C. Camara, P. Peris-Lopez, and J. E. Tapiador (2015). Human Identification Using Compressed ECG Signals. Journal of Medical Systems, 39: 148. Avalaible in https://doi.org/10.1007/s10916-015-0323-2

Human Identification Using Compressed ECG Signals

Carmen Camara[1] · Pedro Peris-Lopez[1] · Juan E. Tapiador[1]

Abstract As a result of the increased demand for improved life styles and the increment of senior citizens over the age of 65, new home care services are demanded. Simultaneously, the medical sector is increasingly becoming the new target of cybercriminals due the potential value of users' medical information. The use of biometrics seems an effective tool as a deterrent for many of such attacks. In this paper, we propose the use of electrocardiograms (ECGs) for the identification of individuals. For instance, for a telecare service, a user could be authenticated using the information extracted from her ECG signal. The majority of ECG-based biometrics systems extract information (fiducial features) from the characteristics points of an ECG wave. In this article, we propose the use of non-fiducial features via the Hadamard Transform (HT). We show how the use of highly compressed signals (only 24 coefficients of HT) is enough to unequivocally identify individuals with a high performance (classification accuracy of 0.97 and with identification system errors in the order of 10^{-2}).

Keywords Healthcare · Biometrics · Human Identification and ECG

✉ Pedro Peris-Lopez
pperis@inf.uc3m.es

Carmen Camara
macamara@pa.uc3m.es

COSEC Lab (Computer Science Department), Carlos III University of Madrid, Avda de la Universidad 30, 28911, Leganes, Spain

Introduction

According to [3], the medical sector is the area that has suffered the major number of hacking incidents over the last two years—43 % of the data breaches in US. Medical companies and hospitals have begun to introduce biometric solutions to mitigate attacks and reduce costs. Furthermore, the proper identification of patients when they walk through the door is a major issue nowadays for all the hospitals around the world. Errors in medical records, or even incorrect treatments, are very costly for the medical centres and harmful for the patient. To avoid this, novel solutions propose to maintain a link between the patient's data biometrics and her medical record. Thus, the biometric signature (monomodal or multimodal) is used as an index to recover the patient's medical record in the standard way: the system compares the master template with the one read locally and, if they match, the associ-ated medical record is retrieved. This process is entirely done locally but may be also done remotely, i.e., the user would provide her biometrics data remotely. In this sense, biometrics could accelerate the transition towards home health care [25].

Home health care allows the treatment of a disease at home. On the other hand it is usually as effective as care at the hospital but less expensive and more convenient for the patient. A wide variety of health care services can be offered (e.g., check your vital signs like temperature or blood pressure remotely or have a video conference with a medical staff). Demographic changes (ageing population), social changes (small family units or mobility across countries), and developments in science and technology are some indicators that help to forecast, in a near future, a spread-use of home health care services [32]. The correct and secure identification of each individual is a key-point for the proper

operation of these systems. We propose the use of a biometric solution for that purpose. For completeness, we next provide a brief introduction to biometrics.

Biometrics refers to the automatic identification of users based on features derived from their physiological and/or behavioural characteristics. The use of such features for identification (authentication) or verification purposes has been thoroughly explored in the last 30 years. In verification, an identity is provided by the user, which is used to retrieve a master template. The master template is then compared with the verification template (1-to-1 comparison) and a matching score is returned by the classifier. Contrarily, in identification systems like the one proposed in this article, the identity of a user rests entirely solely on her biometric information—the classifier performs one-to-many comparisons.

There is a substantial body of knowledge on recogniz-ing subjects by their fingerprint, face, voice, gait, keystroke dynamics, hand, iris, or retina [11]. Depending on the appli-cation and operational context, each one of these features can be used separately [6, 9] or combined in a multi-biometrics setting [33]. The accuracy, measured both as the probability of identifying a correct subject and rejecting a false individual, is possibly the single most important fea-ture of a biometric system. However, in practice there are other properties that can severely limit the use of a particular identification technique [21]. The biometric characteristic must be also universal, stable, and unique. Its acquisition has to be easy and without objections by the users. Finally bio-metrics systems should detect the use of an artefact or sub-stitute. All the mentioned characteristics have been assessed against our proposed system in section "Discussion".

Over the last few years, some works have explored the biometric use of signals that, for different reasons, have traditionally received little attention by the security community. Biosignals—i.e., electrical or chemical signals measuring some activity or parameter of the human body-constitute an important class of such signals, including elec-trocardiograms (ECGs), electroencephalograms (EEGs), and electromyograms (EMGs). These signals have been thoroughly studied for medical applications, on the hypothe-sis that they convey information about different pathologies and, consequently, can be used as a valuable diagnostic tool. For example, automatic classification of ECG signals assists cardiologists to diagnose arrhythmias (i.e., tachycar-dia, bradycardia or atrial fibrillation) [13].

In the last years, several works have demonstrated that many vital signals also contain features unique to the individual and can be used for security purposes. This branch of the biometrics is increasingly referred to as *Intrinsic* or *Hidden Biometrics* [17]. For instance, the electrical activity produced by skeletal muscles can be used for biomet-rics. EMG is the technique used for measurement and the obtained record is called electromyogram. In [30], Suresh et al. proposed the use of electromyograms to generate a signature for human identification. For that, impulsional electrical stimulation is produced over the muscle and its response constitutes the signature. This proposal has been successfully tested over a population of ten individual.

ECG and EEG signals are by far the most commonly studied signals for Hidden Biometrics. EEG records the electrical activity in the brain through a set of electrodes mounted on the scalp. Existing proposals can be grouped according to the classification algorithm used. Linear discriminant classifiers with auto-regressive feature extraction are demonstrated in [20]. In [2], an LVQ neuronal net-work with FFT feature extraction is described, while the work in [29] reports results using a neuronal network with energy feature extraction. On the other hand, EEG records the electrical activity of the heart. The algorithms can be classified according to the features extracted from the ECG signal. Fiducial-based methods extract information from the characteristics points of an ECG wave (e.g., amplitude [7], temporal duration [10]). Non-fiducial methods do not use the characteristics points to extract features. Instead, other features like autocorrelation [1], Fourier [23] or Wavelet coefficients [4] are used. Other solutions (hybrid) combined both methods like in [34] or in [26]. The reader is urged to consult for an exhaustive survey of ECG-based biometrics proposals [18].

