DOI: 10.1109/BIOMS.2014.6951541

HeartCode: a Novel Binary ECG-based Template

Ruggero Donida Labati, Vincenzo Piuri, Roberto Sassi, Fabio Scotti
Department of Computer Science
Università degli Studi di Milano
Milano, 20122, Italy.
{ruggero.donida, vincenzo.piuri, roberto.sassi, fabio.scotti}@unimi.it

***Abstract*—Recent studies on ECG signals proved that they can be employed as biometric traits able to obtain sufficient accuracy in a wide set of applicative scenarios. Most of the systems in the literature, however, are based on templates consisting in vectors of integer or floating point numbers. While any numerical representation is inherently binary, in here we consider as binary templates only those codings in which similarity or distance metrics can be directly applied to the for performing identity comparisons. With respect to templates composed by integer or floating point values, the use of binary templates presents important advantages, such as smaller memory space, and faster and simpler matching functions. Binary templates could therefore be adopted in a wider range of applications with respect to traditional ECG templates, like wearable devices and body area networks. Moreover, binary templates are suitable for most of the biometric template protection methods in the literature. This paper presents a novel approach for computing and processing binary ECG templates (HeartCode). Experimental results proved that the proposed approach is effective and obtains performance comparable to more mature biometric methods for ECG recognition, obtaining Equal Error Rate (EER) of 8.58% on a significantly large database of 8400 samples extracted from Holter acquisitions performed in uncontrolled conditions.**

I. Introduction

Biometric systems based on electrocardiogram (ECG) signals have recently been proposed in the literature. They present important advantages with respect to traditional biometric systems: cardiac signals are difficult to counterfeit since they can only be acquired with specific devices; ECG signals can only be acquired from living people; biometric recognition methods can be adopted in cooperation with healthcare monitoring systems without adding hardware costs; ECG signals are continuously acquired by the sensors, allowing to design continuous verification systems that require a low level of user cooperation.

Most of the systems in the literature, however, are based on feature extraction and matching algorithms adopting templates composed by sets of integer or floating point numbers [1]. While any numerical representation is inherently binary, we consider as binary templates only those codings in which similarity or distance metrics can be directly applied to the bits for performing identity comparisons.

Binary templates present important advantages since they require less storage space and, in principle, they allow to design matching functions requiring smaller computational time. They could therefore enable the use of ECG recognition systems in a wide range of applicative scenarios, encompassing wearable devices [2], and body area networks [3].

Binary templates can also enable the adoption of template protection strategies [4]–[6].

In the literature, there are studies on compression algorithms for ECG signals [7], but they are not designed for biometric recognition. Moreover, many studies do not aim to obtain binary templates and the design of efficient identity comparison methods based on the binary representations obtained by most of the methods is then particularly complex. In fact, most of these representations do not permit the use of simple and fast distance metrics for estimating the similarity between two templates.

At the best of our knowledge, there is only a study on ECG recognition methods based on binary templates [8]. This method, however, needs to store many samples during the training phase since the feature extraction is based on Linear Discriminant Analysis (LDA).

Differently, we propose a biometric recognition approach based on simple knowledge on the mean QRS complex of the population, which can be obtained once from sets of samples not pertaining to the final users, allowing the use of the proposed approach in a wider range of applications (e.g. consumer applications based on low-cost and portable devices).

The proposed approach is based on a strategy similar to the one commonly adopted in iris recognition systems [9], since it is the one that obtained the best results in our experiments.

The main contribution of the paper is two-fold. First, we present strategies to adapt feature extraction and matching algorithms designed for iris recognition to ECG signals. Second, we analyze the properties of the method on a significantly large biometric database of Holter acquisitions performed in uncontrolled conditions during a relevant time interval.

Similarly to [10], the proposed approach only considers the QRS complex since it has been proved to be the part of cardiac signals which changes less over time [11]. The QRS complex represents the depolarization of the right and left ventricles, and is the biggest complex of the ECG (with a duration between 70 and 110 ms in a normal heartbeats).

The approach is able to use single lead ECG signals, but obtains better accuracy using signals acquired from multiple leads.

Since in iris recognition systems the binary templates are usually called IrisCode, in the rest of the paper we refer to binary templates computed from ECG signals as HeartCode. An example of QRS signals extracted from three leads and the corresponding template HeartCode are shown in Fig. 1.

