Signal Processing 106 (2015) 245–252

Contents lists available at ScienceDirect

Signal Processing

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

A sparse and discriminative tensor to vector projection for human gait feature representation

Lefei Zhang [a], Liangpei Zhang [b], Dacheng Tao [c], Bo Du [a,*]

[a] Computer School, Wuhan University, Wuhan 430072, China
[b] State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan 430079, China
[c] Centre for Quantum Computation and Intelligent Systems, Faculty of Engineering and Information Technology, University of Technology, Sydney, NSW 2007, Australia

ARTICLE INFO

Article history:
Received 4 May 2014
Received in revised form
21 July 2014
Accepted 2 August 2014
Available online 13 August 2014

Keywords:
Tensor representation
Multi-linear principal component analysis
Discriminative locality preserving projection
Feature extraction
Gait recognition

ABSTRACT

In this paper, we introduce an efficient tensor to vector projection algorithm for human gait feature representation and recognition. The proposed approach is based on the multi-dimensional or tensor signal processing technology, which finds a low-dimensional tensor subspace of original input gait sequence tensors while most of the data variation has been well captured. In order to further enhance the class separability and avoid the potential overfitting, we adopt a discriminative locality preserving projection with sparse regularization to transform the refined tensor data to the final vector feature representation for subsequent recognition. Numerous experiments are carried out to evaluate the effectiveness of the proposed sparse and discriminative tensor to vector projection algorithm, and the proposed method achieves good performance for human gait recognition using the sequences from the University of South Florida (USF) HumanID Database.
© 2014 Elsevier B.V. All rights reserved.

1. Introduction

Human recognition by analysis of the biometric resources such as face, gait, iris, and palmprint has been thoroughly studied and employed in many applications [1,2]. Among them, gait recognition, i.e., identify a person by the manner of walking patterns extracted from the video, has been reported that could well identify a human subject in low-resolution images that are taken at a distance, has recently become a popular research problem in signal processing community and a number of gait recognition algorithms have been proposed in the recent few years [3–5].

The main challenge of the gait recognition lies in that its performance is affected by many factors, including the environmental factors as well as the physical characteristics of the human subject, such as the camera factors, the observed sequence and preprocessed silhouette quality, the human walking speed, elapsed time, carrying condition, and shoes style. Furthermore, these aforementioned factors are sometimes correlated [2,6,7]. The open source humanID dataset [8] investigated several of these factors and provided the baseline algorithm for human gait recognition, however, how to measure the effect of these factors or to combine these factors is still an open question to date.

Most of the existing approaches for gait recognition could be broadly categorized into two classes: the model based and the silhouette based approaches. The model based methods usually characterise a human subject by some specific models, e.g., the appearance-based models [9,10], the stochastic statistical models [11], the articulated biomechanics models [12,13], and some other parameter-based models such as the method proposed in [14]. In contrast, the silhouette based gait recognition algorithms do not assume certain model of the human subject, but analyze the spatio-temporal shape and motion features of

* Corresponding author.
E-mail address: gunspace@163.com (B. Du).

http://dx.doi.org/10.1016/j.sigpro.2014.08.005
0165-1684/© 2014 Elsevier B.V. All rights reserved.

the silhouette images which are extracted from the sequences [2,15–17]. In this paper, we suppose that the standard silhouette images are available and focus on the silhouette based approaches for human gait feature representation and recognition.

In the computer vision and machine learning research, the effective feature representation of the studied subject is a key issue. Since most of the data dimension reduction algorithms (e.g., the principal components analysis, PCA [18], the linear discriminant analysis, LDA [19], the Sparse PCA [20], and some recently published methods [21,22]) and classification algorithms (e.g., the Support Vector Machines, SVM [23]) in the literature only allow the vector feature as inputs, for the gait recognition task, it is straightforward to resize the silhouette images to the long vectors as the input feature representation. However, the feature dimensionality is usually much larger than the number of gallery sets, which causes an issue that is known as the "small sample-size (SSS) problem" [24,25], and thus results in poor recognition performance [26,27]. The other key shortcoming of the vector feature representation lies in that this scheme has lost many of the original spatial constraints of each pixel in the silhouette images, which hinders the subsequent algorithm to construct the optimal dimension reduction and classification model with only limited training samples.

To address the aforementioned difficulties, some researchers suggest to use matrix and tensor representation instead of vector representation [28,29], as reported in the works of two-dimensional PCA [30,31], two-dimensional LDA [15], tensor subspace analysis [32], multi-linear PCA [16,33], and so on. Especially in the framework of multi-linear PCA, the authors introduce a multi-linear principal component projection that captures most of the original tensorial input variation. However, the standard multi-linear PCA algorithm ignores the discriminative information provided by the gallery set, and its output data is still in a tensor format, which could not be directly processed by a conventional classifier. In this paper, we introduce an efficient tensor to vector projection algorithm for human gait feature representation and recognition. The main contributions of this paper are summarized as follows:

1. We suggest to represent the observed gait sequences as 3-order-tensors by which the data structure of both the spatial and temporal domains is well preserved. Based on the tensor representation, we further introduce the multi-linear PCA algorithm, which performs feature extraction by finding a low-dimensional tensor subspace that captures most of the data variation of original input gait sequence tensors.
2. Followed by the multi-linear PCA algorithm, we adopt a discriminative locality preserving projection with sparse regularization to transform the refined tensor data to the final vector feature representation, by which the class separability is improved and the potential model overfitting is simultaneously avoided.

The output feature representation by our proposed sparse and discriminative tensor to vector projection (SDTTV) could be simply employed for subsequent classification, and numerous experiments indicate that the proposed algorithm achieves good performance for human gait recognition. The remainder of this paper is organized as follows. In the following section, we give a brief description of related tensor algebra, and then present the proposed SDTTV algorithm in detail. After that, the experiments are reported in Section 3, followed by the conclusion.

