# Comparative Analysis and Fusion of Spatiotemporal Information for Footstep Recognition

Ruben Vera-Rodriguez, John S.D. Mason, Julian Fierrez, *Member*, *IEEE*, and Javier Ortega-Garcia, *Senior Member*, *IEEE*

**Abstract**—Footstep recognition is a relatively new biometric which aims to discriminate people using walking characteristics extracted from floor-based sensors. This paper reports for the first time a comparative assessment of the spatiotemporal information contained in the footstep signals for person recognition. Experiments are carried out on the largest footstep database collected to date, with almost 20,000 valid footstep signals and more than 120 people. Results show very similar performance for both spatial and temporal approaches (5 to 15 percent EER depending on the experimental setup), and a significant improvement is achieved for their fusion (2.5 to 10 percent EER). The assessment protocol is focused on the influence of the quantity of data used in the reference models, which serves to simulate conditions of different potential applications such as smart homes or security access scenarios.

**Index Terms**—Biometrics, footstep recognition, gait recognition, pressure analysis, pattern recognition

## 1 Introduction

The growth in biometrics has been very significant in the last few years, not only for the most popular modes such as fingerprint, speech, or face, but as well for the lesser known biometrics such as otoacoustic emissions, palm or footsteps. This paper is focused on the assessment of footstep signals as a relatively new biometric with a comparative analysis and fusion of spatiotemporal information of the signals. In this work, footstep signals are captured from people walking over an instrumented sensing area, in contrast to some works using the sound of the footsteps [1]. It is worth noting that some works in this area [2], [3], [4] refer to footstep recognition as part of gait recognition [5], but using a floor-based approach.

One significant benefit of footsteps over other, better known modes is that footstep signals can be collected unobtrusively with minimal or no person cooperation; this can be very convenient for the user. Other benefits lie in the robustness to environmental conditions, with minimal external noise sources to corrupt the signals. Also, footstep signals do not reveal an identity to other humans, like the face or the voice, making footsteps a less sensitive mode. Footsteps might prove to be an ideal complementary biometric, considering the scenario of a person walking to other arrangement systems such as a passport control, door entry system or a fingerprint scanner. For example, the case of the fusion of footsteps with face and iris systems at a distance in a security gate scenario is very appealing [6].

As described in [7], footstep recognition was first suggested as a biometric in 1977 [8], but it was not until 1997 when the first experiments were reported [9]. Since then the subject has received relatively little attention in the literature compared to other biometrics. A review is presented in Section 2 covering sensors, features, and approaches to classification. The associated results are promising and give an idea of the potential of footsteps as a biometric [10], [11]; however, these results are related to relatively small databases in terms of number of people and footsteps, and this is a limitation of the work to date.

A database is an essential tool to assess any biometric; therefore, this paper reports experimental results of footsteps as a biometric on the largest footstep database to date, with more than 120 people and almost 20,000 signals, enabling assessment with statistical significance.

The main contribution of the present work is the assessment of footsteps in time, in space, and in a combination of the two. Features extracted in the time domain include the ground reaction force (GRF), the spatial average, and the upper and lower contours of the pressure signals, while in the spatial domain, 3D images of the accumulated pressure are obtained. Interestingly, the performance for the two domains proves to be very similar, with equal error rates (EERs) in the range of 5-15 percent for each domain, depending on the experimental setup, and in the range of 2.5-10 percent for their fusion. Results achieved are considerably better compared to other existing methods (e.g., [3], [10]). This paper also considers some important factors for footstep recognition not included in previous related works, such as the influence of the quantity of data used in the training stage of the system, which serves to

- *R. Vera-Rodriguez, J. Fierrez and J. Ortega-Garcia are with the Biometric Recognition Group—ATVS, Universidad Autonoma de Madrid, C\Francisco Tomas y Valiente 11, EPS, Madrid 28049, Spain. E-mail: {ruben.vera, julian.fierrez, javier.ortega}@uam.es.*
- *J.S.D. Mason is with the Speech and Image Research Group, School of Engineering, Swansea University, Singleton Park, Swansea SA2 8PP, United Kingdom. E-mail: j.s.d.mason@swansea.ac.uk.*

*Manuscript received 9 Nov. 2011; revised 15 May 2012; accepted 17 July 2012; published online 1 Aug. 2012.*
*Recommended for acceptance by S. Sarkar.*
*For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number TPAMI-2011-11-0810.*
*Digital Object Identifier no. 10.1109/TPAMI.2012.164.*

simulate various potential applications, and the effect of the sensor density on the performance.

The paper is organized as follows: Section 2 describes related work in order to put ours in context. Section 3 introduces the collection process of the database and the signals obtained. Feature extraction is covered in Section 4, paying special attention to the spatial and temporal components of the signals. Section 5 describes the experimental protocol followed. Experimental results are presented and analyzed in Section 6, and finally, conclusions and future work are drawn in Section 7.

## 2 Related Work

A general classification of biometrics is often made into physiological and behavioral modes (see, for example, [23]). Clearly this is not an orthogonal classification since most of the behavioral modes are affected by physiological characteristics. It is interesting to note that in the case of the behavioral modes the main biometric information is carried along the time axis of the signals.

Footstep signals, and others such as gait or talking face, can be seen as dual biometrics due to the fact that information can be extracted from both the physiological and behavioral components to carry out person recognition. In this paper, we present a new approach to the study of footstep signals, viewing them in the orthogonal dimensions of time and space (named BTime and BSpace, respectively, throughout the remainder of this paper).

This duality of footsteps is reinforced in the literature when reviewing the two main approaches, namely: 1) switch sensors [2], [4], [14], [19], and 2) pressure sensors [9], [12], [13], [17], [20], [24]. Switch sensors have been used with a range of densities from 50 to 1,000 sensors per $\mathrm{m}^2$; this is much higher than the density of approaches using pressure sensors. Approaches based on switch sensors have focused more on the study of the spatial distribution of footstep signals (BSpace), while approaches based on pressure sensors have focused more on the study of the footstep pressure dynamics along the time course (BTime).

As an exception, Jung et al. [15] used a commercial pressure mat with a high sensor density, but only used the spatial domain information. Very recently, Qian et al. [3] also used a commercial pressure mat, extracting the center of pressure (COP) information and adding the pressure information, therefore using time and spatial pressure information only for some selected key points (geometric approach). Section 6.3.1 compares results with our proposed method.

Table 1 presents a comparison of the published work on footsteps as a biometric with reported systems in the rows in a chronological order and parameters in the columns.

The second column of the table shows database sizes. As can be seen, a common characteristic of most of the referenced works is the relatively small size of the databases in relation to other biometric evaluations where people are normally counted in hundreds or thousands and the number of tests perhaps in many thousands. Apart from our initial investigations (41 people and 3,174 footsteps [20]), a maximum number of 16 people [13] and 5,690 footstep examples [3] were gathered. The experiments reported here are carried out over the largest footstep to date, with 9,990 stride footstep signals from 127 persons.

In each case, except for [2], [17], the databases are divided into training and testing sets, but none use independent development and evaluation sets, with the exception of our works [11], [20], [21], [22], a limitation which makes application performance predictions much more difficult. In the work that follows, emphasis is placed on training/test and validation/evaluation sets.

Identification, rather than verification, was the task considered in the majority of the above cases. Identification has the benefit of utilizing the available data to a maximum, but suffers from well-known scalability problems in terms of the number of classes in the set.

It is interesting to point out that some systems present classification results for stride data (consecutive footsteps), e.g., [2], [3], [10], while others are for a single footstep, e.g., [12], [20]. In [10], an identification accuracy of 63 percent using a single footstep as a test was improved to 92 percent when six consecutive footsteps were used, showing the benefits of using stride data compared to single footstep signals.

Regarding the classification, different methods have been used, as can be seen in the table. In [10], Suutala and Roning presented a comparison of performance for various classification methods such as KNN, LVQ, RBF, MLP, and SVM, obtaining the best results for the cases of MLP and SVM, which has been reinforced by Yun [4].

## 3 Data Acquisition

This section describes, on the one hand, the sensor arrangement and the corresponding footstep signals obtained, and on the other hand, the resultant database (SFootBD) that has been collected.

### 3.1 Sensor Arrangement and Footstep Signals

The sensor approach developed here combines the characteristics of a high sensor density and the sampled pressure obtained from piezoelectric sensors. The first characteristic enables the extraction of spatial information regarding shape and position of the foot (BSpace), and the second provides the information of the pressure along the time course (BTime), thus resulting in footstep signals which contain more information than that available from approaches published previously, such as [2], [12], [19]. Exceptions are [3], [15], which used commercial pressure mats with high sensor density (10,000 sensors per $\mathrm{m}^2$), but very low time frequency sampling (30-40 Hz).

