A Large Scale Footstep Database for Biometric Studies Created using Cross-Biometrics for Labelling

Ruben Vera-Rodriguez, Richard P. Lewis, John S.D. Mason
Speech and Image Research Group
Swansea University
Singleton Park, Swansea, SA2 8PP, UK
(r.vera-rodriguez.405831, r.p.lewis, j.s.d.mason)@swansea.ac.uk

Nicholas W.D. Evans
Institut Eurécom
2229 route des Crêtes
06904 Sophia-Antipolis, France
nicholas.evans@eurecom.fr

***Abstract*—This paper describes a semi-automatic system to capture and label a reasonable size biometric database. In our case, the biometric to be assessed are footstep signals, but the system could be extendable to other biometrics. Extra biometric data such as the voice and video recordings of the face and the gait are used to assist the database labelling to minimise the error. Thus, audio identifier recordings are used to automatically label the database with a speaker recognition system achieving results of 0.15% of equal error rate (EER) of person verification using Gaussian mixture models (GMM). Also, a footstep detector system has been developed to reduce the presence of invalid signals from the database having a percentage of less than 1% of correct footsteps miss-classified using features from the ground reaction force (GRF) and using a support vector machine (SVM) classifier. To date, more than 20,000 footstep signals have been collected from more than 100 people, which is well beyond previously reported databases. The database is collected in different sessions which will allow us to study how different factors such as footwear, the person carrying a load or different walking speeds affect the recognition of persons using their footsteps.**

***Index Terms*—Automatic database labelling. Footstep recognition. Pattern recognition.**

I. Introduction

The task of assessing signals for their potential as a biometric invariably involves databases with a large number of labelled examples, coming from many people. For mature biometrics such as fingerprints, face and speech, then such databases exist, containing many thousands of entries. For lesser researched biometrics that might nonetheless have potential, then new larger databases are needed in order to assess accuracy and related practical aspects. Footstep signals as a biometric fall into this category: initial investigations [1–6] indicate potential but with low levels of statistical confidence and with many practical questions remaining unanswered.

As the size of the database grows, then so does the need for automation. In this paper we describe such automation in the context of creating a very large footstep database. For obvious practical reasons, signal captured should be unsupervised and the subsequent labelling of the signals to the appropriate person should encompass a high level of automation and easy cross checking. Here we use speech and video signals to assist in the labelling process, as described below. Accurate labelling is essential. If, for instance, a signal is incorrectly labelled as coming from person X, i.e. the signal belongs to another person, Y, then this will corrupt the experimental assessment whenever the signal is compared with signals from person X or person Y. So correct labelling in this primary sense is manifestly important. Secondary labelling factors, such as the gender, the age or even the date when the biometric was collected can also be important.

In practice, the likelihood of such errors increases with the size of the database and is greatly influenced by the strategies and tools used to create it. There is a clear dichotomy: the larger the database, the greater the chances of mislabelling. However, the larger the database, the greater the levels of confidence that can be ascribed to the experimental results. Doddington's rule of 30 [7] is particulary sobering when conducting biometric experiments on small databases.

This paper describes how these factors have been addressed when creating the world's first reasonably sized footstep biometric database. In particular, we describe facilities designed and implemented to capture footsteps in order to assess the potential of these signals as a biometric. The main database design criteria include: 100 persons or more, 20 or more footsteps per person across several sessions and varying footwear. At the time of writing, the total number of signals captured exceeds 20,000, a number that makes manual labelling wholly impracticable. Due to the fact that footsteps are a relatively new biometric, there are no prerequisites on the representativeness of the population other than to be large. In this scenario the population comprises primarily of university students.

The footstep capture system not only collects footstep signals, but also an audio recording of a spoken identifier and video recordings of the face and gait. Labelling is performed via a three-level protocol. The first level is a footstep detection algorithm to filter invalid signals from the database, the second uses a speaker recognition system to automatically label the database making use of the audio recordings, and the third level is comprised of a web-based application to carry out rapid manual validation of the labels given by the speaker recognition system to minimise possible errors. Whilst the

emphasis of this paper is toward footstep labelling, the primary contribution of this paper relates to a novel cross-biometric labelling strategy of potential to all biometric collection efforts.

