Try Walking in My Shoes, if you can: Accurate Gait Recognition through Deep Learning

Giacomo Giorgi, Fabio Martinelli, Andrea Saracino, Mina Sheikhalishahi

Istituto di Informatica e Telematica, Consiglio Nazionale delle Ricerche, Pisa, Italy,
firstname.lastname@iit.cnr.it

Abstract. Human gait seamless continuous authentication, based on wearable accelerometers, is a novel biometric instrument which can be exploited to identify the user of mobile and wearable devices. In this paper, we present a study on recognition of user identity, by analysis of gait data, collected through body inertial sensors from 175 different users. The mechanism used for identity recognition is based on deep learning machinery, specifically on a convolutional network, trained with readings from different sensors, and on filtering and buffering mechanism to increase the accuracy. Results show a very high accuracy in both recognizing known and unknown identities.

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 [1], 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, which inspired the research interest in using the motion characters of human body for various tasks, spanning from clinical condition monitoring [2], action or gesture categorization [3], to user authentication and identity recognition.

In particular, accelerometer-based identity recognition with the use of body motion seems a promising technique in preventing the misuse of smart devices and the systems linked with them, by ensuring that the device functionalities are not available to other persons different from the owner. However, the majority of current solutions for sensor based authentication, are mainly based on active behavioral mechanisms, which require direct user interaction [4], 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 [5]. 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 study on gait analysis for identity recognition based on inertial sensors and deep learning classification. The presented methodology exploits a public dataset [6] collected on a set of 175 users through five body sensors, presenting the design and implementation of a convolutional neural network for deep learning-based classification. Moreover, it will be detailed the techniques used to filter and augment data for improving the dataset quality, together with the sampling techniques used to improve classification accuracy. Through experimental evaluation, we show the effectiveness of the methodology in recognizing single user on which the 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 rest of the paper is organized as follows: Section 2 reports background notions on the gait analysis and on deep learning. Section 3 describes the presented framework, the used dataset and the design of the convolutional neural network. Section 4 reports the classification results and describes the sampling methodology in operative phase for improved accuracy. Section 5 lists some related work. Finally Section 6 briefly concludes proposing some future directions.

2 Background Knowledge

In this section we present some background notions exploited in the present 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 [6]. 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 [7]. 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 [6]. 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 [8].

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* learning. 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 [9]. 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 [9].

3 Framework

In what follows, we present in detail the process of user identification through wearable accelerometer devices, with the use of deep convolutional neural network.

3.1 Dataset Description

In this study, we utilize the publicly available dataset provided in [10]. This dataset contains the gait acceleration series of records collected from 175 subjects. Out of these 175 series, we use the records of 153 subjects, which are reported 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. The time intervals between first and second data acquisition varies from one week to six months, for different subjects. For each subject, six records are presented in each session, where every record contains 5 gait acceleration series simultaneously measured at the right wrist, left upper arm, right side of pelvis, left thigh, and right ankle, respectively as we can see 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. Figure 2 reports the single readings in the time domain done on each axis for each accelerometer sensor, as specified on the right hand side of the picture.

3.2 Data Processing

The data processing part can be summarized in three main steps, namely *cycles extraction*, *filtering*, and *normalization*, respectively.

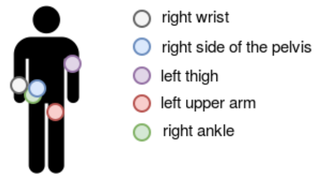


Fig. 1: Sensors considered on the body

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 hell and it continues until the right hell contacts the ground again. When the hell touches the ground, the association between the ground reaction force and inertial force, together make the $z - axis$ signal strongly to change, which forms peaks with the high magnitude. Those peak points are utilized to identify the gait cycles. The ZJU dataset provides the manual annotations of the step cycles of each gait record. 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 have lengths between 90 and 118. For reducing noise and improving dataset quality, the steps have been filtered through a low pass *butterworth* filter [11] and normalized through linear interpolation [12].

3.3 Network Definition and Training

One of the popular deep learning approaches is based on Convolution Neural Networks (CNN). The Deep CNN is an advanced machine learning technique which has inspired many researchers due to the achievement in the state of the art results in several applications of pattern recognitions. More precisely, a CNN is defined as the composition of several convolutional layers and several fully connected layers. Each convolutional layer is, in general, the composition of a non-linear layer and a pooling or sub-sampling layer to get some spatial invariance. Although deep CNN has been successfully used in several difficult pattern recognition problems, to the best of our knowledge, human gait recognition, when data is collected through accelerometer signals, has not been recognized with the use of Deep CNN. Thus, we propose a gait recognition approach based on a specific CNN network to approximate complex functions from high dimensional signals.

