



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 Tjahjedi [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).

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 $\{(\mathbf{w}_i, c_i) | \mathbf{w}_i \in \mathbb{R}^d, c_i \in [1, 2, \dots, C], i = 1, 2, \dots, N\}$, 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 $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_N] \in \mathbb{R}^{d \times N}$. 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 $\mathcal{T} = \{(i_m, j_m, k_m) | c_{i_m} = c_{j_m}, c_{i_m} \neq c_{k_m}, m = 1, 2, \dots, M\}$, where M is the cardinality of $\mathcal{T}$. The SEL_TCB framework learns a set of subspace projection matrices $\{\mathbf{V}_t | \mathbf{V}_t \in \mathbb{R}^{d \times q_t}, t = 1, 2, \dots, l\}$ with the corresponding weight vector $\mathbf{a} = [a_1, a_2, \dots, a_l]^T$, and then projects $\mathbf{w}$ into multiple subspaces and combine them as follows:

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

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:

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

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)

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

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., $\mathbf{V}_t^T \mathbf{V}_t = \mathbf{I}$.

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

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

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

$$\begin{aligned} \min_{\mathbf{u}, \beta} \beta \quad \text{s. t.} \quad & \sum_{m=1}^M u_m h_{mt} \leq \beta, \quad t = 1, 2, \dots, l \\ & \sum_{m=1}^M u_m = 1, \quad 0 \leq u_m \leq D, \quad m = 1, 2, \dots, M \end{aligned} \quad (5)$$

The detail derivation of (5) is provided in Appendix.

By the column generation algorithm [33], the subproblem for learning the projection matrix $\mathbf{V}_t$ is computed as

$$\max_{\mathbf{V}_t} \sum_{m=1}^M u_m h_{mt} = \text{tr} \left(\mathbf{V}_t^T \left[\sum_{m=1}^M u_m (\Delta \mathbf{w}_{imk_m} \Delta \mathbf{w}_{imk_m}^T - F \Delta \mathbf{w}_{imj_m} \Delta \mathbf{w}_{imj_m}^T) \right] \mathbf{V}_t \right) \quad (6)$$

where $\text{tr}(\cdot)$ denotes the trace operator. Let matrix

$$\Phi_t = \sum_{m=1}^M u_m (\Delta \mathbf{w}_{imk_m} \Delta \mathbf{w}_{imk_m}^T - F \Delta \mathbf{w}_{imj_m} \Delta \mathbf{w}_{imj_m}^T) \quad (7)$$

where $\Phi_t \in \mathbb{R}^{d \times d}$, and then columns of the optimal projection matrix $\mathbf{V}_t$ are the normalized eigenvectors corresponding to the first q_t largest eigenvalues $[\lambda_1, \lambda_2, \dots, \lambda_{q_t}]$ of Φ_t .

To reduce the computational cost of Φ_t and be convenient for the following analysis, we construct two directed weighted graphs: the intrinsic graph $G = \{\mathbf{W}, \mathbf{S}\}$ and the penalty graph $G^P = \{\mathbf{W}, \mathbf{S}^P\}$. Specifically, $\mathbf{W}$ is the vertex set, and $\mathbf{S} \in \mathbb{R}^{d \times d}$ and $\mathbf{S}^P \in \mathbb{R}^{N \times N}$ respectively denote the similarity matrix of the corresponding graph. According to the triplet set $\mathcal{T}$ and the corresponding weight $\mathbf{u}$, elements S_{ij} , S_{ij}^P of $\mathbf{S}$ and $\mathbf{S}^P$ are defined as follows:

$$S_{ij} = \begin{cases} \sum_{m \in \phi_{ij}} u_m, & \text{if } \phi_{ij} \neq \emptyset \\ 0, & \text{else} \end{cases}, \quad S_{ij}^P = \begin{cases} \sum_{m \in \phi_{ij}^P} u_m, & \text{if } \phi_{ij}^P \neq \emptyset \\ 0, & \text{else} \end{cases} \quad (8)$$

where $\phi_{ij} = \{m | i_m = i, j_m = j\}$ and $\phi_{ij}^P = \{m | i_m = i, k_m = j\}$. Because G and G^P are directed graphs, $\mathbf{S}$ and $\mathbf{S}^P$ can be asymmetrical matrices. We define the diagonal matrix $\mathbf{D}$ and the Laplacian matrix $\mathbf{L}$ of G and the diagonal matrix $\mathbf{D}^P$ and the Laplacian matrix $\mathbf{L}^P$ of G^P as follows:

$$\begin{aligned} \mathbf{L} &= \mathbf{D} - \mathbf{S}, \quad D_{ii} = \sum_j S_{ij} + S_{ji}, \quad \forall i \\ \mathbf{L}^P &= \mathbf{D}^P - \mathbf{S}^P, \quad D_{ii}^P = \sum_j S_{ij}^P + S_{ji}^P, \quad \forall i \end{aligned} \quad (9)$$

where $\mathbf{S} = \mathbf{S} + \mathbf{S}^T$ and $\mathbf{S}^P = \mathbf{S}^P + \mathbf{S}^{P^T}$. Then, Φ_t in (7) can be rewritten in the form of matrix as

$$\begin{aligned} \Phi_t &= \sum_{i=1}^N \sum_{j=1}^N S_{ij}^P (\mathbf{w}_i - \mathbf{w}_j)(\mathbf{w}_i - \mathbf{w}_j)^T - F \sum_{i=1}^N \sum_{j=1}^N S_{ij} (\mathbf{w}_i - \mathbf{w}_j)(\mathbf{w}_i - \mathbf{w}_j)^T \\ &= \mathbf{W} \mathbf{L}^P \mathbf{W}^T - F \mathbf{W} \mathbf{L} \mathbf{W}^T = \mathbf{W} (\mathbf{L}^P - F \mathbf{L}) \mathbf{W}^T \end{aligned} \quad (10)$$

Note when the dimensionality d is high, directly solving the eigenvalue problem of $d \times d$ matrix Φ_t is intractable. Fortunately, we can determine the first q_t eigenvectors by solving eigenvalue problem of a much smaller $N \times N$ matrix and making linear combinations of the resulting vectors. Let matrix $\Psi_t = (\mathbf{L}^P - F \mathbf{L}) \mathbf{W}^T \mathbf{W} \in \mathbb{R}^{N \times N}$. After calculating the eigenvector $\mathbf{v}$ corresponding the eigenvalue λ of Ψ_t , the eigenvector $\mathbf{v}'$ and the corresponding eigenvalue λ' of Φ_t can be determined as $\mathbf{v}' = \mathbf{W} \mathbf{v}$, $\lambda' = \lambda$.

Algorithm 1. Subspace Ensemble Learning via Totally-corrective

Boosting.

Input: The training samples $\mathbf{W}$, the triplet index set $\mathcal{T}$ and the maximum iteration number L .

Initialization: Initialize u_m to be $1/M$.

for $l=1$ **to** L **do**

- Based on $\mathcal{T}$ and $[u_m]$, construct matrices $\mathbf{S}$ and $\mathbf{S}^P$ according to (8);

- Calculate the optimal projection matrix $\mathbf{V}_l$ by solving the eigenvectors of Φ_l in (10);

- Calculate $[h_{ml}]$ according to (2)

if $\sum_{m=1}^M u_m h_{ml} \leq \beta$ **then** Break; **end if**

- Add h_{ml} to the dual problem (5);

- Minimize β and update $[u_m]$ according to (5) using the primal-dual interior-point LP solver;

end for

Calculate the weight vector $\mathbf{a}$ according to (3).

Output: The weight vector $\mathbf{a}$ and the projection matrix set $\{\mathbf{V}_l | l = 1, 2, \dots, L\}$.

As a summary, the algorithm for the proposed SEL_TCB framework is presented in Algorithm 1. We define the stopping criterion $\sum_{m=1}^M u_m h_{ml} \leq \beta$ that guarantees that the optimal subspace ensemble has been found. Specifically, in each iteration (l), if after learning projection matrices, $\sum_{m=1}^M u_m h_{ml} \leq \beta$ or reach the maximum number of iterations, the algorithm will terminate. The criteria $\sum_{m=1}^M u_m h_{ml} > \beta$ guarantees that the value of the objective $\rho - D \sum_{m=1}^M \zeta_m$ can be increased, and the maximum number of iterations can prevent over-fitting. We iteratively learn multiple subspaces based on the column generation technique in linear programming. At each iteration, the weights of all the learned subspaces are updated. Finally, we explicitly give the weighted combination of the learned subspaces.

It is worthwhile noting that at each iteration, the learned projection matrix $\mathbf{V}_t$ subjects to:

$$\max_{\mathbf{V}_t} \text{tr}(\mathbf{V}_t^T \mathbf{W} \mathbf{L}^P \mathbf{W}^T \mathbf{V}_t) - F \text{tr}(\mathbf{V}_t^T \mathbf{W} \mathbf{L} \mathbf{W}^T \mathbf{V}_t) \quad (11)$$

In the above formula, learning the projection matrix $\mathbf{V}_t$ can essentially be regarded as linearization of the graph embedding [36], which finds a linear low-dimensional subspace to preserve the similarity property of the intrinsic graph G and at the same time suppress the similarity characteristics described by the penalty graph G^P .

2.2. Building the triplets

According to the given class labels $\{c_i\}$, we can totally obtain $\sum_{c=1}^C n_c(n_c - 1)(N - n_c)$ triplets. If all the triplets are used to train SEL_TCB, then the objective function in (6) is updated as

$$\max_{\mathbf{V}_t} \text{tr} \left(\mathbf{V}_t^T \left(\sum_{c=1}^C \sum_{i \in \pi_c, k \notin \pi_c} u_{ik} \Delta \mathbf{w}_{ik} \Delta \mathbf{w}_{ik}^T - F \sum_{c=1}^C \sum_{i, j \in \pi_c, i \neq j} u'_{ij} \Delta \mathbf{w}_{ij} \Delta \mathbf{w}_{ij}^T \right) \mathbf{V}_t \right) \quad (12)$$

where $u_{ik} = \sum_{i_m=i, k_m=k} u_m$ and $u'_{ij} = \sum_{i_m=i, j_m=j} u_m$.

According to (12), the purpose of learning matrix $\mathbf{V}_t$ is to maximize the covariance matrix of intrapersonal subspace while to minimize covariance matrix of extrapersonal subspace in the projective feature space, based on the training samples with different weight distributions [37]. However, when the number of the training samples is large, the total number of triplets would be so large that solving the problem (5) becomes computationally infeasible. Therefore, when dealing with large training set, we should build a reasonable triplet set to effectively learn subspaces, similar to (12), while avoid the number of triplets from becoming too large.

LDA-based triplet set: LDA finds the most discriminant subspace $\mathbf{B}$ by maximizing the ratio between the interclass scatter matrix S_b and the intraclass scatter matrix S_w , as follows:

$$\begin{aligned} \mathbf{B} &= \arg \max_{\mathbf{B}} \frac{\mathbf{B}^T S_b \mathbf{B}}{\mathbf{B}^T S_w \mathbf{B}} \\ S_w &= \sum_{c=1}^C \sum_{i \in \pi_c} (\mathbf{m}_c - \mathbf{w}_i)(\mathbf{m}_c - \mathbf{w}_i)^T, \\ S_b &= \sum_{c=1}^C n_c (\mathbf{m}_c - \mathbf{m})(\mathbf{m}_c - \mathbf{m})^T \end{aligned} \quad (13)$$

where $\mathbf{m}_c = \frac{1}{n_c} \sum_{i \in \pi_c} \mathbf{w}_i$ is the sample center of the c -th class and $\mathbf{m} = \frac{1}{C} \sum_{c=1}^C \mathbf{m}_c$ is the center of all samples. For simplicity, we assume that each class has the same sample number: $n_c = n_0$, $c = 1, 2, \dots, C$. S_b is identical to [38]

$$S_b = \frac{n_0}{2C} \sum_{i=1}^C \sum_{j=1}^C (\mathbf{m}_i - \mathbf{m}_j)(\mathbf{m}_i - \mathbf{m}_j)^T \quad (14)$$

It may be seen from S_w in (13) and S_b in (14) that the purpose of LDA is to maximize distances between different class centers and minimize distances between each sample in a class and its class center. Therefore, we add class centers $\mathbf{M} = [\mathbf{m}_1, \mathbf{m}_2, \dots, \mathbf{m}_C]$ into the training set $\mathbf{W}$ and obtain the extended training set $\mathbf{W}' = [\mathbf{M}, \mathbf{W}]$. Then, based on the set $\mathbf{W}'$, the LDA-based triplet set is constructed as $\mathcal{T} = \{(i_m, j_m + C, k_m) | i_m \in \{1, \dots, C\}, j_m \in \pi_{i_m}, k_m \neq i_m \in \{1, 2, \dots, C\}\}$. The cardinality of the LDA-based triplet set is reduced from $\sum_{c=1}^C n_c(n_c - 1)(N - n_c)$ to $(C - 1)N$.

