J. Vis. Commun. Image R. 25 (2014) 1842–1855

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

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

Cross-view gait recognition based on human walking trajectory

Xin Chen [a,*], Tianqi Yang [a], Jiaming Xu [b]

[a] Department of Information Science and Technology, Jinan University, Guangzhou 510000, China
[b] School of Electrical and Information, Jinan University, Guangzhou 510000, China

ARTICLE INFO

Article history:
Received 10 June 2014
Accepted 4 September 2014
Available online 22 September 2014

Keywords:
Gait recognition
GCT
3-D projection
Cross-view
CCA
View variance
View calculation
Walking trajectory

ABSTRACT

We propose in this paper a novel cross-view gait recognition method based on projection of gravity center trajectory (GCT). We project the coefficients of 3-D GCT in reality to different view planes to complete view variation. Firstly, we estimate the real GCT curve in 3-D space under different views by statistics of limb parameters. Then, we get the view transformation matrix based on the projection principle between curve and plane, and estimate the view of a silhouette sequence by this matrix to complete view variance of gait features. We calculate the body part trajectory on silhouette sequence to improve recognition accuracy by using correlation strength as similarity measure. Lastly, we take nested match method to calculate the final matching score of two kinds of features. Experimental results on the widely used CASIA-B gait database demonstrate the effectiveness and practicability of the proposed method.

1. Introduction

As gait is mainly applied in remote recognition without subject cooperation, the walking direction in gait images is random. So, cross-view is unavoidable in gait recognition field [20,24,30]. In recent years, several approaches have been proposed to address this problem and they can be classified into two categories: model-based [6,7,19] and motion-based [1,2,14,15]. In the model-based methods, a statistic or generic model is usually applied to characterize the dynamic of human movements to attain view-invariance to some extent. For example, Bouchrika and Nixon [6] proposed to represent virtual images in arbitrary views using Fourier descriptor for gait recognition. Bodor et al. [5] learned a 3-D model of the walking person from samples acquired in four different views and used image-based rendering techniques to reconstruct gait for any required viewing angle. Although these methods can achieve relatively high recognition rates for large view variation, they usually require a well-calibrated multi-camera system and are computationally expensive. Therefore, many researchers in this area have also resorted to motion-based methods recently.

For motion-based approaches, view-invariant features are usually sought by some image processing or statistical learning techniques for recognition. For example, Goffredo et al. [2] proposed a markerless joint motion system based on 2-D silhouettes of human gait sequences, and rectified these features across different views to achieve view-invariance. While reasonably good recognition rates can be attained for large view variation, its performance degrades under a small view variation. Moreover, it cannot handle extreme views. Lu and Tan [18] proposed to project gait samples collected from two different views into a low-dimensional feature subspace, so that intra-class geometrical structures are well preserved and interclass distances of gait sequences are maximized simultaneously. However, the performance still leaves much to be desired when the view variation is relatively large.

Recently, the view transformation approach has been proposed for view-invariant recognition task [9,12,17,22]. The basic idea is to transform a gait feature from one view to another by learning a view relationship between the two corresponding views, and subsequently the transformed virtual feature is used for recognition. However, as gait features extracted by current methods lack strong regularity and simplicity, the transformation of a feature to another view could generate large errors and compromise the discriminative information, and methods like learning relationship(regression tool et al.) could not consider the characteristic of each pedestrian because they got relationship of gait features under different views just by training a limited amount of samples, these shortages would degrade the recognition performance. This problem could be solved by extracting features with fixed regularity under various views.

Different from the above mentioned methods, we propose in this paper a method which combines the advantages of model-based and motion-based for gait recognition across different views.

* Corresponding author.
E-mail addresses: 843597029@qq.com (X. Chen), 740345085@qq.com (T. Yang), 596049806@qq.com (J. Xu).

http://dx.doi.org/10.1016/j.jvcir.2014.09.002
1047-3203/

Our method is inspired by the fact that if a GCT is a regular function in 3-D real world, then the GCT detected from other views can be represented by the projection of the GCT to other VPs. Therefore, we propose to use the series after view-invariant as gait features for recognition. To obtain the series, we firstly estimate the parameter of 3-D GCT using statistical method. Then we calculate view of probe sequence based on the relationship between 3-D GCT in real world and 2-D GCT on view plane, and complete variance to gallery view. Experimental results on the widely used CASIA-B gait database are presented to demonstrate the effectiveness of the proposed method. Fig. 1 represents the shooting scene of camera in reality. Cross-view method in this paper has advantages compared with previous work. The most important contribution of this method is, it could construct 3-D GCT just by single-view gait sequence and realize accuracy view variance without gaps using transformation matrix. Besides, it combines GCT and body part trajectory to make up shortage of using only one kind of feature and reduces effect of error in 3-D curve estimation by first matching more accurate features. Also, it has lower computational cost than multi-camera systems.

