MIT Open Access Articles

Non-contact, automated cardiac pulse measurements using video imaging and blind source separation

The MIT Faculty has made this article openly available. ***Please share*** how this access benefits you. Your story matters.

Citation	Poh, Ming-Zher, Daniel J. McDuff, and Rosalind W. Picard. "Non-contact, automated cardiac pulse measurements using video imaging and blind source separation." Optics Express 18 (2010): 10762. ©2011 Optical Society of America.
As Published	http://dx.doi.org/10.1364/OE.18.010762
Publisher	Optical Society of America
Version	Final published version
Accessed	Tue Sep 26 04:11:34 EDT 2017
Citable Link	http://hdl.handle.net/1721.1/66243
Terms of Use	Article is made available in accordance with the publisher's policy and may be subject to US copyright law. Please refer to the publisher's site for terms of use.
Detailed Terms	

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

DSpace@MIT

Non-contact, automated cardiac pulse measurements using video imaging and blind source separation.

Ming-Zher Poh[1,2,*], Daniel J. McDuff[2], and Rosalind W. Picard[2]

[1]*Divison of Health Sciences and Technology, Harvard-MIT, 77 Massachusetts Avenue, Cambridge, MA 02139, USA*
[2]*The Media Laboratory, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA*
**zher@mit.edu*

Abstract: Remote measurements of the cardiac pulse can provide comfortable physiological assessment without electrodes. However, attempts so far are non-automated, susceptible to motion artifacts and typically expensive. In this paper, we introduce a new methodology that overcomes these problems. This novel approach can be applied to color video recordings of the human face and is based on automatic face tracking along with blind source separation of the color channels into independent components. Using Bland-Altman and correlation analysis, we compared the cardiac pulse rate extracted from videos recorded by a basic webcam to an FDA-approved finger blood volume pulse (BVP) sensor and achieved high accuracy and correlation even in the presence of movement artifacts. Furthermore, we applied this technique to perform heart rate measurements from three participants simultaneously. This is the first demonstration of a low-cost accurate video-based method for contact-free heart rate measurements that is automated, motion-tolerant and capable of performing concomitant measurements on more than one person at a time.

©2010 Optical Society of America

OCIS codes: (170.0170) Medical optics and biotechnology; (280.0280) Remote sensing and sensors.

References and links

1. S. Cook, M. Togni, M. C. Schaub, P. Wenaweser, and O. M. Hess, "High heart rate: a cardiovascular risk factor?" Eur. Heart J. **27**(20), 2387–2393 (2006).
2. I. Pavlidis, J. Dowdall, N. Sun, C. Puri, J. Fei, and M. Garbey, "Interacting with human physiology," Comput. Vis. Image Underst. **108**(1-2), 150–170 (2007).
3. M. Garbey, N. Sun, A. Merla, and I. Pavlidis, "Contact-free measurement of cardiac pulse based on the analysis of thermal imagery," IEEE Trans. Biomed. Eng. **54**(8), 1418–1426 (2007).
4. J. Fei, and I. Pavlidis, "Thermistor at a Distance: Unobtrusive Measurement of Breathing," IEEE Trans. Biomed. Eng. **57**(4), 988–998 (2010).
5. F. P. Wieringa, F. Mastik, and A. F. van der Steen, "Contactless multiple wavelength photoplethysmographic imaging: a first step toward "SpO2 camera" technology," Ann. Biomed. Eng. **33**(8), 1034–1041 (2005).
6. K. Humphreys, T. Ward, and C. Markham, "Noncontact simultaneous dual wavelength photoplethysmography: a further step toward noncontact pulse oximetry," Rev. Sci. Instrum. **78**(4), 044304 (2007).
7. C. Takano, and Y. Ohta, "Heart rate measurement based on a time-lapse image," Med. Eng. Phys. **29**(8), 853–857 (2007).
8. S. Hu, J. Zheng, V. Chouliaras, and R. Summers, "Feasibility of imaging photoplethysmography," in *Proceedings of IEEE Conference on BioMedical Engineering and Informatics* (IEEE, 2008), pp. 72–75.
9. W. Verkruysse, L. O. Svaasand, and J. S. Nelson, "Remote plethysmographic imaging using ambient light," Opt. Express **16**(26), 21434–21445 (2008).
10. J. Allen, "Photoplethysmography and its application in clinical physiological measurement," Physiol. Meas. **28**(3), R1–R39 (2007).
11. S. Rhee, B. H. Yang, and H. H. Asada, "Artifact-resistant power-efficient design of finger-ring plethysmographic sensors," IEEE Trans. Biomed. Eng. **48**(7), 795–805 (2001).
12. M. Z. Poh, N. C. Swenson, and R. Picard, "Motion-tolerant magnetic earring sensor and wireless earpiece for wearable photoplethysmography," IEEE Trans Inf Technol Biomed (Epub 2010 Feb).

#124173 - $15.00 USD Received 12 Feb 2010; revised 30 Apr 2010; accepted 6 May 2010; published 7 May 2010

