

Received July 19, 2019, accepted August 4, 2019, date of publication August 13, 2019, date of current version August 23, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2933851

Robust Multistage ECG Identification for Astronaut Spacesuits With IoT Applications

KUO-KUN TSENG¹, LINLIN LIU¹, CHAO WANG¹, K. L. YUNG², (Senior Member, IEEE),
W. H. IP², AND CHIH-YU HSU³

¹School of Computer Science and Technology, Harbin Institute of Technology, Shenzhen 518055, China

²Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University, Hong Kong

³School of Information Science and Engineering, Fujian University of Technology, Fuzhou 350118, China

Corresponding authors: W. H. Ip (wh.ip@polyu.edu.hk) and Chi-Yu Hsu (61201903@fjut.edu.cn)

This work was supported in part by the Shenzhen Government for the project under Grant KQJSCX20170726104033357, Grant JCYJ20150513151706567, and Grant JCYJ20160531191837793, and in part by a grant from the Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University, under Project H-ZG3K.

ABSTRACT Today, the Internet of Things (IOT) concept is gaining much attention and popularity; The related technologies as spacesuits and embedded ECG acquisition device is already existed. However, there are important issues to be resolved when an application is in a space environment. The ECG signal may be measured by different mobile conditions when embedded in spacesuits, requiring a more robust algorithm to remove exercise and noise issues. Thus, we propose a more complete architecture with a new storing polymorphic average template (SPAT) and a multistage identification algorithm (MIA) to improve the robustness of ECG identification in motion. In addition, we select better combinations of de-noising and feature extractions to create a better and more complete architecture. According to our experimental results, our proposed architecture offers better performance than previous adaptive boosting (AdaBoost) methods; thus, it is also suitable for application in astronaut spacesuits.

INDEX TERMS ECG signal, multistage identification algorithm, storing polymorphic average template, spacesuit, MIT-BIH database.

I. INTRODUCTION

In medical settings, ECG can be used in the diagnosis of health conditions, and ECG identification combined with diagnosis is a very promising direction. The ECG signal waveform is related to an individual's heart structure and characteristics of the body; everyone ECG signal is different and not easy to be imitated and hacked. In addition, the collection of ECG signals is convenient and related hardware costs are not prohibitive. ECG signals can also be used to enhance other biometric security measures by enabling more robust biometric features, in turn providing a significant improvement in the identification accuracy and preventing duplication and forgery.

With the advance of space technology, the pace of space exploration will gather speed. Thousands of astronauts will go beyond earth to explore unknown phenomena for mankind. Therefore, it is worth investing more effort in the production of spacesuits to protect astronauts. Spacesuits

worn by astronauts are expensive high-tech products which can help them in adapting to the space environment and accomplish various space missions. Generally, spacesuits are embedded with various sensors for health monitoring and identification of security for astronauts, including an ECG sensor. However, identification technology with mobile ECG signals poses a challenge in terms of developing wearable devices for spacesuits, which is the major research issue in this paper.

At present, most of the existing ECG identification algorithms are based on standard medical monitoring equipment, providing high precision and requiring individuals to lie quietly, as well as making the collection of good ECG signals relatively easily. As a wearable device in a spacesuit is usually in a state of motion, collecting ECG signals would experience interference or deformation. Therefore, collecting ECG signals from astronauts requires a more robust algorithm.

From the general literature on ECG identification, we have investigated the topic from three aspects: (1) signal preprocessing, (2) feature extraction and (3) matching identification.

The associate editor coordinating the review of this article and approving it for publication was Mu-Yen Chen.

Fig. 1 is a classification tree which shows the related algorithm for ECG identification.

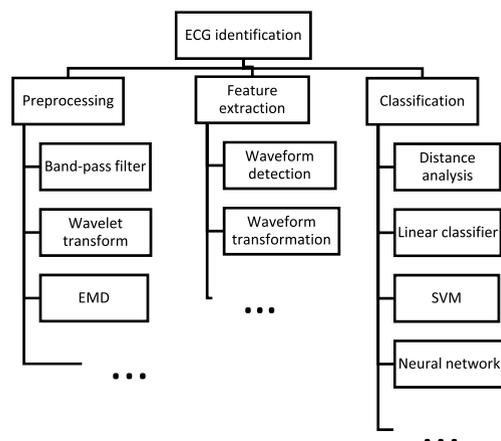


FIGURE 1. Classification tree for the ECG identification algorithm.

According to the above classification tree, we can see that signal preprocessing mainly deals with the noise from various ECG signals in order to provide clear and accurate data signals in the feature extraction phase. An ECG signal has three common sources of noise: baseline drift, power frequency interference and EMG interference. While observing the useful signal in ECG and the noise spectrum distribution, we can find that the baseline drift is in the low-frequency band, while the frequency interference is in the high-frequency band. Israel *et al.* [1] argues that the noise generated by 60 Hz electrical noise and the baseline offset potential of 0.06 Hz is easily recognizable. The author used a 2-40 Hz band-pass filter to filter out noise at 0.06 Hz and 60 Hz while retaining useful ECG signals between 1.10 Hz and 40 Hz. The EMG interference spectrum covers the entire ECG signal. The elimination of myoelectric interference involves a certain degree of difficulty. Many researchers have made attempts to eliminate ECG noise. Kabir and Shahnaz [2] combines Wavelet transform with an EMD algorithm. Palanisamy *et al.* [3] used three kinds of Wavelet function, i.e., “db4”, “coif5” and “sym7”, in a four-threshold selection method for ECG signal de-noising. Zhao *et al.* [4] divided the de-noising process into two steps. First, the trend is used to eliminate the baseline drift. Then, the signal is subjected to a three-layer bi-orthogonal spline of Wavelet transform, with the maximum and minimum threshold method used to filter the noise.

Related work on feature extraction has primarily involved the extraction of the only attribute of the ECG signal which distinguishes it from others. The main features of extraction algorithms include wave detection and waveform transform. Waveform detection methods typically extract the characteristic attributes of some important waveforms in the ECG signal waveform, such as the starting and ending position, amplitude, slope and area. The QRS complex is the most prominent waveform in the ECG signal. Traditional QRS wave detection, proposed by Pan and Tompkins [5], is the

classic differential threshold method. Firstly, the ECG signal is preprocessed in order to eliminate some of the noise and offset, after which some degree of nonlinear transformation is applied to highlight the shape of the QRS waveform, such as a square. Finally, the QRS wave position is located, based on the threshold. In turn, researchers have made further improvements. For example, Chen and Chen [6] improved the adaptive threshold expression to further increase the detection accuracy of QRS complexes. Mallat and Hwang [7] first applied Wavelet transform to detect QRS complexes. Later, scholars proposed many QRS complex detection methods based on Wavelet transform. In [8], morphological and Wavelet transform was fused to locate the QRS wave. This algorithm has the characteristics of strong anti-interference ability, high accuracy and short running time. In addition, some new algorithms are based on other theories [9]–[14].

Waveform transformation methods do not require the detection of important waveforms and boundary points, but directly convert a segment of the ECG signal to extract feature parameters. Plataniotis *et al.* [15] proposed the autocorrelation/discrete cosine transform algorithm in 2006, with an identification result of 100%. Coutinho *et al.* [16] proposed a Ziv-Merhav cross-analysis algorithm. Hegde *et al.* [17] proposed converting a one-dimensional ECG signal into a two-dimensional image, then performing a random transformation on the image and calculating the distance of each row vector in the image, with a final accuracy of 99.85%. Wang *et al.* [18] used sparse decomposition to segment the ECG signal. This method, which does not need to detect a single heartbeat cycle or test a reference point, achieves 99.48% accuracy on a standard open database.

The matching identification stage analyzes and determines the category of characteristic parameters. A variety of distance methods is available to determine the similarity, such as Euclidean distance, Mahalanobis distance, Gaussian likelihood distance and other custom distance methods. In addition, Shen *et al.* [19] used dynamic time warping to find a similar distance between two ECG signals. Among the many ECG matching and classification methods are CNN [20], SVM [21], neural networks [22], [23] and other algorithms.

In addition to the above, we have also investigated ECG applications for astronauts. Some related technologies for spacesuits with ECG acquisition devices already exist [24]. However, applications of this kind seldom focus on ECG algorithms. Charvat *et al.* [25] combined ECG with other detecting results to analyze the cardiovascular status of NASA astronauts. Mccutcheon *et al.* [26] investigated astronauts’ physiological responses by using ECG signals as an emergency aid.

Further to the related general ECG research, there are some aspects which can improve ECG collection in motion, including removing motion artifacts through sensors or filters, and improving feature extraction and principal architecture to improve the robustness of motion in ECG. In terms of sensor improvement methods, two different principles have been proposed: cancellation of de-noising with a capacitive

sensor [27] and de-noising motion bias with an accelerator and gyroscope sensor [28]. In addition, some papers discuss filters for removing motion bias, such as Wavelet [29], ICA [30] and feed-forward combined adaptive [31] filters, which use software-only algorithms to remove noise artifacts.

In conventional ECG research, the topic related to motion ECG is heart rate estimation: [32] examines multichannel heart rate estimation, while [33] investigates compressive sensing used in heart rate estimation. These studies involved research on heart rate estimation rather than on how to solve the problem of identification at different heart rates. Further, applying the overall optimization of the algorithm and architecture for motion ECG, one study has considered DWT filter, PCA and BP neural network recognition [34], but its algorithm and architecture are quite different to ours.

In this paper, we propose a more complete integration architecture, along with a selection of improved algorithms and proposed new methods. Our contribution is summarized as follows:

1. We carry out experiments to find a better combination of algorithms for de-noising filter and feature extraction, although they are relatively mature in ECG processing.
2. We propose the SPAT method to improve the problem of heart rate change.
3. We also propose MIA to enhance robustness, as it is more suitable than the AdaBoost algorithm given the limited scope of ECG data.

In ECG research, there are mature algorithms including signal preprocessing, feature extraction and feature matching. However, for an ECG application in a spacesuit, the signals are usually collected in motion, and the comprehensive architecture and robust algorithm are highly demanding. The identification algorithm is faced with the challenge of signal noise and heart rate change. Therefore, this research is intended to improve the robustness of ECG identification for astronaut spacesuits.

The rest of this paper is organized as follows. The methods and architecture are presented in Section 2. Experimental results and algorithm comparisons are illustrated in Section 3. The application and hardware are set out in Section 4. A discussion is given in Section 5. Finally, a conclusion is made in Section 6.

