

# Coupled Patch Alignment for Matching Cross-View Gaits

Xianye Ben<sup>ID</sup>, *Member, IEEE*, Chen Gong<sup>ID</sup>, *Member, IEEE*, Peng Zhang<sup>ID</sup>, Xitong Jia, Qiang Wu, *Senior Member, IEEE*, and Weixiao Meng<sup>ID</sup>, *Senior Member, IEEE*

**Abstract**—Gait recognition has attracted growing attention in recent years, as the gait of humans has a strong discriminative ability even under low resolution at a distance. Unfortunately, the performance of gait recognition can be largely affected by view change. To address this problem, we propose a coupled patch alignment (CPA) algorithm that effectively matches a pair of gaits across different views. To realize CPA, we first build a certain amount of patches, and each of them is made up of a sample as well as its intra-class and inter-class nearest neighbors. Then, we design an objective function for each patch to balance the cross-view intra-class compactness and the cross-view inter-class separability. Finally, all the local-independent patches are combined to render a unified objective function. Theoretically, we show that the proposed CPA has a close relationship with canonical correlation analysis. Algorithmically, we extend CPA to “multi-dimensional patch alignment” that can handle an arbitrary number of views. Comprehensive experiments on CASIA(B), USF, and OU-ISIR gait databases firmly demonstrate the effectiveness of our methods over other existing popular methods in terms of cross-view gait recognition.

**Index Terms**—Coupled patch alignment, gait recognition, cross-view gait, multi-dimensional patch alignment.

## I. INTRODUCTION

OVER the past few years, biometrics has been broadly applied to social security and personal safety due to

Manuscript received October 14, 2017; revised September 8, 2018 and November 18, 2018; accepted January 10, 2019. Date of publication January 23, 2019; date of current version May 6, 2019. This work was supported in part by the National Key R&D 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 the NSF of Jiangsu Province under Grant BK20171430, in part by the Fundamental Research Funds for the Central Universities under Grant 30918011319, in part by the Summit of the Six Top Talents 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. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Wen Gao. (Corresponding author: Xianye Ben.)

X. Ben and X. Jia are with the School of Information Science and Engineering, Shandong University, Qingdao 266237, China (e-mail: benxianye@mail.com; sdujiaxitong@foxmail.com).

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

P. Zhang and Q. Wu are 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; 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).

Digital Object Identifier 10.1109/TIP.2019.2894362

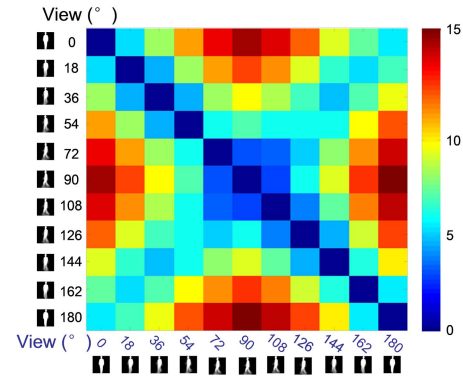


Fig. 1. The Euclidean distance between different pairs of view angles from 0° ~ 180° in terms of the GEIs of a single subject. Different colors indicate different distance values.

a growing necessity in recognition and authentication of individuals. Compared with other biometrics such as face, fingerprint, vein, iris, ear, hand shape, palm print, retina and lip, gait [1]–[5] can be efficiently recognized at a distance without subjects’ cooperation, so it has gained much attention for a variety of practical usages.

The viewing angle is a challenging factor for accurate gait recognition [6]. The perceptions of the same subject’s gait from different viewing directions may differ dramatically. As a result, the performances of most existing cross-view gait recognition techniques [6]–[12] are far from perfect, and the larger the variation of viewing direction is, the worse the recognition performance will be. For example, Gait energy Image (GEI) [13] is an effective and well-known gait feature representation. However, under large view variations GEIs will generate very different representations for the same subject as shown in Fig. 1. To be specific, the Euclidean distance between GEIs of one view and other views shows that the GEI feature is very sensitive to view change. Besides, the GEI similarity of a single subject sharply decreases as the angle between different views becomes larger. Therefore, how to find robust and discriminative representations that enlarge the margin between different subjects and meanwhile narrow the variations of the same subject is a critical but challenging problem in cross-view gait recognition.

Up to now, a range of advanced methods have been proposed to solve the cross-view recognition problem, which can be attributed into three categories. The first category is to construct 3D gait information via panoramic cameras or multiple calibrated cameras [14], [15].

However, these 3D-based methods require a complicated setup of controlled multi-cameras, which may not be available in practice. Even such cooperative system is available, the computational burden is quite heavy that further limits its practical usage. The second category is based on view transformation model (VTM). Representative techniques such as single value decomposition (SVD) and regression have been intensively used to generate the gait feature by utilizing the information from the other view [11], [16]–[21]. Although VTM minimizes the error between the transformed gait features and the original ones, it does not take the discriminative factor into consideration. The third category is to extract view-invariant gait feature. The general idea of view-invariant methods is to generate a view-invariant gait feature among cross-view gait data. The most representative method on generating view-invariant gait feature was developed by Goffredo *et al.* [7] who proposed a self-calibration of limbs' pose in the image reference system. However, this method cannot accurately estimate the limbs' pose when the view of input query gait is very different from the gallery gaits or the gait is captured from a frontal view. To alleviate the above problems, techniques of domain transformation [8]–[10], [12], [22], metric learning [23], and deep CNNs [24] have been introduced recently, among which deep CNNs [24] have achieved encouraging recognition accuracy on the cross-view task due to its powerful representation ability. However, the usage of deep CNNs is often hampered by the insufficient labeled cross-view gait data for model training.

The model proposed in this paper belongs to the last category as the methods of this category have exhibited the state-of-the-art performance up to now. All the methods belonging to this category can reduce the gap caused by view difference to some extent, but most of them do not consider the neighboring relationship between the gait features in separate views, so the information potentially shared by different views have not been sufficiently exploited. To address this shortcoming, we propose to use the patch alignment framework [25] that contains part optimization and whole alignment to explore the neighboring information of different gait samples. Specifically, in part optimization, each patch is formed by a gait sample and also its intra-class and inter-class nearest-neighbors. As a result, all these patches can be used to depict the compactness of intra-class gaits and the separability of inter-class gaits under cross-view situations. In the stage of whole alignment, the established patches are further combined to render an optimization problem for cross-view gait recognition. Based on this motivation, we propose a novel supervised algorithm called Coupled Patch Alignment (CPA) to explicitly build the relationship between cross-view gait data. CPA projects the gait features in two different views onto a common subspace such that the distance between the projected samples that are originally neighbors in the same class can be minimized, and the distance between the projected samples in different classes (i.e. subjects) is maximized. Moreover, we generalize CPA to Multi-dimensional Patch Alignment (MPA) to handle the gait recognition with an arbitrary number of views.

It is worthwhile to highlight our main contributions:

- (1) An efficient CPA algorithm which consists of cross-view part optimization and whole alignment is proposed to address cross-view gait recognition. It can be extended to a more general MPA that learns the common discriminant subspace characterizing both the intra-class compactness and the inter-class separability.
- (2) The proposed CPA is theoretically related to the existing Canonical Correlation Analysis (CCA) algorithm.
- (3) The proposed CPA performs superiorly to existing methods under large view variations, even when the observation angle between the gallery and the probe is as large as  $90^\circ$ .

It is worth noting that the differences between Marginal Fisher Analysis (MFA) [26] and CPA are: (1) In terms of intrinsic and penalty graphs design: MFA characterizes the intra-class compactness and the inter-class separability through the intrinsic and penalty graph functions respectively in a single view. However, CPA models the cross-view gait data as a cross-view patch and conducts alignment among heterogeneous data. (2) In terms of motivation: The motivation from MFA is that the paper [26] builds intrinsic graph and penalty graph in order to maintain intra-class compactness and inter-class separability respectively. The proposed CPA not only achieves the best discriminability as marginal fisher criterion in MFA but also eliminates the variance caused by camera view change in gait recognition. (3) In terms of objective function: MFA minimizes the Marginal Fisher Criterion which is quotients of intra-class compactness and inter-class separability. However, the proposed CPA minimizes the difference between intra-class compactness and inter-class separability with a tuning parameter, which avoids the matrix singularity problem due to no inverse operation over a matrix. Also, for CPA, both intrinsic and penalty graphs are constructed by using adjacent neighbor samples from two views. (4) In terms of application: CPA targets on the cross-view gait recognition problem which treats gait images from different views. Such heterogeneous data cannot be handled by the existing MFA.

This paper is organized as follows: In Section II, we provide an overview of existing cross-view gait recognition technologies and the relationship with CCA. In Section III, the gait feature presentation and CPA algorithm are presented, and the use of CPA for matching cross-view gaits is also described. Section IV details MPA which is the extension of CPA. Experimental results on typical practical databases are presented in Section V. Finally, we conclude this paper in Section VI.

## II. RELATED WORKS

### A. Cross-View Gait Recognition

Generally speaking, gait recognition is a challenging task. A person's gait is often affected by health condition, body weight, clothing, shoe, emotion, carrying condition, and walking surface [27]. The performance of gait recognition is very susceptible to different carrying conditions, camera views, ground surfaces, and the appearance under different time interval between the gallery samples and the query ones [27].

TABLE I  
SUMMARY OF CROSS-VIEW GAIT RECOGNITION METHODS

Category	Representative methods
3D gait information construction	[14], [15].
View transformation model	1) SVD: [16], [17], [19], [20], 2) Regression: [11], [18], [21].
View-invariant gait feature	1) Self-calibration: [7], 2) Body part trajectory normalization: [35], 3) Domain transformation: [8]–[10] [12], [22], [36], 4) Sparse reconstruction based metric learning: [23], 5) View-independent representation: GII [37], 6) Deep CNNs: [24], [38], [39].

Matovski *et al.* [28] claimed that if certain factors (*e.g.* clothing and shoes) can be controlled, the time interval will not affect the recognition performance significantly. To address the issue of clothing changes, the recovery model of static parameters regarding the body [29] can be utilized. For the issue of cross walking speed, Kusakunniran *et al.* [30] proposed a higher order shape configuration which can preserve speed-invariant and discriminative information. Huang *et al.* [31] extracted a non-affine component as the cross-speed gait feature from an affine component which was learned by a thin plate spline kernel-based RBF interpolation. To overcome the issue of shoes that may impair the gait recognition performance, Wang *et al.* [32] proposed a new gait signature termed Chrono-Gait Image (CGI). CGI is a multichannel temporal description, and it can be viewed as a pseudo-color image when the number of channels is 3. In addition, several methods have been proposed to overcome the gait recognition problem in the spatially [33] or temporally [34] low resolution situations.

