A Regression Approach for Robust Gait Periodicity Detection with Deep Convolutional Networks

Kejun Wang[(✉)], Liangliang Liu, Xinnan Ding, Yibo Xu,
and Haolin Wang

College of Automation, Harbin Engineering University, Harbin, China
{wangkejun,liuliangliang,
dingxinnan,xuyibo}@hrbeu.edu.cn, WHLhrbeu@163.com

Abstract. This paper presents a regression approach to gait periodicity detection via fitting gait sequence to a sine function by deep convolutional neural networks. The key idea is to model the gait fluctuation as a sinusoidal function because of similar periodic regularity. Each frame of the gait video corresponds to a function value that can represent its periodic features. Convolutional network serves to learn and locate a frame in a gait cycle. To the best of our knowledge, it is the first work based on deep neural networks for gait period detection in the literature. An extensive empirical evaluation is provided on the CASIA-B dataset in terms of different views and network architectures with comparison to the existing works. The results show the good accuracy and robustness of the proposed method for gait periodicity detection.

Keywords: Gait period detection · Deep convolutional neural networks
Gait recognition · Biometrics technology

1 Introduction

As one of the biometric identification methods, gait recognition is particularly suitable for human identification at a long distance [1]. It requires no contact or explicit cooperation by subjects, compared with other biometric features such as face and fingerprint. Therefore, gait recognition has good prospects for application in many fields such as safety monitoring, human-computer interaction and entrance guard. In recent years, it has attracted wide attention of researchers and many effective algorithms have been proposed.

Periodicity detection is an essential step for vision-based gait recognition. Unlike other biometric techniques, it is not suitable to use a single image of the silhouette for gait recognition because of the wobble of the body in walking. Thus, the input of gait recognition is a video sequence rather than a gait silhouette. Gait period detection is the process of making the suitable length of the input video. A gait cycle can include complete gait features with the least frames. The shorter detected gait periods may miss the effective gait features while the longer ones may contain redundant data and need

more computation. The gait recognition, based on silhouette sequences or class energy images, is directly affected by the accuracy of the periodic detection [2, 3].

Many of the previous researches on gait period detection are based on the width and height of the human body [4, 5], which is usually easy and straightforward. These methods achieve high accuracy near the side view of 90°. But they are not robust to the different condition such as various views and clothes. Recently, the convolutional neural network (CNN) [6] has become the common workhorse for feature learning from images [7]. For the gait period detection task, CNN can extract periodic features of gait silhouette sequences automatically instead of a single artificial feature by the traditional methods. A CNN-based gait periodicity detection approach can be workable to get better effectiveness and robustness.

In this paper, we make the following contributions:

- We propose a regression approach based on a fitting method for gait periodicity detection. The gait sequences are modeled as the sinusoidal function due to the similar periodicity. The function value represents the periodic features of the corresponding frame.
- A CNN-based method is presented to gait period detection. The networks will learn the periodic features of silhouette sequences and locate their position in the period automatically. To the best of our knowledge, it is the first to use deep CNNs for gait periodicity detection in the literature.
- We conduct an extensive evaluation in terms of different views and network architectures. The proposed method shows high accuracy in various views, compared with the existing works.

In the remaining part of this paper, Sect. 2 presents more related works on gait periodicity detection and CNNs. And then Sect. 3 describes the proposed method in detail. An experimental evaluation is conducted on the CASIA-B gait database and results are shown in Sect. 4. Finally, conclusions are drawn in Sect. 5.

2 Related Work

2.1 Gait Periodicity Detection

The existing gait period detection methods mainly base on height and width of body, because the height and the size of footstep change periodically in walking. Collins et al. [4] proposed a method by using the width and height of the body for the period detection, but it is greatly affected by the change of the distance between the person and the camera; Lee et al. [8] utilized the width of silhouettes to detect gait period after the normalization of the pedestrians solving the problem of the changing distance; Wang et al. [5] considered silhouettes changing in size from different views and proposed a method based on the ratio of height to width, leaving out the process of normalization; Wang et al. [9] chose the average width of the legs as a feature to detect periodicity avoid the influence of bags and clothes. Moreover, the area of the body is effective to represent periodic features. Sarkar et al. [10] used the area of the legs as the feature to detect period.

