IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received December 18, 2018, accepted December 23, 2018, date of publication January 10, 2019, date of current version February 4, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2891971

Driver Drowsiness Detection Using Multi-Channel Second Order Blind Identifications

CHAO ZHANG[1], XIAOPEI WU[1], XI ZHENG[2], AND SHUI YU[3], (Senior Member, IEEE)
[1]School of Computer Science and Technology, Anhui University, Hefei 230601, China
[2]Department of Computing, Macquarie University, Sydney, NSW 2109, Australia
[3]School of Software, University of Technology Sydney, Ultimo, NSW 2007, Australia

Corresponding author: Xiaopei Wu (iiphci_ahu@163.com)

This work was supported in part by the Anhui Provincial Natural Science Research Project of Colleges and Universities under Grant KJ2017A012, in part by the Open Project of Key Lab of Optic-Electronic Information Acquisition and Manipulation Ministry of Education, Anhui University, under Grant OEIAM201401, and in part by the Ph.D. Research Startup Foundation of Anhui University.

ABSTRACT It is well known that blink, yawn, and heart rate changes give clue about a human's mental state, such as drowsiness and fatigue. In this paper, image sequences, as the raw data, are captured from smart phones which serve as non-contact optical sensors. Video streams containing subject's facial region are analyzed to identify the physiological sources that are mixed in each image. We then propose a method to extract blood volume pulse and eye blink and yawn signals as multiple independent sources simultaneously by multi-channel second-order blind identification (SOBI) without any other sophisticated processing, such as eye and mouth localizations. An overall decision is made by analyzing the separated source signals in parallel to determine the driver's driving state. The robustness of the proposed method is tested under various illumination contexts and a variety of head motion modes. Experiments on 15 subjects show that the multi-channel SOBI presents a promising framework to accurately detect drowsiness by merging multi-physiological information in a less complex way.

INDEX TERMS Yawn, blink, blood volume pulse (BVP), drowsiness detection, second-order blind identification (SOBI).

I. INTRODUCTION

It is widely accepted that drowsiness (i.e., momentary nodding off) has become one of the main causes for traffic accidents [1]. Some reports have revealed that almost millions of people were injured in traffic accidents every year all over the world and serious traffic accidents caused by driver fatigue and drowsiness became increasingly common in recent years [2], [3]. There are too many factors that may cause drowsiness or fatigue in driving including long time driving, lack of sleep, hypnotized by monotonous driving pattern and circumstance, etc [4]. Drowsy driving is now regarded to be as dangerous as drunk driving [5].

Nowadays intelligent automotive systems have gained increasing popularity and many vehicle companies have produced smart vehicles. For instance, Tesla has launched intelligent cars [6] which are equipped with autonomous features like autopiloting, path planning and remote control *et al.* One of the key objectives of these features is to assist and protect drivers [7]. In order to achieve these objectives, it is essential, first of all, to recognize driver's driving state to provide autonomous safety measures and provide intuitive people-vehicle interaction. Because of this reason, more and more researches have been carried out to develop methods and devices that help to identify the drivers' condition and give some proactive interventions as part of the autonomous features provided in the vehicle systems [8], [9]. Since signals that are generated by daily physical activities can be used to determine the physiological or psychological state of a human, most of the existing drowsiness detection algorithms/systems are based on bio-electrical signal processing [10] or image/video processing [11] to measure the driver's physiological and behavior information. However, many of the existing drowsiness detection methods are prohibitively complex in terms of algorithm and required devices. Even if we don't consider the complexity of drowsiness detection algorithms, the existing methods often rely on nodes attached on driver's body [12] or use cameras connected to a PC [13] to get the original data. Those wearable sensors and required complex wire connections make these systems impractical for daily use.

The usage of smart phone greatly reduces the complexity of video acquisition, store and transmission [14]–[16]. However, since driver's driving environment is highly dynamic, most of the smart phone based algorithms are very susceptible to the dynamics in driving environment because the mobile platform cannot afford the computation and storage requirement for very sophisticated algorithms. The existing algorithms in mobile devices need further improvement to identify the driver's driving state in a more robust way.

In recently years, a strategy based on fusion of multiple information was proposed and improved detection results were reported [17]–[19]. Driver's multiple physiological information can be extracted and combined from the data collected by multiple contact/non-contact sensors, which brings an effective way to obtain a more robust detection performance comparing to those algorithms focusing on a single physiological source. For instance, yawn is an obvious signal of being sleepy and drowsy in everyday life. Besides yawn, studies have also concluded that blink frequency/duration and heart rate (HR) are very closely related to drivers' drowsy state [20], [21]. A more reliable estimation of people's drowsy state could be obtained if yawn, blink and heart rate information can be integrated together to make a combined analysis. However, most of the existing data fusion methods were implemented by directly stacking several algorithms on top of each other to build a multi-dimensional algorithm [22], [23]. This kind of fusion will definitely increase the complexity of the fusion method. If multiple information can be obtained in the same algorithm flow, the complexity of the fusion algorithm for gaining a better detection performance will be much lower.

Second-order blind identification (SOBI) [24] is a representative algorithm under the framework of independent component analysis (ICA) to solve the problem of blind source separation (BSS) [24]. In recent years, ICA has been extensively used as a typical tool for multi-source separation. ICA utilizes high order statistics to recover a set of independent sources from their linearly or nonlinearly mixed observations with the assumption that the sources are statistically independent. Treating multiple physiological information as mutually independent sources, SOBI has great potentialities to provide a balance between acceptable performance and low algorithm complexity by simultaneously extracting the multiple physiological information from the observations.

Our goal is to present a simple yet efficient framework to detect driver's drowsy state. The contributions of our work can be briefly summarized as follows:

1) A framework for simultaneously extracting multiple-physiological information from facial videos is proposed. With the help of multi-channel SOBI, multiple outputs corresponding to different sources are obtained at the same time. Compared to previous fusion methods, our approach is simpler in algorithm structure (See Figure 1);

2) Due to ICA's fundamental limitation of permutation ambiguity, a source identification algorithm combining short-term energy and spectral kurtosis is also proposed to discriminate target signals from the outputs;

3) Drowsiness is finally determined by analyzing the extracted yawn, blink and BVP signals in parallel and making a compound determination. Benefiting from merging the multiple information, the proposed method performs more stably to complex background dynamics in vehicles compared to methods based on one single source.

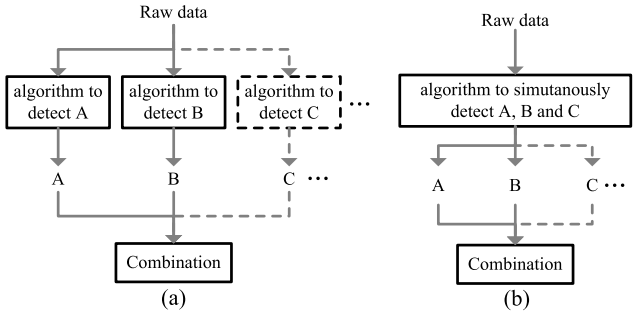


FIGURE 1. Comparisons of the algorithm structures (a) Algorithm structure of the state of the art fusion methods [18], [22], [23]. (b) Algorithm structure of the proposed method.

The rest of this paper is organized as follows. In Section II, we shortly describe some highly relevant work. Next, in Section III, we briefly introduce the key concepts of ICA and SOBI. Section IV gives the details of the proposed method. In Section V, we make a large number comparative experiments and quantitative evaluations to prove the feasibility and advantage of the proposed method. Section VI gives the limitations of the proposed method and proposes some extensions for future work. Finally, a short conclusion summarizes the paper in Section VII.

II. RELATED WORK

Yawn is the most obvious signal of drowsy state and it can be easily detected by visually based method [25], [26]. As for detecting blinks for drowsy state, the most reliable detection method is the electrocardiogram (EOG), which is a measurement of eye bio-electricity and has been widely used to detect drowsiness [27], [28]. However, a set of contact electrodes used for EOG acquisition makes this kind of method uncomfortable and inconvenient for drivers in real driving scenarios. In [29] and [30], studies have revealed that the heart rate (HR) varies remarkably between awake and drowsy state. So heart rate variability (HRV), which refers to the variation of time interval between heartbeats, can be used to detect drowsiness [31], [32]. Electrocardiogram (ECG), which often gains from a pulse oximeter, is thought to be the most accurate way to estimate HR and HRV [33]. But the contact electrodes used in the ECG acquisition also bring too much limitation in the applications of drowsiness detection.

Traditional bio-electric signal acquisition methods rely on a set of sophisticated equipments which are often in direct contact with human skin. That brings both complex operations and uncomfortable experience for the users. Unlike the

contact methods, visual based methods have gain popularity because it provides a noncontact way of sensing physiological state without the usage of electrodes. Visual information of yawn, eye blink, and HR can be monitored from facial videos to determine the driving state. Some previous researches have contributed to this work [34]–[37].

Photoplethysmography (PPG) is a non-invasive and low-cost optical technique that can be used to detect cardiovascular blood volume pulse (BVP) signal in the microvascular bed of tissue [38]. PPG presents pulsatil waveform attributed to cardiac synchronous changes in the blood volume with each heart beat. Based on the detected BVP signal, heart rate [39] and some other features such as oxygen saturation and respiratory rate [40], [41] can be extracted. Previous studies focused mostly on BVP, PPG based blink and yawn detection are rarely reported. Over the last years, in contrast to original PPG technology, new PPG algorithms, called imaging PPG (iPPG) [42] or camera-based PPG (cbPPG) [20] (hereinafter called extended-PPG), became more popular due to the rapid development of powerful image sensors. Extended-PPG is capable of monitoring the changes of the image sequence without concerning the specific content. In extended-PPG, daily used cameras are usually adopted as optical sensors and no special illumination precondition is required. An image stream containing a variety of physiological activities will be recorded. Physiological processes recorded in the video will directly cause the pixel value to change in specific modes, hence can be extracted and identified. Although BVP gained much attention, the fluctuation of pixel value caused by breath, eye blink, yawn or some other kinds of physiological activities also forms pulsatile signals [43], which can be identified and measured in the same way.