2. Related work

Current research on gait recognition under various viewing angles falls into three categories. Approaches in the first category [23,10,11,13,5,27,32] are to construct 3-D gait information through multiple calibrated cameras. Methods in this category belong to multi-view gait recognition because gaits from multiple views are acquired to reconstruct 3-D gait model. An image-based visual hull (IBVH) was invented in [11] to render visual views for gait recognition. IBVH was computed from a set of monocular views captured by multiple calibrated cameras. In this approach, canonical visual camera positions were estimated. Then, rendered images obtained from these viewpoints were used for view normalization.

Zhang and Troje [32] introduced a view-independent gait recognition method based on a 3-D linear model and Bayesian rule. The 3-D linear model was constructed using PCA from a set of Fourier represented examples. The sets of coefficients were used as signature to describe gait, which were derived from projecting 2-D gait sequences under different views onto a 3-D model by means of a maximum of posterior estimate. Zhao et al. [23] reconstructed a 3-D gait model from video sequences captured by multiple cameras. Motion trajectories of lower limbs that were extracted from 3-D models, were used as dynamic features and linear time normalization was exploited for matching and recognition. In general 3-D analysis, at least two cameras are required. However, because of occlusion, gaits from at least four cameras were required for sufficient 3-D gait analysis [23].

Ref. [44] used stochastic learning method to integrate multiview features, which produced a new feature vector using alternative optimization. This method makes sense when multiview features are easy to achieve. But in reality we can only obtain

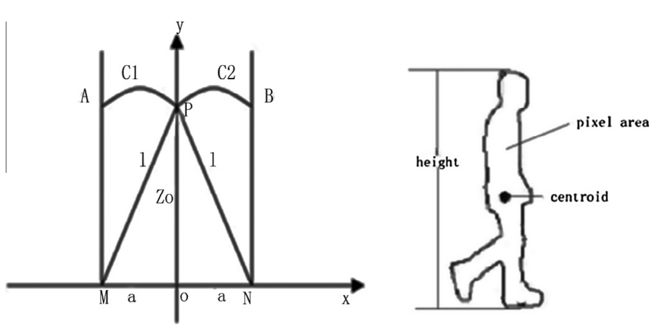

Fig. 1. Transection of gravity center motion.

gait features under single view, because shooting pedestrian from multiple views needs multiple cameras shooting at the same time, which is quite expensive.

There are a few constraints in this category: (1) it is only suitable for a fully controlled and cooperative multi-camera environment such as a biometric tunnel [33]; (2) deploying cooperative multi-camera setup into current surveillance systems is costly and complicated; (3) acquiring complex calibration information for 3-D reconstruction and/or for 2-D rendering processes may involve expensive computation; and (4) gait is a kind of nonrigid dynamic feature such that its analysis is sensitive to several factors such as occlusion and shadow, so the 3-D modeling is not stable.

The second category [2,34,28,35] is to extract gait feature which is invariant to viewing angle change. These methods are applied for *cross-view gait recognition*. Kale et al. [34] developed a method to generate a side-view of gait from any arbitrary view. Two techniques were proposed, which are based on perspective projection model and optical flow structure. Performance of this method was significantly dropped when an angle between image plane and sagittal plane was large. Jean et al. [28] proposed a method to compute view normalized trajectories of body parts which were obtained from monocular video sequences. The normalized feet and head 2-D trajectories from tracked silhouettes were used as view-invariant gait features since they always appeared like being seen from a frontal parallel viewpoint. However, this method efficiently works only for a limited range of views. Han et al. [35] extracted view-invariant features from GEI. In such a way, only parts of gait sequences that overlap between views were selected for constructing a representation of *cross-view* gait matching. Goffredo et al. [2] proposed self calibrating view-invariant gait recognition based on model-based gait features. Lower limbs' poses were estimated based on markerless motion estimation. Then, they were reconstructed in the sagittal plane using viewpoint rectification under an assumption that articulated leg motion is approximately planar. Angular measurements and trunk spatial displacement were derived from the rectified limbs' poses and used as a view-invariant gait feature. An advantage of this method is that it can efficiently perform *cross-view gait recognition* when view difference is large. However, it also has a number of limitations: (1) limbs' poses estimation is not robust from markerless motion; and (2) it is not applicable for frontal view because limbs' poses become untraceable.

The third category [3,4,36,37,21,29,1] relies on learning mapping/projection relationship of gaits under various viewing angles. The trained relationship may normalize gait features from different viewing angles into shared feature spaces before gait similarity can be carried out. Based on the literature review, methods in this category can be applied for both *cross-view* and *multi-view gait recognitions*, although this has not been mentioned in [21]. Methods in this category use a simpler single-camera system when being compared with a complicated multi-camera system of the first category. Besides, it creates a more efficient and stable gait recognition system which is not sensitive to noise when being compared with the system obtained from the second category. The method proposed in this paper belongs to this category. The common challenges among the methods in the third category are: (1) performance of *cross-view gait recognition* drops when viewing angle change is bigger; and (2) since the methods rely on supervised learning, it will be difficult for recognizing gait under untrained/unknown viewing angle.

The method in [1] learned LDA-subspaces to extract discriminative information from gait features under each viewing angle in the training dataset. In testing phase, each gait feature was projected onto each of the subspaces separately. Then, final gait distance was a weighted sum of matching results from each subspace. Bashir et al. [8] modeled the correlation of gait sequences from