The rest of the paper is organized as follows. In section "Methods" the general architecture of the biometric identification system is presented. After that, we review and explain each of its forming components. The results are pre-sented in section "Results". Then, in section "Discussion", we evaluate the main properties of the proposed system. Finally, in section Conclusions and Implications", we pro-vide reasoning about the implications of our proposal and extract some conclusions.

Methods

In Fig. 1 we show the general architecture of a biometric identification system. The first step consists of the data acquisition-—one or several signals take part depending on whether the system is mono-modal or multi-modal, respectively. Usually a set of sensors are placed over the subject (e.g., chest or head) to read the biosignals. Once acquired, the raw data must be prepared for its analysis. Techniques such as normalization, re-sampling or smoothing are commonly used procedures during the pre-processing step. After that, the more relevant information of the signal is represented by a set of numerical o nominal parameters. This step is usually known as feature extraction and is crucial for the success of the whole process. The generated dataset

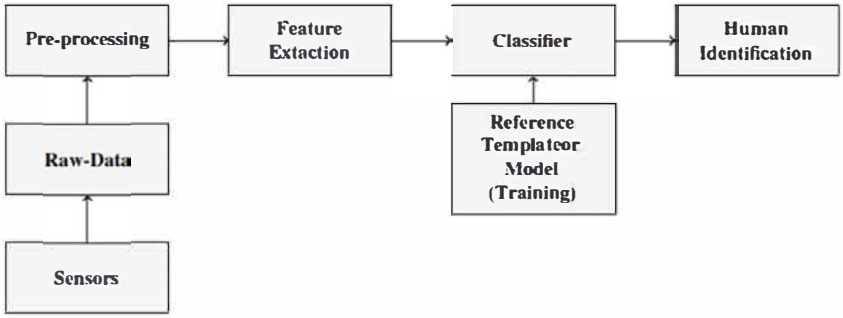

Fig. 1 General structure of a biometric identification system

consists of a number of instances, each one formed by a set of features and a label corresponding to an individual. The aforementioned dataset is split into two subsets for training and testing, respectively. The training set is employed to build the model and the unseen samples (testing set) are used to evaluate the model. That is, for each instance the model outputs a label that is compared to the ground truth. Depending on its success, the classification accuracy will be higher or lower. In the following, each one of these building blocks are explained in more detail taking into consideration the particular procedure used in this article.

Raw Data and Pre-processing

The electrocardiogram (abbreviated as ECG and sometimes EKG) consists of a measurement over the skin surface to record the electrical activity of the heart. The conduction of ions through the myocardium (heart muscle) change with each heart beat. The ECG represents the sum of the action potentials of millions of cardiomyocyte (heart cells).

For our experimentation, we have chosen a well-known dataset. In particular, the MIT-BIH Normal Sinus Rhythm Database is used [8]. It includes long-term recordings of 18 subjects treated at Boston's Beth Israel Hospital. The decision of using this dataset was motivated for the fact that no significant arrhythmias were detected in the subjects. Therefore, the subjects do not present any bias that could help in the identification task.

The heart rate of a person at resting varies from 60 to 100 beats per minute. In order to pre-process the signal, at the fist step the DC components are eliminated. After that, each ECG signal is filtered using a passband filter. The passband range is often governed by the intended application: for instance, [0.05Hz – 150Hz] for diagnostic and [0.67Hz – 40Hz] for patient monitoring. In our cause, we use passband filter with passband rage between 0.67 and 45 Hz. The

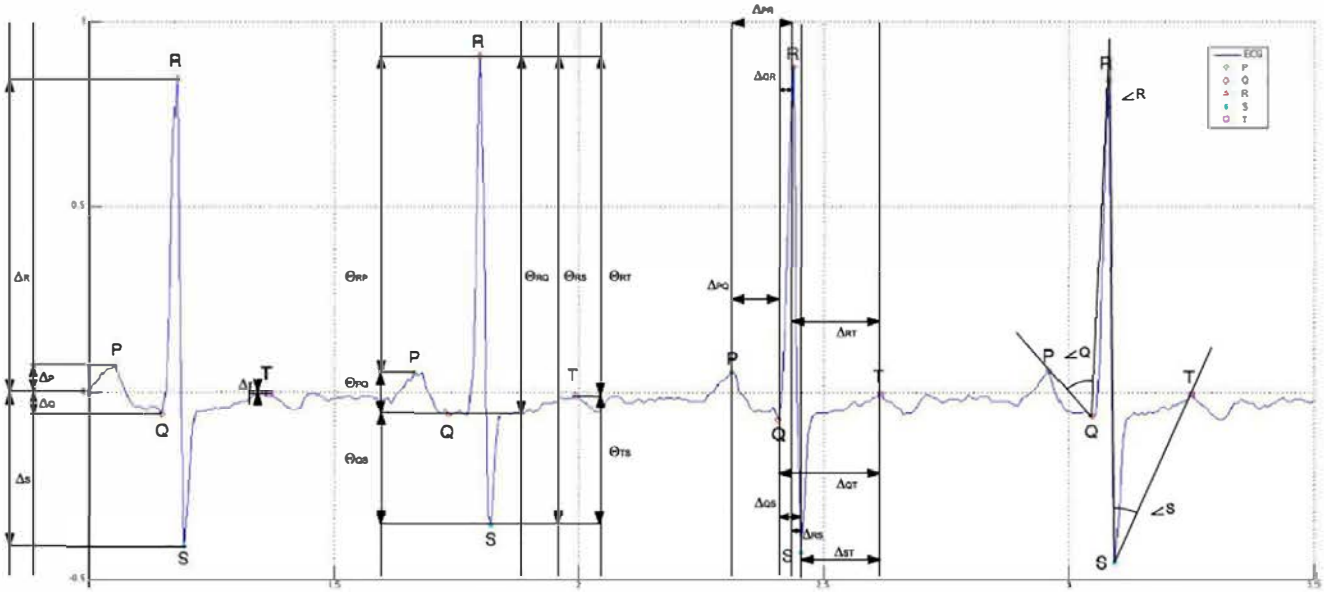

Fig. 2 ECG wave: characteristic points and features

lower cut-off frequency is set 0.67 to eliminate the noise introduced by the respiration of the subject. The upper cut-off frequency is set to 45 Hz to keep as much information as possible and to eliminate the power line noise.