MFA-based triplet set: Based on the graph embedding theory, margin fisher analysis (MFA) [36] devises the similarity matrix $\mathbf{S}$ of G and $\mathbf{S}^p$ of G^p to characterize intraclass compactness and interclass separability with the margin criterion as follows:

$$\begin{aligned} S_{ij} &= \begin{cases} 1, & \text{if } i \in N_{k_1}^+(j) \text{ or } j \in N_{k_1}^+(i) \\ 0, & \text{else} \end{cases} \\ S_{ij}^p &= \begin{cases} 1, & \text{if } (i, j) \in P_{k_2}(c_i) \text{ or } P_{k_2}(c_j) \\ 0, & \text{else} \end{cases} \end{aligned} \quad (15)$$

where $N_{k_1}^+(j)$ indicates the index set of the k_1 nearest neighbors of $\mathbf{w}_j$ in the same class and $P_{k_2}(c)$ is a set of data pairs that are the k_2 nearest pairs among the set $\{(i, j), i \in \pi_c, j \notin \pi_c\}$.

Similar to MFA, we redefine $\mathbf{S}$ and $\mathbf{S}^p$ as follows:

$$S_{ij} = \begin{cases} 1, & \text{if } j \in N_{k_1}^+(i) \\ 0, & \text{else} \end{cases}, \quad S_{ij}^p = \begin{cases} 1, & \text{if } j \in N_{k_2}^-(i) \\ 0, & \text{else} \end{cases} \quad (16)$$

where $N_{k_2}^-(i)$ indicates the index set of the k_2 nearest neighbors of $\mathbf{w}_i$ from different classes. According to $\mathbf{S}$ and $\mathbf{S}^p$ in (16), for any (i, j, k) , $i, j, k \in \{1, 2, \dots, N\}$, if $S_{ij} = 1$ and $S_{ik}^p = 1$, then we select (i, j, k) as an element of the triplet set, i.e., $\mathbf{w}_j$ is one of k_1 nearest neighbor samples of $\mathbf{w}_i$ in the same class and $\mathbf{w}_k$ is one of k_2 nearest neighbor samples of $\mathbf{w}_i$ in the different class. The cardinality of the MFA-based triplet set is $k_1 k_2 N$. In addition, our proposed framework may also combine nonlinear discriminant subspaces such as LDE [39] and DLLE [24] by constructing different triplet sets, according to the corresponding the intrinsic and penalty graphs.

Special triplet set: In the test phase, the gait sequences of the gallery set are commonly captured in the same condition, which is reasonable to guarantee the adequate recognition accuracy and is easy to be available in the most applications. In this case, the test sample only needs to be compared with samples from different individuals in one certain condition. According to the trait of the gallery set, we construct the triplet set. For each triplet (i, j, k) , the corresponding instance $\mathbf{w}_k$ and $\mathbf{w}_i$ are from different individuals but captured in the same condition with the gallery set, and the instance $\mathbf{w}_j$ and $\mathbf{w}_i$ are from the same individual but captured in the different conditions. The cardinality of the triplet set is $(C - 1)(N - C)$.

3. Tensor-based and local patch-based extensions for SEL_TCB

To enhance discrimination and robustness to noise [40], Gabor features of the gait template are extracted. Specifically, we perform convolution transform on the gait image template $\mathbf{I}$ with a set of 2D Gabor functions $\{\psi_{\mu, \nu}(x, y)\}$ at different scale ν and different orientation μ and extract magnitude parts as follows [23]:

$$\begin{aligned} O_\mu(x, y) &= \left| \mathbf{I}(x, y) * \frac{1}{n_\nu} \sum_{\nu} \psi_{\mu, \nu}(x, y) \right| \\ O_\nu(x, y) &= \left| \mathbf{I}(x, y) * \frac{1}{n_\mu} \sum_{\mu} \psi_{\mu, \nu}(x, y) \right|, \\ O(x, y) &= \left| \mathbf{I}(x, y) * \frac{1}{n_\nu n_\mu} \sum_{\nu} \sum_{\mu} \psi_{\mu, \nu}(x, y) \right| \end{aligned} \quad (17)$$

where $|\cdot|$ denotes magnitude operator, $*$ denotes convolution operator, the variable μ and ν are respectively orientation factor and scale factor of Gabor function $\psi_{\mu, \nu}(x, y)$, the variable n_μ and n_ν are the number of orientations and scales, and $O_\mu(x, y)$, $O_\nu(x, y)$ and $O(x, y)$ are transformation outputs. In this paper, as discussed in [23], five scales $\nu \in \{0, 1, 2, 3, 4\}$ and eight orientations $\mu \in \{0, 1, 2, 3, 4, 5, 6, 7\}$ are adopted, and the number $N_2 = 14$ of Gabor features $\{O_\nu, O_\mu, O\}$.

3.1. Tensor SEL_TCB

Based on tensor representation, the multilinear analysis methods can effectively preserve the intrinsic structure information (such as matrix structure information of a 2D image data and time information contained in a video sequence data) of the multidimensional data [41,42]. A tensor is a multidimensional generalization of a vector or a matrix [43]: an n -order tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_n}$ is an object with n indices, with elements denoted by $\mathcal{A}_{i_1 \dots i_s \dots i_n}$ or $a_{i_1 \dots i_s \dots i_n}$, where $1 \leq i_s \leq I_s$. We first introduce some basic terminology on tensor operations, which is related to this paper.

The mode- s product of a tensor $\mathcal{A}$ and a matrix $\mathbf{B} \in \mathbb{R}^{I_s \times I_s}$ is a generalization of matrix multiplication for tensors and is defined as $(\mathcal{A} \times_s \mathbf{B})_{i_1 \dots i_s \dots i_n} = \sum_{l_s=1}^{I_s} a_{i_1 \dots i_s \dots i_n} m_{l_s i_s}$, where $(\mathcal{A} \times_s \mathbf{B}) \in \mathbb{R}^{I_1 \times \dots \times I_{s-1} \times I_{s+1} \times \dots \times I_n}$. The mode- s matricizing of $\mathcal{A}$ is a matrix $\mathcal{A}_{(s)} \in \mathbb{R}^{I_s \times (I_1 \times \dots \times I_{s-1} \times I_{s+1} \times \dots \times I_n)}$ whose j -th row is obtained by concatenating all the elements of $\mathcal{A}$ of the form $a_{i_1 \dots i_{s-1} j i_{s+1} \dots i_n}$. The mode- s matricizing of $(\mathcal{A} \times_s \mathbf{B})$ can be expressed as $(\mathcal{A} \times_s \mathbf{B})_{(s)} = \mathbf{B} \mathcal{A}_{(s)}$. The norm of $\mathcal{A}$ is defined as $\|\mathcal{A}\| = \sqrt{\sum_{i_1 \dots i_n} a_{i_1 \dots i_n}^2}$ and can also be computed as $\|\mathcal{A}\|^2 = \text{tr}(\mathcal{A}_{(s)} \mathcal{A}_{(s)}^T)$, $s \in [1, \dots, n]$.

For the tensor representation, training samples are extended as $\{\mathcal{W}_i | \mathcal{W}_i \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_n}, i = 1, \dots, N\}$. At each iteration t ($t = 1, 2, \dots, l$), tensor SEL_TCB (TSEL_TCB) is to learn a set of subspace project matrices $\{\mathbf{V}_i^t | \mathbf{V}_i^t \in \mathbb{R}^{q_i^t \times I_i}, i = 1, \dots, n\}$ for each order of tensor data, where $q_i^t < I_i$. Along each order, the projection of $\mathcal{W}$ into multiple subspaces is denoted as

$$f(\mathcal{W}) = \{\sqrt{a_i} \mathcal{W} \times_1 \mathbf{V}_1^t \times_2 \mathbf{V}_2^t \dots \times_n \mathbf{V}_n^t | t = 1, 2, \dots, l\}. \quad (18)$$

The variable h_{mt} in (2) is redefined as

$$h_{mt} = \|\Delta \mathcal{W}_{imkm} \times_1 \mathbf{V}_1^t \times_2 \dots \times_n \mathbf{V}_n^t\|^2 - \|\Delta \mathcal{W}_{imjm} \times_1 \mathbf{V}_1^t \times_2 \dots \times_n \mathbf{V}_n^t\|^2 \quad (19)$$

where $\Delta \mathcal{W}_{imkm} \triangleq \mathcal{W}_{im} - \mathcal{W}_{km}$. Substituting Eq. (19) into Eq. (6), we obtain the objective function of $\{\mathbf{V}_i^k\}_{k=1}^n$ as follows:

$$\max_{\mathbf{V}_i^k} \sum_{k=1}^n \sum_{j=1}^N (S_{ij}^p - F \cdot S_{ij}) \|\Delta \mathcal{W}_{ij} \times_1 \mathbf{V}_1^k \times_2 \mathbf{V}_2^k \dots \times_n \mathbf{V}_n^k\|^2 \quad (20)$$

However, in the most situation, there is no closed-form optimal solution for the objective function in (20) [44]. Alternatively, we can obtain the closed-form suboptimal solution of $\mathbf{V}_i^k$ by iteratively solving

the following objective function, according to [23]

$$\max_{\mathbf{V}_t^k} \text{tr} \left(\mathbf{V}_t^k \left(\sum_{i=1}^N \sum_{j=1}^N (L_{ij}^P - F \cdot L_{ij}) \overline{\mathbf{W}}_{i(k)} \overline{\mathbf{W}}_{j(k)}^T \right) \mathbf{V}_t^{kT} \right), \quad k = 1, 2, \dots, n \quad (21)$$

where $\overline{\mathbf{W}}_{i(k)}^T \triangleq (\mathbf{W}_i \times_1 \mathbf{V}_1^1 \times_{k-1} \mathbf{V}_t^{k-1} \times_{k+1} \mathbf{V}_t^{k+1} \times_n \mathbf{V}_t^n)_{(k)}$. Tao et al. [23] have proved the iterative solution of (21) is convergent and indicated that the difference form in (21) is easier to converge than the ratio form. Let matrix

$$\Phi_t^k = \sum_{i=1}^N \sum_{j=1}^N (L_{ij}^P - F L_{ij}) \overline{\mathbf{W}}_{i(k)} \overline{\mathbf{W}}_{j(k)}^T, \quad k = 1, \dots, n \quad (22)$$

then columns of the matrix $\mathbf{V}_t^{kT}$ are the normalized eigenvectors corresponding to the first q_t^k largest eigenvalues of Φ_t^k . Details of the TSEL_TCB framework are shown in Algorithm 2. In this study, we arrange Gabor features $\{F_p, F_e, F\}$ in the third dimension to a three-order tensor data with $I_3 = 14$ and use it as the tensor-based extension of the gait image template $\mathbf{I}$.

Algorithm 2. Tensor Subspace Ensemble Learning via Totally-corrective Boosting.

Input: The training samples $\{\mathbf{W}_i\}$, the triplet index set $\mathcal{T}$ and the maximum iteration number L .
Initialization: Initialize u_m to be $1/M$.
for $l=1$ **to** L **do**
 - Based on $\mathcal{T}$ and $[u_m]$, construct matrices $\mathbf{S}$ and $\mathbf{S}^P$ according to (8);
 - Calculate the projection matrix set $\{\mathbf{V}_t^k\}_{k=1}^n$ by iteratively solving the eigenvectors of Φ_t^k in (22);
 - Calculate $[h_{ml}]$ according to (19)
 if $\sum_{m=1}^M u_m h_{ml} \leq \beta$ **then** Break; **end if**
 - Add h_{ml} to the dual problem (5);
 - Minimize β and update $[u_m]$ according to (5) using the primal-dual interior-point LP solver;
end for
 Calculate the weight vector $\mathbf{a}$ according to (3)
Output: The weight vector $\mathbf{a}$ and the projection matrix sets $\{\mathbf{V}_t^k | k = 1, 2, \dots, n; t = 1, 2, \dots, L\}$

3.2. Local patch-based SEL_TCB

Local patch-based matching has been widely applied on visual pattern recognition [30,45,46]. There are two main steps, patch selection and local feature ensemble, in the most local patch-based recognition methods. In the patch selection step, the main purpose is to seek out the most effective regions and discard redundant regions for recognition. In the local feature ensemble step, local features are first extracted from selected patches. Then, these local features can be concatenated into a global feature vector, or based on the local features from each patch, multiple individual classifiers are trained and then combine them for classification. Local patch-based matching has many attractive properties, including its robust to outliers and its superior discriminant power for recognition [45]. In this section, we propose a novel local patch-based SEL_TCB (LPSEL_TCB) framework, compared with the holistic appearance-based SEL_TCB framework. Fig. 1 summarizes the process diagram of the proposed LPSEL_TCB for gait recognition.