References [44] and [45] first proposed to extract the heartbeat signal and estimate the heart rate (HR) from image sequences. Pelegris *et al.* [46] revealed that video-based HR estimation has approximate accuracy compared to methods based on pulse oximeter. Poh *et al.* [47] employed ICA [48] to de-mix the heart beat signals from facial video recordings. In his method, the R, G, and B component of the video forms a 3-channel observation signal which is treated as a mixture of three independent source signals. The BVP signal can be finally found in one of the three output channels. His work is very encouraging that it demonstrates the possibility of ICA to extract BVP signals from facial videos. As a modified version of Poh's method, Daniel *et al.* proposed remote BVP detection (at a distance of 3 m) [49] using 3-channel ICA. That distance covers almost all the possible application scenarios in daily life, including in-car scenarios for our work. In a recent report on ICA based PPG method, Alghoul *et al.* [50] compared ICA with eulerian video magnification (EVM). Their work shown that ICA and EVM have their own advantages and disadvantages in different frequency bands. But they ignored a fact that ICA has much stronger potential in extracting more sources from the observed signals.

In a previous work [51], we have confirmed that extended-PPG can not only detect heartbeat signals but also detect blinks. In this work, we focus on detecting drowsiness in driving scenarios by using SOBI based extended-PPG to simultaneously extract BVP, blink and yawn signals from smart phone videos. A compound judgment is made by analyzing the separated multiple source signals in parallel to determine driver's driving state in a more robust way.

III. PRELIMINARY

Independent component analysis is a multi-dimensional probabilistic analysis method for searching a linear/nonlinear transform to recover components that are maximally independent to each other [52], [53]. The commonly used linear mixing model of ICA is

$$\mathbf{x} = \mathbf{A}\mathbf{s} \tag{1}$$

where $\mathbf{s} = [s_1(t), \ldots, s_N(t)]^{\mathrm{T}}$ is a random vector representing the statistically independent sources. $\mathbf{x} = [x_1(t), \ldots, x_M(t)]^{\mathrm{T}}$ is the mixture signal which is obtained by multiplying $\mathbf{s}$ with a $M \times N$ scalar mixing matrix $\mathbf{A}$. M is the number of channels and N is the number of the original independent sources. (To simplify the expression, we ignore the noise that may exist in the real world.) Usually, the mixture signal will be normalized below:

$$z_i(t) = \frac{x_i(t) - \mu_i}{\sigma_i} \tag{2}$$

where i means the *ith* component of $\mathbf{x}$, μ_i and σ_i are the mean and standard deviation of $x_i(t)$ respectively. The task of ICA is to estimate the source signal $\mathbf{s}$ by adaptively learning the de-mixing matrix $\mathbf{W}$ (the inverse of the mixing matrix $\mathbf{A}$). The source signal $\mathbf{s}$ can be recovered by simply multiplying the observation signal with the de-mixing matrix, i.e.

$$\hat{\mathbf{s}} = \mathbf{W}\mathbf{z} \tag{3}$$

where $\hat{\mathbf{s}}$ is the estimation of the source signal $\mathbf{s}$. To learn the de-mixing matrix, ICA maximizes the statistical independence of the potential independent components in each iteration.

As an important branch of ICA, second-order blind identification (SOBI) is employed in our work to build a framework in which multiple physiological processes can be simultaneously detected. A detailed analysis of SOBI is given in [24]. As a necessary pre-processing step of SOBI, the observation signal is firstly whitened by applying $\mathbf{x}(t)$ to a whitening matrix $\mathbf{V}$. To estimate the whitening matrix, covariance matrix R(0) with no time delay of $\mathbf{x}(t)$ is first computed. The whitened signal $\mathbf{z}(t)$ can be computed by

$$\mathbf{z}(t) = \mathbf{V}\mathbf{x}(t) \tag{4}$$

The whitening matrix $\mathbf{V}$ in (4) has the following structure:

$$\mathbf{V} = diag(\lambda_i^{-1/2})\mathbf{U}^{\mathrm{T}} \tag{5}$$

where $\lambda_i (i = 1, \ldots, N)$ are the eigenvalues of R(0) and $\mathbf{U}$ is the matrix whose columns are the corresponding eigenvectors. Rather than using a unique covariance matrix, SOBI

calculates a set of covariance matrices with different time delays:

$$\mathrm{R}(\tau) = E[\mathbf{z}(t+\tau)\mathbf{z}^T(t)]\tau \in \{\tau_j | j = 1, \ldots, K\} \quad (6)$$

An unitary matrix $\mathbf{B}$ is obtained by applying joint approximate diagonalization [54] to $\mathrm{R}(\tau)$, the de-mixing matrix $\mathbf{W}$ is finally estimated as $\mathbf{W} = \mathbf{B}^{-1}\mathbf{V}$.

IV. METHODS

Researchers have developed methods to monitor drivers' level of vigilance and alerting drivers when they are not paying adequate attention to the driving. In Section I, we have mentioned that merging different sources of information is a way to increase the reliability of drowsiness detection. Some of the recent works improve the accuracy and reliability of the existing methods in this way. But most of this kind of fusion techniques are implemented by directly stacking different algorithms on top of each other and often bring unacceptable computational complexity [17]–[19], [22], [23]. Our solution is to extract and combine multiple information more naturally into one integrated algorithm thus effectively reducing the complexity while preserving high accuracy of detection.

In this paper, with the help of SOBI, we propose a framework in which the BVP, blink and yawn analysis can be obtained and processed in parallel to detect drowsiness.

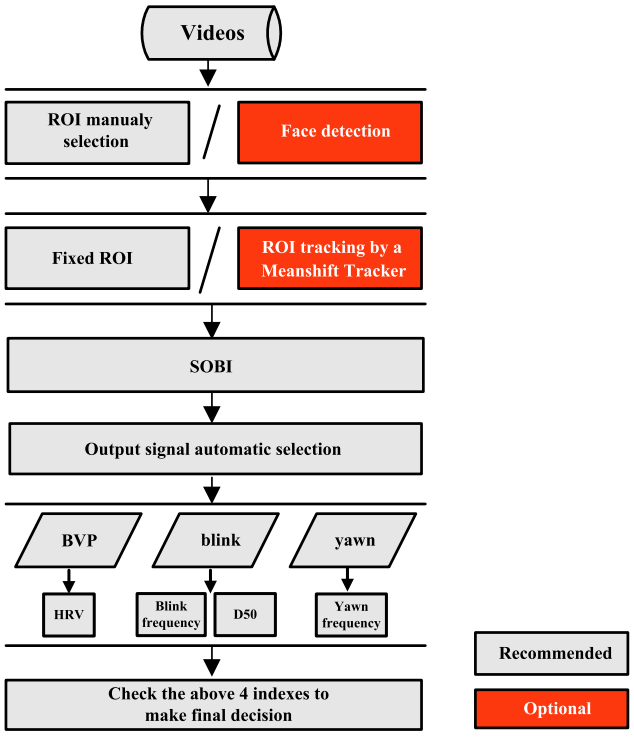


FIGURE 2. Framework of the proposed algorithm.

Figure 2 presents the flow chart of the proposed framework. ROI (Region of Interest) selection is the first step. Face detection and Meanshift tracker are optional to gain a more stable ROI with the cost of increased complexity. Then SOBI is used to estimate the underlying sources corresponding to different physiological activities. Following the SOBI, a source identification algorithm automatically distinguishes the yawn, blink, and BVP signals from the multiple outputs. Further analysis is made to estimate the HRV, blink duration, blink frequency, and yawn frequency in parallel. To make the proposed method as simple as possible, drowsiness is finally determined if any of the above-mentioned three positive results are detected. We will walk through how we use smart phones to acquire video streams in Section IV-A, explain how we generate multi-channel input signal in Section IV-B, illustrate how we identify difference sources from the output signals in Section IV-C, and explain how we extract drowsiness-related features respectively based on BVP, blink and yawn signals to finally make a determination of drowsiness in Section IV-D.

A. SMART PHONE BASED VIDEO ACQUISITION

Traditional PPG requires special light sources and devices, which makes the systems too sophisticated for daily use. With the development of modern electronic technology, some new devices, such as LED light source and high speed cameras were employed to make PPG more convenient to use [29], [55]. But users still have to consider the process of data storage and transmission. Today, smart phones have gained wide popularity. They have powerful CPU and optical sensors to capture and store high quality video streams. Smart phones have significantly reduced the complexity of video acquisition. In our work, all the devices we need for data acquisition are ordinary smart phones. The front cameras of the smart phones are used to capture subjects' facial videos so that the subjects can easily adjust their face positions by themselves. The distance between the smart phone and the subject's face is generally controlled within the range of 15-45 cm to make sure that the facial videos are as clear as possible. That distance can be set to a bigger value in the automotive scenario to have fewer disturbances to the driving tasks. For the smart phones with optical zoom, the distance can be set even bigger to 50 cm to 80 cm.

The demonstration of video acquisition is given in Figure 3. Figure 3 (a) and (b) show the demonstration of video acquisition in in-car and in-lab environment. A facial area that contains eyes and mouth is selected as the ROI, as shown in Figure 3 (c). Figure 3 (d) shows the in-lab driving simulation system in which we used to collect data and develop the algorithms. The videos recorded using the in-lab simulation system is more convincing than videos recorded by letting the users to imitate driving state by themselves. Meanwhile, a serial of sensors were used to synchronously collect nonvisual physiological data (EOG, pulse signal) which serves as reference signals (see Figure 3 (d)).

In the video recording of our experiment, subjects are allowed to have spontaneous head motions. The angle of the subject's facial area facing the smart phone is not strictly limited. But the subject should make sure that his/her eyes and mouth must be covered in the ROI. The ROI in our

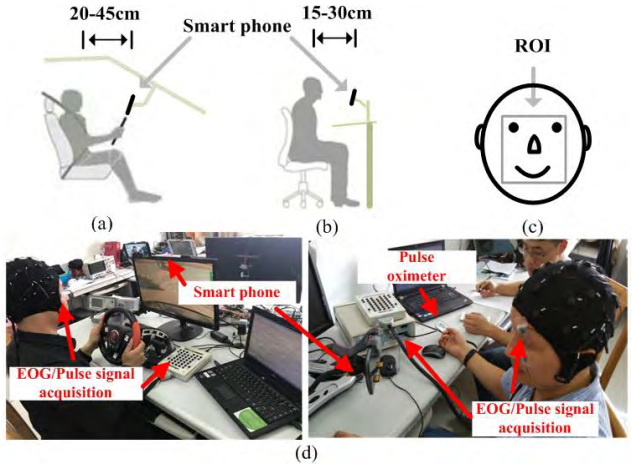


FIGURE 3. Demonstration of the video acquisition. (a) Illustration of in-car scenario for video acquisition. (b) Illustration of video acquisition in lab scenario. (c) Illustration of facial region selected as ROI. (d) Simulation of in-car driving environment with multiple sensors synchronously recording multiple physiological signals.

method can be either manually selected for subsequent calculations or fixed by an automatic face detection algorithm applied to identify the facial area [56]. Meanshift tracker [57] can be used to cover the current position of the face in each frame after ROI being firstly determined. In practice, as long as the subject's movement is not very strenuous, the algorithm gets acceptable performance even without a Meanshift tracker. The Meanshift algorithm that can be used as a face tracking module will give the capability to deal with severe vibrations in real driving scenario and bring more comfortable user experience to the monitored drivers.