Besides, model fitting is also an effective algorithm. Ben et al. [11] proposed a dual-ellipse fitting approach. Two regions of the whole silhouette divided by the centroid are fitted into two ellipses and the gait fluctuation is constructed as a periodic function depending on the eccentricities of two halves of the silhouette over time.

2.2 Deep Convolutional Neural Networks

CNN has shown many advantages in feature learning since it was submitted. The Lenet-5 model is first network designed by CNN [7]. And CNN composes the kernel parts of all the outstanding algorithms in the ImageNet large scale visual recognition challenge since 2012, when Krizhevsky et al. won the championship with the AlexNet [12]. VGG is one of the networks that have achieved excellent results in the ImageNet competition after Alexnet [13]. It has inherited and deepened some of the frameworks of Lenet and Alexnet. And the following year, the deep GoogLeNet won the first with 22 trainable layers and has reduced the top-5 classification error rate down to 6.67% [14]. These successful applications of CNNs motive us to develop gait periodicity detection methods based on CNN.

3 Method

3.1 Overview

We present a novel method to determine a gait cycle via fitting gait sequence to a sine function by deep CNNs. As shown in Fig. 1, the gait silhouette sequence is sent to the deep CNN to extract periodic features after normalization. Then the output is filtered to find the key frame of the gait period, and the frames between two peaks or troughs contribute a periodicity of gait.
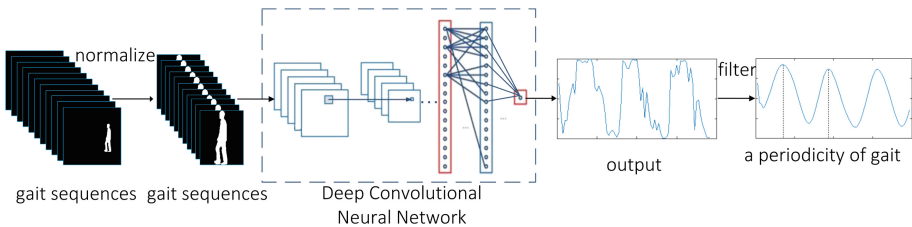

Fig. 1. Process flow of the proposed method. After normalization, deep CNN extracts periodic features of the gait sequence and outputs a waveform. (The horizontal axis shows the number of frames, and the vertical axis shows the value of output). A gait periodicity can be found through locating peaks or troughs after filtering.

The aim of normalization is to make the size of all silhouettes equal to avoid the influence of the change of distance and angle between person and camera. Each frame of the gait sequences should be cropped and resized. We locate the top and bottom pixels of the silhouettes to pick up the areas of pedestrians, and then compute their

gravity center. With the gravity center, the height of silhouettes and the aspect ratio (11/16), the frames are cropped off and rescaled into 88×128.

After normalization, the gait silhouette is input into a trained network in sequence, and output of each frame is a value that can represent its periodic features through learning of CNN. And a waveform similar to the sinusoidal function consists of the output values of a gait sequence.

Finally, filtering is an important step. The mean filtering is applied in this work. Because of the errors, the output value corresponding to a frame is an approximation of the actual value. As long as the most output of frames are relatively accurate values, the rest of fluctuation can be avoided by filtering.

3.2   Modeling as a Sinusoidal Function

The purpose of modeling is to quantify the gait periodic features of each frame into a numerical value. A sinusoidal function as a low dimensional signal is used to represent the periodic fluctuation of the gait sequence. Because the sine function is continuous and periodic with a peak and a trough within one cycle, fitting the characteristics of a gait period with maximum footsteps twice. And it is not difficult to locate the peaks and troughs, which is helpful to find the key frames to determine a gait cycle.

