Gait Recognition Based on Dual-channel Dynamic and Static Fusion Network

Zhang Wenjing [*1], Zhao Xiaoyan [*1,3], Zhang Zhaohui [*1,2], and Zhang Tianyao[*1,2]

[*1] School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing 100083, China
E-mail: 18813058656@163.com
[*2] Beijing Engineering Research Center of Industrial Spectrum Imaging, University of Science and Technology Beijing, Beijing 100083,China
E-mail: zhangzhaohui@ustb.edu.cn
[*3] Shunde Graduate School of University of Science and Technology Beijing, Fo Shan 528399, China
E-mail:zhaoxiaoyan@ustb.edu.cn

Abstract: As a biometric identification method that can be realized under a distance, gait recognition has a broad application scope. The majority of existing gait recognition methods utilize the Gait Energy Image (GEI) for feature extraction. However, GEI ignores the dynamic information of gait, which causes the recognition effect to be significantly affected by variations in carrying objects and clothes. A novel strategy is proposed in this paper to overcome this limitation based on the dual-channel dynamic and static fusion network. The proposed method trains a neural network to achieve feature extraction. The static features are initially extracted from GEI, followed by the extraction of dynamic features from the image sequence. The static and dynamic features are then fused for classification with the nearest neighbor classifier. The application of the proposed method on the CASIA -B dataset presents higher recognition accuracy than the conventional gait recognition method.

Keywords: Gait Recognition, Gait Energy Image, Gait Sequence, Neural Networks

1. INTRODUCTION

Gait recognition refers to identifying people by analyzing their walking posture. Psychological research has found that everyone has their own unique gait, which makes gait recognition possible [1]. Compared with fingerprints and faces, gait is a biometric feature which is capable of identifying individuals at distance without any contact [2]. Moreover, gait recognition does not require the cooperation of the monitored person, and the image resolution requirement is not high, these advantages make gait recognition have a broad application prospect. For example, in the security system, when criminals conduct criminal activities through disguise, they will not deliberately hide their way of walking, and their gait can be acquired under their unconscious conditions.

At present, the recognition methods commonly used in gait recognition research are divided into two categories: model-based and non-model-based. Approaches in the former category first model the human body structure. Then, gait features are characterized by the model parameters such as changes in joint position and joint angle. Finally, classification is completed by some classifier, such as nearest neighbor classifier [3-6]. Such methods frequently require a large amount of computation when modeling the human body. The second category mainly extract gait features from the human silhouettes. GEI [7] is a classic non-model-based method, which is widely used in gait recognition. First, human silhouettes of each frame are obtained by background subtraction. Second, GEI is generated by rendering pixel level operators on the aligned silhouettes. Third, gait features are extracted by approaches such as Linear Discriminant Analysis (LDA) [8], Canonical Correlation Analysis (CCA) [9], and deep learning. Finally, classifier is used to complete the classification. Recently, as deep learning develops rapidly in the field of image processing, it has been widely employed on gait recognition tasks. In [10]-[12], taking the GEI as input, using convolutional neural network (CNN) to extract features for gait recognition. In [13], CNN is used to extract the joint heat map, and the long short-term memory (LSTM) network is used for gait recognition. There are also some gait recognition approaches that directly input the gait silhouette sequence, and then utilize LSTM [6] or 3D convolutional neural network (3D CNN) [14] for feature extraction.

The gait features that can be used for gait recognition can be divided into static feature and dynamic feature [15]. GEI has been widely used in gait recognition, but GEI does not consider the change of gait in the time dimension and lacks gait motion information. This paper proposes a gait recognition method based on dynamic and static fusion, using the dual-channel dynamic and static fusion network combines the static features extracted from the GEI with the dynamic features extracted from the gait sequence. The comparative experiment proves that the method proposed in this paper has higher recognition accuracy.