2. The proposed sparse and discriminative tensor to vector projection algorithm

In this section, we first review some multi-dimensional or tensor signal processing rules that are related to the proposed SDTTV algorithm, then we introduce the multi-linear PCA approach, which aims to find a low-dimensional tensor subspace of original input tensors while most of the data variation could be well preserved. Finally, we provide the optimization of SDTTV algorithm by combining the discriminative locality preserving projection and a sparse regularization to transform the refined tensor data to the final vector feature representation for subsequent gait recognition.

2.1. Related tensor algebra

The notationsused in this paper are followed by convention in the tensor papers, e.g., vectors are denoted by lowercase boldface and italic letters, such as $\boldsymbol{x}$, matrices by uppercase boldface and italic, such as $\boldsymbol{W}$, and tensors by calligraphic letters, such as $\mathcal{A}$. For a M-order-tensor $\mathcal{A} \in \mathbb{R}^{K_1 \times K_2 \times \cdots \times K_M}$, in which K_i suggests the size of this tensor in each mode, and the elements of $\mathcal{A}$ are denoted with indices in lowercase letters, i.e., $\mathcal{A}_{k_1,k_2,\ldots,k_M}$, in which each k_i addresses the i-mode of $\mathcal{A}$, and $1 \le k_i \le K_i$, $i \in (1,2,\ldots,M)$. Unfolding tensor $\mathcal{A}$ along the i-mode is defined by keeping the index k_i fixed and varying the other indices, the result of which is denoted as a matrix $\boldsymbol{A}_{(i)} \in \mathbb{R}^{K_i \times \prod_{j \ne i} K_j}$. The i-mode product of a tensor $\mathcal{A}$ by a matrix $\boldsymbol{W} \in \mathbb{R}^{J_i \times K_i}$ is a tensor with entries $(\mathcal{A} \times_i \boldsymbol{W})_{k_1,\ldots,k_{i-1},j_i,k_{i+1},\ldots,k_M} = \sum_{k_i} \mathcal{A}_{k_1,\ldots,k_M} \boldsymbol{W}_{j_i,k_i}$. The Frobenius norm of a tensor $\mathcal{A}$ is given by $\|\mathcal{A}\| = \sqrt{\sum_{k_1} \cdots \sum_{k_M} \mathcal{A}^2_{k_1,k_2,\ldots,k_M}}$, and the Euclidean distance between two tensors $\mathcal{A}$ and $\mathcal{B}$ could be measured by $\|\mathcal{A}-\mathcal{B}\|$. For more detailed information, refer to [34,35]. As a summary, Table 1 lists the fundamental symbols defined in tensor algebra related to this paper.

2.2. Multi-linear principal components analysis

Suppose we have N gait tensors in the gallery set, i.e., $\{\mathcal{A}_n \in \mathbb{R}^{K_1 \times K_2 \times K_3}, n = 1,2,\ldots,N\}$, in which K_1 and K_2 are the height and width of each grey-level gait frame image from a subject sequence, respectively, and K_3 is the number of frames in a sequence. In order to achieve the objective of finding a low-dimensional tensor subspace of original input tensors in which most of the data variation could be well preserved, the multi-linear PCA algorithm suggests to adopt a series of

Table 1
List of notations in tensor algebra.

$\mathcal{A}_n$	The nth gait tensor
$\boldsymbol{A}_{n(i)}$	The i-order unfolded matrix of $\mathcal{A}_n$ $(i=1,2,3)$
K_1, K_2	The height and width of each gait image
K_3	The number of frames in a sequence
N	The number of gait tensors
$\mathcal{B}_n$	The nth principal component gait tensor
J_1, J_2, J_3	The size of principal component gait tensor
$\boldsymbol{W}_i$	The projection matrix in ith mode, $(i=1,2,3)$
$\overline{\mathcal{A}}$	Mean tensor of a gait tensor set $\{\mathcal{A}_n, n=1,2,\dots,N\}$
$\psi_{\mathcal{A}}$	Total scatter of a gait tensor set $\{\mathcal{A}_n, n=1,2,\dots,N\}$

projection matrices, i.e., $\boldsymbol{W}_i \in \mathbb{R}^{J_i \times K_i}$ $(i=1,2,3)$, thus the principal component gait tensor $\mathcal{B} \in \mathbb{R}^{J_1 \times J_2 \times J_3}$ $(J_i \le K_i, i=1,2,3)$ of an arbitrary gait tensor $\mathcal{A} \in \mathbb{R}^{K_1 \times K_2 \times K_3}$ is acquired by the following multi-linear projection:

$$\mathcal{B} = \mathcal{A} \times_1 \boldsymbol{W}_1 \times_2 \boldsymbol{W}_2 \times_3 \boldsymbol{W}_3 \tag{1}$$

Note that in multi-linear PCA, the projection matrices $\boldsymbol{W}_i \in \mathbb{R}^{J_i \times K_i}$ $(i=1,2,3)$ are learned by the gallery set $\{\mathcal{A}_n, n=1,2,\dots,N\}$, while Eq. (1) is employed for multi-linear dimension reduction of the gallery as well as the probe sets.

