

# Analysis of spatial domain information for footstep recognition

R. Vera-Rodriguez<sup>1,2</sup> J.S.D. Mason<sup>2</sup> J. Fierrez<sup>1</sup> J. Ortega-Garcia<sup>1</sup>

<sup>1</sup>Biometric Recognition Group – ATVS, Universidad Autonoma de Madrid, Avda. Francisco Tomas y Valiente, 11–28049 Madrid, Spain

<sup>2</sup>Speech and Image Research Group, Swansea University, Singleton Park, SA2 8PP Swansea, UK  
E-mail: ruben.vera@uam.es

**Abstract:** This study reports an experimental analysis of footsteps as a biometric. The focus here is on information extracted from the spatial domain of signals collected from an array of piezoelectric sensors. Results are related to the largest footstep database collected to date, with almost 20 000 valid footstep signals and more than 120 persons. A novel feature approach is proposed, obtaining three-dimensional images of the distribution of the footstep pressure along the spatial course. Experimental work is based on a verification mode with a holistic approach based on principal component analysis and support vector machines, achieving results in the range of 6–10% equal error rate (EER) depending on the experimental conditions of quantity of data used in the client models (200 and 40 signals per model, respectively). Also, this study includes the analysis of two interesting factors affecting footstep signals and especially spatial domain features, namely, sensor density and the special case of high heels.

## 1 Introduction

Footstep recognition is a relatively new biometric that aims at discriminating persons using walking characteristics extracted from floor-based sensors. One significant benefit of footsteps over other, better-known modes is that footstep signals can be collected unobtrusively with minimal or no personal cooperation, which can be very convenient for the users. 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 compromising mode.

Different techniques have been developed using different sensors, features and classifiers as described in [1]. The identification rates achieved of around 80–90% are promising and give an idea of the potential of footsteps as a biometric [2, 3]. However, these results are related to relatively small databases in terms of number of persons and footstep signals, typically around 15 people and perhaps 20 footsteps per person [4]. In this paper, results relate to the largest footstep database collected to date, with more than 120 people and almost 20 000 signals, enabling assessment with statistical significance.

Regarding the sensors employed to capture the footstep signals, two main approaches have been followed in the literature: switch sensors [5–7] have been used with a relatively high sensor density (ranging from 50 to 1024 sensors per m<sup>2</sup>) in order to detect the shape and position of the foot. On the other hand, different types of sensors that capture transient pressure [4, 8–11] have been used with

relatively low sensor density (typically 9 sensors per m<sup>2</sup>), more focused on the transient information of the signals along the time course.

The capture system considered here uses a high density of approximately 650 piezoelectric sensors per m<sup>2</sup>, which gives a good spatial information and measures transient pressure, in contrast to previous works.

This paper is focused on the analysis of the spatial information of the footstep signals. A novel feature approach is proposed, obtaining 3D images of the distribution of the footstep pressure along the spatial course. Verification results achieved are in the range of 6–10% of EER depending on the quantity of data used to train the client models (200 and 40 signals per model for the given results, respectively). A similar analysis was presented in [12], but focusing on the temporal information of the signals. In addition, we consider in this paper the effect of two factors affecting footstep signal performance, namely, sensor density and the case of high heel shoes.

The paper is organised as follows. Section 2 describes the footstep signals and the collection of the database. Section 3 presents the feature extraction process, focused on spatial information. Section 4 describes the experimental protocol followed, Section 5 presents the experimental results; and finally conclusions are drawn in Section 6.

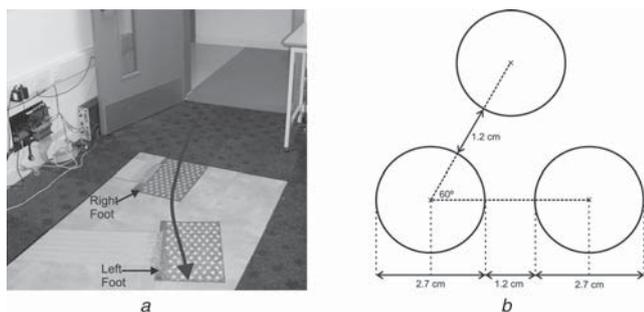
## 2 Footstep signals and database collection

As mentioned above, the main objective regarding the footstep signals was to obtain signals with biometric information in both time and spatial domains. Therefore the

capture system developed to collect the footstep database uses piezoelectric sensors with a relatively high density, this way footstep signals collected contain information in both time and spatial domains. This is in contrast to previous related works, for example, [6–8]. Piezoelectric sensors have some properties that make them very appealing for the application including low cost, robustness and a very thin profile that is ideal for under-floor concealment. As stated in [9], ‘piezoelectric sensors seem perfectly adequate for this application even with their low cost’. Piezoelectric sensors provide a differential voltage output that is directly proportional to the applied pressure.

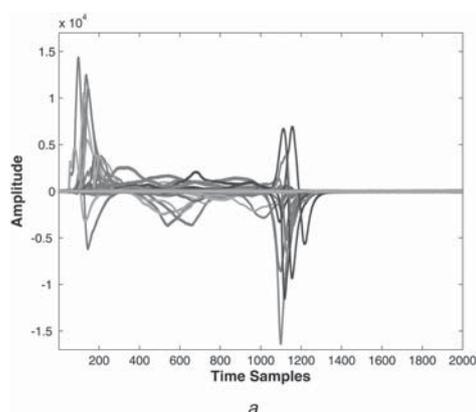
The sensors were mounted on a large printed circuit board and placed under a conventional mat. There are two such mats positioned appropriately to capture a typical (right, left) stride footstep. Each mat contains 88 piezoelectric sensors in an area of  $30 \times 45$  cm, with a sampling frequency of 1.6 kHz. The footstep sensor area, illustrated in Fig. 1a, is positioned at the entrance of a laboratory.

