Aalborg Universitet

AALBORG UNIVERSITY
DENMARK

Estimation of heartbeat peak locations and heartbeat rate from facial video

Haque, Mohammad A.; Nasrollahi, Kamal; Moeslund, Thomas B.

Published in:
Image Analysis

DOI (link to publication from Publisher):
10.1007/978-3-319-59129-2_23

Publication date:
2017

Link to publication from Aalborg University

Citation for published version (APA):
Haque, M. A., Nasrollahi, K., & Moeslund, T. B. (2017). Estimation of heartbeat peak locations and heartbeat rate from facial video. In Image Analysis: 20th Scandinavian Conference, SCIA 2017, Tromsø, Norway, June 12–14, 2017, Proceedings, Part II (pp. 269-281). Springer. https://doi.org/10.1007/978-3-319-59129-2_23

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal -

Take down policy
If you believe that this document breaches copyright please contact us at vbn@aub.aau.dk providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from vbn.aau.dk on: March 29, 2024

Estimation of Heartbeat Peak Locations and Heartbeat Rate from Facial Video

Mohammad A. Haque, Kamal Nasrollahi, and Thomas B. Moeslund

Visual Analysis of People (VAP) laboratory, Aalborg University, Denmark
Emails: {mah, kn, tbm}@create.aau.dk

Abstract. Available systems for heartbeat signal estimations from facial video only provide an average of Heartbeat Rate (HR) over a period of time. However, physicians require Heartbeat Peak Locations (HPL) to assess a patient's heart condition by detecting cardiac events and measuring different physiological parameters including HR and its variability. This paper proposes a new method of HPL estimation from facial video using Empirical Mode Decomposition (EMD), which provides clearly visible heartbeat peaks in a decomposed signal. The method also provides the notion of both color- and motion-based HR estimation by using HPLs. Moreover, it introduces a decision level fusion of color and motion information for better accuracy of multi-modal HR estimation. We have reported our results on the publicly available challenging database MAHNOB-HCI to demonstrate the success of our system in estimating HPL and HR from facial videos, even when there are voluntary internal and external head motions in the videos. The employed signal processing technique has resulted in a system that could significantly advance, among others, health-monitoring technologies.

Keywords: heartbeat rate, facial video, head motion, facial skin color, empirical mode decomposition, multimodal fusion.

1 Introduction

Heartbeat signals represent Heartbeat Peak Locations (HPLs) in a temporal domain and help physicians assess the condition of a human cardiovascular system by detecting cardiac events and measuring different important physiological parameters such as Heartbeat Rate (HR) and its variability [1]. When the human heart pumps blood, subtle chromatic changes in the facial skin and slight head motion occur periodically. These changes and motion are associated with the periodic heartbeat signal and can be detected in a facial video [2].

Takano et al. first utilized the trace of skin color changes in facial video to extract heartbeat signal and estimate HR [3]. They recorded the variations in the average brightness of the Region of Interest (ROI) – a rectangular area on the subject's cheek – to estimate HR. About two years later, Poh et al. proposed a method that used ROI mean color values as color traces from facial video acquired by a simple webcam, and employed Independent Component Analysis (ICA) to separate the periodic signal sources and a frequency domain analysis of an ICA component to measure HR [4]. Kwon et al. improved Poh's method in [4] by using merely green color channel instead of all three Red-Green-Blue (RGB) color channels [5]. Wei et al. employed a Laplacian Eigenmap (LE) and Cheng et al. employed an Empirical Mode Decomposition (EMD), rather than ICA, to obtain the uncontaminated heartbeat signal and demonstrated better

© Springer-Verlag Berlin Heidelberg 2011

performance than the ICA-based method [6], [7]. Other articles contributed a peripheral improvement of the color-based HR measurement by using a better estimation of ROI [8], adding a supervised machine learning component to the system [9], analyzing the distance between the camera and the face during data capture [10], and selecting best facial patches using majority voting [11].

Color-based methods suffer from tracking sensitivity to color noise and illumination variation. Thus, Balakrishnan et al. proposed a method for HR estimation which was based on invisible motion in the head due to pulsation of the heart muscles, which can be obtained by a Ballistocardiogram [12]. In this approach, some feature points were automatically selected on the ROI of the subject's facial video frames. These feature points were tracked by the Kanade–Lucas–Tomasi (KLT) feature tracker [13] to generate some trajectories, and then a Principle Component Analysis (PCA) was applied to decompose trajectories into a set of orthogonal signals based on Eigen values. Selection of the heartbeat rate was accomplished by using the percentage of the total spectral power of the signal, which accounted for the frequency with the maximum power and its first harmonic. A semi-supervised method in [14] was proposed to improve the results of Balakrishnan's method by using the Discrete Cosine Transform (DCT) along with a moving average filter rather than the Fast Fourier Transform (FFT) employed in Balakrishnan's work. The method in [15] also utilized motion information; however, unlike [12] it used ICA (previously used in color-based methods) to decompose the signal.

