A Study on a Bio-signal Biometric Algorithm on the Ubiquitous Environments

Sangjoon Lee, Sung Yun Park, Sung Jae Kim, Jae Hoon Joeng, and Sung Min Kim*

Department of Medical Biotechnology, Dongguk University-Seoul, 100715 Seoul, Republic
{sangjoon,sypark,jjh,joy4110,smkim}@dongguk.edu

Abstract. This paper is about the personal identification algorithm for adapting ubiquitous environment using electrocardiogram (ECG) that has been studied by a few researchers recently. The main characteristic of proposed algorithm uses together features analysis and morphological analysis method. The Principle Component Analysis (PCA) algorithm was applied for morphological analysis method and the features analysis method adapting to Support Vector Machine (SVM) classifier algorithm. We choose 18 ECG files from MIT-BIH Normal Sinus Rhythm Database for estimating algorithm performance. The algorithm extracts 100 heartbeats from each ECG file, and use 40 heartbeats for training and 60 heartbeats for testing. The proposed algorithm shows clearly superior performance in all ECG data, amounting to 93.89% heartbeat recognition rate and 100% ECG recognition rate.

Keywords: Biometric, Ubiquitous, Bio-signal recognition, Pattern recognition.

1 Introduction

AS internet and ubiquitous technology have rapidly developed recently, there is an increasing demand for biometric technologies for information security and for the prevention of invasion of privacy. Conventional biometric systems employ the unique appearance and characteristics of the human body including fingerprints, iris, face, hand vein, and gait. Additionally, biometric systems that have been commercialized up until now have had the trend that identification is performed based on only the features appearing on a human body, and most of them employ a camera as a sensor to measure the features of a human body. The trend causes problems including an increased erroneous identification ratio due to a change in appearance of the identification target, an increased identification time due to complex biometric algorithms, an increased erroneous identification ratio due to variations in the surrounding environment, an increased number of identification targets, and the identification of duplicated images (non-living target identification). To minimize these problems, conventional biometric systems focus on increasing the identification ratio by employing two or more identification elements rather than one[1]. Up until

* Corresponding author.

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*, Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_88, © Springer-Verlag Berlin Heidelberg 2014

now in studies on ECGs and biosignals, algorithms have been developed to detect abnormal cardiac activity that accounts for 0.1% of the 24-hour cardiac activity. However, the biometric system is characterized by a method that extracts and identifies the features in the ECGs that take place in a steady state. A biometric system that uses biosignals is in its early stages of development, and there are no commercialized products yet. The initial approach to an ECG biometric system was introduced by Lena Biel [2] in 2001, and later, it was studied by a few scientists. Lena Biel measured the ECG of 20 subjects using the Megacart ECG signal measurement equipment from SIMENS and provided 30 features automatically from the ECG signal measurement equipment. Individual identification was done based on the extracted features using the SIMCA model [3], and 100% ECG signals identification ratio was obtained as a result. John M. Irvine [4] and Steven A. Israel [5] proved the validity that ECG can be used for individual identification even under a stressed heart state (from excitement or physical exercise), and designed the ECG identification method based on not the fiducial point but the ECG shape analysis. Although the recognition of the entire ECG signal had been studied even before the algorithm was proposed by John M. Irvine, the method was not appropriate for real-time identification of individuals. The morphological analysis method distinguishes the heart beat signals (the signals between RRIs) from the ECG data stream and performs identification by creating the Eigen pulse of the heart beat signals by means of PCA. The individual heart beat signals are identified from the entirety of ECG signals. Besides, Yongjin Wang [6] distinguished a few seconds of ECG signals from the entirety of the ECG signals and identified the ECG signals using the AC (Auto Correlation) and DCT (Discrete Cosine Transform) methods. Moreover, This paper proposes a hybrid type algorithm which uses morphological analysis and features analyzing method. It extracts candidate subset into the training data using morphological analysis and then compares between extracted candidate subset and testing data using features analyzing method.

