PHYSIOLOGICAL PARAMETER MONITORING OF DRIVERS BASED ON VIDEO DATA AND INDEPENDENT VECTOR ANALYSIS

Zhenyu Guo and Z. Jane Wang, Senior Member, IEEE

University of British Columbia
Department of Electrical and Computer Engineering
{zhenyug, zjanew}@ece.ubc.ca

Zhiqi Shen

Nanyang Technological University
School of Computer Engineering
zqshen@ntu.edu.sg

ABSTRACT

Although modern cars are equipped with advanced technologies to be faster, more comfortable and safer, one essential piece of the driving system, the driver, is missing in the picture. Among the physiological measures used for wellness purposes, heart rate variability has been shown to be directly associated with mental and physical status, and is easy to measure. In this paper, to maintain the driver's comfort and enhance the driving safety, we propose a non-contact, video-based approach to continuously monitor the driver's heart rate variability under real-world driving circumstances. Previously, several methods were proposed for similar goals under laboratory conditions, where simple face detectors and independent component analysis approach were used, and they may fail in both image understanding and signal processing steps under real-world circumstances in driving. Here we propose using advanced facial landmark and pose estimation, and independent vector analysis to extract heart rate variability. Our preliminary experimental results demonstrated that the proposed approach works better than the previous state-of-arts.

Index Terms— Driver Monitoring, Heart Rate Variability, IVA

1. INTRODUCTION

Tremendous efforts have been spent to make the cars safer, more comfortable, and more powerful, since the day when cars were invented. Recent cars become more and more intelligent by adopting advanced technologies. However, an important piece of the driving process, the driver, is missing in the picture. Many aspects of the drivers' conditions affect the safety, for example sleepiness and sleep disorder, clinical depression, alcohol consumption, emergencies(myocardial dysfunction), and even their emotions (anger, excitement, etc.). These physiological and mental conditions are highly correlated with the drivers' capability for safe driving. For example, a clear association between sleep-disordered breathing (SDB) and traffic accidents has been previously demonstrated[1]. If we could continuously monitor physiological parameters that can reveal useful information about the drivers, the driving safety could be further improved.

For mental and physiological status monitoring, brain imaging (such as CT, fMRI), EEG, ECG, and etc. are widely used in clinics and research labs. While such monitoring requires be non-intrusive, non-disturbing and comfortable during driving, we decide to use Heart Rate Variability (HRV) as an indicating physiological parameter to monitor. HRV is defined as fluctuations in interbeat interval length which reflect the hearts response to extracardiac factors that affect heart rate. HRV allows simultaneous assessment of both sympathetic and parasympathetic activity and the interplay between them. Increased HRV can be associated with exercise and aerobic fitness, while decreased HRV can be associated with aging, chronic stress, and a wide variety of medical and psychiatric disorders[2]. Decreased HRV can be used to predict mortality in general population samples and patients with myocardial infarction and used as an indicator of the altered autonomic function. Significant negative correlations were found between HRV and both the severity of depression and the duration of the depressive episode. Therefore, HRV makes a good candidate for monitoring drivers' physiological and mental conditions. Usually HRV can be calculated from Blood Volume Pulse (BVP) signals.

Different devices can measure Heart Rate Variability. A chest belt with motion sensors can measure the heart beat signal and is widely used for wellness and fitness purpose. It is obvious that a chest belt is not convenient and comfortable to use during daily driving. Another popular way is based on Photoplethysmography (PPG) to measure BVP through variations in transmitted or reflected light, which usually comes as a form of finger clipper with a dedicated LED light source. [3] showed that pulse measurements from the human face are attainable with normal ambient light as the illumination source, which supports the idea of extracting bvp signals out from facial video data. It is clear that a non-contact, video camera based system is preferred for real-world driving monitoring. In this paper, we propose a video processing based framework for continuously monitoring drivers physiological parameters, where Independent Vector Analysis (IVA) is explored and spectral clustering is proposed to select common sources after IVA.

