ECG-BASED CONTINUOUS AUTHENTICATION SYSTEM USING ADAPTIVE STRING MATCHING

David Pereira Coutinho
Instituto Superior de Engenharia de Lisboa, Instituto de Telecomunicações and Instituto Superior Técnico, Lisboa, Portugal

Ana L. N. Fred, Mário A. T. Figueiredo
Instituto de Telecomunicações and Instituto Superior Técnico, Lisboa, Portugal

Keywords: Biometrics, ECG, String matching, LZ77, Data compression.

Abstract: Conventional access control systems are typically based on a single time instant authentication. However, for high-security environments, continuous user verification is needed in order to robustly prevent fraudulent or unauthorized access. The electrocardiogram (ECG) is an emerging biometric modality with the following characteristics: (i) it does not require liveliness verification, (ii) there is strong evidence that it contains sufficient discriminative information to allow the identification of individuals from a large population, (iii) it allows continuous user verification. Recently, a string matching approach for ECG-based biometrics, using the Ziv-Merhav (ZM) cross parsing, was proposed. Building on previous work, and exploiting tools from data compression, this paper goes one step further, proposing a method for ECG-based continuous authentication. An adaptive way of using the ZM cross parsing is introduced. The use of the Lloyd-Max quantization is also introduced to improve the results with the string matching approach for ECG-based biometrics. Results on one-lead ECG real data are presented, acquired during a concentration task, from 19 healthy individuals.

1 INTRODUCTION

The problem of user verification has concentrated much effort of the research community in the last few years. This has resulted in many approaches for personal authentication based on biometric features. Conventional access control systems are typically based on a single time instant authentication. But the demand for robustly preventing fraudulent or unauthorized access to systems led us to another problem. How to guarantee that the initially authenticated person is the same that is using the system? This problem is addressed by continuous user verification, where biometric authentication (verification) (Jain et al., 2007) is useful, and several studies on this topic have been published (Monrose and Rubin, 2000)(Sim et al., 2007)(Niinuma and Jain, 2010).

The electrocardiogram (ECG) is an emerging biometric measure which exploits a physiological feature that exists on every human. There is strong evidence that the ECG is sufficiently discriminative to identify individuals within a large population. The ECG allows intrinsic liveliness verification, personal identification and authentication (Biel et al., 2001) (Chiu et al., 2008) (Pereira Coutinho et al., 2010b), and different stress or emotion states detection (Medina and Fred, 2010). Furthermore the ECG is a behavioral biometric trait that can be used with other biometric measures (Ross et al., 2006), as a complementary feature, for fusion in a multimodal biometric authentication system (Boulgouris et al., 2009, Ch. 18) and for continuous authentication where biological signatures are continuously monitored (easily done by using new signal acquisition technologies like the Vital Jacket (Cunha et al., 2007), (Leonov, 2009)) in order to guarantee the identity of the operator throughout the whole process (Damousis et al., 2008).

Recently, a string matching approach for ECG-based biometrics using the Ziv-Merhav (ZM) cross parsing was proposed (Pereira Coutinho et al., 2010a). Built on previous work and exploring principles from data compression, in this paper we go one step further, proposing a method for ECG-based continuous authentication. We introduce the use of ZM cross parsing with adaptive models, in a way similar to what is used by dictionary-based text data compression algorithms, such as the LZ77 (Ziv and Lempel, 1977). The use of the Lloyd-Max quantization is also

Pereira Coutinho D., L. N. Fred A. and A. T. Figueiredo M..
ECG-BASED CONTINUOUS AUTHENTICATION SYSTEM USING ADAPTIVE STRING MATCHING.
DOI: 10.5220/0003292003540359
In *Proceedings of the International Conference on Bio-inspired Systems and Signal Processing* (BIOSIGNALS-2011), pages 354-359
ISBN: 978-989-8425-35-5

proposed to improve the results with the string matching approach for ECG-based biometrics. Results on one-lead ECG real data are presented, acquired during a concentration task, from 19 healthy individuals.

This paper is organized as follows: Sec. 2 makes a brief review of the related works on ECG-based biometrics systems and provides some details about a recently proposed string matching approach. Sec. 3 discusses how to improve the efficacy of the string matching approach and describes the proposed continuous authentication method, detailing the use of the ZM cross parsing with adaptive models. Sec. 4 discusses and shows the experimental results. Finally, Sec. 5 finalizes the paper drawing the main conclusions.

