Coupled Bilinear Discriminant Projection for Cross-View Gait Recognition

Xianye Ben, *Member, IEEE*, Chen Gong, *Member, IEEE*, Peng Zhang, Rui Yan, Qiang Wu, *Senior Member, IEEE*, and Weixiao Meng, *Senior Member, IEEE*

***Abstract*— A problem that hinders good performance of general gait recognition systems is that the appearance features of gaits are more affected-prone by views than identities, especially when the walking direction of the probe gait is different from the register gait. This problem cannot be solved by traditional projection learning methods because these methods can learn only one projection matrix, and thus for the same subject, it cannot transfer cross-view gait features into similar ones. This paper presents an innovative method to overcome this problem by aligning gait energy images (GEIs) across views with the coupled bilinear discriminant projection (CBDP). Specifically, the CBDP generates the aligned gait matrix features for two views with two sets of bilinear transformation matrices, so that the original GEIs' spatial structure information can be preserved. By iteratively maximizing the ratio of inter-class distance metric to intra-class distance metric, the CBDP can learn the optimal matrix subspace where the GEIs across views are aligned in both horizontal and vertical coordinates. Therefore, the CBDP is also able to avoid the under-sample problem. We also theoretically prove that the upper and lower bounds of the objective function sequence of the CBDP are both monotonically increasing, so the convergence of the CBDP is demonstrated. In the terms of accuracy, the comparative experiments on the CASIA (B) and OU-ISIR gait databases show that our method is superior to the state-of-the-art cross-view gait recognition methods. More impressively, encouraging performance is obtained by our method even in matching a lateral-view gait with a frontal-view gait.**

***Index Terms*— Gait recognition, coupled bilinear discriminant projection, image alignment, cross-view gait recognition.**

I. Introduction

HUMAN gait is one of the well-known perceptible biometrics at a distance. They can be captured from an unconscious and uncooperative subject compared to other biometrics (such as faces, fingerprints, palms, veins, etc), and thus the research on gait recognition has been conducted extensively during recent decades. There are two main categories of the state-of-the-art gait recognition techniques, i.e. model-based [1] and motion-based approaches [2]–[4]. Model-based approaches extract the gait features robustly and avoid the noise interference problem. Motion-based approaches characterize the motion patterns of human body without fitted model parameters.

However, their performances will drop due to the changes in clothing, shoes, carrying condition, walking surface, walking speed, and elapsed time [5]. More seriously, the view change will lead to a dramatic change of the gait appearance, which poses a great difficulty for accurate gait recognition. Therefore, cross-view gait recognition is quite challenging which deserves further study.

Cross-view gait recognition is tackled by feature description methods [6]–[11] and machine learning-based methods [12]–[20]. The difference is that the former focuses on how to estimate a gait property that is robust to the view change, and the latter investigates how to infer and understand the underlying relationship between cross-view gaits. According to the selection of former gait models, feature description methods include 3D gait model and trajectories estimation. 3D gait model needs high computational complexity, and trajectories estimation is very difficult under a frontal view. However, machine learning-based methods can overcome these problems wonderfully well.

In this paper, we present a Coupled bilinear discriminant projection (CBDP) for aligning gait images across views. The method proposed in this paper belongs to the machine learning-based methods, which has demonstrated the best performance up to now. Although all the methods in the

Manuscript received April 2, 2018; revised September 30, 2018 and December 9, 2018; accepted January 9, 2019. Date of publication January 17, 2019; date of current version March 5, 2020. This work was supported in part by the National Key Research and Development Program of China under Grant 2017YFC0803401, in part by the Natural Science Foundation of China under Grant 61571275, Grant 61602246, and Grant 61201370, in part by NSF of Jiangsu Province under GrantBK20171430, in part by the Fundamental Research Funds for the Central Universities under Grant 30918011319, in part by the Summit of the Six Top Talent Program under Grant DZXX-027, in part by the Young Scholars Program of Shandong University, in part by the Innovative and Entrepreneurial Doctor Program of Jiangsu Province, in part by the Lift Program for Young Talents of Jiangsu Province, and in part by the CAST Lift Program for Young Talents. This paper was recommended by Associate Editor H. Lu *(Corresponding author: Xianye Ben.)*

X. Ben is with the School of Information Science and Engineering, Shandong University, Qingdao 266237, China (e-mail: benxianye@gmail.com).

C. Gong is with the Key Laboratory of Intelligent Perception and Systems for High-Dimensional Information, School of Computer Science and Engineering, Ministry of Education, Nanjing University of Science and Technology, Nanjing 210094, China (e-mail: chen.gong@njust.edu.cn).

P. Zhang is with the School of Electrical and Data Engineering, University of Technology Sydney, Sydney, NSW 2007, Australia (e-mail: peng.zhang-2@student.uts.edu.au).

R. Yan is with Microsoft AI and Research, Bellevue, WA 98004, USA (e-mail: raymondino.yan@gmail.com).

Q. Wu is with the School of Electrical and Data Engineering, University of Technology Sydney, Sydney, NSW 2007, Australia (e-mail: qiang.wu@uts.edu.au).

W. Meng is with the School of Electronics and Information Engineering, Harbin Institute of Technology, Harbin 150080, China (e-mail: wxmeng@hit.edu.cn).

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

Digital Object Identifier 10.1109/TCSVT.2019.2893736

TABLE I
THE SUMMARY OF SPATIO-TEMPORAL GAIT REPRESENTATION APPROACHES

Name	Description	S	D	T
MEI [25]	It can record the dynamically changed location of gait in the image.	N	Y	N
MHI [25]	The intensity value of each pixel is proportional to the duration of continuous movement at that location.	N	Y	Y
GHI [26]	The time duration is controlled by a period resolution of 1/4 of a gait cycle.	Y	Y	Y
MSI [27]	It can record the temporal history silhouette of continuous movement.	Y	N	Y
GEI [28]	It can reflect the duration of each state throughout in each period.	Y	Y	N
EGEI [29]	GEI is adjusted based on the distribution of dynamic area; however, it is very difficult to determine the enhanced parameters.	Y	Y	N
GEnI [30]	Shannon entropy measures the uncertainty of silhouette density at different locations.	Y	Y	N
GFI [31]	GFI embeds irrelevant information when the quality of the gait silhouette image is poor.	N	Y	Y
SEI [32]	SEI can extract the shifted region of GEI.	Y	Y	N
CGI [33]	CGI is a multichannel representation of gait silhouette images.	Y	Y	Y

S, D, T denote static, dynamic, time-varying, respectively, and Y and N mean yes and no.

machine learning-based category can reduce the gap caused by view difference to some extent, most of them [12]–[14], [16]–[22] need concatenating gait energy image (GEI) into a 1D vector, which leads to a high-dimensional feature space for gaits. They do not consider to directly deal with the matching cross-view GEIs, so the underlying structure, such as the correlations among the rows and columns, cannot be exploited and the spatial information of GEIs have not been deployed at the same time. To address this shortcoming, we propose to use the CBDP for directly handling GEIs as 2D data, since GEIs are intrinsically a second-order matrix. CBDP iteratively maximizes the ratio of inter-class distance metric to intra-class distance metric. Specifically, CBDP can learn the optimal matrix subspace where GEIs across views are aligned in both horizontal and vertical coordinates. In particular, we make the following contributions:

(1) We formulate cross-view gait recognition as an image alignment problem based on direct coupled bilinear discriminant projection without losing the spatial information of GEIs. Under the learnt coupled distance metric, the intra-class local geometry and inter-class discriminative information are preserved, which facilitates the nearest neighbor classifier.
(2) By iteratively maximizing the ratio of inter-class distance metric to intra-class distance metric, we theoretically prove that the upper and lower bounds of the objective function sequence of CBDP are both monotonically increasing, so the optimization process is guaranteed to converge.
(3) We extensively compare the proposed method with the state-of-the-art methods on the popular cross-view gait recognition databases such as CASIA(B) [23] and OU-ISIR [24]. Experimental results show the proposed method outperforms other exiting approaches; more impressively, the encouraging performance is obtained even in matching a lateral-view gait with a frontal-view gait (the most challenging case of the largest difference between view points).

The rest of this paper is structured as follows: Section II reviews related literature on spatio-temporal templates and cross-view gait recognition. Section III describes the proposed Coupled bilinear discriminant projection (CBDP). In Section IV, the convergence analyses of CBDP are carefully discussed. Experimental results are given in Section V. Section VI concludes the entire paper.

II. RELATED WORKS

In this section, the popular motion-based approaches for gait recognition and the typical cross-view gait recognition methods will be reviewed.

A. Motion-Based Approaches

The most popular motion-based methods are spatio-temporal templates which are able to express the static, dynamic or time-varying information of gaits, such as Motion Energy Image (MEI) [25], Motion History Image (MHI) [25], Gait History Image (GHI) [26], Moving Silhouette Image (MSI) [27], Gait Energy Image (GEI) [28], Enhanced GEI (EGEI) [29], Gait Entropy Image (GEnI) [30], Gait Flow Image (GFI) [31], Shifted Energy Image (SEI) [32], and Chrono-Gait Image (CGI) [33]. Table I summarizes their technical details.

Reference [34] shows the quality of being a competitive advantage of GEI. In particular, with the purpose of learning a discriminative subspace for GEI data, Xu et al. [35] used Coupled subspaces analysis (CSA) and Discriminant analysis with tensor representation (DATER) to directly deal with the gait energy images (GEIs). Li *et al.* [36] proposed a new supervised manifold learning algorithm called Discriminant locally linear embedding (DLLE) to extract the feature of the GEI. Xu *et al.* [37] developed a matrix-based Marginal fisher analysis (MFA) to directly handle the GEI, which aims to characterize the intra-class compactness through the distance between each positive sample and its neighboring positive samples. Besides, sparse discriminant projection learning [38], multilinear graph embedding [39], and patch distribution compatible semi-supervised dimension reduction [40] were proposed for gait feature extraction. However, the above-mentioned dimension reduction methods are only suitable for identical or approximately identical view gait recognition.

B. Cross-View Gait Recognition Methods

In recent years, several methods have been presented for cross-view gait recognition. According to the different approaches used to generate view-invariance features, the cross-view gait recognition methods can be grouped into two categories: feature description methods and machine learning-based methods.

TABLE II
KEY NOTATIONS USED IN THIS STUDY

Symbol	Explanation
$\mathbf{X}, \mathbf{Y}$	GEIs from views θ and ϑ, respectively
$\mathbf{X}_i^{(c)}, \mathbf{Y}_i^{(c)}$	The i-th GEI of the c-th class from $\mathbf{X}, \mathbf{Y}$, respectively
N_c	Number of GEIs in the c-th class
C	Number of classes
$\bar{\mathbf{X}}^{(c)}, \bar{\mathbf{Y}}^{(c)}$	Mean matrices of class c from views θ and ϑ, respectively
$\mathbf{P}_x, \mathbf{Q}_x$	Mapping matrices for GEIs from view θ
$\mathbf{P}_y, \mathbf{Q}_y$	Mapping matrices for GEIs from view ϑ

Feature description methods usually describe a gait property which is robust to the view change. Stemming from the 3D gait model, Zhao *et al.* [6] successfully extracted the static lengths of key segments features and the dynamic trajectory features of lower limbs to achieve 3D gait recognition. Luo et al. [7] proposed a clothes-independent 3D human pose and shape estimation method to extract 3D human body parameters. Tang *et al.* [41] used of 3D human pose estimation and shape deformation to reconstruct parametric 3D body. Chen *et al.* [8] estimated the real gravity center trajectory (GCT) curve through the statistics of limb parameters, and then projected GCT based on the projection principle between curve and plane. Other related works inspired by gait feature extraction from 3D space for cross-view gait recognition include 3D morphological analysis [9], self-calibration [10] and view-normalized body part trajectories [11]. However, one disadvantage of 3D gait model is that it needs a much more complicated setup environment; in addition, the computational cost involved in this system is very large. Trajectories estimation is very difficult under the frontal view; therefore, it may limit the application of a certain scenarios.

Most machine learning-based methods assume that the target views can be generated from the registered views or the target and the registered views share the common features in the unified space. Zhang *et al.* [42] employed list-wise constraints to learn a projection, which can map gait features from different views into a common discriminative subspace. The underlying relationship between gaits observed from different views can be accurately approximated by radial basis function (RBF) neural networks [12], multi-layer perceptron [13], sparse reconstruction based metric learning [14], deep CNNs [34], [43], and GEINet [15]. View transformation model (VTM) is a well-known model that transfers the gait feature from one view into another view. Recently, several VTMs based on SVD have been proposed, for instance, Kusakunniran *et al.* [16] presented a VTM by using Truncated SVD. Incorporating quality measures, Muramatsu *et al.* [17] proposed a VTM with analysis of part-dependent transformation bias. However, they are based on the assumption that view-independence and individual-independent information can be completely separated. In contrast, the regression technologies such as support vector regression (SVR) [18], sparse regression [19] and CCA regression [20] can learn VTMs, and naturally avoid the problem above. However, these kinds of VTMs can't obtain reasonably good recognition rates when the difference between the target view and the source view is more than 54 degrees. Instead, common subspace methods [21], [22] can learn the joint subspace where the gait features across views for a certain subject are very similar. However, References [21], and [22] take vectors as input, which causes the missing of spatial information carried by GEI.

