



Repositorio Institucional de la Universidad Autónoma de Madrid

<https://repositorio.uam.es>

Esta es la **versión de autor** del artículo publicado en:

This is an **author produced version** of a paper published in:

IEEE Transactions on Pattern Analysis and Machine Intelligence, 30 January
(2018)

DOI: <http://doi.org/10.1109/TPAMI.2018.2799847>

Copyright: © 2018 IEEE

El acceso a la versión del editor puede requerir la suscripción del recurso
Access to the published version may require subscription

Analysis of Spatio-temporal Representations for Robust Footstep Recognition with Deep Residual Neural Networks

Omar Costilla-Reyes, Ruben Vera-Rodriguez, Patricia Scully, *Member, IEEE* and Krikor B Ozanyan, *Senior Member, IEEE*

Abstract—Human footsteps can provide a unique behavioural pattern for robust biometric systems. We propose spatio-temporal footstep representations from floor-only sensor data in advanced computational models for automatic biometric verification. Our models deliver an artificial intelligence capable of effectively differentiating the fine-grained variability of footsteps between legitimate users (clients) and impostor users of the biometric system. The methodology is validated in the largest to date footstep database, containing nearly 20,000 footstep signals from more than 120 users. The database is organized by considering a large cohort of impostors and a small set of clients to verify the reliability of biometric systems. We provide experimental results in 3 critical data-driven security scenarios, according to the amount of footstep data made available for model training: at airports security checkpoints (smallest training set), workspace environments (medium training set) and home environments (largest training set). We report state-of-the-art footstep recognition rates with an optimal equal false acceptance and false rejection rate of 0.7% (equal error rate), an improvement ratio of 371% from previous state-of-the-art. We perform a feature analysis of deep residual neural networks showing effective clustering of clients footstep data and provide insights of the feature learning process.

Index Terms—biometric system, verification system, deep learning, footstep recognition, floor sensor system



1 INTRODUCTION

1.1 Background

SECURITY is an inherent human need to protect assets from a threat. Traditionally, security systems have been based on passwords or security access cards. Biometric recognition deals with the design of security systems for automatic identification or verification [1] of a human subject (client) based on physical and behavioural characteristics. Physical biometric traits include, among others, fingerprints, facial features and the iris. On the other hand, behavioural biometrics, such as gait recognition, are intended to capture the unique signatures delivered by client's natural behavioural patterns. This approach is effective since the complexity in reproducing such patterns by an impostor (intruder) is quite high. Biometric recognition by gait is based on the study of human locomotion to obtain a unique biometric signature of a client. The wide range of biological factors influencing a gait signature has prompted up to date studies of gait for healthcare applications [2], but the

biometric security domain has recently drawn attention as well. More than 24 unique factors have been shown to affect human gait [3], resulting in a singular gait pattern for every individual. A biometric system based on gait requires users to exert minimum effort for appraisal. Furthermore, a gait biometric system can be deployed in a substantial number of access scenarios, ranging from airport entry checkpoints and entry to buildings to a home-based security system.

The advantage of gait as a biometric modality is that it allows natural gait signals to be obtained unobtrusively from a distance, which are difficult to forge by an impostor. However, gait analysis as a biometric modality has raised privacy concerns since it allows footstep signals to be acquired without the client's consent or knowledge. Further difficulties in the acquisition of reliable gait biometric signatures include variations in the client's footwear, their weight or emotional state, among other factors. Research on gait as a biometric modality has focused on gait analysis from video streams [4]. This approach has the disadvantage of being highly vulnerable to noisy environmental conditions, such as illumination and cross-view [5]. An effective alternative to video stream data is biometric identification and verification from floor sensor systems [6], [7], [8], [9], [10], [11], [12], [13]. Footstep recognition uses the ground reaction force (GRF) induced by a client's footsteps on a floor sensor system to construct biometric systems for client identification or verification. In contrast to gait analysis by video streams, it is non-intrusive, can operate in darkness and is less prone to noise in environmental conditions that might degrade the performance of the biometric system. As footstep GRF patterns tend to contain a high degree of fine-grained GRF variability they are difficult to visualise

- Omar Costilla-Reyes and Krikor B Ozanyan are with the School of Electrical and Electronic Engineering, The University of Manchester, Manchester M13 9PL, United Kingdom, Email: omar.costilla.reyes@gmail.com, k.ozanyan@manchester.ac.uk.
- R. Vera-Rodriguez is with the Biometrics and Data Pattern Analytics (BiDA) Lab - ATVS, Universidad Autonoma de Madrid, Avda. Francisco Toms y Valiente, 11 - 28049 Madrid, Spain, Email: ruben.vera@uam.es.
- Patricia Scully is with the School of Chemical Engineering and Analytical Science, Faculty of Engineering and Physical Sciences The University of Manchester, Manchester M13 9PL, United Kingdom, Email: patricia.scully@manchester.ac.uk.

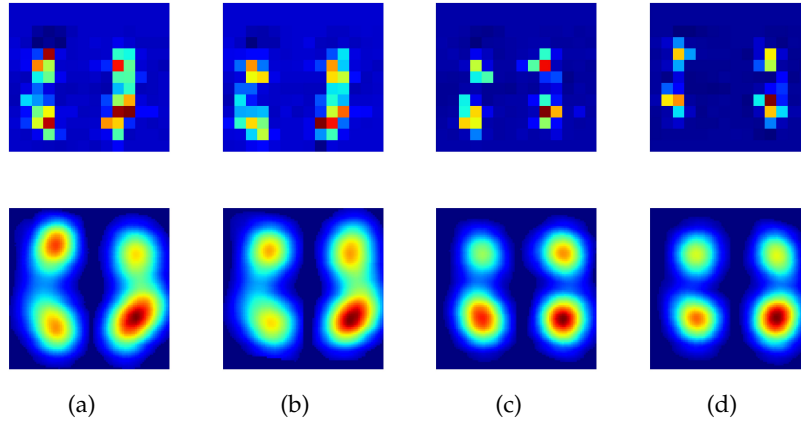


Fig. 1: Spatial raw (top) and spatial processed (bottom) footstep representations of 2 clients of the SFootBD. (a) and (b) footstep samples of user 1. (c) and (d) footstep samples of user 2. Top representation dimension is 13x14 pixels, bottom representation is 88x88 pixels. Solid red: maximum pressure, solid blue: minimum pressure.

for assessment by humans. Figure 1 shows a side by side comparison of raw (top) and processed (bottom) spatial footstep representations from 2 clients of the SFootBD, considering 2 samples per user. The comparison implies that effective footstep recognition based only on visual perception is a very challenging problem as there can be a high intra-user variability and low inter-user variability in some cases. Moreover, humans are not accustomed to recognize naturally detail in this type of images, as opposed to other biometric traits such as facial features.

Advanced computational models, such as machine learning have been used in an attempt to solve the differentiation of the fine-grained GRF variability between clients and impostors. Footstep feature extraction and feature engineering have been central in automatic footstep recognition research [14]. Feature engineering involves the careful selection and design of complex and time-consuming hand-crafted features from footstep data, employing geometric, spectral, wavelet and holistic feature engineering approaches to name a few [6]. Footstep recognition systems for biometric applications have focused on very small footstep datasets [9], [10], sometimes also constrained by the user's not being allowed to change footwear [6]. Deep learning models [15], for biometric footstep recognition using floor sensor systems, have not been sufficiently studied yet. Deep learning models have the ability to learn from raw sensor data and are effective to solve problems from image recognition to natural language processing [15]. Performance analysis of raw and feature engineered data on deep and shallow machine learning models have not been sufficiently addressed in the literature.

1.2 Aims and objectives

The aim of this work is to effectively differentiate the fine-grained GRF variability of footsteps between clients and impostors of a floor sensor biometric system in 3 critical data-driven security scenarios.