Table 1. The values corresponding to the periodic features of gait

Gait silhouette								
The value corresponding to the periodic feature	0	0.270	0.520	0.731	0.888	0.979	0.998	0.942
Gait silhouette								
The value corresponding to the periodic feature	0.816	0.630	0.397	0.134	-0.136	-0.400	-0.633	-0.818
Gait silhouette								
The value corresponding to the periodic feature	-0.943	-0.998	-0.979	-0.887	-0.729	-0.517	-0.267	0

We choose the sinusoidal function with a period and an amplitude of 1 to be fitted. In order to keep the consistency of evaluation periodic features, we define a standard that corresponds to the output value and the periodic feature of gait. In a silhouette

sequence, we set the value of the image where the legs are closed together and the right foot is forward as 0, i.e. the beginning. The period terminates when the legs are closed and the right foot has a forward trend once more. It is the end of the last period and the beginning of the next period. After locating the beginning and the end, the interval is the period (i.e. 1) divided by the number of frames between them. The values are obtained by accumulative and sinusoidal calculation. For example as the periodicity and the corresponding values shown in Table 1, this gait cycle contains 24 frames, so the interval is 1/24 (1 divided by the number of frames), and the position of the each frame of the period is 0, 1/24, 2/24, 3/24,…, 1 respectively. Then the values can be produced easily by sin(0), sin(1/24), sin(2/24), sin(3/24),…, sin(1). In this way, the gait fluctuation is modeled as a sinusoidal function.

3.3   Network Architectures

Deep CNN is the tool to fit the gait frames to a sinusoidal function. And it is used to learn the periodic feature of a frame and locate it in a gait cycle. Thus the input of the network is a silhouette in a frame ($128 \times 88 \times 1$), and the output is a regression value. We present 3 networks architectures for gait periodicity detection with different depths and widths. Thanks to good performance of Alexnet, VGG and GoogLeNet in images classification, similar structures are adopted at bottom layers for feature extraction. The mean squared error (MSE) is applied as their loss functions in this paper.

Basic Network for Gait Periodicity Detection. Table 2 shows the structure of the network in detail. Conv7 represents the convolution layer with 7×7 kernels. Similarly, Conv5 is with 5×5 kernels, and Conv3 is with 3×3 kernels. Larger convolution kernels are used to extract periodic features preliminarily. The number of neurons in the last layer is 1 because the output should be a regression value. And all of the activation functions are Relu.

Table 2. Detailed architecture of basic network

Layer	Output shape
Conv7-Relu-Batch_norm	30,20,48
Maxpool	15,10,48
Conv5-Relu-Batch_norm	11,6,128
Maxpool	5,3,128
Conv3-Relu	5,3,192
Conv3-Relu	5,3,192
Conv3-Relu	5,3,128
Maxpool	2,1,128
FC-Relu	1024
FC-Relu	512
FC-Relu	1

Deep Network for Gait Periodicity Detection. Is has a deeper network structure than the previous, and its structure that we use is shown in Table 3 where Conv3 is also the convolution layer with 3×3 kernels. And the output layer is the same, a neuron with Relu activation function. The difference is that smaller convolutional kernels are adopted totally. A convolutional sequence can simulate the larger receptive fields to reduce computation.