Though estimation of heartbeat signal from facial video was investigated a lot with different applications in the literature [2], [16]–[18], in view of the previous methods described above we summarized the following demands/challenges to investigate in this paper:

i. Previous methods provide an average HR over a certain time period, e.g. 30-60 seconds. Average HR alone is not sufficient to reveal some conditions of the cardiovascular system [12]. Health monitoring personnel often ask for a more detailed view of heartbeat signals with visible peaks that indicate the beats. Moreover, an important vital sign, heartbeat rate variability, cannot be obtained without locating heartbeat peaks in the signal. However, employing frequency domain decompositions along with some filters on the extracted color or motion traces from the facial video, as used in state-of-the-art methods of [4], [5], [8], [12], [14], [18], [19], does not provide visible HPLs for further analysis.
ii. The accuracy of HR estimation from facial video has yet to reach the level of ECG-based HR estimation. This compels investigations of a more effective signal processing method than the methods used in the literature to estimate HR.
iii. When a facial video is available, the beating of a heart typically shows in the face through changing skin color and head movement. Thus, merely estimating HR from color or motion information may be surpassed in accuracy by a fusion of these two modalities extracted from the same video.
iv. Most of the facial video-based fully automatic HR estimation methods, including color-based [3]–[5], [19] and motion-based [12], assume that the head is static (or close to) during data capture. This means that there is neither internal facial motion nor external movement or rotation of the head during the data acquisition phase.

© Springer-Verlag Berlin Heidelberg 2011

We ascribe internal motion to facial expression and external motion to head pose. However, in real life scenarios there are, of course, both internal and external head motion. Current methods, therefore, provide less accuracy in realistic scenarios.

In this paper we address the aforementioned demands/challenges by proposing a novel Empirical Mode Decomposition (EMD)-based approach to estimate HPL and then HR. Unlike previous methods, the proposed EMD-based decomposition of raw heartbeat traces provides a novel way to look into the heartbeat signal from facial video and generates clearly visible heartbeat peaks that can be used in, among others, further clinical analysis. We estimate the HR from both the number of peaks detected in a time interval and inter-beat distance in a heartbeat signal from HPLs obtained by employing the EMD. We then introduce a multimodal HR estimation from facial video by fusion of color and motion information and demonstrate the effectiveness of such a fusion in estimating HR. We report our results through a publicly available challenging database MAHNOB-HCI [20] in order to demonstrate the success of our system in estimating HR, even when there are voluntary internal and external head motions in the videos.

The rest of the paper is organized as follows. Section 2 describes the proposed system. Section 3 describes HR estimation from HPLs, and an approach to fusing color and motion. Section 4 presents the experimental results. Section 5 contains the conclusions.

2 The Proposed System

This section describes the steps of the proposed EMD-based HPL estimation method from color or motion traces as shown in Fig. 1.

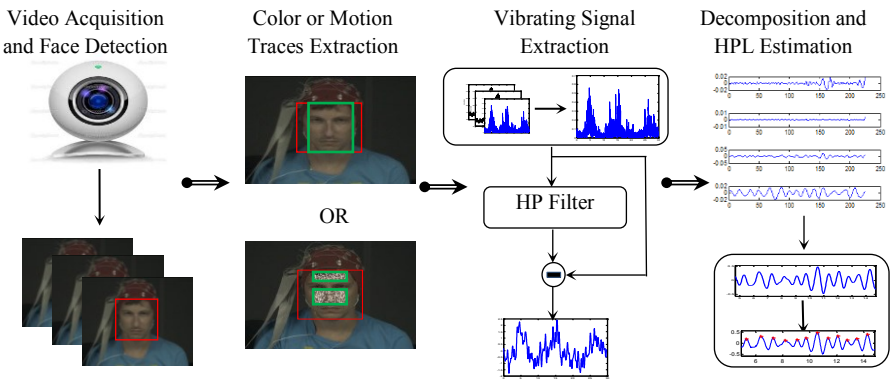


Fig. 1. Steps of the proposed HPL estimation method using skin color or head motion information from facial video.

2.1 Video Acquisition and Face Detection

The first step of the proposed HPL estimation system is face detection from facial video acquired by a simple RGB camera. By following [21], [22], we employ Haar-like feature-based Viola and Jones object detection framework [23] for face detection.

