



A general tensor representation framework for cross-view gait recognition



Xianye Ben^{a,*}, Peng Zhang^{a,b}, Zhihui Lai^c, Rui Yan^d, Xinliang Zhai^a, Weixiao Meng^e

^aSchool of Information Science and Engineering, Shandong University, Qingdao 266237, China

^bSchool of Electrical and Data Engineering, University of Technology Sydney, Sydney, NSW 2007, Australia

^cCollege of Computer Science and Software Engineering, Shenzhen University, Shenzhen 518060, China

^dMicrosoft AI & Research, Bellevue, WA 98004, USA

^eSchool of Electronics and Information Engineering, Harbin Institute of Technology, Harbin 150080, China

ARTICLE INFO

Article history:

Received 28 February 2018

Revised 18 December 2018

Accepted 7 January 2019

Available online 15 January 2019

Keywords:

Gait recognition

Cross-view gait

Tensor representation

Framework

ABSTRACT

Tensor analysis methods have played an important role in identifying human gaits using high dimensional data. However, when view angles change, it becomes more and more difficult to recognize cross-view gait by learning only a set of multi-linear projection matrices. To address this problem, a general tensor representation framework for cross-view gait recognition is proposed in this paper. There are three criteria of tensorial coupled mappings in the proposed framework. (1) Coupled multi-linear locality-preserved criterion (CMLP) aims to detect the essential tensorial manifold structure via preserving local information. (2) Coupled multi-linear marginal fisher criterion (CMMF) aims to encode the intra-class compactness and inter-class separability with local relationships. (3) Coupled multi-linear discriminant analysis criterion (CMDA) aims to minimize the intra-class scatter and maximize the inter-class scatter. For the three tensor algorithms for cross-view gaits, two sets of multi-linear projection matrices are iteratively learned using alternating projection optimization procedures. The proposed methods are compared with the recently published cross-view gait recognition approaches on CASIA(B) and OU-ISIR gait database. The results demonstrate that the performances of the proposed methods are superior to existing state-of-the-art cross-view gait recognition approaches.

© 2019 Elsevier Ltd. All rights reserved.

1. Introduction

The demand to recognize and authenticate individuals using biometrics has been rising due to its broad applications in security and surveillance. During past decades, many biometrics have been applied to practice such as face, fingerprint and iris. These biometrics are unique from person to person, which is essential to quickly identify the target's identity. Different from these biometrics, i.e., face, fingerprint, vein, iris, ear, hand shape, palm print, retina and lip, gait is a kind of soft biometric which aims to recognize one's identity by his unique walking patterns. It is more potential than the biological characteristics in the surveillance field due to the advantage of gait recognition [1,2] lies in the fact that it can be efficiently recognized at a distance without subjects' cooperation.

The popular gait recognition methods can be roughly classified into the following two categories: model-based [3] and motion-

based approaches [4–6]. Model-based approaches can extract the gait features robustly and avoid the noise interference problem. The changes all over the body can be characterized by a short vector. It is possible for gait recognition to obtain a good performance if the model is established accurately. However, the modelling and its matching processes are both complex. Compared to model-based methods, motion-based approaches avoid the complex modelling which can characterize the motion patterns of human body without fitted model parameters. Due to the benefits, motion-based approaches attract more attention recent years. However, a challenge to motion-based gait recognition is the cross-view issue [7,8]. This is because motion pattern changes dramatically as the viewing directions move, even though when it is the same subject's gait. This is the main reason why most state-of-the-art motion-based gait recognition method do not perform well [9–12]. Basically, there is a trend that the larger the variation of viewing direction is, the worse the recognition performance is. Thus, the core research question of this paper is how can we find robust and discriminative representations, such that they can enlarge the

* Corresponding author.

E-mail address: benxianye@gmail.com (X. Ben).

discrimination between different subjects, and meanwhile compact the variations of the same subject?

Previous methods try to bridge the view biases by constructing 3D model and performing view transformation model, however, the former is computationally complex and the latter does not consider discriminability. Data-driven CNN approaches achieve significant success in many fields [13–16], which can be also applied to gait recognition. Due to its powerful representation ability, CNN can extract view invariant features. However, limited labelled data available easily causes the CNN model over-fitting. In another aspect, CNN model highly relies on expensive GPU hardware to accelerate the training speed. Different from them, tensors are higher order generalizations of matrices [17], which is helpful to reduce the small sample size problem in discriminative subspace selection. They have been successfully applied to gait recognition under a fixed angle of view.

Inspired by success of tensor representation, this paper presents a novel general tensor representation framework for cross-view gait recognition. Our formulations model the gait data as a tensor and seek three robust and tensorial discriminative representations by tensor analysis. Our framework can leverage structure information and reduce the number of parameters used to model the cross-view gait recognition. We present three novel criteria of tensorial coupled mappings. First, by preserving local information, we obtain a common subspace that best detects the essential gait manifold structure. Second, by encoding the intra-class compactness and inter-class separability with local relationships, we present coupled multi-linear marginal fisher criterion. Third, by minimizing the intra-class scatter of cross-view gaits, and simultaneously maximizing the inter-class scatter, we propose coupled multi-linear discriminant analysis criterion. These three tensor alignment algorithms of cross-view gaits are achieved by alternating projection optimization procedures. The flourishing of cross-view gait recognition methods depends largely on well-established multi-view gait databases, such as CASIA(B) [18] and OU-ISIR [19]. To the best of our knowledge, our work is the first attempt to address the cross-view gait recognition within a framework of tensor representations. The key contributions of our work can be summarized as follows:

- (1) We propose to model cross-view gait data as tensors and develop a novel framework of cross-view gait recognition by tensor representations.
- (2) We present three novel criteria of tensorial coupled mappings with their tensor alignment algorithms of cross-view gaits.
- (3) We systematically evaluate our methods on both the largest number of cross-views gait database and the largest population gait database.

The remainder of this paper is organized as follows: Section 2 briefly reviews some related works. Section 3 presents our general tensor representation framework for cross-view gait recognition. After that, Section 4 proposes 3 criteria of tensorial coupled mappings with their tensor alignment algorithms of cross-view gaits. Then, Section 5 demonstrates the experimental results on both CASIA(B) and OU-ISIR gait database. Finally, this paper is concluded in Section 6.

2. Related work

In this section, we give a brief literature review of related topics to our work, i.e., cross-view gait recognition and tensor representation for gait analysis.

2.1. Cross-view gait recognition

Several related work tries to tackle this cross-view gait recognition problem, which can be categories into three classes. The first class of work focuses on constructing 3D gait information via panoramic or multiple calibrated cameras [10,20–22]. These 3D-based methods are usually set-up with complicated environment of controlled multi-cameras, which may not be available in practice. Even if it is available, its practical application can be adversely impacted by the computation complexity. The second category is based on view transformation model (VTM). This includes single value decomposition (SVD) and regression, which have been massively deployed to generate gait features with the information from the other view [23–33]. Although these methods minimize the errors between the transformed features and original gaits features, they do not consider the discrimination capability. The third category is to extract view-invariant gait feature. Generally, it infers a view-invariant gait feature among cross-view gait data. For example, in [34], Goffredo et al. proposed a self-calibration of limbs' pose in the image reference system. However, this method can only coarsely estimate the limbs' pose when the view of input gait is very different from the registered or front view gaits. To alleviate this problem, domain transformation [35–38], metric learning [39], and deep CNNs [40–42] have been introduced recently. Especially, deep CNNs have achieved encouraging recognition accuracy on the cross-view task due to its powerful representation ability. The premise of using deep CNNs requires to a large quantity of labelled training data efficiently, however, the limited gait data available restricts its application.

2.2. Tensor representation for gait analysis

A variety of multi-linear subspace learning approaches based on tensor representation have been applied to gait analysis, which can not only extract spatial-temporal gait information but also avoid small size sample problem. For example, Lu et al. [43] proposed multi-linear principal component analysis (MPCA) to capture most of the original tensorial input variation. Then, they extend MPCA to Uncorrelated multi-linear principal component analysis (UMPCA) [44], which can produce uncorrelated features while capturing most of the variation in the original tensorial input. However, the above algorithms only concentrate on unsupervised dimension reduction instead of discriminative feature extraction and classification. Therefore, multi-linear discriminant analysis (MLDA) approaches [45–47] are proposed for gait feature extraction and classification. Tao et al. proposed the general tensor discriminant analysis (GTDA) [45] via maximizing the differential between inter-class scatters and the weighted intra-class scatters. Yan et al. proposed discriminant analysis with tensor representation (DATER) [46] via maximizing the ratio of inter-class scatters to the intra-class scatters. Lu et al. [47] developed uncorrelated multi-linear discriminant analysis (UMDA) to explore uncorrelated discriminative features for gait recognition. Li et al. [48] applied locally linear embedding (LLE) criterion and separability between different classes to formulate tensorization of Discriminant LLE. Aiming at extracting discriminative geometry-preserving features from the original tensorial data, Ben et al. [49] proposed maximum margin projection with tensor representation. By characterizing the multi-factor variation, Chen and Hsu [50] proposed multilinear graph embedding (MGE) to adequately characterize local variations. Zhao et al. [51] adopted sparse constraint in tensor discriminative locality alignment to select gait features. However, the existing tensor analysis on gait recognition do not focus on extracting view-invariant gait feature, and the tensor representation framework for cross-view gait recognition is still lack of study.

3. A general tensor representation framework for cross-view gait recognition

3.1. Cross-view gait data multi-linear transformation

Human gait samples are usually represented by second-order tensor or higher-order tensor. Given two sets of training gait tensorial samples $\{\mathcal{X}_i \in \mathbb{R}^{H_1 \times H_2 \times \dots \times H_N}, i = 1, 2, \dots, M_\theta\}$ and $\{\mathcal{Y}_j \in \mathbb{R}^{L_1 \times L_2 \times \dots \times L_N}, j = 1, 2, \dots, M_\vartheta\}$ from two views θ and ϑ , where H_n and L_n are mode- n dimensions for views θ and ϑ , respectively. M_θ and M_ϑ are the numbers of gait tensorial samples for views θ and ϑ , respectively. Generally, both share a consistent one-to-one match between two views, namely, $M_\theta = M_\vartheta = M$. Now our goal is to find transformation functions $f_\theta(\mathcal{X}_i)$ and $f_\vartheta(\mathcal{Y}_j)$ to make gait data under different views project into a common space and measure their similarity. The objective of a general tensor representation framework for cross-view gait recognition is to find a pair of multi-linear transformation matrices $\{\mathbf{U}_n \in \mathbb{R}^{H_n \times F_n}, n = 1, \dots, N\}$ and $\{\mathbf{V}_n \in \mathbb{R}^{L_n \times F_n}, n = 1, \dots, N\}$ to project cross-view gait samples into a common lower-dimensional tensorial subspace $F_1 \otimes F_2 \otimes \dots \otimes F_N$ from both of original tensorial space $H_1 \otimes H_2 \otimes \dots \otimes H_N$ and $L_1 \otimes L_2 \otimes \dots \otimes L_N$ respectively. Thus,

$$\mathcal{A}_i = \mathcal{X}_i \times_1 \mathbf{U}_1^\top \times_2 \mathbf{U}_2^\top \cdots \times_N \mathbf{U}_N^\top = \mathcal{X}_i \prod_{k=1}^N \times_k \mathbf{U}_k^\top, \quad (1)$$

$$\mathcal{B}_j = \mathcal{Y}_j \times_1 \mathbf{V}_1^\top \times_2 \mathbf{V}_2^\top \cdots \times_N \mathbf{V}_N^\top = \mathcal{Y}_j \prod_{k=1}^N \times_k \mathbf{V}_k^\top, \quad (2)$$

where \mathcal{A}_i and \mathcal{B}_j are projected tensor features for the views θ and ϑ , and $\mathcal{A}_i, \mathcal{B}_j \in \mathbb{R}^{F_1 \times F_2 \times \dots \times F_N}$, $i, j = 1, \dots, M$. F_n is the mode- n dimension of the projected tensor features, and $F_n \leq \min(H_n, L_n)$.

3.2. The similarity of cross-view gait data

To simplify the calculation on the similarity between \mathcal{X}_i and \mathcal{Y}_j across view, we define

$$\text{sim}(\mathcal{X}_i, \mathcal{Y}_j) = \|\mathcal{A}_i - \mathcal{B}_j\|_F^2 w_{ij}, \quad (3)$$

w_{ij} is the similarity relationship between \mathcal{X}_i and \mathcal{Y}_j , which can be calculated according to side information constraint, Gaussian similarity or Cosine similarity [52]. The similarity presented in Table 1 can be used in this paper.

The mode- n unfoldings of \mathcal{A}_i and \mathcal{B}_j can be derived as

$$\mathbf{A}_{i(n)} = \mathbf{U}_n^\top \mathbf{X}_{i(n)} (\mathbf{U}_N \otimes \dots \otimes \mathbf{U}_{n+1} \otimes \mathbf{U}_{n-1} \otimes \dots \otimes \mathbf{U}_1) = \mathbf{U}_n^\top \mathbf{X}_{i(n)} \tilde{\mathbf{U}}_n \quad (4)$$

$$\mathbf{B}_{j(n)} = \mathbf{V}_n^\top \mathbf{Y}_{j(n)} (\mathbf{V}_N \otimes \dots \otimes \mathbf{V}_{n+1} \otimes \mathbf{V}_{n-1} \otimes \dots \otimes \mathbf{V}_1) = \mathbf{V}_n^\top \mathbf{Y}_{j(n)} \tilde{\mathbf{V}}_n \quad (5)$$

where $\tilde{\mathbf{U}}_n = \prod_{k=1, k \neq n}^N \mathbf{U}_k$, $\tilde{\mathbf{V}}_n = \prod_{k=1, k \neq n}^N \mathbf{V}_k$. (9) can be written as

$$\begin{aligned} \text{sim}(\mathcal{X}_i, \mathcal{Y}_j) &= \|\mathcal{A}_i - \mathcal{B}_j\|_F^2 w_{ij} \\ &= \text{Tr} \left(\begin{bmatrix} \mathbf{U}_n \\ \mathbf{V}_n \end{bmatrix}^\top \begin{bmatrix} \mathbf{X}_{i(n)} \tilde{\mathbf{U}}_n & \mathbf{0} \\ \mathbf{0} & \mathbf{Y}_{j(n)} \tilde{\mathbf{V}}_n \end{bmatrix} \begin{bmatrix} w_{ij} \cdot \mathbf{I} & -w_{ij} \cdot \mathbf{I} \\ -w_{ij} \cdot \mathbf{I} & w_{ij} \cdot \mathbf{I} \end{bmatrix} \right. \\ &\quad \left. \begin{bmatrix} \mathbf{X}_{i(n)} \tilde{\mathbf{U}}_n & \mathbf{0} \\ \mathbf{0} & \mathbf{Y}_{j(n)} \tilde{\mathbf{V}}_n \end{bmatrix}^\top \begin{bmatrix} \mathbf{U}_n \\ \mathbf{V}_n \end{bmatrix} \right) \\ &= \text{Tr} \left(\mathbf{P}_n^\top \mathbf{Z}_{ij}^{(n)} \mathbf{G}_{ij} (\mathbf{Z}_{ij}^{(n)})^\top \mathbf{P}_n \right), \quad (6) \end{aligned}$$

where $\mathbf{P}_n = \begin{bmatrix} \mathbf{U}_n \\ \mathbf{V}_n \end{bmatrix}$, $\mathbf{Z}_{ij}^{(n)} = \begin{bmatrix} \mathbf{X}_{i(n)} \tilde{\mathbf{U}}_n & \mathbf{0} \\ \mathbf{0} & \mathbf{Y}_{j(n)} \tilde{\mathbf{V}}_n \end{bmatrix}$, $\mathbf{G}_{ij} = \begin{bmatrix} w_{ij} \cdot \mathbf{I} & -w_{ij} \cdot \mathbf{I} \\ -w_{ij} \cdot \mathbf{I} & w_{ij} \cdot \mathbf{I} \end{bmatrix}$. The detailed mathematical deductions are put into the Appendix A.