13. P. Comon, “Independent component analysis, a new concept?” Signal Process. **36**(3), 287–314 (1994).
14. C. J. James, and C. W. Hesse, “Independent component analysis for biomedical signals,” Physiol. Meas. **26**(1), R15–R39 (2005).
15. M. P. Chawla, H. K. Verma, and V. Kumar, “Artifacts and noise removal in electrocardiograms using independent component analysis,” Int. J. Cardiol. **129**(2), 278–281 (2008).
16. T. P. Jung, S. Makeig, C. Humphries, T. W. Lee, M. J. McKeown, V. Iragui, and T. J. Sejnowski, “Removing electroencephalographic artifacts by blind source separation,” Psychophysiology **37**(2), 163–178 (2000).
17. J.-F. Cardoso, “Multidimensional independent component analysis,” in *Proceedings of IEEE Conference on Acoustics, Speech and Signal Processing* (IEEE, 1998), pp. 1941–1944.
18. M. J. McKeown, S. Makeig, G. G. Brown, T.-P. Jung, S. S. Kindermann, A. J. Bell, and T. J. Sejnowski, “Analysis of fMRI data by blind separation into independent spatial components,” Hum. Brain Mapp. **6**(3), 160–188 (1998).
19. Y. Jianchu, and S. Warren, “A Short Study to Assess the Potential of Independent Component Analysis for Motion Artifact Separation in Wearable Pulse Oximeter Signals,” in *Proceedings of IEEE Conference of the Engineering in Medicine and Biology Society* (IEEE, 2005), pp. 3585–3588.
20. B. S. Kim, and S. K. Yoo, “Motion artifact reduction in photoplethysmography using independent component analysis,” IEEE Trans. Biomed. Eng. **53**(3), 566–568 (2006).
21. W. G. Zijlstra, A. Buursma, and W. P. Meeuwsen-van der Roest, “Absorption spectra of human fetal and adult oxyhemoglobin, de-oxyhemoglobin, carboxyhemoglobin, and methemoglobin,” Clin. Chem. **37**(9), 1633–1638 (1991).
22. H. Trotter, “An elementary proof of the central limit theorem,” Arch. Math. **10**(1), 226–234 (1959).
23. A. Noulas, and B. Kr□se, “EM detection of common origin of multi-modal cues,” in *Proceedings of ACM Conference on Multimodal Interfaces* (ACM, 2006), pp. 201–208.
24. P. Viola, and M. Jones, “Rapid object detection using a boosted cascade of simple features,” in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 2001), p. 511.
25. R. Lienhart, and J. Maydt, “An extended set of Haar-like features for rapid object detection,” in *Proceedings of IEEE Conference on Image Processing* (IEEE, 2002).
26. J.-F. Cardoso, “High-order contrasts for independent component analysis,” Neural Comput. **11**(1), 157–192 (1999).
27. J. M. Bland, and D. G. Altman, “Statistical methods for assessing agreement between two methods of clinical measurement,” Lancet **1**(8476), 307–310 (1986).
28. R. C. Wuerz, D. Travers, N. Gilboy, D. R. Eitel, A. Rosenau, and R. Yazhari, “Implementation and refinement of the emergency severity index,” Acad. Emerg. Med. **8**(2), 170–176 (2001).
29. S. Munder, and D. M. Gavrila, “An experimental study on pedestrian classification,” IEEE Trans. Pattern Anal. Mach. Intell. **28**(11), 1863–1868 (2006).
30. B. Heisele, T. Serre, and T. Poggio, “A component-based framework for face detection and identification,” Int. J. Comput. Vis. **74**(2), 167–181 (2007).
31. H.-Y. Chen, C.-L. Huang, and C.-M. Fu, “Hybrid-boost learning for multi-pose face detection and facial expression recognition,” Pattern Recognit. **41**(3), 1173–1185 (2008).
32. W. J. Cui, L. E. Ostrander, and B. Y. Lee, “In vivo reflectance of blood and tissue as a function of light wavelength,” IEEE Trans. Biomed. Eng. **37**(6), 632–639 (1990).

1. Introduction

Regular and non-invasive assessments of cardiovascular function are important in surveillance for cardiovascular catastrophes and treatment therapies of chronic diseases. Resting heart rate, one of the simplest cardiovascular parameters, has been identified as an independent risk factor (comparable with smoking, dyslipidemia or hypertension) for cardiovascular disease [1]. Currently, the gold standard techniques for measurement of the cardiac pulse such as the electrocardiogram (ECG) require patients to wear adhesive gel patches or chest straps that can cause skin irritation and discomfort. Commercial pulse oximetry sensors that attach to the fingertips or earlobes are also inconvenient for patients and the spring-loaded clips can cause pain if worn over a long period of time. The ability to monitor a patient’s physiological signals by a remote, non-contact means is a tantalizing prospect that would enhance the delivery of primary healthcare. For example, the idea of performing physiological measurements on the face was first postulated by Pavlidis and associates [2] and later demonstrated through analysis of facial thermal videos [3,4]. Although non-contact methods may not be able to provide details concerning cardiac electrical conduction that ECG offers, these methods can now enable long-term monitoring of other physiological signals such as heart rate or respiratory rate by acquiring them continuously in an unobtrusive and comfortable manner. Beyond that, such a technology would also minimize the amount of cabling and clutter associated with neonatal ICU

#124173 - $15.00 USD Received 12 Feb 2010; revised 30 Apr 2010; accepted 6 May 2010; published 7 May 2010
(C) 2010 OSA

monitoring, long-term epilepsy monitoring, burn or trauma patient monitoring, sleep studies, and other cases where a continuous measure of heart-rate is important.

The use of photoplethysmography (PPG), a low cost and non-invasive means of sensing the cardiovascular pulse wave (also called the blood volume pulse) through variations in transmitted or reflected light, for non-contact physiological measurements has been investigated recently [5–9]. This electro-optic technique can provide valuable information about the cardiovascular system such as heart rate, arterial blood oxygen saturation, blood pressure, cardiac output and autonomic function [10]. Typically, PPG has always been implemented using dedicated light sources (e.g. red and/or infra-red wavelengths), but recent work [7,9] has shown that pulse measurements can be acquired using digital camcorders/cameras with normal ambient light as the illumination source. However, all these previous efforts lacked rigorous physiological and mathematical models amenable to computation; they relied instead on manual segmentation and heuristic interpretation of raw images with minimal validation of performance characteristics. Furthermore, PPG is known to be susceptive to motion-induced signal corruption [11,12] and overcoming motion artifacts presents one of the most challenging problems. In most cases, the noise falls within the same frequency band as the physiological signal of interest, thus rendering linear filtering with fixed cut-off frequencies ineffective. In order to develop a clinically useful technology, there is a need for ancillary functionality such as motion artifact reduction through efficient and robust image analysis.

One technique for noise removal from physiological signals is blind source separation (BSS). BSS refers to the recovery of unobserved signals or "sources" from a set of observed mixtures with no prior information about mixing process. Typically, the observations are acquired from the output of a set of sensors, where each sensor receives a different combination of the source signals. There are several methods of BSS and in this paper we will focus on BSS by Independent Component Analysis (ICA) [13]. ICA is a technique for uncovering the independent source signals from a set of observations that are composed of linear mixtures of the underlying sources. The use of this fairly new technique in biomedical signal analysis is rapidly expanding [14], e.g. in noise removal from electrocardiogram (ECG) [15] and electroencephalogram (EEG) recordings [16], separation of fetal and maternal ECGs recorded simultaneously [17], as well as detection of event related regions of activity in functional magnetic resonance imaging (fMRI) experiments [18]. ICA has also been applied to reduce motion artifacts in PPG measurements [19,20].