Fig. 1b shows a diagram of the distribution of the sensors in each array. The sensors have a diameter of 2.7 cm and the distance between two adjacent sensors is 1.2 cm. The geometry ( $60^\circ$  cellular layout) ensures a compact layout with uniform inter-sensor distance. Fig. 2 shows an example of footstep signal with information in both time and spatial domains obtained with the sensor distribution described. Fig. 2a shows the amplitude of the 88 pressure signals of one footstep against time, and Fig. 2b shows the



**Fig. 1** Footstep sensors location and distribution

a Two footstep sensor arrays in the laboratory where the capture system is installed. Sensors are shown here with the mats removed  
b Dimension of the sensors and geometry between adjacent sensors



**Fig. 2** Example of a footstep signal with time and spatial information

a Time domain signal: Sensor signals against time

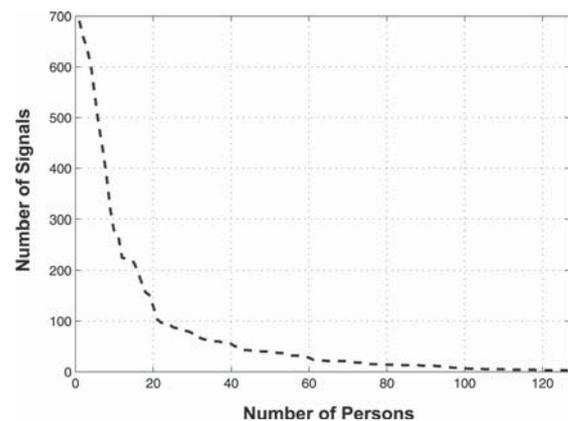
b Spatial domain signal: Accumulated pressure for the 88 sensors for the  $X$  and  $Y$  spatial axis

accumulated pressure in time for the 88 sensors for the  $X$  and  $Y$  spatial axis.

Regarding the collection of the database, one of the main objectives was to collect a database as large as possible. The first session of each person was a supervised enrolment, where a supervisor explained how to provide the footstep data. In this sense persons were asked to walk at a natural speed a few metres before the sensor mats (see Fig. 1a) in order to produce more realistic signals. Persons 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. The enrolment of persons in the system was continuous during the collection period (16 months). Also, different people provided data during different periods of time and in different number of sessions (different days), because as stated before the objective was to obtain a large database.

The main characteristic of the database collected is that it contains a large amount of data for a small subset of people ( $>200$  signals for 15 people) and a smaller quantity of data for a larger group of people ( $>10$  signals for 60 people). This reflects the mode of capture which was voluntary and without reward.

Fig. 3 shows the number of footstep signals per person in the database. There is a total of 9990 stride footstep signals,



**Fig. 3** Number of footstep signals against number of persons in the footstep database

that is, 19 980 single (right, left) footstep signals from the 127 persons enrolled. Fig. 6 shows two diagrams of the divisions of the data into different datasets for the experiments.

Regarding the type of footwear employed, persons were free to walk with different types of footwear such as shoes, trainers, boots, flip-flops, barefoot and even high heels (as reported in Section 5.2). Also, people were allowed to carry weights such as office bags. The associated biometric data from the different conditions were absorbed in the experiments reported, meaning that the results are more realistic in terms of the breadth of conditions encompassed.

The population of the database is mainly constrained to university students (undergraduates and postgraduates), as shown in Fig. 4a. The mean age value is 23.7 years and the ratio male/female is of 65% of males and 35% of females.

Fig. 4b shows the distribution of the height, having an overall average of 174 cm. Fig. 4c shows the distribution of the weight, which is mainly between values of 50 and 90 kg with a mean of 69.8 kg. Fig. 4d shows the distribution of the shoe sizes of the population, which is quite broad, having a mean value of 8.1 UK size.

More information about the collection and labelling of the database can be found in [13, 14].