In the following section, we will introduce 3 criteria with tensorial coupled mappings to obtain \mathbf{U}_n and \mathbf{V}_n . Table 2 shows the alignment representation of each criteria used in the proposed framework.

3.3. Classification

We denote \mathbf{Q}_1 and \mathbf{Q}_2 be projection matrices trained by the improved metric learning approach [52] respectively for the vectorized cross-view gait data $\text{vec}(\mathcal{A}_i), \text{vec}(\mathcal{B}_j)$, $i, j = 1, \dots, M$.

In the testing stage, with the learned multi-linear transformation matrices $\{\mathbf{U}_n \in \mathbb{R}^{H_n \times F_n}, n = 1, \dots, N\}$ and $\{\mathbf{V}_n \in \mathbb{R}^{L_n \times F_n}, n = 1, \dots, N\}$, the class label of a query gait tensor $\mathcal{Y} \in \mathbb{R}^{L_1 \times L_2 \times \dots \times L_N}$ under the view ϑ is determined by π_{i^*}

$$i^* = \arg \min_i \text{dis} (f_\theta(\mathcal{X}_i), f_\vartheta(\mathcal{Y})), \quad (7)$$

where π_{i^*} denotes the class label of the tensor gait sample which has smallest distance to the query gait sample \mathcal{Y} , and $\text{dis}(\cdot, \cdot)$ denotes a distance metric function of transformed gait tensor data. $f_\theta(\cdot)$ and $f_\vartheta(\cdot)$ are transformation functions from tensor spaces $H_1 \otimes H_2 \otimes \dots \otimes H_N$ and $L_1 \otimes L_2 \otimes \dots \otimes L_N$ to a lower dimensional vector space, where

$$f_\theta(\mathcal{X}_i) = \mathbf{Q}_1^\top \text{vec} \left(\mathcal{X}_i \prod_{k=1}^N \times_k \mathbf{U}_k^\top \right), \quad (8)$$

$$f_\vartheta(\mathcal{Y}) = \mathbf{Q}_2^\top \text{vec} \left(\mathcal{Y} \prod_{k=1}^N \times_k \mathbf{V}_k^\top \right). \quad (9)$$

4. Three criteria of tensorial coupled mappings

In this section, we introduce three different criteria under the unified tensorial framework in the last section and analyse the relationship between them.

4.1. Coupled multi-linear locality-preserved criterion (CMLP)

Preserving local information, tensorial coupled mappings with CMLP criterion aim to learn a couple of multi-linear projection matrices for views θ and ϑ to obtain a common subspace that best detects the essential gait manifold structure. The objective function is defined as

$$\begin{aligned} \{\mathbf{U}_n^*, \mathbf{V}_n^*, n = 1, \dots, N\} &= \underset{\mathbf{U}_n, \mathbf{V}_n, n=1, \dots, N}{\text{argmin}} \sum_{ij} \left\| \mathcal{X}_i \prod_{k=1}^N \times_k \mathbf{U}_k^\top \right. \\ &\quad \left. - \mathcal{Y}_j \prod_{k=1}^N \times_k \mathbf{V}_k^\top \right\|_F^2 w_{ij}. \quad (10) \end{aligned}$$

The above objective function does not have a closed-form solution, so an iterative procedure (see Algorithm 1) is proposed to solve it.

Its mode- n objective function can be written in a trace form as

$$\begin{aligned} \arg \min_{\mathbf{U}_n, \mathbf{V}_n} J(\mathbf{U}_n, \mathbf{V}_n) &= \sum_{ij} \|\mathbf{U}_n^\top \mathbf{X}_{i(n)} \tilde{\mathbf{U}}_n - \mathbf{V}_n^\top \mathbf{Y}_{j(n)} \tilde{\mathbf{V}}_n\|_F^2 w_{ij} \\ &= \text{Tr} \left(\begin{bmatrix} \mathbf{U}_n \\ \mathbf{V}_n \end{bmatrix}^\top \begin{bmatrix} \mathbf{X}_{i(n)} \tilde{\mathbf{U}}_n & \mathbf{0} \\ \mathbf{0} & \mathbf{Y}_{j(n)} \tilde{\mathbf{V}}_n \end{bmatrix} \right. \\ &\quad \left. \begin{bmatrix} \mathbf{D}_1 \otimes \mathbf{I} & -\mathbf{W} \otimes \mathbf{I} \\ -\mathbf{W}^\top \otimes \mathbf{I} & \mathbf{D}_2 \otimes \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{X}_{i(n)} \tilde{\mathbf{U}}_n & \mathbf{0} \\ \mathbf{0} & \mathbf{Y}_{j(n)} \tilde{\mathbf{V}}_n \end{bmatrix}^\top \begin{bmatrix} \mathbf{U}_n \\ \mathbf{V}_n \end{bmatrix} \right), \quad (11) \end{aligned}$$

Table 1
Similarity relationship between \mathcal{X}_i and \mathcal{X}_j .

Side information	$w_{ij} = \begin{cases} 1, & \{\mathcal{X}_i, \mathcal{X}_j\} \in \mathbb{C} \\ 0, & \text{else} \end{cases}$	\mathcal{X}_i and \mathcal{X}_j belong to the identical individual \mathbb{C} .
Cosine similarity	$w_{ij} = \begin{cases} \frac{\text{vec}(\mathcal{X}_i) \cdot \text{vec}(\mathcal{X}_j)}{\ \text{vec}(\mathcal{X}_i)\ \ \text{vec}(\mathcal{X}_j)\ }, & \text{if } \mathcal{X}_i \in N_k(\mathcal{X}_j) \text{ or } \mathcal{X}_j \in N_k(\mathcal{X}_i) \\ 0, & \text{else} \end{cases}$	$\text{vec}(\cdot)$ denotes a vectorization operator. $N_k(\cdot)$ denotes k nearest neighbour.
Gaussian similarity	$w_{ij} = \begin{cases} \exp(-\ \text{vec}(\mathcal{X}_i) - \text{vec}(\mathcal{X}_j)\ ^2/t), & \text{if } \mathcal{X}_i \in N_k(\mathcal{X}_j) \text{ or } \mathcal{X}_j \in N_k(\mathcal{X}_i) \\ 0, & \text{else} \end{cases}$	t denotes Gaussian variance parameter.

Table 2
The criteria, the alignment representation and the objective function used in the proposed framework.

Criteria	Alignment representation	Objective function for mode n
Coupled multi-linear locality-preserved criterion (CMLP)	$\begin{bmatrix} \mathbf{D}_1 \otimes \mathbf{I} & -\mathbf{W} \otimes \mathbf{I} \\ -\mathbf{W}^\top \otimes \mathbf{I} & \mathbf{D}_2 \otimes \mathbf{I} \end{bmatrix}$	$\arg \min_{\mathbf{U}_n, \mathbf{V}_n} \sum_j \ \mathbf{U}_n^\top \mathbf{X}_{(n)} \tilde{\mathbf{U}}_n - \mathbf{V}_n^\top \mathbf{Y}_{(n)} \tilde{\mathbf{V}}_n\ _F^2 w_{ij}$
Coupled multi-linear marginal fisher criterion (CMMF)	$\begin{bmatrix} \tilde{\mathbf{D}}_1 \otimes \mathbf{I} & -\tilde{\mathbf{W}} \otimes \mathbf{I} \\ -\tilde{\mathbf{W}}^\top \otimes \mathbf{I} & \tilde{\mathbf{D}}_2 \otimes \mathbf{I} \end{bmatrix}$	$\arg \min_{\mathbf{U}_n, \mathbf{V}_n} \frac{\sum_{i=1}^{N_c} \ \mathbf{U}_n^\top \mathbf{X}_{(n)} \tilde{\mathbf{U}}_n - \mathbf{V}_n^\top \mathbf{Y}_{(n)} \tilde{\mathbf{V}}_n\ _F^2 w_{ij}}{\sum_{i=1}^{N_c} \ \mathbf{U}_n^\top \mathbf{X}_{(n)} \tilde{\mathbf{U}}_n - \mathbf{V}_n^\top \mathbf{Y}_{(n)} \tilde{\mathbf{V}}_n\ _F^2 w_{ij}}$
Coupled multi-linear discriminant analysis criterion (CMDA)	$\begin{bmatrix} \mathbf{I} - \mathbf{K} \otimes \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} - \mathbf{K} \otimes \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{K} \otimes \mathbf{I} - \frac{1}{M} (\mathbf{e}\mathbf{e}^\top) \otimes \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{K} \otimes \mathbf{I} - \frac{1}{M} (\mathbf{e}\mathbf{e}^\top) \otimes \mathbf{I} \end{bmatrix}$	$\arg \max_{\mathbf{U}_n, \mathbf{V}_n} \frac{\sum_{i=1}^{N_c} \left(N_c \ \mathbf{U}_n^\top \tilde{\mathbf{X}}_{(n)}^{(c)} \tilde{\mathbf{U}}_n - \mathbf{U}_n^\top \tilde{\mathbf{X}}_{(n)} \tilde{\mathbf{U}}_n\ _F^2 \right)}{\sum_{i=1}^{N_c} \left(N_c \ \mathbf{V}_n^\top \tilde{\mathbf{Y}}_{(n)}^{(c)} \tilde{\mathbf{V}}_n - \mathbf{V}_n^\top \tilde{\mathbf{Y}}_{(n)} \tilde{\mathbf{V}}_n\ _F^2 \right)}$

Algorithm 1: Projection of cross-view gaits with CMLP criterion.

Input: Two sets of training gait tensor samples $\{\mathcal{X}_i \in \mathbb{R}^{H_1 \times H_2 \times \dots \times H_N}, i = 1, 2, \dots, M\}$ and $\{\mathcal{Y}_j \in \mathbb{R}^{L_1 \times L_2 \times \dots \times L_N}, j = 1, 2, \dots, M\}$ from two views θ and ϑ , dimensionality of the transformed gait tensor $F_1 \times F_2 \times \dots \times F_N$, the maximum iteration T_{\max} ;
Output: The projection matrices $(\mathbf{U}_n, \mathbf{V}_n), n = 1, \dots, N$, the aligned tensors $\{\mathcal{A}_i \in \mathbb{R}^{F_1 \times \dots \times F_N}, i = 1, \dots, M\}$ and $\{\mathcal{B}_j \in \mathbb{R}^{F_1 \times \dots \times F_N}, j = 1, \dots, M\}$;

- 1 Initialize $\mathbf{U}_n^{(0)} = \mathbf{I}_{H_n}, \mathbf{V}_n^{(0)} = \mathbf{I}_{L_n}, n = 1, \dots, N$, where their superscript denotes iteration number;
- 2 Calculate $\mathbf{W}, \mathbf{D}_1, \mathbf{D}_2$ according to Table 1 and (12);
- 3 **while** not converged or $t \leq T_{\max}$ **do**
- 4 **for** mode $n = 1 : N$ **do**
- 5 Calculate
- 6 $\mathbf{X}_{i(n)} \tilde{\mathbf{U}}_n^{(t-1)} \leftarrow_n \mathcal{X}_i \times_1 (\mathbf{U}_1^{(t-1)})^\top \dots \times_{n-1} (\mathbf{U}_{n-1}^{(t-1)})^\top \times_{n+1} (\mathbf{U}_{n+1}^{(t-1)})^\top \dots \times_N (\mathbf{U}_N^{(t-1)})^\top$;
- 7 $\mathbf{Y}_{i(n)} \tilde{\mathbf{V}}_n^{(t-1)} \leftarrow_n \mathcal{Y}_i \times_1 (\mathbf{V}_1^{(t-1)})^\top \dots \times_{n-1} (\mathbf{V}_{n-1}^{(t-1)})^\top \times_{n+1} (\mathbf{V}_{n+1}^{(t-1)})^\top \dots \times_N (\mathbf{V}_N^{(t-1)})^\top$;
- 8 Calculate auxiliary matrices $\mathbf{X}_{(n)}$ and $\mathbf{Y}_{(n)}$ in (11), $\mathbf{Z}_{(n)}$ and \mathbf{G} in (13);
- 9 Calculate the eigenvalues and eigenvectors of $\mathbf{Z}_{(n)} \mathbf{G} \mathbf{Z}_{(n)}^\top \mathbf{p}^{(t)} = \lambda (\mathbf{Z}_{(n)} \mathbf{Z}_{(n)}^\top) \mathbf{p}^{(t)}$ and $\mathbf{p}_n^{(t)*} = [(\mathbf{U}_n^{(t)})^\top \quad (\mathbf{V}_n^{(t)})^\top]^\top = [\mathbf{p}_1^{(n,t)}, \mathbf{p}_2^{(n,t)}, \dots, \mathbf{p}_{F_n}^{(n,t)}]^\top$ where the eigenvectors $\mathbf{p}_1^{(n,t)}, \mathbf{p}_2^{(n,t)}, \dots, \mathbf{p}_{F_n}^{(n,t)}$ corresponding to F_n smallest eigenvalues;
- 10 **end**
- 11 Check convergence $\left\| \left\| \mathbf{p}_n^{(t)*} \left(\mathbf{p}_n^{(t-1)*} \right)^\top - \mathbf{I} \right\| \right\| \leq \varepsilon$.
- 12 **end**
- 13 Project the high-dimensional gait tensor to matrices $(\mathbf{U}_n, \mathbf{V}_n), n = 1, \dots, N$, and obtain the low-dimensional tensors $\mathcal{A}_i = \mathcal{X}_i \times_1 \mathbf{U}_1^\top \times_2 \mathbf{U}_2^\top \dots \times_N \mathbf{U}_N^\top$ and $\mathcal{B}_j = \mathcal{Y}_j \times_1 \mathbf{V}_1^\top \times_2 \mathbf{V}_2^\top \dots \times_N \mathbf{V}_N^\top$;

where $\mathbf{X}_{(n)} = [\mathbf{X}_{1(n)}, \mathbf{X}_{2(n)}, \dots, \mathbf{X}_{M(n)}]$, $\mathbf{Y}_{(n)} = [\mathbf{Y}_{1(n)}, \mathbf{Y}_{2(n)}, \dots, \mathbf{Y}_{M(n)}]$, \otimes denotes Kronecker product, and \mathbf{I} is a unit matrix, similarity matrix \mathbf{W} is made of w_{ij} , and diagonal matrices \mathbf{D}_1 and \mathbf{D}_2 are

$$\mathbf{D}_1 = \begin{bmatrix} \sum_j w_{1j} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \sum_j w_{Mj} \end{bmatrix}, \mathbf{D}_2 = \begin{bmatrix} \sum_i w_{i1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \sum_i w_{iM} \end{bmatrix}. \quad (12)$$

The detailed mathematical deductions of (11) are put into the Appendix B.