II. Footstep Data Capture System

The system captures two different types of data, the biometric signals to be assessed and auxiliary signals to assist the labelling and the validation of the database.

In our previously published work [8–10], footstep recognition was carried out with a database comprised of more than 3500 footstep signals from 55 people from two piezoelectric sensors, the largest footstep database to date and freely available at [11]. Here, building on this previous work, we have developed a new capture system comprised of two sensor mats of 45 cm x 35 cm each containing 88 piezoelectric sensors to capture two consecutive footstep signals. The sensors provide a differential voltage output according to pressure upon them. Figure 1 illustrates a diagram of the capture system. Twelve integrated data acquisition boards are used in total, six per mat, each board captures signals from 16 sensors.

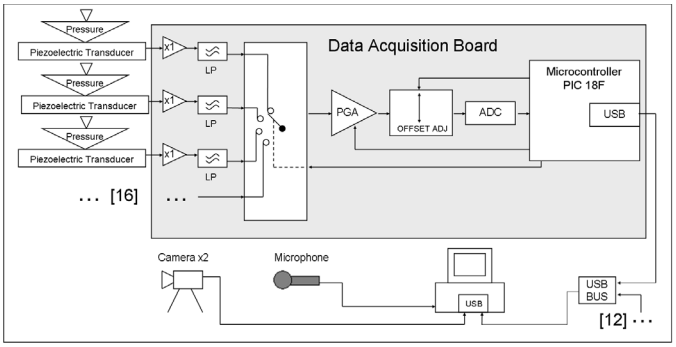

Fig. 1. Screenshot of the footstep capture system software.

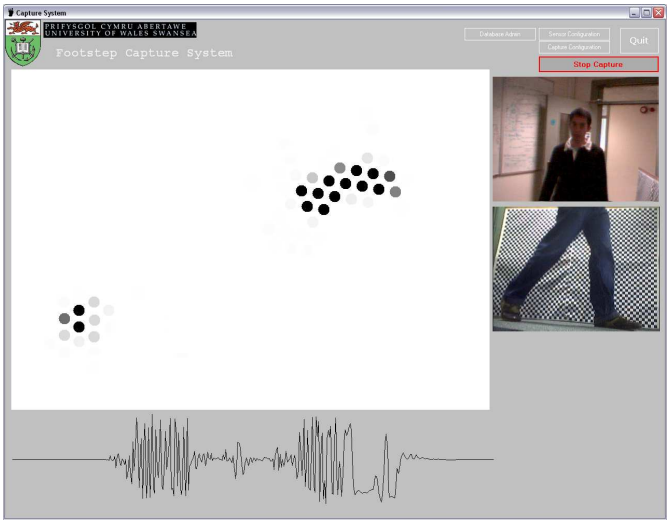

Fig. 2. Screenshot of the footstep capture system software.

A microphone situated a few steps ahead of the sensing area captures a 4-digit spoken identifier, whilst ensuring no disturbance in the natural walking process. Two video cameras capture images of the face and the stride. Figure 2 shows a screenshot of the footstep capture system user interface. A distribution of the sensors activated in the stride footstep is illustrated in the middle of the figure. The bottom part shows the microphone output, while the images on the right show frames from the videos that are captured during the footstep data collection, the top one shows the face and the bottom one the gait. This information is displayed in real-time.

III. Footstep Database Design and Data Labelling

The design of the database is a key issue to take into account, as all the experimental work to be carried out depends upon it. To date we have collected footstep data from more than 100 people and more than 20,000 signals. Data is collected in different sessions to make more realistic models of each client. Different conditions are encouraged including different footwear, barefoot, extra weight by carrying a load and different walking speeds. These conditions can be deduced from the auxiliary signals.

