Multimedia Tools and Applications (2020) 79:23023–23043
https://doi.org/10.1007/s11042-020-09075-2

Exploring the feasibility of seamless remote heart rate measurement using multiple synchronized cameras

Juan Cheng[1] · Xingmao Wang[1] · Rencheng Song[1] · Yu Liu[1] · Chang Li[1] · Xun Chen[2]

Received: 26 June 2019 / Revised: 23 March 2020 / Accepted: 15 May 2020 /
Published online: 5 June 2020
© Springer Science+Business Media, LLC, part of Springer Nature 2020

Abstract

Heart rate (HR) measurement and monitoring is of great importance to determine the physiological and mental status of individuals. Recently, it has been demonstrated that HR can be remotely retrieved from facial video-based photoplethysmographic signals captured using consumer-grade cameras. However, in existing studies, subjects are mostly required to keep their facial regions of interest (ROIs) within one single camera. To make this technique usable in a daily life situation where subjects move around freely, we launch a preliminary simulated study of seamless remote HR measurement using multiple synchronized cameras by combining ensemble empirical mode decomposition (EEMD) with time-delay canonical correlation analysis (TDCCA), termed as EEMD-TDCCA. At each time point, a target ROI with the largest area is first determined from all the ROIs provided by all the cameras. Then, the RGB time sequence is formed by taking average of all pixels within each target ROI. Afterwards, the green channel time sequence is decomposed into several intrinsic mode functions (IMFs) and only the IMF candidates, whose frequency corresponding to the maximum amplitude falling into the interested HR range will be further processed by TDCCA. Finally, the first pair of the canonical variables having the largest correlation coefficient is the HR source and the corresponding HR is derived by peak detection or frequency analysis. Thirty subjects were recruited and four state-of-the-art methods were employed for comparison. The best performance was achieved by using the proposed EEMD-TDCCA followed by frequency analysis, with the mean absolute error 4.11 bpm, mean percentage error 5.26%, root mean square error 5.37 bpm, the Pearson's correlation coefficient 0.90 and the intra-class correlation coefficient 0.89, demonstrating the feasibility of our proposed seamless remote HR measurement framework. This study will provide a promising solution for practical and robust non-contact HR measurement applications.

✉ Rencheng Song
rcsong@hfut.edu.cn

Juan Cheng
chengjuan@hfut.edu.cn

[1] Department of Biomedical Engineering, Hefei University of Technology, Hefei, 230009, China

[2] Department of Electronic Engineering, Information Science, University of Science and Technology of China, Hefei 230026, China

Keywords Remote photoplethysmography (rPPG) · Multi-cameras · Non-contact · Target region of interest (ROI) · Heart rate

1 Introduction

Since heart is one of the most important organs of the body, the measurement and monitoring of heart rate (HR) are essential for clinical diagnosis and daily healthcare, e.g., the surveillance of cardiovascular catastrophes and the treatment therapies of chronic diseases. Various methods have been developed to estimate HR in contact or non-contact manners, and a recent review summarizes the relevant advances [42].

Photoplethysmography (PPG) is a well-known and non-invasive optical method to measure blood volume changes in the skin or in a small part of the body [28]. Traditionally, the light source of PPG is typically provided by a light-emitting diode (LED, either in the VIS and/or NIR frequency range) while the sensor is a photodetector (PD) that is contact to the body, *i.e.*, finger tip. Due to the contact characteristic, it will cause discomfort when wearing them for a long time, and is not suitable for the trauma or infectious patients to wear. Recently, a novel non-contact sensing method, namely remote PPG (rPPG), has spurred a remarkable number of studies. The rPPG measurement is based on the similar principle to that of the conventional PPG, which the pulsatile blood propagating in the cardiovascular system changes the blood volume in the microvascular tissue bed beneath the skin within each heart beat and consequently produces a quasi-periodical fluctuation [12]. Such pulse-induced fluctuation can be reflected by the subtle color variations measured from a distance of up to several meters using cameras under ambient illumination conditions [3, 20, 39]. The rPPG has been proven to be superior because of the no wearing convenience, low cost and widespread that may be suitable for neonatal intensive care unit monitoring, trauma patient monitoring, driver status assessment, and affective state assessment [1, 19, 21, 38, 53], *etc*. However, the rPPG signals are prone to be contaminated by motion artifact, a lot of efforts have been made to eliminate the influence. [2, 16, 32, 46, 47]

However, most of the aforementioned rPPG methods aiming to eliminate motion artifacts use one single camera, and the single camera may miss capturing the facial region of interest (ROI) when someone moves around. In this paper, with the aim to widen the range of rPPG applications, we launch a preliminary simulated study to employ multiple cameras to explore the feasibility of seamless remote HR measurement by combining ensemble empirical mode decomposition (EEMD) with time-delay canonical correlation analysis (TDCCA), termed as EEMD-TDCCA. CCA is able to find linear transformations and extract pairs of canonical variables (CVs) to reflect the correlations between two datasets, while the CVs extracted from each data set are mutually uncorrelated in transformation space [13]. The number of CV pairs can be equal to or lower than the smaller rank of the two datasets. Due to that the rPPG is quasi-periodical, the original HR source will be highly correlated to its time-delay version. In contract, the correlation between the noise source and its corresponding time-delay version is lower. Thereby, the time-delay CCA can be employed to extract CVs corresponding to HR source. Besides, considering that the types of noise can be multiple including motions, illuminations and the mixture caused by rotation, the three RGB channels are far from extract the inherent CVs. Thus, before TDCAA, we utilized EEMD to help decompose the green channel into a couple of intrinsic mode functions (IMFs) for pre-denoising. The corresponding performance is compared with other four state-of-the-art methods.

The main contributions of the proposed framework are as follows:

1. To our knowledge, it is the first time that a synchronized multi-camera framework is proposed to remotely and seamlessly measure HR during the simulated moving-around situation. The situation is that a subject passively rotates 180 degrees with an uniform speed by standing on a rotating turntable. It should be noted that at least one camera will capture the facial ROI at a certain time, but no matter how long it will last.
2. We proposed to select the ROI with the largest area from the cameras as the target ROI at each time point. Then, multichannel RGB time sequence can be calculated by taking average of all pixels within each target ROI. By this means, the changing information of the distance/angle between the face and the camera along with the moving of the subject can be solved.
3. We proposed an algorithm framework of EEMD-TDCCA to evaluate HR during our simulated moving-around experiments and the best performance was achieved by the proposed EEMD-TDCCA followed by frequency analysis, compared to other four methods, demonstrating the effectiveness of the proposed framework.

The organization of the rest of the paper is as follows: In Section 2, we will briefly review the advances on rPPG. Section 3, we will introduce relevant methods with respect to target ROI determination, HR source extraction based on EEMD-TDCCA and HR estimation. Section 4 will present the experiments and related results. Extensive discussions will be shown in Section 5. Section 6 will conclude the paper.

2 Related work

Consumer-level-camera-based rPPG was first proposed by Verkruysse et al. [45] in 2008. They demonstrated that HR could be measured from video recordings of the subject's face under ambient light using an ordinary digital camera. Later, Poh et al. [36] proposed a linear combination of RGB channels to estimate HR by employing blind source separation (BSS) methods. As an alternative, Sun et al. [41] proposed a framework of rPPG measurement based on joint time-frequency analysis. However, almost all above studies have been conducted under relatively well-controlled situations, e.g., stable illumination and unmoved subjects in front of cameras. In realistic situations, the cardiac pulse signals measured from facial regions could be contaminated by a number of factors, typically as illumination variations and motion artifacts [30, 50].