For each gait image template $\mathbf{I}$, N_2 Gabor-based gait feature representations are first extracted, according to (17). For simplicity, we uniformly divide each Gabor-based gait feature representation into N_1 nonoverlapping patches with fixed size, as shown in Patch Segmentation module in Fig. 1, and total $N_1 \times N_2$ patches are obtained.

Then each patch is rearranged in column to acquire local patch-based feature representations $\mathbf{w}^j \in \mathbb{R}^{d'}$ ($j = 1, 2, \dots, N'$), where d' is the number of pixels in a patch and $N' = N_1 \times N_2$. Based on the local patch-based feature representation, training samples are extended as $\{\mathbf{w}_i^1, \mathbf{w}_i^2, \dots, \mathbf{w}_i^{N'} | i = 1, \dots, N\}$. To normalize intensities of different local patches, $\mathbf{w}_i^j$ is normalized to zero mean and unit variance as follows:

$$\mathbf{w}_i^j = \frac{\mathbf{w}_i^j - \text{mean}(\mathbf{w}_i^j) \cdot \mathbf{1}}{\sigma(\mathbf{w}_i^j)}, \quad i = 1, 2, \dots, N; \quad j = 1, 2, \dots, N' \quad (23)$$

where $\text{mean}(\mathbf{w}_i^j)$ and $\sigma(\mathbf{w}_i^j)$ respectively denote the average value and the standard deviation of elements of $\mathbf{w}_i^j$ and $\mathbf{1}$ is the vector of ones.

At each iteration t ($t = 1, 2, \dots, L$), LPSEL_TCB first selects the patch $k_t \in \{1, 2, \dots, N'\}$ and then learns the corresponding subspace projection matrix $\mathbf{V}_t \in \mathbb{R}^{q_t \times d'}$ with $q_t < d'$. According to selected patches k_t , $t = 1, 2, \dots, L$, the projection of $\{\mathbf{w}^j\}$ into multiple subspaces is denoted as

$$f(\{\mathbf{w}^j\}) = [(\sqrt{a_1} \mathbf{V}_1^T \mathbf{w}^{k_1})^T (\sqrt{a_2} \mathbf{V}_2^T \mathbf{w}^{k_2})^T \dots (\sqrt{a_L} \mathbf{V}_L^T \mathbf{w}^{k_L})^T]^T. \quad (24)$$

We redefine the variable h_{mt} in (2) as

$$h_{mt} = \|\mathbf{V}_t^T \Delta \mathbf{w}_{lmk_m}^{k_t}\|^2 - F \|\mathbf{V}_t^T \Delta \mathbf{w}_{lmj_m}^{k_t}\|^2. \quad (25)$$

Compared with SEL_TCB, the new definition of h_{mt} measures the difference of distances among samples based on one of local patches instead of the whole template. As done in SEL_TCB, we formulate LPSEL_TCB into the linear programming, according to (3) and (5). Then, substituting Eq. (25) into Eq. (6), we obtain the subproblem for learning k_t and $\mathbf{V}_t$ as

$$\max_{\mathbf{V}_t, k_t} \text{tr}(\mathbf{V}_t^T \Phi_t \mathbf{V}_t) \Phi_t = \mathbf{W}^{k_t} (\mathbf{I}^P - F \mathbf{L}) \mathbf{W}^{k_t T} \quad (26)$$

where $\mathbf{W}^{k_t} = [\mathbf{w}_1^{k_t}, \mathbf{w}_2^{k_t}, \dots, \mathbf{w}_N^{k_t}]$. First, we test each patch k ($1, \dots, N'$) and calculate the corresponding subspace projection matrix whose columns are the eigenvectors corresponding to the first q eigenvalues of $\mathbf{W}^k (\mathbf{I}^P - F \mathbf{L}) \mathbf{W}^{kT}$ and the corresponding value of (26). Then, we determine k_t with the maximum value of (26), and $\mathbf{V}_t$ is the corresponding optimal subspace projection matrix, i.e., the eigenvectors corresponding to the first q eigenvalues of Φ_t in (26). As a summary, the algorithm for the LPSEL_TCB framework is presented in Algorithm 3.

Algorithm 3. Local Patch-based Subspace Ensemble Learning via Totally-corrective Boosting.

Input: The training samples $\mathbf{W}$, the triplet index set $\mathcal{T}$ and the maximum iteration number L .

Initialization: Initialize u_m to be $1/M$. Extract local patch-based feature representations and normalize them, according to (23).

for $l=1$ **to** L **do**

 - Based on $\mathcal{T}$ and $[u_m]$, construct matrices $\mathbf{S}$ and $\mathbf{S}^P$ according to (8);

 - Select the local patch k_t and calculate the corresponding subspace projection matrix $\mathbf{V}_t$ by maximizing the objection function in (26);

 - Calculate h_{mt} according to (25);

if $\sum_{m=1}^M u_m h_{ml} \leq \beta$ **then** Break; **end if**

 - Add h_{mt} to the dual problem (5);

 - Minimize β and update $[u_m]$ according to (5) using the primal-dual interior-point LP solver;

end for

 Calculate the weight vector $\mathbf{a}$ according to (3)

Output: The local patches $\{k_t\}_{t=1}^L$, and the corresponding projection matrix set $\{\mathbf{V}_t\}_{t=1}^L$ and the weight vector $\mathbf{a}$.

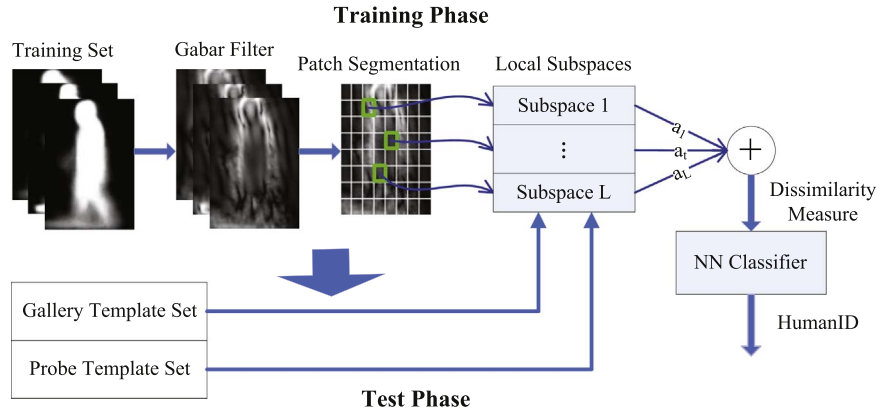


Fig. 1. Process diagram of gait recognition using LPSEL_TCB approach.

4. Evaluation experiments

To evaluate performance of the SEL_TCB, TSEL_TCB and LPSEL_TCB frameworks for the gait recognition, we derived and tested 9 methods from them by respectively using the LDA-, MFA- and SPE-based triplet sets on the USF HumanID gait database [8] and CASIA-B gait database [34]. Specifically, they are respectively LDA_SEL_TCB, MFA_SEL_TCB, SPE_SEL_TCB, LDA_TSEL_TCB, MFA_TSEL_TCB, SPE_TSEL_TCB, LDA_LPSEL_TCB, MFA_LPSEL_TCB and SPE_LPSEL_TCB. In this section, we first introduce the experiment settings on the USF database and CASIA-B database, and then evaluate and analyze the performance of the proposed method for gait recognition.

4.1. Experimental settings

In algorithm training, we adopted the average of GEI images in a gait silhouette image sequence as the gait image template $\mathbf{I}$:

$$\mathbf{I}(x, y) = \frac{1}{N_T} \sum_{i=1}^{N_T} \mathbf{GEI}_i(x, y), \quad \mathbf{GEI}_i(x, y) = \frac{1}{N_S} \sum_{t=(i-1)N_S+1}^{iN_S} \mathbf{BS}_t(x, y) \quad (27)$$

where N_T is number of periods in a sequence, N_S is period length of a sequence, $\mathbf{GEI}_i$ denotes GEI image over the i -th period of a gait sequence, and $\mathbf{BS}_t$ denotes the silhouette image at phase t in a sequence. The use of GEI representation is mainly based on the following reasons [9]: 1) GEI has low computation complexity and is robust to silhouette noise in individual frames; 2) GEI does not need to locate the starting stance of the gait cycle.

In algorithm test, given the gallery sequence $\mathcal{I}_G = \{\mathbf{GEI}_G(i), i = 1, \dots, N_G\}$ and the probe sequence $\mathcal{I}_P = \{\mathbf{GEI}_P(i), i = 1, \dots, N_P\}$, we respectively project each $\mathbf{GEI}_P(i)$ and $\mathbf{GEI}_G(i)$ into multiple subspaces and combine them, denoted by $f(\mathbf{w}_P(i))$ and $f(\mathbf{w}_G(i))$, according to (1), (18) or (24) by the learned projection matrices and the corresponding weights in training stage, where $\mathbf{w}_P(i)$ and $\mathbf{w}_G(i)$ are vector forms of $\mathbf{GEI}_P(i)$ and $\mathbf{GEI}_G(i)$. We adopted a nearest neighbor (NN) classifier for gait recognition. Specifically, the dissimilarity measure between $\mathcal{I}_P$ and $\mathcal{I}_G$ is defined as

$$d(\mathcal{I}_P, \mathcal{I}_G) = \frac{1}{N_P} \sum_{i=1}^{N_P} \|f(\mathbf{w}_P(i)) - \bar{f}(\mathbf{w}_G)\|_2 \quad (28)$$

where $\bar{f}(\mathbf{w}_G) = \frac{1}{N_G} \sum_{i=1}^{N_G} f(\mathbf{w}_G(i))$. Finally, the label of probe sequence $\mathcal{I}_P$ is identified as follows:

$$z = \underset{G}{\operatorname{argmind}}(\mathcal{I}_P, \mathcal{I}_G), \quad G = 1, 2, \dots, C \quad (29)$$

To evaluate the recognition accuracy, rank-1 and rank-5 recognition rates are used. The rank-1 recognition rate is the percentage of the number of the correct people ranked as the first candidate, and the rank-5 recognition rate is the percentage of the number of the correct people ranked among the first five candidates. All experiments were

carried out on the USF database and the CASIA-B database. Fig. 2 illustrates some samples from two databases.

4.1.1. USF HumanID gait database

In the USF HumanID gait database, 1870 sequences are captured from 122 subjects in an outdoor uncontrolled environment. Specifically, the challenge gait database contains gait variations due to the following covariates: 1) surface types (G-Grass/C-Concrete); 2) viewpoints (L-Left/R-Right); 3) shoe types (A/B); 4) carrying conditions (BF-Briefcase or NB-No Briefcase); and 5) captured time (M-May/November in the same year). In November, N_1 -new subjects and N_2 -repeat subjects are captured. By combining these covariates, some individuals can acquire up to 32 sequences captured in different conditions. There are one gallery set and 12 standard probe subsets designed for algorithm evaluation and comparison. Gallery set refers to prior images of people to be recognized and probe set refers to unknown images to be identified. The gallery set and probe sets consist of only one sequence per subject, and there are no common sequences among them. Table 1 lists detailed information about probe sets.

In the experiments on the USF database, as designed by [14], the training set is composed of six sequences per subject of 33 subjects, which are subsets (G, A, R, NB, M), (C, A, L, BF, M), (C, A, L, BF, M), (C, A, L, NB, N₁), (G, A, L, NB, N₁) and (G, A, R, BF, N₁). Note that those chosen subsets for training were captured in different conditions with probe sets, i.e., there are no common sequences between training set and probe sets.

4.1.2. CASIA-B gait database

The CASIA-B gait database consists of 124 subjects. For each subject, ten gait sequences were captured under three different conditions. Specifically, six gait sequences were captured in normal condition (named NM-01 to NM-06), two gait sequences were captured in the condition of subject wearing a bag (named BG-01 and BG-02), the other two sequences were captured in the condition of subject wearing a coat (named CL-01 and CL-02). Each gait sequence was captured from 11 different views. In our experiments, the only gait sequences captured from side view (90 degrees) were used.

