Walking Through the Deep: Gait Analysis for User Authentication Through Deep Learning

Giacomo Giorgi, Fabio Martinelli, Andrea Saracino, Mina Sheikhalishahi

HAL Id: hal-02023725
https://inria.hal.science/hal-02023725

Submitted on 21 Feb 2019

Walking Through the Deep: Gait Analysis for User Authentication through Deep Learning

Giacomo Giorgi[1]✉, Fabio Martinelli, Andrea Saracino, Mina Sheikhalishahi
`firstname.lastname@iit.cnr.it`

Istituto di Informatica e Telematica, Consiglio Nazionale delle Ricerche

Abstract. Seamless authentication is a desired feature which is becoming more and more relevant, due to the distribution of personal and wearable mobile devices. With seamless authentication, biometric features such as human gait, become a way to control authorized access on mobile devices, without actually requiring user interaction. However, this analysis is a challenging task, prone to errors, with the need to dynamic adapt to new conditions and requirements, brought by the dynamic change of biometric parameters. In this paper we present a novel deep-learning based framework for gait-based authentication. The paper presents an in depth study of the building and training of a Recurrent Convolutional Neural Network with a real dataset based on gait reading performed through five body sensors. We introduce methodologies to further increase the classification accuracy based on data augmentation and selective filtering. Finally we will present a complete experimental evaluation performed on more than 150 different identities.

Keywords: Gait recognition · Behavioral analysis · Seamless continuos authentication · Deep learning.

1 Introduction

Wearable technology is advancing at a fast pace, with a large interest in industrial and research world. More and more additional computing capacity and sensors are incorporated into smartphones, tablets, (smart)watches, but also shoes, clothes, and other wearable items. These enhanced objects act as enablers of pervasive computing [12], collecting data used to provide additional smart services to their users.Several of these smart devices come equipped with built-in accelerometers and gyroscopes, which can be exploited to register the body motion of users [11], [10], monitoring the unique movement pattern of a user might become an excellent instrument for seamless authentication [9]. However, the majority of current solutions for sensor based authentication, are mainly based on active behavioral mechanisms, which require direct user interaction [6], having thus limited advantages compared to classical authentication mechanisms, such as PIN, passwords, or finger pattern recognition. Considering that each individual person has a unique manner of walking, *gait* can be interpreted as a biometric trait and consequently, the aforementioned inertial sensors have great potential

to play an important role in the field of biometry [9]. If correctly exploited, the gait can be used as a method for seamless continuous authentication, able to authenticate users of wearable devices continuously during time, without requiring any active interaction. In this paper, we present a deep study on gait analysis for identity recognition based on inertial sensors and deep learning classification. The presented methodology exploits a public dataset [17] collected on a set of 175 users through five body sensors, presenting the design and implementation of a recurrent convolutional neural network for deep learning-based classification. Through experimental evaluation, we show the effectiveness of the methodology in recognizing single user on which the recurrent convolutional network has been trained on and also the ability of the presented system to understand if the monitored gait belongs to an unknown person. The results show an accuracy close to 1, demonstrating the feasibility of the presented approach as a methodology for seamless continuous authentication, which can be exploited by mobile and wearable smart devices. The contributions of this paper are:

- The presentation of a deep-learning method based on Recurrent-Convolutional-Neural-Network for identity classification through gait analysis;
- An analysis of the sensor orientation problem solved considering the magnitude of the 3 axis acceleration vectors;
- It will be detailed the process of designing the network and consecutive training based on analysis of a variable number of body sensors, implemented through the *Keras* framework for deep learning;
- The paper will introduce a methodology for data augmentation which aims at increasing the classification accuracy and prevent overfitting, by generating well instrumented artificial data;
- An extended analysis of the gait recognition will be discussed by proposing a threshold-based method to filter out outliers and increase the overall accuracy. Moreover a study on sensor filtering to demonstrate the high recognition accuracy with less sensor will be presented;
- A study of cross session classification to understand the capability of the network to learn the different walking patterns of a person is also introduced;

   This paper extends the one presented in [2] by presenting as new contributions (i) the introduction of a new deep learning network based on Recurrent Convolutional Neural Network, which improves the accuracy in recognizing identities, (ii) the application of a sensor data orientation invariance method (iii) the sensor filtering approach and related experiments, (iv) a study on cross session classification. The rest of the paper is organized as follows: Section 2 reports background notions on gait analysis and deep learning. Section 3 describes the used dataset and the data preprocessing steps. Section 4 reports the description of the design and implementation of the network and its training metodology. Section 5 reports the classification results for different scenarios and reducing the number of sensors consideres. Section 6 lists some related work. Finally Section 7 briefly concludes proposing some future directions.