B. MULTI-CHANNEL MIXTURE SIGNAL GENERATION

The number of channels selected in ICA will, to some extent, determine what ICA's separated outputs are and whether the outputs are good estimations of the original independent sources. It is well known that in RGB model, data can be naturally divided into three channels. From the perspective of ICA, this three-channel data can be treated as the observed mixture signals and there are three source signals to be estimated. However, if we assume that the observation is composed of more underlying signals (which is very common in real life), then this channel number (i.e., 3) would not be enough.

Figure 4 gives the waveforms of observed mixture signals reflecting the pixel value fluctuation caused by yawns, blinks and BVP in different facial regions. Peaks corresponding to yawns and blinks are labelled with "Y" and "B" respectively in Figure 4. The waveforms also contain BVP component which appears as tiny peaks in Figure 4. In fact, many other physiological processes that can bring about changes in pixel values occur simultaneously under facial skin, for example, respiratory, blood oxygen saturation, or other electromyographic (EMG) signals. The number of potential independent sources is definitely more than three. If you use a small number of channels in a multi-source separation task,

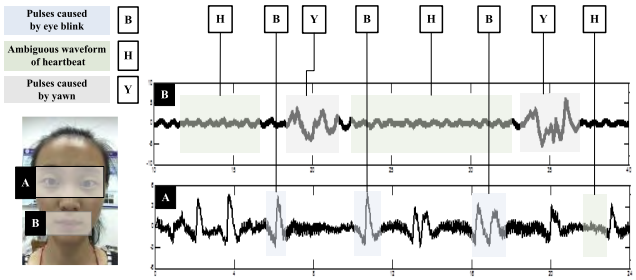


FIGURE 4. Potential sources in observations.

the separated signals would be treated more like another mixture of sources than a separated independent source [58]. Since extracting more sources from the user's facial region is required in our work, a 3-channel input data would not be enough. More channels of data should be properly involved to extract more physiological sources from the facial videos.

In this paper, we take 9-channel SOBI to simultaneously detect blink, yawn and BVP signal. Although 9-channel input will increase the computation load, the outputs, however, are much more rewarding and less noisy than 3-channel outputs because there are enough channels for the potential sources in the inputs to be identified and separated into independent sources that corresponds to different physiological sources. Since the data in SOBI's each channel is one-dimensional, it is necessary to convert each ROI into a (set of) one-dimensional vector to construct the input for SOBI. In 9-channel SOBI, the ROI in each frame (with the size of $M \times N$) is evenly divided into three parts from 1/3 and 2/3 of the height. Each part produces three values corresponding to the R, G, and B components by spatially averaging all the pixels in the part. If the video contains T frames, we have:

$$\mathbf{x}=\left\{\begin{array}{ll}\frac{1}{3MN}\sum_{x\in R}x_{i,j}(t) & i=1,...,M/3;\ j=1,...,N\\ \frac{1}{3MN}\sum_{x\in G}x_{i,j}(t) & i=1,...,M/3;\ j=1,...,N\\ \frac{1}{3MN}\sum_{x\in B}x_{i,j}(t) & i=1,...,M/3;\ j=1,...,N\\ \frac{1}{3MN}\sum_{x\in R}x_{i,j}(t) & i=M/3+1,...,2M/3;\ j=1,...,N\\ \frac{1}{3MN}\sum_{x\in G}x_{i,j}(t) & i=M/3+1,...,2M/3;\ j=1,...,N\\ \frac{1}{3MN}\sum_{x\in B}x_{i,j}(t) & i=M/3+1,...,2M/3;\ j=1,...,N\\ \frac{1}{3MN}\sum_{x\in R}x_{i,j}(t) & i=2M/3+1,...,M;\ j=1,...,N\\ \frac{1}{3MN}\sum_{x\in G}x_{i,j}(t) & i=2M/3+1,...,M;\ j=1,...,N\\ \frac{1}{3MN}\sum_{x\in B}x_{i,j}(t) & i=2M/3+1,...,M;\ j=1,...,N\end{array}\right\}=\begin{bmatrix}x_{R1}(t)\\ x_{G1}(t)\\ x_{B1}(t)\\ x_{R2}(t)\\ x_{G2}(t)\\ x_{B2}(t)\\ x_{R3}(t)\\ x_{G3}(t)\\ x_{B3}(t)\end{bmatrix} \tag{7}$$

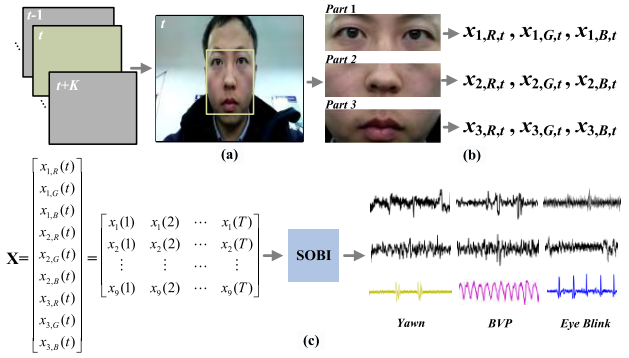


FIGURE 5. Overview of the proposed method to extract the yawn, BVP, and blink waveforms. (a) Original image and the selected ROI. (b) ROI partition to obtain a 9-channel input signal. (c) Multiple physiological sources are extracted from the 9-channel data using ICA. The yellow, red and green signals correspond to yawn, blink and BVP signals, respectively.

Figure 5 shows an overview of our method. Figure 5 (b) provides an explanation of ROI partition. Figure 5 (c) describes the separation process of 9-channel SOBI. Especially, we want to point out that the specific number of channels is, to some extent, task-dependent and even subject-dependent. Any two of the three parts shown in Figure 5 (b) can be associated to form 6-channel observation which leads to lower computational load and obtains an even better performance. But here in our task of yawn, blink and BVP detection, a multi-channel observation is set to cover all the 3 parts.

C. IDENTIFICATION OF THE SEPARATED SOURCES

After many rounds of iterations of SOBI, the output signals finally become mutually independent, they are therefore considered to be the estimation of the original independent sources. Since ICA has a fundamental limitation that the order of its outputs is undetermined. There is not a consistent one-to-one relationship between each output channel and each ICA separation [48]. In off-line circumstance, you can simply select the yawn, BVP and blink waveform by visual inspection. But in on-line circumstance, a source identification algorithm is needed to automatically select the target sources from the multi-channel outputs for drowsiness detection. Here, we use short-term energy with spectral kurtosis to automatically extract separated yawn, blink and BVP signals from the outputs. Short-term energy is to measure the absolute energy of the outputs in a sliding window. The size of the sliding window is set to be one to two times of the frame rate depending on individual subject. Generally speaking, yawn seldom appears in awake state, even in drowsy state, its occurrence is not so frequent comparing to blink and heartbeat. So in most cases, besides several scattered pulses corresponding to yawn, most of the points in the separated yawn waveform are close to zero if they are correctly separated. That brings advantage to distinguish yawn from the multi-channel outputs. The channel which yields most minimal short-term energy is the channel that outputs yawn signal. Figure 6 presents a demonstration of 9-channel short-term energy calculated for yawn signal identification. The red channel, which yields most of the value close to zero, is the channel yawn exists.

The kurtosis [48] of random variable z is defined by:

$$kurt(z) = \frac{E(z^4)}{[E(z^2)]^2} - 3 \tag{8}$$

Here z is zero-mean. Kurtosis is often used to measure the nongaussianity of a random variable. It is a reflection of a signal's pulse characteristic. The spectrum of well estimated BVP signal has a distinct peak at the position corresponding to HR, hence can be easily distinguished using spectral kurtosis. Meanwhile, people's blink signal in drowsy state has a set of sharp pulses, which brings a large value of time-domain kurtosis.

Figure 7 shows the demonstration of blink and BVP signal identification using spectral kurtosis and time-domain kurtosis. The highest value of time-domain kurtosis observed in *Ch4* and the highest values of spectral kurtosis observed in *Ch6* indicated that the outputs in *Ch4* and *Ch6* are individually blink and BVP signals. Pre-processing such as band-pass filtering is often needed before ICA to remove the interference of noise and baseline drift. After the sources were automatically identified, filtering is still needed as post-processing to modify the waveforms for further calculation of the drowsiness-related features.

It should be noted that for online mode, automatic identification of the target sources is very important. The effectiveness of the proposed fusion framework is fundamentally based on successfully identifying the blink, yawn, and BVP signals respectively from the SOBI's 9-channel output. The proposed method will be partially or even entirely invalid if not all the target sources are correctly identified. For example, if one of the three target signals is not successfully identified, the proposed method will degenerate into a drowsiness detection algorithm based on two kinds of physiological information. Even more, if two of the signals are not correctly detected, the proposed method will probably degenerates into a drowsiness detection method based on single physiological parameter.

D. FEATURES CALCULATION FOR DROWSINESS DETERMINATION

The separated BVP can be considered as the substitution for ECG, which is the base for HRV analysis. HRV is often evaluated using time-domain methods or frequency-domain methods. In this paper, we choose frequency domain methods. Frequency domain methods transfer the signals from time domain to frequency domain by FFT to get the power spectral density (PSD).