One of the most challenging issues for gait recognition is the dramatic view change for gait feature matching. To overcome this difficulty, a range of methods have been proposed which can be summarized into three categories (see Table I).

The first category, *i.e.* 3D gait information construction, is to construct the 3D gait information via panoramic cameras or multiple calibrated cameras. Zhao *et al.* [15] captured the video sequences from multiple cameras to achieve 3D gait recognition. Sugiura *et al.* [14] reconstructed a 3D gait model from multi-view observation with an omnidirectional camera.

The second category is based on the view transformation model (VTM) which enables a gait feature from one view to be transferred into another view. A lot of techniques have been adopted in this branch, such as SVD and regression. For example, Kusakunniran *et al.* [20] utilized the truncated SVD to transform the gallery gait data and probe gait data into the same direction. Due to the biased dissimilarity score naturally inherited by VTM, some specific measures were incorporated into VTM to quantify the degree of bias [19]. Considering that the training data are embedded in a potential Grassmann manifold, Connie *et al.* [17] proposed the Grassmann View Compensation (GVC) with randomized kernel extreme learning machine to generate a virtual view, so that the missing view

in the sets can be estimated. When the view of the probe gait is very different from that of the training samples, these VTMs often cannot perform well. Therefore, Muramatsu *et al.* [16] proposed Arbitrary VTM (AVTM), which is trained on the 2D gait features generated from 3D gait volume that has the same view with the target. These methods construct a large matrix of which each column contains the gait information of the same individual with different viewpoints, and each row represents the gait information for different individuals but from the same viewpoint. The basic assumption is that such matrix can be decomposed into two matrices which model the view independence and individual independence, respectively. As this assumption lacks theoretical foundation, Kusakunniran *et al.* [21] showed that this assumption can be formulated as a regression problem, so they constructed VTM by Support Vector Regression (SVR) based on the GEI feature. In addition to SVR, Kusakunniran *et al.* [11] also adopted the sparse regression and multilayer perceptron [18] as VTM construction regression functionals.

The third category is to find view-invariant gait feature. Several approaches have been used to extract the gait representation which is robust to view change, such as self-calibration [7], body part trajectory normalization [35], domain transformation [8]–[10], [12], [22], [36], sparse reconstruction based metric learning [23], and deep CNNs [24], [38], [39]. Goffredo *et al.* [7] estimated the lower limbs' pose in the image reference system and reconstructed the limbs' pose for identification in the sagittal plane. Jean *et al.* [35] proposed an approach to estimate fronto-parallel view of body part trajectories. Different from the geometry-based feature [7], [35], Kusakunniran *et al.* [12] transformed each gait sequence from a certain view onto the common canonical view by domain transformation. Domain transformation can not only obtain the canonical view, but also obtain the prototypes of different views, therefore each sample can be expressed as a linear combination of these prototypes in the corresponding views. Recently, several methods have emerged for this purpose, such as joint subspace learning [10], uncorrelated discriminant simplex analysis [22], uncorrelated multilinear sparse local discriminant CCA [8], and Multi-view discriminant analysis (MvDA) [36]. Hu [9] used the Regularized locally tensor discriminant analysis to extract the feature output by enhanced Gabor GEI, and then adopted an aggregation scheme to fuse different features generated by various regularizers of local tensor discriminant analysis. Lu *et al.* [23] proposed a sparse reconstruction based metric learning approach to learn the discriminative gait feature for identification. Zhang *et al.* [37] proposed a novel view-independent gait representation named gait individuality image (GII), and employed list-wise constraints to learn a projection, which can map gait features from different views into a common discriminative subspace. So far, deep CNNs [24] have yielded the highest accuracy on the cross-view gait recognition task. The cross-view pairs can be processed by deep CNNs with an end-to-end manner, so that the CNNs can predict whether the input pairs belong to the same subject or not [24], [38]. Yu *et al.* [39] generated a fake/real discriminator and a fake/real discriminator to obtain



view-invariant gait images containing human identification information.

### B. Relationship With CCA

In this section, we will show that the proposed CPA is related to the existing CCA [40]. Suppose that  $\mathbf{X}^1$  and  $\mathbf{X}^2$  are centralized, CCA aims to find a pair of linear combination coefficients (also called projection directions) in order to maximize the correlation between the linear combinations of the two sets  $\mathbf{X}^1$  and  $\mathbf{X}^2$ . Let  $\mathbf{P}^1$  and  $\mathbf{P}^2$  denote the two projection directions for  $\mathbf{X}^1$  and  $\mathbf{X}^2$ , respectively, therefore the objective function of CCA is

$$\max_{\mathbf{P}^1, \mathbf{P}^2} \frac{(\mathbf{P}^1)^\top \mathbf{X}^1 (\mathbf{X}^2)^\top \mathbf{P}^2}{\sqrt{(\mathbf{P}^1)^\top \mathbf{X}^1 (\mathbf{X}^1)^\top \mathbf{P}^1} \sqrt{(\mathbf{P}^2)^\top \mathbf{X}^2 (\mathbf{X}^2)^\top \mathbf{P}^2}}. \quad (1)$$

Two constraints  $(\mathbf{P}^1)^\top \mathbf{X}^1 (\mathbf{X}^1)^\top \mathbf{P}^1 = 1$ ,  $(\mathbf{P}^2)^\top \mathbf{X}^2 (\mathbf{X}^2)^\top \mathbf{P}^2 = 1$  are introduced to (1) so as to make the optimization results invariant to scale change. Then this problem can be written as

$$\begin{bmatrix} \mathbf{X}^1 (\mathbf{X}^1)^\top \mathbf{P}^1 - \mathbf{X}^1 (\mathbf{X}^2)^\top \mathbf{P}^2 \\ \mathbf{X}^2 (\mathbf{X}^2)^\top \mathbf{P}^2 - \mathbf{X}^2 (\mathbf{X}^1)^\top \mathbf{P}^1 \end{bmatrix} = \lambda \begin{bmatrix} \mathbf{X}^1 (\mathbf{X}^1)^\top \mathbf{P}^1 \\ \mathbf{X}^2 (\mathbf{X}^2)^\top \mathbf{P}^2 \end{bmatrix}. \quad (2)$$

Furthermore, we can decompose (2) into the form

$$\begin{bmatrix} \mathbf{X}^1 \\ \mathbf{X}^2 \end{bmatrix} \begin{bmatrix} \mathbf{I} & -\mathbf{I} \\ -\mathbf{I} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{X}^1 \\ \mathbf{X}^2 \end{bmatrix}^\top \begin{bmatrix} \mathbf{P}^1 \\ \mathbf{P}^2 \end{bmatrix} = \lambda \begin{bmatrix} \mathbf{X}^1 \\ \mathbf{X}^2 \end{bmatrix} \begin{bmatrix} \mathbf{X}^1 \\ \mathbf{X}^2 \end{bmatrix}^\top \begin{bmatrix} \mathbf{P}^1 \\ \mathbf{P}^2 \end{bmatrix}, \quad (3)$$

which is exactly the Lagrangian expression of the following (4)

$$J(\mathbf{P}^1, \mathbf{P}^2) = \text{tr} \left( \begin{bmatrix} \mathbf{P}^1 \\ \mathbf{P}^2 \end{bmatrix}^\top \begin{bmatrix} \mathbf{X}^1 \\ \mathbf{X}^2 \end{bmatrix} \begin{bmatrix} \mathbf{I} & -\mathbf{I} \\ -\mathbf{I} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{X}^1 \\ \mathbf{X}^2 \end{bmatrix}^\top \begin{bmatrix} \mathbf{P}^1 \\ \mathbf{P}^2 \end{bmatrix} \right) \\ \text{s.t.} \quad \begin{bmatrix} \mathbf{P}^1 \\ \mathbf{P}^2 \end{bmatrix}^\top \begin{bmatrix} \mathbf{X}^1 \\ \mathbf{X}^2 \end{bmatrix} \begin{bmatrix} \mathbf{X}^1 \\ \mathbf{X}^2 \end{bmatrix}^\top \begin{bmatrix} \mathbf{P}^1 \\ \mathbf{P}^2 \end{bmatrix} = \mathbf{I}. \quad (4)$$

We can see that CCA can be expressed as the trace function with cross-view data concatenation. Although (4) and (10) share an analogical expression form, they differ in the meaning of related notations and the imposed constraints. CCA uses the global sample sets, and does not take the local geometry into consideration. CPA, by contrast, sorts and rearranges the sets based on local patches measured by intra-class and inter-class neighbors.

Different from CCA, the alignment matrix of CPA contains  $\mathbf{W}_i = \text{diag}(\omega_i)$  which imposes different weights to different intra-class and inter-class neighbors. The nearest neighbors in the same class rely on ‘1’ to strengthen the intra-class compactness, while the nearest neighbors from different classes are punished by adjusting the parameter  $\zeta$ . Besides, we carry out the comparison experiments between the proposed CPA and Complete Canonical Correlation Analysis (C3A, an algorithm of reducing computational complexity for CCA) [41]. The results in Table IV, V and VI demonstrate the performance improvement provided by CPA.

TABLE II  
LIST OF IMPORTANT MATHEMATICAL NOTATIONS

Symbols	Explanation
$\mathbf{x}_i^v$	The $i$ -th ( $i = 1, 2, \dots, N$ ) gait sample in $\mathbf{X}^v$ for View $v$ ( $v = 1, 2, \dots, V$ ) with $V$ being the total number of views, where $\mathbf{X}^v = [\mathbf{x}_1^v, \mathbf{x}_2^v, \dots, \mathbf{x}_N^v]$ and $N$ is the total number of gait samples for each view
$\mathbf{P}^v$	The projection matrix for View $v$ ( $v = 1, 2, \dots, V$ )
$\mathbf{y}_i^v$	The projected features of $\mathbf{x}_i^v$ ( $i = 1, 2, \dots, N; v = 1, 2, \dots, V$ ) by $\mathbf{P}^v$
$\tilde{\mathbf{X}}_i^v$	The gait sample patch $\tilde{\mathbf{X}}_i^v$ constructed by $\mathbf{x}_i^v$ , i. e. $\tilde{\mathbf{X}}_i^v = \begin{bmatrix} \mathbf{x}_i^v, \dots, \mathbf{x}_i^v \end{bmatrix}_{k_1+k_2}$
$(\mathbf{x}_i^v)_j, (\tilde{\mathbf{x}}_i^v)_j$	The $j$ -th ( $j = 1, \dots, k_1$ ) intra-class and inter-class nearest neighbors of $\mathbf{x}_i^v$ , respectively
$(\mathbf{y}_i^v)_j, (\tilde{\mathbf{y}}_i^v)_j$	The $j$ -th ( $j = 1, \dots, k_1$ ) transformed intra-class and inter-class nearest neighbors of $\mathbf{x}_i^v$ projected by $\mathbf{P}^v$ , respectively
$\tilde{\mathbf{X}}_i^v$	The gait sample patch $\tilde{\mathbf{X}}_i^v$ is constructed by $k_1$ intra-class nearest neighbors of $\mathbf{x}_i^v$ together with $k_2$ inter-class nearest neighbors of $\mathbf{x}_i^v$ , i. e. $\tilde{\mathbf{X}}_i^v = \begin{bmatrix} (\mathbf{x}_i^v)_1, (\mathbf{x}_i^v)_2, \dots, (\mathbf{x}_i^v)_{k_1}, (\tilde{\mathbf{x}}_i^v)_1, (\tilde{\mathbf{x}}_i^v)_2, \dots, (\tilde{\mathbf{x}}_i^v)_{k_2} \end{bmatrix}_{k_1+k_2}$

### III. COUPLED PATCH ALIGNMENT FOR MATCHING CROSS-VIEW GAITS

This section introduces gait feature representation and the proposed CPA for cross-view gait recognition.

