Neural Computing and Applications (2020) 32:14275–14285
https://doi.org/10.1007/s00521-019-04524-y

EXTREME LEARNING MACHINE AND DEEP LEARNING NETWORKS

Gait recognition using multichannel convolution neural networks

Xiuhui Wang[1] · Jiajia Zhang[1] · Wei Qi Yan[2]

Received: 30 October 2018 / Accepted: 5 October 2019 / Published online: 19 October 2019

Abstract

Human gait recognition has a wide range of applications in multiple fields, such as video surveillance, digital security, and forensics. In this paper, we investigate the challenging problem of cross-view gait recognition and propose a novel gait recognition scheme by utilizing the strong expression of convolution neural networks (CNN). First, instead of using gait energy images in traditional gait recognition, we will design a new gait feature representation, trituple gait silhouettes, constructed by using consecutive gait silhouette pictures. Second, we will construct a multichannel CNN network to tackle a set of sequential images in parallel. Each of the image datasets is treated as one input channel, and a different convolutional kernel is used. Finally, the proposed approach is evaluated extensively based on the CASIA gait dataset A/B for cross-view gait recognition, and further on the OU-ISIR large population gait dataset to verify its generalization capability with large-scale data. To the best of our knowledge, this is the first time that this gait recognition scheme is presented. All our experimental results show that the proposed method obtains better performance when compared to those existing methods.

Keywords Gait recognition · Multichannel CNN · Biometrics recognition · Video-based surveillance

1 Introduction

As one of the most promising biometrics, gait embodies human walking habit and its dynamic characteristics. Compared to other biometrics, gait has many advantages, such as non-offensive extraction and low-resolution compatibility [1, 2]. Recently, human gait recognition has gained tremendous attention, and a slew of approaches [3–5] have been proposed to improve gait recognition. Nowadays, the focus of existing methods mainly is on different view angles of gait, namely, cross-view problem. However, the problem is still far away from being completely solved because gait appearance of one person may be dramatically altered when the view angles are changed.

This work was completed when the Xiuhui Wang was a visiting scholar with the Auckland University of Technology, New Zealand.

✉ Xiuhui Wang
wangxiuhui@cjlu.edu.cn

Wei Qi Yan
wyan@aut.ac.nz

1 No. 258, Xueyuan Street, Hangzhou 310018, China

2 No. 2-14, Wakefiled Street, Auckland 1010, New Zealand

Fortunately, with successful applications of deep learning (DL) [6] in object detection, segmentation, and recognition from images and videos, some researchers [7, 8] have applied deep learning to human gait recognition. As an outstanding representative with optimized network structure by combining local perceptions, weights sharing, and spatial downsampling to make good use of the characteristics of samples, convolution neural networks (CNNs) guarantee the consistency of displacement and outperform than most of the state-of-the-art methods.

Furthermore, CNNs can easily adapt to GPU implementations. A number of renowned companies such as NVIDIA, Intel, and Samsung are developing their chips to enable real-time vision applications for smartphones, cameras, and autonomous cars, etc. The models based on CNNs provide an effective way to extract features from images and frame sequences. Theoretically, gait recognition, as a research problem using picture sequences as its input, is also suitable to be solved by using CNN and other techniques in deep learning.

In this paper, in order to get better classification, we propose a new gait feature representation, trituple gait silhouettes (TTGS); we also propose a novel multichannel convolutional neural network (MCNN) to achieve better gait recognition. To the best of our knowledge, this is the first time that this gait recognition scheme is presented. The contributions of the proposed method are described as follows.

1. *A new gait representation, namely tritupple gait silhouettes (TTGS), is proposed* Instead of adopting the traditional gait energy image (GEI) which has been used broadly in many existing recognition methods, we design a new gait feature representation TTGS. By using three consecutive gait silhouette images, TTGS further explores the dynamic gait features that implicit in a gait cycle or across multiple gait cycles. Furthermore, because each TTGS is constructed by using three consecutive silhouette images of gait and the silhouette images are circularly employed, the problem of sharp data-scale reduction caused by GEI construction process may be relieved.
2. *A novel multichannel CNN (MCNN) network* We reconstruct the traditional CNN network and apply it to the new gait feature representation. The MCNN network can tackle a set of sequential images in parallel. Each member of TTGS is fed into one channel of MCNN, respectively. In conjunction with the new TTGS gait feature representation, MCNN can make full use of the feature extraction and representation ability of convolution and pooling operations to obtain more intrinsic features and thus improve the correct rate of gait recognition.
3. *A comprehensive evaluation using the CASIA gait database and OU-ISIR gait database* We conduct three experiments with open-accessed gait datasets to extensively test the proposed gait recognition framework. Experimental results demonstrate that the performance of our method is better than many existing approaches.

The remaining sections of this paper are organized as follows. In Sect. 2, related work on gait recognition will be reviewed. In Sect. 3, the proposed MCNN will be presented and demonstrated on details including its construction, training, and test. Our experiments are carried out based on the CASIA gait dataset and the OU-ISIR gait database to evaluate the proposed algorithms in Sect. 4; our conclusion and future work will be finally stated in Sect. 5.

2 Related work

According to whether deep learning (DL) is used, most of the existing approaches for gait recognition can be grouped into two categories: DL-free methods and DL-based methods. The former uses various techniques to extract gait features and achieves gait classification by using traditional classifiers, such as SVM and KNN; meanwhile, the latter utilizes deep learning in gait classification process.

2.1 DL-free methods

From research viewpoint, DL-free methods mainly focused on one or more specific problems in gait recognition: well-designed features and related dimensionality reduction [9–15], 3D reconstruction methods [16–20], ingenious artificial view-invariant gait features [21, 22], and view transform models [23–25].

In [9], feature presentation of GEI is first proposed, which is computed by averaging properly aligned human silhouettes in video frame sequences. This gait feature presentation and its varieties [10, 11] have been broadly used in multiple studies [3, 4, 12–15].

The work in [16–19] reconstructs 3D structure of each gait to generate arbitrary 2D views by projecting the 3D model. Generally, those methods can achieve promising recognition rate, but they often require multiple calibrated cameras which are unavailable in most real gait databases and practical applications. The work in [21, 22] uses artificial view-invariant features for gait recognition, which can perform very well in some specified cases, but often is hard to be generalized for all applications.