2. RELATED WORKS

Using PPG to measure BVP and then HRV has been widely used in clinics and research labs, due to its simplicity, convenience and accuracy. Very recently, measuring BVP and HRV signals from facial video data recorded by webcam and smart phones have been proposed in [4, 5]. A summery about using smart phones to measure several physiological parameters can be found in [6]. Similar methods were used in [4, 5, 6], including a Viola & Jones face detector[7] to find the facial area, and a step of Independent Component Analysis (ICA)[8] to find the underlying BVP signal as a common source beneath the color intensities from R, G, B three channels. For our driver monitoring application, previous methods can be improved in both the face detection and signal processing steps, since the Viola & Jones detector cannot detect faces with certain pose angle in real world situations, and ICA on the three R, G, B color intensity signals may not be sufficient to discover the latent BVP source signals. Therefore, in this paper, we propose using more advanced facial landmark and pose estimator [9] to find more accurate facial sub-regions, and provide grouped color intensity signals for different facial subregions and color channels. Accordingly, Independent Vector Analysis (IVA)[10] is performed to extract common sources from different groups of mixed signals.

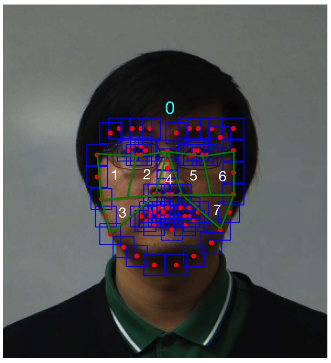


Fig. 1. Examples of the output of the algorithm in [9], where the red dots indicate the facial landmarks, and the blue 0 indicates that the face is 0 degree towards the camera. The green lines describe sub-regions proposed by this paper, and the white numbers are the indices of the sub-regions.

3. PROPOSED METHOD

In this section, we describe the proposed method for monitoring BVP and HRV signals from video data. [4, 5, 6] assume that the color intensity of the facial skin is changing along time, which is driven by the BVP signal. Therefore, they assume the BVP signal is one of the source signals for the observed mixed signals: color intensities from R, G, B three channels. The previous method recovers the BVP signal as one of the independent components obtained by applying ICA on the R, G, B intensity signals. Since three channels may not be statistically sufficient to extract the underlying sources, to overcome this problem, we propose detecting the facial landmarks, instead of only the entire facial region, and then dividing the facial region into seven sub-regions. After obtaining the sub-regions, we construct grouped multivariate time series to represent the observed mixed signals. At last, IVA is applied to extract the common sources shared by different groups of mixed signals. In the following sections, we will describe the details of our proposed algorithm.

3.1. Facial Landmarks Estimation and Sub-region Construction

[9] presented a unified model for face detection, pose estimation, and landmark estimation in real-world, cluttered images, which works better than the Viola & Jones algorithm even only for face detection. The outputs of the facial landmark algorithm are illustrated in Fig. 1, where we can see that the landmarks and the pose are estimated accurately.

After obtaining the facial landmarks, we divide the whole facial region into 7 sub-regions, whose boundaries are indicated by the green lines in Fig. 1. The sub-regions are predefined using the landmarks in the detection model. For computing the average RBG color intensities of each sub-region, we need to find the pixels lying inside the sub-regions. We adopt the algorithm proposed by [11] for finding points in polygon by providing the coordinates of the vertices in order. Therefore, for each image frame we have 7 sub-regions and 3 channels (RGB) of color intensity values. In the following section we will discuss how to group the multivariate time series.

3.2. Independent Vector Analysis

Independent Vector Analysis was recently proposed to solve Joint Blind Source Separation (JBSS) under certain assumptions. We first formulate the JBSS problem. There are K groups of signals, each of which contains T samples, formed from linear mixtures of N independent sources[10]

$$\mathbf{x}^{[k]}(t) = \mathbf{A}^{[k]}\mathbf{s}^{[k]}(t), 1 \leq k \leq K, 1 \leq t \leq T. \tag{1}$$

