Driving Reform: Digital Health is Everyone's Business
A. Georgiou et al. (Eds.)

doi:10.3233/978-1-61499-558-6-146

Improving Video Based Heart Rate Monitoring

Jian LIN[a], David ROZADO[b], Andreas DUENSER[b]
[a] *School of Information Technol ogy and Electrical Engineering*
University of Queensland, Australia
[b]*CSIRO, Digital Productivity, Australia*

Abstract. Non-contact measurements of cardiac pulse can provide robust measurement of heart rate (HR) without the annoyance of attaching electrodes to the body. In this paper we explore a novel and reliable method to carry out video-based HR estimation and propose various performance improvement over existing approaches. The investigated method uses Independent Component Analysis (ICA) to detect the underlying HR signal from a mixed source signal present in the RGB channels of the image. The original ICA algorithm was implemented and several modifications were explored in order to determine which one could be optimal for accurate HR estimation. Using statistical analysis, we compared the cardiac pulse rate estimation from the different methods under comparison on the extracted videos to a commercially available oximeter. We found that some of these methods are quite effective and efficient in terms of improving accuracy and latency of the system. We have made the code of our algorithms openly available to the scientific community so that other researchers can explore how to integrate video-based HR monitoring in novel health technology applications. We conclude by noting that recent advances in video-based HR monitoring permit computers to be aware of a user's psychophysiological status in real time.

Keywords. Heart rate monitoring, Webcam, Psycho-physiology, Signal processing

Introduction

Cardiovascular pulse measurements have been utilized as a regular physiological signal in health studies for years. Resting heart rate (HR), as one of the typical factors, has been found to be associated with psycho-physiological status and cardiovascular disease mortality [6]. HR normally is obtained by taking a pulse manually and / or attaching special devices or electrodes to the subject (to the chest, wrist, finger, etc.). This can be inconvenient for the user in terms of movement restrictions and setup times or may require a break in task performance. Recently, various methods using non-contact, non-invasive devices to obtain pulse measurement have been developed with varying degrees of accuracy. Poh et al. [9] described a video-based method to measure HR by applying the Blind Source Separation algorithm on time series data. A different technique, Eulerian Video Magnification, was developed by [13] for the same purpose. Such contactless HR measurements may overcome some of the limitations and associated inconveniences present in other measurement tools that require some sort of physical contact. They offer exiting new ways of monitoring physiological signals with potential application areas ranging from medical applications to new ways of selfmonitoring. Similar to other emerging monitoring technologies such as gaze tracking [10], body gestures [3] or brain monitoring systems [2], video based HR or HR variability may be another useful tool for designing new health technologies and Human-Computer-Interaction paradigms. Furthermore, video cameras are relatively cheap and already available on many consumer devices.

The key methodology proposed by [9] uses each channel (Red, Green, and Blue) of the image from a video stream and processes it as an independent time-series signal. The algorithm used, Independent Component Analysis (ICA), reveals statistically independent source signals from the set of observations that are composed of linear mixtures of the underlying sources. The three RGB channels are treated as three independent signal sources. The output of ICA does not follow the order that it was fed in. In other words, the outcomes will be in random order and the hidden HR signal can be extracted out from the three components. The application of ICA in signal analysis is growing fairly quickly recently with numerous applications in EEG, speech recognition and other signal classification problems [4, 7, 12].

A recurrent problem of using the technique is the high level of noise in the source and the ICA derived signals. One technique for noise removal against false positives is the Kalman filter [5], which is a set of mathematical equations that provides an efficient recursive solution of the least-squares method. It provides estimations of past, present and future states, even without any prior information about the precise nature of the modeled system [11].

This work aims to investigate the performance of ICA in non-contact HR measurement using a video stream as suggested by Poh et al. [9], and to explore how to improve its accuracy and reliability by introducing several enhancements: skin color filter, raw data detrending and a Kalman filter into the methodology. By testing the different algorithms on actual video recordings, the performance of the original algorithm and our proposed modified methodology can be compared in a quantifiable way. Furthermore, an open-source application has been built for real-time measurement using the explored algorithm based in C++ and using the libraries OPENCV and QT[1].

