Motion Robust Imaging Ballistocardiography Through a Two-Step Canonical Correlation Analysis

Rencheng Song, *Member, IEEE*, Jiji Li, Juan Cheng, *Member, IEEE*, Chang Li, *Member, IEEE*, Yu Liu, *Member, IEEE*, and Xun Chen, *Senior Member, IEEE*

***Abstract*— Imaging ballistocardiography (iBCG) is a video-based noncontact technique to detect heart rate (HR) from weak mechanical head movements caused by heart beating. However, rigid motions caused by voluntary movements and nonrigid motion resulted from facial expressions can easily distort the iBCG measurements. In this article, we propose a novel method, called robust iBCG (RiBCG), to suppress motion artifacts in iBCG with a two-step canonical correlation analysis (CCA). First, feature points are determined and tracked within two regions of interest (ROIs) from the face, where the vertical traces are taken as raw iBCG signals. Next, the first CCA is taken to separately remove the shared rigid motion artifacts between the horizontal and vertical traces in each ROI, where the obtained rigid-motion-free iBCG signals are further compressed by principal component analysis (PCA). Then, CCA is applied again to two sets of principal components to suppress nonrigid motion artifacts with low spatial correlations. Finally, the target HR value is determined as the one with the highest peak of power spectrums among all canonical variates (CVs). Besides, an improved version of RiBCG, termed RiBCG-C, is also proposed to reduce the HR outliers considering the continuity of HR variations. The proposed methods, as well as several other typical video-based HR measurement methods, are evaluated on two public databases, UBFC-RPPG and COHFACE, where the proposed RiBCG-C method achieves overall the best performance. The study provides a promising scheme for RiBCG measurements under realistic application scenarios.**

***Index Terms*— Canonical correlation analysis (CCA), heart rate (HR) estimation, imaging ballistocardiography (iBCG), motion artifact removal.**

I. Introduction

VIDEO-BASED heart rate (HR) measurement techniques attract extensive attention in recent years. Currently, there are two typical ways of measuring HR from facial videos including remote photoplethysmography (rPPG) [1], [2] and imaging ballistocardiography (iBCG) [3]. The rPPG measures HR using periodic facial color changes caused by blood volume variations, while the iBCG measures HR with rhythmic head motions due to cyclical movements of blood. The accuracy and stability of video-based HR measurements are mainly disrupted by ambient light changes and motion artifacts.

For rPPG, researchers have done excellent works to eliminate interferences of illumination variations and motion artifacts [1]. In terms of against ambient light changes, there are generally two types of schemes in rPPG. One scheme directly separates the pulse signal from ambient light changes using blind source separation (BSS) techniques [4] or other signal decomposition methods [5]. The other scheme takes the background region as the noise reference [6]. As for motion artifact elimination, there are also several types of methods. Since the motion is usually assumed to be linearly mixed with the target pulse signal, the BSS-based methods can be used to decouple the related sources. For example, Poh *et al.* [7] took independent component analysis (ICA) to obtain the pulse signal from RGB signals. Another way is to explicitly remove motion artifacts based on the optical reflection model [8]. For instance, de Haan and Jeanne [9] proposed a chrominance-based method (CHROM) to measure HR. Wang *et al.* [8] introduced the "plane-orthogonal-to-skin" (POS) method using a different projection from CHROM to remove the motion-induced noise.

Recently, inspired by the success of deep learning in computer vision areas, many learning-based rPPG methods have also been developed. These data-driven methods use the powerful mapping capabilities of neural networks, combined with a large amount of training data, to establish the complex mapping between inputs and targets, so as to achieve the purpose of denoising. The learning-based rPPG methods can be generally divided into two types according to the differences of inputs. The end-to-end rPPG methods directly build the mapping between the video frames and the target HR information. For example, Chen and Mcduff [10] proposed an end-to-end network by combining a convolutional neural network (CNN) with an attention mechanism, which can perform robust measurement under challenging environments. In contrast, the feature-decoder rPPG methods usually recover the HR information from well-designed feature maps using

Manuscript received August 15, 2020; revised October 11, 2020; accepted October 19, 2020. Date of publication October 30, 2020; date of current version December 28, 2020. This work was supported in part by the National Natural Science Foundation of China under Grant 61922075 and Grant 41901350, in part by the Provincial Natural Science Foundation of Anhui under Grant 2008085QF285, and in part by the Fundamental Research Funds for the Central Universities under Grant JZ2019HGBZ0151. The Associate Editor coordinating the review process was Lihui Peng. *(Corresponding authors: Juan Cheng; Xun Chen.)*

Rencheng Song, Jiji Li, Juan Cheng, Chang Li, and Yu Liu 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; chengjuan@hfut.edu.cn; changli@hfut.edu.cn; yuliu@hfut.edu.cn).

Xun Chen is with the Department of Neurosurgery, The First Affiliated Hospital of USTC, Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei 230001, China, and also with the Department of Electronic Engineering and Information Science, University of Science and Technology of China, Hefei 230026, China (e-mail: xunchen@ustc.edu.cn).

Digital Object Identifier 10.1109/TIM.2020.3034974

a decoder network. For instance, Niu *et al.* [11] extracted HR values from spatiotemporal representation images by a CNN. Song *et al.* [12] introduced a different way to construct spatiotemporal feature images considering the intrinsic characteristics of pulse signals. More developments of rPPG can be referred to the latest reviews, such as [1] and [13].

Unlike rPPG, the iBCG methods measure HR information from rhythmic head movements caused by heart beating. In 2013, Balakrishnan *et al.* [3] first proposed to measure HR values using BCG signals extracted from videos. This method tracks feature points from facial regions, where the vertical traces are processed to determine HRs. Compared with rPPG, the iBCG has the advantage of not necessarily using skin regions. On the other hand, the iBCG is not so sensitive to different types of skin colors or ambient light variations. The main challenge of iBCG is devoted to eliminating motion artifacts. This is attributed to the fact that the movements caused by periodic heart activity are quite weak, and they are easily dominated by voluntary head motions or facial expressions. Head motions can be considered as rigid motions that globally affect all feature points within different regions of interest (ROIs), while facial expressions are considered as nonrigid motions that only cause local interference to feature points within some of the ROIs. Existing iBCG studies to denoise motion artifacts are mainly divided into two categories. The first type is to select proper ROIs that are less affected by motion artifacts [14], [15]. This kind of method can only suppress local motion interference. The second type is to decouple the pulse signal from the raw vertical traces using complicated signal processing algorithms [16]. Although these methods have achieved good results, more explorations are still needed to deal with complex motion interferences in iBCG.

In this article, we propose a novel method, called robust iBCG (RiBCG), to remove both rigid and nonrigid motion artifacts in iBCG with a two-step canonical correlation analysis (CCA). The feature points are determined and tracked within two different ROIs from the face, where we derive the vertical traces of feature points as raw iBCG signals following the same way as existing works. It is observed that the vertical and horizontal traces are both affected by rigid motions and, thereby, have strong correlations. The CCA is a typical method to extract relevant components between the two data sets. Accordingly, the first CCA is applied to both the vertical and horizontal traces within each ROI to remove the shared rigid motion artifacts. The first few canonical variates (CVs) with higher correlation coefficients are set to zero to reconstruct the rigid-motion-free vertical traces. The principal component analysis (PCA) is then separately applied to the reconstructed vertical traces to derive the corresponding principal components (PCs) for dimension reduction. Since the motions caused by facial expressions have low spatial correlations, the second CCA is used to extract the common iBCG pulse signals from the two sets of PCs. Finally, HR is determined through a power spectrum analysis for all candidate CVs derived from the second CCA. In order to ensure the continuity of the estimated HRs, an improved version of RiBCG, termed RiBCG-C, is also proposed. Both the proposed methods are compared with several typical video-based HR measurement methods on two public databases, where our methods achieve superior performance, especially for the case with obvious motion artifacts.