Spatio-temporal raw and processed footstep data representations are designed and evaluated on deep machine learning models, by using the SFootBD database [11], the largest footstep database to date with more than 120 people

and almost 20,000 footstep signals acquired from two rectangular arrays of 88 piezoelectric sensors each, to test the performance of our biometric system. Our deep machine learning models are based on the state-of-the-art resnet architecture [16] and the spatio-temporal two-stream architecture [17] [18] illustrated in Figure 2.

We compare the performance of our deep learning approach to the existing state-of-the-art: a hand-made feature learning approach of processed footstep data, used in a high-performing Support Vector Machine (SVM) model [11]. Moreover, in an effort to obtain a robust footstep recognition system, we design an ensemble [19] of the deep resnet and SVM machine learning models, using both processed and unprocessed footstep data.

Finally, we present a feature learning analysis of the optimal resnet models to observe client's footstep data clustering in an aim to understand the feature learning process of the resnet models.

The rest of the paper is organised as follows: section 2 presents the background. The floor sensor system and database are described in section 3. The spatio-temporal footstep representations are presented in section 4. The resnet model is described in section 5. The methodology and experiments are presented in section 6 and section 7 respectively. The feature analysis for the resnet models is presented in section 8. Finally, the summary and conclusions of this research are given in section 9.

2 RELATED WORK

2.1 Footstep data as a biometric

Footstep recognition from floor sensor systems has been proposed previously [6], [7], [9], [10], [11], [12], [13]. However, the studies rely heavily on complex hand-made feature engineering using shallow machine learning models such as Linear discriminant analysis (LDA) [6] or hidden Markov models (HMM) [9]. Related early studies of footstep data as a biometric, collected footstep signals from two main types of sensors: *i*) switch sensors [8], [20] analysing the spatial distribution of the footstep signals, and *ii*) pressure sensors [7], [9], [10], focusing on analysing the dynamic

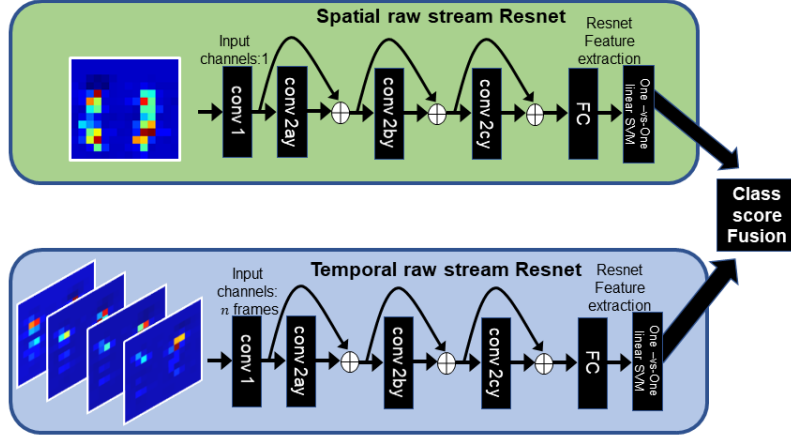


Fig. 2: Two-stream spatio-temporal resnet architecture for raw footstep representation.

pressure information in the signals, but with low spatial resolution. A commercial pressure mat with high resolution was used by Qian *et al.* [21] in order to extract the centre of pressure information, therefore using temporal and spatial pressure information only for some selected key points (geometric approach). Recently, footstep signals were analysed considering both temporal and spatial information contained in the signals [11], reporting experiments on the SFootBD database. The spatial information was extracted from accumulated pressure images, while temporal information was extracted from the average GRF and from other hand-crafted features. In both cases, Principal Component Analysis (PCA) was used to reduce the dimensionality of the footstep data and a non-linear SVM is used for biometric verification. Results in the range of 2.5% to 10% Equal Error Rate (EER) are achieved depending on the application setting.

Very recently, in [22] we reported a pilot study of a convolutional neural network model to learn processed spatial footstep features of the SFootBD database, suggesting significant improvements of footstep recognition performance compared to existing work [11]. Table 1 shows the recognition performance of our approach in comparison to other known biometric verification systems based on floor sensor data only. However, the other studies do not use the SFootBD database, thus cannot be directly compared to this work in terms of performance, since the experiments differ in the number of clients and footstep signals. We are using a much larger database in terms of the number of clients and footstep signals and therefore our results are statistically more significant.

2.2 Spatio-temporal deep learning and gait

Automatic spatio-temporal feature learning in machine learning research has focused on human action recognition from videos [5], [17], [18], [23]. Convolutional neural networks have been widely used in an attempt to solve the spatio-temporal human recognition problem from spatio-temporal data [18]. The two-stream deep learning architecture [17], [18] utilises an end-to-end learning approach for analysing the spatial and temporal streams of videos in two separate deep networks. The spatio-temporal information is then fused at a feature or score level after the last layers of

the network. However, this approach sometimes involves computationally heavy calculations such as optical flow [17].

Wu *et al.* [5] presented a study of cross-view gait for human identification, using deep convolutional neural networks models in three gait datasets. The results show a substantial increase in average recognition rate performance when compared with the previous state-of-the-art. For example, in the CASIA-B dataset, the average recognition rate reaches 94.1% with a deep network. This compares favourably with the previously achieved best recognition rate of 65% by hand-made feature engineering.

3 FLOOR SENSOR SYSTEM AND DATABASE

The floor sensor system used in this work obtains the pressure magnitude from piezoelectric sensors deployed at high spatial density. The system consists of two identical and rectangular sensor mats positioned to capture one stride footsteps: The right foot is captured first then the left. Each sensor mat has dimensions of 45×30 cm and comprises 88 piezo-electric sensors. This yields a resolution of 650 sensors per m^2 . The floor sensor system is capable of obtaining footstep signals at a frame rate of 1.6 kHz.

3.1 SFootBD Database

The SFootBD database¹ consists of a total of 19980 right and left footstep samples (9990 stride footsteps). A footstep sample lasts less than one second (1600 frames). The database footstep signals are synced 50 frames previous to the heel strike, as the sensors information was temporally stored in a buffer. The samples are captured from 127 volunteers in an 18-month period [11]. During this time, some subjects provided plenty of footstep signals in different sessions while other subjects only provided a small number of signals. A natural walking stride footstep was performed by the volunteers and they were allowed to wear any type of footwear to emulate a real-world scenario, as this characteristic has been proven to affect footstep recognition performance [6].

1. Available to the research community from <https://atvs.ii.uam.es/sfootbd.html>

TABLE 1: Verification footstep recognition systems

Research group	Database	Subjects	Samples	Model	Multi-shoe	Norm.	Results (EER)
Cattin, 2002 [7]	470	16	6 step cycles	Euclidean distance	Yes	No	9.45
Stevenson et al., 2007 [9]	88	8	85 step cycles	HMM	Yes	Yes	20
Mostayed et al., 2008 [10]	18	6	5 step cycles	Histogram Similarity	No	No	3.3 to 16
Vera-Rodriguez et al., 2013 [11]	9900	5	500 step cycles	SVM	Yes	Yes	2.6 val 4 eval
Mason et al., 2016 [6]	399	10	30 step cycles	LDA	Yes	Yes	1.52 val 3.1 eval
This study, 2017	9900	127	500 step cycles	resnet and SVM	Yes	Yes	0.7 val 1.70 eval

Three benchmark datasets, B1, B2 and B3 are constructed from the SFootBD database in order to simulate three distinct scenarios (airport, workplace and home scenario) with a different number of clients and available training footstep samples per client. Each benchmark dataset is further divided into training, validation and evaluation sets, as shown in Table 2. A large cohort of impostors is used in all benchmarks.

Benchmark B1 utilises 40 stride footstep samples per 40 clients and 763 impostor footstep samples for training the machine learning models. This is the smallest dataset considered for training, thus representing a security scenario such as at an airport entry checkpoint. Benchmark B2 has 200 stride footstep samples from 15 clients and 2697 impostor samples for training, thus representing a middle-level amount of footstep data available such as a workplace real-world scenario. Benchmark B3 contains 500 stride footstep signals from 5 clients and 5603 impostor samples for training. This is the largest training dataset in our experiments, which resembles a home environment, where a large number of footstep signals can be acquired due to easy access. The evaluation set is comprised of 550 footstep signals, and it is the same for all the three benchmark datasets, for B1 there are 40 clients, for B2 there are 15 clients, and for B3 there are 5 clients. The users of the database chosen as clients or impostors were selected based on the total number of available footstep signals. At each benchmark scenario, the clients were the maximum number of users with a certain amount of footstep signals, and the impostors were the remaining users. The description of the benchmarks is shown in Table 2.

