

## Research Article

# Sparse Matrix for ECG Identification with Two-Lead Features

Kuo-Kun Tseng,<sup>1</sup> Jiao Luo,<sup>1</sup> Robert Hegarty,<sup>2</sup> Wenmin Wang,<sup>3</sup> and Dong Haiting<sup>1</sup>

<sup>1</sup>Shenzhen Key Laboratory of Internet Information Collaboration, Shenzhen Graduate School, Harbin Institute of Technology, Shenzhen, Guangdong 518052, China

<sup>2</sup>School of Computing and Mathematical Sciences, Liverpool John Moores University, Liverpool L3 3AF, UK

<sup>3</sup>School of Electronic and Communication, Shenzhen Graduate School, Peking University, Shenzhen, Guangdong 518052, China

Correspondence should be addressed to Wenmin Wang; wangwm@pkusz.edu.cn

Received 26 March 2014; Revised 6 October 2014; Accepted 27 October 2014

Academic Editor: Zhaojie Ju

Copyright © 2015 Kuo-Kun Tseng et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Electrocardiograph (ECG) human identification has the potential to improve biometric security. However, improvements in ECG identification and feature extraction are required. Previous work has focused on single lead ECG signals. Our work proposes a new algorithm for human identification by mapping two-lead ECG signals onto a two-dimensional matrix then employing a sparse matrix method to process the matrix. And that is the first application of sparse matrix techniques for ECG identification. Moreover, the results of our experiments demonstrate the benefits of our approach over existing methods.

## 1. Introduction

Electrocardiogram (ECG) has become a popular tool in analyzing heart disease with the use of telemedicine and home care techniques [1, 2]. However, ECG not only is useful as a diagnostic tool but also has been applied on information watermarking [3–5], data compression [3, 6], and human identification [7–15]. ECG techniques have the potential to play a role in biometric identification.

Existing biometric identification techniques have focused on the use of fingerprints, facial geometry, and voice analysis. We may be able to apply ECG techniques to protect health care systems from data leakage.

In this work, we propose a new algorithm using two leads of ECG signals for human identification. This algorithm uses the sparse matrix for dimensionality reduction that mapped two-lead data into one coordinate. We take advantage of the sparse matrix for identification. Our algorithm is sparse matrix correlation coefficient (SMCC).

Through the experiment, we demonstrate that our approach is more accurate for human identification and verification than existing techniques. In a summary, compared to the previous ECG identification, our approach has the following advantages.

(1) Using sparse matrix to store data that contains a large number of zero-valued elements can both save a significant amount of memory and speed up the processing of that data.

(2) The algorithm performs rapidly with lower computational complexity than the PCA method to process two-lead signals.

The remainder of this paper is organized as follows: Section 2 contains an overview of related work in ECG identification; Section 3 introduces the proposed ECG identification algorithm; the experimental results are presented in Section 4; finally, some concluding remarks are stated in Section 5.

## 2. Related Work

For ECG identification, research has focused on areas such as signal preprocessing, feature extraction, data classification, data reduction, and intelligence optimization.

Based on our survey, the ECG feature extraction algorithms can be classified into two categories: transform-based [9, 13, 15] and waveform-based [7, 8, 11, 12]. The transform-based algorithms consist of transforms in wavelet [15] and frequency domain. Since the wavelet transform contains information in both the time and frequency domains, it is more popular than the frequency based techniques which include Fourier transform [13] and discrete cosine transform (DCT). Waveform-based method measures the distance and amplitude difference between wave peaks and valleys. These attributes represent certain characteristics of the signal, such

as in [8]; morphological characteristics are first extracted through the wavelet transform.

But some approaches are hybrid, for example [8], using morphological characteristics which are extracted through the wavelet transform. The feature extraction of our approach could be a hybrid approach as well; it transfers two waveform signals into a two-dimensional space then measures their similarity in the two-dimensional space.

Most approaches to ECG identification only use one-lead signal [7, 10, 14]. Lead systems allow you to look at the heart from different angles. Each different angle is called a lead. The different leads can be compared to radiographs taken from different angles. So we can use more than one feature to classify discrimination. In many characteristics classification, there are many kinds of features.

Two-dimensional processing of ECG data has been applied in compression and diagnostic areas. The authors of [16] proposed diagnosis of acute myocardial infarction using two-dimensional echocardiography. The authors of [17, 18] implemented ECG data compression on two-dimensional data. These approaches have however to our knowledge not been applied to ECG identification.

In a summary, most of the work on ECG biometrics made use of only one lead and ignored the other leads that may contain additional information. The ECG signals from the two leads are essentially two observations of the same physiological activity from two different perspectives. Thus, we proposed a new two-lead algorithm for ECG identification.

Data computation is another area that we reviewed when considering existing approaches to ECG identification. A common approach is correlation of coefficients for measurements of feature distance, such as the wavelet distances that have been used in matching acquired ECG signals for identification [9, 15]. The work [19] applied the feature set evaluation (FSE) with  $k$ -nearest neighbor ( $k$ -NN) algorithm to improve low recognition rates and used the eigen-space method to reduce data dimensions; however, this approach is both complicated and time consuming. By using typical neural classifier, the research [11, 20] is applying the neural network in ECG identification.

Further, one popular approach is PCA [21] which is an analogue of the principal axes theorem in mechanics, and it was later independently developed by [22]. A recent application of PCA in ECG signal processing is useful feature reduction of various ECG properties [1, 2, 8, 23].

