Remote Photoplethysmography with An EEMD-MCCA Method Robust Against Spatially Uneven Illuminations

Rencheng Song, *Member, IEEE,* Jiji Li, Minda Wang, Juan Cheng, *Member, IEEE,* Chang Li, *Member, IEEE,* and Xun Chen, *Senior Member, IEEE*

***Abstract*— Remote photoplethysmography (rPPG) is an attractive video-based technique to monitor heart rate (HR) information in telehealth screening. The subjects are required to stay stationary in the remote screening scenario and ambient lights become the main interference source of measuring HR with rPPG. In this paper, we introduce a novel approach robust against spatially uneven illumination interference in rPPG by combining ensemble empirical mode decomposition (EEMD) with multiset canonical correlation analysis (MCCA). We adopt the following procedures to ensure that the pulses dominate the correlations across multiple signal sets while the illumination noises are as diverse as possible. Specifically, a group of optimal regions of interest (ROIs) are selected according to the quality indicators defined from the green channel in each candidate ROI. A multi-channel signal set is then constructed by decomposing the green signal with EEMD. Only those intrinsic mode functions (IMFs) with the dominant frequencies falling into the interested HR range are utilized as the input of MCCA. The canonical variables (CVs) with the highest cross correlations are derived as the underlying candidate pulses. Finally, fast Fourier transform (FFT) is employed to calculate the dominant frequency and the target HR is determined based on both the quasi-periodic property and the continuity of HR. The proposed EEMD-MCCA method is validated on both the in-house BSIPL-RPPG and the public COHFACE databases, which achieves superior performance over several typical rPPG methods. This study will provide a promising tool for realistic rPPG applications in telehealth screening.**

***Index Terms*— Remote photoplethysmography, heart rate estimation, joint blind source separation, uneven illumination**

I. Introduction

Telehealth systems provide medical services to remotely detect, monitor, and diagnose patients with the help of modern communication technology. One of the popular telehealth applications is remote health screening. The accurate measurement of heart rate (HR) is usually required in telehealth systems, since it is closely relevant to the status of human health, mood, and stress etc. The remote photoplethysmography (rPPG) has attracted much attention for non-contact video-based HR measurement in telehealth systems. It extracts pulsatile information from facial videos [1], [2], where the blood volume variations caused by cardiac activities can be measured through changes of skin colors.

In common scenarios of rPPG-based health screening, subjects are asked to put their faces within a predefined region on the screen and they need to keep stationary during the screening. This greatly avoids motion artifacts of rPPG and illumination noise becomes the main interference of the pulsatile signal in such case. For example, the spatially-uneven distribution of lights due to the skewed angle of the face and light sources can cause inconsistent intensities of pulse waveforms in different facial regions. The ambient light variations due to light sources can also severely disturb the subtle color changes caused by heart beating. The two kinds of illumination noise, especially the former one, which is more common for rPPG applications, both need to be eliminated for reliable remote HR screening.

Many efforts have been made to suppress illumination noise from both hardware and algorithms. Infrared cameras are considered to be useful against illumination variations with visible wavelengths. However, the infrared light is not so sensitive to the pulsatile information as the visible light and the popularity of infrared cameras is also not as wide as the RGB ones [3], especially for phone-based screening. Various algorithms have also been proposed to address illumination interference for RGB cameras.

1) Since the videos record simultaneously the pulsatile information and illumination artifacts, one common way of designing illumination-robust methods is to directly decouple the pulse from mixed observation signals. For example, Lin *et al.* [4] introduced the ensemble empirical mode decomposition (EEMD) to decompose the pulse from the green channel. Lam *et al.* [5] took

This work was supported in part by the National Natural Science Foundation of China (Grants 61922075 and 41901350), in part by the Provincial Natural Science Foundation of Anhui (Grant 2008085QF285), and in part by the Fundamental Research Funds for the Central Universities (Grant JZ2019HGBZ0151). (Corresponding author: Juan Cheng).

R. Song, J. Li, M. Wang, J. Cheng, and C. Li are with the Department of Biomedical Engineering, Hefei University of Technology, Hefei 230009, China (e-mail: rcsong@hfut.edu.cn; lijj@mail.hfut.edu.cn; 931000907@qq.com; chengjuan@hfut.edu.cn; changli@hfut.edu.cn).

X. Chen is with the Department of Electronic Engineering and Information Science, University of Science and Technology of China, Hefei, 230001, Anhui, China (e-mail: xunchen@ustc.edu.cn).

the blind source separation (BSS) to estimate HR from multiple patch regions of interest (ROIs). Liu *et al.* [6] employed the homomorphic filtering to linearly decouple the illumination noise, and the pulse was further extracted through an adaptive spatiotemporal filter.

2) Another kind of method takes the background as a reference to estimate the temporal ambient lighting variation and then remove it from the mixture in facial ROI to get the pulse. For example, Cheng *et al.* [7] took the temporal lighting variations from background ROI as a reference and used joint blind source separation (JBSS) [8] to eliminate noise interference caused by illumination variations. Later, Lee *et al.* [9] also employed a similar idea to deal with the illumination-induced challenges under driving conditions.
3) Recently, learning-based methods [10]–[12] have also been proposed to determine pulsatile information from large facial video data sets containing illumination-induced noise. The neural networks learn to build a mapping between the pulse-related input and the target HR using a large and diverse training data to ensure the robustness of the mapping against illumination-induced noise.

Compared to existing studies on temporal illumination variations, less attention is paid to the issue of spatially uneven illuminations on facial skins. In most existing methods, the pulse is still extracted from a large ROI that possibly covers skins with quite different illumination intensities. Particularly, those methods designed using reference from background can not deal with uneven illuminations, because the ambient lights in the background may be different from that of the facial region. On the other hand, the generalization capability for learning-based methods has always been a challenge for limiting their practical applications. It is well known that the illumination can be assumed to be uniform within a small patch. However, the area of the skin ROI is preferred not to be too small, in case of bringing interference of quantization noise. A compromise solution is to extract the HR from multiple small patches. For example, the target HR in [5] was voted from the histogram of all HR estimations independently obtained from small patch ROIs. But its performance strongly depends on the quality of HR estimation on each single patch. Considering these challenges, further research is still required for removing artifacts caused by spatially uneven illuminations in rPPG. This is especially important for remote HR screening in telehealth systems, where other interferences are already minimized when cooperative subject is required to keep stationary.

In this paper, we introduce a novel method applied with multiple patch ROIs to extract pulse signals robust against spatially uneven illuminations. In detail, we take the multi-set canonical correlation analysis (MCCA)-based JBSS [13] technique to extract the shared pulsatile information from multiple patch ROIs. To guarantee the quality of each signal set, two schemes are introduced to prepare the input signal sets. First, multiple patch ROIs are defined and the illumination is considered to be uniform on each patch. We select the four best ones according to the quality indices defined with light intensity, light variation, and signal-to-noise ratio (SNR) of the green signal. This ensures that the ROIs corrupted by noise are removed as much as possible before the pulse extraction. The EEMD is then taken to decompose the single-channel green signal, which contains the strongest pulsatile information [14], from each selected patch. And only those intrinsic mode functions (IMFs) with the dominant frequencies falling into the interested HR range are utilized to construct the multi-channel signal set. This guarantees that the signal set is created using the sensitive signals closely related to HR. The above two procedures are expected to remove the majority of artifacts and make the most correlated information across the signal sets be the HR related ones. In such a way, the four signal sets corresponding to the four optimal patches are sent to MCCA-based JBSS to derive the underlying pulse canonical variables (CVs). Finally, the dominant frequencies of all CVs from the largest cross-correlation coefficients are taken to calculate the target HR. Meanwhile, the HR continuity property is also used to remove HR outliers for a further improved performance of HR measurement.

The main contributions of this paper are summarized as follows. The designed method fully takes use of the spatial illumination variations and the strong cross correlations of pulsatile information in different patches. It adopts reasonable preprocessing techniques, including the use of sensitive green signals from optimal ROIs and the selection of reliable IMFs with EEMD, to ensure the quality of input signal sets, thereby achieving a robust HR extraction against spatially uneven illuminations through MCCA. The proposed method is expected to be useful for remote HR screening in telehealth systems, where uneven illumination is very likely to be the main source of noise for subjects undergoing a cooperative examination.