In the experiments on the CASIA-B database, the training set is composed of three sequences NM-01, BG-01 and CL-01 per subject of 33 subjects. In addition, sequences NM-01 of all the subjects are used as gallery set, and sequences NM-05, NM-06, BG-02 and CL-02 are used as probe sets, respectively.

4.1.3. Parameter settings

In all the experiments, the size of each GEI template is aligned to 128×88 [9]. For Gabor features, to speed up algorithms, the size of each GEI template is downsampled from 128×88 to 64×44. According to (17), the size of each Gabor-based tensor representation is 64 × 44 × 14. To build local patch-based feature pole, the size of each

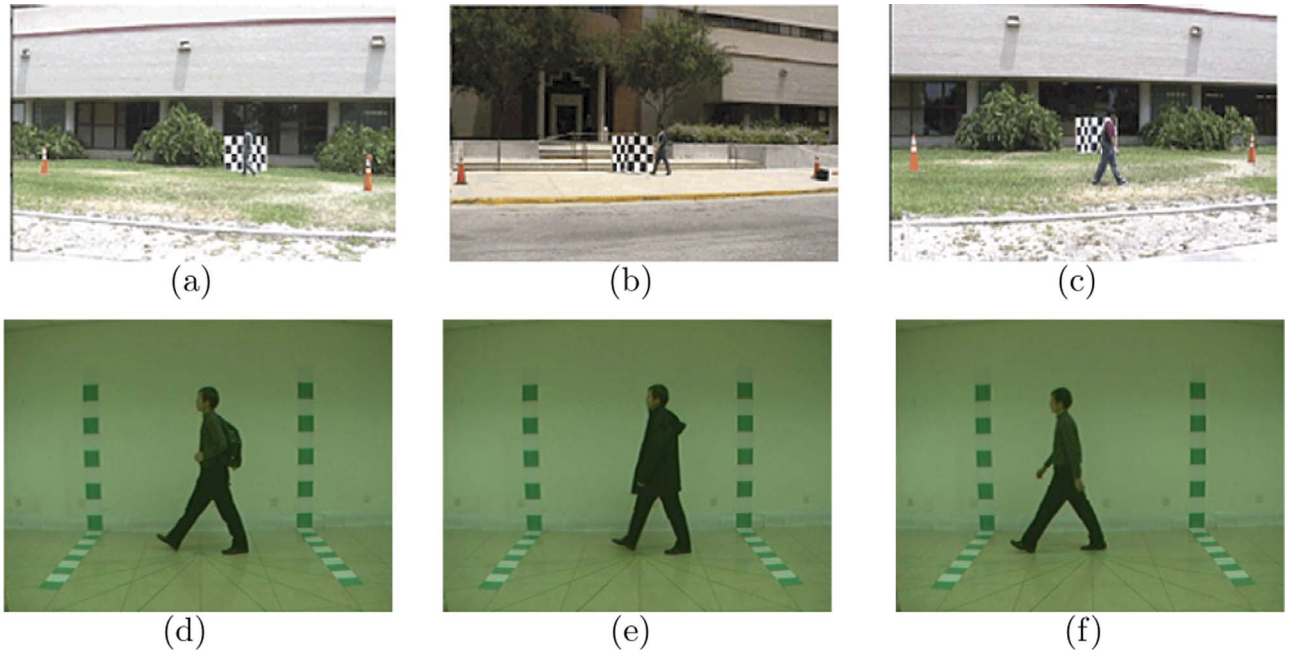


Fig. 2. Example samples from the USF HumanID gait database [8] and the CASIA-B gait database [34].

Table 1

Details of probe sets for challenge experiments in the USF database.

Experiment (Probe)	# of Probe Sets	Gallery/Probe Difference
A (G,A,L, NB,M/N)	122	View
B (G,B,R, NB,M/N)	54	Shoe
C (G,B,L, NB,M/N)	54	View and Shoe
D (C,A,R, NB,M/N)	121	Surface
E (C,B,R, NB,M/N)	60	Surface and Shoe
F (C,A,L, NB,M/N)	121	Surface and View
G (C,B,L, NB,M/N)	60	Surface, Shoe and View
H (G,A,R, BF,M/N)	120	Briefcase
I (G,B,R, BF,M/N)	60	Briefcase and Shoe
J (G,A,L, BF,M/N)	120	Briefcase and View
K (G,A/B,R, NB,N)	33	Time, Shoe, and Clothing
L (C,A/B,R, NB,N)	33	Time, Shoe, Clothing and Surface

patch is set to 11×16 .

In the MFA-based triplet set building, two parameters, namely, the number k_1 of nearest neighbors in the same class and the number k_2 of nearest neighbors from the different classes need to be chosen. First, by empirically setting $k_2 = 20$, we evaluate $k_1 = [1, 2, 3, 4]$ in MFA_SEL_TCB on the USF dataset and choose the optimal one. Then, based on the selected k_1 , we run MFA_SEL_TCB with k_2 from 10 to 100 with a step of 10. The average Rank-1 recognition rates w.r.t. k_1 and k_2 are shown in Fig. 3(a) and (b), where $k_1 = 3$ and $k_2 = 20$ achieve the highest Rank-1 recognition rates. Thus, we fix $k_1 = 3$ and $k_2 = 20$ throughout the experiments.

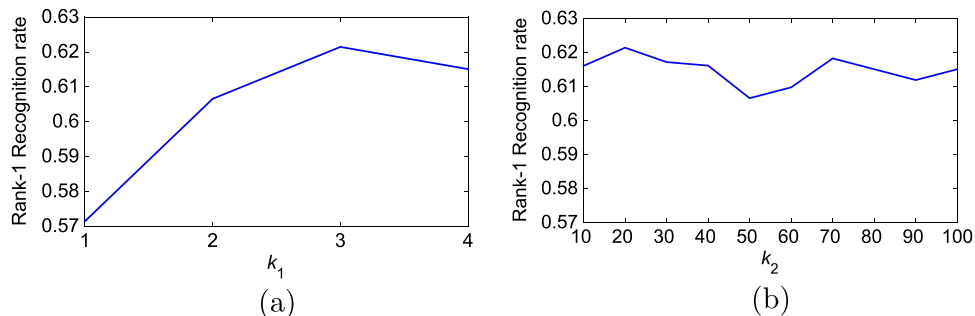


Fig. 3. Average Rank-1 recognition rates w.r.t different k_1 and k_2 by MFA_SEL_TCB on the USF database. (a) k_1 is the number of nearest neighbors in the same class, (b) k_2 is the number of nearest neighbors from different classes.

There are four parameters in the SEL_TCB framework and its TSEL_TCB and LPSEL_TCB extensions, namely, the tuning parameter F , the regularization parameter D , the dimension q_t of combined subspaces and the number of iterations L . Table 2 lists the main parameters in our method in the experiments on the USF and CASIA-B databases. Although choosing good training parameters is critical, it is problem and data dependent. In practice, parameters are chosen empirically. Next, we explain the way to determine the parameters in our method in the experiments on the USF database. For great classification ability, the iteration number L should be enough large, since, as indicated in [33], the linear programming boosting method is not easy to overfit. Like most boosting methods, the performance should not be sensitive to the single classifier, i.e., the subspace dimension q_t . Therefore, we empirically preset $L=50$ and each $q_t=60$, $L=100$ and each $q_t^1 \times q_t^2 \times q_t^3 = 10 \times 6 \times 3$ and $L=100$ and each $q_t=4$ for the SEL_TCB, TSEL_TCB and LPSEL_TCB frameworks, respectively. We empirically preset moderate $\eta = 0.06$, since, as discussed in [33], very small or very large η could cause the training extreme instances. The range of F is $[1, 10]$ for the SEL_TCB framework and $[1, 4]$ for the TSEL_TCB and LPSEL_TCB frameworks. The average Rank-1 recognition rates w.r.t. F are shown in Figs. 4, 5 and 6(a). According to the highest value in each evolution curve of the average Rank-1 recognition rates, we therefore fix F as shown in the second column in Table 2. Then, we perform our method with η from 0.03 to 0.11 with a step of 0.01. The average Rank-1 recognition rates w.r.t. η are shown in Figs. 4, 5 and 6(b). The optimal ones are chosen as shown

Table 2

Main parameters in the experiments on the USF dataset and CASIA-B database.

Database	USF HumanID				CASIA-B			
Method	F	η	q_t	L	F	η	q_t	L
LDA_SEL_TCB	8	0.06	60	40	7	0.06	60	20
MFA_SEL_TCB	3	0.06	80	40	7	0.06	60	20
SPE_SEL_TCB	5	0.04	60	60	7	0.06	60	20
LDA_TSEL_TCB	1	0.03	$10 \times 6 \times 3$	80	1	0.06	$10 \times 6 \times 3$	40
MFA_TSEL_TCB	1	0.09	$10 \times 6 \times 3$	120	1	0.06	$10 \times 6 \times 3$	40
SPE_TSEL_TCB	1	0.07	$10 \times 6 \times 3$	80	1	0.06	$10 \times 6 \times 3$	40
LDA_LPSEL_TCB	1	0.03	2	80	1	0.06	4	40
MFA_LPSEL_TCB	1	0.06	7	80	1	0.06	4	40
SPE_LPSEL_TCB	1	0.08	4	100	1	0.06	4	40

in the third column in Table 2.

Given the optimized parameters F and η , we examine the iteration number L and the subspace dimension q_t in our method. We vary q_t from 20 to 120 with a step of 20 for the SEL_TCB framework, $q_t^1 \times q_t^2 \times q_t^3$ from $5 \times 3 \times 2$ to $25 \times 15 \times 10$ with a step of $5 \times 3 \times 2$ for the TSEL_TCB framework and q_t from 1 to 8 with a step of 1 for the LPSEL_TCB framework. The average Rank-1 recognition rates w.r.t. different q_t are shown in Figs. 4, 5 and 6(c). The range of L is [1,100] for the SEL_TCB framework and [1,200] for the TSEL_TCB and LPSEL_TCB frameworks. The average Rank-1 recognition rates w.r.t. different q_t are shown in Figs. 4, 5 and 6(d). Finally, we fix q_t and L as shown in the fourth and fifth columns in the Table 2. In the similar way, the main parameters in our method in the experiments on the CASIA-B database are empirically chosen.

To evaluate the performance stability of our method with default parameters on the Table 2, we run the experiment 10 times by randomly evenly split training and test sets on the gallery set and the probe sets A, D, F, H and J of the USF database. In each round of the experiment, the gait sequences of half subjects are used as the training set, and the gait sequences of the remaining subjects are still used as the gallery set and probe sets. We report the recognition performance statistics of our methods in Table 3. As shown in the Table 3, our methods have achieved the small standard deviation values, which indicates that our methods with given parameters have the stable performance.

4.1.4. Time complexity analysis

We analyze the time complexity for training the proposed SEL_TCB, TSEL_TCB and LPSEL_TCB frameworks with different triplet sets. In l th iteration ($l = 1, 2, \dots, L$), the primal-dual interior point method is used to solve the linear programming of (5), which converges to an ϵ -accurate solution with a worst case bound of $O(\sqrt{2M} + l + 2 \ln 1/\epsilon)$ on the number of iterations of the method [47]. M is the number of triplets. In addition, in each iteration for SEL_TCB in Algorithm 1 it takes $O(dN^2)$ for computation of Ψ_t and $O(N^3)$ for eigenvalue decomposition of Ψ_t . For TSEL_TCB in Algorithm 2, it takes $O(Kn(n-1)N^2q_l)$ for K times of computation of Φ_t^k , $k = 1, 2, \dots, n$ and $O(K \sum_{k=1}^n l_k^3)$ for K times of eigenvalue decomposition of Φ_t^k , $k = 1, 2, \dots, n$, where $q = \max\{q_t^1, q_t^2, \dots, q_t^n\}$ and

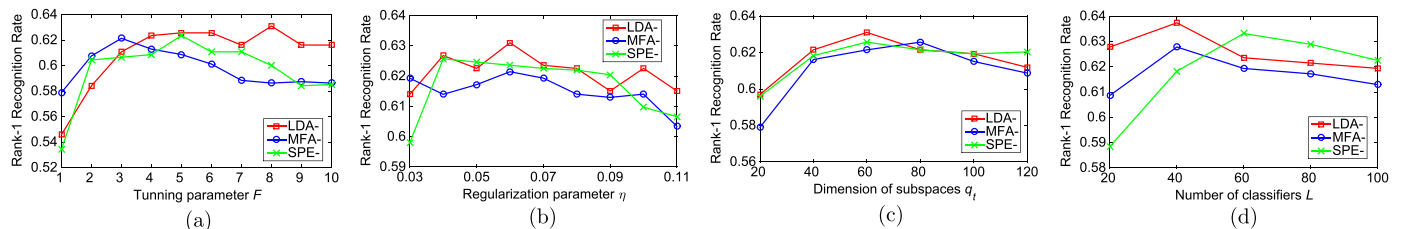


Fig. 4. Average Rank-1 recognition rates w.r.t (a) the tuning parameter F , (b) the regularization parameter η , (c) the subspace dimension q_t and (d) the iteration number L by SEL_TCB with the LDA-, MFA- and SPE-based triplet sets on the USF database.

