Dominant Set Approach to ECG Biometrics

André Lourenço[1,2], Samuel Rota Bulò[3], Carlos Carreiras[2], Hugo Silva[2],
Ana L.N. Fred[2], and Marcello Pelillo[4]

[1] Instituto Superior de Engenharia de Lisboa, Lisbon, Portugal
[2] Instituto de Telecomunicações, Instituto Superior Técnico, Lisbon, Portugal
[3] FBK-irst, via Sommarive, 18, I-38123, Trento, Italy
[4] DAIS, Università Ca' Foscari Venezia, Venice, Italy

Abstract. Electrocardiographic (ECG) signals are emerging as a recent
trend in the field of biometrics. In this paper, we propose a novel ECG
biometric system that combines clustering and classification methodolo-
gies. Our approach is based on dominant-set clustering, and provides a
framework for outlier removal and template selection. It enhances the
typical workflows, by making them better suited to new ECG acquisi-
tion paradigms that use fingers or hand palms, which lead to signals with
lower signal to noise ratio, and more prone to noise artifacts. Preliminary
results show the potential of the approach, helping to further validate
the highly usable setups and ECG signals as a complementary biometric
modality.

Keywords: Biometrics, ECG, Clustering, Dominant Set, Outlier De-
tection, Template Selection.

1 Introduction

The ECG is a graphical record of the bioelectrical signal generated by the heart
during each cardiac cycle. Typically, it exhibits both persistent features (such
as the average P-QRS-T morphology, and the short-term average heart rate or
average RR interval), and nonstationary features (such as the individual RR and
QT intervals, and long-term heart rate trends) [1].

A recent application for the ECG is its use for biometric recognition [2–6].
In this paradigm, the development of highly usable setups that use the fingers,
or hand palms, as the signal source, is one of the recent trends, leading to new
challenges [6]. With these atypical setups, the acquired signal has lower signal
to noise ratio, and is more prone to noise artifacts. Figure 1 illustrates an ECG
time series, its main complexes, and examples of two noise types of artifacts
(powerline noise and motion artifacts), which need to be detected and removed,
so that the recognition performance is not deteriorated.

An ECG biometric system prepared to process such signals requires the im-
plementation of new blocks, namely an outlier removal step to distinguish noisy
segments from normal heartbeats. In this paper we present a novel approach
based on *dominant set clustering* [7]. Our approach enables the simultaneous
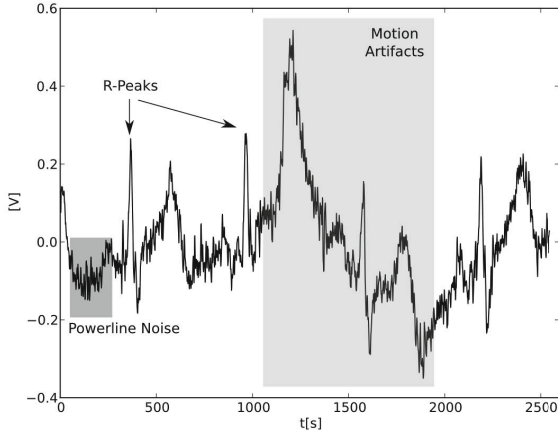removal of outlier segments, and the extraction of representative templates.

Fig. 1. Example of ECG signals acquired at the fingers with heartbeat waveforms with several R-peak complexes annotated, and corrupting noise (motion artifacts and powerline noise)

The remainder of the paper is organized as follows. In Section 2 we present an overview of ECG biometrics and a summary of our system. In Section 3 we review the dominant set clustering paradigm, and show how it can be used for outlier detection and template extraction. Finally, in Sections 4, 5 and 6 we present the experimental setup, a summary of the results, and outline the main conclusions.

2   ECG Biometric Systems

ECG-based biometric methods can be classified as fiducial, non-fiducial or partially fiducial. Fiducial methods use anchor points (called fiducia) as references [3, 5], where the R-peak of the ECG is typically the main reference, since it is the easiest to distinguish [2, 8]. Non-fiducial methods extract information from the ECG signals without having any reference point [4, 5]. Partially fiducial approaches use fiducial information only for ECG segmentation [3, 6]. We follow a partially fiducial approach, and Figure 2 depicts the block diagram of our biometric system: it starts with the acquisition of raw data, using a custom one-lead sensor with virtual ground and dry Ag/AgCl electrodes [6]; the acquired signal is then converted from analog to digital, and submitted to a data preprocessing block which performs a digital filtering step (band-pass FIR filter order 300, and cutoff frequencies 5-20Hz), and the QRS complex detection [6]; finally, the signal is segmented into individual heartbeats.