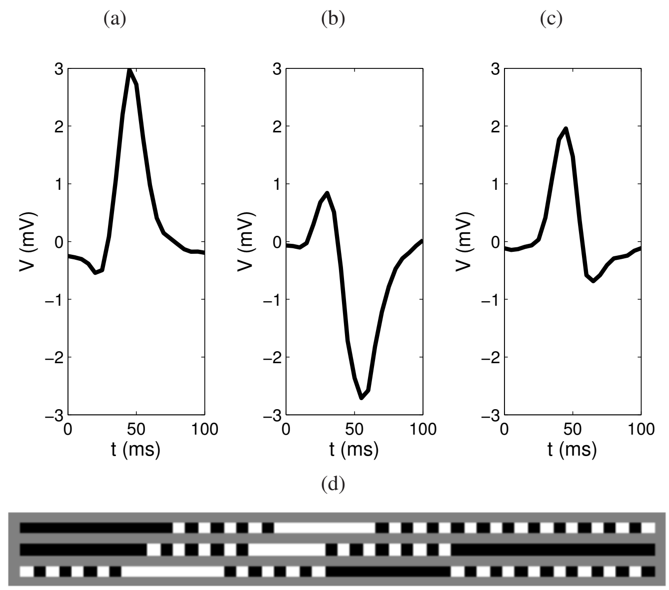

Fig. 1. Example of QRS signals related to three leads of a Holter acquisition and corresponding template HeartCode: (a) lead 1; (b) lead 2; (c) lead 3; (d) template HeartCode.

The obtained results are encouraging, giving the feasibility of the biometric recognition results and the fact that the properties of the HeartCode make it suitable for cryptographic applications.

The paper is structured as follows. Section II describes the related works. Section III presents the proposed method, detailing all the steps of the implemented biometric recognition system. The performed experiments and a comparison with biometric recognition methods based on non-quantized templates are described in Section IV. Finally, Section V concludes the work.

II. Related Works

Most of the studies on ECG signals in biometric systems consider acquisitions performed by a single lead [12]. Other studies use multiple leads in order to increase the recognition accuracy [13], [14].

ECG recognition systems can also be based on fiducial or non-fiducial features [15]. Systems based on fiducial features extract points of interest within the heartbeat wave, called fiducial points. Systems based on non-fiducial features do not consider fiducial points and usually extract features in a transformed domain (frequency or wavelet).

Another characteristic of ECG recognition systems consists in the evaluated portion of the signal. Many recognition algorithms use the full ECG signal [1]. Differently, other methods use only the QRS complex [10], [16], [17] since it is the most stable-in-time part of the signal.

Most of the methods in the literature build templates composed by integer or floating point numbers. Examples of features are the coordinates of fiducial points [15], portions of the ECG signals [10], frequency characteristics [18].

With respect to templates composed by integer or floating point numbers, binary templates present important advantages in terms of computational efficiency and storage space. They can also be easily adopted in template protection schemes [4]–[6]. However, the use of binary templates is not always possible with all the biometric traits since in many cases it might lead to an unsatisfactory accuracy.

Most of the iris recognition systems are based to the binary template IrisCode [9] since it permits to obtain high accuracy with low computational time.

Biometric recognition methods using binary templates and matchers based on the computation of the Hamming distance between templates have also been studied for different biometric traits, like fingerprint [19] and palmprint [20].

In the literature, there are studies on compression methods for ECG signals [7]. However, they are not designed to be applied in biometric systems and do not permit to apply fast and accurate matching algorithms based on simple similitude metrics.

At the best of our knowledge, the only study on ECG recognition systems based on binary templates is described in [8]. The feature extraction is performed by the AC/LDA method, which first computes autocorrelations of portions of ECG signals with fixed time duration, and then performs a dimensionality reduction using the LDA. The SPEC-Hashing algorithm is then applied to the LDA coefficients. The matching scores are finally obtained by computing the Hamming distance between templates.

III. The Proposed Approach

We propose a novel approach for ECG recognition able to obtain fast identity comparisons based on compact binary templates suitable for cryptographic applications and privacy compliant biometric systems. The proposed template HeartCode requires a small amount of memory (150 bits in our tests, with no additional compression) and is designed to be applied in consumer applications like wearable devices and body area networks. The method proposed in this work for processing biometric templates is derived from the demodulation technique used to compute the template IrisCode (widely adopted in commercial iris recognition systems), and applied for the first time to ECG signals acquired by multiple leads. Similarly to [10], the proposed approach considers samples with a fixed duration of Δ_t seconds.