The system developed is comprised of two sensor mats positioned to capture one stride for each signature, i.e., signals from two consecutive footsteps (right foot then left foot). Each mat measures $45 \times 30$ cm and contains 88 piezoelectric sensors, giving a sensor resolution of approximately 650 sensors per $\mathrm{m}^2$; the sampling frequency associated with each sensor is 1.6 kHz, therefore having a capture system with high time and high spatial resolution for the first time. Fig. 1 shows a diagram of the spatial distribution of the piezoelectric sensor mats employed to capture the footstep signals.

TABLE 1
Comparison of Approaches to Footstep Recognition

| Method | Database | Technology | Features | Classifier | Results |
|---|---|---|---|---|---|
| **Addlesee *et al.*, (UK) 1997 (ORL Active Floor)** [9] | 300 footsteps, 15 persons | Load cells | GRF | HMM | ID rate: 91% |
| **Orr and Abowd, (USA) 2000 (Smart Floor)** [12] | 1680 footsteps, 15 persons | Load cells | Geometric from GRF | NN | ID rate: 93% |
| **Cattin, (Switzerland) 2002** [13] | 480 footsteps, 16 persons | Piezoelectric sensors | PSD from derivative GRF | NN | Verif EER: 9.5% |
| **Yun *et al.*, (Korea) 2003 (Ubifloor)** [14] | 500 walking seq., 10 persons | Switch sensors | Position of stride footsteps | MLP neural net. | ID rate: 92% |
| **Jung *et al.*, (Korea and Japan) 2004** [15] | 440 footsteps, 11 persons | Pressure mat | 2D COP trajectories | HMM-NN | ID rate: 80% |
| **Suutala *et al.*, (Finland) 2004/05 (EMFi Floor)** [16] | 440 footsteps, 11 persons | Electro mechanical film | Geometric from GRF, and FFT | MLP neural net. and LVQ | Best ID rate of 92% using 3 footsteps as test |
| **Middleton *et al.*, (UK) 2005** [2] | 180 walking seq., 15 persons | Switch sensors | Stride length, cadence and heel-to-toe ratio | NN | ID rate: 80% |
| **Gao *et al.*, (UK) 2006** [17] | 400 footsteps, 11 persons | Load cells | Geometric from GRF | NN | ID rate: 94% |
| **Stevenson *et al.*, (USA) 2007** [18] | 85 footsteps, 8 persons | Piezoelectric sensors | Derivative GRF | HMM | Verif EER: 20% |
| **Suutala *et al.*, (Finland and Japan) 2008** [19] | 180 walking seq., 9 persons | Switch sensors | length, width, stride length and duration | GP | ID rate: 64% for 1 footstep, 84% for sequence |
| **Suutala and Roning, (Finland) 2008 (EMFi Floor)** [10] | 150 walking seq., 10 persons | Electro mechanical film | Geometric from GRF and FFT | SVM | ID rate: 63% for 1 footstep, 92% for 6 footsteps |
| **Vera-Rodriguez *et al.*, (UK) 2007/09** [20] | 3174 footsteps, 41 persons | Piezoelectric sensors | Geometric and holistic from derivative GRF | SVM | Verif EER: 9.5% for Devel, 13.5% for Eval |
| **Qian *et al.*, (USA) 2010** [3] | 5690 stride footsteps, 11 persons | FSR mat | Geometric from pressure and 2D COP trajectories, stride length of stride | FLD (LDA) | ID rate: 92% |
| **Vera-Rodriguez *et al.*, (UK and Spain) 2010** [21] | 9990 stride footsteps, 127 persons | Piezoelectric sensor mat | Holistic pressure-time info | SVM | Verif EER: 5-15% depending on exp. setup |
| **Yun, (Korea) 2011 (UbifloorII)** [4] | 500 walking seq., 10 persons | Photo interrupter sensors | Foot centers, heel-to-toe time, geometric from footstep sequence | MLP | ID rate: 96% |
| **Vera-Rodriguez *et al.*, (UK and Spain) 2011** [22] | 9990 stride footsteps, 127 persons | Piezoelectric sensor mat | Holistic pressure-space info | SVM | Verif EER: 5-15% depending on exp. setup |
| **Proposed method** | 9990 stride footsteps, 127 persons | Piezoelectric sensor mat | Fusion of time and spatial holistic pressure info | SVM | Verif EER: 2.5-10% depending on exp. setup |

Footstep signals collected here contain information in four dimensions, namely, pressure magnitude, time, and spatial positions $X$ and $Y$. Fig. 2 shows three different 3D plots for an example of a footstep signal, reflecting its three stages: Fig. 2a shows the differential pressure for an instant in the first stage of the footstep, i.e., when the heel strikes the sensor mat, Fig. 2b shows the same but for an instant in the second stage of the footstep, i.e., when the whole foot rests over the sensors, and Fig. 2c the same but for an instant in the third stage of the footstep, i.e., when the heel leaves the surface and the toes push off the sensor mat. It is worth noting that the output of the piezoelectric sensors is the differential pressure in time; thus, it can be seen in Fig. 2c that there are negative values.

In this paper, footstep signals are studied in the time and spatial domains separately, which enables comparison with previous related work and gives an indication of the discriminative power of the signals in each domain.

### 3.2 Database Collection and Labelling

The main objective regarding database size was to collect a database as large as possible; therefore, an automatic capture system was developed in order to collect the biometric data without the need for human intervention. The first session of each person was a supervised enrolment, where a supervisor explained how to provide the footstep data. In this sense, people were asked to walk at a natural speed a few meters before the sensor mats in order to produce more realistic signals. People were encouraged to return as often as they could to provide further sample signals. These following sessions, and therefore the majority of the database, were collected on an unsupervised mode.

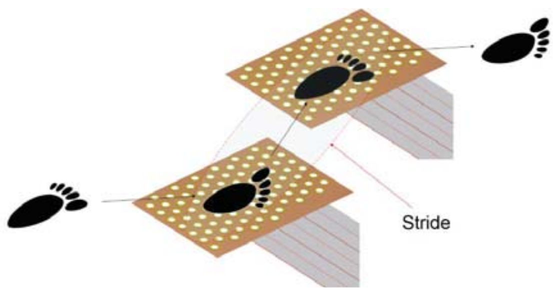



Fig. 1. Spatial distribution of the piezoeletric sensors.

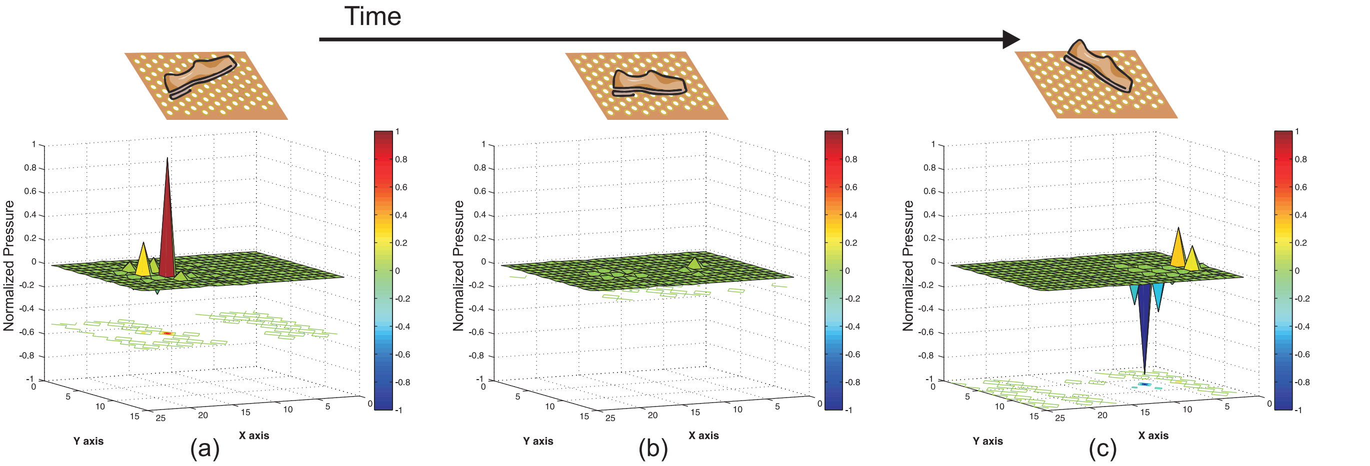



Fig. 2. Spatiotemporal footstep signal in different stages. (a) The time derivative of the pressure against the position $X$ and $Y$ at the first stage of footstep; (b) and (c) the same but for second and third stages of the footstep signal.

The enrollment of people in the system was continuous during the collection period. Also, different people provided data during different periods of time and in different numbers of sessions because, as stated before, the objective was to obtain a large database. More details about the database can be found in [22].