Once the signal is filtered, we split the signal in chunks without overlapping. The chunk size is set to 2 seconds, which means that each chunk consists on 2-3 heart beats. We have chosen this size inspired by the fact that algorithms based on fiducial features often use two beats as chunk length. We attempt to obtain similar information by using a compressed version of the signal. The non-fiducial features used in our experimentation are explained in the next subsection.

Feature Extraction

Features derived from biosignals are effective in the design of human identification systems [14, 16]. ECG signals are one of the most used for this purpose [12]. Generally, exist-ing algorithms can be classified into two main groups [18]. On the hand hand, the algorithms based on fiducial fea-tures use characteristic points (e.g. PQRST peaks) from a ECG trace to extract a set of features (e.g., time intervals between peaks or angles). In Fig. 2 we show the main characteristic points together with the most common features of an ECG wave. Contrarily, algorithms based on non-fiducial features do not employ characteristic points for generating the feature set.

In this article, we propose the use of a non-fiducial based algorithm. In particular, the Hadamard Transform (HT) is used to extract the features of an ECG wave. Figure 3 shows the ECG signal in the time domain and in the Hadamard domain, respectively. In the same way as the Fourier Transform (FT) consists of a projec-tion onto a set of orthogonal sinusoidal waveforms, the Hadamard Transform (HT) lies in a projection onto a set of square waves called Walsh functions. In fact, the Hadamard transform is often called Walsh-Hadamard trans-form, since the base of the transformation consists of Walsh functions.

The Discrete Walsh-Hadamard Transform (DWT) of a data sequence $x(n)$ and $n = \{1 \cdots N\}$ is given by:

$$X_w(k) = \sum_{n=0}^{N-1} x(n) \prod_{i=0}^{M-1} (-1)^{n_i K_{M-1-i}}, k = 0, 1, \cdots, N-1 \quad (1)$$

where N is the number of samples of the data and restricted to be a power-of-2, and $M = \log_2 N$. Therefore, in a simply way, the transform (X_w) consists on the product of the

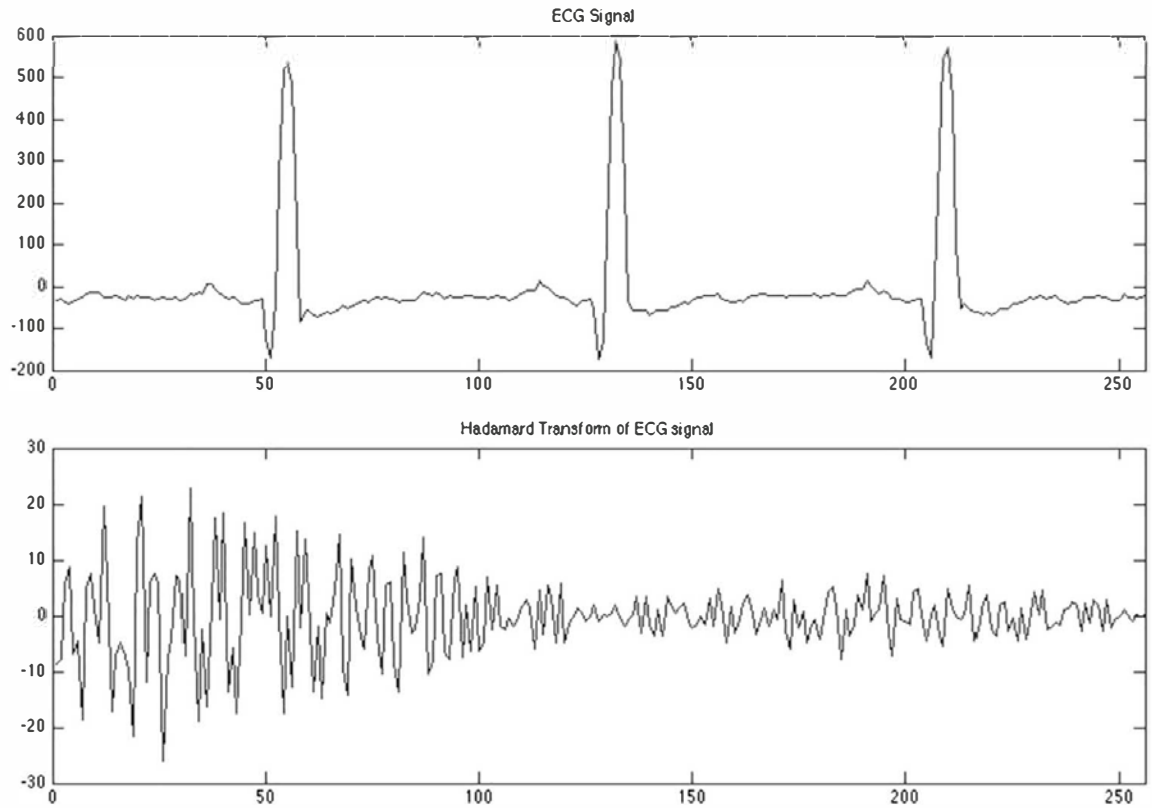

Fig. 3 ECG wave in the time domain and its Hadamard spectrum

sequence (x) of length $1 \times N$ by the Walsh matrix (H) with length $N \times N$:

$$X_w = Hx \tag{2}$$

The inverse of the transform can be easily calculated with the next analogous expression that only differs in the constant divisor:

$$x(n) = \frac{1}{N} \sum_{k=0}^{N-1} X_w(K) \prod_{i=0}^{M-1} (-1)^{n_i K_{M-1-i}},$$
$$n = 0, 1, \cdots, N-1 \tag{3}$$

One advantage of using this transform is that it is computationally more efficient than others, such as the Fourier Transform or the Wavelet Transform. This is important in constrained devices with limited computational capabilities. On the other hand, the usage of this transform facilitates that a compressed version of the signal could be stored, while this compressed signal preserves all the informational of the ECG signal and allows the reconstruction of the signal in the time domain.