To simplify (11), several auxiliary matrices are defined as follows

$$\mathbf{P}_n = \begin{bmatrix} \mathbf{U}_n \\ \mathbf{V}_n \end{bmatrix}, \mathbf{Z}_{(n)} = \begin{bmatrix} \mathbf{X}_{(n)} \tilde{\mathbf{U}}_n & \mathbf{0} \\ \mathbf{0} & \mathbf{Y}_{(n)} \tilde{\mathbf{V}}_n \end{bmatrix}, \mathbf{G} = \begin{bmatrix} \mathbf{D}_1 \otimes \mathbf{I} & -\mathbf{W} \otimes \mathbf{I} \\ -\mathbf{W}^\top \otimes \mathbf{I} & \mathbf{D}_2 \otimes \mathbf{I} \end{bmatrix}. \quad (13)$$

Hence, we have the following objective function for mode- n

$$\{\mathbf{P}_n^*, n = 1, \dots, N\} = \arg \min_{\mathbf{P}_n} \text{Tr}(\mathbf{P}_n^\top \mathbf{Z}_{(n)} \mathbf{G} \mathbf{Z}_{(n)}^\top \mathbf{P}_n), \quad (14)$$

where \mathbf{P}_n^* denotes the optimal solution for mode- n . Therefore, (11) has been decomposed into N different sub-optimization problems although $2N$ optimized variables are coupled in a single objective function.

With the orthogonal constraints introduced into the (14) to make $\{\mathbf{P}_n^*, n = 1, \dots, N\}$ unique, we derive

$$\{\mathbf{P}_n^*, n = 1, \dots, N\} = \arg \min_{\mathbf{P}_n} \text{Tr}(\mathbf{P}_n^\top \mathbf{Z}_{(n)} \mathbf{G} \mathbf{Z}_{(n)}^\top \mathbf{P}_n) \quad (15)$$

s.t. $\mathbf{P}_n^\top \mathbf{P}_n = \mathbf{I}$,

where $\mathbf{I} \in \mathbb{R}^{F_n \times F_n}$ is a unit matrix, $e \in \mathbb{R}^{2MK_n \times 1}$ is a column vector of ones, $K_n = F_{n+1} \times \dots \times F_N \times F_{n-1} \times \dots \times F_1$, the size of a column vector of zeros $\mathbf{0}$ is $F_n \times 1$. The solution of (15) are listed in Appendix C.

4.2. Coupled multi-linear marginal fisher criterion (CMMF)

We proposed CMMF criterion to encode the intra-class compactness and inter-class separability with local relationships. As a

result, for each mode n , the objective function is

$$\{\mathbf{U}_n^*, \mathbf{V}_n^*, n = 1, \dots, N\} = \underset{\mathbf{U}_n, \mathbf{V}_n, n=1, \dots, N}{\operatorname{argmin}} \frac{\sum_{\pi_i=\pi_j} \|\mathcal{X}_i \times_1 \mathbf{U}_1^\top \cdots \times_N \mathbf{U}_N^\top - \mathcal{Y}_j \times_1 \mathbf{V}_1^\top \cdots \times_N \mathbf{V}_N^\top\|_F^2 \tilde{w}_{ij}}{\sum_{\pi_i \neq \pi_j} \|\mathcal{X}_i \times_1 \mathbf{U}_1^\top \cdots \times_N \mathbf{U}_N^\top - \mathcal{Y}_j \times_1 \mathbf{V}_1^\top \cdots \times_N \mathbf{V}_N^\top\|_F^2 \tilde{w}_{ij}} \quad (16)$$

where π_i and π_j denote the class labels of samples i and j , \tilde{w}_{ij} denotes the similarity of intra-class gait data, and \tilde{w}_{ij} denotes the similarity of inter-class gait data, which are defined as

$$\tilde{w}_{ij} = \begin{cases} 1, & \text{if } i \in N_{k_1}^+(j) \text{ or } j \in N_{k_1}^+(i), \\ 0, & \text{else} \end{cases}, \quad (17)$$

$$\tilde{w}_{ij} = \begin{cases} 1, & \text{if } i \in N_{k_2}^-(j) \text{ or } j \in N_{k_2}^-(i), \\ 0, & \text{else} \end{cases},$$

where $N_{k_1}^+(\cdot)$ denotes a set of k_1 intra-class nearest neighbours, and $N_{k_2}^-(\cdot)$ denotes a set of k_2 inter-class nearest neighbours.

Like CMLP, the alternating projection optimization procedure can also decompose (16) into N sub-optimization problems as follows

$$\operatorname{argmin}_{\mathbf{U}_n, \mathbf{V}_n} J(\mathbf{U}_n, \mathbf{V}_n) = \frac{\sum_{\pi_i=\pi_j} \|\mathbf{U}_n^\top \mathbf{X}_{i(n)} \tilde{\mathbf{U}}_n - \mathbf{V}_n^\top \mathbf{Y}_{j(n)} \tilde{\mathbf{V}}_n\|_F^2 \tilde{w}_{ij}}{\sum_{\pi_i \neq \pi_j} \|\mathbf{U}_n^\top \mathbf{X}_{i(n)} \tilde{\mathbf{U}}_n - \mathbf{V}_n^\top \mathbf{Y}_{j(n)} \tilde{\mathbf{V}}_n\|_F^2 \tilde{w}_{ij}}, \quad (18)$$

$$n = 1, \dots, N.$$

Eq. (18) can be rewritten as

$$\operatorname{argmin}_{\mathbf{U}_n, \mathbf{V}_n} J(\mathbf{U}_n, \mathbf{V}_n) = \frac{\operatorname{Tr} \left(\begin{bmatrix} \mathbf{U}_n \\ \mathbf{V}_n \end{bmatrix}^\top \begin{bmatrix} \mathbf{X}_{(n)} \tilde{\mathbf{U}}_n & \mathbf{0} \\ \mathbf{0} & \mathbf{Y}_{(n)} \tilde{\mathbf{V}}_n \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{D}}_1 \otimes \mathbf{I} & -\tilde{\mathbf{W}} \otimes \mathbf{I} \\ -\tilde{\mathbf{W}}^\top \otimes \mathbf{I} & \tilde{\mathbf{D}}_2 \otimes \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{X}_{(n)} \tilde{\mathbf{U}}_n & \mathbf{0} \\ \mathbf{0} & \mathbf{Y}_{(n)} \tilde{\mathbf{V}}_n \end{bmatrix}^\top \begin{bmatrix} \mathbf{U}_n \\ \mathbf{V}_n \end{bmatrix} \right)}{\operatorname{Tr} \left(\begin{bmatrix} \mathbf{U}_n \\ \mathbf{V}_n \end{bmatrix}^\top \begin{bmatrix} \mathbf{X}_{(n)} \tilde{\mathbf{U}}_n & \mathbf{0} \\ \mathbf{0} & \mathbf{Y}_{(n)} \tilde{\mathbf{V}}_n \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{D}}_1 \otimes \mathbf{I} & -\tilde{\mathbf{W}} \otimes \mathbf{I} \\ -\tilde{\mathbf{W}}^\top \otimes \mathbf{I} & \tilde{\mathbf{D}}_2 \otimes \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{X}_{(n)} \tilde{\mathbf{U}}_n & \mathbf{0} \\ \mathbf{0} & \mathbf{Y}_{(n)} \tilde{\mathbf{V}}_n \end{bmatrix}^\top \begin{bmatrix} \mathbf{U}_n \\ \mathbf{V}_n \end{bmatrix} \right)}, \quad (19)$$

where $\tilde{\mathbf{W}}$ and $\tilde{\mathbf{W}}$ are intra-class similarity matrix and inter-class penalty similarity matrix respectively and both of their i th row j th column elements are \tilde{w}_{ij} and \tilde{w}_{ij} . Four diagonal matrices $\tilde{\mathbf{D}}_1$, $\tilde{\mathbf{D}}_2$, $\tilde{\mathbf{D}}_1$ and $\tilde{\mathbf{D}}_2$ are

$$\tilde{\mathbf{D}}_1 = \begin{bmatrix} \sum_j \tilde{w}_{1j} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \sum_j \tilde{w}_{Mj} \end{bmatrix}, \tilde{\mathbf{D}}_2 = \begin{bmatrix} \sum_i \tilde{w}_{i1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \sum_i \tilde{w}_{iM} \end{bmatrix},$$

$$\tilde{\mathbf{D}}_1 = \begin{bmatrix} \sum_j \tilde{w}_{1j} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \sum_j \tilde{w}_{Mj} \end{bmatrix}, \tilde{\mathbf{D}}_2 = \begin{bmatrix} \sum_i \tilde{w}_{i1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \sum_i \tilde{w}_{iM} \end{bmatrix}. \quad (20)$$

To simplify (19), two alignment matrices are defined as follows

$$\tilde{\mathbf{G}} = \begin{bmatrix} \tilde{\mathbf{D}}_1 \otimes \mathbf{I} & -\tilde{\mathbf{W}} \otimes \mathbf{I} \\ -\tilde{\mathbf{W}}^\top \otimes \mathbf{I} & \tilde{\mathbf{D}}_2 \otimes \mathbf{I} \end{bmatrix}, \tilde{\mathbf{G}} = \begin{bmatrix} \tilde{\mathbf{D}}_1 \otimes \mathbf{I} & -\tilde{\mathbf{W}} \otimes \mathbf{I} \\ -\tilde{\mathbf{W}}^\top \otimes \mathbf{I} & \tilde{\mathbf{D}}_2 \otimes \mathbf{I} \end{bmatrix}. \quad (21)$$

Then, (19) reduces to

$$\operatorname{argmin}_{\mathbf{P}_n} J(\mathbf{P}_n) = \frac{\operatorname{Tr}(\mathbf{P}_n^\top \mathbf{Z}_{(n)} \tilde{\mathbf{G}} \mathbf{Z}_{(n)}^\top \mathbf{P}_n)}{\operatorname{Tr}(\mathbf{P}_n^\top \mathbf{Z}_{(n)} \tilde{\mathbf{G}} \mathbf{Z}_{(n)}^\top \mathbf{P}_n)}. \quad (22)$$

Like CMLP, a regularizer $\tau \mathbf{I}$, which can be viewed as a small disturbance, can be also imposed on the item $\tilde{\mathbf{G}}$ to avoid over fitting. Then we have the following criterion

$$\operatorname{argmin}_{\mathbf{P}_n} J(\mathbf{P}_n) = \frac{\operatorname{Tr}(\mathbf{P}_n^\top \mathbf{Z}_{(n)} \tilde{\mathbf{G}} \mathbf{Z}_{(n)}^\top \mathbf{P}_n)}{\operatorname{Tr}(\mathbf{P}_n^\top \mathbf{Z}_{(n)} (\tilde{\mathbf{G}} + \tau \mathbf{I}) \mathbf{Z}_{(n)}^\top \mathbf{P}_n)}. \quad (23)$$

The above problem can be converted to solving the generalized eigen-decomposition problem. As a summary, the iterative procedure for the projection of cross-view gaits with CMMF criterion is presented in Algorithm 2.