A characteristic of the footstep signals is that they do not reveal human identity directly or, in other words, they cannot be used by humans to carry out recognition. Thus, apart from the footstep signals, other support biometric modes were captured simultaneously with same timestamps. These support modes are speech, face, and gait. They were used to assist with both manual and automatic database labeling, and in the cross-checking of apparent label anomalies (suspected errors that might arise during experiments).

Another benefit of having the extra modes is the opportunity to assess them as biometrics in a multimodal manner. Fig. 3 illustrates the setup of the capture system with four biometric modalities which are all acquired with the same timestamps.

The speech mode was deemed to be very appealing for the automatic labeling of the database using speaker recognition [25]. This procedure has resulted in the largest footstep database to date, with over 120 people and almost 20,000 valid footstep signals (i.e., 10,000 stride signals), leading to a more reliable assessment of footsteps as a biometric. The database was collected in various sessions per user during a period of 16 months.

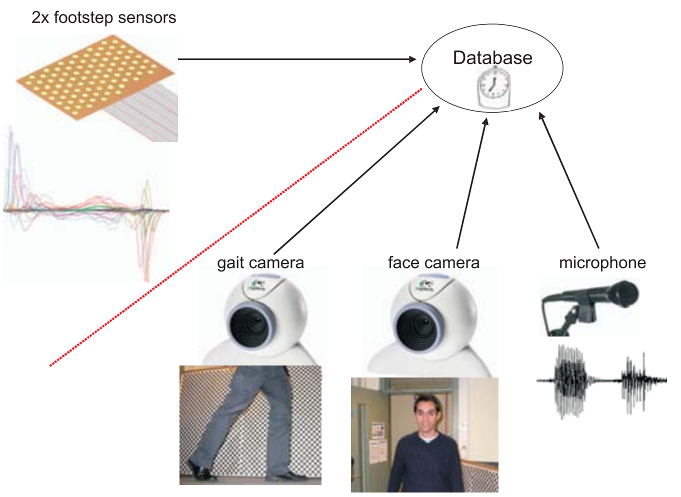



Fig. 3. Four sources of data captured, clockwise from top left: footstep signals, speech identifier, face, and gait videos. Footstep is the primary mode, the other three are support modes. The four modes are connected with a common timestamp.

A characteristic of the labeled database considered here is that it contains a large amount of data for a small subset of subjects ($>200$ signals per subject, for 15 subjects), which could serve to simulate a smart home scenario; and a smaller quantity of data for a larger group of subjects ($>20$ signals per subject, for 54 subjects), which could serve to simulate a security access scenario. This reflects that the mode of capture which was voluntary and unsupervised. Fig. 4 shows the number of stride footstep signals per person in the database. The footstep database (SFootBD) is available to the research community in [26].

It is interesting to note that people were allowed to walk over the footstep sensors with different types of footwear (shoes, trainers, boots, barefoot, etc.) and carrying weights such as office bags, which makes signals collected more realistic. This is in contrast to most databases in the field which only contain people walking barefoot.

## 4 Feature Extraction

This section describes the feature extraction process with a special emphasis on the time and spatial information

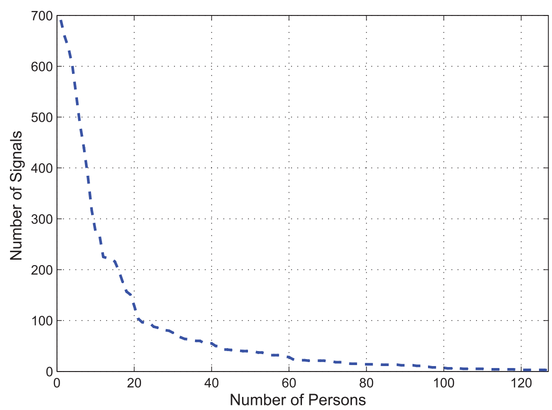
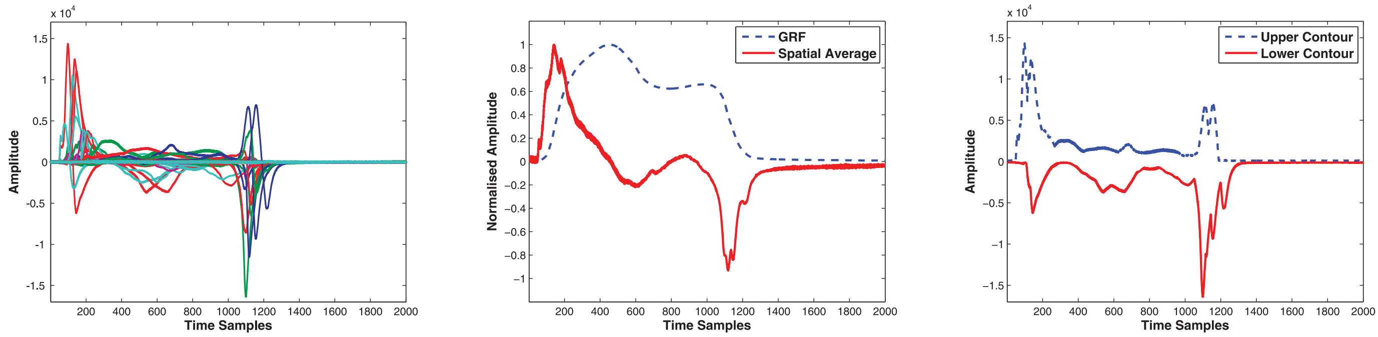



Fig. 4. Number of footstep signals against number of people in the footstep database.




(a) Footstep signal (b) GRF/Spatial Average (c) Upper/Lower contour

Fig. 5. Time domain (BTime) feature extraction. (a) Differential pressure directly from the 88 sensors against time. (b) Normalized ground reaction force profile from (a) as defined in (2), and normalized spatial average of the 88 sensors as defined in (3). (c) Upper and lower contour profiles from (a) as defined in (4) and (5), respectively.

contained in the footstep signals. These two types of features are extracted independently in order to be compared, and are fused in a later stage described in Section 6.2.

## 4.1 Time Domain Features (BTime)

The most popular time domain feature in related works is the ground reaction force [3], [9], [10], [12], [16], [17]. In most of the cases, some key points are extracted from the GRF profile using the time and the pressure value coordinates as features, together with some distance measures between some of them (geometric approach) [7].

In the case considered here, the time domain information of the footstep signals is extracted from the differential pressure of the sensors along the time axis without considering their spatial distribution, similar to the work in [21]. Fig. 5a shows an ensemble of signals from an example single footstep. Each profile represents the differential pressure against time for each of the 88 sensors across one footstep. In the preprocessing stage, an energy detector across the 88 sensors of the signals is used to obtain the beginning of each footstep in order to align the signals to a common time position.

Fig. 5b shows the global GRF profile for the example footstep considered here. In this case, as the piezoelectric sensors provide the differential pressure, the global GRF is obtained by accumulating each sensor signal across time (individual GRF, $\mathrm{GRF}_i$); then an average of the 88 single profiles is computed to provide a global GRF ($\mathrm{GRF}_T$).

Formally, $s_i[t]$ is the output of the piezoelectric sensor $i$, $i = 1, \dots, 88$, and $t = 1, \dots, T_{max}$ are the time samples. $T_{max}$ was set to a value of 2,000 time samples, large enough for all footstep signals considered. Then, the individual GRF ($\mathrm{GRF}_i$) and the global GRF ($\mathrm{GRF}_T$) are defined by

$$GRF_i[t] = \sum_{\tau=0}^{t}(s_i[\tau]), \tag{1}$$

$$GRF_T[t] = \frac{1}{88}\sum_{i=1}^{88}(GRF_i[t]). \tag{2}$$

Apart from the $\mathrm{GRF}_T$, two other feature approaches are studied here. The first comes from a spatial average of the 88 sensors of the mat to produce a single profile. Similar features were also extracted in [13] and [18]. An example is shown in Fig. 5b.

$$s_{ave}[t] = \frac{1}{88}\sum_{i=1}^{88}(s_i[t]). \tag{3}$$

The second is a novel approach which uses the upper and lower contour coming from the maxima and minima of the sensors for each time sample independently of the spatial distribution of the sensors. An example is shown in Fig. 5c. This also gives valuable discriminative information, as can be seen in the results shown in Section 6.1.