2   Background

In this section we present some background notions exploited in this work.

2.1   Gait Analysis

Gait is the motion of human walking, whose movements can be faithfully reflected by the acceleration of the body sections [17]. Human gait recognition has been recognized as a biometric technique to label, describe, and determine the identity of individuals based on their distinctive manners of walking [14]. Basically, due to the fact that walking is a daily activity, human gait can be measured, as a user identity recognition technique, in daily life without explicitly asking the users to walk. This fact distinguishes gait from other accelerometer measurable actions, like gestures, as well as other commonly used biometrics, such as fingerprints, signatures, and face photos, whose data collection usually interrupts the users from normal activities for explicit participation [17]. Moreover, since portable or wearable accelerometers are able to monitor gait continuously during arbitrary time period, accelerometer-based gait recognition would be especially great tool in continuous identity verification [7].

2.2   Deep Learning

A neural network is a class of machine learning algorithms, in which a collection of *neurons* are connected with a set of *synapses*. The collection is designed in three main parts: the input layer, the hidden layer, and the output layer. In the case that neural network has multiple hidden layers, it is called *deep network*. Hidden layers are generally helpful when neural network is designed to detect complicated patterns, from contextual, to non obvious, like image or signal recognition. Synapses take the input and multiply it by a weight, where it represents the strength of the input in determining the output [3]. The output data will be a number in a range like 0 and 1. In forward propagation, a set of weights is applied to the input data and then an output is calculated. In back propagation, the margin of error of the output is measured and then the weights accordingly are adjusted to decrease the error. Neural networks repeat both forward and back propagation until the weights are calibrated to accurately predict an output [3]. Network with many layers and many neurons that are fully connected can become computationally infeasible to train.

Convolutional Neural Networks Convolutional Neural Networks (CNN), is born with the task to reduce the number of parameters to train, limiting the number of connections of the neurons in the hidden layer to only some of the input neurons. This connections are called local receptve field of the convolutional layer and the weights of that region are shared. Each group is generally composed by as many neurons as needed to cover the entire input data. This way, it is as if each group of neurons in the hidden layer calculated a convolution of the input data with their weights. The results of the convolution is a feature. Commonly a pooling layer is applied to the result of the convolutional

layer. It permits to provide translation invariance of the learned features and to reduce the dimensionality of the neurons. The result is a smaller version of the input features. These steps can be repeatedly applied as many times as desired: a new convolutional layer can be applied on the pooled layer, followed by another pooling layer, and so forth. The majors advantages of the CNN are the reduction of the network parameters thanks to sharing weights and the automatic features extraction at different semantic concept layer from the lower to the higher level representation, which provides a better data representation than the hand crafted feature descriptor. Recently, CNNs have been used as very powerful technique to solve and advanced the state-of-the-art accuracies in computer vision tasks such as face recognition [19], object recognition [20].

Recurrent Neural Networks Recurrent Neural Networks (RNNs) are successfully applied to model sequential informations such as speech recognition [21], language translation [22] and so on. Different from the traditional neural networks it assumes the values of the input sequence dependent between them. RNNs perform the same computation for every element of the input sequence and the output is dependent on the previous computation. Bidirectional RNNs is a variant of RNN based on the idea that the output at a certain time is dependent not only on the previous element but also on the future element of the sequence.

3   DATASET DESCRIPTION AND PROCESSING

In what follows, we present in detail the dataset description and the preprocessing steps done in order to prepare the data for the classification process.

3.1   Dataset Description

In this study, we exploit the ZJU-gaitAcc dataset that is public available and described in [8]. This dataset contains the gait acceleration series of records collected from 175 subjects. Out of these 175 series, we consider the records related to 153 subjects, which are divided in two sessions, such that the first session represents the first time that data has been collected, while the second session shows the second time that the data has been recorded. For the remaining 22 identities only a single session has been recorded, hence they have been discarded for the classification task, but they have been considered as unknown subjects to estimate the ability of the network to understand if a monitored gait belongs to an unknown person. For each subject, the time interval between first and second data acquisition varies from one week to six months. For each subject, six records are presented in each session, where every record contains 5 gait acceleration series (Normally composed by 7-14 full step cycles) simultaneously measured with 5 sensor placed on the right wrist, left upper arm, right side of pelvis, left thigh, and right ankle as depicted in Fig. 1.

