

Efficient detection of eye blinking and yawn based on facial video utilizing IPPG technique

Bing Wei^{1, 2, a}, Xue Lu^{1, a}, Chao Zhang^{1, b} and Xiaopei Wu^{1, c}

¹School of Computer Science and Technology, Anhui University, China

²Computer department, Hefei Normal Collage, China

^aweibing8572@qq.com, ^b514207713@qq.com, ^cwxp2001@ahu.edu.cn

Keywords: Eye blinking; Yawn; IPPG; Region of interest; Spatial averaging.

Abstract. For purpose of efficient detection of eye blinking and yawn of human beings, a facial expressions detection method based upon Imaging Photoplethysmography (IPPG) technique was proposed. In the proposed method, facial videos were transformed into R/G/B signals through spatial pixels averaging on related regions of interest (ROI). Then, the signs of eye blinking and yawn were extracted from G-channel signals that contain more stable eye blinking and yawn components by adaptive energy thresholds. The experiments carried out on eight participants demonstrated that the proposed method had an impressive accuracy in detection. It might provide an efficient and succinct approach for applications of facial expression detection, such as fatigue detection, etc.

Introduction

Eye blinking and yawn are vital signs in human facial expressions, which usually indicate physiology and emotion status of human beings [1, 2]. Currently, numerous researches have focused on themes of human physiology monitoring and emotion computation. These researches frequently involved facial expression detection, for instance, eye blinking and yawn detection are significant for studies on human fatigue [1-5]. The detection of eye blinking and yawn in these researches majorly employs Bio-electrical measurement and computer vision technology. Bio-electrical measurement relies on measuring change of Electrooculogram (EOG) or Electromyography (EMG) caused by motions of eye blinking and yawn by bio-electrodes placed on surrounds of related tissues [6, 7]. However, in this approach, metallic electrodes commonly bring discomforts for human body, and the measurements might easily be disturbed by muscular movement or electrostatic noise. As for computer vision technology, it recognizes motions of eye blinking and yawn by using video tracking and detection algorithms on related regions in facial videos [8, 9]. This non-contact technology has obvious advantages in detection and been followed by many studies, but still needs constant researches and explorations on robustness and applicability of relevant algorithms. The computational complexities of many sophisticated facial video tracking and detection algorithms should be taken into account for applications of this technology on different platforms.

IPPG is a new intelligent information processing technique developed in recent years, which could restructure hidden physiological information (such as blood volume pulses) from videos of body surface captured by ordinary camera [10]. In IPPG technique, it commonly transforms Red/Green/Blue (R/G/B) video into three one-dimensional R/G/B signals that have better signal-to-noise ratio (SNR) by using spatial pixels averaging on ROI of video frames [11]. Through ROI spatial averaging, some physiological parameters and motion artifacts could be extracted from generated R/G/B signals efficiently and succinctly. This approach has aroused much interest in physiology and emotion monitoring field [12, 13]. Thus, it deserves to explore eye blinking and yawn detection based upon these related researches. In this study, we proposed a method based on approach of ROI spatial averaging on facial video to detect features of eye blinking and yawn. It is the main contribution of this study that the proposed method employed the theory of IPPG technique could achieve efficient and succinct detection of eye blinking and yawn. In the remainder of the paper we will sequentially explain the details of the method, validation experiments and the results.

Methods

In this section, the details of our method are described. For facial video captured, we utilized mean-shift tracking algorithm to track the two ROI (regions of eyes and mouth) on video frames, and then transformed the video into two groups of one-dimensional R/G/B signals by using spatial pixels averaging on these two ROI. After that, the motion features of eye blinking and yawn were extracted from the relevant G-channel signals by adaptive energy thresholds respectively. The flow chart of the method is shown in Fig.1.