2. GAIT CYCLE AND CEI

Gait is a periodic manifestation of human walking process. The gait cycle can be determined by observing the curve of the lower limb width changing with time in the gait sequence. The cycle curve of a gait sequence is shown in Figure 1, where the sequence between three consecutive peaks is a gait cycle. The silhouette sequence in a gait cycle is shown in Figure 2.

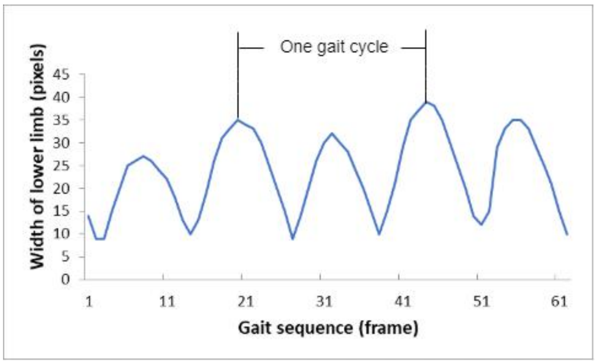


Fig.1 Periodic curve

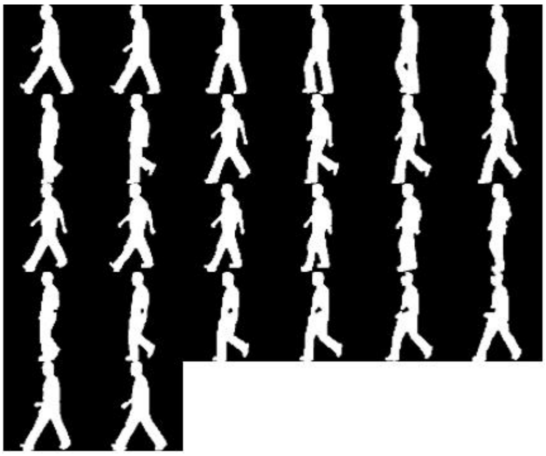


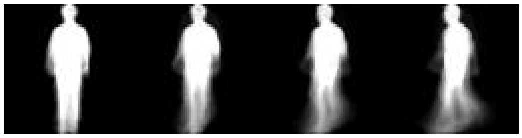
Fig.2 The complete cycle silhouette of a target in the CASIA-B dataset

GEI refers to the combination of silhouette sequences in a gait cycle into an image, and its calculation formula is as shown in (1):

$$G(x,y)=\frac{1}{N}\sum_{i=1}^{N}f_i(x,y) \quad (1)$$

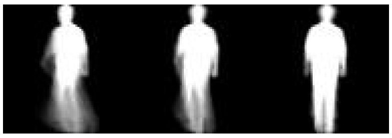
In the formula, N represents the length of the gait period; $f(x,y)_i$ represents the pixel value at the coordinates (x, y) in the gait silhouette at time i; G represents GEI. GEIs for each angle extracted on the CAISA-B dataset is shown in Figure 3.

GEI was first proposed by Han [7]. The advantages of GEI are that: its extraction method is simple; it can reduce the impact of noise; it can well represent the static features of the gait. However, GEI ignores the time information of the gait, and cannot well represent the features of human motion. It is greatly affected by the variations in carrying objects and clothes. Therefore, this paper fuses the features extracted from the GEI and the silhouette sequence, and then uses the fused features for classification and recognition. This method reduces the impact of the variations in carrying objects and clothes on the classification results.



(a)0° (b)18° (c)36° (d)54°

(e) 72 ° (f) 90° (g) 108° (h) 126°



(i) 144° (j) 162° (k) 180°

Fig.3 GEIs for each angle extracted from the CASIA-B dataset

3. DUAL-CHANNEL DYNAMIC AND STATIC FUSION NETWORK

Gait recognition is to identify a person utilizing the features that can uniquely represent the person during walking. The overall process of gait recognition in this paper is shown in Figure 4.