$I = \max\{I_1, I_2, \dots, I_n\}$. K is the number of iterations in the optimization of (20). For LPSEL_TCB in Algorithm 3, it takes $O(N'dN^2)$ for computation of Φ_t and $O(N'd^3)$ for eigenvalue decomposition of Φ_t of N' local patches, where N' is the number of local patches and d' is the dimension of the patch. We run the matlab code of our method on a PC with an Intel Core i7 2.70 GHz processor and 8 GB RAM. Table 4 reports the training time and test time per probe sequence of our method on the USF HumanID dataset.

4.2. Performance evaluation

4.2.1. Comparisons with state-of-the-art gait recognition algorithms

In the experiments on the USF database, the proposed algorithms are evaluated and compared with state-of-the-art gait recognition algorithms including GEI+Real [9], CEI+Real [16], DNGRA [14], DLLE/L [24], GTDA+LDA [23], MFA-1 [21], Gabor-PDF-LGSR [48], SRML [49], SBDA+LDA [50] and VI-MGR [17]. All the experiments are conducted on the 12 probe sets (Probe A-L). Since different experimental settings have been used with different methods, for fair comparisons, we also reimplemented four well-known gait recognition methods ourselves: GEI+Real* [9], MFA-1* [21], DLLE/L* [24], GaborD+GTDA(H)* [23] by using the same training set and dissimilarity measure with our methods. For MFA-1* and DLLE/L*, k_1 is set to $\{1, 2, 3, 4\}$ and k_2 is set from 50 to 500 at a step of 50. For GEI+Real*, MFA-1* and DLLE/L*, the dimension of PCA is chosen according to a 98% energy criterion as in [51]. For GaborD+GTDA(H)*, the lower dimensions of the three-order tensor representation are respectively set from 2 to 32 with a step 4, from 2 to 22 with a step 4 and from 2 to 10 with a step 2, and then we test all their possible combinations and choose the optimal one. In the discriminant analysis step, we test all the possible lower dimensions and choose the optimal one for each probe set.

Table 5 shows the Rank-1 recognition rates of our methods on the USF dataset. Table 6 tabulates the Rank-1 recognition rates and Rank-5 recognition rates of different methods on all the probe subsets in the USF HumanID database. As shown in Table 6, our proposed MFA_TSEL_TCB method outperforms other compared methods. From these comparison results in Table 5 and Table 6, we make the following observations:

- Experimental results show that when the number of training samples is enough to learn the gait manifold, subspace learning algorithms have the potential to achieve the excellent recognition performance. Specifically, GEI+Real*, MFA-1*, DLLE/L* and GTDA+LDA* respectively outperform GEI+Real, MFA-1, DLLE/L and GTDA+LDA on most probe sets. When the training set only consists of samples of the gallery set with only one condition, it is not enough to minimize the intrapersonal variations of the gait.
- LDA_SEL_TCB, MFA_SEL_TCB and SPE_SEL_TCB obtain better recognition performances than GEI+Real*, MFA-1* and DLLE/L*. These results imply that it is effective to learn the multiple subspaces in SEL_TCB framework for improving recognition performance of one single subspace.
- MFA_TSEL_TCB obtains a surprise increase in recognition perfor-

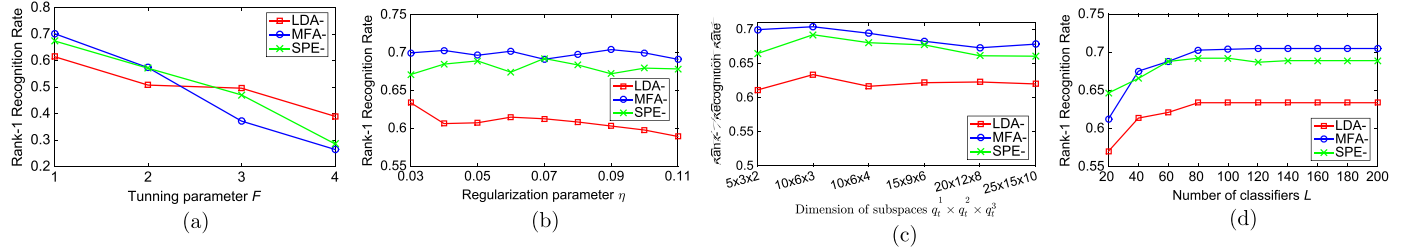


Fig. 5. Average Rank-1 recognition rates w.r.t (a) the tuning parameter F , (b) the regularization parameter η and (c) the iteration number L by TSEL_TCB with the LDA-, MFA- and SPE-based triplet sets on the USF database.

mance compared with MFA_SEL_TCB. These results demonstrate that the senior feature representation such as Gabor features and the tensor representation are important for gait recognition.

- Compared with GTDA+LDA*, SPE_LPSEL_TCB achieves the higher recognition accuracies on the H, J, K probe sets with the briefcase difference, which indicates the joint usage of local patches can be robust to the carrying change.
- For the SEL_TCB, TSEL_TCB and LPSEL_TCB frameworks, LDA-, MFA- and SPE-based triplet sets achieve the best recognition performance on most probe set, respectively. It demonstrates that it is hard to decide which triplet building method is best for recognition.

Based on CASIA gait database (Data set B), we performed experimental evaluation for our methods and compared with state-of-the-art gait recognition algorithms including CAS [34], GEI [9,16], SEIS [15] and CEI [16]. We also conducted GEI+LDA*, MFA-1*, DLLE/L* and GTDA+LDA* methods. For fair comparison, their recognition accuracies with optimal parameters are reported. The Rank-1 recognition rates of our methods are shown in Table 7. Moreover, for LDA_SEL_TCB, MFA_SEL_TCB, SPE_SEL_TCB, the results without applying subspace ensemble, i.e., the number of iteration $L=1$, are shown. It can be seen that the proposed subspace ensemble learning methods can effectively improve the recognition performance of the single subspace on all the probe subsets. Table 8 lists the rank-1 and rank-5 recognition rates of compared methods on NM-5, NM-6, BG-2, and CL-2 probe subsets in the CASIA-B database. As shown in Table 8, our proposed methods outperform the other state-of-the-art methods on BG-2 and CL-02 probe subsets. Due to huge appearance variation, it is harder for gait recognition on BG-02 and CL-02 probe subsets than normal NM-05 and NM-06 probe subsets. Particularly, MFA_TSEL_TCB method achieved a remarkable increase in rank-1 recognition rate of 12% and 25% compared with CEI method and 15% and 5% compared with SELS method on BG-02 and CL-02 probe subsets. Rank-1 recognition rate of MFA_LPSEL_TCB method is also 4% and 27% higher than CEI method and 7% and 7% higher than SEIS method on BG-02 and CL-02 probe subsets. Although rank-1 recognition rates of MFA_TSEL_TCB and MFA_LPSEL_TCB methods are slightly lower than CEI method on NM-06 probe subsets, it is worth noting that CEI method used five normal sequences (NM-01 to NM-05) per subjects as training set and meanwhile as gallery set, which generally achieves better performance than those adopted by our

Table 3

Mean rank-1 recognition rate, standard deviation, median, maximum and minimum results found by our methods for 10 runs using the gallery and probe sets A, D, F, H and J on the USF dataset.

Algorithms	Mean	Std	Max	Min
LDA_SEL_TCB	78.62	2.52	84.61	76.00
MFA_SEL_TCB	79.96	2.17	84.26	75.81
SPE_SEL_TCB	80.54	1.88	83.53	76.79
LDA_TSEL_TCB	78.98	3.12	86.78	73.90
MFA_TSEL_TCB	83.04	2.83	88.79	78.60
SPE_TSEL_TCB	82.15	2.9	86.78	78.18
LDA_LPSEL_TCB	75.73	2.92	82.92	74.08
MFA_LPSEL_TCB	82.29	1.41	84.94	80.96
SPE_LPSEL_TCB	81.45	1.65	85.60	80.73

Table 4

Training time\test time per seq. (Seconds) on the USF HumanID dataset.

Algorithms	LDA-	MFA-	SPE-
SEL_TCB	98.95/0.003	389.72/0.0046	90.77/0.0035
TSEL_TCB	666.66/0.0387	1347.16/0.0571	887.88/0.0394
LPSEL_TCB	189.29/0.0108	259.39/0.011	194.38/0.0112

methods on normal NM-05 and NM-06 probe subsets. In addition, compared with the GEI+LDA*, MFA-1*, DLLE/L*, GTDA+LDA* methods, our MFA TSEL TCB and SPE LPSEL TCB methods consistently achieve the higher accuracies on every probe subsets.

Based on the comparative experiments on the USF and CASIA-B databases, we can observe the following advantages and disadvantages of our proposed method over other state-the-art gait recognition algorithms. Specifically, advantages of using SEL_TCB include.

- SEL_TCB has the ability to learn the complex gait manifold. By maximizing the soft margin among different classes, as defined in (3), SEL_TCB is robust to intrapersonal variations and avoid the sub-optimum problem of single subspace learning algorithms.
- SEL_TCB can be developed using multiple different triplet building methods. In this paper, we have presented three different triplet sets with different recognition performances.
- SEL_TCB can easily be extended to TSEL_TCB in order to solve the tensor data. Many researches have shown that in the subspace

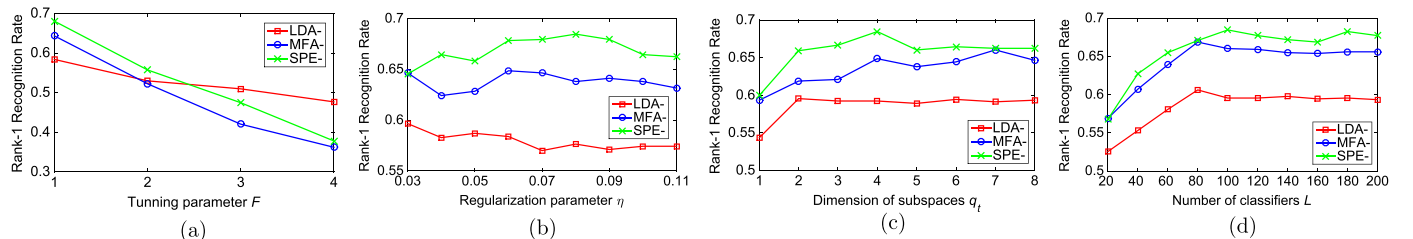


Fig. 6. Average Rank-1 recognition rates w.r.t (a) the tuning parameter F , (b) the regularization parameter η , (c) the subspace dimension q_t and (d) the iteration number L by LPSEL_TCB with the LDA-, MFA- and SPE-based triplet sets on the USF database.

Table 5

Rank-1 Recognition rates (percent) of our methods on twelve probe sets of the USF database.

Probe Set	A	B	C	D	E	F	G	H	I	J	K	L	Avg.
LDA_SEL_TCB	88	91	76	51	45	44	45	87	84	44	64	42	63.74
MFA_SEL_TCB	89	94	78	44	47	38	43	90	91	43	58	39	62.78
SPE_SEL_TCB	89	94	76	46	43	41	47	88	86	46	67	33	63.32
LDA_TSEL_TCB	87	87	69	66	64	42	36	85	76	55	18	21	63.38
MFA_TSEL_TCB	95	91	78	66	59	46	52	93	88	69	30	27	70.49
SPE_TSEL_TCB	93	89	81	64	69	41	52	91	83	70	21	24	69.21
LDA_LPSEL_TCB	90	87	76	48	48	29	33	84	86	61	27	15	60.62
MFA_LPSEL_TCB	93	94	82	54	48	37	38	91	91	70	40	24	66.88
SPE_LPSEL_TCB	93	93	85	53	59	43	43	92	90	71	27	24	68.47

analysis keeping the structure of tensor data is important for gait recognition [23,24,52].