In contrast, our CBDP follows the idea of CCA regression, which maps GEIs across views into a common subspace where the data biases caused by view variation are mitigated. Compared to the unsupervised CCA, the proposed approach follows Fisher criterion which maximizes the inter-class scatter and minimizes the intra-class scatter. Benefit from specific graph embedding technique for both inter-class and intra-class samples, our proposed approach significantly outperforms CCA-based methods. Moreover, our proposed approach maps the raw GEIs directly, rather than using vectorized GEIs, so our method can completely preserve the structure information of the GEIs. We utilize the bilinear projection instead of the conventional linear projection, therefore the transformation can be completed in higher-order space. This is different from above projection-based methods for gait recognition, i.e., VTM-like regression and CCA-like regression.

III. Coupled Bilinear Discriminant Projection for Cross-View Gait Recognition

Unfortunately, traditional projection learning methods will fail when the walking direction of the probe and registered gaits significantly differ, since they can only deal with the identical view gait recognition. In this section, a new Coupled bilinear discriminant projection (CBDP) algorithm is developed as a solution for aligning the GEIs across different views while preserving the spatial information.

A. Problem Statement

Considering two GEIs $\mathbf{X}_i, \mathbf{Y}_j$ from different views, the goal of CBDP is to measure the similarity by directly calculating the generalized Mahalanobis distance of two GEIs across views, which can be denoted as

$$\begin{aligned}\mathbf{D}(\mathbf{X}_i, \mathbf{Y}_j) &= \mathbf{D}_C(f_x(\mathbf{X}_i), f_y(\mathbf{Y}_j)) \\ &= \sqrt{\mathrm{tr}\{[f_x(\mathbf{X}_i) - f_y(\mathbf{Y}_j)]^\top \mathbf{C}[f_x(\mathbf{X}_i) - f_y(\mathbf{Y}_j)]\}},\end{aligned} \tag{1}$$

where $f_x(\cdot)$ and $f_y(\cdot)$ are two bilinear transformations, which map GEIs from different views, i.e., $\mathbf{X}_i$ and $\mathbf{Y}_j$, into a common subspace, respectively. In the subspace, the data biases caused by view variation are mitigated and the transformed features can be directly compared by using distance measurement. Since $\mathbf{C}$ is positive semi-definite, it can be decomposed by $\mathbf{C} = \mathbf{W}_c\mathbf{W}_c^\top$. By defining the bilinear transforms $f_x(\mathbf{X}_i) = \mathbf{U}_x^\top \mathbf{X}_i \mathbf{Q}_x$, $f_y(\mathbf{Y}_j) = \mathbf{U}_y^\top \mathbf{Y}_j \mathbf{Q}_y$ and also using a pair of mapping matrices $\mathbf{U}_x, \mathbf{Q}_x$ and $\mathbf{U}_y, \mathbf{Q}_y$, the distance $\mathbf{D}(\mathbf{X}_i, \mathbf{Y}_j)$ can then be converted to (2), shown at the bottom of the next page.

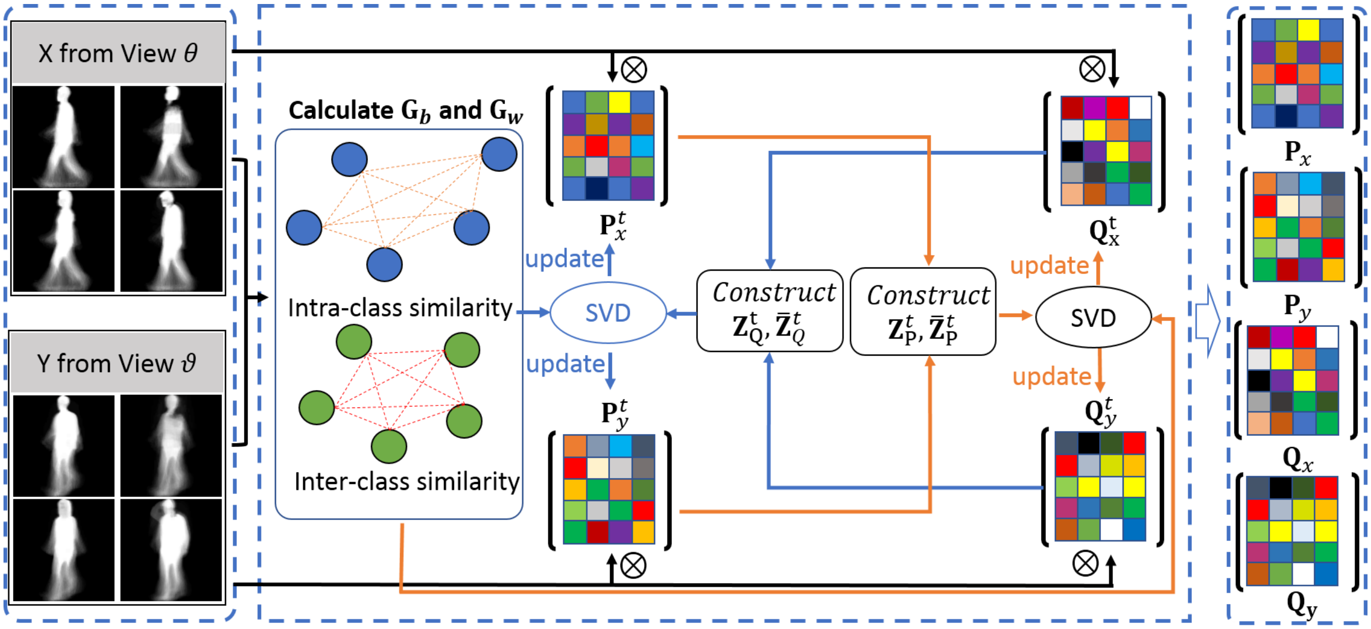


Fig. 1. Alternative optimization procedure of the proposed CBDP. When fixing $\mathbf{Q}_x^{t-1}$ and $\mathbf{Q}_y^{t-1}$ at the t-th iteration, $\mathbf{P}_x^t$ and $\mathbf{P}_y^t$ can be obtained by solving the SVD problem in Eq. 6, similarly, when fixing $\mathbf{P}_x^t$ and $\mathbf{P}_y^t$, $\mathbf{Q}_x^t$ and $\mathbf{Q}_y^t$ can be obtained by solving the SVD problem in Eq. 8.

Then, by denoting $\mathbf{P}_x = \mathbf{U}_x\mathbf{W}_c$ and $\mathbf{P}_y = \mathbf{U}_y\mathbf{W}_c$, (2) can be further rewritten as

$$\mathbf{D}(\mathbf{X}_i, \mathbf{Y}_j) = \sqrt{\mathrm{tr}\{[\mathbf{P}_x^\top\mathbf{X}_i\mathbf{Q}_x - \mathbf{P}_y^\top\mathbf{Y}_j\mathbf{Q}_y]^\top[\mathbf{P}_x^\top\mathbf{X}_i\mathbf{Q}_x - \mathbf{P}_y^\top\mathbf{Y}_j\mathbf{Q}_y]\}}. \quad (3)$$

CBDP aims to find two sets of bilinear transformation matrices $\{\mathbf{P}_x, \mathbf{Q}_x\}$ and $\{\mathbf{P}_y, \mathbf{Q}_y\}$ respectively for GEIs across two different views, and map them into a shared subspace in which features can be directly measured. Therefore, (3) serves as the CBDP for measuring the difference of two arbitrary cross-view GEIs.

B. Coupled Bilinear Discriminant Projection (CBDP)

As shown in Fig. 1, the proposed CBDP learns the common lower-dimensional discriminant matrix subspace that most efficiently links gaits across views. It can directly deal with GEIs, therefore the spatial structure of GEIs can be preserved, and at the same time, the under-sample problem can be avoided.

For ease of representation, Table II lists important notations defined in the paper. Two sets of GEIs across views can be denoted as $\mathbf{X} = \left\{\mathbf{X}_i^{(c)} \in \mathbb{R}^{D_{xm}\times D_{xn}}, i = 1, \ldots, N_c, c = 1, \ldots, C\right\}$ and $\mathbf{Y} = \left\{\mathbf{Y}_j^{(c)} \in \mathbb{R}^{D_{ym}\times D_{yn}}, j = 1, \ldots, N_c, c = 1, \ldots, C\right\}$, respectively, where $D_{xm} \times D_{xn}$ and $D_{ym} \times D_{yn}$ denote the size of GEIs.

To facilitate the subsequent discussion, mean matrices of class c from views θ and ϑ are respectively given by $\bar{\mathbf{X}}^{(c)} = \frac{1}{N_c}\sum_{i=1}^{N_c}\mathbf{X}_i^{(c)}, \bar{\mathbf{Y}}^{(c)} = \frac{1}{N_c}\sum_{j=1}^{N_c}\mathbf{Y}_j^{(c)}$. In order to preserve the inter-class local geometry and intra-class discriminative information, the objective function of the proposed CBDP can be defined as

$$\underset{\mathbf{P}_x,\mathbf{P}_y,\mathbf{Q}_x,\mathbf{Q}_y}{\arg\max} \frac{\sum_{i,j=1}^{C}\left\|\mathbf{P}_x^\top\bar{\mathbf{X}}^{(i)}\mathbf{Q}_x - \mathbf{P}_y^\top\bar{\mathbf{Y}}^{(j)}\mathbf{Q}_y\right\|_F^2\mathbf{S}(i,j)}{\sum_{c=1}^{C}\sum_{i,j=1}^{N_c}\left\|\mathbf{P}_x^\top\mathbf{X}_i^{(c)}\mathbf{Q}_x - \mathbf{P}_y^\top\mathbf{Y}_j^{(c)}\mathbf{Q}_y\right\|_F^2\mathbf{W}^{(c)}(i,j)}, \quad (4)$$

where $\mathbf{S}$ is the inter-class similarity matrix weighted by $\mathbf{S}(i,j) = \exp\left(-\left\|\bar{\mathbf{X}}^{(i)} - \bar{\mathbf{X}}^{(j)}\right\|_F^2/t\right)$ (t is a heat kernel parameter), $\mathbf{W}^{(c)}$ is the intra-class similarity matrix for class c weighted by $\mathbf{W}^{(c)}(i,j) = \exp\left(-\left\|\mathbf{X}_i^{(c)} - \mathbf{X}_j^{(c)}\right\|_F^2\Big/t\right)$, and $\|\cdot\|_F$ denotes Frobenius norm.

Eq. (4) often has no closed-form solution, so an iterative solution is adopted. When fixing $\mathbf{Q_x}$ and $\mathbf{Q_y}$, $\mathbf{P_x}$ and $\mathbf{P_y}$ can be optimized by

$$\underset{\mathbf{P}}{\arg\max}\, J(\mathbf{P}) = \frac{\mathrm{tr}\left(\mathbf{P}^\top\bar{\mathbf{Z}}_Q\mathbf{G}_b\bar{\mathbf{Z}}_Q^\top\mathbf{P}\right)}{\mathrm{tr}\left(\mathbf{P}^\top\mathbf{Z}_Q\mathbf{G}_w\mathbf{Z}_Q^\top\mathbf{P}\right)}, \quad (5)$$

where

$$\mathbf{P} = \left[\mathbf{P}_x^\top\ \mathbf{P}_y^\top\right]^\top, \quad \bar{\mathbf{Z}}_Q = \begin{bmatrix}\bar{\mathbf{A}}_Q & \mathbf{0}\\ \mathbf{0} & \bar{\mathbf{B}}_Q\end{bmatrix},$$

$$\mathbf{D}(\mathbf{X}_i, \mathbf{Y}_j) = \sqrt{\mathrm{tr}\{[\mathbf{W}_c^\top\mathbf{U}_x^\top\mathbf{X}_i\mathbf{Q}_x - \mathbf{W}_c^\top\mathbf{U}_y^\top\mathbf{Y}_j\mathbf{Q}_y]^\top[\mathbf{W}_c^\top\mathbf{U}_x^\top\mathbf{X}_i\mathbf{Q}_x - \mathbf{W}_c^\top\mathbf{U}_y^\top\mathbf{Y}_j\mathbf{Q}_y]\}}. \quad (2)$$

