Neurocomputing 224 (2017) 119–127

Contents lists available at ScienceDirect

Neurocomputing

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

Subspace ensemble learning via totally-corrective boosting for gait recognition

Guangkai Ma[a,1,*], Yan Wang[b,1], Ligang Wu[a]

[a] *Space Control and Inertial Technology Research Center, Harbin Institute of Technology, Harbin 150001, China*
[b] *College of Computer Science, Sichuan University, Chengdu, China*

ARTICLE INFO

Communicated by Jiwen Lu

Keywords:
Subspace learning
Totally-corrective boosting
Ensemble learning
Gait recognition

ABSTRACT

Human identification at a distance has recently become a hot research topic in the fields of computer vision and pattern recognition. Since gait patterns can operate from a distance without subject cooperation, gait recognition has most widely been studied to address this problem. In this paper, a subspace ensemble learning using totally-corrective boosting (SEL_TCB) framework and its kernel extension are proposed for gait recognition. In this framework, multiple subspaces are iteratively learned with different weight distributions on the triplet set using totally-corrective technology, in order to preserve the proximity relationships among instance triplets. Further, we extend the SEL_TCB framework to the kernel SEL_TCB (KSEL_TCB) framework which can deal with the nonlinear manifold of data. We compare our method with the recently published gait recognition approaches on USF HumanID Database. Experimental results indicate that the proposed method achieves highly competitive performance against the state-of-the-art gait recognition approaches.

1. Introduction

Computer vision-based gait recognition is a task that enables computers to automatically identify different people from image sequences by characterizing their walking patterns, termed gait. In comparison with current physiological biometrics such as face, iris and fingerprint, gait has the irreplaceable advantages in visual surveillance. It can perform at a distance without user cooperation, while other biometrics generally require a cooperative subject, ideal environmental condition, fixed view angle and physical contact or close proximity. So far, gait is probably the only available biometric feature from a great distance. However, owning the above loose requirements means that gait will easily suffer from various condition changes, such as shoes, clothing, view angle, load carrying, road surface, etc. In addition, special physical conditions such as injury can also change gait. Therefore, gait recognition is still a challenging task.

Over the last decade, a large number of gait recognition methods have been proposed [1–8]. These methods can be roughly classified into two categories: model-based and appearance-based approaches. Generally, model-based methods utilize the parameters of the body structure, while appearance-based approaches extract the feature representation directly from gait sequences regardless of the underlying structure. Compared with model-based methods, the main advantages of appearance-based methods are that they can be free from recovering the structural or motion model of gait, are insensitive to the quality of silhouettes and have low computational costs. For example, Han and Bhanu [3] proposed gait energy image (GEI) as the feature representation by averaging silhouettes over one gait cycle, which is not sensitive to silhouette noises. Huang and Boulgouris [7] shifted horizontal centers of different parts of GEI to enhance its robustness to carrying condition. In Ref. [9], Wang et al. proposed chrono-gait image (CGI), with the temporal information of the gait sequence. In the appearance-based gait recognition, feature representations are often high-dimensional data with an underlying low-dimensional structure. For understanding and capturing it, subspace learning has recently attracted a lot of interest in pattern recognition. Han and Bhanu [3] adopted linear discriminant analysis (LDA) to extract discriminant features. Lu et al. [8] proposed a sparse reconstruction based metric learning (SRML) to extract view-invariant features by learning a distance metric to minimize the large intra-class sparse reconstruction errors and maximize the inter-class sparse reconstruction errors simultaneously.

Existing in the majority of subspace learning techniques, a fundamental problem is that their recognition performance heavily depends on the selected training set. It is great difficult to train a single optimal subspace using the limited-size training samples in a high-dimensional image space. Specifically, if a training set contains significant transformation differences caused by clothing, view angle, load carrying and

* Corresponding author.
[1] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.neucom.2016.10.047
Received 31 March 2016; Received in revised form 15 September 2016; Accepted 30 October 2016
Available online 04 November 2016
0925-2312/

road surface variants, the intrapersonal variation manifold will become highly nonconvex and complex so that these linear, holistic appearance-based methods such as PCA [10] and LDA [11] fail to describe the protrusions and concavities of nonconvex manifolds, thereby loosing useful information for gait recognition. Although nonlinear subspace methods such as LLE [12] and LPP [13,14] achieve good performance on the training samples, they easily become overfitting and hence do not generalize well to unseen data [15]. To address this, several subspace ensemble methods [16–18] have been proposed to learn a complex manifold by assembling the multiple local linear subspaces using random sampling, boosting and clustering techniques in the previous work. However, these methods often generate a large number of local linear subspaces and evenly combine them, which can be inefficient.