In this paper, we present a novel methodology for non-contact, automated, and motion-tolerant cardiac pulse measurements from video images based on blind source separation. Firstly, we describe our approach and apply it to compute heart rate measurements from video images of the human face recorded using a simple webcam. Secondly, we demonstrate how this method can tolerate motion artifacts and validate the accuracy of this approach with an FDA-approved finger blood volume pulse (BVP) measurement device. Thirdly, we show how this method can be easily extended for simultaneous heart rate measurements of multiple persons.

2. Methods

2.1 Study description and experimental setup

We used a basic webcam embedded in a laptop (built-in iSight camera on a Macbook Pro by Apple Inc.) to record the videos for analysis. All videos were recorded in color (24-bit RGB with 3 channels × 8 bits/channel) at 15 frames per second (fps) with pixel resolution of 640 × 480 and saved in AVI format on the laptop. 12 participants (10 males, 2 females) between the ages of 18-31 years were enrolled for this study that was approved by the Massachusetts Institute of Technology Committee On the Use of Humans as Experimental Subjects (COUHES). Our sample featured participants of both genders, different ages and with varying skin colors (Asians, Africans and Caucasians). Informed consent was obtained from all the participants prior to the start of each study session.

#124173 - $15.00 USD Received 12 Feb 2010; revised 30 Apr 2010; accepted 6 May 2010; published 7 May 2010
(C) 2010 OSA

For all experiments, an FDA-approved and commercially available blood volume pulse (BVP) sensor (Flexcomp Infiniti by Thought Technologies Ltd.) was used to measure the participant's BVP signal via a finger probe at 256 Hz for validation. The experiments were conducted indoors and with a varying amount of sunlight as the only source of illumination. Figure 1 show the experimental setup. Participants were seated at a table in front of a laptop at a distance of approximately 0.5 m from the built-in webcam. Two videos, each lasting one-minute, were recorded for all participants. During the first video recording, participants were asked to sit still and stare at the webcam. For the second video recording, participants were asked to move naturally as if they were interacting with the laptop, but to avoid large or rapid motions and to keep the hand wearing the finger BVP sensor still. In addition, we recorded a single, one-minute video of three participants sitting together at rest.

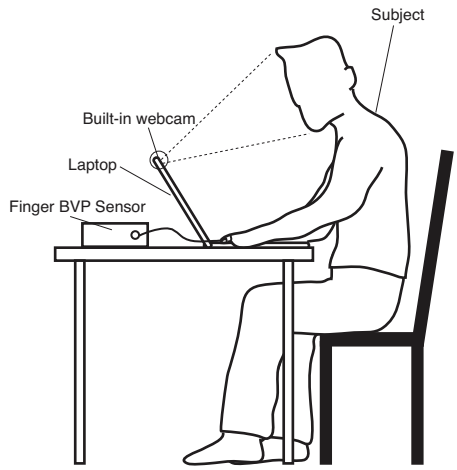

Fig. 1. Experimental setup.

2.2 Independent component analysis (ICA)

In this study, the underlying source signal of interest is the cardiovascular pulse wave that propagates throughout the body. Volumetric changes in the facial blood vessels during the cardiac cycle modify the path length of the incident ambient light such that the subsequent changes in amount of reflected light indicate the timing of cardiovascular events. By recording a video of the facial region with a webcam, the RGB color sensors pick up a mixture of the reflected plethysmographic signal along with other sources of fluctuations in light due to artifacts such as motion and changes in ambient lighting conditions. Given that hemoglobin absorptivity differs across the visible and near-infrared spectral range [21], each color sensor records a mixture of the original source signals with slightly different weights. These observed signals from the red, green and blue color sensors are denoted by $x_1(t)$, $x_2(t)$ and $x_3(t)$ respectively, which are amplitudes of the recorded signals (averages of all pixels in the facial region) at time point *t*. In conventional ICA the number of recoverable sources cannot exceed the number of observations, thus we assumed three underlying source signals, represented by $s_1(t)$, $s_2(t)$ and $s_3(t)$. The ICA model assumes that the observed signals are linear mixtures of the sources, i.e. $x_i(t) = \sum_{j=1}^{3} a_{ij} s_j(t)$ for each $i = 1,2,3$. This can be represented compactly by the mixing equation

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) \tag{1}$$

where the column vectors $\mathbf{x}(t) = [x_1(t), x_2(t), x_3(t)]^{\mathrm{T}}$, $\mathbf{s}(t) = [s_1(t), s_2(t), s_3(t)]^{\mathrm{T}}$ and the square 3×3 matrix $\mathbf{A}$ contains the mixture coefficients a_{ij}. The aim of ICA is to find a separating or

#124173 - $15.00 USD Received 12 Feb 2010; revised 30 Apr 2010; accepted 6 May 2010; published 7 May 2010
(C) 2010 OSA

demixing matrix **W** that is an approximation of the inverse of the original mixing matrix **A** whose output

$$\hat{\mathbf{s}}(t) = \mathbf{W}\mathbf{x}(t) \tag{2}$$

is an estimate of the vector $\mathbf{s}(t)$ containing the underlying source signals. According to the central limit theorem [22], a sum of independent random variables is more Gaussian than the original variables. Thus, to uncover the independent sources, **W** must maximize the non-Gaussianity of each source. In practice, iterative methods are used to maximize or minimize a given cost function that measures non-Gaussianity such as kurtosis, negentropy or mutual information.

2.3 Pulse measurement methodology

Post processing and analysis of both the video and physiological recordings were done using custom software written in MATLAB (The MathWorks, Inc.). An overview of the general steps in our approach to recovering the blood volume pulse is illustrated in Fig. 2. First, an automated face tracker was used to detect faces within the video frames and localize the measurement region of interest (ROI) for each video frame [Fig. 2(a)]. We utilized a free MATLAB-compatible version of the Open Computer Vision (OpenCV) library to obtain the coordinates of the face location [23]. The OpenCV face detection algorithm is based on work by Viola and Jones [24], as well as Lienhart and Maydt [25]. A cascade of boosted classifier uses 14 Haar-like digital image features trained with positive and negative examples. The pre-trained frontal face classifier available with OpenCV 2.0 was used. The cascade nature uses a set of simple classifiers that are applied to each area of interest sequentially. At each stage a classifier is built using a weighted vote, known as boosting. Either all stages are passed, meaning the region is likely to contain a face, or the area is rejected. The dimensions of the area of interest are changed sequentially in order to identify positive matches of different sizes. For each face detected, the algorithm returns the x- and y-coordinates along with the height and width that define a box around the face. From this output, we selected the center 60% width and full height of the box as the ROI for our subsequent calculations. To prevent face segmentation errors from affecting the performance of our algorithm, the face coordinates from the previous frame were used if no faces were detected. If multiple faces were detected when only one was expected, then our algorithm selected the face coordinates that were the closest to the coordinates from the previous frame.