2 Material and Method

2.1 Introducing the Proposed Algorithm

This paper introduces a hybrid type algorithm for ECG biometric. Previously published ECG biometric algorithms can be classified as two methods. One is the method that analyzes ECG features and the other is the morphological analysis of ECG waveform. The main characteristic of the hybrid type algorithm uses together two methods. Firstly, the morphological analysis can obtain the candidate subset among the training subset using PCA(principal component analysis). Secondly, the R-R intervalsegmentation of ECG was classified by Down Slope Tracing Waveform (DSTW) and extracted 7- features among the object heartbeat (testing subset) and candidate heartbeat (training subset), respectively. Lastly, the Support Vector Machine (SVM) classifier was performed to recognition among candidate subsets. The total block diagram of proposed algorithm is shown in Figure 1.

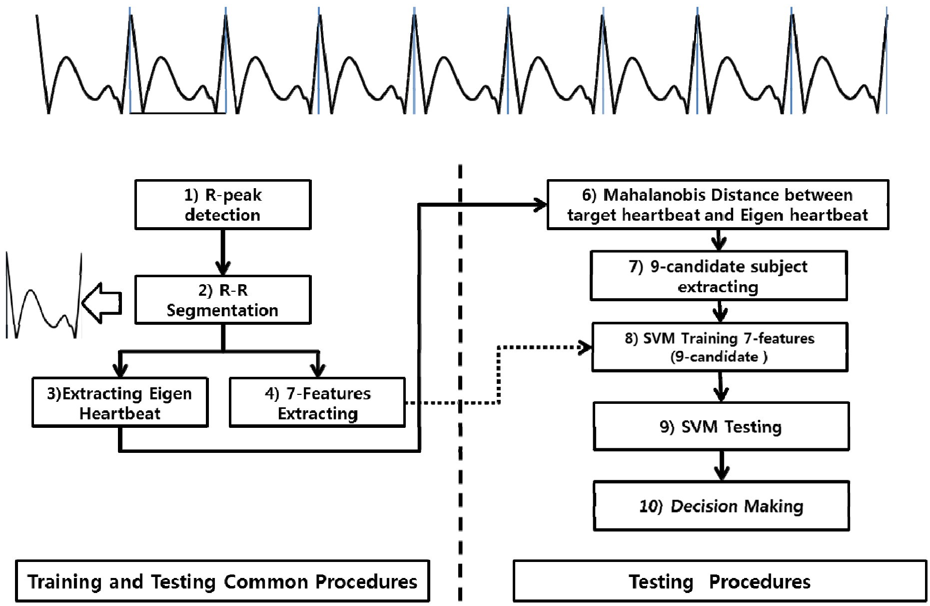

Fig. 1. The main block diagram of propose algorithm

2.2 Peak Detection and Feature Extraction

One of the most unique characteristic of ECG is its periodicity. A typical ECG signal has 3 important features, the P-wave, the QRS complex, and the T-wave. In addition, the R-peak can use a reference point to extract a periodic ECG. A DSTW (down slope trace waveform) is applied for detecting the R peaks of the ECG signal, easily detecting the peaks of the signal in real time. The DSTW algorithm can be used in various fields which detect the (P,Q,R,S,T) segments of the ECG, such as with fibrillation, reducing a power source noise, and a baseline wandering filter, showing an excellent peak detection rate in a signal with mixed amounts of various types of noise [7]. We surveyed the performance of the peak detection of the DSTW algorithm for all 48 instances in the MIT-BIH arrhythmia database. It was shown that 97.42% of the sensitivity and the 95.13% specificity could be determined. The proposed algorithm segments between current R peak and previous R peak when it finished R detecting procedures which is called RR segmentation data and then detects S, T peaks among RR segmentation data. The S, T detecting rule can be defined equation (1) and (2), respectively.

$$\mathrm{S(k)} = \mathrm{Index}(Arg_{min}[x(n)]),\ R(k) \le n \le R(k) + 30mS \tag{1}$$

$$\mathrm{T(k)} = \mathrm{Index}(Arg_{max}[x(n)]),\ S(k) \le n \le S(k) + 30ms \tag{2}$$

The k indices of detected peaks from 1 to M ($1 \le \mathrm{k} \le \mathrm{M}$), n also indices of sampled signal of $x(n)$. The actual obtained data is sampled with 125Hz.