4 SPATIO-TEMPORAL FOOTSTEP REPRESENTATIONS

4.1 Temporal downsampling

A temporal down-sampling factor of 6 is applied to the raw and processed representations by using an order 8 Chebyshev type I filter [24]. The factor is determined by using a grid search hyper-parameter optimisation technique [25]. Downsampling the signals by a factor of 6 showed an average 5% improvement of footstep recognition performance.

4.2 Raw footstep representation

4.2.1 Spatial component

The data from the 88 sensors available per footstep frame are reshaped in a 2D matrix of 13×7 pixels, a pixel value corresponds to a sensor amplitude value. Three zero values are inserted in the matrix, in order to make it rectangular due to the irregular shape of the sensor pad (see [11] for

sensor design specifications). Therefore, a sensor-derived matrix is obtained, of 13×14 pixels per stride footstep for the spatial component. The spatial footstep representation is expressed as the accumulated pressure AP_i of the i th sensor described as:

$$AP_i = \sum_{t=0}^{T_{max}} (GRF_i[t]), \quad (1)$$

here T_{max} is the maximum number of time samples of the footstep signals, set to $T_{max} = 1600$ frames (1 second), as no signal had a longer duration. GRF_i is the Ground Reaction Force (GRF) of the i th sensor in the mat array:

$$GRF_i[t] = \sum_{\tau=0}^t (s_i[\tau]), \quad (2)$$

Where $s_i[t]$ is the output signal of the piezoelectric sensor i . The representation can be observed in Figure 3a.

4.2.2 Temporal component

A set of t frames of the AP_{1-88} (Equation 1) of the footstep sensor signals are considered. The frames index t ranges from 1 to 267 per footstep after the temporal downsampling procedure is applied. The range $50 \leq t < 150$ is found to be the fewer number of frames that maximises performance, by a grid hyperparameter search procedure [26] considering random subsets. The temporal representation can be observed in Figure 3b. Although this method considers less than half the data per footstep, it is due to a feature selection procedure, which increases performance and decreases training time. The selected frames correspond to the flat foot dynamics as well as the end of the heel strike and the start of the heel off. The selected interval is shown in Figure 4a and Figure 4b enclosed in a black box for the sensor raw representation and upper contour respectively.

4.3 Processed footstep representation

4.3.1 Spatial component

The GRF_i (Equation 2) of the sensor signals is considered for this component, integrated along the frame axes. The sensor derived images are first smoothed using a Gaussian filter, then aligned and rotated to a common central position to allow better pattern discrimination by the machine learning models [11]. After processing, a single image of dimensions 88×44 pixels is obtained per footstep by up-sampling. Consequently, an image of dimensions 88×88 pixels is obtained per stride footstep. The representation can be observed in Figure 3c.

TABLE 2: SfootBD Database description

Benchmark	Benchmark B1 (40 clients)			Benchmark B2 (15 clients)			Benchmark B3 (5 clients)		
dataset	Train	Valid	Eval	Train	Valid	Eval	Train	Valid	Eval
Signals/client	40	170*	5	200	210*	5	500	121 *	5
Impostor signals	763	380	350	2697	630	475	5603	730	525
Total signals/set	2363	7077	550	5697	3743	550	8103	1337	550
Total	9990			9990			9990		

* average value

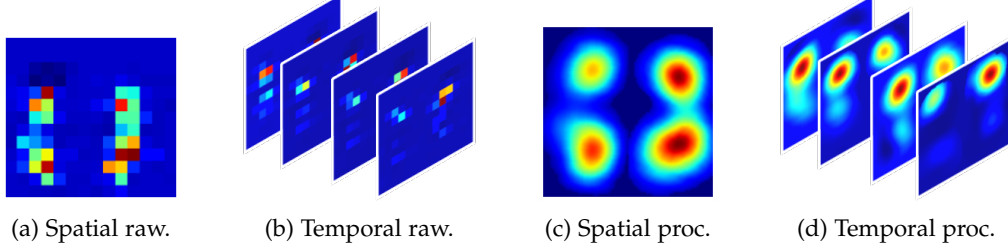


Fig. 3: Raw and processed spatio-temporal footstep representations.

4.3.2 Temporal component

A set of t frames per stride footstep sample is selected from the AP_{1-88} of the stride footstep signals. Similar to the raw data case, the optimal number of frames is again 100 in the range $50 \leq t < 150$ confirmed to be optimal by a grid hyperparameter search procedure [26]. After the AP is calculated, the processing techniques of the spatial component are applied. The temporal representation can be observed in Figure 3d. As in the case for the raw representation, the selected frames correspond to the flat foot dynamics in both footsteps. The effect can be observed in Figure 5c for frame 69 and Figure 5d for frame 78. The frames contain more spatial dynamic information of the footstep signals than the remaining frames as observed in Figure 5a, Figure 5b, Figure 5e and Figure 5f, for frames 32, 40, 181 and 189 respectively. These frames are outside the range found to be optimal.

4.3.3 Spatial-integrated temporal features for SVM model

The spatial-integrated temporal features included the following: A) *Spatial average representation*: the average output value of the 88 sensor signals per footstep are calculated at each t as in:

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

B) *Upper and lower contour representation*: here the maximum (upper) and minimum (lower) footstep AP are considered. These are mathematically expressed as:

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

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

C) *Global GRF[t] profile*: is computed as the average of the 88 individual $GRF_i[t]$, and mathematically expressed as:

$$GRF[t] = \frac{1}{88} \sum_{i=1}^{88} (GRF_i[t]). \quad (6)$$

To illustrate the temporally processed features, Figure 4b shows the representation of the upper contour representation for a stride footstep of client 1 (Figure 1a) of the database.

In previous work [11] the spatially-integrated temporal features and the spatially processed representation presented in section 4.3.1 were trained in a One-Vs-One SVM classifier with an RBF kernel for spatio-temporal fusion. Further details of the SVM methodology and experiments can be found in [11]. We replicated the experiments in a python environment. The experiments are presented in section 7.

4.4 Additional spatio-temporal raw representations

The spatio-temporal footstep data is also modified in systematic ways to observe the effect of changes in raw footstep data arrangements on deep learning models, including the following:

- 1) *Sensor raw representation*: The raw left and right footstep frames are concatenated at the X-axis and the 88 sensors per footstep are plotted at the Y-axis. The shape of a stride footstep sample is $(88, 534)$. The representation is shown in Figure 4a.
- 2) *Sensor raw accumulated pressure*: Same spatio-temporal configuration as sensor raw representation but considering GRF (Equation 2). The shape of a stride footstep sample is $(88, 534)$. The representation is shown in Figure 4c.
- 3) *Sensor raw matrix*: The 88 sensors of the left and right footsteps are concatenated at the Y-axis. The frames are shown at the X-axis. The shape of a stride footstep sample is $(176, 267)$.
- 4) *Sensor raw matrix sorted sensors*: spatio-temporal configuration same as sensor raw matrix, but the AP_i is sorted according to the maximum footstep pressure. The shape of a stride footstep sample is $(176, 267)$.

Due to low biometric verification performance and limitations of space, the additional representations results are not included in the results analysis of this paper. However, we include them to motivate further research.