There are two main schemes to tackle illumination variations. The first scheme is based on signal de-noising methods to separate illumination variation signals from pulse signals. For instance, Chen et al. [8] separated the real cardiac pulse signals from the environmental illumination noise by applying ensemble empirical mode decomposition (EEMD) to the single green channel. The second scheme is mainly based on the assumption that both facial region of interest (ROI) and background ROI have similar illumination variations and the background ROI can be considered as a noise reference to denoise the facial signals. These illumination variation source can be both extracted from the background ROI and the facial ROI by using autoregressive (AR) modeling [43], normalized least mean square (NLMS) [30], least square curve fitting (LSCF) [27], joint blind source separation (JBSS) [14], or partial least squares (PLS) algorithms [49].

As for motion artifact elimination solutions, there are two mainly types based on whether knowing the prior information of components or not. The first type is called blind source separation (BSS)-based methods which extract multi-channel unobserved sources from a

set of observed mixed signals without prior information regarding the mixing process. For example, Poh et al. [35] separated the source signal from the raw signals that contaminated by motion artifacts by using independent component analysis (ICA) (a typical method of BSS). Sun et al. [51] proposed a new artifact-reduction methods with the combination of planar motion compensation and BSS. Monkaresi et al. [33] proposed to eliminate the impact of motion artifacts by applying both ICA and machine learning approach. Since conventional BSS techniques are originally designed to handle one single data set at a time, when the availability of multi-datasets increases, various joint BSS (JBSS) methods are now proposed. Guo et al. [19] first employed independent vector analysis (IVA) to jointly analyze color signals derived from multiple facial sub-region datasets. Qi et al. [37] explored correlations among facial sub-region datasets to improve motion robustness by using JBSS. The second type is model-based methods which utilize color vector information to control the demixing for component derivation. For instance, De Haan and Jeanne [16] developed a chrominance-based method (CHROM) to eliminate the influence of motion artifacts. CHROM considers both diffuse reflection components and specular reflection contributions, and by the linear combination of the individual R, G and B channels to eliminate the impact of motion artifacts. Wang et al. [47] proposed a Plane-Orthogonal-to-Skin (POS) method, which defined a POS tone in the temporally normalized RGB space to extract HR while reducing the impact of motion artifacts. Wang et al. [46] also proposed a rPPG algorithm that referred to Spatial Subspace Rotation (2SR), which estimated a spatial subspace of skin-pixels and measured the corresponding temporal rotation to extract HR.

However, all the aforementioned motion-robust solutions are successful under the situations that the complete facial ROI can be detected and tracked by a single camera with face alignment algorithms. However, there still remains one big challenge that the single camera may miss capturing the facial region of interest (ROI) when someone moves around. To address this problem, several studies have proposed multi-camera frameworks. On the one hand, the performance of HR estimation in the presence of motion artifacts can be improved when adopting multi-camera rather than the single-camera. For instance, J. R. Estepp et al. [17] utilized multi-cameras, instead of the single camera, to mitigate the HR measurement error caused by motion artifacts by increasing the dimensionality of the decomposed channel space prior to employing blind source separation. Similar conclusion can be drawn by E. B. Blackford et al. [4] that the HR estimation from the fusing data acquired by the triple-camera array was better than that from the single camera. On the other hand, the problem of the facial ROI missing can also be solved by the multi-camera scheme. In 2017, D. J. McDuff et al. [32] proposed a method combining BSS with AR model to remotely measure HR from moving subjects by fusing partial camera signals from 9 cameras covering a range of semi-circle. In their study, the subjects only moved their head horizontally. Besides, the camera with the duration of capturing the facial ROI less than 20% of the whole analysis window would be discarded. In this case, if fewer cameras are adopted and only one single camera can capture the facial ROI during a shorter period (less than 20% of the whole analysis window), the data from this camera would be discarded, causing the proposed method in [32] to derive a local evaluated HR instead of a global one within the whole analysis window.

3 Method

The flowchart of our proposed synchronized multi-camera framework is illustrated in Fig. 1. First, multiple cameras are employed to synchronously capture facial videos. For each

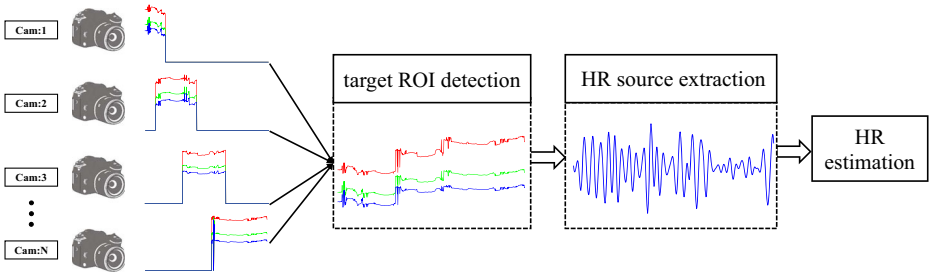


Fig. 1 Flowchart of the synchronized multi-camera framework for non-contact HR measurement

video, face alignment and tracking algorithms are utilized to detect and track facial ROIs respectively. At each time point, the target facial ROI can be determined as the one that has the largest area among all the detected facial ROIs. Then, the spatial means of all the pixels within each target ROI are subsequently calculated and temporally concatenated to form a time sequence. Second, the proposed method EEMD-TDCCA is adopted to obtain the sources containing the blood volume pulse (BVP) information, called HR source extraction. Finally, the extracted HR source will be utilized to measure HR by adopting frequency analysis and peak detection techniques. The details will be elaborated in the following subsections.

3.1 Target ROI detection

3.1.1 ROI detection and tracking

Reliable ROI detection and tracking are crucially important for subsequent rPPG-based HR estimation [18, 30, 34]. It is suggested that the boundary of the detected facial ROI should be within the face boundary. Besides, the eye region is better to be excluded from the facial ROI due to blinking interferes. With the consideration of above aspects, facial landmark localization is a natural approach for facial ROI detection and tracking. In this study, an approximated structured output learning approach in constrained local models technique is employed to efficiently detect 66 facial landmarks [54], which will be tracked by tracking learning detection (TLD) scheme [26]. Besides, only nine of the detected landmarks are selected to define our facial ROI. Figure 2 shows the coordinates of all the 66 facial landmarks in terms of small green circles, and the selected 9 ones, termed as 1 to 9. The area within the red polygon is the facial ROI. Suppose there are N_i ($1 \leq N_i \leq 4$) cameras that can synchronously capture the facial ROIs at the i-th image frame (totally I image frames during each analyzed window $1 \leq i \leq I$), $ROI_{n_i,i}$ is used to denote the facial ROI captured from the n-th camera at the i-th image frame ($1 \leq n_i \leq N_i$) and the corresponding area size can be expressed as area $\left(ROI_{n_i,i}\right)$.

3.1.2 Target ROI detection

For the i-th image frame, the target ROI is defined as the n_i^* one, which has the largest area size among all the detected facial ROI. The n_i^* can be expressed in:

$$n_i^* = \arg\max \left(\text{area}\left(ROI_{n_i,i}\right)\right), \quad 1 \leq n_i \leq N_i \tag{1}$$

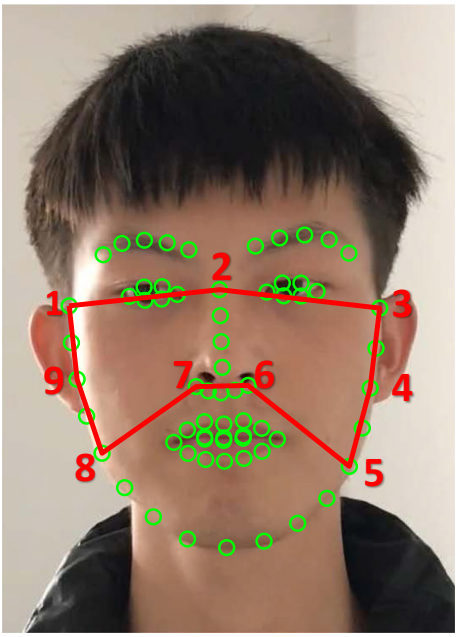


Fig. 2 Illustration of all the detected facial landmarks and the selected ones to form our facial ROI

and the target ROI is marked as $TROI_{n_i^*}$. Thereby, for all the I image frames, the selected target ROIs can be denoted as $TROI = \left\{TROI_{n_1^*}, TROI_{n_2^*}, \ldots, TROI_{n_I^*}\right\}$, and the corresponding area sizes are $AreaTROI = \left\{\text{area}\left(TROI_{n_1^*}\right), \text{area}\left(TROI_{n_2^*}\right), \ldots, \text{area}\left(TROI_{n_I^*}\right)\right\}$.