The structure of the proposed method is depicted in Fig. 2, and can be divided into the sequent steps:

- preprocessing;
- heartbeat selection;
- template computation;
- matching.

A. Preprocessing

This step first performs a noise reduction and then extracts a vector V of QRS signals from the ECG sample.

First, the noise introduced by the 50 Hz electrical component is reduced by applying a notch IIR filter, and then the

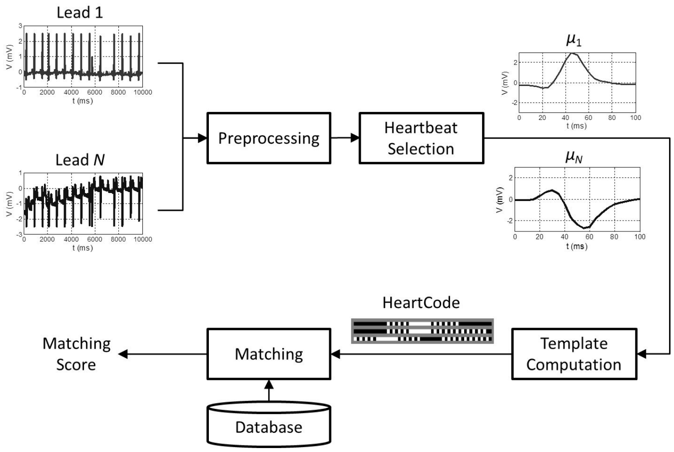

Fig. 2. Schema of the proposed biometric recognition approach.

signal's baseline is normalized using a Butterworth filter with cutoff frequency of 0.5 Hz [21].

In order to obtain the vector V of QRS signals, a list of the R fiducial points is extracted from the ECG signals by using an automatic labeling software (Vision Premier, SpaceLab-Burdick Inc.). For each detected heartbeat, the signal representing the QRS complex is computed as a fixed time window centered in R (from -50 ms to + 70ms), obtaining the vector V. The use of fixed time windows is justified by the fact that the temporal duration of the QRS complex does not depend on the heart rate (if not very minimally). In fact, the QRS complex is produced by the spread of the depolarization front in the myocardial tissue, which is guided by a specialized tissue (the Purkinje fibers). The length of our template was selected to be 120 ms, in order to contain most of the QRS complex (a QRS complex is pathologically wide when longer than 0.12 s).

B. Heartbeat Selection

This step first computes a matrix Q composed by a maximum of m QRS signals. The selected heartbeats are the ones that present the maximum normalized correlation with the mean QRS signal of V. The computation is performed by using the iterative algorithm described in [10], [22].

For each lead i, the mean heartbeat μ_i is then estimated from the matrix Q_i of QRS signals. The signal μ_i is computed as:

$$\mu_i = \sum_{j=1}^{m} Q_i(j)/m. \tag{1}$$

C. Template Computation

Templates can be computed considering only the QRS signals or by applying a template normalization strategy for increasing the recognition accuracy. The template normalization strategy uses simple knowledge on the shape of the average QRS complex of the population in order to extract more discriminant features from the input signals.

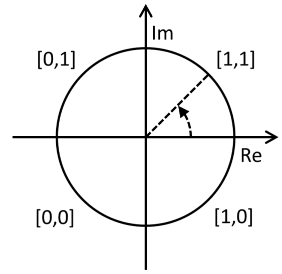

Fig. 3. Phase-quadrant demodulation code [9].

1) Basic template computation algorithm: every signal μ_i is transformed using quadrature 1D Log-Gabor wavelets [9], [23]. Log-Gabor filters are used because they have 0 DC component for arbitrary large bandwidth. The frequency response of a Log-Gabor filters is given as:

$$G(f) = \exp\left\{ \frac{-\left[log\left(f/f_0\right)\right]^2}{2\left[log\left(\sigma/f_0\right)\right]^2} \right\}, \tag{2}$$

where f_0 is the center frequency, and σ represents the bandwidth of the filter. The complex result obtained by convolving the Log-Gabor filter with the signal μ_i is then converted to a binary string S_i by identifying in which quadrant of the complex plane each resultant phasor lies. The coding scheme is shown if Fig. 3.