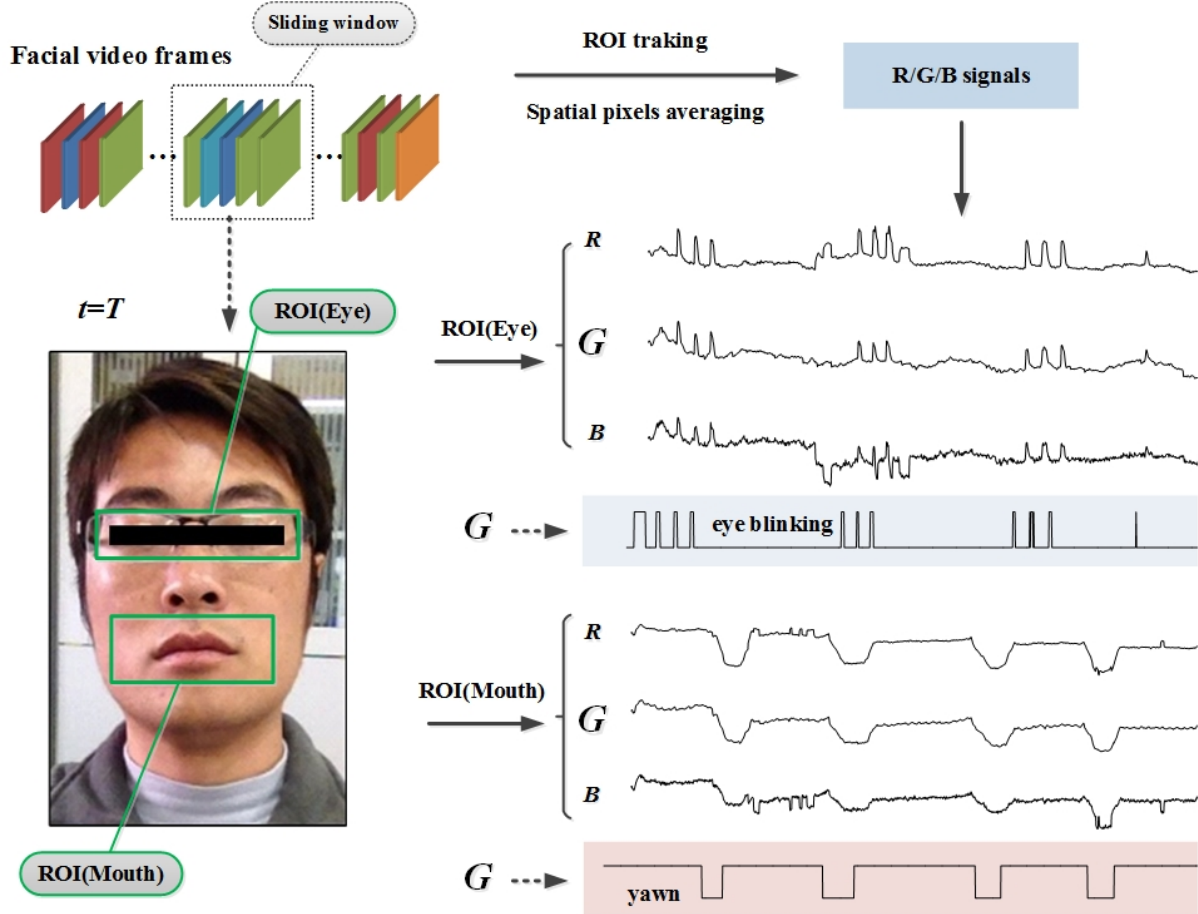


Fig.1. The flow chart of eye blinking and yawn detection

R/G/B signals generation. In this study, the facial videos are captured at a frame rate of 30fps with pixel resolution of 640×480 and saved in MOV format. We selected iPhone4s as camera device, and analyzed the video data on MATLAB2015a platform. For each video, we fixed the two ROI (regions of eyes and mouth shown in Fig.1) on the first video frame, and tracked them in the video utilizing mean-shift tracking algorithm which could compensate for the global motion of the whole face. Then, based on sliding window, these two ROI sequences obtained were transformed into two groups of one-dimensional R/G/B signals by using spatial pixels averaging. Assume R/G/B signals were denoted by the expression $\{x_R(t), x_G(t), x_B(t)\}$, the calculation formula was briefly given as follow:

$$\begin{cases} x_R(t) = \frac{1}{N \cdot M} \sum I_R(x, y, t) \\ x_G(t) = \frac{1}{N \cdot M} \sum I_G(x, y, t) ; \quad 1 \leq x \leq N, 1 \leq y \leq M, \\ x_B(t) = \frac{1}{N \cdot M} \sum I_B(x, y, t) \end{cases} \quad t = 1, 2, \dots, T \quad (1)$$

Where $I_R(x, y, t)$, $I_G(x, y, t)$, and $I_B(x, y, t)$ are R/G/B pixels respectively, N and M are the height and width of the selected ROI, T is the number of the frames in the sliding window.

Detection of eye blinking and yawn. Commonly, R/G/B signals yielded from facial videos variably contains many noises caused by complex factors, such as interaction of light with complicated facial tissues, or even weak ambient light changes. These noises could hardly be attenuated thoroughly only utilizing mean-shift tracking algorithm. Therefore, for noise suppression, we sequentially preprocessed the R/G/B signals by three steps, namely high pass filtering with cutoff at 0.15Hz, detrending and normalization.

For illustrative purposes, we randomly picked a facial video as example, and gave out related results. In Fig.2, there are two segments of R/G/B signals generated based on the two ROI (regions of eyes and mouth) respectively, and the features of eye blinking and yawn could be identified from the waveform of the R/G/B signals [see Fig.2 (a) and Fig.2 (d)]. Then, after preprocessing, these features became more distinguished [see Fig.2 (b) and Fig.2 (e)]. Moreover, among the R/G/B signals, the features of eye blinking and yawn components contained in G channels are more reliable than those in R or B channels.