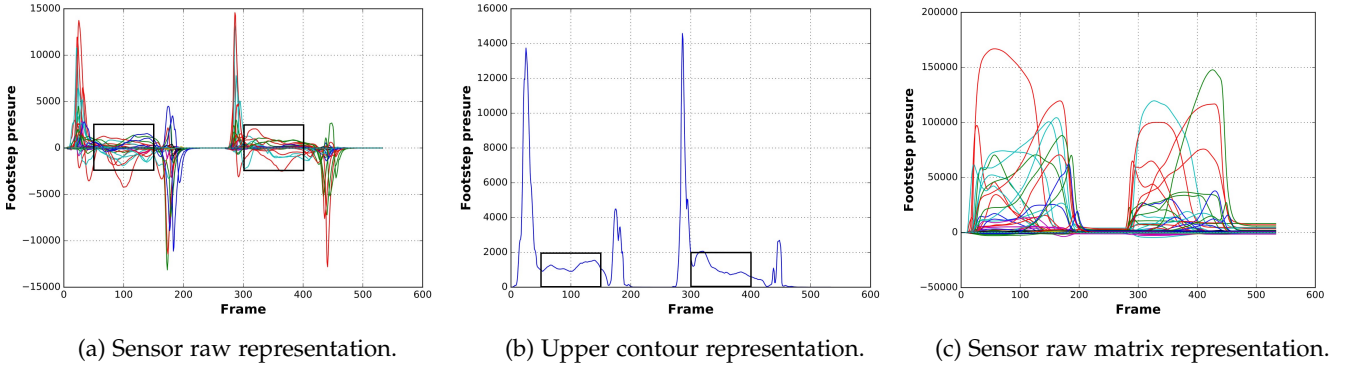


Fig. 4: Three spatio-temporal stride footstep representations. Right footstep first (frames 1 to 267) then left footstep (frames 267 to 534).

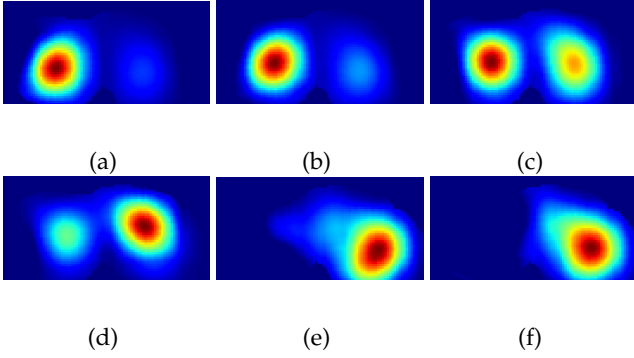


Fig. 5: Set of frames for the footstep processed representation. (a) frame 32, (b) frame 40, (c) frame 69, (d) frame 78, (e) frame 181, (f) frame 189.

5 DEEP RESIDUAL NETWORK MODEL

5.1 Overview

The deep machine learning models used in this work are based on the state-of-the-art resnet architecture [16]. An overview of the resnet architecture is illustrated in Figure 2 consisting of spatial and temporal streams for the raw representation. From input to output, each stream consists of the following layers: firstly there is a resnet configuration 1 block (2ay) (Figure 6 right), followed by resnet configuration 2 block (x2) (2by and 2cy) (Figure 6 left), then an average pooling layer, fully connected layer (FC) and finally a softmax layer. The blocks consist of convolutional layers, batch normalization [27] and ReLU activation functions [28]. The residual units can be expressed in general form as:

$$y_l = h(x_l) + G(x_l, W_l), \quad (7)$$

$$x_{l+1} = f(y_l), \quad (8)$$

where x_l is the input to the l -th residual block, and x_{l+1} is its corresponding output and G is a non-linear residual function. $h(x_l) = x_l$ is an identity mapping and f is a RELU activation [28] function. $W_l = \{W_{l,k} | 1 \leq k \leq K\}$ is the set of weights and biases of the l -th residual block. K is the number of layers in a residual unit. If f is an

identity mapping, then $x_{l+1} \equiv y_l$, therefore Equation 8 can be expressed as:

$$x_{l+1} = x_l + G(x_l, W_l). \quad (9)$$

For any unit of L and shallow unit l , the forward propagation of the feature x_L can be expressed as an additive output:

$$x_L = x_l + \sum_{i=l}^{L-1} G(x_i, W_i). \quad (10)$$

Therefore, during forward propagation, x_l is propagated to any x_L plus the residual factor. If the loss function is expressed as γ , the backpropagation of errors in the network can be expressed as the result of the following chain rule [29]:

$$\frac{\partial \gamma}{\partial x_l} = \frac{\partial \gamma}{\partial x_L} \frac{\partial x_L}{\partial x_l} = \frac{\partial \gamma}{\partial x_L} \left(1 + \frac{\partial}{\partial x_L} \sum_{i=l}^{L-1} G(x_i, W_i)\right). \quad (11)$$

Our resnet model is a shorter version of resnet-50 [16], with only 3 merge layers and 40 layers total, since these are the optimal number of layers found by using a layer-wise optimisation algorithm [26]. This reduces computation time by considering fewer parameters during training than in resnet-50.

5.2 Spatial and temporal architectures

The spatial and temporal footstep representations presented in section 4 share the same baseline resnet architecture shown in Figure 2. The input footstep representations affect the dimensions of the first convolutional (conv.) layer of the resnet model, that takes as input a stride footstep tensor of shape (n, m, c) where $n \times m$ is the 2D footstep sensor matrix and c the frames. $c = 1$ for the spatial case and $c = 100$ for the temporal case. The filter size of the convolutional layers of the resnet blocks (Figure 6) and channels change according to the input footstep tensor dimensions. The widely-used deep network design introduced by the VGG network [30] is adopted for the resnet models. The methodology decreases the spatial component at the conv. layers as a function of increasing the number of filter maps, from the left (input) to the right (output) layers of the network.

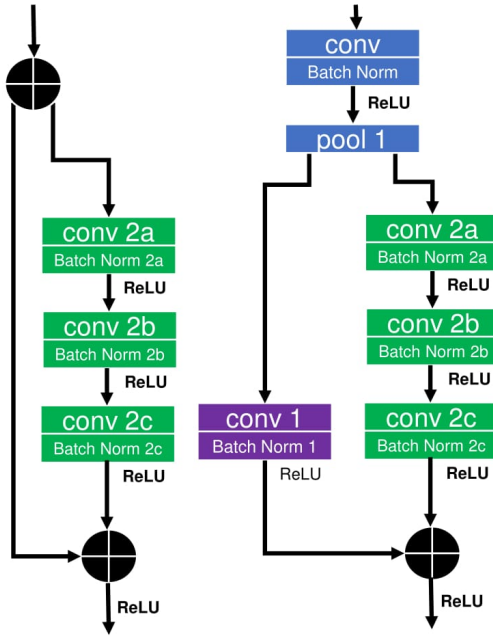


Fig. 6: Resnet model building blocks. Right: resnet configuration 1. Left: resnet configuration 2.

5.3 Data augmentation for raw footstep representations

Data augmentation methods are applied only to the raw spatio-temporal representation at training time to improve footstep recognition performance. Hence this procedure is not applied to the processed representation. The data augmentation techniques consider possible natural variations of the users footsteps when walking on the floor sensor. The rotation method consists of randomly drawing rotation angle values from a uniform distribution ranging from 0 to 20°. A shift method is also applied. The method draws random values from 0 to 20% shift in the X and Y axis. The dataset is augmented 10 times per stride footstep sample in the training set for both the raw spatial and temporal components using the shift and the rotation augmentation technique.

5.4 Loss function

The multi-class negative log-likelihood loss is used to evaluate the performance of the learning model predictions. The normalized loss function, for a single training sample is defined as:

$$P(y_i | x_i; W) = \frac{e^{f_{y_i}}}{\sum_j e^{f_j}} \quad (12)$$

5.5 Training

The resnet models are trained as in a multi-class client manner for a biometric identification scenario [1], in order to reduce substantially computation time. This is done by training one model per spatial and temporal component, instead of training u models for u clients per component. A single class is assigned to the impostor dataset. At training

time, the resnet model's weights are updated by the RM-Sprop [31] optimizer due to its stability. All resnet models are trained with a batch size of 32 stride footstep samples. Transfer learning is considered to initialize the weights of the first 37 layers with the ImageNet resnet-50 model [16] weights, before the fully connected layer. The learning rate is set initially at 0.001 and then decreased by a factor of 10 once the learning error plateaus. An early stopping procedure is implemented: training is stopped once the validation error stopped decreasing. Batch normalization [27] is used to normalise the activation of the convolutional and fully connected layers to prevent internal covariate shift during training.

