Biometrics Method for Human Identification Using Electrocardiogram

Yogendra Narain Singh[1] and P. Gupta[2]

[1] Institute of Engineering & Technology, Lucknow, India
singhyn@gmail.com
[2] Indian Institute of Technology Kanpur, Kanpur, India
pg@iitk.ac.in

Abstract. This work exploits the feasibility of physiological signal electrocardiogram (ECG) to aid in human identification. Signal processing methods for analysis of ECG are discussed. Using ECG signal as biometrics, a total of 19 features based on time interval, amplitudes and angles between clinically dominant fiducials are extracted from each heartbeat. A test set of 250 ECG recordings prepared from 50 subjects ECG from Physionet are evaluated on proposed identification system, designed on template matching and adaptive thresholding. The matching decisions are evaluated on the basis of correlation between features. As a result, encouraging performance is obtained, for instance, the achieved equal error rate is smaller than 1.01 and the accuracy of the system is 99%.

1 Introduction

Poets for long have been using it as a parameter to measure human emotions, where as for doctors it has been a health indicator. But now research also identifies heartbeat as a biometrics measure which can be used for human verification and identification.

Biometrics technology is an automated use of unique patterns of physiological (e.g., face, fingerprint, iris and hand-geometry) and behavioral (e.g., signature, gait) characteristics present in human beings to determine or verify the identity of individuals. Security is a prime concern of the modern society. From local in-house setting to more global scope the concerns of identity theft problems are growing in today's interconnected world. To ensure safe and secure environment, biometrics technology is now being used in many commercial, government and forensic applications. The success of using biometrics technology is that the biometrics characteristics exhibit unique patterns which are difficult to copy, share or distribute. The technology is more reliable and user friendly. However, every biometrics has its strengths and weaknesses. None of them meets the requirements of all the applications efficiently. This research shows the biometrics use of ECG can aid to more accurate identification results among all biometrics characteristics used at present.

The Electrocardiogram (ECG) is a physiological signal generated by electrical activity of heart. It is a non-invasive tool used by cardiologists to diagnose

M. Tistarelli and M.S. Nixon (Eds.): ICB 2009, LNCS 5558, pp. 1270–1279, 2009.

cardiac diseases such as to recognize electrolyte abnormalities and electrical or structural cardiac dysfunctions [1]. In the recent past, only few studies treating ECG as biometrics can be found in the literature. Biel *et al.* [2] have conducted ECG comprising biometrics experiment on 20 subjects using twelve features of their heartbeats. Shen *et al.* [3] have investigated the feasibility of ECG as a new biometrics for identity verification. They have evaluated 20 individuals on seven heartbeat features, mostly extracted from QRS complex. Template matching and decision-based neural network techniques have been used for identity verification. Issues on these studies are mainly the extraction of ECG features and accuracy, selection of consistent features and investigation of ECG to change in physiology of heart. In a recent study, Israel *et al.* [4] have focused on more physiologically found features describing the characteristic timings of ECG signal. On 29 subjects, they have investigated the influence of different anxiety states on the identification of humans by their heartbeat characteristics.

The objectives of this work are two fold: firstly, it discusses the signal preprocessing methods for the automatic delineation of ECG and secondly, it quantifies the potential of ECG as a biometrics for individual identification. ECG delineation process usually concerns signal analysis and diagnostic classification [5] that carries in a series of steps. These steps are: (1) signal acquisition, includes digitization, sampling and filtering, (2) data transformation, includes finding of complexes, (3) data representation, a process of waveform delineation and (4) feature extraction, the measurement of amplitude and interval features between different diagnostic fiducials.

Once the ECG is delineated, features of classes time interval, amplitude and angle between different diagnostic fiducials are extracted from each heartbeat. The extracted features are normalized and to make them independent from changes in heart rate. The biometrics experiment is conducted on the consistent features using proposed identification model, designed on pattern recognition and adaptive thresholding techniques.

The reminder of this paper is organized as follows. The schematic of biometrics system for individual identification on ECG is presented in Section 2. A high level summary of the techniques utilized for automatic ECG delineation to prepare ECG feature set including normalization of features are given in Section 3. A detailed description is next given for proposed identification model, in Section 4. In order to evaluate the performance of identification system, biometrics results are summarized in Section 5. Finally, concluding remarks are presented in Section 6.