As mentioned above, the optimization of multi-linear PCA algorithm is to capture the data variation of the original gait tensors. In order to reach this objective, similar to PCA, the mean tensor of the set $\{\mathcal{A}_n, n=1,2,\dots,N\}$ is defined as

$$\overline{\mathcal{A}} = \sum_{n=1}^{N} \mathcal{A}_n \tag{2}$$

and the total scatter of these tensors is defined as

$$\psi_{\mathcal{A}} = \sum_{n=1}^{N} \|\mathcal{A}_n - \overline{\mathcal{A}}\|^2 \tag{3}$$

Then, the optimal multi-linear projection matrices in (1) could be acquired by the following optimization:

$$\begin{aligned} &\arg \max_{\boldsymbol{W}_i\ (i=1,2,3)} \psi_{\mathcal{B}} \\ &= \arg \max_{\boldsymbol{W}_i\ (i=1,2,3)} \sum_{n=1}^{N} \|\mathcal{A}_n \times_1 \boldsymbol{W}_1 \times_2 \boldsymbol{W}_2 \times_3 \boldsymbol{W}_3 \\ &\quad - \overline{\mathcal{A}} \times_1 \boldsymbol{W}_1 \times_2 \boldsymbol{W}_2 \times_3 \boldsymbol{W}_3\|^2 \end{aligned} \tag{4}$$

Since the objective function of multi-linear PCA (4) is a multi-parameter optimization problem, there is no known optimal solution which allows for the simultaneous optimization of all the projection matrices. Note that the projection to an ith order tensor space consists of i projections to i vector subspaces, thus i optimization sub-problems can be solved by finding the $\boldsymbol{W}_i$ that maximizes the scatter in the i-mode vector subspace [16]. That is to say, if $\{\boldsymbol{W}_j, j=1,\dots,i-1,i+1,\dots,M\}$ is given, then the only left projection matrix $\boldsymbol{W}_i$ could be globally optimized by combining the top J_i eigenvectors which is corresponding to the largest J_i eigenvalues of the covariance matrix:

$$\boldsymbol{C}^i = \sum_{n=1}^{N} (\boldsymbol{A}_{n(i)} - \overline{\boldsymbol{A}}_{(i)})(\boldsymbol{A}_{n(i)} - \overline{\boldsymbol{A}}_{(i)})^{\mathrm{T}} \tag{5}$$

where $\boldsymbol{A}_{n(i)}$ is the i-order unfolded matrix of $\mathcal{A}_n^{(i)}$, and $\mathcal{A}_n^{(i)}$ is defined as

$$\mathcal{A}_n^{(i)} = \mathcal{A}_n \times_1 \boldsymbol{W}_1 \cdots \times_{i-1} \boldsymbol{W}_{i-1} \times_{i+1} \boldsymbol{W}_{i+1} \cdots \times_N \boldsymbol{W}_N \tag{6}$$

Therefore, if we set a suitable initial value of the projection matrices, the objective function (4) could be locally optimized. Furthermore, the convergence of such iterative approach is guaranteed as reported in some tensor papers [6,36].

2.3. SDTTV algorithm

The multi-linear PCA algorithm reveals a feature mapping from the original high-dimensional tensor space $\mathbb{R}^{K_1} \otimes \mathbb{R}^{K_2} \otimes \cdots \otimes \mathbb{R}^{K_M}$ to the reduced low-dimensional subspace $\mathbb{R}^{J_1} \otimes \mathbb{R}^{J_2} \otimes \cdots \otimes \mathbb{R}^{J_M}$ by which the scatter of the tensor set is preserved as much as possible. However, its output data is still in the tensor format, which could not be directly processed by a conventional classifier. Also, the standard multi-linear PCA algorithm is an unsupervised technique without taking class labels into account. Hence, a tensor to vector feature selection strategy is introduced here to select the elements from the refined tensor data to the plain vector feature representation [37,16].

Consider Q as the number of classes in the gallery set, N_q $(q=1,\dots,Q)$ is the number of subjects in class q (i.e., $\sum_{q=1}^{Q} N_q = N$), $\overline{\mathcal{B}}$ is the mean tensor of the set $\{\mathcal{B}_n, n=1,2,\dots,N\}$, and $\overline{\mathcal{B}}_q$ is the class mean tensor of class q. Then, the discriminability index for each element in $\mathcal{B}$ is given by

$$\delta_{k_1,k_2,\dots,k_M} = \frac{\sum_{q=1}^{Q} N_q (\overline{\mathcal{B}}_{q,k_1,k_2,\dots,k_M} - \overline{\mathcal{B}}_{k_1,k_2,\dots,k_M})^2}{\sum_{n=1}^{N} (\mathcal{B}_{n,k_1,k_2,\dots,k_M} - \overline{\mathcal{B}}_{k_1,k_2,\dots,k_M})^2} \tag{7}$$

For the tensor to vector projection, the elements in an arbitrary tensor $\mathcal{B}_n$ are rearranged into a feature vector $\boldsymbol{x}_n$, ordered according to $\delta_{k_1,k_2,\dots,k_M}$ in the descending order, and only the top most discriminative l components are accepted for subsequent feature representation, e.g., $\boldsymbol{x}_n \in \mathbb{R}^l$.

In order to further promote the discriminability of vector feature representation, in this paper, we propose a discriminative locality preserving projection with sparse regularization to transform the gallery feature set $\{\boldsymbol{x}_n \in \mathbb{R}^l, n=1,2,\dots,N\}$ to a more efficient but discriminative feature representation $\{\boldsymbol{y}_n \in \mathbb{R}^d, n=1,2,\dots,N\}$, where $d \ll l$ in practice. This motivation could be arrived by the linear feature mapping $\boldsymbol{y}_n = \boldsymbol{U}^{\mathrm{T}} \boldsymbol{x}_n$ where $\boldsymbol{U} \in \mathbb{R}^{l \times d}$. Specifically, for each subject $\boldsymbol{x}_n$ in the gallery set, we have the following local optimization of the projection matrix $\boldsymbol{U}$:

$$\arg\min_{\boldsymbol{U}} \sum_{j=1}^{\alpha_1} \|\boldsymbol{U}^{\mathrm{T}}\boldsymbol{x}_n - \boldsymbol{U}^{\mathrm{T}}\boldsymbol{x}_j^+\|^2 - \beta \sum_{j=1}^{\alpha_2} \|\boldsymbol{U}^{\mathrm{T}}\boldsymbol{x}_n - \boldsymbol{U}^{\mathrm{T}}\boldsymbol{x}_j^-\|^2 \tag{8}$$