In summary, the main contribution of this article is that we introduce a framework to effectively remove both rigid and nonrigid motion artifacts in iBCG. The proposed methods can achieve more reliable iBCG measurements compared with other conventional ones that are usually unable to work under significant motion artifact situations. Considering that iBCG does not necessarily require skin regions as rPPG, it is more robust against illumination variations, which can potentially extend the application scope of video-based remote HR measurements.

The rest of this article is prepared as follows. Section II introduces some related works of conventional iBCG methods. The proposed method is given in Section III. The experimental setup and comparison results are presented in Section IV. Finally, we conclude this article in Section V.

II. Related Works

As early as in 1877, Gordon [17] found that the blood pumping from the heart caused slight mechanical movements of the body. Under this principle, the BCG measurement was first introduced in the late 1930s [18]. Conventional BCG techniques usually measure vital movements of the human body with sensors, such as electromechanical films integrating with beds or chairs. The obtained BCG signals can be further employed in health monitoring, subject identification, and so on. For example, Rao *et al.* [19] introduced a hybrid deep neural network (DNN) to classify sleep stages using BCG signals. Cimr and Studnička [20] employed the BCG signals to detect disordered breathing with CNN. Zhang *et al.* [21] investigated the effect of BCG signals for subject identification using different structures of recurrent neural networks (RNNs). A brief review of conventional BCG methods with BCG sensors can be referred to [22] and [23].

Since first proposed by Balakrishnan *et al.* [3] in 2013, many studies of iBCG have been proposed to overcome the interference of motion artifacts. One common way is to select proper ROIs that are less disturbed by motion artifacts. Accordingly, the forehead and nose regions are usually chosen as the ROIs in existing iBCG techniques. For example, Shan and Yu [14] extracted the iBCG signal using traces of a single feature point from a small region of the forehead. Haque *et al.* [24] combined facial landmarks with the commonly used feature points to overcome the tracking difficulty under motion interference. Cai *et al.* [15] used feature points from the ear instead of the ones from the face, which completely avoided local motion artifacts due to facial expressions.

Besides choosing a proper ROI, there are also various algorithms proposed to decouple the pulse signal from motion-contaminated vertical traces. Hassan *et al.* [16] introduced a method to identify motion artifacts and prune them directly from the raw iBCG signals. The HR was then estimated by fitting all candidate HRs obtained from the denoised traces under a normal probability distribution. Wang *et al.* [25] took a three-layer temporal filtering on the vertical traces to eliminate

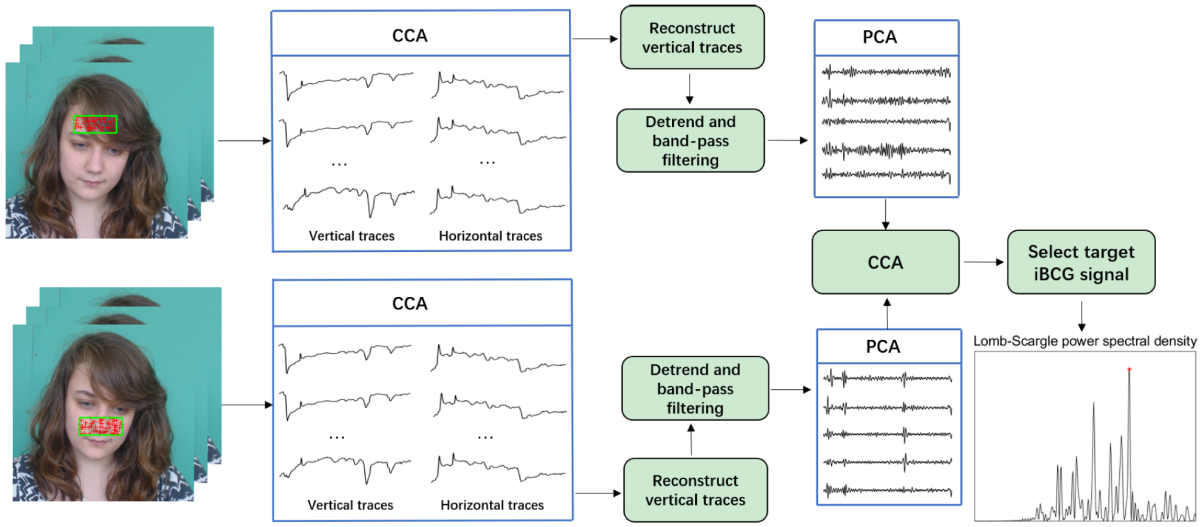

Fig. 1. Flowchart of the proposed two-step CCA method.

motion artifacts. The temporal filters included a limiting filter, a moving averaging, and a Butterworth bandpass filter, which were proved effective to suppress the voluntary head movements if the corresponding spectrum was not overlapped with that of the iBCG pulse signal.

According to the aforementioned studies [3], [16], [24], [25], the ROIs are usually determined to reduce the local influence of facial expression, while filtering or cropping of raw iBCG signals is taken to remove the global rigid motion artifacts. In this article, we will introduce a novel method to remove both rigid and nonrigid motion artifacts under a two-step CCA framework.

III. Methods

The flowchart of the proposed method is illustrated in Fig. 1. First, two ROIs are located, corresponding to the forehead and nose regions, respectively. A two-step CCA is then employed to remove both local and global motion artifacts. Finally, the iBCG pulse signal is selected among all CVs, and the HR value is estimated by a power spectral density (PSD) analysis.

Before introducing the detailed framework, the following notations are listed to avoid ambiguity. The scalar values, vectors, and matrices are denoted by lowercase italic letters, lowercase boldface letters, and capital boldface letters, respectively.

A. ROI Selection and Feature Point Tracking

The traces of feature points are essential to determine the iBCG signals. Fig. 2 shows the location of the two determined ROIs. A bounding box, locating the facial region, is first derived by the well-known Viola–Jones face detector [26]. Based on the generated bounding box, two ROIs corresponding to the forehead and the nose region are defined. Next, the feature points within each ROI are determined by the method of good features to track [27] and then tracked by the Kanade–Lucas–Tomasi (KLT) algorithm [28]. Both the vertical and horizontal raw traces can be generated. To avoid tracking failures, we only retain the most stable feature points as in [3].

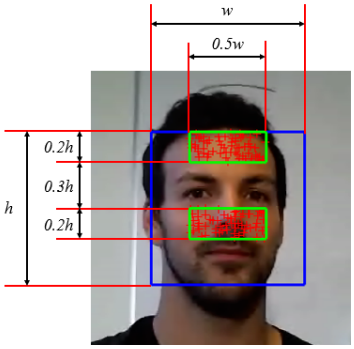

Fig. 2. Definition of two separate ROIs: an example from the COHFACE database.

Suppose that $\mathbf{X}(t) = [\mathbf{x}_1(t), \mathbf{x}_2(t), \ldots, \mathbf{x}_n(t), \ldots, \mathbf{x}_N(t)]^T$, and $t = 1, 2, \ldots, K$ is the data matrix composed by the horizontal traces of feature points, where K is the total number of frames, T is the transpose, and N is the total number of feature points. Each column of $\mathbf{X}(t)$ indicates a trace vector of a single point, and there are total N columns in $\mathbf{X}(t)$. Similarly, $\mathbf{Y}(t) = [\mathbf{y}_1(t), \mathbf{y}_2(t), \ldots, \mathbf{y}_n(t), \ldots, \mathbf{y}_N(t)]^T$, and $t = 1, 2, \ldots, K$ represents the data matrix composed by the vertical traces of feature points. The vertical traces are also considered as raw iBCG signals that are contaminated with local or global motion artifacts.