© Springer-Verlag Berlin Heidelberg 2011

2.2 Facial Color and Motion Traces Extraction

As mentioned earlier, periodic circulation of the blood from the heart through the body causes facial skin to change color, and the head to move or shake in a cyclic motion. The proposed system for HPL estimation can utilize either of the modalities (skin color variations and head motions) as shown in Fig. 1. Recoding RGB values of pixels in facial regions to generate the color trace and tracking some facial feature points to generate the motion trace help to record such skin color variation and head motion from facial video, respectively. In order to obtain the traces of either of the two modalities, we first select a ROI in the detected face. For the color-based approach, the ROI contains 60% of the facial area width (following [2]) detected by the automatic face detection method. We take the average of the RGB values of all pixels in the ROI in each frame instead of only either of the red, green or blue channels to defend the effect of the external lighting condition, as single color channel may prone to increased noise to specific lighting conditions.

For the motion-based approach the ROI (following [14]) contains two areas of forehead and cheek. We divide the ROI into a grid of rectangular regions of pixels and detect some feature points in each region by employing a method called Good Features to Track (GFT) [24], [25]. The GFT works by finding corner points from the minimal Eigen values of the windows of pixels in the ROI. When the head moves due to heart pulse, the feature points also move in the pixel coordinates. We employ a KLT tracker to track the feature points in consecutive video frames and obtain a single trajectory of each feature points in the video by measuring Euclidian distance of the point-locations in consecutive frames. We then fuse all trajectories into a single motion trace.

The next steps follow the same procedure for both color and motion and hereafter we refer to the raw heartbeat signal as $\bar{S}$.

2.3 Vibrating Signal Extraction

The raw heartbeat signal ($\bar{S}$) contains other extraneous high and low frequency cyclic components than heart beat due to ambient color and motion noise induced from the data capturing environment. It also exhibits non-cyclical trendy noise due to voluntary head motion, facial expression, and/or vestibular activity. Thus, to remove/reduce the extraneous frequency components and trends from the signal we decompose it using Hodrick-Prescott (HP) filter [26]. The filter breaks down the signal into the following components with respect to a smoothing penalty parameter, τ:

$$S_{\tau}^{log}(t) = T_{\tau}(t) + C_{\tau}(t) \quad (1)$$

where S_{τ}^{log} is the logarithm of $\bar{S}$, T_{τ} is the trend component, and C_{τ} denotes the cyclical component of the signal with t as the time index (video frame index). We empirically follow two times the decomposition of the trajectory by using two smoothing parameter values $\tau = \infty$ and $\tau = 400$ to obtain all of the cyclic components (C_{∞}) and high frequency cyclic components (C_{400}), respectively. A detailed description of the HP filter can be obtained from [26]. We completely overlook the trend components (T_{τ}) because these are not characterized by cyclic pattern of heartbeat. We then obtain the difference between the cyclical components to get the vibrating signal as follows:

© Springer-Verlag Berlin Heidelberg 2011

$$V(t) = C_{\infty}(t) - C_{400}(t) \quad (2)$$

2.4 EMD-based Signal Decomposition for the Proposed HPL Estimation

The vibrating signal (V) cannot clearly depict the heartbeat peaks (which will be shown in the experimental evaluation section). This is because of the contamination of heartbeat information by the other lighting and motion sources, which the HP filter alone cannot restore for visibility. Previous methods in [4], [5], [8], [12], [14], [19] moved to the frequency domain and filtered the signal by different bandpass filters and/or calculating the power spectrum of the signal to determine the HR in the frequency domain. However, this cannot provide a heartbeat signal with visible peaks, i.e. no possible HPL estimation, and thus cannot be useful for clinical applications where inter-beat intervals are necessary or signal variation needs to be observed over time. Thus, we employ a derivative of EMD to address the issue. EMD usually decomposes a nonlinear and non-stationary time-series into functions that form a complete and nearly orthogonal basis for the original signal [27]. The functions into which a signal is decomposed are all in the time domain and of the same length as the original signal. However, the functions allow for varying frequency in time to be preserved. When a signal is generated as a composite of multiple source signals and each of the source signals may have individual frequency band, calculating IMFs (Intrinsic Mode Functions) using EMD can provide illustratable source signals.

In the case of processing the heart signal obtained from facial video, the obtained vibrating signal (V) is a nonlinear and nonstationary time-series that comes as a composite of multiple source signals from lighting change, and/or internal and external head motions along with heartbeat. The basic EMD, as defined by Huang [28], breaks down a signal into IMFs satisfying the following two conditions:

i. In the whole signal, the number of extrema and the number of zero-crossings cannot differ by more than 1.
ii. At any point, both means of the envelopes defined by the local maxima and local minima are zero.