In this work, we adapt sparse matrix [24] for ECG identification, and it was invented as early as a century ago; CF Gauss, CGJ Jacobi, and others have studied the use of matrix sparse in some ways. Linear programming and numerical solution of boundary value problems had been apply for sparse problems in 1950s. DM Young and RS Varga on iterative research process can also be seen as the result of high-level sparse problem. But modern sparse matrix technology is mainly developed since the 1960s, and in the early and mid-60's some researchers studied the direct method as a starting point. Sparse matrix has penetrated into many areas of research. For example, in structural analysis, network theory, power distribution systems, chemical engineering, photography, surveying and mapping, and other aspects of management

science studies have appeared until hundreds of thousands of rank-order sparse matrix.

But according to our survey, we did not find any one to transform ECG signal into two-dimensional space and fuse with sparse matrix. In this research, we also found that they work well for the similarity measurement in ECG identification.

### 3. Algorithm

In this work, we target the two-lead ECG signal to be transformed into two-dimensional coordinates and perform the identification using sparse matrix. Figure 1 shows the flow of utilizing sparse matrix in ECG human identification system, which consists of three steps. First, we map the ECG two-lead signals into two-dimensional coordination that forms a matrix. Then, we reduce dimensions of the matrix using a special mask, the size of which depends on how many dimensions we want to reduce. We then transfer the matrix into a sparse matrix so that it can be stored and addressed easily. The sparse matrix is regarded as the fusion features of ECG two-lead signals. Finally, the feature data for various individuals are used to train the sparse matrix classifier. Figure 1 is the detailed formula for the procedure.

(1) *Obtaining Two-Lead ECG Signals.* Consider two-lead ECG signals, respectively, as

$$\begin{aligned} S^{(1)} &= \{x_1^{(1)}, x_2^{(1)}, \dots, x_i^{(1)}, \dots, x_n^{(1)}\}, \\ S^{(2)} &= \{x_1^{(2)}, x_2^{(2)}, \dots, x_i^{(2)}, \dots, x_n^{(2)}\}, \end{aligned} \quad (1)$$

where real-valued  $x_i^{(j)}$  ( $j = 1, 2$ ) corresponds to the  $i$ th of the  $j$ th ECG leads signal.

(2) *Transforming ECG Signals into Matrix.* Then, we initialize a matrix  $M$  to zero, and then compare each pair of consecutive input signals at time  $t$ , and set

$$M[x_t^{(1)}, x_t^{(2)}] = 1. \quad (2)$$

In this procedure, we converted each lead signals of ECG to a matrix  $M$  whose size is  $N \times N$ . Here,  $N$  is the scopes of signal values.

(3) *Reduced Matrix.* Next, we defined a mask matrix  $M^{*m \times m}$ , ( $m < N$ ),

$$M^* = \begin{bmatrix} 0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 0 \end{bmatrix}, \quad (3)$$

where  $m$  means the dimensions that we want to reduce. For  $x$  coordination  $k_x = 1, 1 + m, 1 + 2m, \dots, N - m + 1$ , and  $y$  coordination  $k_y = 1, 1 + m, 1 + 2m, \dots, N - m + 1$ , we follow the rule:

$$MR \left[ \frac{k_x + m - 1}{m}, \frac{k_y + m - 1}{m} \right]$$

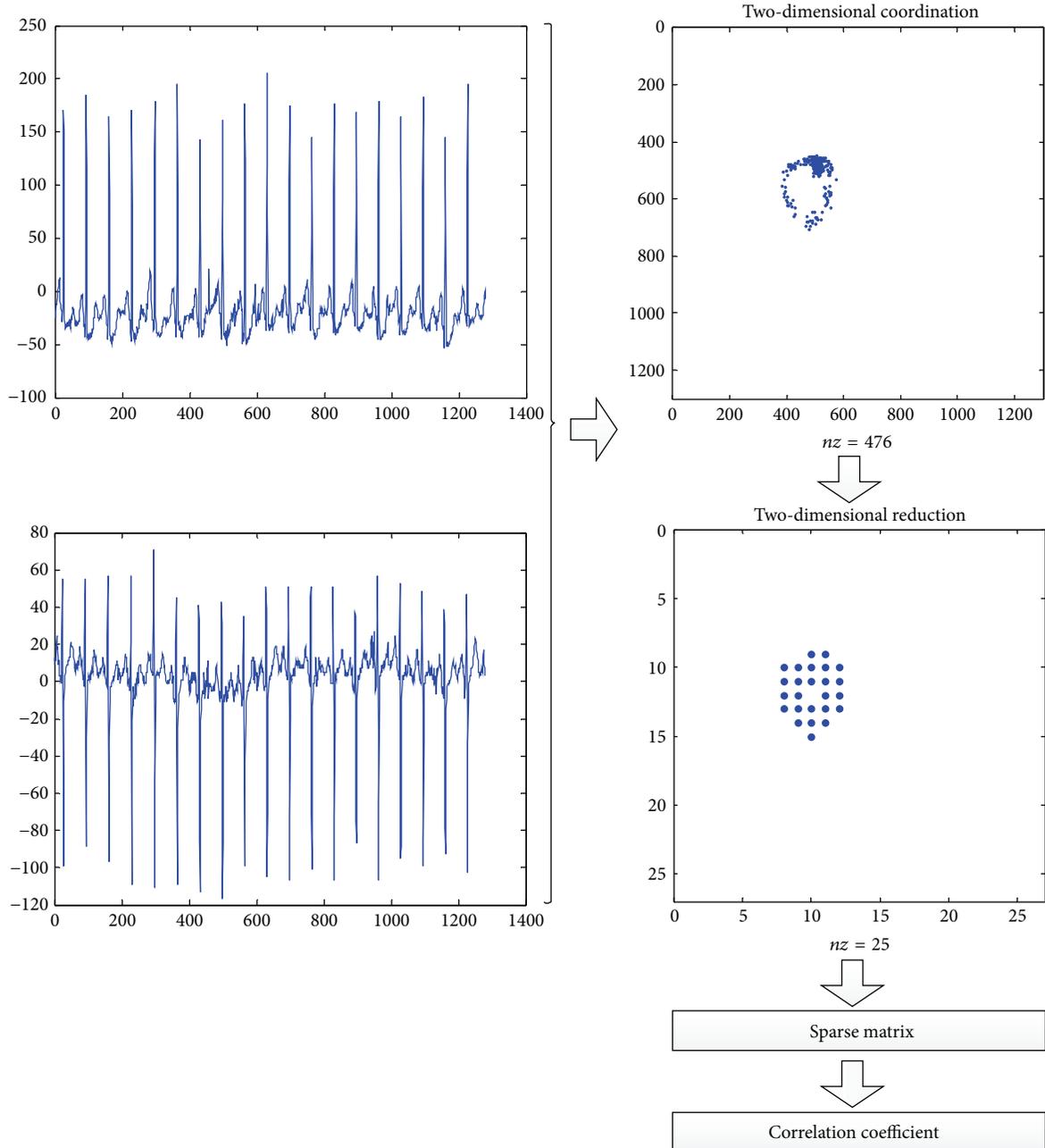


FIGURE 1: Procedure of ECG identification system with sparse matrix.

$$= \begin{cases} 0, & M[k_x : k_x + m - 1, k_y : k_y + m - 1] + M^* \leq 0, \\ 1, & M[k_x : k_x + m - 1, k_y : k_y + m - 1] + M^* > 0. \end{cases} \quad (4)$$

(4) *Storing as a Sparse Matrix.* And then we store the matrix as sparse matrix for the following processing.

Here, we use the coordinate format (COO) to store the sparse matrix; that is, we just store three parameters, row, column, and value. As most elements of the matrix are zero and we set the corresponding point as “1,” we can just only use (row, column) sparse matrix format to express the full matrix of extracted ECG two-lead signals data. For an example, after

processing, we get the ECG data that we extracted from the two-lead signals as Figure 2 shows, where *nz* means number of elements of this sparse matrix.

We can express the sparse matrices as

$$\text{val} = \{(5, 4), (6, 4), (7, 4), (5, 5), (6, 5), (7, 5), (8, 5), (5, 6), (6, 6), (7, 6)\}, \quad (5)$$

$$\text{val} = \{(6, 4), (5, 5), (6, 5), (7, 5), (5, 6), (6, 6), (7, 6)\}.$$

The relative sparse matrix *val* is considered as the features of ECG two-lead signals and will be the input to the

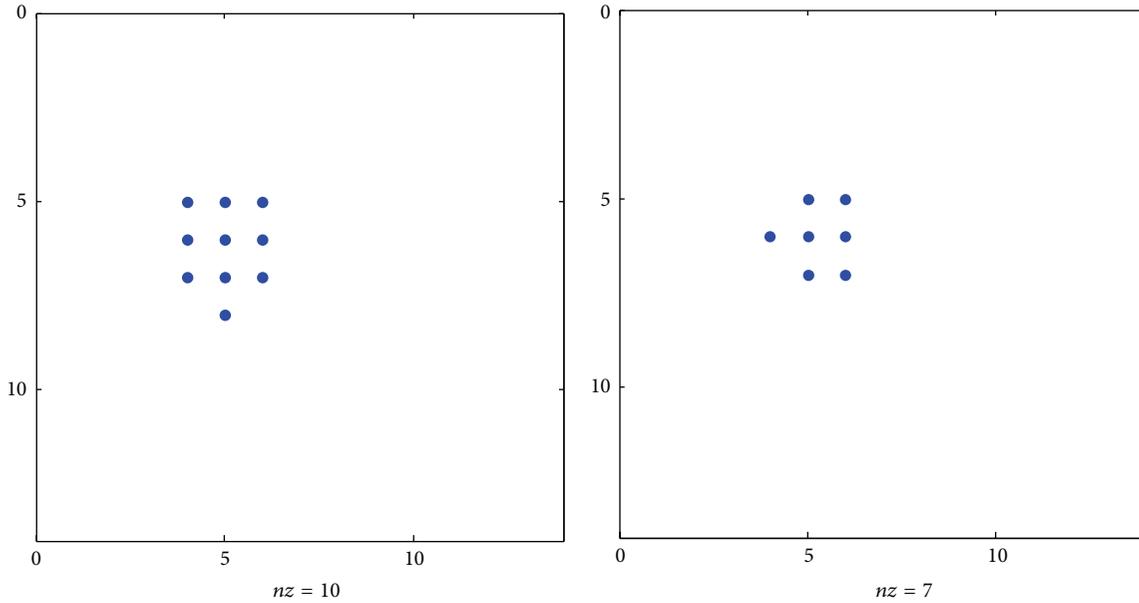


FIGURE 2: Extracted ECG two-lead signals data.

correlation coefficient classifier for training purposes and individual identification.

(5) *Computing Correlation Coefficient.*  $R = \text{corrcoef}(X)$  returns a matrix  $R$  of correlation coefficients calculated from an input matrix  $X$  whose rows are observations and whose columns are variables. The matrix  $R = \text{corrcoef}(X)$  is related to the covariance matrix  $C = \text{cov}(X)$  by

$$R(i, j) = \frac{C(i, j)}{\sqrt{C(i, i) C(j, j)}}. \quad (6)$$

$\text{cov}$  removes the mean from each column before calculating the result.

The covariance function is defined as

$$\text{cov}(\vec{x}_1, \vec{x}_2) = E[(\vec{x}_1 - \mu_1)(\vec{x}_2 - \mu_2)], \quad (7)$$

where  $\vec{x}_i$  is a vector,  $E$  is the mathematical expectation, and  $\mu_i = E\vec{x}_i$ .

