Robust Heart Rate Estimation from Facial Video Using Project_ICA

Lin Qi[1], Huidong Yu[1], Lisheng Xu[1,2]*, Ramadhani Selemani Mpanda[1] and Stephen E. Greenwald[3]

[1] Sino-Dutch Biomedical and Information Engineering School, Northeastern University, Shenyang, Liaoning, China

[2] Key Laboratory of Medical Image Computing, Shenyang, Ministry of Education, China

[3] Blizard Institute, Barts & The London School of Medicine &Dentistry, Queen Mary University of London, United Kingdom

* Corresponding author

E-mail: xuls@bmie.neu.edu.cn

Keywords: remote photoplethysmography, dichromatic reflection model, independent component analysis

Abstract

Remote photoplethysmography (rPPG) can achieve non-contact measurement of heart rate (HR) from a continuous video sequence by scanning the skin surface. However, practical applications are still limited by factors, such as non-rigid facial motion and head movement. In this work, a detailed system framework for remotely estimating heart rate from facial video under various movement conditions is described. After the rPPG signal is obtained from a defined region of the facial skin, a method, termed "Project_ICA", based on a skin reflection model, is employed for extracting the pulse signal from the original signal. To evaluate the performance of the proposed algorithm, a dataset was created from 28 participants, containing 112 videos that include the challenges of various skin tones, body motion and HR recovery after exercise. The results show that Project_ICA, when evaluated by several criteria provides a more accurate and robust estimate of HR than most existing methods, although problems remain in obtaining reliable measurements from subjects with dark skin.

1. Introduction

Heart rate is one of the most important indicators of cardiovascular health. Studies have confirmed that an increase in resting HR is an independent risk factor for predicting the onset, progression and mortality of cardiovascular disease (CVD). Thus its daily monitoring can

contribute greatly to the prevention of cardiovascular disease as well as assessing progress in the rehabilitation of chronic disease(Cook *et al.*, 2006). Currently, the gold standard for HR measurement is ECG, which is highly accurate but requires electrodes to be attached to the subject. To avoid the discomfort and inconvenience of this procedure, optical methods such as photoplethysmography (PPG) have been developed(Allen, 2007). The technique of PPG is based on the difference in optical properties between blood and surrounding tissue and hence, by detecting changes in transmission through, for example, a finger; or by measuring changes in reflectance of the skin(Verkruysse *et al.*, 2008), variations in blood volume, synchronous with the heart rate, can be measured. Furthermore, research has found that the PPG signal can be detected using a video camera to measure changes in facial color during the cardiac cycle. (Poh *et al.*, 2010b; Sun *et al.*, 2012; Sun *et al.*, 2013).

Using this technology, minute variations in light reflected from the skin can be used to extract physiological signals(Takano and Ohta, 2007), and the field of remote physiological measurement has developed extensively over the past decade. The methodology can be divided into two categories: ballistocardiography (BCG) and PPG. For the BCG approach, HR is measured by estimating the minute head motion produced by the pulse during each cardiac cycle. It was first applied to measure HR by Balakrishnan et al. in 2013(Balakrishnan *et al.*, 2013). Subsequently, Haque et al.(Haque *et al.*, 2016) developed a new approach to improve the system shortcomings, which extracted facial landmark points using a supervised descent method and generated stable trajectories of these points for HR estimation. However, the signal was strongly affected by the participants' overall movement, thus seriously impeding further development of this method.

PPG, the alternative optical approach, has on the other hand, shown more promise and has been more widely adopted. In 2008, Verkruysse et al.(Verkruysse *et al.*, 2008) were the first to propose that rPPG could detect the pulse using ambient light alone. The favored analysis approach has been Blind Source Separation (BSS), using Independent Component Analysis (ICA)(Poh *et al.*, 2010b) or Principal Components Analysis (PCA)(Lewandowska *et al.*, 2011) to extract a periodic pulse signal from the original rPPG signal. De Haan and Jeanne (De and Jeanne, 2013) found that multiple color channels can be combined with chrominance weight to enhance the pulse signal quality. Based on that idea, De Haan and Arno (De Haan and Arno, 2014) extended the work using the 'blood volume pulse signature' derived from knowledge of camera characteristics, skin reflectance and ambient illumination, to refine the rPPG signal. These two methods were influential on subsequent research. In 2014, McDuff et al.(Mcduff *et al.*, 2014) proposed an improvement to the optical properties of the digital camera using a five-band lens. Subsequently, Li et al.(Li *et al.*, 2014) proposed a framework which utilized face tracking and adaptive filtering methods to rectify background illumination and improve precision. Meanwhile, several supervised learning methods, such as Support Vector Regression (SVR)(Hsu *et al.*, 2014) and the K-Nearest Neighbors algorithm (KNN)(Monkaresi *et al.*, 2014), were used to generate accurate HR values by using signal features to inform regression

analysis. However, these user-independent models require retraining with new participants. Subsequently Kumar et al.(Kumar *et al.*, 2015) proposed a new camera-based HR estimation algorithm named DistancePPG which can use multiple regions of interest (ROI) to optimize the source signal and address some challenges pertaining to low lighting conditions and subjects with dark skin. Following this, algorithms like Joint Blind Source Separation (JBSS)(Qi *et al.*, 2017) and Ensemble Empirical Mode Decomposition (EEMD)(Cheng *et al.*, 2016) were used to decompose the source signal and derive an accurate estimate of the pulse rate. Wang et al.(Wang *et al.*, 2015; Wang *et al.*, 2016; Wang *et al.*, 2017a) proposed a series of PPG-based core algorithms for HR estimation, such as spatial subspace rotation (2SR) and plane orthogonal to skin (POS). 2SR defines a spatial subspace of skin-pixels and measures its temporal rotation for pulse signal extraction; POS is designed to extract the pulse signal based on a skin reflection model. Furthermore, Wang et al. extended their work to develop the Sub-band rPPG method to suppress different distortion-frequencies arising from subject motion during exercise(Wang *et al.*, 2017b).

Although many camera-based methods for measuring HR with high accuracy have been reported, there are still many challenges that affect measurement results, such as motion artifact and variation in illumination(Hassan *et al.*). Thus, there is scope for further work to solve these problems and thus to improve the technology for clinical monitoring.

This paper proposes a non-contact, robust and real-time system framework for estimating HR using facial video, which builds on some of the previous research (Feng *et al.*, 2015; Prakash and Tucker, 2018; Tulyakov *et al.*, 2016). Our key contribution is a new algorithm, termed *Project_ICA*, which can robustly extract the pulse signal in the face of movement artefacts, poor illumination and dark skin colour. We identify three contributions. Firstly, a fast method to obtain the raw rPPG signal based on face tracking and skin region detection, which significantly improves the speed of face detection compared to other methods, even under illumination variation, dark skin and subject movement. Secondly, a new algorithm to extract an accurate pulse signal from the raw rPPG signal. This method projects the original three-channel rPPG signals onto a predefined plane to form two-channel signals with independent dimensions, and uses a skin reflection model to eliminate illumination variation. Following this, the projected signal is used to extract an accurate pulse signal using the Project_ ICA algorithm. Thirdly, a dataset containing various challenges, such as different skin tones and several subject motion scenarios is used to evaluate the performance of the algorithms.

The remainder of this paper is structured as follows. Signal acquisition and processing are presented in section 2. The experimental setup and validation tests are described in sections 3 and 4. Finally, some conclusions are offered in section 5.

2. Signal acquisition and processing

A framework for overcoming challenges, such as motion artifacts, illumination changes and camera noise, is outlined in Figure 1. The details of each part are presented in the following sections.

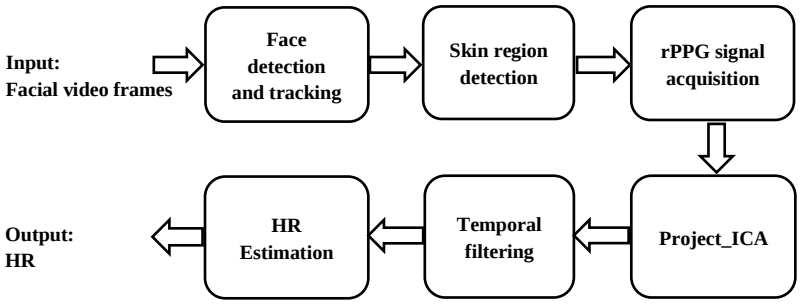


Figure 1 Outline of the system framework.