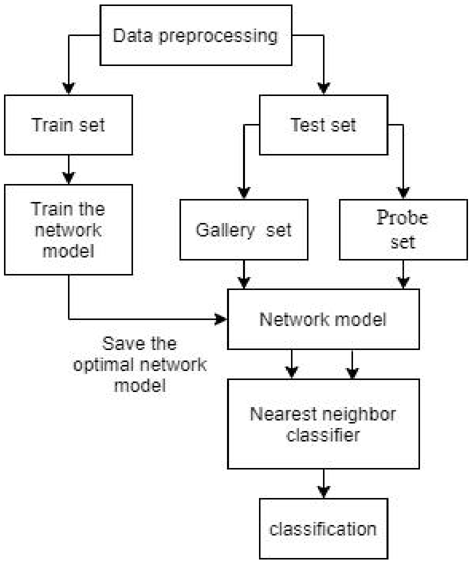


Fig.4 The overall process of gait recognition

As show in Fig4, first, the data is preprocessed, that is, the gait cycle sequence and GEI are extracted from the walking video of human. Second, the preprocessed data is divided into a train set and a test set, and the train set is input to the network to train network. After multiple iterations, the network parameters are continuously optimized and the optimal network model is trained. Third, the gallery set and probe set are input to the optimal network model to extract gait features. Finally, the gait features are classified using the nearest neighbor classifier.

3.1. Network Model

In order to better extract static features and dynamic features of gait, this paper proposes a dual-channel dynamic and static fusion network. The overall structure of the network is shown in Figure 5. The network is mainly composed of two parts: static feature extraction based on GEI and dynamic feature extraction based on gait sequence. Then the features of each channel are fused, and the fused features are input into the fully connected layer (FC) for dimensionality reduction.

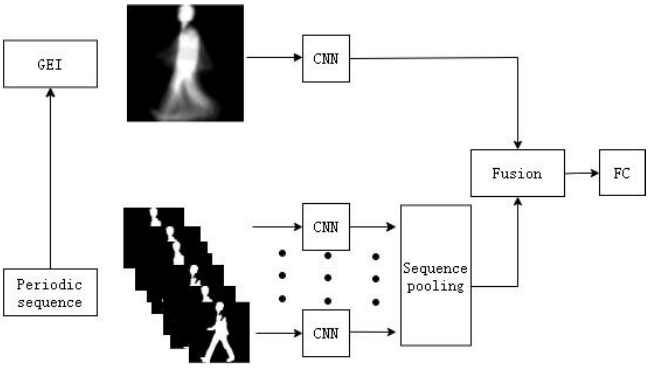


Fig.5 The overall structure of the network

For feature extraction based on the GEI, the CNN used in this paper is shown in Figure 6, in which the Conv layer 3*3*32 represents the size of the convolution kernel is 3*3, the output channel is 32, and the Maxpool layer 2*2/2 represents the size of pooling kernel is 2*2, and the step size is 2. The receptive field obtained by stacking two layers of 3*3 convolution kernels is equivalent to the receptive field of a layer of 5*5 convolution kernels, and the total amount of parameters and calculations will be greatly reduced. More importantly, the deeper network has better nonlinearity, stronger learning ability, and higher performance. Therefore, three convolutional blocks are used for feature extraction here, and each convolutional block contains two convolutional layers with a convolution kernel size of 3*3.

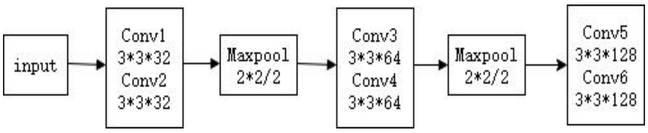


Fig.6 CNN used in this paper

For feature extraction based on gait sequence, this paper uses sequence pooling, that is, using statistical methods to aggregate frame-level features to generate sequence-level features. This operation can not only extract the dynamic features in the gait but also has a low computational cost. Its structure is shown in Figure 7. First, the above CNN is used to independently extract frame-level features from each silhouette in the gait sequence. Then, the statistical method max is used in sequence dimension to aggregate the frame-level features into sequence-level features. In Figure 7, n represents the number of samples, s represents the number of frames, c represents the number of channels, h represents the image height, and w represents the image width.

