This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2019.2912708, IEEE Journal of Biomedical and Health Informatics

Robust Heart Rate Monitoring for Quasi-periodic Motions by Wrist-Type PPG Signals

Wenwen He, *Student Member, IEEE,* Yalan Ye*, *Member, IEEE,* Li Lu, *Member, IEEE,* Yunfei Cheng, *Student Member, IEEE,* Yunxia Li, and Zhengning Wang

***Abstract*—Heart rate (HR) monitoring using photoplethysmography (PPG) is a promising feature in modern wearable devices. PPG is easily contaminated by motion artifacts (MA), hindering estimation of HR. For quasi-periodic motions, previous works generally focused on a few specific motions, such as walking and fast running. However, they may not work well for many different quasi-periodic motions where MA are very complex. In this paper, a robust HR monitoring scheme for different quasi-periodic motions using wrist-type PPG is proposed, which consists of dictionary learning for signal characteristics learning, human motion recognition for the current motion recognition and dictionary selection, sparse representation-based MA elimination for denoising, and spectral peak tracking for HR-related spectral peak tracking. The proposed scheme is robust to MA caused by different motions and has high accuracy. Experiments on six common quasi-periodic motions showed that the average absolute error of heart rate estimation was 2.40 beat per minute, and also showed that the proposed method is more robust than some state-of-the-art approaches for different motions.**

***Index Terms*—Heart rate monitoring, photoplethysmography (PPG), motion artifacts, sparse representation, wearable sensors.**

I. Introduction

HEART rate (HR) monitoring based on modern wearable devices is of interest due to its useful features in controlling training load or health monitoring during physical exercise. Photoplethysmography (PPG) signals [1] have shown its potential in HR monitoring during physical exercise because of its simpler hardware implementation and lower cost over traditional Electrocardiograph (ECG) signals [2]. However, PPG is susceptible to motion artifacts (MA) [3], making HR monitoring based on PPG a difficult problem.

Some methods, which were proposed for small motions, have been investigated. One is independent component analysis (ICA) [4], [5]. However, statistical independence (the key assumption in ICA) does not hold in PPG signals contaminated by MA [6], resulting in unsatisfactory performance of MA elimination. Another method is adaptive filtering (used for small motions) [7], [8]. But adaptive filtering is sensitive to reference signal. Since these techniques were proposed for some small motions, such as finger movements [5], walking [7] and slow running [8], where MA was not strong, these techniques may not work well when monitoring HR during physical exercise [3].

Other techniques, which were proposed for strong motions in physical exercise, were recently investigated, such as adaptive filtering (used for strong motions in physical exercise) [9], [10], [11], [12], empirical mode decomposition (EMD) [13], [14], singular spectrum analysis (SSA) [3], spectrum subtraction (SS) [15], [16], winer filtering [17] and particle filter [18]. Most of them utilized acceleration signals as the reference signals of MA, showing good performance in some scenarios. However, these techniques only focused on a few specific quasi-periodic motions, such as walking and fast running. For specific motions, these techniques can adjust their parameters to work well. However, the adjusted parameters may not work well for different types of intensive physical exercises in the real world including many different quasi-periodic motions, such as elliptical trainer, where the hand movements have different directions which may result in more complex MA.

In this paper, a robust HR monitoring scheme using wrist-type PPG during different types of quasi-periodic motions is proposed, being composed of four key parts: dictionary learning, human motion recognition, sparse representation-based MA elimination, and spectral peak tracking. Dictionary learning aims to obtain different PPG and MA dictionaries to represent PPG signals and different MAs caused by different quasi-periodic motions, where one kind of motion corresponds to one PPG dictionary and one MA dictionary. Human motion recognition aims to recognize the current motion and thereby select the corresponding PPG and MA dictionaries. Using the selected PPG dictionary and MA dictionary, sparse representation-based MA elimination aims to eliminate MA caused by the current motion. Using the cleansed PPG signal after sparse representation, spectral peak tracking aims to locate the HR-related spectral peak. Experiments on many common quasi-periodic motions showed that the proposed HR monitoring approach has satisfactory accuracy, and is robust to different MAs caused by different strong quasi-periodic motions. The detailed contributions of this paper are as follows:

- For intensive physical exercises in the real world, there are many different strong quasi-periodic motions, such as walking, fast running, beckoning, swing arm, elliptical trainer and deep keen bend, leading to very strong and

Wenwen He, Yalan Ye, Li Lu and Yunfei Cheng are with the School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China (e-mail: heww.mq@gmail.com; yalanye@uestc.edu.cn; luli2009@uestc.edu.cn; yunfeicheng@hotmail.com).

Yunxia Li is with the School of Automation Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China (e-mail: yunxiali@uestc.edu.cn).

Zhengning Wang is with the School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China (e-mail: zhengning.wang@uestc.edu.cn).

∗ Corresponding author: Yalan Ye (yalanye@uestc.edu.cn).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2019.2912708, IEEE Journal of Biomedical and Health Informatics

complex MA. To monitor HR during intensive physical exercises, it is necessary to design a robust HR monitoring technique for many different strong quasi-periodic motions. At present, for quasi-periodic motions, most HR monitoring techniques only considered a few specific motions, such as walking and fast running. However, they may not work well during many other different quasi-periodic motions which may cause very complex MA. To overcome the above issue, a robust HR monitoring technique for many different quasi-periodic motions using PPG is proposed in this paper, which is more suitable for practical application scenario.

- To eliminate different MAs caused by different strong quasi-periodic motions, we propose combining human motion recognition and dictionary learning-based sparse representation. Human motion recognition can recognize the motion that the current subject is performing, and then automatically select the dictionaries corresponding to the motion. Using the corresponding dictionaries, dictionary learning-based sparse representation can represent and separate PPG and MA caused by this motion, and then can obtain the clean PPG signal.
- The performance on different types of intensive physical exercises being composed of six typical strong quasi-periodic motions (walking, fast running, beckoning, swing arm, elliptical trainer and deep keen bend) related to hand movements were shown in our experiments, because the main source of MA is from hand movements [15]. Experiments on the six typical motions proved that the proposed approach can achieve an accurate estimate of HR, and it is also robust to different MAs caused by different quasi-periodic motions in different types of physical exercises, indicating that the proposed approach has potential to be used in HR computation of wearable sensors during physical exercises.

II. Motivation

PPG recorded during physical exercise is easily contaminated by MA. One major source of MA is the voluntary or involuntary subject movement which can make the gap (between the sensor and the skin) easily enlarged by hand movements during physical exercises [15]. Removal of MA can not be easily performed because of the likely spectral and temporal similarity between PPG and MA [19]. Acceleration signals are shown to be powerful in eliminating MA because of the strong correlation between acceleration signals and MA [20]. There are mainly three kinds of methods using acceleration signals as the reference signals of MA, as follows.

One kind of technique is signal decomposition assisted by acceleration signals [3], [14], [21]. For example, in [3], a method using singular spectrum analysis (SSA) assisted by acceleration signals was proposed to eliminate MA. First, SSA is exploited to decompose a raw PPG into many components. Then a spectral peak associated with MA in a component is identified by checking whether there is also a spectral peak in the spectrum of acceleration signals at the same frequency bin.

Another kind of technique is spectrum subtraction (SS) assisted by acceleration signals [14], [15], [22]. It removed MA by subtracting the spectrum (calculated by Periodogram) of the acceleration signal from the spectrum of the PPG signal.

Adaptive noise cancellation (ANC) assisted by acceleration signals was also exploited to remove MA [12], [17]. For example, one work that we recently proposed in [12] used nonlinear adaptive filtering with acceleration signals being the reference signals of MA. However, ANC is sensitive to the predefined reference signal. If the predefined reference signal was not chosen properly, the performance of MA elimination may be affected. To overcome this issue, another reference [24] proposed using ANC with four predefined reference signals, instead of only one predefined reference signal.

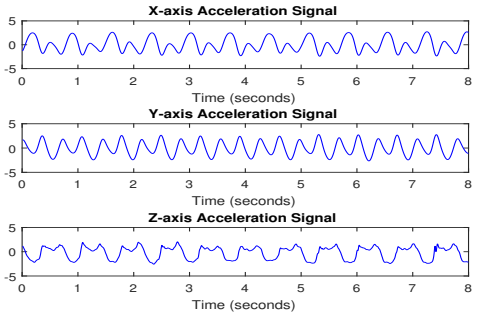
The three kinds of techniques mentioned above obtained satisfactory results to some extent. Unfortunately, we notice that the performance of most of these techniques were only evaluated using specific physical exercises including a few specific motions, such as walking and fast running. In practice, since there are many different types of intensive physical exercises that include many different quasi-periodic motions, such as deep keen bend and elliptical trainer, we have to cope with HR monitoring under many different quasi-periodic motions, where hand movements have different directions and thus cause more complex MA.

To visually see different quasi-periodic motions, an example in Fig. 1 shows segments of three-axis acceleration signals from three different quasi-periodic motions: fast running, elliptical trainer and deep keen bend. The example shows twofold information.

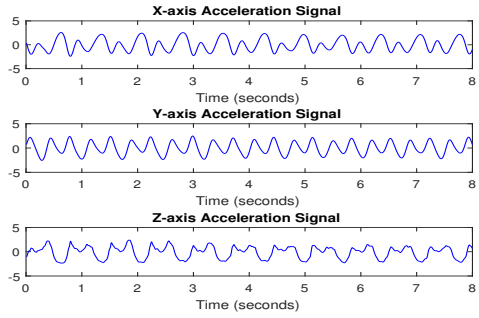
- From Fig. 1(a), Fig. 1(c) and Fig. 1(e) where the signal segments are from one subject, (or Fig. 1(b), Fig. 1(d) and Fig. 1(f) where the signal segments are from another subject), we can see that the acceleration signals from different motions have different characteristics because of different directions of hand movements, which may cause different characteristics of MA in PPG signals recorded by different motions. It indicates that MA caused by different motions are different and very complex, resulting in that it is difficult to eliminate MA caused by different motions.
- From the acceleration signals of fast running (Fig. 1(a) and Fig. 1(b)) from two different subjects, we can see that the acceleration signals from the same kind of motion have similar characteristics. In other words, the acceleration signals from the same kind of motion are regular. Since acceleration signals have strong correlation with MA [20], MA from the same kind of motion may have similar characteristics. If the characteristics of MA caused by one kind of quasi-periodic motion can be learned, MA caused by this kind of motion can be eliminated.