5.6 Feature extraction, validation and evaluation

In order to construct a biometric verification system, the trained resnet models are used as feature extractors as shown in Figure 2. The feature vectors learnt by the model are extracted before the last softmax layer. This procedure delivered v feature vectors of shape $u + 1$. Interestingly, apart from acting as a feature transformer, the resnet models performed feature compression: a trained resnet model evaluated a mini-batch of high dimensional stride footstep tensors of shape $(32, n, m, c)$. Then, at the layer before the softmax, the model outputs a mini-batch tensor with shape $(32, u + 1)$, where $u + 1 \ll (n \times m \times c)$ for all scenarios. The validation set is used to test the model generalisation error at training time for the 3 benchmarks, delivering validation performance measures for an early stopping procedure. By using the resnet model as a feature extractor, we ease the evaluation of the biometric verification system. This allows saving computational resources and time. A One-Vs-One linear SVM classifier is selected, due to its high biometric verification performance when compared with other linear classifiers that were tried. During SVM model training, validation and evaluation, the samples not belonging to the i th client's class u , are treated as out-of-class samples. Then positive (clients) and negative (impostors) scores are created to calculate biometric verification error rates.

5.7 Hyperparameter optimization

The Tree of Parzen Estimator optimisation technique [26] is used to tune the spatio-temporal resnet model hyperparameters presented in Table 3, including value range and layer location in the resnet model. Figure 7 demonstrates the

TABLE 3: Resnet model hyperparameters

Hyperparameters	Hyperparameter space	resnet layer
Kernel size	1, 2, 3, 4, 5, 10, 15, 20, 25	8, 17, 27
Filter size	1, 2, 3, 5, 7	8, 17, 27
Conv. channels	1, 2, 4, 8, 16, 32, 64, 128	1
2D conv. window filter	1x1, 2x2, 3x3, 5x5, 7x7	1
Maxpooling 2D filter	1x1, 2x2, 3x3, 5x5, 7x7	4
Avg. Pooling 2D filter	1x1, 2x2, 3x3, 5x5, 7x7	37

hyperparameter optimisation procedure. The loss value and the average pooling kernel hyperparameter are shown per optimisation epoch. Here the spatial processed resnet representation of benchmark B3 validation dataset is used. The Figure shows how the loss error diminishes as the kernel of

the average pooling layers increases. A large pooling kernel tends to increase the reduction in dimensions of the footstep data from the previous layer, this averages a larger patch of the 2D feature maps. The observed strong fluctuations in the loss value are due to dynamic change in other model hyperparameters (Table 3) at each optimisation step.

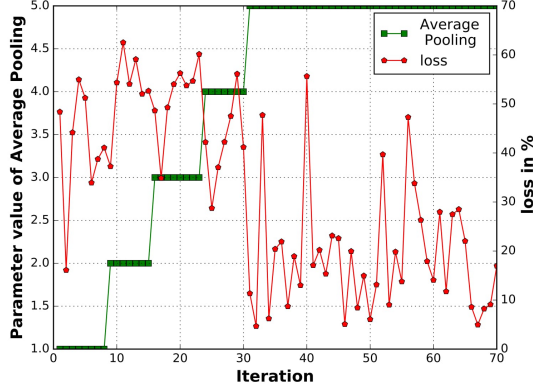


Fig. 7: Average pooling hyperparameter optimization. The average pooling parameter is on the left Y axis and the loss in percentage value is on the right Y axis.

6 METHODOLOGY

6.1 Data preprocessing

As data preprocessing procedure, the mean is removed and scaled to unit variance of the training sets footstep samples, the transformation parameters are then transferred to the validation and evaluation datasets for all experiments.

6.2 Evaluation metrics

The verification system performance is evaluated by using detection error trade-off (DET) curves [32]. The EER is the intersection in the DET curve where the False Rejection Rate (FRR) and the False Acceptance Rate (FAR) are equal. Therefore, equal importance is given to FRR and FAR for the evaluation of our experiments. The EER measure is also used to represent the biometric verification performance of the system with a single value as shown in Figure 9.

6.3 Comparison of feature and score level fusion of spatio-temporal footstep representations

The spatial and temporal streams of the resnet models are fused at the feature level and at the score level to compare performance, the latter being the most common approach [33]. The score level fusion shows the best performance in our experiments, therefore, a spatio-temporal fusion at the score level is adopted for all experiments. Figure 8 shows the biometric verification performance of the following score combination rules: product, sum, max and min [33], as DET curves and EER, for the validation set of benchmark B1. The product fusion rule yields the optimal EER and is selected for all experiments.

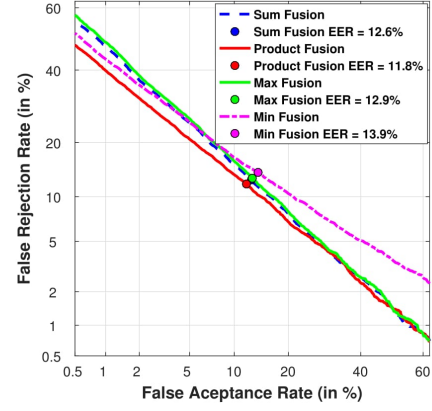


Fig. 8: Spatio-temporal raw resnet fusion score rule performance for benchmark B1.

6.4 Implementation details

Open source software is used for the development and evaluation of the research experiments. Python scientific computing packages [34] are used. Tensorflow numerical computation library [35] are used to develop and train the resnet models. The model evaluation algorithms and SVM models are developed using the scikit-learn machine learning library [36]. Bob package [37] is used to calculate DET curves and EER. Finally, Hyperopt [25] is used for hyperparameter optimisation. They are trained using a Titan X GPU for fast parallel optimisation of the learning models. The SVM model training and evaluation are parallelized at the CPU level.

7 EXPERIMENTS

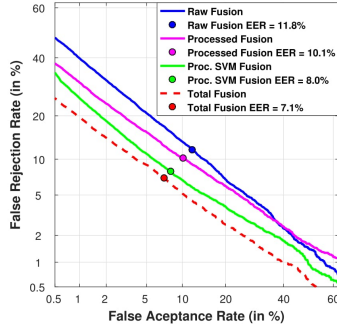
Table 4 shows the EER scores for benchmarks B1, B2 and B3 validation and evaluation datasets. The Table results are organised according to model, representation and representation combination, the performance is reported with EER metrics. For each representation-model combination, EER is shown for the temporal and spatial streams and for the fusion of them. The last two rows of the table show the performance of the fusion of models and representations. Results are shown for the three benchmarks validation and evaluation datasets. Also, Figure 9 shows the verification performance of the system in the form of DET curves for the optimal spatio-temporal models displayed by the benchmark used. Results below are expressed as validation and evaluation pairs: (validation % and evaluation %). The evaluation dataset verification performance is considered as the metric for determining the optimal models since these scores provide biometric verification performance for a held-out dataset.