The acceleration readings have been measured at $100\ Hz$ in straightly walks, through a level floor of $20m$ length. The raw data for each recording are composed by the x, y and z acceleration series during time.

Fig. 1: Body sensors

3.2   Data Processing

The data processing part can be summarized in four main steps, namely *cycles extraction*, *filtering*, and *normalization*, respectively. In addition in this section is described a data augmentation process in order to generate more sintetic data to improve the accuracy and prevent overfitting.

Cycles Extraction. The *gait cycle* is used to simplify the representation of complex pattern of human walking. It starts with initial contact of the right heel and it continues until the right heel contacts the ground again. When the heel touches the ground, the association between the ground reaction force and inertial force make the $z-axis$ signal strongly sensitive to change, forming peaks with high magnitude. Those peak points are utilized to identify the gait cycles. The ZJU dataset provides the manual annotations of the step cycles.

Cycles Normalization. Each gait cycle differs in terms of duration, due to the different speed which varies during walking, but not in shape. In ZJU dataset, the majority of cycles are constituted by a length between 90 and 118 samples. The features extraction phase performed by the CNN requires in input a fixed number of samples for each gait cycle. For this reason each gait cycle is normalized to a length of 118 samples through linear interpolation [4].

Noise reduction. The data collected from accelerometer sensors are affected by several noise sources due to the nature of the sensor. To reduce it and improving dataset quality, a filtering step is required. To this end a low pass *butterworth* filter [5] is applied to smooth the signal and remove high peaks.

Magnitude computation. Most gait recognition studies that employ wearable sensors, consider as unreal assumption that the position and the orientation of the sensors do not change over time. Altough the ZJU dataset not suffers of this problem, in order to reproduce a more realistic result, we applied a sensor data transformation to remove the effect of sensor orientation from the raw sensor data. To this end, instead to consider, as input of the network, the 3-axis accelerometer vectors, we simply considered the magnitude of the acceleration vectors computed as the euclidean norm as: $magnitude = \sqrt{x^2 + y^2 + z^2}$.

Data augmentation In order to improve the performance of the deep learning network and to prevent overfitting, we have artificially increased the number of training examples by data augmentation. It is the application of one or more deformations applied to the labeled data without change the semantic meaning of the labels. In our case, the augmentation is produced varying each signal sample with translation drawn from a uniform distribution in the range $[-0.2, 0.2]$. The process produces a copy of the original gait cycle different in values but with an

equal semantic of the walking cycle. Starting from, approximately 95 gait cycles per identity, with augmentation we reached 190 gait cycle per identity, passing from 14.573 training data to 29.146.

4   DATA ANALYSIS

In this section we describe the design and implementation of the recurrent convolutional neural network, how it has been trained and the metrics used to evaluate the proposed method are given.