The rest of the paper is organized as follows. Section II introduces related work. The main procedures of the EEMD-MCCA method are described in Section III. The details of the experiment setup and results are shown in Section IV. Finally, we conclude the work in Section V.

II. Related Work

A. ROI Determination and Optimization

The quality of ROI has a significant influence on the performance of HR measurement for rPPG [15]. For instance, some ROIs are easily blocked by hairs or beards, which contain no pulsatile information [16]. Some ROIs can be easily disturbed by non-rigid motions like talking or eye blinking. Besides, due to varied distributions of facial capillaries, some ROIs like cheeks and forehead, result in stronger pulse signals [17]. It is also demonstrated in [17] that the skew angle between face and light source could lead to spatially uneven illumination across the skin surface. Consequently, the determination of high-quality facial ROIs is of great significance to enhance the performance of HR measurement.

In the early rPPG research, the ROI was usually determined as a bounding box covering facial skin region. For instance, Verkruysse *et al.* [14] and Lewandowska *et al.* [18] manually selected a rectangle on forehead as the ROI. Lempe *et al.*

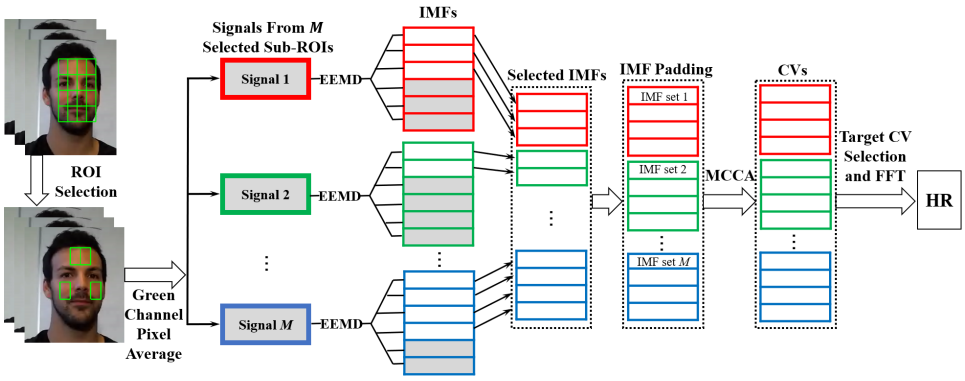


Fig. 1. Flowchart of the proposed method.

[19] found that the cheek ROIs facilitated the pulse extraction after comparing the performance of rPPG on several manually selected ROIs. Later, Kwon *et al.* [20] also showed that the forehead and both cheek ROIs were more appropriate to derive the pulse signal with stronger strength.

According to existing studies, the optimal ROIs are usually selected based on smaller patches from the whole forehead or cheek regions. Quality indices are defined on all patches to evaluate and determine the candidate ROIs. For example, Feng *et al.* [21] divided a fixed ROI into non-overlapped patches, and they selected the high-quality ones according to both the SNR and the cross correlation (CC) coefficient of pulse waveforms from two neighboring windows. Po *et al.* [22] determined the target facial ROIs through comparing the SNR map of rPPG pulse signals on different candidate ROIs. Kumar *et al.* [23] extracted the pulse signals by weighting all candidate rPPG signals with weight coefficients determined from the SNR of pulse extracted on each patch. From these works, it can be seen that the SNR is an important quality indicator and it has often been used to obtain high-quality ROIs. Besides SNR, Bousefsaf *et al.* [24] also indicated that the lightness criteria was also a critical indicator to derive reliable ROI for designing illumination resistant algorithms. They introduced a way to select high-quality ROIs considering both lightness and SNR.

B. BSS and JBSS Based rPPG Method

It is known that the pulsatile information is mixed with specular and diffuse reflections as a linear combination [25], [26] in RGB channels. The BSS-based methods aim to retrieve the pulse signal according to the statistical properties of target signals. The BSS methods can work with a single signal set or multiple signal sets, where the latter is also called the JBSS, which can further take use of the correlation of sources across the multiple sets.

For a single observation set, the ICA-type BSS methods become the mainstream. It extracts the pulse from a single set considering the independence of source signals. The observation set of ICA is usually built from RGB signals [27], [28] in defined ROIs. Since the quality of observation set seriously affects the performance of BSS, the observation set can also be built with monochrome channels from multiple ROIs [5], [29] or a monochrome channel using the delay-coordinate transformation [30], [31]. The ICA-based methods have an inherent permutation issue that the orders of the extracted components are random. However, the selection of target pulse signal from all separated sources is not a trivial task. Besides, the number of sources to be separated needs to be predefined in ICA. The inconsistency between the assumed number of sources and the actual one will also degrade the performance of ICA.

For multiple observation sets, the JBSS [32], [33] extracts underlying sources within each signal set while keeping a consistent ordering of the extracted sources across multiple signal sets. For example, Qi *et al.* [33] utilized MCCA to extract the common pulse signals in multiple facial ROIs. Similar to BSS methods, the performance of the JBSS methods is also affected by the quality of input sets. Besides, the source components in multiple sets need to be consistent with the assumptions of the used JBSS method. For a special case of two observation sets, the canonical correlation analysis (CCA) can be used to extract the shared pulse information. For example, Al-Naji *et al.* [34], [35] decomposed the green-channel signal by the complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN), and the CCA was further taken to obtain shared pulse signals. The quality of observation sets in [34] was effectively improved compared to the direct usage of RGB ones. However, the two input sets obtained from two large ROIs can be easily contaminated with common noises, which affects the accuracy of the results.

In this study, we introduce the EEMD-MCCA framework to work with multiple observation sets defined from optimal patch ROIs. We will verify that the selection of proper ROIs, construction of data sets with IMFs from green channels, have a great impact on the performance of JBSS. It makes the pulse signals be more relevant while the noise signals are more diverse across observation data sets. So the extracted signals by MCCA are more prone to be pulse signals.

III. METHODS

The flowchart of the proposed EEMD-MCCA framework is illustrated in Fig. 1. First, a full ROI is determined and divided into multiple small patches. A green-channel signal is

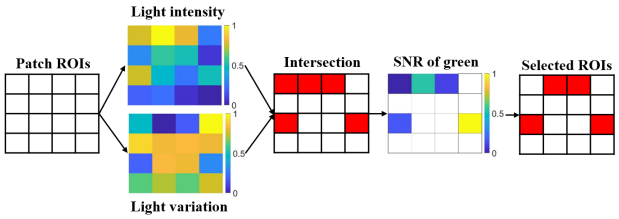


Fig. 2. The details of ROI selection.

then obtained through a frame-by-frame pixel averaging within each patch. Second, the four best patches (termed as sub-ROIs) are selected using quality indicators for high illumination intensity, low spatial variation and high SNR. Third, the green signal of each sub-ROI is decomposed into IMFs by EEMD algorithm and only those with the dominant frequencies falling into the interested HR range are retained to construct a multi-channel signal set for each sub-ROI. All the four signal sets are sent to MCCA to derive the CVs, which usually include the pulses in the first group with the largest cross correlation coefficients. Finally, the pulse signal is selected from the target CV and the corresponding HR is estimated by fast Fourier transform (FFT).

A. Green Signal Extraction

The Viola-Jones face detector [36] is used to get a rectangular bounding box, which is scaled to 60% to remove non-skin areas such as background and hair. The full ROI is then divided into $N \times N$ patches, where the four vertexes of each patch are tracked by using the Kanade-Lucas-Tomasi algorithm [37] along video frames. A green-channel signal is obtained by frame-by-frame pixel averaging within each patch. It is then detrended and band-pass filtered within [0.7 2.5] Hz for further processing.

B. Selection of High-Quality Sub-ROIs

As mentioned above, the quality of sub-ROIs has an important influence on the rPPG performance. It is known that average light intensity on each sub-ROI will affect the intensity of corresponding rPPG pulse. The variations of light intensity within a sub-ROI usually indicate uneven illuminations. These two indicators help to identify whether the current sub-ROI tends to be uniform with possible strong pulse. Additionally, we also employ the SNR of the green signal as a direct indicator of signal quality. These three indicators are combined to determine high-quality sub-ROIs are shown in Fig. 2.