2.1 Acquisition of rPPG signal

During the first stage of the proposed data processing procedure the rPPG signal is obtained from the customized ROI as green boundary as shown in the rightmost panel of Figure 2 (a). In most previous research, face detection and tracking have been implemented by using the Viola-Jones (VJ) face detector applied to each frame(Viola and Jones, 2001). However, this process has proved to be time consuming and prone to misdetection(Tran *et al.*, 2015). In this study, we have developed a fast and efficient method to extract the rPPG signal, which reduces the computational burden. Figure 2 illustrates the process.

Firstly, the VJ algorithm is used to detect only an initial rectangular facial region which is regarded as the ROI in the first video frame. The feature points in the ROI are detected using the minimum eigenvalue algorithm and are shown as the green points within the yellow facial box in the middle panel of Figure 2 (a).

Secondly, to produce a stable ROI for subsequent frames, the Kanade-Lucas-Tomasi feature tracker (KLT)(Tomasi, 1991) is used to track detected feature points and produce matching point pairs between the previous and the current frames. For the n-th frame, the locations of the detected feature points, $\boldsymbol{P_n} \in R^{2\times k}$ (green crosses), are shown in Figure 3 (a), where k is the number of detected feature points in this frame. The corresponding feature points $\boldsymbol{P_{n+1}}$ (red circles) in the (n+1)-th frame can be found using KLT as Figure 3 (b) shows. Next, a geometric transformation matrix $\boldsymbol{A} \in R^{2\times 2}$ is defined between $\boldsymbol{P_n}$ and $\boldsymbol{P_{n+1}}$, such that the relationship can be denoted as $\boldsymbol{P_{n+1}} = \boldsymbol{AP_n}$. Subsequently, the ROI coordinates in the (n+1)-th frame can be computed using the matrix product of $\boldsymbol{A}$ and the coordinates of the ROI boundary points in the n-th frame. This method improves face detection and tracking speed significantly, compared to other methods.

Thirdly, the ROI is redefined by detecting skin pixels within the detected ROI. To achieve rapid skin detection, the detected facial region is converted from RGB to YCrCb(Tsouri and Li,

2015) and the skin region is extracted using a predefined threshold of skin color (as shown by the area between the outer and inner green lines in the rightmost panel of Figure 2 (a)). With skin pixels detected, background interference is removed and only skin pixels are further analysed.

Finally, the redefined ROI is separated into its RGB channels as shown in Figure 2 (b), and the raw rPPG signal $C(t) = \{R(t),\ G(t),\ B(t)\}$ is formed by averaging all skin pixel values within the ROIs from the RGB channels for each frame as,

$$C(t) = \{R(t),\ G(t),\ B(t)\} = \left\{\frac{\sum_{x,y \in S_{skin}} I_{ch}(x,y,t)}{M}\right\} \quad (1)$$

where S_{skin} represents the redefined skin ROI, $I_{ch}(x,y,t)$ is the pixel value in S_{skin} from the R/G/B channel at time t, and M is the number of detected skin pixels in S_{skin}.

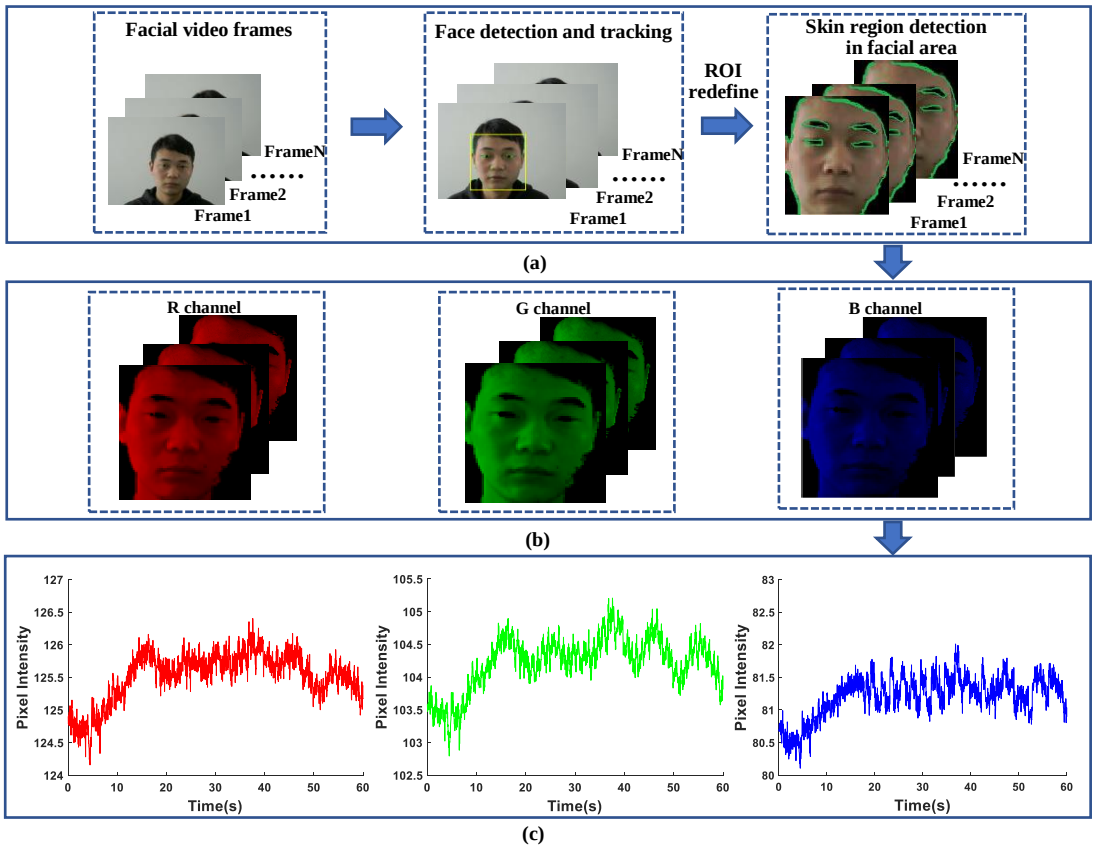
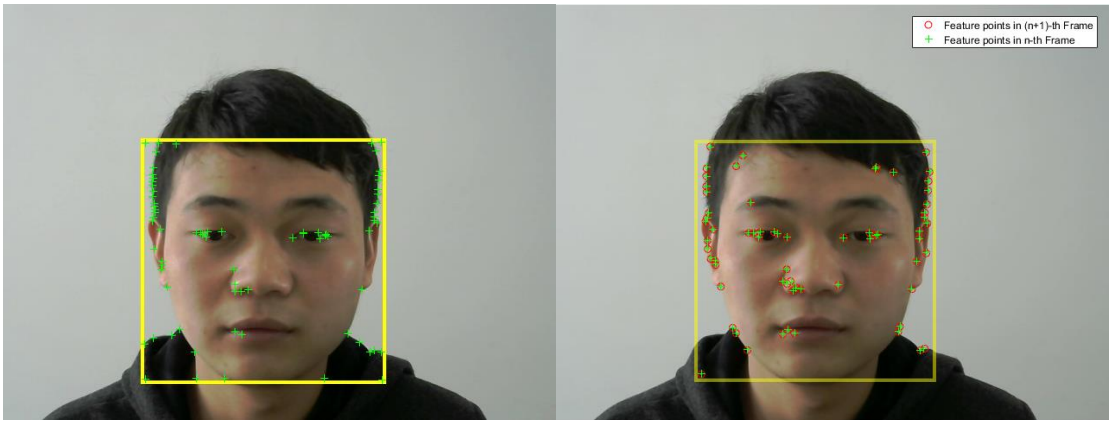


Figure 2 Extraction of raw rPPG signal by spatial pixel averaging. (a) Face detection tracking and skin region detection in the facial area (Green lines represent the detected skin boundary). (b) Frame images are separated into RGB channels. (c) Raw rPPG signal in each colour channel.

Referring to a previous study(Poh *et al.*, 2010b), the longer the duration of the raw rPPG signal, the more pulse information there is in it. Therefore, in subsequent analysis, the raw rPPG signal was processed by using a moving window of duration 30s and a 29s overlap (with 1s moving increment). Thus, 31 moving time windows will be produced for each minute of signal acquisition. Figure 4 illustrates the procedure.



(a) (b)

Figure 3 (a) Green crosses represent detected feature points $\boldsymbol{P_n}$ in the n-th frame. (b) The red circles indicate the feature points $\boldsymbol{P_{n+1}}$ in the (n+1)-th frame generated by KLT, which match those in the n-th frame.