For an ECG data signal  $S = \{s_1, s_2, \dots, s_n\}$ , which is from one unidentified individual. After transferring  $S$  into a sparse matrix  $\mathcal{M}^{\sim}$ , we calculated the correlation coefficient of  $\mathcal{M}^{\sim}$ ,  $\mathcal{M}_i$ , respectively, where  $\mathcal{M}_i$  is the template sparse matrix that came from a special individual, and we get  $r_i = \text{corrcoef}(\mathcal{M}^{\sim}, \mathcal{M}_i)$ . If we have  $n$  individuals, then we have an output result  $\vec{R} = \{r_1, r_2, r_3, \dots, r_n\}$ . In sparse matrix correlation coefficient classification, we define  $\varepsilon_i$  as a threshold of  $i$ th individual. If  $r_i < \varepsilon_i$ , ( $i \in 1, 2, \dots, n$ ), it means we can classify  $S$  to this corresponding target individual type  $i$ ; that is,  $S$  can be classified into the right type only if it belongs to a unique right type.

(6) *Setting Threshold for Identification.* To verify how efficient this algorithm is in human identification by ECG two-lead

signals, we train the sparse matrix which represents a segment of ECG two-lead sample points to get thresholds of each individual. In these experiments, we define the threshold  $\varepsilon_i$  of the  $i$ th individual as

$$\varepsilon_i = R_i^{\min} - \delta, \quad (8)$$

where  $R_i^{\min}$  represents the minimum correlation coefficient in training set of the  $i$ th individual.  $\delta$  is a variable that will be determined in the testing stage. To achieve the optimal identification result, we use increasing circulation  $\delta$  to test result.

## 4. Experimental Results

We conducted a comprehensive experiment on public ECG databases, and we selected the MIT-BIH normal sinus rhythm database [25]. This database includes 18 long-term ECG recordings of subjects referred to the Arrhythmia Laboratory at Boston's Beth Israel Hospital. The subjects included in this database were found to have had no significant arrhythmias. These ECG data have a sampling rate of 128 Hz and a 12-bit binary representation.

For each individual, 8 segments of 10 sample periods long are obtained from the record of the ECG signal in the database. Thus, 1280 sample points in each segment are selected for frequency and rank order statistics. We set the matrix with 1300 rows and 1300 columns for reducing dimension easily. As we know, some of the ECG sample points are negative number, so every sample point value add 500 to get non-negative number. This modification of data set can avoid the problem for mapping them into the matrix. And then mapping those sample points into a  $1300 \times 1300$  matrix. Next, we reduce the dimension of the matrix and store it as a sparse matrix.

In training stage, for each individual, there are 18 data sets for training via calculating the correlation coefficients. The process of training the neural network is from the MIT-BIH database for our sparse matrix experiment.

After training the ECG two-lead signals data with sparse matrix correlation coefficient of 18 individuals, we can obtain the threshold of each individual. The test data for identification are also acquired from the same MIT-BIH database. For each individual, we recapture 10 segments of 1280 sample points of each lead in length. Note that these 10 segments are obtained at different locations of the ECG signal; that is, none of them overlap with previously selected segments used in the training process. Each segment is passed through the sparse matrix correlation coefficient classifier for the identification matching testing. Thus, there are 10 matching tests for each individual.

**4.1. Measurement Approaches.** We used two approaches to evaluate the algorithms.

**4.1.1. Success Rate.** This is a metric used for accuracy measurement. Based on the results of comparisons between the individuals, when the correlation coefficient is smaller than the threshold correlation coefficient, we considered it as an identification error. Summing up these errors gives us the total number of errors; then we divided this figure by the total number of comparisons to give the success rate.

**4.1.2. False Acceptance (FA) and False Rejection (FR) Rates.** These are also the metrics used for accuracy performance. The FR denotes the relative ratio of subjects which should be accepted but are actually rejected by the classifier; similarly, the FA is the ratio of subjects which should be rejected but are actually accepted by the classifier. The threshold which for FA/FR is obtained from the training set, was aimed to minimize.

**4.2. Success Rate Results.** Then, we use these to compute the correlation coefficient between testing data and template matrix of each individual to classify and identify the ECG testing data. As the  $\delta$  of threshold defined is initialized by some random value, the performance might not be good enough; therefore, the classification should be trained a large number of times; for example, it is five times in our experiment.  $\delta$  is initialized to zero, and then it increases by 0.01 to test the identification results. When  $\delta$  increases to 0.20, we can find out the most appropriate  $\delta$  for classification.

We use maximum correlation coefficient as prior method to calculate success rate. For a sparse matrix  $\mathcal{M}$  which came from one unidentified individual, we compute the correlation coefficient of this sparse matrix  $\mathcal{M}$  and each individual to get corresponding  $R_i$ . Since there are 18 individuals in this experiment, we find out the  $\max R_i, i \in (1, 18)$ . And then if  $R_i > T_i$ , where  $T_i$  is the  $i$ th individual threshold that can refuse data not belonging to its own, we have  $\mathcal{M}$  which belongs to the  $i$ th individual. Else,  $\mathcal{M}$  does not belong to any individuals of those 18 individuals.

As we have 8 segments data for training, every segment can be a sparse matrix template for comparison during testing

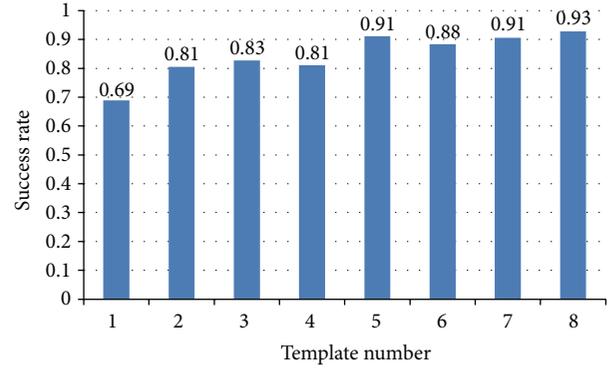


FIGURE 3: Maximum  $R$  prior success rate.

at the testing stage. Figure 3 shows the success rate by using maximum correlation coefficient prior method to identify human with each sparse matrix template.