$$
\begin{aligned}
\bar{\mathbf{A}}_Q &= \left[\bar{\mathbf{X}}^{(1)}\mathbf{Q}_x, \bar{\mathbf{X}}^{(2)}\mathbf{Q}_x, \ldots, \bar{\mathbf{X}}^{(C)}\mathbf{Q}_x\right],\\
\bar{\mathbf{B}}_Q &= \left[\bar{\mathbf{Y}}^{(1)}\mathbf{Q}_y, \bar{\mathbf{Y}}^{(2)}\mathbf{Q}_y, \ldots, \bar{\mathbf{Y}}^{(C)}\mathbf{Q}_y\right],\\
\mathbf{G}_b &= \begin{bmatrix} \mathbf{E}_1 \otimes \mathbf{I} & -\mathbf{S} \otimes \mathbf{I} \\ -\mathbf{S}^\top \otimes \mathbf{I} & \mathbf{E}_2 \otimes \mathbf{I} \end{bmatrix},\\
\mathbf{E}_1(i,i) &= \sum\nolimits_j \mathbf{S}(i,j), \quad \mathbf{E}_2(j,j) = \sum\nolimits_i \mathbf{S}(i,j),\\
\mathbf{Z}_Q &= \begin{bmatrix} \mathbf{A}_Q^{(1)} & \mathbf{0} & \cdots & \mathbf{A}_Q^{(C)} & \mathbf{0} \\ \mathbf{0} & \mathbf{B}_Q^{(1)} & \cdots & \mathbf{0} & \mathbf{B}_Q^{(C)} \end{bmatrix},\\
\mathbf{A}_Q^{(c)} &= \left[\mathbf{X}_1^{(c)}\mathbf{Q}_x, \mathbf{X}_2^{(c)}\mathbf{Q}_x, \ldots, \mathbf{X}_{N_c}^{(c)}\mathbf{Q}_x\right],\\
\mathbf{B}_Q^{(c)} &= \left[\mathbf{Y}_1^{(c)}\mathbf{Q}_y, \mathbf{Y}_2^{(c)}\mathbf{Q}_y, \ldots, \mathbf{Y}_{N_c}^{(c)}\mathbf{Q}_y\right],\\
&\qquad\qquad\qquad \text{for } c = 1, \ldots, C,\\
\mathbf{G}_w &= diag\left(\begin{bmatrix} \mathbf{D}_1^{(1)} \otimes \mathbf{I} & -\mathbf{W}^{(1)} \otimes \mathbf{I} \\ -(\mathbf{W}^{(1)})^\top \otimes \mathbf{I} & \mathbf{D}_2^{(1)} \otimes \mathbf{I} \end{bmatrix}, \ldots,\right.\\
&\qquad \left.\begin{bmatrix} \mathbf{D}_1^{(C)} \otimes \mathbf{I} & -\mathbf{W}^{(C)} \otimes \mathbf{I} \\ -(\mathbf{W}^{(C)})^\top \otimes \mathbf{I} & \mathbf{D}_2^{(C)} \otimes \mathbf{I} \end{bmatrix}\right),\\
\mathbf{D}_1^{(c)}(i,i) &= \sum\nolimits_j \mathbf{W}^{(c)}(i,j), \quad \mathbf{D}_2^{(c)}(j,j) = \sum\nolimits_i \mathbf{W}^{(c)}(i,j).
\end{aligned}
$$

The constraint $\mathbf{P}^\top\mathbf{P} = \mathbf{I}$ is imposed on (5) to uniquely determine the transformation matrices $\mathbf{P}_x$ and $\mathbf{P}_y$. In addition, to avoid overfitting, a regularization $\tau\mathbf{I}$ is imposed on $\mathbf{G}_w$, where τ is a small positive number, such as $\tau = 10^{-6}$. The objective function can be reformulated as a more tractable ratio trace optimization problem as follows [44]

$$
\begin{aligned}
&\arg\max_{\mathbf{P}^\top\mathbf{P}=\mathbf{I}} J(\mathbf{P})\\
&= \operatorname{tr}\left[\left(\mathbf{P}^\top\mathbf{Z}_Q\left(\mathbf{G}_w + \tau\mathbf{I}\right)\mathbf{Z}_Q^\top\mathbf{P}\right)^{-1}\left(\mathbf{P}^\top\bar{\mathbf{Z}}_Q\mathbf{G}_b\bar{\mathbf{Z}}_Q^\top\mathbf{P}\right)\right], \quad (6)
\end{aligned}
$$

which can be easily solved by Lagrangian multiplier method.

Similarly, when fixing $\mathbf{P}_x$ and $\mathbf{P}_y$, $\mathbf{Q}_x$ and $\mathbf{Q}_y$ can be optimized by

$$
\arg\max_{\mathbf{Q}} J(\mathbf{Q}) = \frac{\operatorname{tr}\left(\mathbf{Q}^\top\bar{\mathbf{Z}}_P\mathbf{G}_b\bar{\mathbf{Z}}_P^\top\mathbf{Q}\right)}{\operatorname{tr}\left(\mathbf{Q}^\top\mathbf{Z}_P\mathbf{G}_w\mathbf{Z}_P^\top\mathbf{Q}\right)}, \quad (7)
$$

where $\mathbf{Q} = \left[\mathbf{Q}_x{}^\top\ \mathbf{Q}_y{}^\top\right]^\top$, $\bar{\mathbf{Z}}_P = \begin{bmatrix} \bar{\mathbf{A}}_P & \mathbf{0} \\ \mathbf{0} & \bar{\mathbf{B}}_P \end{bmatrix}$, $\bar{\mathbf{A}}_P = \left[(\bar{\mathbf{X}}^{(1)})^\top\mathbf{P}_x, (\bar{\mathbf{X}}^{(2)})^\top\mathbf{P}_x, \ldots, (\bar{\mathbf{X}}^{(C)})^\top\mathbf{P}_x\right]$, $\bar{\mathbf{B}}_P = \left[(\bar{\mathbf{Y}}^{(1)})^\top\mathbf{P}_y, (\bar{\mathbf{Y}}^{(2)})^\top\mathbf{P}_y, \ldots, (\bar{\mathbf{Y}}^{(C)})^\top\mathbf{P}_y\right]$, $\mathbf{Z}_P = \begin{bmatrix} \mathbf{A}_P^{(1)} & \mathbf{0} & \cdots & \mathbf{A}_P^{(C)} & \mathbf{0} \\ \mathbf{0} & \mathbf{B}_P^{(1)} & \cdots & \mathbf{0} & \mathbf{B}_P^{(C)} \end{bmatrix}$, $\mathbf{A}_P^{(c)} = \left[(\mathbf{X}_1^{(c)})^\top\mathbf{P}_x, (\mathbf{X}_2^{(c)})^\top\mathbf{P}_x, \ldots, (\mathbf{X}_{N_c}^{(c)})^\top\mathbf{P}_x\right]$, $\mathbf{B}_P^{(c)} = \left[(\mathbf{Y}_1^{(c)})^\top\mathbf{P}_y, (\mathbf{Y}_2^{(c)})^\top\mathbf{P}_y, \ldots, (\mathbf{Y}_{N_c}^{(c)})^\top\mathbf{P}_y\right]$, for $c = 1, \ldots, C$.

Also, by adding the constraint $\mathbf{Q}^\top\mathbf{Q} = \mathbf{I}$ and regularization $\tau\mathbf{I}$ to (7), the objective function can be rewritten as

$$
\begin{aligned}
&\arg\max_{\mathbf{Q}^\top\mathbf{Q}=\mathbf{I}} J(\mathbf{Q})\\
&= \operatorname{tr}\left[\left(\mathbf{Q}^\top\mathbf{Z}_P\left(\mathbf{G}_w + \tau\mathbf{I}\right)\mathbf{Z}_P^\top\mathbf{Q}\right)^{-1}\left(\mathbf{Q}^\top\bar{\mathbf{Z}}_P\mathbf{G}_b\bar{\mathbf{Z}}_P^\top\mathbf{Q}\right)\right]. \quad (8)
\end{aligned}
$$

The entire alternating projection optimization procedure for CBDP is summarized in *Algorithm* 1.

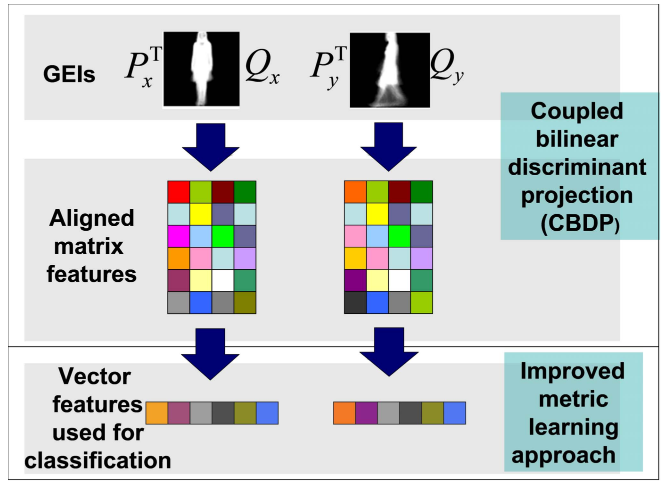


Fig. 2. Overview of the procedure for finding the potential common features across views.

Denoting T as the number of iterations, the increase of the time complexity is proportional to iterations, but the space complexity is not altered with iterations. For convenience, the size of GEI denotes $\frac{L}{2} \times \frac{L}{2}$ for both views. The time and space complexities are $O(TL^3)$ and $O(L^2)$, respectively.

C. Classification

GEIs across views of the same individual are still bridged closely by some potential common features. The overview of the procedure to find the potential common features across views for a certain individual is depicted in Fig. 2. As shown in the upper part of Fig. 2, the goal of CBDP is to find an optimal common matrix space for cross-view GEIs and generate the aligned matrix features. In contrast, the improved metric learning approach [45] can obtain the best common vector space where the cross-view features used for classification are similar.

In the training stage of CBDP, two sets of bilinear transformation matrices $\{\mathbf{P}_x, \mathbf{Q}_x\}$ and $\{\mathbf{P}_y, \mathbf{Q}_y\}$ are learnt respectively for two different views θ and ϑ. The aligned matrix features for cross-view gaits are expressed by

$$
\mathbf{F}_x = \mathbf{P}_x^\top\mathbf{X}_i\mathbf{Q}_x, \mathbf{F}_y = \mathbf{P}_y^\top\mathbf{Y}_j\mathbf{Q}_y. \quad (9)
$$

Then, the gait features extracted from the vectorized features $\mathbf{F}_x$ and $\mathbf{F}_y$ by using the improved metric learning approach [45] are denoted as $\mathbf{f}_x$ and $\mathbf{f}_y$ for two views θ and ϑ. When a query GEI $\mathbf{X}'$ with view θ is received, its feature can be denoted as $\mathbf{f}_x{}'$, while GEIs with another view ϑ are registered. The nearest neighbor classifier is used to determine the class label of $\mathbf{X}'$. If the distance between $\mathbf{f}_y$ and $\mathbf{f}_x{}'$ is minimum, $\mathbf{X}'$ belongs to the class of $\mathbf{Y}_j$.

IV. Convergence Analyses

Here we prove that the CBDP's objective function sequence is monotonically bounded at each iteration. Two sets of bilinear transformation matrices $\{\mathbf{P}_x^0, \mathbf{Q}_x^0\}$ and $\{\mathbf{P}_y^0, \mathbf{Q}_y^0\}$ are respectively initialized as mapping matrices at iteration 0 for two different views. Suppose at iteration t, the concatenation

Algorithm 1 Alternating Projection Optimization Procedure for CBDP

Input:

Training sets of GEIs across views $\mathbf{X} = \left\{\mathbf{X}_i^{(c)} \in \mathbb{R}^{D_{xm}\times D_{xn}}, i=1,\ldots,N_c, c=1,\ldots,C\right\}$ and $\mathbf{Y} = \left\{\mathbf{Y}_j^{(c)} \in \mathbb{R}^{D_{ym}\times D_{yn}}, j=1,\ldots,N_c, c=1,\ldots,C\right\}$, the dimensionality $D_m \times D_n$ of the aligned matrix features, and the number of iterations T.

Output:

Two sets of bilinear transformation matrices $\{\mathbf{P}_x, \mathbf{Q}_x\}$ and $\{\mathbf{P}_y, \mathbf{Q}_y\}$, the aligned matrix features for cross-view gaits $\left\{\mathbf{F}_{xi}^{(c)} \in \mathbb{R}^{D_m\times D_n}, c=1,\ldots,C, i=1,\ldots,N_c\right\}$ and $\left\{\mathbf{F}_{yj}^{(c)} \in \mathbb{R}^{D_m\times D_n}, c=1,\ldots,C, j=1,\ldots,N_c\right\}$.

1: Initialize $\mathbf{Q}_x$ and $\mathbf{Q}_y$ as $\mathbf{Q}_x^0 = \begin{bmatrix}\mathbf{I}_{D_n}\\ \mathbf{0}_{(D_{xn}-D_n)\times D_n}\end{bmatrix}, \mathbf{Q}_y^0 = \begin{bmatrix}\mathbf{I}_{D_n}\\ \mathbf{0}_{(D_{yn}-D_n)\times D_n}\end{bmatrix}$, where $\mathbf{I}_{D_n} \in \mathbb{R}^{D_n\times D_n}$ is the identity matrix, and $\mathbf{0}_{(D_{xn}-D_n)\times D_n} \in \mathbb{R}^{(D_{xn}-D_n)\times D_n}$ is a matrix with all zeros.

2: Calculate $\mathbf{G}_b = \begin{bmatrix}\mathbf{E}_1\otimes\mathbf{I} & -\mathbf{S}\otimes\mathbf{I}\\ -\mathbf{S}^\top\otimes\mathbf{I} & \mathbf{E}_2\otimes\mathbf{I}\end{bmatrix}$ and $\mathbf{G}_w =$ $diag\left(\begin{bmatrix}\mathbf{D}_1^{(1)}\otimes\mathbf{I} & -\mathbf{W}^{(1)}\otimes\mathbf{I}\\ -(\mathbf{W}^{(1)})^\top\otimes\mathbf{I} & \mathbf{D}_2^{(1)}\otimes\mathbf{I}\end{bmatrix}, \ldots, \begin{bmatrix}\mathbf{D}_1^{(C)}\otimes\mathbf{I} & -\mathbf{W}^{(C)}\otimes\mathbf{I}\\ -(\mathbf{W}^{(C)})^\top\otimes\mathbf{I} & \mathbf{D}_2^{(C)}\otimes\mathbf{I}\end{bmatrix}\right)$.