The ROI was then separated into the three RGB channels [Fig. 2(b)] and spatially averaged over all pixels in the ROI to yield a red, blue and green measurement point for each frame and form the raw traces $x_1(t)$, $x_2(t)$ and $x_3(t)$ respectively [Fig. 2(c)]. Subsequent processing was performed using a 30 s moving window with 96.7% overlap (1 s increment). We normalized the raw RGB traces as follows:

$$x_i'(t) = \frac{x_i(t) - \mu_i}{\sigma_i} \tag{3}$$

for each $i = 1,2,3$ where μ_i and σ_i are the mean and standard deviation of $x_i(t)$ respectively. The normalization transforms $x_i(t)$ to $x_i'(t)$ which is zero-mean and has unit variance.

The normalized raw traces are then decomposed into three independent source signals using ICA [Fig. 2(d)]. In this report, we used the joint approximate diagonalization of eigenmatrices (JADE) algorithm developed by Cardoso [26]. This approach by tensorial methods uses fourth-order cumulant tensors and involves the joint diagonalization of cumulant matrices; the solution of this approximates statistical independence of the sources (to the fourth order). Although there is no ordering of the ICA components, the second

#124173 - $15.00 USD Received 12 Feb 2010; revised 30 Apr 2010; accepted 6 May 2010; published 7 May 2010
(C) 2010 OSA

component typically contained a strong plethysmographic signal. For the sake of simplicity and automation, we always selected the second component as the desired source signal.

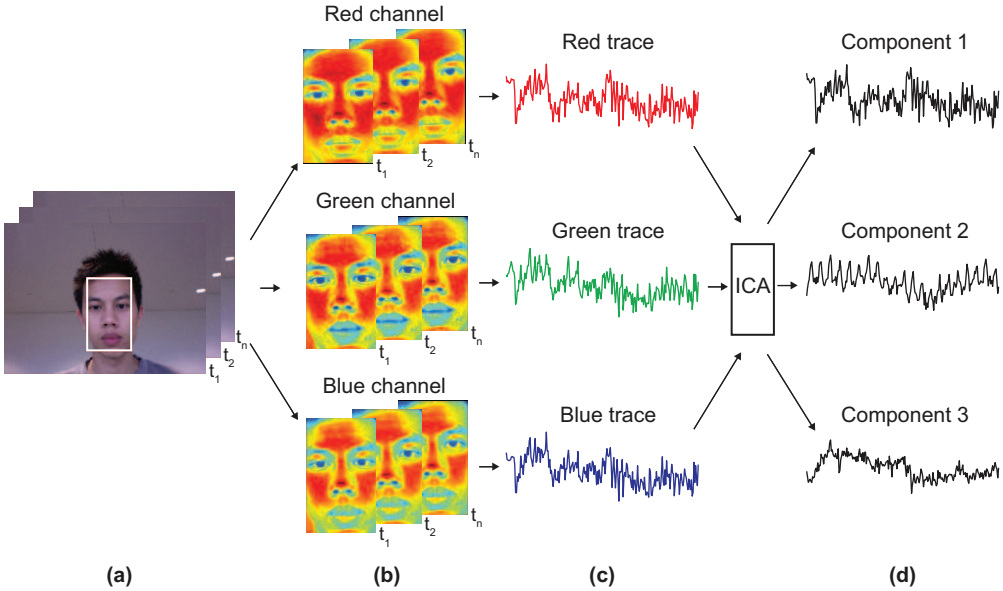

Fig. 2. Cardiac pulse recovery methodology. (a) The region of interest (ROI) is automatically detected using a face tracker. (b) The ROI is decomposed into the RGB channels and spatially averaged to obtain (c) the raw RGB traces. ICA is applied on the normalized RGB traces to recover (d) three independent source signals.

Finally, we applied the fast Fourier transform (FFT) on the selected source signal to obtain the power spectrum. The pulse frequency was designated as the frequency that corresponded to the highest power of the spectrum within an operational frequency band. For our experiments, we set the operational range to [0.75, 4] Hz (corresponding to [45, 240] bpm) to provide a wide range of heart rate measurements. Similarly, we obtained the reference heart rate measurements from the recorded finger BVP signal using the same steps.

Despite the application of ICA in our proposed methodology, the pulse frequency computation may occasionally be affected by noise. To address this issue, we utilize the historical estimations of the pulse frequency to reject artifacts by fixing a threshold for maximum change in pulse rate between successive measurements (taken 1 s apart). If the difference between the current pulse rate estimation and the last computed value exceeded the threshold (we used a threshold of 12 bpm in our experiments), the algorithm rejected it and searched the operational frequency range for the frequency corresponding to the next highest power that met this constraint. If no frequency peaks that met the criteria were located, then the algorithm retained the current pulse frequency estimation.

2.4 Statistics

Bland Altman plots [27] were used for combined graphical and statistical interpretation of the two measurement techniques. The differences between estimates from ICA and the Flexcomp finger BVP sensor were plotted against the averages of both systems. The mean and standard deviation (SD) of the differences, mean of the absolute differences and 95% limits of agreement (± 1.96 SD) were calculated. The root mean squared error (RMSE), Pearson's correlation coefficients and the corresponding p-values were calculated for the estimated heart rate from ICA and the finger BVP. In addition, we calculated the false positive rate as the total number of segmentations yielding more than one face over the total number of frames segmented in the single-participant experiments. The false negative rate was computed

#124173 - $15.00 USD Received 12 Feb 2010; revised 30 Apr 2010; accepted 6 May 2010; published 7 May 2010
(C) 2010 OSA

as the total number of segmentations failing to return a face over the total number of frames segmented (all frames contained one face).