3.1.3 Spatial RGB means

Due to the fact that the spatial pixel averaging can reduce the camera quantization error [47]. In this paper, we employs the same strategy. Specifically, for each image frame, the spatial RGB means of all the pixels within the target ROI is calculated. Since visible cameras have three color channels R, G and B, 3 averaged pixel intensities within the target ROI can be formed as £°

$$RGB_{c,i} = \frac{\sum_{x,y \in TROI_{n_i^*}} I(x, y, i)}{\text{area}\left(TROI_{n_i^*}\right)} \tag{2}$$

with

$$RGB_i = \left\{RGB_{c,i}\right\}, \quad c = R, G, B \tag{3}$$

where $I(x, y, i)$ denotes the pixel value of coordinate (x, y) within the target ROI at the i-th image frame. After processing all the I image frames, a temporal RGB trace matrix $RGB = \{RGB_i\} \in \Re^{3 \times I}, i = 1, 2, \ldots, I$ is formed, with each row representing the averaged pixel values within the target ROI from a color channel.

3.2 EEMD-TDCCA-based HR source extraction

Recent study has demonstrated that EEMD is a good choice to decompose the single channel time sequence, typically the green channel time sequence, into several intrinsic mode functions (IMFs), among which the HR information is included [8]. Thus, EEMD helps to separate the real cardiac pulse signals from the environmental artifacts [8]. Besides, researchers have proposed several methods to deal with the RGB traces with the aim of extracting the source that contains the pulse-induced information. The RGB spatial means can be composed of linear mixtures of several underlying sources including the HR source. In this case, blind source separation (BSS) technique is a feasible option to separate the temporal RGB traces into independent signal sources using different criteria and will take the most periodic source as the HR source [23, 29, 52]. However, when the amplitude of the motion is large, the BSS-based method might also be challenged. Recently, canonical correlation analysis (CCA) has been proposed as an alternative BSS approach for HR source extraction [2]. With the combination of EEMD and CCA, the muscular artifacts can be isolated from electroencephalogram signals [11, 13]. In this paper, we address the difference between rPPG signals and the artifacts in the form of autocorrelation. Unlike rPPG quasi-periodical characteristic, while the artifacts have relatively low autocorrelation for their broad frequency spectrum. Using raw rPPG signal as the first dataset and its temporally delayed version as the second, CCA seeks the sources maximally autocorrelated and mutually uncorrelated by exploiting second-order statistics (SOS). Thereby, in this paper, EEMD followed by TDCCA employs this unique characteristic to isolate motion artifacts from rPPG signals. The performance of the CHROM [16], POS [47], EEMD [8] and EEMD-CCA (without time delay) methods will be compared to demonstrate the feasibility of our proposed EEMD-TDCCA framework.

3.2.1 Ensemble Empirical Mode Decomposition (EEMD)

EEMD is a noise assisted data analysis method which is an optimized version of empirical mode decomposition (EMD). EMD, first proposed by Huang [25], is a time space analysis method to analysis nonlinear and non-stationary data. In EMD method, the color trace $X_c = \left\{RGB_{c,i}\right\}, i = 1, 2, \ldots, I$ is decomposed into a finite number of intrinsic mode functions (IMFs):

$$X_c = \sum_{j=1}^{N} IMF_j + r_N \tag{4}$$

where r_N is the residue of X_c, after N number of IMFs are extracted. IMFs are simple oscillatory functions with varying amplitude and frequency, and hence have the following properties [25]: 1. Throughout the whole length of a single IMF, the number of extrema and the number of zero-crossings must either be equal or differ at most by one (although these numbers could differ significantly for the original data set); 2. At any data location, the mean value of the envelope defined by the local maxima and the envelope defined by the local minima is zero. In practice, the EMD is implemented through a sifting process that uses only local extrema, which will lead to mode mixing [48]. To overcome this disadvantage, a new noise-assisted data analysis method, called the ensemble EMD (EEMD), is proposed. EEMD defines the true IMF components as the mean of an ensemble of trials, each consisting of the signal plus a white noise of finite amplitude.

The main steps of EEMD are shown in Algorithm 1. By this means, the added white noise will be averaged out with a sufficient number of trials, and the only persistent part that

survives the averaging process is the component of the signal. According to [48] and [7], an ensemble number of a few hundreds will generally lead to a reasonable result, and the noise standard deviation has been suggested to be 20% the standard deviation of the original signal.

Algorithm 1 The major steps of EEMD.

Input: The Green channel time sequence RGB_G.
Output: The intrinsic mode functions IMF_j, $j = 1, 2, \ldots, N$.

1: **repeat**
2: Add a white noise to the signal (different white noise series for each time).
3: Find all the local maxima and minima over the full length of the signal.
4: The local maxima are connected using a cubic spline creating an upper envelope. The local minima repeat the same process to calculate the lower envelope.
5: Compute the average of the two envelopes and subtract this average from the original signal, resulting in a new signal.
6: **until** Obtain the (ensemble) means of corresponding IMFs of the decompositions as the final result

After IMF extraction, fast Fourier transform (FFT) was adopted to calculate the corresponding frequency spectrum for each IMF. For each IMF, the frequency corresponding to its maximal amplitude is identified. If the frequency falls within the possible frequency range of HR (ranging from 0.7 Hz to 3 Hz, covering the normal range of human HR from 42 beats per minute (bpm) to 180 bpm), the corresponding IMF will be determined as an IMF candidate. Then the IMF candidates can be obtained, denoted as $X \in \mathbb{R}^{D\times I} = \{X_1, X_2, \ldots, X_D\}^T$, where D is the number of the IMF candidates.

3.2.2 Canonical Component Analysis (CCA)

Canonical component analysis (CCA) is a popular technique to find two pairs of canonical variables (CVs), one for each set, such that the correlation matrix between the two sets is diagonal and the correlations on the diagonal are maximized which is a second-order multivariate statistical analysis method proposed by Hotelling [24]. CCA has been extensively utilized in many fields due to it's low complexity, fixed number of model parameters, and correlation representation in multiview learning [31], such as biomedical signal processing [6, 10], including medical data fusion [15], physiological parameter detection [37], removal of muscle artifacts from electroencephalogram signals [9], analysis of gene expression [40] *etc*. Considering that the rPPG is quasi-periodical, the original HR source will be highly correlated to its time-delay version and the correlation between the noise source and its corresponding time-delay version is lower. Thereby, we propose to employ the time-delay CCA to extract CVs corresponding to HR source.