3: for $t = 1 : T$

Calculate $\bar{\mathbf{Z}}_Q^t = \begin{bmatrix}\bar{\mathbf{A}}_Q^t & \mathbf{0}\\ \mathbf{0} & \bar{\mathbf{B}}_Q^t\end{bmatrix}$,

$\bar{\mathbf{A}}_Q^t = \left[\bar{\mathbf{X}}^{(1)}\mathbf{Q}_x^t, \bar{\mathbf{X}}^{(2)}\mathbf{Q}_x^t, \ldots, \bar{\mathbf{X}}^{(C)}\mathbf{Q}_x^t\right]$,

$\bar{\mathbf{B}}_Q^t = \left[\bar{\mathbf{Y}}^{(1)}\mathbf{Q}_y^t, \bar{\mathbf{Y}}^{(2)}\mathbf{Q}_y^t, \ldots, \bar{\mathbf{Y}}^{(C)}\mathbf{Q}_y^t\right]$,

$\mathbf{Z}_Q^t = \begin{bmatrix}(\mathbf{A}^{(1)})_Q^t & \mathbf{0} & \cdots & (\mathbf{A}^{(C)})_Q^t & \mathbf{0}\\ \mathbf{0} & (\mathbf{B}^{(1)})_Q^t & \cdots & \mathbf{0} & (\mathbf{B}^{(C)})_Q^t\end{bmatrix}$,

$(\mathbf{A}^{(c)})_Q^t = \left[\mathbf{X}_1^{(c)}\mathbf{Q}_x^t, \mathbf{X}_2^{(c)}\mathbf{Q}_x^t, \ldots, \mathbf{X}_{N_c}^{(c)}\mathbf{Q}_x^t\right]$, $(\mathbf{B}^{(c)})_Q^t = \left[\mathbf{Y}_1^{(c)}\mathbf{Q}_y^t, \mathbf{Y}_2^{(c)}\mathbf{Q}_y^t, \ldots, \mathbf{Y}_{N_c}^{(c)}\mathbf{Q}_y^t\right]$, for $c = 1, \ldots, C$.

4: Optimize (6) by SVD on $\left(\mathbf{Z}_Q^t\,(\mathbf{G}_w + \tau\mathbf{I})\,(\mathbf{Z}_Q^t)^\top\right)^{-1}\left(\bar{\mathbf{Z}}_Q^t\mathbf{G}_b(\bar{\mathbf{Z}}_Q^t)^\top\right)$ to obtain $\mathbf{P}^t = \left[\,(\mathbf{P}_x^t)^\top\ (\mathbf{P}_y^t)^\top\,\right]^\top$.

5: Calculate $\bar{\mathbf{Z}}_P^t = \begin{bmatrix}\bar{\mathbf{A}}_P^t & \mathbf{0}\\ \mathbf{0} & \bar{\mathbf{B}}_P^t\end{bmatrix}$,

$\bar{\mathbf{A}}_P^t = \left[(\bar{\mathbf{X}}^{(1)})^\top\mathbf{P}_x^t, (\bar{\mathbf{X}}^{(2)})^\top\mathbf{P}_x^t, \ldots, (\bar{\mathbf{X}}^{(C)})^\top\mathbf{P}_x^t\right]$,

$\bar{\mathbf{B}}_P^t = \left[(\bar{\mathbf{Y}}^{(1)})^\top\mathbf{P}_y^t, (\bar{\mathbf{Y}}^{(2)})^\top\mathbf{P}_y^t, \ldots, (\bar{\mathbf{Y}}^{(C)})^\top\mathbf{P}_y^t\right]$,

$\mathbf{Z}_P^t = \begin{bmatrix}(\mathbf{A}^{(1)})_P^t & \mathbf{0} & \cdots & (\mathbf{A}^{(C)})_P^t & \mathbf{0}\\ \mathbf{0} & (\mathbf{B}^{(1)})_P^t & \cdots & \mathbf{0} & (\mathbf{B}^{(C)})_P^t\end{bmatrix}$,

$(\mathbf{A}^{(c)})_P^t = \left[(\mathbf{X}_1^{(c)})^\top\mathbf{P}_x^t, (\mathbf{X}_2^{(c)})^\top\mathbf{P}_x^t, \ldots, (\mathbf{X}_{N_c}^{(c)})^\top\mathbf{P}_x^t\right]$,

$(\mathbf{B}^{(c)})_P^t = \left[(\mathbf{Y}_1^{(c)})^\top\mathbf{P}_y^t, (\mathbf{Y}_2^{(c)})^\top\mathbf{P}_y^t, \ldots, (\mathbf{Y}_{N_c}^{(c)})^\top\mathbf{P}_y^t\right]$, for $c = 1, \ldots, C$.

6: Optimize (8) by SVD on $\left(\mathbf{Z}_P^t\,(\mathbf{G}_w + \tau\mathbf{I})\,(\mathbf{Z}_P^t)^\top\right)^{-1}\left(\bar{\mathbf{Z}}_P^t\mathbf{G}_b(\bar{\mathbf{Z}}_P^t)^\top\right)$ to obtain $\mathbf{Q}^t = \left[\,(\mathbf{Q}_x^t)^\top\ (\mathbf{Q}_y^t)^\top\,\right]^\top$.

7: $t = t + 1$

8: **Until** $Error = \left\|\mathbf{P}^t(\mathbf{P}^{t-1})^\top - \mathbf{I}\right\| + \left\|\mathbf{Q}^t(\mathbf{Q}^{t-1})^\top - \mathbf{I}\right\| \le \varepsilon$.

9: Calculate $\mathbf{F}_{xi}^{(c)} = \mathbf{P}_x^\top\mathbf{X}_i^{(c)}\mathbf{Q}_x$, $\mathbf{F}_{yj}^{(c)} = \mathbf{P}_y^\top\mathbf{Y}_j^{(c)}\mathbf{Q}_y$, $i, j = 1, 2, \ldots, N_c$, $c = 1, 2, \ldots, C$.

matrix $\mathbf{Q}^t$ (composed of $\mathbf{Q}_x^t \in \mathbb{R}^{D_{xn}\times D'}$ and $\mathbf{Q}_y^t \in \mathbb{R}^{D_{yn}\times D'}$) with the size of $\left(D_{xn} + D_{yn}\right) \times D'$ is computed by SVD, and only the first D' columns are preserved.

Denote

$$\bar{\mathbf{Z}}^{(i)} = \begin{bmatrix}\bar{\mathbf{X}}^{(i)} & \mathbf{0}\\ \mathbf{0} & \bar{\mathbf{Y}}^{(i)}\end{bmatrix}, \quad \mathbf{Z}_i^{(c)} = \begin{bmatrix}\mathbf{X}_i^{(c)} & \mathbf{0}\\ \mathbf{0} & \mathbf{Y}_i^{(c)}\end{bmatrix},$$

$$\bar{\mathbf{Z}} = \begin{bmatrix}\bar{\mathbf{A}} & \mathbf{0}\\ \mathbf{0} & \bar{\mathbf{B}}\end{bmatrix}, \quad \text{where } \bar{\mathbf{A}} = \left[\bar{\mathbf{X}}^{(1)}, \bar{\mathbf{X}}^{(2)}, \ldots, \bar{\mathbf{X}}^{(C)}\right],$$

$$\bar{\mathbf{B}} = \left[\bar{\mathbf{Y}}^{(1)}, \bar{\mathbf{Y}}^{(2)}, \ldots, \bar{\mathbf{Y}}^{(C)}\right],$$

$$\mathbf{Z} = \begin{bmatrix}\mathbf{A}^{(1)} & \mathbf{0} & \cdots & \mathbf{A}^{(C)} & \mathbf{0}\\ \mathbf{0} & \mathbf{B}^{(1)} & \cdots & \mathbf{0} & \mathbf{B}^{(C)}\end{bmatrix},$$

where

$$\mathbf{A}^{(c)} = \left[\mathbf{X}_1^{(c)}, \mathbf{X}_2^{(c)}, \ldots, \mathbf{X}_{N_c}^{(c)}\right], \quad \mathbf{B}^{(c)} = \left[\mathbf{Y}_1^{(c)}, \mathbf{Y}_2^{(c)}, \ldots, \mathbf{Y}_{N_c}^{(c)}\right] \quad \text{for } c = 1, \ldots, C,$$

$$\bar{\mathbf{Z}}' = \begin{bmatrix}\bar{\mathbf{A}}' & \mathbf{0}\\ \mathbf{0} & \bar{\mathbf{B}}'\end{bmatrix},$$

where

$$\bar{\mathbf{A}}' = \left[(\bar{\mathbf{X}}^{(1)})^\top, (\bar{\mathbf{X}}^{(2)})^\top, \ldots, (\bar{\mathbf{X}}^{(C)})^\top\right],$$

$$\bar{\mathbf{B}}' = \left[(\bar{\mathbf{Y}}^{(1)})^\top, (\bar{\mathbf{Y}}^{(2)})^\top, \ldots, (\bar{\mathbf{Y}}^{(C)})^\top\right],$$

$$\mathbf{Z} = \begin{bmatrix}(\mathbf{A}')^{(1)} & \mathbf{0} & \cdots & (\mathbf{A}')^{(C)} & \mathbf{0}\\ \mathbf{0} & (\mathbf{B}')^{(1)} & \cdots & \mathbf{0} & (\mathbf{B}')^{(C)}\end{bmatrix},$$

where

$$(\mathbf{A}')^{(c)} = \left[(\mathbf{X}_1^{(c)})^\top, (\mathbf{X}_2^{(c)})^\top, \ldots, (\mathbf{X}_{N_c}^{(c)})^\top\right],$$

$$(\mathbf{B}')^{(c)} = \left[(\mathbf{Y}_1^{(c)})^\top, (\mathbf{Y}_2^{(c)})^\top, \ldots, (\mathbf{Y}_{N_c}^{(c)})^\top\right] \quad \text{for } c = 1, \ldots, C.$$

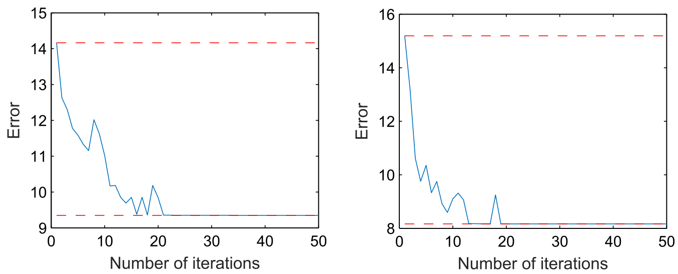
Thus, the corresponding optimal $\mathbf{P}^t = \begin{bmatrix}\mathbf{P}_x^t\\ \mathbf{P}_y^t\end{bmatrix}$ at iteration t is yielded by

$$\begin{aligned}\mathbf{P}^{t*} &= \arg\max_{\mathbf{P}^\top\mathbf{P}=\mathbf{I}} \frac{\sum_{i,j=1}^{C}\left\|\mathbf{P}_x^\top\bar{\mathbf{X}}^{(i)}\mathbf{Q}_x - \mathbf{P}_y^\top\bar{\mathbf{Y}}^{(j)}\mathbf{Q}_y\right\|_F^2\mathbf{S}(i,j)}{\sum_{c=1}^{C}\sum_{i,j=1}^{N_c}\left\|\mathbf{P}_x^\top\mathbf{X}_i^{(c)}\mathbf{Q}_x - \mathbf{P}_y^\top\mathbf{Y}_j^{(c)}\mathbf{Q}_y\right\|_F^2\mathbf{W}^{(c)}(i,j)}\\ &= \arg\max_{\mathbf{P}^\top\mathbf{P}=\mathbf{I}} \frac{\mathrm{tr}\left(\mathbf{P}^\top(\bar{\mathbf{Z}}\mathbf{G}_b\bar{\mathbf{Z}}^\top)\mathbf{P}\right)}{\mathrm{tr}\left(\mathbf{P}^\top(\mathbf{Z}\mathbf{G}_w\mathbf{Z}^\top)\mathbf{P}\right)}. \end{aligned} \tag{10}$$

The detailed mathematical deductions are put into the Appendix.