Least square is another method for identifying human. For ECG signals which are extracted from MIT-BIH normal sinus rhythm database, we can get the correlation coefficient  $R_i$  corresponding to  $i$ th individual. After that, we compute the square  $\Delta_i$ :

$$\Delta_i = (R_i - \epsilon_i)^2, \quad (9)$$

where  $\epsilon_i$  means the average correlation coefficient of  $i$ th individual at training sets. And then we find out the minimum  $\Delta_{\min} = \min\{\Delta_i \mid i \in (1, 18)\}$ , where the  $\Delta_{\min}$  corresponding individual is  $j$ . That is,  $\Delta_{\min} = \Delta_j, j \in (1, 18)$ . As a result, this ECG signal belongs to  $j$ th individual. Figure 4 shows the success rate by using least squares correlation coefficient prior method to identify human with each template sparse matrix.

**4.2.1. FA and FR Rates Results.** Figure 5 summarizes the FA and FR change with the  $\delta$  changes. It shows that the FA/FR ratio for the matching test lies between 0.047 and 0.207 for 18 individuals with 10 segments each, which is acceptable for multiple subject classification.

The FA/FR rate of Figure 5 is compared with the 7th template of each individual. Next, we will show experiment result of different templates.

Figure 6(a) displays different FA rates of those eight templates of the sparse matrices. And we can figure out that the FA rate is smaller for all templates when the threshold  $\delta$  is small. The best is  $\delta = 0$ .

Figure 6(b) displays different FR rates of those eight templates of sparse matrices. And we can figure out that the FR rate is smaller for all templates when the threshold  $\delta$  is bigger. We have that when  $\delta \geq 0.01$ , all templates FR rate is zero.

The Acc accuracy of sparse matrix correlation coefficient algorithm can be calculated by the FA rate and FR rate. In common, the calculation function would be

$$\text{Acc} = 1 - \frac{\text{FA} + \text{FR}}{2}. \quad (10)$$

Although FA rate and FR rate have different tendency when changing threshold  $\delta$ , we might find out a suitable  $\delta$  by an iterative loop to find a better accuracy Acc.

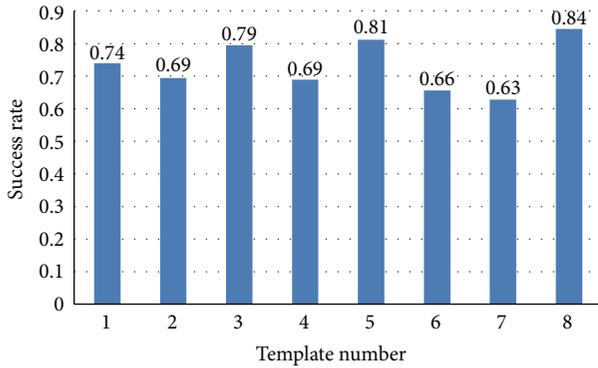


FIGURE 4: Least squares identification result.

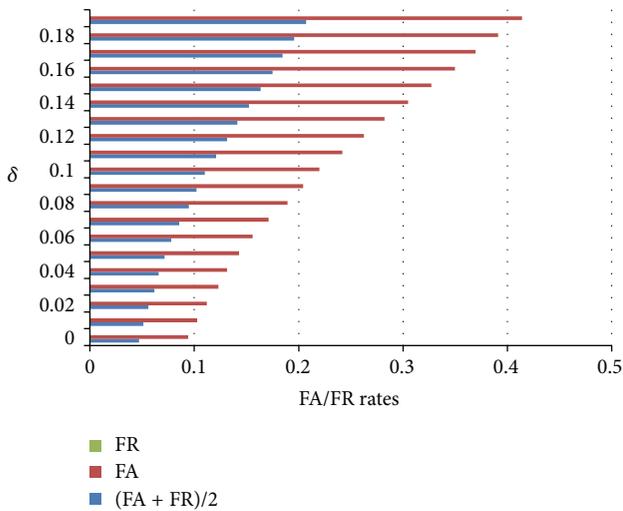


FIGURE 5: FA and FR rates.

According to formula (10), we can get higher Acc when  $(FA + FR)/2$  is smaller.

As shown in Figure 7, we have eight templates matrix for training and testing ECG data in our SMCC algorithm. When  $\delta = 0$ , the result can reach the better performance. So we choose templates  $T = 7$  and  $\delta = 0$  to show FA and FR of each individual by sparse matrix correlation coefficient algorithm here, as Figure 7 shows.

## 5. Comparison

To compare with the common one-dimensional algorithm, three ECG identification algorithms are compared in this experiment with the same database and comparison method. The four common identification algorithms are described below.

First, we list the ECG identification algorithms with single lead signal data.

**5.1. Comparing with Reduced Binary Pattern (RBP) Algorithm.** This algorithm uses the frequency and rank order statistics of the input ECG signal [26]. For any ECG signal, we can express it as  $\{x_1, x_2, x_3, \dots, x_n\}$ , where  $x_i$  represents the  $i$ th signal from the input data. According to the decrease or increase

of two consecutive  $x$  values, the two-state function,  $R_n$ , is mapped onto the values of 0 and 1, respectively:

$$R = \begin{cases} 0, & x_n \leq x_{n-1}, \\ 1, & x_n > x_{n-1}. \end{cases} \quad (11)$$

Through formula (11), the reduced binary pattern is simply represented by one binary sequence consisting of digits 0 and 1.