II. BIOMETRIC-BASED DIGITAL WATERMARKING ALGORITHM

For astronaut spacesuits with IOT applications, our proposed method can be divided into four steps, which includes the selection of two combination algorithms in Step 1 and Step 4, and two new improved algorithms in Step 2 and Step 3, as follows:

- Step 1. Signal de-noising: We have carried out many experiments and selected a better combination of ECG de-noising algorithms.
- Step 2. Heart rate variability reduction: We propose the SPAT method for reducing exercise issues and finding a

better detection and positioning of ORS complex waves to improve the performance of the SPAT.

- Step 3. MIA: MIA can cascade each weak classifier into a strong classifier. It is also better able to fuse multiple features to multistage architecture.
- Step 4. Feature extraction algorithm: We have carried out many experiments to select the most suitable combination of feature extractions for the corresponding stage in the multistage algorithm.

A. SIGNAL DE-NOISING

A variety of noises will be mixed with ECG signals during their collection, especially in the case of portable chip acquisition devices. These noises greatly affect the accuracy of identification. Common ECG noises mainly contain baseline drift, power frequency interference and EMG interference. Noise and the corresponding de-noising algorithm are described in detail below.

1) POWER LINE INTERFERENCE

50-Hz power line interference is generated by electromagnetic radiation and electromagnetic coupling. In the course of research, we found that, regardless of hardware settings or software filters, there are several mature and effective ways to remove power line interference. In this paper, the notch filter [35] and Butterworth band-stop filter are proposed for the removal of 50-Hz power line noise, and the de-noising results are compared. Notch filter is a filter that can rapidly attenuate the input ECG signal at a certain frequency point in order to achieve the filtering effect that selectively rejects a portion of the spectrum. An adaptive notch filter is a system with a notch filter that has a transfer function controlled by variable parameters and a means to adjust those parameters according to an optimization algorithm.

The adaptive notch filter in consideration is as follows:

$$H(z) = \frac{1 + \alpha}{2} \frac{1 - 2\beta(k)z^{-1} + z^{-2}}{1 - \beta(k)(1 + \alpha)z^{-1} + z^{-2}} \quad (1)$$

The coefficients $\beta(k)$ and α depend on the notch frequency w_o and the 3dB attenuation bandwidth Ω given that:

$$\beta(k) = \cos(w_o), \quad \alpha = \frac{1 - \tan(\Omega/2)}{1 + \tan(\Omega/2)} \quad (2)$$

After observing the band diagram of the butterworth band-stop filter, we can see that the spectrum curve in this filter passband is smooth and aligned with the reference line, while the curve in the stopband is slowly reduced to 0. The passband cutoff lower-limit frequency of the band rejection filter is set at 49Hz and the upper limit frequency is 51Hz. The equation for the butterworth principle is as follows:

$$|H(\omega)|^2 = \frac{1}{1 + (\frac{\omega}{\omega_c})^{2n}} = \frac{1}{1 + \epsilon^2 \left(\frac{\omega}{\omega_p}\right)^{2n}} \quad (3)$$

where: n = filter order; ω_c = cutoff frequency; ω_p = passband edge frequency.

2) EMG INTERFERENCE

The frequency range of EMG interference is 5~2,000Hz, which is completely overlapped with the spectrum of useful ECG signals. Noise and the useful signal in the Wavelet domain are different. With the increase in the Wavelet decomposition scale, the noise energy becomes smaller and smaller; however, the energy distribution of the original signal is concentrated at the critical position of the signal mutation and does not weaken with the increase in the decomposition scale [36]. This provides a theoretical basis for using the Wavelet coefficient threshold de-noising algorithm to remove EMG interference. The Wavelet threshold de-noising flow chart is shown in Fig. 2.

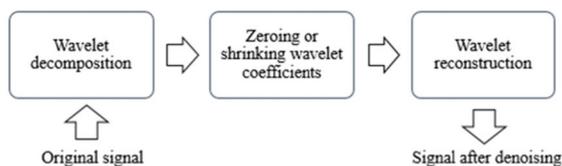


FIGURE 2. Wavelet threshold de-noising flow chart.

The Wavelet threshold de-noising algorithm involves the following:

- Decompose the multi-layer Wavelet of the ECG signal to obtain Wavelet coefficients on different scales.
- Shrink the Wavelet coefficients or adjust them to 0 according to the threshold.
- Multiscale reconstruction of the Wavelet coefficient after being shrunken or adjusted to 0.

In this experiment, the signal is decomposed by three layers, with the detail coefficient of the signal extracted at the scale of 1, 2 and 3. The extracted detail coefficient should be disposed after the threshold is estimated. Common threshold selection rules include the Stein unbiased likelihood estimation threshold, the fixed criterion threshold, the heuristic criterion threshold and the minimum-maximum criterion threshold. Throughout the experiment, we found that different threshold selection strategies have little effect on eliminating the signal-to-noise ratio (SNR) of the signals before and after EMG interference. Thus, we choose the Stein unbiased likelihood estimation threshold for the ECG signals in the self-acquisition database. The mother Wavelet function chooses db8 [37]. Finally, the processed Wavelet coefficients are reconstructed to remove the influence of EMG interference.

3) BASELINE DRIFT

The movement of the human body will cause a baseline drift in the ECG signal during the acquisition process of ECG data using a portable device. The baseline offset belongs to low-frequency interference [38], whose spectrum is generally below 0.5 Hz. The baseline drift signal is shown in Fig. 3.

After the original ECG signal is subjected to multi-layer Wavelet decomposition, the baseline drift is filtered in the

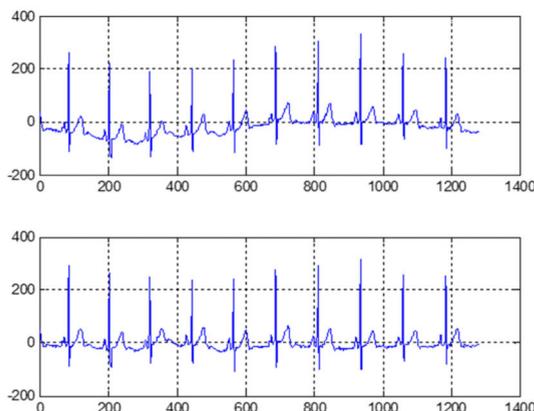


FIGURE 3. ECG signal before and after removing baseline drift.

low-frequency component. Zeroing the low-frequency component of the baseline drift can remove baseline drift in the Wavelet domain. We also select the mobile window median filter method to remove the baseline drift [39] by comparison with the SNR and the root-mean-square error (RMSE) in order to choose which is optimal.

4) COMBINATION OF DE-NOISING ALGORITHMS

SNR and RMSE are two commonly used criteria for measuring noise reduction. SNR reflects the proportional relationship between noise and the true signal. The larger the value, the smaller the proportion of the noise energy to the overall signal energy. RMSE represents the amount of deviation between the observed value and the true value. The smaller the RMSE, the smaller the distortion rate of the signal before and after de-noising. Table 1 shows the performance comparison using the combination of de-noising algorithms described above.

TABLE 1. Comparison of different de-noising algorithms combination.

Combination of de-noising algorithms	SNR	RMSE
Wavelet threshold de-noising + wavelet reconstruction de-noising + notch filter	82.253	9.66E-04
Wavelet threshold de-noising + moving window median filter + notch filter	89.9935	0.0076
Wavelet threshold de-noising + wavelet reconstruction de-noising + Butterworth band-stop filter	78.736	2.37E-04
Wavelet threshold de-noising + moving window median filter + Butterworth band-stop filter	79.4153	2.32E-04

According to the de-noising experiment, we choose a better combination of filter algorithms. The combination of the

Wavelet threshold of de-noising, the moving window median filter and the Butterworth band-stop filter is better when compared with the SNR and RMSE values in Table 1; therefore, this combination is adopted as the de-noising scheme.

B. HEART RATE VARIABILITY REDUCTION

As ECG signals, based on the motion problem, cause heart rate variability to be large, we take a series of measures to enhance the robustness of the algorithm. The following describes the storing polymorphic average template, QRS complex wave detection and positioning, and fractional sampling rate conversion.

1) STORING POLYMORPHIC AVERAGE TEMPLATE (SPAT)

The problem of movement is a challenge for ECG in real life, because the individual cannot always be in a state of calm during the acquisition process. If a person slightly moves or is faced with great emotional ups and downs, his/her heart rate will be significantly accelerated. Therefore, the time-domain period of the ECG waveform will become shorter, and its amplitude will cause tensile deformation in the vertical direction. Fig. 4 shows the ECG signal waveform acquired by the subjects in a quiet state and after motion. We can see that the single cycle length of the ECG waveform in the motion state is shorter than in the quiet state; further, the T wave amplitude is larger than in the quiet state.

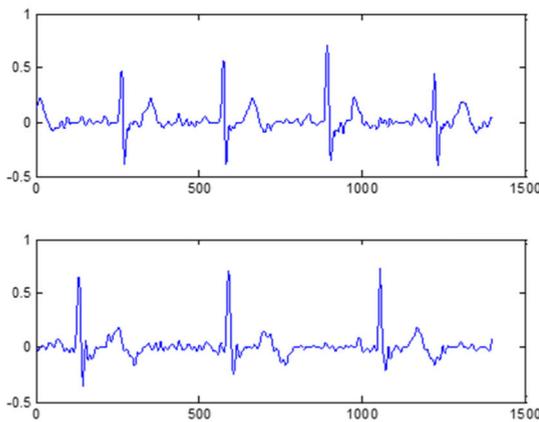


FIGURE 4. ECG waveform comparison between exercise state and clam state.

In this paper, the SPAT optimization method is used to solve the problem of heart rate variability caused by exercise. The heart rate of normal adults is in the 60~100 bpm range, while the exercise heart rate of normal people is 170 bpm on average. In this experiment, in order to cover the extreme heartbeat range, ECG data within the range of 50~140 bpm were selected. Then, those selected ECG data were divided by 10 bpm, 5 bpm and 2 bpm for storage in the corresponding template. At the time of testing, the process involves finding the closet template for each single heartbeat cycle and to make a comparison. Fig. 5 is a graphical representation of how the

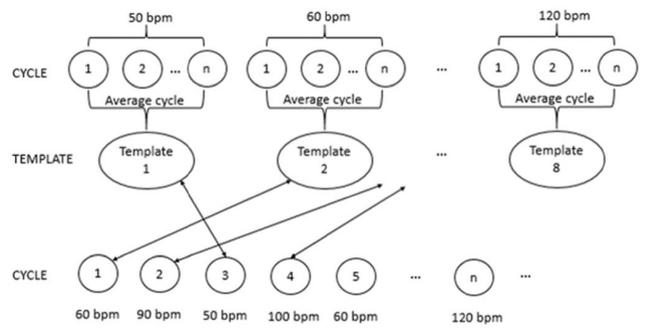


FIGURE 5. Process of storing polymorphic average templates.

polymorphic average templates are stored, as well as how their similarity is calculated.

