Pattern Recognition 66 (2017) 280–294

Contents lists available at ScienceDirect

Pattern Recognition

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

A general subspace ensemble learning framework via totally-corrective boosting and tensor-based and local patch-based extensions for gait recognition

Guangkai Ma[a], Ligang Wu[a,*], Yan Wang[b]

[a] Research Institute of Intelligent Control and Systems, Harbin Institute of Technology, Harbin 150001, China
[b] College of Computer Science, Sichuan University, Chengdu, China

ARTICLE INFO

Keywords:
Subspace learning
Boosting
Ensemble learning
Pattern recognition
Gait recognition

ABSTRACT

Subspace learning methods have played an important role on handling the high-dimensional gait template for human identification. Particularly, linear discriminant analysis (LDA) method has been widely applied to find one discriminant low-dimensional subspace for gait recognition. However, when gait templates changed by clothing, view angle, load carrying and road surface variants, only learning one single subspace is prone to dropping into local optimum. In this paper, a subspace ensemble learning using totally-corrective boosting (SEL_TCB) framework and its tensor-based and local patch-based extensions are proposed for gait recognition. In this framework, multiple discriminant subspaces are iteratively learned using totally-corrective boosting technology to preserve the proximity relationships described by instance triplets. Meanwhile, by constructing different triplet set, the presented framework can deal with complex application environments. Further, we extend SEL_TCB framework to tensor SEL_TCB (TSEL_TCB) framework which effectively preserves the structural information of the gait template. Meanwhile, compared with the holistic appearance-based SEL_TCB framework, a local patch-based SEL_TCB (LPSEL_TCB) framework is proposed, which iteratively learns multiple discriminant subspaces corresponding to several local patches selected from the gait template. The proposed method is compared with the recently published gait recognition approaches on USF HumanID database and CASIA Gait database. Experimental results indicate that the proposed method achieves highly competitive performance against state-of-the-art gait recognition approaches.

1. Introduction

The gait analysis has widely applied to a variety of applications. For example, in human identification applications the walking patterns, termed gait as a behavioral biometric are characterized to identify people from image sequences [1–3]. For the diagnosis of neuro-degenerative diseases such as Parkinson's disease, biomedical signals are extracted from the gait in a non-invasive way [4–7].

As an effective tool for human identification at a distance, gait recognition has recently attracted wide attention in the computer vision field [8–10]. In particular, gait has the irreplaceable advantage for the availability at a distance in visual surveillance [8]. Due to limitation in environmental condition, camera view, hardware device and people compatibility, some physiological biometrics such as face, iris and fingerprint are inadequate ability to identify people at a distance. Although gait recognition can perform at a distance without user cooperation, it is still a challenging task, due to various condition changes, such as shoes, clothing, view angle, load carrying, road surface, etc [1].

Silhouette images of walking human have proven to be very useful when used to identify people at a distance [9,11–13]. During the last decade, many silhouette based methods have been proposed [9,14–17]. For instance, Han and Bhanu [9] extracted gait energy image (GEI) by averaging gait silhouettes across a gait cycle for individual recognition. In Liu and Sarkar [14], a population Hidden Markov Model (pHMM) is trained using manually extracted silhouettes for gait dynamics normalization. Huang and Boulgouris [15] proposed shifted energy image (SEI) to normalize the horizontal centers of different areas in GEI. Wang et al. [16] proposed a temporal template, named Chrono-Gait Image (CGI), which utilizes a multichannel mapping function to extend GEI template for preserving more temporal information in a gait sequence. Choudhury and Tjahjadi [17] applied Gaussian filter to a GEI to generate a multiscale gait image in order to improve the robustness to carrying conditions and clothing variation.