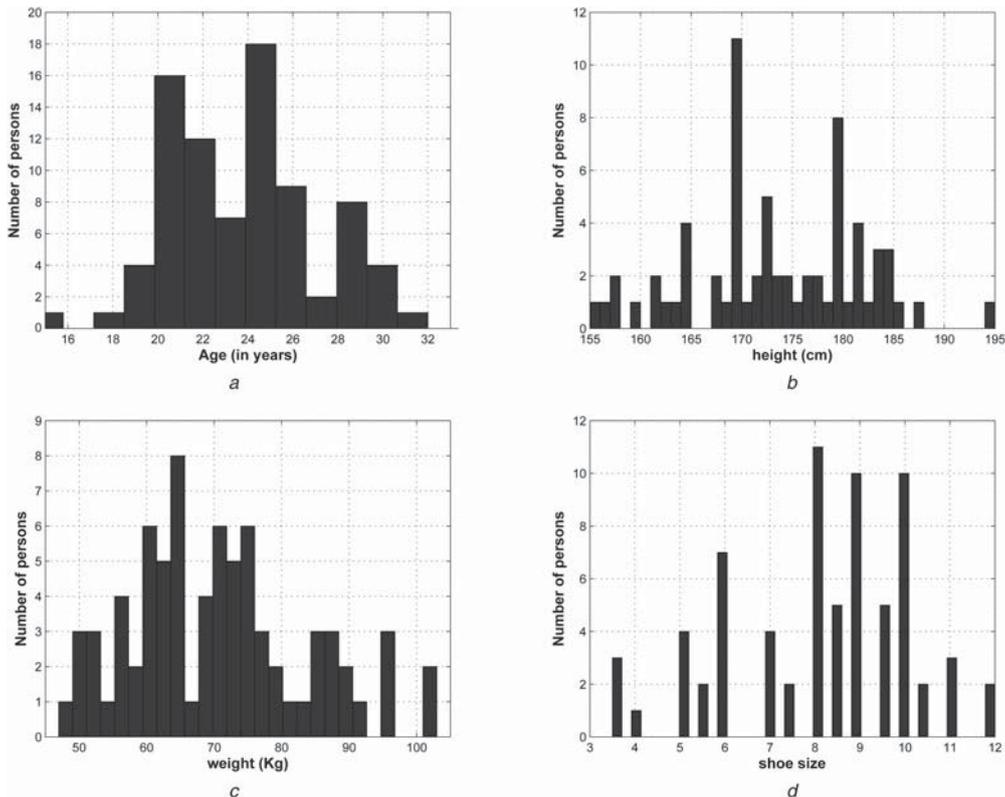
### 3 Feature extraction and matching

This section describes the spatial domain features that are used to assess the footstep signals as a biometric. A feature approach based on time-domain information was proposed in [12]. As a brief description, three features were extracted

from the time-domain information of the signals. The first was the popular ground reaction force (GRF) used previously in [2, 4, 8, 11, 15], in this case an average across all sensors was carried out to obtain a global profile for the GRF. The other two features were the spatial average of the sensors, which results in a single average profile of all sensors of the footstep signal; and finally the upper and lower contour profiles of the time domain signal. These three features were fused at the feature level, data dimensionality was reduced using principal component analysis (PCA) and finally, support vector machines (SVMs) were used to carry out the matching.

In this paper, the feature extraction is carried out over the spatial-domain information contained in the footstep signals. In this case, the time-domain information is not considered, so a single value of the pressure of each sensor of the mat is obtained by integrating the signals across the time axis. It is worth noting that owing to the differential nature of the footstep signals obtained from the piezoelectric sensors, a simple integration of the signal across the time would produce an approximately zero value. To solve this, the integration is carried out over the related GRF signal of each sensor ( $GRF_i$ ). This way we obtain the accumulated pressure ( $AP_i$ ), which is the measure used to study the distribution of the pressure across the spatial domain of the signals, as shown in Fig. 2b.

In the pre-processing 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. Formally,  $s_i[t]$  is the output of the piezoelectric sensor  $i$ ,  $i = 1, \dots, 88$  and  $t = 1, \dots, T_{max}$  are



**Fig. 4** Statistics of the population of the database  
 a Distribution of the age of the population, the mean being 23.7 years  
 b Distribution of the height of the population, the mean being 174 cm  
 c Distribution of the weight of the population, the mean being 69.8 kg  
 d Distribution of the shoe size of the population, the mean being 8.1 UK shoe size

the time samples.  $T_{max}$  was set to a value of 2000 time samples large enough for all footstep signals considered. Then, the  $GRF_i$  and  $AP_i$  are defined by

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

$$AP_i = \sum_{t=0}^{T_{max}} (GRF_i[t]) \quad (2)$$

The  $GRF_i[t]$  results in a profile per sensor, which is the integration of the output signal from the piezoelectric sensor. The  $AP_i$  gives a single value of the accumulated pressure for each sensor of the mat. Fig. 2b represents the 88 values of  $AP_i$  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  $mm^2$ , giving the values  $AP_i$  to the positions with sensors and zeros to the rest of the image, keeping this way the original geometry of the sensors. This resolution was chosen for simplicity, but in a real-time application it is likely that similar recognition results could be obtained with a lower image resolution.

The following step is the alignment and rotation of the spatial images to a fixed central position, but before, the images were smoothed using a Gaussian filter (defined in (3)) in order to obtain a continuous image as if we had a

much higher sensor resolution. Bicubic spline interpolation was also tried but better results were obtained using the Gaussian filter. Figs. 5a and b show the result image for the given example after the Gaussian filter from a lateral and a top view, respectively. Best result images were obtained using values of  $x, y = 1, \dots, 100$  and  $\sigma = 14$  for the filter.

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2/2\sigma^2)} \quad (3)$$

These images are then aligned and rotated based on the points with maximum pressure, corresponding with the toe and the heel areas, respectively. The aligned and rotated image is shown in Fig. 5c, which is used to carry out the biometric classification.

The rows of the resulting image, which has a dimension of  $280 \times 420$  pixels, are concatenated to form a feature vector of dimension 117 600. Data dimensionality are also reduced using PCA [16], retaining more than 96% of the original information by using the first 140 principal components. For the case of the stride (right and left) footstep, the feature vector comprises the concatenation of the 140 component feature vectors for the right and left foot plus the relative angle and length of the stride, that is, 282 features. Regarding the classifier, an SVM [17] was adopted with a radial basis function (RBF) as the kernel, because of very good performance in previous studies in this area [2, 3].