Let $\boldsymbol{x}_i$ be an individual heartbeat obtained after segmenting the ECG signal. Each heartbeat is described by a vector of features $\boldsymbol{x}_i = (x_{1i}, \ldots, x_{\mathsf{m}i})^\top \in \mathbb{R}^{\mathsf{m}}$,
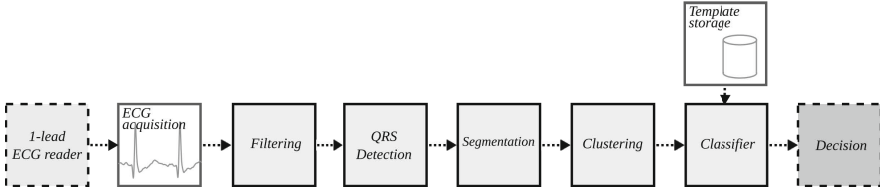
Fig. 2. Block diagram of our ECG biometric recognition approach

whose values, in this paper, are the amplitudes of the ECG heartbeat wave form. Also, consider $\mathsf{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n]$ as the $\mathsf{m} \times \mathsf{n}$ matrix formed by n individual heartbeats. The clustering block assigns each of the instances of X to a cluster. Previous work in the context of ECG biometrics has already approached the topic of outlier detection, evaluating two algorithms [9]: 1) distance-based detection (DMEAN); and 2) clustering with integrated outlier detection criterion, using DBSCAN. The approach presented in this paper is more integrated, and also enables the generation of representative templates on a cluster basis.

For classification, we compare an instance-based learning template-matching approach, through a k-Nearest Neighbor (k-NN) classifier, and with a Support Vector Machine (SVM) classifier, which have been found in the literature to perform adequately [10].

3   Robust Template Selection Using Dominant Sets

In order to select a set of ECG signal templates for a given person in a way to be robust to outliers, we make use of a graph-based clustering approach introduced in [7], known as *dominant set clustering*.

First of all, we cast the ECG clustering task for a given individual into a graph clustering problem. To this end, let $G = (V, \omega)$ be a complete, edge-weighted graph without self-loops, where $V = \{1, \ldots, n\}$ is the set of vertices, each vertex $i \in V$ being related to an individual heartbeat $\boldsymbol{x}_i$ as defined in Section 2, and $\omega : V \times V \to \mathbb{R}_+$ being a similarity function providing each edge in $(i, j) \in V \times V$ with a nonnegative weight $\omega_{ij} = \omega(i, j) \geq 0$. The weight function is defined to measure the similarity between two ECG heartbeats, assuming $\omega_{ii} = 0$ for all $i \in V$. In this work, we employ the following similarity measures:

- *Euclidean (Eucl)*: $\omega_{ij} = e^{-d_{ij}^2/\sigma^2}$, $d_{ij} = \|\boldsymbol{x}_i - \boldsymbol{x}_j\|$,
- *Cosine (Cos)*: $\omega_{ij} = \boldsymbol{x}_i^\top \boldsymbol{x}_j \big/ (\|\boldsymbol{x}_i\| \|\boldsymbol{x}_j\|)$,
- *Discrete Fourier Transform (DFT)*: $\omega_{ij} = e^{-d_{ij}^2/\sigma^2}$,
  $d_{ij} = \|\mathsf{S}^{-1}[\mathcal{F}(\boldsymbol{x}_i) - \mathcal{F}(\boldsymbol{x}_i)]\|$,

where $\sigma > 0$ is a parameter of the Euclidean and DFT similarity measures that we set to the median of the distances among all pairs of ECG signals, *i.e.* $\sigma = \mathrm{median}(\{d_{ij} : i, j \in V\})$. Moreover, $\mathcal{F}$ is the discrete Fourier transform

operator and S is a diagonal matrix, in which the diagonal holds the row-wise standard deviations of X.

In the clustering framework proposed in [7], clusters are characterized in terms of dominant sets, *i.e.* subsets of vertices generalizing the notion of maximal clique to edge-weighted graphs that encode two fundamental properties of a cluster: internal coherency and external incoherency. From a combinatorial perspective, a dominant set is defined as a subset of vertices $D \subseteq V$ satisfying the following conditions:

$$w_D(i) > 0, \quad \forall i \in D,$$
$$w_D(i) < 0, \quad \forall i \in V \setminus D.$$