* Corresponding author.
E-mail addresses: maguangkai@gmail.com (G. Ma), ligangwu@hit.edu.cn (L. Wu), ywang@scu.edu.cn (Y. Wang).

http://dx.doi.org/10.1016/j.patcog.2017.01.003
Received 28 May 2016; Received in revised form 6 December 2016; Accepted 2 January 2017
Available online 10 January 2017
0031-3203/

In the silhouette based gait recognition, silhouette images or the gait templates such as GEI are often high-dimensional. To deal with those high-dimensional data, an efficient feature reduction algorithm is necessary [18,19]. Veres et al. [18] analyzed the importance for gait recognition of image information by using two statistical analysis methods but not for the purpose of classification. Dupuis et al. [19] ranked features' importance by using the Random Forest algorithm to select the most discriminant features. In Han and Bhanu [9], linear discriminant analysis (LDA) is used to learn a discriminant low-dimensional subspace where between-class scatter matrix is maximized while within-class scatter matrix is minimized. To preserve the matrix structure of an image, the vector-based discriminant analysis is extended to the matrix-based one by [20–22]. Some methods further perform the multilinear analysis on third-order tensor to fully preserve the spatial and temporal correlations of the silhouette sequence [23–25]. However, when silhouettes suffer to significant transformation differences caused by clothing, view angle, load carrying or road surface variants, these methods generally obtain a suboptimal subspace rather than a global optimal one [26].

To address these problems, several subspace ensemble learning methods [26–29] have been proposed to learn a complex manifold by combining the multiple linear subspaces by using random sampling, boosting and clustering techniques. In [27], Adaboost algorithm is utilized to combine a number of low-rank positive semidefinite (PSD) matrices for metric learning. However, for high-dimensional data such as GEI, training each rank-one PSD matrix is computationally expensive. In [28], to reduce the effect of covariate factors in gait recognition, Guan et al. proposed a random subspace learning method to combine amount of local enhancing classifiers in the randomly extracted subspaces, which achieve the excellent recognition performance. They also proposed the hybrid decision-level fusion strategy to further improve the accuracy. However, it often needs a large number of random subspaces and evenly combine them, which is very inefficient. On the other hand, they indicated that local regions have different discriminative abilities to different application conditions, and in some conditions certain local region is even more effective than the whole gait in [30,31]. Rida et al. [32] proposed to utilize group Lasso learning algorithm to select the most discriminative human body part.

In this paper, we present a novel subspace ensemble learning using totally-corrective boosting (SEL_TCB) framework and its tensor and local patch-based extensions to learn multiple discriminant subspaces for gait recognition. We aim to learn and calculate the weighted optimum combination of the discriminant features in the multiple subspaces. First, we construct the triplet set to describe the class relationships among training samples. Specifically, for each instance triplet (i, j, k), the class label of instance i is the same with instance j and is different from instance k. Then, by using totally-corrective technique [33], we iteratively learn the multiple subspaces based on the different weight distributions on the triplet set, in order to preserve the proximity relationships among instance triplets: instance i is closer to instance j than to instance k. Finally, we give the weighted combination of the learned subspaces for recognition. Further, we propose the tensor SEL_TCB (TSEL_TCB) framework to cover the tensor data. Meanwhile, the local patch-based SEL_TCB (LPSEL_TCB) framework is proposed to solve the local patch-based subspace ensemble learning problem. The experimental results on USF HumanID database [8] and CASIA gait database [34] clearly demonstrate the efficacy of the proposed algorithms. To the best of our knowledge, the totally-corrective boosting technique is first applied to subspace ensemble learning for biometric authentication.

We presented a preliminary version of patch-based extension of the new algorithm in [35]. Compared with the previous work, we present several novelties in the method as described below. Moreover, we extensively evaluate the sensitivity of our method with respect to different parameters, and also validate our method with an additional CASIA gait dataset [34].

- We propose a general subspace ensemble via totally-corrective learning framework based on the vector form of the feature representation.
- This paper presents three triplet building methods which increase the training efficiency of the proposed method.
- We propose a tensor extension of the proposed subspace ensemble learning algorithm which can effectively learn and combine multiple subspaces for the feature representation in the tensor form.

The rest of the paper is organized as follows: Section 2 details the SEL_TCB framework and the building method of the triplet set. Its tensor-based and local patch-based extensions are presented in Section 3. Experiments are performed and analyzed in Section 4. Finally, the conclusion is drawn in Section 5.