The decomposition can be formulated as follows:

$$V(t) = \sum_{i=1}^{m} M_i + r \quad (3)$$

where M_i presents the mode functions satisfying the aforementioned conditions, m is the number of modes, and r is the residue of the signal after extracting all the IMFs. The procedure of extracting such IMFs (M_i) is called shifting. The shifting process starts by calculating the first mean ($\mu_{i,0}$) from the upper and lower envelopes of the original signal (V in our case) by connecting local maxima. Then a component is calculated as the first component ($I_{i,0}$) for iteration as follows:

$$I_{i,0} = V(t) - \mu_{i,0} \quad (4)$$

The component $I_{i,0}$ is then considered the data signal for an iterative process, which is defined as follows:

$$I_{i,j} = I_{i,j-1} - \mu_{i,j} \quad (5)$$

© Springer-Verlag Berlin Heidelberg 2011

The iteration stops when a predefined value exceeds the following parameter (δ) calculated in each step:

$$\delta_{i,j} = \sum_{k=1}^{l} \frac{(I_{i,j-1}(k) - I_{i,j}(k))^2}{I_{i,j-1}^2(k)} \quad (6)$$

where l is the number of samples in I (in our case, the number of video frames used).

The basic model of EMD described above, however, exhibits some problems such as the presence of oscillations of very disparate amplitudes in a mode and/or the presence of very similar oscillations in different modes. In order to solve these problems an enhanced model of EMD was proposed by Torres et al. [29], which is called Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN). CEEMDAN adds a particular noise at each stage of the decomposition and then computes residue to generate each IMF. The results reported by Torres showed the efficiency of CEEMDAN over EMD. Therefore, we decompose our vibrating signal (V) into IMFs (M_i) by using the CEEMDAN. The total number of IMFs depends on the vibrating signal's nature. As a normal adult's resting HR falls within the frequencies $[0.75 - 2.0]$ Hz (i.e. $[45 - 120]$ bpm) [12] and merely 6-th IMF falls within this range, we selected the 6-th IMF as the final uncontaminated (or less contaminated) form of the heartbeat signal of all experimental cases.

We employ a local maxima-based peak detection algorithm on the selected heartbeat signal (the 6-th IMF) to estimate the HPL. The peak detection process was restricted by a minimum peak distance parameter to avoid redundant peaks in nearby positions. The obtained peak locations are the HPLs estimated by the proposed system.

3 HR Calculation Using the Proposed Multi-Modal Fusion

The HPLs we obtained in the previous section can be utilized to measure the total number of peaks and peak distances in a heartbeat signal. These can be obtained for either case of the color and motion information from facial video. We calculate the HR in bpm for both approaches in two different ways – from the total number of peaks and average peak distance – as follows:

$$HR_{numPeak} = \left(\frac{N \times F_{rate}}{F_{total}}\right) \times 60 \quad (7)$$

$$HR_{distPeak} = \left(\frac{F_{rate}}{\frac{1}{(N-1)}\sum_{k=1}^{(N-1)} d_k}\right) \times 60 \quad (8)$$

where N is the total number of peaks detected, F_{rate} is the video frame rate per second, F_{total} is the total number of video frames used to generate the heartbeat signal, and d_k is the distance between two consecutive peaks.

As we stated in the first section of this article, facial video contains both color and motion information that denote heartbeat. Along with the proposed EMD-based method, the applications of color information for HR estimation were shown in [4], [5], [8], [9], [19], and the applications of motion information were shown in [12], [14]. None of these

© Springer-Verlag Berlin Heidelberg 2011

methods exploited both color and motion information. We assume that, since color and motion information have different notions of heartbeat representation, a fusion of these two modalities in estimating HR can include more deterministic characteristics of heart pulses in the heartbeat signal.

There are three levels that can be considered for the fusion of modalities: raw-data level, feature level, and decision level [30]. Although the extracted raw heartbeat signals in color and motion-based approaches have the same dimensions, they are mismatched due to the nature of the data they present. Thus, instead of raw-data level and feature level fusion, we propose a rule-based decision level (HR estimation results) fusion in this paper for exploiting the HR estimation results from both modalities. For each of the modalities, we obtain two results using the total number of peaks and average peak distance. Thus, we have four different estimates of the HR: $HR_{numPeak}^{color}$, $HR_{distPeak}^{color}$, $HR_{numPeak}^{motion}$, and $HR_{distPeak}^{motion}$. We employed four feasible rules, listed in TABLE I, to find the optimal fusion.