The time-delay version of the candidate IMF X is defined as $X_D \in \mathbb{R}^{D\times I}$, where$X_D(i) = X(i + \tau)$, τ denote the points of the time delay. Here $\tau = 1$ and $i = 1, 2, \ldots, I - 1$. CCA is to find pairs of CVs $U = \{U_1, U_2, \ldots, U_J\}^T$ and $V = \{V_1, V_2, \ldots, V_J\}^T$, $J \leq \min(\text{rank}(X, X_D))$, which tries to maximize the correlation between the two matrices by computing two canonical coefficient vectors or weight vectors $\alpha \in \mathbb{R}^{D\times J} = (\alpha_1, \alpha_2, \ldots, \alpha_J)$ and $\beta \in \mathbb{R}^{D\times J} = (\beta_1, \beta_2, \ldots, \beta_J)$. Thus, the *j-th* pair of CVs can be expressed as :

$$U_j = \alpha_j^T X \tag{5}$$

$$V_j = \beta_j^T X_D \tag{6}$$

Then, the corresponding weight vectors can be obtained by maximizing the correlations between U and V by optimizing the objective function:

$$\max_{\alpha_j, \beta_j} \rho(X, X_D) = \frac{\mathrm{Cov}\left(U_j, V_j\right)}{\sqrt{\mathrm{Var}\left(U_j\right)\mathrm{Var}\left(V_j\right)}} = \frac{\alpha_j^T C_{XX_D} \beta_j}{\sqrt{\left(\alpha_j^T C_{XX} \alpha_j\right)\left(\beta_j^T C_{X_D X_D} \beta_j\right)}} \tag{7}$$

where C_{XX_D} is the cross-covariance matrix of X and X_D , C_{XX} and $C_{X_D X_D}$ are the autocovariance matrices of X and X_D , respectively

Through a deflationary procedure and using the method of Lagrange multipliers, J pairs of weight vectors and their corresponding canonical variates can be derived. Consequently, the first pair of CVs (U_1 and V_1) having the largest canonical correlation, the following pair of CVs having the second largest canonical correlation, and until the last pair of CVs have the smallest canonical correlation. Thereby, the first pair of CV (either U_1 or V_1) will be determined as the HR source.

3.3 HR estimation

The HR can be estimated after the extraction of HR source. Currently, two main types of methods can be utilized to estimate HR, called peak detection and frequency analysis based on fast Fourier transform (FFT). The peak detection algorithm is performed using a custom algorithm with a moving time window δ, since there is a refractory period for an excitable cardiac muscle membrane to be ready for a second stimulus. The HR is calculated as $\mathrm{N}\cdot 60/\mathrm{T}$ beat per minute (bpm), where N is the number of peaks, and T is the analysis time in terms of seconds. Frequency analysis means that first the HR source signals are first transformed from time domain into frequency domain by FFT. Then, the HR frequency is defined as the frequency having the maximum amplitude in the frequency domain, marked as $f_{\max}$. The HR is finally calculated as $60 \cdot f_{\max}$ (bpm). The main steps of the EEMD-TDCCA-based HR estimation is described in Algorithm 2.

Algorithm 2 The main steps of the EEMD-TDCCA-based HR estimation.

Input: The Green channel time sequence RGB_G.
Output: The estimated HR

```
Apply EEMD to RGB_G and derive a sets of IMFs. The number of the IMFs is N. ← EEMD
for n = 1 to N do
    Find the frequency corresponding to its maximal amplitude F_n.
    if 0.7Hz ≤ F_n ≤ 3Hz then
        Obtain one IMF candidate.
    end if
    All IMF candidates are obtained, denoted as X.
end for
Define the time-delay version of X, X_D = [X(:, 2 : I), X(:, 1)].
Solve the maximization problem in (7) and obtain the HR source (U_1). ← CCA
Estimate the HR by applying peak detection or frequency analysis to the HR source.
```

3.4 Performance measures

Several metrics have been used to evaluate the performance of HR estimation, such as mean absolute error (MAE), standard deviation (SD_e), mean percentage error (MPE), root mean square error (RMSE), the Pearson's correlation coefficient r and intraclass correlation coefficient (ICC).

Mean absolute error measures the magnitude of the absolute average error and the corresponding formula is:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|HR_{nci} - HR_{gti}\right| \tag{8}$$

where HR_{nci} denotes the i-th HR estimation obtained from rPPG signals, HR_{gti} is the i-th reference one and n denotes the total number of participants.

Standard deviation is the standard deviation of the heart rate differences between the observed values and the reference values.

Mean percentage error is the average of the percentage error between HR_{nci} and HR_{gti}, expressed as:

$$\mathrm{MPE} = \frac{100\%}{n}\sum_{i=1}^{n}\frac{\left|HR_{nci} - HR_{gti}\right|}{HR_{gti}} \tag{9}$$

Root mean square error is defined as the square root of the average of squared differences between the observed value (HR_{nci}) and the reference value (HR_{gti}). It can be computed as:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(HR_{nci} - HR_{gti}\right)^2} \tag{10}$$

The Pearson's correlation coefficient r between the observed value (HR_{nci}) and the reference value (HR_{gti}) is also considered:

$$r = \frac{\sum_{i=1}^{n}\left(HR_{nci} - \overline{HR_{nc}}\right)\left(\left(HR_{gti} - \overline{HR_{gt}}\right)\right.}{\sqrt{\sum_{i=1}^{n}\left(HR_{nci} - \overline{HR_{nc}}\right)^2}\sqrt{\sum_{i=1}^{n}\left(HR_{gti} - \overline{HR_{gt}}\right)^2}} \tag{11}$$

where $\overline{HR_{nc}}$ and $\overline{HR_{gt}}$ are the mean values of the estimated and the reference HR values, respectively.

Intraclass correlation coefficient (ICC) a statistical parameter that measures absolute agreement between two continuous variables which is defined as:

$$Y_{ij} = \mu + \alpha_j + \varepsilon_{ij} \tag{12}$$

where Y_{ij} is the j-th observation of the i-th individual, μ is the mean of the overall observation, α_j is an unobserved random effect shared by all values in group j, and ε_{ij} is an unobserved noise term. Thus ICC can be defined as:

$$ICC = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_\varepsilon^2} \tag{13}$$

where σ_α^2 is the variance of α_j and σ_ε^2 is the variance of ε_{ij} . The ICC values below 0.4 represent poor agreement, it is good agreement when the values between 0.4 and 0.75 , and values above 0.75 represent an excellent agreement of the measurements [22].

4 Experiments and results

4.1 Experimental setup

Our experimental setup is shown in Fig. 3. Four iPhone7s (Apple Inc, USA) with ios 11 system were employed to synchronously collect facial videos. Each iPhone was fixed on one tripod. Totally four tripods were placed on the semicircular arc with a radius of 0.8 (m), while an electric turntable (ComXim, MT380WL80H) was placed on the center of the circle. In order to simulate the move-around situation, participants stood on the turntable and moved around by controlling the corresponding switch to rotate with a relatively uniform speed. The rotation speed was controllable ranging from 4.5 degrees per second (dps) to 9 dps. During the video recording, participants were asked to stand on the turntable, and the height of the tripods were adjustable, making sure that the face were in the middle of the camera view. The angle between each two cameras was set to 45°. Meanwhile, the ECG acquisition system ECG6951D (Nihon Kohden Co., Shinjuku-ku, Tokyo, Japan) was utilized to acquire the HR ground truth, which was synchronized with the recorded videos. With the approval of the Ethics Review Committee of Hefei University of Technology, 30 informed Asian-skin-color subjects (22 males and 8 females), with the ages ranging from 20 to 30 years old were recruited in this experiment. The participants' HR range from 60 bpm to 110.4 bpm. All the videos were recorded with a frame rate of 30 frames per second (fps) and a resolution of 1920×1080.

To cover the range of the semicircle, the rotation speed was selected as 8.18 dps. In other word, each semicircle rotation took about 22 seconds. Each participant performed once facial video collection, producing 30 facial videos in total. During the following process of HR estimation, a duration about 22 seconds of all the facial videos were selected as the processing window length.