Denote $\lambda_d\left(\mathbf{Z}'\mathbf{G}_w(\mathbf{Z}')^\top\right)$ and $\lambda_d\left(\bar{\mathbf{Z}}'\mathbf{G}_b(\bar{\mathbf{Z}}')^\top\right)$ as the d-th largest eigenvalues of $\mathbf{Z}'\mathbf{G}_w(\mathbf{Z}')^\top$ and $\bar{\mathbf{Z}}'\mathbf{G}_b(\bar{\mathbf{Z}}')^\top$, respectively. The minimum bound ($\alpha_{\min}$) and the maximum bound ($\alpha_{\max}$) of the term

$$\frac{\sum_{d=1}^{D'}\mathrm{tr}\left((\mathbf{q}_d^t)^\top(\mathbf{Z}'\mathbf{G}_w(\mathbf{Z}')^\top)\mathbf{q}_d^t\right)}{\sum_{d=D'+1}^{D_{xn}+D_{yn}}\mathrm{tr}\left((\mathbf{q}_d^t)^\top(\bar{\mathbf{Z}}'\mathbf{G}_b(\bar{\mathbf{Z}}')^\top)\mathbf{q}_d^t\right)}, \tag{11}$$



(a) (b)

Fig. 3. Convergence error on the CASIA(B) gait database. (Red broken lines indicate upper and the lower bounds). (a) Dimension $= 15 \times 10$, (b) Dimension $= 20 \times 15$.

satisfy that

$$\alpha_{\min} \leq \frac{\sum_{d=D_{xn}+D_{yn}-D'+1}^{D_{xn}+D_{yn}} \lambda_d \left(\mathbf{Z}'\mathbf{G}_w(\mathbf{Z}')^\top\right)}{\sum_{d=1}^{D_{xn}+D_{yn}-D'} \lambda_d \left(\bar{\mathbf{Z}}'\mathbf{G}_b(\bar{\mathbf{Z}}')^\top\right)},$$

$$\alpha_{\max} \leq \frac{\sum_{d=1}^{D'} \lambda_d \left(\mathbf{Z}'\mathbf{G}_w(\mathbf{Z}')^\top\right)}{\sum_{d=D'+1}^{D_{xn}+D_{yn}} \lambda_d \left(\bar{\mathbf{Z}}'\mathbf{G}_b(\bar{\mathbf{Z}}')^\top\right)}. \tag{12}$$

Therefore, the lower and upper bounds of the objective function are respectively represented by (13) and (14), shown at the bottom of the next page.
Absolutely, both $J\left(\mathbf{P}^t\right)^{lower}$ and $J\left(\mathbf{P}^t\right)^{upper}$ are monotonically increasing functions. With the increase of iteration times, both $J\left(\mathbf{P}^t\right)^{lower}$ and $J\left(\mathbf{P}^t\right)^{upper}$ converge monotonically and approach their maximum values when

$$\mathbf{P} = \arg \max_{\mathbf{P}^\top\mathbf{P}=\mathbf{I}} \frac{\mathrm{tr}\left(\mathbf{P}^\top\bar{\mathbf{Z}}\mathbf{G}_b\bar{\mathbf{Z}}^\top\mathbf{P}\right)}{\mathrm{tr}\left(\mathbf{P}^\top\mathbf{Z}\mathbf{G}_w\mathbf{Z}^\top\mathbf{P}\right)},$$

$$\mathbf{Q} = \arg \max_{\mathbf{Q}^\top\mathbf{Q}=\mathbf{I}} \frac{\mathrm{tr}\left(\mathbf{Q}^\top\bar{\mathbf{Z}}'\mathbf{G}_b(\bar{\mathbf{Z}}')^\top\mathbf{Q}\right)}{\mathrm{tr}\left(\mathbf{Q}^\top\mathbf{Z}'\mathbf{G}_w(\mathbf{Z}')^\top\mathbf{Q}\right)}. \tag{15}$$

Similarly, given $\mathbf{P}^t = \begin{bmatrix} \mathbf{P}_x^t \\ \mathbf{P}_y^t \end{bmatrix} \in \mathbb{R}^{(D_{xm}+D_{ym})\times D'}$, the corresponding function $J\left(\mathbf{Q}^t\right)$ is monotonically convergent with lower bound $J\left(\mathbf{Q}^t\right)^{lower}$ and upper bound $J\left(\mathbf{Q}^t\right)^{upper}$. Since $J\left(\mathbf{P}^t\right)^{lower} < J\left(\mathbf{Q}^t\right)^{lower}$ and $J\left(\mathbf{P}^t\right)^{upper} < J\left(\mathbf{Q}^t\right)^{upper}$, both lower and upper objective function sequences are monotonically bounded as follows

$$J\left(\mathbf{P}^t\right)^{lower} < J\left(\mathbf{Q}^t\right)^{lower} < J\left(\mathbf{P}^{t+1}\right)^{lower} < \cdots \leq J^{lower},$$
$$J\left(\mathbf{P}^t\right)^{upper} < J\left(\mathbf{Q}^t\right)^{upper} < J\left(\mathbf{P}^{t+1}\right)^{upper} < \cdots \leq J^{upper}. \tag{16}$$

As a result, when the difference of J between two successive iterations is smaller than the threshold $\varepsilon = J^{upper} - J^{lower}$ or J reaches the peak, the iterative optimization procedure can be stopped.

In order to empirically check the convergence of the CBDP, we test the error mentioned in Fig. 3 under different reduced dimensions for GEI matrices on the CASIA(B) gait database. $\{\mathbf{P}_x^0, \mathbf{Q}_x^0\}$ and $\{\mathbf{P}_y^0, \mathbf{Q}_y^0\}$ are initialized to be a concatenation matrix composed of both an identity matrix and a matrix with all zeros at iteration 0 for two views. Fig. 3a and 3b show the error with respect to the number of iterations when GEIs of view 90° and view 72° are aligned and respectively transformed into matrices with the dimensions of 15×10 and 20×15. In both cases, CBDP converges over iterations, and the convergence error sequence of CBDP is lower and upper bounded by two monotonically decreasing sequences, which is equivalent to increasing lower and upper bounded objective function sequences.

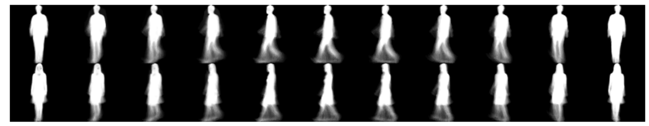


Fig. 4. GEIs from the CASIA(B) database.

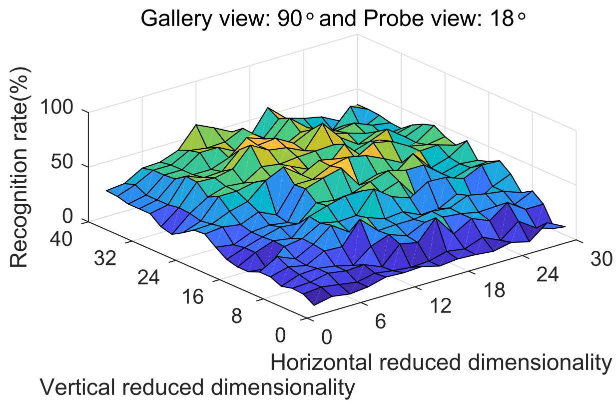
V. Experiments

In this paper, GEI is used as the feature representation from gait silhouettes within a complete walking period [46]. In order to reduce the redundancy of GEIs, 2DPCA [47] can be employed to project the GEIs into a lower 2D subspace. Furthermore, the proposed CBDP method is used to align GEIs across views. Finally, the classification is achieved by using the improved metric learning approach. In this section, we compare the proposed method with the state-of-the-art cross-view gait recognition methods by using all the sequences in a normal walking condition on both CASIA(B) and OU-ISIR gait databases. Our experiments are conducted using Matlab running on a desktop with Intel(R) Core(TM) i5-6300U CPU@2.40Hz and 8GB RAM.

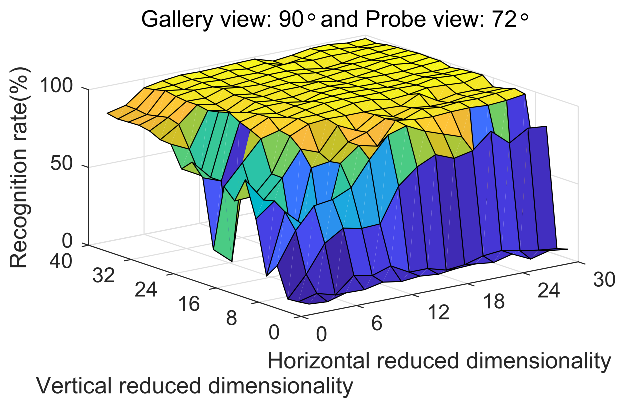
A. Experiments on CASIA(B) Database

CASIA(B) database is the largest multi-view gait dataset up to now, and it contains 13640 sequences from 124 subjects in total. For each subject, gaits are recorded by the cameras from 11 views. In our experiments, the size of GEI is normalized to 64×64 pixels. The GEIs from 11 viewing angles for two subjects are shown in Fig. 4. We repeat the experiments with different data setup for 10 times and the average recognition rate (performance of cross-view gait recognition) is recorded. In each experiment, the data splits are as follows: we randomly separate the database into two non-overlapped groups, i.e. the first group, which contains 3 sequences covering all views of subjects, is taken as the probe set, and the remaining sequences form the second group that is treated as the gallery set. In the gallery set, 60 subjects are randomly selected for training. On average, it takes 9.7s to train the CBDP by using 360 training samples on the CASIA(B) database.

Fig. 5 shows the recognition performance of the proposed method. The size of aligned cross-view gait matrix features are determined by different settings of horizontal and vertical reduced dimensionality. Fig. 5a and 5b correspond to the



(a)



(b)

Fig. 5. The recognition performance with different settings of horizontal reduced dimensionality and vertical reduced dimensionality. (a) Gallery view is 90° and probe view is 18°, (b) Gallery view is 90° and probe view is 72°.

results of the probe views 18° and 72° respectively, while the registered view is 90°. The results suggest that the selections of horizontal reduced dimensionality and vertical reduced dimensionality have a significant effect on the recognition performance especially in the case of a large difference

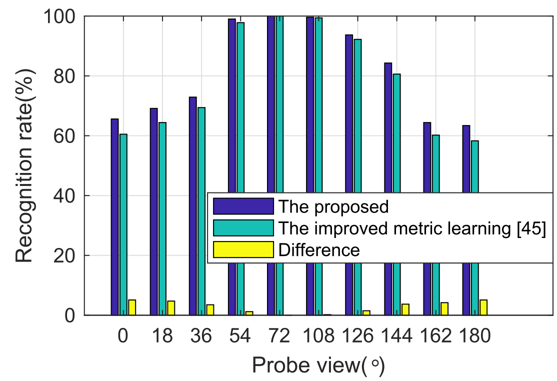


Fig. 6. Performance comparisons of the proposed method and the improved metric learning [45] when the gallery view is 90°.

between the probe and gallery views, since the undulation of the recognition rates is fierce. Therefore, the parameters of horizontal reduced dimensionality and vertical reduced dimensionality are chosen through the cross-validation. We also test the accuracies of the proposed method with different selections of reduced dimensionality for the preprocessing of 2DPCA, and we can find that the results are generally satisfactory when the reduced dimensionality is larger than 20×15. Therefore, we empirically choose the reduced dimensionality of 20×15 for 2DPCA to balance the computational efficiency and computational accuracy.

We also assess the significance of cross-view GEIs alignment. That is, we testify the role of proposed method and the improved metric learning [45]. Fig. 6 shows the comparisons of recognition rates for each probe view of {0°, 18°, 36°, 54°, 72°, 108°, 126°, 144°, 162°, 180°} when the gallery view is 90°."Difference" means the difference between the recognition rate of the proposed method and the improved metric learning [45]. The results demonstrate that GEIs alignment (with CBDP) can greatly improve the recognition accuracy under a larger view difference.

$$J\left(\mathbf{P}^t\right)^{lower} = \frac{\dfrac{\mathrm{tr}\left(\mathbf{P}^\top(\bar{\mathbf{Z}}\mathbf{G}_b\bar{\mathbf{Z}}^\top)\mathbf{P}\right)}{\mathrm{tr}\left(\mathbf{P}^\top(\mathbf{Z}\mathbf{G}_w\mathbf{Z}^\top)\mathbf{P}\right)} - \left\{\alpha_{\min} + \dfrac{\sum_{d=D'+1}^{D_{xn}+D_{yn}} \mathrm{tr}\left((\mathbf{q}_d^t)^\top(\mathbf{Z}'\mathbf{G}_w(\mathbf{Z}')^\top)\mathbf{q}_d^t\right)}{\sum_{d=D'+1}^{D_{xn}+D_{yn}} \mathrm{tr}\left((\mathbf{q}_d^t)^\top(\bar{\mathbf{Z}}'\mathbf{G}_b(\bar{\mathbf{Z}}')^\top)\mathbf{q}_d^t\right)}\right\}^{-1}}{1 - \left\{\alpha_{\max}\dfrac{\sum_{d=D'+1}^{D_{xn}+D_{yn}} \mathrm{tr}\left((\mathbf{q}_d^t)^\top(\bar{\mathbf{Z}}'\mathbf{G}_b(\bar{\mathbf{Z}}')^\top)\mathbf{q}_d^t\right)}{\sum_{d=D'+1}^{D_{xn}+D_{yn}} \mathrm{tr}\left((\mathbf{q}_d^t)^\top(\mathbf{Z}\mathbf{G}_w\mathbf{Z}^\top)\mathbf{q}_d^t\right)} + 1\right\}^{-1}}, \tag{13}$$

$$J\left(\mathbf{P}^t\right)^{upper} = \frac{\dfrac{\mathrm{tr}\left(\mathbf{P}^\top(\bar{\mathbf{Z}}\mathbf{G}_b\bar{\mathbf{Z}}^\top)\mathbf{P}\right)}{\mathrm{tr}\left(\mathbf{P}^\top(\mathbf{Z}\mathbf{G}_w\mathbf{Z}^\top)\mathbf{P}\right)} - \left\{\alpha_{\max} + \dfrac{\sum_{d=D'+1}^{D_{xn}+D_{yn}} \mathrm{tr}\left((\mathbf{q}_d^t)^\top(\mathbf{Z}'\mathbf{G}_w(\mathbf{Z}')^\top)\mathbf{q}_d^t\right)}{\sum_{d=D'+1}^{D_{xn}+D_{yn}} \mathrm{tr}\left((\mathbf{q}_d^t)^\top(\bar{\mathbf{Z}}'\mathbf{G}_b(\bar{\mathbf{Z}}')^\top)\mathbf{q}_d^t\right)}\right\}^{-1}}{1 - \left\{\alpha_{\min}\dfrac{\sum_{d=D'+1}^{D_{xn}+D_{yn}} \mathrm{tr}\left((\mathbf{q}_d^t)^\top(\bar{\mathbf{Z}}'\mathbf{G}_b(\bar{\mathbf{Z}}')^\top)\mathbf{q}_d^t\right)}{\sum_{d=D'+1}^{D_{xn}+D_{yn}} \mathrm{tr}\left((\mathbf{q}_d^t)^\top(\mathbf{Z}\mathbf{G}_w\mathbf{Z}^\top)\mathbf{q}_d^t\right)} + 1\right\}^{-1}}. \tag{14}$$