4.3. Coupled multi-linear discriminant analysis criterion (CMDA)

General tensor discriminant analysis (GTDA) is a linear discriminant analysis extended in the tensor space, which introduce supervised information into multi-linear analysis. Motivated by GTDA, the mode n intra-class scatter and inter-class scatter matrices in the projected tensor space are defined as follows:

$$J_w(\mathbf{U}_n, \mathbf{V}_n) = \sum_{c=1}^C \left(\sum_{i=1}^{N_c} \|\mathbf{U}_n^\top \mathbf{X}_{i(n)}^{(c)} \tilde{\mathbf{U}}_n - \mathbf{U}_n^\top \bar{\mathbf{X}}_{(n)}^{(c)} \tilde{\mathbf{U}}_n\|_F^2 + \sum_{j=1}^{N_c} \|\mathbf{V}_n^\top \mathbf{Y}_{j(n)}^{(c)} \tilde{\mathbf{V}}_n - \mathbf{V}_n^\top \bar{\mathbf{Y}}_{(n)}^{(c)} \tilde{\mathbf{V}}_n\|_F^2 \right),$$

$$J_b(\mathbf{U}_n, \mathbf{V}_n) = \sum_{c=1}^C (N_c \|\mathbf{U}_n^\top \bar{\mathbf{X}}_{(n)}^{(c)} \tilde{\mathbf{U}}_n - \mathbf{U}_n^\top \bar{\mathbf{X}}_{(n)} \tilde{\mathbf{U}}_n\|_F^2 + N_c \|\mathbf{V}_n^\top \bar{\mathbf{Y}}_{(n)}^{(c)} \tilde{\mathbf{V}}_n - \mathbf{V}_n^\top \bar{\mathbf{Y}}_{(n)} \tilde{\mathbf{V}}_n\|_F^2), \quad (24)$$

where $\bar{\mathbf{X}}_{(n)}^{(c)}$, $\bar{\mathbf{Y}}_{(n)}^{(c)}$ are the means of samples from the class c from two views θ and ϑ , respectively, $\bar{\mathbf{X}}_{(n)}$ and $\bar{\mathbf{Y}}_{(n)}$, are the means of total data sets from two views θ and ϑ respectively.

$$\bar{\mathbf{X}}_{(n)}^{(c)} = \frac{1}{N_c} \sum_{i=1}^{N_c} \mathbf{X}_{i(n)}^{(c)}, \bar{\mathbf{Y}}_{(n)}^{(c)} = \frac{1}{N_c} \sum_{j=1}^{N_c} \mathbf{Y}_{j(n)}^{(c)}, \bar{\mathbf{X}}_{(n)} = \frac{1}{C} \sum_{c=1}^C \bar{\mathbf{X}}_{(n)}^{(c)},$$

$$\bar{\mathbf{Y}}_{(n)} = \frac{1}{C} \sum_{c=1}^C \bar{\mathbf{Y}}_{(n)}^{(c)}. \quad (25)$$

Again we wish to construct a large scalar when the mode n intra-class scatter is small and the mode n inter-class scatter is large.

$$\{\mathbf{U}_n^*, \mathbf{V}_n^*\} = \operatorname{argmax}_{\mathbf{U}_n, \mathbf{V}_n} \frac{J_b(\mathbf{U}_n, \mathbf{V}_n)}{J_w(\mathbf{U}_n, \mathbf{V}_n)} = \frac{\sum_{c=1}^C (N_c \|\mathbf{U}_n^\top \bar{\mathbf{X}}_{(n)}^{(c)} \tilde{\mathbf{U}}_n - \mathbf{U}_n^\top \bar{\mathbf{X}}_{(n)} \tilde{\mathbf{U}}_n\|_F^2 + N_c \|\mathbf{V}_n^\top \bar{\mathbf{Y}}_{(n)}^{(c)} \tilde{\mathbf{V}}_n - \mathbf{V}_n^\top \bar{\mathbf{Y}}_{(n)} \tilde{\mathbf{V}}_n\|_F^2)}{\sum_{c=1}^C \left(\sum_{i=1}^{N_c} \|\mathbf{U}_n^\top \mathbf{X}_{i(n)}^{(c)} \tilde{\mathbf{U}}_n - \mathbf{U}_n^\top \bar{\mathbf{X}}_{(n)}^{(c)} \tilde{\mathbf{U}}_n\|_F^2 + \sum_{j=1}^{N_c} \|\mathbf{V}_n^\top \mathbf{Y}_{j(n)}^{(c)} \tilde{\mathbf{V}}_n - \mathbf{V}_n^\top \bar{\mathbf{Y}}_{(n)}^{(c)} \tilde{\mathbf{V}}_n\|_F^2 \right)} \quad (26)$$

(26) can be rewritten as

Algorithm 2: Projection of cross-view gaits with CMMF criterion.

Input: Two sets of training gait tensor samples $\{\mathcal{X}_i \in \mathbb{R}^{H_1 \times H_2 \times \dots \times H_N}, i = 1, 2, \dots, M\}$ and $\{\mathcal{Y}_j \in \mathbb{R}^{L_1 \times L_2 \times \dots \times L_N}, j = 1, 2, \dots, M\}$ from two views θ and ϑ , dimensionality of the transformed gait tensor $F_1 \times F_2 \times \dots \times F_N$, the maximum iteration T_{\max} ;

Output: The projection matrices $(\mathbf{U}_n, \mathbf{V}_n), n = 1, \dots, N$, the aligned tensors $\{\mathcal{A}_i \in \mathbb{R}^{F_1 \times \dots \times F_N}, i = 1, \dots, M\}$ and $\{\mathcal{B}_j \in \mathbb{R}^{F_1 \times \dots \times F_N}, j = 1, \dots, M\}$;

- 1 Initialize $\mathbf{U}_n^{(0)} = \mathbf{I}_{H_n}, \mathbf{V}_n^{(0)} = \mathbf{I}_{L_n}, n = 1, \dots, N$, where their superscript denotes iteration number;
- 2 Calculate $\tilde{\mathbf{W}}, \tilde{\mathbf{D}}_1, \tilde{\mathbf{D}}_2, \tilde{\mathbf{W}}, \tilde{\mathbf{D}}_1$ and $\tilde{\mathbf{D}}_2$ according to (17) and (20);
- 3 **while** not converged or $t \leq T_{\max}$ **do**
- 4 **for** mode $n = 1 : N$ **do**
- 5 Calculate
- 6 $\mathbf{X}_{i(n)} \tilde{\mathbf{U}}_n^{(t-1)} \leftarrow_n \mathcal{X}_i \times_1 (\mathbf{U}_1^{(t)})^\top \dots \times_{n-1} (\mathbf{U}_{n-1}^{(t)})^\top$
 $\times_{n+1} (\mathbf{U}_{n+1}^{(t-1)})^\top \dots \times_N (\mathbf{U}_N^{(t-1)})^\top$
- 7 $\mathbf{Y}_{i(n)} \tilde{\mathbf{U}}_n^{(t-1)} \leftarrow_n \mathcal{Y}_i \times_1 (\mathbf{V}_1^{(t)})^\top \dots \times_{n-1} (\mathbf{V}_{n-1}^{(t)})^\top$
 $\times_{n+1} (\mathbf{V}_{n+1}^{(t-1)})^\top \dots \times_N (\mathbf{V}_N^{(t-1)})^\top$.
- 8 Calculate auxiliary matrices $\mathbf{X}_{(n)}, \mathbf{Y}_{(n)}$ and $\mathbf{Z}_{(n)}$ as Step 7 in Algorithm. 1, $\tilde{\mathbf{G}}$ and $\tilde{\mathbf{C}}$ in (21);
- 9 Calculate the generalized eigen-decomposition problem on $(\mathbf{Z}_{(n)}(\tilde{\mathbf{G}} + \tau \mathbf{I})\mathbf{Z}_{(n)}^\top)^{-1} \mathbf{Z}_{(n)} \tilde{\mathbf{C}} \mathbf{Z}_{(n)}^\top$,
 $\mathbf{P}_n^{(t)*} = [(\mathbf{U}_n^{(t)})^\top \quad (\mathbf{V}_n^{(t)})^\top]^\top = [\mathbf{p}_1^{(n,t)}, \mathbf{p}_2^{(n,t)}, \dots, \mathbf{p}_{F_n}^{(n,t)}]$,
 where the eigenvectors $\mathbf{p}_1^{(n,t)}, \mathbf{p}_2^{(n,t)}, \dots, \mathbf{p}_{F_n}^{(n,t)}$ correspond to F_n smallest eigenvalues;
- 10 **end**
- 11 Check convergence $\sum_{n=1}^N \left\| \mathbf{P}_n^{(t)*} (\mathbf{P}_n^{(t-1)*})^\top - \mathbf{I} \right\| \leq \varepsilon$;
- 12 **end**
- 13 Project the high-dimensional gait tensor to matrices $(\mathbf{U}_n, \mathbf{V}_n), n = 1, \dots, N$, and obtain the low-dimensional tensors $\mathcal{A}_i = \mathcal{X}_i \times_1 \mathbf{U}_1^\top \times_2 \mathbf{U}_2^\top \dots \times_N \mathbf{U}_N^\top$ and $\mathcal{B}_j = \mathcal{Y}_j \times_1 \mathbf{V}_1^\top \times_2 \mathbf{V}_2^\top \dots \times_N \mathbf{V}_N^\top$.

Then, (27) reduces to

$$\arg \min_{\mathbf{P}_n} J(\mathbf{P}_n) = \frac{\text{Tr}(\mathbf{P}_n^\top \mathbf{Z}_{(n)} \mathbf{G} \mathbf{Z}_{(n)}^\top \mathbf{P}_n)}{\text{Tr}(\mathbf{P}_n^\top \mathbf{Z}_{(n)} \mathbf{G}' \mathbf{Z}_{(n)}^\top \mathbf{P}_n)}. \quad (29)$$

The above problem can be converted to solve the generalized eigen-decomposition problem. As a summary, the iterative procedure for the projection of cross-view gaits with CMDA criterion is presented in Algorithm 3.

4.4. Analysis

Complexity analysis. For all of the three criteria, the time complexity includes three aspects: n -mode projection in step 5 of Algorithms 1 and 2, step 4 in Algorithm 3, the scatter calculation and general eigenvalue decomposition. In each iteration, the time complexity of the n -mode projection and general eigenvalue decomposition can be respectively computed by $O(M \sum_{n=1}^N H_n^2 \tilde{H}_n + L_n^2 \tilde{L}_n)$ and $O(\sum_{n=1}^N (H_n + L_n)^3)$ for all the three variations, where $H_n = \prod_{i=1, i \neq n}^N H_i$ and $L_n = \prod_{i=1, i \neq n}^N L_i$. For CMLP, the complexity to compute the scatter matrices in the optimization procedure is $O(\sum_{i=1}^N 2MK_n(H_n + L_n)(2MK_n + H_n + L_n))$. CMMF and CMDA consider both inter-class scatter and intra-class scatter, which double the computation cost, i.e., $O(\sum_{i=1}^N 4MK_n(H_n + L_n)(2MK_n + H_n + L_n))$. For simplicity, we assume that $H_n = L_n = (\prod_{n=1}^N H_n)^{1/n} = (\prod_{n=1}^N L_n)^{1/n} = I$, thus the computing complexity of CMLP can be denoted as $O(2MN \cdot I^{(N+1)} + 8N \cdot I^3 + 8MNI(M \cdot F(N-1) + I) * F(N-1))$ where it is assumed the dimensionality of transformed tensor $\forall F_n = F, n = 1, 2, \dots, N$. Similarly, CMMF and CMDA double the computational complexity.

Relationship analysis. All of the proposed three variations under the unified framework aim to learn a shared multi-linear subspace in which the data biases caused by view differences are alleviated. Then, gait across different views can be directly measured in the shared subspace. However, the three criteria construct different manifold structure in the subspace. As in Table 2, different alignment representation are embedded into the manifold graph, which makes the subspace discriminable. In detail, CMLP constructs a graph incorporating neighbourhood information across gait views spanned on the dataset. From perspective of Laplacian, we asym-

$$\arg \min_{\mathbf{U}_n, \mathbf{V}_n} J(\mathbf{U}_n, \mathbf{V}_n)$$

$$= \frac{\text{Tr} \left(\begin{bmatrix} \mathbf{U}_n \\ \mathbf{V}_n \end{bmatrix}^\top \begin{bmatrix} \mathbf{X}_{(n)} \tilde{\mathbf{U}}_n & \mathbf{0} \\ \mathbf{0} & \mathbf{Y}_{(n)} \tilde{\mathbf{V}}_n \end{bmatrix} \begin{bmatrix} \mathbf{I} - \mathbf{K} \otimes \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} - \mathbf{K} \otimes \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{X}_{(n)} \tilde{\mathbf{U}}_n & \mathbf{0} \\ \mathbf{0} & \mathbf{Y}_{(n)} \tilde{\mathbf{V}}_n \end{bmatrix}^\top \begin{bmatrix} \mathbf{U}_n \\ \mathbf{V}_n \end{bmatrix} \right)}{\text{Tr} \left(\begin{bmatrix} \mathbf{U}_n \\ \mathbf{V}_n \end{bmatrix}^\top \begin{bmatrix} \mathbf{X}_{(n)} \tilde{\mathbf{U}}_n & \mathbf{0} \\ \mathbf{0} & \mathbf{Y}_{(n)} \tilde{\mathbf{V}}_n \end{bmatrix} \begin{bmatrix} \mathbf{K} \otimes \mathbf{I} - \frac{1}{M}(\mathbf{e}\mathbf{e}^\top) \otimes \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{K} \otimes \mathbf{I} - \frac{1}{M}(\mathbf{e}\mathbf{e}^\top) \otimes \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{X}_{(n)} \tilde{\mathbf{U}}_n & \mathbf{0} \\ \mathbf{0} & \mathbf{Y}_{(n)} \tilde{\mathbf{V}}_n \end{bmatrix}^\top \begin{bmatrix} \mathbf{U}_n \\ \mathbf{V}_n \end{bmatrix} \right)}. \quad (27)$$

The detailed mathematical deductions of (27) are put into the Appendix D.

To simplify (27), several auxiliary matrices are defined as follows

$$\mathbf{P}_n = \begin{bmatrix} \mathbf{U}_n \\ \mathbf{V}_n \end{bmatrix}, \mathbf{Z}_{(n)} = \begin{bmatrix} \mathbf{X}_{(n)} \tilde{\mathbf{U}}_n & \mathbf{0} \\ \mathbf{0} & \mathbf{Y}_{(n)} \tilde{\mathbf{V}}_n \end{bmatrix},$$

$$\mathbf{G} = \begin{bmatrix} \mathbf{I} - \mathbf{K} \otimes \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} - \mathbf{K} \otimes \mathbf{I} \end{bmatrix},$$

$$\mathbf{G}' = \begin{bmatrix} \mathbf{K} \otimes \mathbf{I} - \frac{1}{M}(\mathbf{e}\mathbf{e}^\top) \otimes \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{K} \otimes \mathbf{I} - \frac{1}{M}(\mathbf{e}\mathbf{e}^\top) \otimes \mathbf{I} \end{bmatrix}. \quad (28)$$

metrically map the tensor data points from different gait views into a shared subspace by a couple of multi-linear projection matrices, which the multi-linear transformation optimally preserves local neighbourhood structure in the shared subspace. CMMF also blurs the data bias across views by optimally learning asymmetrical projections based on local neighbourhood information. However, CMMF is a multi-linear tensor discriminative model, which encodes both the intra-class compactness and inter-class separability with local neighbourhood relationships in the manifold. CMDA is also a discriminative model which make the samples from the same class compact and samples from the different class separate in the shared subspace. In contrast to CMLP and CMMF, CMDA encodes not only local neighbourhood information but also global relationships between classes. Though the three criteria follow the same framework, each of them build unique manifold.

Algorithm 3: Projection of cross-view gaits with CMMF criterion.

