2013 6th IAPR International Conference on Biometrics, ICB. IEEE 2013. 1 - 6

DOI: http://dx.doi.org/10.1109/ICB.2013.6613014

A Novel Approach of Gait Recognition Through Fusion with Footstep Information

Ruben Vera-Rodriguez[a], Julian Fierrez[a], John S.D. Mason[b], and Javier Ortega-Garcia[a]
[a]Biometric Recognition Group - ATVS, EPS, Universidad Autonoma de Madrid
Avda. Francisco Tomas y Valiente, 11 - Campus de Cantoblanco - 28049 Madrid, Spain
[b]Speech and Image Research Group, Swansea University, Singleton Park SA2 8PP, Swansea, UK
{ruben.vera, julian.fierrez, javier.ortega}@uam.es, j.s.d.mason@swansea.ac.uk

Abstract

This paper is focused on two biometric modes which are very linked together: gait and footstep biometrics. Footstep recognition is a relatively new biometric based on signals extracted from floor sensors, while gait has been more researched and it is based on video sequences of people walking. This paper reports a directly comparative assessment of both biometrics using the same database (SFootBD) and experimental protocols. A fusion of the two modes leads to an enhanced gait recognition performance, as the information from both modes comes from different capturing devices and is not very correlated. This fusion could find application in indoor scenarios where a gait recognition system is present, such as in security access (e.g. security gate at airports) or smart homes. Gait and footstep systems achieve results of 8.4% and 10.7% EER respectively, which can be significantly improved to 4.8% EER with their fusion at the score level into a walking biometric.

1. Introduction

Gait and footsteps are two biometric modes which are very linked together as they both extract discriminative information from the way people walk. In the biometric context, gait aims to discriminate persons using walking characteristics extracted from video recordings, while footstep recognition is based on signals captured from persons walking over an instrumented sensing area. An advantage of gait is that it offers potential for recognition at a distance or at low resolution in situations where other biometrics might not be possible [1]. However, some disadvantages are that gait can suffer from occlusions, differences in lighting conditions and background movements [2]. On the other hand, footstep is a more controlled biometric, but can be collected unobtrusively and is very robust to environmental conditions.

Gait has received far more attention in the literature than footsteps, perhaps for the ready availability of video cameras in different everyday situations in contrast to the dedicated pressure floor sensors used to capture footstep signals. In this paper gait and footsteps are considered as coming from a normal walking sequence. Thus, in this context footsteps and gait are inextricably linked. They are two modes sufficiently independent to hypothesize that they would be complementary in person classification and hence enhance biometric performance. It is interesting to note the parallel case of visual speech [3].

A preliminary fusion of gait and footstep signals was reported by Cattin in 2002 [4] achieving very good results of 1.6% EER, but for a very small database with only 16 people. Thus, this paper reports results of the first meaningful fusion between gait and footsteps as it is based on the largest footstep database to date, SFootBD [5]. A dataset of this database comprised of 7147 gait and footstep signals from 122 persons has been considered here. Also, this database was collected on an unsupervised and uncontrolled manner, i.e., factors providing variability in each biometric mode such as illumination or clothing for gait, footwear for footsteps or speed for both were not controlled, making this a very challenging problem and results achieved are realistic in terms of the breadth of conditions encompassed.

The fusion of gait and footstep modes is carried out at the score level following a product rule. The same database structure and protocols are followed for both biometrics enabling a direct performance comparison of the two biometrics for the first time. The gait recognition system developed is based on the appearance, using the silhouette of the persons walking to extract the discriminative information following two approaches: EGEI [6] and MPCA [7]. On the other hand, the footstep recognition system developed is based on spatio-temporal information from the pressure signals [5]. Individual results achieved for gait and footstep modes are 8.4% and 10.7% EER respectively. A very sig-

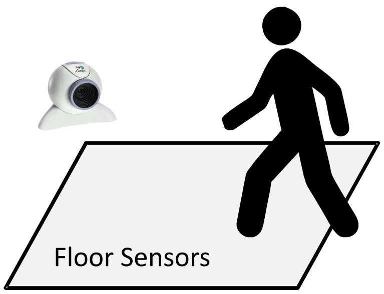


Figure 1. Arrangement of the gait and footstep capturing system.