In  $w_k 2^m - 1 w_k S_1 S_2$  counting and ranking process, the frequency of each  $w_k$  whose value ranges between 0 and  $2^m - 1$  is calculated in the counting process. Therefore, we incorporate a weighted distance formula to define the measurement of similarity between  $S_1$  and  $S_2$ :

$$D_m(S_1, S_2) = \frac{\sum_{k=0}^{2^m-1} |R_1(w_k) - R_2(w_k)| p_1(w_k) p_2(w_k)}{(2^m - 1) \sum_{k=0}^{2^m-1} p_1(w_k) p_2(w_k)}, \quad (12)$$

where  $p_i(w_k)$  and  $R_i(w_k)$  represent the probability and ranking of  $w_k$  in the sequence  $S_i$ ,  $i = 1$  or  $2$ . The absolute difference between two rankings is multiplied by the normalized probabilities as a weighted sum; the factor  $2^m - 1$  in the denominator is to ensure all values of  $D_m$  lie within the scope of  $[0, 1]$ .

**5.2. Comparing with Waveform Algorithm.** In a waveform-based study [8], a total of 19 features are extracted from the four classes: amplitude (PQ, RQ, TQ, RT, PS, RP, TS, RS, PT, and QS), duration (QS, PR, QR, ST, and QT), slope (RS, ST, and QR), and area (area of the QRS triangle). These features form a feature vector  $S$ .

After obtaining some waveform feature for the individual difference, we use formula (12), a similarity algorithm to evaluate difference between two individuals. The closeness between two feature-vectors  $S_1$  and  $S_2$  is considered as their distance  $d(S_1, S_2)$ ; the intra- and intergroup distances can be evaluated through (12).

**5.3. Comparing with Wavelet Transform Algorithm.** Wavelet analysis or wavelet transformation is the finite or rapid attenuation of oscillation waveform signals, which is called the mother wavelet.

The procedures of the wavelet-based algorithm [9] in comparison include the following: each R-R cardiac cycle is obtained through R-R detection; an interpolation is performed on the R-R interval so each R-R cardiac cycle holds 284 data points; every R-R cycle is cut into three parts, each containing 85, 156, and 43 points; the first 85 and the last 43 points in each R-R cycle are assembled to form a 128-point segment; every four segments are grouped and an  $n$ -level discrete wavelet transform (DWT) is performed to obtain the corresponding wavelet coefficients. Four of the computed wavelet coefficients are gathered as a wavelet vector and expressed as

$$S = [a_n, d_n, d_{n-1}, d_1]. \quad (13)$$

The Euclidean distance between two wavelet vectors  $S_1$  and  $S_2$  is regarded as their distance  $d(S_1, S_2)$ ; the intra- and intergroup distances can then be calculated through (12).

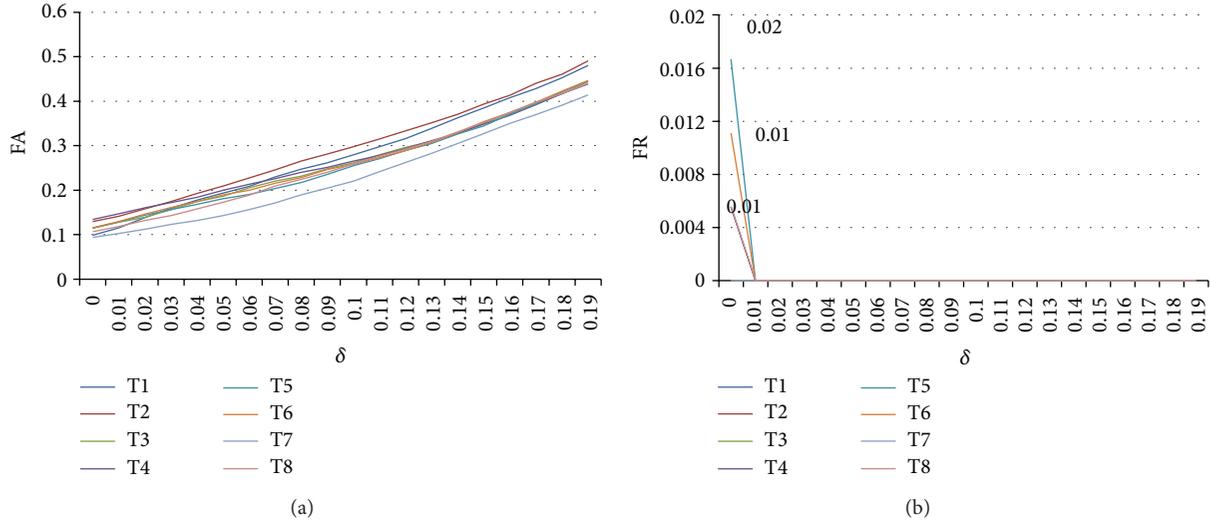


FIGURE 6: (a) FA rates of eight templates with different threshold and (b) FR rates of eight templates with different threshold.

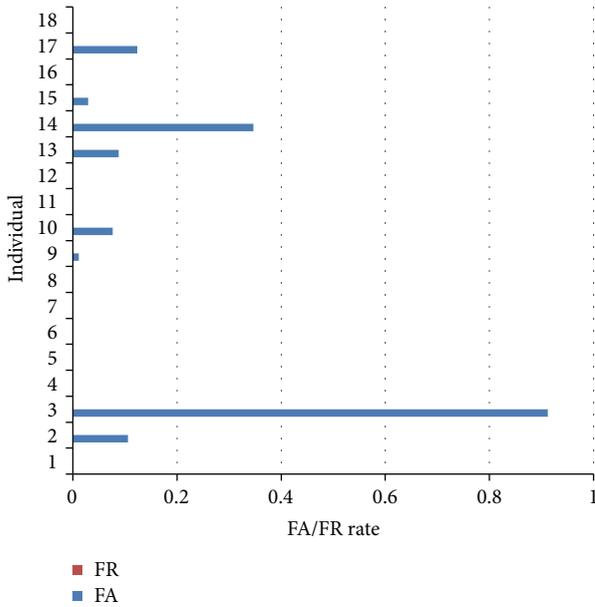


FIGURE 7: FA and FR of sparse matrix with correlation coefficient computation.