#### A. Gait Feature Representation

Regular gait can be regarded as a cyclic motion at a stable frequency, therefore there are two shortcomings if the entire gait video is used for person recognition. First, this large amount of data contains redundant information, leading to the phenomenon of ‘data rich but information poor’. Second, the massive data will increase the computational burden. Therefore, gait cycle estimation plays an important role in the gait recognition task. In this paper, as shown in Fig. 2, we employ the dual-ellipse fitting approach [42] based on the local maximum eccentricity points to estimate the gait cycle of input cross-view gait sequences. This approach is robust to scale and view variations, therefore it is very suitable for solving the gait cycle estimation problem in cross-view gait recognition. Then, GEI [13] is used as the gait feature representation because it is computationally efficient, storage space-saving, and robust to the noise in silhouette images. After that, CPA developed in this paper is proposed to find the corresponding bases in separate views. The GEI in an arbitrary view can be written as a linear combination of its corresponding view bases. The coefficients in linear combination under each view are taken as features for the subsequent gait recognition.

#### B. Coupled Patch Alignment (CPA)

For ease of representation, the important notations that will be later used are listed in Table II.

As discussed in Section I, GEIs from different views for the same individual may not be similar in the original data space.

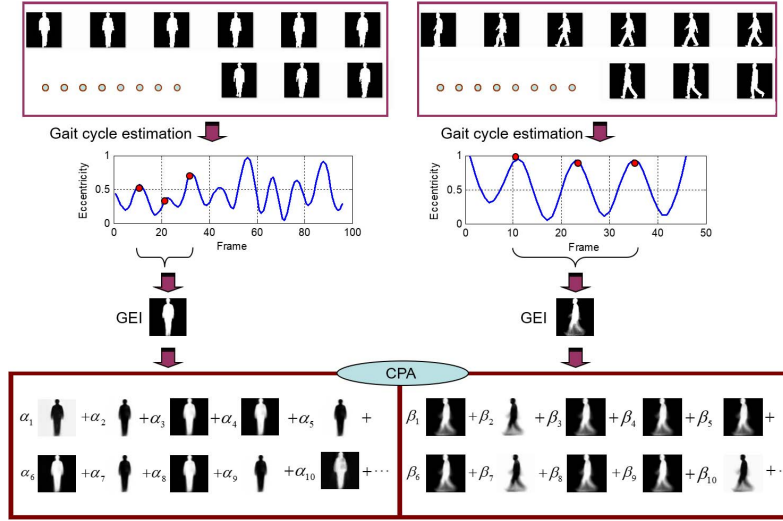


Fig. 2. The pipeline of the proposed cross-view gait recognition approach. We first estimate the gait cycle from the input images, and then represent the gait sequence via GEI. The GEI of a certain view can be written as a linear combination of bases of the corresponding views by the proposed CPA. Therefore,  $[\alpha_1, \alpha_2, \dots]$  for frontal GEI and  $[\beta_1, \beta_2, \dots]$  for lateral GEI are final feature representations used for the subsequent cross-view gait recognition.

Therefore, CPA is proposed in this paper to fill the information gaps across different views.

Without loss of generality, we assume that there are  $2N$  cross-view gait samples (GEIs) in the training set, and then the samples of each view can be denoted as  $\mathbf{X}^1 = [\mathbf{x}_1^1, \mathbf{x}_2^1, \dots, \mathbf{x}_N^1]$  and  $\mathbf{X}^2 = [\mathbf{x}_1^2, \mathbf{x}_2^2, \dots, \mathbf{x}_N^2]$ , respectively, where the associated superscripts indicate different views, and  $N$  is the number of samples for each view. Given the sample pairs  $\{(\mathbf{x}_i^1, \mathbf{x}_i^2) | \mathbf{x}_i^1 \in \mathbb{R}^h, \mathbf{x}_i^2 \in \mathbb{R}^l, i = 1, 2, \dots, N\}$  ( $h$  and  $l$  are the feature dimensionalities in two different views, correspondingly) denoting the cross-view gait samples for the same subject, our goal is to learn two projection matrices  $\{(\mathbf{P}^1, \mathbf{P}^2) | \mathbf{P}^1 \in \mathbb{R}^{h \times d}, \mathbf{P}^2 \in \mathbb{R}^{l \times d}\}$ , by which the transformed cross-view gait samples are projected onto a common subspace, namely

$$\mathbf{y}_i^1 = (\mathbf{P}^1)^\top \mathbf{x}_i^1, \quad \mathbf{y}_i^2 = (\mathbf{P}^2)^\top \mathbf{x}_i^2. \quad (5)$$

Assuming that the nearest neighbors of a gait sample under different views are identical, the goal is to transform a pair of cross-view gait samples to the common subspace such that the distance between the projected nearest neighbors in the same class is minimized while the distance between the projected nearest neighbors in different classes is maximized. Thus we have

$$\arg \min_{\mathbf{y}_i^1} \sum_{j=1}^{k_1} \|\mathbf{y}_i^1 - (\mathbf{y}_i^2)_j\|^2, \quad (6)$$

$$\arg \max_{\mathbf{y}_i^1} \sum_{j=1}^{k_2} \|\mathbf{y}_i^1 - (\tilde{\mathbf{y}}_i^2)_j\|^2. \quad (7)$$

The designed optimization expression (6) characterizes the intra-class compactness across two different views. A sample  $\mathbf{y}_i^1$  is connected to  $k_1$  intra-class nearest neighbors of  $\mathbf{y}_i^2$  which is the counterpart of  $\mathbf{y}_i^1$  in View 2. Similarly, (7) characterizes the inter-class separability across two different views.

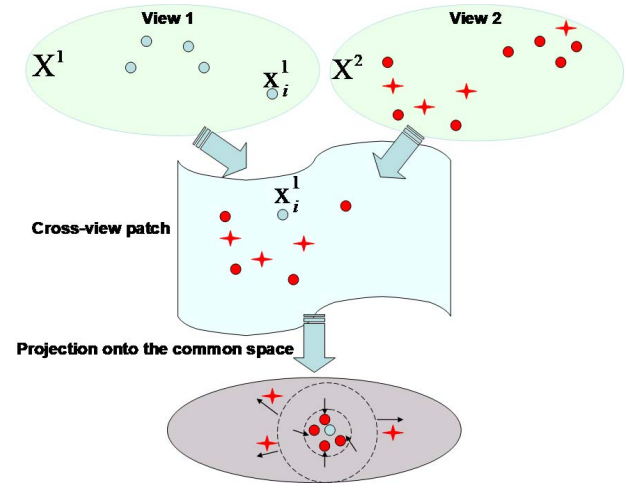


Fig. 3. Diagram of CPA.  $\mathbf{X}^1$  and  $\mathbf{X}^2$  represent two sets of cross-view gait samples. The red circles in View 2 represent the intra-class nearest neighbors of cross-view counterpart  $\mathbf{x}_i^1$ . The red four-point stars in View 2 represent the inter-class nearest neighbors of  $\mathbf{x}_i^1$ . The aim of the proposed CPA is to project the cross-view gait samples onto the common space, so that the projected gait features of  $\mathbf{x}_i^1$  (i.e.  $\mathbf{y}_i^1$ ) and its intra-class cross-view nearest neighbors are gathered together, while  $\mathbf{y}_i^1$  and the projected inter-class cross-view nearest neighbors are dispersed.

A sample  $\mathbf{y}_i^1$  is connected to  $k_2$  intra-class nearest neighbors of  $\mathbf{y}_i^2$  in View 2.

As shown in Fig. 3, a cross-view patch is constructed by the  $i$ -th sample  $\mathbf{x}_i^1$  from View 1 and its counterpart nearest neighbors from the other View 2. After CPA, both of the cross-view gait samples are projected onto the common space. For example, we expect that the transformed  $(\mathbf{y}_i^2)_1, (\mathbf{y}_i^2)_2, \dots, (\mathbf{y}_i^2)_{k_1}$  are as close to  $\mathbf{y}_i^1$  as possible, while  $(\tilde{\mathbf{y}}_i^2)_1, (\tilde{\mathbf{y}}_i^2)_2, \dots, (\tilde{\mathbf{y}}_i^2)_{k_2}$  are as far away from  $\mathbf{y}_i^1$  as possible in the common space.

By following the idea of the differential scatter discriminant criterion [1], we balance the intra-class compactness and inter-class separability by using a tuning parameter  $\zeta \in [0, 1]$  and

then the objective functions of (6) and (7) can be unified as

$$\arg \min_{\mathbf{y}_i^1} \left( \sum_{j=1}^{k_1} \left\| \mathbf{y}_i^1 - (\mathbf{y}_i^2)_j \right\|^2 - \zeta \sum_{j=1}^{k_2} \left\| \mathbf{y}_i^1 - (\tilde{\mathbf{y}}_i^2)_j \right\|^2 \right). \quad (8)$$

It can be seen that the objective function of CPA is the difference between the sum of squared distance of  $k_1$  intra-class nearest neighbors weighted by 1 and the sum of squared distance of  $k_2$  inter-class nearest neighbors weighted by  $\zeta$ . That is to say, the parameter  $\zeta$  plays an important role in balancing the influence of the intra-class and inter-class nearest neighbors. More discusses about  $\zeta$  are given in Section V-D.

