Human Gait Recognition via Sparse Discriminant Projection Learning

Zhihui Lai, Yong Xu, *Member, IEEE*, Zhong Jin, and David Zhang, *Fellow, IEEE*

***Abstract*— As an important biometric feature, human gait has great potential in video-surveillance-based applications. In this paper, we focus on the matrix representation-based human gait recognition and propose a novel discriminant subspace learning method called sparse bilinear discriminant analysis (SBDA). SBDA extends the recently proposed matrix-representation-based discriminant analysis methods to sparse cases. By introducing the L_1 and L_2 norms into the objective function of SBDA, two interrelated sparse discriminant subspaces can be obtained for gait feature extraction. Since the optimization problem has no closed-form solutions, an iterative method is designed to compute the optimal sparse subspace using the L_1 and L_2 norms sparse regression. Theoretical analyses reveal the close relationship between SBDA and previous matrix-representation-based discriminant analysis methods. Since each nonzero element in each subspace is selected from the most important variables/factors, SBDA is potential to perform equivalent to or even better than the state-of-the-art subspace learning methods in gait recognition. Moreover, using the strategy of SBDA plus linear discriminant analysis (LDA), we can further improve the performance. A set of experiments on the standard USF HumanID and CASIA gait databases demonstrate that the proposed SBDA and SBDA + LDA can obtain competitive performance.**

***Index Terms*— Feature extraction, gait recognition, linear discriminant analysis (LDA), sparse regression.**

I. Introduction

RECENT research [1], [2] has shown that individuals have distinctive and special ways of walking, which can be used to identify different individuals. Therefore, human gait recognition has attracted much attention in the field of pattern recognition and it can be widely used in video-surveillance-based identification and safety monitoring. The advantage of gait recognition is that it can be captured from a distance and recognize the individuals without being obtrusive.

There are numerous schemes proposed for gait recognition in the past decade, which can be roughly divided into two categories: 1) model-based and 2) motion-based techniques. The model-based techniques [3]–[5] used the model parameters fitted to describe the human body structure, while the motion-based techniques employed a low-dimensional compact representation to describe the motion patterns of the human body. Compared with the time-series-based silhouette representation [1], previous research [6], [7] had demonstrated that the gray-level average silhouette over a gait cycle was an efficient and effective representation for gait recognition. As a result, motion-based techniques were developed very fast and a lot of effective algorithms were proposed to obtain the low-dimensional representation of the gait for recognition. In particular, subspace learning methods have attracted increasing attention in recent years.

As two classical subspace learning methods to obtain the low-dimensional compact representation, principal component analysis (PCA) [8]–[10] and linear discriminant analysis (LDA) [11]–[13] were also used for gait recognition [14]–[16]. Since the average silhouette was directly represented in the form of a matrix, Xu *et al.* [17] proposed to use the coupled subspace analysis (CSA) [18] and discriminant analysis with tensor representation (DATER) [19] for gait recognition. Tao *et al.* [20] proposed the general tensor discriminant analysis (GTDA) [20] for gait recognition. By exploiting the local structure information of the data set, discriminant locally linear embedding [21] and matrix marginal Fisher analysis (MMFA) [22] can also achieve competitive performance on gait recognition. Recently, the new distance measure [23] based on the integer programming method, multilevel features [24], and spatio-temporal neighborhood topology features [25] were also proposed for gait recognition.

With the fast development of the subspace learning methods, the recently proposed sparse subspace learning methods attracted much attention, owing to the nature of the L_1 norm minimization for sparse feature selection [26], [27] and the robustness in recognition tasks [28]. The least absolute shrinkage and selection operator [29], least angle regression [30], and the elastic net [31] were widely used for feature selection and obtained better performances than

Manuscript received May 27, 2013; revised August 29, 2013 and October 27, 2013; accepted November 4, 2013. Date of publication February 11, 2014; date of current version October 1, 2014. This work was supported in part by the Natural Science Foundation of China under Grants 61203376, 61203243, 61202276, 61005005, 61071179, 61125305, 61170077, 61362031, and 61373063; in part by the Central Fund from the Hong Kong Polytechnic University (K.14.37.ZP43) as well as the China Post-Doctoral Science Foundation under Project 2013T60370, 2012M511479; in part by the Guangdong Natural Science Foundation under Project S2012040007289; and in part by the Shenzhen Municipal Science and Technology Innovation Council under Grants JC201005260122A, JCYJ20120613153352732, and JCYJ201206131348 43060. This paper was recommended by Associate Editor C. Shan. *(Corresponding author: Y. Xu)*

Z. Lai and Y. Xu are with the Bio-Computing Research Center, Shenzhen Graduate School, Harbin Institute of Technology, Shenzhen 518055, China, and also with the College of Computer Science and Software Engineering, Shenzhen University, Shenzhen 518060, China (e-mail: lai_zhi_hui@163.com; yongxu@ymail.com).

Z. Jin is with the School of Computer Science, Nanjing University of Science and Technology, Nanjing 210094, China (e-mail: zhongjin@njust.edu.cn).

D. Zhang is with the Department of Computing, Biometrics Research Centre, Hong Kong Polytechnic University, Kowloon, Hong Kong (e-mail: csdzhang@comp.polyu.edu.hk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCSVT.2014.2305495

the ordinary least squares regression. Using the L_1 norm sparse learning models, Zou *et al.* [26] reformulated the PCA as regression-type optimization and proposed sparse PCA (SPCA) [26] to achieve the goal of spare feature selection and dimensionality reduction with interpretabilities. Using the label information, sparse discriminant analysis (SDA) [32] and sparse LDA (SLDA) [33] were proposed for learning a sparse discriminant subspace for feature extraction and classification in biological or medical data analysis, such as microarrays. Based on maximal marginal criterion, sparse tensor discriminant analysis (STDA) [34] was proposed for face and action feature extraction. These sparse regression based methods have been proved to be effective for classification and prediction.

Until now, though sparse subspace learning methods have been used in many fields, gait recognition with sparse manner has not been investigated and how to extend the discriminant analysis algorithms with sparseness to the natural representation of gait recognition is unsolved. These sparse learning methods mentioned above may not be suitable for gait recognition due to the following two reasons: first, the gait silhouette is directly represented in the form of matrix but these sparse subspace methods cannot operate on the silhouette matrix; second, if these methods are used for learning the discriminant subspace on larger than 10 000 ($128 \times 88 = 11264$)-dimensional vector space concatenated by the silhouette matrix, the computation burden will be very large and the underlying structure information of the second-order silhouette matrix will be lost. Therefore, it is desirable to design new algorithms for this special case of gait feature extraction and recognition.

In this paper, inspired by recent advances in matrix-based dimensionality reduction algorithms [18]–[21] and the sparse subspace learning algorithms, we propose matrix-based sparse bilinear discriminant analysis (SBDA) directly handling gray-level averaged images as 2-D data for feature extraction and classification. Our starting point is to introduce the L_1 and L_2 norms penalty on the projection vectors/matrices and use sparse regression method to select the most discriminant features/variables to form the projections. Thus, two interrelated sparse projection vectors/matrices formed by the important discriminant variables embedded in the gait matrices are obtained. Therefore, it is expected to explore the more powerful discriminant subspaces than other methods for gait recognition.

SBDA has two significantly different properties from the existing tensor/matrix-based feature extraction methods. First, different from the existing tensor-based methods, such as the MMFA, DATER, and GTDA, in which the discriminant subspaces are nonsparse, all the projection vectors/matrices of SBDA are sparse. Second, the optimal bilinear sparse discriminant subspaces of SBDA are obtained using the elastic net regression, whereas previous matrix-based subspace learning methods solve the eigenequations. Although both STDA and SBDA use the elastic net regression to compute the sparse solutions, it is worthy to note that the proposed SBDA in this paper is not the bilinear version/case of STDA and the image matrix-based STDA cannot degrade to be SBDA and vice versa. The key ideas, the models, and the optimization procedures of STDA and SBDA are all different.

The main contributions of this paper are as follows: first, we define a novel divergence bilinear scatter value, which is integrated to the matrix-representation-based discriminant analysis. By introducing the L_1 and L_2 norms penalty terms, the sparse discriminant subspace can be obtained from an alternately iterative procedures using elastic net. Second, the relationships between SBDA and other algorithms, i.e., DATER and 2-D LDA, are theoretically analyzed. Third, methodologically speaking, this paper provides a novel idea on how to design the different scatter matrices to solve the sparse multicorrelated constraint learning model in discriminant analysis, which is significantly different from SPCA, SLDA, and STDA.

The rest of this paper is organized as follows. In Section II, SBDA algorithm and related analyses are presented. In Section III, theoretical analysis is presented for investigating the relationships between SBDA and the previous methods, and the convergence and the computational complexity are also given. Experiments are carried out to evaluate SBDA algorithm in Section IV for gait recognition, and the conclusion is given in Section V.