In detail, we first calculate the indicators of light intensity and variation on a single sub-ROI as the mean and standard deviation (SD) of the intensity channel L in LAB color space. The mean value of channel L for each patch is then sorted in a descending order, while that of the SD of light variations is sorted in an ascending order. Accordingly, the first $Q(Q < N^2)$ patches are determined respectively with the highest intensities and the lowest light variations. The SNRs of the green signals from those intersection patches are then sorted in a descending order, and the first M intersected patches are determined as high-quality sub-ROIs for subsequent processing.

In this study, the whole facial ROI derived by the face detector is divided into 4×4 patches (N equals to 4), Q equals to 10, and M is set as 4, which means four high-quality sub-ROIs will be determined for the following processing.

C. EEMD-MCCA algorithm

The JBSS requires signal sets composed by multiple-channel signals. As a result, the single-channel green signal on each sub-ROI needs to be converted into a multi-channel signal set. The ensemble empirical mode decomposition (EEMD) [38] is such a kind of method to decompose a non-stationary single-channel signal into the multi-channel signals in terms of multiple IMFs. The derivation of an IMF requires the following two conditions: (1) the number of extreme values and the number of zero crossing must either be equal or differ at most by one; and (2) at any point, the mean value of the envelope defined by the local maxima and the envelope defined by the local minima is zero [39]. The decomposed signal $\mathbf{g}$ can be expressed as

$$\mathbf{g} = \sum_{i=1}^{I} \mathbf{IMF}_i + \mathbf{r}_I, \tag{1}$$

where $\mathbf{IMF}_i$ is the ith IMF, and $\mathbf{r}_I$ is the residual. Here, each green signal is decomposed into several IMFs by EEMD, and those IMFs with dominant frequencies falling into the interested HR range [0.7 2.5] Hz will be retained to form the multi-channel signal set. The purpose is to remove noisy components for the next MCCA processing. Since the number of retained IMFs may be different among the M signal sets, we further pad the set with a small number of signals using its first reserved IMF. The prepared signal sets will be taken as the input of MCCA-based JBSS technique.

The MCCA is a method to deal with cross correlations between multiple signal sets. Given M signal sets $\mathbf{X}_i(t) = \{x_i^1(t), x_i^2(t), ..., x_i^K(t)\}^{\mathrm{T}}, i = 1, 2, ..., M$, where K is the total number of channels within each signal set, and the superscript T denotes the transpose operation. Each signal set is assumed to be a linear mixture of several underlying uncorrelated source signals as follows

$$\mathbf{X}_i(t) = \mathbf{A}_i \mathbf{S}_i(t), i = 1, 2, ..., M \tag{2}$$

where $\mathbf{S}_i(t) = \{s_i^1(t), s_i^2(t), ..., s_i^K(t)\}^T, i = 1, 2, ..., M$ are the unknown sources and $\mathbf{A}_i$ is the mixing matrix.

Suppose $\mathbf{y}_i$ is a linear combination for the ith signal set $\mathbf{X}_i(t)$ as

$$\mathbf{y}_i(t) = \mathbf{w}_i^{\mathrm{T}} \mathbf{X}_i(t), \tag{3}$$

where $\mathbf{w}_i$ is the unknown combination coefficient vector. The MCCA [13], [40] aims to maximize the correlations of $\mathbf{y}_i(t)$ across all sets by searching the coefficients $\mathbf{w}_i$ for $i = 1, 2, \ldots, M$. The obtained $\{\mathbf{y}_i(t)\}_{i=1}^M$ are also known as canonical variables (CVs). The calculation of CVs needs to be carried out repeatedly on all sets until reaching the dimension K. The coefficient vectors $\mathbf{w}_i$ can be obtained

through different types of objective functions. In this paper, we employ the objective function 'SSQCOR', which maximizes the sum of squares of all cross-correlation coefficients of CVs. More details can be found in [13].

We finally get the de-mixing matrix $\mathbf{W}_i$ composed by all of the obtained $\mathbf{w}_i$ for the ith signal set $\mathbf{X}_i(t)$, such that

$$\mathbf{Y}_i(t) = \mathbf{W}_i\mathbf{X}_i(t), i = 1, 2, ..., M \tag{4}$$

where $\mathbf{W}_i$ is the inverse of unknown mixing matrix $\mathbf{A}_i$ and $\mathbf{Y}_i(t)$ is the obtained approximation of sources $\mathbf{S}_i(t)$.

D. HR Estimation

The final procedure is to determine the target CV that contains the pulse and calculate the corresponding HR. The pulse signals are considered to be highly correlated, while the illumination noise is varied across among different sub-ROIs. It implies that the target pulses are usually included within the first group of CVs. Accordingly, for the current processing segment, we select the candidate pulse as the one with the highest ratio of energy around the fundamental frequency and first harmonic in the first group of CVs. The FFT is then taken to calculate the HR.

The above procedure can well determine most of the HR measurements. However, there are still few outliers that destroy the stability of HR measurement. For example, if the noise components are shared in multiple input sets, it can lead to the real HR appearing in the second group of CVs. In order to achieve a more stable HR measurement, we further introduce a two-step flowchart in Fig. 3 to remove outliers according to HR continuity under the assumption that obtained HRs do not dramatically change during a short period. The core idea is to first find a reference HR_{ref} to determine the outliers (Fig. 3(a)), and then replace the outliers with other candidate HRs from the other CVs (Fig. 3(b)). The details of the flowchart are described as below.

1) First, we determine the reference HR to identify outliers as shown in Fig. 3(a). For a one-minute video, we get seven segments if the processing window is set as 30 seconds with an overlapping of 5 seconds. Accordingly, we get $\{\mathrm{HR_n}\}_{n=1}^{7}$ for each segment from the first group of CVs. The absolute errors between each $\mathrm{HR_n}$ and the remaining ones are calculated and compared with the predefined threshold $Th1$ to determine a number P. The current $\mathrm{HR_n}$ will be treated as a valid HR candidate to calculate the HR reference $\mathrm{HR_{ref}}$ if $P > Th2$. Finally, the $\mathrm{HR_{ref}}$ is determined as the mean value of all valid HR candidates.
2) Second, we remove the outliers with the help of reference $\mathrm{HR_{ref}}$ as shown in Fig. 3(b). Similar to the first step, we calculate absolute error between the current $\mathrm{HR_n}$ and the reference $\mathrm{HR_{ref}}$ and compare it with the threshold $Th1$. If it is identified as an outlier, the $\mathrm{HR_n}$ will be replaced by the $\mathrm{HR_{n,new}}$ obtained from the rest CVs as the one closet with $\mathrm{HR_{ref}}$.

The reliability of the reference HR in Fig. 3(a) depends on two premises. First, most of the candidate $\{\mathrm{HR_n}\}_{n=1}^{7}$ need to be valid. Second, the true HR needs to be continuous for stationary subjects. It has been reported in [41] that the resting HR has small variations during one minute. And we will verify the first condition in the experiment.

In the current study, the error threshold $Th1$ is set to 9.0 beats per minute (bpm) in the above flowchart, while the threshold $Th2$ of P is set to 3. Here $Th2 = 3$ implies that the selected $\mathrm{HR_n}$ is close to at least half of the other HR measurements within one minute. Although the above settings are suitable for subjects who are generally stationary, the parameters still need to be carefully set according to actual distributions of heart rates in different applications.

IV. Experiments and Results

In this section, we will evaluate the proposed EEMD-MCCA and compare it with other typical methods on the in-house BSIPL-RPPG database and the public COHFACE database [42] to illustrate its effectiveness. In detail, we first introduce the experimental setup, and then the results are presented. we also take ablation study to further understand the influence of ROI selection and the HR continuity.

A. Experiment Setup

The BSIPL-RPPG database was collected indoors [43] with natural light sources. With the approval of the ethics review committee of Hefei University of Technology, 37 healthy student subjects (24 males and 13 females) participated in the experiment and a total of 37 videos were collected. The subjects were asked to sit by the window which caused uneven illuminations. During a 4.5-minute video recording, the subject needs to stay still for the first 2 minutes and then imitate head shaking while driving for the rest 2.5 minutes. The videos were recorded by Logitech C920 pro HD with a resolution of 640 x 480 with a frame rate of 30 frames per second (fps). The PPG signal was synchronously acquired at a 60 Hz sampling rate by a Contec CMS50E pulse oximeter. All videos were stored in the MPEG-4 format.

