Fusion of Footsteps and Face Biometrics on an Unsupervised and Uncontrolled Environment

Ruben Vera-Rodriguez, Pedro Tome, Julian Fierrez and Javier Ortega-Garcia

Biometric Recognition Group - ATVS, Universidad Autonoma de Madrid,
Avda. Francisco Tomas y Valiente, 11 - 28049 Madrid, Spain

ABSTRACT

This paper reports for the first time experiments on the fusion of footsteps and face on an unsupervised and not controlled environment for person authentication. Footstep recognition is a relatively new biometric based on signals extracted from people walking over floor sensors. The idea of the fusion between footsteps and face starts from the premise that in an area where footstep sensors are installed it is very simple to place a camera to capture also the face of the person that walks over the sensors. This setup may find application in scenarios like ambient assisted living, smart homes, eldercare, or security access. The paper reports a comparative assessment of both biometrics using the same database and experimental protocols. In the experimental work we consider two different applications: smart homes (small group of users with a large set of training data) and security access (larger group of users with a small set of training data) obtaining results of 0.9% and 5.8% EER respectively for the fusion of both modalities. This is a significant performance improvement compared with the results obtained by the individual systems.

Keywords: Multimodal biometric, fusion, footstep recognition, gait recognition, face recognition

1. INTRODUCTION

Unobtrusive biometric systems are receiving recently great attention from the research community due to the high degree of acceptability from the users in different applications. Footsteps and face are two good examples of unobtrusive biometrics, which could be fused. Footstep recognition is a relatively new biometric, which aims to discriminate persons using walking characteristics extracted from floor-based sensors. Footstep signals are very robust to environmental conditions, with minimal external noise sources to corrupt the signals.[1] On the other hand, face is a modality with better individual performance compared to footsteps, but it is strongly affected by external factors such as illumination, pose, subject-to-camera distance or appearance.[2]

This paper is focused on the fusion of footsteps and face on an unsupervised and uncontrolled environment. Some previous related works have carried out the fusion between face and gait[3,4] achieving very good recognition results due mainly to the uncorrelation of both biometrics. This is a similar case to ours, although footsteps is a more controlled mode compared to gait, but signals are more robust to environmental conditions, with minimal external noise sources to corrupt the signals. In our case, footstep signals and face images are very easy to be collected together by simply placing a camera to capture the face of the person that walks over the footstep sensors. This is a good example of an unobtrusive and transparent multimodal biometric system as the person walks freely over an area without having to interact with any device. Figure 1 shows a diagram of the arrangement of the footstep sensors and the face camera in the case considered here.

The database considered in this paper[5] was collected on an unsupervised and uncontrolled manner, i.e., factors providing variability in each biometric mode such as illumination, pose, etc. for face, and footwear or speed for the case of footsteps were not controlled, which is a much more challenging problem and results achieved are more realistic in terms of the breadth of conditions encompassed.

Further author information: (Send correspondence to R.V.-R.)
R.V.-R.: E-mail: ruben.vera@uam.es
P.T.: E-mail: pedro.tome@uam.es
J.F.: E-mail: julian.fierrez@uam.es
J.O.-G.: E-mail: javier.ortega@uam.es

Sensing Technologies for Global Health, Military Medicine, Disaster Response, and Environmental Monitoring II; and Biometric Technology for Human Identification IX, edited by Sárka O. Southern, et al., Proc. of SPIE Vol. 8371, 83711U
 · doi: 10.1117/12.918550