4.1   Network description and Training

In this paper, we proposed a deep neural network architecture applied to the problem of gait classification of 153 persons. Given a gait cycle, the task is to determine to which person the cycle belongs. We designed and implemented two network architectures suitables respectively for single sensor and multiple sensors experiments. As reported in Figure 2, both the single sensor and the multiple sensors architectures are based on the same core. It extracts, from a single input gait cycle magnitude, features of two different abstraction level and applies a temporal aggregation on the features extracted in the second level. The first two level features are extracted automatically from the input data through two stacked 1D convolutional layers, which compute respectively 128 and 256 features vectors with kernel size 2 and 3. The second features vectors level is passed to a bidirectional recurrent layer based on Gated Recurrent Units (GRU) [1] with 256 neurons. It produces a temporal aggregation feature vector that is passed to two different pooling layers, which compute a feature subsampling respectively using the average and the maximum pooling. The functionality of that layer is to reduce the spatial size of the representation reducing the amount of parameter to train. The final result is the concatenation of the pooling results that represent the feature vector extracted from the input gait cycle. Each one of the convolutional layer output are passed through a batch normalization layer to regularize the model then trhough a Rectified Linear Unit (ReLu). In the single sensor scenario, the feature vector extracted are passed directly to a fully connected classifier containing 153 softmax units which compute the probability of the input gait cycle to belong to a specific subject. Thus, in the multiple sensor scenario, the feature vector is calculated for each sensor, aggregated with a concatenation operation and finally passed to the fully connected classifier. Our recognition problem is posed as a classification problem. Training data are groups of accelerometer data labeled with the owner identity. The optimization objective is average loss over all the identities in the data set. The loss is used in backpropagation steps, in order to update the weights. We used Adam optimization algorithm [23] to update network weights iterative based in training data. We start with a base learning rate of 0.001 and gradually decrease it as the training progresses. We use a momentum of $\mu = 0.9$ and weight decay $\lambda = 5 \cdot 10^{-4}$. With more passes over the training data, the model improves until it converges.
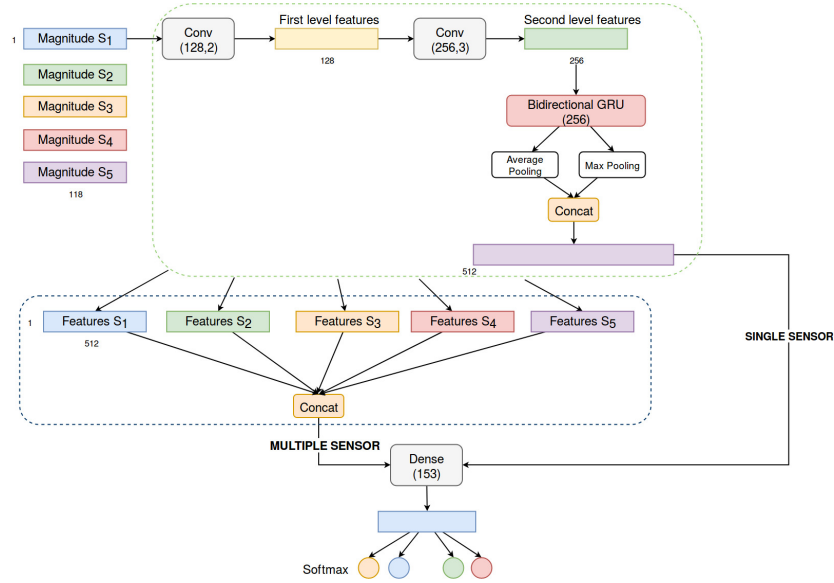
Fig. 2: Recurrent Convolutional gait recognition network

The hyperparameter tuning (number of epochs, learning rate, number of layers, number of neurons per layer) is made through a manually search of the best hyperparameter settings. Using knowledge you have about the problem guess parameters and observe the result. Based on that result tweak the parameters. Repeat this process until you find parameters that work well or you run out of time.

4.2 Evaluation metrics

Gait recognition is the process of assign a given waling gait pattern to its own identity. In our case we consider as gait pattern a walking step cycle. As described in Section 4.1, the gait recognition network returns a probability vector, which reports the belonging probabilities of the given gait cycle for each subject class. Sorting the resulting probability vector, can be possible determine in which rank the given gait cycle has been assigned. Thanks to that it is possible, not only compute the recognition accuracy on the first rank (1-rank), but also the accuracy in recognizing the identity within the top k-ranks. More in detail the recognition accuracy of unseen walking record at 1-rank is given by the number of step cycles correctly identified at 1-rank, divided by the total number of step cycles in the walking record. In the same manner the accuracy at 2-rank is given by the number of step cycles correctly identified at 1-rank or 2-rank divided by the total number of step cycles in the walking record. In the same way is computed the accuracy for the rest of ranks. In order to evaluate the recognition capacity of the method presented, the following metrics are introduced (i)

the overall recognition accuracy of unseen walking record on the 1-rank; (ii) the overall recognition accuracy of unseen walking record on the top k ranks, with $1 < k < 153$. Finally another important statistics to consider are the mean value and the standard deviation of the correct and wrong probabilities predicted at 1-rank: $P_{T1rank}, StDev_{T1rank}, P_{F1rank}, StDev_{F1rank}$. They provide us an estimation of the difference between the resulting probabilities of the corrects and wrong predictions that can be useful to determine the probability threshold as discussed in Section 5.1.

5   EXPERIMENTAL ANALYSIS

In this section we report the description and results of the performed experiments to evaluate the effectiveness of the proposed methodology. It has been analyzed the accuracy of the gait recognition network on the ZJU dataset in cross and single session, first, considering all the sensors, then reducing the number of sensors considered. Finally we proposed metodologies to improve the overall accuracy based on data augmentation and filtering threshold.

Single session. In this scenario, we explored the capacity of the network in recognize a subject in one single session (walking gaits recorded in the same day). To this end, we consider only session 1 splitting the data as reported in Figure 3(a). For each subject we considered the first five walking records as training set and the sixth walking record as testing set. This setting is better suited for training because it uses 80% of dataset, 5.850, for training (about 38 gait cycle per identities), and roughly 20% for testing, 1.453 testing samples. After the augmentation, which is applied only on the training set, the number of training samples becomes 17.550 (about 114 gait cycle per identities). Figure 4(a), shows the CMC curve that reports the recognition accuracy for the single session scenario using 5 sensors. At 1-rank the accuracy is 99.06% with augmentation and 98.86% without augmentation. Furthermore an accuracy of 100% at rank 17 and 28 is achieved respectively for the augmentation and non-augmentation experiments.

