Spatial Footstep Recognition by Convolutional Neural Networks for Biometric Applications

Omar Costilla-Reyes*, Ruben Vera-Rodriguez†, Patricia Scully* and Krikor B Ozanyan*
*School of Electrical and Electronic Engineering and Photon Science Institute
The University of Manchester, Manchester M13 9PL, United Kingdom
† Biometric Recognition Group - ATVS, Escuela Politecnica Superior, Universidad Autonoma de Madrid
Avda. Francisco Tomas y Valiente, 11 - Campus de Cantoblanco - 28049 Madrid, Spain
Correspondence email: omar.costilla.reyes@gmail.com, ruben.vera@uam.es

***Abstract*—We propose a Convolutional Neural Network model to learn spatial footstep features end-to-end from a floor sensor system for biometric applications. Our model's generalization performance is assessed by independent validation and evaluation datasets from the largest footstep database to date, containing nearly 20,000 footstep signals from 127 users. We report footstep recognition performance as Equal Error Rate in the range of 9% to 13% depending on the test set. This improves previously reported footstep recognition rates in the spatial domain up to 4% EER.**

***Index Terms*—pattern recognition, machine learning, convolutional neural networks, gait analysis, floor sensor system.**

I. Introduction

The analysis of human gait has been used in many different applications such as medicine, sports, surveillance, smart homes, multimedia and biometrics [1]. Gait analysis can be measured from different types of sensors such as video cameras which normally record people walking (application in CCTV analysis among others), motion sensors attached to the lower body or contact sensors placed on the floor to capture footstep signals. Floor sensor systems have the advantage of being non-invasive and have a high adoption rate in home environments [2]. This work focuses on the analysis of footstep signals acquired from piezoelectric sensors placed under a carpet floor. We propose a machine learning model based on Convolutional Neural Networks (CNNs) to automatically learn spatial footstep features to construct a biometric system.

II. Background

Footstep signals were first assessed in [1], [3] as biometric with statistical significance, reporting experiments on SFootBD, the largest public database with more than 120 people and almost 20,000 signals acquired from a mat array of piezoelectric sensors. The discriminative information contained in the footstep signals was analysed in both time and spatial domains [1], as well as their fusion. Spatial and temporal information was extracted from Principal Component Analysis (PCA) features of accumulated pressure images, and a non-linear Support Vector Machine (SVM) was used for classification.

On the other hand, CNNs models, as one of the architectures of deep learning [4], have recently shown state-of-the-art performance for image recognition tasks, including object detection and localization. In this work, the CNN application is expanded to the problem of footstep recognition, proposing a CNN model to automatically learn footstep features, end-to-end, from spatial footstep signals acquired from a floor sensor system. Results show significant improvements of performance compared to previously published works on the same database.

III. Floor Sensor System and Database

A. Floor Sensor System

The system combines the characteristics of a high sensor density and pressure magnitude obtained from piezoelectric sensors. The system is comprised of two sensor mats positioned to capture one stride footstep, i.e. signals from two consecutive footsteps (right foot then left foot). Each mat measures 45 $\times$ 30 cm and contains 88 piezoelectric sensors, giving a sensor resolution of approximately 650 sensors per m^2; the sampling frequency associated with each sensor is 1.6 kHz, and therefore having a capture system with high time and high spatial resolution.

B. SFootBD Database

The SFootBD database [1] is the largest publicly available database to date of pressure based footstep signals, with 127 subjects and almost 20,000 valid footstep signals (i.e. 10,000 stride signals). The volunteers were allowed to wear any type of footwear, or remain barefoot and could also carry bags, in order to emulate a real world home scenario.

The experimental work is carried out following the available baseline benchmark [3], which divides the database (9900 samples) into training (2363 samples), validation (7077 samples) and evaluation datasets (550 samples), having the training set comprised of 40 clients with 40 stride footstep signals each, and 87 impostor subjects. The database reflects a real world recognition system by dividing the datasets according to date. This enables the training set to use the first set of signals and the evaluation set to contain the later signals that were last or more recently acquired.

IV. Methods

A. Footstep Spatial Representation

We use a spatial representation of the pressure of the footstep signals. The footstep profile is expressed as the accumulated pressure AP_i of the ith sensor described as:

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

Where GRF_i is the Ground Reaction Force (GRF) of the ith sensor in the array:

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

This representation considers the distribution of the accumulated pressure along a footstep signal in the spatial domain. In this case, the individual GRF (GRF_i) of each footstep sensor is integrated along the time axis, obtaining a single value of the accumulated pressure (AP_i) for each sensor of the array for a footstep signal.