TABLE I FUSION RULES INVESTIGATED

Parameter	Definition	Parameter	Definition
$HR_{numPeak}^{Fuse}$	$mean\left(HR_{numPeak}^{color}, HR_{numPeak}^{motion}\right)$	HR_{all}^{Fuse}	$mean\begin{pmatrix} HR_{numPeak}^{color}, HR_{distPeak}^{color}, \\ HR_{numPeak}^{motion}, HR_{distPeak}^{motion} \end{pmatrix}$
$HR_{distPeak}^{Fuse}$	$mean\left(HR_{distPeak}^{color}, HR_{distPeak}^{motion}\right)$	$HR_{nearestTwo}^{Fuse}$	$mean_{nearest\ two}\begin{pmatrix} HR_{numPeak}^{color}, HR_{distPeak}^{color}, \\ HR_{numPeak}^{motion}, HR_{distPeak}^{motion} \end{pmatrix}$

4 Experiments and Obtained Results

4.1 Experimental Environment

The proposed methods were implemented in Matlab (2013a). Most of the previous methods provided their results on local datasets, which makes the methods difficult to compare with the other methods. In addition, most of the previous methods did not report the results on a challenging database that includes realistic illumination and motion changes. In order to overcome such problems and show the competency of our methods, we used the publicly available MAHNOB-HCI database [20] for the experiment. The database is recorded in realistic Human-Computer Interaction (HCI) scenarios which was treated as a realistic and highly challenging dataset in the literature [8] because it contains facial videos recorded in realistic scenarios, including challenges from illumination variation and internal and external head motions. It contains videos of 491 sessions with 25 subjects that are longer than 30 seconds, and subjects who consent attribute 'YES'. Both males and females participated; they were between 19 and 40 years of age. Among the sessions, 20 sessions of subject '12' do not have ECG ground truth data and 20 sessions of subject '26' are missing video data. Excluding these sessions, we used the remainder as the dataset for our experiment. As the original videos are of different lengths, we use 30 seconds (frame 306 to 2135) of each video and the corresponding ECG from EXG3 sensor for the ground truth defined in [20].

We show the experimental results for HPL estimation in a qualitative manner and HR estimation through four statistical parameters used in the previous literature [4], [8]. The

© Springer-Verlag Berlin Heidelberg 2011

first one is mean error, defined as follows:

$$M_E = \frac{1}{N}\sum_{k=1}^{N}\left(HR_k^{video} - HR_k^{groundTruth}\right) \quad (9)$$

where HR_k^{video} is the calculated HR from the k-th video of a database, $HR_k^{groundTruth}$ is the corresponding HR from the ECG ground truth signal, and N is the total number of videos. The second parameter is the standard deviation of M_E, defined as follows:

$$SD_{M_E} = \sqrt{\left(\frac{1}{N}\sum_{k=1}^{N}\left(HR_k^{video} - M_E\right)^2\right)} \quad (10)$$

The third parameter is the root mean squared error, defined as follows:

$$RMS_E = \sqrt{\left(\frac{1}{N}\sum_{k=1}^{N}\left(HR_k^{video} - HR_k^{groundTruth}\right)^2\right)} \quad (11)$$

The fourth parameter is the mean of error rate in percentage, defined as follows:

$$M_{ER} = \frac{1}{N}\sum_{k=1}^{N}\left(\frac{HR_k^{video} - HR_k^{groundTruth}}{HR_k^{groundTruth}}\right) \times 100 \quad (12)$$

4.2 Experimental Evaluation

The proposed method tracks color change and head motion due to heartbeat in a video. Fig. 2 shows the average trajectory ($\bar{S}$) calculated from the individual trajectories of the feature points and corresponding vibrating signal (V in eq. (2)) obtained after employing the HP filter for a video. We observe that the vibrating signal is less noisy than the previous signal due to the application of the HP filter. We obtain similar results in the color-based approach.

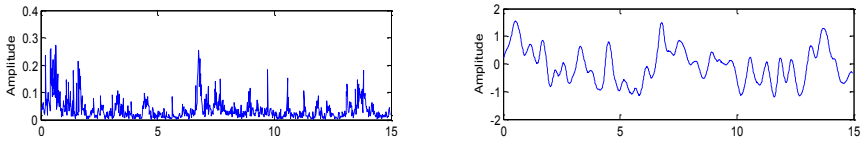


Fig. 2. Vibrating signal extraction by HP filtering: average signal from motion trajectories (left) and filtered vibrating signal (right).

The CEEMDAN, a derivative of EMD, decomposes the vibrating signal into IMFs (M_i) to provide an uncontaminated form of heartbeat signal. Fig. 3 shows first eight IMFs obtained from the signal by eq. (3)-(6). The IMFs are separated by different frequency components as discussed in Section 2.4, and we selected M_6 as the final heartbeat signal to employ the peak detection algorithm. The result of peak detection on M_6 is also shown in Fig. 3. One can observe that the final heartbeat signal has more clearly visible beats than the raw heartbeat signal obtained from motion traces. After employing peak detection we obtained all HPL that can be used in further medical analysis. The qualitative and quantitative comparison of the estimated HPL with the beat locations in ground truth ECG is shown in the performance comparison section.