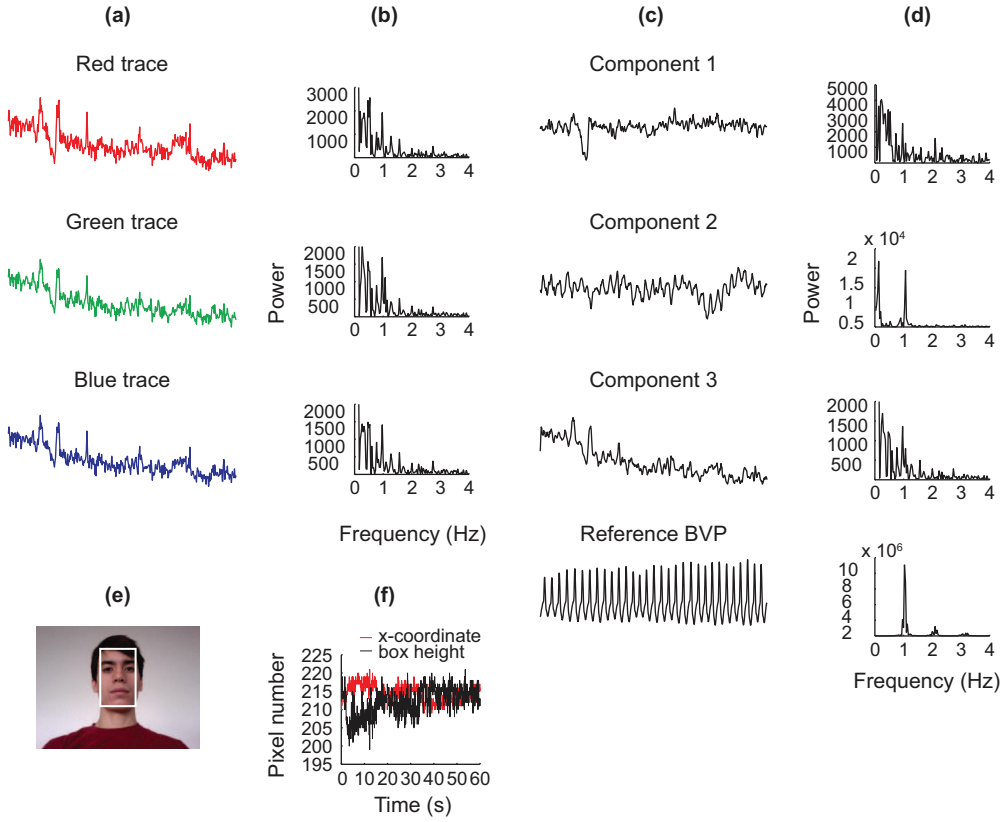

Fig. 3. Recovery of the cardiac pulse from a webcam video recording of a participant at rest. (a) 30 s raw RGB traces and (b) their respective power spectra. (c) The independent components recovered using ICA along with the reference finger BVP signal and (d) their respective power spectra. (e) (Media 1) A single-frame excerpt from the webcam video recording with localized ROI (white box). (f) Evolution of the localized ROI over 1 min.

3. Results

3.1 Heart rate measurements at rest

An example of recovering the cardiac pulse rate from a webcam video recording of a participant at rest is shown in Fig. 3 along with a 10 s portion of the corresponding video (Media 1). Figure 3(a) shows 30 s of the raw RGB traces obtained from the webcam video recording. We did not observe plethysmographic information in either the red, green or blue raw traces. However, from the three independent sources recovered by ICA [Fig. 3(c)], the cardiovascular pulse wave was clearly visible in the second component and in close agreement with the reference BVP signal. In the power spectrum of the reference BVP signal [Fig. 3(d)], a clear peak corresponding to the pulse frequency was seen at 1.03 Hz along with the 2nd and 3rd harmonics. The power spectrums of the raw RGB traces and ICA components all contained peaks around that frequency, but the power spectrum of the second ICA component yielded the highest SNR and the closest estimate of the pulse frequency (1.07 Hz). Figure 3(f) shows the evolution of the face tracker's x-coordinate and box height used to localize the ROI. It can be seen that both the location and height of the ROI fluctuated even though the participant tried to keep still during the experiment. Overall, the false positive and

#124173 - $15.00 USD Received 12 Feb 2010; revised 30 Apr 2010; accepted 6 May 2010; published 7 May 2010
(C) 2010 OSA

false negative rate was 0% over all one-minute recordings from 12 participants (a total of 10,800 frames).

To illustrate the effect of ICA, we first evaluated the accuracy of heart rate measurements obtained directly from the raw traces by designating the pulse frequency as the frequency that corresponded to the highest power (within the operational frequency band) of the raw green trace spectrum [Fig. 3(b)]. The green channel trace was chosen because it reportedly contains the strongest plethysmographic signal among all three channels [9]. When the agreement between 372 pairs of measurements from 12 participants was tested by Bland-Altman analysis [Fig. 4(a)], the mean bias $\bar{d}$ was 0.09 bpm with 95% limits of agreement −11.68 to 11.86 bpm. Using the proposed method with ICA to recover the heart rate reduced the error; $\bar{d}$ was −0.05 bpm with 95% limits of agreement −4.55 to 4.44 bpm [Fig. 4(b)]. The root mean square error (RMSE) was reduced nearly threefold from 6.00 (obtained from raw green channel trace before ICA) to 2.29 bpm and the correlation coefficient r increased from 0.89 to 0.98 ($p < 0.001$ for both).