in which $\boldsymbol{x}_j^+$ $(j=1,2,\dots,\alpha_1)$ is the jth subject from the same class, and $\boldsymbol{x}_j^-$ $(j=1,2,\dots,\alpha_2)$ is the jth subject from different classes, sorted by their Euclidean distance to $\boldsymbol{x}_n$. β is a trade-off parameter. Similar to the strategy in the patch alignment framework [38], the discriminative local patch is defined by

$$\boldsymbol{X}^{(n)} = [\boldsymbol{x}_n, \boldsymbol{x}_1^+, \boldsymbol{x}_2^+, \dots, \boldsymbol{x}_{\alpha_1}^+, \boldsymbol{x}_1^-, \boldsymbol{x}_2^-, \dots, \boldsymbol{x}_{\alpha_2}^-] \tag{9}$$

By defining a coefficient vector $\boldsymbol{u}$ as (10), local optimization (8) can be reduced to (11):

$$\boldsymbol{u}=[1,1,\dots,1,-\beta,-\beta,\dots,-\beta] \tag{10}$$

$$\begin{aligned}\arg\min_{\boldsymbol{U}} \sum_{j=1}^{\alpha_1+\alpha_2} \boldsymbol{u}_j \|\boldsymbol{U}^{\mathrm{T}}\boldsymbol{X}_1^{(n)}-\boldsymbol{U}^{\mathrm{T}}\boldsymbol{X}_{j+1}^{(n)}\|^2 \\ =\arg\min_{\boldsymbol{U}} \operatorname{tr}(\boldsymbol{U}^{\mathrm{T}}\boldsymbol{X}^{(n)}\boldsymbol{L}^{(n)}\boldsymbol{X}^{(n)\mathrm{T}}\boldsymbol{U})\end{aligned} \tag{11}$$

in which $\boldsymbol{L}^{(n)}$ is the local alignment matrix

$$\boldsymbol{L}^{(n)}=\begin{bmatrix}-\boldsymbol{e}_{\alpha_1+\alpha_2}^{\mathrm{T}} \\ \boldsymbol{I}_{\alpha_1+\alpha_2}\end{bmatrix}\operatorname{diag}(\boldsymbol{u})\left[-\boldsymbol{e}_{\alpha_1+\alpha_2}\boldsymbol{I}_{\alpha_1+\alpha_2}\right] \tag{12}$$

and $\boldsymbol{e}_{\alpha_1+\alpha_2}$ and $\boldsymbol{I}_{\alpha_1+\alpha_2}$ are identity vector and identity matrices, respectively. Then, the full optimization is obtained by summing all the local optimizations of $\boldsymbol{x}_n$ (11). Because the columns in patch $\boldsymbol{X}^{(n)}$ are selected from the gallery set $\boldsymbol{X}$, we introduce a selection matrix $\boldsymbol{S}^{(n)}\in \mathrm{R}^{N\times(\alpha_1+\alpha_2+1)}$ to align all the samples in $\boldsymbol{X}^{(n)}$ together into a unified index system in $\boldsymbol{X}$, i.e., $\boldsymbol{X}^{(n)}=\boldsymbol{X}\boldsymbol{S}^{(n)}$ [38,39]. We then sum the local optimizations (11) over all the subjects in the gallery set and obtain the whole optimization of discriminative locality preserving projection:

$$\begin{aligned}\arg\min_{\boldsymbol{U}} \sum_{n=1}^{N} \operatorname{tr}(\boldsymbol{U}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{S}^{(n)}\boldsymbol{L}^{(n)}\boldsymbol{S}^{(n)\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{U}) \\ =\arg\min_{\boldsymbol{U}} \operatorname{tr}(\boldsymbol{U}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{L}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{U})\end{aligned} \tag{13}$$

in which $\boldsymbol{L}=\sum_{n=1}^{N}\boldsymbol{S}^{(n)}\boldsymbol{L}^{(n)}\boldsymbol{S}^{(n)\mathrm{T}}$.

The solution of (13) provides a discriminative subspace in which each basis is a linear combination of all the original features, moreover, $\boldsymbol{U}$ is a dense matrix in which most elements are nonzero. In fact, it is necessary to control the model complexity according to the regularization theory [40,41]. Since the dense projection provided by (13) is often difficult to interpret [42], in this paper, we propose to control the model complexity according to a sparse regularization [43]. Sparsity could not only provide a good interpretation of the feature projection model, but also decrease the variance brought about by possible overfitting [44]. In the SDTTV algorithm, we put forward the l_2-norm of the matrix $\boldsymbol{U}$ to control the number of its nonzero elements, i.e.,

$$\arg\min_{\boldsymbol{U}} \operatorname{tr}(\boldsymbol{U}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{L}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{U})+\gamma\|\boldsymbol{U}\|^2 \tag{14}$$

in which γ is the regularization parameter. Eq. (14) could be further written to (15) derived from the fact that $\|\boldsymbol{U}\|^2=\operatorname{tr}(\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U})$

$$\arg\min_{\boldsymbol{U}} \operatorname{tr}[\boldsymbol{U}^{\mathrm{T}}(\boldsymbol{X}\boldsymbol{L}\boldsymbol{X}^{\mathrm{T}}+\gamma\boldsymbol{I})\boldsymbol{U}] \tag{15}$$

The solution of (15) is obtained by solving the eigenvalue decomposition of the matrix $(\boldsymbol{X}\boldsymbol{L}\boldsymbol{X}^{\mathrm{T}}+\gamma\boldsymbol{I})$ and combining its top d eigenvectors associated with the d smallest eigenvalues.