3.4 Identification Algorithm

In this paper, we propose a deep neural network architecture applied to the problem of gait identification of a dataset containing 153 persons. Given a gait cycle, the task is to determine to which person the cycle belongs. Our network architecture has been shown in Figure 4. The network consists of the following

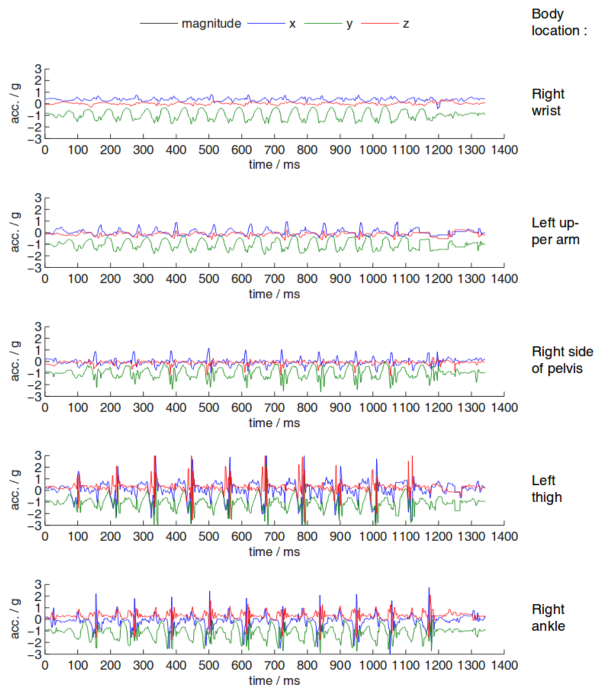


Fig. 2: Acceleration series of a gait record

layers: two layers of convolution, followed by a single max pooling and two fully connected layer followed by a softmax layer used to estimate the belonging class.

For the input layer, the different signals filtered and normalized belonging to each sensor are stacked in order to a matrix of dimension 15×118, in which each group of three rows correspond respectively to the x, y and z axes of a single sensor. 118 is the length of the normalized cycles.

The first two layers of our network are convolution layers, which we use to compute higher-order features. As shown in Figure 4, in the first convolution layer we pass the 2D input data of 15×118 through 128 learned filters of size (1×3) performing a 2D convolution on each single component of each sensor. This convolution is useful to extract the low level feature maintaining separated each component of each sensor. The resulting shape obtained after convolution is three-dimensional feature maps of size $12 \times 15 \times 61$.
In the second convolution the result obtained is passed to 512 convolutional filters of size (3×3) which perform a 2D convolution used to extract the mid level features combining the components within each sensor. $512 \times 5 \times 33$ is the resulting shape obtained. Each one of convolution layer are passed through a rectified linear unit (ReLu).
The second convolutional feature maps are passed through a max-pooling kernel that halves the width and height of features with filters of size (1×3). The

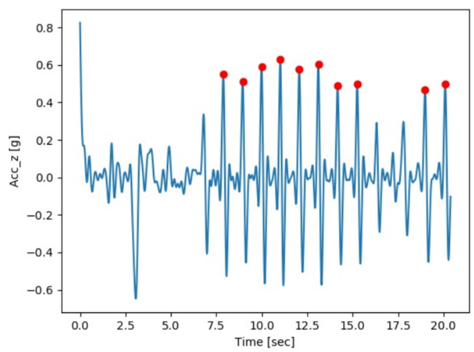


Fig. 3: Gait cycles of a walking record

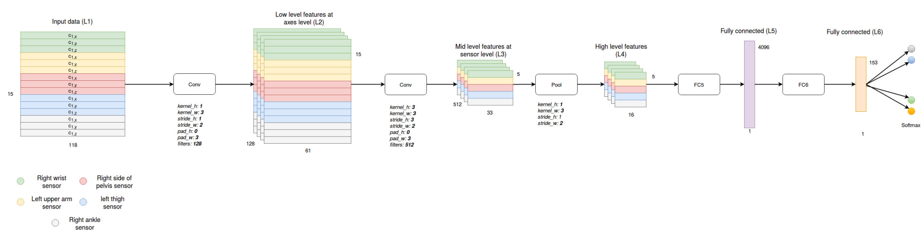


Fig. 4: Deep convolutional gait recognition network

functionality is to reduce the spatial size of the representation reducing the amount of parameters of the network.

Finally, we apply two fully connected layers: the first one uses 4096 nodes providing a feature vector of size 4096, these outputs are then passed to another fully connected layer containing 153 softmax units which represent the probability of similarity for each indentity to recognizes.