Usually, a biometric database is collected to be representative for a specific application. Due to the fact that footsteps are a relatively new biometric, there are no prerequisites on the representativeness of the population other than to be large. The database being collected is primarily comprised of university students. Once validated, evaluation protocols are designed to represent applications such as the usage of footstep biometric in smart homes or as a generic footstep verification system. Due to our past experience in speaker verification, in our previously published work [9], [10], evaluation protocols were designed following the format of the international NIST speaker recognition evaluations (SRE) [12], and the database was divided into independent development and evaluation datasets following best practice.

In the related work only a few footstep recognition systems have been developed, using different sensors, features and classifiers to research footsteps as a biometric [1–6]. Results achieved are promising and give an idea of the potential of footstep signals; however, they relate to small databases in number of persons and footsteps and this is a limitation of the work to date. Furthermore, the task of automating database labelling has not been taken into account in the related work due to database sizes that could be labelled manually.

A three-layer protocol has been developed to assist the database labelling, since manual labelling is impracticable in this case. The first layer is made up of an algorithm to remove invalid signals from the database. A total of 3134 footstep signals were manually examined to derive the ground truth and then the whole database was processed. Some examples of bad signals present in the database can be seen

in Figure 3 such as noisy signals Figure 3(a), partial footsteps in time Figure 3(b) or footsteps partially inside the sensor mat Figure 3(c). This algorithm is described in more detail in Section 4.A.

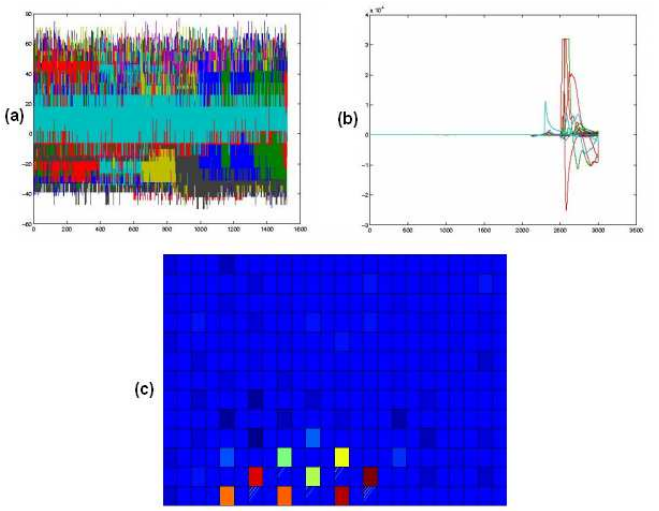

Fig. 3. Examples of invalid footstep signals. (a) Noisy signal. (b) Partial footstep in time. (c) Partial footstep in space.

The second layer uses a speaker recognition system to automatically label the database. To carry out this procedure it is necessary to manually label a few footsteps per person to train the speaker recognition system. This is described in more detail in Section 4.B.

The third layer is a web-based application designed to manually validate the labels assigned by the speaker recognition system and correct possible errors. This application is necessary to label extra information such as the type of footwear, if the person is carrying a load, or the walking speed, etc. Figure 4 shows a screenshot of the web application with the list of the footsteps in the database. Each row corresponding to the column 'Label' shows the identity of the person, the columns 'Decision Right' and 'Decision Left' show the decision taken by the footstep detector, as described in Section 4.A. If the system decides the footstep is good, a green tick is assigned, otherwise a red cross is set. All footsteps labelled as valid are inserted in the 'Correct Footsteps' list, which is then made available to the web application to manually label footsteps. The columns 'Valid Right' and 'Valid Left' show the decision of good or bad footstep taken manually which serves as the ground truth for the footstep detector algorithm. The column 'Extra weight' shows if the person is carrying a load or not, and the left columns show the type of footwear, the speed and finally if the footstep has been labelled by a user of the database or by the speaker recognition system.