This example motivates using sparse representation based on dictionary learning to eliminate different MAs caused by different motions. Since the acceleration signals from the same kind of quasi-periodic motion are regular and are sparse in certain transform domain, the characteristics of MA, which have strong correlation with acceleration signals, from the same kind of quasi-periodic motion can be learned by dictionary

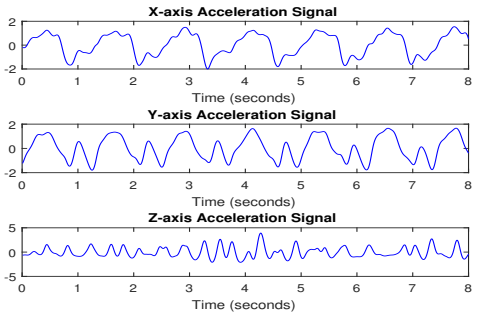
This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2019.2912708, IEEE Journal of Biomedical and Health Informatics



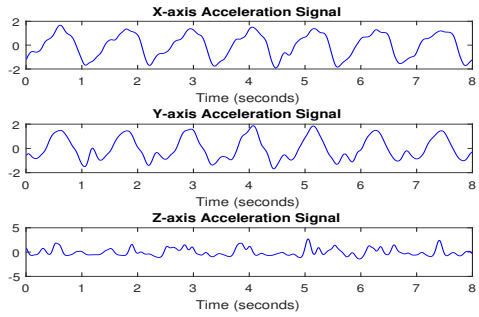
(a) Segments of one subject from fast running.



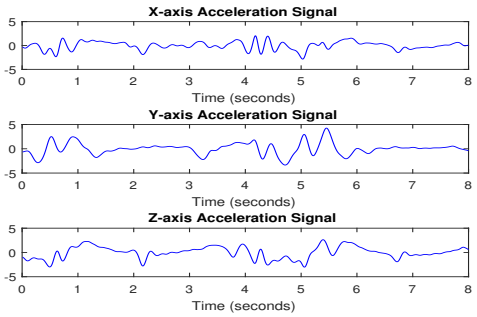
(b) Segments of another subject from fast running.



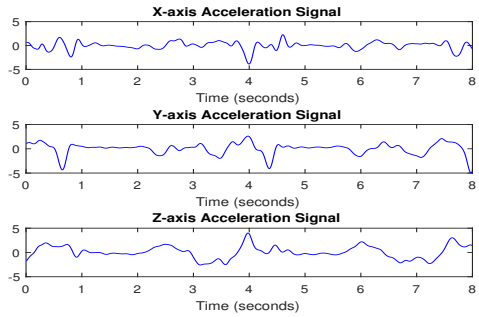
(c) Segments of one subject from elliptical trainer.



(d) Segments of another subject from elliptical trainer.



(e) Segments of one subject from deep keen bend.



(f) Segments of another subject from deep keen bend.

Fig. 1: An example showing segments of simultaneously recorded raw acceleration signals from three different quasi-periodic motions: fast running, elliptical trainer and deep keen bend. Fig. 1(a) and Fig. 1(b) show the signals recorded during fast running from two different subjects. Fig. 1(c) and Fig. 1(d) show the signals recorded during elliptical trainer from two different subjects. Fig. 1(e) and Fig. 1(f) show the signals recorded during deep keen bend from two different subjects.

learning. In other words, dictionary learning can be used to learn a MA dictionary for representing MA caused by this kind of motion through the acceleration signals. Then different MA dictionaries can be learned through the acceleration signals from different motions, used for representing different MAs caused by different motions. Similarly, different PPG dictionaries can be learned through clean PPG signals from different motions, used for representing PPG recorded from different motions. Using the learned PPG and MA dictionaries, sparse representation can remove different MAs caused by different motions. Based on the proposed MA elimination method, we can effectively eliminate different MAs caused by different motions.

III. Proposed Heart Rate Monitoring Approach

The proposed HR monitoring approach is composed of four key parts: dictionary learning for learning different PPG and MA dictionaries, human motion recognition used to recognize which kind of motion one subject is performing and thereby selecting the corresponding dictionaries, sparse representation-based MA elimination for eliminating MA with the selected PPG dictionary and MA dictionary, and spectral peak tracking used to locate the spectral of HR value via a algorithm we proposed in [12]. The flowchart of the proposed HR monitoring approach is shown in Fig. 2.

Before starting the proposed HR monitoring approach, we exploit a preprocessing step called band passing to remove noise and MA outside of the frequency band of interest. The

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2019.2912708, IEEE Journal of Biomedical and Health Informatics

proposed HR monitoring approach uses raw PPG signals and simultaneously recorded acceleration signals. A time window of 8 seconds is sliding on the signals with incremental step of 2 seconds. The proposed HR monitoring approach estimates HR in each time window.

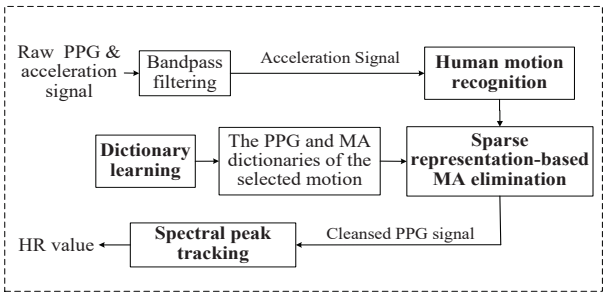


Fig. 2: The flowchart of the proposed HR monitoring approach.

A. Dictionary learning

The goal of this part is to learn the PPG dictionaries and the MA dictionaries corresponding to different motions, preparing for the part of sparse representation-based MA elimination. Predefined transform basis and dictionary learning are two main methods to obtain dictionary. Dictionary learning is used in this paper because it has encouraging potential to provide a sparser representation compared with predefined transform basis [25], [26]. For PPG signals used for arterial blood pressure estimation, dictionary learning was adopted in [27]. The reference [27] proved that the characteristic of PPG signals can be learned by dictionary learning. In our work, dictionary learning is implemented in feature transform domain in order to make PPG and acceleration signals (the reference signals of MA) sparse. Short-time Fourier transform (STFT) is used because STFT is simple and easy to implement.

For dictionary learning, training data are required. Specifically, clean PPG signals from one motion are required to learn the PPG dictionary of one motion, and acceleration signals from one motion are required to learn the MA dictionary of one motion. Acceleration signals for one motion are easily to be obtained, whereas clean PPG signals for one motion are difficult to obtain because the PPG signals recorded during motions often contain MA. There are mainly two ways to get clean PPG signals for one motion. The first way is to use the signals after being processed by some other MA elimination methods. For example, one MA elimination method can be selected to process the raw PPG signals used for training; if the signals after being processed by the method are clean, the signals after being processed by the method would constitute training data. The second way is to select the PPG signals being approximately clean from some other motions. For example, the approximately clean PPG signals from one motion (denoted by Motion1) can be used as the training signals of another motion (denoted by Motion2). At this time, the HR values of the signals from Motion1 should be similar to the HR values of the signals from Motion2. If they have similar HR values, the clean PPG signals for the two motions are similar, and thus the approximately clean PPG signals from Motion1 can be used. In our experiments, the first way is adopted. Specifically, one method proposed in our previous work [12] is used to process the raw PPG signals used for training.

By means of dictionary learning, a PPG dictionary $\tilde{\mathbf{R}}$ (used to represent PPG signals) can be learned from a training set $\tilde{\mathbf{S}}$ which is assemble of STFT magnitudes of clean PPG signals from many time windows. Also, via dictionary learning, a MA dictionary $\mathbf{R}$ (used to represent the MA caused by certain kind of quasi-periodic motion) can be learned from a training set $\mathbf{S}$ which is assemble of STFT magnitudes of acceleration signals corresponding to this motion from many time windows.

$\tilde{\mathbf{R}}$ can be obtained by solving the following optimization problem:

$$\min_{\tilde{\mathbf{R}},\tilde{\mathbf{C}}} \left\| \tilde{\mathbf{S}} - \tilde{\mathbf{R}}\tilde{\mathbf{C}} \right\|_F^2 \quad \text{s.t.} \quad \forall i, \|\tilde{\mathbf{c}}_i\|_0 \leq \tau, \tag{1}$$

where $\|\cdot\|_F$ denotes the Frobenius norm, $\|*\|_0$ denotes the number of elements which are not zero. τ denotes a small positive integer. In equation(1), the training set $\tilde{\mathbf{S}} \in \mathbb{R}^{D\times W}$ (being assemble of training samples) stands for $\{\tilde{\mathbf{s}}_i\}_{i=1}^{W}$. Here D denotes the length of one training sample, and W denotes the number of training samples. A dictionary $\tilde{\mathbf{R}} \in \mathbb{R}^{D\times M}$ stands for $\{\tilde{\mathbf{r}}_i\}_{i=1}^{M}$ (where $\tilde{\mathbf{r}}_i$ denotes one atom), and a coding matrix $\tilde{\mathbf{C}} \in \mathbb{R}^{M\times W}$ stands for $\{\tilde{\mathbf{c}}_i\}_{i=1}^{W}$. Here M denotes the number of atoms in one dictionary.

$\mathbf{R}$ can be obtained by solving the following optimization problem:

$$\min_{\mathbf{R},\mathbf{C}} \|\mathbf{S} - \mathbf{R}\mathbf{C}\|_F^2 \quad \text{s.t.} \quad \forall i, \|\mathbf{c}_i\|_0 \leq \rho, \tag{2}$$

where $\|\cdot\|_F$ denotes the Frobenius norm, ρ denotes a small positive integer, the training set $\mathbf{S} \in \mathbb{R}^{D\times N}$ stands for $\{\mathbf{s}_i\}_{i=1}^{N}$. Here D denotes the length of one training sample, and N denotes the number of training samples. A dictionary $\mathbf{R} \in \mathbb{R}^{D\times L}$ stands for $\{\mathbf{r}_i\}_{i=1}^{L}$, and a coding matrix $\mathbf{C} \in \mathbb{R}^{L\times N}$ stand for $\{\mathbf{c}_i\}_{i=1}^{N}$. Here L denotes the number of atoms in one dictionary.