To show the effectiveness of the HT with ECG signals, we have studied the effect of compressing the signal. To illustrate this, an ECG wave of 256 samples has been used. The HT is computed over this signal and 256 coefficients are obtained. After that, we have taken fractions of these coefficients (i.e., $\{X_w(0), \cdots X_w(P)\}$ and $P = \{256, 128, 32, 16, 8\}$) and calculated the inverse of the transform to reconstruct the signal. The results of the reconstructed signals are shown in Fig. 4, which illustrate how the signal can be highly compressed while preserving the signal' s main characteristics.

In Fig. 5 we sketch the feature extraction procedure. As shown, the features used in our proposed ECG-system are mainly based on the coefficients of the HT. In particular, the 24 lower sequencing coefficients has been used—see section "Results" for details. Furthermore two additional features have been computed over the whole set of HT coefficients. Shannon entropy (E_{SH}) and Log-Energy entropy (E_{LE}) are the two features chosen in our experimentation-other features like statistical metrics were tested but finally discarded. Let x a signal and $X(n)$ the coefficients of x in a orthogonal base, both entropies can be calculated as follows:

$$E_{SH} = -\sum_n X(n)^2 \log(X(n)^2) \tag{4}$$

$$E_{LE} = -\sum_n \log(X(n)^2) \tag{5}$$

Classifier

Inductive machine learning uses the concept of learning by example. A system infers a set of rules from a set of input instances (training set). Once the model is generated, the built model can be used to classify unseen instances (testing set). There is a wide range of classification algorithms and the choice of one or another is determined by the nature of the problem, the dataset characteristics and the application where it will be used. Taking into consideration is function or form, classifiers can be categorized in numerous types,

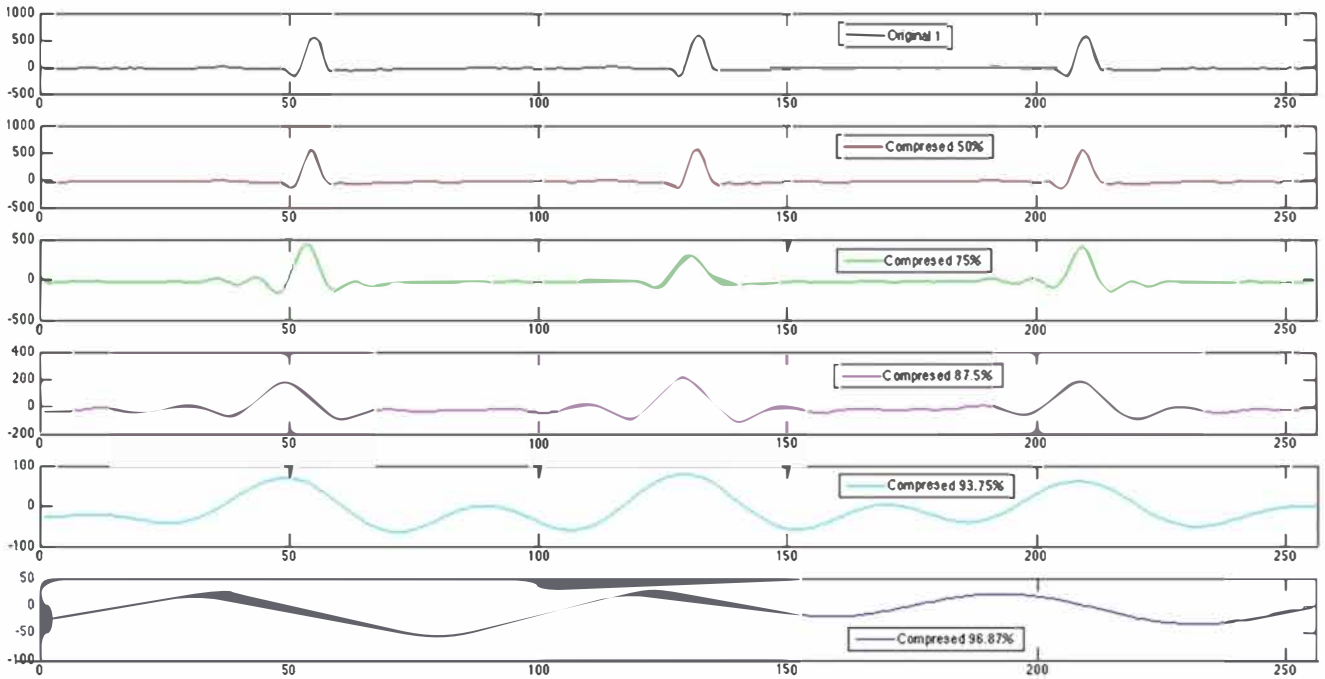

Fig. 4 Reconstructed ECG signal after compression via the HT

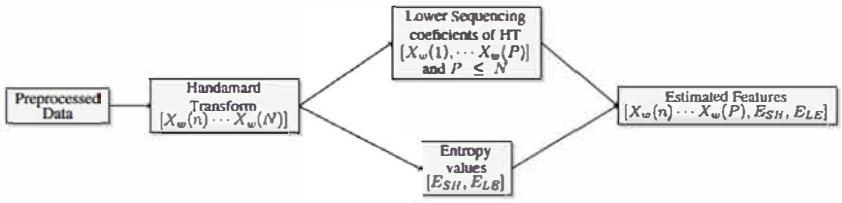

Fig. 5 Feature extraction procedure

including decision tree learning algorithms, kernel methods, lazy learning algorithms, etc.