$$s_{up}[t] = \max_{i=1}^{88}(s_i[t]), \tag{4}$$

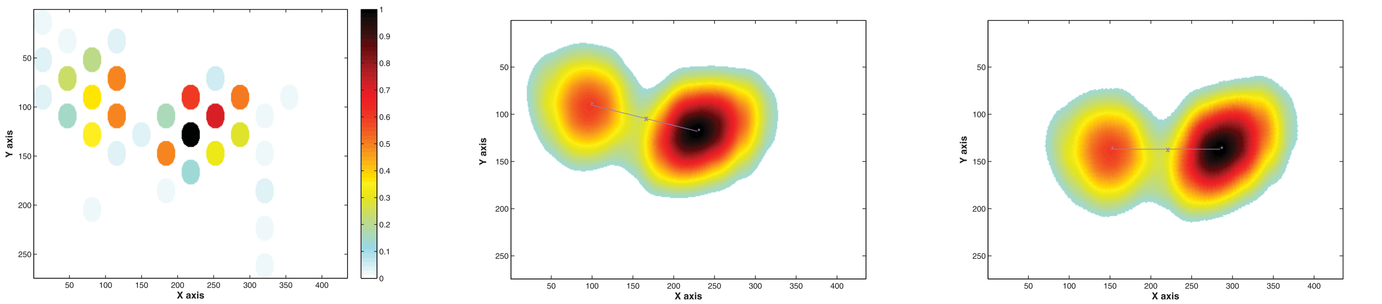
$$s_{lo}[t] = \min_{i=1}^{88}(s_i[t]). \tag{5}$$

These two signals are then concatenated into one contour signal $s_{con}[t] = [s_{up}[t], s_{lo}[t]]$. Equations (2) to (5) lead to a high dimensionality in the time domain with 2,000 samples per footstep signal (1,25 seconds). Data dimensionality is further reduced using principal component analysis (PCA). Empirically, good development results are obtained when retaining approximately 96 percent of the original information by using the first 120 principal components for each feature approach.

## 4.2 Spatial Domain Features (BSpace)

Related works have extracted spatial domain information from capture systems working with switch sensors and therefore no pressure information is included. The most common features are the length and width of the footstep signals and the length and the relative positions of the stride footsteps [2], [14], [19].

The spatial domain information extracted here is a novel approach which considers the distribution of the accumulated pressure along a footstep signal in the spatial domain. In this case, the individual GRF ($\mathrm{GRF}_i$) of each footstep sensor is integrated along the time axis, obtaining a single value of the accumulated pressure ($\mathrm{AP}_i$) for each sensor of the array for a footstep signal, similar to the work in [22]. The accumulated pressure ($\mathrm{AP}_i$) is the measure used to study the distribution of the pressure across the spatial domain of the signals, and it is defined by




(a) Accumulated Pressure (AP) (b) Same as (a) after smoothing (c) Same as (b) after alignment and rotation

Fig. 6. Spatial domain (BSpace) feature extraction. (a) Accumulated pressure (AP) of each sensor across one footstep. (b) Result after smoothing the image from (a) with a Gaussian filter. (c) Result after alignment and rotation to a common center of signals from (b).

$$AP_i = \sum_{t=0}^{T_{max}} (GRF_i[t]). \tag{6}$$

Fig. 6a represents the 88 values of $\mathrm{AP}_i$ in the $X$ and $Y$ spatial axes for an example footstep signal. In this case, we have used an image resolution of one pixel per $\mathrm{mm}^2$, giving the values of $\mathrm{AP}_i$ to the positions with sensors and zero values to the rest of the image, in this way keeping the original geometry of the sensors.

It is worth noting that (2) and (6) use the 88 profiles of $\mathrm{GRF}_i$ to extract time and spatial information, respectively. In the first case, the 88 profiles of $\mathrm{GRF}_i$ are averaged to produce a global GRF ($\mathrm{GRF}_T$), i.e., pressure information in the time domain (see Fig. 5b). In the second case, the total sum of the time values of each $\mathrm{GRF}_i$ profile give 88 values of the accumulated pressure ($\mathrm{AP}_i$), i.e., pressure information in the space domain (see Fig. 6a).

The sensor-derived images are then processed to give a form suitable for subsequent pattern matching, i.e., rotation and alignment to a common position is needed for all the spatial images extracted from the database. The spatial sensor resolution of the array only allows a $\pm 60$ degrees rotation for perfect matching, which is too large; therefore, the images were smoothed using a Gaussian filter (defined in (7)) in order to obtain a continuous image. Bicubic spline interpolation was also tried, but better results were obtained using the Gaussian filter. Fig. 6b shows the resultant image for the given example after the Gaussian filter. The best results were obtained using values of $\mathrm{x}, \mathrm{y} = 1, \ldots, 100$ and $\sigma = 14$ for the filter:

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}. \tag{7}$$

These images are then aligned and rotated based on the points with maximum pressure, corresponding with the toe and the heel areas, respectively. The resultant image is shown in Fig. 6c, which is used to carry out the biometric classification. For the case of experiments that make use of the stride footstep, the relative angle between the left and the right footstep is also considered as a feature.

The above process results in an image with high dimension (necessary for accurate rotation and alignment). The next step is then to reduce this dimension prior to the classification. The resulting images have a dimension of $280 \times 420 = 117{,}600$ pixels for the single footstep, and this is reduced using principal component analysis to 140 principal components (keeping 96 percent of the original information).

Regarding the classifier, in both cases of time and spatial domain features a support vector machine (SVM) [27] was adopted with a radial basis function (RBF) as the kernel, due to very good performance in previous studies in this area [10], [11].

## 5 Experimental Protocol

This section describes the experimental protocol followed to assess footsteps as a biometric. Special attention has been paid to the partitioning of the data into three sets, namely, training, validation, and evaluation sets, the first being used for model training and the second two for testing.

As an assessment protocol of the footstep recognition evaluation, index files were created to provide a list of the footstep signals to use in each one of the training and test datasets following the structure utilized by the international NIST SRE [28].

The training set is comprised of a set of in-class data used to train one model per client, and a set of out-class data from a cohort of impostors which is used in the training process to obtain better statistical models. PCA transformation is only carried out with the data from the training set, and the coefficients of the PCA transformation are then applied to the data of the validation and evaluation sets to reduce their dimensionality too. Also, SVM is used in the training stage to train a model per client.

Validation and evaluation sets are two test sets, the main difference being that the evaluation set is a balanced set comprised of the last five footstep signals provided by persons P1 to P110, while the validation set is an unbalanced set which contains a larger number of test signals for subjects included in the training data. The validation set is used to tune the system, i.e., type of features, number of PCA components, SVM parameters, etc., in order to obtain the best results. The evaluation set is comprised of unseen data, not used in the development of the system.

It is to be stressed that these sets reflect the chronological order of the data capture. Therefore, the training data is comprised of the first data provided by each user, and the data used in the evaluation set is the last collected. This is a realistic approach reflecting actual usage, in contrast to previous related works, e.g., [2], [3], [11], [14], which

TABLE 2
Database Configuration for the Case of
40 Models and 40 Reference Signals per Model

| | Training Set | Validation Set | Evaluation Set |
|---|---|---|---|
| Clients | P1 – P40 | P1 – P40 | P1 – P40 |
| Signals per client | 40 | 170 (8-650) | 5 |
| Total signals clients | 1,600 | 6,697 | 200 |
| Cohort impostors | P41 - P127 | P41 - P78 | P41 - P110 |
| Total signals impostors | 763 | 380 | 350 |
| Total signals per set | 2,363 | 7,077 | 550 |
| Total | 9,990 | | |

randomly divide the data into training and test sets, or use a leave-one-out approach.

Results described in Sections 6.1 and 6.2 relate to a benchmark division of the database which is set to use 40 footstep signals per client to train the models, having available a group of 40 clients (and therefore 40 models). Table 2 shows the numbers of data in each set in more detail, and Fig. 7 shows a diagram of the partitioning of the database. Each signal from the test sets is matched against all the trained models (40 models in this case). As can be seen in the table, the total number of stride signals in the database is 9,990, i.e., 19,980 single (right and left) signals in total.

Sections 6.3 and 6.4 report experimental results in the form of error rates against quantity of data used to train each model. In this case, the database is divided into 11 configurations of different numbers of clients and data per client. The influence of the quantity of data used to train and test the system is a key factor in any performance assessment; while common in more established biometric modes, this aspect is not considered in many cases of footstep studies, for example, in [3], [9], [12], and [13], due to limited amounts of data per person in the databases. Different applications can be simulated using different quantities of data in the client models. In the present work, we consider applications such as smart homes and access control scenarios. In the case of a smart home there would potentially be a very large quantity of training data available for a small number of clients, while in security access scenarios such as a border control, limited training data would be available, but potentially for a very large group of clients.

## 6 Experimental Results

This section describes the experimental results obtained for the assessment of footsteps as a biometric. Results relate to the comparative analysis of both BTime and BSpace feature approaches, their fusion, and some other important aspects such as the influence of the quantity of data used to train the models in the performance, the difference between using single signals (right, left) or stride signals (connected footsteps), the influence of the sensor density, and a final evaluation of the recognition system.