The two constrained problems expressed by equation (1) and (2) can be solved by many algorithms, such as method of optimal directions (MOD) [28], [30] and k - singular value decomposition (K-SVD) [26]. For illustration, we chose MOD in our experiments. Since equation (1) and (2) are two similar problems, we only describe how to solve equation (2) using MOD as follows.

To update $\mathbf{R}$ and $\mathbf{C}$ iteratively, MOD uses a two phase approach: sparse coding stage and dictionary update stage. The two steps of MOD are as follows.

In sparse coding stage, $\mathbf{R}$ is fixed. Any pursuit algorithm, such as orthogonal matching pursuit (OMP) [31], can be used to compute each column $\mathbf{c}_i$ by approximating the solution of

$$\min_{\mathbf{c}_i} \left\{ \|\mathbf{s}_i - \mathbf{R}\mathbf{c}_i\|_2^2 \right\} \quad \text{s.t.} \quad \|\mathbf{c}_i\|_0 \leq \rho. \tag{3}$$

where $\mathbf{c}_i$ stands for one column in $\mathbf{C}$, $\mathbf{s}_i$ stands for one column in $\mathbf{S}$, and ρ denotes the max number of coefficients for each signal $\mathbf{s}_i$.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2019.2912708, IEEE Journal of Biomedical and Health Informatics

After the sparse coding stage, $\mathbf{c}_i$ for each example $\mathbf{s}_i$ is known. Thus errors $\mathbf{e}_i = \mathbf{s}_i - \mathbf{R}\mathbf{c}_i$ can be obtained. The overall representation mean square error can be represented by

$$\|\mathbf{E}\|_F^2 = \|\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_N\|_F^2 = \|\mathbf{S} - \mathbf{R}\mathbf{C}\|_F^2 . \tag{4}$$

where $\|\cdot\|_F$ denotes the Frobenius norm.

In dictionary update stage, $\mathbf{C}$ is fixed. We can find an update to $\mathbf{R}$ such that the above error (represented by equation(4)) is minimized. $\mathbf{R}$ can be obtained by taking the derivative of equation(4) with respect to $\mathbf{R}$, and the obtained $\mathbf{R}$ is

$$\mathbf{R} = \mathbf{S}\mathbf{C} \cdot (\mathbf{C}\mathbf{C}^{\mathrm{T}})^{-1}. \tag{5}$$

Repeat the sparse coding stage and the dictionary update stage until convergence. Finally, MOD can produce a PPG dictionary $\tilde{\mathbf{R}}$ that approximates $\tilde{\mathbf{S}}$ sparsely and accurately. Similarly, MOD can produce a MA dictionary $\mathbf{R}$ that approximates $\mathbf{S}$ sparsely and accurately. $\tilde{\mathbf{R}}$ and $\mathbf{R}$ will be used in the part of sparse representation-based MA elimination.

B. Human motion recognition

The purpose of this part is to use acceleration signals to recognize which motion one subject is performing, in order to automatically choose a PPG dictionary and a MA dictionary corresponding to this motion. One commonly-used algorithm is XGBoost [32], which is a scalable end-to-end tree boosting system. Since XGBoost has achieved state-of-the-art results on many machine learning challenges [32], it is selected as the classifier in our experiments.

Note that, to train XGBoost-based classifier, training data (acceleration signals from many time windows) are required. When collecting training data, like the part of dictionary learning, we used the leave-one-subject-out method. Specifically, the data from one subject in one data set were used as testing data, while the data from the remaining subjects in one data set were used as training data.

1) Classifier training:

Feature extraction. Before training the classifier, feature vectors being the input of the classifier need to be extracted from the training data. A feature vector is composed of the features: kurtosis, skewness, energy, spectral mean, spectral kurtosis, spectral skewness, the sum of energy for each sub-band signal obtained by wavelet transform, the sum of the mean for each sub-band signal, and the sum of the standard deviation for each sub-band signal. One feature vector is normalized so that the range of features is between -1 and 1. Finally, a lot of feature vectors are obtained, forming a training set.

Classifier training. The training set was used to train the XGBoost-based classifier to obtain the classifier parameters for predicting.

2) Human motion recognition using the trained classifier:

After training the XGBoost-based classifier, this step is to determine which kind of motion one subject is performing. The features mentioned in the last step are extracted from a three-axis acceleration signal in the current time window. The features will be used as the input of the classifier, used to recognize the motion one subject is performing in one time window. Once the motion one subject is performing has been recognized, the PPG and the MA dictionaries corresponding to this motion would be selected.

C. Sparse representation-based MA elimination

In this subsection, a correlation decision is first introduced. Correlation decision is used to determine whether the raw PPG signal contains very strong MA. Then sparse representation-based MA elimination is introduced.

1) Correlation decision:

This step aims to determine whether the raw PPG signal contains very strong MA by calculating the correlation between the STFT magnitudes of the raw PPG signal and the acceleration signal. If the Pearson correlation value between the STFT magnitudes of the raw PPG signal and the acceleration signal is not very high, the raw PPG signal does not contain very strong MA. At this time, to avoid the disturbance of the MA dictionary which comes from the acceleration signals, there is no need to use the MA dictionary $\mathbf{R}$.

Before the correlation decision, STFT is used to transform the raw PPG signal and the acceleration signal after bandpass filtering, $\mathbf{s}_{raw}$ and $\mathbf{a}_{raw}$, into signals in STFT magnitude domain, expressed by

$$\begin{aligned} \mathbf{X} &= |\mathrm{STFT}(\mathbf{s}_{raw})| , \\ \mathbf{A} &= |\mathrm{STFT}(\mathbf{a}_{raw})| . \end{aligned} \tag{6}$$

where $\mathrm{X} \in \mathbb{R}^{m \times n}$ denotes the STFT amplitude of $\mathbf{s}_{raw}$, $\mathrm{A} \in \mathbb{R}^{m \times n}$ denotes the STFT amplitude of $\mathbf{a}_{raw}$. m denotes a frequency range from 0 to 5 Hz, which is selected because the frequency of HR exists in the frequency range from 0 to 5 Hz. n denotes the number of time windows. Here the acceleration signal after bandpass filtering, $\mathbf{a}_{raw}$, is a summation of three-axis acceleration signals after bandpass filtering.

Then $\mathbf{X}$ is transformed to into a column vector, $\mathbf{x} \in \mathbb{R}^{(m \cdot n) \times 1}$, which is composed of the column vectors of $\mathbf{X}$. $\mathbf{A}$ is transformed to into a column vector, $\mathbf{a} \in \mathbb{R}^{(m \cdot n) \times 1}$, which is composed of the column vectors of $\mathbf{A}$. Here $m \cdot n$ denotes the multiplication of m and n.

Next we measure whether the Pearson correlation value ρ_{corr} between $\mathbf{x}$ and $\mathbf{a}$ satisfies

$$|\rho_{corr}| \leq \Delta_{corr}, \tag{7}$$

where Δ_{corr} is a preset threshold.

If ρ_{corr} can satisfy (7), the raw PPG signal is identified as a clean signal. At this time, only the PPG dictionary will be used. This can be expressed by

$$\mathbf{D} = \tilde{\mathbf{R}}, \tag{8}$$

where $\mathbf{D}$ denotes the dictionary that will be used in sparse representation-based MA elimination.

If ρ_{corr} can not satisfy (7), the raw PPG signal is identified as a not clean signal. At this time, both the PPG dictionary and the MA dictionary will be used. This can be expressed by

$$\mathbf{D} = [\tilde{\mathbf{R}}\ \mathbf{R}], \tag{9}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2019.2912708, IEEE Journal of Biomedical and Health Informatics

where $\mathbf{D}$ denotes the dictionary that will be used in sparse representation-based MA elimination.

2) Sparse representation-based MA elimination:

Sparse representation-based MA elimination is to eliminate MA in a raw PPG signal using sparse representation. Sparse representation (using sparse as a condition) is an emerging signal processing technique, showing great potentials in many application fields [33], [34], [35]. PPG and acceleration signals (the reference signals of MA) are sparse in a specific domain. Moreover, these two kinds of signals can be well represented by sparse representation due to the regularity of PPG and acceleration signals (the reference signals of MA).

If one raw PPG signal contains strong MA, we assume that the STFT magnitude of raw PPG signal, $\mathbf{x} \in\in \mathbb{R}^{(m\cdot n)\times 1}$, can be approximately a sum of the STFT magnitude of clean PPG signal, $\mathbf{s} \in\in \mathbb{R}^{(m\cdot n)\times 1}$, and the STFT magnitude of MA, $\mathbf{m} \in\in \mathbb{R}^{(m\cdot n)\times 1}$. The assumption can be expressed by:

$$\mathbf{x} = \mathbf{s} + \mathbf{m}. \tag{10}$$

The assumption used in the paper is similar to the assumption used in the reference [36] for speech enhancement. The assumption used by [36] is that one raw speech signal is approximately a sum of spectral magnitudes of clean speech signal and noise.

The purpose of the assumption is that: if one raw PPG signal contains strong MA, $\mathbf{D} = [\tilde{\mathbf{R}}\,\mathbf{R}]$; at this time, if the assumption can be satisfied, the constrained problem (introduced by the next equation (11)) of the proposed MA elimination approach can be used to eliminate MA.

In sparse representation-based MA elimination, the aim of sparse representation algorithm is to approximate $\mathbf{x}$ with low error using linear combination of a few atoms (chosen from dictionary $\mathbf{D}$) that are correlative to $\mathbf{x}$. Note that the aim of sparse representation algorithm depends on dictionary $\mathbf{D}$. If $\mathbf{D} = \tilde{\mathbf{R}}$, the aim of sparse representation algorithm is to approximate $\mathbf{x}$ with low error using linear combination of a few atoms from $\tilde{\mathbf{R}}$. If $\mathbf{D} = [\tilde{\mathbf{R}}\,\mathbf{R}]$, sparse representation algorithm aims to approximate $\mathbf{x}$ with low error using a sum of a linear combination of atoms from the PPG dictionary $\tilde{\mathbf{R}}$ and a linear combination of atoms from the MA dictionary $\mathbf{R}$.

