtering algorithm," in *Proc. Int. Conf. Cloud Comput. Intell. Syst.*, Nov. 2015, pp. 247–252.
[17] Y. Peng and B.-L. Lu, "Discriminative extreme learning machine with supervised sparsity preserving for image classification," *Neurocomputing*, vol. 261, pp. 242–252, Oct. 2017.

XIAOJUAN DING received the B.E. degree in network engineering from Jianghuai College, Anhui University, Hefei, China, in 2015. She is currently pursuing the M.E. degree with Anhui University. Her current research interests include biomedical signal processing and human-computer interaction.

ZHAO LV received the Ph.D. degree in computer application technology from Anhui University, Hefei, China, in 2011. He is currently an Associate Professor with the School of Computer Science and Technology, Anhui University. He is also a key member with the Laboratory of Intelligent Information Processing and Human Computer Interaction. His research focuses on intelligent information processing, biomedical signal processing, and speech signal processing. He is the author of over 40 articles. He applied for nine national invention patents, of which five were authorized. His research is supported by the National Nature Science Foundation of China and Natural Science Foundation of Anhui Province.

CHAO ZHANG received the M.E. and Ph.D. degrees in computer application technology from Anhui University, Hefei, China, in 2009 and 2014, respectively. He is currently a Lecturer with the School of Computer Science and Technology, Anhui University, Hefei, China. He research interests include intelligent information processing and human–computer interaction.

XIANGPING GAO received the bachelor's degree in 1998, and the M.E. degree in signal and information processing from the School of Electronic and Information Engineering, Anhui University in 2006. She is currently a Lecturer with the Anhui University of Computer Science and Technology, Hefei, China. Her research interests include intelligent information processing and pattern recognition.

BANGYAN ZHOU is currently pursuing the Ph.D. degree with Anhui University, Hefei, China. She is also a Lecturer with the Computer Studies Department, Anhui University. Her research interests include Biomedical Signal Processing and Pattern Recognition.