By defining a weight multiplier vector as

$$\boldsymbol{\omega}_i = \left[ \underbrace{1, \dots, 1}_{k_1}, \underbrace{-\zeta, \dots, -\zeta}_{k_2} \right]^\top, \quad (9)$$

(8) can be simplified as

$$\begin{aligned} & \sum_{j=1}^{k_1+k_2} \left\| \mathbf{y}_{Idx(i)}^1 - \mathbf{y}_{Idx(j+1)}^2 \right\|^2 (\boldsymbol{\omega}_i)_{Idx(j)} \\ &= \text{tr} \left( \sum_{j=1}^{k_1+k_2} \left( (\mathbf{P}^1)^\top \mathbf{x}_{Idx(i)}^1 - (\mathbf{P}^2)^\top \mathbf{x}_{Idx(j+1)}^2 \right) \right. \\ & \quad \times \left. \left( (\mathbf{P}^1)^\top \mathbf{x}_{Idx(i)}^1 - (\mathbf{P}^2)^\top \mathbf{x}_{Idx(j+1)}^2 \right)^\top (\boldsymbol{\omega}_i)_{Idx(j)} \right) \\ &= \text{tr} \left( \begin{bmatrix} \mathbf{P}^1 \\ \mathbf{P}^2 \end{bmatrix}^\top \begin{bmatrix} \tilde{\mathbf{X}}_i^1 \\ \tilde{\mathbf{X}}_i^2 \end{bmatrix} \begin{bmatrix} \mathbf{W}_i & -\mathbf{W}_i \\ -\mathbf{W}_i & \mathbf{W}_i \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{X}}_i^1 \\ \tilde{\mathbf{X}}_i^2 \end{bmatrix} \begin{bmatrix} \mathbf{P}^1 \\ \mathbf{P}^2 \end{bmatrix} \right) \quad (10) \end{aligned}$$

where  $Idx(i) = \left\{ i, (i)_1, \dots, (i)_{k_1}, (\tilde{i})_1, \dots, (\tilde{i})_{k_2} \right\}$ ,  $Idx(i)$

and  $Idx(j+1)$  denotes the first and  $(j+1)$ -th item of  $Idx(i)$ .  $(\boldsymbol{\omega}_i)_{Idx(j)}$  denotes the  $j$ -th item of  $\boldsymbol{\omega}_i$ . Besides,  $\mathbf{W}_i = \text{diag}(\boldsymbol{\omega}_i)$ ,  $\tilde{\mathbf{X}}_i^1 = \left[ \mathbf{x}_{Idx(i)}^1, \dots, \mathbf{x}_{Idx(i)}^1 \right] = \left[ \mathbf{x}_i^1, \dots, \mathbf{x}_i^1 \right]$ ,

and  $\tilde{\mathbf{X}}_i^2 = \left[ \mathbf{x}_{Idx(2)}^2, \mathbf{x}_{Idx(3)}^2, \dots, \mathbf{x}_{Idx(k_1+k_2+1)}^2 \right] = \left[ \mathbf{x}_i^2, \mathbf{x}_i^2, \dots, \mathbf{x}_i^2 \right]$ . Here the

matrix  $\mathbf{P} = \begin{bmatrix} \mathbf{P}^1 \\ \mathbf{P}^2 \end{bmatrix}$  is a concatenation of the projection matrices  $\mathbf{P}^1$  and  $\mathbf{P}^2$ . Let  $\mathbf{F}_i = \begin{bmatrix} \tilde{\mathbf{X}}_i^1 \\ \tilde{\mathbf{X}}_i^2 \end{bmatrix}$  and  $\boldsymbol{\Omega}_i = \begin{bmatrix} \mathbf{W}_i & -\mathbf{W}_i \\ -\mathbf{W}_i & \mathbf{W}_i \end{bmatrix}$ , (10) can be rewritten as

$$J_i(\mathbf{P}) = \text{tr} \left( \mathbf{P}^\top \mathbf{F}_i \boldsymbol{\Omega}_i \mathbf{F}_i^\top \mathbf{P} \right). \quad (11)$$

In order to achieve the whole alignment for all  $2N$  cross-view gait samples in the training set,  $N$  local optimizations

### Algorithm 1 CPA-Based Cross-View Gait Recognition

#### Input:

The training gait sample sets  $\{\mathbf{X}^1, \mathbf{X}^2\}$  for two views, where  $\{\mathbf{x}_i^1, \mathbf{x}_i^2\}_{i=1}^N$  is the  $i$ -th cross-view sample pair of  $\{\mathbf{X}^1, \mathbf{X}^2\}$ , the tuning parameter  $\zeta$ , the number of intra-class nearest neighbors  $k_1$ , the number of inter-class nearest neighbors  $k_2$ , and the given gallery set  $\{\mathbf{x}_i^2 \in \mathbb{R}^l | i = 1, \dots, N_G\}$ .

#### Output:

Projection matrices  $\{\mathbf{P}^1, \mathbf{P}^2\}$  for two views, and the class label of a query gait sample  $\mathbf{x}_q^1$  in View 1.

#### Training:

##### Step 1

Select  $k_1$  intra-class and  $k_2$  inter-class nearest neighbors for each gait sample pairs  $\{\mathbf{x}_i^1, \mathbf{x}_i^2\}_{i=1}^N$  respectively, and then construct their corresponding patches by  $\tilde{\mathbf{X}}_i^1 = \left[ \mathbf{x}_i^1, \dots, \mathbf{x}_i^1 \right]$ , and  $\tilde{\mathbf{X}}_i^2 = \left[ (\mathbf{x}_i^2)_1, (\mathbf{x}_i^2)_2, \dots, (\mathbf{x}_i^2)_{k_1}, (\tilde{\mathbf{x}}_i^2)_1, (\tilde{\mathbf{x}}_i^2)_2, \dots, (\tilde{\mathbf{x}}_i^2)_{k_2} \right]$ .

##### Step 2

Construct auxiliary matrices  $\mathbf{F}_i$  and  $\boldsymbol{\Omega}_i$  by using  $\mathbf{F}_i = \begin{bmatrix} \tilde{\mathbf{X}}_i^1 & \mathbf{O} \\ \mathbf{O} & \tilde{\mathbf{X}}_i^2 \end{bmatrix}$ ,  $\boldsymbol{\Omega}_i = \begin{bmatrix} \mathbf{W}_i & -\mathbf{W}_i \\ -\mathbf{W}_i & \mathbf{W}_i \end{bmatrix}$ ,  $\mathbf{W}_i = \text{diag}(\boldsymbol{\omega}_i)$ , and (9).

##### Step 3

Construct matrices  $\mathbf{F}$  and  $\boldsymbol{\Omega}$  via  $\mathbf{F} = [\mathbf{F}_1, \mathbf{F}_2, \dots, \mathbf{F}_N]$  and  $\boldsymbol{\Omega} = \text{diag}(\boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2, \dots, \boldsymbol{\Omega}_N)$  correspondingly.

##### Step 4

Solve the eigenvalue decomposition problem  $\mathbf{F} \boldsymbol{\Omega} \mathbf{F}^\top \mathbf{p} = \lambda \mathbf{p}$  to obtain a concatenation projection matrix  $\mathbf{P}^* = \begin{bmatrix} \mathbf{P}^{1*} \\ \mathbf{P}^{2*} \end{bmatrix} = [\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_d]$ , in which  $\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_d$  are  $d$  eigenvectors associated with the  $d$  smallest eigenvalues.

#### Testing:

$\mathbf{x}_q^1$  is assigned to the  $i^*$ -th class, where  $i^* = \arg \min_i \text{dis} \left( (\mathbf{P}^1)^\top \mathbf{x}_q^1, (\mathbf{P}^2)^\top \mathbf{x}_q^2 \right)$ .

are combined into a single objective function, which is

$$\begin{aligned} & \arg \min_{\mathbf{P}} \sum_{i=1}^N \text{tr} \left( \mathbf{P}^\top \mathbf{F}_i \boldsymbol{\Omega}_i \mathbf{F}_i^\top \mathbf{P} \right) \\ &= \arg \min_{\mathbf{P}} \text{tr} \left( \mathbf{P}^\top \left( \sum_{i=1}^N \mathbf{F}_i \boldsymbol{\Omega}_i \mathbf{F}_i^\top \right) \mathbf{P} \right) \\ &= \arg \min_{\mathbf{P}} \text{tr} \left( \mathbf{P}^\top \mathbf{F} \boldsymbol{\Omega} \mathbf{F}^\top \mathbf{P} \right), \quad (12) \end{aligned}$$

where  $\mathbf{F} = [\mathbf{F}_1, \mathbf{F}_2, \dots, \mathbf{F}_N]$  and  $\boldsymbol{\Omega} = \text{diag}(\boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2, \dots, \boldsymbol{\Omega}_N)$ . In order to uniquely determine  $\mathbf{P}$ , the constraint  $\mathbf{P}^\top \mathbf{P} = \mathbf{I}$  is imposed on (12), where  $\mathbf{I}$  is the identity matrix. Therefore, the objective function can be written as

$$\arg \min_{\mathbf{P}} \text{tr} \left( \mathbf{P}^\top \mathbf{F} \boldsymbol{\Omega} \mathbf{F}^\top \mathbf{P} \right) \quad \text{s.t.} \quad \mathbf{P}^\top \mathbf{P} = \mathbf{I}. \quad (13)$$

By using Lagrangian multiplier method, (13) can be converted to the following eigenvalue decomposition problem

$$\mathbf{F}\mathbf{\Omega}\mathbf{F}^\top \mathbf{p} = \lambda \mathbf{p}, \quad (14)$$

where  $\mathbf{F}\mathbf{\Omega}\mathbf{F}^\top$  is called coupled-alignment matrix, and  $\lambda$  is the eigenvalue of this matrix. Let  $\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_d$  be the  $d$  eigenvectors corresponding to the  $d$  smallest eigenvalues. The optimal solution  $\mathbf{P}^*$  to (14) is

$$\mathbf{P}^* = \begin{bmatrix} \mathbf{P}^{1*} \\ \mathbf{P}^{2*} \end{bmatrix} = [\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_d], \quad (15)$$

where  $\mathbf{P}^{1*}$  and  $\mathbf{P}^{2*}$  constitute the 1st~ $h$ -th rows and the  $(h+1)$ -th~ $(h+l)$ -th rows of  $\mathbf{P}^*$ , respectively.

### C. Cross-View Gaits Matching

In this paper, we use the proposed CPA to learn the projection matrices  $\{\mathbf{P}^1, \mathbf{P}^2\}$  for Views 1 and 2, and use the nearest neighbor classifier for matching the cross-view gaits. Given  $N_G$  samples  $\{\mathbf{x}_i^2 \in \mathbb{R}^l | i = 1, \dots, N_G\}$  for View 2 in the gallery set, the query gait sample  $\mathbf{x}_q^1$  for View 1 can then be assigned to the class  $\pi_{i^*}$ , where  $\pi_{i^*}$  is the class label of the  $i^*$ -th sample in the gallery set

$$\begin{aligned} i^* &= \arg \min_i \text{dis} \left( (\mathbf{P}^1)^\top \mathbf{x}_q^1, (\mathbf{P}^2)^\top \mathbf{x}_i^2 \right) \\ &= \arg \min_i \left\{ -\frac{[(\mathbf{P}^1)^\top \mathbf{x}_q^1]^\top [(\mathbf{P}^2)^\top \mathbf{x}_i^2]}{\|(\mathbf{P}^1)^\top \mathbf{x}_q^1\| \|(\mathbf{P}^2)^\top \mathbf{x}_i^2\|} \right\}. \end{aligned} \quad (16)$$

Based on the above analysis, the procedure for CPA-based cross-view gait recognition is presented in *Algorithm 1*.