We count the number of peaks and measure average peak distance from HPLs. The associated results are shown in TABLE II. We have indicated some of the best cases in

© Springer-Verlag Berlin Heidelberg 2011

bold. From the results we observe that counting the number of peaks provides better results than measuring peak distance for both color and motion information. This is because, unlike counting peaks, heartbeat rate variability in the signal can contribute negatively to the average peak distance. The overall error rates (M_{ER}) are less than 10% for HR estimation by counting the number of peaks for both motion and color signals. The fusion results show that, similar to the individual use of motion or color information, the number of peaks fusion generates the best results out of the four fusion rules. Simple arithmetic mean in decision level fusion, as we used, shows a strong correlation with the corresponding color and motion-based results. While comparing the results to the individual motion and color-based estimations, the fusion shows greater accuracy.

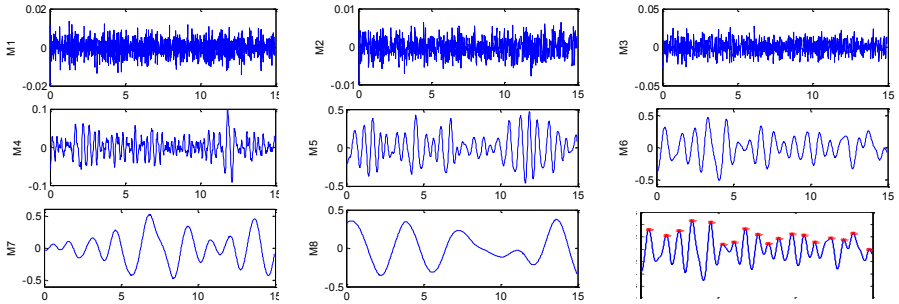


Fig. 3. Obtained IMFs (M_1-M_8) after employing CEEMDAN on the vibrating signal (V) and the detected heartbeat peaks in M_6.

TABLE II HR ESTIMATION RESULTS OF THE PROPOSED EMD-BASED METHODS USING COLOR, MOTION AND FUSION

No.	Method	M_E (bpm)	SD_{M_E} (bpm)	RMS_E (bpm)	M_{ER} (%)
1.	$HR_{numPeak}^{motion}$	***-0.90***	***8.28***	***8.32***	***8.65***
2.	$HR_{distPeak}^{motion}$	*-1.33*	*10.77*	*10.84*	*11.51*
3.	$HR_{numPeak}^{color}$	***0.21***	***8.55***	***8.54***	***9.26***
4.	$HR_{distPeak}^{color}$	*0.95*	*10.25*	*10.29*	*11.59*
5.	$HR_{distPeak}^{Fuse}$	*-0.19*	*10.08*	*10.07*	*11.00*
6.	HR_{all}^{Fuse}	*-0.27*	*9.04*	*9.03*	*9.79*
7.	$HR_{nearestTwo}^{Fuse}$	*-0.29*	*8.47*	*8.46*	*8.92*
8.	$HR_{numPeak}^{Fuse}$	***-0.35***	***8.08***	***8.08***	***8.63***

4.3 Performance Comparison

The performance of the proposed method has been compared with the relevant state of the art methods in two respects: i) presentation of visible heartbeat peaks in the extracted heartbeat signal in time domain and ii) accuracy of HR estimation. Fig. 4 shows the extracted heartbeat signals using the proposed EMD-based method from motion trajectories of two videos (subject ID-1, session 14 and subject ID-20, session 26) from the MAHNOB-HCI database next to the extracted signals using the motion-based method of [12]. The second video represents the case of voluntary facial motions (right). We have also included ground truth ECG for these videos in overlapping plots. From the figures we observe that the final time domain signal extracted by [12] is not plausible to

© Springer-Verlag Berlin Heidelberg 2011

comprehend for visual analysis. This is also true for the other methods because of similar filters and ICA/PCA/DCT-based decomposition. On the other hand, the final time domain signal generated by the proposed method not only shows heartbeat peaks but also preserves a good correspondence to the ECG ground truth.

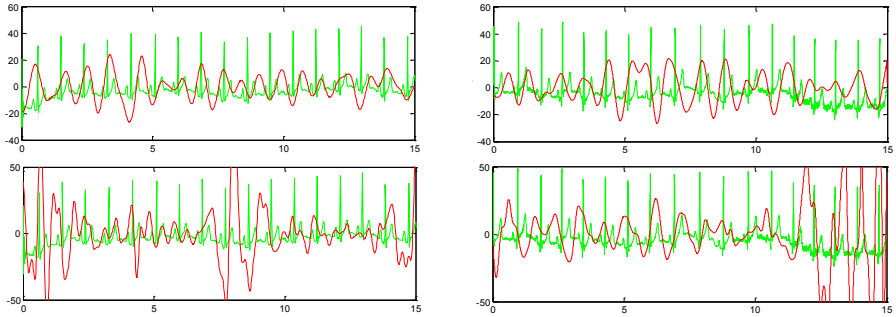


Fig. 4. Illustrating heartbeats obtained by different methods for two facial videos from two different subjects for normal case (left) and challenging case with voluntary motion (right): first row presents the results for our method and the second row presents the results for Guha2013 and Ramin2014 [12], [14]. Ground truth ECG is shown in green.