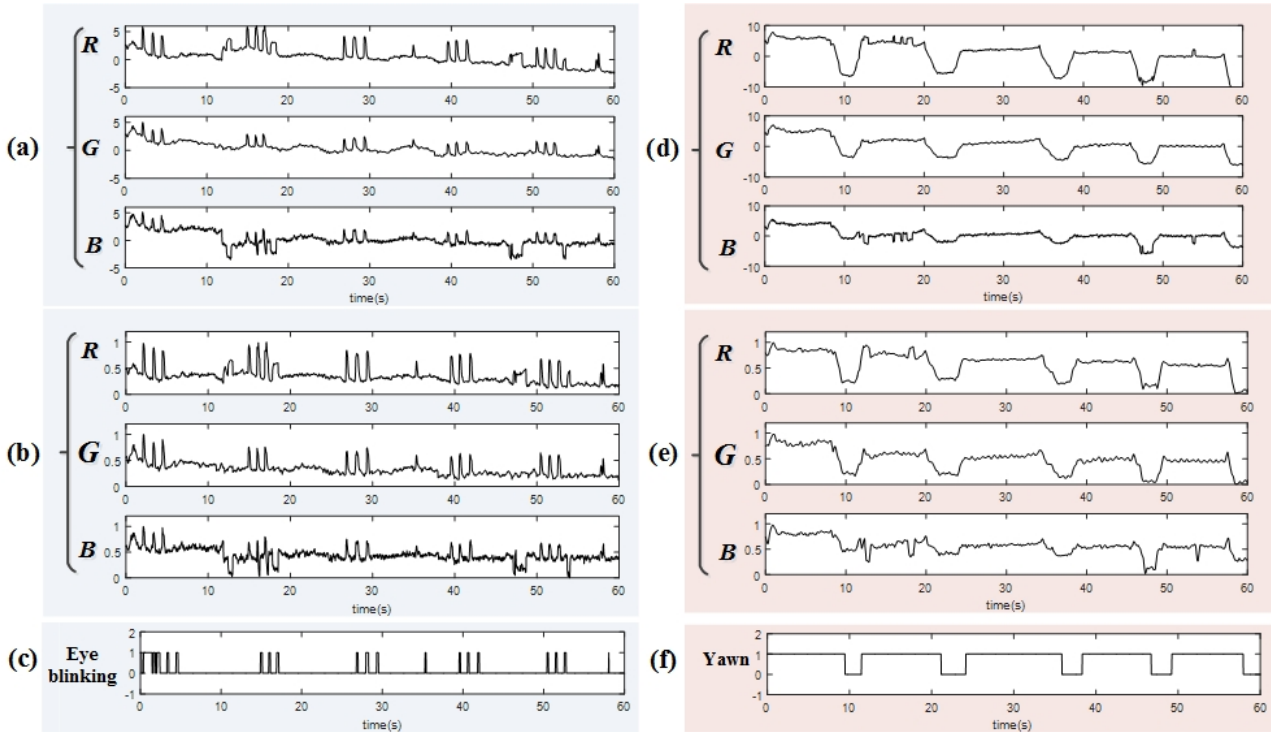


Fig.2 Detection of eye blinking and yawn based upon R/G/B signals. (a) and (d) are two segments of R/G/B signals generated based on the two ROI respectively, their preprocessing results are shown in (b) and (e), further, the eye blinking and yawn detected from G channels of preprocessed signals are given in (c) and (f).

Follow the reasons indicated above, we detected eye blinking and yawn from G-channel of preprocessed signals by adaptive energy thresholds. Firstly, we smoothed the G-channel signals by three-point average smoothing filter, and extracted the peaks of waveform using the following formulas:

$$D(i) = S(i) - S(i-1); \quad 2 \leq i \leq N. \quad (2)$$

$$Peaks(j, p) = \begin{cases} rising(j) = \sum_{i=k_1}^p D(i), & D(i) \geq 0 \\ falling(j) = \sum_{i=p}^{k_2} D(i), & D(i) < 0 \end{cases}; \quad \begin{matrix} 2 \leq i \leq N, & 1 \leq j < N-1, \\ 2 \leq k_1 < p < k_2 \leq N \end{matrix} \quad (3)$$

Where $S(i)$ and $D(i)$ are G-channel signal and its differences respectively. We structured $Peaks(j,p)$ comprised of rising edge or falling edge by accumulating the consecutive $D(i)$ values with the same signs, where p is index of the peak in G-channel signal. Then, all the peaks denoted by $S(p)$ could be extracted from G-channel signal. Based upon a moving time window of 20s, we clustered the peaks $S(p)$ within the window into two clusters by K-means clustering method. For the two clusters, in eye blinking detection, the peaks caused by eye blinking are more prominent compared with other thorns, the cluster with larger peak values is eye blinking cluster. Thus we took the minimum one of the eye blinking cluster as the energy threshold. Similarly, as for yawn detection, the cluster with smaller peak values is yawn cluster, and the maximum one of the yawn cluster was taken as the energy threshold. Consequently, the eye blinking and yawn contained in G-channel signals are detected effectively by the adaptive energy thresholds devised [see Fig.2 (c) and Fig.2 (f)].

Experiments and results

The experiments were carried out on eight subjects aged between 20 and 40 have no physical disability on eyes and mouth. During the experiments, the lab illumination was kept adequate and stable, the subjects sited still in front of the camera (iphone4s) at a distance of 0.6m and did eye blinking and yawn according to guidance. Table1 shows the statistic results of experimental data. The experimental results demonstrated that the eye blinking and yawn done by subjects could be accurately detected from G-channel signals by the adaptive energy thresholds based on the moving time window.