## IV. EXTENSION TO MULTI-VIEW CASE

In many real-world scenarios, the gallery gait samples can be captured under multiple views, simultaneously. Without loss of generality, we extend CPA to a more general Multi-dimensional Patch Alignment (MPA) framework to better characterize the shared information in an arbitrary number of views. To formulate the MPA for  $V$  ( $V > 2$ ) views, we may follow the similar derivation as CPA with a few modifications.

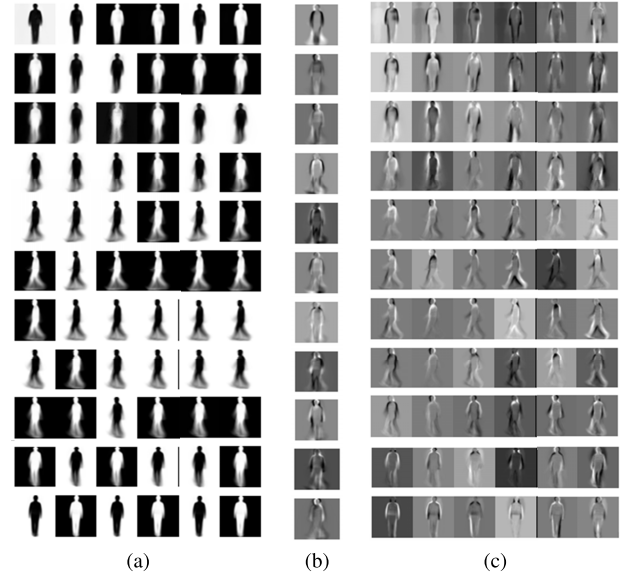


Fig. 4. The visualization of eigenvectors of our MPA and PCA. (a) MPAgaits trained by multiple view GEIs, (b) Eigengaits trained by multiple view GEIs, (c) Eigengaits trained by each view GEIs.

### A. Multi-Dimensional Patch Alignment (MPA)

Let  $\mathbf{X}^v = [\mathbf{x}_1^v, \mathbf{x}_2^v, \dots, \mathbf{x}_N^v]$  ( $v = 1, \dots, V$ ) be  $V$  gait sample sets collected from different views. Moreover,  $\{\mathbf{x}_i^1, \mathbf{x}_i^2, \dots, \mathbf{x}_i^V\}, i = 1, 2, \dots, N\}$  represents the corresponding multi-view gait samples for the same subject. Let  $\mathbf{P}^v$  ( $v = 1, \dots, V$ ) be  $V$  transformation matrices, which project multi-view gait samples into the unified subspace. Then the objective function for MPA can be formulated as<sup>1</sup>

$$\arg \min_{\mathbf{P}} \sum_{i=1}^N J_i(\mathbf{P}) = \arg \min_{\mathbf{P}} \text{tr}(\mathbf{P}^\top \mathbf{G} \mathbf{P}) \quad s.t. \quad \mathbf{P}^\top \mathbf{P} = \mathbf{I}, \quad (17)$$

where  $\mathbf{P} = \begin{bmatrix} \mathbf{P}^1 \\ \vdots \\ \mathbf{P}^V \end{bmatrix}$ , and  $\mathbf{G} = \sum_{i=1}^N \mathbf{G}_i$  is called multi-dimensional alignment matrix.  $\mathbf{G}_i$  is formulated by (18), as shown at the bottom of this page.

<sup>1</sup>The detailed mathematical deductions are put into the Appendix.

$$\mathbf{G}_i = \begin{bmatrix} (V-1)\tilde{\mathbf{x}}_i^1 \mathbf{w}_i(\tilde{\mathbf{x}}_i^1)^\top & -\tilde{\mathbf{x}}_i^1 \mathbf{w}_i(\tilde{\mathbf{x}}_i^2)^\top & \dots & -\tilde{\mathbf{x}}_i^1 \mathbf{w}_i(\tilde{\mathbf{x}}_i^V)^\top \\ -\tilde{\mathbf{x}}_i^2 \mathbf{w}_i(\tilde{\mathbf{x}}_i^1)^\top & (V-2)\tilde{\mathbf{x}}_i^2 \mathbf{w}_i(\tilde{\mathbf{x}}_i^2)^\top + \tilde{\mathbf{x}}_i^2 \mathbf{w}_i(\tilde{\mathbf{x}}_i^2)^\top & \dots & -\tilde{\mathbf{x}}_i^2 \mathbf{w}_i(\tilde{\mathbf{x}}_i^V)^\top \\ -\tilde{\mathbf{x}}_i^3 \mathbf{w}_i(\tilde{\mathbf{x}}_i^1)^\top & -\tilde{\mathbf{x}}_i^3 \mathbf{w}_i(\tilde{\mathbf{x}}_i^2)^\top & (V-3)\tilde{\mathbf{x}}_i^3 \mathbf{w}_i(\tilde{\mathbf{x}}_i^3)^\top + 2\tilde{\mathbf{x}}_i^3 \mathbf{w}_i(\tilde{\mathbf{x}}_i^3)^\top & \dots & -\tilde{\mathbf{x}}_i^3 \mathbf{w}_i(\tilde{\mathbf{x}}_i^V)^\top \\ \vdots & \dots & \dots & \ddots & \vdots \\ -\tilde{\mathbf{x}}_i^V \mathbf{w}_i(\tilde{\mathbf{x}}_i^1)^\top & -\tilde{\mathbf{x}}_i^V \mathbf{w}_i(\tilde{\mathbf{x}}_i^2)^\top & \dots & \dots & (V-V)\tilde{\mathbf{x}}_i^V \mathbf{w}_i(\tilde{\mathbf{x}}_i^V)^\top + (V-1)\tilde{\mathbf{x}}_i^V \mathbf{w}_i(\tilde{\mathbf{x}}_i^V)^\top \end{bmatrix} \quad (18)$$



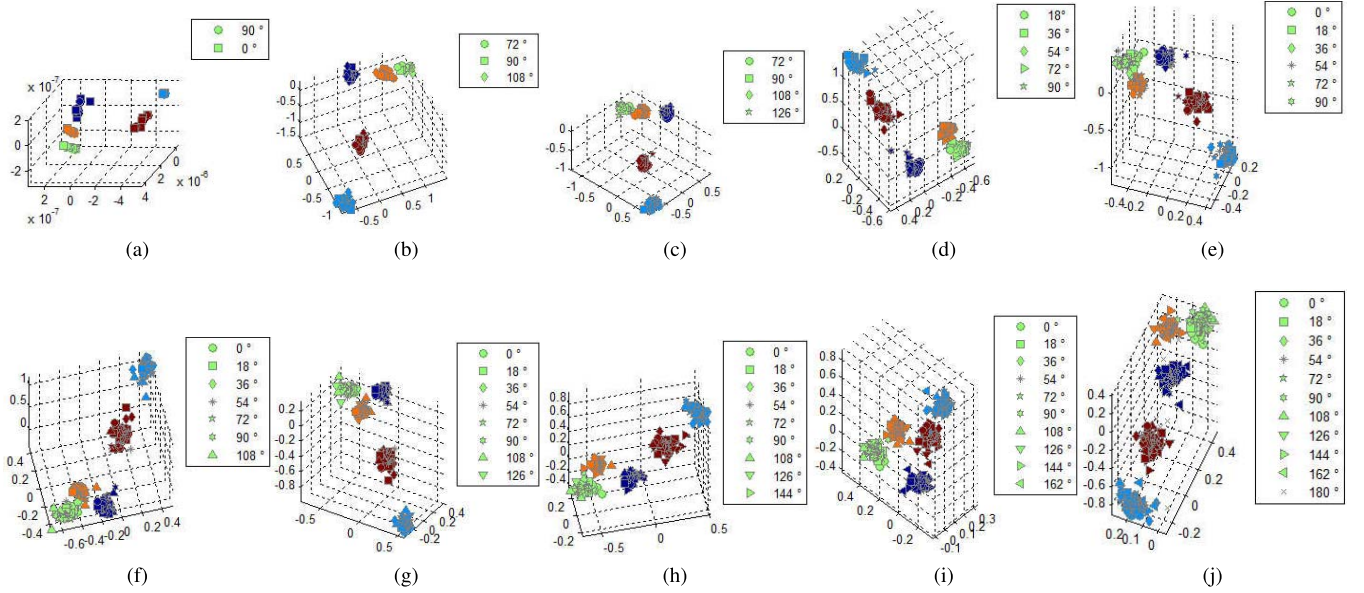


Fig. 5. Projection results of MPA under different numbers of views. Different colors denote different subjects, and different mark symbols represent different views. (a) 2 views. (b) 3 views. (c) 4 views. (d) 5 views. (e) 6 views. (f) 7 views. (g) 8 views. (h) 9 views. (i) 10 views. (j) 11 views.

The obtained eigenvectors of  $\mathbf{P}^v$  ( $v = 1, \dots, V$ ) are then called multi-dimensional-patch-alignment-gaits (MPAgaits). In order to present the visualization of MPAgait, the examples of the first 6 MPAgait for each of the views  $\{0^\circ, 18^\circ, \dots, 180^\circ\}$  are shown in Fig. 4a by simultaneously training the GEIs of 11 views with MPA. We see that the gait silhouette images generated by MPAgait are quite clear. Moreover, MPAgait have a good performance on view separation. The Eigengaits trained by multiple view GEIs together and by each view [43] are respectively shown in Fig. 4b and 4c for comparison. We find that the Eigengait generated by PCA are confounded by multiple views. On the contrary, Eigengaits trained by single view GEIs are distinct. They have the same view orientation properties as MPAgait, which further explain that MPA can distinguish among different views.

To illustrate the effectiveness of MPA, we use five subjects in the CASIA(B) gait database [44], and present the projection results of MPA under different numbers of views (see Fig. 5). We see that the MPA performs satisfactorily in distinguishing different individuals with various views. This is consistent with our initial anticipation, namely, MPA transforms the multi-view gait feature data points such that the distance between the projected nearest neighbors in the same class is minimized while the distance between the projected nearest neighbors in different classes can be maximized.