- SEL_TCB can effectively select and learn for the recognition task from local patches by extending to LPSEL_TCB. When silhouettes of some body parts are changed by the carrying or coat conditions, effective selection of local patches can avoid the degradation of recognition performance.
- SEL_TCB can provide the sparse combination of multiple subspaces. Compared with random subspace ensemble learning algorithms which generally generate amount of subspaces and just evenly combine them, SEL_TCB give the weighted combination of less subspaces based on the defined stop criteria for the iterative learning. In the following section, we have also illustrated the sparsity of the weights in SEL_TCB in Fig. 8 and Table 9.
- SEL_TCB are intended to combine discriminant features in multiple subspaces, which can further be used as input of other classifiers. In the following section, we have evaluated the performance of the

Table 7

Rank-1 recognition rates (percent) of our methods on probe sets of the CASIA-B database.

Probe Set	NM-5	NM-6	BG-2	CL-2	Avg.
LDA_SEL_TCB(L=1)	92.52	87.85	57.01	50.41	65.87
MFA_SEL_TCB(L=1)	96.26	97.2	67.29	57.94	73.99
SPE_SEL_TCB(L=1)	90.65	90.65	59.81	49.53	66.66
LDA_SEL_TCB	98.13	97.2	76.64	66.36	80.38
MFA_SEL_TCB	98.13	97.2	73.83	69.16	80.22
SPE_SEL_TCB	98.13	97.2	71.96	66.36	78.66
LDA_TSEL_TCB	96.26	92.52	82.24	47.66	74.76
MFA_TSEL_TCB	99.07	99.07	87.85	69.16	85.36
SPE_TSEL_TCB	98.13	98.13	85.98	61.68	81.93
LDA_LPSEL_TCB	97.2	95.33	76.64	66.36	79.76
MFA_LPSEL_TCB	98.13	98.13	79.44	70.09	82.55
SPE_LPSEL_TCB	99.07	98.13	79.44	71.96	83.33

Table 6

Rank-1 Recognition rates (percent) of different methods on twelve probe sets of the USF HumanID database.

Probe	A	B	C	D	E	F	G	H	I	J	K	L	Avg.
Rank-1 Recognition Rates													
GEI+Real [9]	89	87	78	36	38	20	28	62	59	59	3	6	51.04
CGI+Real [16]	90	91	80	33	33	18	22	84	80	61	3	6	54.49
DNGRA [14]	85	89	72	57	66	46	41	83	79	52	15	24	62.94
DLLE/L [24]	90	89	81	40	50	27	26	65	67	57	12	18	54.78
GTDA+LDA [23]	93	95	88	39	47	28	33	82	82	63	19	19	60.24
MFA-1 [21]	89	89	83	38	43	25	29	58	59	56	9	18	52.4
Gabor-PDF-LGSR [48]	95	93	89	62	62	39	38	94	91	78	21	21	70.07
SRML [49]	93	94	85	52	52	37	40	86	85	68	18	15	66.50
SBDA+LDA [50]	93	94	85	51	50	29	36	85	83	68	18	24	61.35
VI-MGR [17]	95	96	86	54	57	34	36	91	90	78	31	28	68.13
GEI+Real*	85	89	78	42	41	34	41	87	79	41	33	33	58.54
MFA-1*	88	93	83	44	53	33	40	89	88	39	42	36	61.08
DLLE/L*	89	94	85	46	55	35	40	90	90	39	42	33	62.04
GTDA+LDA*	91	85	76	58	57	48	48	85	78	58	48	33	65.9
MFA_TSEL_TCB(Ours)	95	91	78	66	59	46	52	93	88	69	30	27	70.49
SPE_LPSEL_TCB(Ours)	93	93	85	53	59	43	43	92	90	71	27	24	68.47
Rank-5 Recognition Rates													
GEI+Real [9]	93	93	89	65	60	42	45	88	79	80	6	9	68.68
CGI+Real [16]	95	94	93	66	65	48	52	96	97	87	24	27	75.05
DNGRA [14]	96	94	89	85	81	68	69	96	95	79	46	39	82.05
DLLE/L [24]	95	96	93	74	78	50	53	90	90	83	33	27	76.60
GTDA+LDA [23]	98	99	97	68	68	50	56	95	99	84	40	40	77.58
MFA-1 [21]	95	96	93	64	67	44	53	89	88	81	24	27	72.5
Gabor-PDF-LGSR [48]	99	94	96	89	91	64	64	99	98	92	39	45	85.31
SBDA+LDA [50]	98	98	94	74	79	57	60	95	95	84	40	40	79.93
VI-MGR [17]	100	98	96	80	79	66	65	97	95	89	50	48	83.75
GEI+Real*	96	96	89	69	71	60	67	95	93	67	64	61	77.99
MFA-1*	98	98	94	75	78	58	69	97	98	68	64	61	80.54
DLLE/L*	97	98	93	74	78	58	69	97	98	68	64	58	80.01
GTDA+LDA*	97	96	89	85	78	69	69	92	93	86	91	70	84.54
MFA_TSEL_TCB(Ours)	100	96	93	84	83	73	74	95	96	89	64	52	86.09
SPE_LPSEL_TCB(Ours)	99	94	94	86	78	72	72	98	97	92	61	55	86.52

Methods labeled * are trained using the same training set with the proposed methods, where training samples of each individual consist of 6 gait image sequences from different conditions.

Table 8

Rank-1 and Rank-5 recognition rates (percent) of different methods on probe sets of the CASIA-B database.

Recognition Rate Algorithms\ Probe	Rank-1					Rank-5				
	NM-5	NM-6	BG-2	CL-2	Avg.	NM-5	NM-6	BG-2	CL-2	Avg.
CEI (k=5) [16]	–	100	75	44.44	73.15	–	100	89.81	72.22	87.34
SEIS (LDA) [15]	99	–	72	64	78.33	–	–	–	–	–
GEI (k=5) [9,16]	–	98.15	51.85	25.93	58.64	–	100	75.93	61.11	79.01
CAS [34]	98	–	51.85	25.93	58.59	–	100	75.93	61.11	79.01
GEI+Real [†]	98.13	98.13	71.03	58.88	76.01	99.07	99.07	86.92	81.31	89.10
MFA-1 [†]	98.13	98.13	69.16	65.42	77.57	100	100	87.85	81.31	89.72
DLLE/L [†]	97.20	98.13	69.16	61.68	76.17	100	100	87.85	79.44	89.10
GTDA+LDA [†]	97.20	96.26	78.50	57.01	77.41	100	100	93.46	78.50	90.65
MFA_TSEL_TCB(Ours)	99.07	99.07	87.85	69.16	85.36	100	100	96.26	85.98	94.09
SPE_LPSEL_TCB(Ours)	99.07	98.13	79.44	71.96	83.33	100	100	94.39	83.18	92.52

Table 9

Percentage of the number of subspaces with nonzero weight in the total learned subspaces on the USF HumanID dataset.

Algorithms	LDA-	MFA-	SPE-
SEL_TCB	45% (18/40)	50% (20/40)	43.33% (26/60)
TSEL_TCB	45% (36/80)	49.17% (59/120)	46.25% (37/80)
LPSEL_TCB	67.5% (54/80)	68.75% (55/80)	64% (64/100)

extracted features together with other classifiers, and Table 10 shows the comparative results with different classifiers.

Disadvantages of using SEL_TCB include.

- When dealing with a large number of triplets, along with increase of the number of iterations, optimizing LP in SEL_TCB can become very time-consuming.
- There are four parameters that need empirically tuning for recognition tasks in SEL_TCB.
- SEL_TCB requires an auxiliary set of data to learn subspaces for gait recognition.

4.2.2. Parameter analysis

In this section, we mainly investigate the effect on recognition performance of the parameters of the representative SPE_LPSEL_TCB method, including tuning parameter F , regularization parameter η , number of iterations L , dimensionality q_t of each subspace and image patch size ($w \times h$). To highlight the effect on performance of each parameter, when one parameter is tested, other parameters are fixed. Specifically, we set η from 0.03 to 0.11 with a step of 0.01, q_t from 1 to 8 with a step of 1 and F from 1 to 4 with a step of 1. The range of L is set

to [1, 200]. For the image patch size $w \times h$ test, h is set as [4 8 16 32], w is set as [41122], and then all the combinations of w and h are considered. Fig. 7 plots rank-1 recognition rates of SPE_LPSEL_TCB with different parameters on probe subsets from A to L of the USF database. From the results showed in Fig. 7, we obtain the following observations:

- In the experiments on probe subsets D, E, F, G, K and L, the effects of parameters of SPE_LPSEL_TCB are more obvious than other probe subsets, which keeps consistent with conclusions drawn in [8].
- When the tuning parameter F is set as 1, SPE_LPSEL_TCB obtains the best recognition performance on most probe subsets. Because the feature representation of image patch has been normalized, it is reasonable to set $F=1$ for good performance.
- When the regularization parameter η is set from 0.03 to 0.11 and dimensionality of each subspace q_t is set as $q_t > 1$, SPE_LPSEL_TCB shows the stable recognition performances on the most probe subsets. It is useful for SPE_LPSEL_TCB to easily adjust η and q_t to obtain good performance.
- After number of iterations L arrives at 60, rank-1 recognition rates of SPE_LPSEL_TCB on probe subsets A, B, C, H, I and J exceed 70%, but rank-1 recognition rates on the remaining probe sets all the time are under 60%. It indicates that the effect on recognition performance of surface and time differences is more severe than view, shoe and briefcase differences, as shown in [8]. Along with increase of L , the recognition accuracy of SPE_LPSEL_TCB experiences the rapid growth at the beginning, the stable increase trend in the middle and maintain at a high recognition performance of level in the last. The results show that, SPE_LPSEL_TCB is able to converge to the best recognition performance.
- For the image patch size, SPE_LPSEL_TCB with the image patch

Table 10

Rank-1 recognition (%) of different classifiers on the USF HumanID dataset.

Algorithms	LDA_SEL_TCB			MFA_TSEL_TCB			SPE_LPSEL_TCB		
	NN	NFS	SRC	NN	NFS	SRC	NN	NFS	SRC
Probe A	88	89	84	95	95	97	93	98	95
Probe B	91	96	94	91	91	91	93	94	94
Probe C	76	85	78	78	80	85	85	83	83
Probe D	51	56	50	66	64	75	53	58	55
Probe E	45	53	48	59	62	62	59	53	48
Probe F	44	42	41	46	47	48	43	38	36
Probe G	45	45	48	52	50	53	43	40	38
Probe H	87	90	86	93	92	93	92	91	92
Probe I	84	90	81	88	90	90	90	91	88
Probe J	44	44	48	69	73	73	71	77	75
Probe K	64	52	42	30	18	27	27	24	27
Probe L	42	33	21	27	30	27	24	27	24
Avg.	63.74	65.65	62	70.49	70.7	73.46	68.47	69.48	67.41

