Research Article

Noncontact Heart Rate Measurement Using a Webcam, Based on Joint Blind Source Separation and a Skin Reflection Model: For a Wide Range of Imaging Conditions

Boyuan Zhang,¹ Hengkang Li,¹ Lisheng Xu ^{1,2}, Lin Qi ^{1,2}, Yudong Yao ³,
and Stephen E. Greenwald ⁴

¹College of Medicine and Biological Information Engineering, Northeastern University, Shenyang 110819, China

²Key Laboratory of Medical Image Computing, Ministry of Education, 110169, China

³Electrical and Computer Engineering Department, Stevens Institute of Technology, Hoboken, NJ 07030, USA

⁴Blizard Institute, Barts & The London School of Medicine & Dentistry, Queen Mary University of London, E1 2AD, London, UK

Correspondence should be addressed to Lin Qi; qilin@bmie.neu.edu.cn

Received 6 March 2021; Revised 16 June 2021; Accepted 21 June 2021; Published 14 July 2021

Academic Editor: Carlos Marques

Copyright © 2021 Boyuan Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Remote photoplethysmography (rPPG) can be used for noncontact and continuous measurement of the heart rate (HR). Currently, the main factors affecting the accuracy and robustness of rPPG-based HR measurement methods are the subject's skin tone, body movement, exercise recovery, and variable or inadequate illumination. In response to these challenges, this study is aimed at investigating a rPPG-based HR measurement method that is effective under a wide range of conditions by only using a webcam. We propose a new approach, which combines joint blind source separation (JBSS) and a projection process based on a skin reflection model, so as to eliminate the interference of background illumination and enhance the extraction of pulse rate information. Three datasets derived from subjects with different skin tones considering six environmental scenarios are used to validate the proposed method against three other state-of-the-art methods. The results show that the proposed method can provide more accurate and robust HR measurement for all three datasets and is therefore more applicable to a wide range of scenarios.

1. Introduction

The heart rate (HR) is a widely used indicator of health status [1, 2]. At present, electrocardiography (ECG) or photoplethysmography (PPG) are well-established as reference methods for HR measurement [3, 4], but they require the application of electrodes or transducers to the skin during the measurement, which may cause discomfort in long-term monitoring and do not lend themselves well to personal use.

In recent years, as an alternative and less obtrusive method, remote photoplethysmography (rPPG) has offered a low-cost and contactless approach for measuring HR with a camera. In 2008, Verkruijsse et al. first introduced ambient light photoplethysmography to implement noncontact HR measurement [5, 6]. In contrast to contact PPG, which in

both reflection and transmission modes, uses a built-in light source, rPPG measures the variation with time of ambient light reflected from the skin, making it necessary to distinguish between the specular and diffuse reflections of light from the skin surface. The diffused component is the reflected light that remains after absorption and scattering in the skin, subdermal tissue, and blood. Consequently, it varies with changes in blood volume and the movement of the blood vessel wall [7–9]. The reflected light can be recorded with a variety of optical sensors, including cameras, and the plethysmographic signal can be retrieved to infer the physiological information of interest. Thus, the rPPG approach for noncontact HR monitoring and telemedicine is receiving more and more attention and, especially in recent months, for its potential application in isolation units to reduce the risk of cross-infection.

However, many factors influence the accuracy of rPPG-based HR measurement. These include the subject's skin tone [10], head movement [10], rapid changes during recovery from exercise [11], variation in ambient illumination [12], and, when imaging the head, facial expression changes associated with talking, for example, during human-machine interaction [13]. To address these issues, many approaches have been proposed, the principal ones of which include signal decomposition, model-based methods, and data-driven methods.

1.1. Signal Decomposition Methods. In early studies, blind source separation (BSS) based on independent component analysis (ICA) [8] or principal component analysis (PCA) [14] was widely used to extract pulse information from the rPPG signal. The BSS methods usually assume that the pulse signal satisfies independence from the other components in the mixed signal and can be extracted by analyzing the correlation among its different components. Recently, some new signal decomposition methods have been reported. Cheng et al. proposed a joint blind source separation and ensemble empirical mode decomposition (JBSS_EEMD) method, which was found to be a feasible way to suppress the effects of varying illumination, by extracting the underlying common light sources [15]. Wei et al. used a second-order BSS method based on two regions of interest (ROIs) and obtained plausible estimates of both the heart and respiratory rates, although the ROIs were not detected automatically [16]. Macwan et al. presented a semiblind source method using a multiobjective optimization approach with autocorrelation as a periodicity measure, which focused on the subject's skin tone and a human-machine interaction scenario, wherein subjects were engaged in a mathematical computer game [17]. Song et al. also proposed a new semiblind source approach using fast kernel density independent component analysis (KDICA), which is a nonparametric BSS method, with specific importance attached to different distances between the participant and the camera [18]. However, these signal decomposition methods only achieve high accuracy under a limited range of conditions, whereas the model-based methods outlined below are more robust for a more extensive variety of scenarios.

1.2. Model-Based Methods. At present, model-based methods are based on the physical principles, which govern the rPPG signal generation. There are two typical models: the skin reflection model [19] and the local invariance of HR model [20]. The skin reflection model-based methods use the color vector information of the different components of the rPPG signal to facilitate the separation of specular and diffuse reflections. De Haan and Jeanne presented a chrominance-based signal processing method, which used the chrominance subspace to extract the pulse signal, although they mainly focused on the exercise recovery scenario [11]. Wang et al. proposed a planar orthogonal-to-skin (POS) method based on a skin reflection model, and they considered subjects' skin tone, recovery after exercise, head movement, and illumination by fluorescent lamps of different colors [19]. In a previous study from our laboratory, Qi et al.

employed a skin reflection model with ICA to present the Project_ICA method and considered the effects of skin tone, exercise recovery, head movement, and human-computer interaction [21]. The local invariance of the HR model, developed by Pilz et al., converts rPPG signals to certain specific spaces and uses the local group invariance of HR to estimate its value. They have successively proposed the diffusion process method [13], the local group invariance method [20], and the spherical method [22] and investigated the effect of head movement and person-to-person interaction.

1.3. Data-Driven Methods. More recently, to address the limitations of conventional heuristic man-made features or engineered models, some researchers have used data-driven methods to infer HR through collected signal datasets. For example, Monkaresi et al. proposed a machine learning-based K -nearest neighbor algorithm [23], using the rPPG signal features for regression analysis, considering both stationary and exercise scenarios. In addition, Chen et al. proposed a deep learning-based convolutional attention network: Deep-Phys [24], and they focused on the subject's skin tone, head movement, and person-to-person communication scenarios. However, data-driven methods require a large amount of training data to feed the network, so they have some limitations in real applications.