1. Methodology and experimental design

The goal is to detect the HR of humans by monitoring and amplifying the hidden signal (blood flow) underneath the skin as captured by a webcam, and the region of interest is the human face. Therefore face detection and a skin color filter were implemented in this work for excluding the background and unwanted pixels in the frame. The time series of a captured video are stored in three channels (RGB), each channel is treated as a signal source and thus there are three signal sources as input for the independent component analysis algorithm. After being processed by the ICA algorithm, three new independent components are extracted out of the three RGB channels in random order. A frequency spectrum analysis is applied on each component and according to Poh et al. [9], the maximum amplitude of three components corresponds to the pulse of the subject. However, the result of this approach fluctuates considerably due to non-constant illumination or incomplete Fast Fourier transform, and the maximum amplitude component is not always the best estimator.

[1] https://github.com/PugKing1988/finetuning-bss-based-remote-pulse-measurement-method

Applying a Kalman filter in this methodology could improve the stability. It first generates estimates of the current state variables, along with their errors. Secondly, when the next measurement is computed or produced, these estimates are modified using a weighted average, which gives more weight to estimates with high certainty. Due to its recursive nature, the algorithm is able to run in real time using only two input parameters: one measurement from the previously calculated state, and one for the present measurement.

1.1. Pulse Measurement Improvement Methodology

First, our approach reads images from video-files or webcams (for real-time application) and then an automated face tracker is used to detect faces within the video frames and records the coordinates of the measured region of interest (ROI). We use the Open Computer Vision library to acquire the coordinates of the face location [8].

The ROI is then processed by a skin color filter which only allows certain skin color pixels to be collected. This is also implemented through the use of the OPENCV library. Furthermore, those three RGB traces are normalized. New traces are fed into a data detrending function, which helps to enhance signal strength and pulse detection.

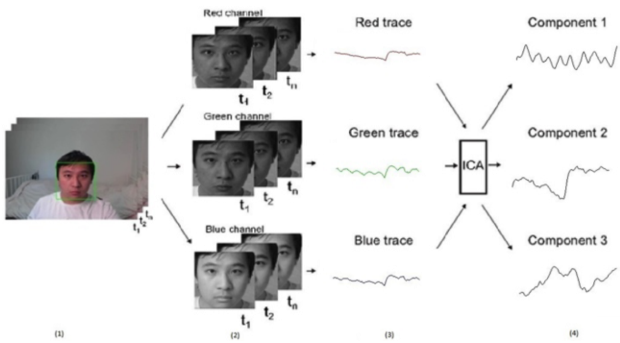

Figure 1: Pulse signal recovery method (1) The region of interest (ROI) is tracked with a face tracker (2) The ROI is separated into RGB channels and is normalized to obtain (3) the raw traces. Independent Component Analysis is applied on the normalized traces to acquire (4) three independent source signals

Independent Component Analysis is used on detrended traces to reveal the hidden signal which can be the subject's pulse [Figure 1]. We use the joint approximate diagonalisation of eigenmatrices (JADE) algorithm developed by Cardoso [1] to perform the ICA calculation. Since the output of ICA is in random order, the fast Fourier transform is performed on all of them to acquire three power spectra, the signal within the operational frequency range, which is 0.75 Hz with highest power density in each spectrum is recorded for each frame. The data are ranked in frequency from high to low and the pulse frequency is designated as the frequency with the highest power density among the top 20 points. This is derived from many observations of data scatter and called: pulse spectral maximum frequency-density selection (PSMFS).

Furthermore, a power density threshold is set to exclude false positive signals, which occurs frequently due to DC offset noise. It is set at 1.0 with a FFT size of 512 and another threshold is set for maximum change in pulse rate between successive measurements, which is taken 1 second apart. If the current pulse estimation has a power density less than 1, the difference between the current and the last computed value cannot exceed 7, or the current estimation will be rejected. Moreover, if the power density exceeded 1, the difference between current and the last computed value cannot exceed 12, or the current measurement will be neglected and the system retains the last pulse measurement. A Kalman filter then is used once a successful measurement was detected to predict the next HR estimation.

1.2. Experiment Setup

The experimental system is based on a desktop computer (Dell Optiplex 990) and a webcam (Microsoft LifeCam HD-5000) for recording video streams of the user sitting in front of the computer. A commercially available Pulse Oximeter (Model CMS60D by CONTEC) served as ground truth data source. We used the external library OPENCV for image processing and QT for its professional GUI development tools.

For our experiments, we captured a 3-mintue long video of human faces using a webcam with 30 frames per second and 640*480 pixels for each frame.