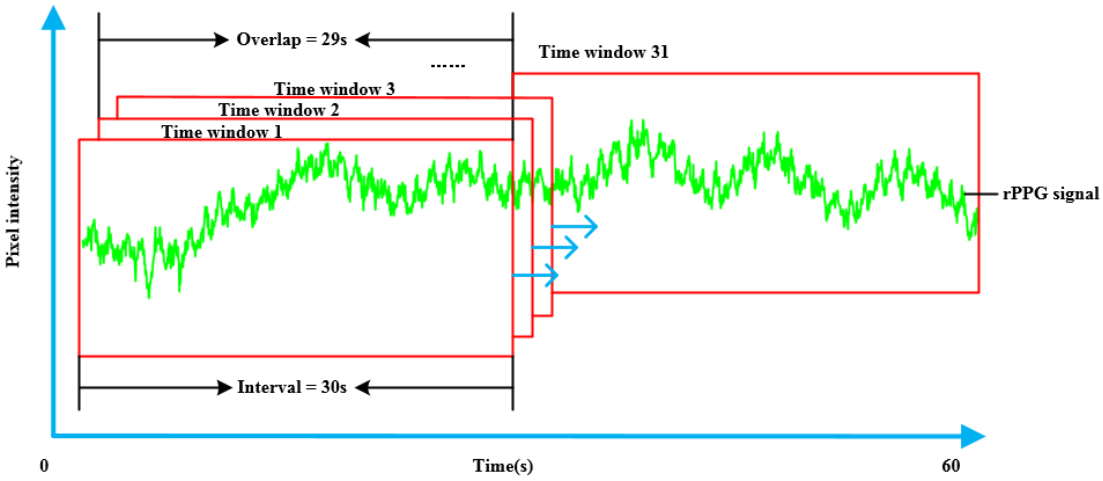


Figure 4 The moving time windows.

2.2 Extraction of the pulse signal

After obtaining the raw rPPG signal, Project_ICA is employed to extract the pulse signal from it. In the next section, for clarity and brevity we use boldface upper- and lower-case **U** and **u** to represent a matrix and column vector, respectively. **1** represents the unit vector$[1, 1, 1]$.

2.2.1 Simplified skin dichromatic reflection model

Skin is a multilayered and inhomogeneous organ, composed of epidermis, dermis and subcutis(Krishnaswamy and Baranoski, 2004). In previous work(Xu *et al.*, 2014; Lam and Kuno, 2016; Wang *et al.*, 2017a), the Lambert-Beer law or Shafer's dichromatic reflection model(Shafer, 1985) have been used to extract physiological variables from skin images. In this study, a simplified skin reflection model(Wang *et al.*, 2017a) was used, as shown in Figure 5.

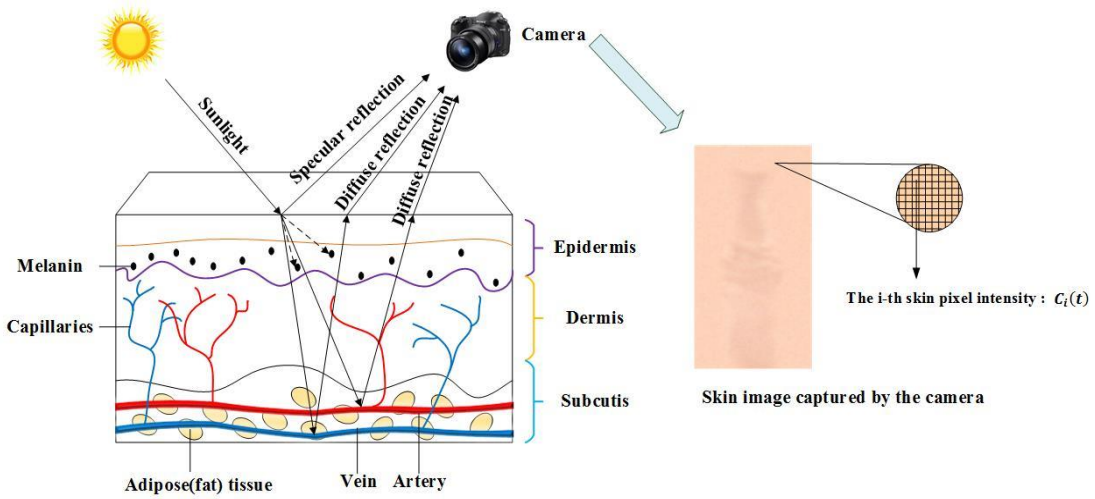


Figure 5 Simplified skin reflection model.

Based on the simplified skin reflection model, the skin pixel value of every frame can be defined as a time-varying function:

$$\boldsymbol{c_i}(t) = I(t) \cdot \big(\boldsymbol{v_s}(t) + \boldsymbol{v_d}(t)\big) + \boldsymbol{v_n}(t) \tag{2}$$

where $\boldsymbol{c}_i(t)$ is the i-th skin pixel intensity, which can be rewritten as $\{R_i(t), G_i(t), B_i(t)\}$; $I(t)$ denotes the luminance intensity level, which depends on light source and distance changes between light source, skin tissue and camera; $\boldsymbol{v_s}(t)$ represents the specular component of the reflections from the skin surface; $\boldsymbol{v_d}(t)$ represents diffuse component caused by scatter and absorption of light in skin tissue; $\boldsymbol{v_n}(t)$ denotes the quantization noise of the camera sensor.

Furthermore, (2) can be expanded to extract the pulse as follows.

$$\boldsymbol{c_i}(t) = I_0 \cdot (1 + i(t))\big(\boldsymbol{u_s} \cdot s(t) + \boldsymbol{u_c} \cdot c_0 + \boldsymbol{u_p} \cdot p(t)\big) + \boldsymbol{v_n}(t) \tag{3}$$

where $I_0 \cdot (1 + i(t))$ denotes illumination containing a stationary and a time-varying part. $\boldsymbol{u_s} \cdot s(t)$ and $\boldsymbol{u_p} \cdot p(t)$ are time-varying signals of the specular and diffuse components, respectively. Also, there are various constant components in the reflected signal, so the color vector of skin reflection is denoted as $\boldsymbol{u_c}$ and$\cdot c_0$ is the reflection strength. Subsequently, quantization noise is eliminated by spatially averaging skin pixels in the RGB channels every frame. Because the time-varying components are much smaller than the static components, the reflection model can be simplified by eliminating the products of the time-varying components.

$$\boldsymbol{c}(t) = I_0 \cdot \boldsymbol{u_c} \cdot c_0 \cdot \big(1 + i(t)\big) + \boldsymbol{u_s} \cdot I_0 \cdot s(t) + \boldsymbol{u_p} \cdot I_0 \cdot p(t) \tag{4}$$

The signal can be processed by temporal normalization to eliminate the static component further thus simplifying the expression to:

$$\boldsymbol{c_n}(t) = \boldsymbol{1} \cdot \big(1 + i(t)\big) + \boldsymbol{D} \cdot \boldsymbol{u_s} \cdot I_0 \cdot s(t) + \boldsymbol{D} \cdot \boldsymbol{u_p} \cdot I_0 \cdot p(t) \tag{5}$$

where $\boldsymbol{D} \in R^{3\times3}$ denotes the diagonal matrix and is used to normalize the steady component, and $p(t)$ presents the pulse signal, from which HR can be estimated. From the foregoing, a

core algorithm, called Project_ICA, is developed to extract $p(t)$ from the normalized rPPG signal. The detailed implementation of the proposed method is discussed in the next section.

2.2.2 Project_ICA

Traditional ICA(Stone, 2004) is a computational method for separating multivariate signals into additive subcomponents. It has been widely used for image-based physiological parameter measurement (Poh *et al.*, 2010b, a; Shan and Yu, 2014).

Based on (5), $i(t)$, $s(t)$ and $p(t)$ can be separated from the raw rPPG signal using ICA. However, there is a correlation between the time varying luminance signal $i(t)$ and the specular component $s(t)$, because they are simultaneously affected by head motion, thus interfering with the performance of ICA. Project_ICA has been developed to resolve this interference and obtain a more robust pulse signal.