As a summary of this section, we provide the detailed procedure of SDTTV algorithm as follows.

Algorithm 1. Procedure for the proposed SDTTV algorithm.

Input: N gait tensors in the gallery set, i.e., $\{\mathcal{A}_n\in \mathrm{R}^{K_1\times K_2\times K_3}, n=1,2,\dots,N\}$;
Step 1: Set the initial value of the projection matrices $\boldsymbol{W}_i$, $i\in(1,2,3)$;
Step 2: Repeat
- Optimize $\boldsymbol{W}_1$ by the eigen-decomposition of $\boldsymbol{C}^1$ in (5);
- Optimize $\boldsymbol{W}_2$ by the eigen-decomposition of $\boldsymbol{C}^2$ in (5);
- Optimize $\boldsymbol{W}_3$ by the eigen-decomposition of $\boldsymbol{C}^3$ in (5);

Until Convergence
Step 3: Find the top most discriminative l components $\boldsymbol{x}_n$ from the principal component gait tensor $\mathcal{B}_n$ by (7);
Step 4: Construct the discriminative locality preserving matrix $\boldsymbol{L}$ by (13);
Step 5: Optimize the sparse feature projection matrix $\boldsymbol{U}$ by (15);
Output: Multi-linear projection matrices $\boldsymbol{W}_i$, $i\in(1,2,3)$ and sparse feature projection matrix $\boldsymbol{U}$. For an arbitrary gait tensor $\mathcal{A}$, the final vector feature representation could be obtained by $\mathcal{B}=\mathcal{A}\times_1\boldsymbol{W}_1\times_2\boldsymbol{W}_2\times_3\boldsymbol{W}_3$ and $\boldsymbol{y}_n=\boldsymbol{U}^{\mathrm{T}}\boldsymbol{x}_n$.

3. Experiments and discussions

In this section, the University of South Florida (USF) HumanID gait database [8] is used to demonstrate the superiority of the proposed SDTTV algorithm compared to several baseline approaches. Here we first briefly describe the standard database and some necessary preprocessing steps, then provide the performance of our algorithm on this database along with some analyses and discussions.

3.1. Experimental configurations

The USF HumanID gait database was built for vision-based gait recognition and it is widely used. This database consists of 452 sequences from 74 subjects (people) walking in elliptical paths in front of the camera. For each subject, there are the following covariates: surface type (grass or concrete), shoe type (two different types), and viewpoint (left or right). There is a gallery set for algorithm training and a set of seven predesigned experiments for algorithm comparisons. For the algorithm training, the provided gallery set is collected with the following covariates: grass surface, shoe type A, and right view (GAR). The capturing condition for each probe set is different from the gallery set, and summarized in Fig. 1 and Table 2; note that there is no common sequences between the gallery set and any of the probe sets, also, all the probe sets are distinct. Detailed information about the probe sets is given in Table 2. More detailed information about the USF HumanID is described in [2,8,16].

Since the proposed SDTTV algorithm belongs to the silhouette based approach, extracting of the accurate silhouettes from the original video data becomes a critical preprocessing step. Fortunately, a reference algorithm is proposed in [8] to extract human silhouettes and recognize individuals in this database. In this paper, we treat each half gait cycle as a gait tensor, and the number of gait tensors in the gallery set and the probe sets are also clarified in Table 2. From this table we learn that each subject has an average of roughly 10 gait tensors available. Although each frame of the grey-level gait silhouette extracted by the method in [8] is of standard size 128×88, the number of frames in each sequence should be also unified in order to let the input gait tensors to be normalized to the same dimension in each mode. In this paper, we use the time-mode normalization technology that is introduced in [16], and thus each input gait tensor has a canonical representation of $128\times 88\times 20$. After that,

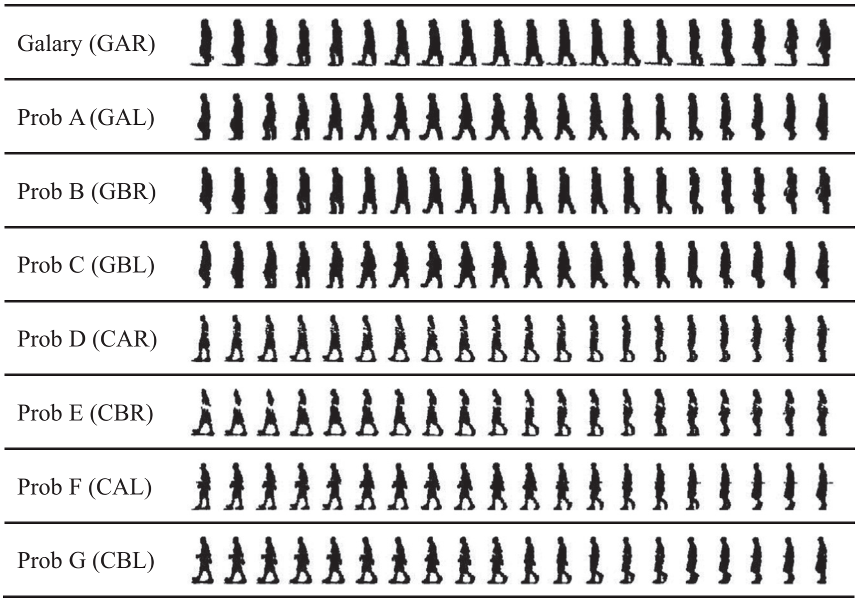

Fig. 1. Illustration of the sequences of the first subject in the gallery gait, probe A gait, probe B gait, probe C gait, probe D gait, probe E gait, probe F gait, and probe G gait, respectively. The C, G, A, B, L, and R stand for cement surface, grass surface, shoe type A, shoe type B, left view, and right view, respectively.

Table 2
Seven probe sets for challenge experiments.