The COHFACE database was also employed to evaluate the performance of our method. It contains 160 RGB videos of approximately one minute in length collected from 40 subjects (28 males and 12 females) under two different light conditions. One light condition employed the ceiling lights and a 400W halogen spotlight, making the illuminations be even on the faces of subjects. The other one was the natural light illuminating from the window, resulting in uneven illumination to the subject. The videos were recorded by Logitech HD Webcam C525 with a resolution of 640 x 480 and a frame rate of 20 fps. The reference PPG signals were synchronously collected by BVP model SA9308M.

In the experiment, six videos from subject 24 and subject 26 in COHFACE database were excluded. The dark skin color of subject 24 makes the processing of all the four videos from this subject very challenging for all comparison methods we have tried. For subject 26, the camera was shaken during video capturing under natural light situations and the related two videos were dropped. Consequently, there are total 154 videos in COHFACE database employed for further analysis.

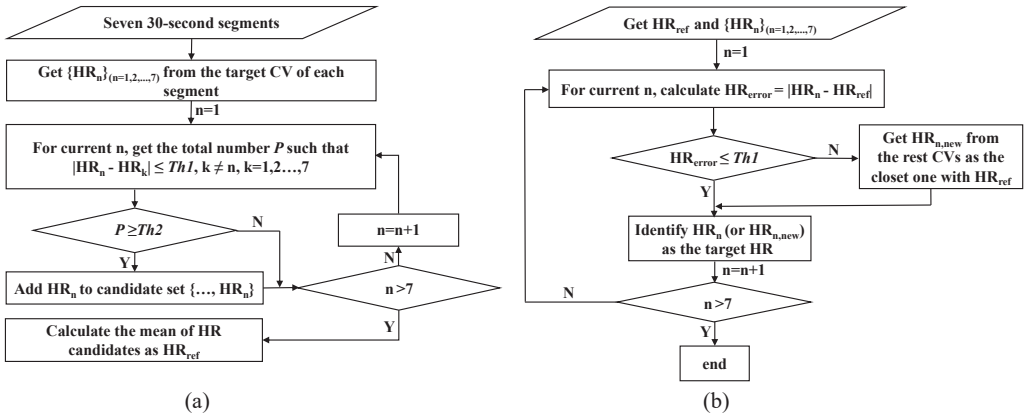


Fig. 3. The flowchart of removing outliers according to HR continuity. (a) Calculate the reference $\mathbf{HR_{ref}}$ (b) Remove HR outliers.

In the BSIPL-RPPG database, we process all videos from 20 seconds to 80 seconds, while the full length of each video is processed in the COHFACE database. The processing segment was set as 30 seconds with an overlapping length of 5 seconds for neighbouring windows. Accordingly, there are total 7 segments to be processed for a one-minute video. Finally, there are 259 video clips obtained from BSIPL-RPPG database and 1082 video clips from COHFACE database.

The proposed EEMD-MCCA algorithm was implemented with MATLAB R2019. Particularly, the EEMD [38] was implemented with the open-source code from github [1]. The MCCA [13] was implemented with the open-source joint multiset decomposition code [2].

B. Results

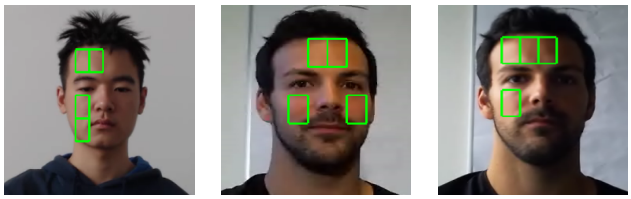
To verify the performance of the proposed method, we compare it with some typical methods on the two databases. These methods including the ICA [27], CHROM [26], POS [25] and iBCG [44] implemented with the 'iPhys' [45]. Besides, to show advantages over existing work, we also compare another four methods listed as follows. All methods were run on a desktop with Intel Core i5 processor (3.4GHz) and 8GB RAM.

(1) To verify the importance of employing multiple patch ROIs against uneven illuminations, the EEMD-CCA is also adopted for comparison. The EEMD-CCA has been implemented with the same procedures as those of EEMD-MCCA, except that the best two sub-ROIs are combined into one single ROI, while the remaining two sub-ROIs are merged into another one. The CCA is then taken to replace the MCCA in the EEMD-MCCA.

(2) To demonstrate the superiority of JBSS over the BSS technique, we compare the proposed method with the MRICA (multiple-region ICA) [29], which directly applies ICA on a signal set composed by green signals obtained from four high-quality sub-ROIs.

(3) To verify the advantage of EEMD-MCCA for using IMFs in MCCA, we also compare it to the MCCA method with input sets composed by RGB signals from each sub-ROI. The latter one is termed as RGB-MCCA. Particularly, we also implement the original MCCA method in [33] for a comparison, which is the same as RGB-MCCA except the selection of target pulse using a spectrum clustering (SC).

Four metrics are employed to evaluate the performance of each method. They are the root mean square error HR_{rmse}; the mean absolute error HR_{mae}; the standard deviation HR_{sd} of the absolute errors, and the Pearson's correlation coefficient r. The detailed definitions and descriptions can refer to [46].



(a) (b) (c)

Fig. 4. Examples of determining high-quality sub-ROIs under different light conditions. (a) Natural light conditions on BSIPL-RPPG database; (b) Good indoor light conditions on COHFACE database; (c) Natural light conditions on COHFACE database.

1) Selection of Optimal Sub-ROIs: Fig. 4 shows an example of selecting high-quality sub-ROIs based on the indicators described in Section III-B. Specifically, Fig. 4(a) indicates sub-ROIs selection under natural light conditions in BSIPL-RPPG database, Fig. 4(b) shows the selected sub-ROIs under indoor lighting conditions in COHFACE database, and Fig. 4(c) shows the selected sub-ROIs under natural lighting situations in COHFACE database. It can be seen from Fig. 4(b) that, when the light shines evenly on the face, the high-quality sub-ROIs are mainly determined from the forehead and both cheek regions with rich capillaries. In contrast, the high-quality ROIs are mainly obtained around the bright parts of the forehead and cheek regions in Fig. 4(a) and Fig. 4(c), respectively. This is in accordance with the fact that the natural light illuminates from the window, making one side of the subject's face more bright than the other side.

2) Results of HR Estimation: The overall HR measurement results of different methods on BSIPL-RPPG and COHFACE databases are summarized in Table I and Table II, respectively.

