

Gait Recognition Based on GF-CNN and Metric Learning

Junqin Wen*

Abstract

Gait recognition, as a promising biometric, can be used in video-based surveillance and other security systems. However, due to the complexity of leg movement and the difference of external sampling conditions, gait recognition still faces many problems to be addressed. In this paper, an improved convolutional neural network (CNN) based on Gabor filter is therefore proposed to achieve gait recognition. Firstly, a gait feature extraction layer based on Gabor filter is inserted into the traditional CNNs, which is used to extract gait features from gait silhouette images. Then, in the process of gait classification, using the output of CNN as input, we utilize metric learning techniques to calculate distance between two gaits and achieve gait classification by k-nearest neighbors classifiers. Finally, several experiments are conducted on two open-accessed gait datasets and demonstrate that our method reaches state-of-the-art performances in terms of correct recognition rate on the OULP and CASIA-B datasets.

Keywords

Convolutional Neural Network, Gait Recognition, Metric Learning, k-Nearest Neighbors

1. Introduction

In the last decades, gait recognition has received considerable attention [1] for its non-invasive to individuals and compatibility from a long distance. Gait recognition determines the identity of an unknown person in a video by means of regularity features extracted from changes in the posture of the person walking. However, due to the influence of various factors such as different dressing condition, weight bearing situation and ground levelling, currently, gait recognition research faces many challenges [2,3]. In addition, variations of the positional relationship between the camera's viewing direction and the walking direction of a target person, as well as the influence of illumination changes, are also unpredictable factors in gait recognition.

An effective way to solve the problem of gait recognition is to make the most use of deep learning techniques [4] to extract the most discriminative features of gait from a large number of gait samples and there are has been extensive research regarding gait recognition. For example, Wu et al. [5] and Takemura et al. [6] demonstrated that convolutional neural networks (CNNs) can be used in cross-view human identification based on gait recognition. Nevertheless, although gait recognition has yielded some exciting experimental results, gait recognition remains at the laboratory stage.

※ This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Manuscript received May 9, 2019; first revision June 26, 2019; accepted October 6, 2019.

Corresponding Author: Junqin Wen ([wenjq@zjtie.edu.cn](mailto:wenjqq@zjtie.edu.cn))

* Digital Information Technology Institute, Zhejiang Technical Institute of Economics, Hangzhou, China ([wenjq@zjtie.edu.cn](mailto:wenjqq@zjtie.edu.cn))

The present paper proposed a novel CNN based on Gabor filter, namely, GF-CNN. In the GF-CNN, we embed a gait feature extraction layer based on Gabor filter, which was used to extract gait features from gait contour images. Furthermore, in the stage of gait classification, to measure the distance between gait features more effectively, we utilized the technique of metric learning to calculate distance and achieved gait classification by k-nearest neighbors (KNN) classifiers. Finally, we conducted several experiments on two well-known gait datasets to evaluate the proposed framework. Experimental results demonstrate that by incorporating GF-CNN network and a metric learning-based (ML-based) gait classification algorithm, we can significantly increase the correct recognition rate (CRR).

The rest of this work is structured as follows: in Section 2, we review the gait recognition approaches that were proposed in recent years. In the next section, we propose a ML-based framework for gait recognition. In Section 4, we conduct two comparative experiments and compare the existing and proposed methods in terms of correct recognition rates. In the final section, we conclude this work.

2. Related Work

Most of the existing gait recognition approaches could be grouped as shallow-learning-based (SL-based) methods and deep-learning-based (DL-based) methods. Generally, both these methods use gait energy images (GEI) or its variants as input, which was firstly proposed in [7]. GEI reflects main information of human walking status. Therefore, in some degree, GEIs are the energy accumulation of gait changing within a gait cycle. The points having larger values in a GEI reflect human body changes faster in the corresponding positions.

SL-based gait recognition methods utilize traditional machine learning technology to extract gait features, design and train classifiers to complete gait classification and recognition. Choi et al. [8] presented an approach of comparing two gaits, which can minimize impacts of such varieties factors as view changing. In [9], the authors proposed a gait recognition method that utilized a (2D)²PCA algorithm to scale down the dimension of gait features. Bakchy et al. [10] proposed a developed technique for gait identification, which used Kohonen self-organizing mapping neural network and utilized robust view transformation models to solve the problem of viewing angle changing. Wang and Yan [11] employed continuous density hidden Markov models to perform gait recognition. Besides, they used an adaptive algorithm based on Cox regression analysis to adaptively adjust parameters of each gait models.