To verify the robustness of rPPG-based HR measurement, most of the previous studies have focused on various common scenarios, e.g., the subject's skin tone, exercise recovery, and head movement, but less attention has been paid to poor illumination and illumination variation, both of which can cause interference in facial rPPG signal acquisition. Therefore, we have addressed these two illumination problems by means of a novel rPPG-based HR measurement method using a webcam. The system also functions well under a variety of other conditions, including exercise recovery, head movement, and human-machine interaction, and is therefore applicable to more scenarios than existing methods.

The main contributions of this paper are twofold: (1) We use both the signal decomposition method and the projection process based on a skin reflection model. This approach can eliminate the common illumination component from the facial and background ROI signals and extract the signal which carries the pulse rate information. (2) After ICA processing, to determine the output signal which carries the pulse information, we use the green channel signal in the facial ROI as a reference to identify the pulse signal. To validate the proposed method under a range of real-life conditions, we introduced for testing three datasets which include six different scenarios. The first dataset is from our previous study [21], the second is the Public Benchmark Dataset [25], and the third is a new dataset assembled for this study. We then compare the proposed method with three recently developed methods, using three metrics: Pearson correlation coefficient (r), Mean Absolute Deviation (MAD) (bpm), and Root Mean Square Error (RMSE) (bpm).

The remainder of the paper is structured as follows. Section 2 describes the proposed method and the datasets. The experimental findings are presented in Section 3. The results

are analyzed and discussed in Section 4, and conclusions are drawn in Section 5.

2. Methods and Materials

The proposed method includes the following steps: (1) The facial video is recorded with a standard RGB camera, and the facial and background ROI signals are extracted from each video frame. (2) The JBSS method is used to remove the common illumination component from the facial and background ROI signals, obtaining a new facial ROI signal. (3) The pulse signal is extracted from the new facial ROI signal using the projection method. (4) The green channel signal from the facial ROI is used to identify the pulse signal, and the HR is calculated using the fast Fourier transform (FFT). This method not only removes the illumination variation to enhance the extraction of the pulse information but also solves the problem of no ordering of the output components after ICA. The whole process is shown in Figure 1, and the specific details of the major steps are described in the following Sections 2.1 to 2.4.

2.1. Acquisition of Facial ROI and Background ROI Signals. In this step, the Viola-Jones face detector is used to detect a rectangular facial region in the first video frame, and the facial feature points in the rectangular facial region are then detected using the minimum eigenvalue algorithm [26]. Next, the Kanade-Lucas-Tomasi (KLT) tracking method [27] is utilized to produce the facial feature matching point pairs between the previous and current frames, and a rectangular facial region is generated in each frame, as Figure 2(a) shows. Then, the YCbCr [28] color space is used to detect the skin region, as shown in Figure 2(b). The skin region is treated as the facial ROI, and the nonskin background region is treated as the background ROI. Finally, we calculate the spatial RGB averages of all the pixels in each facial ROI and background ROI at time t to obtain the facial ROI signal and background ROI signal as

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

where $S(t)$ is the facial ROI signal or the background ROI signal from the RGB channels, S_{skin} is the area of facial ROI or the background ROI, and $M(x, y, t)$ is the pixel value of pixel (x, y) in S_{skin} at time t . K is the total number of pixels in S_{skin} .

2.2. Joint Blind Source Separation. rPPG-based HR measurement is often affected by environmental illumination, which is taken as a common factor in the facial and background ROI signals. We employed the JBSS to remove interference caused by the illumination, separating the underlying common data sources into multiple datasets and maintaining a consistent sequence of the extracted sources in the multiple datasets. This method termed independent vector analysis Gaussian (IVA-G) [29] was selected to implement the JBSS framework, and the process of IVA-G is shown in Figure 3.

For a given set of M matrices ($M \geq 2$) as the datasets, each matrix has P rows and Q columns. The m -th matrix $X^{[m]}$

can be expressed as $X^{[m]} = [x_{(1)}^{[m]}, x_{(2)}^{[m]}, \dots, x_{(p)}^{[m]}]^T$, $x_{(n)}^{[m]} \in \mathbb{R}^{1 \times Q}$. Assuming that each matrix $X^{[m]}$ can be composed of L underlying independent sources as

$$X^{[m]} = A^{[m]} \cdot S^{[m]}, \quad (2)$$

where $A^{[m]} \in \mathbb{R}^{P \times L}$ is the mixing matrix and $S^{[m]} \in \mathbb{R}^{L \times Q}$ can be expressed as $S^{[m]} = [s_{(1)}^{[m]}, s_{(2)}^{[m]}, \dots, s_{(L)}^{[m]}]^T$, and $s_{(n)}^{[m]} \in \mathbb{R}^{1 \times Q}$ is the underlying independent source, as shown in Figure 3(a).

Furthermore, L SCV, as defined in [30], can be obtained by extracting the underlying independent sources in each matrix $S^{[m]}$, and the n -th SCV can be expressed as $\text{SCV}_n = [s_{(n)}^{[1]}, s_{(n)}^{[2]}, \dots, s_{(n)}^{[M]}]^T$. Each SCV is independent of all others, and the components within each SCV are correlated. In order to get the target SCV which is regarded as the common component across the given M matrices for each SCV, we can compute a $M \times M$ correlation matrix with $M \times (M - 1)/2$ distinct correlation values. (The correlation values are determined by the linear correlation between the components within the SCV.) Then, using these correlation values, we can analyze how close the components within each SCV are to each other. We assume that all the correlation values for the target SCV are sufficiently high, and we then determine the target SCV [31] by (1) counting the number of correlation values greater than an empirically determined threshold, (2) calculating the ratio of this number to the total number of correlation values for each SCV, and (3) choosing the maximum of this ratio. The procedure is shown diagrammatically in Figure 3(b).

In this study, the facial and background ROI signals are regarded as two matrices $X \in \mathbb{R}^{P \times Q}$ and $Y \in \mathbb{R}^{P \times Q}$, where $M = 2$ and $P = 3$ for the RGB channels, and Q is the number of video frames (see Figure 4). After IVA-G, the target SCV can be extracted and is assumed to be the common illumination interference. To remove this interference, each component in the target SCV is set to zero, and then, the resulting $S_{\text{new}}^{[1]}$ is substituted into equation (2) to get a new X_{new} , which is the new facial ROI signal [15]. For the signal preprocessing, JBSS is applied to the whole facial ROI signal and the background ROI signal is extracted from each video clip.