TABLE III

RESULTS OF CROSS-VIEW GAIT RECOGNITION ON THE CASIA(B) DATABASE(%)

		Gallery view (°)										
		0	18	36	54	72	90	108	126	144	162	180
Probe view (°)	0	-	96.7	82.2	71.7	60.0	60.5	56.7	68.9	76.1	88.3	96.1
	18	98.9	-	98.9	93.9	71.7	64.4	58.9	63.9	66.7	73.3	90.0
	36	80.0	97.5	-	99.4	80.5	69.4	67.2	76.1	75.0	69.4	71.7
	54	59.4	70.0	98.6	-	98.6	97.8	90.0	80.0	74.4	70.0	62.8
	72	58.9	65.6	92.1	97.5	-	100.0	93.3	91.1	72.2	60.5	59.4
	90	61.1	65.0	67.8	92.8	99.4	-	99.4	96.7	71.9	68.3	57.2
	108	61.7	66.1	70.6	82.7	85.3	99.4	-	99.7	88.9	72.2	63.9
	126	68.9	64.4	70.0	76.7	87.8	92.2	96.4	-	99.7	83.3	60.0
	144	68.3	73.9	73.9	77.2	79.4	80.6	90.6	99.4	-	98.3	90.0
	162	91.1	86.1	75.0	60.8	61.4	60.2	67.8	87.8	98.5	-	90.0
	180	96.4	94.4	79.4	60.5	60.5	58.3	60.6	66.7	81.1	96.1	-

Table. III reports all the possible cross-view recognition rates of the proposed method. The results suggest that large view differences significantly decrease the accuracy of the proposed method. This is because the GEI appearances have a similar visual effect under a small view change, and when the view difference turns large, their similarity becomes suddenly low. The proposed method achieves very high recognition rates when the view difference is small, i.e., less than 36°. Noticeably, the proposed method achieves the recognition rate that is close to 100% in some cases when the view difference is smaller than 18°. Hence, we certainly assure that the proposed method is robust when the view difference is not beyond 36°. Another interesting fact is that good performances can be achieved when the views of the gallery GEIs and probe GEIs are under complementary angles, such as 0° versus 180° as well as 36° versus 144°. Because we can clearly capture more jointly discriminative information when gaits are recorded under the complementary view.

We also compare the proposed method with several existing methods for cross-view gait recognition task: 1) GEI [23], 2) Complete canonical correlation analysis (C3A) [21], 3) Correlated Motion Co-Clustering (CMCC) [20], 4) Truncated SVD (TSVD) [16], 5) VTM+Quality Measures (VTMQ) [17], 6) SVR [18], 7) GEINet [15] and 8) Deep CNNs [34]. For a fair comparison, all the methods are evaluated under the data splits as the above-mentioned for the proposed CBDP. Fig. 7 illustrates the recognition rates for four probe views (0°, 18°, 162°, and 180°) by using nine different methods. From the results, we can observe the following facts:

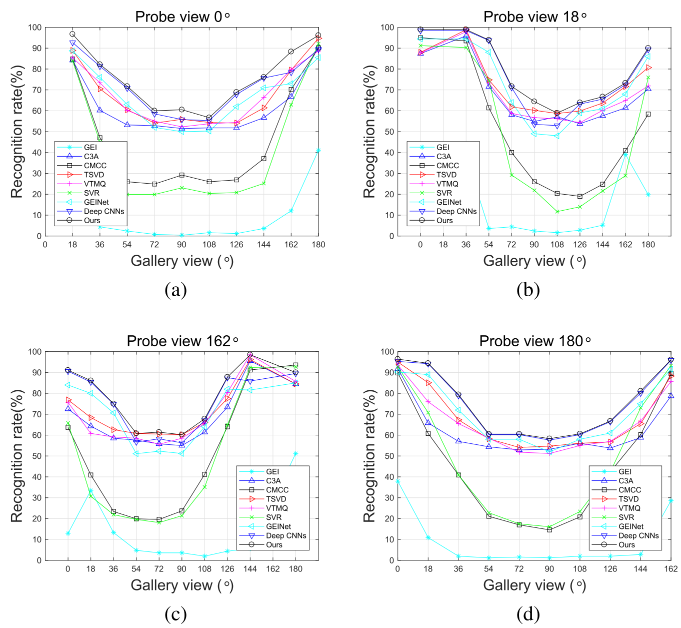


Fig. 7. Performance comparisons on the CASIA(B) gait database. (a) Probe view is 0°, (b) Probe view is 18°, (c) Probe view is 162°, (d) Probe view is 180°.

1) The proposed method is the most robust method, and always achieves higher recognition accuracies than others under both small and large cross-view differences. This indicates that learning direct coupled distance metric for aligning GEIs is beneficial to extracting the common feature from the cross-view GEIs.
2) The proposed method is better than C3A and CMCC. This is because C3A can maximize the correlation of the vectorized GEIs across views, and it ignores the missing of spatial information carried by GEI. However, the proposed method can make full use of this information to align GEIs across different views. Although CMCC considers the clustering relationship of sub-region of GEIs across views, it is difficult to accurately estimate the strict correspondence of optimal sub-regions across views when the view difference is large.
3) The proposed method outperforms VTM methods, such as TSVD, VTMQ and SVR, which are all reconstruction-based methods. In addition, TSVD and VTMQ only decompose view-independence and individual-independent information, and they are lack of discriminant analysis. As a result of that the gait information of virtual view synthesized by another view always appears differently from the reference, and the performance of SVR is worse than the proposed.
4) The proposed method obviously outperforms GEINet [15], and slightly outperforms Deep CNNs [34]. GEINet [15] is based on one of the simplest CNNs, and it has one input GEI, and the number of nodes in the final layer equals to the number of training samples. In Deep CNNs [34], pairs of GEIs are fed into the network to detect the most discriminative changes of

TABLE IV
DETAILED GALLERY AND PROBE DATASETS

	Training	Gallery	Probe
Test1-1	CV01	One angle view GEIs from CV02	Other angle view GEIs from CV02
Test1-2	CV02	One angle view GEIs from CV01	Other angle view GEIs from CV01
Test2-1	CV03	One angle view GEIs from CV04	Other angle view GEIs from CV04
Test2-2	CV04	One angle view GEIs from CV03	Other angle view GEIs from CV03
Test3-1	CV05	One angle view GEIs from CV06	Other angle view GEIs from CV06
Test3-2	CV06	One angle view GEIs from CV05	Other angle view GEIs from CV05
Test4-1	CV07	One angle view GEIs from CV08	Other angle view GEIs from CV08
Test4-2	CV08	One angle view GEIs from CV07	Other angle view GEIs from CV07
Test5-1	CV09	One angle view GEIs from CV10	Other angle view GEIs from CV10
Test5-2	CV10	One angle view GEIs from CV09	Other angle view GEIs from CV09

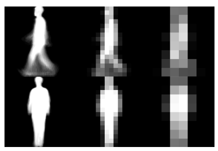


Fig. 8. GEIs with the resolutions of 64×64, 16×16 and 8×8.

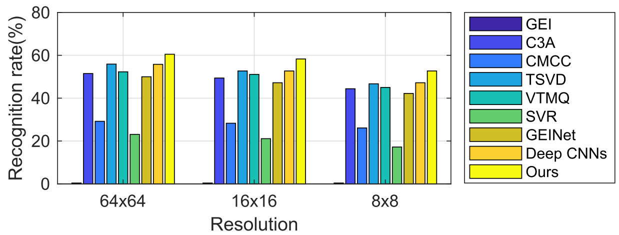


Fig. 9. Comparison of cross-view gait recognition against different resolutions on the CASIA(B) database.

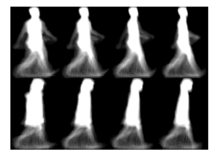


Fig. 10. GEIs from the OU-ISIR database.

TABLE V
RESULTS OF CROSS-VIEW GAIT RECOGNITION ON THE OU-ISIR DATABASE(%)

		Gallery view (°)			
		55	65	75	85
Probe view (°)	55	-	100.0	97.5	90.8
	65	99.8	-	100.0	98.8
	75	100.0	100.0	-	99.9
	85	92.1	98.8	100.0	-

gait features. Such deep learning methods are inferior to the proposed method, and the reason is that the limited labeled data in gait dataset easily lead to the overfitting of CNN model.

In gait recognition, the image is usually captured without targets' cooperation, which usually leads to poor quality samples. Here, we evaluate the robustness of our proposed method to low resolution. We down-sample the GEIs from CASIA(B) database into two low-resolutions, i.e. 64×64, 16×16 and 8×8 (see Fig. 8), and test the performances when taking the samples from 0° as probe and the samples from 90° as gallery. Fig. 9 shows the performances. Our proposed method achieves 60.5%, 58.3% and 52.7% accuracy on the resolutions of 64×64, 16×16, 8×8, respectively. It is easy to observe that the proposed method is robust to low resolution scenarios.

B. Experiments on OU-ISIR Database

The OU-ISIR large population gait database contains 1912 subjects with ages ranging from 1 to 94 years old, and each of them is captured from 4 different observation angles of 55°, 65°, 75° and 85°. This database is equally divided into two sets randomly for 5 times. Thus, the cross-view GEIs from 956 subjects are used for training, and the remaining 956 subjects for testing. It averagely takes 1143.5s to train the CBDP by using 3824 training samples on the OU-ISIR database. Each testing subject's one angle view GEIs are used as register samples, and other angle view GEIs are used as query samples. Table IV lists the detailed gallery and probe datasets used to reliably evaluate the accuracy of the proposed method. For each pair of views, we test the recognition rates for 10 times, and report the average recognition rate over these 10 runs. In our experiments, the size of GEIs is normalized to 64×44 pixels. The GEIs from 4 views for two subjects are shown in Fig. 10. Though the variation of views in the OU-ISIR database is smaller than that of the CASIA(B) database, this database allows us to compare the recognition performance among related cross-view gait recognition methods due to the large number of subjects and its wide range of age variations.

We also empirically choose the reduced dimensionality of 20×15 for 2DPCA and the same aligned dimensionality of 20×15 for the proposed CBDP. Table V reports the recognition rates of the proposed method under all the investigated pairs of views. Noticeably, we obtain recognition accuracies higher than 90%. There are two reasons for better performance in OU-ISIR gait database: 1) OU-ISIR gait database consists of a larger number of training samples which can help avoid the over-fitting problem; 2) OU-ISIR gait database's largest view difference is not larger than 30°(view changes range from 55° to 85°). However, there are some failure cases. For example, the 2-nd sample of the 421-st person is mis-identified as the 680-th person when the gallery view is 75°