Cross session. In this scenario, we explored the capacity of the network to learn different walking pattern of the same user. To this end we considered the two sessions recorded over time. As reported in Figure 3(b) we split the data considering the first five walking records of both sessions as training set and the last walking record of both sessions as testing set. The total amount of data is 11.748 for the training set before augmentation and 2.933 for the testing set.

Figure 4(b), shows the CMC curve that reports the recognition accuracy for the cross session scenario using 5 sensors. At 1-rank the accuracy is 98.70% with augmentation and 97.50% without augmentation. We obtain an accuracy of 100% at rank 20 and rank 35 respectively for the augmentation and non augmentation experiment.
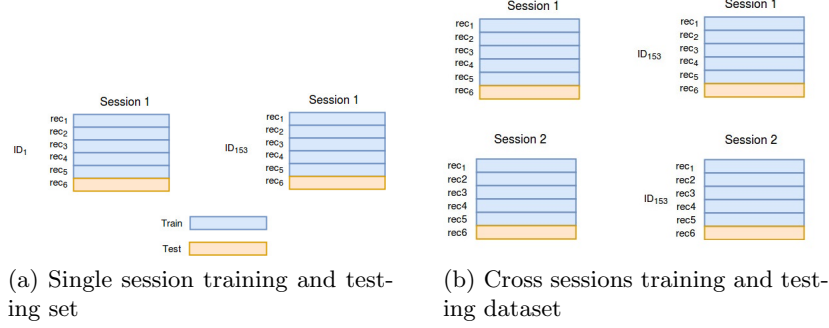
(a) Single session training and testing set

(b) Cross sessions training and testing dataset

Fig. 3: Training and testing sets

(a) CMC single session - 5 sensors

(b) CMC cross session - 5 sensors
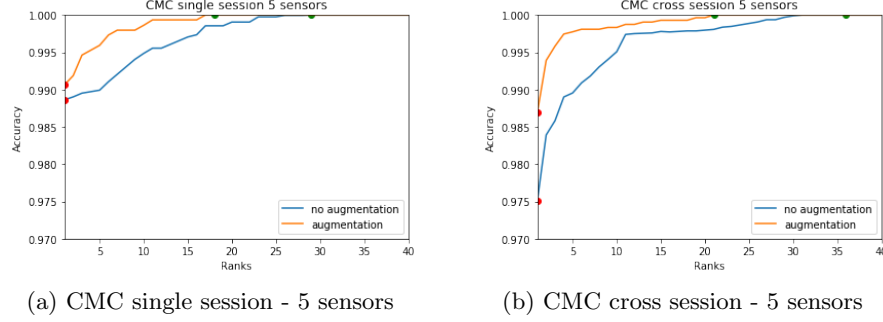
Fig. 4: 5 sensors results.

5.1 Sensor Filtering

As an additional set of experiments, we have evaluated the accuracy results by considering different subsets of the five initial sensors. Conducting this type of experiments we evaluated the behavior of the proposed method in recognizing gait cycles in a less intrusive way. Considering the powerful of the data augmentation in improving accuracy as demonstrated in Figure 4, the results reported for the cross filtering refer only to data augmentation case.

Single session filtering experiments As first experiment we explored the capacity of recognition exploiting a single sensor. The Figure 5(a), shows the CMC curves comparison between the 5 sensors taken standalone in the single session scenario. The best accuracy result is given by the sensor S3 (right side of the pelvis) with 88.75% of recognition accuracy at 1-rank. The obtained results provide us the criterion to create the multi sensors experiments selecting the most promising sensors subsets. The following sensors combinations have been tested: S3-S2 (right side of the pelvis-left upper arm), S3-S2-S4 (right side of the pelvis-left upper arm-left thigh) and S3-S2-S4-S1 (right side of the pelvis-left upper arm-left thigh-right wrist). In addition to this combinations has been tested the sensors combinations S3-S1 (right side of the pelvis-right wrist) that reflect