2.3. Projection Process Based on a Skin Reflection Model. The new facial ROI signal X_{new} contains the specular reflection component due to the direct reflection of illumination from the skin surface, and the diffuse reflection component after the reabsorption of the incident light by the skin and underlying tissue, including blood vessels. In order to extract the diffuse reflection component, which carries the pulse rate information, we use the projection method based on the simplified skin reflection model [19]:

$$X_{\text{new}}(t) = \mathbf{1} \cdot (1 + i(t)) + \mathbf{C} \cdot \mathbf{u}_s \cdot \mathbf{I}_0 \cdot s(t) + \mathbf{C} \cdot \mathbf{u}_p \cdot \mathbf{I}_0 \cdot p(t), \quad (3)$$

where $X_{\text{new}}(t)$ is the skin pixel intensity from the RGB channels at time t , $\mathbf{1} \cdot (1 + i(t))$ represents the illumination

FIGURE 1: Flow chart of the algorithm framework.

FIGURE 2: (a) Feature points in frames n and $n+1$ generated by the KLT algorithm. (b) Extracted facial ROI and background ROI for a representative frame.

FIGURE 3: The process of IVA-G. (a) Extraction of the underlying independent sources. (b) Extraction of the target source component vectors (SCV).

variation component and $\mathbf{1}$ is the unit vector, $u_s \cdot s(t)$ denotes the specular component, $u_p \cdot p(t)$ denotes the diffused component, I_0 denotes the stationary illumination component, and $\mathbf{C} \in \mathbb{R}^{3 \times 3}$ is the diagonal matrix that is used to normalize the steady component.

In order to reduce the signal dimensionality and eliminate the illumination variation component $i(t)$, the time-series signal $X_{\text{new}} \in \mathbb{R}^{3 \times Q}$ is projected onto the specific orthogonal plane $\mathbf{D} \in \mathbb{R}^{2 \times 3}$. The optimal plane $\mathbf{D}$ was selected empirically by studying how the projection axes on the

FIGURE 4: Schematic diagram of JBSS processing.

FIGURE 5: Schematic of the projection process.

projected plane affected the quality of the projected signal, as performed in our previous study [21]:

$$\mathbf{D} = \begin{bmatrix} \mathbf{D}_1 \\ \mathbf{D}_2 \end{bmatrix} = \begin{pmatrix} -0.4082 & -0.4082 & 0.8165 \\ 0.7071 & -0.7071 & 0 \end{pmatrix}, \quad (4)$$

$$\Pr(t) = \begin{bmatrix} \Pr_1(t) \\ \Pr_2(t) \end{bmatrix} = \mathbf{D} \cdot X_{\text{new}}$$

$$= \mathbf{D} \cdot \mathbf{C} \cdot \mathbf{u}_s \cdot \mathbf{I}_0 \cdot s(t) + \mathbf{D} \cdot \mathbf{C} \cdot \mathbf{u}_p \cdot \mathbf{I}_0 \cdot p(t),$$

where $\mathbf{D}_1 \in \mathbb{R}^{1 \times 3}$ and $\mathbf{D}_2 \in \mathbb{R}^{1 \times 3}$ are orthogonal to each other and $\Pr(t) \in \mathbb{R}^{2 \times Q}$ is the projection signal with two components $\Pr_1(t) \in \mathbb{R}^{1 \times Q}$ and $\Pr_2(t) \in \mathbb{R}^{1 \times Q}$ (see the left-hand section of Figure 5). After the projection process, the three-dimensional signal $X_{\text{new}} \in \mathbb{R}^{3 \times Q}$ has been reduced to a two-dimensional signal $\Pr(t)$, and $i(t)$ is eliminated. $\Pr_1(t)$ and $\Pr_2(t)$ in $\Pr(t)$ are assumed to be independent of each other; thus, $\Pr_1(t)$ and $\Pr_2(t)$ can be decomposed using ICA, to retrieve $s(t)$ and $p(t)$, as shown in equation (6):

$$\begin{bmatrix} s(t) \\ p(t) \end{bmatrix} = \mathbf{W} \cdot \Pr(t) = \mathbf{W} \cdot \begin{bmatrix} \Pr_1(t) \\ \Pr_2(t) \end{bmatrix}, \quad (5)$$

where $\mathbf{W} \in \mathbb{R}^{2 \times 2}$ is the unmixing matrix (see the right hand (ICA decomposing section) in Figure 5).

2.4. Calculation of Heart Rate. However, there is no ordering of $s(t)$ and $p(t)$ after ICA processing, so we have to determine which one is the pulse signal. In our previous study [21], we

selected the POS signals [19] for linear correlation analysis with $s(t)$ and $p(t)$ and chose the one more closely correlated to the POS signal as $p(t)$. In contrast, some previous studies have pointed out that the green channel signal in the facial ROI contains more pulse rate information [5, 32]. Lin and Lin also determined that the green channel signal achieved the highest signal-to-noise ratio for a rPPG signal under a variety of different lighting conditions [33]. Consequently, we performed a linear correlation analysis between the green channel signal and the output signals after ICA and then chose the pulse signal $p(t)$ which had the highest correlation with the green channel signal. After $p(t)$ was selected, the components unrelated to HR were removed by applying a band-pass filter with cut-off frequencies of 0.7 Hz and 4 Hz, covering the normal range of human HR from 42 beats per minute (bpm) to 240 bpm [21]. Finally, an FFT was applied to the filtered pulse signal, and the heart rate was taken as the frequency where the spectral power was maximal. The process is shown in Figure 6.

2.5. Experiments and Datasets. To evaluate the algorithm's performance, we selected three datasets for testing: data from a previous study in our laboratory [21], the Public Benchmark Dataset [25], and a dataset collected specifically for this study. All the computations were carried out on a laptop (Intel® Core™ i5-7300HQ CPU with 2.5 GHz processor speed and 8 GB RAM) running MATLAB® 2017a (The MathWorks, Inc.).

2.5.1. Data from a Previous Study in Our Laboratory. This dataset includes 112 different videos from 28 subjects with

FIGURE 6: The process for calculating HR.

TABLE 1: Details of the dataset from a previous study in our laboratory [21].

Scenarios	Description of the scenario
Stationary	All subjects were required to sit still with their faces unobstructed and to look at the camera directly without movement.
Exercise recovery	All subjects were asked to do 20 squats in approximately two minutes. Immediately after the exercise, the subjects were asked to sit still and look directly at the camera.
Swinging head	All subjects were required to swing their heads from side to side at a rate of about 10 times per minute. Some stills from the recording of two subjects with different skin colors are shown in Figure 7.
Human-computer interaction	All subjects were required to play computer games with no constraints on their movement.

FIGURE 7: Snapshots of the swinging head scenario: (a) pale-skinned subject; (b) dark-skinned subject.

various skin colors (18 pale-skinned subjects and 10 dark-skinned subjects) [21]. The subjects were tested in four different scenarios: stationary, exercise recovery, swinging head, and interacting with a computer (playing video games) in a natural light environment. A standard RGB camera (C2070i, Logitech Inc., CA, USA) was used to record the subject's face at 30 frames per second (fps) with a pixel resolu-

tion of 640×480 , and each video record was one-minute long. A transmissive finger pulse oximeter (Yuwell™ YX303, Yuyue Medical Equipment and Supply Co., China.) was used to acquire the reference HR value, averaged over successive 10-second periods. Details of the four scenarios are shown in Table 1. More details of the experimental setup are described in reference [21].