II. Sparse Bilinear Discriminant Analysis

In this section, we first present the idea of SBDA and briefly define some basic bilinear notations, definitions, and operations, and then present the SBDA algorithm.

A. Idea and Preparation for SBDA

Assume that the training samples are represented as the matrix $\{X_i \in R^{m_1 \times m_2}, i = 1, 2, \ldots, N\}$, where N denotes the total number of the training samples. Moreover, let N_c and N_{c_i} denote the total numbers of the classes and samples in the ith class, respectively.

Since previous research shows that the matrix-based feature extraction algorithms can obtain competitive performance on gait recognition, in this way, the goal of the SBDA is also to obtain two projection matrices $U \in R^{m_1 \times d_1}$ ($d_1 \le m_1$) and $V \in R^{m_2 \times d_2}$ ($d_2 \le m_2$) that map the original silhouette image into a low-order feature matrix

$$Y_i = U^T X_i V. \tag{1}$$

Different from previous methods, the two projections of SBDA are sparse. Our idea is to introduce the sparsity in the two projection matrices U and V for feature selection. To enhance the discriminant ability of the algorithm, the sparsity and Fisher criterion are combined together to select the most useful discriminant features for classification.

Let us give an intuitive example to show how the proposed method is suitable for gait recognition. Suppose X_1 and X_2 denote the Gallery gait with shoe and the Probe gait without shoe of somebody, the upper parts of the matrices in X_1 and X_2 should be very similar, but the lower parts should be very different since the only difference between them is with or without shoe. If the proposed sparse learning method can obtain two sparse projection for feature selection (for simplicity, let us take only one projection in U and V for example, that is, $U = u_1 = (0, 0, 1, 1, 0, \ldots, 0)^T$ and $V = v_1 = (0, 1, 0, 1, 0, \ldots, 0)$), then we obtain the low-dimensional feature $Y_1 = u_1^T X_1 v_1$ and $Y_2 = u_2^T X_2 v_2$. From the matrix multiplication, we can find that the bottom part of the matrices X_1 and X_2 have no contribution to Y_1 and Y_2, due to the zero elements at the posterior part of projection u_1 (similarly, v_1 will perform feature selection on the columns of matrices X_1 and X_2). Thus, when we integrate the sparsity with the Fisher criterion, the selected low-dimensional features between the Gallery gait with shoe and the Probe gait without shoe of somebody will be more similar (closer) than the original features, which will be more suitable for classification and thus potentially increase the algorithm's discriminative performance. Therefore, SBDA may be more suitable for gait recognition. However, the nonsparse feature extraction methods have no such function since the projections are not sparse.

To achieve this goal, a set of definitions must be given at first.

Definition 1: The matrix-based within-class bilinear scatter value J_W is defined as

$$J_W(U, V) = \sum_{j=1}^{N_c} \sum_{i=1}^{N_{c_i}} \left\| U^T X_i V - U^T \bar{X}_j V \right\|_F^2 = \sum_{j=1}^{N_c} \sum_{i=1}^{N_{c_i}} \mathrm{tr}\Big[\big(U^T X_i V - U^T \bar{X}_j V\big)^T \times \big(U^T X_i V - U^T \bar{X}_j V\big) \Big] \quad (2)$$

where $\|\cdot\|_F$ denotes the Frobenius norm of a matrix and $\bar{X}_j$ denotes the mean value matrix of the average silhouette in the jth class.

Definition 2: The matrix-based between-class bilinear scatter value is defined as

$$J_B(U, V) = \sum_{i=1}^{N_c} N_{c_i} \left\| U^T \bar{X}_i V - U^T \bar{X} V \right\|_F^2 = \sum_{i=1}^{N_c} N_{c_i} \mathrm{tr}\Big[\big(U^T \bar{X}_i V - U^T \bar{X} V\big)^T \times \big(U^T \bar{X}_i V - U^T \bar{X} V\big) \Big] \quad (3)$$

where $\bar{X}$ denotes the mean value of the samples of all the average silhouette images of the training samples.

To introduce the L_1 and L_2 norms learning into the SBDA algorithm, we also need the following definition.

Definition 3: The matrix-based within-class divergence bilinear scatter value J_D is defined as

$$J_D(U, V, P_U, P_V) = \sum_{j=1}^{N_c} \sum_{i=1}^{N_{c_i}} \left\| (P_U - U)^T (X_i - \bar{X}_j)(P_V - V) \right\|_F^2 \quad (4)$$

where P_U and P_V are also the projective matrices designed for computing the sparse discriminant subspace with the same size as U and V, respectively, and $\bar{X}_j$ denotes the mean value matrix of the average silhouette in the jth class.

With the above preparations, we can directly present the objective function of SBDA in the following section.

B. Objective Function of SBDA

One of the reasonable criteria is to minimize the within-class bilinear scatter value J_W and maximize the between-class bilinear scatter value J_B or let J_B to be a constant, that is

$$\min J_W(U, V) \quad (5)$$

$$\text{subject to } J_B(U, V) = 1. \quad (6)$$

The above optimization problem is the extension of the classical LDA with matrix representation, which has been used in DATER presented in [19]. The optimal solutions of this problem can be obtained by solving a series of generalized eigenequations. However, its solutions are not sparse. Since DATER is very effective for matrix-based gait recognition, a basic idea is to use the model of DATER to develop a more robust algorithm in a sparse manner. To obtain the sparse subspaces, the objective function of SBDA is to minimize the discriminant function of the L_1 and L_2 norms penalty optimization problem with a constraint that $J_B(U, V)$ is a constant

$$\begin{aligned} &\min J(U, V, P_U, P_V) \\ &\quad = \min \mu J_W(U, V) + (1 - \mu) J_D(U, V, P_U, P_V) \\ &\qquad + \alpha(\|P_U\|_F^2 + \|P_V\|_F^2) + \sum_r \beta_r |p_{\mathrm{vr}}| + \sum_l \gamma_l |p_{ul}| \\ &\qquad \text{subject to} \quad J_B(U, V) = 1 \end{aligned} \quad (7)$$

where $1 \geq \mu \geq 0$ is a suitable constant set by the user to balance different terms, p_{vr} is the rth column of P_V, p_{ul} is the lth column of P_U, and $|\cdot|$ denotes the L_1 norm of a vector. α is the constant set by the user and β_rs and γ_ls are the coefficients of the L_1 norm, which can be optimally determined by the elastic net [31]. Similar L_1 and L_2 norms regular techniques were also proposed in [26], [32], and [33], which show the robustness to feature extraction and recognition.

When $\mu \neq 1$, projection matrices P_U and P_V are defined as the optimal projection of SBDA. If $\mu = 1$, projection matrices P_U and P_V have nothing to do with U and V. In this case, P_U and P_V are forced to be zero matrices, and thus U and V are defined as the optimal projections of SBDA.

The above optimization problem intrinsically inherits the idea of classical LDA. Minimizing $J_W(U, V)$ indicates that the within-class bilinear scatter value should be minimized as much as possible with the constraint that the between-class bilinear scatter value is a constant. Minimizing $J_D(U, V, P_U, P_V)$ requires that the divergence within-class bilinear scatter value should approximate to zero, which means $P_U \to U$ and $P_V \to V$. Thus, the sparse projection matrices P_U and P_V can be used for feature extraction, and thus the discrimination performance of SBDA will be at least as good as DATER or even better than DATER if we suitably set the model parameters. Therefore, the projections with such properties will be also powerful for discrimination.

To the best of our knowledge, there exist no closed-form solutions for such complex objective function of SBDA, in which the variables both in the objective function and the constraint are correlated (this is significantly differed from SPCA, SDA, and STDA). Fortunately, the optimization problem can be converted into the following problem: to independently determine two subspaces P_U and P_V to minimize the within-class bilinear scatter value J_W with the constraint that J_B is a constant with the L_1 and L_2 norms penalty. Therefore, we can further obtain the sparse discriminant projections by rewriting the optimization problem as a set of independent regression problems. Thus, the key problem is to convert the sparse optimization model into the existing models that are easy to solve.

C. Solutions of the Optimization Problem

In this section, an iterative method is designed to solve the optimization problem of SBDA. The proposed iterative method includes two main steps and each step contains two substeps. At first, we fix U and P_U to compute V and P_V. Under this case, since there are still two variables, we further fix P_V to compute the optimal V and then fix optimal V to compute the optimal P_V. When the optimal V and P_V are obtained, we compute U and P_U in the same way. Iterative computation of these steps will be not terminated until the algorithm converges. This will obtain the optimal sparse subspaces P_U and P_V for feature extraction.