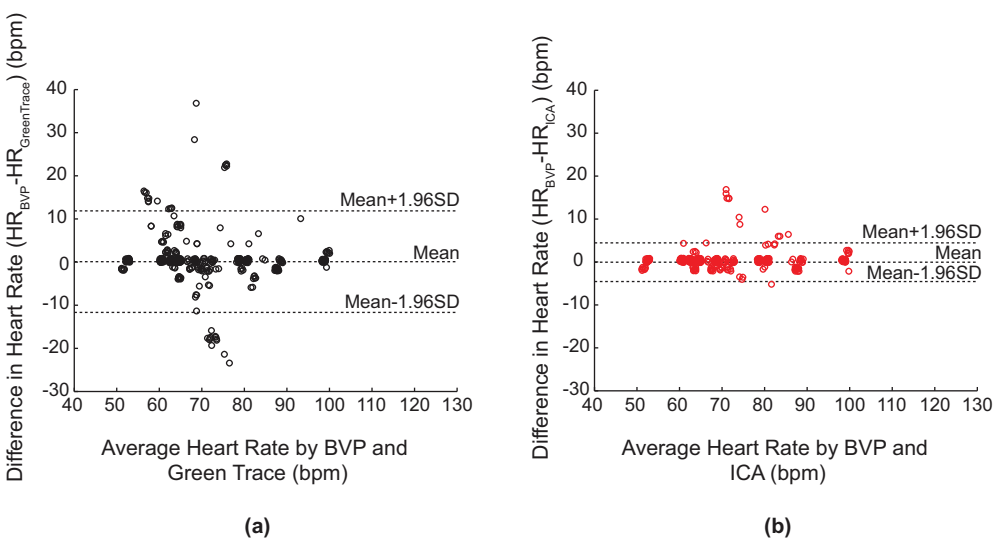

Fig. 4. Bland-Altman plots demonstrating the agreement between 30 s epoch heart rate measurements obtained from participants sitting at rest using finger BVP and (a) the raw green channel trace (black circles), (b) the proposed ICA method (red circles) (a total of 372 measurement pairs from 12 participants). The lines represent the mean and 95% limits of agreement.

3.2 Heart rate measurements during motion

We also evaluated the robustness of the proposed methodology for heart rate measurements in the presence of motion artifacts. In these experiments, participants were free to move their head or body slowly while remaining seated. Typical movements included tilting the head sideways, nodding the head, looking up/down and leaning forward/backward. Several participants also made various facial expressions, talked or laughed during the video recordings. An example of recovering the cardiac pulse rate from a webcam video recording of a moving participant is shown in Fig. 5 along with a 10 s portion of the corresponding video (Media 2). The plethysmographic signal was not visible in all three raw traces [Fig. 5(a)] and their respective power spectra were noisy with no clear indication of the pulse frequency [Fig. 5(b)]. In the power spectrum of the reference BVP signal [Fig. 5(d)], a clear peak corresponding to the pulse frequency was visible at 1.1 Hz along with the 2nd harmonic. From the recovered ICA components [Fig. 5(c)], oscillations similar to the reference BVP signal could be seen in component 2 although the signal was weaker than that obtained at rest. In the frequency domain, both components 2 and 3 exhibited a peak at the pulse frequency

#124173 - $15.00 USD Received 12 Feb 2010; revised 30 Apr 2010; accepted 6 May 2010; published 7 May 2010
(C) 2010 OSA

(1.1 Hz), but component 2 yielded a better SNR. As expected, the random movements of the participant resulted in large fluctuations in the x-coordinate of the ROI [Fig. 5(e)]. Overall, the false positive rate was 0.47% and the false negative rate was 0.01% over all one-minute recordings from 12 participants (a total of 10,800 frames).

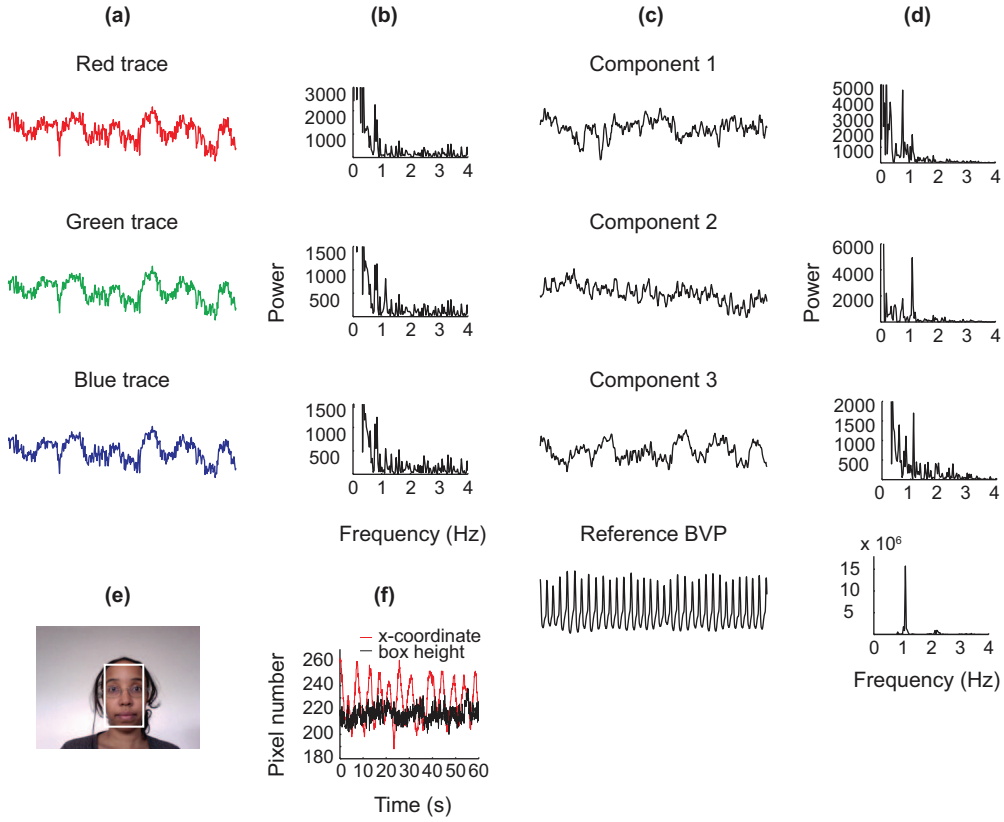

Fig. 5. Recovery of the cardiac pulse from a webcam video recording of a moving participant. (a) 30 s raw RGB traces and (b) their respective power spectra. (c) The independent components recovered using ICA along with the reference finger BVP signal and (d) their respective power spectra. (e) (Media 2) A single-frame excerpt from the webcam video recording with localized ROI (white box). (f) Evolution of the localized ROI over 1 min.

From the Bland-Altman analysis of 372 pairs of measurements from 12 participants, we see a significant difference in the distribution of points before and after ICA (Fig. 6). Using the raw green channel trace without ICA, $\bar{d}$ was 8.16 bpm with 95% limits of agreement −26.31 to 42.62 bpm [Fig. 6(a)], RMSE was 19.36 bpm and r was 0.15 ($p < 0.005$). After applying the proposed method with ICA, the points were distributed closer to zero and $\bar{d}$ was 0.64 bpm with 95% limits of agreement −8.35 to 4.63 bpm [Fig. 6(b)]. The RMSE was reduced fourfold to 4.63 bpm and r increased to 0.95 ($p < 0.001$).