7.1 Airport scenario: Benchmark B1

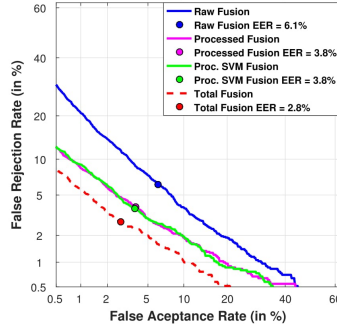
The combination of the raw, processed resnet and processed SVM representations delivers the optimal performance overall with 7.10% and 10.50% EER. thus, the fusion of models and representations improves the previously reported EER optimal performance [11] of 8% and 12.5%

TABLE 4: Biometric verification results in terms of EER (in %) for benchmarks B1, B2 and B3

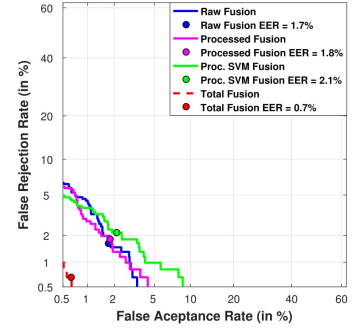
Domain	Model	Benchmark B1 (40 clients)		Benchmark B2 (15 clients)		Benchmark B3 (5 clients)	
		Validation	Evaluation	Validation	Evaluation	Validation	Evaluation
Raw representations							
Temporal	resnet	14.70%	18.00%	8.20%	6.70%	4.60%	8.00%
Spatial	resnet	16.30%	13.40%	11.20%	10.70%	3.40%	12.00%
Fusion Spatio-temporal	resnet	11.80%	11.50%	6.10%	8.00%	1.70%	5.60%
Fusion Spatio-temporal	DNN	27.65%	27.93%	14.33 %	17.33%	5.76%	6.57%
Fusion Spatio-temporal	CNN	31.28%	31.21%	14.26%	14.67%	3.62%	4%
Processed representations							
Temporal	resnet	12.20%	18.00%	6.60%	9.30%	3.90%	2.00%
Spatial	resnet	13.60%	15.50%	5.50%	9.30%	3.00%	6.60%
Fusion Spatio-temporal	resnet	10.10%	14.50%	3.80%	8.00%	1.80%	2.60%
Fusion Spatio-temporal	DNN	17.25%	21%	6.10%	6.66%	2.80%	3%
Fusion Spatio-temporal	CNN	18.1%	23%	6.07%	9.95%	1.61%	3.38%
Processed SVM representations							
Spatial-integrated temporal	SVM	12.10%	16.50%	9.30%	12.00%	6.10%	8.20%
Spatial	SVM	11.70%	17.50%	5.90%	9.20%	3.80%	2.60%
Fusion Spatio-temporal	SVM	8.00%	12.50%	3.80%	6.70%	2.10%	3.20%
Fusion of representations							
Raw & Processed	resnet	8.10%	10.70%	3.20%	5.30%	0.80%	2.10%
Raw & Processed & Processed SVM	resnet and SVM	7.10%	10.50%	2.80%	4.90%	0.70%	1.70%



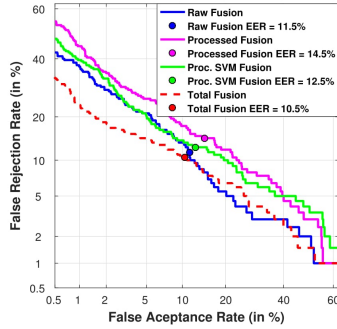
(a) Validation for benchmark B1.



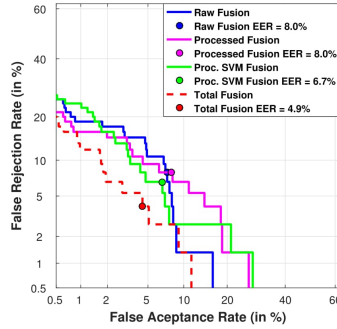
(b) Validation for benchmark B2.



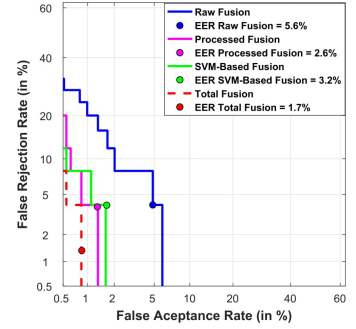
(c) Validation for benchmark B3.



(d) Evaluation for benchmark B1.



(e) Evaluation for benchmark B2.



(f) Evaluation for benchmark B3.

Fig. 9: Validation and evaluation dataset performance for spatio-temporal fusions and fusion of models and representations for B1, B2 and B3 benchmarks. Step-wise evaluation curves are due to small dataset size.

EER, which is an improvement ratio of 12% and 19% on the previous EER. This benchmark considers the least amount of footstep data per client for training from the 3 benchmarks. The benchmark exemplifies a real-world security application, where footstep data is scarce for training, considering the few number of times a client passes a security check-point.

7.2 Workplace scenario: Benchmark B2

The combination of the raw, processed resnet and processed SVM representations deliver the optimal performance overall as well of 2.80% and 4.90% EER for this dataset. This improves the previously reported optimal performance [11] of 4.5% and 7% EER, which is an improvement ratio of 60% and 43% on the previous EER. This benchmark considers a medium amount of footstep data per client for training from the 3 benchmarks. A workplace security environment exemplifies a real-world scenario.

7.3 Home scenario: Benchmark B3

The combination of the raw, processed resnet and processed SVM representations deliver the optimal performance overall as well with 0.70% and 1.70% EER for this dataset and overall in all experiments. This improves the previous reported optimal performance [11] of 2.6% and 4% EER, which is an improvement ratio of 371% and 235% on the previous EER. A home environment exemplifies a real-world security application of this dataset, where a large number of footstep signals can be obtained due to easy accessibility.

7.4 Experimental comparison with baseline neural network architectures

Baseline experimental results were performed for shallow CNN and feed forward neural networks (FNN) in order to train the processed and raw data representations in low complexity models to compare to the performance results of the resnet model (higher complexity).

The CNN model consists of a single CNN layer followed by a fully connected layer. The FNN model consists of a single fully connected layer. As the testing of the deeper feedforward neural networks did not manifest any significant improvement over the results of a FNN with a single layer.

The evaluation results presented in Table 4 show that neither the CNN or FNN of the shallow models outperform the deep residual neural network. In regards to the improvement due to residual connections, for Benchmark B1 the performance results for a single layer CNN is 18.1% and 23% EER, when compared with the results of the resnet model of 10.1% and 14.5% EER. This shows a significant improvement ratio of 79% and 63% in validation and evaluation respectively considering the resnet model with residual connections.

7.5 Discussion

As presented in Table 4, better verification performance can be observed with increased footstep data available per client for training. Fusion of resnet and SVM models, considering both raw and processed footstep representations performs best overall in all benchmark experiments. One of the reasons for the higher verification performance is due to consider the models and representations as in an ensemble learning scenario [19], thus influencing a robust model predictive performance. Also, when combined, the footstep data representations provide unique features from spatio-temporal footstep signals. This behaviour is also supported by evidence in [38] that residual networks can behave as an ensemble of shallow convolutional neural networks.

8 DEEP RESIDUAL NETWORK FEATURE ANALYSIS

Our raw resnet models have the advantage of being able to create features and learn from raw footstep data. In this section we visualise and analyse the features created by such models, to understand the automatic learning process by visualising the features.

8.1 Manifold learning

The t-distributed Stochastic Neighbour Embedding (t-SNE) technique was tested on the trained resnet model of Benchmark B3 for feature learning visualization. The techniques transform the trained resnet models output $u + 1$ -dimensional vectors (clients + impostor class) at the layer before the softmax to a 2-dimensional space to observe the underlying structure of the learned features in the datasets, in an aim to uncover visual cluster patterns between features and classes. The t-SNE representations for the training, validation and evaluation datasets for the spatial processed feature of benchmark B3 can be observed in Figure 10a, Figure 10b and Figure 10c respectively. The Figures show clustering of clients (class 0 to 4) and impostors (class 5), the clusters are easily distinguishable for the training set, then due to generalization performance the clusters become less visible for the validation and evaluation datasets.

8.2 Layer-wise feature maps

We extract feature maps from the trained resnet model of the benchmark B3 dataset spatial processed representation at critical layers of the network, from left (input) to right (output). This is carried out in order to observe the automatic feature learning process that the resnet model performs at critical layers of the network. The feature maps of Figure 11 (client 2 footsteps shown in Figure 1b) show footstep signals twisted, shifted and rotated automatically by the trained resnet model. The feature maps obtained at layer 36, show that the features created at the layer have 4 isolated footstep patterns. This representation is the closest feature map to the softmax classification layer, therefore used for training.