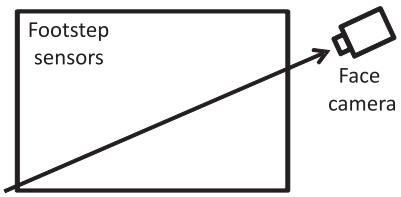
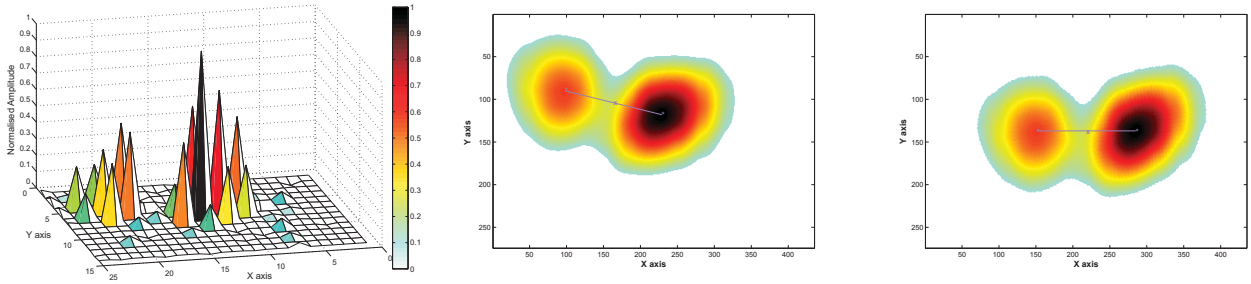


Figure 1. Arrangement of the footstep sensors and the face camera. The arrow shows the direction of the person walking.



(a) Accumulated pressure (b) As (a) after smoothing (c) After alignment and rotation

Figure 2. (a) Example of a footstep signal with the accumulated pressure in the X and Y axes. (b) Same as (a) after smoothing with a Gaussian filter. (c) Footstep image after alignment and rotation to a common centre.

The fusion of footsteps and face is carried out at the score-level with different score normalization techniques in order to make comparable the scores from the two systems. Two different fusion architectures have been considered, an ideal case having exactly the same number of footstep and face signals, and a more realistic case with an adaptive fusion for the case of having more footstep signals than face images. Also, two different applications have been simulated: smart homes (small group of users with a large set of training data) and security access (larger group of users with a small set of training data). Best results of 0.9% and 5.8% EER have been achieved for the fusion of both biometrics for each application respectively.

The paper is organized as follows. Section 2 describes the footstep signals and the footstep recognition system. Section 3 presents the face signals and the face recognition system. Section 4 describes the experimental protocol followed for the fusion, Section 5 presents the experimental results; and finally conclusions are drawn in Section 6.

2. FOOTSTEP RECOGNITION

This section describes the characteristics of the footstep signals and the recognition system developed. The main characteristic of the footstep signals considered here is that contain information in both time and spatial domains, in contrast to previous works.[6–8] In this case, a high density array of piezoelectric sensors (650 sensors per m^2), which capture the transient pressure, are arranged in a regular pattern working at a sampling frequency of 1.6 kHz. The area where the footstep sensors are placed (see Figure 1) is large enough to collect a stride (right to left) footstep signal. The cost of the footstep sensor array used here was around 2000 USD, which of course would be much cheaper for mass production.

In this paper, the features extracted from the footstep signals to carry out person recognition are based on the accumulated pressure of each piezoelectric sensor over a footstep, similar to the work in.[9] Figure 2(a) shows an example footstep signal with the accumulated pressure of each sensor for the X and Y axes. Alignment and rotation is carried out over this type of images to place them into a fixed position, but before, the images are smoothed using a Gaussian filter in order to obtain a continuous image. Figure 2(b) shows the result image for the given example after the Gaussian filter from a top view.

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 result image is shown in Figure 2(c), which is

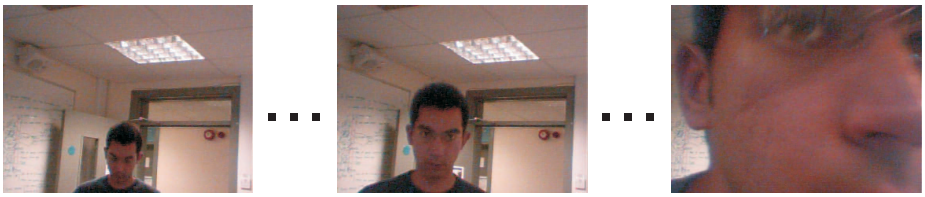


Figure 3. Example of a sequence of face images linked to a stride footstep signal in the database.