15 participants (12 males, 3 females), aged of 20-35, were recruited for this study. All participants were seated in front of the computer at a distance of approximately 0.5 m from the external webcam sitting on top of the computer screen. The participants were asked to stay relatively still and focus on the screen to ensure the face detection routine worked throughout the recording. Meanwhile, the pulse oximeter, which samples at 60Hz, was clipped on the fingers of the participants to monitor and record the actual HR simultaneously with the captured video. The experiments were all conducted indoors and with a constant amount of luminescent light as the only source of illumination. In addition, 2 participants were asked to complete 5 more two-minute recordings right after some intense exercise. We refer readers to a video[2] showing the system performing HR estimation in real time.

2. Results

The different approaches for HR estimation are compared with each other. The original approach as proposed by Poh et al. [9] had the largest errors and the PSMFS approach outweighs the original approach by 0.86 bps [Figure 2]. The PSMFS approach with detrending function works even better than just the PSMFS by itself. We should also note that the PSMFS approach combined with all of the proposed improvements had the best performance, which obtained only 2.73 bps offset from actual pulse. To identify whether there is a significant difference between these methods, an ANOVA showed a significant difference between the different modalities ($F(5,84)=2.04$, $p < .01$). A post-hoc analysis (with Bonferroni adjustment) suggests that there is an improvement in the PSMFS approach with Kalman filter + skin color filter + detrending compared to the single PSMFS approach or the original approach.

False positives caused by DC offset need to be filtered out by setting a signal density threshold, which causes a delay on locking up the true HR. From now on, we refer to this independent variable (the delay) as latency. We compared accuracy and latency with different thresholds (from 1.0 to 2.5) for the PSMFS approach with all the modifications. The accuracy for each threshold was relatively close and an ANOVA revealed no significant difference between the approaches. However, the results for

[2] A video showing the real-time application: http://youtu.be/kHuX550DYD0

latency had a proportional relationship with the power density threshold. The lower the threshold is, the faster it locks into a pulse.

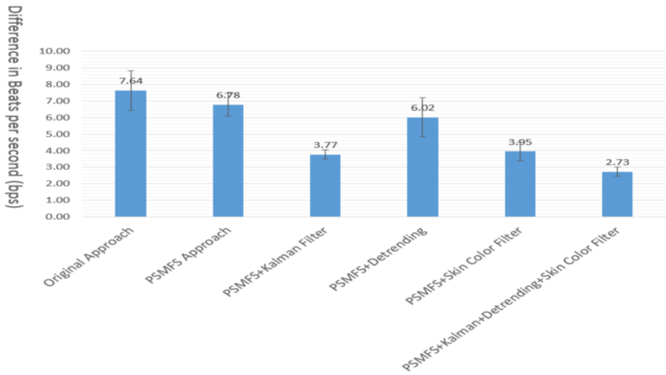

Figure 2: Average difference in beats per second between actual (as measured by a Pulse Oximeter) and estimated pulse for the different methods analyzed in this work.

The HR of people sitting quietly in front of a computer is usually in the range of 50-90 bpm. In order to test the performance of the algorithm to monitor HR of 100 bpm and higher, we measured the HR of two participants who had just completed intensive exercise in order to bring their HR up. The results of these pulse measurements were relatively poor and the average error in HR estimation exceeded 23.40 bps.

3. Discussion and Conclusion

As shown in [Figure 2], the original HR estimation method has the biggest difference in beats per second compared to the actual pulse. This was not expected because the original authors reported estimation within -0.05 bpm. An explanation for the lower performance of our implementation could be that in our case the blue component had two similar amplitude peaks at 0.75Hz and 1.03Hz within the operational frequency range. The actual pulse in the measurement was recorded as 73 bpm, and, according to the original approach, the designated pulse frequency was the peak at 0.75Hz. However, the correct pulse should correspond to the second peak at 1.03 Hz. A possible cause for this confusion may be the DC offset spectral frequency generated by the signal. It cannot be completely eliminated by using a band pass filter since the operational frequency range overlaps with the DC offset frequency range. It is worth pointing out that the original author used only natural light whereas our experiments were done in artificial light. Furthermore, the original authors also used a uniform white background for recording and we used a more cluttered background for all experiments.

Although the PSMFS approach with detrending and the original approach did not differ significantly, detrending had some effects in enhancing signal density. In certain videos, the 2.0 threshold was not exceeded at all throughout the experiment, however, by enabling the detrending function some signals reached levels above 2.0.

One limitation is that the webcam frame rate (30 fps) was relatively low, which requires at least 10 seconds to fill the FFT data array. If the light conditions are poor, the frame rate can drop to around 20 fps, which slows the process of filling up the FFT data array. Possible solutions for this problem are either to increase the sampling rate or decrease the size of FFT data array. Lowering the size of the FFT data array in theory will decrease the accuracy because it will enlarge the beat difference between adjacent bins. On the other hand, increasing frame rate would require the use of a high-speed camera, limiting the applicability of this technique for low-cost applications.