The constrained problem of the proposed MA elimination approach can be represented by

$$\begin{aligned} &\min_{\mathbf{c}} \|\mathbf{x} - \mathbf{D}\mathbf{c}\|_2 \\ &\text{s.t.} \quad \|\mathbf{c}\|_0 \leq K \end{aligned} \tag{11}$$

where $\mathbf{x}$ is the STFT magnitude of a raw PPG signal after bandpass filtering, $\mathbf{D}$ is the dictionary after correlation decision ($\mathbf{D} = \tilde{\mathbf{R}}$ or $\mathbf{D} = [\tilde{\mathbf{R}}\,\mathbf{R}]$), K is a small positive integer. $\mathbf{c}$ is a sparse coefficient vector. $\mathbf{c} = \mathbf{c}_{\tilde{r}}$ or $\mathbf{c} = [\mathbf{c}_{\tilde{r}}\ \ \mathbf{c}_r]^{\mathrm{T}}$, where $\mathbf{c}_{\tilde{\mathbf{r}}}$ and $\mathbf{c}_{\mathbf{r}}$ are sparse coefficients corresponding to $\tilde{\mathbf{R}}$ and $\mathbf{R}$.

The constrained problem (11) can be solved by many approaches, such as orthogonal matching pursuit (OMP) [31] and basis pursuit denoising [37]. These approaches solve the problem based on different criteria and procedures. We chose OMP in our experiments because of its low complexity and simple implementation [38]. OMP is a greedy algorithm used to choose atoms from a dictionary [31].

After using OMP to solve the sparse representation problem (9), we obtain the sparse coefficients vector $\mathbf{c}$, where $\mathbf{c} = \mathbf{c}_{\tilde{r}}$ if $\mathbf{D} = \tilde{\mathbf{R}}$, and $\mathbf{c} = [\mathbf{c}_{\tilde{r}}\ \ \mathbf{c}_r]^{\mathrm{T}}$ if $\mathbf{D} = \left[\tilde{\mathbf{R}}\,\mathbf{R}\right]$. Then the clean speech magnitude is estimated by disregarding the contribution from the MA dictionary, preserving only the linear combination of PPG dictionary atoms. An estimate of the STFT magnitude of a cleansed PPG segment, $\hat{\mathbf{s}}$, can be obtained by

$$\hat{\mathbf{s}} = \tilde{\mathbf{R}}\mathbf{c}_{\tilde{r}}{}^{\mathrm{T}}, \tag{12}$$

where $\hat{\mathbf{s}}$ can be transformed into a cleansed PPG signal.

The vector $\hat{\mathbf{s}} \in \mathbb{R}^{(m\cdot n)\times 1}$ is transformed into one matrix $\hat{\mathbf{S}} \in \mathbb{R}^{m\times n}$, which can be used to get a cleansed PPG signal in the time domain.

Using $\hat{\mathbf{S}}$, a cleansed PPG signal in the time domain can be obtained by inverse STFT transform. Here inverse STFT transform is used to transform the obtained signal in the STFT magnitude domain, $\hat{\mathbf{S}}$, into a cleansed PPG signal in the time domain, $\mathbf{s}_{recon}$, which will be used in spectral peak tracking. Note that, for inverse STFT transform, the phase of the raw PPG signal, $\mathbf{x}$, is used. The process can be expressed by

$$\mathbf{s}_{recon} = \mathrm{ISTFT}(\hat{\mathbf{S}}), \tag{13}$$

where ISTFT denotes inverse STFT transform.

The process of sparse representation-based MA elimination approach is given in **Algorithm 1**. In the OMP algorithm of this process, for dictionary matrix $\mathbf{D}$, the matrix with indices of its columns in Ω is denoted by $\mathbf{D}_\Omega$. In the OMP algorithm, each iteration t of the while-loop consists of two steps: atom selection and update of the coding vector. The step of atom selection selects the atom that is most coherent to the current residual $\hat{\mathbf{r}}$, implemented in first two lines of the while-loop. The step of update of the coding vector sets $\mathbf{c}$ to the orthogonal projection of $\mathbf{x}$ onto the subspace $\mathbf{D}_\Omega$, implemented in the third line of the while-loop. New residual is computed in the fourth line of the while-loop.

D. Spectral peak tracking

Using the cleansed PPG signal $\mathbf{s}_{recon}$, a spectral peak tracking algorithm we proposed in [12] is used to get HR value. The algorithm is performed in the spectrum f (calculated by Periodogram) of $\mathbf{s}_{recon}$.

Before the algorithm starts, some variables are defined. L_p is the frequency location index of HR estimated in the previous time window. $L_{R1} = [L_p - \Delta_s, \cdots, L_p + \Delta_s]$, where L_{R1} is the range of fundamental frequency of HR, and Δ_s is a small positive integer. $L_{R2} = [2(L_p - \Delta_s - 1) + 1, \cdots, 2(L_p + \Delta_s - 1) + 1]$, where L_{R2} is the range of first-order harmonic frequency of HR. $L_i^0 (i = 1, 2)$ denotes the frequency location indexes of two dominant peaks in L_{R1}, and $L_i^1 (i = 1, 2)$ is from L_{R2}. Here dominant peak denotes the spectral peak with an amplitude larger than a threshold of the maximum amplitude.

Depending on the variables, a harmonic pair $(L_i^0, L_j^1)(i, j \in \{1, 2\})$ with a harmonic relation is defined. Also, one group

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2019.2912708, IEEE Journal of Biomedical and Health Informatics

Algorithm 1 Sparse representation-based MA elimination

Input: a raw PPG signal $\mathbf{s}_{raw}$ and the acceleration signal after bandpass filtering $\mathbf{a}_{raw}$, the PPG dictionary $\tilde{\mathbf{R}}$ and the MA dictionary $\mathbf{R}$, and a small positive integer K.

Output: a cleansed PPG signal $\mathbf{s}_{recon}$.

- Perform STFT transform for $\mathbf{s}_{raw}$ and $\mathbf{a}_{raw}$, getting X and A.
  $\mathbf{X} = |\mathrm{STFT}(\mathbf{s}_{raw})|$.
  $\mathbf{A} = |\mathrm{STFT}(\mathbf{a}_{raw})|$.
- Transform X and A into two vectors, $\mathbf{x}$ and $\mathbf{a}$.
- Perform correlation decision that compute correlation between $\mathbf{x}$ and $\mathbf{a}$, getting the dictionary $\mathbf{D}$ used in sparse representation.
  $\mathbf{D} = \tilde{\mathbf{R}}$ or $\mathbf{D} = \left[\tilde{\mathbf{R}}\ \mathbf{R}\right]$.
- Perform OMP algorithm used to compute sparse coefficient vector $\mathbf{c}$.
  Set $\Omega_0 = \{\}$, $\mathbf{c} = \mathbf{0}$, $\hat{\mathbf{r}} = \mathbf{x}$, $t = 0$.
  while $\|\mathbf{c}\|_0 \leq K$ **do**
    $\mathbf{h} = \mathbf{D}^{\mathrm{T}}\hat{\mathbf{r}}$.
    $\Omega_{t+1} = \Omega_t \cup \{\arg\max_l |\mathbf{h}(l)|\}$.
    $\mathbf{c} = \left(\mathbf{D}_{\Omega_{t+1}}^{\mathrm{T}}\mathbf{D}_{\Omega_{t+1}}\right)^{-1}\mathbf{D}_{\Omega_{t+1}}^{\mathrm{T}}\mathbf{x}$.
    $\hat{\mathbf{r}} = \mathbf{x} - \mathbf{D}\mathbf{c}$.
    $t = t + 1$.
  end while
- Get cleansed PPG signals in the STFT magnitude domain.
  $\hat{\mathbf{s}} = \tilde{\mathbf{R}}\mathbf{c}_{\tilde{r}}{}^{\mathrm{T}}$, where $\mathbf{c}_{\tilde{r}}$ (being from $\mathbf{c}$) is sparse coefficients corresponding to $\tilde{\mathbf{R}}$.
- Transform $\hat{\mathbf{s}}$ into one matrix $\hat{\mathbf{S}}$.
- Perform inverse STFT transform for $\hat{\mathbf{S}}$.
  $\mathbf{s}_{recon} = \mathrm{ISTFT}(\hat{\mathbf{S}})$.

$\{L_1^0, L_2^0, \frac{L_1^1-1}{2}+1, \frac{L_2^1-1}{2}+1\}$ is defined. Based on the harmonic pair and the group, three labels are defined. Label 1 mainly represents the frequency location index of a spectral peak selected from the harmonic pair $(L_i^0, L_j^1)(i, j \in \{1, 2\})$. Label 2 represents the frequency location index (from the group $\{L_1^0, L_2^0, \frac{L_1^1-1}{2}+1, \frac{L_2^1-1}{2}+1\}$) that is closet to the frequency location index of HR in the previous time window. Label 3 represents the frequency location index of the spectral peak of HR in the previous time window.

The algorithm mainly consists of two steps: random forest-based classifier training and spectral peak tracking using the trained classifier. The first step aims to train random forest-based classifier. Based on the trained classifier, the second step is to locate the spectral peak of HR in the spectrum f. After the second step, if the class label of the classifier is Label l (l = 1, 2, or 3), the frequency location index of the spectral peak corresponding to Label l would become the frequency location index H associated with HR in the current time window.

The frequency location index H can be transformed to HR value through the following rule:

$$\mathrm{HR} = 60 * \frac{H-1}{N} f_s \ (\mathrm{BPM}), \tag{14}$$

where f_s denotes the sampling rate, BPM denotes beat per minute, and H belongs to $\{1, 2, \cdots, N\}$ which are the location indexes of N equal frequency bins being from the division of location index of the spectrum $[0, f_s]$.

IV. Data Sets and Performance Indexes

A. Data sets

Using different types of physical exercises consisting of many different motions, we evaluated the performance of the proposed HR monitoring approach. In the experiments of this paper, we showed the performance of HR estimation on six data sets recorded under six common motions: walking, fast running, beckoning, swing arm, elliptical trainer and deep keen bend, which were all related to hand movements because hand movements are the main source of MA [15]. Note that the hand movements of these motions are from different directions, because the hand movements in practical application scenario are from different directions. Fig. 3 gives the sketch maps of the six common motions: walking, fast running, beckoning, swing arm, elliptical trainer and deep keen bend.