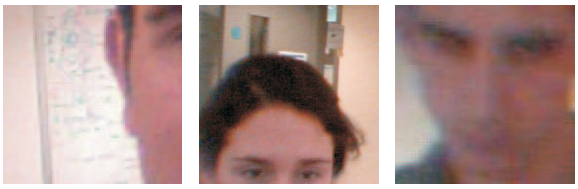


Figure 4. Examples of face images discarded

used to carry out the biometric classification. In this paper, we concatenate the resulting images for a stride (right to left) footstep signal into a feature vector, considering also the relative angle and length of the stride as features. Data dimensionality is reduced using principal component analysis (PCA), retaining more than 96% of the original information by using the first 200 components. Regarding the classifier, a support vector machine (SVM)[10] was adopted with a radial basis function (RBF) as the kernel, due to very good performance in previous studies in this area.[8,11]

3. FACE RECOGNITION

This section describes the face images and the recognition system developed. Face images are collected from a commercial low quality video camera at a frequency of 30 frames per second with a resolution of 640 × 480 pixels. For each stride footstep signal there is a linked face video of the person walking towards the camera. The synchronization between the face images and the footstep signals is made with a common timestamp of the collection of the two biometric modes. Figure 3 shows an example of a sequence of images collected in the database.

In the preprocessing stage, VeriLook SDK v2.0 (commercial system) is applied over all the images collected to locate and segment the face in the image. The resulting face images are size normalized to 64 × 80 pixels (width × height). VeriLook provides a quality index for each face image in the range of 0 to 100. Based on this, the image with the best quality index in each sequence is selected to carry out person recognition. Also, a threshold based on this quality index was set at value 55 to discard low quality images.

Figure 4 shows some examples of discarded face images due to, for example, partial face images (a), (b) or very blurred images (c). Note that these images are already the ones with the highest quality index of their sequence. Figure 5 shows some examples of face images considered in the experiments. Figure 5(a) is an example of a good quality image, (b) is a bit blurred, and (c) and (d) show two images of the same person with different lighting conditions. Figure 5 shows the difficulty of the problem addressed here considering images collected on

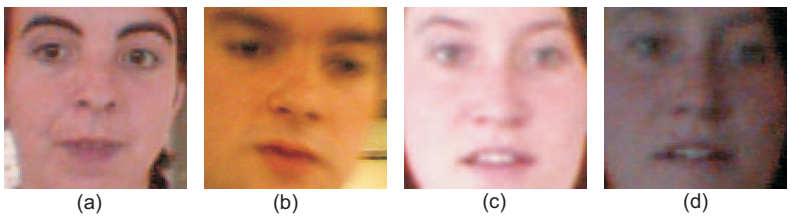


Figure 5. Examples of face images considered. Note it is an uncontrolled scenario.

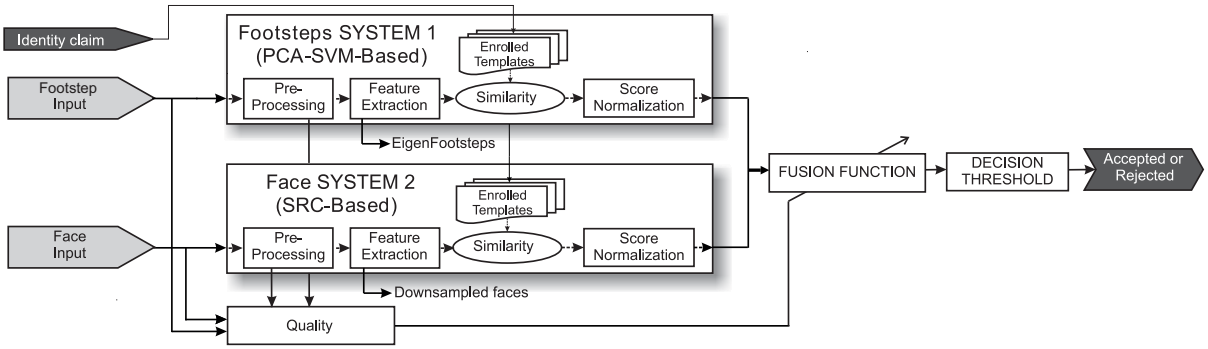


Figure 6. General architecture for the fusion of footsteps and face. The fusion function is different depending on the ideal and realistic cases defined.