2 RELATED WORKS ON ECG-BASED BIOMETRICS

Two question arise concerning the design of an ECG-based biometric system: How to store an ECG sample in a biometric database? How to match one ECG sample with database samples?

Biel et al. (Biel et al., 2001) were the first to answer these questions, by proposing the extraction of fiducial features, which are used for database storage and classification (see Figure 1).

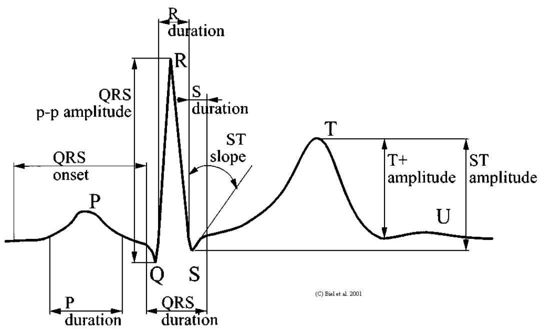

Figure 1: Elements of a typical ECG waveform and features used by Biel et al. for classification.

2.1 Fiducial versus Non-fiducial Approaches

In a broad sense, one can say there are two different approaches in the literature concerning feature extraction from ECG: fiducial and non-fiducial.

Fiducial methods use points of interest within a single heartbeat waveform, such as local maxima or minima; that is, they use references to allow the definition of features like latency times and amplitudes, as shown in Figure 1. Many other examples can be found in the literature.

Non-fiducial techniques aim at extracting discriminative information from the ECG waveform without having to extract fiducia. A global pattern from several heartbeat waveforms may be used as a feature; typically, wavelet or DCT coefficients have been used as features.

Some methods combine these two different approaches or are partially fiducial (Wang et al., 2008) (e.g., they use only the R peak as a reference for segmentation of the heartbeat waveforms). Table 1 summarizes several approaches found in the literature; for more details on each method, see the corresponding publication.

2.2 String Matching Approach

A new non-fiducial approach was recently proposed (Pereira Coutinho et al., 2010a). It is a simple approach where the first step is to convert ECG samples into sequences of symbols (strings) from an alphabet, using 8 bit uniform quantization (256 symbols). Although information is lost due the quantization process, enough discriminative information is preserved (as shown by the experiments). Identification or authentication is based on a 1-NN (nearest neighbor) classifier, using the string similarity measure described next. An application example on person identification is depicted in Figure 2.

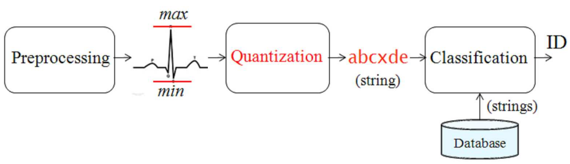

Figure 2: Identification block diagram overview. The acquired ECG signals are preprocessed in order to remove noise, segmented, quantized between its maximum and minimum value, resulting in a string. Test samples classification is based on string matching with database stored strings.

2.2.1 String Similarity Measure

The string similarity measure used in (Pereira Coutinho et al., 2010a) is based on *Ziv-Merhav cross parsing* (ZMCP) (Ziv and Merhav, 1993). Ziv and Merhav introduced an empirical divergence between two sequences $\mathbf{z}$ and $\mathbf{x}$ based on two LZ-type parsing algorithms: the incremental LZ parsing (LZ78) which is a self parsing procedure of a sequence (let the self parsing length of a given a sequence $\mathbf{z}$ be denoted by $c(\mathbf{z})$) and the cross parsing (LZ78 parsing variation)

Table 1: Comparison of related works with string matching approach. The accuracy (Accur.) values shown are the reported results for person identification.

Ref.	Feature	Method	Subjs.	Accur.
(Biel et al., 2001)	Fiducial	PCA	20	100%
(Shen et al., 2002)	Fiducial	Templ. matching+DBNN	20	100%
(Israel et al., 2005)	Fiducial	LDA	29	98 %
(Silva et al., 2007)	Fiducial	FSE	26	99.97%
(Chan et al., 2008)	Non-fiducial	Wavelet Distance	50	95%
(Wang et al., 2008)	Non-fiducial	AC/DCT+KNN	13	97.8%
(Pereira Coutinho et al., 2010a)	Non-fiducial	ZM Cross Parsing+1NN	19	100%

which is a sequential parsing of a sequence with respect to another sequence (let us denote by $c(\mathbf{z}|\mathbf{x})$ the sequential parsing length of a sequence $\mathbf{z}$ given a sequence $\mathbf{x}$).