B. Rigid Motion Artifact Removal

Rigid motions have a global impact on all feature points. Voluntary head movements usually lead to a visible change to all feature points. Fig. 3 shows the correlations of vertical and horizontal traces under situations with some typical rigid head motions. It can be observed from Fig. 3 that the horizontal and vertical traces are highly correlated when the head is voluntarily moved, such as rotating, swing, and nodding.

Under a BSS framework, the $\mathbf{X}(t)$ and $\mathbf{Y}(t)$ trace signals are assumed to be linear mixtures of source signals as follows:

$$\mathbf{X}(t) = \mathbf{A}_X \mathbf{S}_X(t), \mathbf{Y}(t) = \mathbf{A}_Y \mathbf{S}_Y(t) \quad (1)$$

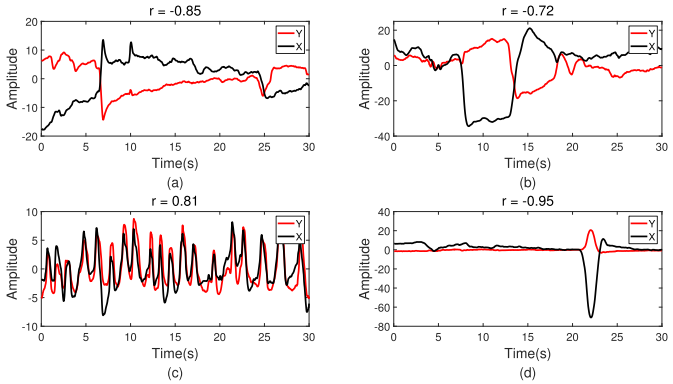

Fig. 3. Correlation of horizontal and vertical traces under rigid head motion situations. (a) Rotating. (b) Swing. (c) Nodding. (d) Turning.

where $\mathbf{S}_X(t) = [\mathbf{s}_X^1(t), \mathbf{s}_X^2(t), \ldots, \mathbf{s}_X^N(t)]$ and $\mathbf{S}_Y(t) = [\mathbf{s}_Y^1(t), \mathbf{s}_Y^2(t), \ldots, \mathbf{s}_Y^N(t)], t = 1, 2, \ldots, K$, are the unknown source signals, respectively. $\mathbf{A}_X$ and $\mathbf{A}_Y$ are the mixing matrices, respectively.

The unknown source signals can also be represented in demixing forms as follows:

$$\widetilde{\mathbf{S}}_X(t) = \mathbf{W}_X\mathbf{X}(t), \widetilde{\mathbf{S}}_Y(t) = \mathbf{W}_Y\mathbf{Y}(t) \tag{2}$$

where $\mathbf{W}_X$ and $\mathbf{W}_Y$ are the corresponding demixing matrices, and $\widetilde{\mathbf{S}}_X(t)$ and $\widetilde{\mathbf{S}}_Y(t)$ are the recovered sources, respectively.

We aim to identify the rigid-motion related sources and reconstruct vertical traces without the influence of these sources. Since $\mathbf{X}(t)$ and $\mathbf{Y}(t)$ are highly correlated, we take a CCA to remove rigid motion artifacts in vertical traces regarding the horizontal ones.

Suppose that $\mathbf{u} = \mathbf{w}_1^T\mathbf{X}(t)$ and $\mathbf{v} = \mathbf{w}_2^T\mathbf{Y}(t)$ are the CVs, where $\mathbf{w}_1$ and $\mathbf{w}_2$ are the weight vectors, respectively. The objective function of CCA is defined as

$$\max_{\mathbf{w}_1,\mathbf{w}_2} r(\mathbf{u}, \mathbf{v}) = \frac{\mathbf{w}_1^T\boldsymbol{\Sigma}_{12}\mathbf{w}_2}{\sqrt{\mathbf{w}_1^T\Sigma_{11}\mathbf{w}_1}\sqrt{\mathbf{w}_2^T\Sigma_{22}\mathbf{w}_2}} \tag{3}$$

where r indicates the correlation coefficient, $\boldsymbol{\Sigma}_{11}$ and $\boldsymbol{\Sigma}_{22}$ are the autocovariance matrices of $\mathbf{X}(t)$ and $\mathbf{Y}(t)$, respectively, and $\boldsymbol{\Sigma}_{12}$ is cross-covariance matrix of $\mathbf{X}(t)$ and $\mathbf{Y}(t)$.

The CCA problem in (3) can be solved through a method of Lagrange multiplier. We obtain a set of CVs, $\{\mathbf{u}_j\}_{j=1}^N$ and $\{\mathbf{v}_j\}_{j=1}^N$, with correlation coefficients in a decreasing order. These CVs are corresponding to the correlated sources in $\widetilde{\mathbf{S}}_X(t)$ and $\widetilde{\mathbf{S}}_Y(t)$.

The first few CVs with higher correlations are considered to be related to rigid motion artifacts, and the target iBCG signals are contained in the remaining CVs with lower correlations. This is because the rigid motion has global impacts on all feature points within different ROIs, and the energy of the rigid motion is much stronger than that of the target iBCG signal. Therefore, it is important to determine the order of noisy CVs and set them to zeros for reconstructing the rigid-motion-free vertical traces. To do this, we propose an L-curve method and choose the turning point to determine the truncation order of CVs. The full L-curve algorithm is listed in Algorithm 1.

In Fig. 4, we visually explain the L-curve algorithm using an example with clear voluntary head movements. The correlation

Algorithm 1 L-Curve Algorithm

1: Check whether rigid motion is contaminated in vertical traces of feature points. The average standard deviation (SD) δ of all vertical traces is calculated and compared with a predefined threshold μ. If $\delta > \mu$, then continue with step 2. Otherwise, it indicates there is no rigid motion.
2: Set $\{\mathbf{v}_1(t), \mathbf{v}_2(t), \ldots, \mathbf{v}_k(t)\}$ to zero for $k = 1, 2, \ldots, N-1$, respectively, and reconstruct vertical traces to obtain $\mathbf{Y}^k(t) = \{\mathbf{y}_1^k(t), \mathbf{y}_2^k(t), \ldots, \mathbf{y}_N^k(t)\}^T$.
3: Compute the correlation coefficients $r(\mathbf{x}_1(t), \mathbf{y}_1^k(t))$ for $k = 1, 2, \ldots, N-1$, and obtain a correlation coefficient vector $\mathbf{r} = [r_1, r_2, \ldots, r_k, \ldots, r_{N-1}]$.
4: Treat $\mathbf{r}$ as a curve and get the turning point to determine the order N_{tp} for the CVs to be set as zero.

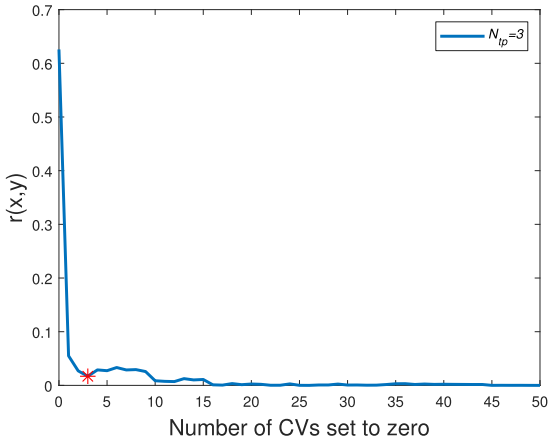

Fig. 4. L-curve plotting of correlation coefficients between the original horizontal trace and the reconstructed vertical trace with different orders of CV cancellation: an example from the first feature point in the forehead ROI.

coefficients between the reconstructed vertical and the original horizontal traces are plotted using the first feature point from a forehead ROI. The reason why we only choose the first feature point is that the rigid motions globally affect all feature points. It can be seen from Fig. 4 that the correlation coefficient can reach about 0.6 without the noisy CV cancellation. However, when the first two pairs of CVs are set to zeros, the correlation coefficient drops below 0.1. It indicates that the correlations will reduce if the rigid-motion-related CVs are sequentially removed from the reconstructed vertical trace until it reaches a balance. For this example, the turning point is determined as the red star marked in Fig. 4, which has an order of $N_{tp} = 3$. It means that the first three orders of CVs will be set to zeros to eliminate the rigid motion artifacts, and $\mathbf{Y}^3(t)$ is the reconstructed data matrix that has removed most of the interference from the rigid motion artifacts. It should be noted that the generation of L-curve only needs a single CCA calculation, and it is fast enough for real-time applications. By the way, if there is no rigid motion detected in algorithm 1, then the first CCA can be omitted, which means that the raw vertical traces will directly serve as the input of the next step.