We compare the accuracy of the proposed method with state of the art color and the motion-based methods of [4], [5], [12], [19]. The results of the accuracy comparison are summarized in TABLE III. From the results, it is clear that the proposed EMD-based methods for both color and motion provide a better estimation of HR than the other state of the art methods. The proposed methods outperformed the other methods in both RMS_E and M_{ER} because EMD can decompose the signal in a better way than the filters and ICA/PCA-based decomposition used in the previous methods. The results of the proposed method demonstrate a high degree of consistency in estimating HR in comparison to the other methods. This, in turn, validates our peak location estimation as well because the peak locations have been used to estimate HR.

5 Conclusions

This paper proposed methods for estimating HPL and HR from color and motion information from facial video by a novel use of an HP filter and EMD decomposition. The paper also proposed a fusion approach to exploit both color and motion information together for multi-modal HR estimation. The contributions of these methods are as follows: i) provided the notion of visually analyzing heartbeat signal in time domain with clearly visible heartbeat peaks for clinical applications, ii) provided better estimations of HR for separate color and motion traces, iii) a decision level fusion further improved the result, and iv) provided a highly accurate HPL and HR estimations method from facial video in the presence of challenging situations due to illumination change and voluntary head motions. The proposed method, however, also imposed some limitations when generating the results. We assume that the camera will be placed in close proximity to the face (about one meter away). Moreover, we did not employ any sophisticated ROI detection and tracking methods, illumination rectification methods, or

© Springer-Verlag Berlin Heidelberg 2011

extraneous motion filtering. Future work should address these points.

TABLE III PERFORMANCE COMPARISON OF THE PROPOSED METHODS WITH THE PREVIOUS METHODS FOR HR ESTIMATION

No.	Method	M_E (bpm)	SD_{M_E} (bpm)	RMS_E (bpm)	M_{ER} (%)
1.	Poh (ICA) [4]	*-8.95*	*24.3*	*25.90*	*25.00*
2.	Kwon (ICA) [5]	*-7.96*	*23.8*	*25.10*	*23.60*
3.	Guha (PCA) [12]	*-14.4*	*15.2*	*21.00*	*20.70*
4.	Poh (ICA) [19]	*2.04*	*13.5*	*13.60*	*13.20*
5.	**Proposed_color (CEEMDAN)**	***0.21***	***8.55***	***8.54***	***9.26***
6.	**Proposed_motion (CEEMDAN)**	***-0.90***	***8.28***	***8.32***	***8.65***
7.	**Proposed_Fusion**	***-0.35***	***8.08***	***8.08***	***8.63***

References

[1] J. Klonovs *et al.*, *Distributed Computing and Monitoring Technologies for Older Patients*, 1st ed. Springer International Publishing, 2015.

[2] M. A. Haque, K. Nasrollahi, and T. B. Moeslund, "Heartbeat Signal from Facial Video for Biometric Recognition," in *Image Analysis*, R. R. Paulsen and K. S. Pedersen, Eds. Springer International Publishing, 2015, pp. 165–174.

[3] 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.

[4] M.-Z. Poh, D. J. McDuff, and R. W. Picard, "Non-contact, automated cardiac pulse measurements using video imaging and blind source separation," *Opt. Express*, vol. 18, no. 10, pp. 10762–10774, May 2010.

[5] S. Kwon, H. Kim, and K. S. Park, "Validation of heart rate extraction using video imaging on a built-in camera system of a smartphone," in *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2012, pp. 2174–2177.

[6] L. Wei, Y. Tian, Y. Wang, T. Ebrahimi, and T. Huang, "Automatic Webcam-Based Human Heart Rate Measurements Using Laplacian Eigenmap," in *Computer Vision – ACCV 2012*, K. M. Lee, Y. Matsushita, J. M. Rehg, and Z. Hu, Eds. Springer Berlin Heidelberg, 2012, pp. 281–292.

[7] J. Cheng, X. Chen, L. Xu, and Z. J. Wang, "Illumination Variation-Resistant Video-Based Heart Rate Measurement Using Joint Blind Source Separation and Ensemble Empirical Mode Decomposition," *IEEE J. Biomed. Health Inform.*, vol. PP, no. 99, pp. 1–1, 2016.

[8] X. Li, J. Chen, G. Zhao, and M. Pietikainen, "Remote Heart Rate Measurement From Face Videos Under Realistic Situations," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014, pp. 4321–4328.