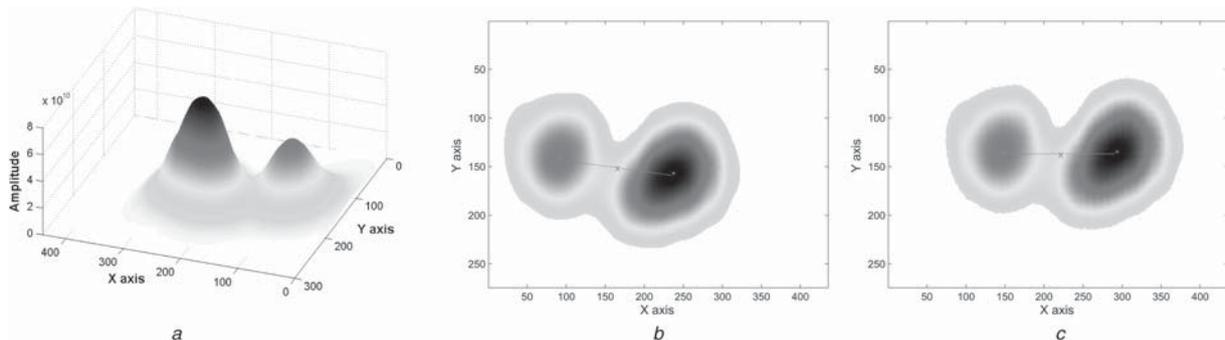


Fig. 5 Feature extraction in spatial domain for a footstep signal

- a 3D view image result for the smoothing of signal from Fig. 2b with a Gaussian filter
- b Same as a but from a top view
- c Footstep spatial image after alignment and rotation to a common centre

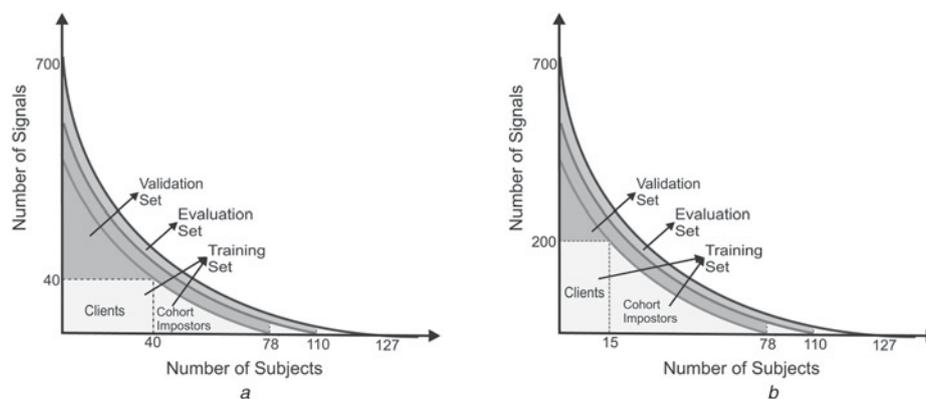


Fig. 6 Number of footstep signals against number of subjects in the database

Diagram of the database with the different divisions of Training, Validation and Evaluation sets for benchmarks B1 and B2. Numbers are described in Table 1

- a Benchmark B1
- b Benchmark B2

## 4 Experimental protocol

Regarding the experimental protocol followed to assess footsteps as a biometric, special attention was paid to the partitioning of the data into three sets, namely Training, Validation and Evaluation sets. Fig. 6 shows two diagrams of the partitioning of the database into three datasets.

The Training set comprises 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 also 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 comprising the last five footstep signals provided by persons P1–P110, while the Validation set is an unbalanced set that contains a larger number of test signals for subjects included in the Training data. The Validation set is used to tune the system, that is, type of features, number of PCA components, SVM parameters etc., in order to obtain the best results. The Evaluation set comprises unseen data, not used in the development of the system.

It is worth noting that in this paper the data used in the different sets keep the chronological time of the collection. Therefore the training data comprises the first data provided by each user, and the data used in the Evaluation set are the last collected. This is a realistic approach reflecting actual usage in contrast to previous related works, for example, [3, 5, 6], which randomly divide the data into training and test sets, or use a leave-one-out approach.

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 [4, 8, 9], owing to limited numbers 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 simulate important applications such as smart homes and access control scenarios. In the case of a smart home, there would be potentially 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.

Two benchmark points have been defined to simulate conditions of different applications, as can be seen in

Fig. 6: benchmark B1, which could simulate a security-access scenario, is set to use 40 footstep signals per client to train the models having a group of 40 clients (and therefore 40 models); and benchmark B2, which could simulate a smart home scenario, is set to use 200 footstep signals per client to train the models having a group of 15 clients.

Table 1 shows the quantity of data used in benchmarks B1 and B2 for the different datasets. Each signal from the test sets is matched against all the trained models (40 models in B1 and 15 in B2). As can be seen in the table, the total number of stride signals in the database is 9990, that is 19 980 single (right and left) signals in total.

## 5 Experimental results

This section describes the assessment of the spatial domain features described in Section 3 following the protocols defined in Section 2.

Fig. 7 shows the detection error tradeoff (DET) curves obtained for the Validation set for the cases of the stride and single (right, left) footstep signals for the spatial features described in Section 3. Fig. 7a shows the results for B1, that is, using 40 client models and 40 signals to train each model. Error rates of 10.5% are achieved for the case of stride footsteps and an average of 13.6% for the case of single footsteps. Fig. 7b shows the DET curves results for B2, that is, using 15 models and 200 signals per model. Error rates of 6.2% are achieved for the case of stride footsteps and an average of 9.6% for the case of single footsteps.