C. Local Motion Artifact Removal

After separately removing rigid motion artifacts from all the vertical traces in each ROI, the two sets of reconstructed

vertical traces are assumed to share the same iBCG signals but different local motion artifacts due to facial expressions or speaking. This is consistent with the intrinsic correlation properties of the pulse information derived from different ROIs. Hence, the second CCA is utilized to extract the common iBCG pulse signals from the two sets of rigid-motion-free vertical traces. Before performing the second CCA, the reconstructed vertical traces are detrended, bandpass filtered, and compressed. The detailed process is given as follows.

1) First, the signal quality of the rigid-motion-free vertical traces is further improved through detrending and bandpass filtering. Especially, the detrending filter in [29] is used with the smoothness parameter as 20.0. The third-order Butterworth bandpass filter is then applied with a frequency interval of [0.75 3.0] Hz to remove the noise components beyond normal HR spectrums.
2) Next, the PCA is applied separately to the filtered vertical traces in each ROI, with the aim of reducing the dimension. The first five PCs are empirically selected as the input for the second CCA.
3) Finally, the second CCA is applied to the two sets of PCs to find the target iBCG pulse signal.

D. Pulse Selection and HR Estimation

In order to determine the target iBCG signal, we derive the spectrums of all candidate CVs after the second CCA by the Lomb–Scargle PSD [30]. The target iBCG signal is then selected as the one with the highest normalized power peak. Finally, the HR value is estimated as $f \times 60$, where f is the dominate frequency in PSD spectrums of the target iBCG signal.

During the experiment, we observe that the selected iBCG signals occasionally (about 5% percentages) correspond to HR outliers. Due to the fact that HR values are continuously changing during a short period of time, thereby, we introduce an improved version of RiBCG, termed RiBCG-C, to further improve the performance of HR measurements. Fig. 5 illustrates the flowchart of RiBCG-C.

As shown in Fig. 5, we first need to identify the HR outliers for each video. The HR vector obtained in a single video is denoted as **h**. The SD of **h** is then used as an indicator to determine whether there is an outlier or not. We compare the SD of **h** with a predefined threshold S_{Th}. If it is higher than S_{Th}, we think that there is an outlier in the estimated HRs as the one with the largest difference from the mean value of **h**. It should be noted that there may be more than one outliers that are wrongly selected. Thus, after finding the first outlier, we remove the outlier from **h** and compute the SD of remaining HRs again until it is less than S_{Th}. The remaining HRs in **h** are considered to be normal. The indices of all outliers are recorded in a vector of $\mathrm{IDX}_{\mathrm{abnor}}$. The next step is to correct the outlier HRs. For the index j in $\mathrm{IDX}_{\mathrm{abnor}}$, we use the remaining HRs in **h** to correct it. In detail, we find two nearest windows of j in **h** and use the average value $\mathrm{HR}_{\mathrm{ref}}$ of the corresponding HRs as a reference. The HR in the jth window is then reselected from all CVs of the same window in the second CCA. Particularly, we choose the one that is closest to $\mathrm{HR}_{\mathrm{ref}}$ to correct the current HR outlier.

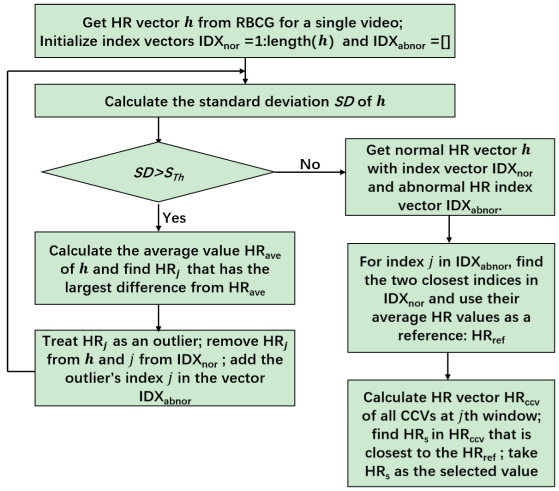

Fig. 5. Flowchart of the RiBCG-C method.

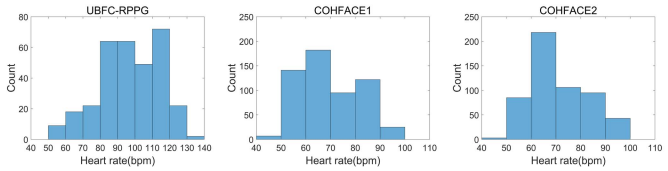

Fig. 6. HR distributions of PPG signals in UBFC-RPPG, COHFACE1, and COHFACE2, respectively.

IV. Experiments and Results

In this section, we will evaluate the performance of the proposed methods on two public databases: UBFC-RPPG [31] and COHFACE [32]. The UBFC-RPPG database contains dominant voluntary motion artifacts, and hence, it is very challenging for iBCG measurements. The COHFACE database contains different illumination conditions and is suitable for validating the robustness of the proposed RiBCG and RiBCG-C methods against illumination variations. The performance will be compared with that of several typical video-based HR measurement methods on the aforementioned two databases.

A. Experimental Settings

1) UBFC-RPPG Database: The UBFC-RPPG database [31] contains 42 videos from 42 subjects. These videos last for about 1 min and were recorded by Logitech C920 HD pro RGB camera using an uncompressed 8-bit format. The resolution of each video is 640 × 480, and the frame rate is 30 frames per second (fps). The reference PPG signals were recorded by Contec Medical CMS50E at a sampling rate of 60 Hz. The HR values of each subject in the experiment were remarkably changing since the subjects were asked to play a time-sensitive mathematical game in front of the computer. The HR distribution is illustrated in Fig. 6, where the HR values in UBFC-RPPG vary from 50 to 140 beats per minute (bpm). Meanwhile, many subjects have obvious voluntary head movements, which makes it difficult to acquire reliable iBCG measurements.

2) COHFACE Database: The COHFACE database [32] contains 160 videos from 40 subjects. The videos were

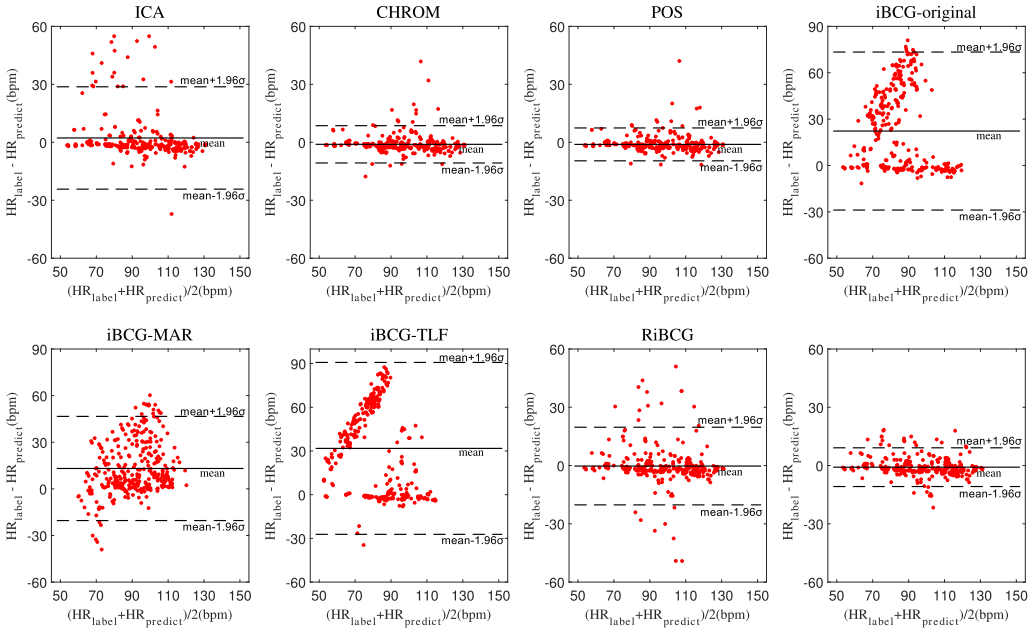

Fig. 7. Bland–Altman plots of HR on the UBFC-RPPG database.