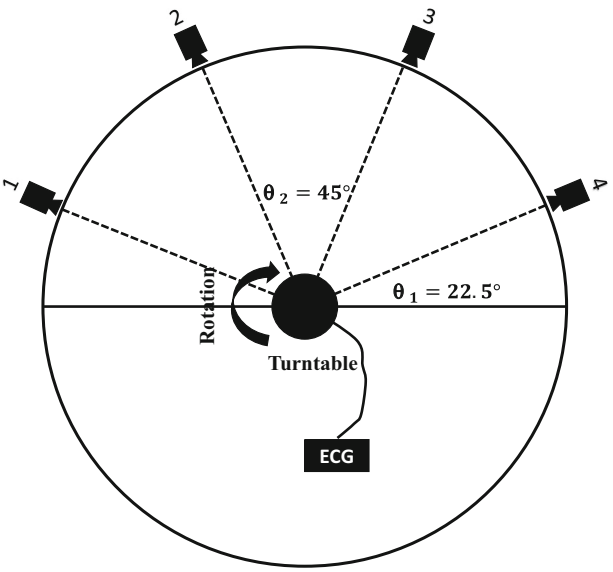


Fig. 3 Experimental setup for facial video acquisition based on multi-cameras and corresponding HR reference acquisition

4.2 Experimental results

4.2.1 HR estimation based on EEMD-TDCCA

Before using EEMD, the green channel spatial means of the target facial ROI are normalized with z-scores to be zero mean and unit variance [35] and then detrended by a smoothness priors approach with the smoothing parameter (λ = 10) [44]. According to [48], [7], an ensemble number of a few hundreds will generally lead to a reasonable result and the noise standard deviation has been suggested to be 20% the standard deviation of the original signal. In this study, the ensemble number was set as 100 and the noise standard deviation was set as 0.2. Since it is usually suggested that the number of IMFs is determined by the equation:

$$N = fix\left(\log_2(I)\right) \tag{14}$$

where N is the number of data points, $fix()$ is a rounding function. In this study, I is equal to 660, according to the equation, the number of the IMFs is determined as 10.

According to the selecting criteria of IMF candidates, the IMF candidates turn to be IMF3, IMF4 and IMF5 with the frequencies corresponding to the largest amplitude 2.31 Hz, 2.02 Hz and 1.15 Hz respectively. Figure 4 shows the IMF candidates and their corresponding frequency spectrums. Among these IMF candidates, the amplitude of IMF2 is the largest. In this case, the estimated HR will be 121.2 bpm if only the EEMD algorithm is adopted, while the real HR is 70.9 bpm. Fortunately, in this study, the three IMF candidates (one dataset) and their corresponding time-delay version (the other dataset) are sent to CCA. The derived U_1 is determined as the HR source, and its frequency corresponding to the largest amplitude is 1.15 Hz. Both the reference HR signal (marked as ECG) and the HR source (marked as rPPG) are shown in Fig. 5. It can be seen that the final estimated HR is 69 (60 $\times$ 1.15) bpm, only 1.9 bpm less than the reference HR, demonstrating the feasibility of the proposed EEMD-TDCCA framework.

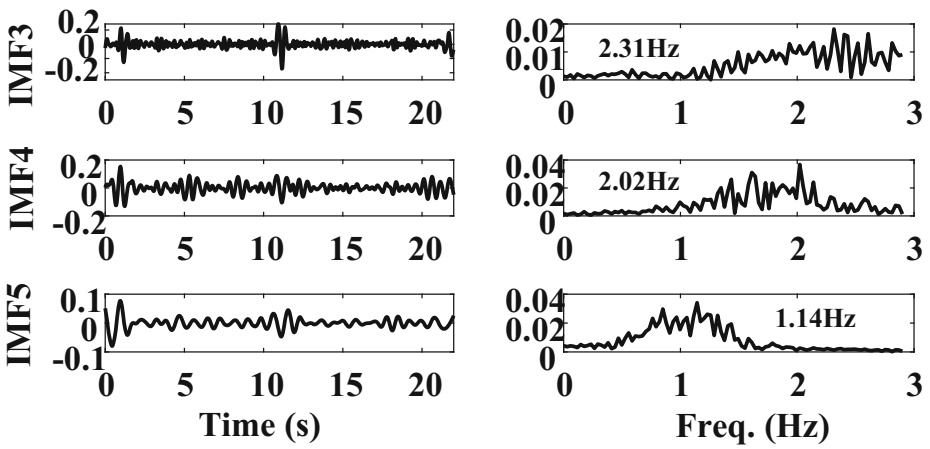


Fig. 4 IMF candidates and their corresponding frequency spectrums

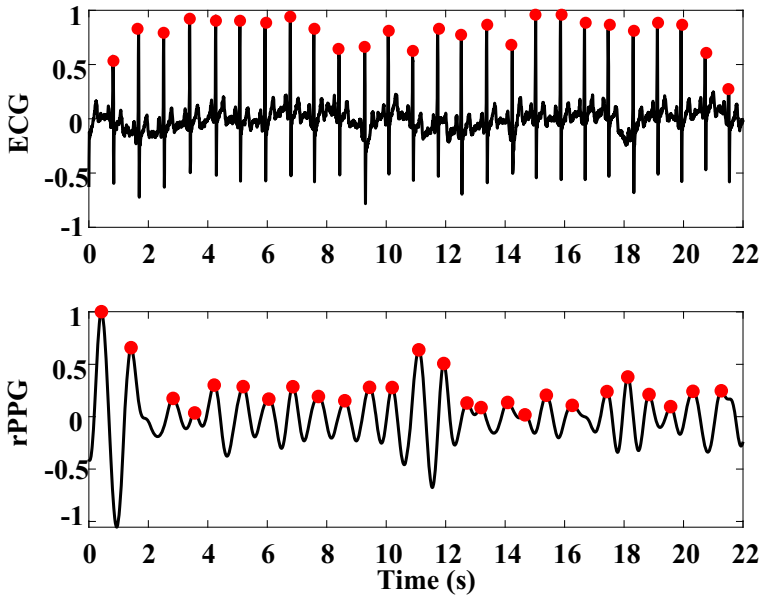


Fig. 5 Illustration of the reference HR signal (top) and the HR source derived from EEMD-TDCCA (bottom)

4.2.2 Performance comparison and summary

Table 1 lists the performance of HR estimation in terms of the above mentioned five performance metrics by adopting the EEMD-TDCCA and other four state-of-the-art methods (CHROM [16], POS [47], EEMD [8] and EEMD-CCA (without time delay)) followed

Table 1 The performance of HR estimation in terms of performance metrics by adopting different methods followed by peak detection and frequency analysis

		MAE (SD_e) (bpm)	MPE (%)	RMSE (bpm)	r	ICC
CHROM	*P*	10.92 (9.20)	15.13	14.28	0.39	0.34
	F	7.84 (6.51)	10.71	10.19	0.65	0.65
POS	*P*	11.78 (7.52)	16.25	13.97	0.31	0.13
	F	8.20 (6.95)	10.57	10.75	0.62	0.62
EEMD	*P*	10.26 (6.41)	13.99	12.10	0.54	0.44
	F	7.87 (5.95)	10.22	9.86	0.62	0.62
EEMD-CCA	*P*	10.29 (5.97)	14.02	11.88	0.58	0.52
	F	12.20 (7.75)	15.91	14.45	0.49	0.47
EEMD-TDCCA	*P*	7.71 (4.14)	10.17	8.75	0.67	0.64
	F	**4.11 (3.46)**	**5.26**	**5.37**	**0.90**	**0.89**

P stands for peak detection and *F* corresponds to frequency analysis

by peak detection (abbreviated as P) and frequency analysis (abbreviated as F). It can be seen from Table 1 among all the method with peak detection, the best performance is achieved by our proposed EEMD-TDCCA. To be specific, the MAE (SD_e) of the estimated HR by using CHROM, POS, EEMD, EEMD-CCA and EEMD-TDCCA with peak detection are 10.92 (9.20) bpm, 11.78 (7.52) bpm, 10.26 (6.41) bpm, 10.29 (5.97) bpm and 7.71 (4.14) bpm, respectively. The MPEs of the estimated HR by using the five methods with peak detection range from 10.17% to 16.25%, and the RMSEs range from 8.75 bpm to 14.28 bpm. The Pearson's correlation coefficient of the estimated HR derived from EEMD-TDCCA with peak detection is the largest with 0.67 while that of the other four methods are 0.54, 0.58 , 0.39 and 0.31, respectively. The ICCs of the estimated HR derived from EEMD, EEMD-CCA and EEMD-TDCCA with peak detection are 0.44, 0.52, and 0.64 respectively, demonstrating a good agreement with the corresponding reference HR (between 0.4 and 0.75). The ICCs of CHROM and POS with peak detection are both below 0.4.