Input: Two sets of training gait tensor samples $\{\mathcal{X}_i \in \mathbb{R}^{H_1 \times H_2 \times \dots \times H_N}, i = 1, 2, \dots, M\}$ and $\{\mathcal{Y}_j \in \mathbb{R}^{L_1 \times L_2 \times \dots \times L_N}, j = 1, 2, \dots, M\}$ from two views θ and ϑ , dimensionality of the transformed gait tensor $F_1 \times F_2 \times \dots \times F_N$, the maximum iteration T_{\max} ;

Output: The projection matrices $(\mathbf{U}_n, \mathbf{V}_n), n = 1, \dots, N$, the aligned tensors $\{\mathcal{A}_i \in \mathbb{R}^{F_1 \times \dots \times F_N}, i = 1, \dots, M\}$ and $\{\mathcal{B}_j \in \mathbb{R}^{F_1 \times \dots \times F_N}, j = 1, \dots, M\}$;

- 1 Initialize $\mathbf{U}_n^{(0)} = \mathbf{I}_{H_n}, \mathbf{V}_n^{(0)} = \mathbf{I}_{L_n}, n = 1, \dots, N$, where their superscript denotes iteration number.
- 2 **while** not converged or $t < T_{\max}$ **do**
- 3 **for** mode $n = 1 : N$ **do**
- 4 Calculate
- 5 $\mathbf{X}_{i(n)} \tilde{\mathbf{U}}_n^{(t-1)} \leftarrow_n \mathcal{X}_i \times_1 (\mathbf{U}_1^{(t)})^\top \dots \times_{n-1} (\mathbf{U}_{n-1}^{(t)})^\top$
 $\times_{n+1} (\mathbf{U}_{n+1}^{(t-1)})^\top \dots \times_N (\mathbf{U}_N^{(t-1)})^\top$
- 6 $\mathbf{Y}_{i(n)} \tilde{\mathbf{U}}_n^{(t-1)} \leftarrow_n \mathcal{Y}_j \times_1 (\mathbf{V}_1^{(t)})^\top \dots \times_{n-1} (\mathbf{V}_{n-1}^{(t)})^\top$
 $\times_{n+1} (\mathbf{V}_{n+1}^{(t-1)})^\top \dots \times_N (\mathbf{V}_N^{(t-1)})^\top$
- 7 Calculate auxiliary matrices $\mathbf{X}_{(n)}, \mathbf{Y}_{(n)}, \mathbf{Z}_{(n)}, \mathbf{G}, \mathbf{G}'$ in (28);
- 8 Calculate the generalized eigen-decomposition problem on $(\mathbf{Z}_{(n)}(\mathbf{G}' + \tau \mathbf{I})\mathbf{Z}_{(n)}^\top)^{-1} \mathbf{Z}_{(n)} \mathbf{G} \mathbf{Z}_{(n)}^\top$,
 $\mathbf{P}_n^{(t)*} = [(\mathbf{U}_n^{(t)})^\top \quad (\mathbf{V}_n^{(t)})^\top]^\top = [\mathbf{p}_1^{(n,t)}, \mathbf{p}_2^{(n,t)}, \dots, \mathbf{p}_{F_n}^{(n,t)}]$,
 where the eigenvectors $\mathbf{p}_1^{(n,t)}, \mathbf{p}_2^{(n,t)}, \dots, \mathbf{p}_{F_n}^{(n,t)}$ correspond to F_n smallest eigenvalues;
- 9 **end**
- 10 Check convergence $\sum_{n=1}^N \left\| \mathbf{P}_n^{(t)*} \left(\mathbf{P}_n^{(t-1)*} \right)^\top - \mathbf{I} \right\| \leq \varepsilon$;
- 11 **end**
- 12 Project the high-dimensional gait tensor to matrices $(\mathbf{U}_n, \mathbf{V}_n), n = 1, \dots, N$, and obtain the low-dimensional tensors $\mathcal{A}_i = \mathcal{X}_i \times_1 \mathbf{U}_1^\top \times_2 \mathbf{U}_2^\top \dots \times_N \mathbf{U}_N^\top$ and $\mathcal{B}_j = \mathcal{Y}_j \times_1 \mathbf{V}_1^\top \times_2 \mathbf{V}_2^\top \dots \times_N \mathbf{V}_N^\top$.

5. Experiments

In this section, the effectiveness of the proposed framework is assessed by extensive experiments conducted on two databases: (1) CASIA(B) gait database and (2) OU-ISIR large population gait database. Videos/images in both databases are collected from multi-view cameras; therefore they are most applicable to evaluate the performance of cross-view gait recognition. For each gait sequence, we use dual-ellipse fitting approach [53] to detect gait periodicity. Then, we adopt gait energy image (GEI) [54] as the gait feature in a gait cycle. After that, GaborD-based, GaborS-based and GaborSD-based gait representations [45] are used as recognized feature. GaborD and GaborS features are respectively obtained by the direction summation and scale summation of Gabor features, and GaborSD feature is obtained by both direction and scale summation of Gabor features. For these three gait representations, the dimension of MPCA is chosen according to a 98% energy criterion as in [43]. Furthermore, the proposed three criteria of tensorial coupled mappings are used for extracting features. Finally, the nearest neighbour classifier is used for classification.

5.1. Databases

5.1.1. CASIA(B) gait database

The CASIA(B) gait database contains 13,640 sequences of 124 subjects. For each subject, gaits are recorded by the cameras from

11 views, i.e., 0°, 18°, 36°, 54°, 72°, 90°, 108°, 126°, 144°, 162°, and 180°. There are 10 gait sequences for each subject: 6 samples under normal condition, 2 samples walking with coats and 2 samples carrying bags. Since the samples of walking with coats and carrying bags are too limited to calculate intra-class scatter, all the 6 samples containing normal walk for each subject are selected for the experiments in this paper. These normal walk samples are divided into training and testing sets, and the first 64 subjects are used for training and the rest 60 subjects are used to test the performance of gait recognition approaches under the view change. All the GEIs are cropped and normalized to 64×64 pixels. Fig. 1 shows the GEIs from 11 viewing angles. After Gabor filter with the above-mentioned 5 scales and 8 directions, the sizes of GaborD-based, GaborS-based and GaborSD-based gait features are $64 \times 64 \times 8$, $64 \times 64 \times 5$ and 64×64 , respectively.

5.1.2. OU-ISIR large population gait database

The OU-ISIR large population gait database has been released recently. It contains 1912 subjects whose gait sequences are captured from 4 different observation angles of 55°, 65°, 75° and 85°. We randomly divide OU-ISIR gait database into two sets equally for 5 times. Therefore, 956 subjects compose the training set, and the rest 956 subjects make up the testing set. In our experiments, the size of GEIs is aligned to 64×44 pixels. Fig. 2 shows GEIs from 4 different views. Compared with the CASIA(B) database, this database contains more subjects with wider range of age variations but narrower range of view variations. The sizes of GaborD-based, GaborS-based and GaborSD-based gait features are $64 \times 44 \times 8$, $64 \times 44 \times 5$ and 64×44 respectively in our experiments.

5.2. Performance evaluation

5.2.1. Evaluation on the CASIA(B) gait database

In this section, we evaluate the effectiveness of the proposed GaborD-CMLP, GaborD-CMMF, GaborD-CMDA, GaborS-CMLP, GaborS-CMMF, GaborS-CMDA, GaborSD-CMLP, GaborSD-CMMF and GaborSD-CMDA using the CASIA(B) gait database. The numbers of iteration of CMLP, CMMF and CMDA are all set to 5. For fair comparison, the recognition accuracies with optimal parameters are reported. Table 3 illustrates the recognition rates of probe view 54°. It can be seen that when the difference increases between the probe and the gallery sample is 18°, the recognition rates of the proposed 9 methods, are above 90%, and the recognition rate of each method is very close. In addition, the recognition rates decrease as the view difference between the gallery and probe views, but slow down the downward trend at a symmetrical gallery view angle.

We compare the proposed methods with state-of-the-art cross-view gait recognition methods including GEI [54], CMCC [31], VTM+QM [23], SVD [24], SVR [32], MvDA [55] and GEI + Deep CNNs [40]. GEI is a kind of spatial-temporal template, which is the state-of-the-art feature representation in gait recognition. We utilized GEI to characterize gait patterns in all of the compared approaches. For CMCC, the computing complexity includes four parts, i.e., bipartite graph modelling, bipartite graph multipartitioning, correlation optimization by CCA and linear approximation processes, which take $O(MI^2)$, $O(I^3)$, $O(I^3)$ and $O(MI)$, respectively, where I' is the dimension of the GEI segment. The complexity of VTM+QM and SVD is dominated by SVD factorization, which is $O(2M^2I)$. For SVR, the complexity is related to the number of support vectors S_v , which is $O(S_v^3 + MS_v^2 + MIS_v)$ with upper bound $O(M^2I)$. For MvDA, computational costs are mainly from matrix inversion and eigenvalue decomposition, and both are $O(8I^3)$. Table 3 tabulates comparison results under Probe view 54°. As shown in Table 3, the proposed methods consistently outperform other state-of-the-art methods. The average recognition rate of the



Fig. 1. GEIs from the CASIA(B) database.

Table 3

Comparison on various methods on CASIA(B) under Probe view 54° (the best record under each Gallery view is marked in bold, the second best is marked by underline '_').

Gallery view	0°	18°	36°	72°	90°	108°	126°	144°	162°	180°
GEI [54]	0.04	0.09	0.30	0.22	0.18	0.17	0.38	0.19	0.02	0.03
CMCC [31]	0.24	0.65	0.97	0.95	0.63	0.53	0.48	0.34	0.23	0.22
VTM+QM [23]	0.21	0.67	0.96	0.97	0.70	0.66	0.39	0.33	0.20	0.22
SVD [24]	0.13	0.46	0.87	0.81	0.49	0.31	0.27	0.19	0.18	0.16
SVR [32]	0.22	0.64	0.95	0.93	0.59	0.51	0.42	0.27	0.20	0.21
MvDA [55]	0.28	0.70	<u>0.98</u>	0.97	0.72	0.68	0.53	0.42	0.25	0.28
GEI + CNNs [40]	0.5	0.67	0.99	0.99	0.93	0.90	0.79	0.73	0.63	0.62
GaborD-CMLP	0.37	0.78	0.99	0.97	<u>0.82</u>	0.69	0.64	0.55	0.33	0.32
GaborD-CMMF	<u>0.43</u>	0.72	0.99	<u>0.98</u>	<u>0.82</u>	0.71	0.66	0.56	0.37	0.30
GaborD-CMDA	0.31	0.51	0.92	0.86	0.57	0.58	0.53	0.30	0.25	0.25
GaborS-CMLP	0.38	<u>0.74</u>	0.99	0.97	0.79	0.73	0.73	0.51	0.38	0.33
GaborS-CMMF	0.42	0.78	0.99	<u>0.98</u>	0.81	0.76	0.65	0.53	0.37	<u>0.35</u>
GaborS-CMDA	0.31	0.66	0.96	0.95	0.74	0.61	0.61	0.42	0.34	0.29
GaborSD-CMLP	0.36	0.70	<u>0.98</u>	0.96	<u>0.82</u>	0.73	0.70	<u>0.57</u>	0.37	0.33
GaborSD-CMMF	0.37	0.72	0.99	<u>0.98</u>	0.81	<u>0.77</u>	<u>0.76</u>	<u>0.57</u>	<u>0.42</u>	0.33
GaborSD-CMDA	0.25	0.52	0.93	0.90	0.64	0.59	0.54	0.42	0.39	0.29

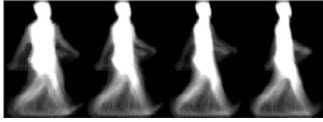


Fig. 2. GEIs from the OU-ISIR database.

proposed methods is 14%, 13%, 27%, 16% and 7% higher than CMCC, VTM+QM, SVD, SVR and MvDA. Particularly, the proposed methods achieve more remarkable increase under large view differences. Compared to GEI + CNNs, the proposed approaches achieve equivalent performances under small view variance, i.e., 18°. But, the performances of the proposed approaches are still posterior when the view differences are enlarged. Though GEI + CNNs achieves better performances, it demands large volume of labelled data and high computational costs which is not practical in gait recognition. In contrast, our proposed approaches overcome the drawbacks which aims at small size sample problem and reduces dependences on computational sources. Especially, our framework can achieve equal performance under small view variances.

5.2.2. Evaluation on the OU-ISIR large population gait database

Since OU-ISIR gait database contains two gait sequences per subject, each query subject's one angle view GEIs are used as gallery samples, and GEIs under other angle view are used as query samples. We repeat the experiment by swapping the samples in the training and testing sets, therefore, we test the recognition rates for 10 times for each cross-view as [55]. The average recognition rates over these 10 runs are reported in this paper.

We evaluate the accuracies of the proposed GaborD-CMLP, GaborD-CMMF, GaborD-CMDA, GaborS-CMLP, GaborS-CMMF, GaborS-CMDA, GaborSD-CMLP, GaborSD-CMMF and GaborSD-CMDA, and also the effect of view angle variations. Table 4 illustrates the recognition rates with the gallery view and probe view of 55°, 65°, 75° and 85°. We observe that the recognition rates decrease monotonically as the view difference between the gallery and probe views. The lowest recognition rate can reach more than 96% even when the maximum view difference is 30° (85° – 55°).

We also compare the proposed methods with those above-mentioned state-of-the-art cross-view gait recognition methods using OU-ISIR database. Because the performance results of the proposed method are very close, the data is retained to the third digit after the decimal point, which is easier to distinguish between the pros and cons of them. Namely, we report the results accurate to one-thousandth for the OU-ISIR database. Table 4 presents the results in terms of recognition rate for the different methods. We can see that the proposed methods yield average increases of 13%, 17%, 19%, 20% and 5% as compared to CMCC, VTM+QM, SVD, SVR and MvDA. These results, again, corroborate the useful tensor representation framework in addressing the cross-view gait recognition. It worth to point out that the proposed approaches achieve slightly prior performance compared to GEI + CNNs. This verifies the conclusion on CASIA(B) dataset that the proposed framework is effective when view difference is relevant small.