### B. Cross-View Gaits Matching

The procedure for cross-view gait recognition using MPA is very similar to the CPA illustrated in *Algorithm 1* except for the alignment matrix  $\mathbf{G}$ . By using the common subspace of various views established by MPA, the identity of the query gait sample can be easily confirmed. For a given gallery set  $\{\mathbf{x}_i^v | i = 1, \dots, N_G, v = 2, \dots, V\}$  for  $V - 1$  views, the query gait sample  $\mathbf{x}_q^1$  for View 1 can



Fig. 6. GEIs from 11 distinct views from CASIA(B) gait database.

be assigned to the class  $\pi_{i^*}$ , where  $i^* = \arg \min_i \left( (\mathbf{P}^1)^\top \mathbf{x}_q^1, \frac{1}{V-1} \left( (\mathbf{P}^2)^\top \mathbf{x}_i^2 + (\mathbf{P}^3)^\top \mathbf{x}_i^3 + \dots + (\mathbf{P}^V)^\top \mathbf{x}_i^V \right) \right)$ .

## V. EXPERIMENTS

This section firstly studies some critical issues of the proposed algorithm (Sections V-A ~ V-D), and then evaluates our method on three popular gait recognition databases including CASIA(B) [44], USF [27], and OU-ISIR [45] (Sections V-E ~ V-G). Finally, we analyze the results of compared methods in Section V-H.

### A. Impact of $k_1$ and $k_2$

We evaluated the impact of the numbers of intra-class nearest neighbors  $k_1$  and inter-class nearest neighbors  $k_2$  on the performance of CPA on the CASIA(B) database [44]. CASIA(B) gait database covers a wide range of observation angles  $\{0^\circ, 18^\circ, \dots, 180^\circ\}$ . This database contains totally 124 subjects under 11 views. There are 6 gait sequences for each subject under each view. In our experiments, the entire database was randomly divided into two non-overlapped groups: the first group, which contains 3 sequences covering all views of subjects, was taken as the probe set, and the remaining sequences form the second group that was treated as the gallery set. In gallery set, 60 subjects were randomly selected for training the CPA. We repeated these trials 10 times and calculated the average recognition results. The sizes of GEI in our experiments are  $64 \times 64$  pixels, and Fig. 6 shows the GEIs of 11 different views.



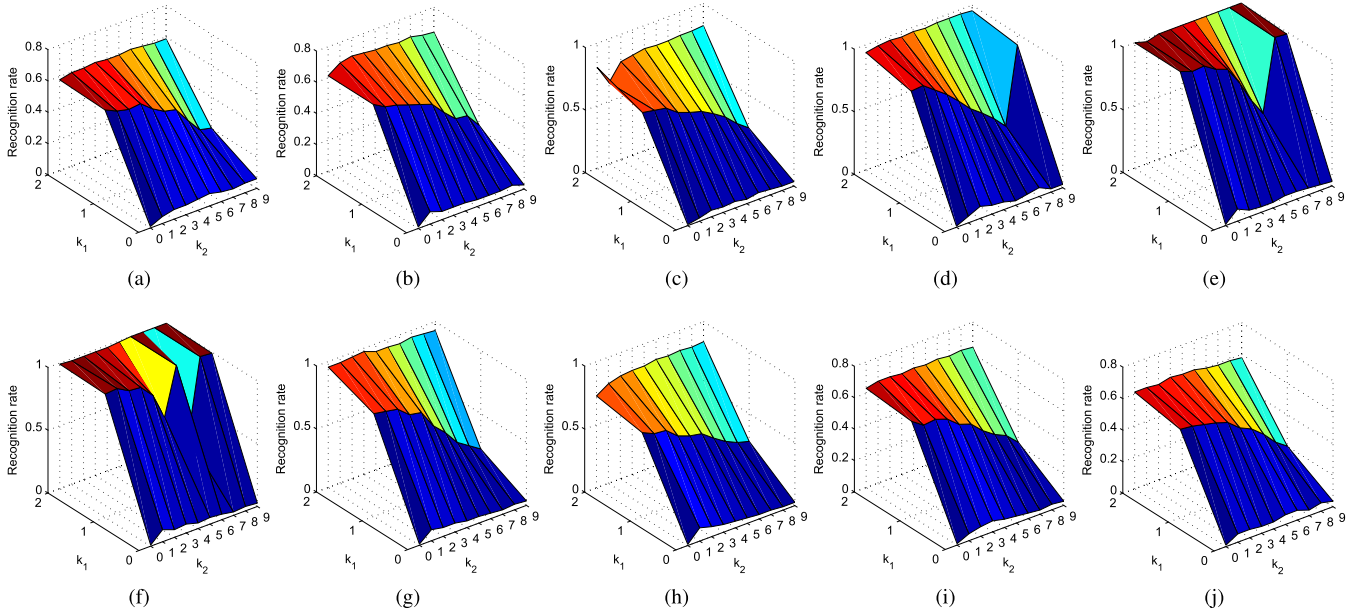


Fig. 7. Impact of  $k_1$  and  $k_2$  to the recognition rate. (a) Probe view:  $0^\circ$ ; (b) Probe view:  $18^\circ$ ; (c) Probe view:  $36^\circ$ ; (d) Probe view:  $54^\circ$ ; (e) Probe view:  $72^\circ$ ; (f) Probe view:  $108^\circ$ ; (g) Probe view:  $126^\circ$ ; (h) Probe view:  $144^\circ$ ; (i) Probe view:  $162^\circ$ ; (j) Probe view:  $180^\circ$ .

TABLE III

THE OPTIMAL SELECTIONS OF  $k_1$  AND  $k_2$  FOR EACH PROBE VIEW

Probe view ( $^\circ$ )	0	18	36	54	72	108	126	144	162	180
$k_1$	2	2	2	2	2	2	2	2	2	2
$k_2$	3	2	9	4	2	6	3	5	4	3

Note that  $k_1$  and  $k_2$  determine the size of local alignment patches, which are important to improve the cross-view gait recognition performance. Suppose that the view  $90^\circ$  was gallery and other views  $\{0^\circ, 18^\circ, \dots, 72^\circ, 108^\circ, \dots, 180^\circ\}$  were probe, then we reported the recognition rates w.r.t. the change of  $k_1$  and  $k_2$  for the proposed CPA. Fig. 7a~j present the results. In these figures, we can see that the accuracy improves with an increasing of  $k_1$ , and reaches the highest record when the number of  $k_1$  equals to 2. Differently, for the number of inter-class nearest neighbors  $k_2$ , the promising results can be obtained when  $k_2$  is selected from  $\{2, 3, 4, 5\}$ , and the accuracy may degenerate when the number of inter-class nearest neighbors increases. The optimal  $k_1$  and  $k_2$  selections are listed in Table III. These results suggest that CPA performs well when the local alignment patches are constructed by a certain number of intra-class samples and relatively few inter-class nearest neighbors.

### B. Impact of View Variance Between the Probe and Gallery

We also evaluated the impact of view variance between the probe and gallery on our CPA on the CASIA(B) database. We conducted the experiments under all pairs of views, and Fig. 8a~k show the recognition rates. From these figures, we can see that the accuracy of the proposed CPA deteriorates with the increasing of view difference between the probe and gallery. CPA can obtain nearly 100% recognition rate when the view difference between the probe and gallery is less than  $18^\circ$ .

This is reasonable since the gait silhouettes in these view point are very similar. The situation becomes quite challenging when the view difference between the probe and gallery is  $90^\circ$ . Therefore, CPA obtains relatively low recognition accuracy under this case.

### C. Impact of the Number of Gallery Views

This section is to evaluate the impact of the number of gallery views for multi-dimensional alignment patches on the cross-view gait recognition performance. These experiments were also carried out on the CASIA(B) gait database. By leveraging lateral view, which is the most different from other views, the regular pattern of the multi-view coupling alignment can be explored. Therefore, the lateral view ( $90^\circ$ ) was selected as the probe view, and the views from  $\{0^\circ, 18^\circ, \dots, 72^\circ, 108^\circ, \dots, 180^\circ\}$  were the gallery views. It is worth noting that these gallery views are deliberately chosen to be strictly adjacent, i.e., the interval between any two views is  $18^\circ$ . Although the recognition rate is as high as 100% or close to 100%, the gallery views with more coupled viewpoints do not bring about higher recognition rate. The high rate is only achieved when the gallery view is equal or close to  $90^\circ$ . This is because the closer the views are, the more shared information there will be. As a result, the gait features projected into the common subspace can achieve good performance. If the gait samples that are far away from  $90^\circ$  are introduced, such as  $0^\circ$  and  $18^\circ$ , the performance of MPA becomes worse when compared to the views that are close to  $90^\circ$  (e.g.  $72^\circ$  and  $108^\circ$ ).

To further explain the impact of number of gallery views for multi-dimensional alignment patches, we assume that the gallery views are adjacent. In other words, when the number of gallery views is equal to 2, only the following 9 combinations are considered:  $\{0^\circ, 18^\circ\}$ ,  $\{18^\circ, 36^\circ\}$ ,