The work in [23–25] needs to learn projection transformation, with which one can transform gait features from different views to one or more general ones. Those methods can compare the normalized gait features extracted from any two videos so as to compute their similarities. In a word, traditional methods can alleviate the influence of various external factors (such as view variation and clothing difference); however, there is still a lack of effective feature extraction and modeling methods to resolve the highly nonlinear correlation between gait features in complex cross-view conditions.

2.2 DL-based methods

At present, DL-based gait recognition methods [7, 8, 26–28] mainly focus on convolutional neural networks (CNN). Wu et al. [8] give an extensive evaluation in terms of cross-view, cross-walking-condition with different preprocessing methods and CNN network architectures. Takemura et al. [26] investigate input/output architectures for CNN-based cross-view gait recognition who discussed verification versus identification and the trade-off between spatial displacement of different subjects and views. In [7, 27], CNN network models are specially designed to address the problem of cross-view gait recognition. Wolf et al. [28] develop a gait recognition method based on a deep-CNN network with 3D convolutions, which uses a specific input format including both the grayscale images and optical flow enhanced color invariance.

The proposed method in this paper belongs to the second category. However, different from existing work based on GEI [11], this paper proposes a new TTGS gait feature representation, which is constructed by using three contiguous gait silhouette images. Compared with the widely used GEI, since TTGS does not require gait cycle segmentation preprocessing and silhouette image merging, it can avoid additional errors in gait cycle segmentation and increase the amount of data available. Furthermore, based on TTGS, we propose a novel multichannel convolutional neural network (MCNN) to achieve human gait recognition. Unlike a traditional CNN network with only one convolution channel, the MCNN network has three processing channels, so it can deal with gait silhouette image sequences in parallel, and extract both static gait features and dynamic gait features from the combining results of multichannels. In addition, unlike existing multichannel CNN networks for pattern recognition, which are primarily designed to use different color channels and/or different CNN parameters, TTGS is aimed to cooperate with the proposed multiple frequency gait features so as to extract more intrinsic dynamic gait features.

3 The proposed methodology

Gait feature representation and classifier design are two key steps in gait recognition. In this paper, instead of using the traditional gait feature representation GEI or its varieties, we employ a trituple image set as the input of our classifier. Furthermore, in order to comply with the new gait feature and simultaneously utilize the capability of powerful feature extraction and classification of CNN, we reconstruct a multichannel CNN model to deal with a sequence of gait silhouette images in this paper.

The proposed method mainly includes four steps:

Step 1 Data preprocessing. Gait silhouette images are normalization and grouped into a serial of TTGS.

Step 2 MCNN construction. The traditional CNN is rebuilt to comply with the new gait feature representation TTGS.

Step 3 MCNN training. In this step, a MCNN model is constructed and trained using the TTGSs from the train set.

Step 4 MCNN-based gait classification. The trained MCNN models are used to classify gait samples in the test dataset.

The diagram of gait recognition using the method proposed in this paper is shown in Fig. 1.

3.1 TTGS representation

The TTGS consists of three frames of gait silhouette images in consecutive time as shown in Fig. 1. In nature, TTGS is an ordered image collection, and each of its members is simultaneously fed into a channel of the MCNN network. Compared with the traditional GEI gait representation, there are at least three advantages by using TTGS as a gait representation:

1. *Better utilizing the inter-frame clues of human gait* By using three consecutive gaits as inputs of the proposed network, we can further explore the dynamic gait features that reflect the essential characteristics of different individuals.

On the one hand, the traditional GEI only takes into consideration of the spatial and temporal accumulative effects in one gait cycle, i.e., a point in GEI with a larger pixel value means that the movement of the corresponding body part in one cycle is more active. As for the temporal distributions and the order of the movements of the corresponding body parts in a gait cycle, it is not reflected in related GEI. For example, when a person is walking, if the left leg is more agile, the half-gait cycle of the left leg movement will take less time. This personalized walking characteristic is not captured by GEI. However, in our solution, because all gait contour images are sequentially input into the MCNN network, consequently the highly discriminative gait characteristics can be reflected by the obtained model parameters after the training.

On the other hand, TTGS can better use the dynamic gait features that span multiple continuous natural cycles. For example, when someone walks, she/he may pause for a little while in every a few natural cycles. This kind of dynamic gait features is usually ignored when using the traditional GEI gait representations based on a natural gait cycle. But sometimes, this kind of dynamic gait features could reflect the particularity of one person's walking habit, therefore have the greatest ability to identify this person.

2. *Providing more available training data* By cycling the original silhouette images to construct a sequence of trituple image sets, the problem of sharp data-scale reduction caused by GEI construction process may be relieved.

Given a gait silhouette sequence: $P = \langle p_1, \ldots, p_M \rangle$, we can construct a set of TTGSs as

$$T = \{\langle p_i, p_{i+1}, p_{i+2} \rangle | i = 1, 2, \ldots, M\}, \tag{1}$$

where M is the number of members in P and T, and p_i refers to the ith gait silhouette image. At time $t = 1$, $\langle p_1, p_2, p_3 \rangle$ is fed into MCNN; then, at time $t = 2$,