Results are presented with DET curves [29], using the equal error rate as the performance measure for verification applications. Also, CMC curves are used in Section 6.2 to show the performance for the case of an identification scenario.

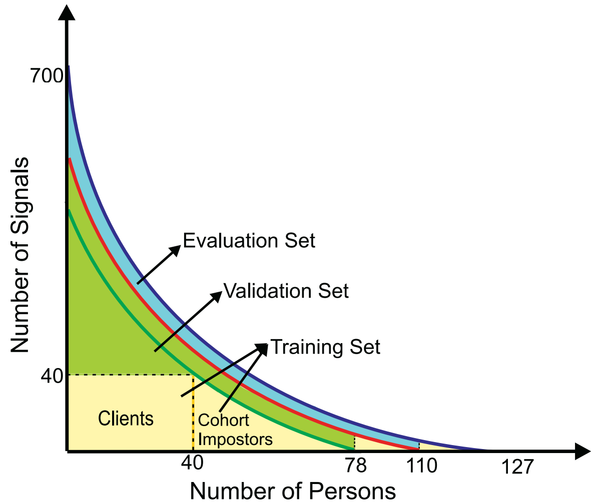



Fig. 7. Number of footstep signals against number of people in the database. Diagram of the database with the different divisions of training, validation, and evaluation sets for the benchmark division. Numbers are described in Table 2.

### 6.1 Comparative Analysis of BTime and BSpace Approaches

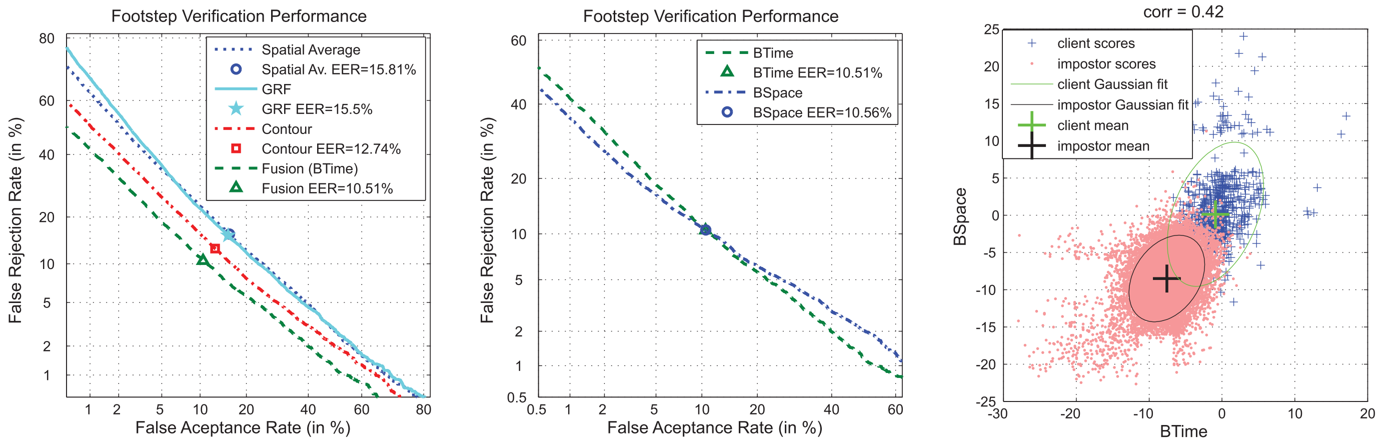
This section presents the comparative analysis of the two BTime and BSpace feature approaches. Fig. 8a shows the DET curve profiles for the BTime approach, which is described in Section 4.1. This is shown for the three features considered, i.e., the global GRF, the spatial average, and the contour. Also, a fourth plot in the figure shows the result of the fusion at the feature level of the three time features, which is carried out by concatenating the features of the single approaches after PCA. These results are generated for stride footsteps, which are comprised of concatenated right and left footstep signals. The benchmark division of the database shown in Table 2 is considered here.

As can be seen, very similar results are obtained for the GRF and spatial average features with EERs of 15.5 and 15.8 percent, respectively. The contour feature approach provides a better result with an EER of 12.7 percent. In any case, the fusion outperforms the three single approaches, obtaining an EER of 10.5 percent.

In a similar way, Fig. 8b shows the DET curve result obtained for the case of the spatial domain approach (BSpace), described in Section 4.2. An EER of 10.6 percent is achieved in this case, which is surprisingly very similar to the result obtained for the fusion of the three time domain features, which can also be seen in Fig. 8b.

This is a directly comparative assessment of the time and spatial information contained in the footstep signals, as the same protocol has been used to carry out the experiments. This implies that the time and spatial information extracted from the footstep signals have similar discriminatory properties.

A further analysis has been carried out to study how statistically relevant are BTime and BSpace feature approaches. First, the Pearson correlation coefficient of the matching scores of both approaches was calculated, obtaining 0.42 (where 1 would mean they are completely correlated and 0 completely uncorrelated). Second, a *t-test* has been carried out at the score level to study the statistical difference between the two approaches. As a result, we obtain $p < 0.001$, which means that scores from both




(a) BTime Approaches (b) Comparison BTime-BSpace (c) Scatter Plots BTime-BSpace

Fig. 8. (a) DET curves for BTime approaches and their fusion. (b) Comparison of performance of BTime and BSpace approaches. (c) Scatter plots of scores for BTime and BSpace.

approaches are statistically different with a 95 percent significance level. Also, the scatter plot for the scores of BTime and BSpace is shown in Fig. 8c. Therefore, we can conclude that the two classes of features are statistically relevant, and fusing the two systems can lead to a performance improvement, as reported in [30].

## 6.2 Fusion of BTime and BSpace

This section is focused on the fusion of the time and spatial information of the footstep signals. In the case considered here, the fusion of BTime and BSpace approaches has been carried out at the feature and score levels. For simplicity, the same weights were applied to the two cases.

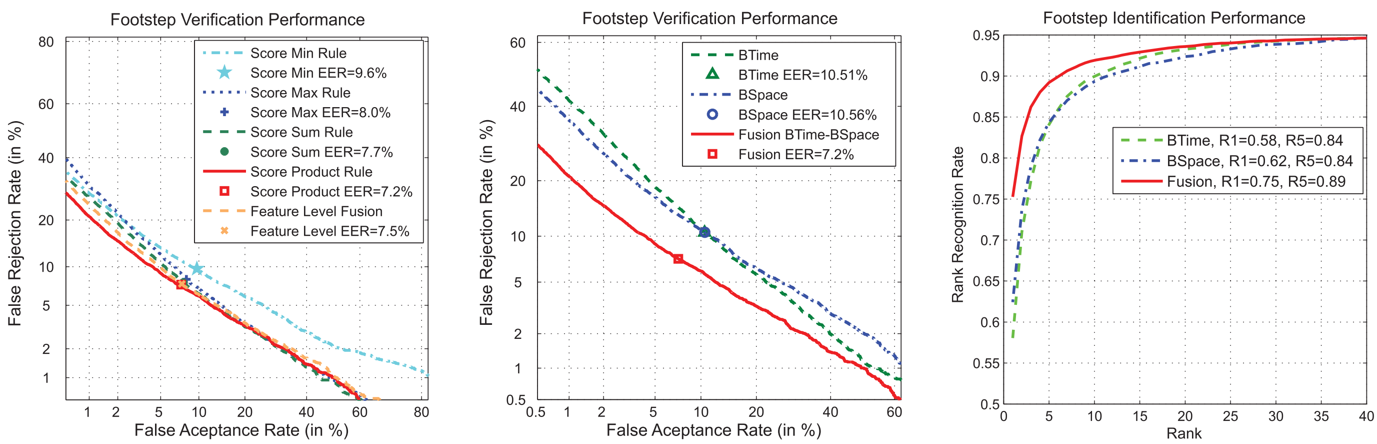
In the feature-level fusion, the feature sets originating from the different algorithms described in Sections 4.1 and 4.2 are fused into a single feature vector. PCA is used in this case to reduce the dimensionality of the feature vector and then it is introduced to an SVM classifier to provide the new matching scores.

In the score-level fusion, the scores originating from the BTime and BSpace approaches are combined to generate a new score. As stated in [31], fusion at the score level is the most common approach due to the ease in accessing and combining the scores generated by different classifiers. In this paper, different combination rules such as the sum, product, max, and min rules have been compared following previous works [32]. These fixed combination rules are simple, computationally fast, and provide good performance compared to other categories such as classifier-based, as stated in [31] and [33].