5.2.3. Ablation study

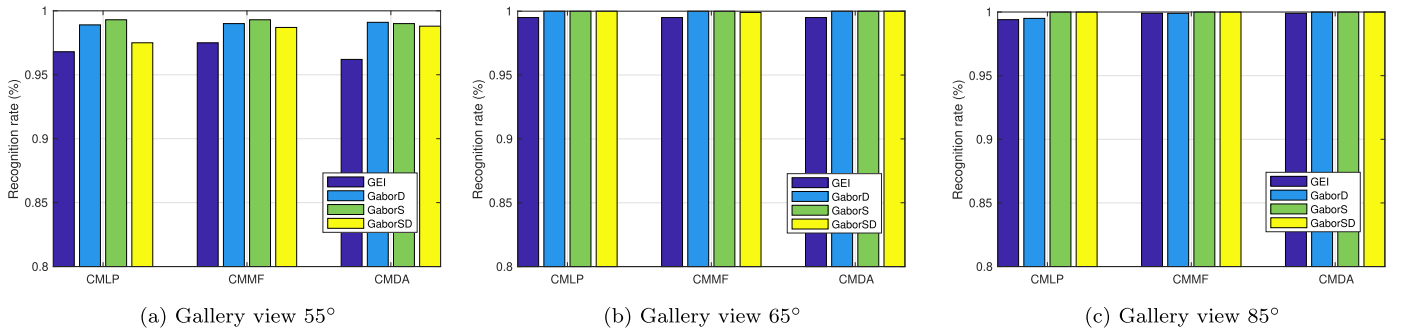
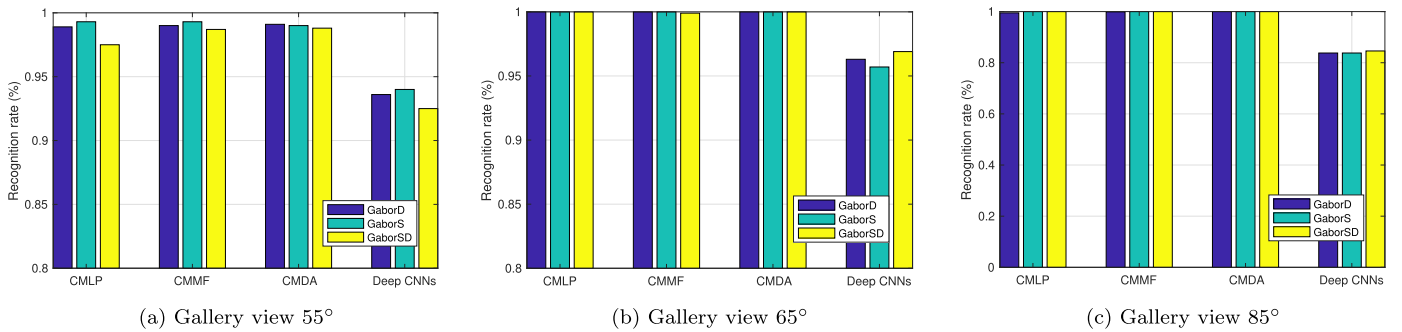
In this section, we evaluate the effectiveness of each component, i.e., Gabor-based features and tensor-based coupled metric learning framework. Fig. 3 compares identification accuracies versus various tensorial coupled mapping methods using GEI and Gabor-based features on OU-ISIR large population gait dataset. We show the recognition results of various gallery views under fixed probe view 75° in sub-figure (a)–(c). From the figures, it can be seen that Gabor-based features achieve superior performances than GEI in most cases when the tensorial coupled mapping approach is fixed. This verifies that Gabor-based feature features boost the performances compared vanilla GEI. It is reasonable since Gabor-based features are learned from GEIs and sensitive to gait pattern variations. The conclusion is also verified in many deep learning studies [56,57], and they prove that CNN learns Gabor features in shallow layers.

Fig. 4 illustrates the effectiveness of the proposed tensorial coupled mapping framework. Sub-figure (a)–(c) show the recognition results of various gallery views, i.e., 55°, 65° and 85°, when the probe view is set to 75°. It is easy to observe that the proposed approaches under the unified tensorial coupled mapping framework outperform the deep CNNs with a large margin when com-

Table 4

Comparison on various methods on OU-ISIR (the best record under each gallery view is marked in bold, the second best is marked by underline ' _ ').

Probe view (°)	55			65			75			85		
Gallery view (°)	65	75	85	55	75	85	55	65	85	55	65	75
GEI [54]	0.284	0.058	0.277	0.277	0.670	0.195	0.507	0.640	0.969	0.262	0.207	0.969
CMCC [31]	0.968	0.785	0.646	<u>0.974</u>	0.963	0.826	0.800	0.975	0.969	0.749	0.785	0.965
VTM+QM [23]	0.941	0.704	0.491	0.957	0.966	0.785	0.756	0.971	0.964	0.555	0.838	0.978
SVD [24]	0.932	0.704	0.523	0.923	0.936	0.771	0.774	0.940	0.947	0.523	0.763	0.925
SVR [32]	0.936	0.710	0.531	0.940	0.943	0.720	0.753	0.943	0.941	0.511	0.711	0.938
MvDA [55]	0.975	0.922	0.858	<u>0.974</u>	<u>0.984</u>	0.949	0.925	0.984	0.985	0.877	0.957	<u>0.988</u>
GEI+CNNs [40]	0.983	0.960	0.805	0.963	0.973	0.833	0.942	0.978	0.851	0.900	0.960	0.984
GaborD-CMLP	0.999	0.990	0.972	1.000	1.000	<u>0.996</u>	0.989	1.000	0.995	0.969	<u>0.996</u>	1.000
GaborD-CMMF	<u>0.998</u>	<u>0.991</u>	0.970	1.000	1.000	<u>0.996</u>	0.990	1.000	<u>0.999</u>	<u>0.971</u>	0.994	1.000
GaborD-CMDA	<u>0.998</u>	0.992	0.967	1.000	1.000	0.995	<u>0.991</u>	1.000	1.000	0.967	0.994	1.000
GaborS-CMLP	0.999	0.992	<u>0.971</u>	1.000	1.000	0.997	0.993	1.000	1.000	0.972	<u>0.996</u>	1.000
GaborS-CMMF	0.999	0.992	0.972	1.000	1.000	0.997	0.993	1.000	1.000	0.970	0.997	1.000
GaborS-CMDA	<u>0.998</u>	0.989	0.968	1.000	1.000	0.997	0.990	1.000	1.000	0.963	0.997	1.000
GaborSD-CMLP	<u>0.998</u>	0.989	<u>0.971</u>	1.000	1.000	0.997	0.975	1.000	1.000	0.970	0.993	1.000
GaborSD-CMMF	0.999	<u>0.991</u>	0.965	1.000	1.000	<u>0.996</u>	0.987	<u>0.999</u>	1.000	0.966	0.994	1.000
GaborSD-CMDA	<u>0.998</u>	0.984	0.967	1.000	1.000	0.994	0.988	1.000	1.000	0.964	0.994	1.000

**Fig. 3.** A comparison of different features on the OU-ISIR database.**Fig. 4.** An illustration of the proposed tensorial coupled mapping framework with Gabor-based features on the OU-ISIR database.

binning with Gabor-based features. It demonstrates the effectiveness of our proposed tensorial coupled mapping framework. In another aspects, it exposes the shortage of deep CNNs that it is easy to be overfitting when the training data is limited. In particular, gait data is simple which also restricts the depth of deep CNNs [40]. It is worth to noting that GEI + CNNs achieves better performance than Gabor-based features + CNNs as in Table 4. We believe that the reason is two-fold: one is that data-driven learned filters using deep CNNs are more appropriate than hand-crafted filters using Gabor; the other one is that it is easier to overfit for Gabor-based features than GEI when using the same depth CNN framework, because the deep CNNs learn Gabor features in shallow layers [56].

Thus it is appropriate to combine Gabor-based features and the proposed tensorial coupled mapping framework together. Each component improves performance of the whole gait recognition system.

5.3. Discussion

Based on the comparative experiments on the CASIA(B) and OU-ISIR gait databases, we discuss and analyse the effects of Gabor gait representation and three criteria of tensorial coupled mappings.

- (1) The experiments in Section 5.2 show that the GaborS-based representation performs slightly better than GaborD-based representations which somewhat outperforms the GaborSD representation for cross-view gait recognition. This observation is consistent with [45]. GaborS and GaborD benefit cross-view gait recognition with Gabor functions over scales and directions representation.
- (2) CMLP criterion has the ability to learn the essential gait manifold structure. By minimizing the distance between gait tensor data under two different observation angles for the

identical subject, simultaneously suppressing the similarity of different subjects, as defined in (10), CMLP algorithm can enlarge the discrimination between different subjects, and meanwhile compact the variations of the same subject. Therefore, GaborD-CMLP, GaborS-CMLP and GaborSD-CMLP generally achieve good performance.

- (3) CMMF criterion minimizes the ratio of intra-class similarity to the inter-class similarity, which ensures the intra-class compactness and inter-class separability. Compared with CMLP criterion, CMMF criterion relies on Fisher discrimination which is more conducive to classification. Therefore, CMMF criterion yields better recognition performance than CMLP criterion.
- (4) Due to the lack of constrained relationship between cross-view gait tensor data, CMDA criterion is worst among the three criteria. Still, it is far superior to other state-of-the-art methods.

6. Conclusion

Cross-view gait recognition is a challenging task because the appearance change caused by view variation. We handle this problem by designing a general tensor representation framework that employs coupled metric learning from cross-view gait tensor data. First, we use GaborD, GaborS and GaborSD to extract Gabor feature with different scales and directions from GEIs. To enhance discrimination between different subjects, and meanwhile compact the variations of the same subject, three criteria of tensorial coupled mappings are proposed to project Gabor-based representations to a common subspace for recognition. Extensive experiments conducted on CASIA(B) and OU-ISIR gait database demonstrate the proposed methods are superior to other state-of-the-art methods. Moreover the proposed methods achieve slightly prior performance compared to GEI + CNNs when view difference is relevant small. When view difference is large, the advantage of CNNs may be more obvious. In the future, we will try a CNN representation framework to select view-invariant gait features for large cross-view gait recognition.

Acknowledgment

This project is supported by the National Key R&D Program of China (Grant No. 2018YFC0831001), the [Natural Science Foundation of China](#) (Grant No. 61571275, 61573248, 61802267, 61201370), the Young Scholars Program of Shandong University, the [Natural Science Foundation of Guangdong Province](#) through the Grant 2017A030313367, and Shenzhen Municipal Science and Technology Innovation Council (Grant No. JCYJ20150324141711637, JCYJ20170302153434048).

Appendix A. Similarity between \mathcal{A}_i and \mathcal{B}_i

$$\begin{aligned} \text{sim}(\mathcal{A}_i, \mathcal{B}_j) &= \|\mathcal{A}_i - \mathcal{B}_j\|_F^2 w_{ij} \\ &= \text{Tr} \left((\mathbf{U}_n^\top \mathbf{X}_{i(n)} \tilde{\mathbf{U}}_n - \mathbf{V}_n^\top \mathbf{Y}_{j(n)} \tilde{\mathbf{V}}_n) \right. \\ &\quad \left. (\mathbf{U}_n^\top \mathbf{X}_{i(n)} \tilde{\mathbf{U}}_n - \mathbf{V}_n^\top \mathbf{Y}_{j(n)} \tilde{\mathbf{V}}_n)^\top w_{ij} \right) \\ &= \text{Tr} \left(\begin{bmatrix} \mathbf{U}_n \\ \mathbf{V}_n \end{bmatrix}^\top \begin{bmatrix} \mathbf{X}_{i(n)} \tilde{\mathbf{U}}_n & \mathbf{0} \\ \mathbf{0} & \mathbf{Y}_{j(n)} \tilde{\mathbf{V}}_n \end{bmatrix} \right. \\ &\quad \left. \begin{bmatrix} w_{ij} \cdot \mathbf{I} & -w_{ij} \cdot \mathbf{I} \\ -w_{ij} \cdot \mathbf{I} & w_{ij} \cdot \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{X}_{i(n)} \tilde{\mathbf{U}}_n & \mathbf{0} \\ \mathbf{0} & \mathbf{Y}_{j(n)} \tilde{\mathbf{V}}_n \end{bmatrix}^\top \begin{bmatrix} \mathbf{U}_n \\ \mathbf{V}_n \end{bmatrix} \right) \end{aligned}$$

Appendix B. Derivation of mode- n objective function of (11)

The mode- n objective function of (11) is obtained as follows:

$$\begin{aligned} &\arg \min_{\mathbf{U}_n, \mathbf{V}_n} J(\mathbf{U}_n, \mathbf{V}_n) \\ &= \sum_{i,j} \|\mathbf{U}_n^\top \mathbf{X}_{i(n)} \tilde{\mathbf{U}}_n - \mathbf{V}_n^\top \mathbf{Y}_{j(n)} \tilde{\mathbf{V}}_n\|_F^2 w_{ij} \\ &= \text{Tr} \left(\sum_{i,j} \begin{bmatrix} \mathbf{U}_n \\ \mathbf{V}_n \end{bmatrix}^\top \begin{bmatrix} \mathbf{X}_{i(n)} \tilde{\mathbf{U}}_n \\ -\mathbf{Y}_{j(n)} \tilde{\mathbf{V}}_n \end{bmatrix} \begin{bmatrix} \mathbf{X}_{i(n)} \tilde{\mathbf{U}}_n \\ -\mathbf{Y}_{j(n)} \tilde{\mathbf{V}}_n \end{bmatrix}^\top \begin{bmatrix} \mathbf{U}_n \\ \mathbf{V}_n \end{bmatrix} w_{ij} \right) \\ &= \text{Tr} \left(\begin{bmatrix} \mathbf{U}_n \\ \mathbf{V}_n \end{bmatrix}^\top \begin{bmatrix} \mathbf{X}_{(n)} \tilde{\mathbf{U}}_n (\mathbf{D}_1 \otimes \mathbf{I}) \tilde{\mathbf{U}}_n^\top \mathbf{X}_{(n)}^\top & -\mathbf{X}_{(n)} \tilde{\mathbf{U}}_n (\mathbf{W} \otimes \mathbf{I}) \tilde{\mathbf{V}}_n^\top \mathbf{Y}_{(n)}^\top \\ -\mathbf{Y}_{(n)} \tilde{\mathbf{V}}_n (\mathbf{W}^\top \otimes \mathbf{I}) \tilde{\mathbf{V}}_n^\top \mathbf{Y}_{(n)}^\top & \mathbf{Y}_{(n)} \tilde{\mathbf{V}}_n (\mathbf{D}_2 \otimes \mathbf{I}) \tilde{\mathbf{V}}_n^\top \mathbf{Y}_{(n)}^\top \end{bmatrix} \begin{bmatrix} \mathbf{U}_n \\ \mathbf{V}_n \end{bmatrix} \right) \\ &= \text{Tr} \left(\begin{bmatrix} \mathbf{U}_n \\ \mathbf{V}_n \end{bmatrix}^\top \begin{bmatrix} \mathbf{X}_{(n)} \tilde{\mathbf{U}}_n & \mathbf{0} \\ \mathbf{0} & \mathbf{Y}_{(n)} \tilde{\mathbf{V}}_n \end{bmatrix} \begin{bmatrix} \mathbf{D}_1 \otimes \mathbf{I} & -\mathbf{W} \otimes \mathbf{I} \\ -\mathbf{W}^\top \otimes \mathbf{I} & \mathbf{D}_2 \otimes \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{X}_{(n)} \tilde{\mathbf{U}}_n & \mathbf{0} \\ \mathbf{0} & \mathbf{Y}_{(n)} \tilde{\mathbf{V}}_n \end{bmatrix}^\top \begin{bmatrix} \mathbf{U}_n \\ \mathbf{V}_n \end{bmatrix} \right) \end{aligned}$$