Here, $w_D(i)$ is a recursive characterization of a weight assigned to each element $i \in V$ with respect to the set $D \subseteq V$ (we refer to the original paper [7] for more details). Another characterization of dominant sets is in terms of strict local maximizers of the following simplex-constrained quadratic optimization problem:

$$\max_{\boldsymbol{\alpha} \in \Delta} \boldsymbol{\alpha}^\top W \boldsymbol{\alpha} \tag{1}$$

where W is a matrix defined as $W = [\omega_{ij}]$ and Δ is the *standard simplex*, which is given by $\Delta = \left\{ \boldsymbol{\alpha} \in \mathbb{R}_+^n : \sum_{i \in V} \alpha_i = 1 \right\}$. It is indeed proven [7] that there exists a one-to-one correspondence between dominant sets of G and strict local maximizers of (1). Additionally, there exists also an interesting game-theoretic characterization of dominant sets in terms of evolutionary stable equilibria of two-person non-cooperative strategic games called *clustering games* [11, 12].

In order to extract a dominant set, we make use of the so-called *Infection and Immunization Dynamics* (InImDyn) [13]. This dynamics modifies over time an initial probability distribution (prior distribution) defined over V. At convergence, the distribution is a local solution of (1) and the elements in V having non-zero probability form a dominant set. This procedure allows to extract one dominant set at a time, and one crucial problem to solve is how to enumerate multiple dominant sets in order to detect multiple objects. A commonly used method consists in a *peeling-off strategy*, *i.e.*, one iteratively removes from the graph the elements belonging to newly extracted dominant sets. Another solution, which turns out to be effective, is the one proposed in [14, Sec. 3.4] (we refer to the paper for the details).

3.1   Extraction of the ECG Signal Templates

The problem of template extraction may be posed as follows [15]: given a set of n heartbeats, acquired during one session, extract k templates that "best" represent the variability, as well as the typically observed patterns. There are two main approaches: a) selection of k heartbeats; and b) generation of k representative heartbeats. This problem has already been tackled in other modalities (e.g. signature and fingerprint) [15, 16], and is mainly based on selection.

Our approach is based on template generation using the dominant sets. As a result of the dominant set extraction procedure mentioned in the previous

section, we obtain a set of probability distributions over the set of heartbeats $\Gamma = \{\boldsymbol{\alpha}^{(1)}, \ldots, \boldsymbol{\alpha}^{(k)}\}$. Each vector $\boldsymbol{\alpha}^{(i)} \in \Gamma \subset \Delta$ corresponds to a dominant set, and each component of $\boldsymbol{\alpha}^{(i)}$ intuitively represents the probability of being a good representative for the ith extracted cluster [7]. Based on this fact, we compute a template $\boldsymbol{t}^{(i)}$ from the ith cluster as the expected ECG heartbeat under the distribution $\boldsymbol{\alpha}^{(i)}$, $i.e.$

$$\boldsymbol{t}^{(i)} = \mathtt{X}\boldsymbol{\alpha}^{(i)} . \tag{2}$$

By repeating this operation for each dominant set extracted from the graph G of ECG signals, we obtain a set of k templates $T = \{\boldsymbol{t}^{(1)}, \ldots, \boldsymbol{t}^{(k)}\}$. Note that a dominant set represents a compact cluster and it is very robust to outliers. Indeed, outliers appear in the vector $\boldsymbol{\alpha}^{(i)}$ as zero components, thus being intrinsically not detrimental to the template computation.

In Figure 3 a) we show the single heartbeats of one of the acquisitions, together with the generated templates. In this case, 3 templates were generated, completely removing the outlier segments. In Figure 3 b) we show the α weights used for this generation (the x-axis represents the partition, and the y-axis represents the weight associated to each partition). Each of the generated templates is based on a portion of the partitions, with several partitions having a zero weight.
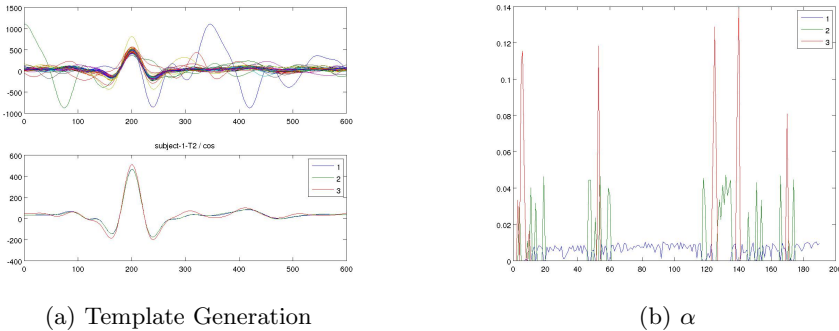

(a) Template Generation   (b) α

Fig. 3. In a) illustration of the single heartbeat waveforms for one of the acquisition sessions and the generated templates; b) α weights used for the generation of these templates