Then, the sensor-derived images are smoothed using a Gaussian filter in order to obtain continuous images and were rotated and aligned to a common position, in a similar way as in [3]. A 88x44 pixel image for the right and left footstep per sample in the database is considered. The two footsteps are also concatenated to create a unique 88x88 pixel image per stride experiment. A spatial representation of one of the stride experiments in the training set after smoothing, centring and rotation is presented in Figure 1.

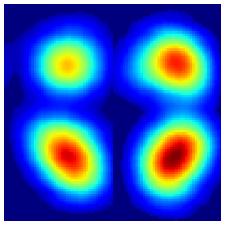

Fig. 1: Concatenated left and right footsteps

B. Data Pre-processing

The training dataset was standardized by removing the mean and scaling to unit variance. These statistics were then transferred to standardize the testing and validation datasets.

C. CNN Model

CNN models were trained per left, right and stride footsteps. The only difference between the models being the input image dimensions. The stride footstep has an input image dimension of 88x88 pixels, while the left and right footsteps have input dimensions of 88x44 pixels.

The CNN models are used as a feature extractor that allows the use of such features in a One-vs-One linear SVM model to construct a biometric verification system. This approach proved to be computationally efficient in comparison of training a CNN model per client.

1) Architecture: Figure 2 shows the architecture of the CNN model proposed in this work. The CNN model has 2 convolutional layers and one fully connected layer. A 7x7 spatial filter was used in the convolutional layers. 20 channels were used per convolutional layer. Each layer is followed by batch normalization and a ReLU activation function as in [5]. The last layer dense uses a softmax activation to obtain class scores. The max and average pooling operations are used as in the Resnet [5] architecture to eliminate the need of extra dense layers at the end of the network. This substantially decreased the number of model parameters and computation time. Our CNN model has less number of layers when compared with other works [4]. Due to its low complexity, this allows an even shorter training time.

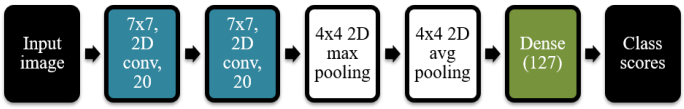

Fig. 2: CNN architecture

2) Training and Testing: The CNNs training and testing computations were performed on a TITAN X Graphic Processing Unit to speed up the network's computations. The CNN was trained end-to-end by backpropagation as in a multiclass problem. The RMSprop optimizer [6] was used with an initial learning rate set at 0.001 and decreased by a factor of 2 when the validation error plateaus. We implemented an early stopping procedure, training was stopped when the validation error did not improve after 10 epochs. A batch size of 16 samples was used per gradient update and the categorical cross entropy was considered as the loss measure. For evaluation of the validation and testing datasets the trained CNN model with optimal set of weights was used for feature extraction.

D. CNN model for Feature Extraction

Our proposed CNN model is used for feature extraction for the complete spatial database. After the CNN is trained, the training, validation and testing datasets are run through the network and the features created by the CNN are extracted at the last layer before the softmax activation. This returns a 127 feature vector per sample in the dataset. Considering that the input spatial footstep image has dimensions of 88x88 pixels for the stride signals or 88x44 pixels for the left or right footsteps, the network also performs dimensionality reduction.

E. Model Evaluation

1) One-Vs-One SVM model: The 127-length feature vector obtained from the CNN model per dataset sample is consequently used in a One-vs-One linear SVM model as in a biometric verification scenario. The model was trained per user in the training set and then evaluated in the validation and evaluation dataset as in [3] for the baseline benchmark. The returned probabilities scores by the SVM model are

separated into true and impostor scores to obtain Detection Error Trade-off (DET) [7] curves and equal error rate (EER) as the experiment's performance metrics. The DET curves are expressed as the False Acceptance Rate (FAR) versus False Recognition Rate (FRR).

V. Results

Results are shown per right, left and concatenated footstep models for the validation and evaluation dataset of the baseline benchmark as presented in [3], that allowed to compare our analysis with previous work.

A. Performance analysis of baseline benchmark

For this experiment 40 One-vs-One linear SVM models were trained for 40 clients by using the features obtained from the CNN model. The dataset split distribution for training, testing and evaluation is presented in subsection III-B.