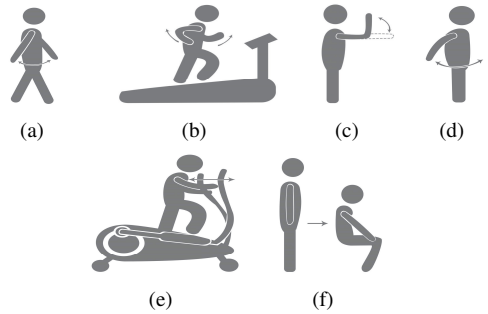


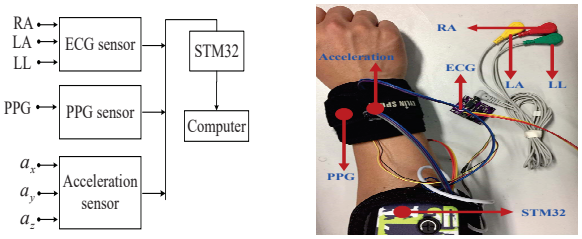
Fig. 3: The sketch maps of six common motions.(a) walking. (b) fast running. (c) beckoning. (d) swing arm. (e) elliptical trainer. (f) deep keen bend.

Note that six data sets used in this paper were recorded during six motions. *Running* was recorded during walking and fast running. *Beckoning* was recorded during beckoning. *Swing Arm* was recorded during swing arm. *Elliptical Trainer* was recorded during elliptical trainer. *Deep Keen Bend* was recorded during deep keen bend. *Mixture* was recorded during beckoning, swing arm and deep keen bend.

The data set (called *Running*) recorded during walking and fast running was first used in [3]. It was recorded from 12 volunteers with age ranged from 18 to 35, being composed of 12 recordings. In each recording, there is a single-channel PPG signal, a three-axis acceleration signal and an ECG signal, recorded simultaneously from a subject. All signals were sampled at 125Hz. In each recording, subjects first walked with 1-2 km/h for 0.5 minute, then ran with 6-8 km/h for 1 minute, next ran with 12-15 km/h for 1 minute, then ran with 6-8 km/h for 1 minute, next ran with 12-15 km/h for 1 minute, and finally walked with 1-2 km/h for 0.5 minute.

The other five data sets (*Beckoning*, *Swing Arm*, *Elliptical Trainer*, *Deep Keen Bend* and *Mixture*) were recorded during beckoning, swing arm, elliptical trainer and deep keen bend. The first four data set was from 9 volunteers with age ranged from 18 to 35, being composed of 9 recordings. The data set of *Mixture* was from 4 volunteers, consisting of 4 recordings.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2019.2912708, IEEE Journal of Biomedical and Health Informatics



(a) The block diagram of the hardware setup. (b) A subject with the hardware.

Fig. 4: (a) The block diagram of the hardware setup for data recording. The ECG sensor was used to obtain a three-lead ECG: LA, RA and LL. The LA and RA were placed at left and right chest, and LL was placed at the left lower abdomen. To collect PPG signals, a reflective pulse oximeter sensor with green LED was used. The PPG sensor was placed at the back of the wrist. To collect three-axis acceleration data, ADXL345 was used. All the data from the sensors were processed by STM32 which is a chip used for configuring ECG, PPG and acceleration sensors. Finally, all the data from STM32 were transmitted to a host computer used for data analysis, processing and presentation. (b) A subject with the hardware.

Each recording lasts for 4 minutes. In each recording, there is a single-channel PPG signal, a three-axis acceleration signal and an ECG signal, recorded simultaneously from a subject. All signals were sampled at 125Hz. Fig. 4(a) shows the block diagram of the hardware setup for data recording. An example of a subject with the hardware is shown in Fig. 4(b).

The details of the other five data sets (*Beckoning*, *Swing Arm*, *Elliptical Trainer*, *Deep Keen Bend* and *Mixture*) are described as follows.

Beckoning was recorded during beckoning. In each recording, subjects rested for the first and last 0.5 minutes. For the other 3 minutes, subjects performed the movement of beckoning. The movement of beckoning was very similar with maneki-neko (literally meaning "beckoning cat") which is a common Japanese lucky figurine that depicts a cat beckoning with an upright paw.

Swing Arm was recorded during swing arm. In each recording, subjects rested for the first and last 0.5 minutes. For the other 3 minutes, the arm wearing sensors swung fore backward with about an angle of sixty degrees.

Elliptical Trainer was recorded during elliptical trainer. In each recording, subjects rested for 0.5 minute, then exercised with 4-6 km/h for 1 minute, next exercised with 7-8 km/h for 1 minute, then exercised with 4-6 km/h for 1 minute, and finally rested for 0.5 minute.

Deep Keen Bend was recorded during deep keen bend. In each recording, subjects rested for the first and last 0.5 minutes. For the other 3 minutes, subjects performed the motion of deep keen bend. Before starting, subjects stood with the waist and back straight and with the knees being in the same direction as the tips of the toes. Note that the movement of squat should be natural and smooth, and hands should be put on the knees when reaching the lowest point of squat so that the hands can give proper support when getting up.

Mixture was recorded during a mixture of three motions (beckoning, swing arm and deep keen bend). The data set consists of four recordings. In each recording, subjects rested for the first and last 0.5 minutes. For the other 3 minutes, subjects performed two types of motions. For the first type, subjects performed beckoning for 1.5 minutes, and then performed swing arm for 1.5 minutes. For the second type, subjects performed beckoning for 1.5 minutes, and then performed deep keen bend for 1.5 minutes. For four recordings in *Mixture*, the first two recordings belong to the first type, and the last two recordings belong to the second type.

B. The performance indexes

1) The performance index for identifying MA intensity:

To better evaluate the performance of HR estimation, an effective way is to see whether one approach can work well on the data set containing strong MA. To achieve this, it is necessary to identify MA intensity in PPG signals, namely to identify whether PPG signals contains strong MA. If one approach can achieve a satisfactory accuracy for HR monitoring using PPG signals containing strong MA, it is robust to strong MA, indicating a good performance of this approach.

In order to identify MA intensity in one recording, namely, to identify whether MA in one recording is very strong, the proportion of time windows identified as being very strong in one recording is calculated. The proportion is denoted by $\mathbf{Intensity}$ and calculated by the following equation:

$$\mathbf{Intensity} = \frac{N_o}{T_o}, \tag{15}$$

where N_o denotes the number of time windows identified as being very strong by the condition expressed in the next equation, and T_o denotes total number of time windows in one recording. The value of $\mathbf{Intensity}$ is between 0 and 1. The larger the value of $\mathbf{Intensity}$ is, the stronger the MA in one recording is.

Table I lists the comparison of six data sets (*Running*, *Beckoning*, *Swing Arm*, *Elliptical Trainer*, *Deep Keen Squat* and *Mixture*) in terms of $\mathbf{Intensity}$. For the six data sets, the averages of $\mathbf{Intensity}$s of all recordings were 0.35, 0.22, 0.43, 0.54, 0.57 and 0.21 respectively. From the results we can see that the MA in *Elliptical Trainer* and *Deep Keen Squat* are more strong than the MA in *Running, Fortune cat* and *Swing Arm*. The results indicates that MA elimination and HR estimation on the data sets of *Elliptical Trainer* and *Deep Keen Squat* are more challenging than those of the data sets of *Running*, *Beckoning* and *Swing Arm*.

Note that, to obtain N_o in equation (15) for one recording, we should design one condition to identify whether the MA in one time window is strong or not. By comparing the spectra of both PPG signals with very strong MA and PPG signals with not strong MA, we observed that the spectrum of the former usually has more spectral peaks than the spectrum of the latter, and also obtained that the distance between the highest peak and the true peak associated with HR (calculated by the simultaneously ECG signal) in the former is usually farther than the distance in the latter. Based on the observations, one condition used to calculate N_o in equation (15) is defined as

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2019.2912708, IEEE Journal of Biomedical and Health Informatics

TABLE I: The comparison of six data sets in terms of **Intensity**. The larger the value of **Intensity** is, the stronger the MA in one recording is. This table shows that *Elliptical Trainer* and *Deep Keen Squat* have more strong MA than *Running*, *Beckoning*, *Swing Arm* and *Mixture*.

Intensity / **Subject** / **Data sets**	Sub.1	Sub.2	Sub.3	Sub.4	Sub.5	Sub.6	Sub.7	Sub.8	Sub.9	Sub.10	Sub.11	Sub.12	Ave
Running	0.64	0.78	0.62	0.15	0.04	0.12	0.00	0.23	0.00	0.71	0.25	0.63	**0.35**
Beckoning	0.23	0.30	0.21	0.32	0.17	0.20	0.19	0.18	0.15	-	-	-	**0.22**
Swing Arm	0.04	0.56	0.35	0.51	0.19	0.34	0.30	0.79	0.79	-	-	-	**0.43**
Elliptical Trainer	0.18	0.06	0.74	0.62	0.74	0.74	0.60	0.42	0.75	-	-	-	**0.54**
Deep Keen Squat	0.57	0.77	0.67	0.75	0.30	0.78	0.51	0.68	0.10	-	-	-	**0.57**
Mixture	0.51	0.13	0.05	0.16	-	-	-	-	-	-	-	-	**0.21**

$$n <= \sigma_1 \quad \& \quad d <= \sigma_2, \tag{16}$$

where n denotes the number of spectral peaks in the spectrum of PPG signal in one time window, d denotes the distance between the highest peak and the true peak associated with HR, and σ_1 and σ_2 denote two small positive numbers being 3 and 2 in our experiments. The process of using (16) to calculate N_o is: set the initial value of N_o to 0; if MA in one time window is identified as being very strong by (16), $N_o = N_o + 1$, otherwise N_o would not change.