The tth sample of the zero-mean source vector, $\mathbf{s}^{[k]} = [s_1^{[k]}(t), ..., \mathbf{s}_N^{[k]}(t)]^T \in \mathcal{R}^N$, is a realization of random real-valued vector $\mathbf{S}^{[k]}$. The invertible mixing matrices, $\mathbf{A}^{[k]} \in \mathcal{R}^{N \times N}$, are unknown real-valued quantities to be estimated. If we concatenated the source vectors in each group to form $\mathbf{s}T = [(\mathbf{s}^{[1]})^T, ..., (\mathbf{s}^{[K]})^T] \in \mathcal{R}^{NK}$, then we can reformulate JBSS as,

$$\mathbf{x} = \mathbf{A}\mathbf{s}, \tag{2}$$

where $\mathbf{A}$ is a block diagonal matrix or $\mathbf{A} = \oplus_{k=1}^{K}\mathbf{A}^{[k]}$. To estimate the mixing matrices $\{\mathbf{A}^{[K]}\}_{k=1}^{K}$ for all groups of singles, [10] proposed to minimize the mutual information among the estimated source component vectors, by assuming that each latent source within a group is both related to a single latent source within each of the other groups and independent of all the other sources within the groups.

In our application, at time t, we have 3 channels and 7 sub-regions of color intensities. We group the mixed signals by sub-region, i.e., we treat each sub-region as a group of mixed signals, which contains 3 time series from R, G, B channels.

3.3. Source Signal Selection via Normalized Cut

After estimating the independent components (ICs), [4, 5] chose the 2nd IC as an estimation of BVP signal based on their empirical analysis. This is not robust since ICA algorithm doesn't return ICs in the same order on different data and different runs. Which ICs to choose is even more problematic in our IVA-based approach, since we will obtain 21 ICs in total. Recall that we assume the ICs are independent within groups, but some of them are highly correlated across groups. Those correlated ones are likely the approximation of BVP signal, since the BVP signal is a common source underlying the mixed signals from all sub-regions and channels. Therefore, we propose using the Normalized Cut [12] algorithm to perform spectral clustering on the similarity matrix of all ICs, which is computed using the normalized cross correlation (NCC) metric. Then we choose the largest cluster that contains most ICs as our candidate set of BVP signals. These signals are all used for further analysis.

We summarize the major steps of the proposed algorithm in Fig.2.

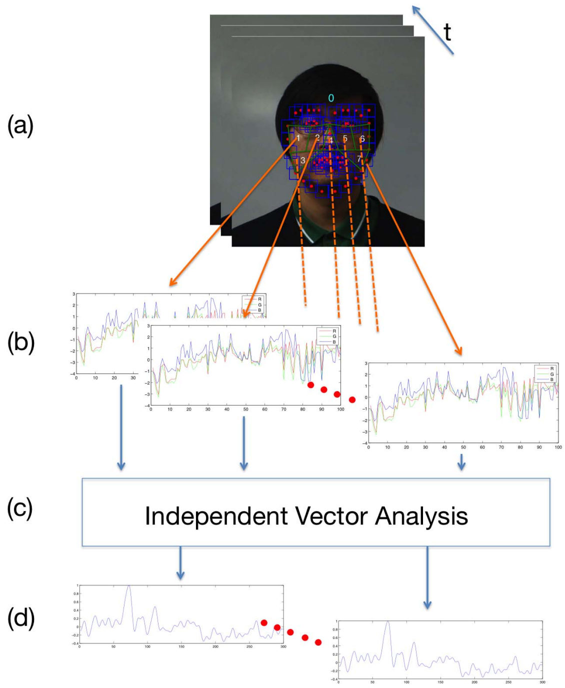


Fig. 2. Illustration of the proposed algorithm. (a) video sequence with detected landmarks and sub-regions. (b) 3 channel signals from 7 sub-regions (i.e. 7 groups). (c) IVA applied on the grouped signals. (d) 21 source signals recovered by IVA.