In this article we use a K-NN algorithm, which fits within the category of non-parametric lazy learning algorithms. Non-parametric refers to the fact that they avoid making assumptions about the data distribution. Lazy means that the training instance are not used to do a generalization, so the training is minimal. The K-NN algorithm makes several assumptions: 1) the instances are in a metric space (i.e., scalars or multidimensional vectors) and distance metrics can be computed between two instances; 2) each instance in the training set is composed of a vector (set features) and a label; and 3) the parameter K determines how many neighbours are considered for classification.

The testing and training phases for the K-NN algorithm are as follows. In the training phase, features vectors with its corresponding class are stored. In the classification phase, let y_j an unseen instance and $\{x_0, \cdots x_k\}$ the K nearest training instances. The label of y_j is determined by majority voting among the labels of its K neighbours.

K-NN has been chosen since it is simple but effective. We have tested several values of the K parameter and finally it has been set to 1. In fact, using higher values for K (i.e., $K = \{3, 5, 9\}$ we do not observe any improvement in the performance while the cost in terms of computational load is significant. Regarding the distance metrics, Euclidean distance (d_E) and Manhattan distance (d_M) have been evaluated. Let two vectors $x = [x(1) \cdots x(N))]$ and $y = [y(1) \cdots y(N))]$, both metrics are defined as follows:

$$d_E = \sqrt{\sum_{i=1}^{N} [x(i) - y(i)]^2} \tag{6}$$

$$d_M = \sum_{i=1}^{N} |x(i) - y(i)| \tag{7}$$

Results

The algorithm proposed in this article fits within the algorithms based on non-fiducial features. The main difference in comparison with its predecessors is that the algorithm works with a compressed version of the original signal via the Hadamard transform. Furthermore, only a small fraction of all the coefficients are necessary for human identification. Since the number of used coefficients—24 coefficients for each ECG chunk—is effective for identification but insufficient to recover the original signal and to preserve its characteristic points, the proposed system is privacy preserving for the user.

The procedure followed for the analysis of the ECG signal is the one explained in section "Methods" and sketched in Fig. 1 . For our experimentation, we use the well-known MIT-BIH Normal Sinus Rhythm database. In particular, ECG signals for two electrodes are available and were pre-processed as explained in section "Raw Data and Pre-processing". Thereafter the same procedure is followed for each electrode signal. The signal is chopped in chunks of 256 samples and for each chunk the DWT is computed. Finally the feature extraction has been evaluated using two approaches:

- Hadamard Coefficients: Only a small fraction of the coefficients are necessary for the identification task. This number has been obtained through experimentation and using the classification accuracy as the metric for comparison between the possible values. After conducting some experiments, we have set this value to 24, which represents less than 10 % of the coefficients. Therefore using these 24 lower sequencing coefficients (48 in total considering the two leads) the system can identify an individual with high accuracy. On the other hand, and considering the worst case in which an attacker would capture these coefficients, she could not reconstruct the original signal as shown in the Fig. 6—only partial information might be retrieved.
- Entropy: Although the system offers a high performance using Hadamard coefficients, we have studied whether additional features are useful for the system. In particular, we have calculated the Shannon and the Log-Energy entropy over the whole set of Hadamard coefficients (i.e., 256 values). It is also worth mentioning that we also tested the inclusion of commonly used statistical metrics (e.g., mean, standard deviation, maximum, minimum, and first derivative, etc.). Nevertheless its benefit over the performance of the system is negligible and for this reason these features were not finally considered in our experimentation.

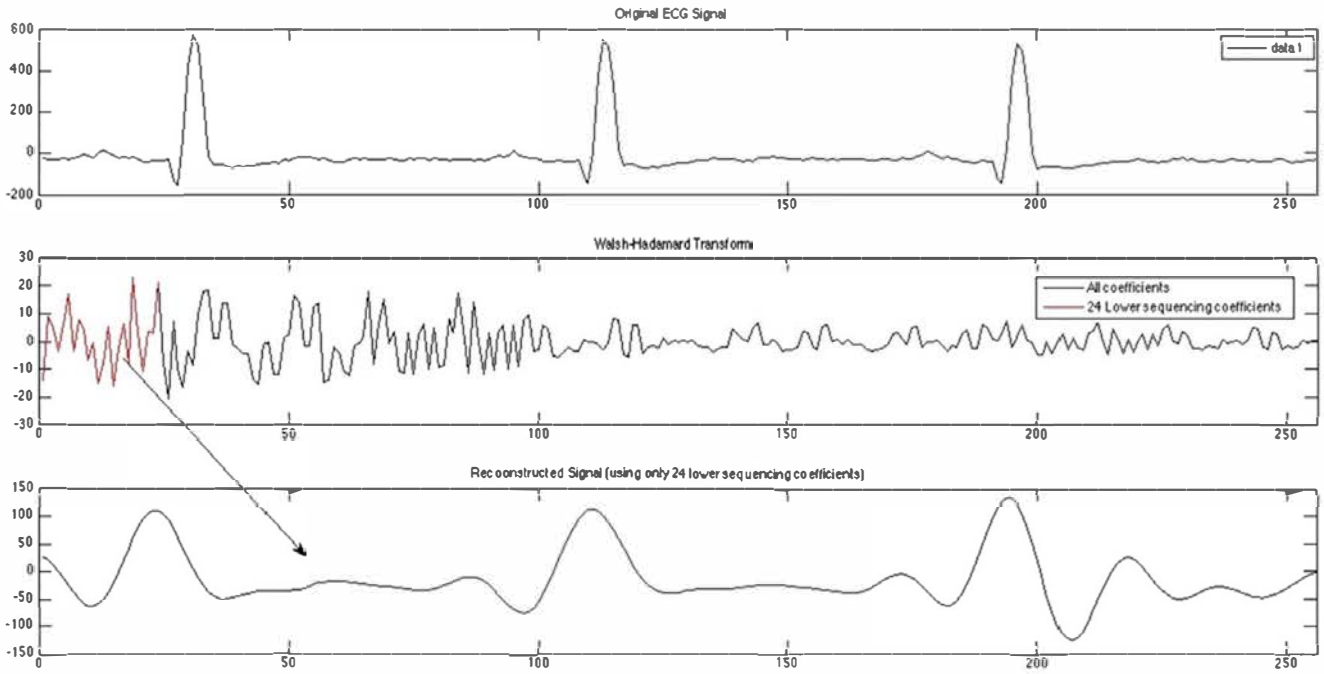

Fig. 6 ECG signal: original, transformed (via HT), reconstructed

Once the features are extracted, we have trained and tested a 1-NN classifier. We have used 10-fold cross-validation in order the classifier can accurately predict unknown data. Each instance consists of a set of features and a label corresponding to the subject (from 1 to 18). Regarding the features employed we have tested two possible configurations: OP-1 only considers 24 lower sequencing coefficients of the HT—in total 48 features taking into consideration the two leads available; and OP-2 considers the same features as OP-1 plus the Shannon and the Log-Energy entropy (4 additional features considering the two leads). For each configuration, the 1-NN classifier has been evaluated using two distances metrics: Euclidean and Manhattan.

