Neurocomputing 224 (2017) 37–44

Contents lists available at ScienceDirect

Neurocomputing

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

Coupled locality preserving projections for cross-view gait recognition

Wanjiang Xu[a,b], Can Luo[a], Aiming Ji[a], Canyan Zhu[a,*]

[a] Institute of Intelligent Structure and System, soochow university, Suzhou 215006, China
[b] Yancheng Teachers University, Yancheng 224002, China

ARTICLE INFO

Communicated by Yue Gao

Keywords:
Gait recognition
Unified subspace
Coupled locality preserving projections

ABSTRACT

Existing methods for gait recognition mainly depend on the appearance of human. Their performances are greatly affected by changes of viewing angle. To achieve higher correct classification rates for cross-view gait recognition, we develop a coupled locality preserving projections (CLPP) method in this paper. It learns coupled projection matrices to project cross-view features into a unified subspace while preserving the essential manifold structure. In the projected subspace, cross-view gait features can be matched directly. By the virtue of structure information, the learnt subspace is more robust to the view change. Experiments based on CASIA and USF gait databases are conducted to verify the efficiency of our approach.

1. Introduction

Gait recognition aims to identify humans at a distance by inspecting their walking manners. It has a wide range of application for visual surveillance in public areas. However, in such environments there are various factors significantly affecting the performance of gait recognition [1]. One of the most challenging factors is the view change [2].

A variety of approaches have been proposed to tackle the view issue. These approaches fall into three categories: 3D gait model reconstruction based approach [3–5], view-invariant approach [6–8], and learning mapping or projection relationships of gait across views approach [9–12]. The third category is further divided into two types: view transformation-based and subspace-based approaches. Among these methods, subspace-based approaches, with the advantage of high efficiency and convenience in application, attract a lot of attentions in recent years. For cross-view gait recognition, CCA [12] is one of the most popular approaches, which projects gait features from different views into two subspaces such that they have the largest correlation. And gait features are matched by CCA correlation strength directly. However, CCA needs exactly same amount of samples under different views in training process, which is impractical in real world.

Coupled metric learning (CML) [13,14] is another typical method. It seeks a common subspace where features of cross-view gait could be measured. However, existing methods based on CML do not take into account each collection's own manifold structure while constructing the relationship between different views. For example, in training process, the originally neighboring points in set X or Y may be apart after being projected; then in testing process, the points may be projected in unexpected direction, since the original local relationship in each data set was not preserved. As shown in Fig. 1, the distance of projected points z_4 and z_5 from different classes may be closer than that of z_2 and z_5 from the same class.

In this work, we manage to seek a coupled projections approach which is more suitable for classification of cross-view gait data. On basis of CML, we attempt to preserve the manifold structure of gait features in each view while simultaneously maximize the relevance of gait features from two different views. The coupled projection matrices are firstly learnt to map original feature spaces into a unified subspace. Then in the learnt subspace, we measure the projected features obtained from gallery and probe gait sequences in Euclidean distance.

The key contributions of this paper can be summarized as follows.

- A novel coupled projections algorithm named Coupled Locality Preserving Projections (CLPP) is proposed, which can project sample points from two different viewing spaces into a unified subspace while preserving the manifold structure of intra-view and correlation of inter-view.
- This paper proposed a practical approach calculating the inter-view similarity matrix by searching the max-weighted path, which avoid being simply substituted by intra-view similarity matrix.
- To further evaluate effectiveness of the proposed method, extensive experiments are carried out for cross-view gait recognition under various scenarios which are quite common in real world.

The rest of this paper is organized as follows. Related works are discussed in Section 2. In Section 3, we briefly review the CML based approach and then introduce our proposed approach formally. Feature presentation, extraction and classification of gait data are described in

* Corresponding author.
E-mail addresses: xuwanjiang0821@163.com (W. Xu), 20134046005@stu.suda.edu.cn (C. Luo), jiaiming@suda.edu.cn (A. Ji), qiwuzhu@suda.edu.cn (C. Zhu).

http://dx.doi.org/10.1016/j.neucom.2016.10.054
Received 27 March 2015; Received in revised form 28 October 2016; Accepted 31 October 2016
Available online 03 November 2016
0925-2312/

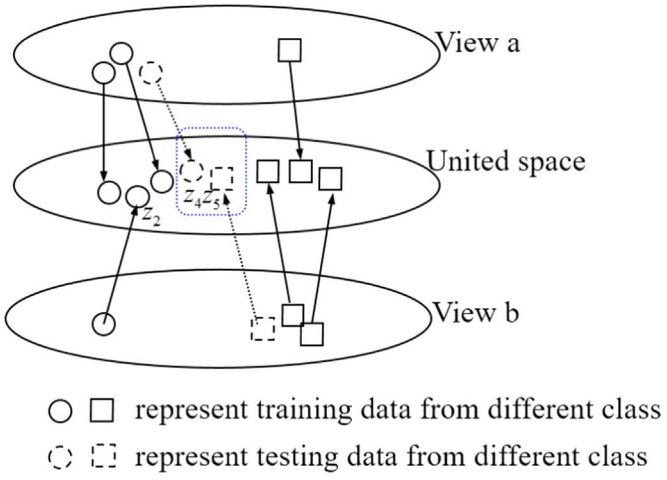

Fig. 1. Neighboring nodes in same view scattered after mapped into a unified space.

Section 4. Experimental results are shown in Section 5 and conclusions are drawn in Section 6.