Researchers have confirmed that the low frequency (LF) component (0.04-0.15 Hz) of the PSD is influenced by both the parasympathetic activity (e.g., neural activities caused by strenuous sports exercise) and sympathetic activity (e.g., neural activities related to peacefully spontaneous breathing) [59] whereas the high frequency (HF) component (0.15-0.4 Hz) is

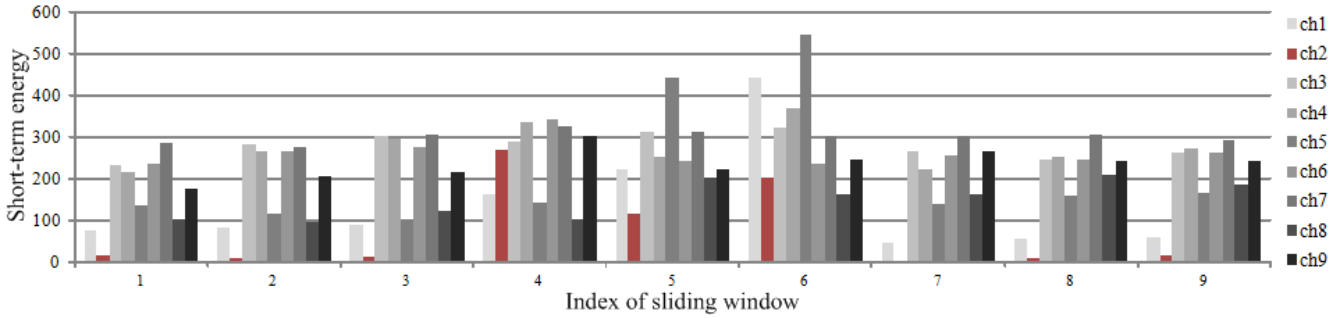


FIGURE 6. Demonstration of short-term energy for yawn signal identification.

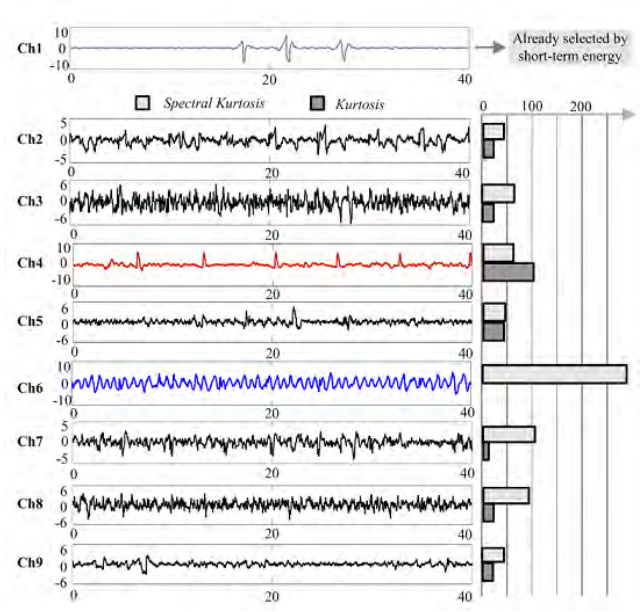


FIGURE 7. Demonstration of blink and BVP signal identification by (time-domain) kurtosis and spectral kurtosis.

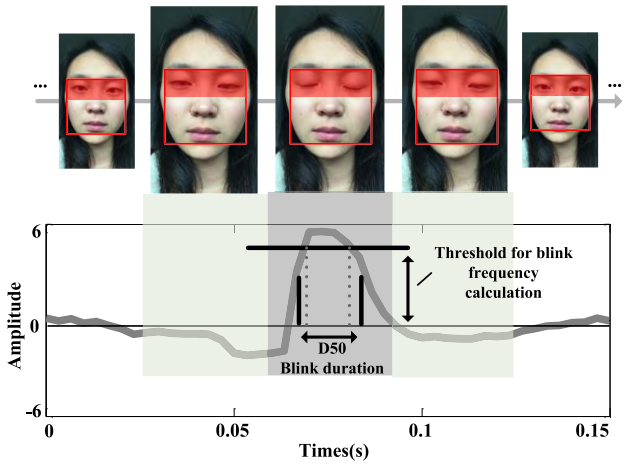


FIGURE 8. Calculation of the blink frequency (BF) and blink duration (D50).

influenced by the parasympathetic activity [60]. So the PSD is divided into low and high frequency bands. For drowsiness detection, the LF to HF ratio is really concerned because it decreases when a person becomes sleepy [31].

Blink related information such as blink duration, blink frequency, PERCLOS [61], [62] are often used to detect drowsiness. When a person is becoming drowsy, his/her eyes will extend the eyelid closure duration involuntarily to protect the eyes and gain more relaxation. So in this work, we focus on long blink durations as well as blink frequency changing to detect drowsiness. We use two features [37]: duration at 50% (D50, the duration of the amplitude higher than 50% of the maximum value) and the blink frequency (BF, the number of blinks in a specific time period).

The separated blink signals have standard pulse shape (see Figure 8) hence bring convenience for D50 and blink frequency calculation. The calculation of D50 and BF is illustrated in Figure 8. In order to precisely count the blinks, the blink pulses are converted into square wave using a threshold. All the data points above the threshold were set to 1 and all the data points below the threshold were set to 0. An open mouth will have a larger visible ''black hole'' comparing to a closed mouth. Hence yawn can be detected by measuring the changing of averaged pixel value in the mouth region. Like blink detection mentioned above, yawn detection in our work is performed in two steps: in the first step we obtain the average pixel value fluctuation caused by yawn by SOBI. In the second step, we estimate the yawn frequency from the separate pulse-like signals.

V. EXPERIMENTS AND DISCUSSION

A. EXPERIMENTAL SETUP

The algorithm was firstly developed and evaluated in a simulated driving environment in the lab. Then, in-car data was collected. To avoid traffic accidents that are likely to happen in real-world for drowsy participating drivers (the participants' sleeping periods are controlled), the in-car data was collected without moving the car. The participants were asked to sit in the car and monotonously step on the brake/gas pedals by turns. Meanwhile, we played hypnotic music in the lab and car to create a drowsy air for the participants.

We collected facial videos from 15 participants (10 males and 5 females) at different times and under different

illumination conditions. Their ages range from 20-35. All the participants have regular sleep pattern and have no sleep disorder. None of them was addicted to drugs or alcohol. The participants were requested to sleep less than their daily sleeping time to increase the chances of becoming drowsy during their video recording. Each of the participants was taken at least 3 videos and the duration of the recorded videos ranges from 10 minutes to 60 minutes. Finally, 1400 minutes of useful videos from participants were selected. The participants were asked to label their own videos for the drowsy and alert part. All the videos used in the experiment were collected from two mobile phones: an iphone 6 and a LG G3. The frame rate of the videos is 30 fps and the resolution is 1280 × 720. The participants were told to relax themselves and sit as naturally as possible while recording the videos. We allowed the participants to have involuntary facial movements and normal head motions. But subjects should make sure that his/her two eyes and mouth were covered in the video. All the facial videos were analyzed offline in MATLAB. Four indicators, namely detection rate (*Sensitivity*,*DR*), false alarm rate (*FAR*), true negative rate

(*Specificity*, *TNR*), and negative predictive rate (*NPR*) [63] are used to make a quantitative analysis for the detection results:

$$DR = \frac{TP}{P} \tag{9}$$

$$FAR = \frac{FP}{TP + FP} \tag{10}$$

$$TNR_F = \frac{TN_F}{TN_F + FP_F} \tag{11}$$

$$NPR_F = \frac{TN_F}{TN_F + FN_F} \tag{12}$$

where TP is the number of true positives, P is the number of all positives that manually labeled and counted by the participants themselves. FP is the number of negatives that are falsely detected as positives. TN is the number of detected true negatives. FN is the number of positives that are falsely detected as negatives. The subscript F in (11) and (12) means frame-level analysis. The calculation results corresponding to the three physiological signals are averaged to get the final values of the indicators. Since the above four indicators can not be clearly defined in HRV estimation (HRV is a dynamic value calculated by measuring a set of heartbeat pulses in a long term period), the peaks in the separated BVP waveform are treated as discrete events like blink and yawn for quantitative evaluation.

TABLE 1 gives the parameters of SOBI and TABLE 2 gives the thresholds used in the experiments. EOG, pulse wave as well as videos were synchronously acquired and compared to validate the proposed method.

B. RESULTS OF THE PROPOSED METHOD AND THE COMPARISON WITH STANDARD METHODS

We compared our method with standard methods to make sure that the signals extracted by SOBI from videos are

TABLE 1. Parameters of SOBI.

Number of time delayed correlation matrices	200
Number of sources	9
Number of sensors	9
Threshold for convergence error	2.5×10^{-4}
Zero mean the data	needed
Pre-whiten the data	needed

TABLE 2. Thresholds used for drowsiness detection.

Yawn detection	Blink detection		HRV calculation
YF	D50	BF	LF/HF
0.1Hz	120ms	0.1Hz	1.2

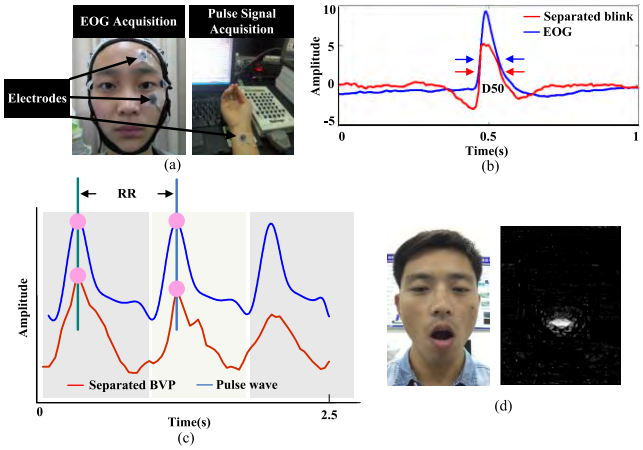
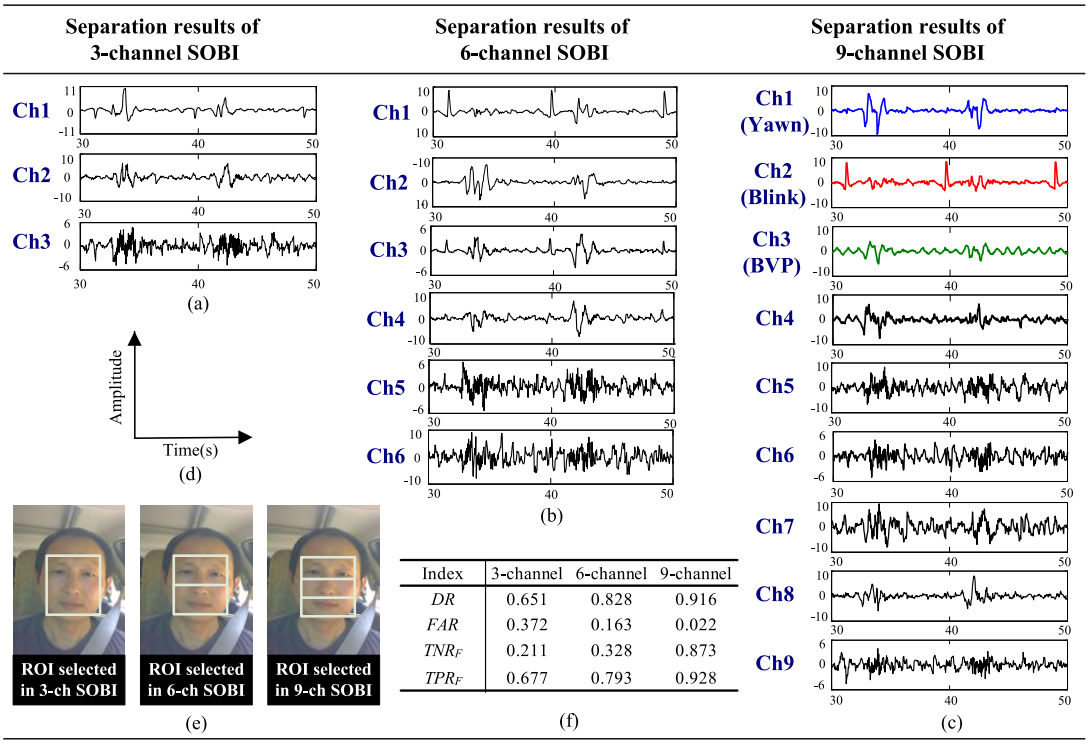