4. EXPERIMENT

In this section, we will describe the experiments and the results. Video data used is taken by the iSight camera on a MacBook Pro under different lightening conditions, with a fixed frame rate at 15 fps. To simulate the real-world conditions during driving, our subjects made different head and mouth movements during video recording, shown in Fig.3

First, we will take the results from one video sequence as an example. According to the algorithm describe in the previous section, we obtain 21 source signals from IVA analysis. The similarity matrix is shown in Fig.4. We set the cluster number to be 15 for the spectral clustering on the similarity matrix, so that most of the resulting clusters contain only 1 signal. Due to the assumption that source signals are independent within one group and several common sources are highly correlated across groups, it is safe to assume that the largest cluster contains common sources, which represent BVP signals in our case. In the example video data, the largest cluster after spectral clustering contains 5 source signals, as shown in Fig.5. From the figure we can see that the source signals are highly correlated, and we choose the minimum value at each time point from the 5 source signals to get a BVP estimation. This BVP signal is further smoothed by FFT filtering with a cutoff frequency at 4Hz(which is equivalent to 240 bmp heart beat rate, which is the maximum of human heart beat rate at normal conditions). The final BVP signal is shown in Fig.6. The red circles are the peaks detected by a peak finding algorithm, and the intervals between peaks are the Inter-Beat Intervals(IBI), which can be used for HRV analysis. The heart

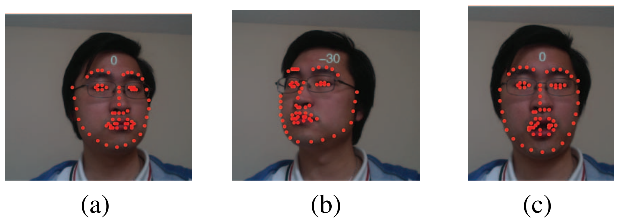


Fig. 3. Example frames from testing videos with head and mouth movements.

beat rate can be simply calculated by $hr = 1/\overline{IBI}$, where $\overline{IBI}$ is the average inter-beat rate of the BVP signal.

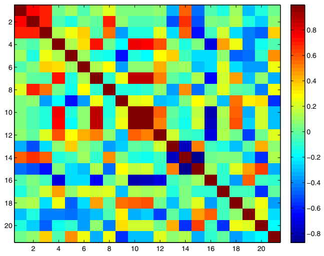


Fig. 4. Similarity matrix of the 21 source signals from IVA.

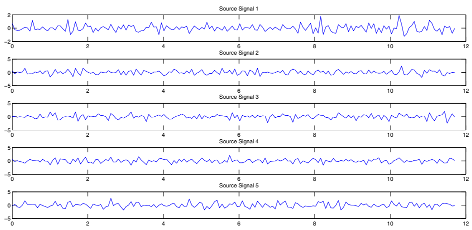


Fig. 5. Plots of 5 source signals from the IVA, which are selected by our proposed spectral clustering method.

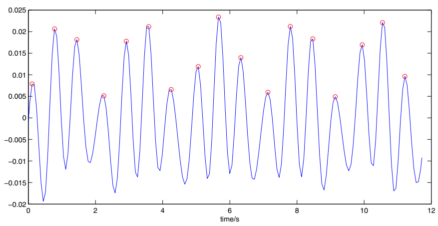
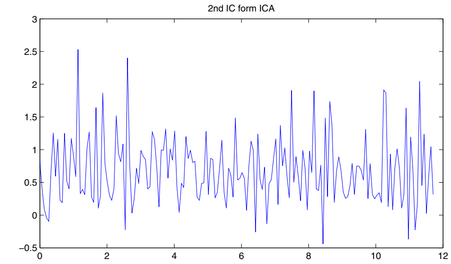


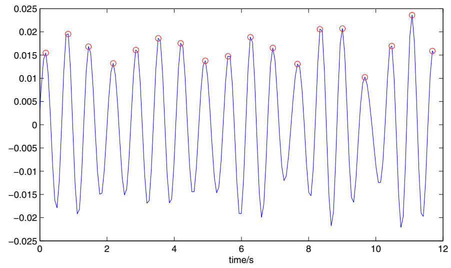
Fig. 6. The BVP estimation from the proposed IVA-based approach.

For comparison, we also estimate the BVP signal by applying ICA on the 3 channel color intensities computed from the whole facial region, as described in [4, 5]. The 2nd IC from ICA and the resulting BVP signal is shown in Fig. 7. Since the focus of this paper is on the estimation of BVP signals, due to the space limit, we didn't report the results of HRV analysis based on the estimated BVP. Please follow [4, 5] for details about HRV analysis.