Figure 5 shows a screenshot of the web application with

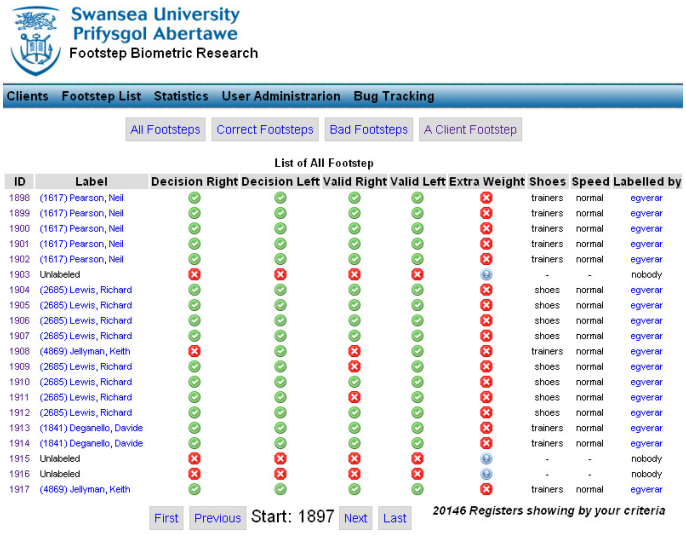

Fig. 4. Screenshot of the web-based application that shows the list of footstep signals in the database.

an example of a footstep signal. The video signals allow various checks including the persons identity and auxiliary information such as if they are carrying a load, the type of footwear and the speed of walking. The output of the pressure sensors can be seen at the bottom of Figure 5. The boxes 'Data Right' and 'Data Left' show the pressure exerted over the sensors against the time, and are useful to check that the right and left footsteps are not cut in time. The boxes 'GRF Right' and 'GRF Left' show the individual GRF (ground reaction force) accumulated for each sensor, and are useful to check that the whole foot is contained within the sensor mat.

IV. Experimental Work

This section first describes the experimental work carried out to filter invalid signals from the database and, second, automatically label the database using speaker recognition.

A. Algorithm to filter invalid footstep signals from the database

This section describes the algorithm developed to detect the valid footsteps to facilitate the manual labelling of the database. 3134 footstep signals were manually labelled as correct/bad signals and were used in the experiments to train the system.

1) Features: For feature extraction, two different feature sets have been used, namely the global GRF of the footstep and the accumulated individual GRF. If we call $s_i(t)$ the output of the piezoelectric sensors, where t = 1.....T_{max} and i = 1.....88, then the global GRF (GRF_T) and the accumulative individual GRF (GRF_{ind}) are determined by equation (1) and (2) respectively:

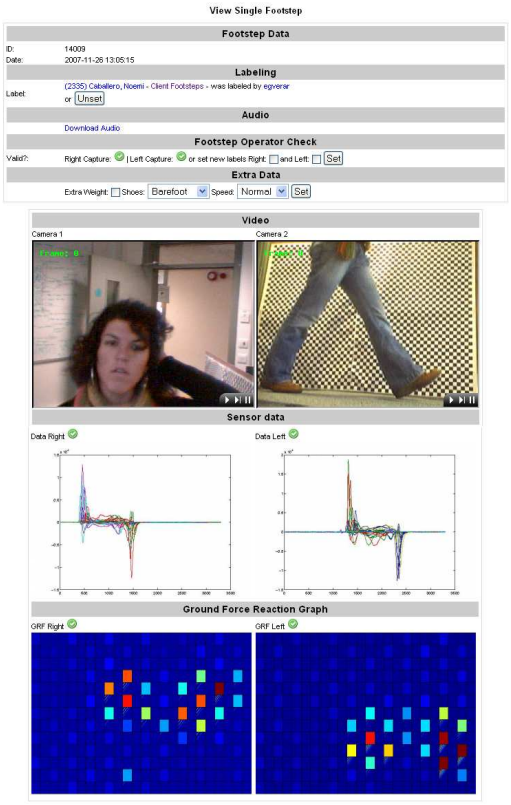

Fig. 5. Screenshot of the web-based application that shows a single footstep signal with its videos and audio recordings.