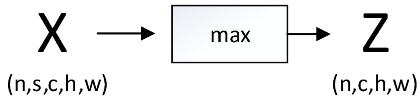


Fig.7 Frame-level feature aggregation

3.2. Classification Recognition

In this paper, the nearest neighbor classifier is adopted to achieve gait classification. The steps of gait recognition are as follows:

i. *Input all data from the gallery set into the dual-channel network model to generate features to form a gallery sample set.*
ii. *Input the data to be verified into the dual-channel network model to generate features, and calculate the Euclidean distance between the data to be verified and each sample in the gallery sample set.*
iii. *Sort each distance, select the sample with the smallest distance, and classify the sample to be verified into this category.*

4. EXPERIMENT AND ANALYSIS

4.1. Data Set and Experimental Design

The data set used in the experiment comes from the CASIA-B gait database created by the Automation Institute of the Chinese Academy of Sciences [16]. It is a large-scale, multi-view gait dataset, containing a total of 124 people, and each person has three walking states, namely 6 groups of normal state (nm), 2 groups of overcoat state (cl), and 2 groups of carrying package state (bg). Each walking state has 11 different viewing angles, namely 0°, 18°, 36°, 54°, 72°, 90°, 108°, 126°, 144°, 162°, 180°.

The experiment uses the first 74 people as the training set, and the remaining 50 people as the validation set. Among them, in the verification set, under the nm state ,the 01-04 sequences are used as the gallery set, and the 05-06 sequences are used as the probe set; under the bg state ,the 01 sequence is used as the gallery set, and the 02 sequence is used as the probe set; under the cl state, the 01 sequence is used as the gallery set, and the 02 sequence is used as the probe set.

The software platform used in this experiment is PyTorch, and the hardware platform is a Linux system server equipped with 4 GPUs. In the network structure training, the number of iterations is 30000, and the network learning rate is 0.0001.

4.2. Experimental Results and Analysis

In this paper, the accuracy of gait recognition for each view in the three states is shown in Table 1. It can be seen from Table 1 that the recognition accuracy of the network in this paper is evenly distributed in 11 different views under three different states and very nice results have been achieved.

Table .1 Dual-channel network recognition accuracy (unit %)

	nm	*bg*	*cl*
0°	100	100	100
18°	100	100	100
36°	100	98	100
54°	100	98	100
72°	100	96	100
90°	100	100	100
108°	100	98	100
126°	100	98	100
144°	100	98	100
172°	100	98	100
180°	100	96	100

Then in order to verify the effectiveness of the network proposed in this paper, the experimental results were compared with IGEI [8], and Deep CNN [11]. The comparison results are shown in Table 2. The data of other methods in Table 2 are directly referenced from their articles, and all the results are averaged on the 11 views. Observing Table 2 can find that the method in this paper achieves a high performance in the three states and exceed the best performance reported so far over 1.7%, 14.2% and 10.9% respectively. The above data shows that under the same view, the network in this paper has a good learning ability, especially under the bg and cl states, the recognition effect has a significant advantage.

Table .2 Comparison of results of each method (unit %)

	nm	*bg*	*cl*
IGEI	95.97	79.03	80.65
Deep CNN	98.3	83.8	89.1
ours	100	98	100

5. CONCLUSION

In this paper, a gait recognition method based on dual-channel dynamic and static fusion network is proposed. The method presented a novel perspective that regards gait features as the fusion of static features extracted from GEI and dynamic features extracted from gait sequence. Compared with using GEI alone, this method also focuses on the changes of gait in time dimension, which reduces the impact of variations in carrying objects and clothes on gait recognition. The comparative experimental results on the CASIA-B dataset show that this method can effectively improve the accuracy of gait recognition, and has achieved high recognition accuracy in different states and different angles, showing its robustness to walking condition variations.