The specific steps for the SPAT and comparison processes are as follows:

- Divide the ECG signal into a single heartbeat cycle using T wave detection and calculate the heart rate of each single heartbeat cycle according to equation (4).

$$HeartRate = \frac{sample_rate}{len0} \times 60 \tag{4}$$

where: *sample_rate* = sampling frequency; *len0* = single heartbeat cycle length.

- After the segmentation step, all the single heartbeat cycles are classified according to the heart rate range [50, 50 + *interval*, 50 + 2 * *interval*, . . . , 140] (bpm), where: *interval* = heart rate interval of 10 bpm, 5bpm and 2 bpm. The classification result is [*Data*₅₀, *Data*_{50+*interval*}, *Data*_{50+2**interval*}, . . . , *Data*₁₄₀].
- Calculate the average value of each column in the single heartbeat cycle which is well classified. A polymorphic average template for the corresponding range of heartbeats is obtained. The process is shown in equation (5).

$$Template(i) = \frac{1}{n} \sum_{k=1}^n Data_i(k) \tag{5}$$

- For each single heartbeat cycle in the training set and test set, find the template closest to its corresponding heart rate in the polymorphic template and calculate their similarity.

2) DETECTION AND POSITIONING OF ORS COMPLEX WAVES

With the aim of storing the polymorphic template storage algorithm, we must separate each single heartbeat cycle and split the single heartbeat cycle for R wave reference point detection. In addition, the extraction of waveform features also needs to detect the location of QRS complex waves. We know from the above research that the difference of threshold method is simple and effective. Here, we use the Pan and Tompkins algorithm [5] to perform R wave detection on the demagnetized ECG signal. A detailed flow chart of the algorithm is shown in Fig. 6.



FIGURE 6. Flow chart of the Pan and Tompkins algorithm.

The basic steps of the Pan and Tompkins algorithm are as follows:

- *Low-pass filter*: The low-pass filter in the Pan and Tompkins algorithm is the integral coefficient, which can reduce the computational complexity according to equation (6):

$$H(z) = \frac{1}{32} \frac{(1 - z^{-6})^2}{(1 - z^{-1})^2} \tag{6}$$

- *High-pass filter*: The role of the high-pass filter in the Pan and Tompkins algorithm is to filter out the baseline drift according to equation (7):

$$H_{hp}(z) = z^{-16} \frac{1}{32} \frac{(1 - z^{-32})}{(1 - z^{-1})} \tag{7}$$

- *Derivative filter*: The derivative filter in the Pan and Tompkins algorithm eliminates the input DC component and enhances the slope of the P wave, T wave and QRS waves. The transfer function is shown in equation (8):

$$H(z) = \frac{1}{8} (2 + z^{-1} - z^{-3} - 2z^{-4}) \tag{8}$$

- *Square filter*: The effect of the square filter is to make the sample value into an integer, highlighting the Q wave and S wave. The relationship between the output and input of the square filter is given by equation (9):

$$y(n) = x^2(n) \tag{9}$$

- *Integral filter*: Set N as 60. The relationship between the output $y(n)$ of the integral filter and the input $x(n)$ is given by equation (10):

$$y(n) = \frac{1}{N} \sum_{i=1}^N x(n - (N - i)) \tag{10}$$

Both the QRS waves of the ECG signal in the self-acquisition database and the standard database were detected by the Pan and Tompkins algorithm. The positioning results are shown in Fig. 7.

3) FRACTIONAL SAMPLING RATE CONVERSION

During exercise, the heart rate becomes larger and the time-domain period of the ECG signal becomes shorter. Moreover, the amplitude and the time interval of the waveform are deformed and compressed. When using polymorphic template optimization, a similar distance is calculated for each single heartbeat cycle in order to find the template closest to the respective heart rate. Therefore, we use the fractional sampling rate conversion method in order to convert the length of two ECG data into one, which will improve the identification algorithm classification rate.

Fig. 8 shows the A/B time process in the sample rate conversion, which input the $x(T_1)$ signal sequence with B time interpolation to output $w(T_2)$, a transform function $h(k)$ may also be applied to obtain different domain signal sequence $v(T_2)$. Finally, sequence $v(T_2)$ can be extracted as sampling frequencies of fractional times $y(T_3)$.

C. MULTISTAGE IDENTIFICATION ALGORITHM (MIA)

Feature fusion can enhance the classification ability of the whole system, and the fusion feature will have the advantage of strong complementarity and low redundancy [40]. The MIA flow chart is shown in Fig. 9.

The main idea behind the algorithm is as follows. First, we divide the ECG signal database into a training database and a testing database. Regardless of whether the assumption produced by the first-stage classification result is strong or weak, the identified sample weight is set to 0, and the unrecognized sample can continue to be input to the second-stage identification process. The feature space is transformed in the second stage, with samples that cannot be identified in the first stage determined according to which individual category they belong from another angle of the eigenvalue. The last stage contains the final classification result and the total misidentification of samples. The pseudo-codes for MIA are shown in Algorithm 1, where: $FeatureExtraction()$ = corresponding feature extraction function for each stage; $Feature_i_stage()$ = classification function in MIA; N = number of stages of this architecture; $Train$ = corresponding training function for each stage; $Best_thre_i$ = optimal threshold for a single stage.

The MIA can cascade each weak classifier into a strong classifier. Each feature parameter is able to analyze the characteristics of ECG data from different angles, meaning that the classifier can achieve high identification accuracy. Since the identification process and feature vectors of each stage are independent of each other, the final identification accuracy of the MIA can be expressed by equation (11):

$$R_n = 1 - (1 - r_1)(1 - r_2) \cdots (1 - r_{n-1}) \tag{11}$$

where: r_i = identification accuracy of each stage. We can obtain equation (12) from equation (11).

$$\lim_{n \rightarrow \infty} R_n = 1 \tag{12}$$

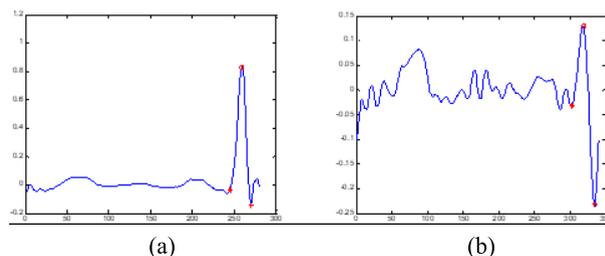


FIGURE 7. ECG signal QRS positioning: (a) QRS positioning in the self-acquisition database; (b) QRS positioning in the standard database.

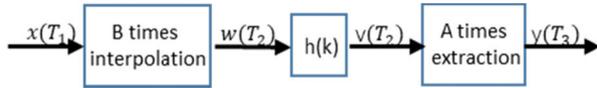


FIGURE 8. Fractional sampling frequency conversion diagram.

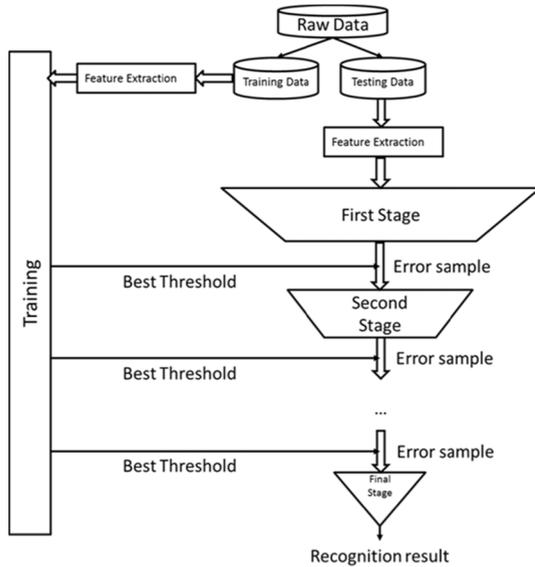


FIGURE 9. MIA flow chart.

The above equation shows us that, when we infinitely increase the number of stages of the MIA, the final identification accuracy can reach 1, which is unrealistic in the actual application, because it must be at the expense of high time complexity. But, for the ECG single-lead one-dimensional signal, the feature extraction algorithm is relatively simple. As such, we can experimentally obtain an MIA within the acceptable range of time complexity.

Algorithm 1 ECG MIA

```

1: Feature_i_train ← FeatureExtraction(trainingData)
2: Feature_i_test ← FeatureExtraction(testingData)
3: Best_thre_i ← 0
4: for i from 1 to N
5: Best_thre_i ← train(Feature_i_train)
6: end
7: ErrorSample ← Feature_i_test
8: for i from 1 to N
9: ErrorSample ← Feature_i_stage(ErrorSample,
    Best_thre_i)
10: end
11: Recog ← 1 - ErrorSample/testingData
    
```

D. FEATURE EXTRACTION ALGORITHM

We know from the above MIA that, theoretically, with the increase in the number of stages, the identification accuracy will reach 100%; but, taking into account time complexity and spatial complexity, we choose three simple and effective

feature extraction algorithms: Wavelet coefficients, reduced binary pattern (RBP) statistics and waveform characteristics. Statistical features and waveform characteristics, based on the RBP algorithm, can come from the time domain, Wavelet domain and statistical characteristics of three angles in order to measure the ECG data. We then adjust and compare the parameters.

1) WAVELET COEFFICIENT CHARACTERISTICS

The k-nearest neighbor (k-NN) classifier is a statistics-based tool. The main goal is to find the k-most similar sample from the test sample and to determine the type of test sample based on the sample type. The similarities among the samples are measured by the Euclidean distance where the distance means the difference. Many studies have used NN or k-NN as their classifier in ECG identification systems, such as [8], [16], [19], [20].

We first detect and position the QRS wave, then divide each single heartbeat cycle. The EMG signals are decomposed into three stages in five cycles, and the obtained detail coefficients $CD_1 \sim CD_3$ and approximate coefficients CA_3 are taken as eigenvectors. We respectively calculate the correlation coefficient of these feature vectors so as to obtain a similar distance between the signal with their own, and with others. The three-stage Wavelet decomposition of ECG signals is shown in Fig. 10.