2 Biometrics Method

The schematic description of individual identification system on ECG is shown in Figure 1. The method is implemented in a series of steps: starting from ECG signal acquisition, preprocessing that include correction of signal from noise and noise-artifacts, ECG delineation that include detection of waveforms and their dominant fiducials from each heartbeat, feature extraction that include

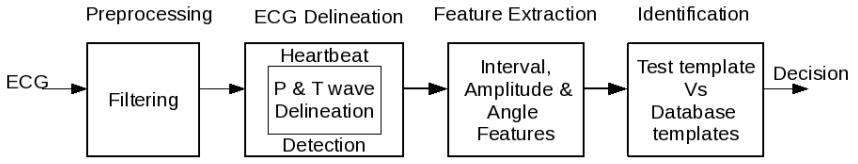

Fig. 1. Schematic of individual identification system on ECG

extraction of time interval, amplitude and angle features from dominant fiducials, followed by identification and decision making. The identification model utilizes template matching and adaptive thresholding technique which is evaluated on the basis of correlation between features.

3 Automated Feature Set Detection

The ECG data is acquired from individuals and subsequently it is digitized. The preprocessing of ECG signal involves correction from low and high frequency noises. Low frequency noise is resulted from baseline oscillations, body movements and respiration while high frequency noise is caused from power line interferences and digitization of analog potential [6]. Digital filters of linear phase characteristics are employed in this experiment. Regarding ECG delineation, it involves the detection of dominant complexes in a heartbeat such as QRS complex, P and T waves from the signal. The heartbeats are detected using a QRS complex delineator. Once a heartbeat is detected a temporal search windows are set before and after a QRS complex to search for other waveforms. The delineation of P and T waveforms are performed by their respective delineators. Found dominant fiducials are later used in preparing the feature set.

3.1 Automated ECG Delineation

QRS Complex Delineation. The QRS complex delineator is implemented using the technique proposed by Pan and Tompkins [7] with some improvements. It employs digital analysis of slope, amplitude and width information of ECG waveforms. The fiducials of QRS complex such as QRS_{onset} and QRS_{offset} are delineated according to the location and convexity of R peak.

P Wave Delineation. P wave is a low amplitude wave and has low signal to noise ratio. It shows atrium function in heart. The delineation technique utilized in this study works directly over the digitized ECG without compromising the accuracy of detected fiducials. It uses first derivative approach for P wave delineation. The first derivative, y_{nT} at time instant T is calculated using the following time difference equation,

$$y_{nT} = -2 * x_{(n-2)T} - x_{(n-1)T} + x_{(n+1)T} + 2 * x_{(n+2)T} \quad (1)$$

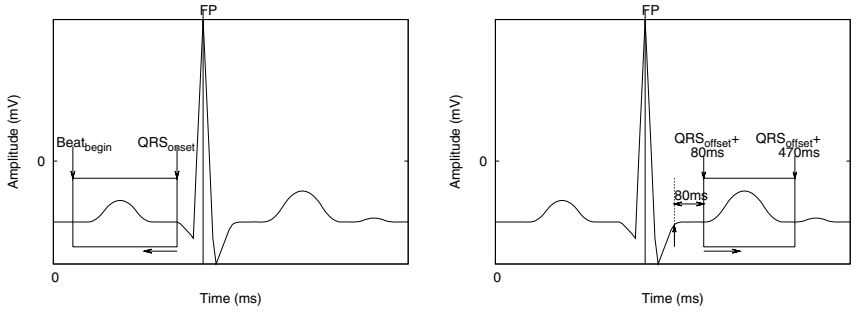

Fig. 2. Setting of search windows for: (a) P wave delineation and (b) T wave delineation

where x_{nT} represent the data sample of size n at discrete instance of time T which is set to $10ms$ at sampling rate of $100Hz$.

In order to determine P wave and its end fiducials, a search window is set prior to the beginning of QRS complex (QRS_{onset}). The search window that approximately contained P wave is set heuristically and extended from QRS_{onset} to the beginning of heartbeat ($Beat_{begin}$) as shown in Figure 2 (a). The $Beat_{begin}$ fiducial can be determined by searching of first isoelectric sample prior to the start of atrium deflection.