FIGURE 9. Comparisons of BVP, blink, and yawn detection between the proposed method and other state of the art methods. (a) Electrode placement for EOG and pulse wave acquisition. (b) Waveforms of separated blink signal and EOG. (c) Waveforms of separated BVP and pulse wave. (d) Yawn detected by visual method.

correct and can be treated as substitution of standard physiological signals. Figure 9 (a) shows the electrode placement for EOG and pulse wave collection. Waveforms separated by multi-channel SOBI and acquired by bio-medical apparatus [64], [65] are compared in Figure 9 (b) and (c). Separated Waveform of SOBI is similar to the waveform collected by bio-medical apparatus both in shape and peak positions, hence resulting in the same HRV, D50 and blink frequency estimations. All yawns happened in the videos were firstly detected using the visual method depicted in [66] and then manually checked. Figure 9 (d) shows the result of yawn detection by the visual method. The method is capable of providing us with an accurate ground truth to ensure that the yawn detected by our algorithm is reliably evaluated.

C. COMPARISON OF SOBI WITH DIFFERENT NUMBERS OF CHANNELS

The more channels used in our work is the key to realize multiple source simultaneous separation. To validate the performance advantages brought by a big channel number,



Index	3-channel	6-channel	9-channel
DR	0.651	0.828	0.916
FAR	0.372	0.163	0.022
TNR_F	0.211	0.328	0.873
TPR_F	0.677	0.793	0.928

FIGURE 10. Results of 3-channel, 6-channel, and 9-channel SOBI. (a) Results of 3-channel SOBI; (b) Results of 6-channel SOBI; (c) Results of 9-channel SOBI; (d) Description of the coordinate system; (e) ROIs used in the experiment; (f) Average *DR*, *FAR*, TNR_F and TPR_F of different channel numbers.

we compared the results of detecting yawn, BVP and blink by utilizing 3-channel, 6-channel, and 9-channel SOBI as shown in Figure 10. Figure 10 (a) and (b) are the results corresponding to SOBI using 3-channel data and 6-channel data. Figure 10 (c) gives the results of 9-channel SOBI. The signals shown in blue, red, and green are extracted yawn, blink and BVP signals respectively in Figure 10 (c). Although blink and yawn can be distinguished by visual inspection in the results of 3-channel SOBI and 6-channel SOBI, the outputs are mixture of the sources rather than ''clean'' source signals. Moreover, 3-channel SOBI and 6-channel SOBI have difficulties in extracting BVP signal. In contrast, the blink and yawn signals separated by 9-channel SOBI are more distinguishable and the separated BVP waveform is more standard which creates good condition for HR and HRV calculation. Although the separated sources of 9-channel SOBI are not completely ''clean'', the signals are much easier to extract comparing to those of 3-channel SOBI and 6-channel. Figure 10 (f) presents the average *DR*, *FAR*, TNR_F and TPR_F of the proposed method with different channel numbers on drowsy detection. From the result, it is clear that although a strategy of combining different sources was used, a small channel number cannot bring an acceptable performance.

D. COMPARISON OF DIFFERENT ICA ALGORITHMS

Different ICA algorithms are suitable for different application scenarios. The specific ICA algorithm adopted in our method will to some extent determine the ''quality'' of the separated signals, hence affecting the estimation accuracy of the physiological features (parameters). Otherwise, adopting an inappropriate ICA algorithm will deteriorate the detection performance. A main contribution of our work is to prove the effectiveness of SOBI based multi-channel ICA algorithm to separate bio-related sources. We compared our proposed approach against three state of the art ICA algorithms, namely FastICA [67], Infomax [68], JADE [69].

Figure 11 (a) - (d) show the waveforms derived from different ICA algorithms with the same channel number. It is clear that SOBI's separation results have the best quality. Although the quality of blink signals provided by the other three algorithms is also acceptable, the separated yawn and BVP signals have too much noise and hence look significantly different from those recovered by SOBI. The waveforms of inaccurately estimated yawn and BVP signal are deformed and contain too much noise which will lead to a reduction in the accuracy of yawn detection and HRV calculation.

Figure 12 presents the *DR*, *FAR*, TNR_F, and TPR_F for the abovementioned four ICA algorithms calculated from 9 subjects. The distances between the maximum and minimum values indicate the fluctuations on the indexes. It is obvious that SOBI approach presents a more stable and accurate result as it has the highest *DR*, TNR_F, TPR_F, and the lowest *FAR* In this task, the other three state of the art ICA algorithms perform not as great as SOBI approach.

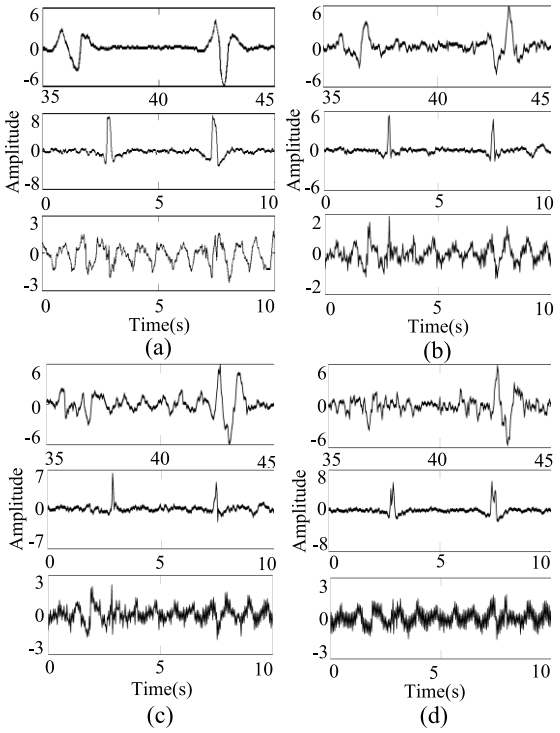


FIGURE 11. Results comparison of four different ICA algorithms.
(a) Waveforms of yawn, blink and BVP separated by SOBI (Our solution).
(b) Waveforms of yawn, blink and BVP separated by FastICA.
(c) Waveforms of yawn, blink and BVP separated by Infomax.
(d) Waveforms of yawn, blink and BVP separated by JADE.

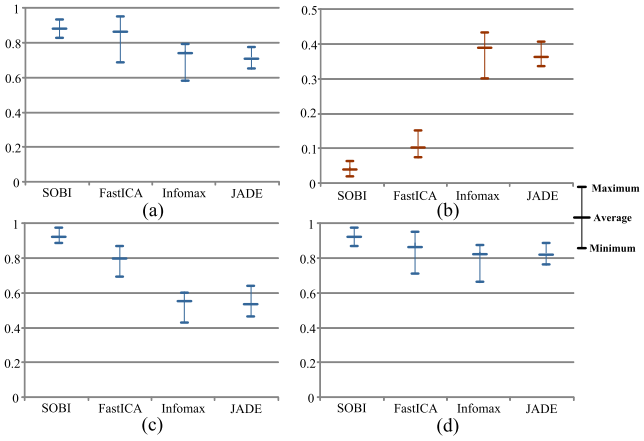


FIGURE 12. *DR, FAR,* TNR_F, and TPR_F of different ICA algorithms.
(a) *DR* of the four ICA algorithms. (b) *FAR* of the four ICA algorithms.
(c) TNR_F of the four ICA algorithms. (d) TPR_F of the four ICA algorithms.

Therefore, the conclusion we obtained through the comparison is that SOBI is more suitable for the task of separating multi-channel physiological sources via facial videos.

E. COMPARISON OF INFORMATION FUSION UNDER VARIOUS DRIVING CONDITION

In real-world driving scenarios, it is very difficult to keep the illumination condition constant for the drivers, and it is impossible for the drivers to sit and drive without any body movements even if they try to do so. Meanwhile, the dramatic change of the images in ROI may cause huge detection errors because PPG is data-driven. Therefore, it is necessary to evaluate the proposed method of merging multiple sources against using only one single source to explore the impact brought by illumination and subjects' head position changes.

FIGURE 13. Images of subjects with different head motions and illumination conditions.

In this section, drowsiness detection algorithms simply based on one specific source (i.e., yawn, blink, HRV) and merging the three sources are compared under different illumination and participant's head motions. Figure 13 shows different head motions and illumination conditions using images from 6 participants. A is a "clean" condition where no significant interferences exist. B is the condition with low illumination. C demonstrates a condition where illumination interference and head moving simultaneously exist. D is a daytime condition with in-plane head rotation of the participant. E is a condition with illumination interference. F is a daytime condition with vertical head rotation of the participant.

In our method, the accuracy of detecting drowsiness fundamentally depends on the correct detection of each individual source. So firstly, we evaluated the proposed method in detecting blink, yawn and BVP in different conditions as shown in Figure 13. The detailed results are summarized in TABLE 3. Although decline in detection accuracy was observed in condition B and C (see the second and third row of TABLE 3), the detection results under condition D, E and F, in which sources were successfully reconstructed without been strongly influenced by head moving and illumination changing, indicates that the proposed method still works pretty well under those conditions. However, as the angle of head rotation increased, the detection result became worse.

The proposed method works well as long as the head rotation angle does not exceed 30 degrees. As to the

TABLE 3. Results of the proposed method to respectively detect Blink, Yawn and BVP in different conditions.