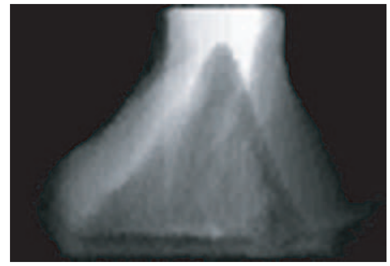


Figure 3. Example of Gait Energy Image (GEI) for SFootBD database.

nificant improvement of performance is obtained with their fusion, achieving an EER of 4.8%, which means a 42.7% of relative improvement compared to the best individual case.

The remainder of the paper is organized as follows. Section 2 describes the database and the signals considered. Section 3 describes the gait recognition system developed. Section 4 describes the footstep recognition system. Section 5 reports the experimental results achieved for the individual modes and their fusion; and finally conclusions are drawn in Section 6.

2. Database and Signals

The database considered for the experimental work presented in this paper is the SFootBD [5]. This is a multimodal database comprising gait, footstep, face and voice signals, with almost 10,000 signals per individual mode and more than 120 persons. This is the largest database of footstep signals to date. The gender distribution of the SFootBD database is 65% males and 35% females. Figure 1 shows a diagram of the arrangement of the capturing system. The person walks over an array of piezoelectric sensors, which trigger the recording of the other biometric modes. The database contains information for different types of footwear such as shoes, boots, trainers, flip-flops and even barefoot. The vast majority of the data was captured in an unsupervised mode, allowing to obtain data of persons walking naturally and similar to what could be captured in a real application.

The main characteristic of the footstep signals considered here is that contain information in both time and spatial domains, in contrast to previous works [2, 8, 9]. In this case, a high density array of piezoelectric sensors ($\sim$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 is large enough to collect a stride (right to left) footstep signal. The gait 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 gait video of the person walking from a side view. It is worth noting that the gait dataset considered from SFootBD contains much less information compared to other more standard gait databases, as the video camera only captures the lower part of the body and less than half of a gait cycle. Figure 2 shows an example sequence of gait images contained in the database. This is a constrained scenario for gait recognition, as the upper part of the human silhouette has been demonstrated to contain very discriminative information [10].

A dataset from the SFootBD database has been used in this paper selecting examples with a total correspondence in both footstep and gait modes. In this case a total of 7147 gait and footstep signals from 122 persons have been considered.

3. Gait Recognition System

During the last few years, many algorithms have been developed to extract the discriminative information for gait recognition, for both appearance-based and model-based main approaches [1]. For the work presented in this paper, only appearance-based feature approaches were considered due to their easier implementation and good results achieved in previous works. A review of the state-of-the-art was conducted selecting six feature approaches, which were implemented and tested with different conditions. These algorithms were: Gait Energy Image (GEI) [11], Enhanced Gait Energy Image (EGEI) [6], Multilinear Principal Component Analysis (MPCA) [7], Active Energy Image (AEI) [12], Gait Flow Image (GFI) [13] and Motion Silhouette Contour Template (MSCT) [13]. The feature approaches that obtained better recognition results for the gait signals considered in this work were EGEI and MPCA, which are described in more detail next.

The first feature approach considered, called Enhanced Gait Energy Image (EGEI) [6], is based on the popular method GEI [11]. For this, an averaged GEI image (see Figure 3) representing each training user class is used to construct a dynamic weight mask by variance analysis. This mask is applied to the original GEI images to obtain the EGEI images. Finally, this method uses a Gabor filter bank considering 5 scales and 8 orientations (i.e. 40 Gabor kernel functions) in order to emphasize the most discrimina-

Figure 2. Examples of a gait image sequence from SFootBD database after image segmentation.

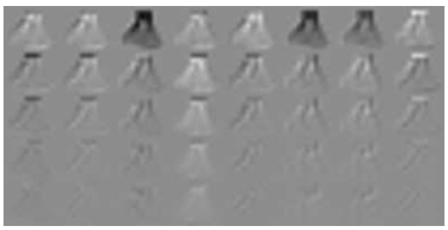


Figure 4. Example of Enhanced Gait Energy Image (EGEI) for SFootBD database.