[1] https://github.com/benpolletta/HHT-Tutorial/blob/master/HuangEMD/eemd.m

[2] http://mlsp.umbc.edu/resources.html#Multiset

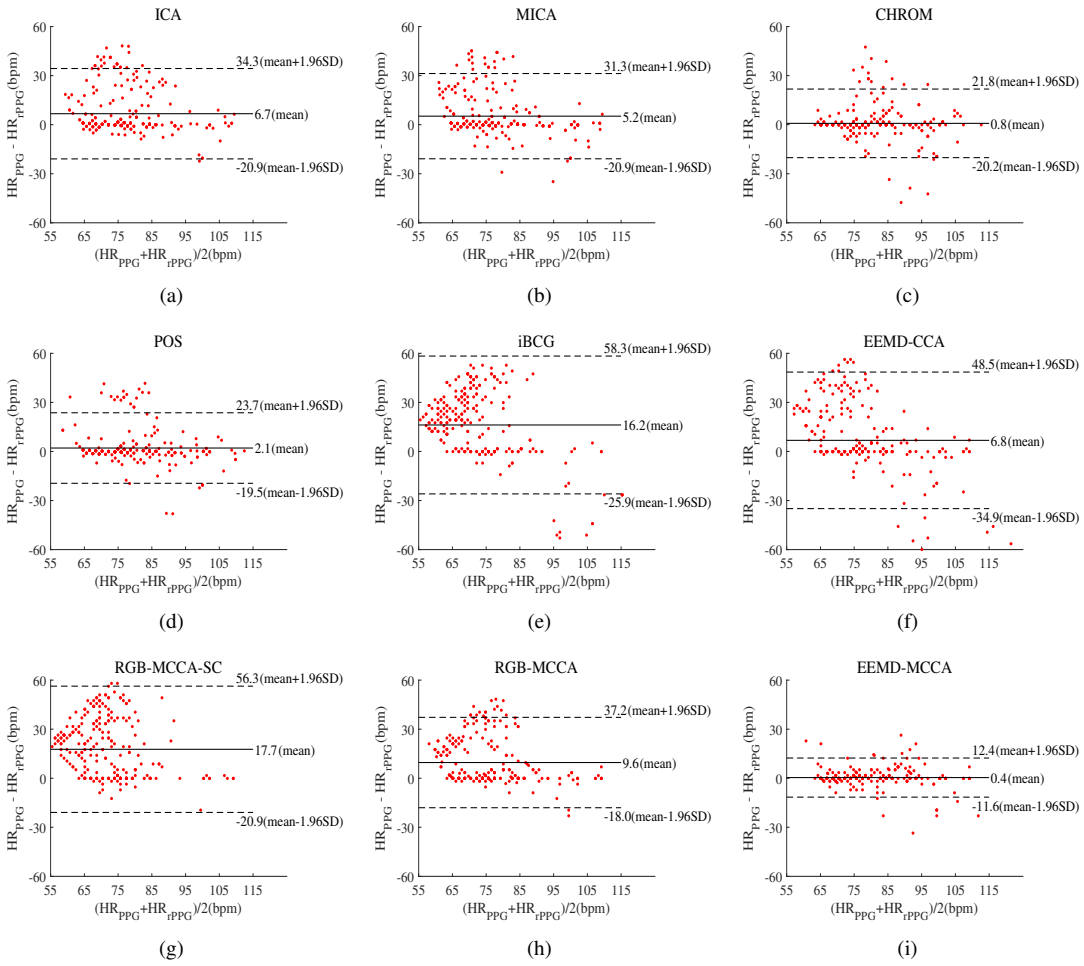


Fig. 5. The Bland-Altman plots of HR measurements on the BSIPL-RPPG database.

TABLE I
RESULTS OF DIFFERENT RPPG METHODS FOR BSIPL-RPPG DATABASE (BEST PERFORMANCE IN BOLD).

Method	HR_{rmse} (bpm)	HR_{mae} (bpm)	HR_{sd} (bpm)	r
ICA [27]	15.56	8.78	12.85	0.40
MRICA [29]	14.27	7.93	11.86	0.51
CHROM [26]	10.69	5.31	9.27	0.60
POS [25]	11.14	5.43	9.73	0.60
iBCG [44]	26.86	21.12	16.59	0.15
EEMD-CCA	22.32	13.82	17.53	0.20
RGB-MCCA-SC [33]	26.43	19.48	17.86	0.06
RGB-MCCA	17.06	10.49	13.46	0.42
EEMD-MCCA	**6.14**	**2.95**	**5.38**	**0.86**

Particularly, existing results (denoted in *italic*) from some latest methods [10]–[12], [47], [48] obtained on the COHFACE database are also listed in Table II for a comparison. It can be observed that the proposed EEMD-MCCA achieves overall superior performance in both the two databases compared to other methods.

In detail, the results of model-based methods including CHROM and POS are not satisfactory in both of the two databases. The probable reason is that the videos of these two databases are stored in a compressed format that may degrade the performance of model-based methods. The performance of

TABLE II
RESULTS OF DIFFERENT RPPG METHODS FOR COHFACE DATABASE: BOTH LIGHTING AND NATURAL CONDITIONS (BEST PERFORMANCE IN BOLD).

Method	HR_{rmse} (bpm)	HR_{mae} (bpm)	HR_{sd} (bpm)	r
ICA [27]	13.99	8.16	11.37	0.36
MRICA [29]	11.26	5.87	9.62	0.55
CHROM [26]	13.74	8.44	10.84	0.34
POS [25]	11.90	6.58	9.92	0.49
iBCG [44]	14.09	8.20	11.47	0.39
HR-CNN [47]	*10.78*	*8.10*	-	*0.29*
Two-stream CNN [10]	*9.96*	*8.09*	-	*0.40*
DeeprPPG [11]	*7.06*	*3.07*	-	*0.86*
Siamese-rPPG [12]	***1.29***	***0.70***	-	*0.73*
MOMBAT [48]	-	*5.89*	*7.38*	*0.62*
EEMD-CCA	16.33	8.58	13.90	0.39
RGB-MCCA-SC [33]	16.03	10.17	12.41	0.24
RGB-MCCA	10.70	5.51	9.17	0.60
EEMD-MCCA	4.80	2.08	**4.33**	**0.91**

Siamese-rPPG [12] is very amazing on this database. However, we need to indicate that all the results of learning-based methods [10]–[12], [47] have been obtained in a within-database configuration. Namely, videos from part of the subjects were taken for training, while the rest ones were employed for testing.

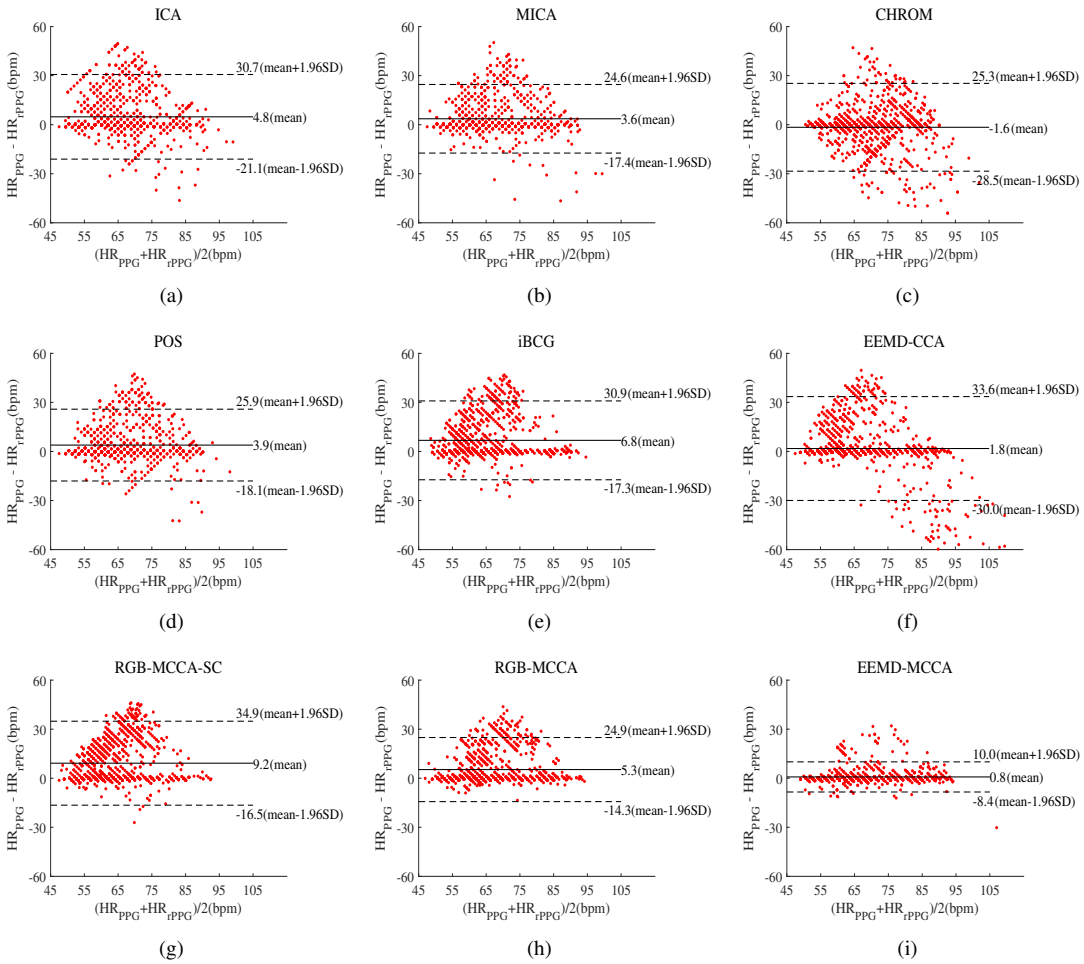


Fig. 6. The Bland-Altman plots of HR measurements on the COHFACE database.