Table 3. Detailed architecture of deep network

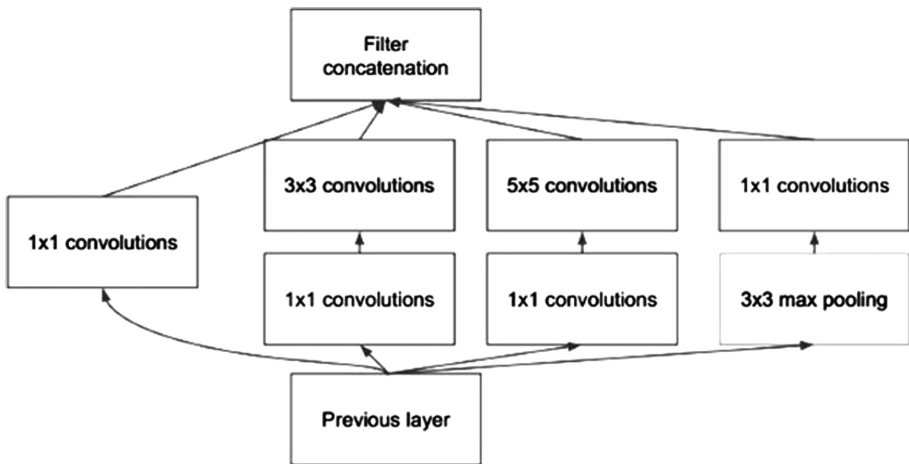| Layer	Output shape
Conv3-Relu	128,88,64
Conv3-Relu	128,88,64
Maxpool	64,44,64
Conv3-Relu	64,44,128
Conv3-Relu	64,44,128
Maxpool	32,22,128
Conv3-Relu	32,22,256
Conv3-Relu	33,22,256
Conv3-Relu	32,22,256
Maxpool	16,11,256
Conv3-Relu	16,11,512
Conv3-Relu	16,11,512
Conv3-Relu	16,11,512
Maxpool	8,5,512
Dropout-FC-Relu	1000
Dropout-FC-Relu	250
FC-Relu	1

Fig. 2. Architecture of Inception [15]

Wide Network for Gait Periodicity Detection. The unique feature of GoogleNet network model,i.e. Inception, is applied in third network architecture. Inception expands the network by improving the width of the network instead of the depth alone as shown in Fig. 2 [14, 15]. Table 4 is the schematic diagram of the third network architecture used in this paper. As in the same way, the output layer transforms the extracted gait periodic features into a regression value.

Table 4. Network architecture based on Inception

Layer	Output shape
9 × Inception	1000
FC-Relu	250
FC-Relu	1

4   Experiments

An empirical evaluation with different network architectures is provided on CASIA-B dataset. There are 124 subjects and 11 views (0, 18, …, 180°) and 10 sequences per subject for each view in the CASIA-B gait dataset [16]. And we evaluate our method with comparison to alternative approaches in terms of different views and network architectures.

4.1   Training

Deep CNN needs to learn from a large number of labeled data. Thus, it is necessary to mark the periodic features of each frame as their labels for training. The function value of each frame represents its periodic features as its label in the training set. We have located the initial position of the periodic sequence manually and have got the label by calculation. More than 96000 images in the dataset are manually marked for training.

The networks are trained using Adam with the MSE loss. Adam parameters are set as default values, i.e. β_1 is 0.9 and β_2 is 0.999. We initialize the weights using a normal distribution with the 0 mean and 0.01 variance. Batch size is set to 128, learning rate to 0.001, and the training is stopped after 75 thousand iterations (100 epochs).

4.2   Evaluation Metric

We define a straightforward metric to evaluate the performance of gait periodicity detection. This metric indicates the ratio of the error and the factual period. It is formally defined as follows:

$$C = \frac{|T - T_s|}{T} \tag{1}$$

where T is the number of the frames in an actual periodicity and T_s is the detected number. The smaller the value of C, the smaller the error is, and the higher accuracy the method gets. Conversely, the larger C value means the worse performance.
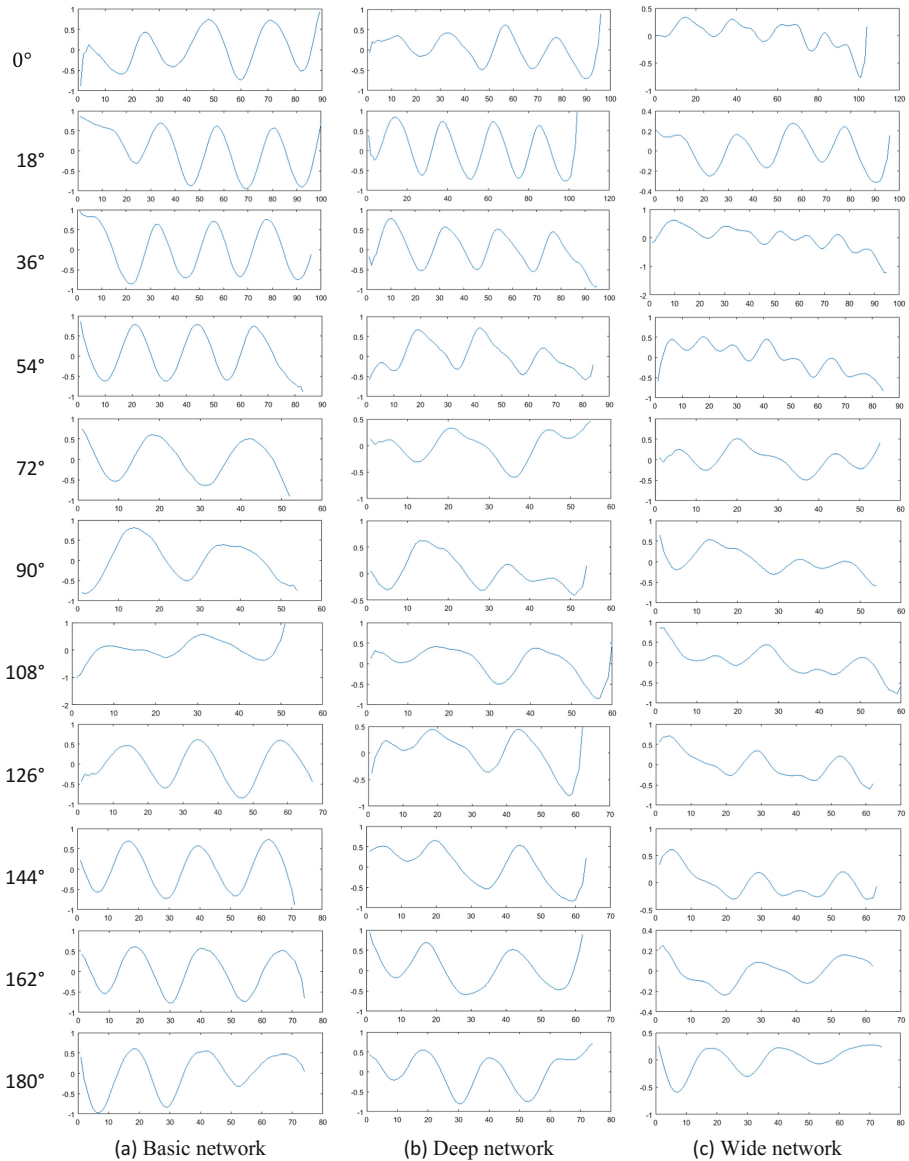

Fig. 3. Output waveforms of different networks and views. Column (a)–(c): The output waveforms in terms of 0°, 18° …, 180° of the basic network, deep network and the network based on Inception respectively. The number of the input frame is shown on the horizontal axis, and the output value is shown on the vertical axis.