TABLE 2: Video labels in the Public Benchmark Dataset.

Label	Description of the label
LC1	Color temperature 6500 kelvin, light intensity 5.2 lux
LC2	Color temperature 6500 kelvin, light intensity 36 lux
LC3	Color temperature 6500 kelvin, light intensity 187 lux
LC4	Color temperature 6500 kelvin, light intensity 720 lux
LC5	Color temperature 6500 kelvin, light intensity 2720 lux
LC6	Color temperature 2700 kelvin, light intensity 35 lux
LC7	Color temperature 6500 kelvin, light intensity 18 lux; the illumination came from the side of the face
M1	Lighting condition: LC4; natural head motions and rotations
M2	Lighting condition: LC4; continuous rhythmic nodding motion 60 times per minute
M3	Lighting condition: LC4; continuous rhythmic nodding motion 90 times per minute
H1	Lighting condition: LC4; high and descending HR (recorded immediately after running)
P1	Subject #1 with light skin tone
P2	Subject #2 with dark skin tone
P3	Subject #3 with dark skin tone

FIGURE 8: Snapshots from the Public Benchmark Dataset: (a) swinging movement of the head; (b) nodding movement of the head (images from reference [25]).

2.5.2. Public Benchmark Dataset. We chose the Public Benchmark Dataset [25] because, unlike many others, this contains recordings under a variety of lighting conditions during both exercise recovery and head movement scenarios. It contains 21 videos recorded under different conditions from three subjects (one light-skinned and two dark-skinned). All videos were recorded with an RGB HD video camera (GZ-VX815BE HD, JVC Inc., Japan) at a frame rate of 30 fps and a pixel resolution of 1080×1920 . The ground-truth HR was measured using a Mobi ECG device which is CE certified (class 2A, type CF). Since the length of each video clip ranged from 3 minutes to 3 minutes and 15 seconds, we processed the first three minutes of each video and treated this as 3 separate one-minute measurements for each subject. Each video has a filename coded by the specific subject and lighting condition, for example, the filename “P2LC3” means the video of subject #2 under lighting condition LC3. The labels

and corresponding lighting conditions are listed in Table 2. More details of the experiment can be found on the dataset website <https://osf.io/rwsx6/>. Some stills from recordings of the head movement scenario are shown in Figure 8. For the other conditions, the subjects were required to look at the camera without moving.

2.5.3. Dataset Recorded Specifically for This Study. The Public Benchmark dataset considers different illumination intensities and color temperatures but does not consider variation in the background illumination. This variation can affect the accuracy of rPPG-based noncontact HR measurements [15]. For example, when a person is looking at the computer screen, the flickering light of the screen will be reflected on their face. Similarly, the illumination on a driver’s face in a traveling car will be continuously varying in the face of oncoming traffic and other changes in external illumination.

FIGURE 9: Schematic of the illumination variation experiment. (a) Timing of the color temperature changes, where the color temperature of the LED lamp was switched between 3300 kelvin and 5600 kelvin every 10 seconds. (b) Experimental setup.

To further investigate the effect of illumination variation on rPPG-based noncontact HR measurements, we established a varying illumination dataset which, in contrast to the Public Benchmark Dataset, used variations in color temperature rather than intensity. It contains 12 videos from 12 subjects (8 pale-skinned and 4 dark-skinned), all of whom gave their informed consent. The experimental protocol was approved by the ethical review board of our institute. A standard RGB camera (C2070i, Logitech Inc., CA, USA) was used to record signals for one minute at 30 fps with 640×480 -pixel resolution.

All subjects remained stationary, exposed to a natural light background upon which was superimposed the additional illumination, switched between two color temperatures. The additional illumination was provided by an LED lamp (LED308W, Godox Inc., China), placed in front of the subjects at a distance of 1 meter, giving a light intensity on the face of 430 lux (manufacturer's figures). The color temperature, set by a manual control unit, was switched between 3300 kelvin and 5600 kelvin every 10 seconds for a period of one minute, according to the schematic in Figure 9(a). A multiphysiological parameter monitor (NSC-M12P, Neusoft Inc., China) was used to acquire the reference HR signal using a pulse oximeter finger probe, and the reference HR value was averaged over successive 10-second intervals. The whole experimental setup is shown in Figure 9(b).

2.6. Comparison with Existing Methods. In order to evaluate the performance of the proposed method, we compared it with three other recently developed ones: JBSS_EEMD [15], POS [19], and Project_ICA [21]. All these methods were applied to the three datasets. To ensure a fair comparison between the different methods in the preprocessing stage, we used the method in Section 2.1 to extract the original facial and background ROI signals from the video. Then, the facial ROI signal was applied using POS and Project_ICA, and both signals were processed by JBSS_EEMD and the proposed method to estimate the HR. A 30 s time window was used to process each video. Adjacent time windows were overlapped by 29 s so that each one-minute video yielded 31 results. The procedure is illustrated in Figure 10. It is worth noting that in the preprocessing stage, the JBSS method was applied to the whole facial ROI signal (1 min long) and background ROI signal (1 min long), and then, the projection process was applied to all time windows (30 s long).

3. Results

To assess and compare the performance of the four methods on the three datasets, three evaluation metrics were adopted: Pearson correlation coefficient (r), Mean Absolute Deviation (MAD) (bpm), and Root Mean Square Error (RMSE) (bpm):

FIGURE 10: The moving time windows for processing a 60 s rPPG recording.

TABLE 3: Performance of four methods on the dataset from a previous study in our laboratory (pale-skinned tone).

Condition	Metric	JBSS_EEMD	POS	Project_ICA	Proposed method
Stationary	MAD (bpm)	4.84	1.32**	1.31*	1.67
	RMSE (bpm)	8.67	1.86*	1.87**	2.80
	r	0.38	0.99*	0.98**	0.96
Exercise recovery	MAD (bpm)	20.33	4.28**	4.01*	4.58
	RMSE (bpm)	26.65	6.02*	6.28**	7.00
	r	-0.07	0.90*	0.90*	0.89**
Swinging head	MAD (bpm)	6.81	4.18*	4.63**	4.86
	RMSE (bpm)	10.82	7.46*	7.59**	7.59**
	r	0.24	0.68	0.71**	0.73*
Human-computer interaction	MAD (bpm)	5.42	1.54*	1.56**	1.94
	RMSE (bpm)	8.84	2.11*	2.34**	3.13
	r	0.34	0.97*	0.96**	0.93

TABLE 4: Performance of four methods on the dataset from a previous study in our laboratory (dark-skinned tone).