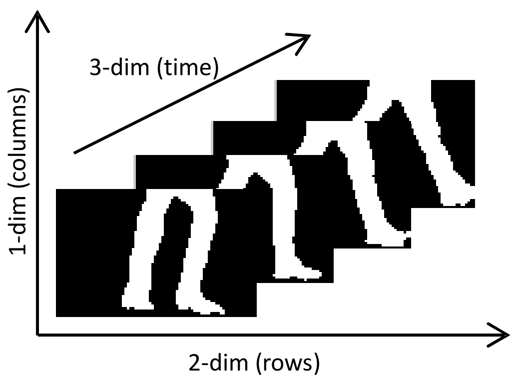


Figure 5. Example of Multilinear Principal Component Analysis (MPCA).

tive parts of the body image as shown in Figure 4. This technique is computationally more expensive than the GEI method, but allows to improve the results in cases of having much noisier environments. Data dimensionality was reduced using PCA plus LDA.

The second approach considered in this paper, called Multilinear Principal Component Analysis (MPCA) [7], is an extension of the popular algorithm PCA. As can be seen in Figure 5, the data is arranged in several dimensions to form a tensor. In fact, two tensors, one for the training data to which is applied MPCA, and another one for the test data to which is applied the MPCA transformation from the training set. In our case, four dimension tensors are used: two spatial dimensions of the images, a time dimension and another dimension for the different data examples. Once the tensor is ready, MPCA can drastically reduce the high dimensionality of the original data into lower dimension feature vectors (300 components were used here). LDA was also applied to further reduce the dimensionality before the classification stage.

4. Footstep Recognition System

For footstep recognition, spatio-temporal features have been extracted from the signals to carry out person recognition, similar to the work described in [5].

The time domain information of the footstep signals (called *BTime*) is extracted from the differential pressure of the sensors along the time axis without considering their spatial distribution. Figure 6(a) 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.

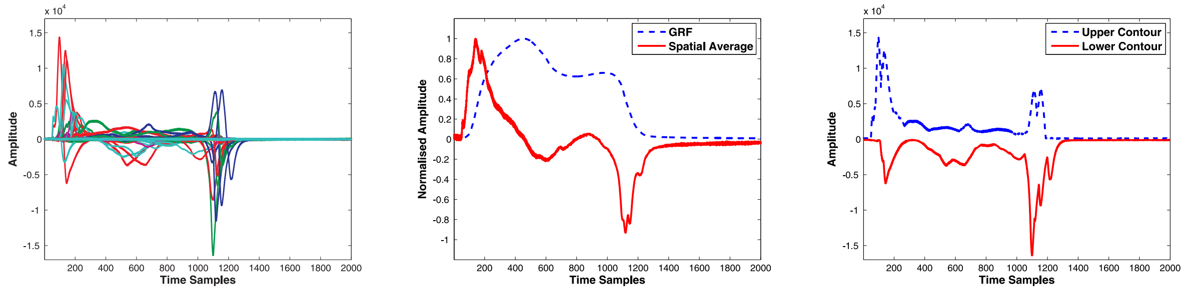
Different approaches can be carried out in order to minimize the effect of the spatial information and extract features of the signals in the time domain. The most popular feature used in the time domain is the ground reaction force (GRF) [2, 14, 15, 16, 17]. Figure 6(b) shows the GRF profile for the example signal considered here. In this case, as the piezoelectric sensors provide the differential pressure, the GRF is obtained by integrating each sensor signal across the time, and then a summation of the 88 single profiles is carried out to provide a global GRF.

Apart from the GRF, two other feature approaches are followed here, the first one is a simple average of the 88 sensors of the array to produce a single profile (spatial average). The other approach is named upper and lower contour of the signal and consists in the maxima and minima of the sensors respectively for each time sample, as can be seen in Figure 6(c).

A combination of these four profiles at the feature level is considered as the time domain information from the signals, following the work described in [5].