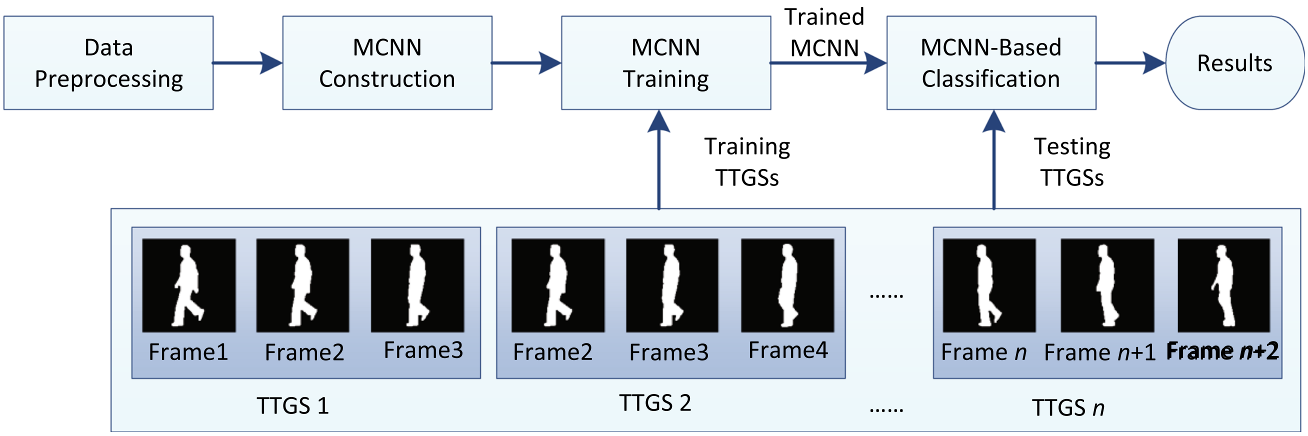

Fig. 1 The diagram of gait recognition by using the proposed method

$\langle p_2, p_3, p_4 \rangle$ is fed into MCNN, and so on. For each channel of the MCNN, such as the first channel, the input data is: p_1, p_2, p_3, p_4, ..., etc. Through this strategy, we can effectively increase the scale of the available gait data.

Currently, as a fact, the model training of deep learning networks generally requires large-scale data, while in current open-accessed gait databases and many practical applications, only relatively few training data are available. Therefore, when we select the current popular deep learning technology for gait recognition, TTGS can provide more favorable training conditions.

3. *Reducing the extra error generated by gait cycle segmentation* Gait cycle segmentation refers to splitting a sequence of gait images into a serial of fixed combinations according to the regularity of people's walking. However, in the process of gait cycle segmentation, additional errors may be introduced in due to complex walking situations and sampling conditions. Gait cycle segmentation is generally an essential pre-step in constructing the traditional gait representation GEI. Because our method utilizes the multiple channels of MCNN network to extract inter-frame clues and directly adopts the most primitive gait silhouette images to construct TTGSs, there is no need to perform gait cycle segmentation in advance. Thus, compared with the traditional GEI gait representation, TTGS can reduce the extra errors generated by gait cycle segmentation.

From the view of training and learning theory, TTGS can improve the classification accuracy in two aspects:

1. *Increase the amount of training data available* Since the deep neural network is a complex architecture composed of many node layers, there are a large number of parameters that need to be optimized through training, including weights, biases, etc. Hence, to achieve sufficient training of a CNN network, a large amount of training data is needed. However, considering that many existing gait databases have just a small amount of data, if we compress the information in a gait cycle into one GEI, the amount of training data will be further reduced. Therefore, by using TTGS as a gait feature representation, we can fully utilize the existing gait data to train a better network model and improve the accuracy of the subsequent recognition process.
2. *Consider more inter-frame clues of human gait to avoid overfitting* Since the traditional GEI only takes into consideration of the spatial and temporal accumulative effects in one gait cycle, many intrinsic gait features, such as temporal distribution and movements order of the corresponding body parts in a gait cycle, are ignored. In addition, GEIs do not take into account cross-cycle gait features. The lack of these gait features is equivalent to reducing the feature dimension that is ultimately used for gait recognition. As a result, even if we collect larger gait data, we can only get more homogeneous and redundant gait characteristics, which is helpless in avoiding overfitting problems in training. Conversely, by using TTGS to consider more gait information, we can improve the diversity of available gait data and reduce the risks of overfitting.

3.2 MCNN construction

In order to accomplish multiclass gait recognition using the new gait representation, namely TTGS, we reconstruct the traditional CNN network by adding more input channels and obtain a multichannel CNN (MCNN). Different from previous methods, hereinafter, multiple channels are used neither for computing similarities between different channels [8] nor for integrating information from different color channel [6]. Our MCNN network consists of five convolution layers, five pooling layers, one fusion layer, one fully connected layer, and one output layer as shown in Fig. 2. Maxpooling is used for all the pooling layers, which can create an invariance to small shifts and distortions. In addition, ReLU activation function is used in all the convolutional layers and the fully connected layer to reduce computational complexity and alleviate gradient vanishing problem. The working flow of MCNN is split into five stages:

Input stage The main task of this layer is to preprocess the input image data. Firstly, the silhouette images are normalized by using object-region segmentation and resizing. Then, the refined silhouette images are organized as TTGSs and fed into the MCNN in parallel. Compared to traditional single-channel CNNs which use gait silhouette one by one, our scheme can utilize more dynamic gait feature information. Furthermore, TTGSs can provide more training data and thus train better DL-based models.

Front-end processing stage There are three convolution layers and three pooling layers for each channel. All the channels use the same pooling strategies, but each channel adopts different convolutional kernels.