Probe	# of Labels	# of Gait tensors	Difference between the gallery set
Gallery	71	731	\
Probe A	71	727	View
Probe B	41	423	Shoe
Probe C	41	420	View and shoe
Probe D	67	682	Surface
Probe E	43	435	Surface and shoe
Probe F	67	685	Surface and view
Probe G	43	424	Surface, shoe and view

Table 3
Rank-1 recognition rate for human gait recognition.

Probe	Baseline	HMM	LTN	MPCA	SDTTV
A (GAL)	79	99	94	92	95
B (GBR)	66	89	83	85	88
C (GBL)	56	78	78	76	71
D (CAR)	29	35	33	39	41
E (CBR)	24	29	24	29	28
F (CAL)	30	18	17	21	26
G (CBL)	10	24	21	21	24
Average	42	53	50	52	53

Table 4
Rank-5 recognition rate for human gait recognition.

Probe	Baseline	HMM	LTN	MPCA	SDTTV
A (GAL)	96	100	99	96	100
B (GBR)	81	90	85	90	93
C (GBL)	76	90	83	81	93
D (CAR)	61	65	65	55	62
E (CBR)	55	65	67	52	66
F (CAL)	46	60	58	58	62
G (CBL)	33	50	48	50	54
Average	64	74	72	69	76

in order to save the space and the time complexity, we further resize this dimension to $32 \times 22 \times 10$, then the preprocessing operations have been completed.

3.2. Experimental results and discussions

The proposed SDTTV algorithm provides an efficient and discriminative vector human gait feature representation for subsequent recognition, thus its performance should be evaluated by the recognition rate. As suggested in [16,45], a matching score which measures a series of test features (i.e., a probe sequence set) against a group of training features (i.e., a gallery sequence set) is employed. This kind of distance metric is symmetric with respect to probe and gallery sequences by which the output matching score would be identical if the probe set and gallery set are interchanged. The behind mechanism of that matching score could refer to the aforementioned literature.

Followed by the numerous gait recognition researches, we use the rank-1 and rank-5 recognition rates to evaluate the performance of human gait recognition. The rank-1 recognition rate is the percentage of the number of correct subjects in the first place of all retrieved subjects and while the rank-5 recognition rate is the percentage of the number of correct subjects in any of the first five places of all retrieved subjects [8]. The proposed SDTTV algorithm is also compared by some of the baseline algorithms, including the baseline algorithm developed by the data publisher [8], the

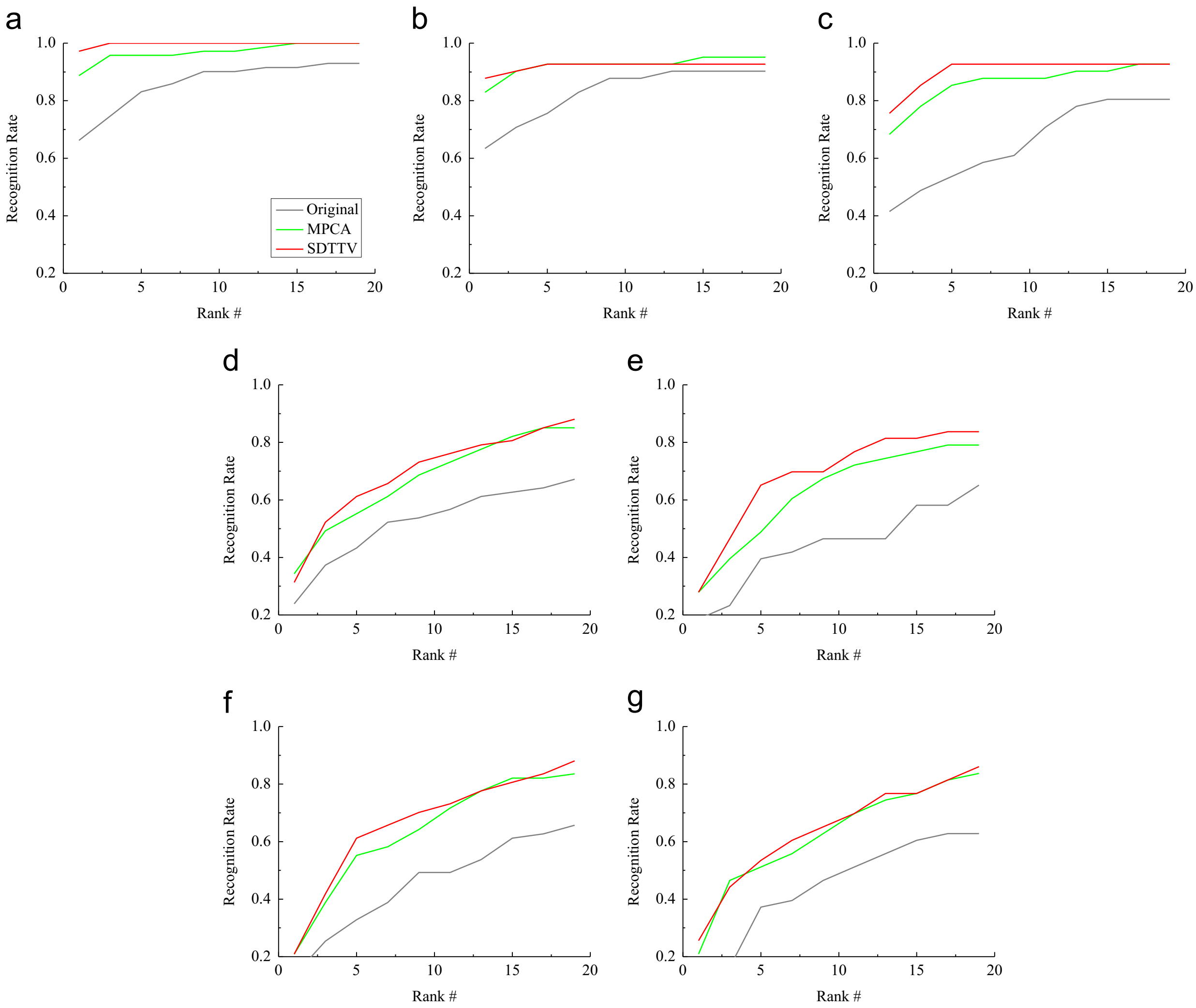

Fig. 2. Gait recognition rate for different k values.

hidden Markov model (HMM) proposed in [11], the linear time normalization (LTN) algorithm suggested in [45], and the pure multi-linear PCA algorithm shared by Lu [16].