ACKNOWLEDGEMENTS

This work was supported by the National Key Research and Development Project (2019YFB2101900) and Found for Beijing Common Construction Project (GJ201804)

REFERENCES:

[1] L. F. Liu, W. Jia , and Y. H. Zhu, Survey of Gait Recognition, International Conference on Emerging Intelligent Computing Technology & Applications with Aspects of Artificial Intelligence Springer-Verlag, 2009.

[2] M. S. Nixon and J. N. Carter, “Advances in automatic gait recognition,” Sixth IEEE International Conference on Automatic Face and Gesture Recognition, 2004. Proceedings., Seoul, South Korea, 2004, pp. 139-144.

[3] L. Lee and W. E. L. Grimson,“Gait analysis for recognition and classification,” Proceeding of Fifth IEEE International Conference on Automatic Face Gesture Recognition,Washington, DC, USA, 2002, pp. 155-162.

[4] B. Dikovski, G. Madjarov, and D. Gjorgjevikj, “Evaluation of different feature sets for gait recognition using skeletal data from Kinect,” International Convention on Information and Communication Technology, Electronics and Microelectronics, Opatija, IEEE,2014, pp. 1304-1308.

[5] M. Deng,C. Wang,F. Chen,and W. Zeng. Deng, “Fusion of spatial-temporal and kinematic features for gait recognition with deterministic learning,” Pattern Recognition the Journal of the Pattern Recognition Society, 2017,67:186-200.

[6] R. Liao,C. Cao , E. B. Garcia, S. Yu, and Y. Huang, “Pose-Based Temporal-Spatial Network (PTSN) for Gait Recognition with Carrying and Clothing Variations,” Chinese Conference on Biometric Recognition,2017,477-483.

[7] J. Han and B. Bhanu, “Individual recognition using gait energy image, ” in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, no. 2, pp. 316-322, Feb. 2006.

[8] R. Imad, A. Somaya, and B. Ahmed, “Improved gait recognition based on gait energy images,” International Conference on Microelectronics,IEEE, 2014,40-43.

[9] C. Luo, W. Xu, and C. Zhu, “Robust gait recognition based on partitioning and canonical correlation analysis, ”International Conference on Imaging Systems and Techniques, Macau, IEEE ,2015, pp. 1-5.

[10] C. Yan, B. Zhang, and F. Coenen, “Multi-attributes gait identification by convolutional neural networks, ” International Congress on Image and Signal Processing, Shenyang, IEEE,2015, pp. 642-647.

[11] M. Alotaibi and A. Mahmood, “Improved gait recognition based on specialized deep convolutional neural network, ” Comput Vision Image Understanding,2017,164:103–110.

[12] K. Shiraga, Y. Makihara, D. Muramatsu, T. Echigo ,and Y. Yagi, “GEINet: View-invariant gait recognition using a convolutional neural network, ” International Conference on Biometrics, Halmstad, IEEE,2016, pp. 1-8

[13] Y. Feng,Y. Li, and J. Luo , “Learning effective Gait features using LSTM, ” International Conference on Pattern Recognition, Cancun, IEEE,2016, pp. 325-330.

[14] T. Wolf, M. Babaee, and G. Rigoll, “Multi-view gait recognition using 3D convolutional neural networks, ” International Conference on Image Processing, Phoenix, AZ, 2016, IEEE ,pp. 4165-4169.

[15] B. Ye and Y. M. Wen, “Review of human identification technology based on gait, ” Computer Application , 2005,25 （11）： 2577-2580.

[16] S. Q. Yu ,D. L. Tan ,and T. N. Tan , “A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition, ” International Conference on Pattern Recognition, Hong Kong, 2006, pp. 441-444.