

Biometric Identification System Based on Electrocardiogram Data

Y. Gahi¹, M. Lamrani¹, A. Zoglat¹, M. Guennoun², B. Kapralos², K. El-Khatib²

¹Département Math-Info, Faculté des Sciences de Rabat, 4 Avenue Ibn Battouta B.P. 1014 RP, Rabat, Maroc

²University of Ontario Institute of Technology, 2000 Simcoe Street North, Oshawa, Ontario, Canada. L1H 7K4.

Recent advancements in computing and digital signal processing technologies have made automated identification of people based on their biological, physiological, or behavioral traits a feasible approach for access control. The wide variety of available technologies has also increased the number of traits and features that can be collected and used to more accurately identify people. Systems that use biological, physiological, or behavioral trait to grant access to resources are called biometric systems. In this paper we present a biometric identification system based on the Electrocardiogram (ECG) signal. The system extracts 24 temporal and amplitude features from an ECG signal and after processing, reduces the set of features to the nine most relevant features. Preliminary experimental results indicate that the system is accurate and robust and can achieve a 100% identification rate with the reduced set of features.

Index Terms—Biometric Identification, ECG, Feature Selection, Mahalanobis Classifier

I. INTRODUCTION

BIOMETRICS is a promising field of technology that is based on measuring some physical, biological, or behavioral characteristics of a person in order to establish identification, to perform identity verification, or to automatically recognize a person through [1]. Recent advances in the computing, signal processing, and medical fields has allowed biometric systems to uniquely identify humans in a computationally feasible manner. Biometric systems can rely on a number of features including fingerprints, voice, face, hand geometry and iris.

Traditionally, password and ID cards have been commonly used as defense measures to access information systems. However, password mechanisms have, on many occasions, proven to be ineffective in providing security protection as they are vulnerable to attack. To overcome some of the limitations variable password policies have been introduced. Such policies include the use of longer passwords, making a mix of lower and upper-case letters, and special characters mandatory when choosing a password, and periodic scheduled password changes. Despite these policies, passwords can still be cracked. As with the use of passwords, ID cards have their share of problems. Specifically, ID cards can be misplaced, shared, stolen, and they can be fraudulently reproduced.

Recently, researchers have begun investigating the use of the heart electrocardiogram (ECG) signal as a biometric trait to identify individuals. The ECG records the electrical activity of the heart over a period of time providing an indication of the overall rhythm of the heart. In contrast to other human biometric traits, the characteristics of the ECG signal can be unique and are difficult to falsify. In addition, ECG data is already available from patients that are connected to an electrocardiogram device, and has used to accurately identify hospital patients awaiting surgery in order to ensure the correct surgical operation is performed on them [2,3]. Furthermore, ECGs can also be used to detect the liveness of the user (e.g., whether or not the user is breathing properly), providing the potential to warn them of a potential medical

problem. The validity of using ECG for biometric recognition is supported by the fact that the physiological and geometrical differences of the heart in different individuals display certain uniqueness in their ECG signals [4]. Studies have shown the stability and distinctiveness of ECG as a biometric trait [5,6].

An ECG based biometric recognition system can be applied in a wide variety of applications including physical access control, medical records management, in addition to government and forensic applications. In this paper, we present an efficient and scalable identification system based on the ECG. The system extracts 24 temporal and amplitude features from an ECG signal and after processing, reduces the set of features to the nine most relevant features which are used to uniquely identify individuals in an efficient manner..

The remainder of the paper is organized as follows: In Section II, previous work involving the use of the ECG within a biometric system is provided. A description regarding the ECG-based biometric system proposed in this paper is provided in Section III while a detailed description regarding the features used to identify the ECG of an individual is provided in Section IV. Experimental results are provided in Section V and finally, conclusions and plans for future work are provided in Section VI.

II. IDENTIFICATION SYSTEMS BASED ON ECG SIGNAL

To date, the use of ECG signals within a biometric system to identify individuals is sparse and preliminary. Signal processing techniques are used to extract the characteristic ECG features of an individual. Existing approaches for biometric recognition from electrocardiogram (ECG) signals are based on temporal and amplitude distances between detected fiducial points. Shen *et al.* [7] extracted seven temporal and amplitude features from the QRST wave. They combined a template matching method with decision based neural network (DBNN) to implement an identity verification system. The template-matching method is first applied to calculate the correlation coefficient for comparison of two QRS complexes. The DBNN is fed with the possible candidates resulting from the first step. Experimental results using this combined method on 20 participants yielded a 100% identification rate. They later performed more extensive

testing by using a larger database, containing 168 individuals and 17 temporal and amplitude features and achieved an identification rate of 95% [8].