Fig. 9a shows the DET curves obtained for the cases of the fusion carried out at the feature-level and at the score-level using the sum, product, max, and min rules. As can be seen, the best performance for the score-level fusion corresponds to the case of the product combination rule with an EER of 7.2 percent. This also gives a slightly better performance that the fusion at the feature-level (7.5 percent EER); therefore results achieved in the following sections relate to the score-level fusion using a product combination rule.

Fig. 9b shows a comparison of performance for the case of the stride footsteps for BTime, BSpace, and their fusion at the score-level. Results improve for the fusion at an absolute average of 3.3 percent of EER, which is very significant.

Fig. 9c shows the corresponding CMC curves which report the expected performance in an identification application. As can be seen, results for BTime and BSpace are similar, obtaining rank 1 of 0.58 and 0.62, respectively, and rank 5 of 0.84 in both cases. When the fusion at the




(a) Fusion Approaches (b) DETs BTime, BSpace and Fusion (c) CMCs BTime, BSpace and Fusion

Fig. 9. (a) Comparison of performance for the four combination rules used in the score-level fusion and the feature-level fusion. (b) Comparison of performance of BTime, BSpace, and their fusion at the score-level. (c) Comparison of CMC curves for an identification application.

TABLE 3
Relation between Number of Signals in the Client Models and Number of Clients Available in Each Case

| Signals/model | 1 | 10 | 20 | 40 | 60 | 80 | 100 | 200 | 300 | 400 | 500 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Client models | 75 | 60 | 54 | 40 | 30 | 24 | 20 | 15 | 9 | 7 | 5 |

score level is applied, the results improve significantly, achieving a rank 1 of 0.75 and a rank 5 of 0.89.

## 6.3 Other Important Aspects

This section is focused on some other important aspects which are not common in previous studies in the field of footsteps as a biometric. The factors studied here are the influence of the quantity of data used to train the models in the performance, the difference between using single signals (right, left) or stride signals (connected footsteps), and the influence of the sensor density in the performance.

### *6.3.1 Quantity of Data*

As described previously, the influence of the quantity of data used to train and test the system is key in any performance assessment, but this aspect is not been considered in many cases of footsteps studies. A characteristic of the database considered here is that it contains a large amount of data for a small subset of subjects ($>200$ signals per subject, for 15 subjects), and a smaller quantity of data for a larger group of subjects ($>20$ signals per subject, for 54 subjects). Therefore, many experiments can be carried out using different divisions of the database in terms of quantity of data used to train the models varying also the number of clients. This could serve to simulate some potential applications for footsteps such as smart homes or security access.

In this sense, the database was divided into 11 different configurations. Table 3 shows the relation between the number of signals used to train the client models and the number of clients available in each case. Note that there are 75 client models with 1 signal as training because at least 10 signals are required for the validation set and at least 5 for the evaluation set. Users 76 to 127 provided a smaller amount of data which was used as data from impostors.

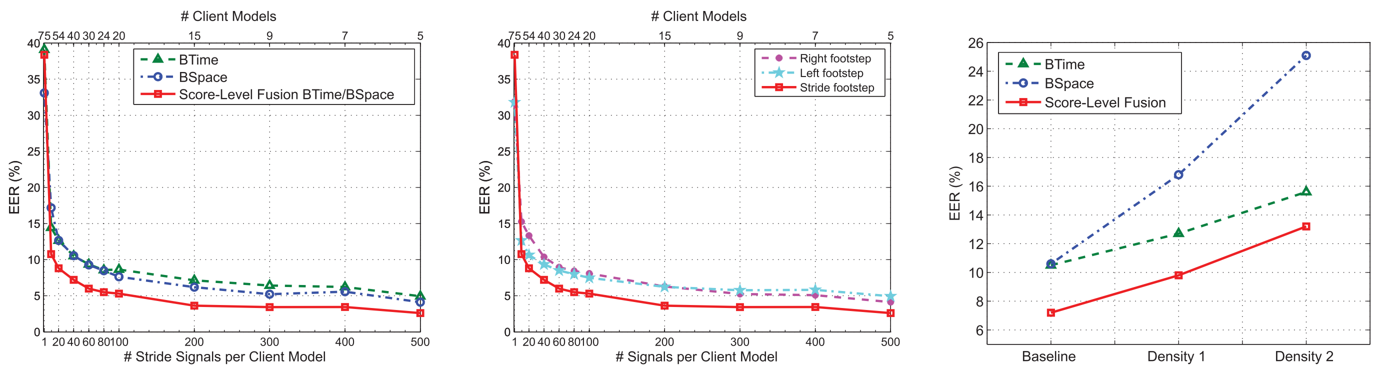
Fig. 10a shows a comparison of performance for the case of the stride footsteps for BTime, BSpace, and their fusion at the score-level. The figure shows the recognition performance in terms of EER against the different quantities of stride footstep signals used to train the models. The top abscissa axis shows the number of client models trained.

All three plots have a similar overall shape with 1) an initial steep fall from approximately 35 percent EER to 15 percent EER for the cases of BTime and BSpace and to 11 percent EER for their fusion when using 1 to 10 footsteps for training, 2) a smooth knee curve when increasing the number of signals used in the models from 20 to 80 where the error rates change less rapidly from 13 to 8.5 percent EER for the cases of BTime and BSpace, and from 9 to 5.5 percent EER for the case of the fusion, and 3) relatively flat profiles where error rates are around 4.5-5 percent EER for BTime and BSpace and around 2.5 percent EER for the case of the fusion when using 500 signals to train the reference models. EER results obtained in the left part of the figure could serve to estimate the performance of footstep recognition in a security access application obtaining results in the range of 5 to 10 percent EER. On the other hand, results in the right part of the figure could serve to estimate the performance of footstep recognition in a smart home application obtaining results in the range of 2.5 to 4 percent EER.

A very recent publication in the field of footstep recognition by Qian et al. [3] achieves recognition results of around 7.5 percent of error rate for the case of using 11 clients and 362 stride footsteps to train each client model. This would be more or less comparable with our case when using nine clients and 300 stride signals for training, in which we achieve 3.5 percent EER. A reason for this better performance in our case could be because we use holistic feature approaches for BTime and BSpace and then we carry out their fusion; on the other hand, they use a geometric approach, selecting a few key points of the center of pressure, using their spatial coordinates and the pressure value, therefore not considering a large amount of information contained in the signals.

### *6.3.2 Single versus Stride*

The study of single versus stride (consecutive footsteps) footsteps has been carried out in previous works such as [2], [4], [10], always obtaining better results for stride compared




(a) Comparison BTime, BSpace and Fusion (b) Score-Level Fusion BTime·BSpace (c) Sensor Density

Fig. 10. (a) Comparison of BTime, BSpace, and their fusion for stride footsteps with EER versus signals used per client model. (b) Comparison of score-level fusion of BTime and BSpace approaches for single and stride footsteps. (c) Comparison of EER against three sensor densities for stride footstep for BTime, BSpace, and their fusion. Baseline density (650 sensors per $\mathrm{m}^2$), density 1 (430 sensors per $\mathrm{m}^2$), and density 2 (220 sensors per $\mathrm{m}^2$).

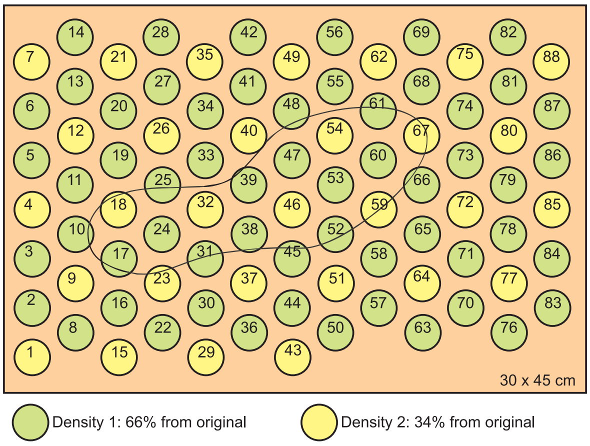


Fig. 11. Density of the array of piezoelectric sensors. Selected sensors for Densities 1 and 2. Example of a 9 UK size foot (27.5 cm long).

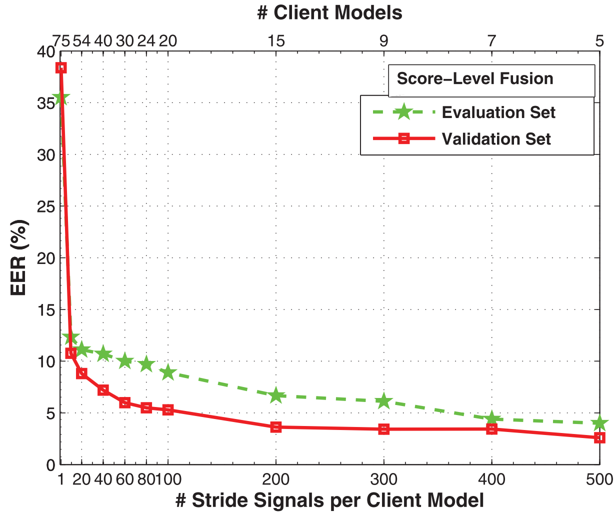


Fig. 12. Evaluation for the score-level fusion of BTime and BSpace and stride footsteps.

to single footsteps. In our case, we can only compare the performance between single and two consecutive footsteps.