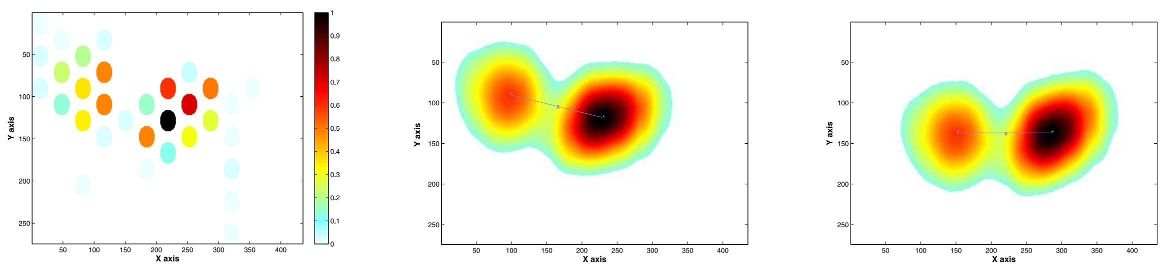
On the other hand, the spatial domain information extracted from the signals (*BSpace*) is based on the accumulated pressure of each piezoelectric sensor over a footstep, as in [5]. Figure 7(a) shows an example footstep signal with the accumulated pressure of each sensor for the X and Y axes. Alignment 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 7(b) shows the result image for the given example after applying 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 7(c), which is used to carry



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

Figure 6. Time domain (BTime) feature extraction: (a) Differential pressure directly from the 88 sensors, against time. (b) Normalised ground reaction force profile from (a), and normalized spatial average of the 88 sensors. (c) Upper and lower contour profiles from (a).



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

Figure 7. Spatial domain (BSpace) feature extraction: (a) Accumulated pressure (AP) of each sensor across one footstep. (b) Result after smoothing image from (a) with a Gaussian filter. (c) Resultant after alignment and rotation to a common centre of signals from (b).

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 was reduced using PCA for both time and spatial approaches.

5. Experimental Work

This section describes the experimental work carried out to analyze both gait and footstep recognition system in a comparative manner, as the same protocols are used for both biometric modes. Then a fusion of gait and footstep systems is carried out at the score level.

5.1. Experimental Protocol

The experimental protocol followed in this paper consists on a division of the data into training and test sets. In this case, the training data was comprised of 59 users with 10 data samples per user. The rest of the data was used for the test set having a total of 122 persons (including the 59 of the training set) with a variable number of data, from just a few to hundreds.

Regarding the classifier, a support vector machine (SVM) [18] was adopted with a radial basis function (RBF) as the kernel, due to very good performance in previous studies in this area [2, 19].

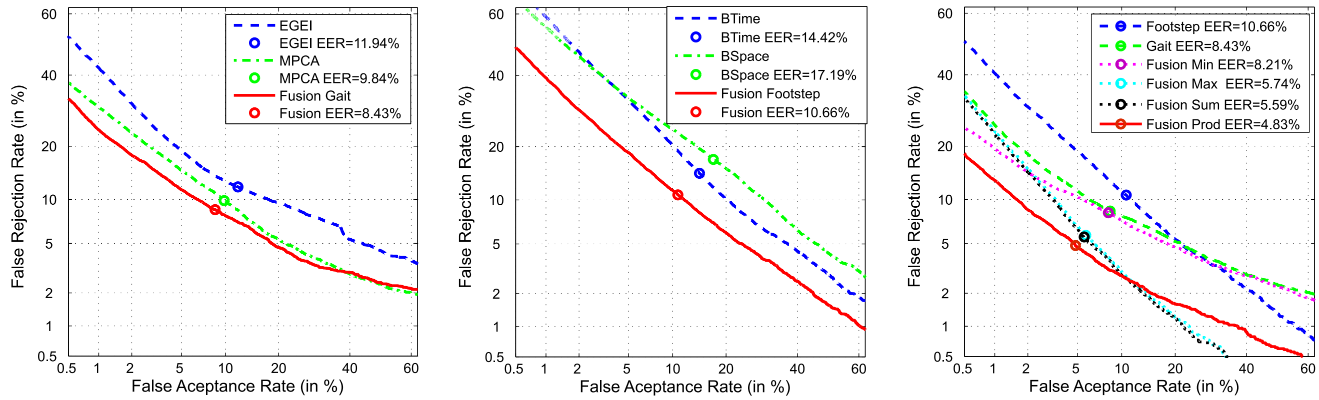
Recognition experiments are carried out in a verification mode, using detection error trade-off (DET) curves and equal error rates (EER) as the measure of performance.

5.2. Gait Results

Figure 8(a) shows the DET curves for the gait feature approaches EGEI and MPCA described in Section 3. As can be seen, results of 11.9% and 9.8% EER are achieved for the cases of EGEI and MPCA respectively. These are quite acceptable results having in mind that only the lower part of the body is available in the images and also for less than half of a gait cycle, which is a fourth of the information present in other more standard gait databases.