9 SUMMARY AND CONCLUSIONS

We have presented a methodology for robust footstep recognition for biometric verification in 3 critical data-driven security scenarios. The methodology is based on spatio-temporal raw and footstep processed representations in resnet and SVM machine learning models. We validate our methodology in a challenging database, the SFootBD, the largest footstep database to date available for footstep research, containing only 2 footsteps per experimental sample. In all security scenarios, we obtain state-of-the-art biometric verification performance for an ensemble of resnet and SVM models using both spatial and temporal components of processed and unprocessed footstep data.

Our three experimental settings are based on varying the amount of footstep data available for training the machine learning models, therefore, varying the number of clients and impostors. The smallest training footstep dataset resembles an airport security checkpoint scenario (benchmark B1) by considering 40 stride footstep samples for 40 clients each and 763 impostor stride footstep samples. The optimal model delivers 10.50% EER in held-out footstep data, an improvement ratio of 19% from previous state-of-the-art. The medium sized training dataset, considers 200 stride footstep samples for 15 clients each and 2697 impostor footstep signals, resembling a workplace scenario (benchmark B2), where we obtain an optimal EER of 4.90% in held-out footstep data, an improvement ratio of 42% from previous state-of-the-art. Finally, the largest training dataset resembles a

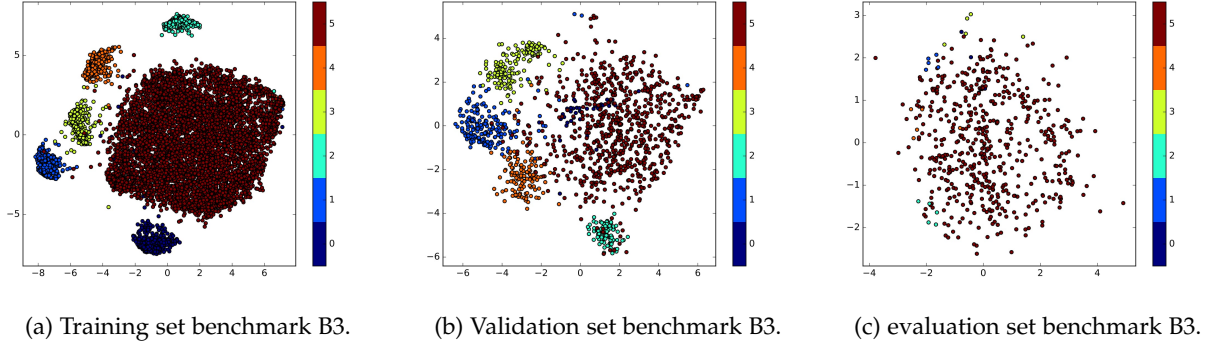


Fig. 10: t-SNE representation of resnet models of benchmark 3 (Temporal raw) dataset. In all figures red colour is the impostor class, other colours are client classes.

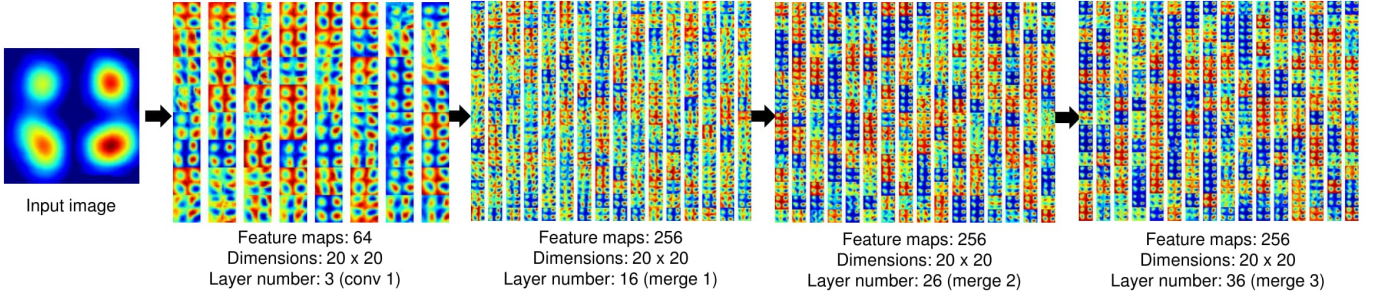


Fig. 11: Feature maps of the benchmark B3 spatial representation training set (a) conv. layer #3, (b) merge layer #16, (c) merge layer #26, (d) merge layer #36.

home-based scenario (benchmark B3) which considers 500 stride footstep samples per 5 clients each and 5603 impostor footstep signals. Here, our experiments deliver the best validation and evaluation verification performance overall with 0.70% EER and 1.70% EER respectively, an improvement ratio of 371% and 235% respectively from previous state-of-the-art [11]. Biometric verification performance tends to improve as a result of increasing the size of footstep training data considered per client and by decreasing the number of clients.

An analysis of the learnt features of the optimal resnet models is undertaken to improve understanding of the feature learning models. Using the t-SNE technique for feature clustering visualisation, we evaluate the optimally trained resnet models of B2 and B3, our results show effective client clustering for the benchmark's datasets.

Finally, we extract feature maps at the merge layers in one optimal resnet models for B3. The feature maps illustrate the automatic feature learning procedure of the resnet model at critical layers: very disperse footstep features are obtained at the leftmost layers, while in contrast, isolated features are shown at the rightmost layers (closer to the classifier), resulting in an effective discrimination between client's and impostor footsteps for robust biometric verification.

In conclusion, we demonstrate robust biometric classification performance from the largest to date footstep database. We have studied 3 representative data-driven real world scenarios resembling biometric tasks at an airport, a workplace and at home. We show notable improvement in the state-of-the-art EER, with up to 371% in one par-

ticular scenario. This is achieved by introducing automatic feature learning and fusion of several representations and models. A feature learning analysis is undertaken to aid the understanding of features created by the deep residual neural networks and to observe client's cluster patterns. The proposed spatio-temporal methodology has potential to be applied in other research areas that deal with similar fine-grained spatio-temporal variability problems outside of biometrics, for example, to classify brain tumours with magnetic resonance imaging scans [39].

ACKNOWLEDGMENTS

We express our gratitude to David H. Foster, Hujun Yin and Bernardino Romera-Paredes for useful discussions. O. Costilla-Reyes would like to acknowledge CONACyT (Mexico) for a studentship. We acknowledge NVIDIA for the donation of the GPU used to perform some of the experiments of this research. This work has been partially supported by Cognimetrics TEC2015-70627-R MINECO/FEDER.

REFERENCES

- [1] J. R. Vacca, *Biometric technologies and verification systems*. Butterworth-Heinemann, 2007.
- [2] M. W. Whittle, "Clinical gait analysis: A review," *Human Movement Science*, vol. 15, no. 3, pp. 369–387, 1996.
- [3] P. D. Tsatsoulis, A. Jaech, R. Batie, and M. Savvides, *Continuous Authentication Using Biometrics*. IGI Global, 2012.
- [4] D. Gafurov, "A survey of biometric gait recognition: Approaches, security and challenges," in *Annual Norwegian computer science conference*, 2007, pp. 19–21.