Training the Network. Our re-identification 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. As the data set can be quite large, in practice we use a stochastic approximation of this objective. Training data are randomly divided into mini-batches. The model performs forward propagation on the current mini-batch and computes the output and loss. Backpropagation is then used to compute the gradients on this batch, and network weights are updated. We perform stochastic gradient descent to perform weight updates. We start with a base learning rate of $\eta^{(0)} = 0.001$ and gradually decrease it as the training progresses using an inverse policy:$\eta^{(i)} = \eta^{(0)}(1 + \gamma \cdot i)^{-p}$. Where $\gamma = 10^{-4}$, $p = 0.75$, and i is the current mini-batch iteration. 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

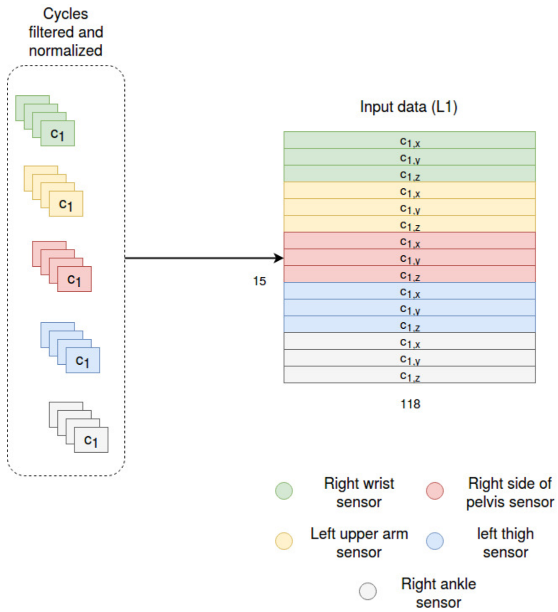


Fig. 5: Input layer

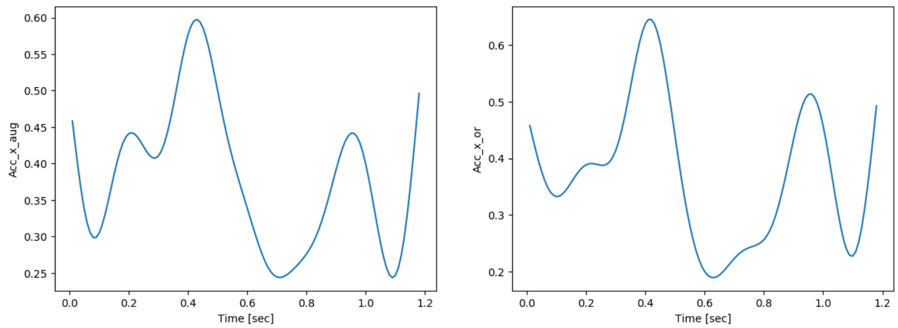
it converges. We use a validation set to evaluate intermediate models and select the one that has maximum performance.

Data augmentation In order to improve the performance of the deep learning network and to prevent overfitting, we have artificially increase the number of training examples by data augmentation. Data augmentation 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 $[-0.2, 0.2]$. As shown in Figure 6 this 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 reach until 190 gait cycle per identity, passing from 14573 training data to 29146.

Network training converges in roughly 30 minutes without augmentation and 1 hour with augmentation on NVIDIA GTX1050 GPU.

4 Experiments

The classification framework has been implemented through the *Caffe* [13] deep learning framework. The experiments are conducted considering one single session of the walking records. The session has six different walking records for each



(a) Original gait cycle (b) Augmented gait cycle

Fig. 6: Augmentation of a gait cycle

one of the 153 identities belonging to the dataset. We used the first five records of each identity for training and the last record for testing. This setting is better suited for deep learning because it uses 84% of the data for training, 7295 training samples (about 47 gait cycle per identities), and roughly 16% for testing, 1465 testing samples. After the augmentation the number of training samples becomes 21885 (about 142 gait cycle per identities). We use a mini-batch size of 120 samples and train the network for 4000 iterations.