In the following, we fix U and P_U to compute V and P_V. In this case, from Definitions 1–3, we have

$$J_W(U,V) = \mathrm{tr}\left(V^T S_W^U V\right) \tag{8}$$

$$J_B(U,V) = \mathrm{tr}\left(V^T S_B^U V\right) \tag{9}$$

$$J_D(U,V,P_U,P_V) = \mathrm{tr}\left[(P_V - V)^T S_D^V (P_V - V)\right] \tag{10}$$

where

$$S_W^U = \sum_{j=1}^{N_c}\sum_{i=1}^{N_{c_i}} \left(U^T X_i - U^T \bar{X}_j\right)^T \left(U^T X_i - U^T \bar{X}_j\right) \tag{11}$$

$$S_B^U = \sum_{i=1}^{N_c} N_{c_i} \left(U^T \bar{X}_i - U^T \bar{X}\right)^T \left(U^T \bar{X}_i - U^T \bar{X}\right) \tag{12}$$

$$\begin{aligned} S_D^U = &\sum_{j=1}^{N_c}\sum_{i=1}^{N_{c_i}} \left[(P_U - U)^T (X_i - \bar{X}_j)\right]^T \\ &\times\left[(P_U - U)^T (X_i - \bar{X}_j)\right]. \end{aligned} \tag{13}$$

From (7) and (11)–(13), it is easy to have the following optimization problem:

$$\begin{aligned} \min_{V,P_V} \; &\mu \mathrm{tr}\left(V^T S_W^U V\right) + (1-\mu)\mathrm{tr}\left[(P_V - V)^T S_D^U (P_V - V)\right] \\ &+ \alpha \left\| P_V \right\|_F^2 + \sum_r \beta_r \left| p_{\mathrm{vr}} \right| \end{aligned} \tag{14}$$

$$\text{subject to} \quad \mathrm{tr}\left(V^T S_B^U V\right) = 1. \tag{15}$$

For the above optimization problem, we first fix V to compute the optimal P_V and then fix optimal P_V to compute the optimal V.

For the given V, the first term of (14) becomes a constant. Then, (14) reduces to

$$\begin{aligned} \min_{P_V} (1-\mu)&\mathrm{tr}\left[(P_V - V)^T S_D^U (P_V - V)\right] \\ &+ \alpha \left\| P_V \right\|_F^2 + \sum_r \beta_r \left| p_{\mathrm{vr}} \right|. \end{aligned} \tag{16}$$

Let the SVD of $S_D^U = \overline{VD}\bar{V}^T = \bar{V}\sqrt{\bar{D}}(\sqrt{\bar{D}}\bar{V}^T) = H_U H_U^T$, then we get the following optimization problem:

$$\begin{aligned} \min_{P_V} \left\| H_U^T V - H_U^T P_V \right\|_F^2 &+ \frac{\alpha}{1-\mu} \left\| P_V \right\|_F^2 \\ &+ \sum_r \frac{\beta_r}{1-\mu} \left| p_{\mathrm{vr}} \right|. \end{aligned} \tag{17}$$

This optimization problem is equal to d_2 independent elastic net problem as

$$\min_{p_i} \left\| H_U^T v_i - H_U^T p_i \right\|_2^2 + \frac{\alpha}{1-\mu} \left\| p_i \right\|_F^2 + \sum_i \frac{\beta_i}{1-\mu} \left| p_i \right| \tag{18}$$

where v_i and p_i are the ith column in V and P_V, respectively.

For the fixed P_V, (14) degrades to (19) with the constraint of (15)

$$\min_V \mu \mathrm{tr}\left(V^T S_W^U V\right) + (1-\mu)\mathrm{tr}\left[(P_V - V)^T S_D^U (P_V - V)\right]. \tag{19}$$

Using Lagrange multiplier method, we obtain

$$\begin{aligned} J(V,\lambda) = \mu \mathrm{tr}\left(V^T S_W^U V\right) &+ (1-\mu)\mathrm{tr}\left[(P_V - V)^T S_D^U (P_V - V)\right] \\ &+ \lambda \left[1 - \mathrm{tr}\left(V^T S_B^U V\right)\right] \end{aligned} \tag{20}$$

where λ is the Lagrange multiplier. Taking the derivative of $J(V,\lambda)$ with respect to V to be zero, we get

$$\mu S_W^U V - (1-\mu) S_D^U (P_V - V) - \lambda S_B^U V = 0 \tag{21}$$

$$\Rightarrow \left(\mu S_W^U + (1-\mu) S_D^U - \lambda S_B^U\right) V = (1-\mu) S_D^U P_V$$

$$\Rightarrow V = (1-\mu)\left[\mu S_W^U + (1-\mu) S_D^U - \lambda S_B^U\right]^{-1} S_D^U P_V. \tag{22}$$

Substituting (22) into (15), we get

$$\begin{aligned} \mathrm{tr}&\begin{pmatrix} \left\{(1-\mu)\left[\mu S_W^U + (1-\mu) S_D^U - \lambda S_B^U\right]^{-1} S_D^U P\right\}^T S_B^U \times \\ (1-\mu)\left[\mu S_W^U + (1-\mu) S_D^U - \lambda S_B^U\right]^{-1} S_D^U P \end{pmatrix} \\ &= 1. \end{aligned} \tag{23}$$

Thus, the optimal λ for (22) should satisfy (23). In fact, it may be not easy to directly solve the optimal λ. One can use the linear search method to get optimal λ from a big range. A tractable and effective method is to simply set λ to be a constant or to determine λ by experiments in applications, which will not significantly affect the performance of the algorithm in gait recognition. More details will be discussed in Section IV.

The above procedures compute the V and P_V when U and P_U are given/fixed.

Similarly, one can construct the matrix-based within-class scatter matrix, between-class scatter matrix, and within-class

TABLE I
SBDA ALGORITHM STEPS

Input: Gait sample matrices $\{X_i \in R^{m_1 \times m_2}, i = 1,2,...,N\}$, the numbers of iterations $T_{\max}$ and T_{EN}, the dimensions $d_i(\le m_i), i = 1,2$
Output: Low-dimensional features $Y_i = P_U^T X_i P_V$ ($i = 1,2,...,N$)
Step 1: Initialize $U^{(0)}$ and $P_U^{(0)}$ as arbitrary random matrices.
Step 2: For $t = 1 : T_{\max}$ do
Step 2: *Compute S_W^U, S_B^U, S_D^U
Step 3: *Perform SVD on S_D^U to obtain H_U
Step 4: * if $t = 1$, initialize $V = V^{(0)}$ as arbitrary columnly-orthogonal matrix, else $V \leftarrow V^{(t-1)}$
Step 5: *For $j = 1 : T_{EN}$ do
-Solve the Elastic Net problem: $P_V^{(t)} \leftarrow \arg\min \left\| H_U^T V - H_U^T P_V^{(t-1)} \right\|_F^2 + \frac{\alpha}{1-\mu} \left\| P_V^{(t-1)} \right\|_F^2 + \sum_r \frac{\beta_r}{1-\mu} \left| p_{vr}^{(t-1)} \right|$
-Compute $V^{(t)} \leftarrow (1-\mu)[\mu S_W^U + (1-\mu) S_D^U - \lambda S_B^U]^{-1} S_D^U P_V^{(t)}$
*End for
Step 6: *Compute S_W^V, S_B^V, S_D^V for given $P_V^{(t)}$ and $V^{(t)}$.
Step 7: *Perform SVD on S_D^V to obtain H_V.
Step 8: *if $t = 1$, $U = U^{(0)}$ else $U \leftarrow U^{(t-1)}$
Step 9: *For $j = 1 : T_{EN}$ do
-Solve the Elastic Net problem: $P_U^{(t)} \leftarrow \arg\min \left\| H_V^T U - H_V^T P_U^{(t-1)} \right\|_F^2 + \frac{\alpha}{1-\mu} \left\| P_U^{(t-1)} \right\|_F^2 + \sum_r \frac{\beta_r}{1-\mu} \left| p_{ur}^{(t-1)} \right|$
-Compute $U^{(t)} \leftarrow (1-\mu)[\mu S_W^V + (1-\mu) S_D^V - \lambda S_B^U]^{-1} S_D^V P_U^{(t-1)}$
*End for
Step 10: *If objective function $J(U^{(t)}, V^{(t)}, P_U^{(t)}, P_V^{(t)})$ never decreases then break.
End for
Step 11: Normalize $P_U^{(t)}$ and $P_V^{(t)}$, i.e. let $P_U(:,s) = P_U^{(t)}(:,s) / \left\| P_U^{(t)}(:,s) \right\|, s = 1 : d_1$
$P_V(:,s) = P_V^{(t)}(:,s) / \left\| P_V^{(t)}(:,s) \right\|, s = 1 : d_2$
Step 12: Project the gait samples into the low-dimensional feature space $Y_i = P_U^T X_i P_V$ ($i = 1,2,...,N$)

divergence bilinear scatter matrix with respect to V and P_V as follows:

$$S_W^V = \sum_{j=1}^{N_c} \sum_{i=1}^{N_{c_i}} \left(X_i V - \bar{X}_j V \right) \left(X_i V - \bar{X}_j V \right)^T$$

$$S_B^V = \sum_{i=1}^{N_c} N_{c_i} \left(\bar{X}_i V - \bar{X} V \right) \left(\bar{X}_i V - \bar{X} V \right)^T$$

$$S_D^V = \sum_{j=1}^{N_c} \sum_{i=1}^{N_{c_i}} \left[(X_i - \bar{X}_j)(P_V - V) \right], \left[(X_i - \bar{X}_j)(P_V - V) \right]^T.$$

Using the very similar procedures, we can compute the optimal U and P_U when V and P_V are given. Since the procedures are almost the same, we omit it in this paper.