One possible reason for poor HR estimation after intensive exercise may be fluctuations in pulse. Because the HR is slightly decreasing over time, the designated frequency is difficult to detect because the fast Fourier transform is filled with different frequency elements. Frequency smearing could easily occur as those frequency elements are adjacent to each other. As a result of this, most recorded data within one second can be totally distorted by one, two bins away or even more.

Our prototype method works best if the user sits still facing the camera. It tolerates some degree of movement and side facing but performance degrades if the user carries out excessive head movements. It needs at least 5 seconds of video feed to start carrying out reliable HR estimations. We tested the method in uncontrolled standard light and we did not empirically evaluate performance under different light conditions. As our method was designed in the context of developing new Human-Machine Interaction paradigms rather than developing clinical applications, current limitations such as reliability and limited accuracy for estimating elevated HR may be of less. Still, these limitations should be addressed in future work.

We have implemented and evaluated several methods of fine-tuning a blind source separation based remote pulse measurement methodology. Moreover, we have shown how this methodology can be implemented in a real-time application with a regular webcam. Overall, our proposed approach shows great potential for reliable, low cost, non-contact HR monitoring.

References

[1] Cardoso, J, 1998 Multidimensional Independent Component Analysis. In Proceedings of the 1998 IEEE International Conference on Acoustics, *Speech and Signal Processing*, 1998 Pp. 1941–1944 vol.4.

[2] Delorme, Arnaud, Tim Mullen, Christian Kothe, et al. 2011. EEGLAB, SIFT, NFT, BCILAB, and ERICA: New Tools for Advanced EEG Processing. *Intell. Neuroscience* 2011: 10:10–10:10.

[3] Gunes, Hatice, and Massimo Piccardi 2007. Bi-Modal Emotion Recognition from Expressive Face and Body Gestures. *Journal of Network and Computer Applications* 30(4). 1334–1345.

[4] James, Christopher J, and Christian W Hesse 2005 Independent Component Analysis for Biomedical Signals. *Physiological Measurement* 26(1): R15–39.

[5] Kalman, RE 1960 A New Approach to Linear Filtering and Prediction Problems. *Transactions of the ASME – Journal of Basic Engineering* (82 (Series D)): 35–45.

[6] Kristal-Boneh, E, H Silber, G Harari, and P Froom 2000 The Association of Resting Heart Rate with Cardiovascular, Cancer and All-Cause Mortality. Eight Year Follow-up of 3527 Male Israeli Employees (the CORDIS Study). *European Heart Journal* 21(2): 116–124.

[7] Lee, Te-Won, M.S. Lewicki, and T.J. Sejnowski 2000 ICA Mixture Models for Unsupervised Classification of Non-Gaussian Classes and Automatic Context Switching in Blind Signal Separation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22(10): 1078–1089.

[8] Noulas, A. K., and B. J. A. Kröse 2006. EM Detection of Common Origin of Multi-Modal Cues. *Proc. 8th International Conf. on Multimodal Interfaces Pp. 201–208. ICMI '06.* New York, USA: ACM.

[9] Poh, Ming-Zher, Daniel J. McDuff, and Rosalind W. Picard 2010. Non-Contact, Automated Cardiac Pulse Measurements Using Video Imaging and Blind Source Separation. *Optics Express* 18(10): 10762.

[10] Rozado, David, Francisco Rodriguez, and Pablo Varona. 2010. Optimizing Hierarchical Temporal Memory for Multivariable Time Series. *Artificial Neural Networks–ICANN* 2010: 506–518.

[11] Welch, Greg, and Gary Bishop 1995. Chapel Hill, USA: University of North Carolina at Chapel Hill.

[12] Wisbeck, John Oersted, Allan Kardec Barros, Allan Kardec Barros Yy, and Renato Garcia Ojeda. Application of ICA in the Separation of Breathing Artifacts in ECG Signals. *The Fifth International Conference on Neural Information Processing, ICONIP'R98,* Kitakyushu, Japan, October 21-23, 1998,

[13] Wu, Hao-Yu, Michael Rubinstein, Eugene Shih, et al. 2012 Eulerian Video Magnification for Revealing Subtle Changes in the World. *ACM Trans. Graph*. 31(4): 65:1–65:8.