The preprocessing procedure is finished for extracting of features by R, S, T peak detecting throughout the DSTW generation. The 7-features were chosen by detected R, S, T peaks as shown in Equation (3), (4), (5), (6), (7), (8), (9).

$$\mathrm{F_1(n) = R_{(n+1)} - R_{(n)},\ 1 \le n \le M} \tag{3}$$

$$\mathrm{F_2(n) = Arg_{max}[X(n)],\ R(n+1) \le n \le R(n)} \tag{4}$$

$$\mathrm{F_3(n) = Arg_{min}[X(n)],\ R(n+1) \le n \le R(n)} \tag{5}$$

$$\mathrm{F_4(n) = R_n - S_n, 1\ \le n \le M} \tag{6}$$

$$\mathrm{F_5(n) = S_n - R_{n+1},\ 1 \le n \le M} \tag{7}$$

$$\mathrm{F_6(n) = R_n - T_n,\ 1 \le n \le M} \tag{8}$$

$$\mathrm{F_7(n)} = {}^{R}/_{T}\ \text{, Ratio of peaks value} \tag{9}$$

2.3 Eigen Heartbeat Extracting by PCA(Principle Component Analysis)

The algorithm should be normalized using previously setting average and standard deviation for all candidate heartbeats data to have same time and amplitude bandwidth by Equation (10).

$$\mathrm{S_k = \left[\frac{1}{n}\sum_{i=1}^{n}\left(x_k^i - Xu_k\right)^2\right]^{\frac{1}{2}},\ 0 \le k \le N-1} \tag{10}$$

Where $\mathrm{Xu_k}$is average value of kth heartbeats data in the Equation (11)

$$\mathrm{Xu_k = \frac{1}{n}\sum_{i=1}^{n} x_k^i,\ 0 \le k \le N-1} \tag{11}$$

The Equation (12) is the amplitude normalization procedure.

$$\mathrm{\overline{y_k} = \frac{\left(x_k^i - Xu_k\right)\times\mu}{S_k+\sigma},\ 1 \le i \le n,\ 0 \le k \le N-1} \tag{12}$$

$$\mathrm{\overline{Y}_k = Spline(\overline{y_k}),\qquad 0 \le k \le N-1} \tag{13}$$

Where σ, μ are previously setting standard distribution and average, respectively. Where $\overline{\mathrm{y_k}}$ is normalized vector data for amplitude, which is inputted heartbeats. The heartbeat data should be normalized for time bandwidth to perform PCA because it has a different time bandwidth although it comes from a same person. The heartbeat data length is arranged 1,000 point and applied to cubic spline interpolation method[8]. Where $\overline{\mathrm{y_k}}$is normalized vector data for amplitude and time bandwidth in Equation (13). Proposed algorithm is implementing the PCA algorithm for extracting Eigen Vector. The Eigen Vector looks similar to heartbeats waveform so it is called Eigen heartbeat in the proposed algorithm. Figure 2 shows the shape of eigen heartbeats waveforms for 18 MIT-BIH normal sinus rhythm database[9]. If new heart beat come into for recognition, proposed algorithm calculates Eigen heartbeat for inputting heartbeat and compares with the distance of prepared 18-MIT-BIH database Eigen heartbeat using the Mahalanobis distance method, and then 9-candidate Eigen heartbeats were selected by descending order of shortest distance.