4.3 Comparison

Different Network Architectures. After training, the networks can be used to determine gait cycle. Gait silhouettes are input into the networks in sequence, and the periodic features are extracted by the networks. For a frame, we can get an output value that represents its periodic features. So for a silhouette sequence, an one-dimensional vector can be got. By locating the adjacent peaks of the waveform, we can determine the gait cycle. The frames between the adjacent peaks are a periodicity of the gait sequence. Figure 3 shows filtered waveforms representing periodicity in terms of various views by 3 networks. And the waveforms can show good periodic characteristics of all views. It is found that the output is an approximation of the sine function, and the basic network has the best performance. All 3 networks work better near the oblique view (such as 18°, 36° and 144°). It may be due to the way of modeling. The value is mainly affected by the step width and the order of the left foot or the right. The bigger width step results in the bigger absolute value, as shown in Table 1. And the order of the left foot or the right results in the sign, which is positive when the right foot goes ahead. The gait silhouettes of 0° and 180° contain fewer features of step width, and the ones of 90° contain fewer features to discriminate the left or right foot. Therefore, the result shows this method is not so well at the view of 0°, 90° and 180°.

Table 5. Performances with different networks

	0°	18°	36°	54°	72°	90°	108°	126°	144°	162°	180°	Mean
Basic network	0.04	0	0	0	0.04	0.08	0.08	0.16	0.08	0.15	0.04	0.06
Deep network	0.04	0	0.04	0	0.42	0.08	0	0.4	0.36	0	0.08	0.13
Wide network	0.08	0	0.08	0.46	0.38	0.08	0.48	0.04	0.24	0	0.04	0.14

Table 5 shows the quantitative performance of each network with the evaluation metric mentioned above. Detected periods of three networks are accurate. And the values of C are close to 0 in various views, which means the performance of detection of gait cycle is well with a small error. The average value of C of basic network is as low as 0.06, and it can be calculated that the average error is about 1.5 frames according to the actual average period of 25 frames. The average C of the other two networks are also not high, 0.13 and 0.14 respectively. The experimental results show the proposed method is effective to detect the gait period and robust to various views.

The basic network is the best one of the models proposed in this paper. It can determine gait periodicity at all views effectively. Because gait silhouettes are binary images and the main features are the edges, the depth of the basic network may be enough to extract periodic features and output accurate values.

Comparison with Other Methods. Figure 4 shows the performances of gait period detection with different approaches. We choose several previous methods that can work in all views to compare (mentioned in Sect. 2). The lines warped on both sides belong to the traditional methods. That means errors of the existing works are nearly a

periodicity of gait at the views near 0° and 180°. By comparison, the proposed method based on the deep CNN can have relatively good effect on the front and back views. At the view near 90°, the errors of our method are larger than the traditional slightly. But the largest C in all views of our method is 0.16. That is to say the error is about 3 to 4 frames. They are acceptable relative to a gait cycle containing about 25 frames. In general, gait periodicity detection by convolutional neural network is feasible in terms of various views, making up for the low accuracy of the previous methods in the front and back view. Besides, the error is reasonable at the side view. Therefore, it is an effective method and robust to various views.
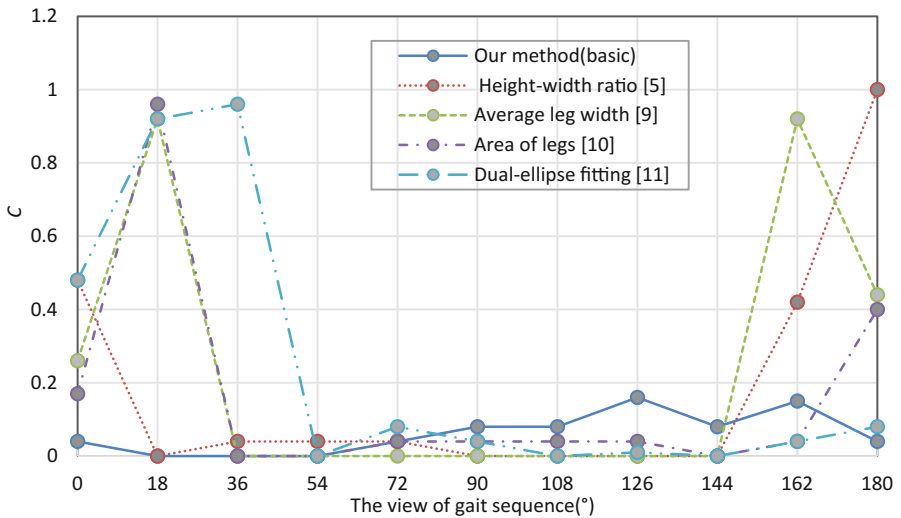

Fig. 4. Comparison with other existent methods. Our approach is shown with the basic nework. And it is compared with the methods of Wang et al. [5], Wang et al. [9], Sarkar et al. [10] and Ben et al. [11].