To reduce the signal dimension and eliminate extra noise $i(t)$, the raw three-dimensional normalized signal would be projected onto a specific plane. In order to determine two projection-axes on this plane, a large number of trials were conducted to study how the projection-axes on the projected plane affected the quality of the projected signal. The results showed that the primary selection criterion was that the projection-axes should be orthogonal to the unit vector $\mathbf{1}$, which can eliminate $i(t)$ in (5). Furthermore, through experiment, although the number of planes orthogonal to **1** is infinite, the HR measurement remained stable with changes of the orthogonal projection plane Therefore, in this study, two projection-axes were defined in this study as:

$$\mathbf{P} = \begin{bmatrix} \boldsymbol{p_1} \\ \boldsymbol{p_2} \end{bmatrix} = \begin{pmatrix} -0.4082 & -0.4082 & 0.8165 \\ 0.7071 & -0.7071 & 0 \end{pmatrix} \tag{6}$$

$\boldsymbol{p_1}$ and $\boldsymbol{p_2}$ are orthogonal to each other and **1**. Through this projection, the residual signal contains only the specular component signal and the pulse signal as follows.

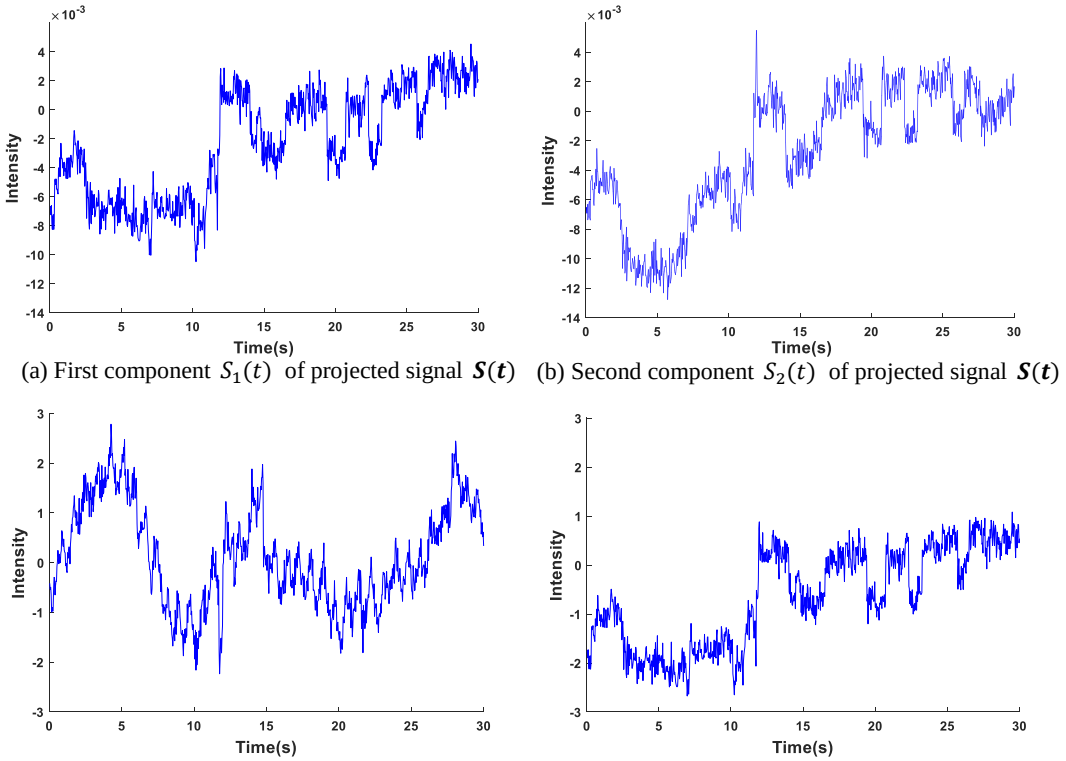
$$\boldsymbol{S(t)} = \begin{bmatrix} S_1(t) \\ S_2(t) \end{bmatrix} = \boldsymbol{P} \cdot \boldsymbol{c_n(t)} = \boldsymbol{P} \cdot \boldsymbol{D} \cdot \boldsymbol{u_s} \cdot I_0 \cdot s(t) + \boldsymbol{P} \cdot \boldsymbol{D} \cdot \boldsymbol{u_p} \cdot I_0 \cdot p(t) \tag{7}$$

where $\boldsymbol{S(t)}$ denotes the projected signal with two components $S_1(t)$ and $S_2(t)$, as shown in Figure 6 (a) and (b). Here the original three-dimensional signal has been reduced to two dimensions with the elimination of the illumination variation component. Most importantly, the generated signal after projection contains two components $S_1(t)$ and $S_2(t)$ that are assumed to be independent. Therefore, they can be decomposed to get $p(t)$ using ICA.

The projected signal contains only the specular component signal $s(t)$ and the pulse signal $p(t)$, which are more strongly independent of each other than the sub-components of the raw rPPG signal. When combined with (7), the ICA model can be rewritten as follows:

$$\begin{bmatrix} s(t) \\ p(t) \end{bmatrix} = \boldsymbol{W} \cdot \boldsymbol{S} = \boldsymbol{W} \cdot \begin{bmatrix} S_1(t) \\ S_2(t) \end{bmatrix} \tag{8}$$

where $\boldsymbol{W} \in R^{2\times2}$ is the demixing matrix used to extract the hidden variables $s(t)$ and $p(t)$. The signal after ICA processing is shown in Figure 6 (c) and (d). However, there is no ordering of the ICA output components. Therefore, in this paper the POS signal(Wang *et al.*, 2017a) is built as a reference pulse signal. Then, the two output ICA components are sorted in descending order according to the magnitude of the correlation coefficients with the reference function. Finally, the first component is selected as the desired pulse signal.



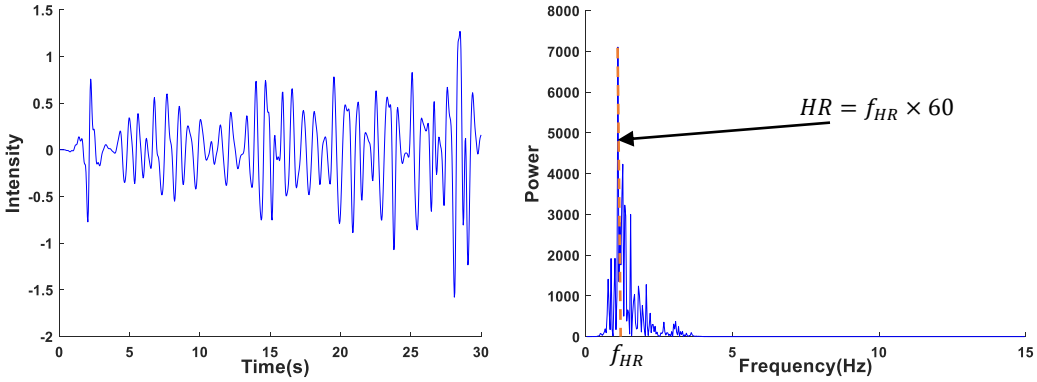
(a) First component $S_1(t)$ of projected signal $\boldsymbol{S}(\boldsymbol{t})$ (b) Second component $S_2(t)$ of projected signal $\boldsymbol{S}(\boldsymbol{t})$

(c) First component of signal after ICA analysis (d) Second component of signal after ICA analysis

Figure 6 Illustration of normalized rPPG signal after projection, panels a and b; and ICA, panels c and d.

2.3 Calculation of heart rate

In this step, two temporal filters are used to refine the pulse signal and exclude the frequencies which are not related to that of HR. Firstly, we used a 5-point moving-average filter, which removes random noise using a temporal average of adjacent frames. In addition, to eliminate the components which are unrelated to that of HR, we employed a Hamming window-based band-pass filter with cutoff frequencies of 0.7 Hz and 4 Hz. This range corresponds to the normal human HR range of 42 to 240 beats per minute (bpm)(Li *et al.*, 2014). Finally, we applied a 1024-point FFT to the filtered pulse signal with a frequency resolution of 0.033 Hz, using 30s rectangular time-window without overlap, and the frequency that corresponds to the maximum power of the spectrum was selected as the HR. The filtered signal and the corresponding FFT result are shown in Figure 7.



(a) (b)

Figure 7 (a) Filtered signal. (b) FFT of the filtered pulse signal.

3. Experiment setup and datasets

The signal processing was implemented in MATLAB® 2016b (The MathWorks, Inc.) and run on a laptop with an Intel Core i5 2.5 GHz CPU and 8 GB RAM.