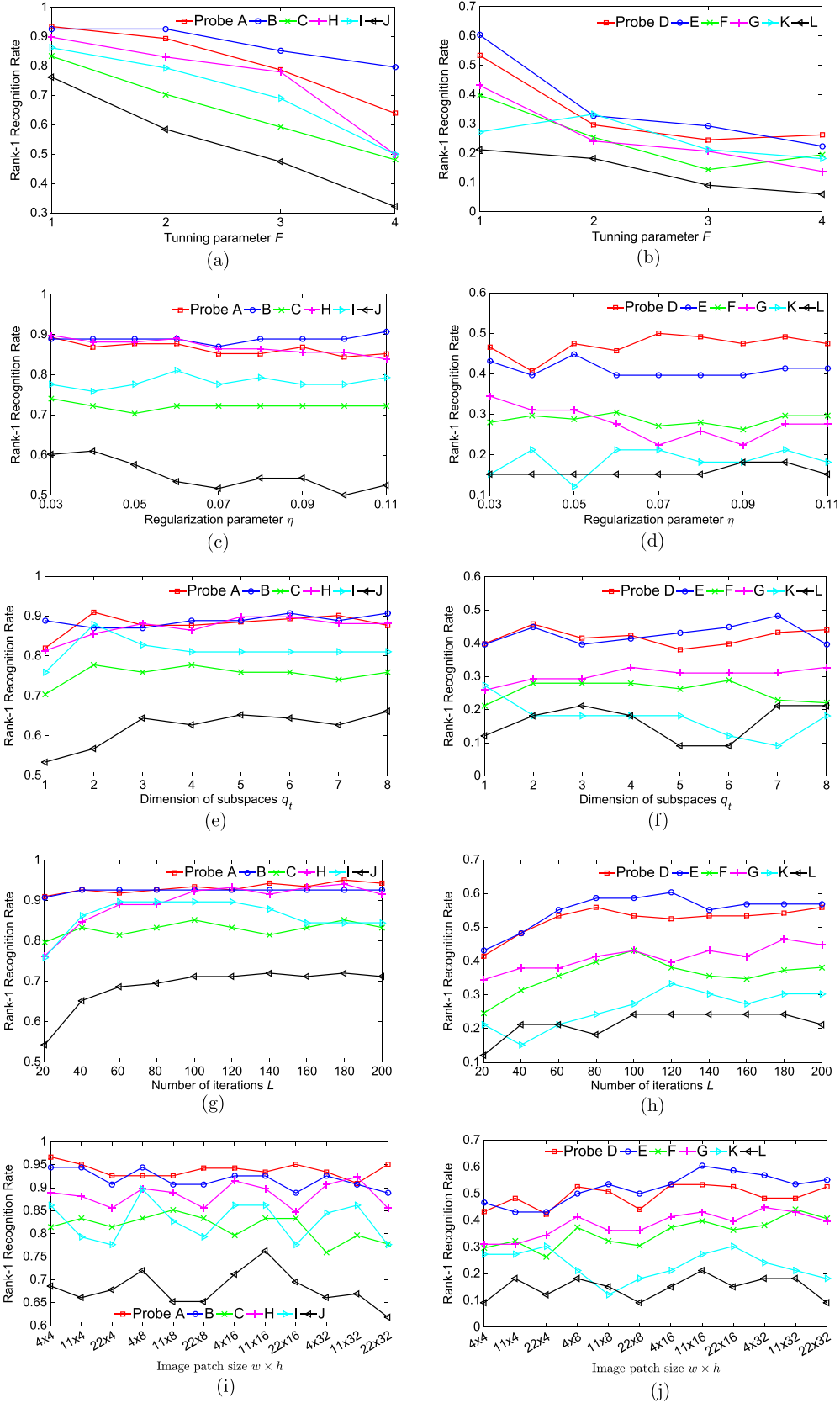


Fig. 7. Parameter analysis of SPE_LP_SEL_TCB on probe sets of the USF HumanID database. Rank-1 recognition rate (percentage) with respective to different (a) and (b) tuning parameter F , (c) and (d) regularization parameter η , (e) and (f) dimensionality of each subspace q_t , (g) and (h) number of iterations L , (i) and (j) image patch size $w \times h$.

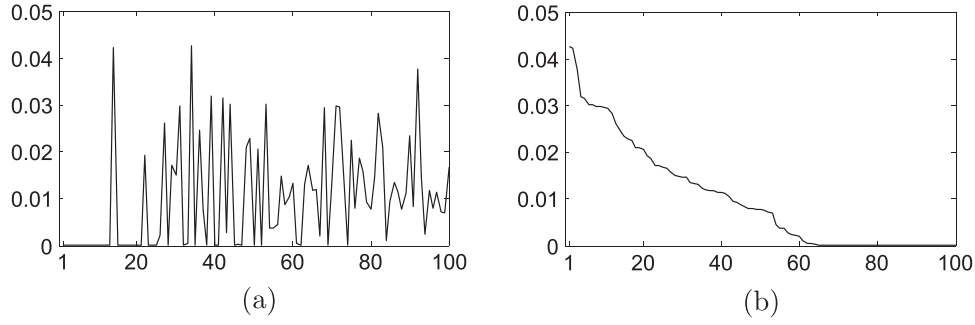


Fig. 8. Weight distribution of 100 learned local patch-based subspaces using SPE_LPSEL_TCB. (a) Original weight distribution, (b) Descended weight distribution.

size of 11×16 achieves marginally lower Rank-1 recognition rates than ones with the optimal image patch size on the most probe subsets and the best performance on the E, J and L probe subsets.

4.2.3. Discussion on sparsity

In this section, we discuss and analyze the sparsity issue of our proposed algorithms. The number of subspaces in combination directly influences computational efficiency of the subspace ensemble learning algorithm in the real application, i.e., the less number of combined subspaces will bring the better real-timeness. Fig. 8(a) shows the original weight distribution of 100 local patch-based subspaces of SPE_LPSEL_TCB and Fig. 8(b) plots the corresponding sorted weight distribution in descending order. In Fig. 8(b), we can observe that 36 subspaces are assigned with zero weight value ($< e^{-10}$), which indicates that these subspaces with zero weight value can be discarded to reduce computational costs. Table 9 lists the percentage of the number of subspaces with nonzero weight value in the total learned subspaces. From the Table 9, it can be seen that our methods can give a relatively sparse combination on the learned subspaces to improve the computational efficiency.

4.2.4. Discussion on classification

The SEL_TCB framework and its extensions are intended to effectively learn and combine features in multiple subspaces. It is interesting that these combined features are applied with variety of classifiers for gait recognition. Besides NN classifier, we further discuss the recognition performance of combined features in multiple subspaces with two other classifiers: nearest feature space (NFS) classifier [53] and sparse representation based classification (SRC) [54]. For the NFS classifier, the dissimilarity between the probe sequence I_P and the gallery sequence I_G is defined as

$$d_{\text{nfs}}(I_P, I_P) = \frac{1}{N_P} \sum_{i=1}^{N_P} \|f(\mathbf{w}_P(i)) - \mathbf{F}_G \alpha_G^i\|_2, \quad (30)$$

where $\mathbf{F}_G = [f(\mathbf{w}_G(1)) f(\mathbf{w}_G(2)) \cdots f(\mathbf{w}_G(N_G))]$ and $\alpha_G^i = (\mathbf{F}_G^T \mathbf{F}_G)^{-1} \mathbf{F}_G^T f(\mathbf{w}_P(i))$. For SRC, the dissimilarity between the probe sequence I_P and the gallery sequence I_G is defined as

$$d_{\text{src}}(I_P, I_P) = \frac{1}{N_P} \sum_{i=1}^{N_P} \|f(\mathbf{w}_P(i)) - \mathbf{F}_G \delta_G^i\|_2, \quad (31)$$

where δ_G^i is a coefficient vector with elements of δ^i associated with I_G , and δ^i is the coefficient vector for $f(\mathbf{w}_P^i)$ which is sparsely reconstructed from all gallery sequences. Table 10 shows the average Rank-1 recognition rate of different classifiers together with features learned from LDA_SEL_TCB, MFA_TSEL_TCB and SPE_TSEL_TCB on the USF HumanID dataset. We observe that MFA_TSEL_TCB with SRC

achieves the highest average Rank-1 recognition rate (73.46%) and the NFS classifier achieves better recognition performance than the NN classifier with three different combined features. It indicates that features extracted from the proposed subspace ensemble learning framework can be applied to improve the performance of the classifier for gait recognition.

5. Conclusion

In this work, we have presented a novel subspace ensemble learning framework and extended it to tensor-based and local patch-based versions to perform gait recognition. First, by building the triplet set, the class relationship among training samples is revealed. Then, based on the triplet set, the subspace ensemble learning problem is presented as linear programming problem by totally-corrective boosting technique. As for the linear programming problem, column generation technique is used to iteratively learn and combine subspaces. At each iteration, a new subspace is learned and added, and then the weight distribution on all the learned subspaces is jointly updated. To enhance the discriminant information of gait template, Gabor features with different scales and orientations are extracted. Meanwhile, the proposed subspace ensemble framework is extended to learn and combine subspaces in the tensor space, which can preserve the structure information of multi-dimensionality gait template. Finally, the proposed framework is extended to solve the local patch-based subspace ensemble learning problem to select the most useful image patches or features, and then learn and combine the corresponding subspaces for recognition. A mass of the comparative experiments on the USF HumanID database and CASIA gait database [34] demonstrate that our proposed algorithms highly outperform the state-of-the-art gait recognition algorithms.

For future work, we will pay attention to sparse combination problem in subspace ensemble learning. The sparseness of the weights largely determines the computational efficiency. Although we have demonstrated that our proposed framework can give the relatively sparse combination of the learned subspaces, explicit control on the sparseness still needs further study.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (61525303), the Top-Notch Young Talents Program of China (Ligang Wu), the Self-Planned Task of State Key Laboratory of Robotics and System (HIT)(201505B) and the Fok Ying Tung Education Foundation (141059).

Appendix: Derivation of Lagrangian dual

Let the partial derivatives of $\mathcal{L}$ in (4) w.r.t. a_t , ζ_m and ρ be equal to 0 as follows:

$$\frac{\partial \mathcal{L}}{\partial a_t} = - \sum_{m=1}^M u_m h_{mt} - q_t - \beta = 0, \quad t = 1, 2, \dots, l$$

$$\frac{\partial \mathcal{L}}{\partial \zeta_m} = D - u_m - g_m = 0, \quad m = 1, 2, \dots, M$$

$$\frac{\partial \mathcal{L}}{\partial \rho} = -1 + \sum_{m=1}^M u_m = 0,$$

then we can obtain

$$\sum_{m=1}^M u_m h_{mt} = -q_t - \beta \leq -\beta, \quad t = 1, 2, \dots, l$$

$$u_m = D - g_m \leq D, \quad m = 1, 2, \dots, M$$

$$\sum_{m=1}^M u_m = 1.$$

By plugging the above equations into $\mathcal{L}$ in (4), $\mathcal{L} = \beta$ is obtained as follows:

$$\begin{aligned} \mathcal{L} &= -\rho + D \sum_{m=1}^M \zeta_m - \sum_{t=1}^l a_t \sum_{m=1}^M u_m h_{mt} + \rho \sum_{m=1}^M u_m - \sum_{m=1}^M u_m \zeta_m - \beta \sum_{t=1}^l a_t + \beta - \mathbf{g}^T \boldsymbol{\zeta} - \mathbf{q}^T \mathbf{a} \\ &= -\rho + D \sum_{m=1}^M \zeta_m - \sum_{t=1}^l a_t (-q_t - \beta) + \rho - \sum_{m=1}^M (D - g_m) \zeta_m - \beta \sum_{t=1}^l a_t + \beta - \mathbf{g}^T \boldsymbol{\zeta} - \mathbf{q}^T \mathbf{a} \\ &= \beta. \end{aligned}$$

Thus, the Lagrangian dual of (4) is updated as $\max_{\mathbf{u}, \beta} \mathcal{L} = \beta$ with $\sum_{m=1}^M u_m h_{mt} \leq -\beta$, $u_m \leq D$ and $\sum_{m=1}^M u_m = 1$. Let $\beta = -\beta$. Finally, we obtain $\min_{\mathbf{u}, \beta} \mathcal{L} = \beta$ with $\sum_{m=1}^M u_m h_{mt} \leq \beta$, $u_m \leq D$ and $\sum_{m=1}^M u_m = 1$, $t = 1, 2, \dots, l$, $m = 1, 2, \dots, M$.