The details of the SBDA algorithm steps are presented in Table I.

III. THEORETICAL ANALYSIS

In this section, some theoretical analyses on SBDA algorithm are presented. In addition, some properties of SBDA are revealed to discuss the relationship between SBDA and pervious discriminative analysis methods.

A. Theoretical Analysis on SBDA

For the optimization problem (7), it is obvious to get the following property.

Property 1: If $\mu = 1$, the optimization problem (7) degrades to (5) and (6). That is, DATER is a special case of SBDA.

Furthermore, the following theorems reveal the close relationship between SBDA and DATER.

Theorem 1: When U and P_U are fixed, for any given μ and V, if $\alpha \to 0_+$ and $\beta_r \to 0_+$, then $P_V \to V$.

Proof: Please see the Appendix.

Similarly, it is easy to prove Theorem 2. ■

Theorem 2: When V and P_V are fixed, for any given μ and U, if $\alpha \to 0_+$ and $\gamma_l \to 0_+$, then $P_U \to U$.

From Property 1 and Theorems 1 and 2, the following conclusion can be drawn.

Corollary 1: Let U^* and V^* be the optimal projections of DATER. If $\alpha \to 0_+$, $\beta_r \to 0_+$ and $\gamma_l \to 0_+$, then $P_V \to V^*$ and $P_U \to U^*$.

Thus, from the above analysis, it could be guaranteed that the optimal projections of the proposed SBDA can approximate to the previous ones derived from DATER, which is proved to be very effective for gait recognition [19]. Therefore, the discrimination performance of SBDA will at least as good as DATER or even better than DATER by suitably tuning the model parameters. Therefore, these properties show that the discriminant ability of SBDA will be also powerful for gait recognition.

B. Case Study on SBDA

This section provides a general case of the image-based sparse subspace learning method. However, if one wants a single subspace for feature extraction, the model of SBDA degrades to be a more simple case. For example, if one only want to compute the P_V for feature extraction, then U and P_U will vanish from the definitions and SBDA optimization model. In this case, we have

$$\min_{V,P_V} \mu \mathrm{tr}\left(V^T S_W V\right) + (1-\mu)\mathrm{tr}\big[(P_V - V)^T S_W (P_V - V)\big] + \alpha \|P_V\|_F^2 + \sum\nolimits_r \beta_r |p_{\mathrm{vr}}| \tag{24}$$

$$\text{subject to } \mathrm{tr}(V^T S_B V) = 1 \tag{25}$$

where

$$S_W = \sum_{j=1}^{N_c}\sum_{i=1}^{N_{c_i}} (X_i - \bar{X}_j)^T (X_i - \bar{X}_j)$$

$$S_B = \sum_{i=1}^{N_c} N_{c_i} (\bar{X}_i - \bar{X})^T (\bar{X}_i - \bar{X}).$$

That is, the matrix-based divergence within-class bilinear scatter matrix S_D^U degrades to be S_W. In fact, S_W and S_B are the within-class and between-class image scatter matrices in 2-D LDA [35]. Since the variables U and P_U vanish in the model, the optimization problem becomes easy to solve.

For the fixed P_V, optimization problem (24), (25) degrades to

$$\min_{V} \mu \mathrm{tr}(V^T S_W V) + (1-\mu)\mathrm{tr}\big[(P_V - V)^T S_W (P_V - V)\big]$$
$$\text{subject to } \mathrm{tr}(V^T S_B V) = 1. \tag{26}$$

Using Lagrange multiplier method, we have

$$(\mu S_W - \lambda S_B)V = (1-\mu) S_W P_V. \tag{27}$$

That is

$$V = (1-\mu)[\mu S_W - \lambda S_B]^{-1} S_W P_V. \tag{28}$$

For the fixed V, optimization problem (24), (25) gives

$$\min_{P_V} (1-\mu)\mathrm{tr}\left[(P_V - V)^T S_W (P_V - V)\right] + \alpha \|P_V\|_F^2 + \sum_r \beta_r |p_{\mathrm{vr}}|. \tag{29}$$

Let the SVD of

$$S_W = \bar{V}_W \bar{D} \bar{V}_W^T = \bar{V}_W \sqrt{\bar{D}_W}\left(\sqrt{\bar{D}_W}\bar{V}_W^T\right) = H_W H_W^T. \tag{30}$$

Then, from (29) and (30), we get the optimization problem

$$\min_{P_V} \left\| H_W^T V - H_W^T P_V \right\|_F^2 + \frac{\alpha}{1-\mu}\|P_V\|_F^2 + \sum_r \frac{\beta_r}{1-\mu}|p_{\mathrm{vr}}|. \tag{31}$$

From the above analysis, we can draw the following conclusion.

Theorem 3: Let V be the optimal projection matrix of 2-D LDA. If $\lim_{\mu\to 1_-} \alpha/1-\mu = 0$ and $\lim_{\mu\to 1_-} \beta_r/1-\mu = 0$, then $P_V \to V$.

Proof: Please see the Appendix. ■

The above theorem shows that when $\mu \to 1_-$, SBDA derives the discriminant subspace approximating to 2-D LDA, which is proved to be very effective and efficient in matrix representation-based face recognition.

C. Computational Complexity and Convergence Analysis for SBDA

1) Computational Complexity: for simplicity, we assume that $m_1 = m_2 = m$. The main computational complexity of SBDA contains two parts: to compute the inverse matrix to get V(or U) and to compute the sparse P_V (or P_U) using the elastic net algorithm. The cost to compute the inverse matrix is $O(m^3)$ and the cost to compute the sparse subspace is at most $O(m^3)$. Thus, the total computational complexity is $O(T_{\max} T_{\mathrm{EN}} m^3)$, where T_{EN} and $T_{\max}$ are the iteration numbers in the inner loop and the outer loop of SBDA, respectively. Although many loops are required for SBDA, it is still computationally much more efficient than the high-dimensional vector based methods, such as LDA [LDA needs $O(m^6)$], since the iteration numbers are usually small.

2) Convergence of SBDA: SBDA converges fast with the similar manners to other matrix-based methods. For the convergence of SBDA, we have the following theorem.

Theorem 4: The iterative procedures of SBDA presented in Table I will converge to a local optimum.

Proof: Please see the Appendix. ■

In Section IV, it will be shown that the proposed SBDA algorithm converges very fast in real-world applications.

IV. Experiment

In this section, a set of gait recognition experiments are presented to evaluate the proposed SBDA algorithm. In addition, some previous subspace learning-based gait recognition algorithms will be compared with the SBDA.

A. USF HumanID Database and Its Matrix Representation

In this paper, we focus on the learning algorithm for gait recognition applications, and thus we start our analysis from the binary image sequences. As shown in [6] and [7], the complete sequence is partitioned into several subsequences according to the gait period length, which is provided in [1] in the USF HumanID database. For each sequence, the binary

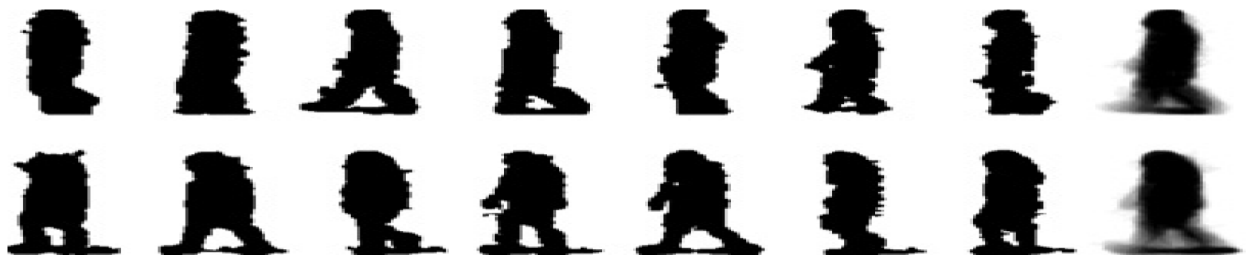

Fig. 1. Normalized and aligned binary silhouettes on USF HumanID database. Each row represents a different image sequence of the same person. The rightmost column shows the corresponding gray-level average silhouettes of the leftmost seven columns.