The detection of P wave is performed using adaptive thresholding technique. It dynamically adjusts the slope threshold with the consideration of high frequency noise present in the beat. The dominant fiducials of P wave: P_{peak} is determined by finding the local maximum in the surrounding region while P_{onset} and P_{offset} are determined by finding the local minimum at the ends within the search region. Some adjustments are needed if there are the existence of inflections near to the ends of the waveform.

T Wave Delineation. T wave is concerned to ventricles repolarization. The problem with T wave delineation is its repolarization cycle which terminates faster and comprising lower stimulation in comparison to the noise artifacts present in the beat. This makes detection of T wave end fiducial (T_{offset}) more cumbersome. In order to achieve the reliable performance of T wave delineation the signal is first corrected from oscillatory patterns of reference potential. A recursive lowpass filter is of following time difference equation,

$$y_{nT} = 2 * y_{(n-1)T} - y_{(n-2)T} + x_n - 2 * x_{(n-4)T} + x_{(n-8)T} \tag{2}$$

is utilized, where x_n represents the data sample of size n at discrete instant of time T. At sampling frequency of 100 Hz, T is found to be 10 ms and processing delay caused by the filter is nearly 30 ms.

Prior to start the delineation process, a search window is defined that approximately contained T wave. The boundaries of search window are set heuristically relative to QRS_{offset} fiducial, extended from $QRS_{offset} + 80ms$ to $QRS_{offset} +$

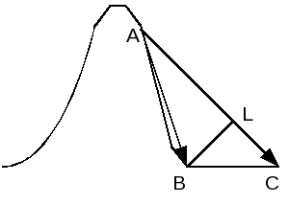

Fig. 3. Determination of T wave end fiducials

$470ms$ as shown in Figure 2 (b). It is observed from the morphology of T wave that a time segment of $80ms$ just after QRS_{offset} fiducial usually concerns the time prior to repolarization of ventricles that shows the negligible stimulation. Similarly, the right boundary of the search window is set according to the duration of depolarization to repolarization of the ventricles. The dominant fiducial corresponding to peak of T wave is determined using the technique of time derivative and adaptive thresholding, as similar to P wave with some adjustments.

The end fiducials of T wave are determined efficiently using the analysis of its waveform curvature. This analysis is based on assumption that the portion of curvature near to T wave ends is convex. The end fiducials of T wave are determined by tracking the signal downhill and finding the location of minimum radius of curvature as shown in Figure 3. By fixing of time difference between A, B and A, C the minimum radius of curvature is found by maximizing BL using vector cross product between two directed line segments i.e., $BL = \frac{|\vec{AC} \times \vec{AB}|}{|\vec{AC}|}$. Using this technique, fiducials are found more robust to local noise present in the beat.

3.2 Feature Set Preparation

In order to carry out the biometrics experiment a feature set is prepared from extracted dominants fiducials of P, QRS complex and T waves. The attributes of feature set are grouped into interval features, amplitude features and angle features. A total of 19 attributes listed in Table 1 are extracted from each heartbeat. The positions of different attributes are shown in Figure 4.

Prior to process the feature set for biometrics experiment, it is required to account for changes in these individual attributes with changes in heart rate. The heart rate varies due to changes in pressure inside heart and ventricular volume. The changes in heart rate consequently changes the duration of P wave, PR interval and QT interval. Thus, attributes related to P and T waves are normalized by dividing the beat length ($PR_I + QT_{CI}$) while the RQ and RS are used as raw attributes. Finally, normalized attributes are represented relative positions of fiducials within a heartbeat. The attributes related to peak fiducials of different waveforms are fairly invariant with changes in heart rate. As a result, raw attributes of amplitude and angle features are used in the feature set preparation.