Fig. 10b shows the results obtained for the case of the score-level fusion of BTime and BSpace for the single (right and left) and the stride footsteps. Plots follow the same trends, but with a significant improvement of the performance for the case of the stride footsteps with an average of 2.5 percent EER compared to the single footsteps. As in related works, better recognition performance could be achieved with the concatenation of a sequence of footstep signals.

### 6.3.3 Sensor Density

This section studies the influence of the sensor density and how it affects the performance, as this has not been considered in related works. It is obvious that if the sensor density is higher, more information can be extracted, but up to a spatial sampling limit.

Fig. 11 shows a diagram of the geometry and density of the piezoelectric sensors, for an example 9 UK size foot (27.5 cm long). A standard 88 sensor density ($\sim$650 sensors per $m^2$) plus two subsampling conditions are considered. The subsampling process is illustrated in the figure with the geometry of the sensors used for Densities 1 and 2.

Density 1 reduces the original sensor density by a 34 percent, i.e., from 88 to 58 sensors ($\sim$430 sensors per $m^2$), and Density 2 reduces the original sensor density by a 66 percent, i.e., from 88 to 30 sensors ($\sim$220 sensors per $m^2$). Density 2 was the optimal sampling distribution, having the sensors in a hexagonal array. In order to have another sampling distribution with a higher sensor density, sensors not used in Density 2 were used to form Density 1.

Fig. 10c shows EER results for the cases of BTime, BSpace, and their fusion for the three densities considered. Results relate to the benchmark division of the database shown in Table 2. As can be seen, the reduction of the sensor density significantly affects the recognition performance. Also, the figure shows a much worse result for the BSpace compared to the BTime approach for the cases of Densities 1 and 2. This due to the fact that BSpace features depend more directly on the sensor density.

The fusion of both feature approaches provides the best results in any case. The trends of EER are similar to the ones obtained for the case of BTime, and are much better than for the case of BSpace. It can be concluded that at least 650 sensors per $m^2$ are required to give the good performance presented in this paper. Given the trends of profiles in Fig. 10c, a higher density might provide even better results.

## 6.4 Evaluation of the System

This section describes the experimental work using the evaluation test set. This evaluation is carried out at this stage for the best working configuration obtained in the validation set. Fig. 12 shows the results for the evaluation test set, which is comprised of the last five signals provided by 110 people. This data was reserved for this trial assessment and has not been used in any form previously.

As could be expected, the performance obtained for the evaluation is worse compared to the validation set. Results of 12.1 percent EER are obtained for the case of 10 signals per model (60 models), which improves to 4 percent EER for the case of 500 signals per model (five client models). A reason for this difference in the performance could be that the signals comprising the evaluation set are the last signals provided by the users and therefore more likely to be more different from the data used in the reference models.

Performance results obtained in the left part of the figure, on the order of 10-12 percent EER, are to be expected in applications such as security access. In applications such as smart homes, better performance results on the order of 3-6 percent EER can be expected, which correspond to the right part of the profiles in the figure.

As an overview of the results achieved, Table 4 shows a comparison of performance in terms of EER for the experiments described in this paper. Three working points have been chosen from profiles in Figs. 10a, 10b, and 12; these are: 40, 100, and 500 signals used to train the client models for 40, 20, and 5 clients, respectively. The table shows EERs for the cases of single and stride footsteps, and for the validation set (BTime, BSpace, and Score-Level fusion) and evaluation set (only for the case of the fusion).

# 7 Conclusions and Future Work

## 7.1 Conclusions

This paper studies footstep signals as a biometric based on the largest footstep database to date, with more than 120 people and almost 20,000 signals.

TABLE 4
Comparison of EER in Percent for BTime, BSpace, and Their Fusion for the Cases of Validation/Evaluation and Single/Stride Footsteps Using 40, 100, and 500 Signals to Train the Client Models

| EER in % | | Single Footstep | | | Stride Footstep | | |
|---|---|---|---|---|---|---|---|
| | | 40 | 100 | 500 | 40 | 100 | 500 |
| Valid | BTime | 14.1 | 11.9 | 8.1 | 10.5 | 8.6 | 5 |
| | BSpace | 13.6 | 11 | 7.1 | 10.6 | 7.6 | 4.1 |
| | Fusion | 9.9 | 7.7 | 4.5 | 7.2 | 5.3 | 2.6 |
| Eval | Fusion | 15.2 | 13.4 | 7.9 | 10.7 | 8.9 | 4 |

The main contribution of the present work is the assessment of footsteps in time, in space, and in a combination of the two. This is in contrast with the great majority of the related works, which either use time or spatial information from the signals due mainly to the limitations of the capture systems. Interestingly, the performance for the two domains proves to be very similar, with equal error rates in the range of 5-15 percent for each domain, depending on the experimental setup, and in the range of 2.5-10 percent for their fusion. To our knowledge, these are the best results achieved for footstep recognition to date.

The experimental protocol is designed to study the influence of the quantity of data used to train the reference models, with a steep fall to 11 percent EER when using 1 to 10 signals, then a smooth knee curve from 9 to 6 percent EER when using 20 to 80 signals, and relatively flat profiles thereafter for the case of the fusion of BTime and BSpace and the stride footstep. This serves to predict the performance for footsteps in applications such as smart homes or border control scenarios for the first time.

Another interesting finding is an average relative improvement of 25 percent EER achieved for stride footsteps compared to single (right or left) footsteps, which suggests that performance could be further improved with the concatenation of a sequence of footstep signals.

Also, this paper studies the influence of the sensor density in the performance, showing a significant increment of the EER when the sensor density is reduced. This is more accentuated for the case of BSpace (average of 2.8 times greater). Note 650 sensors per $m^2$ was the maximum considered here due to physical limitations; results suggest that higher density might improve error rates.

### 7.2 Future Work

As footsteps are a relatively new biometric, there is large amount of research that can be carried out in this field.

Section 6.2 describes the fusion of the time and spatial domain information of the footstep signals. Significant improvements of 3 percent EER are reported with fusion at the feature and score levels. An extension to the work presented would be to combine the spatiotemporal information of the signals from the sensor level.

Experiments reported in this work relate to the situation where the training data and the test data are in logical time sequence to reflect real applications, i.e., all the model data are recorded before the test data. Even better results could be obtained for the case of randomizing the data used in the training and test sets, like some previous works such as [2], [3], [14]. This would produce artificially good results, but it would serve to analyze the effect of the time gap between training and test data.

## Acknowledgments

This work has been supported by projects Contexts (S2009/TIC-1485), Bio-Challenge (TEC2009-11186), and "Catedra UAM-Telefonica." Ruben Vera-Rodriguez is supported by a Juan de la Cierva Fellowship from the Spanish MINECO. The multimodal footstep database was collected with support from United Kingdom EPSRC and in this context the authors wish to acknowledge the major contributions of Nicholas W.D. Evans and Richard P. Lewis.