[9] H. Monkaresi, R. . Calvo, and H. Yan, "A Machine Learning Approach to Improve Contactless Heart Rate Monitoring Using a Webcam," *IEEE J. Biomed. Health Inform.*, vol. 18, no. 4, pp. 1153–1160, Jul. 2014.

[10] A. Shagholi, M. Charmi, and H. Rakhshan, "The effect of the distance from the webcam in heart rate estimation from face video images," in *2015 2nd International Conference on Pattern Recognition and Image Analysis (IPRIA)*, 2015, pp. 1–6.

[11] A. Lam and Y. Kuno, "Robust Heart Rate Measurement from Video Using Select Random Patches," in *IEEE International Conference on Computer Vision (ICCV)*, 2015, pp. 3640–3648.

[12] G. Balakrishnan, F. Durand, and J. Guttag, "Detecting Pulse from Head Motions in Video," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2013, pp. 3430–3437.

© Springer-Verlag Berlin Heidelberg 2011

[13] J. Bouguet, “Pyramidal implementation of the Lucas Kanade feature tracker,” *Intel Corp. Microprocess. Res. Labs*, 2000.
[14] R. Irani, K. Nasrollahi, and T. B. Moeslund, “Improved Pulse Detection from Head Motions Using DCT,” in *9th International Conference on Computer Vision Theory and Applications (VISAPP)*, 2014, pp. 1–8.
[15] L. Shan and M. Yu, “Video-based heart rate measurement using head motion tracking and ICA,” in *2013 6th International Congress on Image and Signal Processing (CISP)*, 2013, vol. 01, pp. 160–164.
[16] M. A. Haque, K. Nasrollahi, and T. B. Moeslund, “Can contact-free measurement of heartbeat signal be used in forensics?,” presented at the 23rd European Signal Processing Conference (EUSIPCO), 2015, pp. 769–773.
[17] K. Nasrollahi, M. A. Haque, R. Irani, and T. B. Moeslund, “Contact-Free Heartbeat Signal for Human Identification and Forensics,” in *Handbook of Biometrics for Forensic Science*, 2017, pp. 289–302.
[18] M. A. Haque, R. Irani, K. Nasrollahi, and T. B. Moeslund, “Heartbeat Rate Measurement from Facial Video,” *IEEE Intell. Syst.*, vol. 31, no. 3, pp. 40–48, May 2016.
[19] 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.
[20] M. Soleymani, J. Lichtenauer, T. Pun, and M. Pantic, “A Multimodal Database for Affect Recognition and Implicit Tagging,” *IEEE Trans. Affect. Comput.*, vol. 3, no. 1, pp. 42–55, Jan. 2012.
[21] M. A. Haque, K. Nasrollahi, and T. B. Moeslund, “Real-time acquisition of high quality face sequences from an active pan-tilt-zoom camera,” in *10th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, 2013, pp. 443–448.
[22] M. A. Haque, K. Nasrollahi, and T. B. Moeslund, “Quality-Aware Estimation of Facial Landmarks in Video Sequences,” in *IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2015, pp. 1–8.
[23] P. Viola and M. J. Jones, “Robust Real-Time Face Detection,” *Int J Comput Vis.*, vol. 57, no. 2, pp. 137–154, May 2004.
[24] J. Shi and C. Tomasi, “Good features to track,” in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1994, pp. 593–600.
[25] M. A. Haque, R. Irani, K. Nasrollahi, and T. B. Moeslund, “Facial Video based Detection of Physical Fatigue for Maximal Muscle Activity,” *IET Comput. Vis.*, vol. 10, no. 4, pp. 323–329, 2016.
[26] T. McElroy, “Exact Formulas for the Hodrick-Prescott Filter.” Statistical Research Division, U.S. Census Bureau, Sep-2006.
[27] H. Ren, Y. Wang, M. Huang, Y. Chang, and H. Kao, “Ensemble Empirical Mode Decomposition Parameters Optimization for Spectral Distance Measurement in Hyperspectral Remote Sensing Data,” *Remote Sens.*, vol. 6, pp. 2069–2083, Mar. 2014.
[28] N. E. Huang, “An Adaptive Data Analysis Method for Nonlinear and Nonstationary Time Series: The Empirical Mode Decomposition and Hilbert Spectral Analysis,” in *Wavelet Analysis and Applications*, T. Qian, M. I. Vai, and Y. Xu, Eds. Birkhäuser Basel, 2006, pp. 363–376.
[29] M. E. Torres, M. A. Colominas, G. Schlotthauer, and P. Flandrin, “A complete ensemble empirical mode decomposition with adaptive noise,” in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2011, pp. 4144–4147.
[30] N. V. Boulgouris, K. N. Plataniotis, and E. Micheli-Tzanakou, *Biometrics: Theory, Methods, and Applications*. John Wiley & Sons, 2009.

© Springer-Verlag Berlin Heidelberg 2011