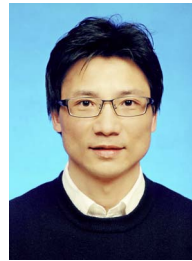
References

- [1] R. Martín-Félez, T. Xiang, Uncooperative gait recognition by learning to rank, *Pattern Recognit.* 47 (12) (2014) 3793–3806.
- [2] D. Muramatsu, A. Shiraishi, Y. Makihara, M.Z. Uddin, Y. Yagi, Gait-based person recognition using arbitrary view transformation model, *IEEE Trans. Image Process.* 24 (1) (2015) 140–154.
- [3] T. Wang, S. Gong, X. Zhu, S. Wang, Person re-identification by discriminative selection in video ranking, *IEEE Trans. Pattern Anal. Mach. Intell.* 38 (12) (2016) 2501–2514.
- [4] M.R. Daliri, Automatic diagnosis of neuro-degenerative diseases using gait dynamics, *Measurement* 45 (7) (2012) 1729–1734.
- [5] M.R. Daliri, Chi-square distance kernel of the gaits for the diagnosis of parkinson's disease, *Biomed. Signal Process. Control* 8 (1) (2013) 66–70.
- [6] A. Khorasani, M.R. Daliri, HMM for classification of parkinson's disease based on the raw gait data, *J. Med. Syst.* 38 (12) (2014) 1–6.
- [7] A. Khorasani, M.R. Daliri, M. Pooyan, Recognition of amyotrophic lateral sclerosis disease using factorial hidden markov model, *Biomed. Eng./Biomed. Tech.* 61 (1) (2016) 119–126.
- [8] S. Sarkar, P.J. Phillips, Z. Liu, I.R. Vega, P. Grother, K.W. Bowyer, The humanid gait challenge problem: data sets, performance, and analysis, *IEEE Trans. Pattern Anal. Mach. Intell.* 27 (2) (2005) 162–177.
- [9] J. Han, B. Bhanu, Individual recognition using gait energy image, *IEEE Trans. Pattern Anal. Mach. Intell.* 28 (2) (2006) 316–322.
- [10] M.R. Aqmar, Y. Fujihara, Y. Makihara, Y. Yagi, Gait recognition by fluctuations, *Comput. Vis. Image Underst.* 126 (2014) 38–52.
- [11] A. Sundaresan, A. Roy-Chowdhury, R. Chellappa, A hidden markov model based framework for recognition of humans from gait sequences, in: *Proceedings of the IEEE International Conference on Image Processing*, Vol. 2, 2003, pp. 143–150.
- [12] M.S. Nixon, J.N. Carter, Automatic recognition by gait, in: *Proceedings of the IEEE* 94 (11) (2006) pp. 2013–2024.
- [13] W. Zeng, C. Wang, F. Yang, Silhouette-based gait recognition via deterministic learning, *Pattern Recognit.* 47 (11) (2014) 3568–3584.
- [14] Z. Liu, S. Sarkar, Improved gait recognition by gait dynamics normalization, *IEEE Trans. Pattern Anal. Mach. Intell.* 1 (6) (2006) 863–876.
- [15] X. Huang, N.V. Boulgouris, Gait recognition with shifted energy image and structural feature extraction, *IEEE Trans. Image Process.* 21 (4) (2012) 2256–2268.
- [16] C. Wang, J. Zhang, L. Wang, J. Pu, X. Yuan, Human identification using temporal information preserving gait template, *IEEE Trans. Pattern Anal. Mach. Intell.* 34 (11) (2012) 2164–2176.
- [17] S.D. Choudhury, T. Tjahjedi, Robust view-invariant multiscale gait recognition, *Pattern Recognit.* 48 (3) (2015) 798–811.
- [18] G.V. Veres, L. Gordon, J.N. Carter, M.S. Nixon, What image information is important in silhouette-based gait recognition?, in: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. 2, 2004, pp. 776–782.
- [19] Y. Dupuis, X. Savatier, P. Vasseur, Feature subset selection applied to model-free gait recognition, *Image Vis. Comput.* 31 (8) (2013) 580–591.
- [20] J. Ye, R. Janardan, Q. Li, Two-dimensional linear discriminant analysis, in: *Advances in neural information processing systems*, 2004, pp. 1569–1576.
- [21] D. Xu, S. Yan, D. Tao, S. Lin, H.-J. Zhang, Marginal fisher analysis and its variants for human gait recognition and content-based image retrieval, *IEEE Trans. Image Process.* 16 (11) (2007) 2811–2821.
- [22] E. Zhang, Y. Zhao, W. Xiong, Active energy image plus 2dpp for gait recognition, *Signal Process.* 90 (7) (2010) 2295–2302.
- [23] D. Tao, X. Li, X. Wu, S.J. Maybank, General tensor discriminant analysis and gabor features for gait recognition, *IEEE Trans. Pattern Anal. Mach. Intell.* 29 (10) (2007) 1700–1715.
- [24] X. Li, S. Lin, S. Yan, D. Xu, Discriminant locally linear embedding with high-order tensor data, *IEEE Trans. Syst., Man, Cybern. B* 38 (2) (2008) 342–352.
- [25] L. Zhang, L. Zhang, D. Tao, B. Du, A sparse and discriminative tensor to vector projection for human gait feature representation, *Signal Process.* 106 (2015) 245–252.
- [26] X. Wang, X. Tang, Random sampling for subspace face recognition, *Int. J. Comput. Vis.* 70 (1) (2006) 91–104.
- [27] J. Bi, D. Wu, L. Lu, M. Liu, Y. Tao, M. Wolf, Adaboost on low-rank psd matrices for metric learning, in: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2011, pp. 2617–2624.
- [28] Y. Guan, C.-T. Li, F. Roli, On reducing the effect of covariate factors in gait recognition: a classifier ensemble method, *IEEE Trans. Pattern Anal. Mach. Intell.* 37 (7) (2015) 1521–1528.
- [29] M. Galar, A. Fernández, E. Barrenechea, H. Bustince, F. Herrera, An overview of ensemble methods for binary classifiers in multi-class problems: experimental study on one-vs-one and one-vs-all schemes, *Pattern Recognit.* 44 (8) (2011) 1761–1776.
- [30] X. Li, S.J. Maybank, S. Yan, D. Tao, D. Xu, Gait components and their application to gender recognition, *IEEE Trans. Syst., Man, Cybern. C* 38 (2) (2008) 145–155.
- [31] S. Yu, T. Tan, K. Huang, K. Jia, X. Wu, A study on gait-based gender classification, *IEEE Trans. Image Process.* 18 (8) (2009) 1905–1910.
- [32] I. Rida, X. Jiang, G.L. Marcialis, Human body part selection by group lasso of motion for model-free gait recognition, *IEEE Signal Process. Lett.* 23 (1) (2016) 154–158.
- [33] A. Demiriz, K.P. Bennett, J. Shawe-Taylor, Linear programming boosting via column generation, *Mach. Learn.* 46 (1–3) (2002) 225–254.
- [34] S. Yu, D. Tan, T. Tan, A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition, in: *Proceedings of the 18th IEEE*

- International Conference on Pattern Recognition, Vol. 4, 2006, pp. 441–444.
- [35] G. Ma, M. Liu, L. Wu, H.R. Karimi, Local patch-based subspace ensemble learning via totally-corrective boosting for gait recognition, in: Proceedings of International Conference on Machine Learning, ICML Workshop, Citeseer, 2013.
 - [36] S. Yan, D. Xu, B. Zhang, H.-J. Zhang, Q. Yang, S. Lin, Graph embedding and extensions: a general framework for dimensionality reduction, *IEEE Trans. Pattern Anal. Mach. Intell.* 29 (1) (2007) 40–51.
 - [37] B. Moghaddam, T. Jebara, A. Pentland, Bayesian face recognition, *Pattern Recognit.* 33 (11) (2000) 1771–1782.
 - [38] X. Wang, X. Tang, A unified framework for subspace face recognition, *IEEE Trans. Pattern Anal. Mach. Intell.* 26 (9) (2004) 1222–1228.
 - [39] H.-T. Chen, H.-W. Chang, T.-L. Liu, Local discriminant embedding and its variants, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. 2, 2005, pp. 846–853.
 - [40] Y. Wang, W. Li, J. Zhou, X. Li, Y. Pu, Identification of the normal and abnormal heart sounds using wavelet-time entropy features based on oms-wpd, *Future Gener. Comput. Syst.* 22 (10) (2013) 3904–3915.
 - [41] H. Lu, K.N. Plataniotis, A.N. Venetsanopoulos, A survey of multilinear subspace learning for tensor data, *Pattern Recognit.* 44 (7) (2011) 1540–1551.
 - [42] Z. Lai, Y. Xu, J. Yang, J. Tang, D. Zhang, Sparse tensor discriminant analysis, *IEEE Trans. Image Process.* 22 (10) (2013) 3904–3915.
 - [43] M.A.O. Vasilescu, D. Terzopoulos, Multilinear analysis of image ensembles: Tensorfaces, in: Proceedings of the 7th European Conference on Computer Vision, Springer, 2002, pp. 447–460.
 - [44] S. Yan, D. Xu, Q. Yang, L. Zhang, X. Tang, H.-J. Zhang, Discriminant analysis with tensor representation, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. 1, 2005, pp. 526–532.
 - [45] J. Zou, Q. Ji, G. Nagy, A comparative study of local matching approach for face recognition, *IEEE Trans. Image Process.* 16 (10) (2007) 2617–2628.
 - [46] Y. Wang, P. Zhang, L. An, G. Ma, J. Kang, F. Shi, X. Wu, J. Zhou, D.S. Lalush, W. Lin, D. Shen, Predicting standard-dose pet image from low-dose pet and multimodal mr images using mapping-based sparse representation, *Phys. Med. Biol.* 61 (2) (2016) 791–812.
 - [47] J. Renegar, A polynomial-time algorithm, based on newton's method, for linear programming, *Math. Program.* 40 (1–3) (1988) 59–93.
 - [48] D. Xu, Y. Huang, Z. Zeng, X. Xu, Human gait recognition using patch distribution feature and locality- constrained group sparse representation, *IEEE Trans. Image Process.* 21 (1) (2012) 316–326.
 - [49] J. Lu, G. Wang, P. Moulin, Human identity and gender recognition from gait sequences with arbitrary walking directions, *IEEE Trans. Inf. Forensics Secur.* 9 (1) (2014) 51–61.
 - [50] Z. Lai, Y. Xu, Z. Jin, D. Zhang, Human gait recognition via sparse discriminant projection learning, 24, 10, 2014, 1651–1662.
 - [51] X. He, S. Yan, Y. Hu, P. Niyogi, H.-J. Zhang, Face recognition using Laplacian faces, *IEEE Trans. Pattern Anal. Mach. Intell.* 27 (3) (2005) 328–340.
 - [52] C. Chen, J. Zhang, R. Fleischer, Multilinear tensor-based non-parametric dimension reduction for gait recognition, in: *Advances in Biometrics*, Springer, 2009, pp. 1030–1039.
 - [53] J.-T. Chien, C.-C. Wu, Discriminant wavelet faces and nearest feature classifiers for face recognition, *IEEE Trans. Pattern Anal. Mach. Intell.* 24 (12) (2002) 1644–1649.
 - [54] J. Wright, A.Y. Yang, A. Ganesh, S.S. Sastry, Y. Ma, Robust face recognition via sparse representation, *IEEE Trans. Pattern Anal. Mach. Intell.* 31 (2) (2009) 210–227.

Guangkai Ma received the B.S. degree in Harbin University of Science and Technology and Technology, China in 2010; the M.E. degree in Space Control and Inertial Technology Research Center, Harbin Institute of Technology, China in 2012. He was a joint-Ph.D. student in IDEA Lab of Department of Radiology and BRIC, University of North Carolina at Chapel Hill, NC, USA in 2013 to 2015. He is currently pursuing his

Ph.D. degree in Inertial Technology Research Center, Harbin Institute of Technology, China. His current research interests include gait recognition, pattern recognition and medical image analysis.



Ligang Wu (M'10-SM'12) received the B.S. degree in Automation from Harbin University of Science and Technology, China in 2001; the M.E. degree in Navigation Guidance and Control from Harbin Institute of Technology, China in 2003; the PhD degree in Control Theory and Control Engineering from Harbin Institute of Technology, China in 2006. From January 2006 to April 2007, he was a Research Associate in the Department of Mechanical Engineering, The University of Hong Kong, Hong Kong. From September 2007 to June 2008, he was a Senior Research Associate in the Department of Mathematics, City University of Hong Kong, Hong Kong. From December 2012 to December 2013, he was a Research Associate in the Department of Electrical and Electronic Engineering, Imperial College London, London, UK. In 2008, he joined the Harbin Institute of Technology, China, as an Associate Professor, and was then promoted to a Professor in 2012. Dr. Wu was the winner of the National Science Fund for Distinguished Young Scholars in 2015. He received China Young Five Four Medal in 2016, and was named to the 2015 & 2016 Thomson Reuters Highly Cited Researchers lists. Dr. Wu currently serves as an Associate Editor for a number of journals, including *IEEE Transactions on Automatic Control*, *IEEE/ASME Transactions on Mechatronics*, *Information Sciences*, *Signal Processing*, and *IET Control Theory and Applications*. He is also an Associate Editor for the Conference Editorial Board, *IEEE Control Systems Society*. Dr. Wu has published 5 research monographs and more than 120 research papers in international referred journals. His current research interests include switched systems, stochastic systems, computational and intelligent systems, multidimensional systems, sliding mode control, and flight control.



Yan Wang received her Bachelor degree, Master degree and Doctoral degree in Electronic and Information Engineering of Sichuan University, Chengdu, China in 2009, 2012 and 2015, respectively. She visited the Department of Radiology, University of North Carolina at Chapel Hill, USA as a visiting scholar in 2014 and 2015. Now, she is a lecturer in College of Computer Science in Sichuan University. Her research interests lies in the field of Biomedical Imaging and Machine Learning.