3.1 Benchmark dataset for the experiment

112 video sequences (with 201600 frames) were recorded using a standard RGB camera C270i (Logitech Inc., Newark, CA, USA) in a 24-bit RGB color format at 30 frames per second with an image size of 640×480 pixels. All videos were saved in uncompressed AVI format on the laptop. After giving written informed consent and receiving instructions, 28 participants (7 females, 21 males) with a range of skin tones were enrolled in the study. In previous studies (Kumar *et al.*, 2015; Wang *et al.*, 2016; Wang *et al.*, 2017a), various skin types resulted in different strengths of rPPG signals and it was confirmed that both dark skin and bright skin were representative subjects to study the accuracy of HR measurement using rPPG. Therefore, in this study, 18 people with “pale” (i.e. East Asian) skin and 10 people with dark skin were employed to create our dataset. In parallel with the rPPG recording, we synchronously recorded pulse-oximeter data from a transmissive finger pulse-oximeter (YUWELL™ YX303, Yuyue medical equipment & supply Co., Ltd., Jiangsu, China.) as the reference HR signal. Experiments were conducted indoors with the subjects seated in front of the camera at a distance of 0.5 m as shown in Figure 8, the only illumination source was external daylight and each recording lasted for 1 minute. Furthermore, any variations in external illumination due for instance to the movement of clouds could have caused transient changes in illumination which the signal processing would have rejected. Slower changes due to gradual changes in weather would have little effect over the 30 seconds measurement time.

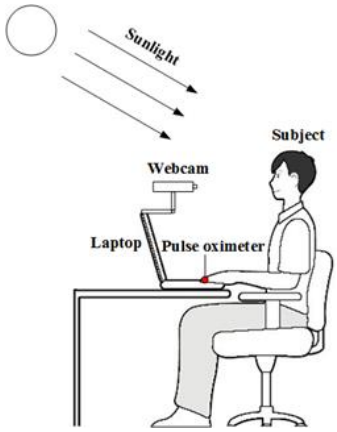


Figure 8 Experimental setup.

To simulate various motion scenarios that might be encountered in daily life, we designed four different experimental protocols based on the set-up shown in Figure 8.

Stationary scenario: In this section, all participants were asked to sit still with their faces unobstructed and to look at the camera without moving for 1 minute. This experiment was considered as the ideal measurement environment, thus providing a base line against which the other movement scenarios can be compared.

Interaction with computer scenario: To simulate a typical scenario in which the subject moves in a limited way, the computer screen in front of subject showed a computer game and the participants were required to freely play the game. They were able to move their heads slightly with freedom to vary their facial expression and to talk to others. In this setup while restricting gross movement of subject, slight head and facial movement were deliberately introduced as a source of noise.

Swinging heads scenario: In this section, subjects were asked to swing their heads from side to side at a rate of around 10 times per minute. This was regarded as a more challenging environment in which to assess the robustness of the HR detection algorithm. Some recording snapshots are shown in Figure 9.

Figure 9 Examples of head swinging.

Exercise recovery scenario: In this section, all participants were required to do 20 squats in approximately 2 minutes to induce a large change in HR. Immediately after the end of the exercise period, the subjects were again seated as shown in Figure 8 and the recording was started. The subjects were asked to keep as still as possible during the recording. The purpose of this scenario was to investigate the accuracy of HR measurements while the rate itself was changing. This ability is essential for investigating and analyzing heart rate variability, a parameter of clinical significance in the study of heart disease.

3.2 Evaluation metrics

To quantitatively assess the accuracy of the proposed method, we used three statistical metrics to investigate the error between the estimated HR and the reference. These were mean absolute deviation (MAD), root mean square error (RMSE) and Pearson correlation coefficient (r). The HR estimated from the facial video and the reference HR measured by pulse oximetry were denoted as HR_{est} and HR_{ref} respectively.

3.3 Comparison with other prevalent methods

The method presented in this paper involves a new framework for estimating HR from rPPG measurements. For a thorough evaluation comparison, we have benchmarked it against the four existing rPPG methods listed in Table 1. All these methods have been applied to the same dataset and computational environment.

Table 1 rPPG methods used for comparison with Project_ICA.

Methods	Description
ICA(Poh *et al.*, 2010b)	A classical method based on BSS
CHROM(De and Jeanne, 2013)	A motion robust method based on chrominance
2SR(Wang *et al.*, 2016)	A data-driven method based on the statistical distribution of the skin-pixels
POS(Wang *et al.*, 2017a)	A method based on signal projection onto the plane orthogonal to the skin-tone vector

4. Results and discussion

To assess the performance of Project_ICA and the other methods mentioned in section 3.3, the metrics in section 3.2 were used for a detailed analysis of the dataset. The HR_{est} results estimated in each time window for each metric, from all 28 subjects measured in each of the 4 environmental scenarios are summarized in Table 2. The optimal values for each condition are shown in bold italics, indicating that Project_ICA performs as well as or better than the POS method for most scenarios and outperforms the other three methods for all scenarios.

Table 2 HR_{est} comparison among different methods on the dataset created in this study.

Conditions	Metrics	ICA	CHROM	2SR	POS	Project_ICA
Stationary	MAD (bpm)	4.96	5.49	32.60	3.36	***3.30***
	RMSE (bpm)	9.18	9.45	59.49	8.64	***7.21***
	r	0.69	0.50	0.12	0.74	***0.76***
Interaction with computer	MAD (bpm)	6.42	6.58	17.42	4.50	***3.93***
	RMSE (bpm)	14.24	10.32	35.19	9.13	***7.10***
	r	0.14	0.46	0.29	0.64	***0.74***
Swinging heads	MAD (bpm)	7.06	7.11	19.47	***5.03***	5.41
	RMSE (bpm)	10.44	10.70	39.72	***8.37***	8.70
	r	0.49	0.41	-0.17	0.68	***0.69***
Exercise	MAD (bpm)	16.86	22.10	27.27	10.07	***9.80***

recovery	RMSE (bpm)	24.89	28.09	47.47	17.90	***17.88***
	r	0.13	0.24	-0.05	***0.49***	0.47

Specifically, Project_ICA is more accurate than the other methods under the stationary, interaction with computer and exercise recovery scenarios. In the other scenarios, the accuracy of the estimation is comparable to the state-of-the-art POS method, and indeed Project_ICA shows the highest correlation between the HR_{est} and HR_{ref}. More detailed comparisons between all methods under different experimental conditions are discussed in the following sections.

4.1 Comparison of subgroups with different skin−tones

The metrics in section3.2 were used to evaluate the performance of the system on subjects with Skin-Tone1 ("dark" skin) and Skin-Tone2 ("pale" skin). The results are summarized in Tables 3 and 4.

Table 3 Analysis of results from subjects with Skin-Tone1.

Conditions	Metrics	ICA	CHROM	2SR	POS	Project_ICA
Stationary	MAD (bpm)	9.70	8.08	83.73	6.69	***6.55***
	RMSE (bpm)	13.92	11.53	96.54	13.83	***11.45***
	r	0.53	0.29	-0.01	***0.57***	0.50
Interaction with computer	MAD (bpm)	12.37	9.29	41.31	8.79	***7.39***
	RMSE (bpm)	21.43	12.61	56.00	14.19	***10.83***
	r	-0.36	0.32	0.31	0.37	***0.48***
Swinging heads	MAD (bpm)	10.31	6.15	50.55	5.62	***5.42***
	RMSE (bpm)	12.78	8.50	69.78	8.33	***8.25***
	r	0.15	0.66	-0.19	0.68	***0.76***
Exercise recovery	MAD (bpm)	28.03	22.89	66.77	20.06	***17.09***
	RMSE (bpm)	32.47	27.51	79.53	28.25	***24.20***
	r	-0.27	***0.26***	-0.02	-0.02	0.15

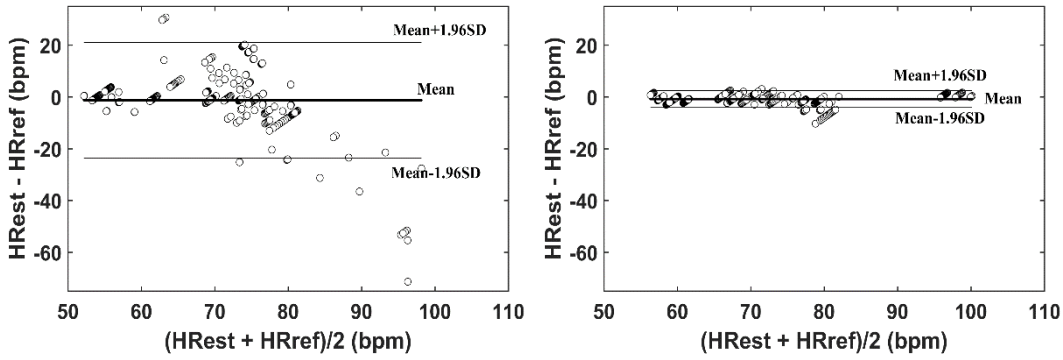
Table 4 Analysis of results from subjects with Skin-Tone2.