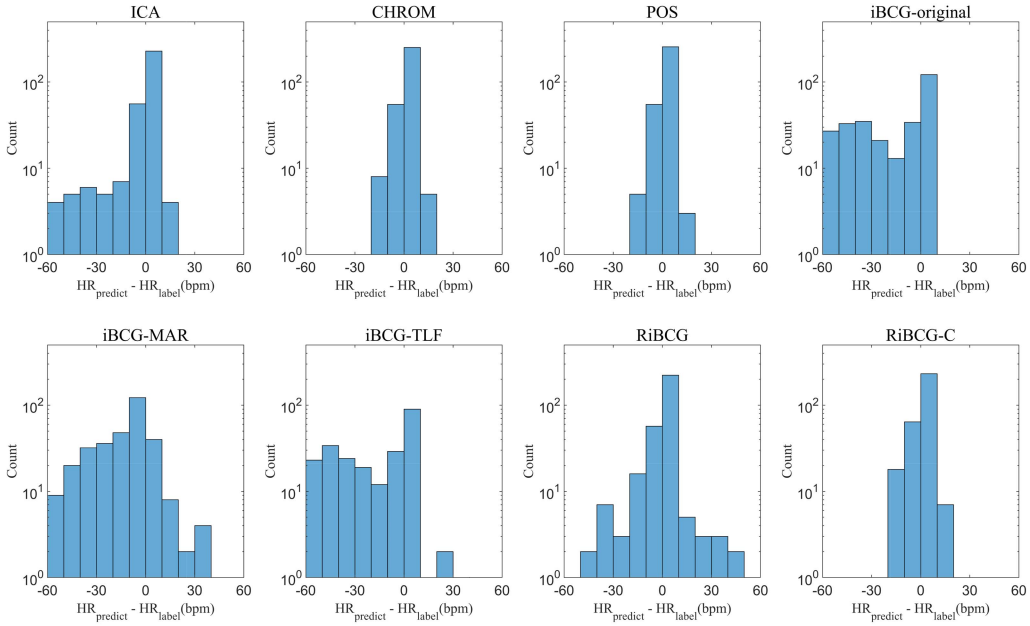

Fig. 8. Statistical histograms of HR errors on the UBFC-RPPG database.

recorded by the Logitech HD C525 Webcam with a resolution of 640 × 480 and a frame rate of 20 fps under realistic conditions. The reference pulse signals have been acquired using a blood-volume pulse (BVP) sensor from Thought Technologies. Each subject had four videos: two videos were recorded under controlled illumination conditions, and the other two were taken under natural illumination conditions. Accordingly, we divide the COHFACE database into two parts: the so-called COHFACE1 with controlled illuminations and the COHFACE2 with natural illuminations. It should be noted that the two ROIs used in COHFACE2 are both defined as the left half of those in Fig. 2 to remove the totally dark regions. This can effectively reduce the failure of tracking feature points. It can be found from Fig. 6 that the distributions of HR values in COHFACE1 and COHFACE2 both change from 45 to 100 bpm.

For each video in the two databases, the processing window was set as 30 s with a 5-s overlapping between the neighboring windows. In the experiment, the threshold μ in the L-curve algorithm was set as 2.5 to determine whether there was a rigid motion in each processing window. The threshold S_{Th} of the RiBCG-C algorithm in Fig. 5 was set as 10.0 and 6.0 for UBFC-RPPG and COHFACE databases, respectively. The threshold S_{Th} for the UBFC-RPPG was higher than that for the COHFACE database due to the fact that the variation of HR values in UBFC-RPPG was remarkably stronger.

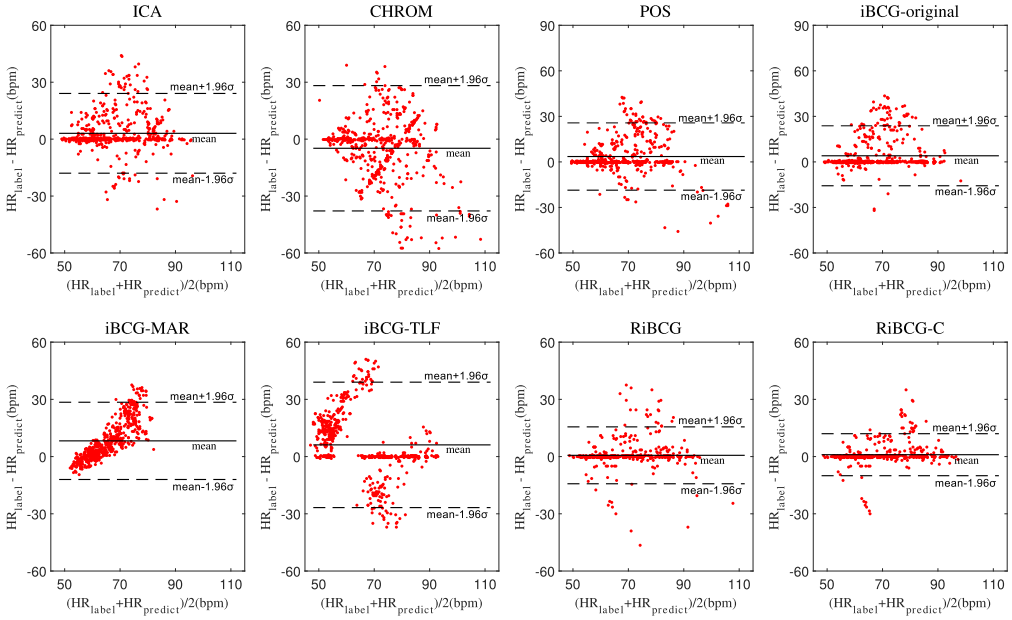

Fig. 9. Bland–Altman plot of HR measurement on the COHFACE database with controlled illuminations: COHFACE1.