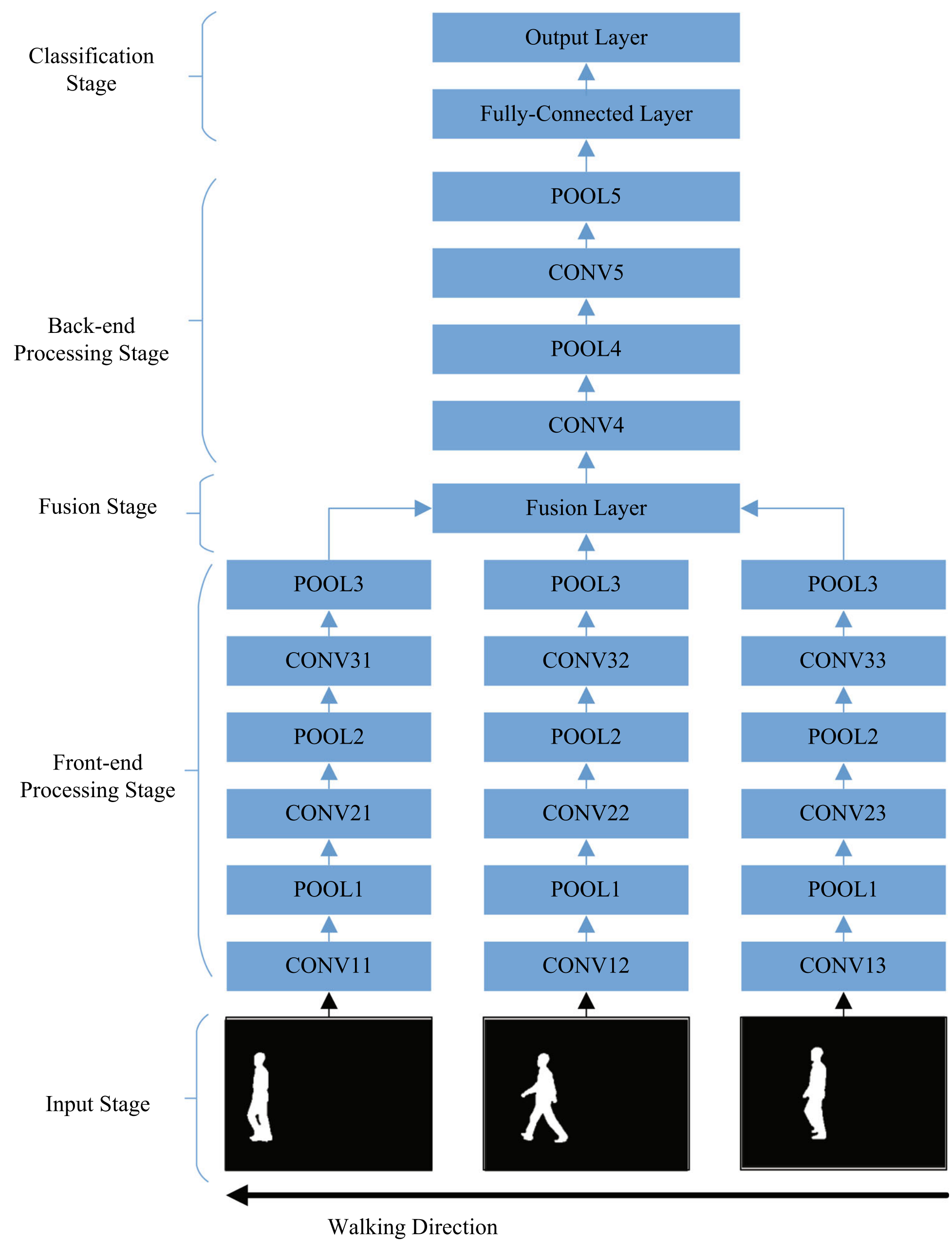

Fig. 2 Architecture of the proposed MCNN

Fusion stage The basic operation of fusion layer is to calculate the differential images from two consecutive channels, respectively, and then conduct a Boolean AND arithmetic operation. Assume that the output images of three channels at time t are $I^t_{c1}(x,y)$, $I^t_{c2}(x,y)$, $I^t_{c3}(x,y)$ respectively, then we obtain

$$d^t_{c2,c1}(x,y) = \left|I^t_{c2}(x,y) - I^t_{c1}(x,y)\right|, \tag{2}$$

$$d^t_{c3,c2}(x,y) = \left|I^t_{c3}(x,y) - I^t_{c2}(x,y)\right| \tag{3}$$

and

$$I^t_F(x,y) = \begin{cases} 1 & d^t_{c2,c1}(x,y) \wedge d^t_{c3,c2}(x,y) = 1 \\ 0 & d^t_{c2,c1}(x,y) \wedge d^t_{c3,c2}(x,y) \neq 1 \end{cases} \tag{4}$$

where $d^t_{c2,c1}(x,y)$ and $d^t_{c3,c2}(x,y)$ are differential images, $I^t_F(x,y)$ is the fusion image at time t, $\wedge$ refers to the Boolean AND operation.

Back-end processing stage Two tuples (convolutional layer and pooling layer) are used in this stage. Convolution and pooling operations are similar to traditional single-channel CNNs.

Classification stage We use Softmax function and cross-entropy to produce the gait predictions. In the training process of deep neural networks, compared to the squared-error cost function, the cross-entropy cost function converges faster, especially when the error between the predicted value and the actual value is larger. Furthermore, for eliminating the joint adaptability of neuron nodes and enhancing the generalization capability of the final model, the fully connected layer uses a dropout technique [29].

3.3 MCNN training

We use backward propagation (BP) [30] to compute the gradients of each weight and bias of each neuron, update the related weights and biases by using the stochastic gradient descent (SGD) algorithm [31]. The training process consists of three iterative steps: forward propagation of the training data, backward propagation of the loss, and weights updating as shown in Fig. 3.

With regard to the forward propagation of the training data, we firstly define the cost function as

$$C = -\frac{1}{N} \cdot \sum_{n=1}^{N} [\hat{y}_n \log y_n + (1-\hat{y}_n)\log(1-y_n)], \tag{5}$$

where N is the number of gait samples, y_n is the actual output, $\hat{y}_n$ is the corresponding label. Then, we calculate the weighted input z^l and the output y^l $(l = 2, 3, \ldots, L)$ of the lth layer.

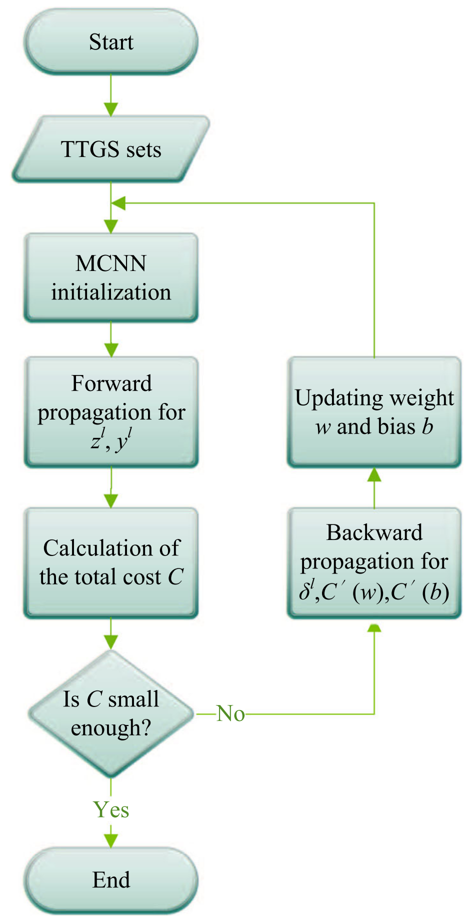

Fig. 3 Training flowchart of the proposed MCNN

$$z^l = w^l y^{l-1} + b^l \tag{6}$$

and

$$y^l = f(z^l), \tag{7}$$

where w^l is the weight, b^l is the bias, and $f(\cdot)$ is the activation function of the lth layer. In particular, because no weight parameters is included in the fusion layer, we only merge the operations based on Eq. 4.