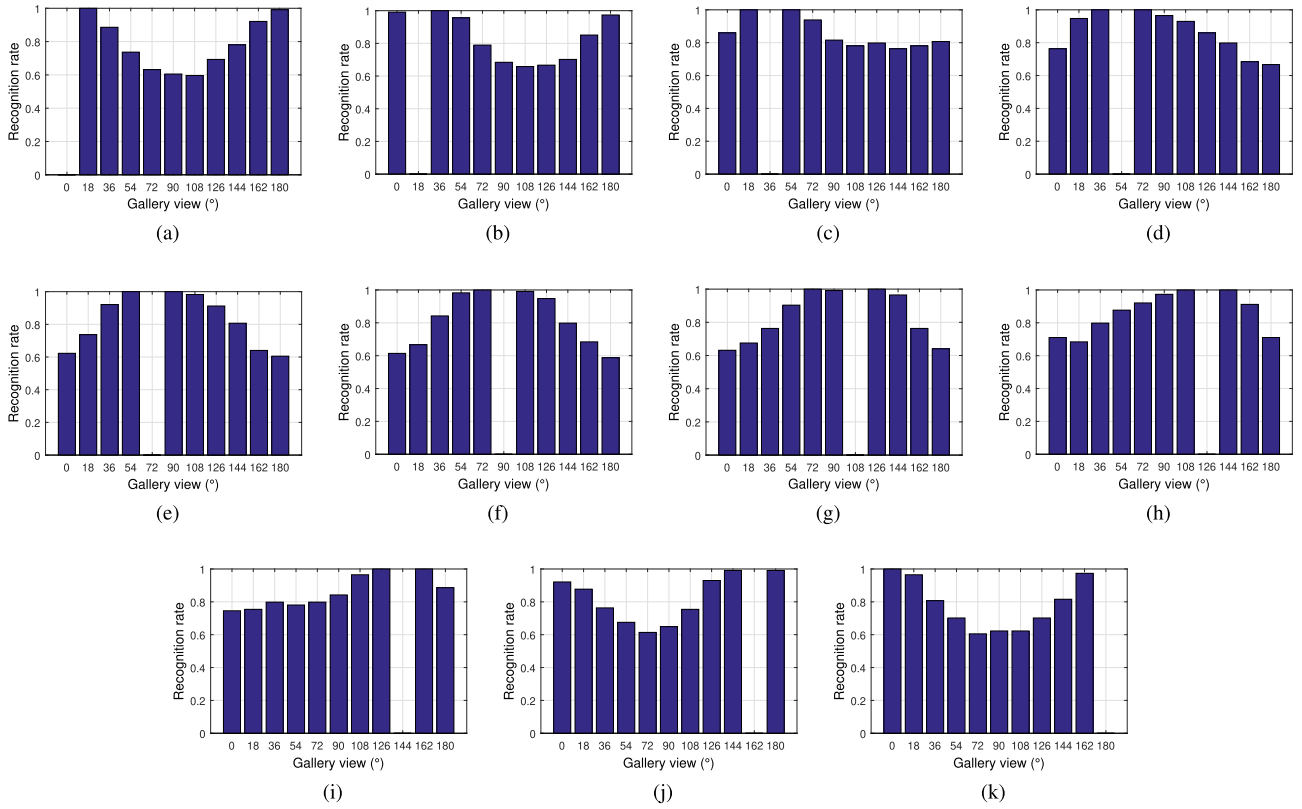


Fig. 8. Recognition rates under different Probe views. (a) Probe view:  $0^\circ$ ; (b) Probe view:  $18^\circ$ ; (c) Probe view:  $36^\circ$ ; (d) Probe view:  $54^\circ$ ; (e) Probe view:  $72^\circ$ ; (f) Probe view:  $90^\circ$ ; (g) Probe view:  $108^\circ$ ; (h) Probe view:  $126^\circ$ ; (i) Probe view:  $144^\circ$ ; (j) Probe view:  $162^\circ$ ; (k) Probe view:  $180^\circ$ .

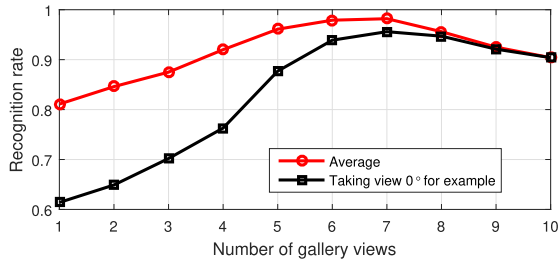


Fig. 9. Impact of the number of gallery views to the recognition rate.

$\{36^\circ, 54^\circ\}$ ,  $\{54^\circ, 72^\circ\}$ ,  $\{72^\circ, 108^\circ\}$ ,  $\{108^\circ, 126^\circ\}$ ,  $\{126^\circ, 144^\circ\}$ ,  $\{144^\circ, 162^\circ\}$  and  $\{162^\circ, 180^\circ\}$ . Fig. 9 illustrates the average recognition rates of MPA when different numbers of gallery views are used. An extreme case is the probe view coupled with 10 available gallery views for MPA, which yields a better precision (an average of 9.3% increase) than the case that the probe view coupled with only one gallery view for CPA. Due to the fact that the most different representation of GEI under  $0^\circ$  is that from  $90^\circ$ , we also visualize the recognition rates of MPA under different numbers of gallery views including  $0^\circ$  (see Fig. 9). We can see that both curves ascend gradually when the number of gallery views is not large. They reach the top value under 7 views and then drop down. This result is reasonable since the frontal viewpoints carry very limited gait information. By contrast, the accuracy can be improved when other oblique viewpoints close to  $90^\circ$  are added to

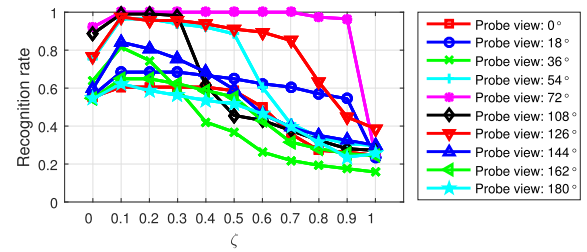


Fig. 10. Impact of the tuning parameter on CPA.

train MPA. Nonetheless, the recognition rate decreases when the gait samples that are deviated from  $90^\circ$  are incorporated.

#### D. Impact of Tuning Parameter $\zeta$

In this section, we evaluate the impact of tuning parameter  $\zeta$  to the performances of CPA and MPA. These experiments are carried out on the CASIA(B) gait database. By tuning  $k_1$  and  $k_2$  as mentioned in Table III, we run CPA when  $\zeta$  changes in the range  $[0, 1]$ . The best recognition rates are generally yielded when  $\zeta = 0.1$ . The performance decreases when  $\zeta$  gradually increases.

Since the penalized item contains  $k_2$  and  $\zeta$  in the objective function of CPA, we further investigate the correlation between  $k_2$  and  $\zeta$ . The performance of the CPA using different numbers of  $k_2$  and  $\zeta$  under the probe view of  $54^\circ$  is shown in Fig. 11. We observe that better results can be obtained when  $\zeta = 0.1$ . By fixing  $k_2$ , the selection of  $\zeta$  has a strong influence on

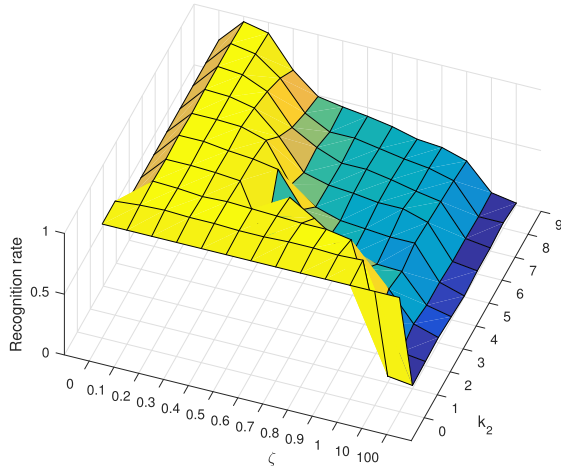


Fig. 11. The recognition rate with respect to  $k_2$  and  $\zeta$ .

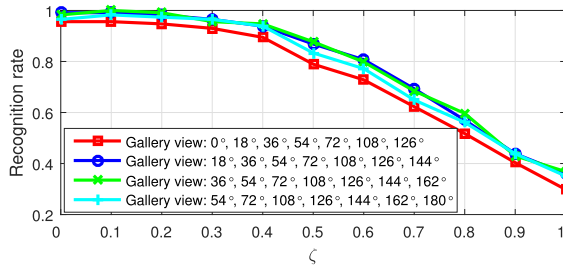


Fig. 12. Impact of the tuning parameter on MPA.

the recognition rate. Instead, by fixing  $\zeta$ , the selection of  $k_2$  has a minor effect on the recognition rate. When changing  $k_2$  and  $\zeta$  at the same time, the best performance can be obtained by cross validation. As reported in Section V-A, the best recognition rate is 96.5% when  $k_2 = 4$  and  $\zeta = 0.1$ . However, altering  $k_2$  can also obtain the best recognition rate of 96.5% when  $\zeta$  does not equal to 0.1, such as  $k_2 = 4$  and  $\zeta = 0.2$ , or  $k_2 = 1$  and  $\zeta = 0.2$ . Therefore, in general the results suggest that the tuning of  $k_2$  and  $\zeta$  correlated.

As discussed in Section V-C, the performance of MPA with 7 gallery views is better than the one under other views. To investigate the effect of different tuning parameter values on the performance of MPA, we report the results for 7 gallery views in Fig. 12. The optimal setting  $k_1$  and  $k_2$  is selected by cross validation. The results shown in Fig. 12 indicate that a large  $\zeta$  leads to a drop of the performance, and the best performance is obtained when  $\zeta = 0.1$ . The finding is similar to CPA.

#### E. Performance Evaluation on CASIA(B) Gait Database

In this section, we compare the performance of the proposed CPA with GEI [44], MFA [26], C3A [41], CMCC [6], VTM+QM [19], SVD [20], SVR [21], MvDA [36], and deep CNNs [24] on CASIA(B) cross-view gait recognition database. For a fair comparison, all the methods are evaluated under the data splits as mentioned in Section V-A. The parameters are set to the recommended values that are provided by the original published work such as GEI [44], C3A [41], CMCC [6],

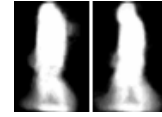


Fig. 13. GEIs of 2 views from USF gait database.

VTM+QM [19], SVD [20], SVR [21], MvDA [36], and deep CNNs [24]. The number of GEI segments is set to 4 for CMCC [6]. For SVR [21], the RBF kernel is adopted, and other parameters are set as  $C = 1$ ,  $\omega = 0.01$ ,  $b = 1$ , and  $g = 0.001$ . For deep CNNs [24], the size of every mini-batch is 128. The learning rate is determined as 0.01.

By varying the observation views from  $0^\circ$  to  $180^\circ$  with an interval of  $18^\circ$  (totally 11 views), all the experimental results are shown in Table IV. Different from most existing papers [6], [20], [21], [24], [36] that only report the results on several selected viewpoints, here we present the accuracies of all methods under all the investigated views. From Table IV, we observe that the proposed method significantly outperforms other methods across a wide range of view changes.

#### F. Performance Evaluation on USF Gait Database

The USF gait database [27] is the most challenging among the existing gait databases, because the gait samples of 122 subjects are collected in a real-world outdoor scene. Based on the official setting, this database includes one gallery set and twelve probe sets (Probe A  $\sim$  Probe L). In our experiments, the gallery is identical to the official Gallery setting. The proposed CPA requires to establish the relationship between two different views, therefore, the first 3 sequences of each subject from both the original Gallery set (left view) and Probe A set (right view) were selected to train the CPA. To verify the effectiveness of CPA for matching cross-view gaits, the remaining samples in Probe A set were used as query samples. The size of GEI in our experiments is  $128 \times 88$ , and Fig. 13 shows the GEIs of 2 views.

The proposed CPA is also evaluated on the real-world outdoor scene database, namely the USF gait database. We set  $k_1 = 2$  and  $k_2 = 4$  via cross validation. We compare GEI [44], MFA [26], C3A [41], CMCC [6], VTM+QM [19], SVD [20], SVR [21], MvDA [36] and deep CNNs [24] with our CPA by treating the cameras L and R as probe set and gallery set, respectively. The experimental results in Table V confirm the effectiveness of CPA when compared with other popular cross-view gait recognition methods. Specifically, CPA obtains a very impressive recognition rate of 94%. In this database our method is slightly worse than deep CNNs; however, it is significantly better than other methods that are not based on deep learning. Specifically, CPA leads the third best method (i.e. CMCC) with a noticeable margin of 5% on recognition rate, which is a very impressive result.