Appendix C. The solution of (15)

Eq. (15) can be solved by a generalized eigen-decomposition of

$$\mathbf{Z}_{(n)} \mathbf{G} \mathbf{Z}_{(n)}^\top \mathbf{p} = \lambda (\mathbf{Z}_{(n)} \mathbf{Z}_{(n)}^\top) \mathbf{p}. \quad (\text{C.1})$$

Taking the eigenvectors $\mathbf{p}_1^{(n)}, \mathbf{p}_2^{(n)}, \dots, \mathbf{p}_{F_n}^{(n)}$ corresponding to F_n ($F_n \leq \max(H_n, L_n)$) smallest eigenvalues $\lambda_1^{(n)} \leq \lambda_2^{(n)} \leq \dots \leq \lambda_{F_n}^{(n)}$, the optimal solution of (22) can be represented as

$$\mathbf{P}_n = \begin{bmatrix} \mathbf{U}_n \\ \mathbf{V}_n \end{bmatrix} = [\mathbf{p}_1^{(n)}, \mathbf{p}_2^{(n)}, \dots, \mathbf{p}_{F_n}^{(n)}]. \quad (\text{C.2})$$

$\mathbf{Z}_{(n)} \mathbf{Z}_{(n)}^\top$ is usually non-invertible. In order to eliminate the singularity and to avoid over fitting, a regularizer $\tau \mathbf{I}$ is imposed on the item $\mathbf{Z}_{(n)} \mathbf{Z}_{(n)}^\top$, where $\mathbf{I} \in \mathbb{R}^{(H_n+L_n) \times (H_n+L_n)}$ and is a small positive constant, such as $\tau = 10^{-6}$ used in this paper. \mathbf{P}_n can be divided into two matrices \mathbf{U}_n and \mathbf{V}_n . \mathbf{U}_n corresponds to the 1st to H_n th rows of \mathbf{P}_n and \mathbf{V}_n corresponds to the $(H_n + 1)$ th to $(H_n + L_n)$ th rows of \mathbf{P}_n .

Appendix D. Derivation of the mode- n objective function of (27)

The denominator of (26) can be rewritten as

$$\begin{aligned} &J_w(\mathbf{U}_n, \mathbf{V}_n) \\ &= \text{Tr} \left(\begin{bmatrix} \sum_{c=1}^C \sum_{i=1}^M (\mathbf{U}_n^\top \mathbf{X}_{i(n)}^{(c)} \tilde{\mathbf{U}}_n - \mathbf{U}_n^\top \tilde{\mathbf{X}}_{(n)}^{(c)} \tilde{\mathbf{U}}_n) (\mathbf{U}_n^\top \mathbf{X}_{i(n)}^{(c)} \tilde{\mathbf{U}}_n - \mathbf{U}_n^\top \tilde{\mathbf{X}}_{(n)}^{(c)} \tilde{\mathbf{U}}_n)^\top \\ + \sum_{c=1}^C \sum_{j=1}^M (\mathbf{V}_n^\top \mathbf{Y}_{j(n)}^{(c)} \tilde{\mathbf{V}}_n - \mathbf{V}_n^\top \tilde{\mathbf{Y}}_{(n)}^{(c)} \tilde{\mathbf{V}}_n) (\mathbf{V}_n^\top \mathbf{Y}_{j(n)}^{(c)} \tilde{\mathbf{V}}_n - \mathbf{V}_n^\top \tilde{\mathbf{Y}}_{(n)}^{(c)} \tilde{\mathbf{V}}_n)^\top \end{bmatrix} \right) \\ &= \text{Tr} \left\{ \begin{bmatrix} \mathbf{U}_n \\ \mathbf{V}_n \end{bmatrix}^\top \begin{bmatrix} \mathbf{S}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_2 \end{bmatrix} \begin{bmatrix} \mathbf{U}_n \\ \mathbf{V}_n \end{bmatrix} \right\}, \quad (\text{D.1}) \end{aligned}$$

where $\mathbf{X}_{c(n)} = [\mathbf{X}_{1(n)}^{(c)}, \dots, \mathbf{X}_{M(n)}^{(c)}]$, for $c = 1, \dots, C$, $\mathbf{X}_{(n)} = [\mathbf{X}_{1(n)}, \dots, \mathbf{X}_{C(n)}]$, $\mathbf{K}_{ij} = \begin{cases} \frac{1}{N_c} & \text{if } \mathcal{X}_i(\mathcal{Y}_i) \text{ and } \mathcal{X}_j(\mathcal{Y}_j) \text{ both belong to the } c\text{-th class} \\ 0 & \text{else} \end{cases}$.

$$\begin{aligned}
\mathbf{S}_1 &= \sum_{c=1}^C \sum_{i=1}^{N_c} \left(\mathbf{X}_{i(n)}^{(c)} \tilde{\mathbf{U}}_n - \bar{\mathbf{X}}_{i(n)}^{(c)} \tilde{\mathbf{U}}_n \right) \left(\mathbf{X}_{i(n)}^{(c)} \tilde{\mathbf{U}}_n - \bar{\mathbf{X}}_{i(n)}^{(c)} \tilde{\mathbf{U}}_n \right)^\top \\
&= \sum_{c=1}^C \left(\mathbf{X}_{c(n)} \tilde{\mathbf{U}}_n \mathbf{X}_{c(n)}^\top - \frac{1}{N_c} \mathbf{X}_{c(n)} \tilde{\mathbf{U}}_n (\mathbf{e}_c \mathbf{e}_c^\top) \otimes \mathbf{I} \mathbf{X}_{c(n)} \tilde{\mathbf{U}}_n^\top \right) \\
&= \mathbf{X}_{(n)} \tilde{\mathbf{U}}_n (\mathbf{I} - \mathbf{K} \otimes \mathbf{I}) \mathbf{X}_{(n)} \tilde{\mathbf{U}}_n^\top, \tag{D.2}
\end{aligned}$$

$$\begin{aligned}
\mathbf{S}_2 &= \sum_{c=1}^C \sum_{j=1}^{N_c} \left(\mathbf{Y}_{j(n)}^{(c)} \tilde{\mathbf{V}}_n - \bar{\mathbf{Y}}_{j(n)}^{(c)} \tilde{\mathbf{V}}_n \right) \left(\mathbf{Y}_{j(n)}^{(c)} \tilde{\mathbf{V}}_n - \bar{\mathbf{Y}}_{j(n)}^{(c)} \tilde{\mathbf{V}}_n \right)^\top \\
&= \mathbf{Y}_{(n)} \tilde{\mathbf{V}}_n (\mathbf{I} - \mathbf{K} \otimes \mathbf{I}) \mathbf{Y}_{(n)} \tilde{\mathbf{V}}_n^\top. \tag{D.3}
\end{aligned}$$

Thus, the alignment expression is obtained as $J_w(\mathbf{U}_n, \mathbf{V}_n) =$

$$\begin{aligned}
&Tr \left(\begin{bmatrix} \mathbf{U}_n \\ \mathbf{V}_n \end{bmatrix}^\top \begin{bmatrix} \mathbf{X}_{(n)} \tilde{\mathbf{U}}_n & \mathbf{0} \\ \mathbf{0} & \mathbf{Y}_{(n)} \tilde{\mathbf{V}}_n \end{bmatrix} \begin{bmatrix} \mathbf{I} - \mathbf{K} \otimes \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} - \mathbf{K} \otimes \mathbf{I} \end{bmatrix} \right. \\
&\left. \begin{bmatrix} \mathbf{X}_{(n)} \tilde{\mathbf{U}}_n & \mathbf{0} \\ \mathbf{0} & \mathbf{Y}_{(n)} \tilde{\mathbf{V}}_n \end{bmatrix}^\top \begin{bmatrix} \mathbf{U}_n \\ \mathbf{V}_n \end{bmatrix} \right)
\end{aligned}$$

The numerator of (26) can be rewritten as

$$\begin{aligned}
J_b(\mathbf{U}_n, \mathbf{V}_n) &= Tr \left(\begin{bmatrix} \sum_{c=1}^C M(\mathbf{U}_n^\top \bar{\mathbf{X}}_{i(n)}^{(c)} \tilde{\mathbf{U}}_n - \mathbf{U}_n^\top \bar{\mathbf{X}}_{i(n)}^{(c)} \tilde{\mathbf{U}}_n) (\mathbf{U}_n^\top \bar{\mathbf{X}}_{i(n)}^{(c)} \tilde{\mathbf{U}}_n - \mathbf{U}_n^\top \bar{\mathbf{X}}_{i(n)}^{(c)} \tilde{\mathbf{U}}_n)^\top \\ + \sum_{c=1}^C M(\mathbf{V}_n^\top \bar{\mathbf{Y}}_{j(n)}^{(c)} \tilde{\mathbf{V}}_n - \mathbf{V}_n^\top \bar{\mathbf{Y}}_{j(n)}^{(c)} \tilde{\mathbf{V}}_n) (\mathbf{V}_n^\top \bar{\mathbf{Y}}_{j(n)}^{(c)} \tilde{\mathbf{V}}_n - \mathbf{V}_n^\top \bar{\mathbf{Y}}_{j(n)}^{(c)} \tilde{\mathbf{V}}_n)^\top \end{bmatrix} \right) \\
&= Tr \left\{ \begin{bmatrix} \mathbf{U}_n \\ \mathbf{V}_n \end{bmatrix}^\top \begin{bmatrix} \mathbf{C}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_2 \end{bmatrix} \begin{bmatrix} \mathbf{U}_n \\ \mathbf{V}_n \end{bmatrix} \right\}, \tag{D.4}
\end{aligned}$$

where

$$\begin{aligned}
\mathbf{C}_1 &= \sum_{c=1}^C N_c (\bar{\mathbf{X}}_{i(n)}^{(c)} \tilde{\mathbf{U}}_n - \bar{\mathbf{X}}_{i(n)}^{(c)} \tilde{\mathbf{U}}_n) (\bar{\mathbf{X}}_{i(n)}^{(c)} \tilde{\mathbf{U}}_n - \bar{\mathbf{X}}_{i(n)}^{(c)} \tilde{\mathbf{U}}_n)^\top \\
&= \sum_{c=1}^C \sum_{i,j=1}^{N_c} \frac{1}{N_c} (\mathbf{X}_{i(n)}^{(c)} \tilde{\mathbf{U}}_n) (\mathbf{X}_{j(n)}^{(c)} \tilde{\mathbf{U}}_n)^\top - M \bar{\mathbf{X}}_{i(n)} \tilde{\mathbf{U}}_n (\bar{\mathbf{X}}_{i(n)} \tilde{\mathbf{U}}_n)^\top \\
&= \mathbf{X}_{(n)} (\mathbf{K} \otimes \mathbf{I} - \frac{1}{M} (\mathbf{e} \mathbf{e}^\top) \otimes \mathbf{I}) \mathbf{X}_{(n)}^\top, \tag{D.5}
\end{aligned}$$

$$\begin{aligned}
\mathbf{C}_2 &= \sum_{c=1}^C N_c (\bar{\mathbf{Y}}_{j(n)}^{(c)} \tilde{\mathbf{V}}_n - \bar{\mathbf{Y}}_{j(n)}^{(c)} \tilde{\mathbf{V}}_n) (\bar{\mathbf{Y}}_{j(n)}^{(c)} \tilde{\mathbf{V}}_n - \bar{\mathbf{Y}}_{j(n)}^{(c)} \tilde{\mathbf{V}}_n)^\top \\
&= \mathbf{Y}_{(n)} (\mathbf{K} \otimes \mathbf{I} - \frac{1}{M} (\mathbf{e} \mathbf{e}^\top) \otimes \mathbf{I}) \mathbf{Y}_{(n)}^\top \tag{D.6}
\end{aligned}$$

Thus, the alignment expression is obtained as follows

$$J_b(\mathbf{U}_n, \mathbf{V}_n) = Tr \left(\begin{bmatrix} \begin{bmatrix} \mathbf{U}_n \\ \mathbf{V}_n \end{bmatrix}^\top \begin{bmatrix} \mathbf{X}_{(n)} \tilde{\mathbf{U}}_n & \mathbf{0} \\ \mathbf{0} & \mathbf{Y}_{(n)} \tilde{\mathbf{V}}_n \end{bmatrix} \\ \mathbf{K} \otimes \mathbf{I} - \frac{1}{M} (\mathbf{e} \mathbf{e}^\top) \otimes \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{K} \otimes \mathbf{I} - \frac{1}{M} (\mathbf{e} \mathbf{e}^\top) \otimes \mathbf{I} \\ \begin{bmatrix} \mathbf{X}_{(n)} \tilde{\mathbf{U}}_n & \mathbf{0} \\ \mathbf{0} & \mathbf{Y}_{(n)} \tilde{\mathbf{V}}_n \end{bmatrix}^\top \begin{bmatrix} \mathbf{U}_n \\ \mathbf{V}_n \end{bmatrix} \end{bmatrix} \right)$$

So the objective function for mode n can be expressed as (27).

References

- [1] X. Ben, C. Gong, P. Zhang, X. Jia, Q. Wu, W. Meng. Coupled Patch Alignment for Matching Cross-view Gait. IEEE Trans. Image Process. doi: 10.1109/TIP.2019.2894362.
- [2] Y. He, J. Zhang, H. Shan, L. Wang, Multi-task gans for view-specific feature learning in gait recognition, IEEE Trans. Inf. Forensics Secur. 14 (1) (2019) 102–113.