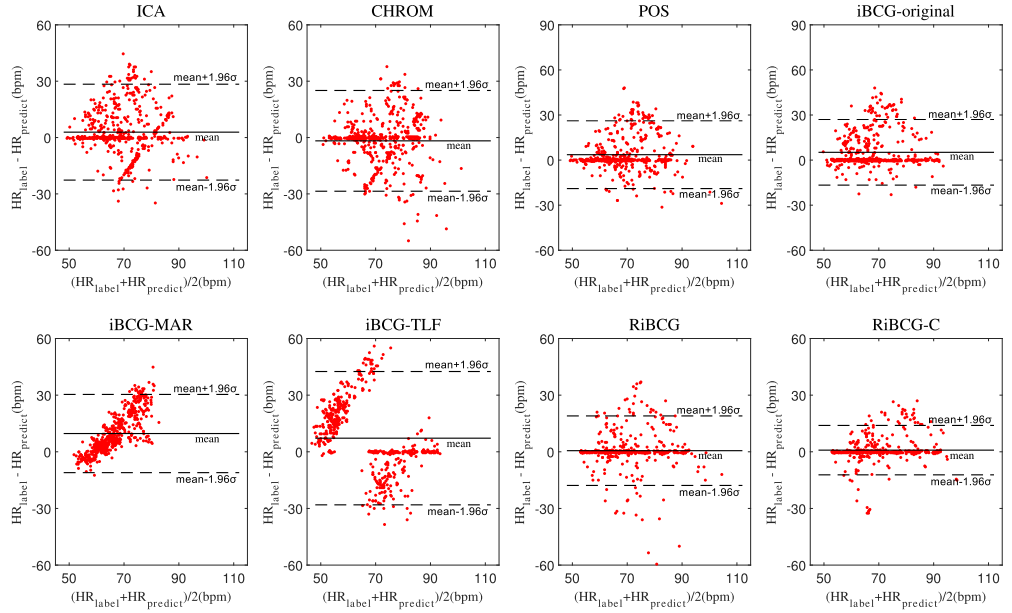

Fig. 10. Bland–Altman plot of HR measurement on the COHFACE database with natural illuminations: COHFACE2.

B. HR Estimation Results

We compare the proposed methods with some other methods, including ICA [7], CHROM [9], POS [8], the original iBCG [3] (iBCG-original), the iBCG with motion artifact removal [16] (iBCG-MAR), and the iBCG with three-layer filter [25] (iBCG-TLF). The implementation of the first four comparison methods has been referred to a public toolbox "iPhys" [33]. All these methods employed the same ROIs as defined in Fig. 2. Several metrics are used to evaluate the performance, including the root mean square error HR_{rmse}, the mean absolute error HR_{mae}, SD HR_{sd}, and Pearson's correlation coefficient r. The detailed definitions of these metrics can be found in [12].

Table I shows the averaged HR results of different methods on the UBFC-RPPG database. Since the POS and CHROM methods are designed to suppress motion artifacts following the skin optical reflection model, these two methods achieve outstanding results on this database with correlation coefficients r as 0.96 and 0.97, respectively. In comparison, the results of all the three existing iBCG methods are not satisfactory. Particularly, the iBCG-MRA achieved better performance than the other two, where HR_{rmse} is 21.56 bpm, the HR_{mae} is 15.72 bpm, and the correlation coefficient r is 0.44. The reason is that rigid motion artifacts seriously affect the performance of those iBCG methods. In contrast, the proposed RiBCG method achieves a much better performance than that of iBCG, where HR_{rmse} is 10.22 bpm, HR_{mae} is 5.43 bpm, and the correlation coefficient r is 0.85. However, the RiBCG has no clear advantage over the two model-based rPPG methods on the UBFC-RPPG database. To further improve the results

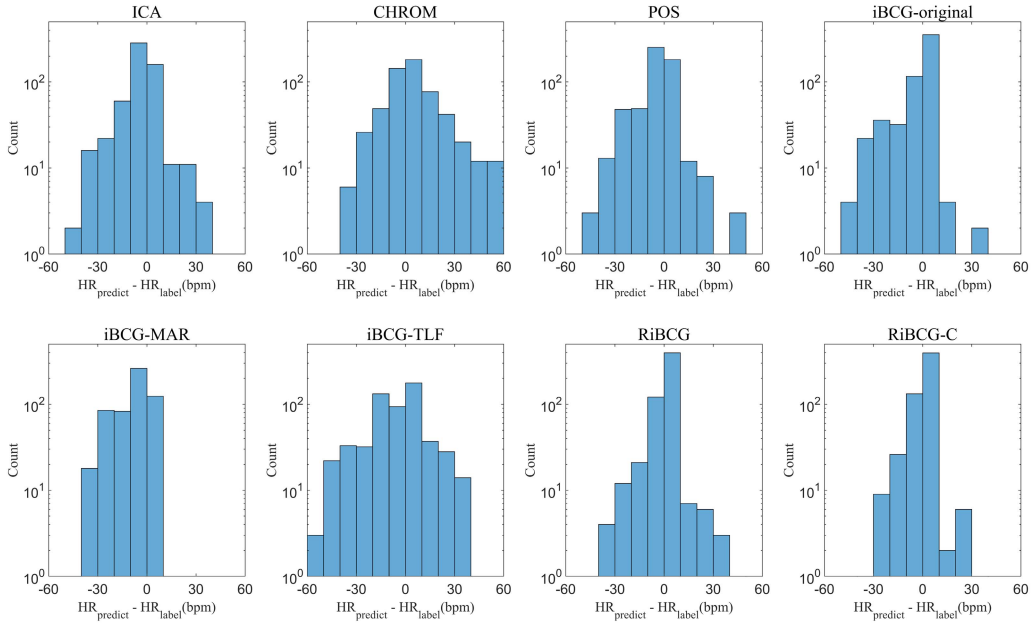

Fig. 11. Statistical histograms of HR errors on the COHFACE database with controlled illuminations: COHFACE1.

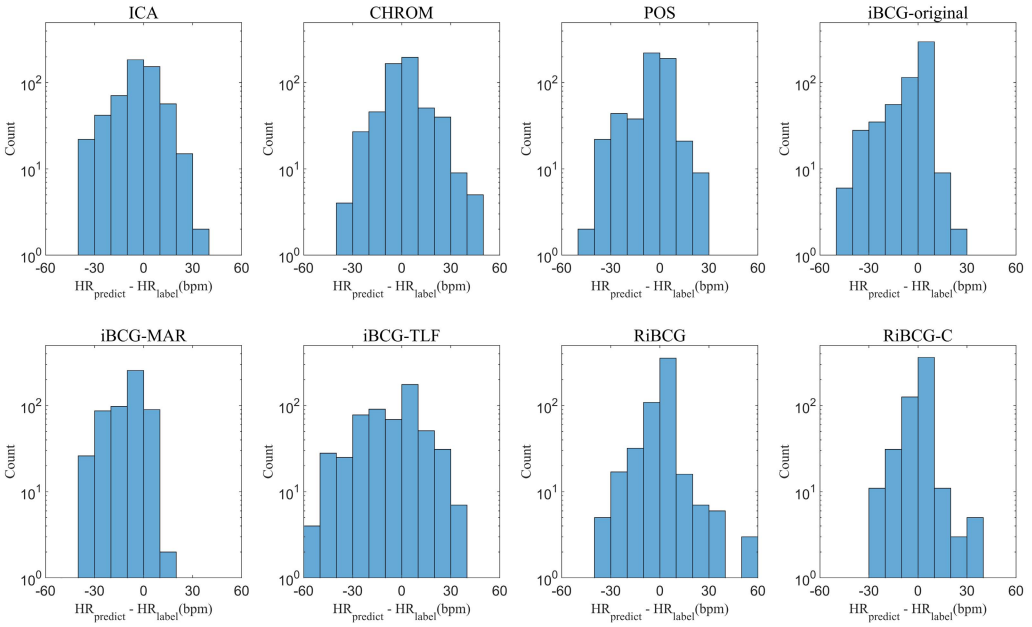

Fig. 12. Statistical histograms of HR errors on the COHFACE database with natural illuminations: COHFACE2.

using the continuity of HR, the RiBCG-C achieves the results of HR_{rmse} as 4.95 bpm, HR_{mae} as 3.52 bpm, and r as 0.96. The performance of RiBCG-C is on the same level as that of the CHROM and POS methods on the UBFC-RPPG database, which proves the effectiveness of the proposed method in the presence of motion artifacts.