The confusion matrix obtained for each configuration can be summarized through the true positives (TP) and false negative (FN) rates and its corresponding complementary values, false positive (FP) and true negative TN rates, respectively. The obtained values are summarized in Table 1. Using these values, the performance of the proposed ECG-based human identification system can be assessed through a number of standard metrics:

Table 1 Overall Performance: False Negative (FN), False Positive (FP), True Positive (TP) and True Negative (TN) rates

Configuration	FNR	FPR	TPR	TNR	
OP-1	d_E	0.0580	0.0582	0.9418	0.9420
	d_M	0.0570	0.0566	0.9434	0.9430
OP-2	d_E	0.0390	0.0386	0.9614	0.9610
	d_M	0.0340	0.0341	0.9659	0.9660

- *Classification Accuracy*. Measures the proportion of correct outputs, both positive and negative:

$$CA = \frac{TP + TN}{TP + FP + FN + TN} \tag{8}$$

- *Sensitivity*. It is simply the true positive rate, i.e., the proportion of actual positives that are correctly identified as such:

$$ST = \frac{TP}{TP + FN} \tag{9}$$

- *Specificity*. Also known as the false positive rate, measures the proportion of actual negatives that are correctly identified as such:

$$SP = \frac{TN}{FP + TN} \tag{10}$$

- *Positive Predictive Value*. Also known as precision, measures the proportion of positive outcomes that are actually positive:

$$PPV = \frac{TP}{TP + FP} \tag{11}$$

- *Negative Predictive Value*. Measures the proportion of negative outcomes that are actually negative:

$$NPV = \frac{TN}{FN + TN} \tag{12}$$

In Table 2 the performance of the proposed system, in its four possible configurations, is summarized. In the next

Table 2 Performance metrics

Configuration	CA	ST	SP	PPV	NPV	
OP-1	d_E	0.9419	0.9420	0.9418	0.9418	0.9420
	d_M	0.9432	0.9430	0.9434	0.9434	0.9430
OP-2	d_E	0.9612	0.9610	0.9614	0.9614	0.9610
	d_M	0.9659	0.9660	0.9659	0.9659	0.9660

section we evaluate the proposed system from a biometric point of view and extract some conclusions about the performance and what is the most recommended configuration.

Discussion

In the above section we have shown the results of our proposal regarding its performance. Seven characteristics (including performance) are commonly demanded to a biometric system [21]: universality, uniqueness, permanence, collectability, acceptability, performance, and resistance to circumvention. In the following each of these characteristics is analyzed for our proposed system:

Universality The biometric characteristic must be universally applicable. In our case, we use the ECG signal, which can be collected from everyone who is alive. Normal values for a person at resting are in the rage of 60 to 100 beats per minute.

Uniqueness The biometric characteristic must be able to unequivocally identify the individuals within the target population. In this article we have proposed the use of the ECG. This signal has already been proved to be effec-tive for biometrics purposes [5, 19]. In our case, we have checked whether features obtained from a compressed ECG signal (via Hadamard transform) can be used to identify individuals. As shown in Table 1, the number of misclassified samples is almost zero for all the con-figurations evaluated. This is a clear indicator about the effectiveness of the Walsh coefficients (lower ones) for the human identification task.

Permanence The biometric characteristic should be invariant over time. Nevertheless, physiological characteristics are not totally invariant during the entire life of an individual. In our case, small variations appear in ECG signals after a five years period. This means that the clas-sifier model would have to be updated after five years since the model was generated. If we compare our sys-tem with other common solutions based on passwords [24], in which the user normally must update the pass-word once per year, our proposed solutions is five times less demanding in terms of updating requirements.

Collectability The biometric characteristic should be quantitatively measurable. In our particular case, ECG signals can be easily gathered through a set of electrodes—3-lead or 12-lead system. Using these electrodes, the electrical activity of the heart can be recorded. More precisely, the ECG represents the potential differences between electrodes.

Acceptability It relates to how the user feels comfortable with the use of the biometric characteristic. We cannot do a strong presumption about this matter since we use a public dataset for our experimentation. Nevertheless, we can predict a high acceptability due to two main reasons: 1) the ECG signal is well-known to deal with heart diseases; and 2) the signal can be easily acquired—just three leads are sufficient for non-medical applications.

Performance Our proposed system offers a high accuracy level. The classification accuracy varies from 0.94 to 0.97 for the two configurations evaluated. Furthermore, and not less important, the identification system errors (i.e., false positive and false negative identification rates) are very low values: of the order of 10^{-2}. In relation with the distance metric, the Manhattan distance seems to offer slightly better results for the 1-NN classifier than the Euclidean distance. From the computational point of view, OP-1 is less demanding since only Hadamard coefficients are necessary and the penalty in performance is small in comparison with OP-2.

