Neurocomputing 239 (2017) 81–93

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Invariant feature extraction for gait recognition using only one uniform model

Shiqi Yu[a,*], Haifeng Chen[a], Qing Wang[a], Linlin Shen[a], Yongzhen Huang[b]

[a] Computer Vision Institute, College of Computer Science and Software Engineering, Shenzhen University, Shenzhen 518060, China
[b] National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing, 100190, China

ARTICLE INFO

Article history:
Received 21 November 2016
Revised 26 January 2017
Accepted 3 February 2017
Available online 9 February 2017

Communicated by Dr Jiwen Lu

Keywords:
Gait recognition
Deep learning
Invariant feature

ABSTRACT

Gait recognition has been proved useful in human identification at a distance. But many variations such as view, clothing, carrying condition make gait recognition is still challenging in real applications. The variations make it is hard to extract invariant feature to distinguish different subjects. For view variation, one view transformation model can be employed to convert the gait feature from one view to another. Most existing models need to estimate the view angle first, and can work for only one view pair. They can not convert multi-view data to one specific view efficiently. Other variations also need some specific models to handle. We employed one deep model based on auto-encoder for invariant gait extraction. The model can synthesize gait feature in a progressive way by stacked multi-layer auto-encoders. The unique advantage is that it can extract invariant gait feature using only one model, and the extracted feature is robust to view, clothing and carrying condition variation. The proposed method is evaluated on two large gait datasets, CASIA Gait Dataset B and SZU RGB-D Gait Dataset. The experimental results show that the proposed method can achieve state-of-the-art performance by only one uniform model.

1. Introduction

Gait, known as human walking style, is a kind biometric feature for human identification at a distance. Compared with other biometric features, such as face, iris, palmprint and fingerprint, gait has great potential in human identification because of its unique advantages such as non-contact, hard to fake and obtainable at a distance. Therefore gait recognition in surveillance attracted increasing attention in computer vision community.

There are many pioneer works on gait recognition. Some of them are model-based methods [1–3], and some are appearance-based ones [4–7]. These works show that gait recognition is feasible in human identification at a distance. But gait recognition is still a challenging task because of view, clothing, occlusion and other variations. These challenges can affect the recognition accuracy greatly. Among these challenges, view variation is one of the most commons because we can not control the walking direction of subjects in real applications. Many existing view invariant gait recognition methods [6,8–12] heavily depend on the accuracy of view angle estimation. For each gallery and probe angle pair, a model is need to be trained, and the model can only transform specific view. Besides of view variation, clothing can also change the human body appearance and shape greatly. Some clothes, such as long overcoats, can occluded the leg motion. Carrying condition is another factor which can effect feature extraction because it is not easy to segment the carried object from a human body in images.

The unique advantage of our work is that only one uniform model is trained which can handle gait data with view, clothing and carrying condition variations. The gait data captured with multiple variations can be transformed into the side view without knowing the specific view angles, clothing type and the object carried. So this method has great potential in real scenes.

The rest of the paper is organized as follows. Section 2 discusses related works. Section 3 describes the proposed invariant feature extraction model. Experiments and evaluation are presented in Section 4. The last section, Section 5, gives the conclusions.

2. Related work

In the following part of this section, we will briefly review gait recognition methods which are invariant to changes.

Some researchers paid close attention to view invariant gait recognition more than a decade ago. Some early methods, such as

* Corresponding author.
E-mail addresses: shiqi.yu@szu.edu.cn, shiqi.yu@gmail.com (S. Yu), chenhaifeng@email.szu.edu.cn (H. Chen), 2009150166@email.szu.edu.cn (Q. Wang), llshen@szu.edu.cn (L. Shen), yzhuang@nlpr.ia.ac.cn (Y. Huang).

http://dx.doi.org/10.1016/j.neucom.2017.02.006
0925-2312/

that in [13], use static body parameters measured from gait images as a kind of view invariant feature. Kale et al. [14] used the perspective projection model to generated side view feature from any other arbitrary view. Actually the relation between two views can not be modeled by a simple linear model, such as the perspective projection model.

Some other researcher employed more complex models to handle this problem. Makihara et al. [8] designed a view transformation model (VTM) in the frequency-domain features nor the spatial domain. The method RSVD-VTM proposed in [9] is in spatial domain. It uses reduced SVD to construct a VTM and optimized Gait Energy Image(GEI) feature vectors based on linear discriminant analysis (LDA), and achieves relative good improvements. According to the great capability of robust principal component analysis (RPRC) in feature extraction, Zheng et al. [6] established a robust VTM via RPCA for view invariant feature extraction. Kusakunniran et al. [10] took the view transformation as a regression problem, and used the sparse regression based on the elastic net as the regression function. Bashir et al. [15] formulated a gaussian process classification framework to estimate view angle in probe set, then uses canonical correlation analysis(CCA) to model the correlation of gait sequences from different views. Luo et al. proposed a gait recognition method based on partitioning and CCA [16]. They separated GEI image into 5 non-overlapping parts, and for each part they used CCA to model the correlation. In [17], Xing et al. also used CCA. But they reformulated the traditional CCA to deal with high-dimensional matrix, and reduce the computational burden in view invariant feature extraction. Lu et al. [18] proposed one method which can handle arbitrary walking directions by cluster-based averaged gait images. But if there is not similar walking direction in the gallery set, the recognition rate will decrease.