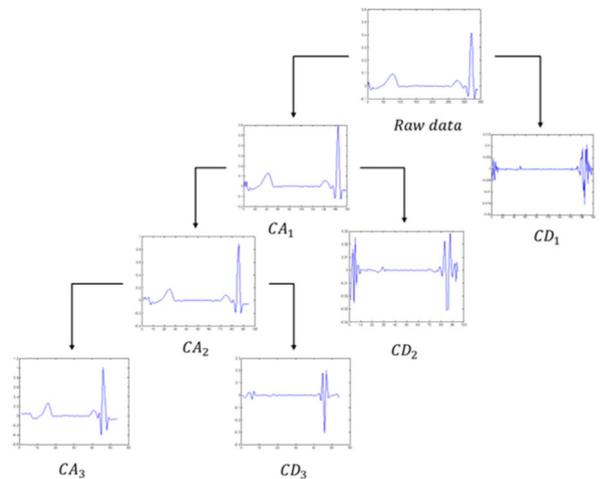


FIGURE 10. Wavelet decomposition of the ECG signal.

2) RBP STATISTICAL CHARACTERISTICS

The RBP algorithm extracts statistical characteristics of local ECG signals [41]. The two successive sample points in the ECG data are treated as a sample pair. We count local characteristics by analyzing the ups and downs of the sample pair. More specifically, when the sample pair is incremented, the sample pair is labelled as 1; when the sample pair is decremented, the sample pair is labelled as 0. The transformation

function is shown in equation (13).

$$y_i = \begin{cases} 1, & x_{i+1} > x_i \\ 0, & x_{i+1} \leq x_i \end{cases} \quad (13)$$

This part of the ECG data will be converted to a 0, 1 sequence, then converted into decimal numbers. If we integrate P bits into decimal numbers, the range of decimal numbers will be $0 \sim 2^P$. We count the number and probability of each decimal number, and use this information as a feature vector to input to the classification algorithm.

ECG data from self-acquisition databases possess much noise and mutated waveforms. When the sampling frequency is high, the smoothness of the ECG signal waveform is not enough. Further, local statistical characteristics of adjacent sample pairs have some limitations. In turn, the characterization effect is not obvious and should be improved. The specific improvement is being able to collect a sample pair from several samples instead of counting adjacent samples by equation (14).

$$y_i = \begin{cases} 1, & x_{i+interval} > x_i \\ 0, & x_{i+interval} \leq x_i \end{cases} \quad (14)$$

where: X is the sampling interval for the sample pair. Experiments show that this improved algorithm offers certain improvements to the MIA. The MIA results, from using different interval values, are shown in Fig. 11.

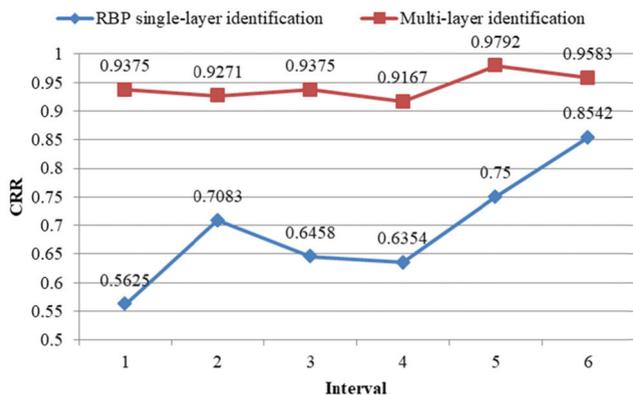


FIGURE 11. Effects of different interval on RBP algorithm identification results.

We observe that, as the interval increases, the identification accuracy increases first and then decreases; and, at the time when $interval = 5$, the CRR increases significantly to 97.92%. Before analyzing and explaining the reason why this situation occurs, we need to describe the calculation equation for the degree of similarity following RBP statistical feature extraction. The similarity function [35] is shown in equation (15).

$$D = \frac{\sum_{k=0}^{2^m-1} |R_1(w_k) - R_2(w_k)| p_1(w_k) p_1(w_k)}{(2^m - 1) \sum_{k=0}^{2^m-1} p_1(w_k) p_1(w_k)} \quad (15)$$

where: $p_1(w_k)$, $p_1(w_k)$ = probability of the occurrence for each decimal integer, respectively; $R_1(w_k)$, $R_2(w_k)$ = ranking of $p_1(w_k)$, $p_1(w_k)$; m = how many bits of binary numbers are integrated into a decimal; w_k = integrated decimal number.

After calculating the similarity of the statistical features, we found that, when $interval = 5$, the similarity has many ranking values. But, when $interval \leq 3$, ranking values almost do not appear. As ranking values lead to fake features, which should increase the identification accuracy, we chose $interval = 3$.

3) WAVEFORM CHARACTERISTICS

Compared with the ECG signal measured by a medical device, the ECG signal waveform in the self-acquisition database is not stable enough. The P wave, T wave and U wave are all distorted except for the QRS complex wave, while the detection error is too large. Therefore, we extracted only eight QRS complex wave-related waveform features: time interval characteristics (RQ, RS, QS), amplitude characteristics (ampRQ, ampRS), slope (RS_slope, QR_slope), and area (QRSarea).

In addition, we take all of the single-heartbeat cycle data as the overall appearance feature vector. Based on the characteristics of the ECG data from the self-acquisition database, wave feature extraction accuracy may not be enough. Moreover, the waveform feature also belongs to the time domain's appearance. Hence, we add the overall appearance of the characteristics as a contrast in order to observe which exterior characteristics of the identification effect are good.

III. ALGORITHM DESIGN

A. THRESHOLD SELECTION

The one-to-one identification mode is used here to confirm whether the user's claimed identity matches his or her identity. During the feature matching process, it is necessary to set a certain threshold for identification. The two threshold selection strategies employed in this paper are described in detail below.

The first strategy is known the scale factor training method. Firstly, we find two subjects' similarity with their own $X(n)$ and their own similarity with others $Y(n)$, then obtain the mean and variance of $X(n)$ using equation (16) and equation (17).

$$Mean = \frac{1}{n} \sum_{i=1}^n X(i) \quad (16)$$

$$Var = \frac{1}{n} \sum_{i=1}^n (X(i) - Mean)^2 \quad (17)$$

Thus, the threshold can be obtained by equation (18).

$$threshold(i) = Mean + j \times Var, \quad i = 1, \dots, k \quad (18)$$

Where: j = scale factor; k = range of scale factor. By adjusting the scale factor, we can obtain a series of thresholds. In the training process, we compare $X(n)$ and $Y(n)$ with this series of thresholds. If x is less than the threshold, then the

classification is wrong; if y is greater than the threshold, the classification is also considered to be wrong. Accordingly, we can calculate the false rejection rate (FRR) and the false acceptance rate (FAR) corresponding to each value. Further, the threshold corresponding to the smallest sum of the FRR and FAR is the best threshold for training.

The second strategy is known as the method of minimum similarity. First, we find subjects' similarity with their own $X(n)$ and their own similarity with others $Y(n)$, then select the minimum value to determine the threshold. The selected equation is shown in equation (19).

$$threshold = \min(X(n)) \tag{19}$$

These two threshold selection strategies apply to the database in the case of different characteristics and different occasions, respectively. The specific experimental comparison results will be given by the following specific algorithm.

B. PERFORMANCE METRICS

This paper takes the false rejection rate (FRR) and the false acceptance (FAR) as the system's identification accuracy metrics. FRR represents the probability of misjudging their own ECG data in others, FAR indicates the probability that others' ECG data are misjudged as their own. We use the equation (20) below to obtain the correct classification rate (CRR) as the final identification accuracy.

$$CRR = 1 - (FAR + FRR/2) \tag{20}$$

The relationship between them is shown in Fig. 12.

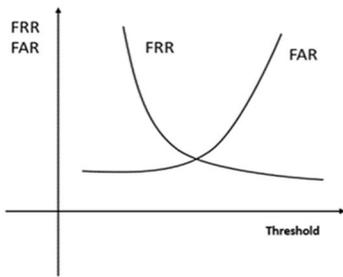


FIGURE 12. Relationship between FRR and FAR .

C. EXPERIMENTAL DATA DESCRIPTION

We used a portable device to collect ECG signals from 12 individuals in a quiet state, and ECG signals from the same 12 individuals after running 4 mins at a speed of 6 km/h on a treadmill. We collected over a period of 2 mins in both motion and quiet states. The sampling frequency was 500 Hz. The device was connected to a PC via a micro USB in order to collect ECG data from a finger-type collection terminal. The collected ECG data were stored on the PC. ECG signals from the MIT-BIH ST Change database were used as comparison data. This database contains 28 individuals' ECG signals, the sampling frequency of which is 360 Hz. Most of them were tested after exercise. Each individual's

heart rate fluctuates in a certain range. We selected 23 individuals' ECG signals as contrastive data based on the motion problem algorithm. Contrasting the experimental results of the two kinds of database should reflect the robustness of the proposed algorithm.

D. COMPARISON OF SPAT OPTIMIZATION

In order to solve the problem that the heart rate variability when collecting individuals' ECG signals is significant after exercise, we applied the SPAT to optimize those heart rate data. More specifically, we extracted each single cycle of heartbeat in the training data and classified these ECGs every 10 bpm, 5 bpm and 2 bpm within the heart rate range of 50~140 bpm, then averaged the periodic cycles to obtain the polymorphic average template. During the training and testing process, the distance between the extracted feature vectors was calculated for each single ECG to find the template closest to the corresponding heart rate. Finally, these distances were used as the input for the MIA in order to obtain the identification accuracy.

The cross-validation algorithm is used for ECG signals without polymorphic average template optimization. Their own distance with their own $X(i)$ and the distance with others $Y(i)$ can be acquired by the cross-validation algorithm. ECG data are divided into SN groups, with each group having one sampling cycle. There will be a $SN - 1$ time comparison with ECG signals from the same individual, and a $(PN - 1) \times SN$ time comparison between different individuals' ECG signals, where: PN = number of individuals. The corresponding equations are shown below.

$$X(i) = \frac{1}{SN - 1} \sum_{j=1}^{SN} d(i, j) \tag{21}$$

$i = 1, \dots, SN; j \in other \cap j \neq i$

$$Y(i) = \frac{1}{PN - 1} * \frac{1}{SN} \sum_{k=1}^{SN} \sum_{m=1}^{PN-1} d_m(i, k) \quad k \in other \tag{22}$$

Fig. 13 presents a comparison of the experimental results of the MIA in the self-acquisition database for the optimization of the SPAT.