Biel *et al.* [9], used specialized hardware equipment to extract 30 ECG features. A simple feature selection algorithm based on the analysis of a correlation matrix was employed to reduce the dimensionality of features to 21. Further reduction of the feature set was based on experimental results. The soft independent modeling of class analogy (SIMCA) method based on principal component analysis (PCA) was used for classification. Experimental results with 20 participants yielded identification rates between 95-100% using empirically selected features. A major drawback of Biel *et al.*'s method is the high number of features used to identify a person given the direct relationship between the number of features and computational requirements (time and potentially memory) This limits the scope of applications where it can be employed to small databases. Furthermore, the best set of features for their observed high identification rates were found experimentally and not proven using a feature selection algorithm.

Wang *et al.* [10], combined analytic and appearance based features to achieve high recognition accuracy. Twenty one (21) analytic features were extracted using known fiducial detectors [11,12]. The appearance features were extracted using unsupervised learning techniques based on principal component analysis (PCA) and linear discriminant analysis (LDA). This work focused more on the extraction of features themselves rather on reducing the dimensionality of the features. Israel *et al.* [11] extracted a total number of 15 features, which are time duration between detected fiducial points. The number of features was reduced to 12 during the feature selection process. Feature selection was provided by a stepwise canonical correlation that used the Wilkes' lambda method and linear discriminant analysis for classification. This system was tested on a database of 29 subjects and yielded a 100% human identification rate and a heartbeat recognition rate of 81%.

Shen proposed a four stage algorithm to select the good features by using the Quartile Discriminant Measurement (QDM) and a correlation matrix [8]. The process starts with the calculation of the correlation matrix between features. Then the computed correlation matrix is used to separate the features into groups, so each group contains highly correlated features. Next, features are ranked from low to high based on a score assigned by the QDM measure. Last, low ranked features are removed from groups containing more than one feature.

III. IDENTIFICATION SYSTEM DESCRIPTION

An identification system based on ECG can be composed of two phases. The *enrollment* phase and *identification* phase. During the enrollment phase, the ECG characteristics of an individual are first captured by a wireless heart monitor. The system then produces a digital representation (template) of the ECG characteristics. A quality check is performed to ensure that the acquired sample can be reliably processed by successive stages. The template is then stored in the central database of the biometric system. In the identification phase,

the extracted features are compared against the stored templates to generate matching scores in order to establish the identity of the individual.

Our identification system follows the generic architecture proposed in [13]. It's composed of the following four modules:

Sensor: captures the ECG data of the user. A 4th order low pass Bessel filter with a cut-off frequency of 30.00 Hz is applied to the raw data to perform noise reduction.

Feature Extractor: processes the ECG signal to extract 24 discriminatory features from the ECG data. It's based on a fiducial detector that extracts the P, PQ, QRS, QT, T and RR time intervals as well as the amplitudes of P, R and T fiducials (see Figure 1). A thresholding method is then applied to remove the outliers that are not appropriate for training and classification. Features that are outside the range of the interval $[m-\sigma, m+\sigma]$ where m is the mean and σ is the variance, are removed. Experimental results indicated that temporal features can vary significantly for the same person.

Matcher: encapsulates a decision making module that establishes the identity of an individual. Mahalanobis distance-based classifier is used to identify the individuals. First, the selected features are compared against the stored templates to generate a matching score (which is initially set to zero). For each heartbeat, the Mahalanobis distance between it and the set of samples (templates) stored in the system database is computed. The template resulting in the smallest distance is considered to be a match and hence results in the score being incremented by one. This process is repeated for 150 heartbeats (approximately two minutes). The template (individual) with the highest score is considered to be a match.

System Database: used to store the biometric templates of registered users. In our system, we store the covariance matrices generated in the enrolment phase.