#124173 - $15.00 USD Received 12 Feb 2010; revised 30 Apr 2010; accepted 6 May 2010; published 7 May 2010
(C) 2010 OSA

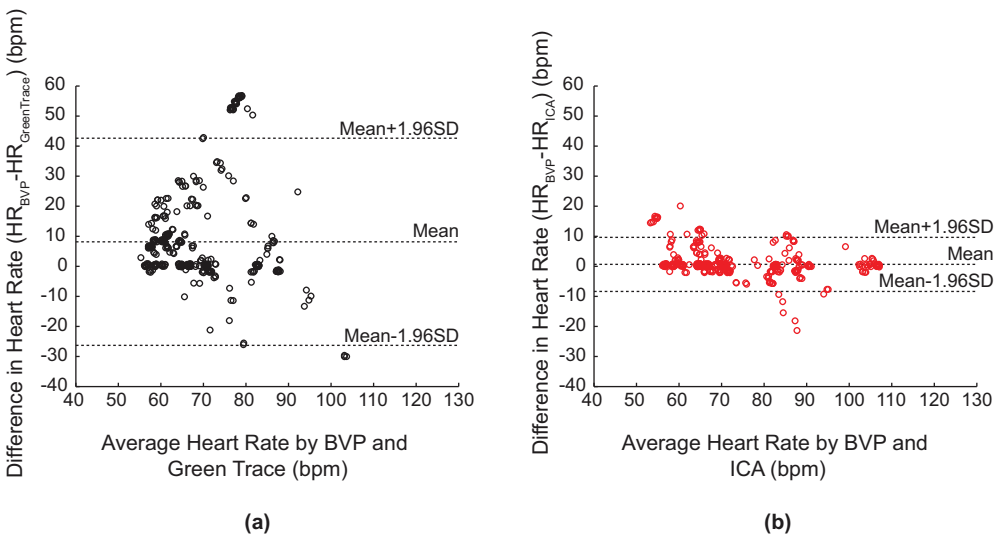

Fig. 6. Bland-Altman plots demonstrating the agreement between 30 s epoch heart rate measurements obtained from moving participants using finger BVP and (a) the raw green channel trace (black circles) and (b) the proposed ICA method (red circles) (a total of 372 measurement pairs from 12 participants). The lines represent the mean and 95% limits of agreement.

3.3 Simultaneous heart rate measurements of multiple participants

In order to demonstrate the capability of this proposed single methodology to perform concomitant heart rate measurements of multiple persons, we recorded a webcam video of three participants within the same field of view. Figure 7 shows the heart rate curves measured by the proposed technique (green, blue and red lines), as well as by reference BVP (black lines) for each individual participant along with a 10 s portion of the corresponding video (Media 3). The heart rate curves produced by our technique closely matched the finger BVP-derived heart rate curve throughout the experiment for all three participants [Fig. 6(b)]. RMSE for participants 1, 2 and 3 were 2.23, 2.66 and 4.56 bpm respectively.

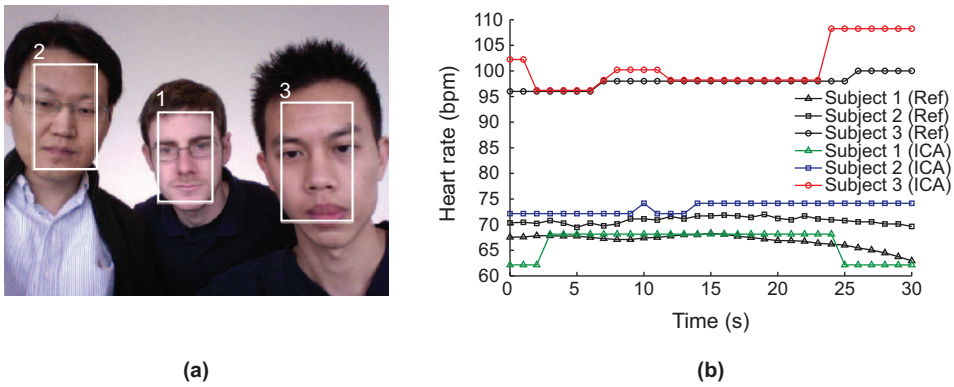

Fig. 7. Simultaneous heart rate measurements of multiple persons in a single webcam video recording. (a) (Media 3) A single-frame excerpt from the webcam video recording with the ROI for each participant highlighted (white boxes). (b) Comparison between heart rate measurements obtained using the proposed methodology (colored lines) and a reference BVP sensor (black lines).

#124173 - $15.00 USD Received 12 Feb 2010; revised 30 Apr 2010; accepted 6 May 2010; published 7 May 2010
(C) 2010 OSA

4. Discussion

It is possible that linearity assumed by ICA is not representative of the true underlying mixture in the signals given that the reflected light intensity varies nonlinearly with distance traveled through the facial tissue according to the Beer-Lambert law. In addition, the physiological changes in blood volume due to motion are not well understood and could also be nonlinear. Nonetheless, given the short time window used for ICA (30 s), a linear model should provide a reasonable local approximation. The results presented in this work verify the effectiveness of the proposed method for removing noise under these assumptions. Table 1 summarizes the descriptive statistics for critical evaluation of the proposed methodology compared to an FDA-approved finger BVP sensor. Overall, our technique showed very high agreement with BVP measurements when participants were sitting still ($\bar{d}$ = -0.05 ± 2.29 bpm) and also in the presence of motion artifacts ($\bar{d} = 0.64 \pm 4.59$ bpm). In both scenarios, ICA reduced the mean bias, standard deviation and RMSE as well as increased the level of correlation. This is significantly higher than the accuracy achieved using thermal imaging of the major superficial vessels ($\bar{d} = 4.74 \pm 9.28$ bpm based on the data presented by Garbey et al. [3]). Moreover, this level of accuracy was reproducible when extended to simultaneous heart rate measurements of multiple persons (RMSE < 5 bpm for all participants).