For example, as shown in Figure 2, the 1-NN based identification of a test sample $\mathbf{z}$ in one of a set of K classes, given the subject models $\mathbf{x}_k$ per class k, is then simply

$$\hat{k}(\mathbf{z}) = \arg\min_{k\in\{1,\ldots,K\}} c(\mathbf{z}|\mathbf{x}_k).$$

The ZMCP implementation consists on a static dictionary pre-loaded with a sequence $\mathbf{x}$ (model), where only the *look ahead buffer* (LAB) slides over the input sequence $\mathbf{z}$ while making the cross parsing (for details see (Pereira Coutinho and Figueiredo, 2005)).

3 PROPOSED STRING MATCHING APPROACH

Our work is built on the basic ideas from (Pereira Coutinho et al., 2010b) proposing the major enhancements described as follows.

3.1 User-tuned Quantization

In the string matching approach, a user tuned authentication scheme was proposed (Pereira Coutinho et al., 2010b). After preprocessing the ECG acquired samples, the system uses uniform quantization to transform the samples into a string. In order to make the system more user-tuned, one can learn a quantizer for each user. That can simply be done using non-uniform quantization, namely the Lloyd-Max quantization (LMQ).

Our proposal is that, after the enrollment process, a Lloyd-Max quantizer be built for each user, and the selected ECG acquired samples be encoded with the associated user-tuned quantizer, and stored in the database, to be used as a model. During the authentication process the selected ECG acquired samples are also encoded with the Lloyd-Max quantizer associated to the subject that the user claims to be, and then compared with the same subject's stored model.

3.2 Continuous Authentication Method

The method that we propose is built on the string matching approach that was briefly described in Section 2.2. The classification stage remains the same, using the ZMCP as the string similarity measure. The major and significant change is in the way user's models are handled. An overview of the proposed system block diagram is depicted in Figure 3.

As in the original work, during the enrollment process the system must learn for each user both the user model and the user threshold. This is the crucial data that will be retrieved from the system database during the whole continuous authentication process. Then, unlike the original work, we propose that every time that the system makes a positive authentication, the system database must be updated with that test sample that led to the positive decision. This adaptive mode of managing user models will enable the system to learn during the authentication process.

Updating the model to store in the system database will result from a simple string concatenation operation. We stress the simplicity of this model updating mode because it is not a time consuming operation, which is an important feature concerning real-time systems. Due to user's models size limitations, a strategy like FIFO (First In First Out), or other criteria, must be adopted when updating these models.

Notice that acquiring ECG signals for continuous authentication will require specific devices that allow continuous monitoring of the operator throughout the whole process. An example of such devices would be a wearable device like the Vital Jacket (Cunha et al., 2007).

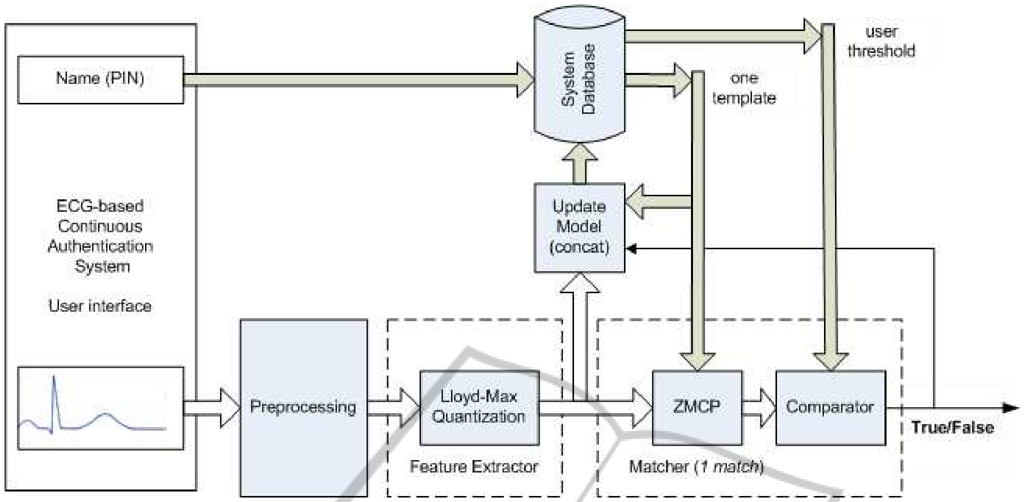

Figure 3: Proposed continuous authentication system block diagram overview. Notice that the system database is updated only if user authentication result is true.