The Figure 7 plots the re-identification accuracy of our model measured as the mean number of gaits recognized correctly for each identity in the first, second and third similarity results returned from the network. Our network reach 0.94% of accuracy in the most similar results returned without augmentation and the 0.95% with data augmentation. In particular the graph shows the accuracy in

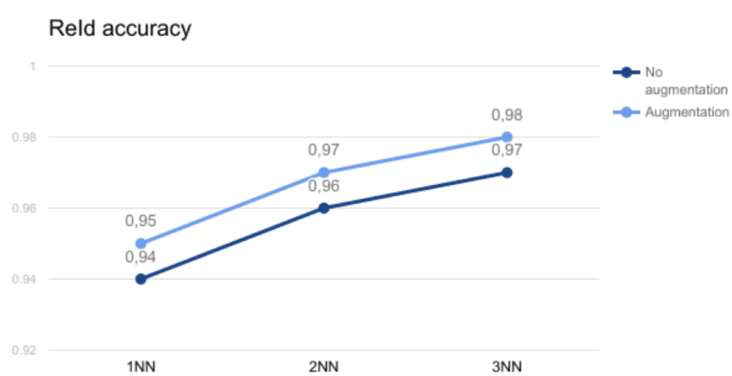


Fig. 7: Accuracy of reidentification.

correctly identifying the identity among the best (1NN), the two best matches (2NN) and the 3 best matches (3NN) extracted by the CNN. In the following only the 1NN case will be considered. It is worth noting the beneficial effect of augmentation for what concerns the accuracy.

Another important statistics to consider are the mean value and standard deviation of the probabilities of similarity related to the first result returned by the network. We computed these values for the true positives (gait cycles identified correctly) $\overline{P_{T1NN}}$, and false positives (gait cycles identified incorrectly) $\overline{P_{F1NN}}$, in a test set composed by 153 known identities with 10 steps each one. The results are reported in Table 1.

Table 1: Similarity classification for known identities.

	$\overline{P_{T1NN}}$	$StDev_{T1NN}$	$\overline{P_{F1NN}}$	$StDev_{F1NN}$
No augmentation	0.92	0.12	0.13	0.08
Augmentation	0.93	0.11	0.13	0.08

Hence, to correctly classify known identities, setting a probability threshold of 0.82, i.e. average similarity value with augmentation, minus the standard deviation, and filtering out all values lower than the threshold, grants an accuracy of 1 for the classification.

However, the CNN 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 CNN will try to match the new gait with a known one. However, we argue that is still possible exploiting the CNN 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. As shown in Table 2, the mean value of similarity for unknown identities is lower than the value for known ones, antecedently shown in Table 1. It is worth noting that in this case the overall accuracy is

Table 2: Similarity for unknown identities

	$\overline{P_{F1NN}}$	$StDev_{F1NN}$
No augmentation	0.82	0.18
Augmentation	0.76	0.22

lower when augmentation is considered. This is because, the altered steps added through the augmentation procedure, increase the likelihood of generating steps which might be similar to the ones of the unknown identities.

Considering ten gait cycles for each identity and imposing again a probability threshold, we obtain the average number of wrongly classified steps with respect to the known gait cycles, classified as unknown (i.e., False Positive) and unknown gait cycles classified as known (i.e. False Negative). The results are reported in Table 3 As shown, the error is quite limited, having slightly more

Table 3: False Positives and False Negatives on ten gait cycles per identity

Threshold	FP	FN
0.93	1.02	3.27
0.95	1.11	2.93

than 1 step out of 10 wrongly considered as unknown for an known identity, whilst for unknown identities slightly more than 3 steps out of 10 are wrongly classified as known. Hence, having a sampling window of 10 steps and exploiting a majority-based approach, is possible to filter away the classification error. It is also possible moving the error toward False Positive or False Negatives by changing the threshold value, in accordance to the application requirement.

5 Related Work

In [14], people are identified in video based on the way they walk (i.e. gait). To this end, convolutional neural networks (CNN) is applied for learning high-level descriptors from low-level motion features (i.e. optical flow components). The average accuracy of the result equals to 88.9 % . Muramatsu et. al. [15] authenticate a person through cross-view gait recognition which exploits a pair of gait image sequences with different observation views. In [16] 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 [17], the three types of sensors , i.e. color sensors, such as a CCD camera, depth sensors, such as a Microsoft Kinect, or inertial sensors, such as an accelerometer, 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 [19] a new method for recognizing humans by their gait using back-propagation neural network. Here, the gait motion is described as rhythmic and periodic motion, and a 2D stick figure is extracted from gait silhouette by motion information with topological analysis guided by anatomical knowledge. A sequential set of 2D stick figures is used to represent the gait signature that is primitive data for the feature extraction based on motion parameters. Then, a back-propagation neural network algorithm is used to recognize humans by their gait patterns.

In [10], an accelerometer-based gait recognition, named *iGait*, is proposed. The core function of iGAIT is exploited to extract 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 25 m along a hallway at their preferred walking speed. The first advantage of our approach comparing to what is proposed by Yang et. al [10] 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 [10] 25 minutes walk is required. At the end, the proposed approach in [10] is evaluated through 15 subjects, whilst our technique is evaluated through 175 persons.