The demodulation scheme produces two bits for each signal's sample. Each sting S_i is therefore composed by 2 n_s bits, where n_s is the number of signal's samples of every QRS.

The template HeartCode H is finally computed as a binary matrix of m rows, in which every row is a binary string S_i.

2) Template normalization: the main idea of the proposed method consists in using a signal representing the ideal average heartbeat of the population to normalize the QRS signals composing the HeartCode. The average QRS signal $\overline{QRS}$ should be computed ones before using the proposed biometric recognition method on a set of available ECG signals.

Given a set of n_t samples, the average signal $\overline{QRS}_i$ of each lead is obtained by first estimating the mean heartbeat μ_{il} of every sample l.

Every signal μ_{il} is then aligned to μ_{i1} by applying a cross-correlation approach. The average signal $\overline{QRS}_i$ is then obtained as:

$$\overline{QRS}_i = \sum_{l=1}^{n_t} \mu_{il}/n_t. \tag{3}$$

During the enrollment and verification / identification phases, the signal μ_i is first aligned to $\overline{QRS}_i$ by applying a cross-correlation approach.

The signal $\hat{\mu}_i$ is then computed as $\hat{\mu}_i = \overline{QRS}_i - \mu_i$.

The template HeartCode H is finally obtained by applying the basic template computation algorithm to $\hat{\mu}_i$.

D. Matching

This step computes the matching score as the Hamming distance between the templates. Given two templates H_A and H_B, the matching score is obtained as:

$$\text{Matching Score} = \|H_A \otimes H_B\|, \tag{4}$$

where $\|\|$ represents the norm, and $\otimes$ is the XOR operator.

IV. Experimental Results

The performed experiments aim to evaluate the effect of the proposed quantization approach on the recognition accuracy and to analyze the entropy of the template HeartCode.

Experiments have been conducted on a significantly large biometric dataset of Holter acquisitions performed in uncontrolled conditions during a relevant time interval. The dataset consists in signals obtained from the E-HOL-03-0202-003 (Intercity Digital Electrocardiogram Alliance – IDEAL) database [24]. This database is composed by digital 24-hours Holter recordings (SpaceLab-Burdick Inc.) from 202 individuals. Data have been collected in a single session. The number of samples acquired from males corresponds to the one acquired from females. The signals have been acquired performed without any restrain or control on the user's activities.

Since performing tests using all the data of the Holter database requires very high computational time, we have selected a subset of samples for our evaluations. The dataset is composed by 8400 samples with a duration $\Delta_t = 300$ seconds, regarding 140 Holter acquisitions from 140 individuals.

The parameters of the Log-Gabor filters used by the demodulation algorithm are: $f_0 = 0.56$ and $\sigma = 0.5$.

The number of QRS signals composing the vectors V is equal to 16.

A. Performance

The performance of the proposed approach has been evaluated in different configurations and compared with our best correlation-based method in order to estimate the effect of the template quantization on the system accuracy. The compared methods are:

- Best correlation-based method - as a reference, we have used the biometric recognition method presented in [10]. For each lead i, this method computes the maximum normalized cross-correlation between every QRS of the vectors V_A and V_B (16×16 normalized cross-correlations). The maximum similarity score obtained on the three leads of the Holter acquisitions is considered as the matching score.
- HeartCode (no normalization) - in order to evaluate the effect of the template normalization, the proposed approach has been applied without performing the template normalization.
- HeartCode - the proposed approach has been applied in its standard configuration, performing the template normalization. For each of the condisedered140 individuals, a single sample not pertaining to the testing dataset has been used to compute the average heartbeat of the population. The shape of the QRS complex is on average very similar across different subjects (in fact, medical doctors use divergences from a common pattern to detect pathologies, which induce heart enlargements). Therefore, we have used a pooled average pattern computed over the entire population.