Table 1. Considered groups of attribute features are selected from ECG dominant fiducials ($\star$: RR is used for QT interval correction)

Group Label	Features	Representation
Interval Features	PR interval	PR_I
	PR segment	PR_S
	corrected-QT interval	QT_{CI}
	ST segment	ST_S
	ST interval	ST_I
	R_{peak} to T_{onset} segment	RT_L
	R_{peak} to P_{peak} segment	RP
	R_{peak} to P_{offset} segment	RP_R
	R_{peak} to Q_{peak} segment	RQ
	R_{peak} to S_{peak} segment	RS
	R_{peak} to P_{onset} segment	RP_L
	R_{peak} to T_{offset} segment	RT_R
	RR interval*	RR
Amplitude Features	RQ amplitude	RQ_A
	RS amplitude	RS_A
	RP amplitude	RP_A
	RT amplitude	RT_A
Angle Features	Angle Q	$\angle Q$
	Angle R	$\angle R$
	Angle S	$\angle S$

4 Identification Model

In order to carry out the biometrics experiment a new identification model is proposed. The model is designed on pattern matching and adaptive thresholding technique which is evaluated on the basis of the correlation between corresponding attributes of the feature vectors. Two ECG records are declared to be matched if their feature sets are matched. Alternatively, if the correlation between corresponding attributes of the feature sets lies above than a threshold then matching decision of two records could be taken. The detection threshold for each attribute of feature set is estimated separately on the basis of interbeat correlation between corresponding attributes of different feature sets extracted from ECG recordings.

The computation procedure of threshold is as follows: Firstly, an ECG dataset (recording of T sec) of k^{th} individual is chipped into n subdata sets (recordings of S sec each, where $S < T$). From each subdata set a pattern of m features, called attributes is extracted. Let P_k be a pattern matrix of size $n \times m$.

$$P^k_{n,m} = \begin{pmatrix} a^k_{1,1} & a^k_{1,2} & \cdots & a^k_{1,m} \\ a^k_{2,1} & a^k_{2,2} & \cdots & a^k_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ a^k_{n,1} & a^k_{n,2} & \cdots & a^k_{n,m} \end{pmatrix} \tag{3}$$

where $a^k_{i,j}$ represents the j^{th} attribute of i^{th} subdata set of an individual k.

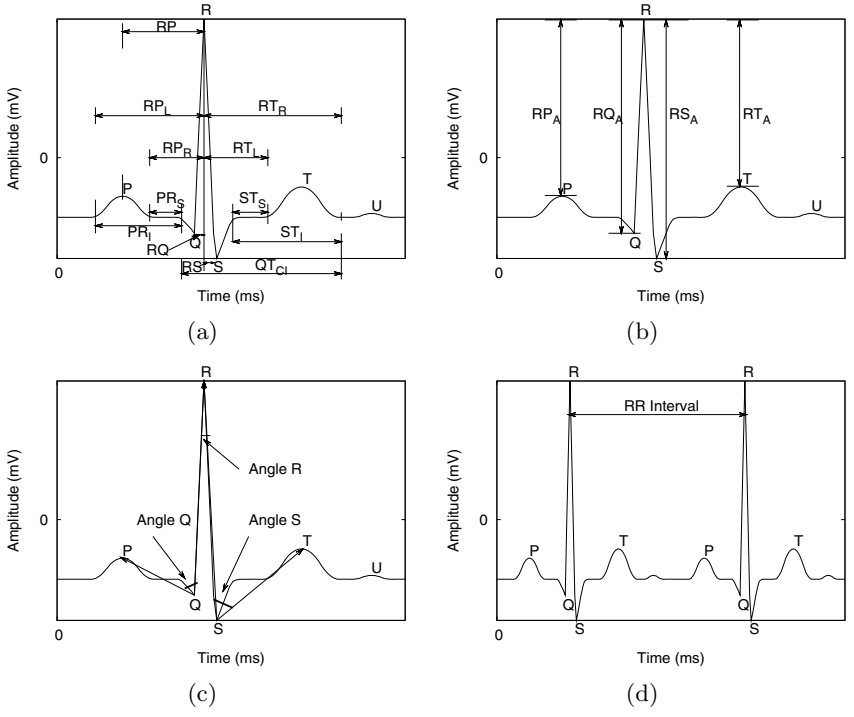

Fig. 4. Considered feature positions of different classes: (a) Time interval, (b) Amplitude and (c) Angles. The RR interval shown in (d) is used for computing QT_I using Bazett's formula [8].