- [5] Z. Wu, Y. Huang, L. Wang, X. Wang, and T. Tan, "A comprehensive study on cross-view gait based human identification with deep cnns," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 2, pp. 209–226, 2017.
- [6] J. E. Mason, I. Traoré, and I. Woungang, *Machine Learning Techniques for Gait Biometric Recognition*. Springer, 2016.
- [7] P. C. Cattin, "Biometric Authentication System Using Human Gait," Ph.D. dissertation, Diss., ETH Zurich, Nr. 14603, 2002.
- [8] L. Middleton, A. Buss, A. Bazin, and M. Nixon, "A Floor Sensor System For Gait Recognition," in *Proc. of Fourth IEEE Workshop on Automatic Identification Advanced Technologies*, 2005, pp. 171–176.
- [9] J. P. Stevenson, S. L. Firebaugh, and H. Charles, "Biometric identification from a floor based pvdv sensor array using hidden markov models," in *Proc. SAS*, vol. 7, 2007.
- [10] A. Mostayed, S. Kim, M. M. G. Mazumder, and S. J. Park, "Foot step based person identification using histogram similarity and wavelet decomposition," in *Proc. of the 2nd International Conference on Information Security and Assurance*. IEEE, 2008, pp. 307–311.
- [11] R. Vera-Rodriguez, J. S. D. Mason, J. Fierrez, and J. Ortega-Garcia, "Comparative analysis and fusion of spatiotemporal information for footstep recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 4, pp. 823–834, 2013.
- [12] S. P. Moustakidis, J. B. Theocharis, and G. Giakas, "Subject recognition based on ground reaction force measurements of gait signals," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 38, no. 6, pp. 1476–1485, 2008.
- [13] O. Costilla Reyes, P. Scully, and K. B. Ozanyan, "Temporal Pattern Recognition in Gait Activities Recorded with a Footprint Imaging Sensor System," *IEEE Sensors Journal*, vol. 16, no. c, dec 2016.
- [14] R. Vera-Rodriguez, N. Evans, and J. S. D. Mason, "Footstep recognition," *Encyclopedia of Biometrics*, pp. 693–700, 2015.
- [15] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [16] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.
- [17] K. Simonyan and A. Zisserman, "Two-stream convolutional networks for action recognition in videos," in *Advances in Neural Information Processing Systems*, 2014, pp. 568–576.
- [18] C. Feichtenhofer, A. Pinz, and R. Wildes, "Spatiotemporal residual networks for video action recognition," in *Advances in Neural Information Processing Systems*, 2016, pp. 3468–3476.
- [19] J. Wu and J. M. Rehg, *Ensemble Machine Learning*. Springer, 2012.
- [20] J. Yun, "User identification using gait patterns on UbiFloorII," *Sensors*, vol. 11, no. 3, pp. 2611–2639, 2011.
- [21] G. Qian, J. Zhang, and A. Kidane, "People identification using floor pressure sensing and analysis," *IEEE Sensors Journal*, vol. 10, no. 9, pp. 1447–1460, 2010.
- [22] O. Costilla-Reyes, R. Vera-Rodriguez, P. Scully, and K. B. Ozanyan, "Spatial footstep recognition by convolutional neural networks for biometric applications," in *Proc. of 2016 IEEE Sensors conference*. IEEE, 2016.
- [23] S. Ji, W. Xu, M. Yang, and K. Yu, "3D Convolutional neural networks for human action recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 221–231, 2013.
- [24] A. Antoniou, *Digital signal processing*. McGraw-Hill Toronto, Canada, 2006.
- [25] J. Bergstra, D. Yamins, and D. D. Cox, "Hyperopt: A python library for optimizing the hyperparameters of machine learning algorithms," in *Proc. of the 12th Python in Science Conference*, 2013, pp. 13–20.
- [26] J. Bergstra, R. Bardenet, Y. Bengio, and B. Kégl, "Algorithms for Hyper-Parameter Optimization," in *Advances in Neural Information Processing Systems*, 2011, pp. 2546–2554.
- [27] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. of the 32nd International Conference on Machine Learning*, 2015, pp. 448–456.
- [28] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proc. of the 14th International Conference on Artificial Intelligence and Statistics*, vol. 15, 2011, pp. 315–323.
- [29] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," in *Proc. of European Conference on Computer Vision*. Springer, 2016, pp. 630–645.
- [30] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.
- [31] T. Tieleman and G. Hinton, "Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude." *COURSERA: Neural Networks for Machine Learning*, 2012.
- [32] A. Martin, G. Doddington, T. Kamm, M. Ordowski, and M. Przybicki, "The det curve in assessment of detection task performance," *DTIC, Tech. Rep.*, 1997.
- [33] A. K. Jain, K. Nandakumar, and A. a. Ross, "Score Normalization in Multimodal Biometric Systems Multimodal Biometric Systems," *Pattern recognition*, vol. 38, no. 12, pp. 4–5, 2005.
- [34] T. E. Oliphant, "Python for scientific computing," *Computing in Science & Engineering*, vol. 9, no. 3, pp. 10–20, 2007.
- [35] M. Abadi, A. Agarwal *et al.*, "Tensorflow: Large-scale machine learning on heterogeneous distributed systems," *arXiv preprint arXiv:1603.04467*, 2016.
- [36] F. Pedregosa, G. Varoquaux *et al.*, "Scikit-learn: Machine learning in python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [37] A. Anjos, L. El-Shafey *et al.*, "Bob: a free signal processing and machine learning toolbox for researchers," in *Proc. of the 20th ACM international conference on Multimedia*. ACM, 2012, pp. 1449–1452.
- [38] A. Veit, M. Wilber, and S. Belongie, "Residual networks are exponential ensembles of relatively shallow networks," *arXiv preprint arXiv:1605.06431*, vol. 1, 2016.
- [39] M. A. Weber, S. Zoubaa *et al.*, "Diagnostic performance of spectroscopic and perfusion MRI for distinction of brain tumors," *Neurology*, vol. 66, no. 12, pp. 1899–1906, 2006.



Omar Costilla-Reyes Received the M.Sc. degree in Electrical Engineering from the University of North Texas, Texas, U.S.A. in 2014. During his master studies, he was a research assistant in projects with funding from the National Science Foundation (NSF) and National Aeronautics and Space Administration (NASA). His M.Sc. dissertation was on dynamic indoor positioning systems using wireless sensor networks. He is currently a PhD candidate in Electrical and Electronics Engineering at the University of Manchester, U.K. He has published papers and has research interest in applications of machine learning for gait analysis in security and healthcare. He received the Best Student Paper Award in Optical Sensing applications at the 2015 IEEE Sensors Conference. He has won scholarships and awards for academic achievement including an academic scholarship for his master's and doctorate studies from the Mexican Science council (CONACyT).



Ruben Vera-Rodriguez Received the M.Sc. degree in telecommunications engineering from Universidad de Sevilla, Spain, in 2006, and the PhD degree in electrical and electronic engineering from Swansea University, U.K., in 2010. Since 2010, he has been with the Biometric Recognition Group, Universidad Autonoma de Madrid, Spain, first as the recipient of a Juan de la Cierva Post-Doctoral Fellowship from the Spanish Ministry of Innovation and Sciences, and as an Assistant Professor since 2013. His research interests include signal and image processing, pattern recognition, and biometrics, with emphasis on signature, face and gait verification and forensic applications of biometrics. Dr Vera-Rodriguez is actively involved in several National and European projects focused on biometrics. He was the recipient of the best paper award at the 4th International Summer School on Biometrics, Alghero, Italy, in 2007.



Patricia Scully Received the PhD degree in Engineering from the University of Liverpool, Liverpool, U.K., in 1992, and reached Reader position with Liverpool John Moores University, Liverpool, in 2000 before joining the University of Manchester as a Senior Lecturer/Associate Professor in Sensor Instrumentation with the University of Manchester in 2002. She is experienced in leading industrial and research council/government funded research projects at national and international levels, and has research

interests in sensors and monitoring for industrial processes, including optical fibre technology and photonic materials for sensors and devices, ranging from functional chemically sensitive optical coatings, to laser inscribed photonic and conducting structures in transparent materials that affect the properties of light.



Krikor B. Ozanyan received the M.Sc. degree in engineering physics (semiconductors) and the PhD degree in solid-state physics in 1980 and 1989, respectively. He has more than 300 publications in the areas of devices, materials and systems for sensing and imaging. He is currently Director of Research at the School of EEE, at the University of Manchester, U.K. He is a Fellow of the Institute of Engineering and Technology, U.K., and the Institute of Physics, U.K. He was a Distinguished Lecturer of the IEEE Sensors

Council in 2009 and 2010, and Guest Editor for the 10th Anniversary Issue of IEEE SENSORS JOURNAL in 2010, as well as the Special Issues on Sensors for Industrial Process Tomography in 2005 and THz Sensing: Materials, Devices and Systems in 2012. He is currently in his second term as Editor-in-Chief of the IEEE SENSORS JOURNAL and is General Co-Chair of the IEEE SENSORS 2017 Conference.