In this paper, we present a novel subspace ensemble learning framework via totally-corrective boosting technique [19] to achieve high performance classifier for gait recognition. First, inspired from the graph embedding [20] we build the triplet set $\{(i, j, k)\}$ of training samples by constructing two graphs, which can effectively control the number of triplets. Then, we utilize totally-corrective technique to iteratively learn the multiple subspaces, based on the different weight distributions on the triplet set, to preserve the proximity relationships among instance triplets: instance i is closer to instance j than to instance k. In each iteration, the objective function of optimal subspaces is formulated as graph-embedding problem, which largely reduce the computational costs on objective function. Finally, we give the weighted combination of the learned subspaces to minimize the distance of points in the intraperson manifold or keep local neighbor relationships of these points and maximize the manifold margins of different people. In addition, we further extend the proposed subspace ensemble learning framework to the kernel version. The experimental results on USF HumanID database [2] clearly demonstrate the efficacy of the proposed algorithms. Our contributions in this work can be summarized as follows:

1. We have proposed a general subspace ensemble learning framework via totally-corrective boosting, which can learn multiple of subspaces to solve the large variation of data. Meanwhile, it gives a linear combination of the learned subspaces for gait recognition.
2. Based on the graph embedding algorithm, a MFA based triplet set building method is presented to construct the triplet set, which effectively describes the class relationships among training samples, and also reduce the number of triplets for the effective and efficient subspace learning.
3. Our proposed framework is extended to the kernel version to deal with the nonlinear manifold of data.
4. We have applied our approach to the widely used USF HumanID gait database and achieved competitive recognition performance with state-of-the-art methods.

The rest of the paper is organized as follows: Section 2 details subspace ensemble learning framework via the totally-corrective boosting technique and the building method of triplet set of training samples. The kernel extension of the framework is presented in Section 3. Experiments are performed and analyzed in Section 4. Finally, the conclusion is drawn in Section 5.

2. Subspace ensemble learning framework

For appearance-based gait recognition, a 2D gait image template $\mathbf{I}$ is usually rearranged, in column or row, to a vector $\mathbf{x}$ with length d in the high-dimensional image space. Consider a training sample set containing N samples $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N]$, $\mathbf{x}_i \in \Re^d$, with corresponding class labels $\{c_i\}_{i=1}^N \in [1, 2, \ldots C]$, where C is the number of classes. Let π_c and n_c respectively denote the index set and number of the samples belonging to the c-th class. To constrain the class relationship among samples, we denote the triplet index set of samples as $\mathcal{T} = \{(i_m, j_m, k_m)\}_{m=1}^M$. Each triplet (i_m, j_m, k_m) indicates that $\mathbf{x}_{i_m}$ is more similar to $\mathbf{x}_{j_m}$ than to $\mathbf{x}_{k_m}$, i.e., $c_{i_m} = c_{j_m}$ and $c_{i_m} \neq c_{k_m}$. Let $M = |\mathcal{T}|$ denote the cardinality of set $\mathcal{T}$. Ideally, SEL_TCB aims to find a weight vector $\mathbf{a} = [a_1, a_2, \ldots, a_l]^T$ and a set of subspace projection matrices $\{\mathbf{V}_t\}_{t=1}^l \in \Re^{d\times q_t}$, where $q_t \ll d$, to make the following hold for any $(i_m, j_m, k_m) \in \mathcal{T}$,

$$\sum_{t=1}^{l} a_t \ \|\mathbf{V}_t^T \Delta\mathbf{x}_{i_m k_m}\|^2 \gg F \sum_{t=1}^{l} a_t \ \|\mathbf{V}_t^T \Delta\mathbf{x}_{i_m j_m}\|^2 \tag{1}$$