The purpose of ECG data set partitioning is to measure the correlation between corresponding collection of attributes of different ECGs. The degree of relationship between attributes of feature vectors are evaluated on the basis of Pearson product-moment correlation coefficient which is computed as follows,

$$R = \left(\left(r_{X_1^i X_1^j}\right)\left(r_{X_2^i X_2^j}\right)\cdots\left(r_{X_m^i X_m^j}\right)\right) \tag{4}$$

where, $X_1^i = [a_{1,1}^i, a_{2,1}^i, \cdots a_{n,1}^i]^T$; $X_2^i = [a_{1,2}^i, a_{2,2}^i, \cdots a_{n,2}^i]^T \cdots X_m^i = [a_{1,m}^i, a_{2,m}^i, \cdots a_{n,m}^i]^T$. These are the column vectors of the pattern matrix for an individual i for $1 \leq i \leq k$. The correlation coefficient $r_{X_1^i X_1^j}$ is calculated using the formula,

$$r_{X_1^i X_1^j} = \frac{n\sum X_1^i X_1^j - \sum X_1^i \sum X_1^j}{\sqrt{n\sum {X_1^i}^2 - (\sum X_1^i)^2}\ \sqrt{n\sum {X_1^j}^2 - (\sum X_1^j)^2}}. \tag{5}$$

The correlation matrix of size $k \times k$ is generated for each column of the matrix R. The maximum is selected from each correlation matrix which shows the optimum degree of relationship between the attributes of feature vectors of different ECGs. Let $\delta_l = Max(r_{X_l^i X_l^j})$ for $l = 1, 2, \ldots, m$, be the maximum

correlation coefficient computed for each column l of matrix R. Then, maximum of correlation for all attributes are obtained as follows,

$$[\delta] = \left(\delta_1\ \delta_2 \cdots\ \delta_m\right) \tag{6}$$

The detection threshold for each attribute of the feature set is estimated on the computed value of corresponding δ. This model estimates detection threshold according to the change in ECG. The power of the proposed model is that all attributes of feature vector participate equally for deciding the threshold level. In decision making, one can set the decision limit on the number of matched attributes according to the detected threshold.

5 Biometrics Results

The performance of identification system on ECG is evaluated on Physionet QT database [9]. The database provides an evaluation tool for validation of experiments based on ECG delineation. The experiment is conducted on 50 first lead individual ECG recordings of the database. Among them 32 recordings are taken from MIT-BIH Arrhythmia database that contains the ECG of mostly inpatient men and women of age between 47 to 84 years. Remaining, 18 recordings are taken from MIT-BIH Normal Sinus database of subjects age between 20 to 50 years those have no significant arrhythmias. During experiment first 8 seconds of each recording is used for training and setting of system parameters.

Table 2. Performance of identification system over test database of size 50 at different threshold levels [GMR is the genuine matching rate, computed as $GMR(\%) = 100 - FNMR(\%)$]

Threshold Level	# FNM	FNMR (%)	GMR (%)	# FM	FMR (%)	Accuracy (%)
14	0	0	100	2	4	98
15	0	0	100	1	2	99
16	0	0	100	1	2	99
17	1	2	98	0	0	99
18	2	4	96	0	0	98
19	2	4	96	0	0	98

Biometrics performance of the identification system is measured on the parameters of false matching rate (FMR) and false non-matching rate ($FNMR$). From the correlation FMR and $FNMR$ are calculated for different threshold values to generate the detection error tradeoff (DET) curve of the identification process. The accuracy (Acc) of the system is also determined using FMR and $FNMR$ as, $Acc(\%) = 100 - \left(\frac{FMR+FNMR}{2}\right)$.

During experiment, five subdata sets of nearly 100 beats are selected randomly from each individual ECG and the feature sets are generated. Thus, a

total of 250 (50×5) samples are prepared in the database for training the identification system. The test of positive identification is conducted on different data samples of an individual ECG recording while the test of negative identification is conducted on data samples of different ECG recordings. The ECG recordings of European ST-T database are used for testing of negative identification.

In order to test the identity of an individual, a test template which is a collection of feature sets consist of five subdata sets of nearly 100 beats, is compared with templates stored in database, e.g. *one-to-many* matching. The decision of best match is taken on the basis of threshold whose level can be set on the number of attributes matched. For example, a test template can be declared matched with its counterpart stored in database if 17 out of 19 attributes are matched. The performance of identification system is examined on different settings of threshold level of attributes between 14 to 19 and results are given in Table 2.