Condition	Metric	JBSS_EEMD	POS	Project_ICA	Proposed method
Stationary	MAD (bpm)	9.74	4.24*	4.58**	6.76
	RMSE (bpm)	12.77	6.52*	6.71**	10.01
	r	0.07	0.76*	0.74**	0.70
Exercise recovery	MAD (bpm)	18.44**	20.49	16.72*	19.35
	RMSE (bpm)	22.26*	27.32	24.34	23.42**
	r	-0.01	-0.03	0.27**	0.28*
Swinging head	MAD (bpm)	15.73	8.61	8.55**	7.26*
	RMSE (bpm)	19.31	11.84**	11.92	9.72*
	r	0.44	0.52**	0.51	0.61*
Human-computer interaction	MAD (bpm)	10.30	6.52**	6.51*	6.86
	RMSE (bpm)	13.32	9.56**	9.39*	10.67
	r	0.04	0.58*	0.58*	0.47**

(a) Stationary condition

(b) Exercise recovery condition

(c) Swinging head condition

FIGURE 11: Continued.

FIGURE 11: Bland-Altman plots of HR_{rppg} against HR_{gt} for the four scenarios. The central solid red line indicates the mean difference, and the outer blue dashed lines show the upper and lower 95% limits of agreement. From (a) to (d), the left and right columns show the plots of the pale- and dark-skinned subjects, respectively.

$$MAD = \frac{1}{n} \sum_{i=1}^n |HR_{rppg}^{(i)} - HR_{gt}^{(i)}|,$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (HR_{rppg}^{(i)} - HR_{gt}^{(i)})^2}{n}}, \quad (6)$$

where $HR_{rppg}^{(i)}$ and $HR_{gt}^{(i)}$ represent the subject's estimated and ground-truth HR values, respectively, at the i -th time window. In order to remove outliers in the estimated HR values, the Boxplot method was employed [34], by choosing 0.8 and 1.2 times the median value as the lower and upper limits, respectively, and removing values outside this range.

3.1. Results from the Dataset Used in a Previous Study from Our Laboratory. This dataset tests the performance of the algorithm under the conditions of no movement, exercise recovery, swinging head, and human-computer interaction. All the measurements were performed on pale- and dark-skinned subjects, and the results are summarized in Tables 3 and 4. The best and second-best agreements are marked with “*” and “**,” respectively.

Tables 3 and 4 indicate that, for most of the scenarios, the results of the proposed method and those of POS and Project_ICA are in good agreement, and these three methods outperform JBSS_EEMD as shown by its higher values of MAD and RMSE and lower values of r . The proposed method had the best performance in two scenarios (exercise recovery and swinging head for dark-skinned subjects) and the second-best in one scenario (swinging head for pale-skinned subjects). Specifically, for the pale-skinned subjects, the POS and Project_ICA methods performed slightly better than the approach proposed here, under the stationary and human-computer interaction scenarios. Furthermore, the results of the proposed method are comparable with POS

and Project_ICA under the exercise recovery and swinging head scenarios. For the dark-skinned subjects, the proposed method is more accurate than POS and Project_ICA under the swinging head scenario. The results of POS and Project_ICA are a little better than that of the proposed method under the stationary and human-computer interaction scenarios. Moreover, the proposed method performs as well as the POS and Project_ICA methods under the exercise recovery conditions.

To evaluate the consistency between the measured HR value and the ground-truth, we constructed Bland-Altman plots comparing HR_{rppg} obtained from the proposed method with HR_{gt} for each of the four scenarios (see Figure 11), considering dark- and pale-skinned subjects separately.

Figure 11 demonstrates that, for the proposed method, the agreement between HR_{rppg} and HR_{gt} under the stationary and human-computer interaction scenarios is better than that under the exercise recovery and swinging head scenarios. The proposed method confirms the superior performance obtained with the pale-skinned subjects under the stationary scenario, with narrower 95% limits of agreement (-6.03 to 4.73 bpm), and with the mean bias close to zero (-0.83 bpm). The results for the pale-skinned subjects under the human-computer interaction scenario show that the 95% limits of agreement are also narrower, ranging from -6.20 to 6.09 bpm with the mean bias close to zero (-0.05 bpm). For all scenarios, the agreement between the proposed method and the ground-truth was closer for the pale-skinned subjects than for the dark-skinned subjects.

3.2. Results from the Public Benchmark Dataset. This dataset has been used mainly to test the performance of the algorithm under various inadequate lighting conditions. In this study, a single light-skinned subject was tested under exercise recovery conditions and with their head moving at different frequencies. The two dark-skinned subjects were only tested under conditions LC1 to LC5, none of which involved the

TABLE 5: Performance of the four methods on the Public Benchmark Dataset (light-skinned tone).

Conditions	Metrics	JBSS_EEMD	POS	Project_ICA	Proposed method
P1LC1	MAD (bpm)	2.91*	31.40	16.92	3.55**
	RMSE (bpm)	3.37*	42.13	29.29	4.41**
	r	0.33**	0.86*	0.03	0.27
P1LC2	MAD (bpm)	15.40**	86.14	87.07	4.25*
	RMSE (bpm)	16.77**	86.33	87.07	5.35*
	r	-0.39	-0.06	0.02**	0.25*
P1LC3	MAD (bpm)	22.76**	59.09	27.33	15.18*
	RMSE (bpm)	24.13**	68.22	32.85	17.26*
	r	-0.17	-0.93	0.16**	0.51*
P1LC4	MAD (bpm)	5.31*	13.24	10.33	5.75**
	RMSE (bpm)	7.35**	14.41	12.09	7.00*
	r	0.04	0.07**	-0.06	0.11*
P1LC5	MAD (bpm)	2.96	15.33	2.81*	2.94**
	RMSE (bpm)	3.75	16.63	3.39*	3.58**
	r	-0.40	-0.40	0.17**	0.23*
P1LC6	MAD (bpm)	10.28	6.46*	7.73	7.67**
	RMSE (bpm)	12.81	7.92*	12.91	9.66**
	r	0.40*	-0.39	-0.28	0.07**
P1LC7	MAD (bpm)	15.65*	88.06	69.18	17.58**
	RMSE (bpm)	17.58*	88.08	75.96	18.22**
	r	0.14**	-0.02	-0.60	0.21*
P1H1	MAD (bpm)	22.17	9.55	8.93*	9.26**
	RMSE (bpm)	29.87	10.31	9.76**	9.45*
	r	-0.74	0.87**	0.88*	0.88*
P1M1	MAD (bpm)	17.92	8.61**	8.87	5.21*
	RMSE (bpm)	11.01	10.44**	10.79	7.00*
	r	-0.14	0.07**	-0.20	0.36*
P1M2	MAD (bpm)	1.08*	2.55	2.09	1.35**
	RMSE (bpm)	1.48*	3.10	2.46	1.64**
	r	-0.32	-0.48	-0.13**	0.06*
P1M3	MAD (bpm)	28.62	28.54**	28.52*	29.12
	RMSE (bpm)	28.63	28.55**	28.52*	29.15
	r	-0.24**	-0.21*	-0.24**	-0.49

Notes: in this and the following table, the conditions are coded as follows: LC1–LC7 refer to the 7 levels of illumination intensities, P1–P3 are the 3 subjects, M1–M3 are head movement rates, and H refers to recordings made immediately after exercise with falling heart rate. (Full details are listed in Table 2.)

head moving. The results for the light-skinned subject are shown in Table 5, and those for dark-skinned subjects are shown in Table 6. The best and the second-best results are marked with “*” and “**,” respectively.