It can be seen from Table 1 that the performance of HR estimation by using a certain method with frequency analysis sometimes is better than that with peak detection except EEMD-CCA. They are at least 2 bpm lower than the frequency analysis accordingly. Besides, our proposed EEMD-TDCCA followed by the frequency analysis achieves the best results. The corresponding MAE is 4.11 bpm, the MPE is 5.26%, the RMSE is 5.37 bpm, the correlation coefficient is 0.90 and the ICC is 0.89, which all indicate the excellent performance of the proposed methods when compared to other four methods. Specifically, the MAE (SD_e) achieved by using EEMD-TDCCA is the smallest with 4.11 (3.46) bpm, while that of the other four methods are 7.84 (6.51) bpm, 8.20 (6.95) bpm, 7.87 (5.95) bpm and 12.20 (7.75) bpm, respectively. The values of MPE and RMSE achieved by using EEMD-TDCCA are also the smallest with 5.26% and 5.37 bpm, respectively, at least 5.31% and 4.82 bpm smaller than those achieved by other methods. Besides, the Pearson's correlation coefficient of EEMD-TDCCA is the largest with 0.90, at least 0.25 higher than that of the other four methods. The ICC of EEMD-TDCCA is the also largest with 0.89, demonstrating the excellent agreement of the methods while other methods are 0.62, 0.47, 0.65, 0.62, respectively.

A one-way ANOVA test was conducted to assess the statistical difference between the proposed EEMD-TDCCA and the other four methods. The level of statistical significance was set to $p < 0.05$. The statistical results show that the HR estimated by the proposed EEMD-TDCCA is significantly better than that of any other method (the maximum p value is 0.0064).

Since the method with frequency analysis is usually better than that of the same method with peak detection, the Bland-Altman plots are utilized to analyze the agreement between the estimated HR by our proposed EEMD-TDCCA framework and the HR reference (with frequency analysis), as well as other three methods accordingly (excluding EEMD-CCA), shown in Fig. 6, with Fig. 6, b, c and d corresponding to CHROM, POS, EEMD and EEMD-TDCCA, respectively. The mean bias of HR estimated by using CHROM-based method is -1.5 bpm and the corresponding 1.96 times SD is 18.6 bpm. The mean bias of HR estimated by using POS-based method is 1.6 bpm and the corresponding 1.96 times SD is 22.8 bpm. The EEMD-based method achieves the mean bias of 1.0 bpm, with the corresponding 1.96 times SD 20.6 bpm. The mean bias of EEMD-TDCCA is only 1.6 bpm with the corresponding 95% limits of agreement ranging only from -8.6 to 11.8 bpm, at least 8.4 bpm improvement compared to other three methods. The Bland-Altman plot demonstrates the feasibility of our proposed EEMD-TDCCA framework followed by the frequency analysis.

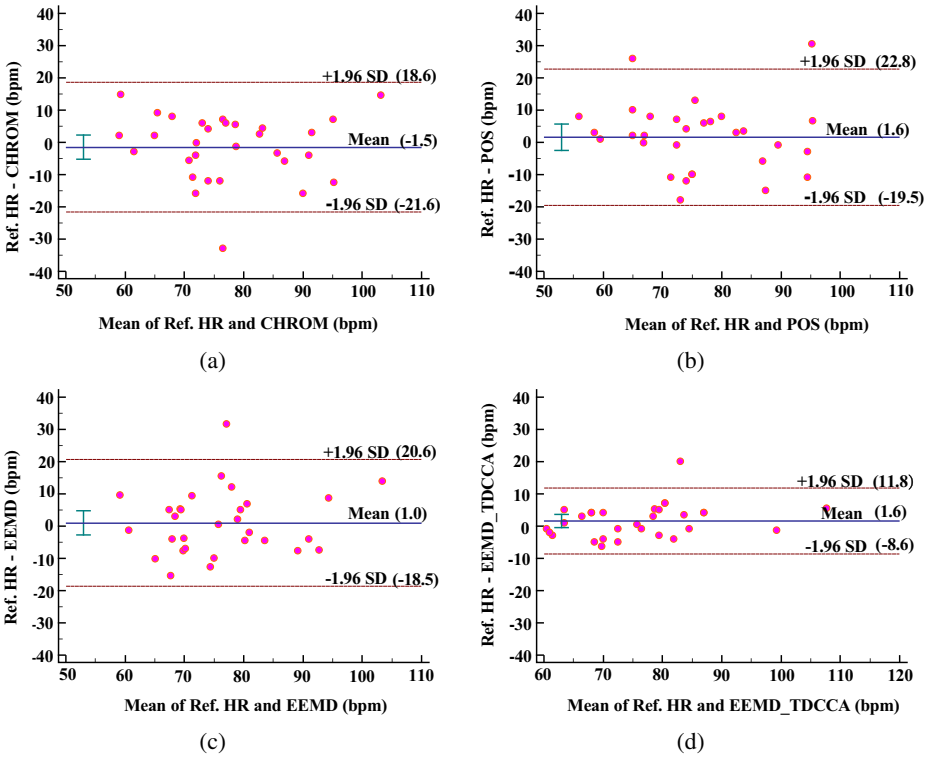


Fig. 6 Bland-Altman plots of the estimated HR by different methods and the reference HR

5 Discussions

According to the mathematical model proposed by Wang et al. in [47], during a constant daytime situation, the skin area measured by each camera has a varying color due to both motion-induced specular variations and pulse-induced subtle color changes. In this study, the motion artifacts are produced by the simulated moving-around situation with the help of turntable. For each camera, the spectral compositions are correspondingly changing since the angle between the skin tissue and the camera is changing along with the rotation. Besides, four cameras placed at different locations had different angles with the sunlight-entering window, which caused corresponding absolute illumination intensities different and led to varying RGB means of ROIs. Consequently, the concatenated RGB means turned out to be non-stationary and non-linear. In this paper, the concatenated RGB means were first normalized with z-scores, with the aim of eliminating the influence of different absolute illumination intensities. Despite of it, the normalized concatenated RGB means still contained pulse-induced subtle color changes, motion-induced variations and other noises. Experimental results demonstrated that CHORM-based method has the ability of extracting HR source from the simulated moving-around situations, since it used the knowledge of color vectors of different components to form the chrominance feature linearly combined by RGB channel signals and the impact of motion artifacts can be roughly eliminated. The reason why the proposed EEMD-TDCCA performed best might be that, although both dominating noises caused by motion artifacts and illumination variations existed, the

EEMD, taking advantages of decomposing the non-linear and non-stationary time series into a finite number of IMFs, can derive the IMF candidates including HR source from the other IMFs including noises, which can be indicated from the fact that the EEMD with frequency analysis can achieve the second best performance. Besides, considering that the HR source is quasi-periodical and has a relatively high autocorrelation while the artifact is random and has a low high autocorrelation, the time-delay CCA can be used to isolate the HR source from the artifacts. As for POS-based methods, since it softened the knowledge required by CHROM for defining the projection plane by using the data-driven approach, it was reported to have the best overall performance in a large benchmark in previous single-camera framework, typically in fitness situations. However, in our multi-camera framework, POS performed not that well, which might result from that a fixed plane orthogonal to the temporally normalized skin-tone direction (four cameras may have four skin-tone directions) can not be defined.

Apart from the above mentioned methods, the popular ICA-based method was also utilized to estimate HR. However, the performance was really bad. It is reasonable that the HR source can be recovered from a set of observed mixtures without prior information by using ICA under the assumption that the RGB means time series are actually a linear combination of the pulse signal and other signals. However, the linear assumption is more appropriate in stationary situation or subtle motion situation. Thereby, the ICA-based method is no longer suitable for our proposed moving-around multi-camera framework. In this paper, we excluded the results derived by ICA-based method.

The time delay points τ can affect the performance of HR estimation. We set this parameter from 1 to 5. Results showed that with the increase of τ, the HR estimation performance of the proposed EEMD-TDCCA decreased. Consequently, the τ was set as 1 in this study.