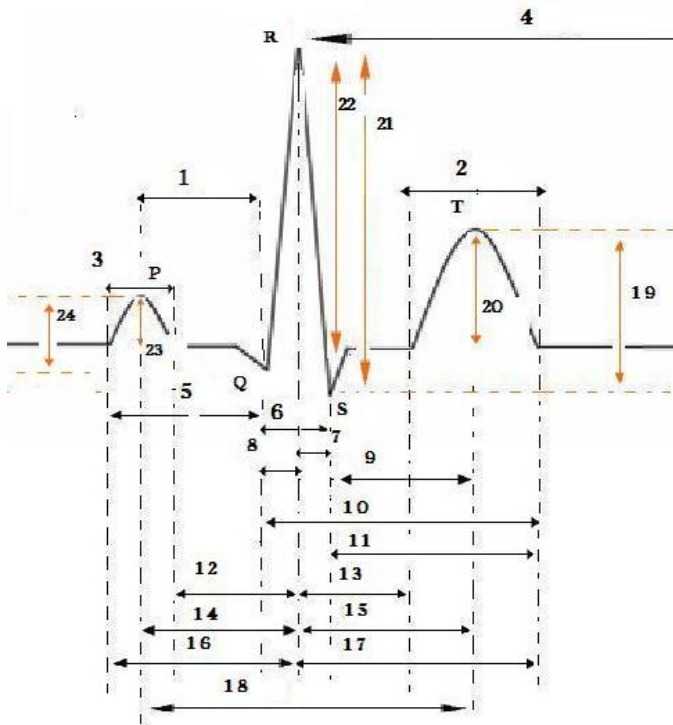


Figure 1: Extracted Features.

IV. FEATURE SELECTION

Feature selection is the most critical step in building any identification system. During this step, the set of attributes or features that deemed to be the most effective attributes are extracted in order to construct suitable detection algorithms (detectors). A key problem faced by many researchers is how to choose the optimal set of features since they are all not necessarily relevant to the learning algorithm. Furthermore, in some cases, irrelevant and redundant features can introduce noisy data that distracts the learning algorithm and therefore severely degrade the accuracy of the detector and cause slow training and testing process. Feature selection was proven to have a significant impact on the performance of the classifiers. Experiments by Yen *et al.* [14] show that feature selection can reduce the building and testing time of a classifier by 50%.

To select the best set of features, we collect, during the enrolment phase, approximately two minutes of heart beat data for 16 participants between 20 to 40 years old. Both women and men have participated in the experiment. 24 features are extracted from the ECG signal (Figure 1 and Table 1). We then use the Information Gain Ratio (IGR) measure [15] to assign a score to each individual feature. As shown in Figure 2, the features are ranked based on this score. We then use a sequential forward selection algorithm to reach the optimal subset of features [16]. The selection algorithm starts with an empty set S of best features, and then proceeds to add features from the ranked set of features F into S sequentially.

Table 1: List of Extracted Features

Features	1.sPQ	2.T	3.P	4.RR
Temporal	5.PQ	6.QRS	7.sRS	8.sRQ
	9.SsT	10.QT	11.SfT	12.sRfP
	13.sRT	14.sRsP	15.sRsT	16.sRdP
	17.sRfT	18.PT		
	19.TS	20.TS'	21.RS	22.QR
	23.PL'	24.amPQ		
Amplitude				

The “goodness” of the resulting set of features S is measured by the Mahalanobis distance-based classifier. The selection process stops when the gained classifier’s accuracy is below a certain selected threshold value or in some cases when the accuracy drops, which indicates that the accuracy of the current subset is below the accuracy of the previous subset.

V. EXPERIMENTAL RESULTS

Using the data set of ECG signals collected from 16 participants, we rank the features according to the score assigned by the IGR measure. The 12 top ranked features are shown in Table 2.

Table 2: Top 12 features

Rank	Feature	IGR
1	RS	0.81
2	QR	0.72
3	sRfP	0.52
4	sRsP	0.52
5	sRdP	0.50
6	sRfT	0.48
7	SfT	0.48
8	PQ	0.41
9	QT	0.40
10	SsT	0.38
11	RR	0.37
12	sPQ	0.36

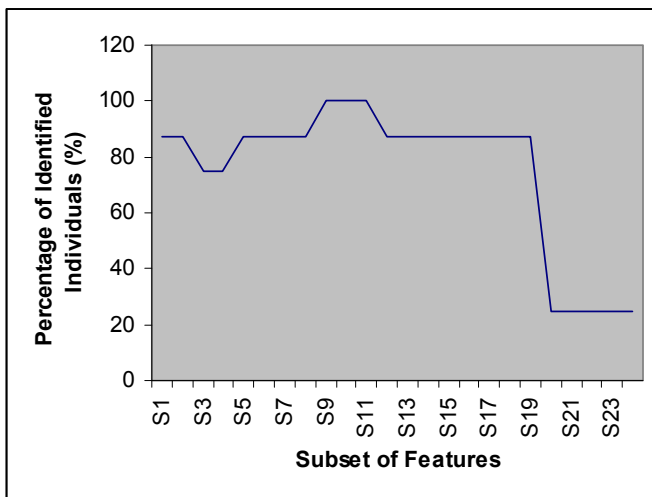


Figure 2: Identification Rate

100% identification rate was achieved using the subset of features S_9 . We can conclude that the first 9 features (RS, QR, sRfP, sRsP, sRdP, sRfT, SfT, PQ, QT) are the best set of features to build a large scale biometric identification system based on the ECG signal. Increasing the number of features does not contribute to the improvement of the accuracy. In contrast, the use of irrelevant features can distract the classifier, leading to a reduction of accuracy as much as 25% with 20 features.