Conditions	Metrics	ICA	CHROM	2SR	POS	Project_ICA
Stationary	MAD (bpm)	2.09	3.90	1.36	1.32	***1.31***
	RMSE (bpm)	4.14	7.91	2.41	***1.86***	1.87
	r	0.90	0.64	0.97	***0.99***	0.98
Interaction with computer	MAD (bpm)	2.57	4.81	1.99	1.73	***1.70***
	RMSE (bpm)	6.04	8.51	3.21	***2.63***	2.65
	r	0.77	0.58	0.94	***0.96***	***0.96***
Swinging heads	MAD (bpm)	5.59	7.52	5.56	***4.76***	5.40
	RMSE (bpm)	9.19	11.53	10.31	***8.37***	8.89
	r	0.55	0.23	0.42	***0.63***	***0.63***
Exercise recovery	MAD (bpm)	10.65	21.65	5.33	4.51	***4.09***
	RMSE (bpm)	19.43	28.41	8.22	7.41	***6.61***

	r	0.20	0.18	0.76	0.88	***0.90***

Tables 3 and 4 show that all the signal processing methods and evaluation metrics reported here perform less effectively with Skin-Tone1 than Skin-Tone2 subjects. Nevertheless, for Skin-Tone2, Project_ICA outperforms the other methods in two of the 4 experimental scenarios, the exception being the stationary and swinging heads scenario, under which conditions the POS method was better. For Skin-Tone1 subjects (Table 4), the results are more variable, although in most cases where Project_ICA is outperformed (predominantly by POS) the differences between the two are minimal. We conclude, therefore, that overall, Project_ICA can reliably estimate HR in individuals irrespective of their skin color, although the errors arising from subjects with dark skin are greater than those from lighter skinned subjects.

To further quantify the accuracy of the Project_ICA approach, Figure 10 shows the results of Bland-Altman analysis to compare HR_{est} and HR_{ref} in Skin-Tone1 and Skin-Tone2 subjects.



(a) HR estimated on Skin-Tone1 (b) HR estimated on Skin-Tone2

Figure 10 Bland-Altman plots of HR_{est} and HR_{ref} under stationary conditions. Center solid line indicates mean difference and outer two thin lines refer to boundaries of 95% limits of agreement.

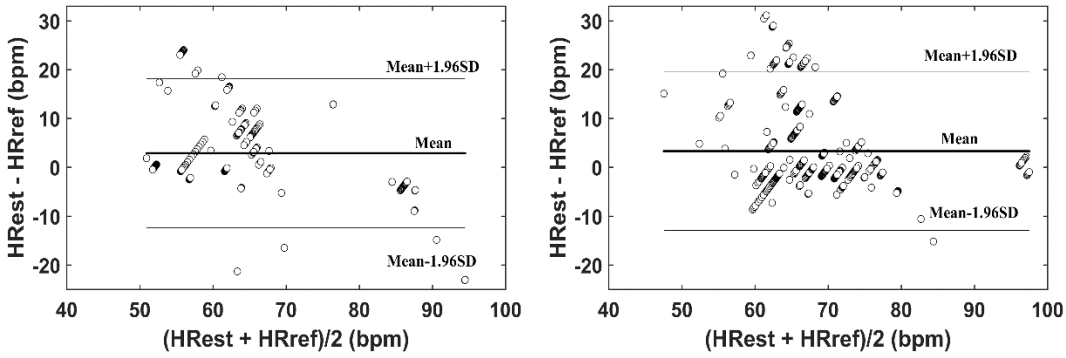
Figure 10 (a) shows that the agreement between HR_{est} and HR_{ref} in Skin-Tone1 subjects is only moderate, with 95% limits of agreement ranging only from −23.65 to 21.02 bpm although the mean bias is close to zero (-1.31 bpm). Results from Skin-Tone2 show that the 95% limits of agreement are narrower, ranging from −4.13 to 2.41 bpm, with mean bias again close to zero (-0.86 bpm) confirming the better performance. This difference can be explained by the optical properties of skin. Most of the incident light captured by the camera is from epidermal reflection. However, this signal does not contain pulsatile information. The remainder of the incident light enters the dermis where some is absorbed by melanin, hemoglobin etc., before emerging from the skin and travelling back to the camera. Only the small fraction of the light that interacts with the blood contains pulsatile information. Since the skin of subjects with Skin-Tone1 contains more melanin, most of the light entering the dermis is absorbed, which results in a pulsatile signal of lower amplitude.

Although, as Table 3 andTable 4 show, most of the signal processing methods described here perform with acceptable accuracy(Hassan *et al.*, 2017) when the subjects are stationary, it is notable that the skin color has a strong effect on HR_{est} when using the 2SR method. This can be explained by the fact that the 2SR method is more sensitive to the spatial distribution of individual skin pixels and for the rPPG signal acquisition process, and skin pixel detection for Skin-Tone1 always introduces more distortion, thus leading to significant differences in measurement error between Skin-Tone1 and Skin-Tone2.

4.2 Performance under interaction with computer and moving head scenarios

Table 2 shows that, for each signal processing method the agreement between HR_{est} and HR_{ref} obtained under the interaction with computer and swinging head conditions was worse than that observed in the stationary scenario, presumably because of motion interference. Nevertheless, the results show that the performance of Project_ICA is comparable to that of the state-of-the-art POS approach, and better than that of the other methods.

Table 3 and 4, again show that under the moving head and interaction with computer scenarios all methods show poorer agreement for Skin-Tone1 subjects than their lighter skinned counterparts, due as mentioned above, to reduced reflection from the darker skin giving less difference between the amplitudes of the RGB color channels. These observations are confirmed by the Bland-Altman plots in Figure 11, which compare HR_{ref} and HR_{est} from the Project_ICA signal analysis, and show that the mean bias for Skin-Tone1 is 2.87 bpm (limits of agreement ranging from -12.36 to 18.10 bpm) and for Skin-Tone2 is 3.27 bpm (limits of agreement ranging from -12.94 to 19.48 bpm).



(a) HR estimated on Skin-Tone1 (b) HR estimated on Skin-Tone2

Figure 11 Bland-Altman plots of HR_{est} and HR_{ref} obtained during the head moving experiments. (a) HR estimated on Skin-Tone1; (b) HR estimated on Skin-Tone2.

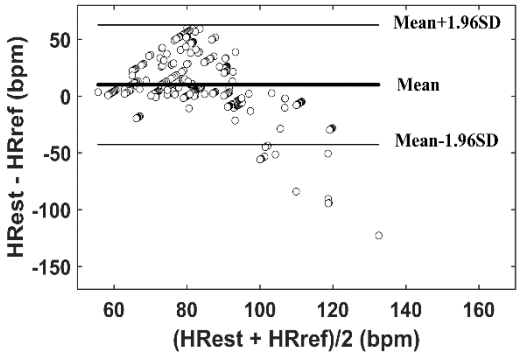
The measurements above the upper 95% confidence line in Figure 11 are assumed to be outliers. Possible reasons for the poor agreement are: firstly, the ROI boundary changes greatly because the feature points move with the head moving. Secondly, with head swing, changes in the area of facial skin blocked by hair at different angles results in errors during the acquisition of the raw signal. Both effects will cause the number of detected skin pixels to vary constantly and will lead to errors both in the original signal and in the subsequent processing. In some

subjects whose head movements are more marked these effects will be more prominent, giving rise to the outliers.

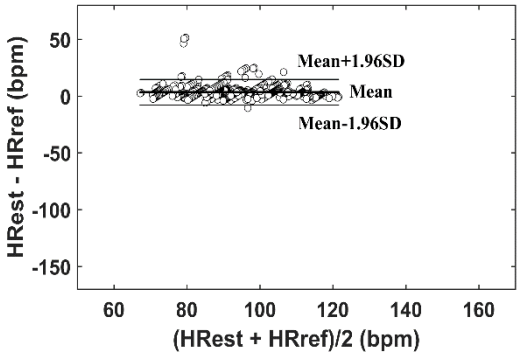
4.3 Effect of exercise recovery

Under the exercise recovery scenario, the HR is expected to change markedly during the measurement period. Consequently, the raw signal was analyzed using 31 overlapping time windows in each minute of recording. Although it can be seen in Table 2 that the two error evaluation metrics for all HR estimation methods are larger in this scenario than the other three scenarios, Project_ICA outperforms the other methods with lower values of MAD and RMSE with HR_{ref}, and better correlation with HR_{ref}.