Tables 3 and 4 report all the experiments, which compare the proposed algorithms with the existing algorithms. The word average in Tables 3 and 4 indicates the averaged recognition rates of all probes (A–G), that is, the ratio of correctly recognized subjects to the total number of subjects in all probes. From the comparison results in Tables 3 and 4, it is clear that the average recognition rates of the seven probes in our algorithm outperforms the previous state-of-the-art algorithms. It should be pointed out that HMM algorithm also achieves the pleased rank-1 and rank-5 recognition rates; however, besides feature extraction and matching, HMM parameter estimation (training) is a major component as well.

Fig. 2(a)–(g) further compares the recognition performance of our algorithm against the original feature and the MPCA algorithm induced feature. As a comprehensive investigation, Fig. 2(a)–(g) plots the identification rates with respect to the various rank k values ranging from 1 to 20, the specific definition of which is similar to the above-discussed rank-1 and rank-5 recognition rates. It is obvious from these figures that the proposed algorithm delivers the most efficient and discriminative feature for human gait recognition. It could also be observed from Tables 3 and 4 and Fig. 2 that the performance of our gait recognition algorithm is satisfactory when there are changes in shoe or/and viewpoint (the corresponding rank-5 recognition rates are higher than 90%). However, in the case of changes in the surface, the performance of all the tested algorithms degrades and thus the recognition rate decreases. In spite of this, in most cases, the correct subject is with high confidence in the top 10 matches of the recognition result.

4. Conclusion

This paper focuses on the feature representation and extraction of the silhouette based models for human gait recognition. By considering the gait sequences as 3-order-tensors, we introduce an efficient tensor to vector projection algorithm to find a low-dimensional tensor subspace of the original input gait sequence tensors while most of the data variation has been well captured. In order to further enhance the class separability and avoid the potential overfitting, we adopt a discriminative locality preserving projection with sparse regularization to transform the refined tensor data to the final vector feature representation for subsequent recognition.

Numerous experiments are carried out on the canonical data set, i.e., the USF HumanID Database. According to the experimental results, the average recognition rates of the seven probes show that our algorithm outperforms the previous baseline algorithms, and the proposed algorithm could find out the correct subject in the top 10 matches of the recognition result in most cases.

Acknowledgments

The authors would like to thank this special issue's editor and the anonymous reviewers for their careful reading and helpful remarks, which have contributed in improving the quality of this paper. This paper is supported in part by the National Natural Science Foundation of China under Grants 61102128, 91338202 and 91338111, and in part by the Fundamental Research Funds for the Central Universities under Grant 211-274175.