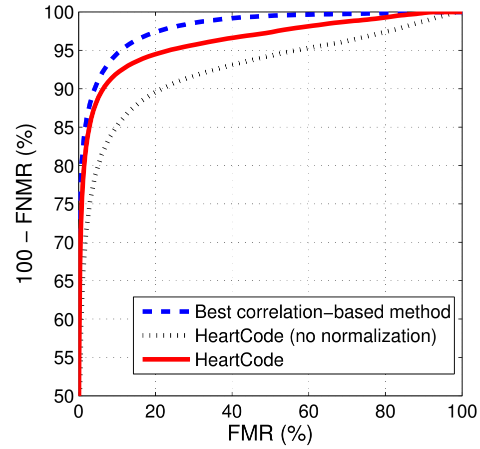

Fig. 4. ROC curves obtained by the correlation-based method described in [10] (Correlation-based) and by the proposed approach based on binary templates (HeartCode) on the same dataset.

TABLE I. Accuracy of the proposed approach.

Method	EER (%)	FNMR (%) @ FMR = 5%	FNMR (%) @ FMR = 1%
HeartCode (no normalization)	12.98	18.55	34.10
HeartCode	8.58	11.11	22.68
Best correlation-based method	7.01	7.85	16.35

The performed tests regard 35,275,800 identity comparisons $(8400 \times 8399/2)$. The obtained Receiver Operating Characteristic (ROC) curves are shown in Fig. 4, and the numerical results are reported in Table I.

It is possible to observe that the proposed template normalization method reduces the EER from 12.98% to 8.58%.

The recognition accuracy, however, is lower than the reference correlation-based method [10]. As an example, the EER is increased from 7.01% to 8.58%. Nevertheless, the obtained accuracy can be considered as satisfactory for a wide range of ECG recognition applications.

An important advantage of the proposed approach consists in the simple and fast matching strategy. The recognition method based on the template HeartCode drastically reduces the matching time with respect to the reference correlation-based technique. The matching time, in fact, has been reduced from 112 μs to 1 μs. The reported values are related to non-optimized implementations in C language executed on a computer with Windows 7, Intel Xeon 3.3 GHz processor, and 8Gb of RAM.

B. Statistical Properties of the HeartCode

As first level of analysis of HartCode templates in cryptographic applications, we have evaluated the probability distribution of the bits pertaining to the obtained binary strings.