According to Fig. 13, for self-acquisition database I, the accuracy will be 1~6% higher after using polymorphic average template optimization. The MIA with waveform features, Wavelet coefficients and RBP statistical features is better than the MIA with overall appearance features, Wavelet features and RBP statistical features, regardless of whether it is optimized. For the polymorphic average template with a heart rate interval of 10 bpm, the best identification accuracy after optimization is 97.92%.

For self-acquisition database II, the cross-validation method is more efficient than polymorphic average template optimization. We compare the classification results of the single classification algorithm, which uses cross validation with the polymorphic average template and analyze the reason for these results. From the above, we know that the results of the MIA have a relationship with the attribute characterization of

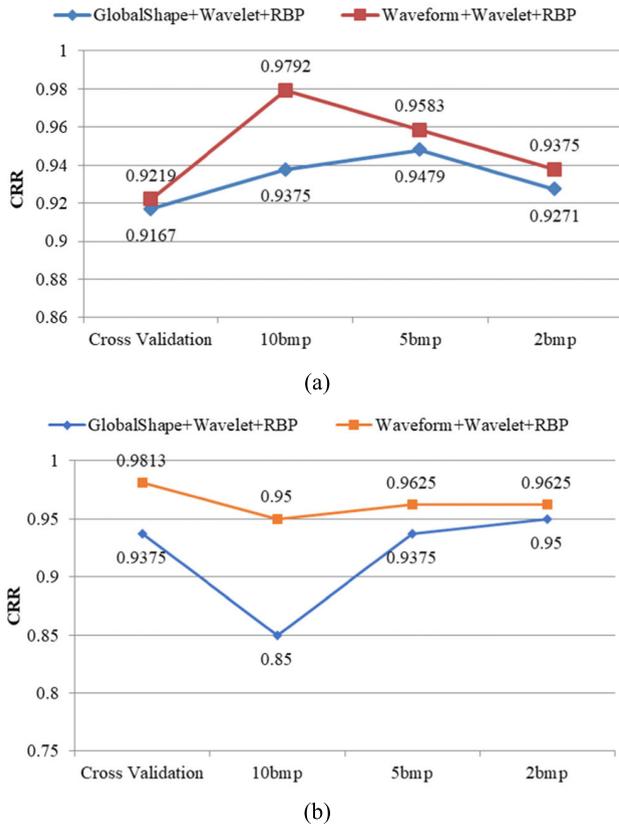


FIGURE 13. MIA comparison of the SPAT and cross validation test results: (a) self-acquisition database I; (b) self-acquisition database II.

the eigenvectors at each stage. Table 2 and Table 3 present a comparison of the experimental results.

TABLE 2. Comparison of the single identification algorithm before and after optimization using self-acquisition database 1.

Algorithms	Cross validation	10 bpm	5 bpm	2 bpm
GlobalShape	0.7604	0.8125	0.875	0.8438
RBP	0.8073	0.6667	0.6458	0.6458
Wavelet	0.7917	0.8333	0.875	0.8438
Waveform	0.7708	0.6667	0.6146	0.5833

According to Table 2, RBP and Waveform (which is susceptible to mutated waveforms) have a higher classification of single accuracy than the polymorphic average template. The reason is that cross validation has some fault tolerance concerning the mutation waveform, which eliminates the effect of the mutation waveform on the overall classification rate to a certain extent. Second, the cross validation processing of RBP and Waveform are, at most, 34% higher than the SPAT in self-acquisition database II and, at most, 14% in self-acquisition database I, which also explains why the accuracy of the cross-validation with MIA is higher than that of the SPAT with the MIA.

TABLE 3. Comparison of the single identification algorithm before and after optimization using self-acquisition database 2.

Algorithms	Cross validation	10 bpm	5 bpm	2 bpm
GlobalShape	0.7563	0.75	0.825	0.8375
RBP	0.85	0.5125	0.55	0.575
Wavelet	0.7875	0.75	0.8375	0.8625
Waveform	0.8625	0.625	0.6125	0.6875

E. MIA APPLICATION TO COMPARE SELF-ACQUISITION DATABASE AND STANDARD DATABASE

The previous experimental results show that the optimized MIA has better identification accuracy in the case of the self-acquisition database with large changes in heart rate after exercise. The standard ECG signal in the MIT-BIH ST Change database is more stable and has less noise than the waveform of the ECG signal acquired using the mobile device. We compared the proposed MIA, based on the motion problem, when applied to the standard database and when applied to the self-acquisition database. For the test group and the training group, those two kinds of databases pursued the same strategy. There were five cycles per group with a total of 20 groups, of which 16 were treated as training groups and four were treated as test groups. The scale factor training method was used to select the best threshold for all groups. The characteristics of the combination were as follows: waveform characteristics, Wavelet coefficient characteristics and RBP statistical characteristics. Fig. 14 shows the results.

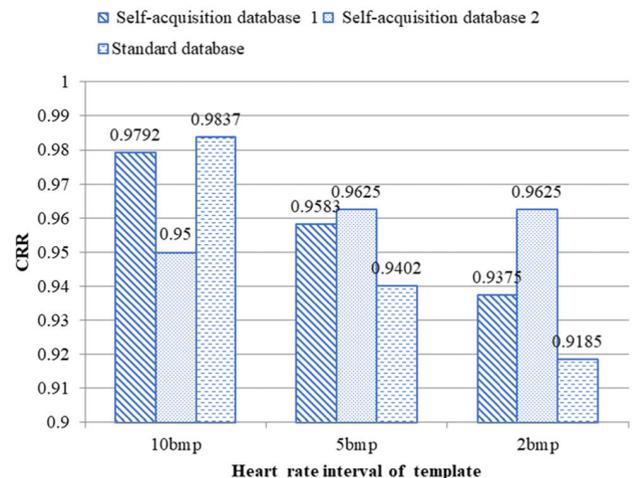


FIGURE 14. MIA application to compare the self-acquisition database and the standard database.

We can see from above Fig. 14 that, when the heart rate interval of the polymorphic template is 10 bpm, the identification accuracy of the standard database is higher than that of the self-acquisition database. However, while the heart rate interval is 5 bpm and 2 bpm, the CRR of the self-acquisition database is higher. According to the previous comparison

of experimental results, the closer the heart rate interval of the polymorphic template, the smaller the distance between templates. The MIT-BIH ST Change database has a relatively large heart rate variability compared to other standard databases. Meanwhile, ECG signals from the self-acquisition database have a much larger heart rate variability than the standard database. Hence, the *CRR* of the self-acquisition database is higher than that of the standard database when the heart rate interval is 5 bpm and 2 bpm; however, when the heart rate interval is 10 bpm, excluding the error factor of the template, the *CRR* of the self-acquisition database is lower. In summary, the proposed MIA, based on the motion problem, has strong robustness.

F. COMPARISON OF THE SINGLE IDENTIFICATION ALGORITHM AND THE *mia*

The feature extraction algorithm used in the MIA can reflect the characteristics of the ECG signal from different angles. The Wavelet coefficients reflect the characteristics of the ECG signals in different frequency domains. The waveform features reflect the time-domain space, area and slope of the ECG signal from the time domain. The RBP algorithm uses a statistical method to obtain the local detail features of the ECG waveform. A single-feature extraction algorithm can only reflect the uniqueness of the ECG signal from a certain point of view; but the MIA is like an X-ray, as it measures the feasibility of ECG identification from all angles. The aim of the so-called single classification algorithm is to select a single-feature extraction algorithm for identification. Table 4 presents the experimental results when adopting the scale factor training method, showing that the accuracy of the MIA is, at most, 31% higher than that of the single identification algorithm, while the effect is significant. The heart rate interval of 5 bpm was selected for Table 4.

TABLE 4. Comparison of the single identification algorithm and the MIA .

Algorithms	<i>FAR</i>	<i>FRR</i>	<i>CRR</i>
RBP	0.25	0.458	0.645
Global Shape	0	0.25	0.875
Wavelet	0	0.25	0.875
Waveform	0.354	0.416	0.614
GlobalShape + Wavelet + RBP	0	0.104	0.947
Waveform + Wavelet + RBP	0	0.083	0.958
		3	3

G. COMPARISON OF THE SINGLE-FEATURE AND MULTIPLE-FEATURE EXTRACTION ALGORITHMS

In this paper, a single-feature extraction algorithm is used for comparison with the proposed multi-feature extraction algorithm based on the MIA. We use Wavelet decomposition

for ECG signals, which are decomposed into high-frequency components and low-frequency components.

The low-frequency components are smoothed, while the polymorphic templates with different heart rate intervals are selected for the high-frequency components and the low-frequency components. As high-frequency components reflect the details of the characteristics, the heart rate interval needs to be larger during the choice of polymorphic template. Meanwhile, as low-frequency components reflect the overall characteristics, the selected state-of-mind template interval is smaller. The experimental results concerning the combination of different templates are shown in Fig. 15.

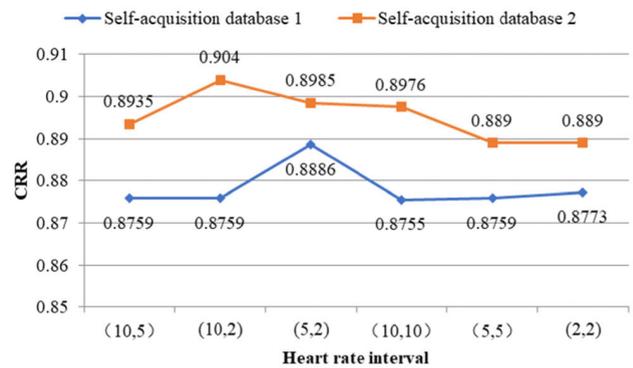


FIGURE 15. Wavelet characteristics of different heart rate intervals.

We can see from Fig. 15 that, after the template with a heart rate interval of 5 bpm was selected for high-frequency components and the template with a heart rate interval of 2 bpm was selected for low-frequency components, the identification accuracy of the self-acquisition database I was up to 88.86%, which is about 1% higher than that of other templates. Similarly, after the template with a heart rate interval of 10 bpm was selected for high-frequency components and the template with a heart rate interval of 2 bpm was selected for low-frequency components, the identification accuracy of the self-acquisition database II was up to 90.4%, which is about 1% higher than that of other templates.