We can compare our system with other proposals with similar results (see Table 3). Most other ECG-based biometric solutions achieve similar performance. Never-theless, the main contribution of this article is the set of features used. Fiducial features has been proven to be effective but its calculation requires moderate computa-tional capabilities [34]. In our case we use non-fiducial features trough the computation of the DWT. A matrix with ones and minus one values has to be stored in mem-ory, in what is called the Wasih matrix. Note that the matrix size is fixed since the length of the ECG chunks does not vary. In our particular case, we have set this parameter through experimentation aiming at optimizing

Table 3 Biosignal-based authentication proposals

System	Correctly Classified Instances
Our system	94 % (OP-1) – 97 % (OP-2) %
ECG [18]	86 % – 100 % (single day data acquisition)
EEG [28]	72 % - 80 % (4-40 individuals)
EEG and ECG [22]	97.9 % (linear boundary)
Pulse-Response [21]	88 % –100 % (small data set)
Finger-vein [35]	98 % (70 individuals)
Iris and Fingerprint [15]	96 % (small dataset)
Face & Iris [27]	99 % (UBIRIS v.2 and ORL)

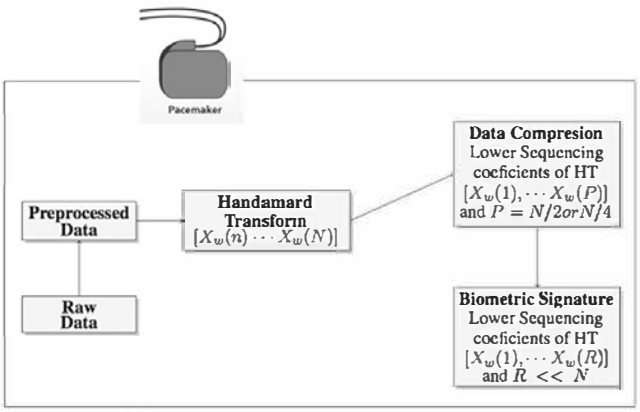

Fig. 7 Pacemaker with data compression and biometrics signature modules

system performance. The Walsh Hadamard coefficients are obtained just by multiplying a vector (an ECG chunk of 256 samples) by the Walsh matrix (a matrix of size 256 x 256 with ones and minus ones). The complexity of this naive algorithm is $O(N^2)$ but this can be reduced to $(N \log N)$ using the Fast Walsh-Hadamard Transform.

Resistance to Circumvention This property is vital for an identification system. The biometric characteristic should prevent an attacker from impersonating an authorized user in the database. In our proposed system, this property is satisfied since the ECG signal (the complete wave) is characteristic of each person. Note that two persons can have identical heart rates but their ECG waves will be different. Previous studies have confirmed this matter and it is commonly assumed that the features of an ECG signal are mainly resistant against counterfeiting [31].

It is clear from all the above that the proposed system satisfied the characteristics required of a biometric system. Therefore the use of compressed ECG signals via Hadamard Transform is robust, effective, and efficient for human identification. We next provide reasoning about the implications of our proposal and extract some conclusions.

Conclusions and Implications

The integration of smart homes and telecare services aims to improve quality life and the possibilities for independent living through the use of new technologies and services. Smart devices at home pursue to increase comfort, energy efficiency, and security. On the other hand telecare services allow people to stay in their homes without prejudicing the quality of the health care services they are getting. The proper identification of the users is crucial to secure the systems. This task can be done through the features extracted from vital signals. Since the ECG signal is often monitored for medical purposes, we can take advantage of this and use also this vital signal for security purposes (e.g., identification or key generation). In our proposal we show how compressed ECG signals are robust and effective to unequivocally identify individuals.

In section Discussion we have evaluated the seven characteristics commonly demanded to biometrics systems. Apart from this, we would like to stress several additional characteristics of the proposed system. On the one hand, the use of compressed signals saves memory space, which could be critical in constrained devices like an implantable medical device such as a pacemaker or a holter monitor. Regarding the computational load, the penalty is very small since a matrix multiplication is only required to obtain the Hadamard coefficients. Furthermore no additional computations are required to extract the signal features---contrary to what occurs in systems based on fiducial features. On the other hand, since only a small fraction of the coefficients (the lower ones) are employed, even if the attacker would acquired these coefficients, she could not reconstruct the original signal. In conclusion, the proposed system is privacy preserving and works with a highly compressed version of the signal. As illustration of how the proposed system might be integrated in implantable medical devices is sketched in Fig. 7, showing how our proposal could contribute to the design of more secure medical applications and devices. For instance, a patient holding this sort of pacemaker could be remotely monitored once she is identified in a secure way using features extracted from her own heart signal.

As a future work, there are several research lines to continue with the ideas presented in this article. The proposal has been only tested with a database (MIT-BIH Normal Sinus Rhythm Database) of healthy individuals. Other databases, which include patients with a heart disease (e.g., MIT-BIH Arrhythmia Database or MIT-BIH Long-Term ST Database) or patients under stress conditions (e.g., MIT-BIH ST Change Database), could be employed to assess the use of compressed ECG signals. In line with this, in our proposal the Hadamard Transform is the core of our system for human identification. It would be interesting to perform a comparative study using a wide set of transforms (e.g., Fourier, Wavelet, Hadamard, etc.). Last but not least, the proposal could be extended to other vital signals like EEG or EMG.

Acknowledgments This work was supported by the MINECO grant TIN2013-46469-R (SPINY: Security and Privacy in the Internet of You) and the CAM grant S2013/ICE-3095 (CIBERDINE: Cybersecurity, Data, and Risks).

Conflict of interests The author declares that they have no conflict of interest.