(a) CMC single session - single sensor

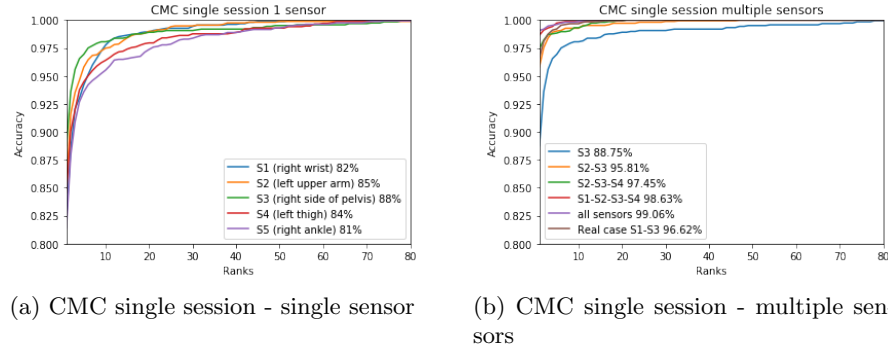(b) CMC single session - multiple sensors

Fig. 5: Single session results.

a real case scenario, combining sensor located on the right wrist representing a smartwatch, and another one located on right side of pelvis representing a smartphone kept in the front pocket. The CMC curve is plotted in Figure 5(b). As aspected, increasing the number of sensors, the 1-rank accuracy increase considerably. The 1-rank accuracy in the real case scenario is 96.62% reaching 100% of accuracy at rank 21.

Cross session filtering experiments The same experiments have been conducted in the cross session scenario. Figure 6(a), shows the CMC curves comparison between the five sensors taken standalone.

(a) CMC cross sessions - single sensor
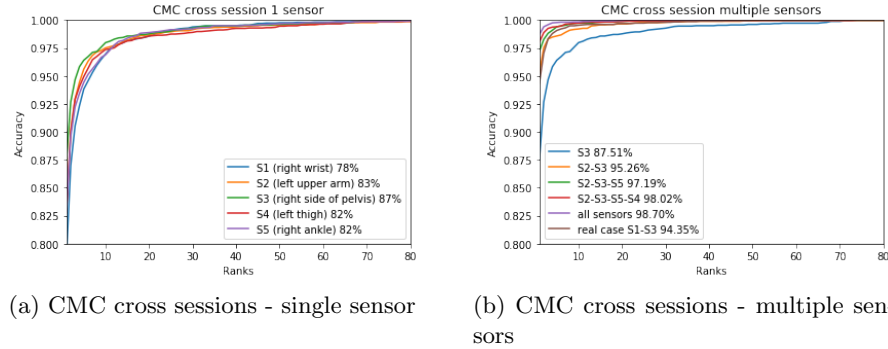
(b) CMC cross sessions - multiple sensors

Fig. 6: Cross session results.

The 1-rank accuracies are slightly lower than the single session scenario due to the fact that the network has to associate different gait patterns to the same identity. However, the most promising sensor remains the right side of pelvis with 87.51% of recognition accuracy at 1-rank. On the basis of single sensor accuracy, we tested the following sensors combinations: S3-S2 (right side of the

pelvis-left upper arm), S3-S2-S5 (right side of the pelvis-left upper- right ankle) and S3-S2-S5-S4 (right side of the pelvis-left upper- right ankle-left thigh). The CMC curve is reported in Figure 6(b). The 1-rank accuracies are approximately 0.60% lower respect to the single sensor scenario. The 1-rank accuracy in the real case is 94.35% with 100% accuracy at rank 68. The differences between the single and cross session scenario is valuable only increasing the 1-ranks. In fact 100% accuracy is reached at very high ranks in the case of cross session respect to the single one. This is due to the fact that in the cross session the network assignes very low probabilities to the false negative gait cycles.

Threshold method based Another important statistics to considering are the mean value probability and its standard deviation of correct and wrong predictions (P_{T1rank},$StDev_{T1rank}$,P_{F1rank},$StDev_{F1rank}$). We computed this values for the sensors combinations reported in the Table 1.

Table 1: Mean probability and standard deviation of correct (TP) and wrong (FP) prediction

Scenario	Sensors	P_{T1rank}	$StDev_{T1rank}$	P_{F1rank}	$StDev_{F1rank}$
Single session	5 sensors	0.99	0.03	0.59	0.19
	real case	0.98	0.06	0.65	0.20
Cross session	5 sensors	0.99	0.04	0.54	0.20
	real case	0.97	0.08	0.62	0.20

Since the probabilities for true positives is much higher than the one for false positives, it is possible to set a probability threshold to distinguish these two values. This leads to improve the overall recognition accuracy. In fact, setting a probability threshold equal to the mean probability of the correct prediction minus its mean standard deviation as: $threshold = P_{T1rank} - StDev_{T1rank}$ and filtering out all values lower than the threshold, grants a recognition accuracy of 100%.