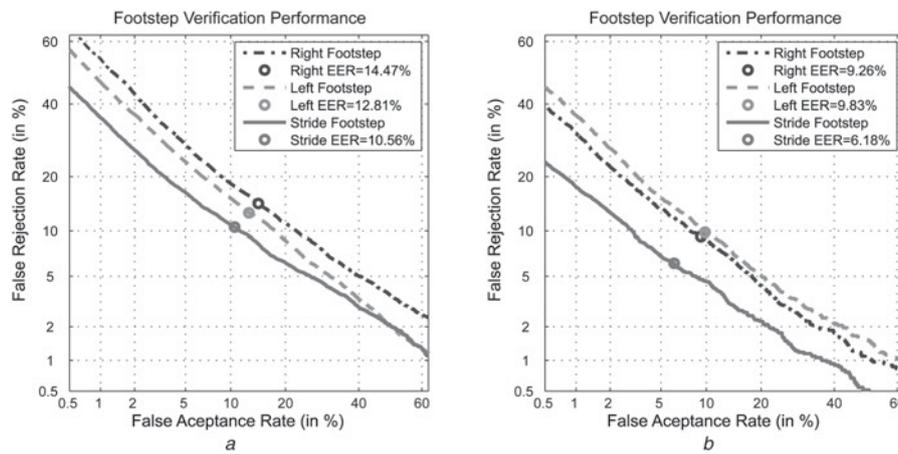
As can be seen in both cases there is an improvement of around 3% EER when single footstep signals are concatenated to produce a stride footstep signal. Previous experiments [2] showed an identification accuracy of 63% using a single footstep signal as a test, and improving to a 92% when six consecutive footstep signals were used. This implies that even better EER results could be obtained in case of concatenating more than two footstep signals. It is worth noting that results in the same range were achieved in [12], which presented a similar experimental protocol but for the case of the time domain information of the signals.

Although the conditions of B1 and B2 are not directly comparative there is an improvement of performance of 4.4% EER for the stride case for B2 compared with B1. This increment of performance could be an effect of the different amounts of signals in the client models or owing to the different number of clients in the two benchmarks.

In order to study this difference in performance, a further experiment was designed keeping the number of client models and varying the number of signals in the models. Fig. 8 shows the EER against different quantities of signals

**Table 1** Database configuration for benchmarks B1 and B2. B1 contains 40 models and 40 signals per model and B2 contains 15 models and 200 signals per model

	Benchmark B1			Benchmark B2		
	Training set	Validation set	Evaluation set	Training set	Validation set	Evaluation set
clients	P1–P40	P1–P40	P1–P40	P1–P15	P1–P15	P1–P15
signals per client	40	170 (8–650)	5	200	210 (15–490)	5
total signals clients	1600	6697	200	3000	3113	75
cohort impostors	P41–P127	P41–P78	P41–P110	P16–P127	P16–P78	P16–P110
total signals impostors	763	380	350	2697	630	475
total signals per set	2363	7077	550	5697	3743	550
total		9990			9990	



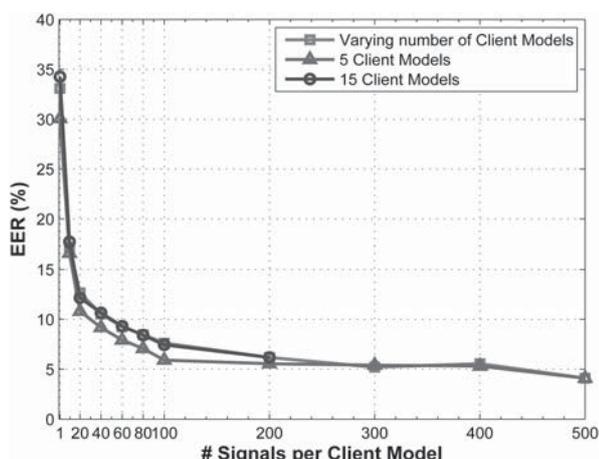
**Fig. 7** DET curves for spatial features extracted from the signals for the Validation set

a Results for benchmark B1 (40 models, 40 signals/model)

b For benchmark B2 (15 models, 200 signals/model)

used to train each client model for the case of stride footsteps. There are three profiles, one considering 15 client models giving values of EER from 1 to 200 signals used to train the models, another profile considering 5 client models but giving values of EER from 1 up to 500 signals used to train the models and the case of using the maximum number of client models available at each condition, that is, 5 models for 500 signals, 20 models for 200 signals, 40 models for 40 signals etc. The three profiles have a similar overall shape with a great improvement in performance when using 1–20 footstep signals for training, falling from an average of 33–11% EER, and then the performance keeps improving slowly to 4% EER with 500 signals to train the system (for the case of using 5 client models). Results obtained in the three cases are very similar, so it can be concluded that the improvement of performance is mainly owing to the number of signals used to train the client model rather than to the number of models considered.

Fig. 9 shows the DET curves obtained for the Evaluation set for benchmarks B1 and B2 for the stride footsteps. In both cases, the result obtained for the Evaluation set is compared with the case of the Validation set as shown in Fig. 7. The data used in the Validation and Evaluation sets



**Fig. 8** EER against number of signals used to train the client models in different benchmarks for the stride footstep for the cases of using 5, 15 client models and using the maximum number of client models available at each condition

are specified in Table 1. In both cases of B1 and B2 there is a superior performance for the case of the Validation set compared with the Evaluation, with an absolute increment of 5.5 and 2.6% EER for B1 and B2, respectively. This degradation of performance for the case of the Evaluation set could be because of the big time gap between the data used for training and test signals because in this case the Evaluation set comprises the last signals collected for each person and the signals comprising the Training set are the first 40 and 200 signals per person for B1 and B2, respectively (see Fig. 6). This effect of the relationship of the time gap between training and test data and performance is an interesting line of further investigation.