2. Related work

2.1. Cross-view gait recognition

Current research on gait recognition across various viewing angles falls into three categories. The approaches in the first category rely on the reconstructed 3D gait model [3–5,15,16]. Bodor et al. [3] computed a 3D visual hull model using gait silhouettes from multiple views. Taking advantage of 3D visual hull model, they could reconstruct gait features under any required view. Zhao et al. [15] reconstructed a 3D human skeleton model by multiple cameras. From 3D models, the lengths of key segments were extracted as static features and motion trajectories of lower limbs were extracted as dynamic features. Linear time normalization was exploited for matching and recognition. Iwashita et al. [16] created a spatio-temporal 3D gait database directly to synthesize gait sequence at each viewpoint. By comparing the gait features with those in the database, the person was identified and his walking direction was estimated. All these approaches achieved high accuracy in their experiments. However, on account of expensive computation and complex camera calibration, this family of approaches is only suitable for a fully controlled and cooperative multi-camera environment such as a bio-metric tunnel [5].

The second category [6–8,17] is to extract view-invariant gait features. Jean et al. [6] proposed a method to compute view normalized feet and head 2D trajectories as view-invariant gait features. By homography based transformation, the trajectories were normalized in a lateral viewpoint for gait recognition. Goffredo et al. [8] proposed self calibrating view-invariant gait recognition based on human joint's position estimation and viewpoint rectification. Angular measurements and trunk spatial displacement were derived from the rectified limbs' poses and used as a view-invariant gait feature. Kusakunniran et al. [17] transformed gait silhouettes from arbitrary view onto the canonical view using domain transformation. Procrustes mean shape was extracted as feature to measure gait similarity. When difference between two views was large, these methods can still perform efficiently. However, these methods in the category are not applicable for front view because the gait feature from front view could not be transformed onto side view.

The third category learns mapping or projection relationship of gaits across views through a training process. The learnt relationships will normalize gait features from different views into shared or associated subspace(s) before gait similarity is measured. Different from the first category, the third category needs only one simple camera. The relationship between gait data from different views is established through the learning process. Compared with the second category, the third category is more efficient and stable. In addition, there is no limitation to the view point. Recent researches in the third category mainly rely on view transform model (VTM) [9–11,18,19] and subspace learning [12,20–24]. VTM was introduced [18] to transform gait features from one view into another view. Frequency-domain gait features from different views were used to form a large matrix. Then the matrix was factorized by adopting singular value decomposition (SVD) to establish the VTM. Kusakunniran et al. [10] created a VTM using support vector regression based on local dynamic feature extraction. Sparse regression-based VTM was also proposed [11] to obtain stable model for VTM construction. Lately, Muramatsu et al. [19] developed an arbitrary view transform model (AVTM) by combining aspects of both first and third categories. 3D gait visual hulls were established and used to generate training gait sequences under any required views. Then VTM was constructed to transform features.

Recently, subspace learning based approaches have been adopted to transform the gait features obtained from various viewing spaces into a shared feature space. Bashir et al. [12], using canonical correlation analysis (CCA), learnt maximally correlated feature subspaces and employed correlation strength to measure gait similarity. A view-invariant discriminative projection (ViDP) method was introduced [22] to learn a low dimensional feature subspace. It was implemented by iteratively learning the low dimensional geometry and finding the optimal projection according to the geometry. Hu [23] proposed an uncorrelated multilinear sparse local discriminant canonical correlation analysis (UMSLDCCA) approach to model the correlations of gait features from different viewing angles. A tensor-to-vector projection (TVP) was adopted to extract gait features for measuring similarity. Xing et al. [24] proposed complete CCA (C3A) to overcome the singular problem of covariance matrix and alleviate the computational burden of high dimensional matrix for typical gait image data. Compared with VTM, subspace learning based methods can cope with feature mismatch across views and are more robust against feature noise.

2.2. Subspace learning

Subspace learning sheds light on various tasks in computer vision and multimedia. It projects the original feature space(s) to a new subspace, wherein specific statistical properties can be well preserved. The representative approaches include principal component analysis (PCA) [25], Linear Discriminant Analysis (LDA) [26] and Locality preserving projections (LPP) [27]. PCA is an unsupervised dimension reduction algorithm, while LDA is a supervised one. LPP can either be supervised or unsupervised by changing similarity matrix. LPP has been applied in many applications. However, it does not work well when data are from two different sets (such as gait data from two different views). The main reason is that the similarity matrix could not be constructed well between two different sets.

CCA is probably the most popular two-view subspace learning method due to its wide-spread use in cross-view biometric recognition [28,29], cross-media retrieval [30,31] and some other problems [32]. Li et al. [29] applied CCA to face recognition based on non-corresponding region matching. An et al. [32] used regularized CCA (RCCA) in conjunction with a reference set for person re-identification. Luo et al. [30] developed tensor CCA (TCCA) by analyzing the high-order covariance tensor over the data from all views. A more reliable common subspace shared by all features could be obtained.

Besides CCA, there are some other methods for two-view subspace learning. Sharma and Jacobs [33] used partial least squares (PLS) to linearly map images from different modalities to a common linear subspace in which they are highly correlated. Tenenbaum and Freeman [28] proposed a bilinear model (BLM) to derive a unified space for cross-modal face recognition. Li et al. [13] proposed coupled metric learning (CML) approach to learn a common space in which degraded