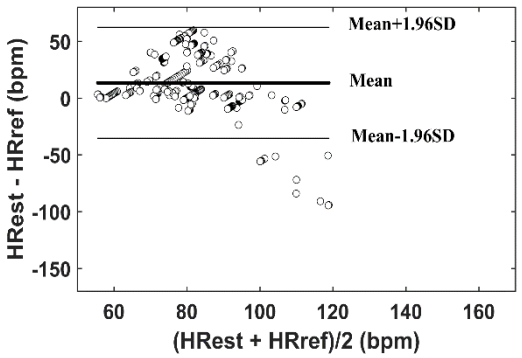
Furthermore, to compare the performance of the HR estimation methods under these conditions of changing HR, Bland-Altman plots for the two skin-tone groups were constructed to show the results for all subjects. Project_ICA, POS and 2SR are selected in the comparison shown in Figure 12.



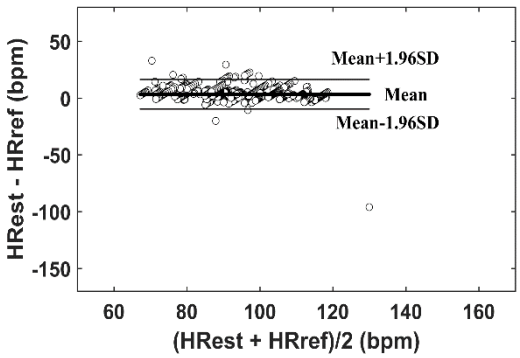
(a) HR estimated for Skin-Tone1 by Project_ICA



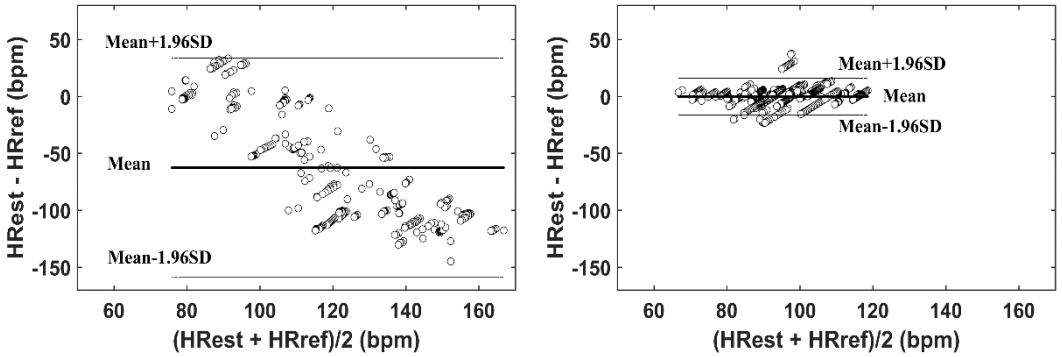
(b) HR estimated for Skin-Tone2 by Project_ICA



(c) HR estimated for Skin-Tone1 by POS



(d) HR estimated for Skin-Tone2 by POS



(e) HR estimated for Skin-Tone1 by 2SR (f) HR estimated for Skin-Tone2 by 2SR

Figure 12. Bland-Altman plots of HR_{est} and HR_{ref} under exercise recovery conditions. (a) and (b) show the limits of agreement between HR_{est} by Project_ICA and HR_{ref} for skin-tone1 and skin-tone2 subjects respectively (c) and (d) show the limits of agreement between HR_{est} by POS and HR_{ref} for the two skin tones (e) and (f) similarly, show the limits of agreement 2SR.

Table 5 summarizes the limits of agreement and bias values derived from the Bland-Altman plots in Figure 12. It can be seen that firstly, for Skin-tone1, the dark-skinned subjects, there is poor agreement with the reference method and marked mean bias with all three analysis methods. On the other hand, for the paler Skin-tone2 subjects the limits of agreement are narrower and the bias values are close to zero. Secondly, the limits of agreement for the Project_ICA and POS analyses are narrower than those of 2SR, and the performance of Project_ICA is marginally better than that of POS. In short, obtaining reliable data at times when HR is changing rapidly is a challenge, especially with dark-skinned subjects.

Table 5 Summary of consistency limits and deviations from Bland-Altman plots in Figures 12.

	Skin-tone1 (bpm)				Skin-tone2 (bpm)			
	Lower limit of agreement	Upper limit of agreement	Range	Bias	Lower limit of agreement	Upper limit of agreement	Range	Bias
Project_ICA	-42.7	62.5	105.2	9.9	-7.9	14.6	22.5	3.3
POS	-35.5	62.2	97.7	13.4	-9.7	16.3	26.0	3.3
2SR	-158.9	33.7	192.6	-62.6	-16.3	15.9	32.2	-0.2

4.4 Effect of short-time HR estimation

To demonstrate the ability of the proposed method for recognizing changes in instantaneous HR, the short-term response of HR_{est} was analyzed by decreasing the length of the time window to 10s with no overlap and recording for 1 minute. As aforementioned, measurements were carried out on Skin-Tone1 and Skin-Tone2 subjects, under the various movement scenarios, and the results were processed using the five methods and three error evaluation metrics described above. The results are summarized in Table 6 and Table 7.

For every participant, HR estimation results were analysed for each time window in one minute and the statistical analysis of overall results is shown in Table 6 and Table 7.

Table 6 Short-time window analysis for Skin-Tone1 subjects.

Conditions	Metrics	ICA	CHROM	2SR	POS	Project_ICA
Stationary	MAD (bpm)	18.10	***8.83***	91.83	22.34	19.25
	RMSE (bpm)	34.02	***11.17***	102.17	35.60	30.81
	r	0.18	0.07	-0.02	***0.29***	0.22
Interaction with computer	MAD (bpm)	20.96	***11.89***	50.38	14.43	16.10
	RMSE (bpm)	31.71	***14.43***	63.46	22.94	24.80
	r	***0.46***	0.15	0.25	0.17	0.07
Swinging heads	MAD (bpm)	16.43	9.09	50.32	10.34	***8.99***
	RMSE (bpm)	21.73	11.42	70.43	14.70	***12.07***
	r	-0.02	0.17	-0.11	***0.49***	0.40
Exercise recovery	MAD (bpm)	19.26	27.24	54.85	18.51	***18.19***
	RMSE (bpm)	22.82	30.51	68.91	***25.11***	25.16
	r	-0.10	-0.09	-0.12	0.21	***0.29***

Table 7 Short-time window analysis for Skin-Tone2 subjects.

Conditions	Metrics	ICA	CHROM	2SR	POS	Project_ICA
Stationary	MAD (bpm)	4.27	8.36	***1.81***	2.12	3.17
	RMSE (bpm)	8.33	12.55	***2.67***	3.02	4.90
	r	0.73	0.08	***0.96***	0.95	0.90
Interaction with computer	MAD (bpm)	5.75	4.16	3.84	3.12	***3.10***
	RMSE (bpm)	8.94	7.39	9.15	***4.78***	4.82
	r	0.61	0.75	0.61	***0.86***	***0.86***
Swinging heads	MAD (bpm)	7.08	10.38	***5.12***	7.40	7.06
	RMSE (bpm)	11.14	13.25	10.98	10.99	***10.68***
	r	0.51	0.11	***0.54***	0.49	0.52
Exercise recovery	MAD (bpm)	18.23	29.99	***4.47***	4.98	5.40
	RMSE (bpm)	25.92	35.18	***6.70***	9.75	10.15
	r	-0.11	-0.08	***0.89***	0.79	0.77

As shown in Table 6 for Skin-Tone1, Project_ICA outperforms the other algorithms under two conditions, the stationary scenario and exercise recovery scenario, giving lower values of MAD and RMSE and higher values of r, in accordance with the conclusions in section 4.1. For Skin-Tone2 (see Table 7), the Project_ICA, POS and 2SR algorithms yielded better results than the other methods for the three evaluation metrics. Previous work (Hassan *et al.*, 2017) has stated that an acceptable margin of error in HR estimation is 5 bpm. Our results show that the accuracy of HR_{est} using Project_ICA meets or comes close to this requirement in Skin-Tone2 subjects; the results of short-time HR estimation on Skin-Tone1 show that Project_ICA performs poorly but better than the other 4 methods under most scenarios. We note that, the 2SR approach is better over the 10s time window than 30s, and the optimal time window length of 2SR is determined by the dynamic range of the subject's HR and the video frame rate (Wang *et al.*, 2016). The results show that Project_ICA can achieve robust short-time HR estimation among individuals with "pale" skin under various scenarios.