## References

[1] A. Itai and H. Yasukawa, "Footstep Classification Using Simple Speech Recognition Technique," *Proc. IEEE Int'l Symp. Circuits and Systems,* pp. 3234-3237, 2008.
[2] L. Middleton, A.A. Buss, A.I. Bazin, and M.S. Nixon, "A Floor Sensor System for Gait Recognition," *Proc. IEEE Workshop Automatic Identification Advanced Technologies,* pp. 171-176, 2005.
[3] G. Qian, J. Zhang, and A. Kidane, "People Identification Using Floor Pressure Sensing and Analysis," *IEEE Sensors J.,* vol. 10, no. 9, pp. 1447-1460, Sept. 2010.
[4] J. Yun, "User Identification Using Gait Patterns on UbiFloorii," *Sensors,* vol. 11, no. 3, pp. 2611-2639, 2011.
[5] M.S. Nixon, T.N. Tan, and R. Chellappa, *Human Identification Based on Gait.* Springer, 2005.
[6] R. Vera-Rodriguez, P. Tome, J. Fierrez, and J. Ortega-Garcia, "Fusion of Footsteps and Face Biometrics on an Unsupervised and Uncontrolled Environment," *Proc. SPIE Biometric Technology for Human Identification IX,* 2012.
[7] R. Vera-Rodriguez, N.W.D. Evans, and J.S.D. Mason, "Footstep Recognition," *Encyclopedia of Biometrics,* pp. 550-557, Springer, 2009.
[8] A. Pedotti, "Simple Equipment Used in Clinical Practice for Evaluation of Locomotion," *IEEE Trans. Biomedical Eng.,* vol. 24, no. 5, pp. 456-461, Sept. 1977.
[9] M.D. Addlesee, A. Jones, F. Livesey, and F. Samaria, "The ORL Active Floor," *IEEE Personal Comm.,* vol. 4, pp. 35-41, Oct. 1997.
[10] J. Suutala and J. Roning, "Methods for Person Identification on a Pressure-Sensitive Floor: Experiments with Multiple Classifiers and Reject Option," *Information Fusion,* vol. 9, pp. 21-40, 2008.
[11] R. Vera-Rodriguez, R.P. Lewis, J.S.D. Mason, and N.W.D. Evans, "Footstep Recognition for a Smart Home Environment," *Int'l J. Smart Home,* vol. 2, pp. 95-110, 2008.
[12] R.J. Orr and G.D. Abowd, "The Smart Floor: A Mechanism for Natural User Identification and Tracking," *Proc. Conf. Human Factors in Computing Systems,* pp. 1-9, 2000.
[13] C. Cattin, "Biometric Authentication System Using Human Gait," PhD dissertation, ETH Zurich, 2002.
[14] J.S. Yun, S.H. Lee, W.T. Woo, and J.H. Ryu, "The User Identification System Using Walking Pattern over the ubiFloor," *Proc. Int'l Conf. Control, Automation, and Systems,* pp. 1046-1050, 2003.
[15] J.-W. Jung, T. Sato, and Z. Bien, "Dynamic Footprint-Based Person Recognition Method Using a Hidden Markov Model and a Neural Network: Research Articles," *Int'l J. Intelligent Systems,* vol. 19, pp. 1127-1141, Nov. 2004.
[16] J. Suutala and J. Roning, "Combining Classifiers with Different Footstep Feature Sets and Multiple Samples for Person Identification," *Proc. Int'l Conf. Acoustics, Speech, and Signal Processing,* vol. 5, pp. 357-360, 2005.
[17] Y. Gao, M.J. Brennan, B.R. Mace, and J.M. Muggleton, "Person Recognition by Measuring the Ground Reaction Force Due to a Footstep," *Proc. Ninth Int'l Conf. Recent Advances in Structural Dynamics,* 2006.
[18] J.P. Stevenson, S.L. Firebaugh, and H.K. Charles, "Biometric Identification from a Floor Based PVDF Sensor Array Using Hidden Markov Models," *Proc. Sensors Applications Symp. Technology Conf.,* 2007.

[19] J. Suutala, K. Fujinami, and J. Röning, "Gaussian Process Person Identifier Based on Simple Floor Sensors," *Proc. European Conf. Smart Sensing and Context*, pp. 55-68, 2008.

[20] R. Vera-Rodriguez, J.S.D. Mason, and N.W.D. Evans, "Assessment of a Footstep Biometric Verification System," *Advances in Pattern Recognition: Handbook of Remote Biometrics,* pp. 313-327, Springer, 2009.

[21] R. Vera-Rodriguez, J.S.D. Mason, J. Fierrez, and J. Ortega-Garcia, "Analysis of Time Domain Information for Footstep Recognition," *Proc. Int'l Symp. Visual Computing,* pp. 489-498, 2010.

[22] R. Vera-Rodriguez, J. Mason, J. Fierrez, and J. Ortega-Garcia, "Analysis of Spatial Domain Information for Footstep Recognition," *IET Computer Vision,* vol. 5, no. 6, pp. 380-388, Nov. 2011.

[23] A.K. Jain, A. Ross, and S. Prabhakar, "An Introduction to Biometric Recognition," *IEEE Trans. Circuits and Systems for Video Technology,* vol. 14, no. 1, pp. 4-20, Jan. 2004.

[24] J. Suutala, S. Pirttikangas, J. Riekki, and J. Roning, "Reject-Optional LVQ-Based Two-Level Classifier to Improve Reliability in Footstep Identification," *Proc. Second Int'l Conf. Pervasive Computing,* vol. 3001, pp. 182-187, 2004.

[25] R. Vera-Rodriguez, J.S.D. Mason, and N.W.D. Evans, "Automatic Cross-Biometric Footstep Database Labelling Using Speaker Recognition," *Proc. Int'l Conf. Biometrics,* pp. 503-512, 2009.

[26] "The Swansea Footstep Biometric Database (SFootBD)," http://atvs.ii.uam.es/databases.jsp, 2012.

[27] V.N. Vapnik, *Statistical Learning Theory.* Wiley, 1998.

[28] NIST, "Speaker Recognition Evaluation Campaign," http://www.itl.nist.gov/iad/mig/tests/sre/, 2010.

[29] A. Martin, G. Doddington, T. Kamm, M. Ordowski, and M. Przybocki, "The DET Curve in Assessment of Detection Task Performance," *Proc. Eurospeech,* vol. 1, pp. 1895-1898, 1997.

[30] N. Poh, "Multi-System Biometric Authentication: Optimal Fusion and User-Specific Information," PhD dissertation, Ecole Polytechnique Federale de Lausanne, 2006.

[31] A. Jain, N. Karthik, and A. Ross, "Score Normalization in Multimodal Biometric Systems," *Pattern Recognition,* vol. 38, no. 12, pp. 2270-2285, 2005.

[32] J. Kittler, M. Hatef, R.P.W. Duin, and J. Matas, "On Combining Classifiers," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 20, no. 3, pp. 226-239, Mar. 1998.

[33] J. Fierrez-Aguilar, "Adapted Fusion Schemes for Multimodal Biometric Authentication," PhD dissertation, Universidad Politecnica de Madrid, May 2006.


**Ruben Vera-Rodriguez** received the MSc degree in electrical engineering from the Universidad de Sevilla, Spain, in 2006, and the PhD degree from Swansea University, United Kingdom, in 2010. Currently, he is working as a postdoctoral researcher in the Biometric Recognition Group—ATVS, Universidad Autonoma de Madrid, Spain, where he currently holds a Juan de la Cierva postdoctoral fellowship. His research interests include signal and image processing, pattern recognition, and biometrics. In 2007, he received the Best Paper Award at the Fourth International Summer School on Biometrics, held in Alghero, Italy, by top international researchers in the field.

**John S.D. Mason** received the MSc and PhD degrees in control and digital signal processing from the University of Surrey, United Kingdom, in 1973. In 1983, he formed the Speech and Image Group, School of Engineering, Swansea University, United Kingdom, where he is currently a senior lecturer. He has supervised more than 30 PhD students in signal processing, mostly in speech and speaker recognition. He has participated in various European projects of particular relevance, such as the Framework 6 Network of Excellence known as Biosecure.

**Julian Fierrez** received the MSc and PhD degrees in telecommunications engineering from the Universidad Politecnica de Madrid, Spain, in 2001 and 2006, respectively. Since 2002, he has been affiliated with the Biometric Recognition Group (ATVS), first at the Universidad Politecnica de Madrid and, since 2004, at the Universidad Autonoma de Madrid, where he is currently an associate professor. From 2007 to 2009, he was a visiting researcher at Michigan State University under a Marie Curie fellowship. His research interests include signal and image processing, pattern recognition, and biometrics, with emphasis on signature and fingerprint verification, multibiometrics, biometric databases, system security, and forensic applications of biometrics. He has been and is actively involved in European projects focused on biometrics (e.g., FP7 TABULA RASA and BEAT), and is the recipient of a number of distinctions for his research, including: the Rosina Ribalta award to the best Spanish PhD proposal in ICT in 2005, the best PhD thesis in computer vision and pattern recognition in 2005-2007 by the IAPR Spanish liaison, the Motorola best student paper at ICB 2006, the EBF European Biometric Industry Award 2006, the IBM best student paper at ICPR 2008, and the EURASIP 2012 Best PhD Award. He is a memeber of the IEEE.

**Javier Ortega-Garcia** received the PhD degree (cum laude) in electrical engineering from the Universidad Politecnica de Madrid, Spain, in 1996. He is the founder and codirector of the Biometric Recognition Group—ATVS. Currently, he is a full professor at the Escuela Politecnica Superior, Universidad Autonoma de Madrid. His research interests include biometrics signal processing. He has published more than 150 international contributions, including book chapters, refereed journal papers, and conference papers. He chaired Odyssey 2004: The Speaker and Language Recognition Workshop, cosponsored by ISCA and IEEE, and cochaired ICB 2009, the Third IAPR International Conference on Biometrics. Very recently, he has been appointed chair of ICB-13, the Fifth IAPR International Conference on Biometrics. He is a senior member of the IEEE.