Table 5 shows the comparison results of the single feature and the multiple features as the input to the MIA. The multiple-features MIA is 9% and 5% higher than the single-feature MIA in those self-acquisition databases I and II, respectively. The experiments show that, in the MIA, each stage which transforms the feature space, in order to improve the final identification accuracy, is effective.

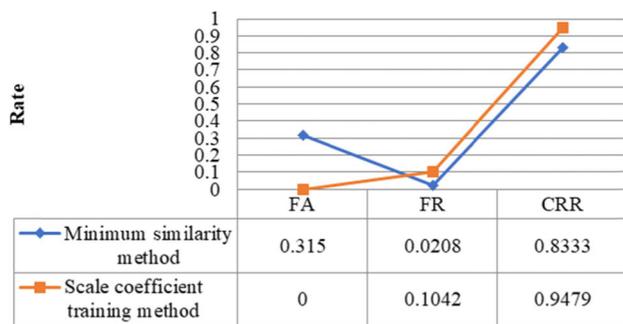
H. COMPARISON OF THRESHOLD SELECTION STRATEGIES

First, the scale factor training method and the minimum similarity method are used to select the optimal threshold for the proposed algorithm. We then compare the results of the experiment in order to analyze and summarize the applicable data and practical applications of these two strategies.

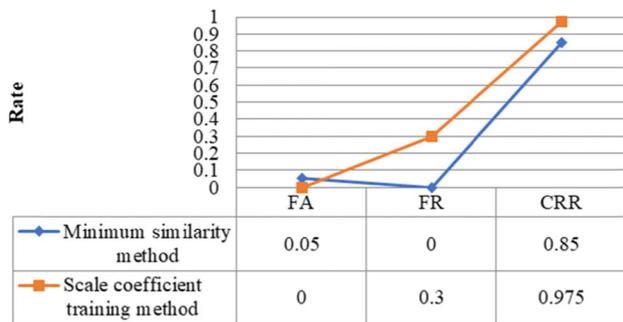
TABLE 5. Comparison of single-feature and multiple-feature MIAs .

	FAR	FRR	CRR
Single-feature MIA	0.0455	0.1773	0.8886
Multiple-features MIA	0	0.0417	0.9792

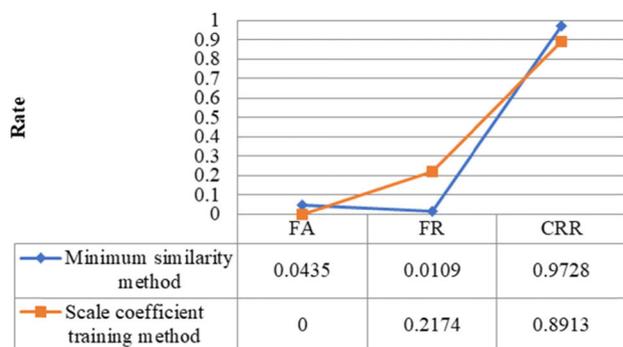
The experimental results are shown in Fig. 16: (a) and (b) present a comparison of the two threshold strategies in the self-acquisition database; (c) is the comparison of the two threshold strategies in the standard database.



(a)



(b)



(c)

FIGURE 16. Comparison of threshold selection strategies: (a) self-acquisition database I; (b) self-acquisition database II; (c) standard database.

We can observe from Fig. 16 that the scale coefficient training method is higher than the minimum similarity method in self-acquisition database I, while the minimum similarity

method is higher than the scale coefficient training method in both the standard database and self-acquisition database II. By comparison, we find that, in these two kinds of databases, the false acceptance rate of the minimum similarity method is higher than that of the scale coefficient training method, but the false rejection rate of the scale coefficient training method is higher than the minimum similarity method. This shows that, for the same feature extraction algorithm, the same data, the threshold selected by the minimum similarity is smaller than that selected by the scale coefficient training method.

We found that data in the standard database and self-acquisition database II have less noise, more stable waveforms and fewer mutation waveforms. Therefore, their own distance with their own and their own distance with others have big differences, while small thresholds can distinguish these differences. While data in self-acquisition database I have more noise, a certain degree of distortion in the waveform makes it more difficult to extract waveform features. Therefore, their own distance with their own and their own distance with others have little difference; further, it is hard to separate them by using small thresholds. Therefore, we should select the appropriate threshold selection strategy for different data characteristics and practical applications.

I. COMPARISON WITH THE AdaBoost ALGORITHM

The AdaBoost algorithm [42] is an iterative algorithm. Each iterative process will obtain a weak assumption, in which the weight of wrong identification samples will be increased. The whole weighted sample is used to train the next weak classification until the specified number of iterations or a sufficiently small error rate is reached. The core idea of the AdaBoost algorithm is to combine weak classifiers as a strong classifier. The MIA idea proposed in this paper has some similarities with it. Hence, we compare the experimental results in order to observe the advantages and disadvantages of those algorithms.

In this experiment, we take the feature vectors obtained by the proposed feature extraction algorithm as the input to the AdaBoost algorithm: $Feature = (Waveform; RBP; Wavelet)$. The ratio of the training set and the test set is the same as that of the MIA, with the number of iterations set to one, three, five, ..., 50; the results are shown in Fig. 17. Thus, when the number of iterations is greater than eight, the identification accuracy of growth tends to be gentle. Considering the time complexity, the number of iterations is set to eight.

When the number of iterations is eight, the experimental results for the self-acquisition database and the standard database are shown in Table 6. According to this table, the CRR of the MIA is 18% and 14% higher than the AdaBoost algorithm in the self-acquisition database, while the CRR of the MIA in the standard database is 13% higher than the AdaBoost algorithm. This is in line with our analysis of the two databases where the AdaBoost standard database identification rate is higher than that of the self-acquisition database. We can find an explanation for this result by

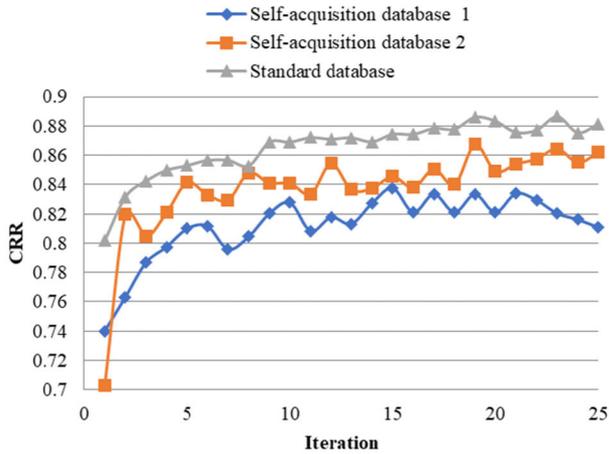


FIGURE 17. Comparison of the number of iterations.

TABLE 6. Comparison between the AdaBoost algorithm and the MIA.

	Self-acquisition database I	Self-acquisition database II	Standard database
AdaBoost algorithm	0.7975	0.817	0.8592
MIA	0.9792	0.95	0.9837

analyzing and comparing the AdaBoost algorithm and the MIA.

First, the AdaBoost algorithm is used to train each eigenvalue in the feature column, before it finds the classifier with the smallest error rate in all the feature columns, which is the optimal weak classifier of the current iteration number. In this paper, we propose an MIA to find the threshold for minimizing the sum of the false rejection rate and the false acceptance rate, which is the optimal classifier for each type of feature. Based on the characteristics of the ECG data in the self-acquisition database: the sampling frequency is large; the mutation waveform is easy to be found; the waveform characteristic accuracy is not high; and the characterization ability of the local statistical feature is also not very strong, meaning that the classifier trained by one eigenvalue compared with that trained by a class of eigenvalues is strongly one-sided.

Secondly, after each iteration in the AdaBoost algorithm, the weights of the misidentified samples are reduced, and the weights of the correctly identified samples are not changed, which is equivalent to highlighting the misidentified samples in the next round of training. In the ‘inverted pyramid’ MIA proposed in this paper, we take each stage in the wrong identification of the sample as the input stage of the next identification process. That is, the weights of the wrong identification sample are unchanged, and the weights of the corrected identification sample will be 0. This process, which filters out the correctly identified samples, when compared with AdaBoost algorithm, will highlight misidentified samples to a greater

extent, resulting in a higher accuracy for the authentication of the MIA.

Therefore, the MIA proposed in this paper is more suited than the AdaBoost algorithm to the relatively small amount of ECG signal data and the simplicity and sensitivity of the ECG signal.

J. TIME COMPLEXITY ANALYSIS AND COMPARISON

In the theoretical analysis of the MIA, we can conclude that, if time complexity and spatial complexity are not taken into account, it is easy to arbitrarily increase the number of stages in the MIA, such that the identification accuracy can reach 100%. However, in the practical application of the proposed algorithm, time complexity is a key factor to consider. To solve the problem of heart rate variability in the motion state, the time complexity of different stages in the MIA, based on the SPAT, is shown in Table 7.

TABLE 7. Time complexity of the MIA at different stages.

Stage	Cross validation	SPAT
getTemplate	—	$o(n^2)$
getDistance	$o(n^3)$	$o(n^2)$
train_K	$o(n^2)$	$o(n^2)$
train_Min	$o(n)$	$o(n)$
test_SingleStage	$o(n^2)$	$o(n^2)$
test_MultiStage	$o(n^2)$	$o(n^2)$

During the process of experiments for the MIA, we found that the consumption of time was mainly concentrated in the feature extraction and that there was similarity in the calculation stage. Time complexity was the same in the training and testing phases regardless of whether or not polymorphic templates were used in the MIA. In the training phase, each group of data for each collected individual must be compared with the threshold. In the training phase, each group of data for each collected individual must be compared with the threshold, so that we can obtain the false rejection rate and the false acceptance and correct identification rate for each collected individual. Hence, the time complexity for the training phase is $o(n^2)$. Meanwhile, in the testing phase, whether a single identification algorithm or an MIA is involved, we compare eigenvectors and optimal thresholds of the group of data for each collected individual in the test database. Hence, the time complexity of the testing phase is also $o(n^2)$. The main difference concerns whether optimization is at the stage of obtaining eigenvector similarity. The polymorphic average template is adopted for the group of data collected from everyone in order to obtain the corresponding template similarity. Hence, the corresponding time complexity is $o(n^2)$. While cross validation is adopted in order to obtain their own distance with their own and their own distance with others for the group of data collected from each individual. Hence, the corresponding time complexity is $o(n^2)$. In summary,

we can observe that the MIA with the SPAT also has some advantages in terms of time complexity.