The comparison with RGB-MCCA demonstrates the superiority of employing IMFs decomposed from green-channel signal to construct multi-channel single set in EEMD-MCCA, which is more sensitive to the pulsatile information and more robust against noise compared to the raw RGB signals. The HR_{rmse} and HR_{mae} are decreased by 10.92 bpm and 7.54 bpm, respectively on the BSIPL-RPPG database, while they are decreased by 5.90 bpm and 3.43 bpm, respectively on the COHFACE database. Besides, the correlation coefficients *r* are improved from 0.42 to 0.86 on the BSIPL-RPPG, and from 0.60 to 0.91 on the COHFACE databases.

The average results of EEMD-MCCA are also much better compared to the EEMD-CCA, which verifies the importance of employing multiple sub-ROIs for BSS methods depending on source correlations. Besides, the comparison with MRICA indicates the advantage of EEMD-MCCA over BSS method.

Since the videos from COHFACE database were recorded under two different illumination conditions, we also summarize the results under different lighting scenarios as shown in Table III and Table IV, respectively. We can see that the performance of different methods is all improved under good illumination conditions compared to that of the natural light conditions. It indicates that the illumination has a great impact on the rPPG methods. However, the proposed method still works well under the natural light conditions on COHFACE database through employing reasonable treatment against spatially uneven Illuminations.

TABLE III
RESULTS OF DIFFERENT RPPG METHODS FOR COHFACE DATABASE UNDER ONLY LIGHTING CONDITIONS (BEST PERFORMANCE IN BOLD).

Method	HR_{rmse} (bpm)	HR_{mae} (bpm)	HR_{sd} (bpm)	r
ICA [27]	12.63	6.42	10.88	0.48
MRICA [29]	10.52	4.95	9.29	0.61
CHROM [26]	13.31	7.69	10.86	0.39
POS [25]	10.19	4.86	8.96	0.63
iBCG [44]	13.55	7.38	11.37	0.42
EEMD-CCA	13.32	5.65	12.06	0.54
RGB-MCCA-SC [33]	14.49	8.14	11.99	0.35
RGB-MCCA	9.80	4.77	8.57	0.68
EEMD-MCCA	**3.26**	**1.41**	**2.94**	**0.96**

The Bland-Altman plots of all methods on the two databases are shown in Fig. 5 and Fig. 6, respectively. In detail, the results in Fig. 5 shows that the proposed EEMD-MCCA method obtains results more consistent with the reference results compared to the other methods. Similar results have also been observed in Fig. 6 for COHFACE database. The box plots are also shown in Fig. 7 to represent the depicting groups of RMSE of different methods through their quartiles,

TABLE IV
RESULTS OF DIFFERENT RPPG METHODS FOR COHFACE DATABASE UNDER ONLY NATURAL CONDITIONS (BEST PERFORMANCE IN BOLD).

Method	HR_{rmse} (bpm)	HR_{mae} (bpm)	HR_{sd} (bpm)	r
ICA [27]	15.30	10.00	11.59	0.25
MRICA [29]	12.20	6.85	9.85	0.49
CHROM [26]	14.18	9.23	10.77	0.29
POS [25]	13.47	8.38	10.55	0.33
iBCG [44]	14.64	9.06	11.50	0.36
EEMD-CCA	19.01	11.67	15.01	0.26
RGB-MCCA-SC [33]	17.52	12.27	12.50	0.11
RGB-MCCA	12.74	7.50	10.30	0.44
EEMD-MCCA	**6.03**	**2.79**	**5.34**	**0.85**

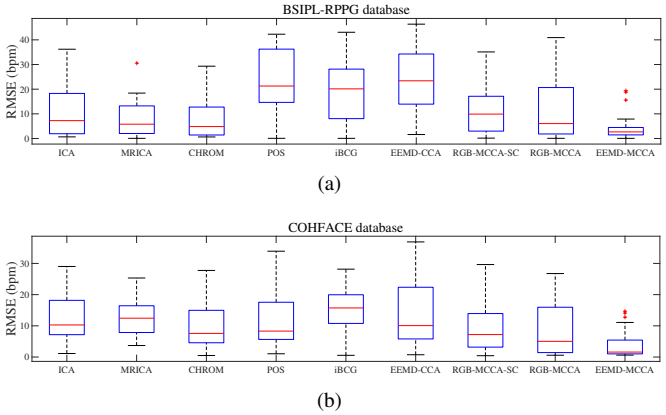


Fig. 7. The box plots of RMSE for different rPPG methods. (a) BSIPL-RPPG database (b) COHFACE database.

where the red line in each box indicates the median value of RMSE obtained by the corresponding method. It can be seen that the median RMSE for the proposed method is 2.69 bpm on BSIPL-RPPG database and 1.59 bpm on COHFACE database, while that of the other methods are all above 4.5 bpm. Besides, the shorter interquartile ranges also indicate there are less dispersed results in the EEMD-MCCA compared to the other methods. In short, the HRs estimated by the proposed method are more accurate on both the BSIPL-RPPG and COHFACE databases.

Fig. 8 shows the detailed HR estimations (blue plus signs) of the proposed method compared with the ground truth HR (red plus signs) on all video clips. It can be seen that most of the blue signs are coincide with the red ones, which demonstrates the effectiveness of the proposed EEMD-MCCA method.

3) Time cost: The time cost of EEMD-MCCA method is mainly determined by two parts, the decomposition of each green signal with EEMD and the MCCA to extract the shared pulsatile information from multi-channel signal sets. We have done repetitive experiments to evaluate the computational cost of the EEMD-MCCA algorithm. It is found that the bottleneck for real-time computation is the complexity of EEMD. It takes 14.891 seconds while only 0.027 seconds are costed by MCCA for processing a single sample.

However, the calculation of EEMD can be accelerated by fast algorithms to achieve real-time performance. It has been reported in [49] that the complexity of fast EEMD can achieve an order of $O(n \log n)$ which is equivalent to FFT. Besides, the

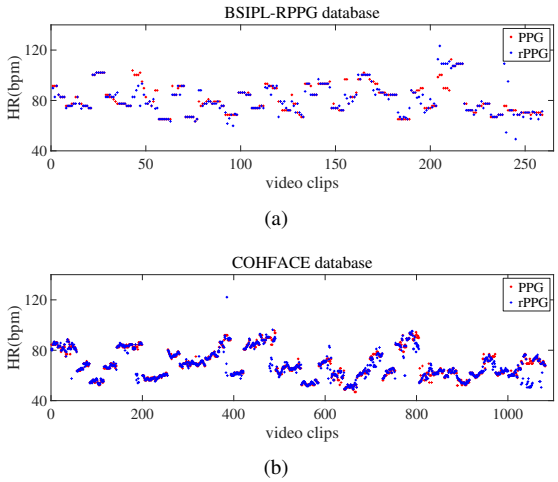


Fig. 8. The estimated HRs for all the video clips both by our proposed method and the ground truth PPG. (a) BSIPL-RPPG database (b) COHFACE database.

EEMD was applied sequentially for decomposing each green signal in our existing implementation. The EEMD-MCCA method can be further accelerated through parallel usage of EEMD on multi-core processors.

C. Ablation study

The above results on two databases verify the superior performance of the proposed method against spatially uneven illuminations. To further understand the factors that affect the performance, we take ablation study on the following aspects. First, we check the effectiveness of employing high-quality sub-ROIs as well as removal of outliers according to HR continuity. Second, we further check the influence of selecting pulses from different groups of CVs to further understand the robustness of the proposed method.