4 EXPERIMENTS

4.1 Data Collection

The ECG waveform dataset used was acquired using one lead, in the context of the Himotion project [1]. The dataset contains ECG recordings from 19 subjects acquired during a concentration task on a computer, designed for an average completion time of 10 minutes. All the acquired ECG signals were normalized and band-pass filtered (2–30Hz) in order to remove noise. Each heartbeat waveform was sequentially segmented from the full recording and then all the obtained waveforms were aligned by their R peaks. From the resulting collection of ECG heartbeat waveforms, the mean wave for groups of 10 consecutive waveforms (without overlap) was computed. Each of these mean waveforms is what it is called a single heartbeat in (Pereira Coutinho et al., 2010b). Notice that an intra-class study (Medina and Fred, 2010) with the dataset, in the context of the exploration of electrophysiological signals for emotional states detection, showed the existence of differentiated states in the data that represent the ECG signal of a subject.

4.2 Experimental Results

The experimental results were obtained with a strategy of leave-one-out cross validation over 50 runs. The samples for the model and for the test sample were chosen randomly from all the acquired single heartbeats waveforms for each user.

[1] https://www.it.pt/auto_temp_web_page_preview.asp?id=305

We evaluated user authentication using non-uniform quantization and Table 2 shows the comparison results of authentication experiments over the same dataset. The results show that Lloyd-Max (rather than uniform) quantization and user-adjusted thresholds clearly improve the performance of the original string matching approach.

The proposed continuous authentication method was evaluated. Table 2 shows the results for the continuous authentication experiments over the same dataset. Test results show that the introduced adaptive capability, that is using ZMCP with adaptive user models, improve also the performance of the original system.

5 CONCLUSIONS

In this paper, the presented work is built on the basic ideas from (Pereira Coutinho et al., 2010b). Two major enhancements on the string matching approach for ECG-based biometrics was proposed. One was the use of non-uniform quantization, namely the Lloyd-Max quantization, when converting an ECG acquired sample into a string. The other was to allow system database update during the continuous authentication process. This adaptive mode depends on a simply string concatenation operation, which is not a time consuming operation, an important feature regarding real-time system.

Experiments carried out on a dataset with 19 healthy subjects, for whom the existence of differentiated states in the ECG data of a subject has been

Table 2: Comparison of authentication related work results over the same dataset.

Reference	Feature	EER
(Oliveira and Fred, 2009)	Fiducial (1-NN classifier)	8.0 %
(Gamboa, 2008)	Fiducial (user-tuned)	1.7 %
(Pereira Coutinho et al., 2010b)	Non-fiducial (uniform quantiz., user-tuned)	1.1 %
Proposed authentication method	Non-fiducial (LLoyd-Max quantiz., user-tuned)	0.37 %
Proposed cont. authentication method	Non-fiducial (LLoyd-Max quantiz., user-tuned, adaptive models)	0.36 %

shown (Medina and Fred, 2010). Results showed that our method improve the performance of the original system, enabling an average EER (equal error rate) of 0.37 % on authentication and 0.36 % on continuos authentication.

Future work will include tests with other datasets for further evaluation of our method, particularly with datasets that have longer ECG samples. This allow a more accurate performance evaluation in the case of continuous authentication. The size of the HiMotion project samples was quite small and this was a drawback in the present work. The user threshold tunning process is another problem that must be addressed in future studies because an adaptive learning strategy is needed.

ACKNOWLEDGEMENTS

We acknowledge the following financial support: Instituto Superior de Engenharia de Lisboa (ISEL), the FET programme, within the EU FP7, under the SIMBAD project (contract 213250); Fundação para a Ciência e Tecnologia (FCT), under grants PTDC/EEA-TEL/72572/2006 and QREN 3475.

REFERENCES

Biel, L., Pettersson, O., Philipson, L., and Wide, P. (2001). ECG analysis – a new approach in human identification. *IEEE Transactions on Instrumentation and Measurement*, 50(3):808–812.

Boulgouris, N. V., Plataniotis, K. N., and (Editors), E. (2009). *Biometrics: Theory, Methods, and Applications*. Wiley-IEEE Press.

Chan, A., Hamdy, M., Badre, A., and Badee, V. (2008). Wavelet distance measure for person identification using electrocardiograms. *Instrumentation and Measurement, IEEE Transactions on*, 57(2):248–253.