Tables 5 and 6 indicate that, for the 12 scenarios which involve the different illumination conditions (P1LC1–LC7, P2LC1–LC5, and P3LC1–LC5), the proposed method and JBSS_EEMD were more accurate than POS and Project_ICA, and that the proposed method has the best performance

in six scenarios and the second-best in five. It is worth noting that under some low illumination scenarios, such as LC1, LC2, LC3, and LC7, the errors of POS and Project_ICA are much larger than those of the JBSS_EEMD and the proposed method. The results of the proposed method are comparable with POS and Project_ICA under the exercise recovery (P1H1) and swinging head scenarios (P1M1), and these three methods are superior to JBSS_EEMD. It should be noted that the results of all four methods were less accurate when

TABLE 6: Performance of the four methods on the Public Benchmark Dataset (dark-skinned tone).

Conditions	Metrics	JBSS_EEMD	POS	Project_ICA	Proposed method
P2LC1, P3LC1	MAD (bpm)	17.41*	25.95**	30.62	28.72
	RMSE (bpm)	20.97*	29.61**	34.97	30.76
	r	0.59*	0.10	0.51	0.52**
P2LC2, P3LC2	MAD (bpm)	4.12*	23.28	25.13	4.66**
	RMSE (bpm)	6.93*	32.39	33.44	8.05**
	r	0.85*	0.82**	0.78	0.80
P2LC3, P3LC3	MAD (bpm)	18.03*	39.01	39.91	18.98**
	RMSE (bpm)	21.56**	44.71	46.39	21.41*
	r	0.35**	-0.93	-0.93	0.51*
P2LC4, P3LC4	MAD (bpm)	19.66**	24.75	24.17	18.38*
	RMSE (bpm)	23.74**	27.98	27.44	20.81*
	r	0.04	0.09**	-0.04	0.56*
P2LC5, P3LC5	MAD (bpm)	19.68*	26.80	28.13	19.95**
	RMSE (bpm)	22.99**	29.50	30.82	22.84*
	r	0.07**	-0.17	-0.28	0.39*

subjects nodded their heads at a rate of 90 times per minute (P1M3) than when subjects nodded their heads 60 times per minute (P1M2).

For the proposed method, we constructed Bland-Altman plots to show the agreement of the HR_{rppg} and HR_{gt} (see Figure 12). Since the two skin-tone groups were both recorded under LC1 to LC5 light intensities, the results of light-skinned and dark-skinned subjects under these five conditions are plotted alongside each other for ease of comparison.

Figure 12 compares the agreement between HR_{rppg} and HR_{gt} for the subjects with different skin tones and shows that this method performs better for the light-skinned subject than the dark-skinned ones under all conditions.

3.3. Dataset Specific to This Study. This dataset allows evaluation of the performance of the four methods when the subjects are stationary, and the background is varied by step changes in the color temperature of the illumination. Experiments were performed on 8 pale-skinned and 4 dark-skinned subjects, and the results are listed in Table 7. As before, the best and the second-best results are marked with “*” and “**,” respectively.

Table 7 shows that when the background color temperature is repeatedly varied between two levels, the proposed method yields the best results both for the pale- and dark-skinned subjects. Figure 13 further demonstrates the consistency of the proposed method and shows the consistency for the pale-skinned subjects, with the 95% limits of agreement from -7.74 to 10.06 bpm and the mean bias of 1.16 bpm, which is marginally better than that for the dark-skinned subjects, where the corresponding values are -12.82 to 9.45 bpm for the 95% limits of agreement and -1.68 bpm for the mean bias.

4. Discussion

The results from the data obtained in our earlier study [21] show that, with sufficient illumination, POS, Project_ICA, and the proposed method, all of which are based on the skin reflection model, have better outcomes than the JBSS_EEMD approach. This implies that the skin reflection model, which relies on the projection process to weaken the contribution of the specular component of the light reflected from the skin, and thus the processed signal, can better retain the pulse rate information [19, 21]. We also note that, under adequate illumination, the performance of the POS and Project_ICA methods are slightly better than that of the proposed method.

The results from the Public Benchmark Dataset show that POS and Project_ICA performed badly when the illumination was poor. In contrast, JBSS_EEMD and the new method proposed here had relatively better outcomes, although the new method was superior to JBSS_EEMD. Not unexpectedly, under adequate illumination, the proposed method was more accurate than when operating with low illumination. For POS and Project_ICA, the projection process is applied directly to eliminate the component of the signal due to variations in background illumination, giving rise to more efficient separation of the diffuse and direct reflection components [19]. However, when there is inadequate background illumination, the camera cannot capture enough illumination and the projection process does not work well. At the same time, the JBSS preprocessing can effectively reduce some of the undesirable effects of variation in the environmental illumination and, therefore, enhance the performance of the subsequent projection process. We also found the same pattern, namely, good agreement between them for each metric, in the exercise recovery and head swinging scenarios as in the dataset from our earlier study [21].

(a) LC1 condition

(b) LC2 condition

(c) LC3 condition

FIGURE 12: Continued.

FIGURE 12: Bland-Altman plots of HR_{rppg} and HR_{gt} under different light intensities (LC1 to LC5). The central solid red line indicates the mean difference, and the outer blue dashed lines show the upper and lower 95% limits of agreement. From (a) to (e), the left and right columns show the plots for the light- and dark-skinned subjects, respectively.

TABLE 7: Performance of four methods on the dataset specifically compiled for this study.

Participant	Metrics	JBSS_EEMD	POS	Project_ICA	Proposed method
Pale-skinned tone	MAD (bpm)	5.88	4.55**	7.08	3.31*
	RMSE (bpm)	7.90	7.44**	11.04	4.68*
	r	0.62**	0.52	0.20	0.79*
Dark-skinned tone	MAD (bpm)	7.38	4.84*	6.62	4.95**
	RMSE (bpm)	9.07	6.46**	10.50	5.90*
	r	0.21**	-0.02	-0.05	0.24*

In the results from the dataset compiled for this study, when the background illumination was repeatedly switched between two color temperatures, it was found that the proposed method had the optimum results. The main reason is that it first uses JBSS to weaken the adverse effects of changes in environmental illumination and then uses the skin reflection model to more consistently obtain the signal component which contains the pulse rate information.