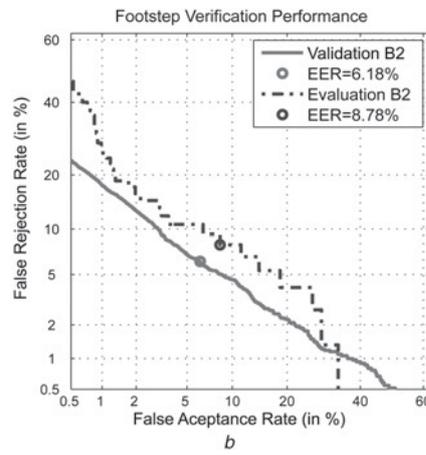
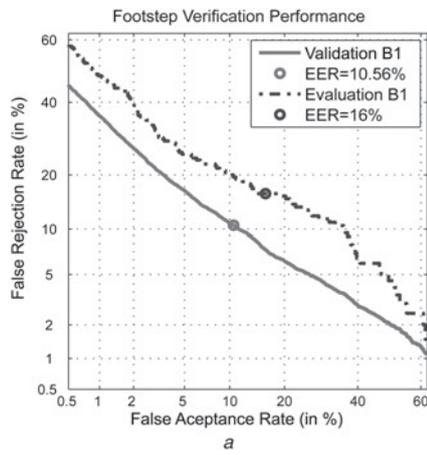
### 5.1 Influence of the sensor density in the performance

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

Fig. 10 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 (650 sensors per  $m^2$ ) plus two sub-sampling conditions are considered. The sub-sampling process is illustrated in the figure with the geometry of the sensors used for Density 1 and for Density 2.

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

Fig. 11 shows EER results for benchmarks B1 and B2 for the three densities for the case of the stride footstep. The trends of EER are very similar for both benchmarks, with an average increment of 7.7% EER for Density 1 compared with the baseline, and an average increment of 13.2% EER for Density 2 compared with the baseline. As can be seen the spatial features are very much affected by the reduction of the sensor density.



**Fig. 9** DET curves for the Evaluation set of stride footstep signals

Comparison of results for the Evaluation set for benchmarks B1 and B2

a Evaluation for benchmark B1

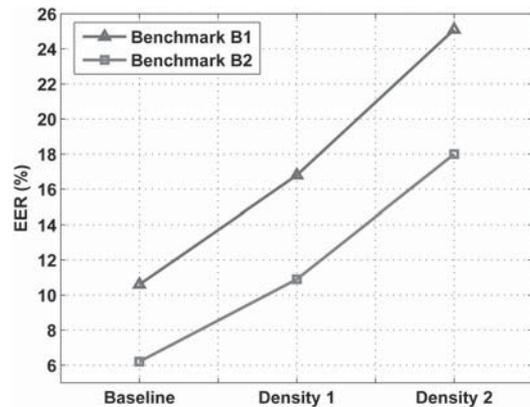
b Evaluation for benchmark B2

It can be concluded that at least 650 sensors per m<sup>2</sup> are required to give the good performance presented in this paper when only spatial information is considered. Given the trends of profiles in Fig. 11, a higher density might provide even better results.

### 5.2 Analysis of the special case of high heels

This section analyses the effect of the special case of high heels on the performance. Persons contributing to the database do so under different conditions such as different types of footwear or extra weight. These conditions are absorbed in the experiments, meaning that the results are more realistic because of the breadth of conditions encompassed.

Here the effect of high heels is analysed using an illustrative example of 40 footstep signals provided by one subject wearing high heels in three different sessions (two different pairs of heels). Fig. 12 shows two footstep examples for this subject; in (a), (b) and (c) the person is wearing trainers and in (d), (e) and (f) the person is wearing high heels. As expected, high-heel data looks completely different, as can be seen in Fig. 12; and therefore higher error rates are to be expected for these signals.



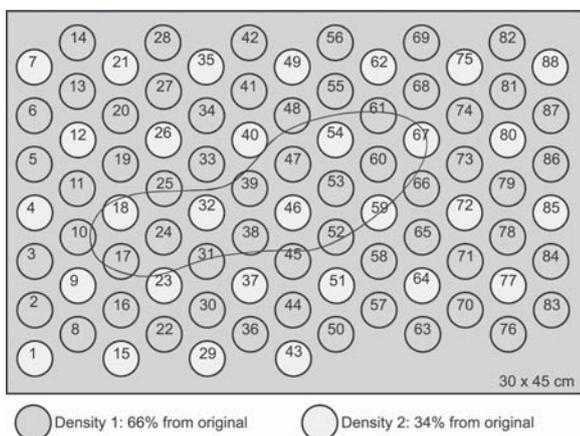
**Fig. 11** EER against three different sensor densities for benchmarks B1 and B2 and stride footstep

Baseline density (650 sensors per m<sup>2</sup>), Density 1 (430 sensors per m<sup>2</sup>) and Density 2 (220 sensors per m<sup>2</sup>)

The analysis was carried out using the same experimental protocol described in Section 2 for B1, that is, having 40 client models, each comprising 40 signals to train each model, and the Validation test set. Two different experiments were considered:

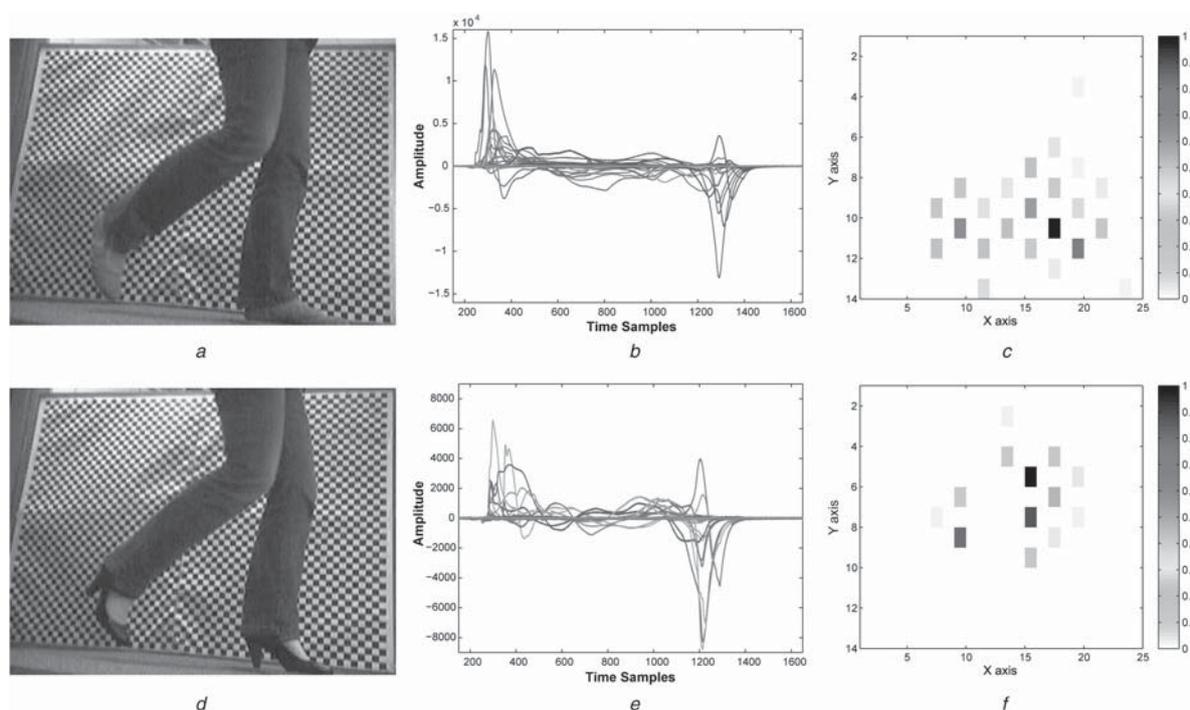
- Experiment 1. In this case no data with high heels were included in the model for the subject under study, and errors produced by 20 test signals with high heels for that person were analysed.
- Experiment 2. In this case 10 signals with high heels were included in the model for the subject under study (25% of the total training data for that subject), and errors produced by the same 20 test signals as in Experiment 1, which were collected in a different session were analysed.

Table 2 shows the results of the error analysis carried out for the two experiments. The 20 test signals analysed with high heels for the subject under study were compared with the 40 client models available, having for each test one genuine comparison (with the model from the subject under study) and 39 false comparisons with the rest 39 client models; so in total there are 20 genuine comparisons and 780 impostor comparisons. This error analysis was carried



**Fig. 10** Density of the array of piezoelectric sensors

Example of a 9 UK size foot (27.5 cm long)



**Fig. 12** Examples of footstep signals with two types of footwear

- a* Person wearing trainers  
*b* Response of the piezoelectric sensors against time for example in *a*  
*c* Accumulated pressure in the spatial domain, for example in *a*  
*d, e* and *f* Same but for the case of the same person wearing high heels

**Table 2** Error analysis for the case of high heels. Results obtained for 20 signals with high heels tested against 40 client models

Condition	Exp 1: no heels in training		Exp 2: heels in training	
	FRR, %	FAR, %	FRR, %	FAR, %
time	70	12.3	0	9.2
space	75	12.8	10	11.3

out for the cases of time-domain features (as in [12]) and spatial domain features (as described in Section 3). Percentages of ‘false rejection rate’ (FRR) and ‘false acceptance rate’ (FAR) are given in Table 2. Percentages are obtained by using the threshold that gave the EER.

As could be expected, results shown in Table 2 for Experiment 1 are significantly worse compared with those obtained for Experiment 2, the case where signals with high-heels condition are present in both training and test sets. In the case of Experiment 1 the FRR is very high as there are a greater number of true tests below the threshold. It is interesting to note that results are slightly worse for the case of spatial domain features compared with the time-domain features, as the distribution of the pressure across the space is more affected by the condition of high heels (as can be seen in Figs. 12c and f). It can be concluded that high heels degrade the performance, but this effect can be significantly reduced when the same condition is included in the training data (as in Experiment 2).

An equivalent analysis of this could be done on other footwear conditions or extra weight, which is proposed as further work.

## 6 Conclusions

This paper studies footstep signals as a biometric focusing on the spatial information of the signals. A novel feature approach extracts biometric information from the distribution of the pressure of the footsteps signals along the spatial domain.

The experimental protocol is designed to study the influence of the quantity of data used in the client models, simulating conditions of possible extreme applications such as smart homes or border control scenarios. Results in the range of 6–10% EER are achieved in the different conditions for the case of the stride footstep. These results are in the same range as those achieved in [12] for a similar approach but considering the temporal information of the signals, which implies that the time and spatial information extracted from the footstep signals have similar discriminatory properties.