To further demonstrate the performance for individual samples, the Bland–Altman plots and the error distribution histograms of all methods are illustrated in Figs. 7 and 8, respectively. It can be observed from Fig. 7 that the results of RiBCG are much more consistent with the reference ones compared with those of the existing iBCG methods. However, there are still some obvious HR outliers using RiBCG, which indicates the reason why RiBCG achieves slightly worse results than those of the two model-based rPPG methods. Nevertheless, the RiBCG-C, considering the continuity of HR during a short period of time, remarkably reduces the HR outliers and, hence, achieves comparable results compared with those of the comparison methods. The error distributions in Fig. 8 demonstrate a similar trend as the Bland–Altman plots. Thus, we will not go into too much detail.

Next, we compare the performance of all methods on the COHFACE database with different illuminations. Table II shows the averaged HR results of the COHFACE1 database

TABLE I
AVERAGED HR RESULTS OF DIFFERENT METHODS ON THE UBFC-RPPG DATABASE

Method	HR_{rmse} (bpm)	HR_{mae} (bpm)	HR_{sd} (bpm)	r
ICA [7]	13.60	6.23	12.09	0.76
CHROM [9]	5.17	3.20	4.06	0.96
POS [8]	4.61	3.00	3.50	0.97
iBCG-oringinal [3]	34.39	24.24	24.28	0.15
iBCG-MRA [16]	21.56	15.72	14.75	0.44
iBCG-TLF [25]	43.87	33.70	28.09	-0.08
RiBCG	10.22	5.43	8.60	0.85
RiBCG-C	4.95	3.52	3.48	0.96

TABLE II
AVERAGED HR RESULTS OF DIFFERENT METHODS IN COHFACE1 DATABASE WITH CONTROLLED ILLUMINATIONS

Method	HR_{rmse} (bpm)	HR_{mae} (bpm)	HR_{sd} (bpm)	r
ICA [7]	11.14	6.01	9.38	0.58
CHROM [9]	17.51	11.64	13.08	0.21
POS [8]	11.84	6.38	9.97	0.53
iBCG-original [3]	10.84	4.90	9.67	0.60
iBCG-MRA [16]	13.02	9.37	9.31	0.53
iBCG-TLF [25]	17.85	12.40	12.84	0.40
RiBCG	7.63	2.81	7.09	0.79
RiBCG-C	5.70	2.22	5.25	0.88

TABLE III
AVERAGED HR RESULTS OF DIFFERENT METHODS IN COHFACE2 DATABASE WITH NATURAL ILLUMINATIONS

Method	HR_{rmse} (bpm)	HR_{mae} (bpm)	HR_{sd} (bpm)	r
ICA [7]	13.32	8.91	9.91	0.38
CHROM [9]	13.78	8.75	10.65	0.34
POS [8]	12.04	6.79	9.94	0.45
iBCG-original [3]	12.27	6.46	10.43	0.51
iBCG-MRA [16]	14.31	10.71	9.50	0.43
iBCG-TLF [25]	19.39	14.16	13.24	0.36
RiBCG	9.40	4.18	8.41	0.67
RiBCG-C	6.73	2.99	6.02	0.82

with controlled illuminations. The model-based CHROM and POS methods do not achieve satisfactory results as they have done on the UBFC-RPPG database. The reason may be that the videos in the COHFACE database were recorded in a compressed format, which may break the assumption of the optical reflection model. We observe that the three existing iBCG methods achieve much better results on this database than the UBFC-RPPG database. This is as expected because the subjects in COHFACE were asked to stay still, and visually, there are very few voluntary head movements. The proposed RiBCG method achieves the best results compared with those of the other methods, where the $\mathrm{HR}_{\mathrm{rmse}}$ is 7.63 bpm, $\mathrm{HR}_{\mathrm{mae}}$ is 2.81 bpm, and r is 0.79, which suggests that although the subjects in COHFACE database are relatively more stationary than those in UBFC-RPPG database, the slight movements still can easily interfere the iBCG measurement since the latter one is really too weak. As expected, the RiBCG-C further improves the performance of RiBCG. The correlation coefficient has been improved by 11.39%, which is increased from 0.79 to 0.88.

Similar results are shown in Table III on the COHFACE2 database with natural illuminations. The proposed RiBCG method gets the $\mathrm{HR}_{\mathrm{rmse}}$ as 9.40 bpm, $\mathrm{HR}_{\mathrm{mae}}$ as 4.18 bpm, and the correlation coefficient r as 0.82. In contrast, we observe that the performance of RiBCG is slightly degraded on COHFACE2 compared with that of COHFACE1. This indicates that the illumination condition also affects the accuracy of iBCG methods. It may increase the difficulty of tracking feature points. The performance degradation due to the change of illumination conditions can also be observed from the Bland–Altman plots in Figs. 9 and 10, respectively. We find that the results of RiBCG have more outliers on COHFACE2 than COHFACE1, where 5%–6% samples have relatively large errors. Consistent with the results on the UBFC-RPPG database, the RiBCG-C method further improves the results of RiBCG and gets overall the best performance. This is also demonstrated by the error distribution histograms in Figs. 11 and 12, where we can find that the errors of RiBCG-C are more concentrated than those of the RiBCG.

Besides the comparison methods listed above, some other latest results of video-based HR measurements on the COHFACE database are also listed here for reference. Gupta *et al.* [34] proposed an rPPG method, termed "MOMBAT," to monitor HR using pulse modeling and Bayesian tracking. The MOMBAT method achieved the performance of $\mathrm{HR}_{\mathrm{mae}}$ as 5.89 bpm, HR_{sd} as 7.38 bpm, and Pearson's correlation coefficient r as 0.62 on the COHFACE database. Hernandez-Ortega *et al.* [35] took a comparison study of several rPPG methods on the COHFACE database, and they found that the learning-based methods outperformed the conventional handcrafted ones. Particularly, the "DeepPhys" method in [10] gets the best $\mathrm{HR}_{\mathrm{mae}}$ as 3.79 bpm. Tsou *et al.* [36] introduced a Siamese-rPPG network to measure the HR following an end-to-end way. The testing results on the COHFACE database are quite nice, where $\mathrm{HR}_{\mathrm{rmse}}$ is 1.29 bpm, $\mathrm{HR}_{\mathrm{mae}}$ is 0.70 bpm, and Pearson's correlation coefficient r is 0.73. However, it should be indicated that the above results of learning-based methods in [35] and [36] were all obtained using the training data from the same database. Therefore, the generalization ability of related learning models has not been fully tested yet on this database.

In summary, the proposed method achieves similar performance in the UBFC-RPPG database compared with the benchmark methods. This verifies the effectiveness of the RiBCG against motion artifacts. It also outperforms the comparison methods in the COHFACE database, which demonstrates the superior performance of iBCG-type methods over rPPG methods for videos with illumination variations.

V. CONCLUSION

In this article, we have introduced a novel iBCG method, termed RiBCG, to measure HR with a two-step CCA in the presence of both rigid and nonrigid motions. We determine feature points from two different ROIs from the face and get the horizontal and vertical traces through a tracking scheme, where the vertical ones are considered as raw iBCG signals. With the help of the first CCA, the rigid motions can be separately removed from the vertical traces of feature points in each ROI according to the high correlations with the horizontal traces. The obtained rigid-motion-free vertical traces are then

compressed by PCA to reduce the dimension. Next, the second CCA is utilized to extract the common target iBCG signals from two sets of obtained PCs based on low spatial correlations of nonrigid motions. Finally, the HR value is selected as the one with the highest peak of power spectrums among all CVs. The improved version of RiBCG, the so-called RiBCG-C, has also been proposed to remove the HR outliers according to the continuity of HR values during a short period of time. The experimental results on two public databases have demonstrated that the proposed methods have achieved significant improvements compared with three existing iBCG methods, where the RiBCG-C method has got overall the best results. The proposed methods have been proven to be effective for measuring HRs under realistic environments, and they are expected to expand the application scope of iBCG techniques.