Under the head movement and exercise recovery scenarios, the results from the Public Benchmark Dataset were similar to those from the dataset used in our previous report [21], and the methods based on the skin reflection model were generally better than the JBSS_EEMD. We also observed that HR errors of subjects under the head movement and exercise recovery scenarios were greater than those of the stationary subjects. We also found that when the frequency of head

FIGURE 13: Bland-Altman plots of HR_{rPPG} and HR_{gt} under repeated stepwise illumination changes. The central solid red line indicates the mean difference, and the outer dashed blue lines refer to the upper and lower 95% limits of agreement. (a) and (b) are the plots of the pale-skinned subjects and the dark-skinned subjects, respectively.

movements is higher, the accuracy of the four methods is lower. The reason for the poor performance in the head movement scenario is that when the subjects' heads are moving rapidly, the detected facial ROI becomes small or is even lost entirely, which leads to the unsatisfactory acquisition of the rPPG signal. The reason for the comparatively poor performance in the exercise recovery scenario is that the subject's HR is expected to change markedly during the measurement period thus perturbing the HR estimation over the sampling period used in this and our previous study [21]. Therefore, this variation cannot be accurately captured, and the average over this period does not represent the exercise HR or the resting value to which it returns. We also observed that POS and Project_ICA performed slightly better than the proposed method under the stationary scenario because, under the conditions of sufficient and invariant illumination, the JBSS is of limited contribution in the preprocessing.

From the results on the three datasets, it is clear that the dark-skinned subjects yield less accurate HR values than those with lighter skins. The reason is that the absorption spectrum of melanin and that of hemoglobin coincide to some extent, so the components of the light reflected from the skin which has interacted with the hemoglobin will be affected by the melanin in the skin [35]. The more melanin, the more significant the impact it has on the HR signal.

Overall, the proposed method combines the advantages of the signal decomposition method and the skin reflection model approach, in which the JBSS is used to remove the common illumination component from the facial ROI and background ROI signals, and the projection process is then used to obtain the signal that retains the pulse rate information. This technique can not only lessen the effect of poor illumination but can also provide an accurate HR measurement for subjects under conditions of inadequate and or repeated changes of illumination. However, the proposed method may be inadequate as a means of processing the facial ROI, resulting in poorer accuracy for the rapid head

movement cases as well as for the subjects with more melanin in the skin. Although we have exploited three datasets that contain a variety of subjects and scenarios, there are still some limitations. The number of subjects in the Public Benchmark Dataset and the number of scenarios in the dataset compiled for this study are still too small to robustly validate the accuracy of noncontact video HR measurements under real-world conditions. In addition, the dataset compiled for this study involves a simple and repeated change in color temperature and is thus only a basic simulation of the more diverse changes in ambient illumination expected in the field. In future studies, we will investigate more realistic changes in illumination and pay further attention to the application of novel sensor technologies, such as fiber optic sensing, perhaps combining them with devices designed to detect force or biomolecules of interest, such as blood glucose or ascorbic acid [36–39], although achieving this in a noncontact device would be a challenging problem.

5. Conclusions

In this study, we have proposed a new method for measuring HR using a webcam, which combines the JBSS and projection methods based on a skin reflection model. The JBSS method can effectively reduce the effect of low-intensity ambient illumination and illumination variations. The projection method based on the skin reflection model requires adequate illumination intensity to effectively extract the pulse rate information. We validated the proposed method on three datasets and compared the results with data from three other studies. Although the proposed method is not optimal under the stationary and human-computer interaction scenarios, differing from the best by a small margin, it can provide better estimates under most conditions, especially under inadequate and varying illumination. Thus, the approach proposed here has the potential to become the method of choice for the noncontact measurement of HR, because it is reasonably accurate and

functions well under a variety of environmental conditions and for subjects with a range of skin tones.

Data Availability

The Public Benchmark Dataset is available at <https://osf.io/rwsx6/>. The datasets used in a previous study from our laboratory and the dataset specific to this study are not available at this time due to participant privacy.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work is supported by the National Key Research and Development Program of China (2020YFC2004400), the National Natural Science Foundation of China (No. 61773110, No. 61374015), the Fundamental Research Funds for the Central Universities (Nos. N181906001, N172008008), and the Open Grant by the National Health Commission Key Laboratory of Assisted Circulation (Sun Yat-sen University) (No. cvclab201901).