A fusion of these two approaches was carried out at the score level following a simple product rule after the scores were normalized between 0 and 1. The fusion improved the results to 8.43% EER, which is a 14.3% of relative improvement regarding the best individual case (MPCA). A fusion of these approaches with other model-based approaches would be likely to further improve the results.

It is worth comparing these results with the findings of Veres *et. al.* [10], that carried out an analysis of what image information is more discriminative on silhouette based gait approaches, concluding that the upper part of the body which corresponds to the most static component contains



(a) Gait Approaches (b) Footstep Approaches (c) Gait and Footstep Fusion

Figure 8. (a) DET curves for Gait approaches (EGEI and MPCA) and their fusion. (b) DET curves for Footstep approaches (BTime and BSpace) and their fusion. (c) DET curves for the score-level fusion of Gait and Footstep modes.

the most discriminative information. In the case considered here, only the lower part of the body is present and good recognition results are achieved, so we can conclude that there is also discriminative information in this part, which could obviously be improved if the upper part of the body was also present.

5.3. Footstep Results

Figure 8(b) shows the DET curves for the spatio-temporal footstep feature approaches described in Section 4. As can be seen, results of 14.4% and 17.2% EER are achieved for the cases of BTime and BSpace respectively, which are a bit worse compared to the results achieved for the gait mode. A similar fusion was carried out for the footstep approaches at the score level following a product rule. In this case the fusion improved the results to 10.66% of EER, which is a 26.1% of relative improvement. This improvement of performance is larger than in the case of gait, due probably to the fact that the spatio-temporal information is not as correlated as the two features approaches used for gait.

5.4. Fusion of Gait and Footstep Systems

This section reports the experimental results obtained for the fusion of the two biometric systems developed for gait and footsteps, which leads to a walking biometric. In this paper gait and footsteps are considered as coming from a normal walking sequence. Then, in this context gait and footsteps are inextricably linked together. They are two modes sufficiently independent to hypothesize that they would be complementary in person classification and hence enhance biometric performance.

The fusion of footsteps and gait has not received much attention in the literature. In 2002, Cattin [4] presented experimental results in this area, fusing data acquired from 3 tiles of 4 piezoelectric sensors each for footsteps and a video camera for gait. A database of 480 footstep signals was collected from 16 persons walking barefoot. A fusion at the score level was carried out for five feature approaches, four for gait and one for footsteps, giving this way more importance to the gait mode. A final result of 1.6% EER was achieved for this fusion.

In this case, the fusion of gait and footsteps is carried out at the score level following four different fusion rules such as max, min, sum and product. The matching scores coming from the two gait and footstep systems were previously normalized in the range between 0 and 1. Figure 8(c) shows the performance for the cases of footsteps (as shown in Figure 8(b)), gait (as shown in Figure 8(a)), and the four score-level fusions of the two of them. As can be seen, the fusion using the product of the scores of the individual modes outperforms the other three fusion rules. In this case, the fusion achieves a very significant improvement of performance, going from 10.7% and 8.43% of EER for footsteps and gait respectively to 4.83% of EER, which means a relative improvement of 42.7% compared to the best individual case corresponding to the gait system.

6. Conclusions and Future Work

This paper describes a new approach for gait recognition based on the fusion of traditional gait information of persons walking extracted from video cameras with footstep information extracted from pressure floor sensors. Both gait and footstep modes are assessed using the same database and protocols, enabling direct performance comparison of the two systems.

Two feature approaches have been followed for gait, EGEI and MPCA, both based on the silhouette information, achieving results of 8.43% of EER for their fusion. This is obtained for a gait database showing only the lower part of the body and less than half of a gait cycle available, so lower error rates are expected for more ideal gait images.

Regarding the footstep mode, a fusion of spatio-temporal information is carried out achieving results of 10.7% EER, which are not as good as for the gait system. A final fusion at the score level was performed for the two modes achieving a very significant relative improvement of performance of 42.7% compared to the best individual system, with an EER of 4.83%.