REFERENCES

[1] 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 Trans. Instrum. Meas.*, vol. 68, no. 10, pp. 3600–3615, Oct. 2019.

[2] Y. Sun and N. Thakor, "Photoplethysmography revisited: From contact to noncontact, from point to imaging," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 3, pp. 463–477, Mar. 2016.

[3] G. Balakrishnan, F. Durand, and J. Guttag, "Detecting pulse from head motions in video," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 3430–3437.

[4] A. Lam and Y. Kuno, "Robust heart rate measurement from video using select random patches," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 3640–3648.

[5] D.-Y. Chen *et al.*, "Image sensor-based heart rate evaluation from face reflectance using Hilbert–Huang transform," *IEEE Sensors J.*, vol. 15, no. 1, pp. 618–627, Jan. 2015.

[6] J. Cheng, X. Chen, L. Xu, and Z. J. Wang, "Illumination variation-resistant video-based heart rate measurement using joint blind source separation and ensemble empirical mode decomposition," *IEEE J. Biomed. Health Informat.*, vol. 21, no. 5, pp. 1422–1433, Sep. 2017.

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

[8] W. Wang, A. C. den Brinker, S. Stuijk, and G. de Haan, "Algorithmic principles of remote PPG," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 7, pp. 1479–1491, Jul. 2017.

[9] G. de Haan and V. Jeanne, "Robust pulse rate from chrominance-based rPPG," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 10, pp. 2878–2886, Oct. 2013.

[10] W. Chen and D. McDuff, "Deepphys: Video-based physiological measurement using convolutional attention networks," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 349–365.

[11] X. Niu, H. Han, S. Shan, and X. Chen, "SynRhythm: Learning a deep heart rate estimator from general to specific," in *Proc. 24th Int. Conf. Pattern Recognit. (ICPR)*, Aug. 2018, pp. 3580–3585.

[12] 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 Trans. Instrum. Meas.*, vol. 69, no. 10, pp. 7411–7421, Oct. 2020.

[13] M. Leo, P. Carcagnì, P. L. Mazzeo, P. Spagnolo, D. Cazzato, and C. Distante, "Analysis of facial information for healthcare applications: A survey on computer vision-based approaches," *Information*, vol. 11, no. 3, p. 128, Feb. 2020.

[14] L. Shan and M. Yu, "Video-based heart rate measurement using head motion tracking and ICA," in *Proc. 6th Int. Congr. Image Signal Process. (CISP)*, vol. 1, Dec. 2013, pp. 160–164.

[15] X. Cai, G. Han, and J. Wang, "Video-based noncontact heart rate measurement using ear features," in *Proc. IEEE Int. Conf. Prog. Informat. Comput. (PIC)*, Dec. 2015, pp. 262–265.

[16] M. A. Hassan, A. S. Malik, D. Fofi, N. M. Saad, Y. S. Ali, and F. Meriaudeau, "Video-based heartbeat rate measuring method using ballistocardiography," *IEEE Sensors J.*, vol. 17, no. 14, pp. 4544–4557, Jul. 2017.

[17] J. Gordon, "Certain molar movements of the human body produced by the circulation of the blood," *J. Anatomy Physiol.*, vol. 11, no. Pt 3, p. 533, 1877.

[18] I. Starr, A. J. Rawson, H. A. Schroeder, and N. R. Joseph, "Studies on the estimation of cardiac ouptut in man, and of abnormalities in cardiac function, from the heart's recoil and the blood's impacts; the ballistocardiogram," *Amer. J. Physiol.-Legacy Content*, vol. 127, no. 1, pp. 1–28, Jul. 1939.

[19] S. Rao, A. E. Ali, and P. Cesar, "DeepSleep: A ballistocardiographic deep learning approach for classifying sleep stages," in *Proc. ACM Int. Joint Conf. Pervas. Ubiquitous Comput. Proc. ACM Int. Symp. Wearable Comput. UbiComp/ISWC*, 2019, pp. 187–190.

[20] D. Cimr and F. Studnička, "Automatic detection of breathing disorder from ballistocardiography signals," *Knowl.-Based Syst.*, vol. 188, Jan. 2020, Art. no. 104973.

[21] X. Zhang, Y. Zhang, L. Zhang, H. Wang, and J. Tang, "Ballistocardiogram based person identification and authentication using recurrent neural networks," in *Proc. 11th Int. Congr. Image Signal Process., Biomed. Eng. Informat. (CISP-BMEI)*, Oct. 2018, pp. 1–5.

[22] E. Vogt, D. MacQuarrie, and J. P. Neary, "Using ballistocardiography to measure cardiac performance: A brief review of its history and future significance," *Clin. Physiol. Funct. Imag.*, vol. 32, no. 6, pp. 415–420, Nov. 2012.

[23] I. Sadek, J. Biswas, and B. Abdulrazak, "Ballistocardiogram signal processing: A review," *Health Inf. Sci. Syst.*, vol. 7, no. 1, p. 10, Dec. 2019.

[24] M. A. Haque, R. Irani, K. Nasrollahi, and T. B. Moeslund, "Heartbeat rate measurement from facial video," *IEEE Intell. Syst.*, vol. 31, no. 3, pp. 40–48, May 2016.

[25] L. Wang, S. Geng, B. Liu, and Y. Jin, "Ballistocardiogram heart rate detection: Improved methodology based on a three-layer filter," *Measurement*, vol. 149, Jan. 2020, Art. no. 106956.

[26] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Dec. 2001, p. 1.

[27] J. Shi and Tomasi, "Good features to track," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 1994, pp. 593–600.

[28] C. Tomasi and T. Kanade, "Detection and tracking of point features," School Comput. Sci., Carnegie Mellon Univ., Pittsburgh, PA, USA, Tech. Rep. CMU-CS-91-132, 1991.

[29] M. P. Tarvainen, P. O. Ranta-aho, and P. A. Karjalainen, "An advanced detrending method with application to HRV analysis," *IEEE Trans. Biomed. Eng.*, vol. 49, no. 2, pp. 172–175, 2002.

[30] W. H. Press and G. B. Rybicki, "Fast algorithm for spectral analysis of unevenly sampled data," *The Astrophysical J.*, vol. 338, pp. 277–280, 1989.

[31] S. Bobbia, R. Macwan, Y. Benezeth, A. Mansouri, and J. Dubois, "Unsupervised skin tissue segmentation for remote photoplethysmography," *Pattern Recognit. Lett.*, vol. 124, pp. 82–90, Jun. 2019.

[32] G. Heusch, A. Anjos, and S. Marcel, "A reproducible study on remote heart rate measurement," 2017, *arXiv:1709.00962*. [Online]. Available: http://arxiv.org/abs/1709.00962

[33] D. McDuff and E. Blackford, "IPhys: An open non-contact imaging-based physiological measurement toolbox," in *Proc. 41st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2019, pp. 6521–6524.

[34] P. Gupta, B. Bhowmick, and A. Pal, "MOMBAT: Heart rate monitoring from face video using pulse modeling and Bayesian tracking," *Comput. Biol. Med.*, vol. 121, Jun. 2020, Art. no. 103813.

[35] J. Hernandez-Ortega, J. Fierrez, A. Morales, and D. Diaz, "A comparative evaluation of heart rate estimation methods using face videos," in *Proc. IEEE 44th Annu. Comput., Softw., Appl. Conf. (COMPSAC)*, Jul. 2020, pp. 1438–1443.

[36] Y.-Y. Tsou, Y.-A. Lee, C.-T. Hsu, and S.-H. Chang, "Siamese-rPPG network: Remote photoplethysmography signal estimation from face videos," in *Proc. 35th Annu. ACM Symp. Appl. Comput.*, Mar. 2020, pp. 2066–2073.