On the other hand, DL-based gait recognition methods make the most use of the latest deep learning technology to achieve gait feature extraction or gait classification. Sokolova and Konushin [12] presented a pose-based CNN network architecture to achieve gait classification, which estimated the optical flow between consecutive frames. In [13], by combining several basic gait learners together, an ensemble learning-based gait recognition method was proposed. In [14], the authors proposed a framework for resolving the problem of cross-view gait recognition by making the most use of GEIs and sparse autoencoders. In [15], the authors introduced a gait identification method on the basis of radar signals, which utilize several processing technologies to extract multiple feature representations. Wang and Feng [16] proposed an ensemble learning framework for gait recognition, which combined two types of base gait classifiers.

3. ML-Based Gait Classification

In this section, we will describe the two key steps in gait recognition, feature extraction and gait classification respectively. Fig. 1 shows the flow chart of our method. Different from traditional CNN-based methods, we have made three improvements. First, we used a Gabor filter in the input layer to pre-process the input gait silhouette images. Secondly, we removed the SoftMax layer at the end of the CNN network. The gait feature vectors extracted by the convolutional layers and the pooling layers are input into a KNN classifier. Thirdly, in the KNN classifier, we employ a ML-based algorithm to calculate the distance between gait vectors. This distance calculation method based on metric learning does not need to set the number of individuals in the gait dataset in advance, so we can conveniently increase the capacity of the gait dataset and incrementally train it.

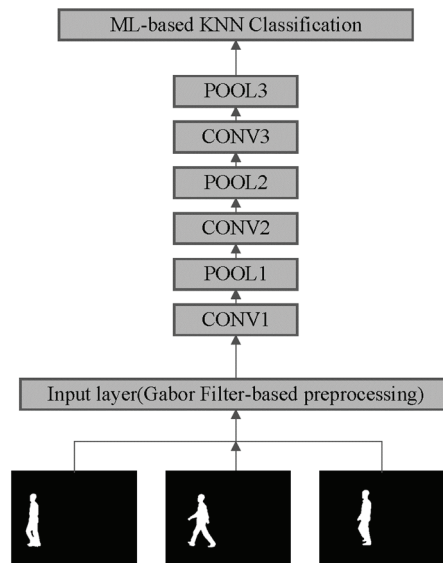


Fig. 1. Flow chart of the proposed method.

3.1 Feature Extraction

The process of feature extraction consists of an input layer and a series of convolutional and pooling layers. We firstly use Gabor filters in the input layer to pre-process the inputting gait silhouette images. In image processing, Gabor filters are special functions to extract features. In constructing the Gabor filter, we select the real part of a Gauss filter two-dimensional wavelet transforms as the Gabor filter, because the real part after the transformation is closer to the features of generally gait image. The Gabor filter is defined as:

$$f(x, y) = \exp\left(\frac{u^2 + v^2}{-2\rho^2}\right) \cos\left(2\pi \frac{u}{\lambda}\right), \quad (1)$$

$$u = x \cos \theta + y \sin \theta, \quad (2)$$

$$v = y \cos \theta - x \sin \theta, \quad (3)$$

where ρ represents the standard deviation of the Gaussian factor in a Gabor function, which follows the bandwidth changes. The smaller the bandwidth, the larger the value of ρ . θ specifies the two-dimensional rotation angle of the Gabor filter, that is, the angle between the gradient direction of the gait image and the walking direction. λ is the wavelength in pixels. γ is the aspect ratio that determines the flatness of a Gabor function.

Then, a series of Conv-Pooling pairs is used to further extract gait characteristics. The convolution layers are responsible for extracting features, while the pooling layers are responsible for filtering features. We used three convolutional layers and three pooling layers. The kernel sizes corresponding to the three convolutional layers (named CONV1, CONV2, and CONV3 as shown in Fig. 1) are 3×3 , 3×3 , and 5×5 respectively. In the pooling layers, we all employ the max-pooling strategy with 2×2 filters and stride 2. In order to train the CNN network in Fig. 1, we added a temporary SoftMax layer to calculate the loss. Backpropagation algorithm is used to compute the gradient of the cost function and update the weights in each layer.

3.2 Metric Learning-Based Gait Recognition

How to calculate the distance between two adjacent samples is an essential step in the process of KNN-based identification. In the stage of gait classification, we use metric learning techniques to calculate the distance between different gait feature vectors. The final goal of metric learning is to reduce intra-class distance, and, at the same time, to increase the inter-class distance. By learning the distance between the gait data, the metric learning can be used to analyse the association information between the gait samples; at the same time, the most distinguishable gait features can be extracted, thereby improving the correct recognition rate of the subsequent classification process.