TABLE II
USF HumanID Gait Database (V—View, S—Shoe, U—Surface, B—Briefcase, and T—Time)

Probe set	A	B	C	D	E	F	G	H	I	J	K	L
Probe size	122	54	54	121	60	121	60	120	60	120	33	33
Gallery/Probe Difference	V	S	VS	U	SU	SV	SUV	B	SB	VB	T	TU

silhouette images within a single gait cycle are averaged to acquire several gray-level average silhouette images AS_i by

$$\mathrm{AS}_i = \frac{1}{N_{\text{Gait}}} \sum_{k=(i-1)N_{\text{Gait}}+1}^{k=iN_{\text{Gait}}} S(k), \;\; i = 1, 2, \ldots, \left\lfloor \frac{T}{N_{\text{Gait}}} \right\rfloor \tag{32}$$

where $S = \{S(T), \ldots, S(T)\}$ denotes the binary images for one sequence, T is the total number of frames, and $[T/N_{\text{Gait}}]$ denotes the floor function, which is the largest integer that is not larger than $[T/N_{\text{Gait}}]$.

On the USF HumanID database, the training set comprises all of the gray-level average silhouette images from all of the sequences in the Gallery set, and the test set comprises the average silhouette images from all of the sequences in the related Probe set. We note that, in the USF HumanID database, there are no overlapped sequences between the Gallery and Probe sets. Some samples are shown in Fig. 1. Table II shows the information of the Probe set and Gallery set. In addition, more details about the database can be found in [1].

B. Parameters' Setting of SBDA and Its Properties

1) Convergent Property: As it is shown in Fig. 2(a), SBDA algorithm converges very fast. Similar to other image-representation-based methods, SBDA usually converges within several iterations. Our empirical research indicates that even if the number of the iteration is equal to one, SBDA can also give acceptable recognition rate compared with previous subspace learning method. As it is shown in Fig. 3(b), the recognition rate of SBDA is almost not affected significantly by the number of iterations, so one can simply set the number of iterations as a small integer.

2) Parameters' Setting: There are four other parameters in SBDA algorithm. The recognition rate versus the variations of the parameters can be observed in Fig. 2. In the experiments, the sparsity parameter cardinality K, i.e., the number of the nonzero elements in the projection, varies in the range of [1, 15] since a larger cardinality cannot achieve higher recognition rates, as it is shown in Fig. 2(b). When using the elastic net, the parameter α is simply set to be 0.01, and thus β can be automatically determined since the elastic net algorithm could provide the optimal solution path of β [31]. The balance parameter μ is selected from $\{0.01, 0.1, \ldots, 0.9, 0.99\}$. The model parameter λ(lambda) is selected from $\{10^{-5}, 10^{-4}, \ldots, 10^{5}\}$. In addition, the optimal sparse subspace dimensions is ranged in [1, 20]. All the parameters are varied in the corresponding ranges to search the best performance. The experimental procedures are the same as in [20], [22], and [36].

Fig. 2(c) shows that when the L_2 norm penalty parameter is set to be zero (i.e., without L_2 norm terms in the model), the performance of SBDA is usually slightly lower than the case with suitable settings. This indicates that combining the L_1 and L_2 norms penalty in SBDA can enhance the performance.

As it is shown in Section III that when $\mu \to 1_{-}$, SBDA approximates to DATER/2DLDA. It can be observed from Fig. 2(d) that SBDA achieves higher recognition rates when μ is far from 0.99, which is approximates to one. This indicates that SBDA will at least perform as good as or even better than DATER/2DLDA in gait recognition. Fig. 3(a) shows the variation of recognition rate versus the value of lambda. It is shown that when lambda takes a smaller value, SBDA always obtains good performance.

C. Feature Extraction and Classification

We compared our algorithm with several other state-of-the-art algorithms in Tables II and III, in which Rank-1 indicates that the correct subject is ranked as the top candidate, while Rank-5 indicates that the correct subject is ranked among the top five candidates, and Average is the recognition accuracy among all the Probe sets (i.e., the ratio of correctly recognized persons to the total number of persons in all the Probe sets). These algorithms include baseline [1] LDA [7] image Euclidean distance (IMED) [37], IMED + LDA, hidden Markov Model [38], LDA + Sync [7], LDA + fusion [7], image-to-class [23], MMFA [22], GTDA [20], CSA + DATER [17], and periodicity feature vector + discriminantive locality alignment [36]. The results for IMED and IMED + LDA are from [20].

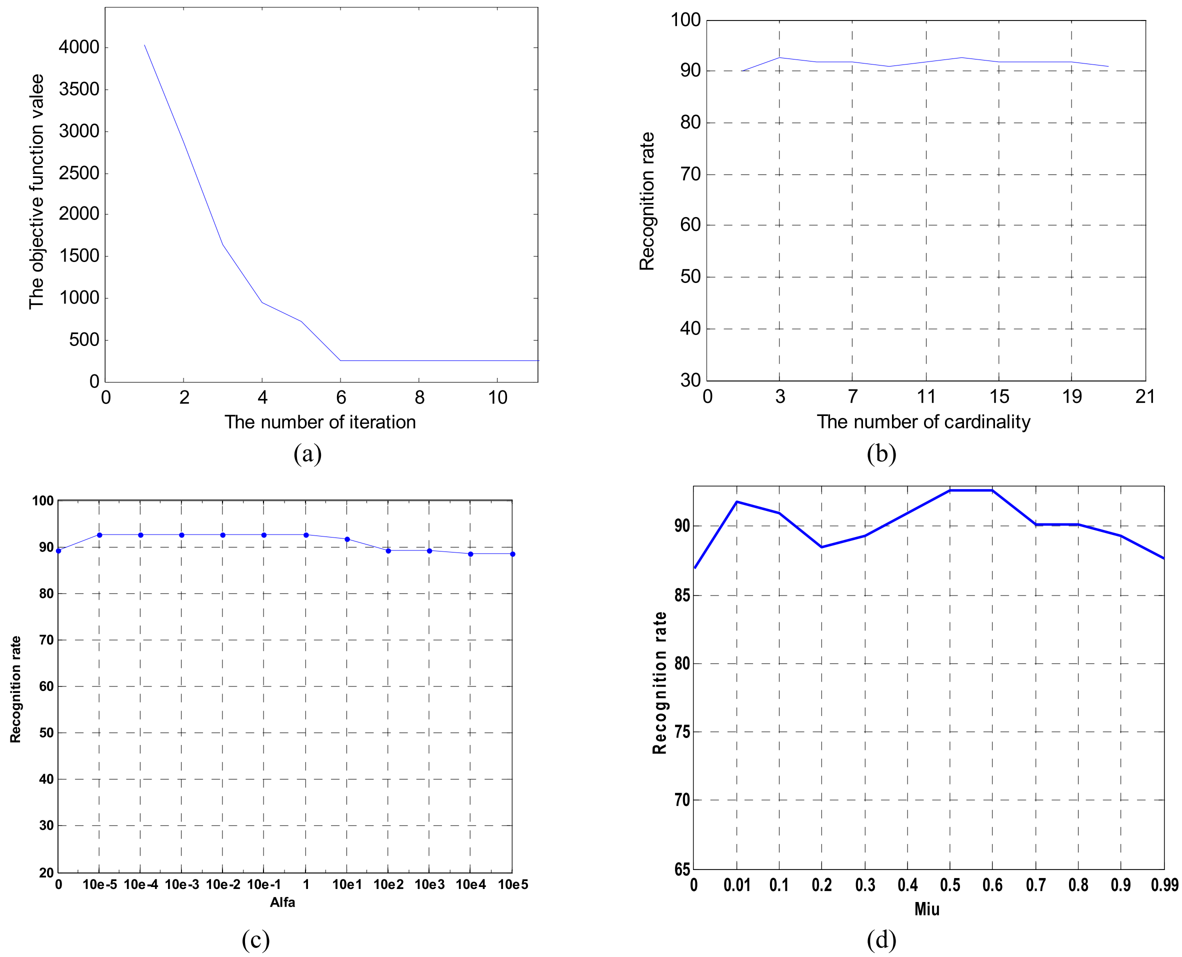

Fig. 2. Some properties of SBDA algorithm for gait recognition. (a) Convergence of SBDA. (b) Recognition rate versus the number of cardinality. (c) Recognition rate versus parameter alpha. (d) Recognition rate versus parameter μ.