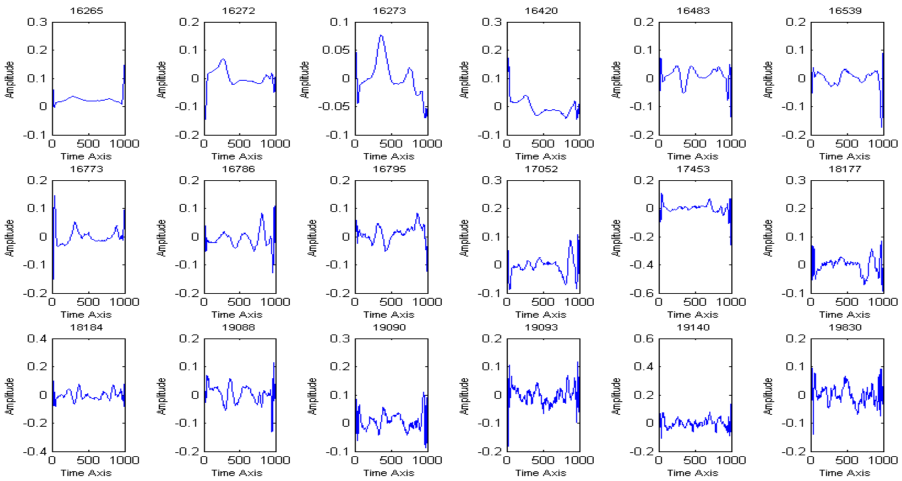

Fig. 2. Theeigen heartbeats of 18 MIT-BIH normal sinus rhythm ECG signals

2.4 Support Vector Machine (SVM)

The SVM (support vector machine) is new paradigm of pattern recognition for learning system which designed by Cortes and Vapnik[10]. The SVM was not noticed by the early, but recently has been widely used in various fields such as bio-informatics, letter, face and fingerprint recognition because of the superior performance in the supervised pattern recognition cases. The aim of SVM designs the hyperplane($w^T x + b = 0$) which having the maximum margin between decision boundary and training data and among the set of patterns that determine the boundary closer to the called support vector. The basic equation form of SVM can define the equation (14)

$$\min_{\mathbf{w},\varepsilon} \frac{1}{2}(\mathbf{w}^{\mathbf{T}}\,\mathbf{w}) + \; C \textstyle\sum_{i=1}^{l} \xi_i \tag{14}$$

Where the C is undecided constant variable, which is called normalized variable, can adjust the balance between classification error and the maximum margin.The two classes SVM case defines $y_i \in \{-1, +1\}$ but One against One SVM cases which has k classes, can define $y_i \in \{1, \dots, k\}$which can expand 2-class SVM classifier to k-class SVM classifier using simple mathematical manipulating [11]. In order to recognize for inputting heartbeat, 7-features were acquired by section 2.2 method and then the k-class SVM algorithm performs among9-candidate subset which were selected by PCA performing. The 9-candidate subsets consist of 7-features subsets of each heartbeat.

3 Experiment

In order to evaluate the performance of proposed algorithm, 100 heartbeats are classified in each 18-MIT normal sinus rhythm and extracted 7-features at the each heartbeats. The experiment method is shown Figure 1. The input parameter of One Against One SVM has C that is the maximize error rate between the hyperplane and the maximize margin, is fixed 13 and evaluating the inter-symbol heartbeats recognition performance. The training data are adjusted from 10 to 50 and surveyed recognition rate about the remaining testing data. The recognition rate calculation is performed by the equation (15).The following table.1 is the recognition rate when the training data are adjusted from 10 to 50. The average recognition rate is shown 49.2% when PCA only performing but the hybrid type ECG biometric algorithm is shown the lowest recognition rate is 88.26% in average when the training data are 20 and the highest recognition rate is 93.89% when the training data are 40.

$$\text{Accuracy}(\%) = \frac{\text{Correct Match}}{\text{Total testing number of heartbeats}} \times 100 \tag{15}$$