where $\Delta\mathbf{x}_{i_m j_m} \triangleq \mathbf{x}_{i_m} - \mathbf{x}_{j_m}$, $\|\cdot\|$ denotes the ℓ_2 norm operator, and $F > 0$ is tuning parameter. If F is large, then the method tends to minimize the distance between samples from the same class but if F is small, then the method prefers to maximizing the distance between samples from different classes. To remove the scalability effect of the distance resulting from the matrix $\mathbf{V}_t$, the norm of each column vector of $\mathbf{V}_t$ is normalized to 1, i.e., $\mathbf{V}_t^T\mathbf{V} = \mathbf{I}$. To avoid notational ambiguity in latter equations, we introduce a vector $\mathbf{h}_m = [h_{mt}]$, where h_{mt} corresponds to $\mathbf{V}_t$ and the m-th triplet (i_m, j_m, k_m), defined as

$$h_{mt} = \|\mathbf{V}_t^T \Delta\mathbf{x}_{i_m k_m}\|^2 - F\,\|\mathbf{V}_t^T \Delta\mathbf{x}_{i_m j_m}\|^2 \tag{2}$$

We formulate the problem Eq. (1) through optimizing the soft margin defined in the following linear programming (LP), similar to the LPBoost [19].

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

where ζ_m is the slack variable to enable soft margin, and $D > 0$ is the regularization parameter which penalizes the slack variables. We further set the regularization parameter $D = \frac{1}{M\eta}$, $\eta \in [\frac{1}{M}, 1]$, where $\frac{1}{M} \leq \eta \leq 1$ is to maintain dual of Eq. (3) feasibility.

The Lagrangian dual problem of Eq. (3) can be written as

$$\max_{\mathbf{u},\beta,\mathbf{g},\mathbf{q}} \min_{\mathbf{a},\zeta,\rho} \mathcal{L} = -\rho + D\sum_{m=1}^{M}\zeta_m - \sum_{m=1}^{M} u_m\left(\sum_{t=1}^{l} a_t h_{mt} - \rho + \zeta_m\right) - \beta\left(\sum_{t=1}^{l} a_t - 1\right) - \mathbf{g}^T\zeta - \mathbf{q}^T\mathbf{a} \tag{4}$$

with $\mathbf{u} = [u_m] \geq 0$, $\mathbf{g} \geq 0$ and $\mathbf{q} \geq 0$. After the first derivative of the Lagrangian Eq. (4) w.r.t. the primal variables $\mathbf{a}$, $\boldsymbol{\zeta}$ and ρ to be equal to 0, the Lagrange dual of Eq. (3) can be written as

$$\min_{\mathbf{u},\beta} \ \beta \text{ s. t. } \sum_{m=1}^{M} u_m h_{mt} \leq \beta, \quad t = 1, \ldots, l \quad \sum_{m=1}^{M} u_m = 1, \quad 0 \leq u_m \leq D, \quad m = 1, \ldots, M \tag{5}$$

By the column generation algorithm [19], the subproblem for learning the projection matrix $\mathbf{V}_t$ is described as

$$\max_{\mathbf{V}_t} \ \sum_{m=1}^{M} u_m h_{mt} = \sum_{m=1}^{M} u_m (\|\mathbf{V}_t^T \Delta\mathbf{x}_{i_m k_m}\|^2 - F\,\|\mathbf{V}_t^T \Delta\mathbf{x}_{i_m j_m}\|^2) \tag{6}$$

After some simple matrix algebraic operations, the objective function of problem Eq. (6) can be rewritten as

$$\max_{\mathbf{V}_t} tr\left(\mathbf{V}_t^T\left[\sum_{m=1}^{M} u_m (\Delta\mathbf{x}_{i_m k_m}\Delta\mathbf{x}_{i_m k_m}^T - F\Delta\mathbf{x}_{i_m j_m}\Delta\mathbf{x}_{i_m j_m}^T)\right]\mathbf{V}_t\right) \tag{7}$$

where $tr(\cdot)$ denotes the trace operator.

Let matrix