References

1. Agrafioti, F., and Hatzinakos, D., ECG based recognition using second order statistics. In: 6th Annual Conference on Communication Networks and Services Research (CNSR), pp. 82–87, 2008.
2. Cempirek, M., and Stastny, J., The optimization of the EEG-based biometric classification. *Applied Electronics*, 25–28, 2007.
3. Identity Theft Resource Center. Data breach report. Technical report, December 2014.
4. Chan, A.D.C., Hamdy, M.M., Badre, A., and Badee, V., Wavelet distance measure for person identification using electrocardiograms. *IEEE Trans. Instrum. Meas.* 57(2):248–253, Feb 2008.
5. Luz, E.J.d. a.S., Menotti, D., and Robson Schwartz, W., Evaluating the use of {ECG} signal in low frequencies as a biometry. *Expert Systems with Applications* 41(5):2309–2315, 2014.
6. Frank, M., Biedert, R., Ma, E., Martinovic, I., and Touchalytics, D.Song., On the applicability of touchscreen input as a behavioral biometric for continuous authentication. *IEEE Transactions on Information Forensics and Security* 8(1):136–148, 2013.
7. Gahi, Y., Lamrani, M., Zoglat, A., Guennoun, M., Kapralos, B., and El-Khatib, K., Biometric identification system based on elec-trocardiogram data. In: Int. Conference on new technologies, mobility and security (NTMS), pp. 1–5, 2008.
8. Goldberger, A.L., Amaral, L.A.N., Glass, L., Hausdorff, J.M., Ivanov, P.Ch., Mark, R.G., Mietus, J.E., Moody, G.B., Peng, C.-K., and Stanley, H.E., PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation* 101(23):e215–e220, June 13. doi:10.1161/01.CIR.101.23.e215. Circulation Elec-tronic Pages: http://circ.ahajournals.org/cgi/content/full/101/23/e215PMID:1085218.
9. Inthavisas, K., and Lopresti, D., Secure speech biometric templates for user authentication. *IET Biometrics* 1(1):46–54, 2012.
10. Israel, S.A., Irvine, J.M., Cheng, A., Wiederhold, M.D., and Wiederhold, B.K., Ecg to identify individuals. *Pattern Recogn.* 38(1):133–142, 2005.
11. Jain, A.K., Ross, A., and Pankanti, S., Biometrics: a tool for information security. *IEEE Transactions on Information Forensics and Security* 1(2):125–143, 2006.
12. Khalifa, W., Salem, A., and Roushdy, M., A survey of eeg based user authentication schemes. In: 8th International Conference on Informatics and Systems, pp. 55–60, May 2012.
13. Kiranyaz, S., Ince, T., Pulkkinen, J., and Gabbouj, M., Personalized long-term ecg classification: A systematic approach. *Expert Systems with Applications* 38(4):3220–3226, 2011.
14. Kumari, P., and Vaish, A., Brainwave based user identification sys-tem: A pilot study in robotics environment. *Robot. Auton. Syst.* 65(0):15–23, 2015.
15. Mehrotra, H., Rattani, A., and Gupta, P., Fusion of iris and fingerprint biometric for recognition. In: Proceedings of the International Conference on Signal and Image Processing, pp. 1–6, 2006.
16. Miller, B., Vital signs of identity [biometrics]. *IEEE Spectrum* 31(2):22–30, Feb 1994.
17. Nait-Ali, A., Beyond classical biometrics: When using hidden biometrics to identify individuals, pp. 241–246, 2011.
18. Odinaka, I., Po-Hsiang, L., Kaplan, A.D., O'Sullivan, J.A., Sirevaag, E.J., and recognition, J.W.Rohrbaugh.E. cg.b. iometric., A comparative analysis. *IEEE Transactions on Information Forensics and Security* 7(6):1812–1824, Dec 2012.
19. Pal, S., and Mitra, M., Increasing the accuracy of {ECG} based biometric analysis by data modelling. *Measurement* 45(7):1927–1932, 2012.
20. Palaniappan, R., Multiple mental thought parametric classification: A new approach for individual identification. *Journal of Information and Communication Engineering* 2(4), 2006.
21. Rasmussen, K.B., Roeschlin, M., Martinovic, I., and Tsudik, G.: Authentication using pulse-response biometrics. In The Network and Distributed System Security Symposium (NDSS), 2014.
22. Riera, A., Dunne, S., Cester, I., and Ruffini, G.: STARFAST: a wireless wearable eeg/ecg biometric system based on the ENOBIO sensor. International Workshop on Wearable Micro and Nanosystems for Personalised Health, 2008.
23. Saechia, S., Koseeyaporn, J., and Wardkein, P., Human identification system based ECG signal. *In IEEE TENCON*, 1–4, 2005.
24. Schneier, B.: Changing passwords. https://www.schneier.com/blog/archives/2010/11/changing_passwo.html, November 2010.
25. Shin, L., How biometrics could improve health security. *Fortune*, 2015.
26. Silva, H., Gamboa, H., and Fred, A. *One lead ecg based personal identification with feature subspace ensembles*, pp. 770–783. Springer: Berlin Heidelberg, 2007.
27. Sim, H.M., Asmuni, H., Hassan, R., and biometrics, R.M.Othman.M. ultimodal., Weighted score level fusion based on non-ideal iris and face images. *Expert Systems with Applications* 41(11):5390–5404, 2014.
28. Singh, Y.N., Singh, S.K., and Ray, A.K., Bioelectrical signals as emerging biometrics Issues and challenges. In: ISRN Signal Processing, Vol. 2012, pp. 1–13, 2012.
29. Sun, S., Multitask learning for eeg-based biometrics. In: 19th International Conference on Pattern Recognition (ICPR), pp. 1–4, 2008.
30. Suresh, M., Krishnamohan, P.G., and Holi, M.S., GMM modeling of person information from EMG signals. *IEEE Recent Advances in Intelligent Computational Systems (RAICS)*, 712–717, 2011.
31. Tantawi, M.M., Revett, K., Tolba, M.F., and Salem, A., On the use of the electrocardiogram for biometrie authentication. In: 8th International Conference on Informatics and Systems, pp. 48–54, May 2012.
32. Tarricone, R., and Tsouros, A.D.: Home Care in Europe: The Solid Facts. WHO Regional Office Europe, 2008.
33. Tresadern, P., Cootes, T.F., Poh, N., Matejka, P., Hadid, A., Levy, C., McCool, C., and Marcel, S., Mobile biometrics: Combined face and voice verification for a mobile platform. *IEEE Pervasive Computing* 12(1):79–87, 2013.
34. Wang, Y., Agrafioti, F., Hatzinakos, D., and Plataniotis, K.N., Analysis of human electrocardiogram for biometric recognition. *EURASIP Journal on Advances in Signal Processing*, 2008, January 2008.
35. Yang, J., Shi, Y., and Yang, J., Personal identification based on finger-vein features. *Comput. Hum. Behav.* 27(5):1565–1570, 2011.