an unsupervised and uncontrolled scenario. The only two controlled parameters are the subject aging (collection was carried out over a period of 16 months) and the camera, which was the same over the collection. From the selected images, pose is compensated with respect to the position of the eyes, and photometric compensation is used for the illumination in order to obtain better recognition results.

Regarding the matching, the verification system used here is a state-of-the-art system based on sparse representation classification (SRC).[12] The features used are simple downsampled images with a ratio of 1/4 going from the original 5120 dimensions to 320 dimensions, which form the feature vectors.

4. EXPERIMENTAL PROTOCOL

4.1 Fusion

This section describes the fusion of the footstep and face modalities. In general, multimodal biometric fusion can be carried out at different levels; in our case the fusion is carried out at the score level, which is the most common approach due to the ease on combining scores originated from the different matchers.[13,14] Figure 6 shows a general diagram of the fusion of the two systems. Two different fusion architectures have been followed:

1. Ideal case. In this case, there would be always a face image to be fused with a stride (right to left) footstep signal. A simple sum rule is adopted in this case as the fusion function shown in Figure 6. Section 5.1 shows the results for this case.

2. Adaptive fusion (realistic case). In this case, due to the rejection of some low quality face images, there would be more footstep signals than face images in the fusion. Therefore an adaptive fusion is carried out by fusing footsteps and face with a sum rule (as in the ideal case) when there are entries for the two modalities, and giving the whole weight to the footstep signal when there is no linked face image. Section 5.2 shows the results for this case.

Before carrying out the fusion of the two biometric modalities, score normalization is needed to transform the scores in order to make them comparable, i.e., put them in a common domain. Following findings from Jain *et al.*,[14] max-min, Z-norm and tanh-norm are compared in Section 5. In the cases of Z-norm and tanh-norm, scores are also transformed with a logit function[15] given by:

$$s' = \text{logit}(s) = \log \frac{s}{1-s} \tag{1}$$

before applying the normalization, where s is the score and s' denotes the "logit" transformed score.

		Security Access (SA)			Smart Home (SH)		
		Train	Test		Train	Test	
	Subjects	P1-P54	P1-P54	P55-P78	P1-P15	P1-P15	P16-P78
Foot	#Signals	1080	7725	250	3000	3113	630
Face	#Signals	809	5932	179	2148	2561	484
	Discarded	25,1%	23,2%	25,4%	28,4%	17,7%	23,2%

Table 1. Database configuration for the two applications considered: security access and smart homes.

4.2 Application Scenario

Two different application scenarios have been designed in the experimental protocol: smarts homes and security access. 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), 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 the mode of capture which was voluntary and no supervised.

In a smart home scenario, the proposed biometric system could be placed in the entrance of a house or high security area of a building such as a bank or an embassy for example. In the case of a security access scenario, the biometric system could be placed to control the access where not much training data is available, for example a security gate at an airport.

Table 1 shows the divisions of the database into training and test sets for each application considered. The number of signals in the footstep and face modes are different due to the rejection of some face images as described before. In particular, 24.5% of the face images are discarded in the case of the security access (SA) application, and 23.1% are discarded in the case of the smart home (SH) application.

5. EXPERIMENTAL RESULTS

This section describes the experimental results of the fusion of footstep and face modes on an uncontrolled and unsupervised environment. Results are shown using ROC curves with EERs and verification rates (VR) working at a FAR=0.001.