6 Conclusions

In this paper, we have demonstrated the feasibility of the synchronized multi-camera framework for seamless remote HR measurement by our proposed algorithm framework EEMD-TDCCA. Meanwhile, the performance has been compared with other four state-of-the-art methods, typically as EEMD, EEMD-CCA, CHROM and POS. To our knowledge, by the target ROI selection strategy, the temporally concatenated spatial RGB means of all the pixels within each target ROI, can be considered as being derived from a "same" camera. Afterwards, the green channel derived from the target ROIs can be pre-denoised by being decomposed into several IMFs by EEMD. Furthermore, taking consideration of the characteristic of both HR source (quasi-periodical) and the artifacts (random), the HR source can be determined as the first pair of CVs (having the largest correlation coefficient) derived by time-delay CCA. Our multi-camera framework will be a promising tool for practical non-contact video-based vital sign measurement applications.

There are some limitations that should be noted. In the experiments, the participants were asked to stand on the turntable peacefully with no facial expressions or head motions to avoid extra motion artifacts. In the future, more realistic moving around situations along with vivid facial or head motions can be studied. In that case, the method in [5] can be utilized to recognize our interesting event. Besides, only the HR parameter has currently been considered. In the future, other physiological parameters, typically heart rate variability, arterial blood oxygen saturation, blood pressure and respiratory rate will be studied.

Acknowledgments This work was supported by the National Key R&D Program of China (Grant 2017YFB1002802), National Natural Science Foundation of China (Grants: 61922075, 81571760, 61701160 and 41901350) and the Fundamental Research Funds for the Central Universities (Grants: JZ2019HGBZ0151 and JZ2020HGPA0111).

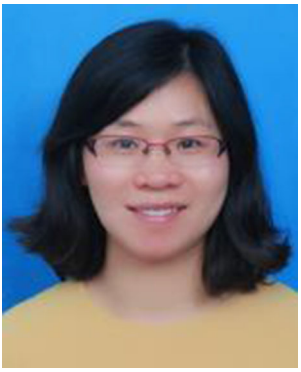
References

1. Aarts LA, Jeanne V, Cleary JP, Lieber C, Nelson JS, Oetomo SB, Verkruysse W (2013) Non-contact heart rate monitoring utilizing camera photoplethysmography in the neonatal intensive care unit a pilot study. Early Hum Dev 89(12):943–948
2. Al-Naji A, Perera AG, Chahl J (2017) Remote monitoring of cardiorespiratory signals from a hovering unmanned aerial vehicle. Biomed Eng Online 16(1):101
3. Amelard R, Scharfenberger C, Kazemzadeh F, Pfisterer KJ, Lin BS, Clausi DA, Wong A (2015) Feasibility of long-distance heart rate monitoring using transmittance photoplethysmographic imaging (ppgi). Sci Rep 5:14637
4. Blackford EB, Estepp JR (2017) Using consumer-grade devices for multi-imager non-contact imaging photoplethysmography. In: Optical Diagnostics and Sensing XVII: Toward Point-of-Care Diagnostics, vol. 10072. International Society for Optics and Photonics, pp 100720P
5. Caruccio L, Polese G, Tortora G, Iannone D (2019) Edcar: a knowledge representation framework to enhance automatic video surveillance. Expert Syst Appl 131:190–207
6. Chen X, Wang ZJ, McKeown M (2013) A three-step multimodal analysis framework for modeling corticomuscular activity with application to parkinson's disease. IEEE J Biomed Health Inf 18(4):1232–1241
7. Chen X, Liu A, Peng H, Ward R (2014) A preliminary study of muscular artifact cancellation in single-channel eeg. Sensors 14(10):18 370–18 389
8. Chen D-Y, Wang J-J, Lin K-Y, Chang H-H, Wu H-K, Chen Y-S, Lee S-Y (2015) Image sensor-based heart rate evaluation from face reflectance using hilbert–huang transform. IEEE Sens J 15(1):618–627
9. Chen X, Liu A, Chiang J, Wang ZJ, McKeown M, Ward RK (2015) Removing muscle artifacts from eeg data: Multichannel or single-channel techniques? IEEE Sens J 16(7):1986–1997
10. Chen X, Wang ZJ, McKeown M (2016) Joint blind source separation for neurophysiological data analysis: Multiset and multimodal methods. IEEE Signal Proc Mag 33(3):86–107
11. Chen X, Chen Q, Zhang Y, Wang ZJ (2018) A novel eemd-cca approach to removing muscle artifacts for pervasive eeg. IEEE Sens J 19(19):8420–8431
12. Chen X, Cheng J, Song R, Liu Y, Ward R, Wang ZJ (2018) Video-based heart rate measurement: Recent advances and future prospects. IEEE Transactions on Instrumentation and Measurement
13. Chen X., Xu X., Liu A., McKeown M. J., Wang Z. J. (2018) The use of multivariate emd and cca for denoising muscle artifacts from few-channel EEG recordings. IEEE Trans Instrum Measur 67(2):359–370
14. Cheng J, Chen X, Xu L, Wang ZJ (2017) Illumination variation-resistant video-based heart rate measurement using joint blind source separation and ensemble empirical mode decomposition. IEEE J Biomed Health Inf 21(5):1422–1433
15. Correa NM, Adali T, Li Y-O, Calhoun VD (2010) Canonical correlation analysis for data fusion and group inferences. IEEE Signal Process Mag 27(4):39–50
16. De Haan G, Jeanne V (2013) Robust pulse rate from chrominance-based rppg. IEEE Trans Biomed Eng 60(10):2878–2886
17. Estepp JR, Blackford EB, Meier CM (2014) Recovering pulse rate during motion artifact with a multi-imager array for non-contact imaging photoplethysmography. In: 2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC). IEEE, pp 1462–1469
18. Feng L, Po L-M, Xu X, Li Y, Ma R (2015) Motion-resistant remote imaging photoplethysmography based on the optical properties of skin. IEEE Trans Circ Syst Video Technol 25(5):879–891
19. Guo Z, Wang ZJ, Shen Z (2014) Physiological parameter monitoring of drivers based on video data and independent vector analysis. In: 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, pp 4374–4378
20. Haque MA, Irani R, Nasrollahi K, Moeslund TB (2016) Heartbeat rate measurement from facial video. IEEE Intell Syst 31(3):40–48
21. Hassan M, Malik A, Fofi D, Saad N, Meriaudeau F (2017) Novel health monitoring method using an rgb camera. Biomed Opt Express 8(11):4838–4854
22. Hills M (1987) The design and analysis of clinical experiments. J R Stat Soc Ser A (Gen) 150(4):400–400