References

- [1] P. Palatini, "Need for a revision of the normal limits of resting heart rate," *Hypertension*, vol. 33, no. 2, pp. 622–625, 1999.
- [2] J. Kranjec, S. Begus, G. Gersak, and J. Drnovsek, "Non-contact heart rate and heart rate variability measurements: a review," *Biomedical Signal Processing and Control*, vol. 13, no. 1, pp. 102–112, 2014.
- [3] Y. Sun and N. Thakor, "Photoplethysmography revisited: from contact to noncontact, from point to imaging," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 3, pp. 463–477, 2016.
- [4] J. Allen, "DFE timing acquisition: analysis and a new approach for fast acquisition," *Physiological Measurement*, vol. 36, no. 5, pp. 2193–2196, 2000.
- [5] W. Verkruysse, L. O. Svaasand, and J. S. Nelson, "Remote plethysmographic imaging using ambient light," *Optics Express*, vol. 16, no. 26, pp. 21434–21445, 2008.
- [6] Y. Sun, S. Hu, V. Azorin-Peris, S. Greenwald, J. Chambers, and Y. Zhu, "Motion-compensated noncontact imaging photoplethysmography to monitor cardiorespiratory status during exercise," *Journal of Biomedical Optics*, vol. 16, no. 7, article 077010, 2011.
- [7] C. Takano and Y. Ohta, "Heart rate measurement based on a time-lapse image," *Medical Engineering & Physics*, vol. 29, no. 8, pp. 853–857, 2007.
- [8] M. Z. Poh, D. J. McDuff, and R. W. Picard, "Non-contact, automated cardiac pulse measurements using video imaging and blind source separation," *Optics express*, vol. 18, no. 10, pp. 10762–10774, 2010.
- [9] Y. Sun, S. Hu, V. Azorin-Peris, R. Kalawsky, and S. Greenwald, "Noncontact imaging photoplethysmography to effectively access pulse rate variability," *Journal of Biomedical Optics*, vol. 18, no. 6, 2013.
- [10] P. V. Rouast, M. T. P. Adam, R. Chiong, D. Cornforth, and E. Lux, "Remote heart rate measurement using low-cost RGB face video: a technical literature review," *Frontiers of Computer Science*, vol. 12, no. 5, pp. 858–872, 2018.
- [11] G. De Haan and V. Jeanne, "Robust pulse rate from chrominance-based rPPG," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 10, pp. 2878–2886, 2013.
- [12] M. A. Hassan, A. S. Malik, D. Fofi, N. Saad, B. Karasfi, and Y. S. Ali, "Heart rate estimation using facial video: a review," *Biomedical Signal Processing and Control*, vol. 38, pp. 346–360, 2017.
- [13] C. S. Pilz, J. Krajewski, and V. Blazek, "On the diffusion process for heart rate estimation from face videos under realistic conditions," in *German Conference on Pattern Recognition*, pp. 361–373, Springer, Cham, 2017.
- [14] M. Lewandowska, J. Ruminski, T. Kocejko, and J. Nowak, "Measuring pulse rate with a webcam - a non-contact method for evaluating cardiac activity," in *Federated Conference on Computer Science and Information Systems - FedCSIS*, pp. 18–21, Szczecin, Poland, 2011.
- [15] J. Cheng, X. Chen, L. Xu, and Z. J. Wang, "Illumination variation-resistant video-based heart rate measurement using joint blind source separation and ensemble empirical mode decomposition," *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 5, pp. 1422–1433, 2017.
- [16] B. Wei, X. He, C. Zhang, and X. Wu, "Non-contact, synchronous dynamic measurement of respiratory rate and heart rate based on dual sensitive regions," *Biomedical engineering online*, vol. 16, no. 1, p. 17, 2017.
- [17] R. Macwan, Y. Benezeth, and A. Mansouri, "Heart rate estimation using remote photoplethysmography with multi-objective optimization," *Biomedical Signal Processing and Control*, vol. 49, pp. 24–33, 2019.
- [18] R. Song, S. Zhang, J. Cheng, C. Li, and X. Chen, "New insights on super-high resolution for video-based heart rate estimation with a semi-blind source separation method," *Computers in biology and medicine*, vol. 116, p. 103535, 2020.
- [19] W. Wang, B. D. Brinker, S. Stuijk, and G. D. Haan, "Algorithmic principles of remote-PPG," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 7, pp. 1479–1491, 2017.
- [20] C. S. Pilz, S. Zaunseder, J. Krajewski, and V. Blazek, "Local group invariance for heart rate estimation from face videos in the wild," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, Salt Lake City, USA, 2018.
- [21] L. Qi, H. Yu, L. Xu, R. S. Mpanda, and S. Greenwald, "Robust heart-rate estimation from facial videos using Project_ICA," *Physiological Measurement*, vol. 40, no. 8, 2019.
- [22] C. Pilz, I. Makhlof, V. Blazek, and S. Leonhardt, "On the vector space in photoplethysmography imaging," in *The IEEE International Conference on Computer Vision (ICCV) Workshops*, Seoul, South Korea, 2019.
- [23] H. Monkaresi, R. A. Calvo, and H. Yan, "A machine learning approach to improve contactless heart rate monitoring using a webcam," *IEEE Journal of Biomedical and Health Informatics*, vol. 18, no. 4, pp. 1153–1160, 2014.
- [24] W. Chen and D. McDuff, "DeepPhys: video-based physiological measurement using convolutional attention networks," in *European Conference on Computer Vision ECCV*, Munich, Germany, 2018.
- [25] Centre for Research Data, "Public benchmark dataset for testing rPPG algorithm performance," 2019.

- [26] P. A. Viola and M. J. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, USA, 2001.
- [27] C. Tomasi and T. Kanade, "Detection and tracking of point features," *International Journal of Computer Vision*, vol. 9, no. 3, pp. 137–154, 1991.
- [28] G. R. Tsouri and Z. Li, "On the benefits of alternative color spaces for noncontact heart rate measurements using standard red-green-blue cameras," *Journal of Biomedical Optics*, vol. 20, no. 4, 2015.
- [29] M. Anderson, T. Adali, and X. L. Li, "Joint blind source separation with multivariate Gaussian model: algorithms and performance analysis," *IEEE Transactions on Signal Processing*, vol. 60, no. 4, pp. 1672–1683, 2012.
- [30] T. Kim, T. Eltoft, and T. W. Lee, "Independent vector analysis: an extension of ICA to multivariate components," in *International conference on independent component analysis and signal separation*, pp. 165–172, Springer, Berlin, Heidelberg, 2006.
- [31] Q. Long, S. Bhinge, V. D. Calhoun, and T. Adali, "Independent vector analysis for common subspace analysis: application to multi-subject fMRI data yields meaningful subgroups of schizophrenia," *Neuroimage*, vol. 216, p. 116872, 2020.
- [32] L. F. C. Martinez, G. Paez, and M. Strojnik, "Optimal wavelength selection for non-contact reflection photoplethysmography," in *proceedings of SPIE, 22nd congress of the International Commission for Optics: Light for the Development of the World*, vol. 8011, p. 8011191, 2011.
- [33] Y. C. Lin and Y. H. Lin, "A study of colour illumination effect on the SNR of rPPG signals," in *IEEE. Engineering in Medicine and Biology Society (EMBC)*, pp. 4301–4304, Seogwipo, South Korea, 2017.
- [34] S. Walfish, "A review of statistical outlier methods," *Pharmaceutical Technology*, vol. 30, no. 11, pp. 82–86, 2006.
- [35] G. Zonios, J. Bykowski, and N. Kollias, "Skin melanin, hemoglobin, and light scattering properties can be quantitatively assessed *in vivo* using diffuse reflectance spectroscopy," *The Journal of Investigative Dermatology*, vol. 117, no. 6, pp. 1452–1457, 2001.
- [36] G. L. Arnaldo, F. Anselmo, and M. Carlos, "Optical fiber specklegram sensors for mechanical measurements: a review," *IEEE Sensors Journal*, vol. 20, no. 2, pp. 569–576, 2020.
- [37] G. L. Arnaldo, M. Carlos, and R. N. R. Moises, "FBG-embedded 3-D printed ABS sensing pads: the impact of infill density on sensitivity and dynamic range in force sensors," *IEEE Sensors Journal*, vol. 18, no. 20, pp. 8381–8388, 2018.
- [38] K. Santosh, S. Ragini, and K. K. Brajesh, "Highly sensitive, selective and portable sensor probe using germanium-doped photosensitive optical fiber for ascorbic acid detection," *IEEE Sensors Journal*, vol. 21, no. 1, pp. 1–70, 2020.
- [39] Y. Qingshan, X. Zhang, and K. Santosh, "Development of glucose sensor using gold nanoparticles and glucose-oxidase functionalized tapered fiber structure," *Plasmonics*, vol. 15, no. 3, pp. 841–848, 2020.