$$
\begin{aligned}
\mathbf{P}^{t*} &= \arg\max_{\mathbf{P}^\top\mathbf{P}=\mathbf{I}} \frac{\sum_{i,j=1}^{C} \left\| \mathbf{P}_x^\top \bar{\mathbf{X}}^{(i)} \mathbf{Q}_x - \mathbf{P}_y^\top \bar{\mathbf{Y}}^{(j)} \mathbf{Q}_y \right\|_F^2 \mathbf{S}(i,j)}{\sum_{c=1}^{C} \sum_{i,j=1}^{N_c} \left\| \mathbf{P}_x^\top \mathbf{X}_i^{(c)} \mathbf{Q}_x - \mathbf{P}_y^\top \mathbf{Y}_j^{(c)} \mathbf{Q}_y \right\|_F^2 \mathbf{W}^{(c)}(i,j)} \\
&= \arg\max_{\mathbf{P}^\top\mathbf{P}=\mathbf{I}} \frac{\mathrm{tr}\left(\sum_{i,j=1}^{C} \mathbf{P}^\top \bar{\mathbf{Z}}^{(i)} \mathbf{Q}\mathbf{Q}^\top (\bar{\mathbf{Z}}^{(j)})^\top \mathbf{P}\mathbf{S}(i,j) \right)}{\mathrm{tr}\left(\sum_{c=1}^{C} \sum_{i,j=1}^{N_c} \mathbf{P}^\top \mathbf{Z}_i^{(c)} \mathbf{Q}\mathbf{Q}^\top (\mathbf{Z}_j^{(c)})^\top \mathbf{P}\mathbf{W}^{(c)}(i,j) \right)} \\
&= \arg\max_{\mathbf{P}^\top\mathbf{P}=\mathbf{I}} \frac{\mathrm{tr}\left\{ \mathbf{P}^\top \left[\sum_{i,j=1}^{C} \bar{\mathbf{Z}}^{(i)} \left(\mathbf{I} - \sum_{d=D'+1}^{D_{xn}+D_{yn}} \mathbf{q}_d^t (\mathbf{q}_d^t)^\top \right) (\bar{\mathbf{Z}}^{(j)})^\top \mathbf{S}(i,j) \right] \mathbf{P} \right\}}{\mathrm{tr}\left\{ \mathbf{P}^\top \left[\sum_{c=1}^{C} \sum_{i,j=1}^{N_c} \mathbf{Z}_i^{(c)} \left(\mathbf{I} - \sum_{d=D'+1}^{D_{xn}+D_{yn}} \mathbf{q}_d^t (\mathbf{q}_d^t)^\top \right) (\mathbf{Z}_j^{(c)})^\top \mathbf{W}^{(c)}(i,j) \right] \mathbf{P} \right\}} \\
&= \arg\max_{\mathbf{P}^\top\mathbf{P}=\mathbf{I}} \frac{\mathrm{tr}\left(\mathbf{P}^\top (\bar{\mathbf{Z}}\mathbf{G}_b \bar{\mathbf{Z}}^\top) \mathbf{P} \right) - \sum_{d=D'+1}^{D_{xn}+D_{yn}} \mathrm{tr}\left((\mathbf{q}_d^t)^\top (\bar{\mathbf{Z}}'\mathbf{G}_b (\bar{\mathbf{Z}}')^\top) \mathbf{q}_d^t \right)}{\mathrm{tr}\left(\mathbf{P}^\top (\mathbf{Z}\mathbf{G}_w \mathbf{Z}^\top) \mathbf{P} \right) - \sum_{d=D'+1}^{D_{xn}+D_{yn}} \mathrm{tr}\left((\mathbf{q}_d^t)^\top (\mathbf{Z}'\mathbf{G}_w (\mathbf{Z}')^\top) \mathbf{q}_d^t \right)} \\
&= \arg\max_{\mathbf{P}^\top\mathbf{P}=\mathbf{I}} \frac{\left(\frac{\mathrm{tr}\left(\mathbf{P}^\top (\bar{\mathbf{Z}}\mathbf{G}_b \bar{\mathbf{Z}}^\top)\mathbf{P}\right)}{\mathrm{tr}\left(\mathbf{P}^\top (\mathbf{Z}\mathbf{G}_w \mathbf{Z}^\top)\mathbf{P}\right)} - \left\{ \frac{\sum_{d=1}^{D'} \mathrm{tr}\left((\mathbf{q}_d^t)^\top (\mathbf{Z}'\mathbf{G}_w (\mathbf{Z}')^\top)\mathbf{q}_d^t\right)}{\sum_{d=D'+1}^{D_{xn}+D_{yn}} \mathrm{tr}\left((\mathbf{q}_d^t)^\top (\bar{\mathbf{Z}}'\mathbf{G}_b (\bar{\mathbf{Z}}')^\top)\mathbf{q}_d^t\right)} + \frac{\sum_{d=D'+1}^{D_{xn}+D_{yn}} \mathrm{tr}\left((\mathbf{q}_d^t)^\top (\mathbf{Z}'\mathbf{G}_w (\mathbf{Z}')^\top)\mathbf{q}_d^t\right)}{\sum_{d=D'+1}^{D_{xn}+D_{yn}} \mathrm{tr}\left((\mathbf{q}_d^t)^\top (\bar{\mathbf{Z}}'\mathbf{G}_b (\bar{\mathbf{Z}}')^\top)\mathbf{q}_d^t\right)} \right\}^{-1} \right)}{1 - \left\{ \frac{\sum_{d=1}^{D'} \mathrm{tr}\left((\mathbf{q}_d^t)^\top (\mathbf{Z}'\mathbf{G}_w (\mathbf{Z}')^\top)\mathbf{q}_d^t\right)}{\sum_{d=D'+1}^{D_{xn}+D_{yn}} \mathrm{tr}\left((\mathbf{q}_d^t)^\top (\bar{\mathbf{Z}}'\mathbf{G}_b (\bar{\mathbf{Z}}')^\top)\mathbf{q}_d^t\right)} \frac{\sum_{d=D'+1}^{D_{xn}+D_{yn}} \mathrm{tr}\left((\mathbf{q}_d^t)^\top (\bar{\mathbf{Z}}'\mathbf{G}_b (\bar{\mathbf{Z}}')^\top)\mathbf{q}_d^t\right)}{\sum_{d=D'+1}^{D_{xn}+D_{yn}} \mathrm{tr}\left((\mathbf{q}_d^t)^\top (\mathbf{Z}\mathbf{G}_w \mathbf{Z}^\top)\mathbf{q}_d^t\right)} + 1 \right\}^{-1}} \\
&= \arg\max_{\mathbf{P}^\top\mathbf{P}=\mathbf{I}} \frac{\mathrm{tr}\left(\mathbf{P}^\top (\bar{\mathbf{Z}}\mathbf{G}_b \bar{\mathbf{Z}}^\top) \mathbf{P} \right)}{\mathrm{tr}\left(\mathbf{P}^\top (\mathbf{Z}\mathbf{G}_w \mathbf{Z}^\top) \mathbf{P} \right)}
\end{aligned}
$$

and the probe view is 65°. The root reasons for the failures are 1) the view difference of positive pairs is slightly larger than that of the negative pairs, though the view difference of positive and negative samples from the same camera is assumed to be same. This is because different people cannot walk strictly along the same path. 2) Human is a non-rigid object whose gait patterns are highly influenced by one's pose. 3) The matching accuracy is highly influenced by the quality of GEI, which is constructed by human silhouettes. Though we can extract good silhouettes in most cases, there are some failures, especially when the image is captured in occluded scenarios.

We compare the proposed method with GEI [23], 2) C3A [21], 3) CMCC [20], 4) TSVD [16], 5) VTMQ [17], 6) SVR [18], 7) GEINet [15] and 8) Deep CNNs [34]. Fig. 11 shows the recognition rates for four probe views (55°, 65°, 75°, and 85°) generated by all nine different methods. From the results, we can observe that the proposed method achieves the highest recognition accuracy in most cases, and it achieves the recognition rate that is larger than 95% in most cases. The proposed method is significantly superior to GEI, CMCC, TSVD, VTMQ and SVR. However, the accuracy of the proposed method is occasionally a little lower than C3A. Because C3A is applicable to cross-view gait recognition with small difference in views, and the advantage of GEIs matrix alignment is less obvious when the cross-view difference is small. Moreover, the proposed approach outperforms deep learning-based methods, i.e., GEINet [15] and Deep CNNs [34], due to its benefits on small-size dataset when compared with deep learning-based approaches.

We also evaluate the performance of the proposed method against low resolution on OU-ISIR database. We down-sample the image samples from OU-ISIR database into two scales: 16×11 and 8×5 (see Fig. 12). Fig. 13 shows the results of which gallery view is 55° and probe view is 85°. Our proposed method achieves the accuracies of 92.1%, 90.9% and 74.2% on the resolutions of 64×44, 16×11 and 8×5, respectively. It shows that our proposed method is not sensitive to the variation of resolution to some extent. However, the performances of all methods drop drastically

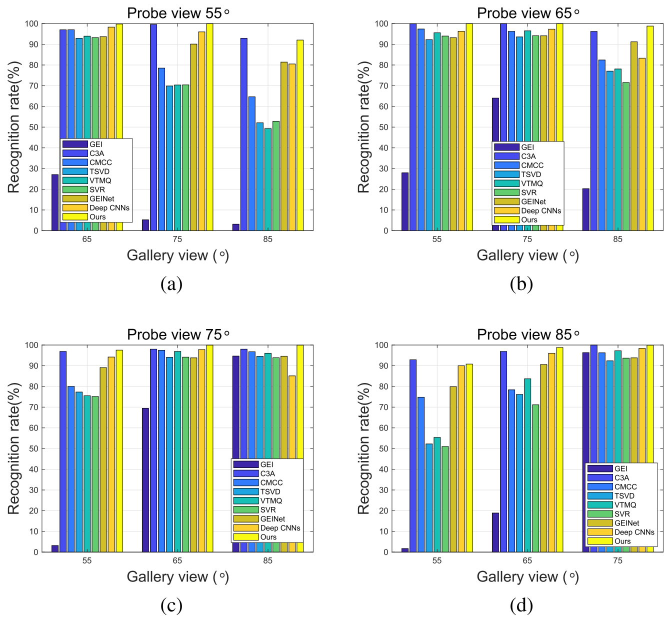


Fig. 11. Performance comparisons of various methods on the OU-ISIR gait database. (a) Probe view is 55°, (b) Probe view is 65°, (c) Probe view is 75°, (d) Probe view is 85°.

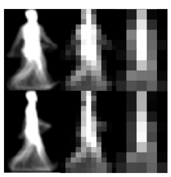


Fig. 12. GEIs with the resolutions of 64×44, 16×11 and 8×5.

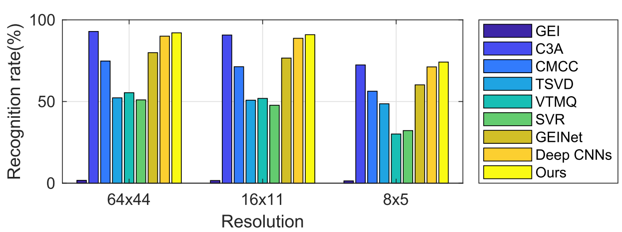


Fig. 13. Comparison of cross-view gait recognition against different resolutions on the OU-ISIR database.

when the resolution is very low. This is because too much useful information is missing in such low resolution GEIs.

VI. Conclusion

In this work, we propose a Coupled bilinear discriminant projection (CBDP) method for aligning gait images across different views. CBDP learns a common lower-dimensional discriminant subspace that effectively links the gaits across views. It can directly deal with GEIs, therefore the spatial structure of GEIs can be preserved, and at the same time, the under-sample problem can be avoided. The upper and lower bounds of the objective function sequence of CBDP are given, and to our best knowledge, this is the first detailed analysis on the convergence of CBDP with a trace ratio and four optimized transformation matrices. Through a series of experiments on the CASIA(B) and OU-ISIR gait databases, we can see that GEIs alignment by the CBDP can significantly improve the results that are without alignment. Furthermore, the proposed method is superior to other state-of-the-art cross-view gait recognition methods. CBDP needs computing $C_n^2 = \frac{n!}{2!(n-2)!}$ projection matrices for n different views. When $n = 11$ in the CASIA(B), CBDP needs $11 \times 10/2 = 55$ projection matrices. In the future, we will potentially extend the proposed CBDP to deep learning-based models such as [43] to handle an arbitrary number (n) of views with certain number projection matrices.

Appendix

In this appendix, we show how to obtain the corresponding optimal $\mathbf{P}^t = \begin{bmatrix} \mathbf{P}_x^t \\ \mathbf{P}_y^t \end{bmatrix}$ at iteration t. The equation can be derived, as shown at the top of the previous page.

References

[1] G. Ariyanto and M. S. Nixon, "Model-based 3D gait biometrics," in *Proc. Int. Joint Conf. Biometrics*, Oct. 2011, pp. 1–7.

[2] X. Ben, P. Zhang, Z. Lai, R. Yan, X. Zhai, and W. Meng, "A general tensor representation framework for cross-view gait recognition," *Pattern Recognit.*, 2019, doi: 10.1016/j.patcog.2019.01.017.

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

[4] H. El-Alfy, I. Mitsugami, and Y. Yagi, "Gait recognition based on normal distance maps," *IEEE Trans. Cybern.*, vol. 48, no. 5, pp. 1526–1539, May 2018, doi: 10.1109/TCYB.2017.2705799.

[5] S. Sarkar, P. J. Phillips, Z. Liu, I. R. Vega, P. Grother, and K. W. Bowyer, "The humanID gait challenge problem: Data sets, performance, and analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 2, pp. 162–177, Feb. 2005.

[6] G. Zhao, G. Liu, H. Li, and M. Pietikainen, "3D gait recognition using multiple cameras," in *Proc. 7th Int. Conf. Autom. Face Gesture Recognit. (FGR)*, Apr. 2006, pp. 529–534.

[7] J. Luo, J. Tang, T. Tjahjadi, and X. Xiao, "Robust arbitrary view gait recognition based on parametric 3D human body reconstruction and virtual posture synthesis," *Pattern Recognit.*, vol. 60, pp. 361–377, Dec. 2016.

[8] X. Chen, T. Yang, and J. Xu, "Cross-view gait recognition based on human walking trajectory," *J. Vis. Commun. Image Represent.*, vol. 25, no. 8, pp. 1842–1855, 2014.

[9] D. López-Fernández, F. J. Madrid-Cuevas, A. Carmona-Poyato, M. J. Marín-Jiménez, R. Muñoz-Salinas, and R. Medina-Carnicer, "Viewpoint-independent gait recognition through morphological descriptions of 3D human reconstructions," *Image Vis. Comput.*, vols. 48–49, pp. 1–13, Apr./May 2016.

[10] M. Goffredo, I. Bouchrika, J. N. Carter, and M. S. Nixon, "Self-calibrating view-invariant gait biometrics," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 40, no. 4, pp. 997–1008, Aug. 2010.