Let $M : X \times X \rightarrow R_0^+$ is a function over vector space Ω , then M can be called a metric only if for all vectors $x_1, x_2, x_3 \in \Omega$, they meet the following relationship [17]:

- 1) $M(x_1, x_2) + M(x_2, x_3) \geq M(x_1, x_3)$;
- 2) $M(x_1, x_2) \geq 0$;
- 3) $M(x_1, x_2) = M(x_2, x_1)$;
- 4) $M(x_1, x_2) = 0 \Leftrightarrow x_1 = x_2$.

In this paper, we utilize a metric model to learn the Mahalanobis distance of different gait vectors. The Mahalanobis distance is a distance based on the sample distribution, which is a normalized format of the Euclidean distance. Such normalized space is the principal component decomposition of the sample data by principal component analysis. Then all the principal component decomposition axes are normalized to form a new coordinate axis. The space formed by these axes is the normalized principal component space. Assuming the gait feature vector set extracted by GF-CNN is $V = \{v_i | i = 1, 2, \dots, N\}$, then the Mahalanobis distance between two gait feature vectors v_j and v_k is defined as:

$$D_{j,k} = \sqrt{(v_j - v_k)^T S^{-1} (v_j - v_k)}, \quad (4)$$

where N refers to the gait vector number, $j, k \in [1, N]$, S represents covariance matrix related to V .

For gait classification based on KNN to succeed, the gait feature neighbors from the same person of each input should be closer than the gait feature vectors from other persons. Therefore, to increase the

robustness of KNN-based gait classification, we utilize an even more strict constraints for learning the Mahalanobis distance, i.e., to keep a large distance between impostors and the perimeters established by the neighbors from the same person [18].

4. Experimental Results

For evaluating the GF-CNN-based gait recognition method, two comparative experiments are conducted on two open-accessed gait datasets.

4.1 Experimental Configurations

In our experiments, we used the two open-accessed gait datasets:

1) Gait database from CASIA (tagged as CASIA-B) [19]. CASIA dataset B is a large-scale cross-view dataset including 124 individuals, which provides clear separation of view angles and other external factors, such as clothing changes, carrying, etc. CASIA dataset B enables our evaluations due to other effects, such as camera axis changing, different dressing, with or without a bag. The walking videos were sampled using eleven cameras set at different positions, distributed from 0° to 180° .

2) LP gait database from OU-ISIR (tagged as OU-ISIR-LP) [20]. OU-ISIR-LP was sampled by Osaka University, which includes gait videos from two cameras at different positions. The OU-ISIR large population dataset includes subset A and subset B. Subset A is a set of two sequences per object and subset B is a set of one sequence per object. Besides, each of these subsets is further divided into five groups with view angles 55° , 65° , 75° , and 85° separately. The dataset consists of more than 4,000 individuals captured by two cameras at 30 fps with 640×480 pixels.

Furthermore, we used CMC curves [21] to display the experimental data and compare different correction recognition rate of different methods. A CMC curve is a precision curve that provides recognition precision for each rank, which expresses the performance of a recognition system and returns ranked lists of candidates. The x-axis of CMC curves is rank of recognition, and the y-axis is precision percent. To estimate the CMC of a recognition system, the quantitative similarity degrees between an unknown gait and gaits in test set are ranked from large to small order. Generally, if the quantitative similarity degree related to the correct identification sample from a given enrolment dataset is smaller, the recognition process is more successful. Besides, in our comparative experiment, six existing methods were considered, namely, Original GEI [7], RVTM [10], CDHMM [11], 2D2PCA [9], and a simple CNN method.

4.2 Experiment on CASIA Gait Dataset B

In this experiment, we divided all gait silhouette images in CASIA-B into training sets and test sets according to the ratio of 8:2. The four existing methods and the simple CNN method we implemented are involved in the comparison. The CMC curves are shown in Fig. 2.

The results shown in Fig. 2 demonstrate that GF-CNN-based approach outperforms other methods in terms of correct recognition rates. This is mainly because the method in this paper adopts a more effective distance calculation method based on metric learning. In addition, Fig. 2 also shows that the DL-based

methods outperform the SL-based methods. The main reason is because the gait features extracted by deep learning techniques are more discriminative than those extracted by traditional methods.