2) The performance indexes for HR monitoring:

To evaluate the performance of our proposed HR monitoring approach, the simultaneous ECG signal was used to calculate the ground-truth. Using the ground-truth, two performance indexes used in [3] were exploited. One was the average absolute error defined as:

$$\mathbf{Error} = \frac{1}{\tilde{W}} \sum_{i=1}^{\tilde{W}} |HR_{est}(i) - HR_{true}(i)|, \tag{17}$$

where HR_{true} represents the ground truth of HR in the i-th time window, HR_{est} denotes the estimated HR values, and $\tilde{W}$ denotes the total number of time windows.

The second one was the Bland-Altman plot used to verify agreement between the ground-truth of HR and the estimated HR values. In the plot, the horizontal axis represents the average of two measures (the ground-truth and the estimates), and the vertical axis represents the difference between the two measurements, namely, the difference of the estimates and the ground-truth. The Limit of Agreement (LOA) expressed by $[\mu - 1.96\sigma, \mu + 1.96\sigma]$ was also calculated here, where μ is the average of the differences between the estimates and the ground-truth, and σ is the standard deviation of the differences. In this range, 95% of all differences are inside.

V. Experiments

In this simulation, we evaluated the performance of HR estimation for six data sets of *Running, Beckoning, Swing Arm, Elliptical Trainer, Deep Keen Squat* and *Mixture*, respectively, which were recorded during six motions of walking, fast running, beckoning, swing arm, elliptical trainer and deep keen bend. In the part of dictionary learning, the PPG dictionaries and the MA dictionaries corresponding to the motions in the six data sets were learned by MOD [28], separately. To obtain the training data used for dictionary learning and human motion recognition, the first five data sets (*Running, Beckoning, Swing Arm, Elliptical Trainer,Deep Keen Squat*) adopt the leave-one-subject-out method, which is chosen because it resembles the real world situation [29]. In the real world situation, one method is trained by the data from all the available subjects, and then the method will be tested on new subjects whose data did not exist in the data during training. In the method of leave-one-subject-out, each time the data from one of the subjects are used as testing signals, and the data from the remaining subjects are used as training signals. Note that, for the data set of *Mixture*, training data are obtained by another way, instead of the method of leave-one-subject-out. Specifically, the data in the other five data sets are used as training data, and the data in *Mixture* are tested. When testing the performance on *Mixture*, this way is similar to the real world situation where the proposed method is trained by the data from many motions, and then the proposed method will be tested on a mixture of trained motions.

First, we present the efficacy of sparse representation (SR)-based MA elimination. Fig. 5 shows the elimination of MA caused by three different motions: swing arm, elliptical trainer and deep keen bend. In Fig. 5(a), (or Fig. 5(b), or Fig. 5(c)), from the spectrum of raw PPG we can see that the spectral peak associated with HR is not prominent due to the effect of MA. In contrast, the spectrum of PPG after SR-based MA elimination clearly presents the spectral peak associated with HR. The experiment indicates that SR-based MA elimination can eliminate MA caused by three different motions (swing arm, elliptical trainer and deep keen bend), making the spectral peak associated with HR more prominent, which can help to obtain a more accurate HR value.

We then present the results of HR monitoring for some state-of-the-art HR monitoring approaches. Table II lists the comparison of performances of some state-of-the-art HR monitoring approaches (RandomForest [12], temko [17], TROIKA [3], JOSS [15] and EEMD [14]) in terms of **Error** on six data sets. Averaged across all the recordings in six data sets, **Error** of the proposed approach was 2.40 $\pm$ 1.30 BPM. These results show that the proposed approach can achieve satisfactory accuracy on all six data sets, even on the two challenging data sets: *Elliptical Trainer* and *Deep Keen Squat*.

Moreover, from Table II we can see that, RandomForest [12] and temko [17] can achieve satisfactory accuracy on the data sets of *Running, Beckoning, Swing Arm* and *Mixture*, whereas they can not work well on two challenging data sets: *Elliptical Trainer* and *Deep Keen Squat*. For TROIKA [3], JOSS [15] and EEMD [14], they can work well on *Running*, whereas

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2019.2912708, IEEE Journal of Biomedical and Health Informatics

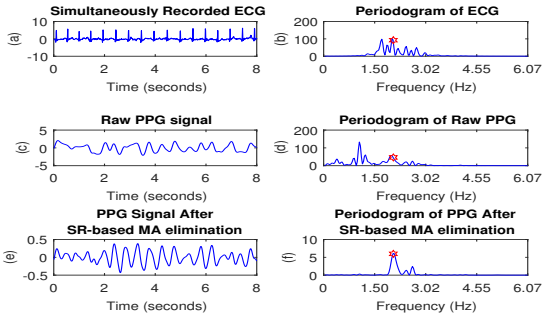
they did not always work well on the other five data sets. In contrast, the proposed approach can achieve satisfactory performance on all the six data sets. The results means that the proposed approach is more robust than the other five approaches for different motions in the six data sets. Further, we can see that the standard deviation (std) of the proposed method (1.30 BPM) is smaller than those of the other five methods: 7.12 BPM for RandomForest [12], 9.73 BPM for temko [17], 10.82 BPM for TROIKA [3], 15.97 BPM for JOSS [15] and 19.36 BPM for EEMD [14]. The results again indicate that the proposed approach is more robust than the other five approaches for different motions in the six data sets.

In the table, for the performances of the other five methods, we found that there are some outliers, such as 70.18 in TROIKA [3]. Considering the **Error**s without outliers larger than 15 BPM, the average **Error**s on all the recordings for the five methods are 3.25 BPM, 3.95 BPM, 3.52 BPM, 4.94 BPM and 2.95 BPM. If some outliers in the other five methods are removed, we also remove the corresponding errors in the proposed method. Since the outliers in the other five methods are different, the proposed method removes different errors, obtaining five different average **Error**s: 2.35 BPM, 2.27 BPM, 2.40 BPM, 2.19 BPM and 2.07 BPM. The results show that the performance of the proposed method is still better than that of the other five methods, even though the outliers have been removed.

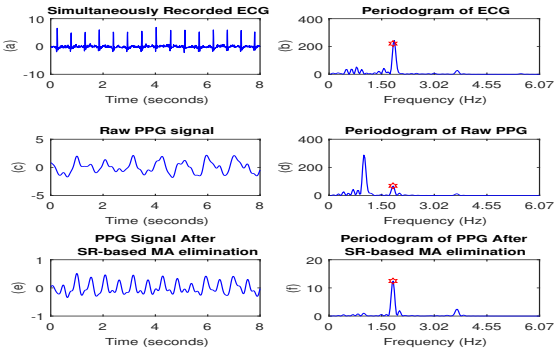
To better compare the methods in Table II, t-test is adopted to test whether the estimation errors of the proposed method are significantly different from those of the other five methods. Here we consider average **Error**s without outliers larger than 15 BPM. Based on t-test, the estimation errors of the proposed method was significantly different from those of the other five methods at the significant level α=0.01. The p values were 7.99×10^{-31} for the proposed method and randomForest [12], 2.22×10^{-6} for the proposed method and temko [17], 1.61×10^{-73} for the proposed method and TROIKA [3], 6.86×10^{-76} for the proposed method and JOSS [15], and 7.21×10^{-23} for the proposed method and EEMD [14].

To further show the performance of the proposed approach, Fig. 6 gives the Bland-Altman plot. In the figure, the absolute value of μ for six data sets were 0.52 BPM, 0.46 BPM, 0.29 BPM, 0.88 BPM, 0.23 BPM and 0.09 BPM. From the results we can see that μ in the six data sets are all close to zero. We also can see that most of the points are close to μ, which means that the estimates are close to the ground truth. The results mean that the proposed approach can achieve satisfactory accuracy on all six data sets, indicating that it is robust to different MAs caused by different motions and is robust to strong MAs in the data sets of *Elliptical Trainer* and *Deep Keen Squat*. This figure indicates a good estimation performance of the proposed approach.

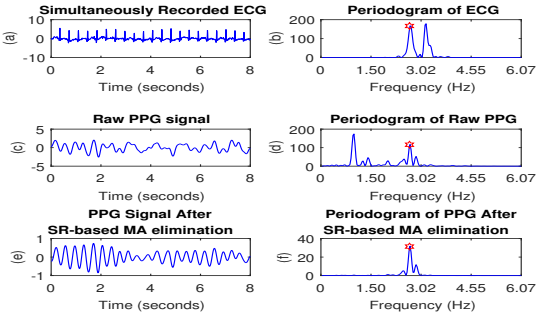
Fig. 6 can also indicate the range of HR where MA is very strong. For example, from Fig. 6(e) we can see that large errors (outside the red line) occur frequently in a low HR range. This indicates that, for *Deep Keen Squat* in Fig.6(e), the proposed method made large errors in a low HR range. This phenomenon may indicate that MA is very strong in a low HR range for *Deep Keen Squat*.



(a) Swing arm.



(b) Elliptical trainer.



(c) Deep keen bend.

Fig. 5: Experiments showing the benefit of SR-based MA elimination using PPG signal segments during three motions: swing arm, elliptical trainer and deep keen bend. The figure shows simultaneously recorded ECG and its spectrum (calculated by Periodogram), simultaneously recorded raw PPG and its spectrum, and PPG after SR-based MA elimination and its spectrum. The red circles indicate the spectral peak associated with HR calculated from simultaneously recorded ECG.