At the step of backward propagation for updating the loss and weights, with layer by layer, we will use different strategies to calculate the sensitivity δ of each layer and partial derivatives of the cost function $\frac{\partial C}{\partial w}$ and $\frac{\partial C}{\partial b}$

$$\delta^l = \frac{\partial C}{\partial z^l}. \tag{8}$$

Notably, in our MCNN network, the convolution layers and pooling layers are one-to-one correspondence, here we treat them as a whole.

1. *Output layer* The activation function in the output layer is typically defined as

$$f(z)_i = \frac{e^{z_i}}{\sum_{j=1}^{D} e^{z_j}}, \tag{9}$$

where z is the output vector with D dimension, z_i is the ith element. According to the chain rule and following Eqs. 5–8, we obtain sensitivity δ^s of the output layer and the corresponding partial directives

$$\delta^s = y^s - \hat{y}^s, \tag{10}$$

$$\frac{\partial C}{\partial w^s} = y^{s-1}\delta^s \tag{11}$$

and

$$\frac{\partial C}{\partial b^s} = \delta^s, \tag{12}$$

where w^s and b^s are the weight and bias of the output layer respectively, y^{s-1} is the output before the output layer.

2. *Fully connected layer* We use ReLU function as the activation function of the fully connected layer, which is defined as

$$f(x) = \begin{cases} x & x > 0 \\ 0 & x \leq 0 \end{cases}. \tag{13}$$

The backward propagation process of the fully connected layer is similar with that of the output layer, except that the activation function is ReLU function.

3. *Convolution-pooling layers* As the convolution-pooling 2-tuple or pair at both the front-end processing stage and back-end processing stage have similar strategies in backward propagation of the loss, suppose that y_i^l is the output of the ith feature map in the lth convolution layer, then

$$y_i^l = f\left(\sum_j \left(y_j^{l-1} \cdot k_{i,j}^l + b_i\right)\right), \tag{14}$$

where $f(\cdot)$ is activation function of the lth layer, y_j^{l-1} is the output of the $(l-1)$th layer, $k_{i,j}^l$ is the convolution kernel used by the connection from the lth feature map in the lth layer to the jth feature map in the $(l-1)$th layer, and b_i is the corresponding bias. Given the error term δ^{l+1} of the next pooling layer, the error term δ_i^l with respect to the lth output feature map of the current convolution layer is expressed as

$$\delta_i^l = U(\delta_i^{l+1}) \circ f'(z_i^l), \tag{15}$$

where $U(\cdot)$ is an operation which saves δ^l to the size before pooling operation, and then put the values of each pooling area to the places where we obtain the maximum elements, z_i^l is the corresponding weighted input, and the symbol '$\circ$' refers to Hadamard product which denotes an elementwise product of two vectors. Then, we can further calculate the gradients of the loss function with respect to each element in the convolution kernel and bias of the lth layer as

$$\frac{\partial C}{\partial b_i^l} = \sum_{u,v} \left(\delta_i^l\right)_{u,v} \tag{16}$$

and

$$\frac{\partial C}{\partial k_{i,j}^l} = \left(y_i^l\right)_{u,v} \circ \left(\delta_i^l\right)_{u,v}, \tag{17}$$

where $\left(y_i^l\right)_{u,v}$ is the patch which is used to conduct convolution operation with $k_{i,j}^l$, (u, v) is the center of the patch, $\left(\delta_i^l\right)_{u,v}$ is the corresponding patch in δ_i^l. Finally, because there is no trainable parameter in each pooling layer, we just consider the problem of backward propagation of the loss. Assume that the error term of the previous convolution layer is δ^{l+1}, the error term δ^l of the current pooling layer is defined as

$$\delta^l = \delta^{l+1} \cdot R(W^{l+1}) \circ f'(z^l), \tag{18}$$

where $R(\cdot)$ is an operation that rotates the convolution kernel by 180°.

4. *Fusion layer* As the fusion layer has not parameters to be updated, we only need to consider the backward propagation of the loss. The basic operation of the fusion layer is to calculate the differential images of adjacent two channels respectively and then conduct a Boolean AND arithmetic operation, as shown in Eq. 4. When dealing with backward propagation of the loss, we averagely allocate the loss to the three channels.

4 Experiments

Based on the CASIA gait database [32] and OU-ISIR gait dataset [33], we design and implement three experiments for evaluating the proposed gait recognition framework. Experiment I is used to evaluate the cross-view gait recognition on CASIA Dataset B which contains the videos of 124 working individuals including 11 views, distributed from 0° to 180° under three kinds of walking conditions, i.e., normal conditions, wearing coats, and carrying bags. Silhouette images and TTGS examples from CASIA Dataset B as shown in Fig. 4.

Experiment II is to verify the generalization ability of the proposed MCNN network with large data, which takes use of the OU-ISIR Large Population Dataset consisting of two subsets: A and B. Dataset A is a set of two sequences per subject, while Dataset B is a set of one sequence per subject. In addition, each of the main subsets is further

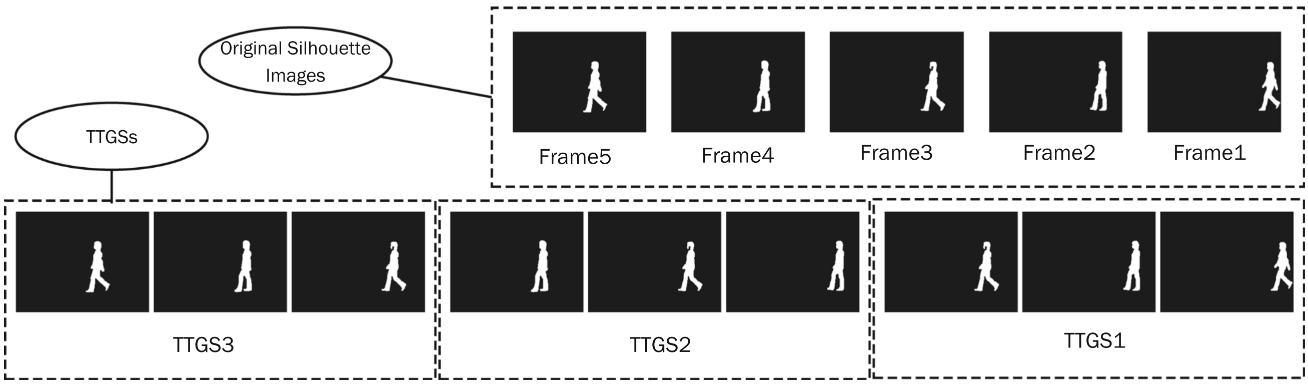

Fig. 4 Normal condition examples with 90° angle view of the No. 8 person in CASIA dataset B