4.3 Experiment on OU-ISIR Large Population Dataset

This experiment makes use of the gait silhouette images from OU-IRIS large population dataset. The training set contains 2,068 samples of 1,034 persons. For each person, one sample is under his/her own carried objects condition, another sample is without his/her carried objects. The test set contains 1,036 persons with carried objects that is disjoint from the 1,034 subjects in the training set. Similar to the previous experiment, this comparative experiment still uses four existing methods, as well as a simple CNN method that is implemented by us. The experimental results are shown in Fig. 3.

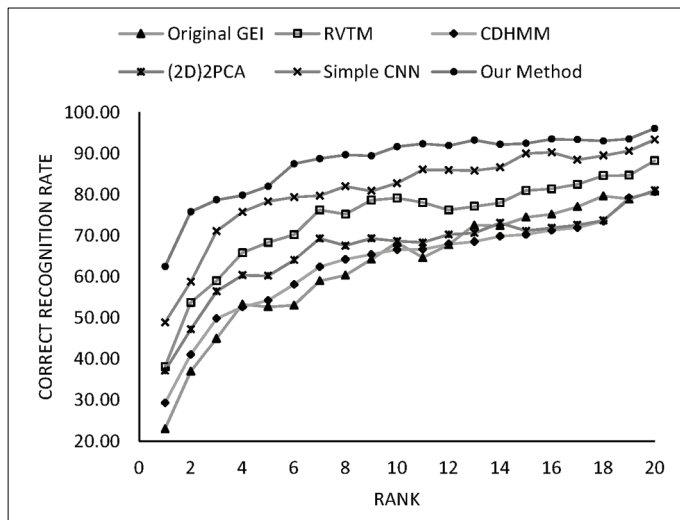


Fig. 2. Comparison of six approaches on CASIA-B dataset.

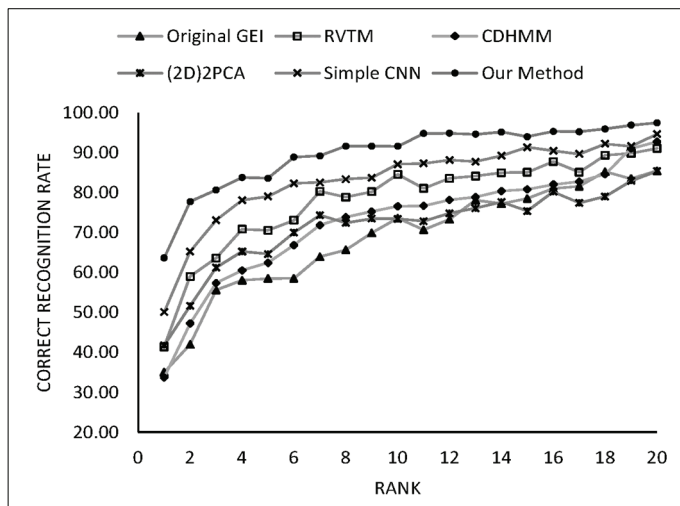


Fig. 3. Comparison of six approaches on OU-ISIR-LP dataset.

Fig. 3 demonstrates that the proposed method in this paper works better than other methods. The reason for this result is that the proposed method in this work uses Gabor filter-based gait feature pre-processing algorithm, and the calculation method of distance between gait feature vectors makes the most of metric learning techniques. In addition, compared with the previous experiment, all the methods in this experiment are generally improved. This is mainly because the external interference conditions of the gait samples used in this experiment are relatively simple.

5. Conclusion

In this work, a GF-CNN-based gait classification method is presented, which improve the traditional CNN from two aspects: (1) introduce a new type of feature extraction layer by utilizing Gabor filters, and (2) replace the traditional distance calculation method in gait classification with a new one based on metric learning technique. With the proposed method, we have carried out comparative experiments with the CASIA-B and OU-ISIR-LP datasets, showing its robustness to cross-view conditions, and its generalization ability to large-scale datasets.

As one possible limitation of our work, our method has only been evaluated on two gait databases with indoor scenes. With outdoor scenes in more complex backgrounds, the performance of the proposed method needs further verification. In future work, we will focus on how to extend the proposed method on other gait datasets with outdoor scenes.