References

[1] 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.
[2] 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.
[3] N.V. Boulgouris, D. Hatzinakos, K.N. Plataniotis, Gait recognition: a challenging signal processing technology for biometric identification, IEEE Signal Process. Mag. 22 (6) (2005) 78–90.
[4] 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.
[5] S.D. Choudhury, T. Tjahjadi, Silhouette-based gait recognition using procrustes shape analysis and elliptic Fourier descriptors, Pattern Recognit. 45 (9) (2012) 3414–3426.
[6] X. Li, S.J. Maybank, S. Yan, D. Tao, D. Xu, Gait components and their application to gender recognition, IEEE Trans. Syst. Man Cybern. Part C: Appl. Rev. 38 (2) (2008) 145–155.
[7] N. Liu, J. Lu, Y.-P. Tan, Joint subspace learning for view-invariant gait recognition, IEEE Signal Process. Lett. 18 (7) (2011) 431–434.
[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, Statistical feature fusion for gait-based human recognition, in: IEEE CVPR, vol. 2, 2004, pp. 842–847.
[10] R. Tanawongsuwan, A. Bobick, Modelling the effects of walking speed on appearance-based gait recognition, in: IEEE CVPR, vol. 2, 2004, pp. 783–790.
[11] A. Kale, A. Sundaresan, A.N. Rajagopalan, N.P. Cuntoor, A.K. Roy-Chowdhury, V. Krger, R. Chellappa, Identification of humans using gait, IEEE Trans. Image Process. 13 (9) (2004) 1163–1173.
[12] A. Bobick, A. Johnson, Gait recognition using static activity-specific parameters, in: IEEE CVPR, vol. 1, 2001, pp. 423–430.
[13] D. Cunado, M. Nixon, J. Carter, Automatic extraction and description of human gait models for recognition purposes, Comput. Vis. Image Underst. 90 (1) (2003) 1–41.
[14] R. Cutler, L. Davis, Robust periodic motion and motion symmetry detection, in: IEEE CVPR, 2000, pp. 615–622.
[15] Z. Liu, S. Sarkar, Simplest representation yet for gait recognition: averaged silhouette, in: IEEE CVPR, vol. 4, 2004, pp. 211–214.
[16] H. Lu, K.N. Plataniotis, A.N. Venetsanopoulos, Mpca: multilinear principal component analysis of tensor objects, IEEE Trans. Neural Netw. 19 (1) (2008) 18–39.
[17] H. Lu, K.N. Plataniotis, A.N. Venetsanopoulos, Uncorrelated multilinear discriminant analysis with regularization and aggregation for tensor object recognition, IEEE Trans. Neural Netw. 20 (1) (2009) 103–123.
[18] I.T. Jolliffe, Principal Component Analysis, Springer, New York, USA, 2002.
[19] R.A. Fisher, The use of multiple measurements in taxonomic problems, Ann. Hum. Genet. 7 (2) (1936) 179–188.
[20] W. Liu, H. Zhang, D. Tao, Y. Wang, K. Lu, Large-scale paralleled sparse principal component analysis, Multimed. Tools Appl. (2014), http://dx.doi.org/10.1007/s11042-014-2004-4.
[21] J. Yu, D. Tao, J. Li, J. Cheng, Semantic preserving distance metric learning and applications, Inf. Sci. 281 (2014) 674–686, http://dx.doi.org/10.1016/j.ins.2014.01.025.
[22] J. Yu, D. Liu, D. Tao, H.S. Seah, Complex object correspondence construction in two-dimensional animation, IEEE Trans. Image Process. 20 (11) (2011) 3257–3269.
[23] V.N. Vapnik, An overview of statistical learning theory, IEEE Trans. Neural Netw. 10 (5) (1999) 988–999.
[24] D. Tao, X. Li, X. Wu, W. Hu, S.J. Maybank, Supervised tensor learning, Knowl. Inf. Syst. 13 (1) (2007) 1–42.
[25] D. Tao, X. Tang, X. Li, X. Wu, Asymmetric bagging and random subspace for support vector machines-based relevance feedback in image retrieval, IEEE Trans. Pattern Anal. Mach. Intell. 28 (7) (2006) 1088–1099.
[26] D. Xu, S. Yan, L. Zhang, S. Lin, H.-J. Zhang, T.S. Huang, Reconstruction and recognition of tensor-based objects with concurrent subspaces analysis, IEEE Trans. Circuits Syst. Video Technol. 18 (1) (2008) 36–47.
[27] C. Tian, G. Fan, X. Gao, Q. Tian, Multiview face recognition: from tensorface to v-tensorface and k-tensorface, IEEE Trans. Syst. Man Cybern. Part B: Cybern. 42 (2) (2012) 320–333.
[28] L. Zhang, L. Zhang, D. Tao, X. Huang, A multifeature tensor for remote-sensing target recognition, IEEE Geosci. Remote Sens. Lett. 8 (2) (2011) 374–378.
[29] L. Zhang, L. Zhang, D. Tao, X. Huang, Tensor discriminative locality alignment for hyperspectral image spectral-spatial feature extraction, IEEE Trans. Geosci. Remote Sens. 51 (1) (2013) 242–256.
[30] J. Yang, D. Zhang, A.F. Frangi, J. yu Yang, Two-dimensional pca: a new approach to appearance-based face representation and recognition, IEEE Trans. Pattern Anal. Mach. Intell. 26 (1) (2004) 131–137.
[31] Y. Pang, D. Tao, Y. Yuan, X. Li, Binary two-dimensional pca, IEEE Trans. Syst. Man Cybern. Part B: Cybern. 38 (4) (2008) 1176–1180.
[32] X. He, D. Cai, P. Niyogi, Tensor subspace analysis, in: NIPS, vol. 17, 2005, pp. 499–506.
[33] H. Lu, K.N. Plataniotis, A.N. Venetsanopoulos, A survey of multilinear subspace learning for tensor data, Pattern Recognit. 44 (7) (2011) 1540–1551.
[34] L.D. Lathauwer, Signal processing based on multilinear algebra (Ph. D. thesis), Katholike Universiteit Leuven, 1997.
[35] T.G. Kolda, B.W. Bader, Tensor decompositions and applications, SIAM Rev. 51 (3) (2009) 455–500.
[36] D. Tao, X. Li, X. Wu, S. Maybank, Tensor rank one discriminant analysis—a convergent method for discriminative multilinear subspace selection, Neurocomputing 71 (10–12) (2008) 1866–1882.
[37] S. Yan, D. Xu, Q. Yang, L. Zhang, X. Tang, H.-J. Zhang, Discriminant analysis with tensor representation, in: CVPR, vol. 1, 2005, pp. 526–532.
[38] T. Zhang, D. Tao, X. Li, J. Yang, Patch alignment for dimensionality reduction, IEEE Trans. Knowl. Data Eng. 21 (9) (2009) 1299–1313.
[39] J. Yu, D. Liu, D. Tao, H.S. Seah, On combining multiple features for cartoon character retrieval and clip synthesis, IEEE Trans. Syst. Man Cybern. Part B: Cybern. 42 (5) (2012) 1413–1427.
[40] W. Liu, D. Tao, J. Cheng, Y. Tang, Multiview hessian discriminative sparse coding for image annotation, Comput. Vis. Image Underst. 118 (2014) 50–60.
[41] W. Liu, D. Tao, Multiview hessian regularization for image annotation, IEEE Trans. Image Process. 22 (7) (2013) 2676–2687.
[42] D. Cai, X. He, J. Han, Spectral regression: a unified approach for sparse subspace learning, in: ICDM, 2007, pp. 73–82.

[43] L. Zhang, L. Zhang, D. Tao, X. Huang, Sparse transfer manifold embedding for hyperspectral target detection, IEEE Trans. Geosci. Remote Sens. 52 (2) (2014) 1030–1043.
[44] T. Zhou, D. Tao, X. Wu, Manifold elastic net: a unified framework for sparse dimension reduction, Data Min. Knowl. Disc. 22 (3) (2010) 340–371.
[45] N.V. Boulgouris, K.N. Plataniotis, D. Hatzinakos, Gait recognition using linear time normalization, Pattern Recognit. 39 (5) (2006) 969–979.