1) Influence of Sub-ROI Selection and HR Continuity: First, we investigate the effect of selecting high-quality sub-ROIs and the outlier removal through HR continuity. We conduct experiments under four different conditions as shown in Fig. 9. Particularly, the 'woR' here indicates without ROI selection, which means the EEMD-MCCA is directly applied to signal sets obtained from all 16 sub-ROIs. The 'woC' means the target pulse is selected only from the first group of CVs as the one with the highest ratio of energy around the fundamental frequency and first harmonic. It can be seen that the EEMD-MCCA method achieves the best performance under the case of 'wR-wC' on both the two databases. The performance degrades if either the option 'woR' or the option 'woC' is employed. Particularly, the comparison between 'wR-woC' and the 'wR-wC' verifies the effectiveness of HR continuity. Similarly, the quality metrics also improve significantly through comparing the cases of 'woR-wC' and the 'wR-wC', which proves the importance of sub-ROI selection.

2) Influence of Selecting HR from Different Groups of CVs: Table V and Table VI summarize the results of HR estimation when selecting the target HR from different groups of CVs using HR continuity strategy. Here EEMD-MCCA(1) means the target HR is only selected from the first group of CVs,

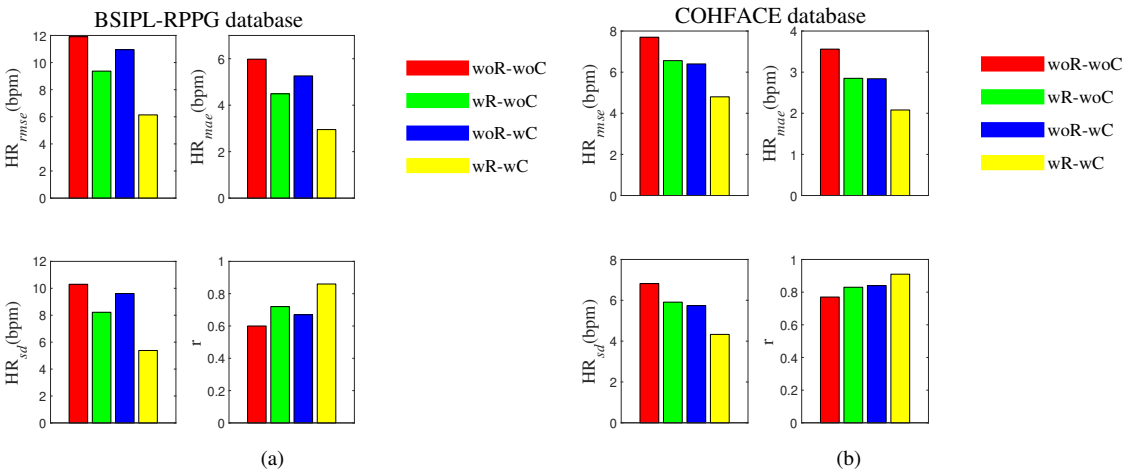


Fig. 9. The experimental results under four conditions: woR-woC means without ROI selection and without HR continuity; wR-woC refers to with ROI selection and without HR continuity; woR-wC means without ROI selection and with HR continuity; wR-wC indicates with ROI selection and with HR continuity. (a) BSIPL-RPPG database (b) COHFACE database.

TABLE V
RESULTS OF EEMD-MCCA WITH TARGET HRS SELECTED FROM DIFFERENT GROUPS OF CVS ON BSIPL-RPPG DATABASE (BEST PERFORMANCE IN BOLD).

Method	HR_{rmse} (bpm)	HR_{mae} (bpm)	HR_{sd} (bpm)	r
EEMD-MCCA(1)	6.91	3.24	6.10	0.83
EEMD-MCCA(2)	6.24	3.01	5.47	0.86
EEMD-MCCA(a)	**6.14**	**2.95**	**5.38**	**0.86**

TABLE VI
RESULTS OF EEMD-MCCA WITH TARGET HRS SELECTED FROM DIFFERENT GROUPS OF CVS ON COHFACE DATABASE (BEST PERFORMANCE IN BOLD).

Method	HR_{rmse} (bpm)	HR_{mae} (bpm)	HR_{sd} (bpm)	r
EEMD-MCCA(1)	5.04	2.20	4.53	0.90
EEMD-MCCA(2)	4.95	2.14	4.48	0.90
EEMD-MCCA(a)	**4.80**	**2.08**	**4.33**	**0.91**

EEMD-MCCA(2) indicates the target HR can be selected from the first two groups of CVs, and EEMD-MCCA(a) means the target HR can be determined from all the groups of CVs.

It can be seen from the two tables that the performance of EEMD-MCCA(1) does not degrade much with the benchmark EEMD-MCCA(a). For example, the HR_{rmse} in Table V slightly decreases from 6.91 bpm to 6.14 bpm, and r increases from 0.83 to 0.86 on the BSIPL-RPPG database. The results imply that almost all target HRs are selected in the first group of CVs. As for BSIPL-RPPG database, there are 98.46% (255/259) clips that the target HR is determined from the first group of CVs. Similarly, for COHFACE database, there are 97.69% (1057/1082) clips that the target HR is obtained from the first group of CVs. These results verify the assumption that the most relevant sources shared among all facial ROIs are the pulsatile signals, thereby laying the foundation for the application of MCCA to extract the HR from the first group of CVs.

V. CONCLUSION

In this paper, we have proposed a method of EEMD-MCCA to estimate HR from facial videos against spatially uneven illuminations. We have adopted two strategies to prepare input signal sets for robustness of HR extraction in MCCA algorithm. The high-quality sub-ROIs are firstly selected using quality indicators defined with the illumination intensity, the illumination variations and the SNR of green signal. The EEMD is then taken to decompose the green signal from each selected sub-ROI, where the resulting IMFs are employed to construct a multi-channel input signal set for MCCA. According to these two strategies, the CVs derived by MCCA with the highest correlations are prone to be candidate pulses. The dominant frequency of each CV is calculated by FFT, and the target HR is determined according to the quasi-periodicity and continuity of HR. The performance of the EEMD-MCCA method has been evaluated on the in-house BSIPL-RPPG database and the public COHFACE database, where the experimental results have demonstrated the superior performance of EEMD-MCCA over several other typical rPPG methods. Besides the ambient light interference, the proposed method is also expected to work for possible non-rigid motion artifacts [2] due to facial expressions. The kind of noise is also considered to be spatially uneven, and the EEMD-MCCA framework is also potentially effective in such case.

REFERENCES

[1] P. V. Rouast, M. T. 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.

[2] X. Chen, J. Cheng, R. Song, Y. Liu, R. Ward, and Z. J. Wang, "Video-based heart rate measurement: Recent advances and future prospects," *IEEE Transactions on Instrumentation and Measurement*, vol. 68, no. 10, pp. 3600–3615, 2018.

[3] S. Kado, Y. Monno, K. Yoshizaki, M. Tanaka, and M. Okutomi, "Spatial-spectral-temporal fusion for remote heart rate estimation," *IEEE Sensors Journal*, vol. 20, no. 19, pp. 11 688–11 697, 2020.

[4] K.-Y. Lin, D.-Y. Chen, and W.-J. Tsai, "Face-based heart rate signal decomposition and evaluation using multiple linear regression," *IEEE Sensors Journal*, vol. 16, no. 5, pp. 1351–1360, 2015.

[5] A. Lam and Y. Kuno, "Robust heart rate measurement from video using select random patches," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 3640–3648.

[6] X. Liu, X. Yang, D. Wang, and A. Wong, "Detecting pulse rates from facial videos recorded in unstable lighting conditions: An adaptive spatiotemporal homomorphic filtering algorithm," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, p. 4001215, 2021.

[7] J. Cheng, X. Chen, L. Xu, and Z. J. Wang, "Illumination variation-resistant video-based heart rate measurement using joint blind source separation and ensemble empirical mode decomposition," *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 5, pp. 1422–1433, 2016.

[8] X. Chen, Z. J. Wang, and M. McKeown, "Joint blind source separation for neurophysiological data analysis: Multiset and multimodal methods," *IEEE Signal Processing Magazine*, vol. 33, no. 3, pp. 86–107, 2016.

[9] K. Lee, J. Lee, C. Ha, M. Han, and H. Ko, "Video-based contactless heart-rate detection and counting via joint blind source separation with adaptive noise canceller," *Applied Sciences*, vol. 9, no. 20, p. 4349, 2019.