grouped into 5 subsets based on the observation angles 55°, 65°, 75°, 85° including all four angles. The dataset consists of over 4000 persons walking on the ground surrounded by 2 cameras at 30 fps with 640 × 480 pixels as shown in Fig. 5. Experiment III is to further verify the generalization ability of the proposed MCNN network with multiview gait data, which was conducted on the CASIA Dataset A. This dataset includes gait data of 20 walking persons, including 12 image sequences, 4 sequences for each of the three directions, i.e., parallel, 45° and 90° to the image plane as shown in Fig. 6.

In addition, for the convenience of quantitative comparison and analysis, we use cumulative match characteristics (CMCs) [34] as the evaluation criterion in our experiments, which is a well-accepted measurement to judge the classification capabilities of closed-set identification systems. The CMC is sometimes called a lift curve, because we can see how much the line representing the model performance is lifted up over the random performance. In the CMC curve, the abscissa is the rank of correct recognition and the ordinate is the precision percent.

Without loss of generality, given a gallery gait set $G = \{g_1, g_2, \ldots, g_n\}$ and a query gait set $X = \{x_1, x_2, \ldots, x_m\}$, the set G is constructed from any individuals of a gait database D, but each identity in set X is presumed to be in set G. In addition, the set X may contain more than one gait samples of a given subject and need not to include a sample of each subject in set G. In order to create the CMC curve, each query gait is matched to every gallery gait, we can get

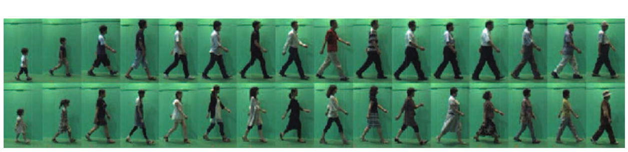

Fig. 5 Examples from OU-ISIR Large Population Dataset

Fig. 6 Examples from CASIA Dataset A

$$S_i = \{s(x_i, g_j) \mid i \in [1, \ldots, n]; j \in [1, \ldots, m]\}, \tag{19}$$

where $s(x_i, g_j)$ is the similarity score between x_i and g_j. $s(x_i, g_j)$ is ordered as

$$s(x_i, g_{(1)}) \geq s(x_i, g_{(2)}) \geq \cdots \geq s(x_i, g_{(m)}), \tag{20}$$

and x_i is assigned a rank $k_i = k$ if the matching gait from G is $g_{(k)}$. Finally, we obtain a set of $n - rank$ estimates $\{k_i | i \in [1, \ldots, n]\}$.

4.1 Benchmarks

Several state-of-the-art and baseline gait recognition approaches are selected in our comparative experiments. Furthermore, we also implement a simple CNN-based method which uses traditional CNN network to achieve gait classification and recognition. Besides, our previous gait recognition method based on 2D2PCA is added to compare with the proposed method [15].

1. *Deep-CNN LB* [8] The deep-CNN LB network is proposed by Wu et al. [8] which presents three network architectures, namely, LB, MT (Mid-Level @ Top), and GT (Global @ Top). The MT network is similar to the LB network, except that two extra nonlinear projections are employed before comparing the differences between the GEI pairs. There is no significant difference between the LB and MT networks in our practical tests. The GT network suffers from severe overfitting and has less satisfactory performance. Thus, we only implemented the LB network as the first method for the purpose of comparisons.

2. *View transformation model (VTM)* [23] VTM is a classical method to resolve the problem of cross-view gait recognition. We implement it as a baseline and compared the performance with other several approaches. The view transformation model is obtained with a set of multiple individuals from multiple views. In the recognition, the model transforms gallery gaits into the same view as that of related input gait, and the gaits can well match each other.
3. *Original GEI* [9] This method uses the original GEI as gait feature vectors and the minimum Euclidean distance classifier for individual recognition. As a widely used gait representation, GEI has been employed in many existing gait recognition approaches. Therefore, we adopt this method as a baseline algorithm for DL-free approaches.
4. *2D2PCA* [15] In this work, a gait recognition algorithm based on Gabor wavelets is proposed, which employs a two-dimensional principal component analysis (2D2PCA) method to reduce the dimension of gait features. Besides, the multiclass support vector machine (SVM) is adopted to classify different gaits.
5. *Simple CNN* This approach is implemented using basic CNN network and fed with classical gait representation GEIs. The parameter initialization process is similar to the proposed MCNN method. We use it as a baseline to obtain the performance after reconstruction.

4.2 Experiment I: comparative experimental results of several methods on CASIA Dataset B

This comparative experiment is conducted on CASIA Dataset B and all the gait samples are randomly grouped into two sets, training set and test set according to the ratio of 80% and 20%. Moreover, in order to comply with the related literature, GEIs are used as inputs of all these methods except for the proposed MCNN-based method which uses TTGSs as its input. The experimental results are shown as CMC curves in Fig. 7, where the horizontal axis is the rank and the vertical axis is the recognition rate.

Figure 7 shows that the proposed MCNN-based method performs better than other methods in terms of correct recognition rate. On the one hand, compared with DL-free methods, our method has a very significant advantage with regard to correct recognition rate. For example, our correct recognition rate at Rank 1 is almost 65%, while the correct recognition rates of most DL-free methods are under 40%. On the other hand, our method performs better than DL-based methods, as the simple CNN method and deep-CNN LB method, in terms of correct recognition rate when the rank value is less than 10. The main reasons why the proposed method achieves promising correct recognition rate have three aspects.