VI. CONCLUSION & FUTURE WORK

In this paper, we have presented a biometric system that accurately identifies individuals based on the electrocardiogram (ECG) signal. Initially, a set of 24 features of the ECG is used but later reduced to nine features. Preliminary results indicate that the system is capable of accurately identifying individuals in a computationally efficient manner. Furthermore, it has also been observed that increasing the number of features beyond the nine used here does not lead to increased accuracy. Future work includes greater, more extensive experiments with human participants. It is anticipated the results of these experiments will confirm the preliminary results presented here. However, these experiments can also provide greater insight to the methods and techniques presented here and potentially lead to further improvements.

VII. REFERENCES

- [1] J. D. Woodward, N. M. Orlans, P. T. Higgins, Biometrics, New York: McGraw-Hill, 2002.
- [2] "Birmingham Heartlands RFID-tags patients to avoid litigation", February 2005. [Retrieved June 6, 2007, from <http://www.bjhc.co.uk/news/1/2005/n502016.htm>]
- [3] J. Best. "RFID comes to European hospitals", 2005. [Retrieved June 2, 2007, from <http://networks.silicon.com/lans/0,39024663,39129743,00.htm>]
- [4] R. Hoekema, G. G. H. Uijen, A. van Oosterom, "Geometrical aspect of the interindividual variability of multilead ECG recordings", IEEE Trans. Biomed. Eng., vol.48, pp.551-559, 2001.

- [5] B. P. Simon and C. Eswaran, "An ECG classifier designed using modified decision based neural network", Comput. Biomed. Res., vol. 30, pp. 257-272, 1997
- [6] G. Wuebbeler, et al., Human Verification by Heart Beat Signals, Working Group 8.42, Physikalisch-Technische Bundesanstalt (PTB).
- [7] T.W. Shen, W. J. Tompkins, and Y. H. Hu, "One-lead ECG for identity verification", Proc. of the 2nd Conf. of the IEEE Eng. in Med. and Bio. Society and the Biomed. Eng. Society, vol. 1, pp. 62-63, 2002
- [8] T.W. Shen, "Biometric Identity Verification Based on Electrocardiogram (ECG)", PHD Dissertation, University of Wisconsin, Madison, 2005
- [9] L. Biel, O. Pettersson, L. Philipson, P. Wide, "ECG analysis: a new approach in human identification", IEEE Trans. on Instrumentation and Measurement, vol. 50, no. 3, pp. 808-812, 2001
- [10] Y. Wang, F. Agrafioti, D. Hatzinakos, K. N. Plataniotis, "Analysis of Human Electrocardiogram (ECG) for Biometric Recognition", EURASIP Journal on Advances in Signal Processing, Volume 2008, Article ID 148658.
- [11] S. A. Israel, J. M. Irvine, A. Cheng, M. D. Wiederhold, and B. K. Wiederhold, "ECG to identify individuals", Pattern Recognition 38 (1): 133-142, 2005
- [12] S. A. Israel, W. T. Scruggs, W. J. Worek, and J. M. Irvine, "Fusing face and ECG for personal identification", Proc. of 32nd Applied Imagery Pattern Recognition Workshop, pp.226-231, 2003.
- [13] A. K. Jain, A. Ross, S. Prabhakar "An Introduction to Biometric Recognition", In IEEE Transactions On Circuits and Systems for Video Technology, VOL. 14, NO. 1, JANUARY 2004.
- [14] Y. Chen, Y. Li, X. Cheng, L. Guo, "Survey and Taxonomy of Feature Selection Algorithms in Intrusion Detection System", Inscrypt 2006.
- [15] J.R. Quinlan, "Induction of decision trees", Machine Learning, 1, p81-106.
- [16] M. Guennoun, A. Lbekkouri, K. El-Khatib, "Selecting the Best Set of Features for Efficient Intrusion Detection in Wireless Networks", 3rd International IEEE Conference on Information and Communication Technologies: From Theory to Applications, 2008.