5.1 Ideal Case

This section presents the performance results for the case of the ideal fusion (as described in Section 4.1). Figure 7 shows the recognition performance of the footstep and face systems when operated as unimodal systems. Thus, there is the same number of signals for footstep and face modes which corresponds to the number of face signals shown in Table 1. ROC curves for both applications considered show significantly better results for the case of smart homes (SH) compared to security access (SA), due mainly to the larger quantity of training data used per subject and the smaller group of subjects in the training. Also, results achieved for face are better than those obtained for footsteps.

The performance of the multimodal biometric system has been studied under different score normalization techniques for the simple sum fusion rule of the scores. The normalized scores were obtained by using one of the following techniques: max-min normalization, logit transformation with Z-norm, tanh-norm, and logit transformation with tanh-norm. ROC curves for SA and SH applications are shown in Figure 8. We observe that among the various normalization techniques, the tanh-norm and the logit with tanh-norm outperform the other techniques at all low and high FAR values. In particular, best results of 66.9% of VR at FAR=0.001 and 5.7% EER are achieved for the case of the security access application, and 95.6% of VR and 0.9% EER for the case of the smart home application, in both cases using the logit with tanh score normalization, which is the technique also used in the realistic scenario (Sect. 5.2). Table 2 shows the comparative results with the individual modalities.

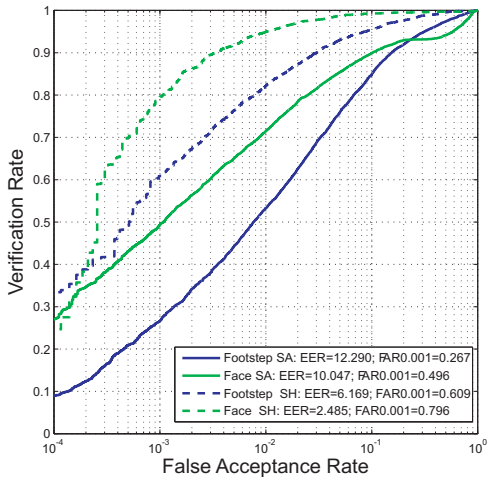
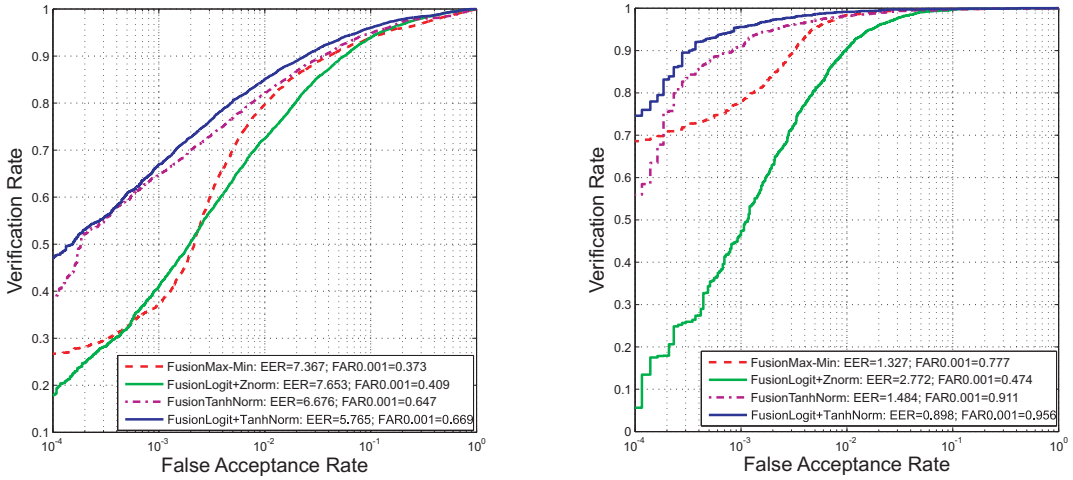


Figure 7. ROC curves for footsteps and face modalities for security access (SA) and smart homes (SH) applications.