$$\text{GRF}_T(t) = \sum_{i=1}^{88}(s_i(t) + s_i(t-1)) \quad (1)$$

$$\text{GRF}_{ind}(i) = \sum_{t=1}^{T_{max}}(s_i(t) + s_i(t-1)) \quad (2)$$

An example of these two different feature approaches can be seen in Figure 6. The global GRF (Figure 6(a)) is a profile which shows the total pressure across the 88 sensors against the time as per equation (1). To compare all the footstep signals, the profiles were aligned in time, and the first 1400 samples were extracted. The global GRF was used in related work [1], [2], [6], [10] for footstep recognition. The second approach is the accumulated individual GRF (Figure 6(b)) where a single value per sensor expresses the accumulated pressure exerted over it during the time the footstep lasts as per equation (2).

Due to the high dimensionality of both feature approaches, principal component analysis (PCA) [13] was used to distil the information content.

2) Results: Both approaches were assessed with support vector machine (SVM) classifier using a radial basis function (RBF) as a Kernel [14], [15] used previously in [8–10]. Results are presented with detection error tradeoff (DET) curves in Figure 7. Three different cases are considered, the first one using only the global GRF after PCA, the second case using only the accumulate individual GRF after PCA, and the third one combining the other two cases.

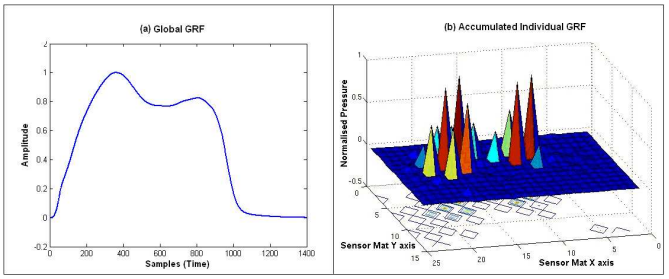

Fig. 6. Features approaches used to detect correct footsteps. (a) Global GRF. (b) Accumulated individual GRF.

The database used to train the system was comprised of 3134 manually labelled footsteps (1567 right and 1567 left), of which 1869 where labelled as correct footsteps and 1265 as bad footsteps. DET curves results for the three different cases are shown in Figure 7. The equal error rate (EER) for the first case using the global GRF after PCA was 9.9%, the EER for the second case using the accumulated individual GRF after PCA was 17.3%, and the EER for the third case as a combination of the other two cases was 9.4%, the best of them, thus showing the benefit, albeit small, of combining the two feature sets.

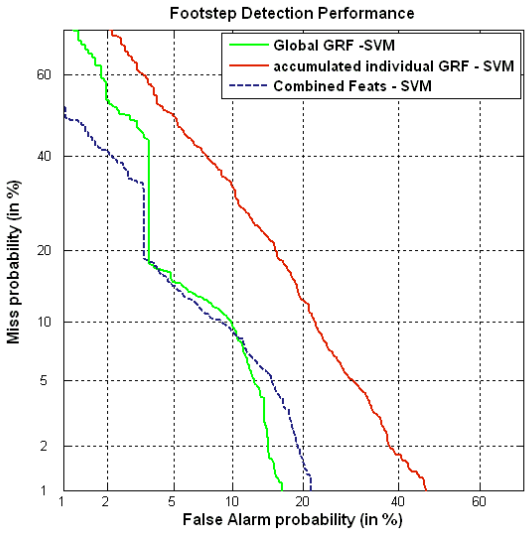

Fig. 7. DET curves result of the classification of correct/bad footsteps.

Figure 8 shows the score distributions for the above combination, the best of the three. The two distribution are quite far apart, but there is a middle part with overlap of the two areas. This is due the manual labelling and for some footstep signals the band between correct and not correct is very narrow.