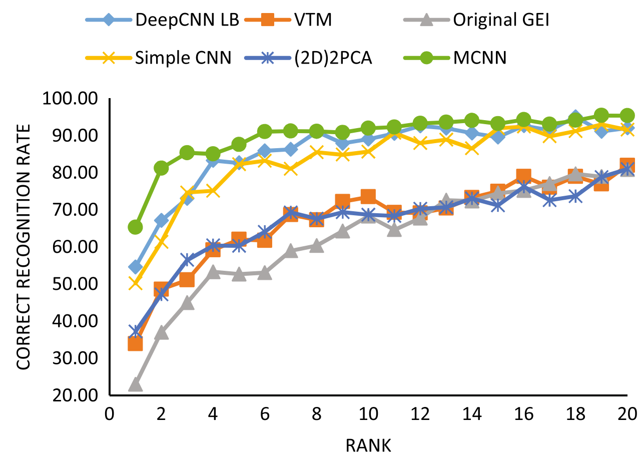

Fig. 7 The CMC curves of different methods in Experiment I

1. *A new gait representation, called TTGS, is used* By exploring the linkages between adjacent gait silhouettes, TTGS can reflect the essential characteristics of different individuals.
2. *The proposed method provides more available training data* Using the original gait silhouettes to construct a serial of tritiple image sets, the problem of sharp data-scale reduction caused by GEI construction process is relieved.
3. *The new recognition framework can reduce the extra error generated by gait cycle segmentation* Gait cycle segmentation is generally a required pre-step in constructing the traditional GEI representation. However, in the process of gait cycle segmentation, additional error is brought into this step due to complex walking action.

4.3 Experiment II: comparative experimental results of several methods on OU-ISIR Large Population Dataset

In this experiment, the generalization capability of the proposed MCNN-based method is assessed on the OU-ISIR Large Population Dataset, named OULP-C1V2. In this experiment, all the subjects are randomly grouped into training set and test set by using the ratio of 80% and 20%. In each run, we train the relevant gait classifier using the training set. Figure 8 shows the CMC curves of six methods.

Figure 8 shows that the proposed method outperforms than those DL-free methods having significant margins. The main reason is that deep neural networks can better explore the essential differences between different human

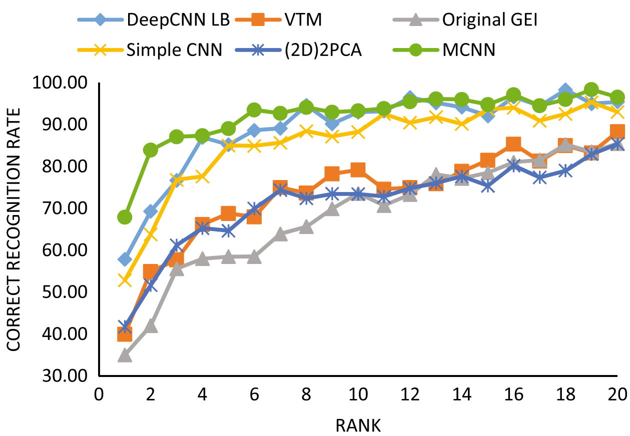

Fig. 8 The CMC curves of different methods in Experiment II

gaits. Furthermore, Fig. 8 also shows that the proposed MCNN-based method and the deep-CNN LB method are comparable in correct recognition rate when the rank values are greater than 10. But when the values are small, the proposed method outperforms other DL-based methods. This is due to more training data when using TTGSs as the input and the refined network structure of MCNN.

4.4 Experiment III: comparative experimental results of several methods on CASIA Dataset A

In Experiment III, two image sequences in CASIA Dataset A are selected as training set, the other sequences are used as testing dataset. Figure 9 shows the variation of the overall recognition rates under different rank numbers using the proposed framework, deep-CNN LB [8], VTM [23], $(2D)^2$PCA [15], original GEI [9], and a simple CNN algorithm.

In Fig. 9, we see that, for rank number less than 10, the proposed MCNN framework has higher recognition rate than other five methods. Figure 9 also illustrates the fact that, compared with the results of Experiment I and Experiment II, the results of all approaches are slightly improved. The main reason is that, in Dataset A, there is no interference, such as carrying a bag or wearing a coat.

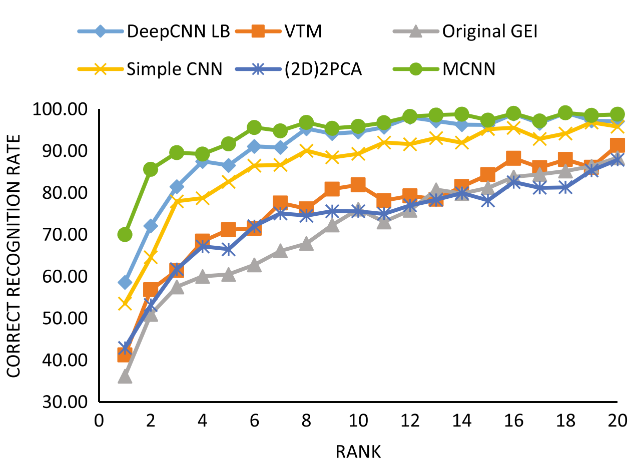

Fig. 9 The CMC curves of different methods in Experiment III

5 Conclusion and future work

In this paper, a well-designed gait representation, namely trituple gait silhouettes (TTGS), has been presented to take full use of local details and release the strict constraint of gait cycle segmentation. Then, a novel gait recognition method based on MCNN has been proposed, which utilizes the powerful ability of CNN to extract the essential features of human gait and simultaneously comply with the new gait representation.

To the best of our knowledge, this is the first time we work for gait recognition using multichannel CNN models. Our contributions are that we (1) have presented a new gait representation, namely TTGS, to preprocess gait data, which contains more local details and does not require strict gait cycle segmentation, (2) have proposed a novel gait recognition method based on MCNN for gait classification and recognition, (3) have greatly advanced the record scores based on the CASIA Dataset B and OU-ISIR Large Population Dataset, which demonstrate that our method can work under different datasets with well generalization capability.

The limitation of the proposed algorithms is that only silhouette images are used, though we have obtained the original gait videos. How to further improve the performance of gait recognition and apply gait-based technologies in practical applications, such as video surveillance, is our future work.

Acknowledgements This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grant Nos. 61303146 and 61602431 as well as a scholarship of the China Scholarship Council (CSC) (Gran No. 201708330459). We thank reviewers for their constructive suggestions and valuable comments.

Compliance with ethical standards

Conflict of interest We declare that we have not financial and personal relationships with other people or organizations that can inappropriately influence our work, there is no professional or other personal interest of any nature or kind in any product, service and/or company that could be construed as influencing the position presented in, or the review of, the manuscript entitled "Gait Recognition Using Multichannel Convolution Neural Networks".