(a) Fusion Security Access (SA) (b) Fusion Smart Homes (SH)

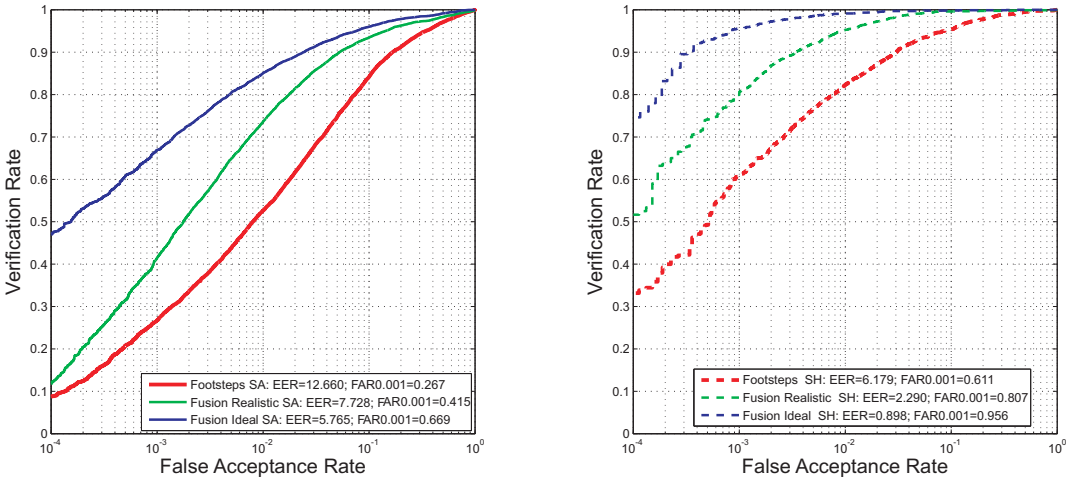
Figure 8. ROC curves for the fusion of footsteps and face in the ideal case.

5.2 Realistic Case

This section shows the performance results for the case of the realistic fusion, which is an adaptive fusion as described in Sect. 4.1. Figure 9 shows the ROC curves for the cases of the two applications comparing the performance of footsteps, the fusion in the realistic case and in the ideal case. As can be seen, the performance for the realistic fusion is not as good as for the ideal case, but shows the improvement achieved compared with having only the footstep signals. Also, as stated before, it is worth noting that around 24% of the available data is discarded in the ideal case, which would not be very sensible in a real scenario. In this case results of 7.7% EER and 41.5% VR are achieved for the SA application, and results of 2.3% EER and 80.7% VR for the SH application. Table 2 shows the comparative results with the individual modalities and the fusion in the ideal case.

6. CONCLUSIONS

This paper reports for the first time experimental results for the fusion of footstep and face biometric modes on an unsupervised and uncontrolled environment. Footstep signals have the benefit over other biometric modes that can be collected covertly, which is very convenient for the users, and also footstep signals are very difficult



(a) Fusion Security Access (SA) (b) Fusion Smart Homes (SH)

Figure 9. ROC curves for the fusion of footsteps and face in the realistic case compared to the ideal case and the individual footsteps.

	Security Access (SA)		Smart Home (SH)	
	VR	EER	VR	EER
Footstep	26.7	12.7	61.1	6.2
Face	49.6	10.0	79.6	2.5
Fusion Ideal	66.9	5.8	95.6	0.9
Fusion Realistic	41.5	7.7	80.7	2.3

Table 2. Comparative results of VR (in % for FAR=0.001) and EER (in %).

to me imitated. Fusion of footstep signals with other unobtrusive biometric modes such as face or gait can be carried out easily. In the experimental protocol, two different applications have been considered: a security access and smart homes scenarios. The fusion is carried out at the score level and following two architecture configurations: (*i*) an ideal case achieving results of 5.8% and 0.9% EER for each application respectively; and (*ii*) a more realistic case when there is not always a face image linked to a footstep signal. In this case an adaptive fusion is carried out obtaining results of 7.7% and 2.3% EER for each application respectively, which is a 39.4% and 62.9% relative improvement of EER compared to the footstep performance individually.