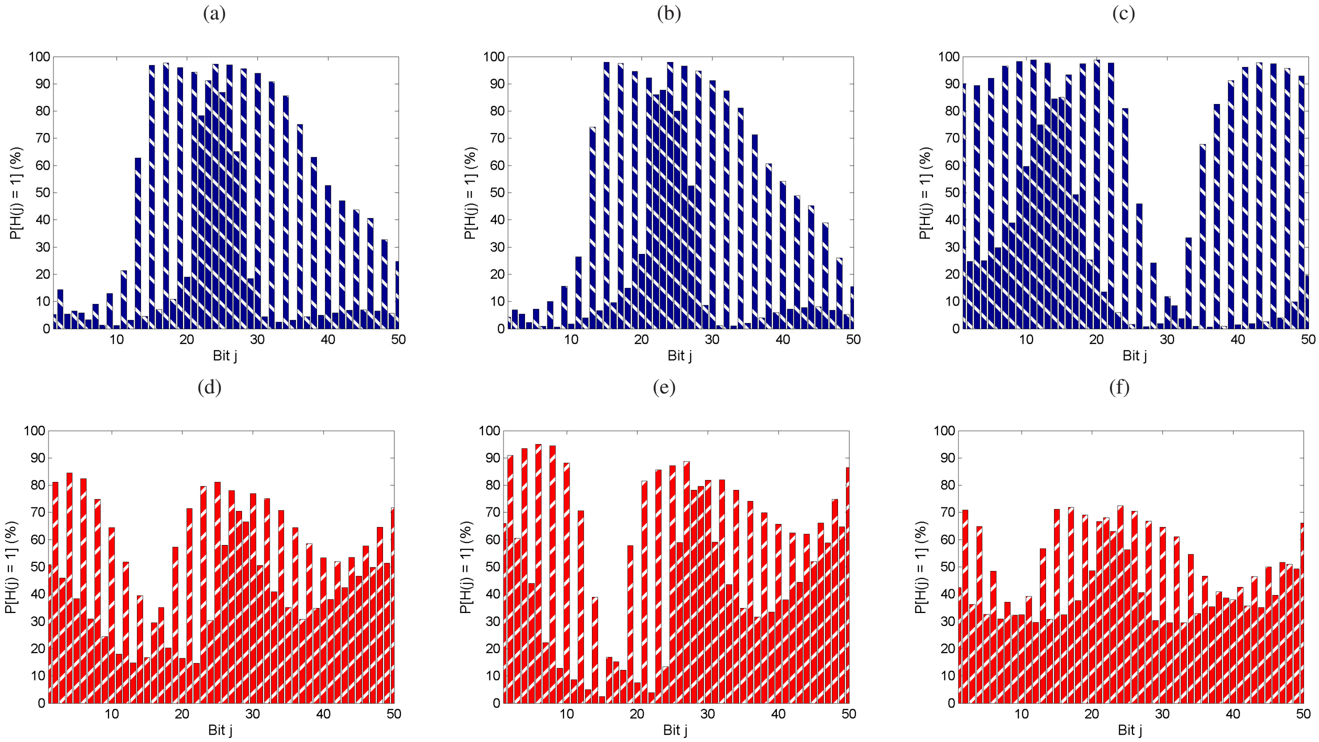

Fig. 5. Probability of obtaining templates HeartCode with bit j equal to 1, $P[H(j) = 1]$: (a) HeartCode (no normalization), lead 1; (b) HeartCode (no normalization), lead 2; (c) HeartCode (no normalization), lead 3; (d) HeartCode, lead 1; (e) HeartCode, lead 2; (f) HeartCode, lead 3. The template normalization strategy improves the statistical properties of the HeartCode.

We have analyzed templates obtained from 8400 samples using both the simplified version of the proposed approch (HeartCode no normalization), and the configuration of the method performing the template normalization (HeartCode). In the first case, we obtained mean $\{P[H(j) = 1]\} = 42.19\%$. The standard configuration of the method obtained mean $\{P[H(j) = 1]\} = 51.20\%$.

For each bit j of the binary strings, we have then evaluated the probability $P(H(j) = 1)$. The histograms of the obtained values are shown in Fig. 5. It is possible to observe that the proposed template normalization strategy allows both to increase the recognition accuracy (Fig. 4) and to improve the statistical properties of the template HeartCode since the probability of a bit to be equal to 1 is closer to 50%.

V. Conclusion

In this paper, we presented a novel approach for biometric recognition systems based on ECG signals, which constructs templates consisting in binary strings (HeartCode) and performs identity comparisons by computing the Hamming distance between templates.

The presented approach obtained results comparable to more mature ECG recognition methods, with an Equal Error Rate (EER) of 8.58% on a dataset of 8400 samples obtained from Holter acquisitions performed in uncontrolled conditions.

The template HeartCode also showed promising statistical properties that make it suitable for cryptographic applications.

Future works should regard the optimization of the family of wavelets adopted in the template creation and the application of the approach in template protection schemes.

Acknowledgment

This work was partly funded by the European Commission under the project ABC4EU (contract n. FP7-312797) and the Italian Ministry of Research within PRIN project "GenData 2020" (2010RTFWBH). Data used for this research were provided by the Telemetric and Holter ECG Warehouse of the University of Rochester (THEW), NY.

References

[1] I. Odinaka, P.-H. Lai, A. Kaplan, J. O'Sullivan, E. Sirevaag, and J. Rohrbaugh, "ECG biometric recognition: A comparative analysis," *IEEE Trans, on Information Forensics and Security*, vol. 7, no. 6, pp. 1812–1824, December 2012.

[2] E. Winokur, M. Delano, and C. Sodini, "A wearable cardiac monitor for long-term data acquisition and analysis," *IEEE Trans.on Biomedical Engineering*, vol. 60, no. 1, pp. 189–192, Jan 2013.

[3] S. Movassaghi, M. Abolhasan, J. Lipman, D. Smith, and A. Jamalipour, "Wireless body area networks: A survey," *IEEE Communications Surveys Tutorials*, vol. PP, no. 99, pp. 1–29, 2014.

[4] R. D. Labati, V. Piuri, and F. Scotti, *Biometric Privacy Protection: Guidelines and Technologies*. Springer, 2012, vol. 314, pp. 3–9.

[5] S. Cimato, R. Sassi, and F. Scotti, "Biometric privacy," in *Encyclopedia of Cryptography and Security (2nd ed.)*, H. van Tilborg and S. Jajodia, Eds. Springer, 2011, pp. 101–104.