To quantitatively demonstrate the improvement of our method, we conduct experiments on 5 testing videos, where the subjects are in different poses, lightening conditions and with different types of movements. For comparison, we also report the results from ICA based methods. Since the Viola & Jones face detector used in previous methods failed in the



(a) 2nd IC form ICA.



(b) BVP estimation from ICA.

Fig. 7. Results from the ICA-based approach applied on 3 channel color intensities of whole facial region.

Video ID	True HR	IVA HR	ICA HR
test1	74	73.6	failed to converge
test2	75	75.8	71.5
test3	72	73.3	69.8
test4	72	71.9	75
test5	71	70.6	79

Table 1. Experiment results for 5 testing videos under differnet pose, movement and lightening conditions. HR is shown in bpm. ICA based method faild to estimate the IC in video test1.

testing video, we use the facial landmarks estimator to get the whole facial area for the ICA algorithm. Results are shown in Table.1, where we can clear see that the proposed IVA based method works better in all cases with an average error rate around 0.7 bpm. And previous ICA based method failed on video test1, due to the head movement and mouth movement of the subject. And in the cases where ICA successfully gave estimations, the performance is much worse than the proposed IVA based method.

5. CONCLUSION

In this paper, we propose using video data to monitor drivers' physiological parameters based on IVA, with source selection by spectral clustering. The proposed algorithm makes improvement over previous ones in both image understanding step and signal processing step.

6. REFERENCES

[1] Rosa Catarino, Jorge Spratley, Isabel Catarino, Nuno Lunet, and Manuel Pais-Clemente, "Sleepiness and sleep-disordered breathing in truck drivers," *Sleep and Breathing*, pp. 1–10, 2013.

[2] Magdalena Romanowicz, John E Schmidt, John M Bostwick, David A Mrazek, and Victor M Karpyak, "Changes in heart rate variability associated with acute alcohol consumption: current knowledge and implications for practice and research," *Alcoholism: Clinical and Experimental Research*, vol. 35, no. 6, pp. 1092–1105, 2011.

[3] Wim Verkruysse, Lars O Svaasand, and J Stuart Nelson, "Remote plethysmographic imaging using ambient light," *Optics express*, vol. 16, no. 26, pp. 21434–21445, 2008.

[4] Ming-Zher Poh, Daniel J McDuff, and Rosalind W Picard, "Non-contact, automated cardiac pulse measurements using video imaging and blind source separation," *Optics Express*, vol. 18, no. 10, pp. 10762–10774, 2010.

[5] M-Z Poh, Daniel J McDuff, and Rosalind W Picard, "Advancements in noncontact, multiparameter physiological measurements using a webcam," *Biomedical Engineering, IEEE Transactions on*, vol. 58, no. 1, pp. 7–11, 2011.

[6] Christopher G Scully, Jinseok Lee, Joseph Meyer, Alexander M Gorbach, Domhnull Granquist-Fraser, Yitzhak Mendelson, and Ki H Chon, "Physiological parameter monitoring from optical recordings with a mobile phone," *Biomedical Engineering, IEEE Transactions on*, vol. 59, no. 2, pp. 303–306, 2012.

[7] Paul Viola and Michael Jones, "Rapid object detection using a boosted cascade of simple features," in *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*. IEEE, 2001, vol. 1, pp. I–511.

[8] Pierre Comon, "Independent component analysis, a new concept?," *Signal processing*, vol. 36, no. 3, pp. 287–314, 1994.

[9] Xiangxin Zhu and Deva Ramanan, "Face detection, pose estimation, and landmark localization in the wild," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 2879–2886.

[10] Matthew Anderson, Tuelay Adali, and Xi-Lin Li, "Joint blind source separation with multivariate gaussian model: algorithms and performance analysis," *Signal Processing, IEEE Transactions on*, vol. 60, no. 4, pp. 1672–1683, 2012.

[11] Kai Hormann and Alexander Agathos, "The point in polygon problem for arbitrary polygons," *Comput. Geom. Theory Appl.*, vol. 20, no. 3, pp. 131–144, 2001.

[12] Jianbo Shi and Jitendra Malik, "Normalized cuts and image segmentation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 22, no. 8, pp. 888–905, 2000.