Table 1. heartbeats recognition rate depending on training number of heartbeats

Recognition Rate depending on the training number(%)						
Training Number		**10**	**20**	**30**	**40**	**50**
Subject	16265	96.67	98.75	98.57	100.00	100.00
	16272	92.22	71.25	95.71	100.00	96.00
	16273	66.67	72.50	75.71	80.00	88.00
	16420	76.67	73.75	92.86	96.67	96.00
	16483	96.67	97.50	95.71	98.33	98.00
	16539	64.44	65.00	78.57	80.00	86.00
	16773	90.00	90.00	91.43	93.33	92.00
	16786	98.89	98.75	98.57	98.33	92.00
	16795	97.78	98.75	98.57	98.33	98.00
	17052	96.67	95.00	90.00	93.33	94.00
	17453	93.33	92.50	91.43	91.67	90.00
	18177	91.11	92.50	80.00	100.00	100.00
	18184	78.89	80.00	84.29	80.00	78.00
	19088	94.44	95.00	91.43	90.00	90.00
	19090	98.89	98.75	97.14	98.33	98.00
	19093	97.78	97.50	98.57	98.33	98.00
	19140	88.89	95.00	95.71	95.00	94.00
	19830	72.22	76.25	100.00	98.33	98.00
Average Rate(%)		88.46	88.26	91.90	93.89	93.67

4 Discussion and Conclusion

We has been tested a new hybrid type of ECG biometric algorithm, which uses together morphology and features analysis method. The recognition performance is shown 93.89% which meaning is possible to uses commercial type biometric system if it would be solve some problems. The reason why only choosing 7-features of heartbeat which shows a good representative of person identity. In the morphological analysis, we find out a high recognition rate when 9-candidates were chose by PCA The biosignal biometric system suggested to be equally evaluated in comparison with other commercialized biometric systems, identification performance should be verified in terms of the convenience of the measurement method using a large number of subjects with a wide range of age distribution. The private service application would be rapidly increasing issue and cyber security would be more important on the ubiquitous environment. To embed the ubiquitous device, the recognition algorithm needs to fast recognition time, less than computation time and ease measurement method. From the perspective of the convenience of the measurement method, a two-lead electrode ECG measurement system[12] needs to be fabricated to measure the ECG signals in a simpler manner.

References

1. Jain, A.K., Ross, A.: Multibiometric systems. Communications of the ACM 47(1), 34 (2004)
2. Biel, L., et al.: ECG analysis: A new approach in human identification. Ieee Transactions on Instrumentation and Measurement 50(3), 808–812 (2001)
3. Esbensen, K., Schönkopf, S., Midtgaard, T.: Multivariate Anal. in Practice, 1st edn., vol. 1. Camo, Trondheim (1994)
4. Irvine, J.M., et al.: eigenPulse: Robust human identification from cardiovascular function. Pattern Recognition 41(11), 3427–3435 (2008)
5. Israel, S.A., et al.: ECG to identify individuals. Pattern Recognition 38(1), 133–142 (2005)
6. Wang, Y., et al.: Analysis of Human Electrocardiogram for Biometric Recognition. EURASIP Journal on Advances in Signal Processing 2008, 1–11 (2008)
7. Kim, J., et al.: Development of an automatic external biphasic defibrillator system. Journal of Biomedical Engineering Research 25(2), 119–127 (2004)
8. Conte, S.D., De Boor, C.: Elementary numerical analysis: an algorithmic approach, 3rd edn., vol. xiii, p. 432. McGraw-Hill, London (1980)
9. The MIT-BIH normal sinus rhythm database, Physionet, `http://www.physionet.org/hysiobank/database/nsrdb/`
10. Cortes, C., Vapnik, V.: Support-vector network. Machine Learning 20, 273–297 (1995)
11. Kreßel, U., et al.: Pairwise classification and support vector machines. In: Advances in Kernel Methods-Support Vector Learning, pp. 255–268. MIT Press, Cambridge (1999)
12. Dobrev, D., Neycheva, T., Mudrov, N.: Simple two-electrode bio signal amplifier. Medical & Biological Engineering & Computing 43, 725–730 (2005)