These interesting results allow us to think of some lines for future work, such as new feature extraction approaches for both gait and footsteps modes, different score normalization and fusion strategies to further improve the results, for example giving different weights to the systems, or weights based on the quality of the signals. Also, giving results for different quantities of training data (less or more data) per person would be of interest to analyze the expected performance of the system for different applications such as security access or smart homes for example.

7. Acknowledgements

R. Vera-Rodriguez is supported by a Juan de la Cierva Fellowship from the Spanish MINECO. This work has been partially supported by projects Bio-Shield (TEC2012-34881), Contexts (S2009/TIC-1485), TeraSense (CSD2008-00068) and "Cátedra UAM-Telefónica".

References

[1] M. Nixon and J. Carter. Automatic Recognition by Gait. *Proceedings of the IEEE*, 94(11):2013–2024, Nov. 2006.

[2] J. Suutala and J. Roning. 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.

[3] W. Karam, H. Bredin, H. Greige, G. Chollet, and C. Mokbel. Talking-Face Identity Verification, Audiovisual Forgery, and Robustness Issues. *EURASIP Journal on Advances in Signal Processing*, 2009.

[4] C. Cattin. *Biometric Authentication System Using Human Gait.* PhD thesis, ETH, Switzerland, 2002.

[5] R. Vera-Rodriguez, J. S. Mason, J. Fierrez, and J. Ortega-Garcia. Comparative Analysis and Fusion of Spatio-Temporal Information for Footstep Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012.

[6] X. Yang, Y. Zhou, T. Zhang, G. Shu, and J. Yang. Gait Recognition Based on Dynamic Region Analysis. *Signal Processing*, 88(9):2350–2356, 2008.

[7] H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos. Multilinear Principal Component Analysis of Tensor Objects for Recognition. In *Proc. of the 18th International Conference on Pattern Recognition*, volume 2, pages 776–779, 2006.

[8] J. S. Yun, S. H. Lee, W. T. Woo, and J. H. Ryu. The User Identification System Using Walking Pattern over the ubi-Floor. In *Proceedings of International Conference on Control, Automation, and Systems*, pages 1046–1050, 2003.

[9] L. Middleton, A. A. Buss, A. I. Bazin, and M. S. Nixon. A Floor Sensor System for Gait Recognition. In *Proceedings of Fourth IEEE Workshop on Automatic Identification Advanced Technologies (AutoID'05)*, pages 171–176, 2005.

[10] G. Veres, L. Gordon, J. Carter, and M. Nixon. What Image Information Is Important in Silhouette-Based Gait Recognition? In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, pages 776–782, Los Alamitos, CA, USA, 2004. IEEE Computer Society.

[11] J. Han and B. Bhanu. Individual Recognition Using Gait Energy Image. *IEEE Trans. Pattern Anal. Mach. Intell.*, 28(2):316–322, Feb. 2006.

[12] E. Zhang, Y. Zhao, and W. Xiong. Active Energy Image Plus 2DLPP for Gait Recognition. *Signal Processing*, 90(7):2295–2302, 2010.

[13] G. Trivino, A. Alvarez-Alvarez, and G. Bailador. Application of the Computational Theory of Perceptions to Human Gait Pattern Recognition. *Pattern Recognition*, 43(7):2572–2581, July 2010.

[14] M. D. Addlesee, A. Jones, F. Livesey, and F. Samaria. The ORL Active Floor. *IEEE Personal Communications*, 4:235–241, 1997.

[15] R. J. Orr and G. D. Abowd. The Smart Floor: A Mechanism for Natural User Identification and Tracking. In *Proceedings of Conference on Human Factors in Computing Systems*, 2000.

[16] Y. Gao, M. J. Brennan, B. R. Mace, and J. M. Muggleton. Person Recognition by Measuring the Ground Reaction Force due to a Footstep. In *Proceedings of 9th International Conference on Recent Advances in Structural Dynamics*, 2006.

[17] J. Suutala and J. Roning. Combining Classifiers with Different Footstep Feature Sets and Multiple Samples for Person Identification. In *Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, volume 5, pages 357–360, 2005.

[18] V. N. Vapnik. Statistical Learning Theory. *Wiley. New York*, 1998.

[19] R. Vera-Rodriguez, J. Mason, J. Fierrez, and J. Ortega-Garcia. Analysis of Spatial Domain Information for Footstep Recognition. *IET Computer Vision*, 5(6):380 –388, november 2011.