Chiu, C.-C., Chuang, C.-M., and Hsu, C.-Y. (2008). A novel personal identity verification approach using a discrete wavelet transform of the ECG signal. In *MUE '08: Proceedings of the 2008 International Conference on Multimedia and Ubiquitous Engineering*, pages 201–206, Washington, DC, USA. IEEE Computer Society.

Cunha, J., Cunha, B., Xavier, W., Ferreira, N., and Pereira, A. (2007). Vital-jacket: A wearable wireless vital signs monitor for patients mobility. In *Proceedings of the Avantex Symposium*.

Damousis, I. G., Tzovaras, D., and Bekiaris, E. (2008). Unobtrusive multimodal biometric authentication: the humabio project concept. *EURASIP J. Adv. Signal Process*, 2008:1–11.

Gamboa, H. (2008). *Multi-Modal Behavioral Biometrics Based on HCI and Electrophysiology*. PhD thesis, Instituto Superior Técnico, Universidade Técnica de Lisboa, Lisboa, Portugal.

Israel, S. A., Irvine, J. M., Cheng, A., Wiederhold, M. D., and Wiederhold, B. K. (2005). ECG to identify individuals. *Pattern Recognition*, 38(1):133–142.

Jain, A. K., Flynn, P., and Ross, A. A. (2007). *Handbook of Biometrics*. Springer-Verlag New York, Inc., Secaucus, NJ, USA.

Leonov, V. (2009). Wireless body-powered electrocardiography shirt. In *Proceedings of the Smart Systems Integration European Conference*.

Medina, L. A. S. and Fred, A. L. N. (2010). Genetic algorithm for clustering temporal data - application to the detection of stress from ecg signals. In *Proc 2nd International Conference on Agents and Artificial Intelligence (ICAART)*, pages 135–142.

Monrose, F. and Rubin, A. D. (2000). Keystroke dynamics as a biometric for authentication.

Niinuma, K. and Jain, A. K. (2010). Continuous user authentication using temporal information. In *Proc. SPIE, Vol. 7667, 76670L*.

Oliveira, C. and Fred, A. L. N. (2009). ECG-based authentication: Bayesian vs. nearest neighbour classifiers. In *Proc International Conf. on Bio-inspired Systems and Signal Processing - Biosignals - INSTICC*, Porto, Portugal.

Pereira Coutinho, D. and Figueiredo, M. (2005). Information theoretic text classification using the Ziv-Merhav method. *2nd Iberian Conference on Pattern Recognition and Image Analysis – IbPRIA'2005*.

Pereira Coutinho, D., Fred, A. L. N., and Figueiredo, M. (2010a). One lead ecg-based personal identification using ziv-merhav cross parsing. In *Proceedings of the 20th International Conference on Pattern Recognition (ICPR)*, Istanbul, Turkey.

Pereira Coutinho, D., Fred, A. L. N., and Figueiredo, M. (2010b). Personal identification and authentication based on one-lead ecg using ziv-merhav cross parsing. In *Proceedings of the 10th International Workshop on Pattern Recognition in Information Systems (PRIS)*, Funchal, Portugal.

Ross, A. A., Nandakumar, K., and Jain, A. K. (2006). *Handbook of Multibiometrics (International Series on Biometrics)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA.

Shen, T., Tompkins, W. J., and Hu, Y. H. (October 2002). One-lead ECG for identity verification. *Proc. of the 2nd Joint Conf. of the IEEE Eng. in Medicine and Biology Soc. and the 24th Annual Conf. and the Annual Fall Meeting of the Biomedical Eng. Soc. (EMBS/BMES '02)*, 1:62–63.

Silva, H., Gamboa, H., and Fred, A. (2007). One lead ECG based personal identification with feature subspace ensembles. In *MLDM '07: Proceedings of the 5th international conference on Machine Learning and Data Mining in Pattern Recognition*, pages 770–783, Berlin, Heidelberg. Springer-Verlag.

Sim, T., Zhang, S., Janakiraman, R., and Kumar, S. (2007). Continuous verification using multimodal biometrics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29:687–700.

Wang, Y., Agrafioti, F., Hatzinakos, D., and Plataniotis, K. N. (2008). Analysis of human electrocardiogram for biometric recognition. *EURASIP J. Adv. Signal Process*, 2008:19.

Ziv, J. and Lempel, A. (1977). A universal algorithm for sequential data compression. *IEEE Transactions on Information Theory*, 23(3):337–343.

Ziv, J. and Merhav, N. (1993). A measure of relative entropy between individual sequences with application to universal classification. *IEEE Transactions on Information Theory*, 39:1270–1279.