[10] Z.-K. Wang, Y. Kao, and C.-T. Hsu, "Vision-based heart rate estimation via a two-stream cnn," in *2019 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2019, pp. 3327–3331.

[11] S.-Q. Liu and P. C. Yuen, "A general remote photoplethysmography estimator with spatiotemporal convolutional network," in *2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020)(FG)*. IEEE Computer Society, pp. 685–692.

[12] Y.-Y. Tsou, Y.-A. Lee, C.-T. Hsu, and S.-H. Chang, "Siamese-rppg network: remote photoplethysmography signal estimation from face videos," in *Proceedings of the 35th Annual ACM Symposium on Applied Computing*, 2020, pp. 2066–2073.

[13] Y.-O. Li, T. Adali, W. Wang, and V. D. Calhoun, "Joint blind source separation by multiset canonical correlation analysis," *IEEE Transactions on Signal Processing*, vol. 57, no. 10, pp. 3918–3929, 2009.

[14] W. Verkruysse, L. O. Svaasand, and J. S. Nelson, "Remote plethysmographic imaging using ambient light." *Optics Express*, vol. 16, no. 26, pp. 21 434–21 445, 2008.

[15] P. Li, Y. Benezeth, K. Nakamura, R. Gomez, C. Li, and F. Yang, "Comparison of region of interest segmentation methods for video-based heart rate measurements," in *2018 IEEE 18th International Conference on Bioinformatics and Bioengineering (BIBE)*. IEEE, 2018, pp. 143–146.

[16] R. Van Luijtelaar, W. Wang, S. Stuijk, and G. de Haan, "Automatic roi detection for camera-based pulse-rate measurement," in *Asian Conference on Computer Vision*. Springer, 2014, pp. 360–374.

[17] M. A. Hassan, A. S. Malik, D. Fofi, N. Saad, B. Karasfi, Y. S. Ali, and F. Meriaudeau, "Heart rate estimation using facial video: A review," *Biomedical Signal Processing and Control*, vol. 38, pp. 346–360, 2017.

[18] M. Lewandowska, J. Rumiński, T. Kocejko, and J. Nowak, "Measuring pulse rate with a webcama non-contact method for evaluating cardiac activity," in *2011 Federated Conference on Computer Science and Information Systems (FedCSIS)*. IEEE, 2011, pp. 405–410.

[19] G. Lempe, S. Zaunseder, T. Wirthgen, S. Zipser, and H. Malberg, "Roi selection for remote photoplethysmography," in *Bildverarbeitung für die Medizin 2013*. Springer, 2013, pp. 99–103.

[20] S. Kwon, J. Kim, D. Lee, and K. Park, "Roi analysis for remote photoplethysmography on facial video," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2015, pp. 4938–4941.

[21] L. Feng, L.-M. Po, X. Xu, Y. Li, C.-H. Cheung, K.-W. Cheung, and F. Yuan, "Dynamic roi based on k-means for remote photoplethysmography," in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2015, pp. 1310–1314.

[22] L.-M. Po, L. Feng, Y. Li, X. Xu, T. C.-H. Cheung, and K.-W. Cheung, "Block-based adaptive roi for remote photoplethysmography," *Multimedia Tools and Applications*, vol. 77, no. 6, pp. 6503–6529, 2018.

[23] M. Kumar, A. Veeraraghavan, and A. Sabharwal, "DistancePPG: Robust non-contact vital signs monitoring using a camera," *Biomedical Optics Express*, vol. 6, no. 5, pp. 1565–1588, 2015.

[24] F. Bousefsaf, C. Maaoui, and A. Pruski, "Automatic selection of webcam photoplethysmographic pixels based on lightness criteria," *Journal of Medical and Biological Engineering*, vol. 37, no. 3, pp. 374–385, 2017.

[25] W. Wang, A. C. den Brinker, S. Stuijk, and G. de Haan, "Algorithmic principles of remote PPG," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 7, pp. 1479–1491, 2016.

[26] 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.

[27] 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. 10 762–10 774, 2010.

[28] 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.

[29] R. Favilla, V. C. Zuccalà, and G. Coppini, "Heart rate and heart rate variability from single-channel video and ica integration of multiple signals," *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 6, pp. 2398–2408, 2018.

[30] Y. Sun, S. J. Hu, V. Azorin-Peris, S. Greenwald, J. Chambers, and Y. S. Zhu, "Motion-compensated noncontact imaging photoplethysmography to monitor cardiorespiratory status during exercise," *Journal of Biomedical Optics*, vol. 16, no. 7, p. 077010, 2011.

[31] F. Zhao, M. Li, Y. Qian, and J. Z. Tsien, "Remote measurements of heart and respiration rates for telemedicine," *PloS One*, vol. 8, no. 10, p. e71384, 2013.

[32] Z. Guo, Z. J. Wang, and Z. Shen, "Physiological parameter monitoring of drivers based on video data and independent vector analysis," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2014, pp. 4374–4378.

[33] H. Qi, Z. Guo, X. Chen, Z. Shen, and Z. J. Wang, "Video-based human heart rate measurement using joint blind source separation," *Biomedical Signal Processing and Control*, vol. 31, pp. 309–320, 2017.

[34] A. Al-Naji, A. G. Perera, and J. Chahl, "Remote monitoring of cardiorespiratory signals from a hovering unmanned aerial vehicle," *Biomedical Engineering Online*, vol. 16, no. 1, p. 101, 2017.

[35] A. Al-Naji and J. Chahl, "Remote optical cardiopulmonary signal extraction with noise artifact removal, multiple subject detection & long-distance," *IEEE Access*, vol. 6, pp. 11 573–11 595, 2018.

[36] P. Viola and M. 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. CVPR 2001*, vol. 1. IEEE, 2001, pp. I511–I518.

[37] C. Tomasi and T. Kanade, "Detection and tracking of point features," Carnegie Mellon University, Tech. Rep., 1991.

[38] Z. Wu and N. E. Huang, "Ensemble empirical mode decomposition: a noise-assisted data analysis method," *Advances in Adaptive Data Analysis*, vol. 1, no. 01, pp. 1–41, 2009.

[39] N. E. Huang, Z. Shen, and S. R. Long, "A new view of nonlinear water waves: the Hilbert spectrum," *Annual Review of Fluid Mechanics*, vol. 31, no. 1, pp. 417–457, 1999.

[40] J. R. Kettenring, "Canonical analysis of several sets of variables," *Biometrika*, vol. 58, no. 3, pp. 433–451, 1971.

[41] H. Kobayashi, "Effect of measurement duration on accuracy of pulse-counting," *Ergonomics*, vol. 56, no. 12, pp. 1940–1944, 2013.

[42] G. Heusch, A. Anjos, and S. Marcel, "A reproducible study on remote heart rate measurement," *arXiv preprint arXiv:1709.00962*, 2017.

[43] R. Song, H. Chen, J. Cheng, C. Li, Y. Liu, and X. Chen, "PulseGAN: Learning to generate realistic pulse waveforms in remote photoplethysmography," *IEEE Journal of Biomedical and Health Informatics*, to be published.

[44] G. Balakrishnan, F. Durand, and J. Guttag, "Detecting pulse from head motions in video," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 3430–3437.

[45] D. McDuff and E. Blackford, "iphys: An open non-contact imaging-based physiological measurement toolbox," in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2019, pp. 6521–6524.

[46] R. Song, S. Zhang, C. Li, Y. Zhang, J. Cheng, and X. Chen, "Heart rate estimation from facial videos using a spatiotemporal representation with convolutional neural networks," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 10, pp. 7411–7421, 2020.

[47] R. Špetlík, V. Franc, and J. Matas, "Visual heart rate estimation with convolutional neural network," in *Proceedings of British Machine Vision Conference*, 2018, pp. 1–12.

[48] P. Gupta, B. Bhowmick, and A. Pal, "Mombat: Heart rate monitoring from face video using pulse modeling and bayesian tracking," *Computers in Biology and Medicine*, p. 103813, 2020.

[49] Y.-H. Wang, C.-H. Yeh, H.-W. V. Young, K. Hu, and M.-T. Lo, "On the computational complexity of the empirical mode decomposition algorithm," *Physica A: Statistical Mechanics and its Applications*, vol. 400, pp. 159–167, 2014.