Environment	Index	Yawn	Blink	BVP
A	*DR*	0.878	0.931	0.953
	FAR	0.111	0.042	0.016
	TNR_F	0..918	0.905	0.893
	NPR_F	0.903	0.937	0.944
B	*DR*	0.707	0.691	0.667
	FAR	0.203	0.235	0.315
	TNR_F	0.726	0.717	0.689
	NPR_F	0.658	0.679	0.655
C	*DR*	0.272	0.223	0.287
	FAR	0.618	0.661	0.703
	TNR_F	0.267	0.275	0.201
	NPR_F	0.303	0.319	0.288
D	*DR*	0.781	0.814	0.827
	FAR	0.165	0.091	0.103
	TNR_F	0.771	0.792	0.724
	NPR_F	0.803	0.811	0.747
E	*DR*	0.674	0.633	0.621
	FAR	0.225	0.212	0.243
	TNR_F	0.596	0.611	0.589
	NPR_F	0.677	0.682	0.656
F	*DR*	0.716	0.692	0.717
	FAR	0.201	0.212	0.168
	TNR_F	0.727	0.732	0.722
	NPR_F	0.731	0.756	0.752

condition C where illumination interference and head rotation simultaneously exist, the performance of our method is poor because of the drastic changing of the pixel value in the ROI. The observation signals derived from the obscure ROI in condition C contain almost no effective information for sources separation, hence leading to a reduction in detection accuracy. A face tracking module is necessary here to guarantee the facial region could be properly covered to tolerant strenuous head motions. Meanwhile an image enhancement algorithm is needed to improve the performance in low-illumination environment.

Although drowsiness can be determined by simply using one of the features that extracted from the outputs of SOBI, a merge of multiple physiological information will provide us with a more robust and accurate drowsiness detection result. TABLE 4 presents the comparison of drowsiness detection results using different feature(s). Single parameter and multi-parameter based methods derived from [17], [18], [21]–[23], [25], [26], and [61] were used to make comparisons with the proposed method. The results are averaged over 80 videos recoded both in-lab and in-car from 15 subjects. Obviously, the proposed method obtained higher *DR*, $TNR_{(F)}$, and $NPR_{(F)}$ in the experiments, which is more effective and stable than the other methods.

F. USING A SMALL ROI

In our experiments, we found that reducing half of the ROI in the vertical direction does not bring much difference into the

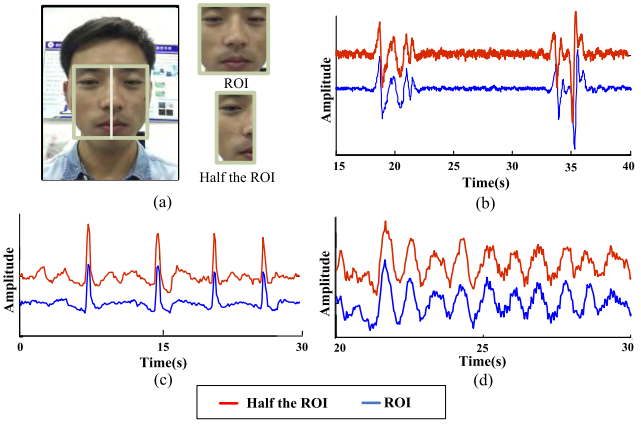


FIGURE 14. Detection results using two ROIs. (a) ROI and half the ROI. (b) Yawn detected using the two ROIs. (c) Blink detected using the two ROIs. (d) BVP detected using the two ROIs.

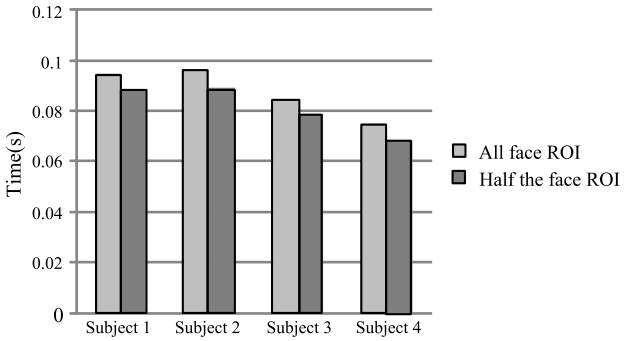


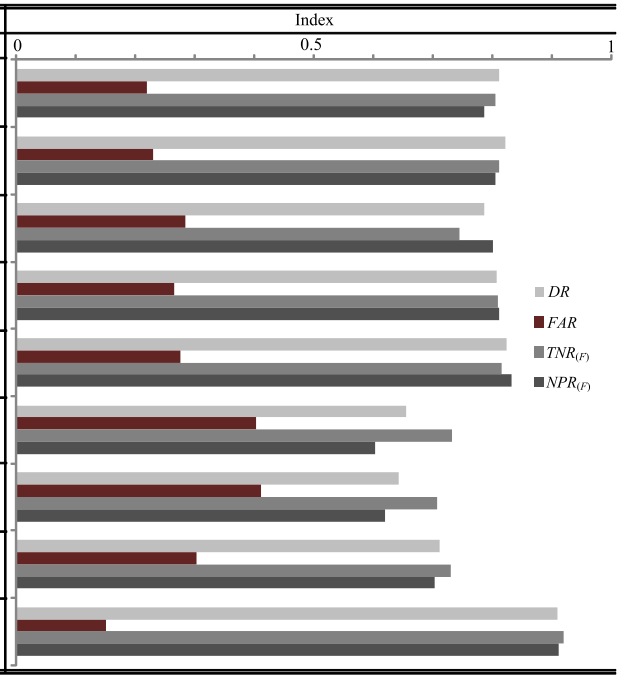
FIGURE 15. Average per-frame processing time of different ROIs.

detection results. A reduction of half the ROI means less time consuming in building the observation matrix, which in turn increases the overall operational efficiency of the proposed method. We therefore think about whether we can use half-the-face ROI with only one eye and half a mouth to obtain the same detection performance as that using a full-face ROI. The two rectangles in Figure 14 (a) shows the two ROIs used in this experiment. The detection results corresponding to the two ROIs are shown respectively in red and blue in Figure 14 (c) and (d). As can be seen in the figure, when half-the-face ROI was used, more tiny peaks were observed in the separated blinks, but the blinks were still distinguishable enough.

Small ROI leads to a reduction of visual information associated with facial cardiovascular activity and mouth motion, and as a result, the separated yawn and BVP waveforms were not as smooth as that from all-face ROI. A post filtering can be used here to eliminate the tiny performance degradation by modifying the waveforms. The experiment revealed that sufficient information can be provided for multiple physiological information extraction even if small ROI which contains only one eye and half the mouth is used. Figure 15 shows the comparison of average per-frame processing time between the two ROIs on videos from four subjects. It can be observed that

TABLE 4. Performance evaluation of different drowsiness detection methods.

Baseline methods	Feature: Yawn	Feature: Blink	Feature: BVP
[17]	•	•	
[18]	•	•	
[21]		•	
[22]		•	•
[23]		•	•
[25]	•		
[26]	•		
[61]		•	
proposed method	•	•	•



half-the-face ROI needs less time consuming than all-face ROI as we expected.

VI. LIMITATIONS AND FUTURE WORK

We have reported in the paper that we found the performance of the proposed method is to some extent subject dependent. The extended-PPG method is built based on measuring the physiological characteristics of subjects. However, people's physiological characteristics vary from person to person. The proposed method may work well on most of the subjects but there are indeed someone who has specific physiological characteristic that our method cannot deal with without further adjustment. For example, videos from a participant who has small eyes is relatively difficult to extracted pulsatil signals corresponding to blink because the pixel value change caused by blink is indistinguishable. Moreover, some common items that are frequently used in daily driving, for example, sunglasses, bring difficulties for our method. When the participants are driving with sunglasses, the lenses between the face and the camera will obscure the real ROI and even make the eyes invisible. So the proposed framework is currently not suitable for sunglasses-wearing drivers.

As an additional part to the experiments mentioned above, we tentatively recorded some data in real driving scenario on two of the subjects whom have valid driving licenses for more than two years and have rich driving experience. The participants were asked to drive following a fixed route again and again and not allowed to stop the car even though they feel sleepy (to maximize the safety we assigned another person to sit next to the driver to provide safety measures). The total driving time was 155 minutes and about 95 minutes data

FIGURE 16. The demo runs on a smart phone. (a) The in-car environment for system testing. (b) Screenshot of the system.

was useful. In the video analysis, we found that severe vibrations happened when the car going across a curve or a rough road. That kind of vibration is too severe that the whole facial area may go out of the camera scene even though a Meanshift tracker had been employed to prevent the ROI from missing in the videos. Under that condition, the structure of the mixture (observation) signal changed drastically and BVP, blink, and yawn signal are almost impossible to be precisely extracted.

In this work, the emphasis is on reporting the SOBI based fusion framework for drowsiness detection. A very simple strategy was used here to merge multiple physiological information. To increase the robustness and adaptability to real driving environment, method that are more sophisticated will

be employed in our future work. Furthermore, a better performance may be achieved if combining the framework with other wearable wireless sensors such as wrist band to overcome the drastic performance degradation in some extreme circumstances.

As another important part of our future work, a demo based on the proposed method is now under development. The system is implemented completely on a smart phone without using any other external hardware. The system can play an audible alert if a driver is detected as in the drowsy state.

VII. CONCLUSIONS

We present a method for drowsiness detection based on simultaneous detection of yawn, blink and BVP in this paper. 9-channel SOBI is proposed to provide the main frame of our method by simultaneously separating multiple physiological sources. Short-term energy and kurtosis are combined to automatically identify the sources from the multi-outputs. Drowsiness is finally determined by analyzing the separated yawn, blink, and BVP signals in parallel.

The experimental results show improvements in *DR*, $TNR_{(F)}$, $NPR_{(F)}$, and waveform quality over existing methods. Furthermore, we tested the methods under different illumination conditions and head motions. The proposed method is motion-tolerant to normal head motions, but is affected by drastic change of illumination or head motion. We also show that accurate BVP, blink and yawn waveforms can be extracted even when half of the original ROI is reduced.

The proposed method is relatively simple in both device and algorithm but still effective. It presents the possibility for measuring the driver state in real driving environment by smart phone. This will enable new applications which protect driving safety.

REFERENCES

[1] A. Sahayadhas, K. Sundaraj, and M. Murugappan, "Detecting driver drowsiness based on sensors: A review," *Sensors*, vol. 12, no. 12, pp. 16937–16953, Dec. 2012. doi: 10.3390/s121216937.