Some other researchers also tried to solve view variance using only one model,such as Hu et al. [19] proposed a method named as ViDP which extracts view invariant features using a linear transform. Wu et al. [20] trained deep convolution neural networks using supervised information and achieved high accuracies.

The clothing invariant gait recognition methods are not as many as view invariant ones in the literature. In [21], clothing invariant gait recognition is implemented by dividing the human body into 8 parts and analyzing the discrimination capability of different parts. In [22], Guan et al. proposed a random subspace method (RSM) for clothing-invariant gait recognition by combining multiple inductive biases for classification.

The variations on gait data can cause the recognition rate decrease greatly. Some methods in the literature can only solve a specific variation, such as view and clothing. A general method which can extract variant gait feature using only one model should be attractive.

3. Proposed method

In gait recognition, when the angle between the walking direction of and the camera is 90° (the side view), it is the best view for gait recognition because of more dynamic information. We would try to transform the gait data from any views, clothing and carrying condition to the side view with normal clothing condition and not carrying objects using one uniform non-linear model, and then extract invariant feature. The proposed model is inspired by the one in [23] where a model based on auto-encoder which is named as Stacked Progressive Auto-Encoders(SPAE). The model in [23] is proposed to deal with multi-view face recognition. We adapt it to deal with the view, clothing and carrying condition challenges. The framework is illustrated in Fig. 1. we will describe the framework in the following sections.

3.1. Gait energy image

Gait energy image [4], an appearance-based recognition method, which is produced by averaging the silhouettes in one gait cycle in a gait sequence as illustrated in Fig. 2, is well known for its robustness to image noise and reduction on computation. The pixel values in a GEI are the probabilities of the positions are occluded by a human body. According to the success of GEI in gait recognition, we take GEI as the input raw data of our method. The silhouettes and energy images used in the experiments are produced as those in [24].

3.2. Auto-encoder for image transformation

Auto-encoder [25] is one of the popular models in recent years. It can be used to extract compact features. As shown in Fig. 3, an auto-encoder usually contains three layers: one input layer, one hidden layer and one output layer. There are two parts in an auto-encoder, encoder and decoder. The encoder can transform the input data into a new representation in the hidden layer. It usually consists of a linear and a nonlinear transformation as follows:

$$y = f(x) = s(Wx + b) \tag{1}$$

where $f(\cdot)$ denotes the encoder, W denotes the linear transformation, b denotes the basis and $s(\cdot)$ is the nonlinear transformation, also called activation function, such as:

$$s(x) = \frac{1}{1 + e^{-x}} \tag{2}$$

or

$$s(x) = \ln(1 + e^{-x}) \tag{3}$$

The decoder can transform the hidden layer representation back to input data as follows:

$$x' = g(y) = s(W'y + b') \tag{4}$$

where $g(\cdot)$ denotes the decoder, W' and b' denote the linear transformation and basis in decoder and x' is the output data.

We usually use the least square error as the cost function to optimize the parameters in W, b, W' and b'.

$$\begin{aligned}[W, b, W', b'] &= min\Sigma_{i=1}^{N} \parallel x_i - x_i' \parallel^2 \\ &= min\Sigma_{i=1}^{N} \parallel x_i - g(f(x_i)) \parallel^2 \end{aligned} \tag{5}$$

where x_i denotes the ith one of the N training samples and x_i' means the correspond output of x_i. In our experiments, we train auto-encoder use Caffe [26,27] with Euclidean loss and Stochastic Gradient Descent (SGD).

The traditional auto-encoder can reconstruct the input. If we replace the output with a different data what distinguishes with the input data, the whole auto-encoder could be regarded as a regression function. But it would be really hard for just one auto-encoder to deal with large angle change, clothing and carrying variations. As shown in Fig. 4(a), the difference between 54° images and 90° ones is much larger than that between 72° images and 90° ones, especially in the leg part. It would be very difficult for just one auto-encoder to transform 54° images to 90° ones. But if we use one auto-encoder to transform 54° images to the 72° ones, and then use another auto-encoder to transform 72° images to 90° ones, it would be much easier. So multiple auto-encoders are needed to deal with gait variations. Some more auto-encoders are needed to handle clothing and carrying condition variations as shown in Fig. 4(b).

3.3. SPAE for gait variations

The main idea of the proposed method is stacked some auto-encoders together to deal with the view, clothing and carrying