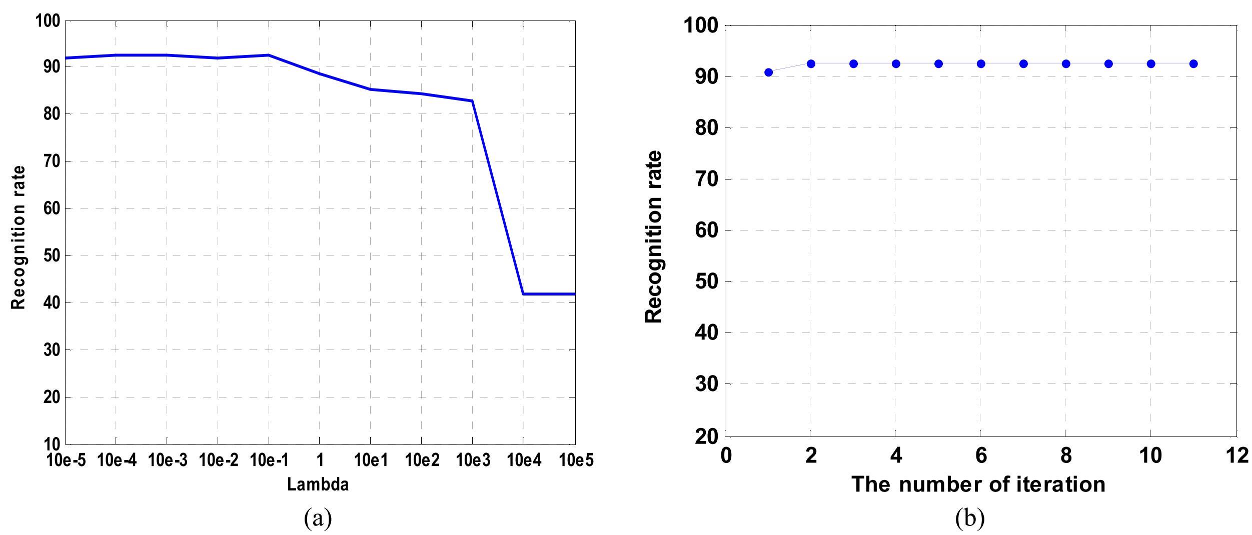

Fig. 3. Some properties of SBDA algorithm for gait recognition. (a) Recognition rate versus the value of lambda. (b) Recognition rate versus the number of iteration.

Considering that the median operation is more robust to noise effects than the traditional minimum operation, we use the same distance measure for Gallery and Probe sequences as in [1], [6], and [17]. The Rank-1 and Rank-5 recognition rates of SBDA and the compared methods are listed in Tables III and IV. Note that the recognition rates of previous methods in

TABLE III
COMPARISON OF RECOGNITION RATE (%) ON THE USF HUMANID GAIT DATABASE (RANK-1)

Probe set	A	B	C	D	E	F	G	H	I	J	K	L	Average
Baseline	73	78	48	32	22	17	17	61	57	36	3	3	40.96
HMM	89	88	68	35	28	15	21	85	80	58	17	15	53.54
IMED	75	83	65	25	28	19	16	58	60	42	2	9	42.87
IMED + LDA	88	86	72	29	33	23	32	54	62	52	8	13	48.64
LDA	87	85	76	31	30	18	21	63	59	54	3	6	48.20
LDA + sync	83	94	61	50	48	22	33	48	52	34	18	12	48.04
LDA + fusion	91	94	81	51	57	25	29	62	60	57	9	12	55.83
DATER	87	93	78	42	42	23	28	80	79	59	18	21	58.17
CSA+DATER	89	93	80	44	45	25	33	80	79	60	18	21	59.58
DLLE/L	90	89	81	40	50	27	26	65	67	57	12	18	51.83
MMFA	89	94	80	44	47	25	33	85	83	60	27	21	59.90
GTDA	91	93	86	32	47	21	32	95	90	68	16	19	60.58
PFV+DLA	94	92	85	46	51	28	34	68	66	62	13	24	60.25
Image-to-class	93	89	81	54	52	32	34	81	78	62	12	9	61.19
SBDA	93	94	80	44	45	25	34	80	83	62	18	21	60.83
SBDA+LDA	93	94	85	51	50	29	36	85	83	68	18	24	61.35

TABLE IV
COMPARISON OF RECOGNITION RATE (%) ON THE USF HUMANID GAIT DATABASE (RANK-5)

Probe set	A	B	C	D	E	F	G	H	I	J	K	L	Average
Baseline	73	78	48	32	22	17	17	61	57	36	3	3	40.96
HMM	-	-	-	-	-	-	-	-	-	-	-	-	-
IMED	91	93	83	52	59	41	38	86	76	76	12	15	65.31
IMED + LDA	95	95	90	52	63	42	47	86	86	78	21	19	68.60
LDA	92	93	89	58	60	36	43	90	81	79	12	12	67.37
LDA + sync	92	96	91	68	69	50	55	80	78	69	39	30	70.85
LDA + fusion	94	96	93	85	79	52	57	89	86	69	39	30	76.18
DATER	96	96	93	69	69	51	52	92	90	83	40	36	72.25
CSA+DATER	96	96	94	74	79	53	57	93	91	83	40	36	74.33
DLLE/L	95	96	93	74	78	50	53	90	90	83	33	27	71.83
MMFA	98	98	94	76	76	57	60	95	93	84	48	39	79.90
GTDA	98	99	97	68	68	50	56	95	99	84	40	40	77.58
PFV+DLA	99	97	94	66	69	52	59	92	91	84	28	35	77.16
Image-to-class	97	98	93	81	74	59	55	94	95	83	30	33	79.33
SBDA	96	96	93	74	69	53	55	92	93	83	40	36	77.15
SBDA+LDA	98	98	94	74	79	57	60	95	95	84	40	40	79.93

Tables III and IV are directly obtained from the corresponding references.

Since the low-dimensional feature matrix $Y_i = P_U^T X_i P_V$ $(i = 1, 2, \ldots, N)$ obtained by SBDA is a $d_1 \times d_2$ matrix, when it is concatenated column by column to be a vector, its dimension may be still high. Motivated by IMED + LDA, CSA + LDA, and the classical PCA + LDA algorithm, we proposed SBDA + LDA for further dimensionality reduction on gait features. Thus, the features' between-class separability and within-class compactness will be further improved in the subspace of LDA, which will increase the recognition rate.

D. Experiments on CASIA Gait Database

The proposed method was tested on the CASIA gait database (data set-B) [39] to evaluate its effectiveness and robustness. The CASIA-B data set contains 124 subjects' gaits sequences collected under different views, clothing and carrying conditions. There are 11 views $(0°, 18°, \ldots, 180°)$ for each subject and ten sequences per view. Among all the ten sequences, two sequences are walking with bag, two on coat, and the rest are normal walking. In this paper, 2750 gait sequences from 25 subjects were used in the experiment. We select the most competitive and correlated methods, i.e., LDA, DATER, CSA + DATER, MMFA, GTDA, and image-to-class, as the compared methods. The experiments were repeated ten times by randomly splitting the database to be the Probe set and the Gallery set. The average recognition rate and the standard deviation of these methods, including the proposed SBDA and SBDA + LDA, are listed in Table V. It can be found that the proposed methods also perform better than the compared subspace learning methods under different views, clothing and carrying conditions.

E. Discussion

Based on the experimental results shown in the above sections, the following observations and conclusions are obtained.

1) It can be found from the experiments that SBDA performs equivalent to or even better than the state-of-the-art subspace learning algorithms on average recognition rates. In most of the cases, SBDA performs better than the previous subspace learning methods, such as DATER and GTDA. It can be found that SBDA has the same or better recognition rates when it is compared

TABLE V
COMPARISON OF AVERAGE RECOGNITION RATE (%) AND STANDARD DEVIATION ON THE CASIA GAIT DATABASE

Methods	LDA	DATER	CSA+DATER	MMFA	GTDA	Image-to-class	SBDA	SBDA+LDA
Rank-1	70.60	59.60	68.20	71.00	71.80	73.60	74.00	78.20
	±6.48	±7.42	±6.24	±6.40	±5.36	±5.04	±4.47	±4.45
Rank-5	86.00	84.80	86.40	88.40	86.80	90.40	91.20	93.20
	±5.38	±5.67	±5.20	±4.20	±4.80	±2.80	±2.49	±1.20

with DATER, which supports the theoretical analysis presented in Section IV that the subspaces of DATER can be approximated by those of SBDA. The better performance of SBDA indicates that combining the L_1 and L_2 norms for sparse subspace learning with matrix representation can obtain more discriminative information than the classical linear methods or matrix representation-based nonsparse methods.