- [3] G. Ariyanto, M.S. Nixon, Model-based 3d gait biometrics, in: Proceedings of the International Joint Conference on Biometrics, 2011, pp. 1–7.
- [4] I.R. Vega, S. Sarkar, Statistical motion model based on the change of feature relationships: human gait-based recognition, IEEE Trans. Pattern Anal. Mach. Intell. 25 (10) (2003) 1323–1328.
- [5] C. Wang, J. Zhang, L. Wang, J. Pu, X. Yuan, Human identification using temporal information preserving gait template, IEEE Trans. Pattern Anal. Mach. Intell. 34 (11) (2012) 2164–2176.
- [6] X. Ben, C. Gong, P. Zhang, R. Yan, Q. Wu, W. Meng. Coupled Bilinear Discriminant Projection for Cross-view Gait Recognition. IEEE Trans. on Circuits Syst. Video Technol. doi: 10.1109/TCSVT.2019.2893736.
- [7] Y. Guan, C.T. Li, F. Roli, On reducing the effect of covariate factors in gait recognition: a classifier ensemble method, IEEE Trans. Pattern Anal. Mach. Intell. 37 (7) (2015) 1521–1528.
- [8] S. Yu, D. Tan, T. Tan, Modelling the effect of view angle variation on appearance-based gait recognition, in: Proceedings of the Asian Conference on Computer Vision, 2006, pp. 807–816.
- [9] S.D. Choudhury, T. Tjahjedi, Robust view-invariant multiscale gait recognition, Pattern Recognit. 48 (3) (2015) 798–811.
- [10] J. Luo, J. Tang, T. Tjahjedi, X. Xiao, Robust arbitrary view gait recognition based on parametric 3d human body reconstruction and virtual posture synthesis, Pattern Recognit. 60 (2016) 361–377.
- [11] X. Ben, P. Zhang, W. Meng, R. Yan, M. Yang, W. Liu, H. Zhang, On the distance metric learning between cross-domain gaits, Neurocomputing 208 (2016) 153–164.
- [12] J. Lu, G. Wang, P. Moulin, Human identity and gender recognition from gait sequences with arbitrary walking directions, IEEE Trans. Inf. Forensics Secur. 9 (1) (2014) 51–61.
- [13] W. Zhang, X. Yu, X. He, Learning bidirectional temporal cues for video-based person re-identification, IEEE Trans. Circuits Syst. Video Technol. (2017), doi:10.1109/TCSVT.2017.2718188.
- [14] B. Du, W. Xiong, J. Wu, L. Zhang, L. Zhang, D. Tao, Stacked convolutional denoising auto-encoders for feature representation, IEEE Trans. Cybern. 47 (4) (2017) 1017–1027, doi:10.1109/TCYB.2016.2536638.
- [15] C. Ding, D. Tao, Trunk-branch ensemble convolutional neural networks for video-based face recognition, IEEE Trans. Pattern Anal. Mach. Intell. 40 (4) (2018) 1002–1014.
- [16] S. Huang, Z. Xu, D. Tao, Y. Zhang, Part-stacked cnn for fine-grained visual categorization, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 1173–1182.
- [17] H. Lu, K.N. Plataniotis, A.N. Venetsanopoulos, A survey of multilinear subspace learning for tensor data, Pattern Recognit. 44 (7) (2011) 1540–1551.
- [18] S. Yu, D. Tan, T. Tan, A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition, in: Proceedings of the International Conference on Pattern Recognition, 2006, pp. 441–444.
- [19] H. Iwama, M. Okumura, Y. Makihara, Y. Yagi, The OU-ISIR gait database comprising the large population dataset and performance evaluation of gait recognition, IEEE Trans. Inf. Forensics Secur. 7 (5) (2012) 1511–1521.
- [20] G. Zhao, G. Liu, H. Li, M. Pietikainen, 3d gait recognition using multiple cameras, in: Proceedings of the International Conference on Automatic Face and Gesture Recognition, 2006, pp. 529–534.
- [21] K. Sugiura, Y. Makihara, Y. Yagi, Gait identification based on multi-view observations using omnidirectional camera, in: Proceedings of the Asian Conference on Computer Vision, 2007, pp. 452–461.
- [22] J. Tang, J. Luo, T. Tjahjedi, F. Guo, Robust arbitrary-view gait recognition based on 3d partial similarity matching, IEEE Trans. Image Process. 26 (1) (2016) 7–22.
- [23] D. Muramatsu, Y. Makihara, Y. Yagi, View transformation model incorporating quality measures for cross-view gait recognition, IEEE Trans. Cybern. 46 (7) (2016) 1602–1615.
- [24] W. Kusakunniran, Q. Wu, H. Li, J. Zhang, Multiple views gait recognition using view transformation model based on optimized gait energy image, in: Proceedings of the IEEE International Conference on Computer Vision Workshops, 2010, pp. 1058–1064.
- [25] W. Kusakunniran, Q. Wu, J. Zhang, H. Li, Multi-view gait recognition based on motion regression using multilayer perceptron, in: Proceedings of the International Conference on Pattern Recognition, 2010, pp. 2186–2189.
- [26] T. Connie, M.K.O. Goh, A.B.J. Teoh, A grassmannian approach to address view change problem in gait recognition, IEEE Trans. Cybern. 47 (6) (2017) 1395–1408.
- [27] D. Muramatsu, A. Shiraishi, Y. Makihara, M.Z. Uddin, Y. Yagi, Gait-based person recognition using arbitrary view transformation model, IEEE Trans. Image Process. 24 (1) (2015) 140–154.
- [28] S. Zheng, J. Zhang, K. Huang, R. He, T. Tan, Robust view transformation model for gait recognition, in: Proceedings of the IEEE International Conference on Image Processing, 2011, pp. 2073–2076.
- [29] Y. Makihara, R. Sagawa, Y. Mukaigawa, T. Echigo, Y. Yagi, Which reference view is effective for gait identification using a view transformation model? in: Proceedings of the Conference on Computer Vision and Pattern Recognition Workshop (CVPRW '06), 2006. 45–45
- [30] Y. Makihara, R. Sagawa, Y. Mukaigawa, T. Echigo, Y. Yagi, Gait recognition using a view transformation model in the frequency domain, in: Proceedings of the European Conference on Computer Vision, 2006, pp. 151–163.
- [31] W. Kusakunniran, Q. Wu, J. Zhang, H. Li, L. Wang, Recognizing gaits across views through correlated motion co-clustering, IEEE Trans. Image Process. 23 (2) (2014) 696–709.

- [32] W. Kusakunniran, Q. Wu, J. Zhang, H. Li, Support vector regression for multi-view gait recognition based on local motion feature selection, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2010, pp. 974–981.
- [33] W. Kusakunniran, Q. Wu, J. Zhang, H. Li, Gait recognition under various viewing angles based on correlated motion regression, *IEEE Trans. Circuits Syst. Video Technol.* 22 (6) (2012) 966–980.
- [34] M. Goffredo, I. Bouchrika, J.N. Carter, M.S. Nixon, Self-calibrating view-invariant gait biometrics, *IEEE Trans. Syst. Man Cybern. Part B Cybern.* 40 (4) (2010) 997–1008.
- [35] Y. Makihara, A. Mansur, D. Muramatsu, Z. Uddin, Multi-view discriminant analysis with tensor representation and its application to cross-view gait recognition, in: Proceedings of the IEEE International Conference and Workshops on Automatic Face and Gesture Recognition, 2015, pp. 1–8.
- [36] N. Liu, J. Lu, Y.P. Tan, Joint subspace learning for view-invariant gait recognition, *IEEE Signal Process. Lett.* 18 (7) (2011) 431–434.
- [37] X. Xing, K. Wang, T. Yan, Z. Lv, Complete canonical correlation analysis with application to multi-view gait recognition, *Pattern Recognit.* 50 (2016) 107–117.
- [38] J. Lu, Y.P. Tan, Uncorrelated discriminant simplex analysis for view-invariant gait signal computing, *Pattern Recognit. Lett.* 31 (5) (2010) 382–393.
- [39] J. Lu, G. Wang, P. Moulin, Human identity and gender recognition from gait sequences with arbitrary walking directions, *IEEE Trans. Inf. Forensics Secur.* 9 (1) (2014) 51–61.
- [40] Z. Wu, Y. Huang, L. Wang, X. Wang, T. Tan, A comprehensive study on cross-view gait based human identification with deep CNNs, *IEEE Trans. Pattern Anal. Mach. Intell.* 39 (2) (2017) 209–226.
- [41] N. Takemura, Y. Makihara, D. Muramatsu, T. Echigo, Y. Yagi, On input/output architectures for convolutional neural network-based cross-view gait recognition, *IEEE Trans. Circuits Syst. Video Technol.* (2018), doi:10.1109/TCSVT.2017.2760835.
- [42] K. Shiraga, Y. Makihara, D. Muramatsu, T. Echigo, Y. Yagi, GEINET: View-invariant gait recognition using a convolutional neural network, in: Proceedings of the International Conference on Biometrics, 2016, pp. 1–8.
- [43] 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.
- [44] H. Lu, K.N. Plataniotis, A.N. Venetsanopoulos, Uncorrelated multilinear principal component analysis for unsupervised multilinear subspace learning, *IEEE Trans. Neural Netw.* 20 (11) (2009) 1820–1836.
- [45] 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.
- [46] S. Yan, D. Xu, Q. Yang, L. Zhang, X. Tang, H.J. Zhang, Discriminant analysis with tensor representation, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005, pp. 526–532.
- [47] 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). 103–23
- [48] X. Li, S. Lin, S. Yan, D. Xu, Discriminant locally linear embedding with high-order tensor data, *IEEE Trans. Syst. Man Cybern. Part B Cybern.* 38 (2) (2008) 342–352.
- [49] X. Ben, P. Zhang, R. Yan, M. Yang, G. Ge, Gait recognition and micro-expression recognition based on maximum margin projection with tensor representation, *Neural Comput. Appl.* 27 (8) (2016) 2629–2646.
- [50] Y.L. Chen, C.T. Hsu, Multilinear graph embedding: representation and regularization for images, *IEEE Trans. Image Process.* 23 (2) (2014) 741–754.
- [51] N. Zhao, L. Zhang, B. Du, L. Zhang, D. Tao, J. You, Sparse tensor discriminative locality alignment for gait recognition, in: Proceedings of the International Joint Conference on Neural Networks, 2016, pp. 4489–4495.
- [52] X. Ben, W. Meng, R. Yan, K. Wang, An improved biometrics technique based on metric learning approach, *Neurocomputing* 97 (2012) 44–51.
- [53] X. Ben, W. Meng, R. Yan, Dual-ellipse fitting approach for robust gait periodicity detection, *Neurocomputing* 79 (2012) 173–178.
- [54] J. Han, B. Bhanu, Individual recognition using gait energy image, *IEEE Trans. Pattern Anal. Mach. Intell.* 28 (2) (2005) 316–322.
- [55] A. Mansur, Y. Makihara, D. Muramatsu, Y. Yagi, Cross-view gait recognition using view-dependent discriminative analysis, in: Proceedings of the IEEE International Joint Conference on Biometrics, 2014, pp. 1–8, doi:10.1109/BTAS.2014.6996272.
- [56] M.D. Zeiler, R. Fergus, Visualizing and understanding convolutional networks, in: Proceedings of the European Conference on Computer Vision, 2014, pp. 818–833.
- [57] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.

Xianye Ben received the Ph.D. degree in pattern recognition and intelligent system from the College of Automation, Harbin Engineering University, Harbin, in 2010. She is currently working as an Associate Professor in the School of Information Science and Engineering, Shandong University, Qingdao, China. She has published more than 80 papers in major journals and conferences, such as IEEE T-IP, IEEE T-CSVT, PR, etc. Her current research interests include pattern recognition and image processing. She received the Excellent Doctorial Dissertation awarded by Harbin Engineering University. She was also enrolled by the Young Scholars Program of Shandong University.

Peng Zhang received the B.S. and M.S. degree in communication engineering from the School of Information Science and Engineering, Shandong University, Jinan, China, in 2013 and 2016, respectively. He is a second-year PhD candidate with Global Big Data and Technologies Centre (GBDTC) in University of Technology Sydney, Sydney, Australia currently. He has published more than 10 academic papers in major conferences and journals and holds several Chinese invention patents. His current research interests include gait recognition, person re-identification and generative adversarial network.

Zhihui Lai received the B.S degree in mathematics from South China Normal University, M.S degree from Jinan University, and the PhD degree in pattern recognition and intelligence system from Nanjing University of Science and Technology (NUST), China, in 2002, 2007 and 2011, respectively. He has been a research associate, Post-doctoral Fellow and Research Fellow at The Hong Kong Polytechnic University. 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. He has published over 60 scientific articles, including 30 papers published on top-tier IEEE Transactions. Now he is an associate editor of International Journal of Machine Learning and Cybernetics. For more information, including all the papers and the Matlab codes, please refer to his website: <http://www.scholart.com/laizhihui>.

Rui Yan graduates from Rensselaer Polytechnic Institute majoring in Computer Science with a Ph.D. degree. His now a data scientist in Microsoft AI & R. His research interests include knowledge graph, machine learning and pattern recognition.

Xinliang Zhai was born in Xilinguole, China, in 1995. He received the B.E. degree from the School of Information Science and Engineering, Shandong University, Jinan, China, 2017 and now is a master candidate of Shandong University. His current research interests include person re-identification and gait recognition.

Weixiao Meng was born in Harbin, China, in 1968. He received his B.S. degree in Electronic Instrument and Measurement Technology from Harbin Institute of Technology (HIT), China, in 1990. And then he obtained the M.S. and Ph.D. degree, both in Communication and Information System, HIT, in 1995 and 2000, respectively. Now he is a professor in School of Electronics and Communication Engineering, HIT. Besides, he is a senior member of IEEE, a senior member of China Institute of Electronics, China Institute of Communication and Expert Advisory Group on Harbin E-Government. His research interests mainly focus on adaptive signal processing. In recent years, he has published 1 authored book and more than 100 academic papers on journals and international conferences, more than 60 of which were indexed by SCI, EI and ISTP. Up to now, he has totally completed more than 20 research projects and holds 6 China patents. 1 standard proposal was accepted by IMT-Advanced technical group.