References

1. Sarkar S, Phillips P, Liu Z (2005) The humanid gait challenge problem: data sets, performance, and analysis. IEEE Trans Pattern Anal Mach Intell 27(02):162–177
2. Wang X, Yan WQ (2019) Cross-view gait recognition through ensemble learning. Neural Comput Appl 1(1):1–13
3. Huang X, Boulgouris N (2012) Gait recognition with shifted energy image and structural feature extraction. IEEE Trans Image Process 21(04):2256–2268
4. Boulgouris N, Huang X (2013) Gait recognition using HMMs and dual discriminative observations for sub-dynamics analysis. IEEE Trans Image Process 22(09):3636–3647
5. Aggarwal H, Vishwakarma D (2017) Covariate conscious approach for gait recognition based upon Zernike moment invariants. IEEE Trans Cogn Dev Syst 10(2):397–407
6. LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521(5):436–445
7. Shiraga K, Makihara Y, Muramatsu D (2016) GEINet: view-invariant gait recognition using a convolutional neural network. In: International conference on biometrics, Halmstad, Sweden
8. Wu Z, Huang Y, Wang L, Wang X, Tan T (2017) A comprehensive study on cross-view gait based human identification with deep CNNs. IEEE Trans Pattern Anal Mach Intell 39(02):209–226
9. Han J, Bhanu B (2006) Individual recognition using gait energy image. IEEE Trans Pattern Anal Mach Intell 28(02):316–323
10. Tao D, Li X, Wu X, Maybank S (2007) General tensor discriminant analysis and Gabor features for gait recognition. IEEE Trans Pattern Anal Mach Intell 29(10):1700–1715
11. Yang X, Zhou Y, Zhang T, Shu G, Yang J (2008) Gait recognition based on dynamic region analysis. Signal Process 88(9):2350–2356
12. Theekhanont P, Kurutach W, Miguet S (2012) Gait recognition using GEI and pattern trace transform. In: Information technology in medicine and education, Hokodate, Hokkaido, Japan, pp 936–940
13. Connie T, Goh M, Teoh A (2017) A Grassmannian approach to address view change problem in gait recognition. IEEE Trans Cybern 47(06):1395–1408
14. Guan Y, Li C, Roli F (2015) On reducing the effect of covariate factors in gait recognition: a classifier ensemble method. IEEE Trans Pattern Anal Mach Intell 37(07):1521–1529
15. Wang X, Wang J, Yan K (2018) Gait recognition based on Gabor wavelets and (2D)2PCA. Multimed Tools Appl 77(10):12545–12561
16. Zhao G, Liu G, Li H, Pietikainen M (2006) 3D gait recognition using multiple cameras. In: International conference on automatic face and gesture recognition, Southampton
17. Ariyanto G, Nixon M (2011) Model-based 3D gait biometrics. In: International conference on biometrics, Washington, DC, USA
18. Abdulsattar F, Carter J (2016) Performance analysis of gait recognition with large perspective distortion. In: IEEE international conference on identity, security and behavior analysis, Sendai, Japan
19. Luo J, Tang J, Tjahjadi T (2016) Robust arbitrary view gait recognition based on parametric 3D human body reconstruction and virtual posture synthesis. Pattern Recogn 60:361–377
20. Tang J, Luo J, Tjahjadi T (2017) Robust arbitrary-view gait recognition based on 3D partial similarity matching. IEEE Trans Image Process 26(1):7–23
21. Goffredo M, Bouchrika I, Carter J, Nixon M (2010) Self-calibrating view-invariant gait biometrics. IEEE Trans Syst Man Cybern Part B 40(4):997–1008
22. Kusakunniran W, Wu Q, Zhang J, Ma Y, Li H (2013) A new view invariant feature for cross-view gait recognition. IEEE Trans Inf Forensics Secur 8(10):1642–1653
23. Makihara Y, Sagawa R, Mukaigawa Y, Echigo T, Yagi Y (2006) Gait recognition using a view transformation model in the frequency domain. In: IEEE ECCV, Graz, Austria, pp 151–163
24. Kusakunniran W, Wu Q, Li H, Zhang J (2009) Multiple views gait recognition using view transformation model based on optimized gait energy image. In: IEEE ICCV, Kyoto, Japan, pp 1058–1064
25. Kusakunniran W, Wu Q, Zhang J, Li H, Wang L (2014) Recognizing gaits across views through correlated motion co-clustering. IEEE Trans Image Process 23(2):696–709
26. Takemura N, Makihara Y, Muramatsu D, Echigo T, Yagi Y (2019) On input/output architectures for convolutional neural network-based cross-view gait recognition. IEEE Trans Circuits Syst Video Technol 29(9):2708–2719
27. Jia N, Sanchez V, Li C (2017) Learning optimized representations for view-invariant gait recognition. In: International joint conference on biometrics, Denver, USA, pp 774–780
28. Wolf T, Babaee M, Rigoll G (2016) Multi-view gait recognition using 3D convolutional neural networks. In: IEEE international conference on image processing, Phoenix, USA, pp 4165–4169
29. Ko B, Kim H, Oh K, Choi H (2017) Controlled dropout: a different approach to using dropout on deep neural network. In: IEEE international conference on big data and smart computing (BigComp), Jeju, South Korea, pp 358–362
30. Zhang R, Xu Z, Huang G, Wang D (2012) Global convergence of online BP training with dynamic learning rate. IEEE Trans Neural Netw Learn Syst 23(2):330–341
31. Li X (2018) Preconditioned stochastic gradient descent. IEEE Trans Neural Netw Learn Syst 29(5):1454–1466
32. Yu S, Tan D, Tan T (2006) A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition. In: International conference on pattern recognition, Hong Kong, China, pp 441–444
33. Iwama H, Okumura M, Makihara Y, Yagi Y (2012) The OU-ISIR gait database comprising the large population dataset and performance evaluation of gait recognition. IEEE Trans Inf Forensics Secur 7(5):1511–1521
34. Bolle RM, Connell JH, Pankanti S, Ratha NK, Senior AW (2005) The relation between the ROC curve and the CMC. In: Fourth IEEE workshop on automatic identification advanced technologies (AutoID'05), Buffalo, NY, USA, USA, pp 15–20