2. Subspace ensemble learning using totally-corrective boosting

2.1. Subspace ensemble learning framework

A 2D image $\mathbf{I}$ is rearranged in column to a vector $\mathbf{w} \in \mathbb{R}^d$, where d is the number of pixels in $\mathbf{I}$. Consider a training set $\{(\mathbf{w}_i, c_i)|\mathbf{w}_i \in \mathbb{R}^d, c_i \in [1, 2, \ldots, C], i = 1, 2, \ldots, N\}$, where $\mathbf{w}_i$ is the i-th training sample, c_i is the corresponding class label, and C and N are respectively the number of classes and samples. Let matrix $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_N] \in \mathbb{R}^{d\times N}$. Let π_c and n_c respectively denote the index set and the number of samples belonging to the c-th class. First, we construct the triplet index set of samples $\mathcal{T} = \{(i_m, j_m, k_m)|c_{i_m} = c_{j_m}, c_{i_m} \neq c_{k_m}, m = 1, 2, \ldots, M\}$, where M is the cardinality of $\mathcal{T}$. The SEL_TCB framework learns a set of subspace projection matrices $\{\mathbf{V}_t|\mathbf{V}_t \in \mathbb{R}^{d\times q_t}, t = 1, 2, \ldots, l\}$ with the corresponding weight vector $\mathbf{a} = [a_1, a_2, \ldots, a_l]^{\mathrm{T}}$, and then projects $\mathbf{w}$ into multiple subspaces and combine them as follows:

$$f(\mathbf{w}) = [(\sqrt{a_1}\mathbf{V}_1^{\mathrm{T}}\mathbf{w})^{\mathrm{T}}(\sqrt{a_2}\mathbf{V}_2^{\mathrm{T}}\mathbf{w})^{\mathrm{T}}\cdots(\sqrt{a_l}\mathbf{V}_l^{\mathrm{T}}\mathbf{w})^{\mathrm{T}}]^{\mathrm{T}}. \quad (1)$$

The goal of the SEL_TCB framework is that for any triplet $(i_m, j_m, k_m) f(\mathbf{w}_{i_m})$ is more similar to $f(\mathbf{w}_{j_m})$ than to $f(\mathbf{w}_{k_m})$.

For each subspace projection matrix $\mathbf{V}_t$, $q_t < d$. We define h_{mt} as follows:

$$h_{mt} = \|\mathbf{V}_t^{\mathrm{T}}\Delta\mathbf{w}_{i_m k_m}\|^2 - F\,\|\mathbf{V}_t^{\mathrm{T}}\Delta\mathbf{w}_{i_m j_m}\|^2, \quad (2)$$

where $F > 0$ is the tuning parameter, $\Delta\mathbf{w}_{i_m j_m} \triangleq \mathbf{w}_{i_m} - \mathbf{w}_{j_m}$ and $\|\cdot\|$ denotes the ℓ_2 norm operator. Then, according to the totally-corrective boosting algorithm, the subspace ensemble problem can be formulated through optimizing the soft margin defined in the following linear programming (LP)

$$\max_{\mathbf{a},\zeta,\rho} \rho - D\sum_{m=1}^{M}\zeta_m \quad \text{s. t.} \quad \sum_{t=1}^{l} a_t h_{mt} \geq \rho - \zeta_m, \quad m = 1, 2, \ldots, M$$
$$\sum_{t=1}^{l} a_t = 1, \quad \zeta_m \geq 0, \quad m = 1, 2\ldots, M; \quad a_t \geq 0, \quad t = 1, 2, \ldots, l \quad (3)$$

where ζ_m is the slack variable to enable soft margin, and $D > 0$ is the regularization parameter which penalizes the slack variables. If F is large, then the subspace tends to minimize the distance between samples from the same class. On the other hand, if F is small, then the subspace prefers to maximizing the distance between samples from different classes. The norm of each column vector of $\mathbf{V}_t$ is normalized to 1, i.e., $\mathbf{V}_t^{\mathrm{T}}\mathbf{V}_t = \mathbf{I}$.

The Lagrangian dual problem of (3) can be defined as