2) SBDA + LDA can obtain better performance than SBDA in gait recognition. The reason is that using the L_1 norm for sparse discriminant learning with matrix representation can avoid the overfitting, and thus it can be more robust or obtain more generalization abilities than the LDA-based methods. With the robust features from SBDA, LDA can further enhance the recognition rates. Similar results can also be found in IMED + LDA and CSA + DATER.
3) Although MMFA, DATER, GTDA, and SBDA all use the very similar discriminant information, adding the L_1 and L_2 norms terms for sparse learning in SBDA, the performance can be enhanced. This proves that introducing the sparsity constraint to learn the discriminant subspace is a tractable method for improving the performance on gait recognition. Although image-to-class method obtains the very similar performance with SBDA, once the discriminant projections of SBDA are obtained from the training steps, it will be more convenient and efficient for SBDA to perform feature extraction and classification than image-to-class method, which has to (frequently) solve the optimal binary flow problem for each Probe gait.
4) Among the most competitive three subspace learning methods, i.e. MMFA, GTDA, and SBDA, no one is the complete winner in all the cases. MMFA obtains the best performance on Probe set K, the main reason may be that the local structure plays an important role when there are time difference between Gallery and Probe sets since GTDA and SBDA do not introduce the locality in the learning step. However, in the Probe set C, GTDA performs best, which indicates that direct feature extraction on the high-order tensor data instead of average silhouettes can also obtain high recognition rate in some special case. However, the disadvantage of GTDA is that its computational complexity is higher than the ones of the average silhouette matrix-based MMFA and SBDA. Usually, SBDA can perform as well as (or even better than) the better results of GTDA and MMFA in most of the cases, which shows the advantages originated from sparse feature selection on the average silhouette matrices. However, the disadvantage of SBDA and GTDA is that they do not make use of the local geometric structure of the data set. Thus, SBDA and GTDA cannot perform better than MMFA in some special cases when the local geometric structure plays an important role in feature extraction and recognition.

V. CONCLUSION

A matrix-representation-based sparse subspace learning method called SBDA was proposed in this paper for gait recognition. The L_1 and L_2 norms were integrated to the discriminant criterion, and thus a novel sparse learning model was obtained. The optimal solutions of this model could be computed by the iterative algorithm using the elastic net regression. Theoretical analyses were presented to explore the properties of SBDA and the relationships among SBDA and previous algorithms were also presented. The experimental results on the USF HumanID and CASIA gait databases show that SBDA performs better than the state-of-the-art subspace learning algorithms for gait recognition. In addition, with the strategy of SBDA + LDA, the gait recognition rates can be further enhanced. The experimental results also show that SBDA is more robust than the compared methods in gait variations and enforcing the sparsity on the projections can obtain more generalization capability and robustness than the previous subspace methods for gait recognition. This indicates that introducing the sparsity in the projection vectors/matrices can enhance the performance of the discriminative method. In the future, we will explore different discriminative criterions in the sparse manner for gait recognition.

APPENDIX
PROOF OF THEOREM 1

From (17), we have

$$\left\| H_U^T V - H_U^T P_V \right\|_F^2 + \frac{\alpha}{1-\mu} \left\| P_V \right\|_F^2 + \sum_r \frac{\beta_r}{1-\mu} |p_{\text{vr}}|$$
$$= \operatorname{tr}\left(V^T S_D^U V - 2V^T S_D^U P_V + P_V^T S_D^U P_V + \frac{\alpha}{1-\mu} P_V^T P_V \right)$$
$$+ \sum\nolimits_r \frac{\beta_r}{1-\mu} |p_{\text{vr}}| .$$

Since $\beta_r \to 0_+$, then $\sum_r \beta_r/(1-\mu)\,|p_{\text{vr}}| \to 0$ and it vanishes. If we take the derivative of above equation with respect to P_V to be zero, the following equation can be obtained:

$$P_V = \left(S_D^U + \frac{\alpha}{1-\mu} I \right)^{-1} S_D^U V = V \ (\alpha \to 0_+).$$

■

Proof of Theorem 3

From (27), if $\mu \to 1_-$, $(1-\mu)S_W P_V \to 0$ and $\mu S_W \to S_W$, then $(\mu S_W - \lambda S_B)V = 0 \Rightarrow S_W V = \lambda S_B V$. This implies that V contains the eigenvetors of the generalized eigenequation of 2-D LDA.

For this given V, similar to the proof of Theorem 1, (31) gives

$$P_V = \left(S_W + \frac{\alpha}{1-\mu} I\right)^{-1} S_W V = V \ , \left(\lim_{\mu \to 1_-} \frac{\alpha}{1-\mu} = 0\right).$$

That is, $P_V \to V$ when these conditions are satisfied. ■

Proof of Theorem 4

We need to prove that the objective function of SBDA is nonincreasing and has a lower bound (at least bigger than a constant $c > 0$). The original objective function of SBDA in each iteration step can be rewritten as follows:

$$\begin{aligned} &J\big(U^{(t-1)}, V^{(t-1)}, P_U^{(t-1)}, P_V^{(t-1)}\big) \\ &\quad = \mu J_W\big(U^{(t-1)}, V^{(t-1)}\big) \\ &\qquad + (1-\mu) J_D\big(U^{(t-1)}, V^{(t-1)}, P_U^{(t-1)}, P_V^{(t-1)}\big) \\ &\qquad + \alpha\big(\big\| P_U^{(t-1)}\big\|_F^2 + \big\| P_V^{(t-1)}\big\|_F^2\big) + \sum_r \beta_r \big|p_{\mathrm{vr}}^{(t-1)}\big| \\ &\qquad + \sum_l \gamma_l \big|p_{ul}^{(t-1)}\big|. \end{aligned}$$

From the inner loop of the iteration procedures, we know that the objective function achieves a local minimum when $U^{(t-1)}$ and $P_U^{(t-1)}$ are given. Therefore, we have

$$\begin{aligned} J&\left(U^{(t-1)}, V^{(t-1)}, P_U^{(t-1)}, P_V^{(t-1)}\right) \\ &\geq J\left(U^{(t-1)}, V^{(t)}, P_U^{(t-1)}, P_V^{(t-1)}\right) \\ &\geq J\left(U^{(t-1)}, V^{(t)}, P_U^{(t-1)}, P_V^{(t)}\right). \end{aligned} \tag{A1}$$

On the other hand, when $V^{(t)}$ and $P_V^{(t)}$ are given and fixed, the value of the objective function can be further reduced from the iteration, that is

$$\begin{aligned} J\left(U^{(t-1)}, V^{(t)}, P_U^{(t-1)}, P_V^{(t)}\right) &\geq J\left(U^{(t)}, V^{(t)}, P_U^{(t-1)}, P_V^{(t)}\right) \\ &\geq J\left(U^{(t)}, V^{(t)}, P_U^{(t)}, P_V^{(t)}\right). \end{aligned} \tag{A2}$$

From (A1) and (A2), we have

$$\begin{aligned} J\left(U^{(t-1)}, V^{(t-1)}, P_U^{(t-1)}, P_V^{(t-1)}\right) &\geq J\left(U^{(t)}, V^{(t)}, P_U^{(t)}, P_V^{(t)}\right) \\ &\geq c \geq 0. \end{aligned}$$

Therefore, the objective function will converge to a local optimum. ■

References

[1] S. Sarkar, P. J. Phillips, Z. Liu, I. R. Vega, P. Grother, and K. W. Bowyer, "The humanID gait challenge problem: Data sets, performance, and analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 2, pp. 162–77, Feb. 2005.

[2] L. Wang, T. Tan, S. Member, H. Ning, and W. Hu, "Silhouette analysis-based gait recognition for human identification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 12, pp. 1505–1518, Dec. 2003.

[3] A. Bobick and A. Johnson, "Gait recognition using static activity-specific parameters," in *Proc. IEEE Conf. CVPR*, Dec. 2001, pp. 423–430.

[4] D. Cunado, M. S. Nixon, and J. N. Carter, "Automatic extraction and description of human gait models for recognition purposes," *Comput. Vis. Image Understand.*, vol. 90, no. 1, pp. 1–41, Apr. 2003.

[5] R. Tanawongsuwan and A. Bobick, "Gait recognition from time-normalized joint-angle trajectories in the walking plane," in *Proc. IEEE Conf. CVPR*, vol. 2. Jan. 2001, pp. 726–731.

[6] Z. Liu and S. Sarkar, "Simplest representation yet for gait recognition: Averaged silhouette," in *Proc. 17th ICPR*, 2004, pp. 211–214.

[7] J. Han and B. Bhanu, "Statistical feature fusion for gait-based human recognition," in *Proc. IEEE Conf. CVPR*, vol. 2. Jul. 2004, pp. 842–847.

[8] L. Sirovich and M. Kirby, "Low-dimensional procedure for characterization of human faces," *J. Opt. Soc. Amer. A, Opt., Image Sci. Vis.*, vol. 4, no. 3, pp. 519–524, Mar. 1987.

[9] M. Kirby and L. Sirovich, "Application of the Karhunen–Loeve procedure for the characterization of human faces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 1, pp. 103–108, Jan. 1990.

[10] M. Turk, "Eigenfaces for recognition," *J. Cognit. Neurosci.*, vol. 3, no. 1, pp. 71–86, Jan. 1991.

[11] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriengman, "Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 711–720, Jul. 1997.

[12] C. Fraley and A. E. Raftery, "Model-based clustering, discriminant analysis and density estimation," *J. Amer. Statist. Assoc.*, vol. 97, no. 458, pp. 611–631, Jun. 2002.