No approach is perfectly immune to all motion artifacts; thus, it is important to recognize the limitations of this study. The primary use of the methodology we described in this paper is likely to be in a home environment (e.g. telemedicine) given the ease of incorporating this novel pulse measurement technique into personal computers. As such, the motion artifacts we considered were in the context of a person interacting with a computer. The motion artifacts evaluated in our experiments were typically slow and relatively small movements such as tilting the head sideways, nodding the head, looking up/down, leaning forward/backward and talking. The average standard deviations of the x- and y- coordinates of the segmented face box were 35.1 and 13.16 pixels respectively in our motion experiments. For comparison, we analyzed one-minute video segments (randomly selected from recordings of 30 – 40 min) of 12 separate participants performing a variety of tasks on a computer including reading articles, watching videos and filling out web-based questionnaires (H. Ahn, unpublished data). In that independent experiment, the average standard deviations of the x- and y-coordinates of the segmented face box were 24.21 and 18.83 pixels respectively. Thus, the range of motion artifacts our methodology can overcome is realistic of a true assessment environment in the context of a person interacting with a computer and represents a significant advancement over previous studies on remote heart rate measurements. The fluctuations in estimated heart rate based on the peak frequency exceeded the threshold 7% of the time when participants were sitting still and 25% of the time in the presence of movement artifacts. These fluctuations usually occurred when there was an abrupt movement causing a large change in the baseline of the raw RGB signals. One possible method of improving performance would be to detrend the RGB signals prior to performing ICA. One could also detect abrupt large movements and ignore that data or use joint modeling of the motion and the signal to reestimate the heart rate at such times. The clinical acceptance of measurement error depends on the application. For example, vital signs play a role in the Emergency Severity Index (ESI) triage and a person with a heart rate above 100 bpm is considered to be in the "danger zone" [28]. In this scenario, a difference of 3 to 5 bpm (3 – 5% error) is likely to be acceptable especially given that the measurements can be performed remotely.

Another source of artifacts arises from the use of the automatic face tracker. From Fig. 2(e) it can be seen that even in the absence of movement artifacts, the localized ROI fluctuates due to tracking irregularities. In spite of this, ICA was able to recover the underlying blood volume pulse. Furthermore, our results indicate that a simple, inexpensive webcam (1.3 megapixels with a plastic lens and fixed focus) is sufficient for video capture. Webcams with similar specifications cost under $40 and many laptops already have built-in

#124173 - $15.00 USD Received 12 Feb 2010; revised 30 Apr 2010; accepted 6 May 2010; published 7 May 2010
(C) 2010 OSA

webcams. In this study, we did not address how the proposed method will fare in low lighting (e.g. in a dark room), but our experiments were conducted at different times of day with different degrees of ambient illumination (the average grayscale luminance of the ROI ranged from 61 to 125). The performance of our technique did not vary significantly within this range of luminance, but we expect the SNR of the recovered plethysmographic signal to decrease in dim light.

There are limitations to identifying faces with the pre-trained frontal face OpenCV classifier. Firstly, the system identified a number of false positives in the video clips due to background artifacts. With limited motion false negatives were uncommon. However, the OpenCV classifier would potentially give rise to a greater number of false negatives if movement were increased significantly. Tilting and turning of the head increased the likelihood of false negatives. In order to improve performance a context specific classifier could be trained. Training with examples specific to this application would allow user-defined performance targets to be set [29]. This could improve both the detection rate of positive and rejection rate of negative examples. Alternatively, face detection algorithms that use skin color and multi-component frameworks could be used to improve detection rates [30,31]. Another area for improvement is in determining which ICA component to choose for spectral analysis. The second component typically, but not always contained the strongest pulse signal. It is unclear why this is the case but one might expect it to be related to the green channel and the fact that hemoglobin absorptivity is highest in green/yellow light. This is consistent with previous reports indicating that green/yellow light (510 – 590 nm) provides the greatest sensitivity to blood pulsations [9,32]. Although the simple method of always selecting the second component yielded good results in our experiments, further development in pattern recognition is needed to establish a more robust method for identifying the ICA component containing the strongest PPG signal. In addition, the recording time for this present work was relatively short and future work needs to extend the time window to enable long-term, continuous measurements.

Table 1. Descriptive Statistics of Heart Rate Measurements by Proposed Methodology and Reference BVP

	Sitting still		With movement artifacts	
Statistic	Before ICA	After ICA	Before ICA	After ICA
No. of measurement pairs	372	372	372	372
Mean bias (bpm)	0.09	−0.05	8.16	0.64
Mean absolute bias (bpm)	2.79	0.91	10.81	2.44
SD of bias (bpm)	6.01	2.29	17.58	4.59
Upper limit (bpm)	11.86	4.44	42.62	9.64
Lower limit (bpm)	−11.68	−4.55	−26.31	−8.35
RMSE	6.00	2.29	19.36	4.63
Correlation coefficient	0.89*	0.98*	0.15‡	0.95*

*Indicates significance at $p < 0.001$, ‡Indicates significance at $p < 0.005$

5. Conclusion

We have described, implemented and evaluated a novel methodology for recovering the cardiac pulse rate from video recordings of the human face and demonstrated an implementation using a simple webcam with ambient daylight providing illumination. To our knowledge, this is the first demonstration of a low-cost method for non-contact heart rate measurements that is automated and motion-tolerant. Moreover, we have shown how this approach is easily scalable for simultaneous assessment of multiple people in front of a camera. Given the low cost and widespread availability of webcams, this technology is promising for extending and improving access to medical care. Although this paper only addressed the recovery of the cardiac pulse rate, many other important physiological

#124173 - $15.00 USD Received 12 Feb 2010; revised 30 Apr 2010; accepted 6 May 2010; published 7 May 2010
(C) 2010 OSA

parameters such as respiratory rate, heart rate variability and arterial blood oxygen saturation can potentially be estimated using the proposed technique. Creating a real-time, multi-parameter physiological measurement platform based on this technology will be the subject of future work.

Acknowledgments

This work was funded by the MIT Media Lab Things That Think Consortium and by the Nancy Lurie Marks Family Foundation (NLMFF). The authors are grateful to Hyungil Ahn for generously sharing his video data on persons interacting with computers. The opinions expressed here are those of the authors and may or may not reflect those of the sponsoring parties.

#124173 - $15.00 USD Received 12 Feb 2010; revised 30 Apr 2010; accepted 6 May 2010; published 7 May 2010
(C) 2010 OSA