ARTICLE IN PRESS
JID: PATREC [m5G;October 27, 2017;11:34]
Pattern Recognition Letters 000 (2017) 1–8
Contents lists available at ScienceDirect
Pattern Recognition Letters
journal homepage: www.elsevier.com/locate/patrec

Multiview max-margin subspace learning for cross-view gait recognition

Wanjiang Xu[a,b,*], Canyan Zhu[a], Ziou Wang[a]

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

ARTICLE INFO

Article history:
Available online xxx

MSC:
41A05
41A10
65D05
65D17

Keywords:
Subspace learning
Gait recognition
Transform matrices

ABSTRACT

Cross-view gait recognition can be regarded as a domain adaption problem, in which, probe gait to be recognized in one view is different from gallery gaits collected in another view. In this paper, we present a subspace learning based method, called Multiview Max-Margin Subspace Learning (MMMSL), to learn a common subspace for associating gait data across different views. A group of projection matrices that respectively map data from different views into the common subspace are optimized via simultaneously minimizing the within-class variations and maximizing the local between-class variations of the low-dimensional embeddings from both inter-view and intra-view. In the learnt subspace, same-class samples from all views cluster together, and each different-class cluster is kept away from its nearest neighbors as far as possible. Experimental results on two benchmark gait databases, CASIA-B and OU-ISIR, demonstrate the effectiveness of the proposed method. Extensive experiments also show that our MMMSL achieves significant improvements compared with related subspace learning based methods.

1. Introduction

Gait is one of the few biometric features that can be measured remotely without physical contact and proximal sensing. Gait recognition is to identify a person by his (her) walking manner. It is useful in many applications such as robot vision and intelligent security system. However, in reality, there are some factors significantly affecting human gait including walking speed, dressing, carrying objects, and viewing angle change [30]. Among these factors, viewing angle change has been regarded as one of the most challenging problems for gait recognition [42]. This is because walking sequences of one person catched by a camera is a 2D gait, and the appearance of 2D gait observed varies a lot from one viewing angle to another one.

Recognizing gaits across different views can be considered as domain adaption problems, in which, gallery gaits (termed as source domain) and probe gaits (termed as target domain) follow different distributions. Domain adaptation is also more generally known as transfer learning [28], which has been applied in various fields, e.g., computer vision [6], natural language processing [35], etc. In principle, domain adaptation attempts to transfer the rich knowledge in a source domain to target domain with limited information to induce a better model. During the learning process, domain adaptation makes use of information coming from both source and target domains. To bridge the two domains, an efficient scheme is exploring the commonality of both domain [20]. For example, a common structure for two different modalities was learned in the work [13] to reduce the sematic gap via pairwise constraints. A common dictionary [43] in the low-dimensional space was formed for domain adaptive sparse representation. To discover the common knowledge, specifically, a common subspace is always acquired, in which the structures of both domains are preserved and the disparity is reduced [20,33].

The most typical approach to obtain a common subspace for multiple views should be the Canonical Correlation Analysis (CCA). CCA attempts to learn two transformations, one for each view, to respectively project samples from two views into a common subspace, by maximizing the cross correlation between two views. In addition, several variants of CCA have been proposed for cross-view gait recognition, such as kernal CCA (KCCA) [3] and discriminant CCA (DCCA) [15]. CCA and its variants are only applicable for two-view scenario. To deal with multi-view (more than two views) cases, the pairwise strategy is usually exploited to convert one common subspace for v views problem to C_v^2 common subspaces problem, which costs more computational resources. A more efficient and robust solution is to learn a unified common subspace shared by all views rather than two views. For this purpose, Multiview CCA (MCCA) [29] was proposed to obtain one common space

* Corresponding author.
E-mail address: xuwanjiang0821@163.com (W. Xu).

https://doi.org/10.1016/j.patrec.2017.10.033
0167-8655/

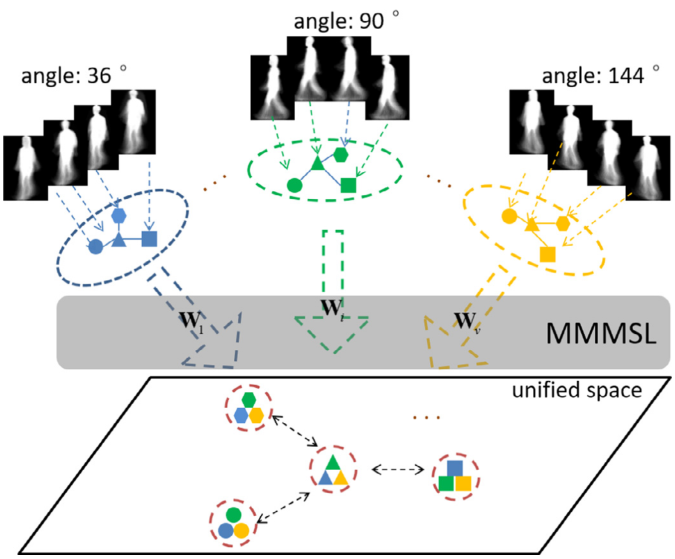

Fig. 1. The overview of multiview max-margin subspace learning for gait recognition. Gait samples are represented by Gait Energy Image (GEI) [12]. Gaits from different views are projected into a common subspace. In the common subspace, gaits of same person from various viewing angles are collapsed into a sphere, and different spheres keep away from each other.

by maximizing the total correlations between any two views. Compared with pairwise CCA, MCCA obtains only v view-specific transforms rather than C_v^2 pairs of two view-specific transforms.