Table 1. The statistic results of experimental data

Recognition ratio Subjects	Eye blinking	Yawn
Subject1	18/18	4/4
Subject2	18/18	4/4
Subject3	19/20	5/5
Subject4	20/20	5/5
Subject5	21/22	6/6
Subject6	20/22	6/6
Subject7	22/24	6/7
Subject8	23/24	7/7

Conclusions

We presented a method for eye blinking and yawn detection from facial videos. The method employed the theory of IPPG technique could achieve efficient and succinct detection in experiments. For the video captured, we transformed it into R/G/B signals by using spatial pixels averaging approach on the eye ROI and mouth ROI tracked. Then we extracted eye blinking and yawn from G-channel signals by the adaptive energy thresholds. We found that the features of eye blinking and yawn could be highlighted through simple ROI spatial pixels averaging. Moreover, the G-channel signals yielded contain more distinguished and reliable eye blinking and yawn components compared with R or B channels. It allowed G-channel signals be appropriately utilized in some facial expression detection, instead of employing complicated video algorithms. The study has a good application prospect in the field of facial expression detection, such as fatigue detection, etc. We would do further exploration on the related issues in follow-up work.

Acknowledgments

This work was financially supported by National Natural Science Fund of China under Grants (61271352), Natural Science Fund of Hefei normal collage under Grants (2015QN16). The authors thank the lab colleagues who volunteered to be the subjects in this study.

References

- [1] U. Yufeng, W. Zengcai, "Detecting driver yawning in successive images", Proc. 1st International Conf. on Bioinformatics and Biomedical Engineering, pp. 581-583, 2007.
- [2] X. Fan, B. Yin, Y. Fun, "Yawning Detection For Monitoring Driver Fatigue", Proc. Sixth International Conf. on Machine Learning and Cybernetics, pp. 664-668, 2007.
- [3] T. Azim, M.A. Jaffar, A.M. Mirza, "Automatic Fatigue Detection of Drivers through Pupil Detection and Yawning Analysis", Proc. Fourth International Conf. on Innovative Computing Information and Control, pp. 441-445, 2009.
- [4] Ilkwon Park, Jung-Ho Ahn, Hyeran Byun, "Efficient Measurement of the Eye Blinking by Using Decision Function for Intelligent Vehicles", ICCS 2007, pp. 546-549, 2007.
- [5] Mandalapu Saradadevi, Preeti Bajaj, "Driver Fatigue Detection Using Mouth and Yawning Analysis", IJCSNS International Journal of Computer Science and Network Security, vol. 8, no. 6, pp. 183-188, June 2008.
- [6] Hari Singh Dhillon, Rajesh Singla, Navleen Singh Rekhi, Rameshwar Jha, "EOG and EMG Based Virtual Keyboard: A Brain-Computer Interface", Computer Science and Information Technology 2009. ICCSIT 2009. 2nd IEEE International Conference on, 2009.
- [7] B. Estrany, P. Fuster, A. Garcia, Y. Luo, "EOG signal processing and analysis for controlling computer by eye movements", PETRA'09, June 2009.
- [8] Wei Zhang, Bo Cheng, Yingzi Lin-"Driver Drowsiness Recognition Based on Computer Vision Technology" Volume 17, Number 3, June 2012 lpp354-362
- [9] X. Fan, B. Yin, Y. Fun, "Yawning Detection For Monitoring Driver Fatigue", Proc. Sixth International Conf. on Machine Learning and Cybernetics, pp. 664-668, 2007.
- [10] Y. Sun, N. Thakor, "Photoplethysmography revisited: From contact to noncontact from point to imaging", IEEE Transactions on Biomedical Engineering, vol. 63, pp. 463-477, Mar. 2016.
- [11] W. Verkruysse et al., "Remote plethysmographic imaging using ambient light," Opt. Exp., vol. 16, no. 26, pp. 21 434–21445, 2008.
- [12] J. Lee et al., "Comparison between red, green and blue light reflection photoplethysmography for heart rate monitoring during motion," in Proc. IEEE Eng. Med. Biol. Soc. Conf., 2013, pp. 1724–1727.
- [13] Y. Sun et al., "Use of ambient light in remote photoplethysmographic systems: Comparison between a high-performance camera and a low cost webcam," J. Biomed. Opt., vol. 17, no. 3, pp. 37005-1–37005-10, 2012.