It is worth noting that the experimental set-up here is the most realistic at least in two factors: (i) it considers the largest footstep database to date and (ii) it keeps the time lapse between training and test data, in contrast to most previous works, for example [3 5, 6], which randomise the time sequence of the data in the experiments.

This paper also analyses two important factors not taken into account in previous studies in the area. Capturing systems have used different sensor densities, but there is no study of the influence of the variation of sensor density in the performance, which is considered here for the first time. Experiments show that reduction of the sensor density affects significantly the recognition performance. This indicates that a relatively high density is necessary and might well contribute to the good recognition performance reported here compared with that in related publications [2, 9].

Also, the special case of the high heels has been analysed, which is a good illustrative example, of the capacity of the database considered here. The error analysis shows that high heels affect significantly the performance; however, this can be reduced when the same conditions are included in the training data.

For further work, it would be interesting to carry out a fusion of both approaches of time and spatial domain information, as well as a holistic feature approach to extract both time and spatial information from the signals simultaneously.

## 7 Acknowledgments

Ruben Vera-Rodriguez, Julian Fierrez and Javier Ortega-Garcia are supported by projects Contexts (S2009/TIC-1485), Bio-Challenge (TEC2009-11186), TeraSense (CSD2008-00068) and 'Catedra UAM-Telefonica'. Postdoctoral work of Ruben Vera-Rodriguez is supported by a Juan de la Cierva Fellowship from the Spanish MICINN. The multimodal footstep database was collected with support from UK EPSRC and in this context the Authors wish to acknowledge the major contributions of Nicholas W.D. Evans and Richard P. Lewis.

## 8 References

- 1 Vera-Rodriguez, R., Evans, N., Mason, J.: 'Footstep recognition', in 'Encyclopaedia of biometrics' (Springer, 2009)
- 2 Suutala, J., Roning, J.: 'Methods for person identification on a pressure-sensitive floor: experiments with multiple classifiers and reject option', *Inf. Fusion*, 2008, **9**, (1), pp. 21–40
- 3 Vera-Rodriguez, R., Lewis, R., Mason, J., Evans, N.: 'Footstep recognition for a smart home environment', *Int. J. Smart Home*, 2008, **2**, pp. 95–110
- 4 Addelee, M.D., Jones, A., Livesey, F., Samaria, F.: 'The ORL active floor', *IEEE Pers. Commun.*, 1997, **4**, pp. 235–241
- 5 Yun, J.S., Lee, S.H., Woo, W.T., Ryu, J.H.: 'The user identification system using walking pattern over the ubiFloor'. Proc. Int. Conf. on Control, Automation, and Systems, 2003, pp. 1046–1050
- 6 Middleton, L., Buss, A.A., Bazin, A.I., Nixon, M.S.: 'A floor sensor system for gait recognition'. Proc. Fourth IEEE Workshop on Automatic Identification Advanced Technologies, (AutoID'05), 2005, pp. 171–176
- 7 Suutala, J., Fujinami, K., Roning, J.: 'Gaussian process person identifier based on simple floor sensors'. Proc. 3rd Eur. Conf. on Smart Sensing and Context, (EuroSSC'08), Zurich, Switzerland, 2008, pp. 55–68
- 8 Orr, R.J., Abowd, G.D.: 'The smart floor: a mechanism for natural user identification and tracking'. Proc. Conf. on Human Factors in Computing Systems, 2000
- 9 Cattin, C.: 'Biometric authentication system using human gait'. PhD thesis, Swiss Federal Institute of Technology, ETH, Zurich, 2002
- 10 Suutala, J., Pirttikangas, S., Riekkii, J., Roning, J.: 'Reject-optional LVQ-based two-level classifier to improve reliability in footstep identification'. Proc. Pervasive Computing, Vienna, Austria, 2004, (*LNCS*, **3001**), pp. 182–187
- 11 Gao, Y., Brennan, M.J., Mace, B.R., Muggleton, J.M.: 'Person recognition by measuring the ground reaction force due to a footstep'. Proc. 9th Int. Conf. on Recent Advances in Structural Dynamics, Southampton, UK, 2006
- 12 Vera-Rodriguez, R., Mason, J., Fierrez, J., Ortega-Garcia, J.: 'Analysis of time domain information for footstep recognition'. Proc. 6th Int. Symp. on Visual Computing, (ISVC'2010), Las Vegas, NV, USA, 2010, (*LNCS*, **6453**), pp. 489–498
- 13 Vera-Rodriguez, R., Lewis, R.P., Mason, J.S.D., Evans, N.W.D.: 'A large scale footsteps database for biometric studies created using cross-biometrics for labelling'. Proc. 10th IEEE Int. Conf. on Control, Automation, Robotics and Vision (ICARCV), Hanoi, Vietnam, 2008, pp. 1361–1366
- 14 Vera-Rodriguez, R., Mason, J., Evans, N.: 'Automatic cross-biometric footstep database labelling using speaker recognition'. Proc. 3rd IAPR/IEEE Int. Conf. on Biometrics (ICB), June 2009, (*LNCS*, **5558**), pp. 503–512
- 15 Suutala, J., Roning, J.: 'Combining classifiers with different footstep feature sets and multiple samples for person identification'. Proc. Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP), 2005, vol. 5, pp. 357–360
- 16 Jolliffe, I.T.: 'Principal component analysis' (Springer Series in Statistics, Springer, 2002)
- 17 Vapnik, V.N.: 'Statistical learning theory' (Wiley, New York, 1998)