1) Validation dataset performance: For single footsteps a similar performance of 14.76% and 14.23% EER was obtained for left and right footsteps models respectively. The stride footstep model obtained a significant EER improvement of 9.392% when compared to single footsteps. The DET curves of the validation dataset performance are shown in Figure 3.

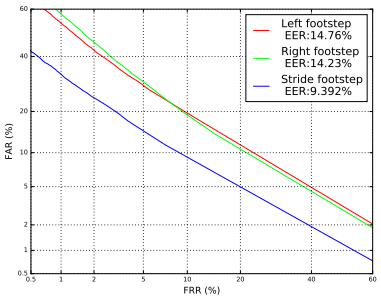

Fig. 3: DET curves for validation dataset

2) Evaluation dataset performance: An EER of 21.30% was obtained for the left footstep, EER of 20.23% for the right footstep and EER of 13.86% for the stride footstep by testing the trained model in the evaluation dataset. A significant EER improvement was obtained when considering stride footsteps as in the validation dataset experiment. The DET curves of the evaluation dataset performance are shown in Figure 4.

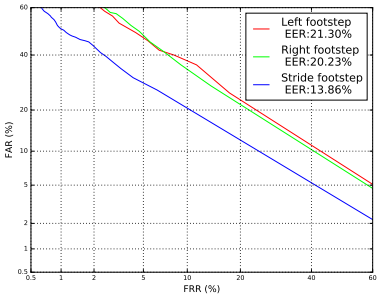

Fig. 4: DET curves for evaluation dataset

The difference in performance of the validation and evaluation datasets is due to the dataset sizes and the model's generalization error. In addition, the footstep signals used for the evaluation set were sampled with a large time gap with the training footstep signals in contrast to the ones used for the validation dataset. For both the validation and evaluation datasets, the EER was improved for stride footsteps. An EER improvement of 5% and 8% was obtained respectively for the validation and evaluation set when compared to single footstep signals.

The best footstep recognition performance obtained with our methodology was for the stride footstep with an EER of 9.392% and 13.86% for the validation and evaluation datasets respectively for the baseline benchmark. This significantly improves previously reported [3] EER performance of 10.56% and 16% for the validation and evaluation datasets respectively.

One of the limitations of the SfootDB database is that only one stride signal is available per footstep experiment. As our experiments have shown, it is evident that if a longer gait cycle is considered, better footstep recognition can be obtained.

VI. Conclusion and Future Work

We have presented a CNN model to learn spatial footstep features from a floor sensor system. Our approach obtained an EER score of 9.392% and 13.83% for the validation and evaluation datasets. This improves previously reported EER of 10.56% and 16% for the validation and evaluation datasets respectively [3] in the spatial domain.

We argue that the performance improvement of our CNN model feature extraction approach (Figure 2) is due to its ability to learn automatically a proper set of spatial features from the footstep images provided (Figure 1). This is in contrast with previous work [3] were a nonlinear SVM model was proposed for footstep recognition from spatial pixel values.

As future work, deep learning architectures will be investigated for their ability to learn spatio-temporal features from the SfootdB database, in an attempt to improve the presented spatial footstep recognition rates. This in order to provide a better model suited for footstep recognition by using only floor sensor systems.

References

[1] R. Vera-Rodriguez, J. S. Mason, J. Fierrez, and J. Ortega-Garcia, "Comparative analysis and fusion of spatiotemporal information for footstep recognition," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 35, no. 4, pp. 823–834, 2013.

[2] M. Ziefle, S. Himmel, and W. Wilkowska, *When your living space knows what you do: Acceptance of medical home monitoring by different technologies*. Springer, 2011.

[3] R. Vera-Rodriguez, J. S. Mason, J. Fierrez, and J. Ortega-Garcia, "Analysis of spatial domain information for footstep recognition," *Computer Vision, IET*, vol. 5, no. 6, pp. 380–388, 2011.

[4] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[5] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *arXiv preprint arXiv:1512.03385*, 2015.

[6] Y. N. Dauphin, H. de Vries, J. Chung, and Y. Bengio, "Rmsprop and equilibrated adaptive learning rates for non-convex optimization," *arXiv preprint arXiv:1502.04390*, 2015.

[7] A. Martin, G. Doddington, T. Kamm, M. Ordowski, and M. Przybocki, "The det curve in assessment of detection task performance," DTIC Document, Tech. Rep., 1997.