VI. Discussions

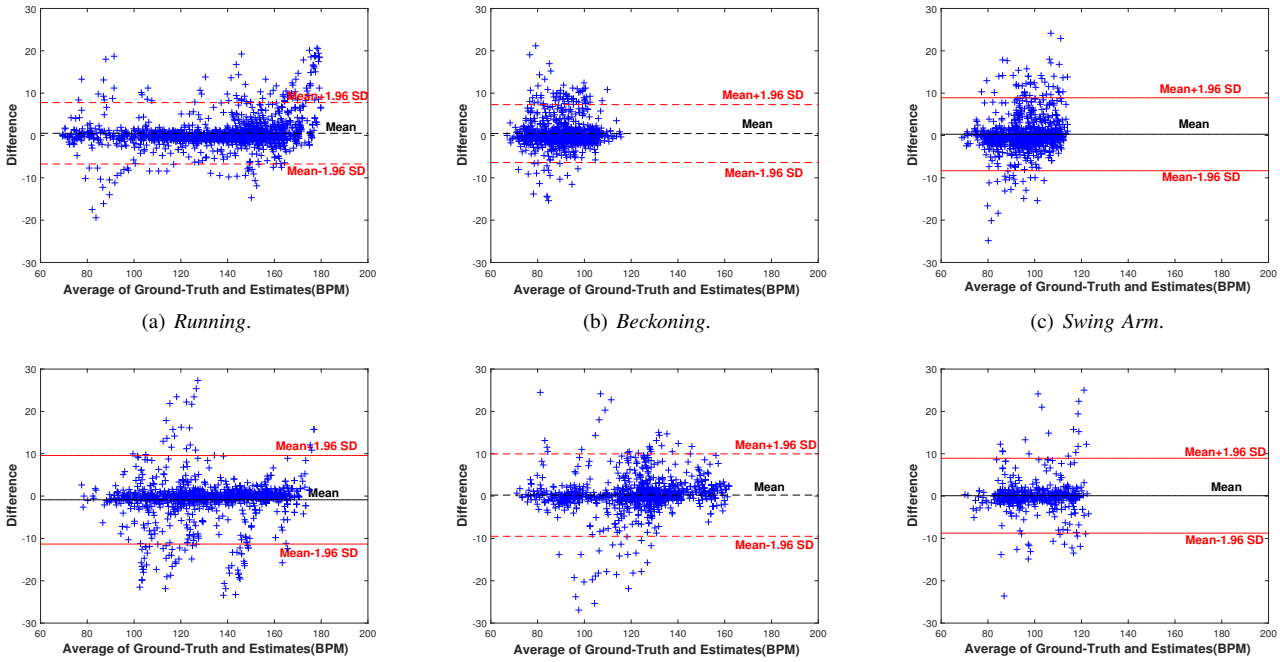
In this paper, we focus on some common quasi-periodic motions that could be modeled. Though quasi-periodic motions are a specific subset of motions, eliminating MA caused by quasi-periodic motions are still challenging, because quasi-periodic motions may be detected as a HR component falsely and thus may result in inaccurate estimation of HR.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2019.2912708, IEEE Journal of Biomedical and Health Informatics

TABLE II: The comparison of HR estimation performances of some HR monitoring approaches (the proposed approach, RandomForest [12], temko [17], TROIKA [3], JOSS [15] and EEMD [14]) in terms of **Error** on six data sets: *Running, Beckoning, Swing Arm, Elliptical Trainer, Deep Keen Squat* and *Mixture* (a mixture of beckoning, swing arm and deep keen bend).

Data Sets	Error: Approaches / Subject	Proposed	RandomForest[12]	temko[17]	TROIKA[3]	JOSS[15]	EEMD[14]
Running [3]	Sub.1	1.94	1.61	1.25	2.87	1.33	2.06
	Sub.2	2.88	1.39	1.41	2.75	1.75	3.59
	Sub.3	0.98	0.73	0.71	1.91	1.47	0.92
	Sub.4	1.54	1.48	0.97	2.25	1.48	1.54
	Sub.5	1.11	0.77	0.75	1.69	0.69	0.97
	Sub.6	1.90	1.34	0.92	3.16	1.32	1.64
	Sub.7	0.67	0.59	0.65	1.72	0.71	2.25
	Sub.8	0.96	0.63	0.97	1.83	0.56	0.63
	Sub.9	0.60	0.57	0.55	1.58	0.49	0.62
	Sub.10	6.58	3.50	2.06	4.00	3.81	4.62
	Sub.11	2.07	1.07	1.03	1.96	0.78	1.30
	Sub.12	1.93	1.04	0.99	3.33	1.04	1.80
	Average in *Running* (mean±std)	**1.93 ± 1.61**	1.23 ± 0.80	1.02 ± 1.25	2.42 ± 0.78	1.28 ± 2.61	1.83 ± 1.21
Beckoning	Sub.1	2.13	2.87	2.20	4.07	17.44	2.04
	Sub.2	1.71	3.80	4.80	2.65	22.60	1.67
	Sub.3	1.97	3.51	2.81	2.81	4.87	2.40
	Sub.4	1.57	2.28	3.42	3.93	32.10	23.03
	Sub.5	2.01	3.23	1.62	70.18	4.75	3.63
	Sub.6	2.66	2.35	2.35	3.81	25.59	2.82
	Sub.7	2.73	3.67	2.59	2.72	8.21	4.26
	Sub.8	2.30	5.86	3.12	2.89	3.59	2.30
	Sub.9	1.59	2.36	1.84	2.49	5.31	1.88
	Average in *Beckoning* (mean±std)	**2.08 ± 0.43**	3.33 ± 1.12	2.75 ± 0.96	10.62 ± 22.34	13.83 ± 10.80	4.89 ± 6.85
Swing Arm	Sub.1	1.57	1.56	1.66	2.65	6.76	2.56
	Sub.2	1.49	2.51	8.59	3.46	19.00	25.58
	Sub.3	3.33	3.00	4.85	4.37	10.04	33.20
	Sub.4	1.55	1.67	1.78	3.96	3.24	37.37
	Sub.5	3.00	2.33	1.78	3.67	31.11	2.94
	Sub.6	2.48	1.81	1.70	2.70	2.68	1.50
	Sub.7	2.80	2.30	1.55	3.86	9.69	3.15
	Sub.8	3.52	5.09	13.77	4.86	19.12	37.97
	Sub.9	3.99	4.25	14.15	11.50	11.11	29.08
	Average in *Swing Arm* (mean±std)	**2.64 ± 0.93**	2.72 ± 1.21	5.54 ± 5.31	4.56 ± 2.70	12.53 ± 9.11	19.26 ± 16.31
Elliptical Trainer	Sub.1	1.31	4.75	1.53	2.65	3.31	3.07
	Sub.2	2.18	1.38	1.25	3.28	1.60	2.81
	Sub.3	2.91	2.09	33.40	42.52	41.33	42.76
	Sub.4	5.42	1.27	3.25	5.73	20.77	9.52
	Sub.5	6.38	14.56	45.47	4.59	40.88	58.81
	Sub.6	1.39	3.40	4.33	2.58	11.83	1.65
	Sub.7	3.78	35.63	2.83	13.64	7.16	3.18
	Sub.8	1.21	0.86	1.30	2.21	3.28	11.24
	Sub.9	1.40	1.88	2.03	3.10	13.25	2.89
	Average in *Elliptical Trainer* (mean±std)	**2.89 ± 1.93**	7.31 ± 11.44	10.60 ± 16.65	8.92 ± 13.09	15.93 ± 15.48	15.10 ± 20.89
Deep Keen Squat	Sub.1	1.83	3.31	2.29	2.49	2.94	44.44
	Sub.2	2.75	2.63	2.23	2.89	74.73	73.08
	Sub.3	3.42	36.56	1.91	3.88	11.75	53.31
	Sub.4	2.75	14.28	31.99	3.90	7.44	25.81
	Sub.5	1.37	1.96	1.55	3.12	15.43	2.66
	Sub.6	3.89	5.23	31.50	3.90	4.59	51.99
	Sub.7	3.14	8.74	21.02	2.58	4.72	59.27
	Sub.8	3.74	14.26	16.73	5.66	70.91	27.03
	Sub.9	1.08	1.46	3.27	1.97	4.03	4.35
	Average in *Deep Keen Squat* (mean±std)	**2.66 ± 1.02**	9.83 ± 11.19	12.50 ± 13.03	3.38 ± 1.10	21.82 ± 29.13	37.99 ± 24.51
Mixture	Sub.1	4.12	3.52	4.29	5.04	9.57	16.26
	Sub.2	2.00	1.90	1.84	2.39	2.88	2.22
	Sub.3	1.36	3.75	1.75	2.22	2.78	1.54
	Sub.4	1.92	2.08	2.35	2.55	1.97	8.02
	Average in *Mixture* (mean±std)	**2.35 ± 1.21**	2.81 ± 0.96	2.56 ± 1.19	3.05 ± 1.33	4.30 ± 3.53	17.40 ± 23.41
	Average of all the subjects in all six data sets (mean±std)	**2.40 ± 1.30**	4.51 ± 7.12	5.86 ± 9.73	5.55 ± 10.82	11.72 ± 15.97	14.33 ± 19.36

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2019.2912708, IEEE Journal of Biomedical and Health Informatics



(a) *Running*. (b) *Beckoning*. (c) *Swing Arm*.

(d) *Elliptical Trainer*. (e) *Deep Keen Squat*. (f) *Mixture*.

Fig. 6: The Bland-Altman plot of the estimates of our proposed approach on the six data sets. In the figure, Mean and SD denote μ and σ, respectively.

As shown in the experiments, the proposed method can work well for many different quasi-periodic motions that have been trained. For one motion that has not been trained, if the motion can be divided into many movements, one scheme depending on motion decomposition and movements comparison could be adopted. Before the scheme, it is necessary to learn movements as many as possible in training stage. The scheme consists of two steps: first, one can divide an untrained quasi-periodic motion into some specific movements (motion decomposition); then one can eliminate MA caused by these specific movements using the dictionaries corresponding to the trained movements which are most similar to these specific movements (movements comparison).

Though the paper focuses on quasi-periodic motions, the paper can be extended into dealing with aperiodic motions, such as shake hands, stretch, push, boxing, which have been considered by many prior studies, such as [12], [17], [39], [40], [41]. In future works, we would like to investigate HR monitoring during aperiodic motions. If one aperiodic motion can be divided into many specific movements associated with arms, one scheme depending on motion decomposition and movements comparison can be adopted. The scheme is similar with the scheme when encountering certain untrained quasi-periodic motion. Before the scheme, one can lean as many movements as possible in advance. The scheme consists of two steps: first, one can divide the aperiodic motion into some specific movements (motion decomposition); then one can eliminate MA caused by these specific movements using the dictionaries corresponding to the trained movements which are most similar to these specific movements (movements comparison).