B. Automatic labelling using speaker recognition system

As stated above a speaker recognition system has been used to automatically label the footstep signals in the database. To train the system, a set of 1950 footstep

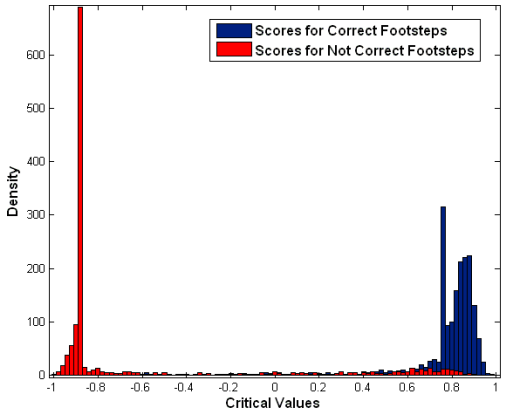

Fig. 8. Histograms of the scores for the classification of footsteps combining the two feature sets, the global GRF and the accumulated individual GRF.

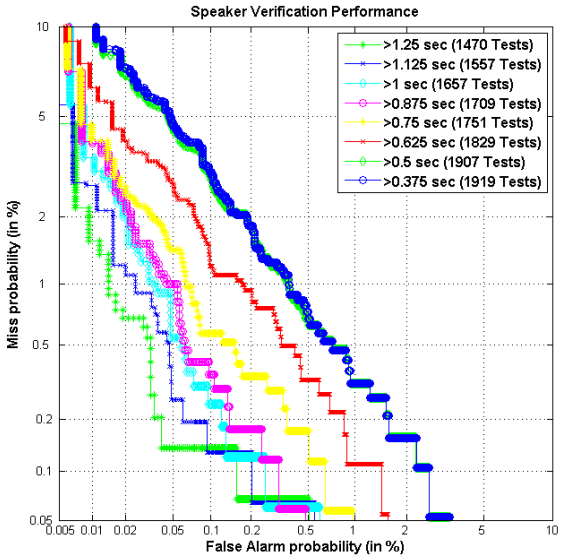

Fig. 9. DET curvess result of the person verification using the audio identifier.

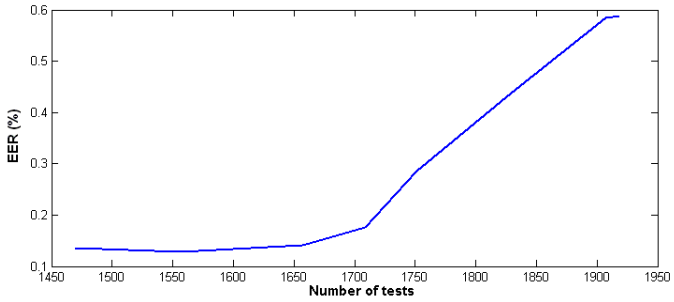

Fig. 10. EER against number of tests for the speaker verification system.

signals from 50 people were labelled using the web-based application described above. Each person in the database has got a different audio identifier comprised of four spoken digits.

For feature extraction, the audio recordings were first processed to extract the useful speech information by setting an energy threshold. A Gaussian mixture model (GMM) was used to assess the person verification as reported in [16]. The length of the speech extracted from the audio files was used as a parameter to carry out several experiments. Thus, the GMM classifier was first tested on signals with more than 0.375 seconds of speech duration, 1919 tests in this case. Then, it was tested on signals with more than 0.5 seconds (1907 tests), and so on with steps of 0.125 seconds until signals with more than 1.25 seconds. DET curves of these experiments are shown in Figure 9. As can be seen, the best result corresponds with the case of using only the speech signals with more than 1.25 seconds of duration (1470 tests). An EER of around 0.15% is achieved in this case which makes the speaker recognition system an excellent method to carry out the automatic labelling of a database by just using a four digits identifier. Figure 10 shows the EER result compared to the number of tests carried out with the speaker verification system. As the number of tests grows, i.e. the speech signal is shorter in size, the EER increases. This process can be combined with human checks of examples near the threshold.