The accelerometer-based gait recognition approach proposed in [6] is evaluated on the same dataset we exploited in our experiments. In this work, Zhang et. al. first addresses the problem of step-cycle detection which suffer from failures and intercycle 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 rank-1 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.

6 Conclusion and Future Work

We have presented in this paper a preliminary study performed by means of deep learning classification, to identify people exploiting the gait collected through inertial sensors. The presented results show that this approach is promising in finding a tool for seamless continuous authentication of users for mobile and wearable devices. As future work, we plan to apply the proposed methodology on data collected through smartphones and smartwatches, also measuring the similarity with the data acquired from 2 out of the 5 sensors considered in this work.

Acknowledgement

This work has been partially funded by EU Funded projects H2020 C3ISP, GA #700294, H2020 NeCS, GA #675320 and EIT Digital HII on Trusted Cloud Management.

References

1. Wu, Z., Pan, G.: SmartShadow: Models and Methods for Pervasive Computing. Springer Publishing Company, Incorporated (2013)
2. Ren, Y., Chen, Y., Chuah, M.C., Yang, J.: User verification leveraging gait recognition for smartphone enabled mobile healthcare systems. IEEE Trans. Mob. Comput. **14**(9) (2015) 1961–1974
3. Bao, L., Intille, S.S. In: Activity Recognition from User-Annotated Acceleration Data. Springer Berlin Heidelberg, Berlin, Heidelberg (2004) 1–17
4. Buriro, A., Crispo, B., Delfrari, F., Wrona, K.: Hold and sign: A novel behavioral biometrics for smartphone user authentication. In: 2016 IEEE Security and Privacy Workshops (SPW). (May 2016) 276–285
5. Sprager, S., Juric, M.B.: Inertial sensor-based gait recognition: A review. Sensors **15**(9) (2015) 22089–22127
6. Zhang, Y., Pan, G., Jia, K., Lu, M., Wang, Y., Wu, Z.: Accelerometer-based gait recognition by sparse representation of signature points with clusters. IEEE Transactions on Cybernetics **45**(9) (Sept 2015) 1864–1875
7. Alotaibi, M., Mahmood, A.: Improved gait recognition based on specialized deep convolutional neural networks. In: 2015 IEEE Applied Imagery Pattern Recognition Workshop (AIPR). (2015) 1–7
8. Gafurov, D., Bours, P., Snekkenes, E.: User authentication based on foot motion. Signal, Image and Video Processing **5**(4) (2011) 457
9. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature **521**(7553) (2015) 436–444
10. Yang, M., Zheng, H., Wang, H., Mcclean, S., Newell, D.: igait: An interactive accelerometer based gait analysis system. Comput. Methods Prog. Biomed. **108**(2) (November 2012) 715–723
11. van Vollenhoven, E., Reuver, H., Somer, J.: Transient response of butterworth filters. IEEE Transactions on Circuit Theory **12**(4) (Dec 1965) 624–626
12. Coursey, C.K., Stuller, J.A.: Linear interpolation lattice. IEEE Transactions on Signal Processing **39**(4) (Apr 1991) 965–967
13. Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R.B., Guadarrama, S., Darrell, T.: Caffe: Convolutional architecture for fast feature embedding. CoRR **abs/1408.5093** (2014)
14. Castro, F.M., Marín-Jiménez, M.J., Guil, N., de la Blanca, N.P.: Automatic learning of gait signatures for people identification. CoRR **abs/1603.01006** (2016)
15. Muramatsu, D., Makihara, Y., Yagi, Y.: View transformation model incorporating quality measures for cross-view gait recognition. IEEE Transactions on Cybernetics **46**(7) (July 2016) 1602–1615
16. Choudhury, S.D., Tjahjadi, T.: Robust view-invariant multiscale gait recognition. Pattern Recognition **48**(3) (2015) 798–811
17. Zou, Q., Ni, L., Wang, Q., Li, Q., Wang, S.: Robust gait recognition by integrating inertial and RGBD sensors. CoRR **abs/1610.09816** (2016)
18. Martinelli, F., Saracino, A., Sheikhalishahi, M.: Modeling privacy aware information sharing systems: A formal and general approach. In: 2016 IEEE Trustcom/BigDataSE/ISPA, Tianjin, China, August 23-26, 2016. (2016) 767–774
19. Yoo, J.H., Hwang, D., Moon, K.Y., Nixon, M.S.: Automated human recognition by gait using neural network. In: Image Processing Theory, Tools and Applications, 2008. IPTA 2008. (2008) 1–6