VII. Conclusion

It is extremely difficult to monitor HR via wrist-type PPG for different types of intensive physical exercises in the real world which includes all kinds of strong quasi-periodic motions, because different quasi-periodic strong motions could cause very strong and complex MA. To alleviate the problem, this study proposed to use dictionary learning-based sparse presentation, which has the ability to design different dictionaries to represent PPG signals and different MAs caused by different strong motions, and thereby improving denoising performance. Experiments on different kinds of quasi-periodic motions demonstrated that the proposed HR monitoring approach is more robust than some state-of-the-art HR monitoring approaches, indicating the potential of the proposed approach in wearable sensors for health monitoring and fitness tracking.

Acknowledgment

This work is supported by a grant from Science & Technology Department of Sichuan Province of China (No. 2019YFG0122), by a grant from the National Key Research and Development Program of China (No. 2017YFB1003003), by a grant from the National Natural Science Foundation of China (No. 61872061 and No. 61872068), by a grant from Science & Technology Department of Sichuan Province of China (No. 2018GZ0071), and by a grant from the Fundamental Research Funds for the Central Universities (No. 2672018ZYGX2018J014).

REFERENCES

[1] J. Allen, "Photoplethysmography and its application in clinical physiological measurement," Physiol. Meas., vol. 28, no. 3, pp. R1-R39, 2007.

[2] A. Johansson, P.A. Oberg, G. Sedin, "Monitoring of heart and respiratory rates in newborn infants using a new photoplethysmographic technique," J. Clin. Monit. Comput. pp. 461-467, 1999.

[3] Zhilin Zhang, Zhouyue Pi, and Benyuan Liu, "TROIKA: A General Framework for Heart Rate Monitoring Using Wrist-Type Photoplethysmographic Signals During Intensive Physical Exercise," IEEE Trans.on Biomedical Engineering, vol. 62, no. 2, pp. 522-531, 2015.

[4] B. S. Kim and S. K. Yoo, "Motion artifact reduction in photoplethysmography using independent component analysis," IEEE Trans. Biomed. Eng., vol. 53, no. 3, pp. 566-568, Mar. 2006.

[5] R. Krishnan, B. Natarajan, and S. Warren, "Two-stage approach for detection and reduction of motion artifacts in photoplethysmographic data," Biomedical Engineering, IEEE Transactions on, vol. 57, no. 8, pp. 1867-1876, 2010.

[6] J. Yao and S. Warren, "A short study to assess the potential of independent component analysis for motion artifact separation in wearable pulse oximeter signals," in Proc. Eng. Med. Biol. Soc. IEEE EMBS 27th Annu. Int. Conf., 2005, pp. 3585-3588.

[7] R. Yousefi, M. Nourani, S. Ostadabbas, and I. Panahi, "A motion-tolerant adaptive algorithm for wearable photoplethysmographic biosensors," IEEE Journal of Biomedical and Health Informatics, vol. 18, no. 2, pp. 670-681, 2014.

[8] M.-Z. Poh, N. C. Swenson, and R. W. Picard, "Motion-tolerant magnetic earring sensor and wireless earpiece for wearable photoplethysmography," Information Technology in Biomedicine, IEEE Transactions on, vol. 14, no. 3, pp. 786-794, 2010.

[9] Lin Z, Zhang J, Chen Y, et al., "Heart rate estimation using wrist-acquired photoplethysmography under different types of daily life motion artifact," Communications (ICC), 2015 IEEE International Conference on. IEEE, pp. 489-494, 2015.

[10] Wijshoff R W, Mischi M, Aarts R M, "Reduction of periodic motion artifacts in photoplethysmography," IEEE Transactions on Biomedical Engineering, vol. 64, no. 1, pp. 196-207, 2017.

[11] Ye Y, Cheng Y, He W, et al., "Combining Nonlinear Adaptive Filtering and Signal Decomposition for Motion Artifact Removal in Wearable Photoplethysmography," IEEE Sensors Journal, vol. 16, no. 19, pp. 7133-7141, 2016.

[12] Ye Y, He W, Cheng Y, et al., "A Robust Random Forest-Based Approach for Heart Rate Monitoring Using Photoplethysmography Signal Contaminated by Intense Motion Artifacts," Sensors, vol. 17, no. 2, pp. 385, 2017.

[13] Khan E, Al Hossain F, Uddin S Z, et al., "A robust heart rate monitoring scheme using photoplethysmographic signals corrupted by intense motion artifacts," IEEE Transactions on Biomedical Engineering, vol. 63, no. 3, pp. 550-562, 2016.

[14] Zhang Y, Liu B, Zhang Z, "Combining ensemble empirical mode decomposition with spectrum subtraction technique for heart rate monitoring using wrist-type photoplethysmography," Biomedical Signal Processing and Control, vol. 21, pp. 119-125, 2015.

[15] Zhang Z, "Photoplethysmography-based heart rate monitoring in physical activities via joint sparse spectrum reconstruction," IEEE transactions on biomedical engineering, vol. 62, no. 8, pp. 1902-1910, 2015.

[16] Sun B, Zhang Z, "Photoplethysmography-based heart rate monitoring using asymmetric least squares spectrum subtraction and bayesian decision theory," IEEE Sensors Journal, vol. 15, no. 12, pp. 7161-7168, 2015.

[17] Temko A, "Accurate heart rate monitoring during physical exercises using ppg," IEEE Transactions on Biomedical Engineering, vol. 64, no. 9, pp. 2016-2024, 2017.

[18] Fujita Y, Hiromoto M, Sato T. PARHELIA: Particle Filter-Based Heart Rate Estimation from Photoplethysmographic Signals during Physical Exercise[J]. IEEE Transactions on Biomedical Engineering, 2018, 65(1): 189-198.

[19] Hayes M J, Smith P R, "A new method for pulse oximetry possessing inherent insensitivity to artifact," IEEE Transactions on Biomedical Engineering, 2001, 48(4): 452-461.

[20] Fukushima H, Kawanaka H, Bhuiyan M S, et al., "Estimating heart rate using wrist-type photoplethysmography and acceleration sensor while running," Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE. IEEE, 2012: 2901-2904.

[21] Zhang Z, "Heart rate monitoring from wrist-type photoplethysmographic (PPG) signals during intensive physical exercise," Signal and Information Processing (GlobalSIP), 2014 IEEE Global Conference on. IEEE, 2014: 698-702.

[22] Sun B, Feng H, Zhang Z, "A new approach for heart rate monitoring using photoplethysmography signals contaminated by motion artifacts," Acoustics, Speech and Signal Processing (ICASSP), 2016 IEEE International Conference on. IEEE, 2016: 809-813.

[23] P. C. Loizou, Speech Enhancement: Theory and Practice. Boca Raton, FL, USA: CRC Press, 2013.

[24] Chowdhury S, Hyder R, Hafiz M S, et al., "Real time robust heart rate estimation from wrist-type PPG signals using multiple reference adaptive noise cancellation," IEEE journal of biomedical and health informatics, 2016.

[25] Zhou Y, Gao J, Chen W, et al., "Seismic simultaneous source separation via patchwise sparse representation", IEEE Transactions on Geoscience and Remote Sensing, vol. 54, no. 9, pp. 5271-5284, 2016.

[26] Aharon M, Elad M, Bruckstein A, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation", IEEE Transactions on signal processing, vol. 54, no. 11, pp. 4311-4322, 2006.

[27] Bose, S. Sree Niranjanaa, and S. Kandaswamy, "Sparse representation of photoplethysmogram using K-SVD for cuffless estimation of arterial blood pressure", Advanced Computing and Communication Systems (ICACCS), 2017 4th International Conference on. IEEE, 2017.

[28] K. Engan, S. O. Aase, and J. H. Hakon-Husoy, "Method of optimal directions for frame design," in IEEE Int. Conf. Acoust., Speech, Signal Process., vol. 5, pp. 2443-2446, 1999.

[29] Ansari, Sardar, et al. "Reduction of periodic motion artifacts from impedance plethysmography", Bioinformatics and Biomedicine Workshops (BIBMW), 2011 IEEE International Conference on. IEEE, 2011.

[30] K. Engan, B. D. Rao, and K. Kreutz-Delgado, "Frame design using focuss with method of optimal directions (mod)," in Norwegian Signal Process. Symp., pp. 65-69, 1999.

[31] Pati Y C, Rezaiifar R, Krishnaprasad P S, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," Signals, Systems and Computers, 1993. 1993 Conference Record of The Twenty-Seventh Asilomar Conference on. IEEE, pp. 40-44. 1993.

[32] Chen T, Guestrin C. Xgboost: A scalable tree boosting system[C]//Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining. ACM, 2016: 785-794.

[33] Sulam J, Ophir B, Elad M, "Image denoising through multi-scale learnt dictionaries[C]//Image Processing (ICIP), 2014 IEEE International Conference on. IEEE, pp. 808-812, 2014.

[34] Deleforge A, Kellermann W, "Phase-optimized K-SVD for signal extraction from underdetermined multichannel sparse mixtures," Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on. IEEE, pp. 355-359, 2015.

[35] Sigg C D, Dikk T, Buhmann J M, "Speech enhancement using generative dictionary learning," IEEE Transactions on Audio, Speech, and Language Processing, vol. 20, no. 6, pp. 1698-1712, 2012.

[36] Sigg C D, Dikk T, Buhmann J M. Speech enhancement with sparse coding in learned dictionaries[C]//Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on. IEEE, pp. 4758-4761, 2010.

[37] Chen S S, Donoho D L, Saunders M A, "Atomic decomposition by basis pursuit," SIAM review, vol. 43, no. 1, pp. 129-159, 2001.

[38] Huang S, Zhu J, "Recovery of sparse signals using OMP and its variants: convergence analysis based on RIP," Inverse Problems, vol. 27, no. 3, pp. 035003, 2011.

[39] Temko A, "Estimation of heart rate from photoplethysmography during physical exercise using Wiener filtering and the phase vocoder", Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE. IEEE, pp. 1500-1503, 2015.

[40] Fallet S, Vesin J M," Robust heart rate estimation using wrist-type photoplethysmographic signals during physical exercise: an approach based on adaptive filtering", Physiological measurement, vol. 38, no. 2, pp. 155., 2017

[41] Mashhadi M B, Farhadi M, Essalat M, et al., "Low Complexity Heart Rate Measurement from Wearable Wrist-Type Photoplethysmographic Sensors Robust to Motion Artifacts", 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, pp. 921-924, 2018.