Acknowledgements

This work has been partially supported by projects Contexts (S2009/TIC-1485), Bio-Challenge (TEC2009-11186) and "Catedra UAM-Telefonica". Postdoctoral work of author R.V.R is supported by a Juan de la Cierva Fellowship from the Spanish MINECO.

REFERENCES

[1] Vera-Rodriguez, R., Lewis, R., Mason, J., and Evans, N., "Footstep recognition for a smart home environment," *International Journal of Smart Home. Special Issue on Future Generation Smart Space (FGSS)* **2**, 95–110 (2008).

[2] Tome, P., Fierrez, J., Alonso-Fernandez, F., and Ortega-Garcia, J., "Scenario-based score fusion for face recognition at a distance," in [*IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*], 67 – 73 (June 2010).

[3] Kale, A., Roychowdhury, A., and Chellappa, R., "Fusion of gait and face for human identification," in [*Acoustics, Speech, and Signal Processing, 2004. Proceedings. (ICASSP '04). IEEE International Conference on*], **5**, V – 901–4 vol.5 (may 2004).

[4] Geng, X., Smith-Miles, K., Wang, L., Li, M., and Wu, Q., "Context-aware fusion: A case study on fusion of gait and face for human identification in video," *Pattern Recognition* **43**(10), 3660 – 3673 (2010).

[5] Vera-Rodriguez, R., Mason, J., and Evans, N., "Automatic cross-biometric footstep database labelling using speaker recognition.," in [*Proceedings of the IAPR/IEEE International Conference on Biometrics (ICB)*], 503 – 512 (2009).

[6] Yun, J. S., Lee, S. H., Woo, W. T., and Ryu, J. H., "The User Identification System Using Walking Pattern over the ubiFloor," in [*Proceedings of International Conference on Control, Automation, and Systems*], 1046–1050 (2003).

[7] Middleton, L., Buss, A. A., Bazin, A. I., and Nixon, M. S., "A floor sensor system for gait recognition," in [*Proceedings of Fourth IEEE Workshop on Automatic Identification Advanced Technologies (AutoID'05)*], 171–176 (2005).

[8] Suutala, J. and Roning, J., "Methods for person identification on a pressure-sensitive floor: Experiments with multiple classifiers and reject option," *Information Fusion. Special Issue on Applications of Ensemble Methods* **9**(1), 21 – 40 (2008).

[9] Vera-Rodriguez, R., Mason, J., Fierrez, J., and Ortega-Garcia, J., "Analysis of spatial domain information for footstep recognition," *Computer Vision, IET* **5**, 380 –388 (november 2011).

[10] Vapnik, V. N., "Statistical learning theory," *Wiley. New York* (1998).

[11] Vera-Rodriguez, R., Mason, J., Fierrez, J., and Ortega-Garcia, J., "Analysis of time domain information for footstep recognition," in [*Proc. 6th International Symposium on Visual Computing (ISVC'2010)*], *Lecture Notes in Computer Science*, 489–498, Springer (2010).

[12] Wright, J., Yang, A. Y., Ganesh, A., Sastry, S. S., and Ma, Y., "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.* **31**, 210–227 (2 2009).

[13] Alonso-Fernandez, F., Fierrez, J., Ramos, D., and Gonzalez-Rodriguez, J., "Quality-based conditional processing in multi-biometrics: application to sensor interoperability," *IEEE Trans. Sys. Man Cyber. Part A* **40**, 1168–1179 (November 2010).

[14] Jain, A., Karthik, N., and Ross, A., "Score normalization in multimodal biometric systems," *Pattern recognition* **38**(12), 2270–2285 (2005).

[15] Poh, N., Kittler, J., Marcel, S., Matrouf, D., and Bonastre, J.-F., "Model and score adaptation for biometric systems: Coping with device interoperability and changing acquisition conditions," in [*Pattern Recognition (ICPR), 2010 20th International Conference on*], 1229 –1232 (aug. 2010).