[6] E. Argones Rua, E. Maiorana, J. Alba Castro, and P. Campisi, "Biometric template protection using universal background models: An application to online signature," *IEEE Trans. on Information Forensics and Security*, vol. 7, no. 1, pp. 269–282, February 2012.

[7] D. Craven, B. McGinley, L. Kilmartin, M. Glavin, and E. Jones, "Compressed sensing for bioelectric signals: A review," *IEEE Journal of Biomedical and Health Informatics*, no. 99, 2014.

[8] S. Hari, F. Agrafioti, and D. Hatzinakos, "Design of a Hamming-distance classifier for ECG biometrics," in *Proc. of the IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, May 2013, pp. 3009–3012.

[9] J. Daugman, "How iris recognition works," in *Proc of the Int.Conf. on Image Processing*, vol. 1, 2002, pp. I–33–I–36.

[10] R. Donida Labati, R. Sassi, and F. Scotti, "ECG biometric recognition: Permanence analysis of QRS signals for 24 hours continuous authentication," in *IEEE Int. Workshop on Information Forensics and Security (WIFS)*, Nov 2013, pp. 31–36.

[11] S. A. Israel, J. M. Irvine, A. Cheng, M. D. Wiederhold, and B. K. Wiederhold, "ECG to identify individuals," *Pattern Recognition*, vol. 38, pp. 133–142, 2005.

[12] L. Biel, O. Pettersson, L. Philipson, and P. Wide, "ECG analysis: a new approach in human identification," in *Proc. of the 16th IEEE Instrumentation and Measurement Technology Conference*, vol. 1, 1999, pp. 557–561.

[13] C. Ye, M. Coimbra, and B. Kumar, "Investigation of human identification using two-lead electrocardiogram (ECG) signals," in *Proc. of the 2010 Fourth IEEE Int. Conf. on Biometrics: Theory Applications and Systems (BTAS)*, 2010, pp. 1–8.

[14] F. Poree, J. Y. Bansard, G. Kervio, and G. Carrault, "Stability analysis of the 12-lead ECG morphology in different physiological conditions of interest for biometric applications," in *Computers in Cardiology*, 2009, pp. 285–288.

[15] D. Pereira Coutinho, H. Silva, H. Gamboa, A. Fred, and M. Figueiredo, "Novel fiducial and non-fiducial approaches to electrocardiogram-based biometric systems," *IET Biometrics*, vol. 2, no. 2, pp. 64–75, June 2013.

[16] I. Khalil and F. Sufi, "Legendre polynomials based biometric authentication using QRS complex of ECG," in *Proc. of the 2008 Int. Conf. on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP)*, 2008, pp. 297–302.

[17] K. Sidek, I. Khalil, and M. Smolen, "ECG biometric recognition in different physiological conditions using robust normalized QRS complexes," in *Computing in Cardiology*, 2012, pp. 97–100.

[18] I. Odinaka, P.-H. Lai, A. Kaplan, J. O'Sullivan, E. Sirevaag, S. Kristjansson, A. Sheffield, and J. W. Rohrbaugh, "ECG biometrics: A robust short-time frequency analysis," in *Proc. of the IEEE Int. Workshop on Information Forensics and Security*, December 2010, pp. 1–6.

[19] H. Xu and R. Veldhuis, "Binary representations of fingerprint spectral minutiae features," in *Proc. of the Int. Conf. on Pattern Recognition*, August 2010, pp. 1212–1216.

[20] A. Kong, *Palmprint Identification Based on Generalization of IrisCode*. PhD thesis, 2007.

[21] J. Proakis and D. Manolakis, *Digital signal processing*. Pearson Prentice Hall, 2007.

[22] A. Bonissi, R. Donida Labati, L. Perico, R. Sassi, F. Scotti, and L. Sparagino, "A preliminary study on continuous authentication methods for photoplethysmographic biometrics," in *Proc. of the IEEE Workshop on Biometric Measurements and Systems for Security and Medical Applications*, September 2013, pp. 28–33.

[23] L. Masek, "Recognition of human iris patterns for biometric identification," 2003.

[24] University of Rocher Medical Center, Telemetric and Holter ECG Warehouse, "E-HOL-03-0202-003," http://thew-project.org/Database/E-HOL-03-0202-003.html.