5. Conclusion

In this study, we present a novel framework for rapidly and accurately recovering the HR signal from video recordings of the human face. The core idea of the proposed method is to project the normalized original signal onto a plane orthogonal to the unit vector and to use the ICA method to unmix the projected signal. In addition, some advanced techniques, such as feature point detection tracking and skin pixel detection, are utilized to suppress motion distortion and obtain a relatively pure raw signal. Experiments demonstrate that the proposed framework outperforms several classical methods, like the popular ICA-based approach, CHROM, 2SR and POS under the challenging conditions of poorly reflective skin, subject motion and heart rate recovery after exercise. Nevertheless, there are still major problems in obtaining reliable remote HR measurements in dark-skinned subjects. Further work will focus on improving the performance of the proposed method during heart rate recovery and for dark-skinned subjects. In addition, more experiments are needed under complex ambient environmental conditions with various light sources.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Nos. 61773110, 61374015, and 61202258), Natural Science Fund Project of Liaoning Province (No. 20170540312), and the Fundamental Research Funds for the Central Universities (Nos. N161904002 and N161604006).

Reference

Allen J 2007 Photoplethysmography and its application in clinical physiological measurement *Physiol. Meas.* **28** R1-R39

Balakrishnan G, Durand F and Guttag J 2013 Detecting Pulse from Head Motions in Video *IEEE Conference on Computer Vision and Pattern Recognition* 3430-7

Cheng J, Chen X, Xu L and Wang Z J 2016 Illumination Variation-Resistant Video-Based Heart Rate Measurement Using Joint Blind Source Separation and Ensemble Empirical Mode Decomposition *IEEE J. Biomed. Health. Inform.* **21** 1422-33

Cook S, Togni M, Schaub M C, Wenaweser P and Hess O M 2006 High heart rate: a cardiovascular risk factor? *Eur. Heart J.* **27** 2387-93

De Haan G and Arno A J 2014 Improved motion robustness of remote-PPG by using the blood volume pulse signature *Physiol. Meas.* **35** 1913-26

De Haan G and Jeanne V 2013 Robust pulse rate from chrominance-based rPPG *IEEE Trans. Biomed. Eng.* **60** 2878-86

Feng L, Po L M, Xu X, Li Y and Ma R 2015 Motion-Resistant Remote Imaging Photoplethysmography Based on the Optical Properties of Skin *IEEE T. Circ. Syst. Vid.* **25** 879-91

Haque M A, Irani R, Nasrollahi K and Moeslund T B 2016 Heartbeat Rate Measurement from Facial Video *IEEE Intell. Syst.* **31** 40-8

Hassan M A, Malik A S, Fofi D, Saad N, Karasfi B, Ali Y S and Meriaudeau F 2017 Heart rate estimation using facial video: A review *Biomed. Signal Proces.* **38** 346-60

Hsu Y C, Lin Y L and Hsu W 2014 Learning-based heart rate detection from remote photoplethysmography features. *IEEE International Conference on Acoustics, Speech and Signal Processing,* pp 4433-7

Krishnaswamy A and Baranoski G V 2004 A study on skin optics *Technical Report* **1** 1-17

Kumar M, Veeraraghavan A and Sabharwal A 2015 DistancePPG: Robust non-contact vital signs monitoring using a camera *Biomed. Opt. Express* **6** 1565-88

Lam A and Kuno Y 2016 Robust Heart Rate Measurement from Video Using Select Random Patches. *IEEE International Conference on Computer Vision,* pp 3640-8

Lewandowska M, Rumiński J, Kocejko T and Nowak J 2011 Measuring pulse rate with a webcam — A non-contact method for evaluating cardiac activity. *Computer Science and Information Systems,* pp 405-10

Li X, Chen J, Zhao G, Pietik and Inen M 2014 Remote Heart Rate Measurement from Face Videos under Realistic Situations. *Computer Vision and Pattern Recognition,* pp 4264-71

Mcduff D, Gontarek S and Picard R W 2014 Improvements in remote cardiopulmonary measurement using a five band digital camera *IEEE Trans. Biomed. Eng.* **61** 2593-601

Monkaresi H, Calvo R A and Yan H 2014 A machine learning approach to improve contactless heart rate monitoring using a webcam *IEEE J. Biomed. Health. Inform.* **18** 1153-60

Poh M Z, Mcduff D J and Picard R W 2010a Advancements in noncontact, multiparameter physiological measurements using a webcam *IEEE Trans. Biomed. Eng.* **58** 7-11

Poh M Z, Mcduff D J and Picard R W 2010b Non-contact, automated cardiac pulse measurements using video imaging and blind source separation *Opt. Express* **18** 10762-74

Prakash S K A and Tucker C S 2018 Bounded Kalman filter method for motion-robust, non-contact heart rate estimation *Biomed. Opt. Express* **9** 873-97

Qi H, Guo Z, Chen X, Shen Z and Jane Wang Z 2017 Video-based human heart rate measurement using joint blind source separation *Biomed. Signal Proces.* **31** 309-20

Shafer S A 1985 Using color to separate reflection components *Color Res. Appl.* **10** 210-8

Shan L and Yu M 2014 Video-based heart rate measurement using head motion tracking and ICA. In: *International Congress on Image and Signal Processing,* pp 160-4

Stone J. V. 2004 Independent component analysis: a tutorial introduction. MIT press.

Sun Y, Hu S, Azorin-Peris V, Kalawsky R and Greenwald S 2013 Noncontact imaging photoplethysmography to effectively access pulse rate variability *J. Biomed. Opt.* **18** 061205

Sun Y, Papin C, Azorin-Peris V, Kalawsky R, Greenwald S and Hu S 2012 Use of ambient light in remote photoplethysmographic systems: comparison between a high-performance camera and a low-cost webcam *J. Biomed. Opt.* **17** 037005

Takano C and Ohta Y 2007 Heart rate measurement based on a time-lapse image *Med. Eng. Phys.* **29** 853-7

Tomasi C 1991 Detection and tracking of point features *Technical Report* **91** 9795-802

Tran D N, Lee H and Kim C 2015 A robust real time system for remote heart rate measurement via camera. In: *IEEE International Conference on Multimedia and Expo,* pp 1-6

Tsouri G R and Li Z 2015 On the benefits of alternative color spaces for noncontact heart rate measurements using standard red-green-blue cameras *J. Biomed. Opt.* **20** 048002

Tulyakov S, Alameda-Pineda X, Ricci E, Yin L J, Cohn J F and Sebe N 2016 Self-Adaptive Matrix Completion for Heart Rate Estimation from Face Videos under Realistic Conditions *IEEE Conference on Computer Vision And Pattern Recognition* 2396-404

Verkruysse W, Svaasand L O and Nelson J S 2008 Remote plethysmographic imaging using ambient light *Opt. Express* **16** 21434-45

Viola P and Jones M 2001 Rapid Object Detection using a Boosted Cascade of Simple Features. In: *IEEE Computer Society Conference on Computer Vision & Pattern Recognition*, p 511

Wang W, Brinker A C D, Stuijk S and Haan G D 2017a Algorithmic Principles of Remote PPG *IEEE Trans. Biomed. Eng.* **64** 1479-91

Wang W, den Brinker A C, Stuijk S and de Haan G 2017b Robust heart rate from fitness videos *Physiol. Meas.* **38** 1023-44

Wang W, Stuijk S and de Haan G 2016 A Novel Algorithm for Remote Photoplethysmography: Spatial Subspace Rotation *IEEE Trans. Biomed. Eng.* **63** 1974-84

Wang W, Stuijk S and De H G 2015 Exploiting Spatial Redundancy of Image Sensor for Motion Robust rPPG *IEEE Trans. Biomed. Eng.* **62** 415-25

Xu S, Sun L and Rohde G K 2014 Robust efficient estimation of heart rate pulse from video *Biomed. Opt. Express* **5** 1124-35