23. Holton BD, Mannapperuma K, Lesniewski PJ, Thomas JC (2013) Signal recovery in imaging photoplethysmography. Physiol Measur 34(11):1499
24. Hotelling H (1935) Relations between 2 sets of variants. Biometrika 28(3–4):312–377
25. Huang NE, Shen Z, Long SR, Wu MC, Shih HH, Zheng Q, Yen N-C, Tung CC, Liu H (1998) The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis. Proc R Soc Lond Ser A: Math Phys Eng Sci 454(1971):903–995
26. Kalal Z, Mikolajczyk K, Matas J (2012) Tracking-learning-detection. IEEE Trans Pattern Anal Mach Intell 34(7):1409–1422
27. Lee D, Kim J, Kwon S, Park K (2015) Heart rate estimation from facial photoplethysmography during dynamic illuminance changes. In: 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, pp 2758–2761
28. Leonhardt S, Leicht L, Teichmann D (2018) Unobtrusive vital sign monitoring in automotive environments a review. Sensors 18(9):3080
29. Lewandowska M, Rumiński J, Kocejko T, Nowak J (2011) Measuring pulse rate with a webcam a non-contact method for evaluating cardiac activity. In: 2011 federated conference on computer science and information systems (FedCSIS). IEEE, pp 405–410
30. Li X, Chen J, Zhao G, Pietikainen M (2014) Remote heart rate measurement from face videos under realistic situations. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 4264–4271
31. Liu H, Sun X (2016) Linear canonical correlation analysis based ranking approach for facial age estimation. In: 2016 IEEE International Conference on Image Processing (ICIP). IEEE, pp 3249–3253
32. McDuff DJ, Blackford EB, Estepp JR (2018) Fusing partial camera signals for noncontact pulse rate variability measurement. IEEE Trans Biomed Eng 65(8):1725–1739
33. Monkaresi H, Calvo RA, Yan H (2014) A machine learning approach to improve contactless heart rate monitoring using a webcam. IEEE J Biomed Health Inf 18(4):1153–1160
34. Po L-M, Feng L, Li Y, Xu X, Cheung TC-H, Cheung K-W (2018) Block-based adaptive roi for remote photoplethysmography. Multimed Tools Appl 77(6):6503–6529
35. Poh M-Z, McDuff DJ, Picard RW (2010) Non-contact, automated cardiac pulse measurements using video imaging and blind source separation. Opt Express 18(10):10 762–10 774
36. Poh M-Z, McDuff DJ, Picard RW (2011) Advancements in noncontact, multiparameter physiological measurements using a webcam. IEEE Trans Biomed Eng 58(1):7–11
37. Qi H, Guo Z, Chen X, Shen Z, Wang ZJ (2017) Video-based human heart rate measurement using joint blind source separation. Biomed Signal Process Control 31:309–320
38. Rasche S, Trumpp A, Waldow T, Gaetjen F, Plötze K, Wedekind D, Schmidt M, Malberg H, Matschke K, Zaunseder S (2016) Camera-based photoplethysmography in critical care patients. Clin Hemorheol Microcirc 64(1):77–90
39. Rouast PV, Adam MT, Chiong R, Cornforth D, Lux E (2018) Remote heart rate measurement using low-cost rgb face video: a technical literature review. Front Comput Sci 12(5):858–872
40. Soneson C, Lilljebjörn H, Fioretos T, Fontes M (2010) Integrative analysis of gene expression and copy number alterations using canonical correlation analysis. BMC Bioinform 11(1):191
41. Sun Y, Azorin-Peris V, Kalawsky R, Hu S, Papin C, Greenwald SE (2012) Use of ambient light in remote photoplethysmographic systems: comparison between a high-performance camera and a low-cost webcam. J Biomed Opt 17(3):037005
42. Sun Y, Thakor N (2016) Photoplethysmography revisited: from contact to noncontact, from point to imaging. IEEE Trans Biomed Eng 63(3):463–477
43. Tarassenko L, Villarroel M, Guazzi A, Jorge J, Clifton D, Pugh C (2014) Non-contact video-based vital sign monitoring using ambient light and auto-regressive models. Physiol Measur 35(5):807
44. Tarvainen MP, Ranta-Aho PO, Karjalainen PA (2002) An advanced detrending method with application to hrv analysis. IEEE Trans Biomed Eng 49(2):172–175
45. Verkruysse W, Svaasand LO, Nelson JS (2008) Remote plethysmographic imaging using ambient light. Opt Express 16(26):21 434–21 445
46. Wang W, Stuijk S, De Haan G (2016) A novel algorithm for remote photoplethysmography: Spatial subspace rotation. IEEE Trans Biomed Eng 63(9):1974–1984
47. Wang W, den Brinker AC, Stuijk S, de Haan G (2017) Algorithmic principles of remote ppg. IEEE Trans Biomed Eng 64(7):1479–1491
48. Wu Z, Huang NE (2009) Ensemble empirical mode decomposition: a noise-assisted data analysis method. Adv Adapt Data Anal 1(01):1–41
49. Xu L, Cheng J, Chen X (2017) Illumination variation interference suppression in remote ppg using pls and memd. Electron Lett 53(4):216–218

50. Yang Z, Yang X, Wu X (2019) Motion-tolerant heart rate estimation from face videos using derivative filter. Multimed Tools Appl 78:26747–26757
51. Yu S, Hu S, Azorin-Peris V, Chambers JA, Zhu Y, Greenwald SE (2011) Motion-compensated noncontact imaging photoplethysmography to monitor cardiorespiratory status during exercise. J Biomed Opt 16(7):077010
52. Yu Y-P, Raveendran P, Lim C-L, Kwan B-H (2015) Dynamic heart rate estimation using principal component analysis. Biomed Opt Express 6(11):4610–4618
53. Zhao F, Li M, Jiang Z, Tsien JZ, Lu Z (2016) Camera-based, non-contact, vital-signs monitoring technology may provide a way for the early prevention of sids in infants. Front Neurol 7:236
54. Zheng S, Sturgess P, Torr PH (2013) Approximate structured output learning for constrained local models with application to real-time facial feature detection and tracking on low-power devices. In: 2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG). IEEE, pp 1–8

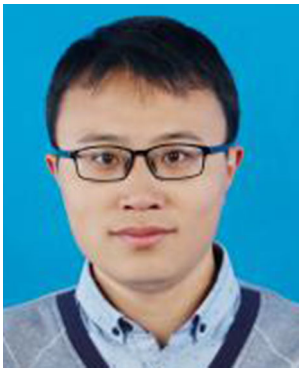
Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



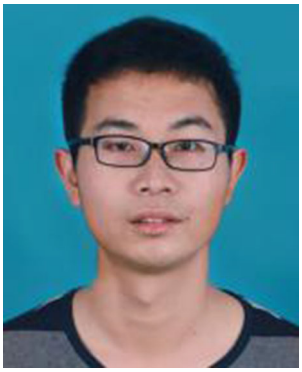
Juan Cheng received her B.S. degree and Ph. D degree from the Department of Electronic Science and Technology, University of Science and Technology of China (USTC) in 2008 and 2013 respectively. She is currently an associated professor with the Department of Biomedical Engineering, Hefei University of Technology (HFUT), Hefei, China. Her research interests include biomedical signal/image processing, non-contact physiological parameter measurement, and machine learning.

Xingmao Wang received his B.S. degree from the Department of biomedical Engineering, Hefei University of Technology (HFUT), Hefei, China, where he is currently working toward the M.S. degree. His research interests include non-contact physiological parameter measurement.

Rencheng Song received his B.S. degrees in Mathematics from Jilin University, Changchun, China, in 2005, and the Ph.D. degree from Zhejiang University, Hangzhou, China in 2010. He is currently an associate professor with the Department of Biomedical Engineering at Hefei University of Technology. His research interests include biomedical signal processing, non-contact physiological parameter measurement, machine learning and electromagnetic imaging.



Yu Liu received his B.S. degree and Ph. D degree from the Department of Automation, University of Science and Technology of China in 2011 and 2016, respectively. He is currently an assistant professor in the Department of Biomedical Engineering at Hefei University of Technology. His research interests include image/signal processing, computer vision, information fusion and machine learning.



Chang Li received his B.S. degree in information and computing science from the Wuhan Institute of Technology, Wuhan, China, in 2012, and the Ph.D. degree from the School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan, in 2018. He is currently a lecturer with the Department of Biomedical Engineering, Hefei University of Technology, Hefei, China. His current research interests include the areas of biomedical signal processing, hyperspectral image analysis, computer vision, pattern recognition, and machine learning.

Xun Chen received the B.S. degree in the Department of Electronic Science and Technology at the University of Science and Technology of China (USTC) in 2009, and received the Ph.D degree in the Department of Electrical and Computer Engineering at the University of British Columbia (UBC) in 2014. He is with the Hefei National Laboratory for Physical Sciences at the Microscale and Department of Electronic Science and Technology at USTC as a professor. His research interests include the broad areas of statistical signal processing and machine learning in biomedical applications. He has published over 70 refereed scientific papers. He is serving as associate editor for IEEE Access, and also serving as guest associate editor for Frontiers in Aging Neuroscience and Sensors.