IV. APPLICATION AND HARDWARE

A. APPLICATION

During exploratory missions, spacecraft will need to identify astronauts. ECG identification is one of the more reliable and safe identification technologies. It is reasonable to embed an ECG device into a spacesuit or provide a portable device (such as a smartwatch) with ECG identification technology to astronauts. Take a smartwatch as an example: the smartwatch is firstly used to collect ECG signals and then upload them to the spacecraft or spacesuit via a wired or wireless medium. After that, the spacecraft receives the signal and begins to process the data. Finally, the spacecraft sends the results to the smartwatch.

Spacesuits are costly high-tech products which help astronauts adapt to the space environment and accomplish a variety of space activities. Embedding ECG signal acquisition devices into spacesuits allows for the physical condition of astronauts to be checked [44], including health status and individual identification. A spacesuit embedded with ECG signal acquisition devices is shown in Fig. 18 (a).

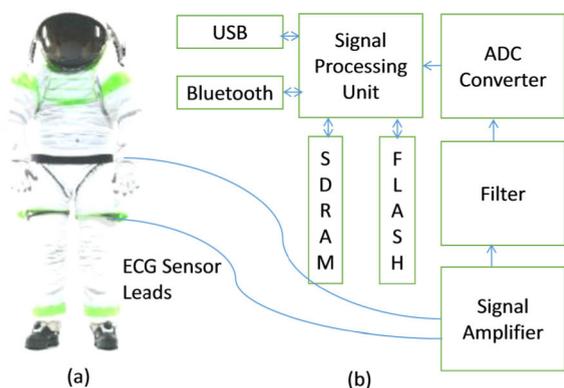


FIGURE 18. (a) A spacesuit embedded with ECG signal acquisition device; (b) a block diagram of a portable ECG device.

In a summary, the reason of ECG is suitable for spacesuits as follows:

1) The spacesuit is a critical equipment, and the tasks performed by the astronauts are also important, so it is more necessary to verify the user's identity than the general equipment.

2) Astronauts already need to monitor and diagnose their physical state in real time, so the integration of ECG identification will not bring the additional burden of users.

Of course, although our current method can be used with spacesuits, and can resist some noise through the algorithm, the problem of missing acquisition signals may occur when astronauts are in motion, which still needs to be improved. In the future, we can research on the compensation and restoration of missing signals.

B. HARDWARE

The block diagram of a portable ECG acquisition device is shown in Fig. 18 (b); the self-acquisition database was acquired by this device. First, the individual is connected to the acquisition terminal, then raw signals are collected and transferred to the signal amplifier for signal amplification. The next step is de-noising: the corresponding frequencies of noise from amplified signals are filtered out by the low-pass filter and high-pass filter. After the de-noising step, amplified signals become purer, but they are still analog signals. They are then converted into digital signals by the ADC converter. Finally, digital signals are processed by the signal processing unit and stored in flash memory. In the meantime, processed signals can be taken out from flash memory via Bluetooth or a micro USB port.

V. DISCUSSION

Numerous mature and effective algorithms have been created by scholars for ECG signal identification. The variety of proposed characteristics demonstrates the possibility that the ECG signal as a biometric representative can identify an individual's unique attributes. However, compared with the standard medical testing equipment, the accuracy of data collected by spacesuits or other wearable devices is greatly reduced. In other words, compared with the accuracy of the standard medical lead system, wearable equipment can only be designed to extend some of the patch contact with skin. This process will inevitably involve movement and loosening, resulting in the generation of a mutation waveform in the ECG data, posing great challenges for identification accuracy. Not only that, as individuals could collect their ECG data at any time, their physical and emotional conditions will also affect the collection of ECG signals, which may have a negative impact on the identification results. Therefore, this paper completed the work described below and achieved some results.

We took a variety of de-noising algorithms to remove ECG signal noise from the self-acquisition database. By comparing results, we found the best algorithms for combination to be the following: Wavelet threshold de-noising method [3], moving window median filter [31] and butterworth band-stop filter. We proposed that the SPAT could solve the problem that the heart rate variability of collected individuals is large after exercise.

We extracted the ECG signal feature by various algorithms and chose those feature combinations which reflect ECG signal characteristics from different angles. Then, we put forward an MIA. The number of identification samples per stage was controlled across a small range, and the samples which were difficult to be identified on the upper stage were identified in the next stage. In this way, the proposed architecture raised the feature identification rate. Besides, this architecture has some similarities with the AdaBoost algorithm [42], the core idea of which is to combine weak classifiers as a strong classifier; we in turn compared similarities.

According to Table 6, the *CRR* of the MIA is much higher than the AdaBoost algorithm.

There are still many problems to be studied and solved in order to improve the robustness of ECG identification. The accuracy of waveform positioning for ECG data in the acquisition database has yet to be improved. Since the data in the self-acquisition database are relatively small and lack comprehensiveness, this may hide the deep internal relations within the data. Besides, our acquisition device is very simple. More advanced devices and comprehensive data are urgent priorities for future work.

Regarding the feasibility of ECG biometric application, we point out in the ECG research article [44] that the integration of diagnosis and identification is a possible application scenario. Other recent papers [45] also integrate heart rate estimation and biometric identification. These can prove the feasibility of our ECG identification in the future.

VI. CONCLUSION

ECG technology for health monitoring embedded in astronaut spacesuits involves mature applications. Thus, ECG identification could be easily integrated with the same hardware. In this paper, we have proposed a comprehensive architecture with a selection of suitable combinations of de-noising filters and feature extraction. Moreover, a new SPAT and MIA are proposed to reduce the effect of heart rate variability and to strengthen identification robustness, based on the motion problem, respectively. Our experimental results show that our architecture improves efficacy and accuracy compared with existing algorithms. According to our experimental results, our proposed algorithm offers much better performance than the AdaBoost algorithm. In particular, we obtained high identification accuracies of 97.92% and 98.37% from the self-acquisition database and the public MIT-BIH database, respectively.

REFERENCES

- [1] S. A. Israel, J. M. Irvine, A. Cheng, M. D. Wiederhold, and B. K. Wiederhold, "ECG to identify individuals," *Pattern Recognit.*, vol. 38, no. 1, pp. 133–142, Jan. 2005.
- [2] M. A. Kabir and C. Shahnaz, "Denosing of ECG signals based on noise reduction algorithms in EMD and wavelet domains," *Biomed. Signal Process. Control*, vol. 7, no. 5, pp. 481–489, Sep. 2012.
- [3] K. Palanisamy, M. Murugappan, and S. Yaacob, "ECG signal denoising using wavelet thresholding techniques in human stress assessment," *Int. J. Elect. Eng. Inform.*, vol. 4, no. 2, pp. 306–319, Jun. 2012.
- [4] Z. Zhao, L. Yang, D. Chen, and Y. Luo, "A human ECG identification system based on ensemble empirical mode decomposition," *Sensors*, vol. 13, no. 5, pp. 6832–6864, May 2013.
- [5] J. Pan and W. J. Tompkins, "A real-time QRS detection algorithm," *IEEE Trans. Biomed. Eng.*, vol. 32, no. 3, pp. 230–236, Mar. 2007.
- [6] H. C. Chen and S. W. Chen, "A moving average based filtering system with its application to real-time QRS detection," in *Proc. Comput. Cardiol.*, Thessaloniki, Chalkidiki, Greece, Sep. 2003, pp. 585–588.
- [7] S. Mallat and W. L. Hwang, "Singularity detection and processing with wavelets," *IEEE Trans. Inf. Theory*, vol. 38, no. 2, pp. 617–643, Mar. 1992.
- [8] S. Liu, L. U. Jilai, H. Li, and H. U. Guangshu, "Detection of QRS complex using mathematical morphology and wavelet transform," *J. Tsinghua Univ.*, vol. 44, no. 6, pp. 852–855, Jun. 2004.
- [9] J. Neves Rodrigues, V. Owall, and L. Sormmo, "QRS detection for pacemakers in a noisy environment using a time lagged artificial neural network," in *Proc. IEEE Int. Symp. Circuits Syst.*, Sydney, NSW, Australia, May 2001, pp. 596–599.
- [10] J. Lee, K. L. Park, K. J. Lee, and H. R. Yoon, "A new QRS detection algorithm using resonance theory," in *Proc. IEEE 20th Annu. Int. Conf. Eng. Med. Biol. Soc.*, Hong Kong, Nov. 1998, pp. 154–155.
- [11] D. S. Benitez, P. A. Gaydecki, A. Zaidi, and A. P. Fitzpatrick, "A new QRS detection algorithm based on the Hilbert transform," in *Proc. Comput. Cardiol.*, Cambridge, MA, USA, Sep. 2000, pp. 379–382.
- [12] F. I. de Oliveira and P. U. Cortez, "A QRS detection based on Hilbert transform and wavelet bases," in *Proc. IEEE 14th Signal Process. Soc. Workshop Mach. Learn. Signal Process.*, Sao Luis, Brazil, Oct. 2004, pp. 481–489.
- [13] L. Gang, Y. Wenyu, L. Ling, Y. Qilian, and Y. Xuemin, "An artificial-intelligence approach to ECG analysis," *IEEE Eng. Med. Biol. Mag.*, vol. 19, no. 2, pp. 95–100, Mar. 2000.
- [14] C. Tu, Y. Zeng, and S. Li, "A new approach to detection of QRS complexes based on histogram," *J. Biomed. Eng.*, vol. 20, no. 2, pp. 325–327, Jun. 2003.
- [15] K. N. Plataniotis, D. Hatzinakos, and J. K. M. Lee, "ECG biometric recognition without fiducial detection," in *Proc. Biometrics Symp., Special Session Res. Biometric Consortium Conf.*, Baltimore, MD, USA, Sep. 2006, pp. 1–6.
- [16] D. P. Coutinho, A. L. N. Fred, and M. A. T. Figueiredo, "One-lead ECG-based personal identification using Ziv-Merhav cross parsing," in *Proc. 20th Int. Conf. Pattern Recognit.*, Istanbul, Turkey, Aug. 2010, pp. 3858–3861.
- [17] C. Hegde, H. R. Prabhu, D. S. Sagar, P. D. Shenoy, K. R. Venugopal, and L. M. Patnaik, "Human authentication based on ECG waves using radon transform," in *Proc. 3rd Int. Conf. Disaster Recovery Bus. Continuity*, Jeju Island, South Korea, Jan. 2010, pp. 197–206.
- [18] J. Wang, M. She, S. Nahavandi, and A. Kouzani, "Human identification from ECG signals via sparse representation of local segments," *IEEE Signal Process. Lett.*, vol. 20, no. 10, pp. 937–940, Oct. 2013.
- [19] J. Shen, S.-D. Bao, L.-C. Yang, and Y. Li, "The PLR-DTW method for ECG based biometric identification," in *Proc. IEEE 33rd Annu. Int. Conf. Eng. Med. Biol. Soc.*, Boston, MA, USA, Sep. 2011, pp. 5248–5251.
- [20] S. M. Mathews, C. Kambhamettu, and K. E. Barner, "A novel application of deep learning for single-lead ECG classification," *Comput. Biol. Med.*, vol. 99, pp. 53–62, Aug. 2018.
- [21] M. Li and S. Narayanan, "Robust ECG biometrics by fusing temporal and cepstral information," in *Proc. 20th Int. Conf. Pattern Recognit.*, Istanbul, Turkey, Aug. 2010, pp. 1326–1329.
- [22] S. Pant and A. K. Wadhvani, "A real time QRS detection system used error back propagation neural network," *Int. J. Eng. Sci. Technol.*, vol. 3, no. 9, pp. 7205–7210, Sep. 2011.
- [23] Q. Zhang, D. Zhou, and X. Zeng, "HeartID: A multiresolution convolutional neural network for ECG-based biometric human identification in smart health applications," *IEEE Access*, vol. 5, pp. 11805–11816, 2017.
- [24] M. Redmond, J. D. Polk, D. Hamilton, M. Schuette, J. Guttromson, T. Guess, and B. Smith, "Space suit electrocardiographic electrode selection: Are commercial electrodes better than the old apollo technology?" in *Proc. 76th Aerosp. Med. Assoc. Annu. Meeting*, Kansas City, MO, USA, 2005.
- [25] J. M. Charvat, S. M. C. Lee, M. L. Wear, M. B. Stenger, and M. Van Baalen, "Cardiovascular disease outcomes among the NASA astronaut corps," in *Proc. NASA Hum. Res. Program Investigators' Workshop*, Galveston, TX, USA, 2018.
- [26] E. P. McCutcheon, C. A. Berry, G. F. Kelly, R. M. Rapp, and R. Hackworth, "Physiological responses of the astronaut," *Results Second U.S. Manned Orbital Space Flight*, pp. 54–62, May 1962.
- [27] M. Choi, J. J. Jeong, S. H. Kim, and S. W. Kim, "Reduction of motion artifacts and improvement of R peak detecting accuracy using adjacent non-intrusive ECG sensors," *Sensors*, vol. 16, no. 5, p. 715, 2016.
- [28] I. D. Castro, R. Morariu, T. Torfs, C. Van Hoof, and R. Puers, "Robust wireless capacitive ECG system with adaptive signal quality and motion artifact reduction," in *Proc. IEEE Int. Symp. Med. Meas. Appl.*, May 2016, pp. 1–6.
- [29] S. Nagai, D. Anzai, and J. Wang, "Motion artefact removals for wearable ECG using stationary wavelet transform," *Healthcare Technol. Lett.*, vol. 4, no. 4, pp. 138–141, Aug. 2017.