[11] F. Jean, A. B. Albu, and R. Bergevin, "Towards view-invariant gait modeling: Computing view-normalized body part trajectories," *Pattern Recognit.*, vol. 42, no. 11, pp. 2936–2949, 2009.

[12] W. Zeng and C. Wang, "View-invariant gait recognition via deterministic learning," *Neurocomputing*, vol. 175, pp. 324–335, Jan. 2016.

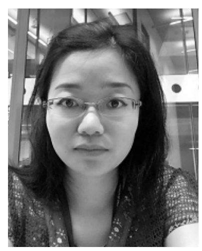
[13] W. Kusakunniran, Q. Wu, J. Zhang, and H. Li, "Cross-view and multi-view gait recognitions based on view transformation model using multi-layer perceptron," *Pattern Recognit. Lett.*, vol. 33, no. 7, pp. 882–889, 2012.

[14] J. Lu, G. Wang, and P. Moulin, "Human identity and gender recognition from gait sequences with arbitrary walking directions," *IEEE Trans. Inf. Forensics Security*, vol. 9, no. 1, pp. 51–61, Jan. 2014.

[15] K. Shiraga, Y. Makihara, D. Muramatsu, T. Echigo, and Y. Yagi, "GEINet: View-invariant gait recognition using a convolutional neural network," in *Proc. Int. Conf. Biometrics (ICB)*, Jun. 2016, pp. 1–8.

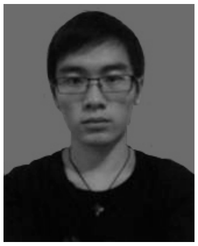
[16] W. Kusakunniran, Q. Wu, H. Li, and J. Zhang, "Multiple views gait recognition using view transformation model based on optimized gait energy image," in *Proc. IEEE 12th Int. Conf. Comput. Vis. Workshops (ICCV Workshops)*, Sep./Oct. 2009, pp. 1058–1064.

[17] D. Muramatsu, Y. Makihara, and Y. Yagi, "View transformation model incorporating quality measures for cross-view gait recognition," *IEEE Trans. Cybern.*, vol. 46, no. 7, pp. 1602–1615, Jul. 2016.
[18] W. Kusakunniran, Q. Wu, J. Zhang, and H. Li, "Support vector regression for multi-view gait recognition based on local motion feature selection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 974–981.
[19] W. Kusakunniran, Q. Wu, J. Zhang, and H. Li, "Gait recognition under various viewing angles based on correlated motion regression," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 6, pp. 966–980, Jun. 2012.
[20] W. Kusakunniran, Q. Wu, J. Zhang, H. Li, and L. Wang, "Recognizing gaits across views through correlated motion co-clustering," *IEEE Trans. Image Process.*, vol. 23, no. 2, pp. 696–709, Feb. 2014.
[21] X. Xing, K. Wang, T. Yan, and Z. Lv, "Complete canonical correlation analysis with application to multi-view gait recognition," *Pattern Recognit.*, vol. 50, pp. 107–117, Feb. 2016.
[22] A. Mansur, Y. Makihara, D. Muramatsu, and Y. Yagi, "Cross-view gait recognition using view-dependent discriminative analysis," in *Proc. IEEE Int. Joint Conf. Biometrics*, Sep. 2014, pp. 1–8.
[23] S. Yu, D. Tan, and T. Tan, "A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition," in *Proc. 18th Int. Conf. Pattern Recognit. (ICPR)*, vol. 4, Aug. 2006, pp. 441–444.
[24] H. Iwama, M. Okumura, Y. Makihara, and Y. Yagi, "The OU-ISIR gait database comprising the large population dataset and performance evaluation of gait recognition," *IEEE Trans. Inf. Forensics Security*, vol. 7, no. 5, pp. 1511–1521, Oct. 2012.
[25] A. F. Bobick and J. W. Davis, "The recognition of human movement using temporal templates," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 3, pp. 257–267, Mar. 2001.
[26] J. Liu and N. Zheng, "Gait history image: A novel temporal template for gait recognition," in *Proc. IEEE Int. Conf. Multimedia Expo*, Jul. 2007, pp. 663–666.
[27] T. H. W. Lam and R. S. T. Lee, "A new representation for human gait recognition: Motion silhouettes image (MSI)," in *Proc. Int. Conf. Adv. Biometrics*, 2006, pp. 612–618.
[28] J. Han and B. Bhanu, "Individual recognition using gait energy image," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 2, pp. 316–322, Feb. 2006.
[29] X. Yang, Y. Zhou, T. Zhang, G. Shu, and J. Yang, "Gait recognition based on dynamic region analysis," *Signal Process.*, vol. 88, no. 9, pp. 2350–2356, 2008.
[30] K. Bashir, T. Xiang, and S. Gong, "Gait recognition without subject cooperation," *Pattern Recognit. Lett.*, vol. 31, no. 13, pp. 2052–2060, 2010.
[31] T. H. W. Lam, K. H. Cheung, and J. N. K. Liu, "Gait flow image: A silhouette-based gait representation for human identification," *Pattern Recognit.*, vol. 44, no. 4, pp. 973–987, Apr. 2011.
[32] X. Huang and N. V. Boulgouris, "Gait recognition with shifted energy image and structural feature extraction," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 2256–2268, Apr. 2012.
[33] C. Wang, J. Zhang, L. Wang, J. Pu, and X. Yuan, "Human identification using temporal information preserving gait template," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2164–2176, Nov. 2012.
[34] Z. Wu, Y. Huang, L. Wang, X. Wang, and T. Tan, "A comprehensive study on cross-view gait based human identification with deep CNNs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 2, pp. 209–226, Feb. 2017.
[35] D. Xu, S. Yan, D. Tao, L. Zhang, X. Li, and H.-J. Zhang, "Human gait recognition with matrix representation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 7, pp. 896–903, Jul. 2006.
[36] X. Li, S. Lin, S. Yan, and D. Xu, "Discriminant locally linear embedding with high-order tensor data," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 38, no. 2, pp. 342–352, Apr. 2008.
[37] D. Xu, S. Yan, D. Tao, S. Lin, and H. J. Zhang, "Marginal Fisher analysis and its variants for human gait recognition and content-based image retrieval," *IEEE Trans. Image Process.*, vol. 16, no. 11, pp. 2811–2821, Nov. 2007.
[38] Z. Lai, Y. Xu, Z. Jin, and D. Zhang, "Human gait recognition via sparse discriminant projection learning," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 10, pp. 1651–1662, Oct. 2014.
[39] Y.-L. Chen and C.-T. Hsu, "Multilinear graph embedding: Representation and regularization for images," *IEEE Trans. Image Process.*, vol. 23, no. 2, pp. 741–754, Feb. 2014.
[40] Y. Huang, D. Xu, and F. Nie, "Patch distribution compatible semisupervised dimension reduction for face and human gait recognition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 3, pp. 479–488, Mar. 2012.
[41] J. Tang, J. Luo, T. Tjahjadi, and F. Guo, "Robust arbitrary-view gait recognition based on 3D partial similarity matching," *IEEE Trans. Image Process.*, vol. 26, no. 1, pp. 7–22, Jan. 2017.
[42] Z. Zhang, J. Chen, Q. Wu, and L. Shao, "GII representation-based cross-view gait recognition by discriminative projection with list-wise constraints," *IEEE Trans. Cybern.*, vol. 48, no. 10, pp. 2935–2947, Oct. 2018.
[43] N. Takemura, Y. Makihara, D. Muramatsu, T. Echigo, and Y. Yagi, "On input/output architectures for convolutional neural network-based cross-view gait recognition," *IEEE Trans. Circuits Syst. Video Technol.*, to be published, doi: 10.1109/TCSVT.2017.2760835.
[44] Y. Yang, F. Nie, D. Xu, J. Luo, Y. Zhuang, and Y. Pan, "A multimedia retrieval framework based on semi-supervised ranking and relevance feedback," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 4, pp. 723–742, Apr. 2012.
[45] X. Ben, W. Meng, R. Yan, and K. Wang, "An improved biometrics technique based on metric learning approach," *Neurocomputing*, vol. 97, no. 1, pp. 44–51, 2012.
[46] X. Ben, W. Meng, and R. Yan, "Dual-ellipse fitting approach for robust gait periodicity detection," *Neurocomputing*, vol. 79, no. 3, pp. 173–178, Mar. 2012.
[47] J. Yang, D. Zhang, A. F. Frangi, and J.-Y. Yang, "Two-dimensional PCA: A new approach to appearance-based face representation and recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 1, pp. 131–137, Jan. 2004.



Xianye Ben received the Ph.D. degree in pattern recognition and intelligent systems from the College of Automation, Harbin Engineering University, Harbin, in 2010. She is currently an Associate Professor with the School of Information Science and Engineering, Shandong University, Qingdao, China. She has published over 80 papers in major journals and conferences such as the IEEE TIP, IEEE TCSVT, and PR. Her current research interests include pattern recognition, digital image processing and analysis, and machine learning. She received the Excellent Doctoral Dissertation Award from Harbin Engineering University. She was also enrolled in the Young Scholars Program of Shandong University.

Chen Gong received the B.E. degree from the East China University of Science and Technology in 2010 and the dual Ph.D. degrees from Shanghai Jiao Tong University (SJTU) and the University of Technology Sydney in 2016 and 2017, under the supervision of Prof. J. Yang and Prof. D. Tao, respectively. He was also enrolled in the Summit of the Six Top Talents Program of Jiangsu Province, China, and the Lift Program for Young Talents of China Association for Science and Technology. He is currently a Full Professor with the School of Computer Science and Engineering, Nanjing University of Science and Technology. He has published over 50 technical papers in prominent journals and conferences such as the IEEE TNNLS, the IEEE TIP, the IEEE TCYB, the IEEE TCSVT, the IEEE TMM, the IEEE TITS, CVPR, AAAI, IJCAI, and ICDM. His research interests mainly include machine learning, data mining, and learning-based vision problems. He received the Excellent Doctoral Dissertation Award from SJTU and the Chinese Association for Artificial Intelligence.



Peng Zhang received the B.S. and M.S. degrees in communication engineering from the School of Information Science and Engineering, Shandong University, Jinan, China, in 2013 and 2016, respectively. He is currently pursuing the Ph.D. degree with the Global Big Data and Technologies Centre, University of Technology Sydney, Sydney, Australia. He has published over 10 academic papers in major conferences and journals. He holds several Chinese invention patents. His current research interests include gait recognition, person re-identification, and generative adversarial networks.

Rui Yan received the Ph.D. degree in computer science from the Rensselaer Polytechnic Institute in 2018. He is currently a Data Scientist with Microsoft AI & R, Bellevue, WA, USA. He has published over 20 conference and journal papers. His research interests include knowledge graph, pattern recognition, and machine learning.

Qiang Wu (M'02–SM'10) received the B.Eng. and M.Eng. degrees in electronic engineering from the Harbin Institute of Technology, Harbin, China, in 1996 and 1998, respectively, and the Ph.D. degree in computing science from the University of Technology Sydney, Sydney, Australia, in 2004. He is currently an Associate Professor with the School of Computing and Communications, University of Technology Sydney. His major research interests include computer vision, image processing, pattern recognition, machine learning, and multimedia processing. He has published over 70 refereed papers, including those published in prestigious journals and top international conferences. He has served as the chair and/or a program committee member for a number of international conferences. He has been a Guest Editor of several international journals such as the *Pattern Recognition Letters* and the *International Journal of Pattern Recognition and Artificial Intelligence*.

Weixiao Meng (SM'10) received the B.Eng., M. Eng., and Ph.D. degrees from the Harbin Institute of Technology (HIT), Harbin, China, in 1990, 1995, and 2000, respectively. From 1998 to 1999, he worked at NTT DoCoMo Inc., where he was involved in adaptive array antenna and dynamic resource allocation for beyond 3G, as a Senior Visiting Researcher. He is currently a Full Professor, and also the Vice Dean of the School of Electronics and Information Engineering, HIT. His research interests include broadband wireless communications, space-air-ground integrated networks, and wireless localization technologies. He has published four books and over 260 papers on journals and international conferences. He is the Chair of the IEEE Communications Society Harbin Chapter, a fellow of the China Institute of Electronics, and a Senior Member of the IEEE ComSoc and the China Institute of Communication. In 2005, he was honored Provincial Excellent Returnee, and was selected into the New Century Excellent Talents Plan by the Ministry of Education, China, in 2008, and the Distinguished Academic Leadership of Harbin. He was a recipient of the Chapter of the Year Award, the Asia Pacific Region Chapter Achievement Award, and the Member & Global Activities Contribution Award in 2018. He has acted as the leading TPC Co-Chair of the ChinaCom2011 and ChinaCom2016, the leading Services and Applications Track Co-Chair of the IEEE WCNC2013, the Awards Co-Chair of the IEEE ICC2015 and the Wireless Networking Symposia Co-Chair of the IEEE Globecom2015, the AHSN Symposia Co-Chair of the IEEE Globecom2018, and the leading Workshop Co-Chair of the IEEE ICC2019. From 2010 to 2017, he has been an Editorial Board Member of the Wileyâs WCMC Journal, an Area Editor of PHYCOM journal, from 2014 to 2016, an Editorial Board of the IEEE COMMUNICATIONS SURVEYS AND TUTORIALS, from 2014 to 2017, and the IEEE WIRELESS COMMUNICATIONS, since 2015.