5   Conclusion

We present a novel approach for robust gait periodicity detection based on regression method by deep CNNs. The networks can learn the periodic features of the gait sequences and output a value that can represent the features of each frame. Experimental results confirm the effectiveness and robustness of the proposed method for gait periodicity detection in terms of various views, compared with the existing works.

References

1. Phillips, P.J.: Human identification technical challenges. In: 2002 International Conference on Image Processing, pp. 49–52. IEEE, Rochester (2002)
2. Makihara, Y., Sagawa, R., Mukaigawa, Y., Echigo, T., Yagi, Y.: Gait recognition using a view transformation model in the frequency domain. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3953, pp. 151–163. Springer, Heidelberg (2006). https://doi.org/10.1007/11744078_12
3. Li, C., Min, X., Sun, S., Lin, W., Tang, Z.: Deepgait: a learning deep convolutional representation for view-invariant gait recognition using joint Bayesian. Appl. Sci. **7**(3), 210 (2017)
4. Collins, R.T., Gross, R., Shi. J.: Silhouette-based human identification from body shape and gait. In: 5th IEEE International Conference on Automatic Face and Gesture Recognition, pp. 366–372. IEEE, Washington (2002)
5. Wang, L., Tan, T., Ning, H., Hu, W.: Silhouette analysis-based gait recognition for human identification. IEEE Trans. Pattern Anal. Mach. Intell. **25**(12), 1505–1518 (2003)
6. LeCun, Y., Boser, B., Denker, J., Henderson, D.: Handwritten digit recognition with a back-propagation network. In: Advances in Neural Information Processing Systems, pp. 396–404. ACM, San Francisco (1990)
7. Krizhevsky, A., Sutskever, I., Hinton, G.: ImageNet classification with deep convolutional neural networks. In: International Conference on Neural Information Processing Systems, pp. 1106–1114. ACM, New York (2012)
8. Lee, C.P., Tan, A.W.C., Tan, S.C.: Gait recognition with transient binary patterns. Vis. Commun. Image Represent. **33**(C), 69–77 (2015)
9. Wang, C., Zhang, J., Wang, L., Pu, J., Yuan, X.: Human identification using temporal information preserving gait template. IEEE Trans. Pattern Anal. Mach. Intell. **34**(11), 2164–2176 (2012)
10. Sarkar, S., Phillips, P.J., Liu, Z., Vega, I.R., Grother, P., Bowyer, K.W.: The humanid gait challenge problem: data sets, performance, and analysis. IEEE Trans. Pattern Anal. Mach. Intell. **27**(2), 162–177 (2005)
11. Ben, X., Meng, W., Yan, R.: Dual-ellipse fitting approach for robust gait periodicity detection. Neurocomputing. **79**(3), 173–178 (2012)
12. Lecun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. Proc. IEEE **86**(11), 2278–2324 (1998)
13. Simonyan, K., Zisserman, A..: Very deep convolutional networks for large-scale image recognition. CoRR (2014). https://arxiv.org/abs/1409.1556
14. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., et al.: Going deeper with convolutions. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–9. IEEE, Boston (2015)
15. Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A.: Inception-v4, Inception-ResNet and the impact of residual connections on learning. CoRR (2016). https://arxiv.org/pdf/1602.07261
16. Yu, S., Tan, D., Tan, T.: A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition. In: 18th International Conference on Pattern Recognition, pp. 441–444. IEEE, Hong Kong (2006)