References

- [1] J. P. Singh, S. Jain, S. Arora, and U. P. Singh, "Vision-based gait recognition: a survey," *IEEE Access*, vol. 6, pp. 70497-70527, 2018.
- [2] Y. He, and J. Zhang, "Deep learning for gait recognition: a survey," *Pattern Recognition and Artificial Intelligence*, vol. 31, no. 5, pp. 442-451, 2018.
- [3] X. Wang and W. Q. Yan, "Human gait recognition based on frame-by-frame gait energy images and convolutional long short-term memory," *International Journal of Neural Systems*, vol. 30, no. 1, article no. 1950027, 2019.
- [4] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 5, pp. 436-445, 2015.
- [5] Z. Wu, Y. Huang, L. Wang, X. Wang, and T. Tan, "A comprehensive study on cross-view gait based human identification with deep CNNs," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 2, pp. 209-226, 2017.
- [6] N. Takemura, Y. Makihara, D. Muramatsu, T. Echigo, and Y. Yagi, "On input/output architectures for convolutional neural network-based cross-view gait recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 29, no. 9, pp. 2708-2719, 2019.
- [7] J. Han and B. Bhanu, "Individual recognition using gait energy image," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 2, pp. 316-322, 2006.
- [8] S. Choi, J. Kim, W. Kim, and C. Kim, "Skeleton-based gait recognition via robust frame-level matching," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 10, pp. 2577-2592, 2019.
- [9] X. Wang, J. Wang, and K. Yan, "Gait recognition based on Gabor wavelets and (2D)²PCA," *Multimedia Tools and Applications*, vol. 77, no. 10, pp. 12545-12561, 2018.

- [10] S. C. Bakchy, M. R. Islam, and A. Sayeed, "Human identification on the basis of gait analysis using Kohonen self-organizing mapping technique," in *Proceedings of the 2nd International Conference on Electrical, Computer & Telecommunication Engineering (ICECTE)*, Rajshahi, Bangladesh, 2016, pp. 1-4.
- [11] X. Wang and K. Yan, "Human gait recognition using continuous density hidden Markov models," *Pattern Recognition and Artificial Intelligence*, vol. 29, no. 8, pp. 709-717, 2016.
- [12] A. Sokolova and A. Konushin, "Pose-based deep gait recognition," *IET Biometrics*, vol. 8, no. 2, pp. 134-143, 2019.
- [13] X. Wang and W. Q. Yan, "Cross-view gait recognition through ensemble learning," *Neural Computing and Applications*, vol. 32, no. 11, pp. 7275-7287, 2020.
- [14] U. Martinez-Hernandez, A. Rubio-Solis, and A. A. Dehghani-Sani, "Recognition of walking activity and prediction of gait periods with a CNN and first-order MC strategy," in *Proceedings of the 7th IEEE International Conference on Biomedical Robotics and Biomechanics (Biorob)*, Enschede, Netherlands, 2018, pp. 897-902.
- [15] I. Cheheb, N. Al-Maadeed, S. Al-Maadeed, and A. Bouridane, "Investigating the use of autoencoders for gait-based person recognition," in *Proceedings of NASA/ESA Conference on Adaptive Hardware and Systems (AHS)*, Edinburgh, UK, 2018, pp. 148-151.
- [16] X. Wang and S. Feng, "Multi-perspective gait recognition based on classifier fusion," *IET Image Processing*, vol. 13, no. 11, pp. 1885-1891.
- [17] R. M. Bolle, J. H. Connell, S. Pankanti, N. K. Ratha, and A. W. Senior, "The relation between the ROC curve and the CMC," in *Proceedings of the Fourth IEEE Workshop on Automatic Identification Advanced Technologies (AutoID'05)*, Buffalo, NY, USA, 2005, pp. 15-20.
- [18] K. Q. Weinberger and L. K. Saul, "Distance metric learning for large margin nearest neighbor classification," *Journal of Machine Learning Research*, vol. 10, no. 9, pp. 207-244, 2009.
- [19] R. Sun, Z. Wang, K. E. Martens, and S. Lewis, "Convolutional 3D attention network for video based freezing of gait recognition," in *Proceedings of Digital Image Computing: Techniques and Applications (DICTA)*, Canberra, Australia, 2018, pp. 1-7.
- [20] S. Yu, D. Tan, and T. Tan, "A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition," in *Proceedings of the 18th International Conference on Pattern Recognition (ICPR'06)*, Hong Kong, China, 2006, pp. 441-444.
- [21] H. Iwama, M. Okumura, Y. Makihara, and Y. Yagi, "The OU-ISIR gait database comprising the large population dataset and performance evaluation of gait recognition," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 5, pp. 1511-1521, 2012.



Junqin Wen <https://orcid.org/0000-0002-4827-2641>

She received M.S. degree in School of Computer Science and Engineering from Zhengzhou University in 2004. Her current research interests include pattern recognition and computer graphics. Since September 2004, she is with the Digital Information Technology Institute of Zhejiang Technical Institute of Economics.