V. Conclusions

This paper presents a semi-automatic system to capture and label a reasonable size biometric database. This system could be extendable to different biometric databases. In our particular case, the biometric to be assessed are footstep signals, and extra biometric data such as the voice and video recordings of the face and the gait are used to assist the database labelling to minimise the error. To date, more than 20,000 footstep signals have been collected from more than 100 people, which is well beyond previously reported databases. The database is being collected in different sessions which will allow us to study how different factors such as footwear, the person carrying a load or different speed affect the recognition performance. The capture and the database verification philosophy could be readily adapted to other biometric signals, dynamic or static.

References

[1] Addlesee, M.D., Jones, A., Livesey, F., Samaria, F.: The ORL Active Floor. IEEE Personal Communications. **4** (1997) 235–241
[2] Orr, R.J., Abowd, G.D.: The Smart Floor: A Mechanism for Natural User Identification and Tracking. In: Proceedings of Conference on Human Factors in Computing Systems. (2000)
[3] Cattin, C.: Biometric Authentication System Using Human Gait. Swiss Federal Institute of Technology, Zurich. PhD Thesis. (2002)
[4] Yun, J.S., Lee, S.H., Woo, W.T., Ryu, J.H.: The User Identification System Using Walking Pattern over the ubiFloor. In: Proceedings of International Conference on Control, Automation, and Systems. (2003) 1046–1050
[5] Middleton, L., Buss, A.A., Bazin, A.I., Nixon, M.S.: A floor sensor system for gait recognition. In: Proceedings of Fourth IEEE Workshop on Automatic Identification Advanced Technologies (AutoID'05). (2005) 171–176
[6] Suutala, J., Roning, J.: Methods for person identification on a pressure-sensitive floor: Experiments with multiple classifiers and reject option. Information Fusion **9** (2008) 21–40
[7] Doddington, G.R., Przybocki, M.A., Martin, A.F., Reynolds, D.A.: The NIST speaker recognition evaluation: Overview methodology, systems, results, perspective. Speech Communication. (2000), **31** 225–254

[8] Vera-Rodriguez, R., Evans, N.W.D., Lewis, R.P., Fauve, B., Mason, J.S.D.: An experimental study on the feasibility of footsteps as a biometric. In: Proceedings of 15th European Signal Processing Conference (EUSIPCO'07), Poznan, Poland (2007) 748–752

[9] Vera-Rodriguez, R., Lewis, R.P., Evans, N.W.D., Mason, J.S.D.: Optimisation of Geometric and Holistic Feature Extraction Approaches for a Footstep Biometric Verification System. In: International Summer School for Advanced Studies on Biometrics for Secure Authentication, Alghero, Italy (2007)

[10] Vera-Rodriguez, R., Lewis, R.P., Mason, J.S.D., Evans, N.W.D.: Footstep Recognition for a Smart Home Environment. To appear in International Journal of Smart Home. Special Issue on Future Generation Smart Space (FGSS). (2008)

[11] Footstep recognition evaluation at Swansea Universtity available at http://eeswan.swan.ac.uk

[12] NIST, Speaker Recognition Evaluation website, *http://www.nist.gov/speech/tests/spk/index.htm*.

[13] Jolliffe, I.T.: Principal Component Analysis. Springer Series in Statistics. (2002)

[14] Vapnik, V.N.: Statistical Learning Theory. Wiley, New York. (1998)

[15] Burges, C.J.C.: A tutorial on support vector machines for pattern recognition. Data Mining and Knowl. Discov. **2**, no.2 (1998) 1–47

[16] Fauve, B., Bredin, H., Karam, W., Verdet, F., Mayoue, A., Chollet, G., Hennebert, J., Lewis, R., Mason, J., Mokbel, C., Petrovska, D.: Some Results From The Biosecure Talking Face Evaluation Campaign. In: Proceedings of ICASSP. (2008)