3.2   Exploitation of ECG Signal Templates for Classification

Our classification methodology follows the approach found in [10], where k-NN and SVM classifiers were tested in the context of ECG recognition. Let $\mathtt{X}^{(u)}$ denote the matrix of heartbeat waveforms of the uth individual, $u \in \{1, \ldots, \mathsf{q}\}$ and let $T^{(u)}$ be the set of ECG signal templates extracted according to the methodology presented in the previous section from $\mathtt{X}^{(u)}$. In general, different subjects can have a different number of templates. We construct a training set from the collected templates as

$$\mathbb{D}_{training} = \bigcup_{u \in \{1,\ldots,q\}} \left\{ (\boldsymbol{t}, u) : \boldsymbol{t} \in T^{(u)} \right\}, \tag{3}$$

and we train a k-NN or SVM classifier based on it.

4   Experimental Setup

For the evaluation of the proposed approach, we used a dataset consisting of the ECG data from 63 subjects (49 males and 14 females) with an average age of 20.68 ± 2.83 years. The subjects were asked to sit for 2 minutes in a resting position with two fingers, one from the left and another from the right hand, placed in each of the dry electrodes (more details in [6]). The signals were acquired using a bioPLUX research acquisition unit (12-bit resolution and 1kHz sampling frequency). The data consists of two independent acquisition sessions separated by a 3-month interval, entitled "T1" and "T2" [17].

The evaluation of our system is based on: a) analysis of the generated templates; b) a quantitative analysis on the recognition performance. We compare against the performance obtained by the DMEAN algorithm [9], when using as templates, means of 5 heartbeats.

Recognition performance was assessed for both the identification and authentication scenarios. For identification, we based the analysis on the Error of Identification (EID), which is the proportion of incorrect identifications with respect to the total number of tests (Rank-1 error). For authentication we compute, for each operating point (the threshold in k-NN, or the fraction of agreeing models in SVM), the False Acceptance Rate (FAR) and the False Rejection Rate (FRR), which are used to determine the Equal Error Rate (EER). Regarding the classifiers, for the k-NN, we used $k = 1$, and for the SVM we used a linear kernel.

5   Results

Table 1 shows the biometric recognition performance (EER and EID) of the proposed approaches (dominant sets with *Cos*, *DFT* and *Eucl*), against baseline (*DMEAN*), within-session T1 (T1-T1), and across-sessions (training in T1, and testing in T2; T1-T2). For the T1-T1 analysis, we employed cross-validation using 30% of T1 as training data and the rest as testing data, over 30 runs.

We can observe that across-sessions results are much worse than within-session results, which was expected given the fact that, in the within-session case, the training and testing data belong to the same acquisition session and, therefore, the extracted ECG segments exhibit greater similarity, especially in regard to the heart rate. In the across-sessions case, the use of the template generation method produces recognition rates with worse performance, but in the same order of magnitude, when compared with what is obtained with the DMEAN method, which only performs outlier removal. This may be a consequence of having fewer templates than DMEAN approach. The same is not true for the within-session

case. In particular, the *Eucl* metric using the SVM classifier outputs the global best results. Concerning the three adopted metrics (*Cos*, *DFT* and *Eucl*), the results are similar (within each of the train and test conditions), with a preference for the *Eucl* metric. Finally, regarding the classification method, the SVM classifier performs better in all studied cases.

These results suggest that, although the template generation method does not improve the recognition performance for the across-sessions case (the more realistic one), we can still gain from its use in the sense that it reduces the number of templates that are required to train the classifiers, thus achieving better computational efficiency (i.e. the mean number of templates used for training in the *Eucl* is 6.9 templates per subject). Additionally, a fine-tuning of the parameters may lead to better performance.

Table 1. Equal Error Rates (EER) and Identification Errors (EID) in %; proposed proposed approach (DS) vs. DMEAN, baseline presented in [9], without template generation; values in bold represent the best score for each column

Train Method		1-NN				SVM		
		T1-T1		T1-T2		T1-T1		T1-T2	
		EER	EID	EER	EID	EER	EID	EER	EID
DMEAN		2.53	**1.01**	**11.90**	**41.57**	1.00	1.22	**9.47**	**38.30**
DS	Cos	2.66	1.90	13.91	47.67	1.26	1.61	11.76	46.45
	DFT	**2.48**	1.28	12.46	44.07	0.99	1.21	11.10	40.63
	Eucl	2.51	1.43	12.31	42.57	**0.75**	**0.86**	10.86	39.91

6   Conclusion

Research on Electrocardiographic (ECG) signals has advanced a long way from its clinical roots, to novel application domains in areas so diverse as biometric recognition. The development of highly usable setups based on fingers or hand palms leads to signals with lower signal to noise ratio, and more artifacts (e.g. motion induced).