5.2   Unknown identities recognition

The proposed method, is only able to classify identities on which it has been trained on. Hence, if presented with a set of steps coming from an unknown identity, the Recurrent Convolutional Neural network will try to match the new gait with a known one. However, we argue that is still possible exploiting the RCNN to understand if a set of steps is belonging to an unknown identity rather than to a known one. It is worth noting that such a feature would be useful in the design of anti-theft applications for mobile and wearable devices. To this scope, we exploited the 22 unknown identites presented on the dataset to only one session and we measured the mean probability of the false positive prediction in cross and single session scenario. Table 2 shows that. It is evident as the predicted probabilities of the unknown identities is highly lower than the known

Table 2: Mean prediction probability and std-dev for unknown identities

Scenario	Sensors	P_{F1rank}	$StDev_{F1rank}$
Single session	5 sensors	0.75	0.15
	real case	0.73	0.23
Cross session	5 sensors	0.78	0.13
	real case	0.76	0.12

Table 3: False positive and False negative varying threshold

Scenario	Sensors	Threshold	FP	FN
Single session	5 sensors	0.98	0	1.02
	real case	0.96	0.07	2.25
Cross session	5 sensors	0.98	0.12	1.54
	real case	0.96	0.91	2.13

ones. Imposing again a probability threshold, we obtain a limitating error in recognizing unknown identities as known. The results reported in Table 3, shows the variation of the False Positive (known gait cycles, classified as unknown) and False Negative per identities (unknown gait cycles, classified as known) varying threshold value.

6   Related Work

In [15] a two-phase view-invariant multiscale gait recognition method (VI-MGR) is proposed, which is robust to variation in clothing and presence of a carried item. In phase 1, VI-MGR uses the entropy of the limb region of a gait energy image (GEI) to determine the matching gallery view of the probe using 2-dimensional principal component analysis and Euclidean distance classifier. In phase 2, the probe subject is compared with the matching view of the gallery subjects using multiscale shape analysis. In [16], the three types of sensors, color sensors, depth sensors and inertial sensors, are combined for gait data collection and gait recognition, which can be used for important identification applications, such as identity recognition to access a restricted building or area. Being based on deep learning, the accuracy of our framework is increased if the training is performed with a larger and diverse dataset. However, real data collection could be an issue which also brings privacy concerns. In [18] a framework for privacy preserving collaborative data analysis is presented, which could be exploited by our framework to increase the accuracy, without violating users' privacy. In [24], an accelerometer-based gait recognition, named *iGait*, is proposed. The core function of iGAIT extracts 31 features from acceleration data, including 6 spatio-temporal features, 7 regularity and symmetry features, and 18 spectral features. The proposed framework has been used to analyze the gait pattern of 15 control subjects, where a (HTC) phone was attached to the back of participants by belts. In each trial, participants walked 25m along a hallway at their preferred walking speed. The first advantage of our approach comparing to what is proposed by Yang et. al [24] is that deep-learning-based approaches learn features gradually. Hence, our methodology finds the most discriminating features through self training process. The second advantage is related to time needed to reach to 100% accuracy. In our approach 10 steps is enough to identify a person while in [24] 25 minutes walk is required. At the end, the proposed approach in [24] is evaluated through 15 subjects, whilst our technique is evaluated through 153 persons. The accelerometer-based gait recognition approach proposed in [17]

is evaluated on the same dataset we exploited in our experiments. The work, first consider the problem of step-cycle detection which suffer from failures and inter-cycle phase misalignment. To this end, an algorithm is proposed which makes use of a type of salient points, named signature points (SPs). Experimental results on the equivalent dataset of our experiment shows 1-rank accuracy of 95.8% for identification and the error rate of 2.2% for user verification. However, this accuracy is obtained on 14 steps, while in our proposed approach 100% is achieved in 10 steps.

7   Conclusion and Future Work

Gait analysis is an enabling technology for seamless user authentication, still it requires fast, accurate and flexible mechanism for an effective classification. In this paper we have presented a classification methodology based on deep learning, to perform accurate user recognition through gait analysis. The reported accuracy on the considered dataset made of more than 150 identities, has proven to be extremely precise, especially when to the standard classification process, we apply sensor filtering, data augmentation and threshold based analysis. Furthermore, we have demonstrated that the present approach is effective in recognizing users in a plausible use case where only sensors representing smartphone and smartwatch have been used, i.e. the authentication process does not require the presence of additional sensors whose only task is to perform the identification, instead it is integrated in popular personal items. As future work, we plan to consider a real use case, where the framework is directly installed on personal devices and the training and classification are performed at runtime. In addition, in order to obtain a more general architecture for authentication, we plan to explore a siamese neural network architecture training it starting from the features extracted by the network presented.