- [30] S. Tian, J. Han, J. Yang, L. Zhou, and X. Zeng, "Motion artifact removal based on ICA for ambulatory ECG monitoring," in *Proc. IEEE Int. Conf. ASIC*, Nov. 2016, pp. 1–4.
- [31] H. Zhang, S. Zhang, Q. Jin, X. Liu, Q. Li, and J. Yang, "Motion artifact suppression in ambulatory ECG with feed forward combined adaptive filter," in *Proc. IEEE Comput. Cardiol. Conf.*, Sep. 2016, pp. 1–4.
- [32] F. Andreotti, F. Grasser, H. Malberg, and S. Zaunseder, "Non-invasive fetal eeg signal quality assessment for multichannel heart rate estimation," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 12, pp. 2793–2802, Dec. 2017.
- [33] G. Da Poian, C. J. Rozell, R. Bernardini, R. Rinaldo, and G. D. Clifford, "Matched filtering for heart rate estimation on compressive sensing ECG measurements," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 6, pp. 1349–1358, Jun. 2018.
- [34] D. Bhoraniya and R. Kher, "Machine intelligence based identification of body movements in Ambulatory ECG (A-ECG)," in *Proc. IEEE Int. Conf. Med. Imag., M-Health Emerg. Commun. Syst. (MedCom)*, Nov. 2015, pp. 80–85.
- [35] S. S. Dhillon and S. Chakrabarti, "Power line interference removal from electrocardiogram using a simplified lattice based adaptive IIR notch filter," in *Proc. IEEE 23rd Annu. Int. Conf. Eng. Med. Biol. Soc.*, Istanbul, Turkey, Oct. 2001, pp. 3407–3412.
- [36] E. Stalberg, J. Chu, V. Brill, S. Nandedkar, S. Stalberg, and M. Ericsson, "Automatic analysis of the EMG interference pattern," *Electroencephalogr. Clin. Neurophysiol.*, vol. 56, no. 6, pp. 672–681, Dec. 1983.
- [37] J. Yao and Y. Wan, "A wavelet method for biometric identification using wearable ECG sensors," in *Proc. 5th Int. Summer School Symp. Med. Devices Biosensors*, Hong Kong, Jun. 2008, pp. 297–300.
- [38] E. Plesnik, O. Malgina, J. F. Tasič, and M. Zajc, "ECG baseline drift correction through phase space for simple R-point detection," in *Proc. 25th Int. Symp. Comput.-Based Med. Syst.*, Rome, Italy, Jun. 2012, pp. 1–4.
- [39] W. Zhu and C. Qi, "An applied baseline filtering algorithm," *J. Soochow Univ.*, vol. 26, no. 1, pp. 62–64, Feb. 2006.
- [40] H. Zhang, L. Chong, and D. Guo, "A fusion of ECG signal identification method," in *Proc. 5th Int. Conf. Educ., Manage., Inf. Med.*, Shenyang, China, 2015, pp. 33–38.
- [41] K.-K. Tseng, H.-N. Huang, F. Zeng, and S.-Y. Tu, "ECG sensor card with evolving RBP algorithms for human verification," *Sensors-Basel*, vol. 15, no. 8, pp. 20730–20751, Aug. 2015.
- [42] P. Viola and M. J. Jones, "Robust real-time face detection," *Int. J. Comput. Vis.*, vol. 57, no. 2, pp. 137–154, Jan. 2004.
- [43] M. Canina, D. J. Newman, and G. L. Trotti, "Preliminary considerations for wearable sensors for astronauts in exploration scenarios," in *Proc. IEEE/EMBS 3rd Int. Summer School Med. Devices Biosensors*, Cambridge, MA, USA, Sep. 2006, pp. 16–19.
- [44] K.-K. Tseng, L. Fu, L. Liu, D. Lee, C. Wang, L. Li, and Y. Meng, "Human identification with electrocardiogram," *Enterprise Inf. Syst.*, vol. 12, no. 7, pp. 798–819, 2018.
- [45] D. Biswas, L. Everson, M. Liu, M. Panwar, B.-E. Verhoef, S. Patki, C. H. Kim, A. Acharyya, C. Van Hoof, M. Konijnenburg, and N. Van Helleputte, "CorNET: Deep learning framework for PPG-based heart rate estimation and biometric identification in ambulant environment," *IEEE Trans. Biomed. Circuits Syst.*, vol. 13, no. 2, pp. 282–291, Apr. 2019.



KUO-KUN TSENG was born in 1974. He received the Ph.D. degree in computer information and engineering from National Chiao Tung University, Taiwan, in 2006. Since 2004, he has many years of research and development experience, has been long engaged in biometric systems and algorithms research. He is currently an Associate Professor and Shenzhen Peacock B-level talent. His current research results are published in 65 articles of which about 20 is a high impact factor of the SCI magazine or the famous ACM/IEEE series of magazines.

LINLIN LIU is currently pursuing the Ph.D. degree with the Harbin University of Technology (Shenzhen). She was a Graduate Student with the Harbin University of Technology (Shenzhen). Her research interests include machine learning and bio-signal processing.

CHAO WANG is currently pursuing the master's degree with the Harbin University of Technology, Shenzhen. His research interests include deep learning algorithm and bio-signal processing.



K. L. YUNG received the B.Sc. degree in electronic engineering from Brighton University, in 1975, the M.Sc. and DIC degrees in automatic control systems from the Imperial College of Science, Technology, and Medicine, University of London, in 1976, and the Ph.D. degree in microprocessor applications in process control from Plymouth University, U.K., in 1985. He worked in U.K. for companies, such as BOC Advanced Welding Co. Ltd., the British Ever Ready Group, and the

Cranfield Unit for Precision Engineering (CUPE). In 1986, he returned to Hong Kong to join the Hong Kong Productivity Council as a Consultant and subsequently switched to academia to join the Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University. He is currently an Associate Head and a Chair Professor with the Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University. He became a Chartered Engineer and a member of IEE in 1982.



W. H. IP received the M.Sc. degree in industrial engineering from Cranfield University, the LL.B. degree (Hons.) from the University of Wolverhampton, and the Ph.D. degree from Loughborough University, U.K., and Brunel University. He has more than 30 years of experience in teaching, research, industry, and consulting. After finishing postgraduate degrees in industrial engineering in U.K., his engineering research career began at that time with novel work in industrial engineering, information, and sensor systems.



CHIH-YU HSU received the B.S. degree in mechanical engineering from National Chiao Tung University, Hsinchu, Taiwan, in 1985, and the M.S. and Ph.D. degrees in applied mathematics from National Chung Hsing University, Taichung, Taiwan, in 1993. From 2009 to 2018, he was an Associate Professor with the Department of Information and Communication Engineering, Chaoyang University of Technology, Taichung. He is currently a Full Professor with the School of Information Science and Engineering, Fujian University of Technology, Fuzhou, China. His research interests include developing innovative numerical methods and machine learning algorithm to solve partial differential equations (PDEs) and analyses algorithms for applying PDEs on various areas of engineering and image processing. He has been a member of the Phi Tau Phi Honor Society, since 1997.