So far, we have introduced three algorithms for human identification by using one lead of ECG signal. Now we will design an experiment to conduct a comparison between the three algorithms and our sparse matrix algorithm.

In the evaluation using the MIT-BIH normal databases, it is obvious from the comparison of outcomes shown in Table 1 that the RBP, waveform-based, and wavelet transform algorithms perform well but our advanced sparse matrix with two-lead algorithm still excels them, and it has a better accuracy rate in the MIT-BIH normal public database.

5.4. *Comparison Result with One-Lead Methods.* From Table 1 we know that using two-lead ECG signal for human

TABLE 1: FA/FR of the compared result.

Item	RBP	Waveform	Wavelet	SMCC
FA	0.3748	0.3092	0.3734	0.0941
FR	0.2500	0.1222	0.0833	0
(FA + FR)/2	0.3124	0.2157	0.2283	0.0471
Accuracy	68.76%	78.43%	77.17%	95.29%

identification can enhance the identification accuracy. The result demonstrates that there is a great potential of our proposed method in the ECG biometrics system. Next, we will compare two typical two-dimension algorithms with our two-lead ECG algorithm.

5.5. *Comparing with Basic Two-Dimensional Method.* In Section 3, we described the basic flow of our sparse matrix algorithm with two-lead ECG signal. We reduce the sample points dimension directly through formula (4). Now we propose the basic method to deal with the similarity of two sparse matrices.

As we know, the baseline to measure similarity between two matrices is subtracted for two sparse matrices  $SM_1$  and  $SM_2$ , let

$$b_m = (SM_1 - SM_2)^2, \quad (14)$$

and then calculate the correlation of  $b_i$  and  $b_j$ . The following steps to calculate the FA and FR are the same as our sparse matrix.

5.6. *Comparing with PCA Method.* In this comparison, we use principal component analysis (PCA) to fusion two-lead ECG signal for identification. PCA is a statistical technique whose purpose is to condense the information of a large set of correlated variables into a few variables as principal components, while not throwing overboard the variability present in the data set [27].

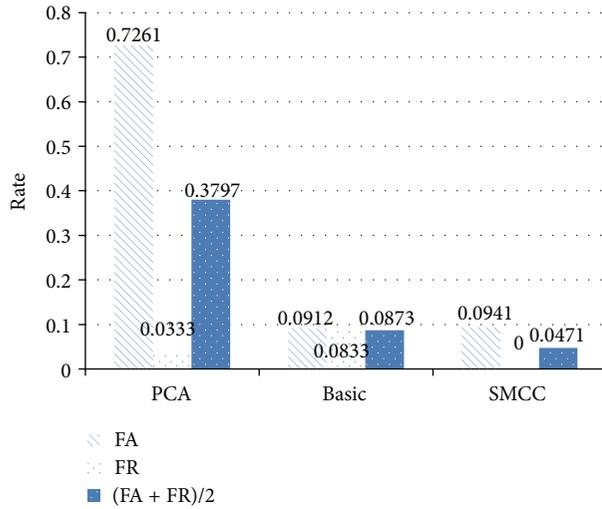


FIGURE 8: Comparison of ECG two-lead algorithms.

For a matrix  $X$  which is consisting of sample data. The linear transformation of converting  $X$  to  $Y$  is

$$Y = PX, \quad (15)$$

where  $Y$  is considered as extracting the principal components of the original matrix  $X$ . And  $P$  is a linear transformation matrix. Each row of  $P$  is the eigenvector of matrix  $C_X$ , and

$$C_X = \frac{1}{n}XX^T. \quad (16)$$

As our method, in this PCA algorithm, we read two-lead signal with the same steps. The difference is that PCA algorithm adds the second lead data behind the first lead and then use PCA method to reduce the features. Choose the main feature to train and test for human identification.

**5.7. Comparison Result with Two-Lead Methods.** From our experiment, we determined that when we choose the main 19 features, the result of identification is improved over existing approaches.

Figure 8 shows the comparison of those two algorithms and our sparse matrix algorithm.

We can get our SMCC algorithm which have the better FA/FR rate compared with other two two-lead ECG identification methods.

## 6. Conclusions

In this paper, a new ECG identification method is proposed with two-dimensional sparse matrix algorithm, in which two-lead ECG signals are fused using a sparse matrix approach. Using experimentation, we demonstrate that two-lead identification offers improvements over one-lead identification. Two typical two-dimensional classifications were compared with our method. The performance of our sparse matrix has around 95.3% accuracy which is better than basic two-dimensions 90.1% and PCA 63.2%. The results show that our sparse matrix has a good performance for ECG identification.

## Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

## References

- [1] T. Stamkopoulos, K. Diamantaras, N. Maglaveras, and M. Strintzis, "ECG analysis using nonlinear PCA neural networks for ischemia detection," *IEEE Transactions on Signal Processing*, vol. 46, no. 11, pp. 3058–3067, 1998.
- [2] H. Zhang and L.-Q. Zhang, "ECG analysis based on PCA and support vector machines," in *Proceedings of the IEEE International Conference on Neural Network and Brain*, vol. 2, 2005.
- [3] S.-G. Miaou, H.-L. Yen, and C.-L. Lin, "Wavelet-based ECG compression using dynamic vector quantization with tree codevectors in single codebook," *IEEE Transactions on Biomedical Engineering*, vol. 49, no. 7, pp. 671–680, 2002.
- [4] M. S. Nambakhsh, A. Ahmadian, M. Ghavami, R. S. Dilmaghani, and S. Karimi-Fard, "A novel blind watermarking of ECG signals on medical images using EZW algorithm," in *Proceedings of the 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS '06)*, pp. 3274–3277, New York, NY, USA, September 2006.
- [5] A. Ibaida, I. Khalil, and R. van Schyndel, "A low complexity high capacity ECG signal watermark for wearable sensor-net health monitoring system," in *Proceedings of the Computing in Cardiology (CinC '11)*, pp. 393–396, September 2011.
- [6] K. Ranjeet, A. Kumar, and R. K. Pandey, "ECG signal compression using different techniques," *Communications in Computer and Information Science*, vol. 125, pp. 231–241, 2011.
- [7] L. Biel, O. Pettersson, L. Philipson, and P. Wide, "ECG analysis: a new approach in human identification," *IEEE Transactions on Instrumentation and Measurement*, vol. 50, no. 3, pp. 808–812, 2001.
- [8] C. Ye, M. T. Coimbra, and B. V. K. V. Kumar, "Investigation of human identification using two-lead Electrocardiogram (ECG) signals," in *Proceedings of the 4th IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS '10)*, pp. 1–8, September 2010.
- [9] C.-C. Chiu, C.-M. Chuang, and C.-Y. Hsu, "Discrete wavelet transform applied on personal identity verification with ECG signal," *International Journal of Wavelets, Multiresolution and Information Processing*, vol. 7, no. 3, pp. 341–355, 2009.
- [10] T.-W. Shen and W. J. Tompkins, "Biometric statistical study of one-lead ECG features and body mass index (BMI)," in *Proceedings of the 27th Annual International Conference of the Engineering in Medicine and Biology Society (EMBS '05)*, vol. 2, pp. 1162–1165, September 2005.
- [11] Y. Wang, K. N. Plataniotis, and D. Hatzinakos, "Integrating analytic and appearance attributes for human identification from ECG signals," in *Proceedings of the Biometrics Symposium: Special Session on Research at the Biometric Consortium Conference*, pp. 1–6, 2006.
- [12] S. A. Israel, J. M. Irvine, A. Cheng, M. D. Wiederhold, and B. K. Wiederhold, "ECG to identify individuals," *Pattern Recognition*, vol. 38, no. 1, pp. 133–142, 2005.
- [13] S. Saechia, J. Koseyaporn, and P. Wardkein, "Human identification system based ECG signal," in *Proceedings of the 27th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 1–4, 2005.

- [14] H. Silva, H. Gamboa, and A. Fred, "One lead ECG based personal identification with feature subspace ensembles," in *Machine Learning and Data Mining in Pattern Recognition*, pp. 770–783, Springer, New York, NY, USA, 2007.
- [15] A. D. C. Chan, M. M. Hamdy, A. Badre, and V. Badee, "Wavelet distance measure for person identification using electrocardiograms," *IEEE Transactions on Instrumentation and Measurement*, vol. 57, no. 2, pp. 248–253, 2008.
- [16] R. S. Horowitz, J. Morganroth, C. Parrotto, C. C. Chen, J. Soffer, and F. J. Pauletto, "Immediate diagnosis of acute myocardial infarction by two-dimensional echocardiography," *Circulation*, vol. 65, no. 2, pp. 323–329, 1982.
- [17] I. M. Rezazadeh, M. H. Moradi, and A. M. Nasrabadi, "Implementing of SPIHT and sub-band energy compression (SEC) method on two-dimensional ECG compression: a novel approach," in *Proceedings of the 27th Annual International Conference of the Engineering in Medicine and Biology Society (IEEE-EMBS '05)*, pp. 3763–3766, Shanghai, China, September 2005.
- [18] I. B. Ciocoiu, "ECG signal compression using 2D wavelet foveation," *International Journal of Advanced Science and Technology*, vol. 13, pp. 576–580, 2009.
- [19] A. C.-C. Yang, S.-S. Hseu, H.-W. Yien, A. L. Goldberger, and C.-K. Peng, "Linguistic analysis of the human heartbeat using frequency and rank order statistics," *Physical Review Letters*, vol. 90, no. 10, Article ID 108103, 2003.
- [20] O. Boumbarov, Y. Velchev, and S. Sokolov, "ECG personal identification in subspaces using radial basis neural networks," in *Proceedings of the IEEE International Workshop on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS '09)*, pp. 446–451, Rende, Italy, September 2009.
- [21] K. Pearson, "LIII. On lines and planes of closest fit to systems of points in space," *Philosophical Magazine Series 6*, vol. 2, no. 11, pp. 559–572, 1901.
- [22] H. Hotelling, "Analysis of a complex of statistical variables into principal components," *Journal of Educational Psychology*, vol. 24, no. 6, pp. 417–441, 1933.
- [23] F. Castells, P. Laguna, L. Sörnmo, A. Bollmann, and J. M. Roig, "Principal component analysis in ECG signal processing," *EURASIP Journal on Advances in Signal Processing*, vol. 2007, Article ID 74580, 2007.
- [24] I. S. Duff, "A survey of sparse matrix research," *Proceedings of the IEEE*, vol. 65, no. 4, pp. 500–535, 1977.
- [25] A. L. Goldberger, L. A. Amaral, L. Glass et al., "Physiobank, physiotookit, and physionet components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. E215–E220, 2000.
- [26] F. Zeng, K.-K. Tseng, H.-N. Huang, S.-Y. Tu, and J.-S. Pan, "A new statistical-based algorithm for ECG identification," in *Proceedings of the 8th International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIHMSP '12)*, pp. 301–304, July 2012.
- [27] I. Jolliffe, *Principal Component Analysis*, Wiley Online Library, 2005.