Acknowledgements

References

1. Chung, J., Gulcehre, C., Cho, K., Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. arXiv preprint arXiv:1412.355
2. Giorgi, Giacomo, et al. "Try Walking in My Shoes, if You Can: Accurate Gait Recognition Through Deep Learning." International Conference on Computer Safety, Reliability, and Security. Springer, Cham, 2017
3. LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. "Deep learning." nature 521.7553 (2015): 436
4. Coursey, Cameron K., and John A. Stuller. "Linear interpolation lattice." IEEE Transactions on signal processing 39.4 (1991): 965-967.

5. Van Vollenhoven, E., H. Reuver, and J. Somer. "Transient response of Butterworth filters." IEEE Transactions on Circuit Theory 12.4 (1965): 624-626.
6. Buriro, Attaullah, et al. "Hold and sign: A novel behavioral biometrics for smartphone user authentication." Security and Privacy Workshops (SPW), 2016 IEEE. IEEE, 2016.
7. Gafurov, Davrondzhon, Patrick Bours, and Einar Snekkenes. "User authentication based on foot motion." Signal, Image and Video Processing 5.4 (2011): 457.
8. Zhang, Yuting, et al. "Accelerometer-based gait recognition by sparse representation of signature points with clusters." IEEE transactions on cybernetics 45.9 (2015): 1864-1875.
9. Sprager, Sebastijan, and Matjaz B. Juric. "Inertial sensor-based gait recognition: a review." Sensors 15.9 (2015): 22089-22127.
10. Bao, Ling, and Stephen S. Intille. "Activity recognition from user-annotated acceleration data." International Conference on Pervasive Computing. Springer, Berlin, Heidelberg, 2004.
11. Ren, Yanzhi, et al. "User verification leveraging gait recognition for smartphone enabled mobile healthcare systems." IEEE Transactions on Mobile Computing 14.9 (2015): 1961-1974.
12. Wu, Zhaohui, and Gang Pan. Smartshadow: models and methods for pervasive computing. Springer, 2013.
13. Muramatsu, Daigo, Yasushi Makihara, and Yasushi Yagi. "View transformation model incorporating quality measures for cross-view gait recognition." IEEE transactions on cybernetics 46.7 (2016): 1602-1615.
14. Alotaibi, Munif, and Ausif Mahmood. "Improved gait recognition based on specialized deep convolutional neural network." Computer Vision and Image Understanding 164 (2017): 103-110.
15. Choudhury, Sruti Das, and Tardi Tjahjadi. "Robust view-invariant multiscale gait recognition." Pattern Recognition 48.3 (2015): 798-811.
16. Zou, Qin, et al. "Robust gait recognition by integrating inertial and RGBD sensors." IEEE transactions on cybernetics 48.4 (2018): 1136-1150.
17. Zhang, Yuting, et al. "Accelerometer-based gait recognition by sparse representation of signature points with clusters." IEEE transactions on cybernetics 45.9 (2015): 1864-1875.
18. Martinelli, Fabio, Andrea Saracino, and Mina Sheikhalishahi. "Modeling privacy aware information sharing systems: A formal and general approach." Trustcom/BigDataSE/I SPA, 2016 IEEE. IEEE, 2016.
19. Parkhi, Omkar M., Andrea Vedaldi, and Andrew Zisserman. "Deep Face Recognition." BMVC. Vol. 1. No. 3. 2015.
20. Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." Advances in neural information processing systems. 2012.
21. Graves, Alex, Abdel-rahman Mohamed, and Geoffrey Hinton. "Speech recognition with deep recurrent neural networks." Acoustics, speech and signal processing (icassp), 2013 ieee international conference on. IEEE, 2013.
22. Cho, Kyunghyun, et al. "Learning phrase representations using RNN encoder-decoder for statistical machine translation." arXiv preprint arXiv:1406.1078 (2014).
23. Kingma, Diederik P., and Jimmy Ba. "Adam: A method for stochastic optimization." arXiv preprint arXiv:1412.6980 (2014).
24. Yang, Mingjing, et al. "iGAIT: an interactive accelerometer based gait analysis system." Computer methods and programs in biomedicine 108.2 (2012): 715-723.