[2] D. Dawson, A. K. Searle, and J. L. Paterson, "Look before you (s)leep: Evaluating the use of fatigue detection technologies within a fatigue risk management system for the road transport industry," *Sleep Med. Rev.*, vol. 18, no. 2, pp. 141–152, Apr. 2014, doi: 10.1016/j.smrv.2013.03.003.

[3] G. Zhang, K. K. Yau, X. Zhang, and Y. Li, "Traffic accidents involving fatigue driving and their extent of casualties," *Accident Anal., Prevent.*, vol. 87, pp. 34–42, Feb. 2016, doi: 10.1016/j.aap.2015.10.033.

[4] P. Philip and T. Akerstedt, "Transport and industrial safety, how are they affected by sleepiness and sleep restriction?" *Sleep Med. Rev.*, vol. 10, no. 5, pp. 347–356, Oct. 2006, doi: 10.1016/j.smrv.2006.04.002.

[5] J. C. Verster, J. Taillard, B. Olivier, P. Philip, and P. Sagaspe, "Prolonged nocturnal driving can be as dangerous as severe alcohol-impaired driving," *J. Sleep Res.*, vol. 20, no. 4, pp. 585–588, Jan. 2011, doi: 10.1111/j.1365-2869.2010.00901.x.

[6] J.-Y. Bae and K. Yong, "The core technical trends of TESLA EV(electric vehicle) motors," *Trans. Korean Inst. Power Electron.*, vol. 22, pp. 414–422, Oct. 2017. [Online]. Available: http://koreascience.or.kr/journal/JRJJC3.page

[7] C. M. Martinez, M. Heucke, B. Gao, D. Cao, and F.-Y. Wang, "Driving style recognition for intelligent vehicle control and advanced driver assistance: A survey," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 3, pp. 666–676, Mar. 2018, doi: 10.1109/TITS.2017.2706978.

[8] M. Tanha and H. Seifoory, "Morphological drowsy detection," in *Proc. ICSIPA*, Kuala Lumpur, Malaysia, 2011, pp. 63–65.

[9] X. Wang and C. Xu, "Driver drowsiness detection based on non-intrusive metrics considering individual specifics," *Accident Anal. Prevention*, vol. 95, pp. 350–357, Oct. 2016, doi: 10.1016/j.aap.2015.09.002.

[10] M. Hajinoroozi, Z. Mao, and Y. Huang, "Prediction of driver's drowsy and alert states from EEG signals with deep learning," in *Proc. CAMSAP*, Cancún, Mexico, 2016, pp. 493–496.

[11] S.-Y. Shi, W.-Z. Tang, and Y.-Y. Wang, "A review on fatigue driving detection," in *Proc. ITA*, Guangzhou, China, 2017, p. 01019.

[12] L. B. Leng, L. B. Giin, and W.-Y. Chung, "Wearable driver drowsiness detection system based on biomedical and motion sensors," in *Proc. IEEE SENSORS*, Busan, South Korea, Nov. 2015, pp. 1–4.

[13] W.-C. Li, W.-L. Ou, C.-P. Fan, and C.-H. Huang, "Near-infrared-ray and side-view video based drowsy driver detection system: Whether or not wearing glasses," in *Proc. APCCAS*, Jeju, South Korea, 2017, pp. 429–432.

[14] M. F. Maysoon, H. A. Hesham, and F. Lamiaa, "Using mobile platform to detect and alerts driver fatigue," *Int. J. Comput. Appl.*, vol. 123, no. 8, pp. 27–35, Aug. 2015, doi: 10.5120/ijca2015905428.

[15] L. Xu, S. Li, and K. Bian, "Sober-drive: A smartphone-assisted drowsy driving detection system," in *Proc. ICNC*, Honolulu, HI, USA, 2014, pp. 398–402.

[16] H. Eren, S. Makinist, E. Akin, and A. Yilmaz, "Estimating driving behavior by a smartphone," in *Proc. IEEE Intell. Vehicles Symp.*, Alcala de Henares, Spain, Jun. 2012, pp. 234–239.

[17] K. Churiwala, R. Lopes, A. Shah, and N. Shah, "Drowsiness detection based on eye movement, yawn detection and head rotation," *Int. J. Appl. Inf. Syst.*, vol. 2, no. 6, pp. 0868–2249, May 2012.

[18] V. B. Hemadri and U. P. Kulkarni, "Detection of drowsiness using fusion of yawning and eyelid movements," in *Advances in Computing, Communication, and Control*, vol. 361. Berlin, Germany: Springer, 2013, pp. 583–594.

[19] S. Samiee, S. Azadi, A. Nahvi, A. Eichberger, and R. Kazemi, "Data fusion to develop a driver drowsiness detection system with robustness to signal loss," *Sensors*, vol. 14, no. 9, pp. 17832–17847, Sep. 2014, doi: 10.3390/s140917832.

[20] J. D. Fuletra and D. Bosamiya, "A survey on driver's drowsiness detection techniques," *Int. J. Recent Innov. Trends in Comput. Commun.*, vol. 1, no. 11, pp. 816–819, 2013.

[21] T. Hayami, K. Matsunaga, K. Shidoji, and Y. Matsuki, "Detecting drowsiness while driving by measuring eye movement—A pilot study," in *Proc. IEEE Int. Conf. Intell. Transp. Syst.*, Singapore, Sep. 2002, pp. 156–161.

[22] A. Noda *et al.*, "Simultaneous measurement of heart rate variability and blinking duration to predict sleep onset and drowsiness in drivers," *J. Sleep Disorders Therapy*, vol. 38, no. 1, pp. 12–16, 2015, doi: 10.4172/2167-0277.1000213.

[23] A. Tsuchida, M. S. Bhuiyan, and K. Oguri, "Estimation of drowsiness level based on eyelid closure and heart rate variability," in *Proc. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Minneapolis, MN, USA, Sep. 2009, pp. 2543–2546.

[24] A. Belouchrani, K. Abed-Meraim, J.-F. Cardoso, and E. Moulines, "A blind source separation technique using second-order statistics," *IEEE Trans. Signal Process.*, vol. 45, no. 2, pp. 434–444, Feb. 1997, doi: 110.1109/78.554307.

[25] N. Alioua, A. Amine, and M. Rziza, "Driver's fatigue detection based on yawning extraction," *Int. J. Veh. Technol.*, vol. 2014, no. 1, Aug. 2014, At. no. 678786, doi: 10.1155/2014/678786.

[26] S. Abtahi, B. Hariri, and S. Shirmohammadi, "Driver drowsiness monitoring based on yawning detection," in *Proc. I2MTC*, Binjiang, China, 2011, pp. 1–4.

[27] J. Ahmed, J.-P. Li, S. A. Khan, and R. A. Shaikh, "Eye behaviour based drowsiness detection system," in *Proc. ICCWAMTIP*, Chengdu, China, 2015, pp. 268–272.

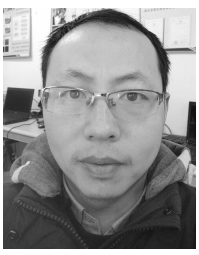
[28] N. Galley, R. Schleicher, and L. Galley, "Blink parameters as indicators of driver's sleepiness–possibilities and limitations," *Vis. Vehicles*, vol. 10, pp. 189–196, Jan. 2004.

[29] F. Andreotti, A. Trumpp, H. Malberg, and S. Zaunsed, "Improved heart rate detection for camera-based photoplethysmography by means of Kalman filtering," in *Proc. ELNANO*, Kiev, Ukraine, 2015, pp. 428–433.

[30] W. J. Jiang, S. C. Gao, P. Wittek, and L. Zhao, "Real-time quantifying heart beat rate from facial video recording on a smart phone using Kalman filters," in *Proc. Healthcom*, Natal, Brazil, 2014, pp. 393–396.

[31] J. Vicente, P. Laguna, A. Bartra, and R. Bailon, "Drowsiness detection using heart rate variability," *Med. Biol. Eng. Comput.*, vol. 54, no. 6, pp. 927–937, Jun. 2016, doi: 10.1007/s11517-015-1448-7.