[13] Q. Liu, H. Lu, and S. Ma, "Improving kernel Fisher discriminant analysis for face recognition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 1, pp. 42–49, Jan. 2004.

[14] H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "MPCA: Multilinear principal component analysis of tensor objects," *IEEE Trans. Neural Netw.*, vol. 19, no. 1, pp. 18–39, Jan. 2008.

[15] C. BenAbdelkader and R. Cutler, "Motion-based recognition of people in EigenGait space," in *Proc. 5th IEEE Int. Conf. Autom. Face Gesture Recognit.*, May 2002, pp. 254–259.

[16] J. Han and B. Bhanu, "Individual recognition using gait energy image," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 2, pp. 316–322, Feb. 2006.

[17] D. Xu, S. Yan, D. Tao, L. Zhang, X. Li, and H. Zhang, "Human gait recognition with matrix representation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 7, pp. 896–903, Jul. 2006.

[18] D. Xu, S. Yan, L. Zhang, S. Lin, H. Zhang, and T. S. Huang, "Reconstruction and recogntition of tensor-based objects with concurrent subspaces analysis," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 1, pp. 36–47, Jan. 2008.

[19] S. Yan, D. Xu, Q. Yang, L. Zhang, X. Tang, and H. Zhang, "Discriminant analysis with tensor representation," in *Proc. IEEE Conf. CVPR*, Jun. 2005, pp. 526–532.

[20] D. Tao, X. Li, X. Wu, and S. J. Maybank, "General tensor discriminant analysis and Gabor features for gait recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 10, pp. 1700–1715, Oct. 2007.

[21] X. Li, S. Lin, S. Yan, and D. Xu, "Discriminant locally linear embedding with high-order tensor data," *IEEE Trans. Syst., Man, Cybern., B, Cybern.*, vol. 38, no. 2, pp. 342–352, Apr. 2008.

[22] D. Xu, S. Yan, D. Tao, S. Lin, and H.-J. Zhang, "Marginal Fisher analysis and its variants for human gait recognition and content-based image retrieval," *IEEE Trans. Image Process.*, vol. 16, no. 11, pp. 2811–2821, Nov. 2007.

[23] I. Distance, Y. Huang, D. Xu, and T. Cham, "Face and human gait recognition using image-to-class distance," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 3, pp. 431–438, Mar. 2010.

[24] X. Wu, D. Xu, L. Duan, J. Luo, and Y. Jia, "Action recognition using multilevel features and latent structural SVM," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 8, pp. 1422–1431, Aug. 2013.

[25] A. J. Ma, P. C. Yuen, W. W. W. Zou, and J. Lai, "Supervised spatio-temporal neighborhood topology learning for action recognition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 8, pp. 1447–1460, Aug. 2013.

[26] H. Zou, T. Hastie, and R. Tibshirani, "Sparse principal component analysis," *J. Comput. Graph. Statist.*, vol. 15, no. 2, pp. 265–286, Jun. 2006.

[27] H. Chun and S. Keleş, "Sparse partial least squares regression for simultaneous dimension reduction and variable selection," *J. R. Statist. Soc. Ser. B, Statist. Methodol.*, vol. 72, no. 1, pp. 3–25, Jan. 2010.

[28] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, Feb. 2009.

[29] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. R. Statist. Soc., Ser. B, Statist. Methodol.*, vol. 58, no. 1, pp. 267–288, 1996.

[30] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least angle regression," *Ann. Statist.*, vol. 32, no. 2, pp. 407–499, 2004.

[31] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *J. R. Statist. Soc., Ser. B, Statist. Methodol.*, vol. 67, no. 2, pp. 301–320, 2005.

[32] L. Clemmensen, T. Hastie, D. Witten, and B. Ersboll, "Sparse discriminant analysis," *Technometrics*, vol. 53, no. 4, pp. 406–413, 2011.

[33] Z. Qiao, L. Zhou, and J. Z. Huang, "Sparse linear discriminant analysis with applications to high dimensional low sample size data," *IAENG Int. J. Appl. Math.*, vol. 39, no. 1, pp. 48–60, 2009.

[34] Z. Lai, Y. Xu, J. Yang, J. Tang, and D. Zhang, "Sparse tensor discriminant analysis," *IEEE Trans. Image Process.*, vol. 22, no. 10, pp. 3904–3915, May 2013.

[35] J. Yang, D. Zhang, X. Yong, and J. Yang, "Two-dimensional discriminant transform for face recognition," *Pattern Recognit.*, vol. 38, no. 7, pp. 1125–1129, 2005.

[36] R. Hu, W. Shen, and H. Wang, "Recursive spatiotemporal subspace learning for gait recognition," *Neurocomputing*, vol. 73, nos. 10–12, pp. 1892–1899, Jun. 2010.

[37] L. Wang, Y. Zhang, and J. Feng, "On the Euclidean distance of images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 8, pp. 1334–1339, Aug. 2005.

[38] A. Kale, A. Sundaresan, A. N. Rajagopalan, N. P. Cuntoor, A. K. R. Chowdhury, V. Kruger, *et al.*, "Identification of humans using gait," *IEEE Trans. Image Process.*, vol. 13, no. 9, pp. 1163–1173, Sep. 2004.

[39] S. Yu, D. Tan, and T. Tan, "A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition," in *Proc. 18th ICPR*, Aug. 2006, pp. 441–444.

Zhihui Lai received the B.S degree in mathematics from South China Normal University, Guangzhou, China; the M.S degree from Jinan University, Guangzhou; and the Ph.D. degree in pattern recognition and intelligence system from Nanjing University of Science and Technology, Nanjing, China, in 2002, 2007, and 2011, respectively.

He was a Research Associate and a Post-Doctoral Fellow with The Hong Kong Polytechnic University, Hong Kong, from 2010 to 2014. He is currently a Post-Doctoral Fellow with the Bio-Computing Research Center, Shenzhen Graduate School, Harbin Institute of Technology, Harbin. His research interests include face recognition, image processing and content-based image retrieval, pattern recognition, compressive sense, human vision modelization, and applications in the fields of intelligent robot research.

Yong Xu (M'06) received the B.S. and M.S. degrees from Air Force Institute of Meteorology, Xi'an, China, and the Ph.D. degree in pattern recognition and intelligence system from Nanjing University of Science and Technology, Nanjing, China, in 1994, 1997, and 2005, respectively.

He was a Post-Doctoral Research Fellow with Shenzhen Graduate School, Harbin Institute of Technology, Harbin, China, from 2005 to 2007, where he is currently a Professor. He was a Research Assistant Researcher with The Hong Kong Polytechnic University, Hong Kong, from 2007 to 2008. He has published more than 40 scientific papers. His research interests include pattern recognition, biometrics, and machine learning.

Zhong Jin received the B.S. degree in mathematics, the M.S. degree in applied mathematics, and the Ph.D. degree in pattern recognition and intelligence system, all from Nanjing University of Science and Technology (NUST), Nanjing, China, in 1982, 1984, and 1999, respectively.

He was a Research Assistant with the Department of Computer Science and Engineering, the Chinese University of Hong Kong, Hong Kong, from 2000 to 2001. He was with the Laboratoire HEUDIASYC, Universite de Technologie de Compiegne, Compiègne, France, from 2001 to 2002. He was the Ramon y Cajal Program Research Fellow with the Centre de Visio per Computador, Universitat Autonoma de Barcelona, Bellaterra, Spain, from 2005 to 2005. He is currently a Professor with the Department of Computer Science, NUST. His research interests include pattern recognition, computer vision, face recognition, facial expression analysis, and content-based image retrieval.

David Zhang (F'06) received the degree in computer science from Peking University, Beijing, China; the M.Sc. degree in computer science and a Ph.D. degree from Harbin Institute of Technology (HIT), Harbin, China; and a second Ph.D. degree in electrical and computer engineering from University of Waterloo, Waterloo, ON, Canada, in 1982, 1986, and 1994, respectively.

He was a Post-Doctoral Fellow with Tsinghua University, Beijing, from 1986 to 1988, and an Associate Professor with the Academia Sinica, Beijing. He is currently a Head with the Department of Computing, and a Chair Professor with The Hong Kong Polytechnic University, Hong Kong. He is the Visiting Chair Professor with Tsinghua University and an Adjunct Professor with Peking University, Beijing; Shanghai Jiao Tong University, Shanghai; HIT; and University of Waterloo. He is the Founder and Editor-in-Chief of *International Journal of Image and Graphics*, the Book Editor of *Springer International Series on Biometrics*, the Organizer of the International Conference on Biometrics Authentication, the Associate Editor of more than ten international journals including IEEE TRANSACTIONS ON PATTERN RECOGNITION, and the author of more than 10 books and 200 journal papers.

Prof. Zhang is a Croucher Senior Research Fellow, Distinguished Speaker of the IEEE Computer Society, and a Fellow of IAPR.