It shows that at threshold level 15, only one individual (out of 50) is matched wrongly while none of them is non-matched wrongly. Alternatively, the system achieved $GMR = 100\%$ at $FMR = 2\%$. At threshold level 17, no one is matched wrongly while only one individual is non-matched wrongly. Thus, 99% of genuine subjects are matched correctly while none of them is forged to system. Therefore, system achieved the accuracy of 99% at this setting of threshold.

The detection error tradeoff (DET) curve of the identification process is shown in Figure 5. It is worthwhile to mention that an equal error rate (EER) smaller than 1.01 is obtained. The obtained results are encouraging and demonstrate the potential of heartbeat signals for human identification.

The concern of identification system on ECG is the changes of ECG dominant fiducials during aging. Study shows that these changes are only seen upto the age of adolescence (∼14 years) [10]. After adolescence the ECG dominant features are relatively consistent. This may be a minor concern because biometrics applications are mainly employed to identify the individuals those have passed the adolescence age.

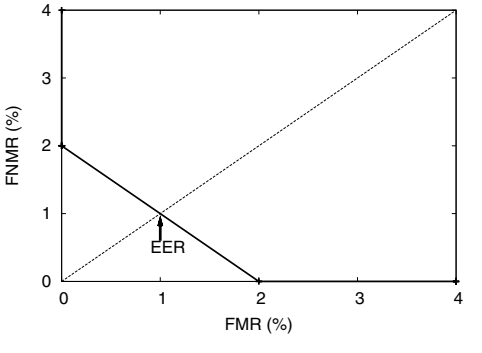

Fig. 5. Detection error trade-off curve (DET) for identification of individuals using ECG. The location of equal error rate (ERR) is indicated.

6 Conclusions

In this paper, it has shown that dominant fiducials delineated from ECG recording exhibits features that are unique to an individual. A series of experiments have been conducted for individual identification on ECG on bench mark database. The database consists of normal and inpatient men and women of age 50 ± 23 years. Biometrics results have shown that the ECG features are useful for distinguishing different individuals. From these observations it has been concluded that each individual ECG has a unique set of heartbeat features that hold sufficient information which can be used as potential biometrics for individual identification.

References

1. Kligfield, P.: The Centennial of the Einthoven Electrocardiogram. Journal of Electrocardiology 35, 123–129 (2002)
2. Biel, L., Pettersson, O., Lennart, P., Peter, W.: ECG Analysis: A New Approach in Human Identification. IEEE Transaction on Instrumentation and Measurement 50(3), 808–812 (2001)
3. Shen, T.W., Tompkins, W.J.: One-Lead ECG for Identity Verification. In: Proceedings of the Second Joint EMBS/BMES Conference, pp. 62–63 (2002)
4. Israel, S.A., Irvine, J.M., Andrew, C., Mark, D.W., Brenda, K.W.: ECG to Identify Individuals. Pattern Recognition 38(1), 133–142 (2005)
5. Kors, J.A., Bemmel, J.H., Zywietz, C.: Signal analysis for ECG interpretation. Methods Inf. Med. 29(4), 317–329 (1990)
6. Friesen, G.M., Thomas, C.J., Manal, A.J., Stanford, L.Y., Stephen, R.Q., Troy, N.: A Comparison of the Noise Sensitivity of Nine QRS Detection Algorithms. IEEE Transaction on Biomedical Engineering 37(1), 85–98 (1990)
7. Pan, J., Tompkins, W.J.: A Real Time QRS Detection Algorithm. IEEE Transactions on Biomedical Engineering 33(3), 230–236 (1985)
8. Bazett, H.C.: An Analysis of the Time-Relations of Electrocardiograms. Heart 7, 353–370 (1920)
9. Laguna, P., Mark, R.G., Goldberger, A.L., Moody, G.B.: A Database for Evaluation of Algorithms for Measurement of QT and Other Waveform Intervals in the ECG. In: Computers in Cardiology, pp. 673–676 (1997)
10. Rijnbeek, P.R., Witsenburg, M., Schrama, E., Hess, J., Kors, J.A.: New Normal Limits for the Pediatric Electrocardiogram. European Heart Journal 22, 702–711 (2001)