[32] A. Murata and Y. Hiramatsu, "Evaluation of drowsiness by HRV measures-basic study for drowsy driver detection," in *Proc. IEEE IWCIA*, Apr. 2008, pp. 99–102.
[33] T. Štula, "Evaluation of heart rate variability from ECG signal," *IFAC Proc. Volumes*, vol. 36, no. 1, pp. 429–432, Feb. 2003, doi: 10.1016/S1474-6670(17)33789-8.
[34] P. Smith, M. Shah, and N. da Vitoria Lobo, "Determining driver visual attention with one camera," *IEEE Trans. Intell. Transp. Syst.*, vol. 4, no. 4, pp. 205–218, Dec. 2003, doi: 10.1109/TITS.2003.821342.
[35] T. D'Orazio, M. Leo, C. Guaragnella, and A. Distante, "A visual approach for driver inattention detection," *Pattern Recognit.*, vol. 40, no. 8, pp. 2341–2355, Aug. 2007, doi: 10.1016/j.patcog.2007.01.018.
[36] R. Oyini Mbouna, S. G. Kong, and M.-G. Chun, "Visual analysis of eye state and head pose for driver alertness monitoring," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 3, pp. 1462–1469, Sep. 2013, doi: 10.1109/TITS.2013.2262098.
[37] A. Picot, S. Charbonnier, and A. Caplier, "On-line detection of drowsiness using brain and visual information," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 42, no. 3, pp. 764–775, May 2012, doi: 10.1109/TSMCA.2011.2164242.
[38] J. Allen, "Photoplethysmography and its application in clinical physiological measurement," *Physiol. Meas.*, vol. 28, no. 3, p. R1, Feb. 2007, doi: 10.1088/0967-3334/28/3/R01.
[39] C. Takano and Y. Ohta, "Heart rate measurement based on a time-lapse image," *Med. Eng. Phys.*, vol. 29, no. 8, pp. 853–857, Oct. 2007, doi: 10.1016/j.medengphy.2006.09.006.
[40] L. Kong *et al.*, "Non-contact detection of oxygen saturation based on visible light imaging device using ambient light," *Opt. Express*, vol. 21, no. 15, pp. 17464–17471, Jul. 2013, doi: 10.1364/OE.21.017464.
[41] Y.-D. Lin, Y.-H. Chien, and Y.-S. Chen, "Wavelet-based embedded algorithm for respiratory rate estimation from PPG signal," *Biomed. Signal Process.*, vol. 36, pp. 138–145, Jul. 2017, doi: 10.1016/j.bspc.2017.03.009.
[42] Y. Sun, S. Hu, R. Kalawsky, S. Greenwald, and V. Azorin-Peris, "Non-contact imaging photoplethysmography to effectively access pulse rate variability," *J. Biomed. Opt.*, vol. 18, no. 6, p. 061205, Oct. 2012, doi: 10.1117/1.JBO.18.6.061205.
[43] J. Lee, K. Matsumura, K.-I. Yamakoshi, P. Rolfe, S. Tanaka, and T. Yamakoshi, "Comparison between red, green and blue light reflection photoplethysmography for heart rate monitoring during motion," in *Proc. EMBC*, Jul. 2013, pp. 1724–1727.
[44] V. Ntziachristos, J. Ripoll, L. V. Wang, and R. Weissleder, "Looking and listening to light: The evolution of whole-body photonic imaging," *Nature Biotechnol.*, vol. 23, pp. 313–320, Mar. 2005, doi: 10.1038/nbt1074.
[45] J. Zheng and S. Hu, "The preliminary investigation of imaging photoplethysmographic system," *J. Phys., Conf. Ser.*, vol. 85, no. 1, p. 012031, Oct. 2007.
[46] P. Pelegris, K. Banitsas, K. Marias, and T. Orbach, "A novel method to detect Heart Beat Rate using a mobile phone," in *Proc. EMBC*, Nov. 2010, pp. 5488–5491.
[47] M.-Z. Poh, D. J. McDuff, and R. W. Picard, "Advancements in noncontact, multiparameter physiological measurements using a webcam," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 1, pp. 7–11, Jan. 2011, doi: 10.1109/TBME.2010.2086456.
[48] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, vol. 46. Hoboken, NJ, USA: Wiley, 2004.
[49] D. McDuff, S. Gontarek, and R. W. Picard, "Remote detection of photoplethysmographic systolic and diastolic peaks using a digital camera," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 12, pp. 2948–2954, Dec. 2014, doi: 10.1109/TBME.2014.2340991.
[50] K. Alghoul, S. Alharthi, A. El Saddik, and H. Al Osman, "Heart rate variability extraction from videos signals: ICA vs. EVM comparison," *IEEE Access*, vol. 5, pp. 4711–4719, 2017, doi: 10.1109/ACCESS.2017.2678521.
[51] C. Zhang, X. Wu, L. Zhang, Z. Lv, and X. He, "Simultaneous detection of blink and heart rate using multi-channel ICA from smart phone videos," *Biomed. Signal Process. Control*, vol. 33, pp. 189–200, Mar. 2017, doi: 10.1016/j.bspc.2016.11.022.
[52] M. R. H. Samadi and N. Cooke, "VOG-enhanced ICA for SSVEP response detection from consumer-grade EEG," in *Proc. EUSIPCO*, Lisbon, Portugal, Nov. 2014, pp. 2025–2029.
[53] S. H. Sardouie, L. Albera, M. B. Shamsollahi, and I. Merlet, "An efficient jacobi-like deflationary ICA algorithm: Application to EEG denoising," *IEEE Signal Process. Lett.*, vol. 22, no. 8, pp. 1198–1202, Aug. 2015, doi: 10.1109/LSP.2014.2385868.
[54] A. Belouchrani, K. Abed-Meraim, J. F. Cardoso, and E. Moulines, "Second-order blind separation of temporally correlated sources," in *Proc. Int. Conf. Digital. Signal Process.*, Nicosia, Cyprus, May 1993, pp. 346–351.
[55] R. Stojanovic and D. Karadaglic, "A LED–LED-based photoplethysmography sensor," *Physiol. Meas.*, vol. 28, no. 6, p. N19, May 2007, doi: 10.1088/0967-3334/28/6/N01.
[56] S. Zafeiriou, C. Zhang, and Z. Zhang, "A survey on face detection in the wild: Past, present and future," *Comput. Vis. Image Understand.*, vol. 138, pp. 1–24, Sep. 2015, doi: 10.1016/j.cviu.2015.03.015.
[57] D. Comaniciu, V. Ramesh, and P. Meer, "Real-time tracking of non-rigid objects using mean shift," in *Proc. CVPR*, Hilton Head Island, SC, USA, Jun. 2000, pp. 142–149.
[58] I. Rejer and P. Górski, "Benefits of ICA in the case of a few channel EEG," in *Proc. EMBC*, Milan, Italy, Aug. 2015, pp. 7434–7437.
[59] A. Daffonchio, C. Franzelli, P. Castiglioni, G. Mancia, A. U. Ferrari, and R. M. Di, "Sympathetic, parasympathetic and non-autonomic contributions to cardiovascular spectral powers in unanesthetized spontaneously hypertensive rats," *J. Hypertension*, vol. 13, no. 13, pp. 1636–1642, Dec. 1995.
[60] H.-B. Kang, "Various approaches for driver and driving behavior monitoring: A review," in *Poc. ICCV*, Sydney, NSW, Australia, Dec. 2013, pp. 616–623.
[61] S. Boverie and A. Giralt, "Driver vigilance diagnostic based on eyelid movement observation," *IFAC Proc. Volumes*, vol. 41, no. 2, pp. 12831–12836, Jul. 2008, doi: 10.3182/20080706-5-KR-1001.02170.
[62] M. Chau and M. Betke, "Real time eye tracking and blink detection with USB cameras," Dept. Comput. Sci., Boston Univ., Boston, MA, USA, CAS Comput. Sci., Tech. Rep. 2005-12, May 2005.
[63] A. Panning, A. Al-Hamadi, and B. Michaelis, "A color based approach for eye blink detection in image sequences," in *Proc. IEEE SIGPRO*, Kuala Lumpur, Malaysia, Nov. 2011, pp. 40–45.
[64] R. Barea, L. Boquete, E. López, J. M. Rodríguez-Ascariz, and S. Ortega, "EOG-based eye movements codification for human computer interaction," *Expert Syst. Appl.*, vol. 39, no. 3, pp. 2677–2683, Feb. 2012, doi: 10.1016/j.eswa.2011.08.123.
[65] J. E. Dotter, "Apparatus and method for measuring heart rate and other physiological data," U.S. Patent 2008 0 171 945 A1, Jul. 17, 2008.
[66] G. Jing, C. E. Siong, and D. Rajan, "Foreground motion detection by difference-based spatial temporal entropy image," in *Proc. TENCON*, Chiang Mai, Thailand, Nov. 2004, pp. 379–382.
[67] A. Hyvarinen, "Fast ICA for noisy data using Gaussian moments," in *Proc. ISCAS*, Orlando, FL, USA, May 1999, pp. 57–61.
[68] T.-W. Lee, M. Girolami, and T. J. Sejnowski, "Independent component analysis using an extended infomax algorithm for mixed subgaussian and supergaussian sources," *Neural Comput.*, vol. 11, no. 2, pp. 417–441, Feb. 1999, doi: 10.1162/089976699300016719.
[69] K. Nordhausen, J.-F. Cardoso, H. Oja, E. Ollila, S. Taskinen, and J. Miettinen, "JADE: Blind source separation methods based on joint diagonalization and some BSS performance criteria," *J. Stat. Softw.*, vol. 76, no. 2, pp. 1–31, Mar. 2018, doi: 10.18637/jss.v076.i02.



CHAO ZHANG received the B.S. degree in electronic information engineering from the Army Officer Academy of PLA, Hefei, China, in 2006, and the M.E. and Ph.D. degrees in computer application technology from Anhui University, Hefei, in 2009 and 2014, respectively, where he is currently a Lecturer with the School of Computer Science and Technology.

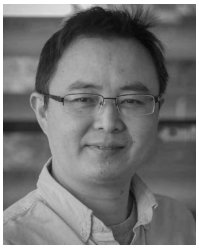
His research interests include intelligent information processing and human–computer interaction.

XIAOPEI WU received the B.S. degree from the Radio Engineering Department, Anhui University, Hefei, China, in 1985, the M.S. degree in communications and electronic engineering from the University of Electronic Science and Technology, Chengdu, China, in 1988, and the Ph.D. degree in electronic science and technology from the University of Science and Technology of China, Hefei, in 2002.

He is currently a Full Professor with the School of Computer Science and Technology, Anhui University. He has published more than 100 high-quality publications in journals and conferences and has held more than ten patents. His research interests include statistical signal processing, blind signal processing, biomedical signal processing, and human–computer interaction.

XI ZHENG received the bachelor's degree in computer information system from Fudan University, the master's degree in computer and information science from UNSW, and the Ph.D. degree in software engineering from UT Austin. From 2005 to 2012, he was a Chief Solution Architect with Menulog, Australia, the company sold for 8.55 billion U.S. Dollars. He is currently an Assistant Professor/Lecturer in software engineering with Macquarie University.

He has published more than 30 high-quality publications in top journals and conferences (PerCOM, ICSE, ICCPS, the IEEE Systems Journal, and the *ACM Transactions on Embedded Computing Systems*). His research interests include service computing, the IoT security, and reliability analysis. He received the Deakin Research Outstanding Award, in 2016, and the Best Paper Award in the Australian Distributed Computing and Doctoral Conference, in 2017. He is a Reviewer for top journals and conferences (the IEEE Systems Journal, the *ACM Transactions on Design Automation of Electronic Systems*, *Pervasive and Mobile Computing*, the IEEE Transactions on Cloud Computing, and PerCOM).



SHUI YU is currently a Full Professor with the School of Software, University of Technology Sydney, Australia. His research interests include security and privacy, networking, big data, and mathematical modeling.

He has edited two books and has published two monographs and more than 200 technical papers, including top journals and top conferences, such as the IEEE TPDS, TC, TIFS, TMC, TKDE, TETC, ToN, and INFOCOM. He initiated the research field of networking for big data, in 2013. His h-index is 32. He is a Senior Member of the IEEE, a member of the AAAS and the ACM, the Vice Chair of the Technical Committee on Big Data of the IEEE Communication Society, and a Distinguished Lecturer of the IEEE Communication Society. He actively serves his research communities in various roles. He is currently serving the Editorial Boards of the IEEE Communications Surveys and Tutorials, the *IEEE Communications Magazine*, the IEEE Internet of Things Journal, the IEEE Communications Letters, the IEEE Access, and the IEEE Transactions on Computational Social Systems. He has served more than 70 international conferences as a member of the Organizing Committee, such as the Publication Chair for the IEEE GLOBECOM, in 2015, and the IEEE INFOCOM, in 2016 and 2017, as the TPC Chair for the IEEE BigDataService, in 2015, and as the General Chair for ACSW, in 2017.