Although MCCA can learn a common subspace shared by all views, the learnt subspace may be unfavorable for cross-view gait recognition. This is because MCCA does not consider discriminant information, e.g., class label, which is beneficial for recognition or classification. To address this problem, we propose a Multiview Max-Margin Subspace Learning (MMMSL) method that can learn a discriminant common subspace. As seen in Fig. 1, gait features from different views can be mapped into the common subspace then they could be matched directly. In addition, the learnt subspace is a discriminant subspace, in which same-class gait samples cluster together and keep away from neighboring clusters. Compared with previous works, the proposed method has the following properties. 1) A common subspace shared by multiple views is obtained for cross-view gait recognition by jointly optimizing v view-specific transforms. 2) In the learning process, the max margin constraint between neighboring different-class samples brings a more discriminant subspace. 3) The proposed method achieves pretty performances on CASIA-B [42] and OU-ISIR [24] gait datasets. The scores achieved by the proposed method on two datasets are better than that of related subspace learning methods.

The rest of this paper is organized as follows. Related works are reviewed in Section 2. In Section 3, problem statement and formulation are analyzed. Experimental results are shown in Section 4, and conclusion is drawn in Section 5.

2. Related works

In the recent literature, approaches to gait recognition can be grouped into two categories, model-based methods [1,36,40] and appearance-based methods [18,22,38]. The model-based methods generally characterize kinematics of human joints in order to measure physical gait parameters such as trajectories, limb lengths, and angular speeds. These methods were prevail in the early researches. However, human body is a highly flexible structure, and it is difficult to accurately locate the joints position. In addition, only a small amount of parameters may be not enough to uniquely identify a person. Instead, appearance-based methods extract gait representations directly from video. Compared with the former, the latter has demonstrated better performance on the common databases. However, all these methods meet with the difficulty of view change.

A variety of approaches have been proposed to solve the problem of viewing angle change, which can be generally classified into three main categories. The first category [5,17,31,44] is to construct 3D gait model through cooperative multi-camera. Sufficient 3D information guarantees the promising performance for gait recognition. However, the methods based on 3D reconstruction require complex camera calibration and expensive computation. This restricts greatly their application in many scenarios. Approaches in the second category [11,18,23] aim to extract gait features that are invariant to view change. These approaches can perform well for their specific scenarios. But usually it is hard to generalize for other cases. The third category [21,37,38] relies on learning projection relationships of gait across views. Through a training process, gait features from different viewing spaces are projected into one or more common canonical subspaces before gait similarity is measured. Compared with the former two, the third category only needs a simple non-cooperative camera system and has more efficient and stable if sufficient training samples are supplied. The method proposed in this paper belongs to this category.

In the third category, there are two kinds of methods. One is based on view transform model (VTM), and the other is based on subspace learning. VTM can transform gait features from one view into another. Makihara et al. [25] used frequency-domain gait features from different views to form a large matrix. Then they factorized the matrix by adopting singular value decomposition (SVD) to establish the VTM. Kusakunniran et al. [21] created a VTM using support vector regression based on local dynamic feature extraction. Lately, Muramatsu et al. [26] 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.

Different from VTM, subspace learning based approaches attempt to transform features from various viewing spaces into a shared feature subspace. Bashir et al. [3] using CCA learned maximally correlated feature subspaces and employed correlation strength to measure gait similarity. Hu [15] 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. [38] 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.

On the other hand, subspace learning methods for multiview data have widely been applied in many other scenarios such as multiview face recognition [8,10,19], text-image retrieval [32]. Among them, one of the best known approach is canonical correlated analysis (CCA). CCA is a two-view subspace learning method. For multi-view (more than two views) gait recognition, they can be extended by using pairwise strategy. However, such a pairwise manner is neither efficient nor optimal for classification across different views. What we need is a unified semantic common space, which should embody invariant features or attributes that can identify the underlying object, commonly shared by all the views rather than only two views. For this purpose, Multiview CCA (MCCA) [27] was proposed to obtain one common subspace for all views by maximizing the total correlations between any

Please cite this article as: W. Xu et al., Multiview max-margin subspace learning for cross-view gait recognition, Pattern Recognition Letters (2017), https://doi.org/10.1016/j.patrec.2017.10.033