The biometric recognition sytems adapted to this new paradigm require the inclusion of outlier detection blocks; furthermore, given that heartbeat waveforms show high intra-subject morphological similarity, template generation can also be a nice add-on to ECG-based biometric systems. In this paper we presented a novel approach based on dominant set clustering. The described system enables the simultaneous generation of templates and removal of outliers. Experimental results have shown that our approach enables high recognition rates comparable to other approaches that also integrate outlier detection, but no template extraction step. Therefore, the main advantage the proposed approach is the improvement of the computational efficiency of the biometric recognition system, given that the number of templates that have to be stored per subject is much smaller, significantly reducing the number of comparisons that have to be made.

References

1. Clifford, G.D.: ECG Statistics, Noise, Artifacts, and Missing Data. In: Clifford, G.D., Azuaje, F., Mcsharry, P. (eds.) Advanced Methods and Tools for ECG Data Analysis. Artech House Publishers (2006)
2. Biel, L., Petterson, O., Phillipson, L., Wide, P.: ECG analysis: A new approach in human identification. IEEE Trans. Inst. and Measurement 50(3), 808–812 (2001)
3. Wang, Y., Agrafioti, F., Hatzinakos, D., Plataniotis, K.N.: Analysis of human electrocardiogram for biometric recognition. EURASIP J. Adv. S. Processing (2008)
4. Chan, A.D.C., Hamdy, M.M., Badre, A., Badee, V.: Wavelet distance measure for person identification using electrocardiograms. IEEE Trans. on Instrumentation and Measurement 57(2), 248–253 (2008)
5. Odinaka, I., Lai, P.H., Kaplan, A., O'Sullivan, J., Sirevaag, E., Rohrbaugh, J.: ECG biometric recognition: A comparative analysis. IEEE Trans. on Information Forensics and Security 7(6), 1812–1824 (2012)
6. Silva, H., Lourenço, A., Canento, F., Fred, A., Raposo, N.: ECG biometrics: Principles and applications. In: Proc. of the 6th Int'l Conf. on Bio-Inspired Systems and Signal Processing, BIOSIGNALS (2013)
7. Pavan, M., Pelillo, M.: Dominant sets and pairwise clustering. IEEE Trans. Pattern Analysis and Machine Intelligence 29(1), 167–172 (2007)
8. Silva, H., Gamboa, H., Fred, A.: One lead ECG based personal identification with feature subspace ensembles. In: Perner, P. (ed.) MLDM 2007. LNCS (LNAI), vol. 4571, pp. 770–783. Springer, Heidelberg (2007)
9. Lourenço, A., Silva, H., Carreiras, C., Fred, A.: Outlier detection in non-intrusive ECG biometric system. In: Kamel, M., Campilho, A. (eds.) ICIAR 2013. LNCS, vol. 7950, pp. 43–52. Springer, Heidelberg (2013)
10. Lourenço, A., Silva, H., Fred, A.L.N.: ECG-based biometrics: A real time classification approach. In: IEEE Int'l W. Machine Learning for Signal Proc. (2012)
11. Torsello, A., Rota Bulò, S., Pelillo, M.: Grouping with asymmetric affinities: A game-theoretic perspective. In: IEEE Conf. Computer Vision and Patt. Recogn., pp. 292–299 (2006)
12. Rota Bulò, S., Pelillo, M.: A game-theoretic approach to hypergraph clustering. IEEE Trans. Patt. Analysis Machine Intell. 35(6), 1312–1327 (2013)
13. Rota Bulò, S., Pelillo, M., Bomze, I.M.: Graph-based quadratic optimization: A fast evolutionary approach. Comp. Vis. and Image Understanding 115, 984–995 (2011)
14. Kontschieder, P., Rota Bulò, S., Donoser, M., Pelillo, M., Bischof, H.: Evolutionary hough games for coherent object detection. Comp. Vis. and Image Understanding 116, 1149–1158 (2012)
15. Uludag, U., Ross, A., Jain, A.: Biometric template selection and update: a case study in fingerprints. Pattern Recognition 37(7), 1533–1542 (2004)
16. Liu, N., Wang, Y.: Template selection for on-line signature verification. In: Proc. of the 19th Int. Conf. on Pattern Recognition (ICPR), pp. 1–4 (December 2008)
17. Silva, H., Lourenço, A., Fred, A.L.N.: Finger ECG signal for user authentication: Usability and performance. In: IEEE BTAS (September 2013)