#### G. Performance Evaluation on OU-ISIR Gait Database

In order to study the performance of our method on the large population cases, we conducted the experiments under



TABLE IV  
COMPARISON OF VARIOUS METHODS ON CASIA(B) GAIT DATABASE (THE BEST RECORD UNDER EACH VIEW IS MARKED IN BOLD)

Probe View (°)	0									
Gallery View (°)	18	36	54	72	90	108	126	144	162	180
GEI [44]	0.238	0.044	0.024	0.008	0.004	0.016	0.012	0.036	0.121	0.411
MFA [26]	0.924	0.345	0.029	0.032	0.018	0.023	0.061	0.070	0.316	0.626
C3A [41]	0.842	0.602	0.532	0.529	0.515	0.518	0.518	0.567	0.667	0.904
CMCC [6]	0.848	0.471	0.260	0.249	0.292	0.260	0.269	0.371	0.702	0.901
VTM+QM [19]	0.854	0.734	0.602	0.550	0.523	0.538	0.544	0.664	0.798	0.886
SVD [20]	0.889	0.705	0.605	0.541	0.559	0.544	0.541	0.614	0.795	0.947
SVR [21]	0.839	0.450	0.199	0.199	0.231	0.204	0.208	0.252	0.629	0.921
MvDA [36]	0.939	0.795	0.617	0.506	0.532	0.447	0.532	0.626	0.854	0.956
Deep CNNs [24]	0.927	0.813	0.708	0.585	0.558	0.553	0.678	0.757	0.784	0.895
CPA (ours)	<b>1.000</b>	<b>0.886</b>	<b>0.737</b>	<b>0.632</b>	<b>0.605</b>	<b>0.597</b>	<b>0.693</b>	<b>0.781</b>	<b>0.921</b>	<b>0.991</b>
Probe View (°)	18									
Gallery View (°)	0	36	54	72	90	108	126	144	162	180
GEI [44]	0.319	0.379	0.036	0.044	0.024	0.016	0.028	0.052	0.391	0.198
MFA [26]	0.728	0.933	0.143	0.029	0.026	0.012	0.099	0.058	0.436	0.289
C3A [41]	0.874	0.959	0.716	0.581	0.549	0.571	0.538	0.576	0.614	0.704
CMCC [6]	0.950	0.935	0.614	0.400	0.260	0.204	0.190	0.248	0.409	0.584
VTM+QM [19]	0.877	0.979	0.733	0.587	0.567	0.561	0.543	0.602	0.649	0.719
SVD [20]	0.880	0.988	0.748	0.619	0.602	0.587	0.599	0.637	0.716	0.807
SVR [21]	0.912	0.903	0.745	0.292	0.219	0.117	0.140	0.216	0.289	0.760
MvDA [36]	0.953	0.983	0.813	0.646	0.573	0.567	0.626	0.629	0.728	0.830
Deep CNNs [24]	0.984	0.984	0.937	0.711	0.535	0.529	0.632	0.656	0.725	0.893
CPA (ours)	<b>0.991</b>	<b>1.000</b>	<b>0.956</b>	<b>0.790</b>	<b>0.684</b>	<b>0.658</b>	<b>0.667</b>	<b>0.702</b>	<b>0.851</b>	<b>0.974</b>
Probe View (°)	36									
Gallery View (°)	0	18	54	72	90	108	126	144	162	180
GEI [44]	0.093	0.399	0.290	0.073	0.048	0.020	0.060	0.282	0.157	0.081
MFA [26]	0.208	0.816	0.667	0.096	0.053	0.053	0.254	0.149	0.202	0.088
C3A [41]	0.657	0.988	0.967	0.836	0.628	0.605	0.584	0.596	0.596	0.643
CMCC [6]	0.672	0.938	0.959	0.643	0.377	0.233	0.175	0.160	0.219	0.321
VTM+QM [19]	0.687	0.953	0.892	0.696	0.585	0.567	0.547	0.559	0.559	0.617
SVD [20]	0.722	0.985	0.985	0.819	0.687	0.649	0.602	0.649	0.678	0.670
SVR [21]	0.775	0.936	0.939	0.731	0.409	0.219	0.149	0.117	0.228	0.415
MvDA [36]	0.807	0.988	0.985	0.825	0.681	0.649	0.673	0.678	0.670	0.714
Deep CNNs [24]	0.798	0.953	0.992	0.789	0.680	0.672	0.760	0.744	0.690	0.708
CPA (ours)	<b>0.860</b>	<b>1.000</b>	<b>1.000</b>	<b>0.939</b>	<b>0.816</b>	<b>0.781</b>	<b>0.798</b>	<b>0.763</b>	<b>0.781</b>	<b>0.807</b>
Probe View (°)	54									
Gallery View (°)	0	18	36	72	90	108	126	144	162	180
GEI [44]	0.040	0.089	0.298	0.218	0.177	0.169	0.375	0.185	0.024	0.032
MFA [26]	0.061	0.096	0.594	0.594	0.231	0.307	0.541	0.193	0.064	0.015
C3A [41]	0.558	0.772	0.968	0.991	0.871	0.754	0.655	0.620	0.582	0.585
CMCC [6]	0.249	0.649	0.936	0.901	0.629	0.500	0.450	0.310	0.219	0.211
VTM+QM [19]	0.573	0.746	0.971	0.962	0.769	0.673	0.605	0.591	0.547	0.544
SVD [20]	0.599	0.784	0.980	0.994	0.892	0.790	0.722	0.673	0.591	0.614
SVR [21]	0.202	0.623	0.936	0.921	0.588	0.509	0.418	0.263	0.178	0.199
MvDA [36]	0.611	0.804	0.988	0.988	0.696	0.830	0.766	0.690	0.602	0.594
Deep CNNs [24]	0.591	0.696	0.985	0.985	0.930	0.895	0.790	0.728	0.629	0.620
CPA (ours)	<b>0.763</b>	<b>0.947</b>	<b>1.000</b>	<b>1.000</b>	<b>0.965</b>	<b>0.930</b>	<b>0.860</b>	<b>0.798</b>	<b>0.684</b>	<b>0.667</b>
Probe View (°)	72									
Gallery View (°)	0	18	36	54	90	108	126	144	162	180
GEI [44]	0.032	0.044	0.117	0.230	0.823	0.714	0.335	0.044	0.016	0.020
MFA [26]	0.020	0.023	0.155	0.538	0.988	0.787	0.561	0.175	0.035	0.006
C3A [41]	0.550	0.579	0.819	0.971	0.991	0.959	0.830	0.640	0.573	0.553
CMCC [6]	0.205	0.409	0.643	0.912	0.953	0.614	0.439	0.190	0.146	0.219
VTM+QM [19]	0.529	0.623	0.807	0.974	0.974	0.883	0.737	0.629	0.535	0.553
SVD [20]	0.582	0.635	0.807	0.985	1.000	0.930	0.792	0.652	0.591	0.550
SVR [21]	0.231	0.439	0.708	0.930	0.942	0.746	0.439	0.219	0.123	0.146
MvDA [36]	0.518	0.649	0.857	0.994	0.965	0.918	0.787	0.614	0.582	0.512
Deep CNNs [24]	0.583	0.649	0.909	0.921	0.939	0.933	0.901	0.722	0.602	0.586
CPA (ours)	<b>0.623</b>	<b>0.737</b>	<b>0.921</b>	<b>1.000</b>	<b>1.000</b>	<b>0.983</b>	<b>0.912</b>	<b>0.807</b>	<b>0.640</b>	<b>0.605</b>
Probe View (°)	90									
Gallery View (°)	0	18	36	54	72	108	126	144	162	180
GEI [44]	0.032	0.036	0.069	0.165	0.815	0.879	0.222	0.016	0.008	0.008
MFA [26]	0.018	0.026	0.067	0.211	0.965	0.991	0.401	0.155	0.041	0.026
C3A [41]	0.544	0.558	0.643	0.886	0.988	0.980	0.913	0.667	0.573	0.535
CMCC [6]	0.178	0.240	0.363	0.599	0.950	0.930	0.614	0.395	0.178	0.099
VTM+QM [19]	0.521	0.567	0.690	0.860	0.974	0.983	0.830	0.684	0.535	0.541
SVD [20]	0.559	0.605	0.675	0.889	0.994	0.983	0.816	0.708	0.579	0.564
SVR [21]	0.152	0.214	0.339	0.617	0.936	0.933	0.643	0.351	0.175	0.117
MvDA [36]	0.462	0.559	0.652	0.901	0.962	0.965	0.886	0.766	0.538	0.509
Deep CNNs [24]	0.608	0.647	0.675	0.906	0.995	<b>0.995</b>	0.930	0.660	0.681	0.569
CPA (ours)	<b>0.614</b>	<b>0.667</b>	<b>0.842</b>	<b>0.983</b>	<b>1.000</b>	0.991	<b>0.947</b>	<b>0.798</b>	<b>0.684</b>	<b>0.588</b>

TABLE IV

(Continued.) COMPARISON OF VARIOUS METHODS ON CASIA(B) GAIT DATABASE (THE BEST RECORD UNDER EACH VIEW IS MARKED IN BOLD)

Probe View (°)	108									
Gallery View (°)	0	18	36	54	72	90	126	144	162	180
GEI [44]	0.002	0.036	0.081	0.214	0.681	0.823	0.480	0.032	0.008	0.016
MFA [26]	0.012	0.035	0.079	0.307	0.532	0.933	0.874	0.354	0.044	0.018
C3A [41]	0.553	0.588	0.640	0.754	0.933	0.982	0.974	0.833	0.614	0.576
CMCC [6]	0.205	0.181	0.249	0.442	0.629	0.901	0.921	0.643	0.433	0.234
VTM+QM [19]	0.497	0.541	0.640	0.746	0.851	0.974	0.962	0.845	0.605	0.538
SVD [20]	0.576	0.617	0.646	0.804	0.906	0.983	0.971	0.833	0.617	0.582
SVR [21]	0.190	0.199	0.257	0.395	0.687	0.915	0.921	0.678	0.401	0.260
MvDA [36]	0.442	0.557	0.643	0.807	0.906	0.965	0.971	0.854	0.591	0.477
Deep CNNs [24]	0.609	0.656	0.690	0.819	0.845	0.984	0.942	0.880	0.722	0.636
CPA (ours)	<b>0.632</b>	<b>0.675</b>	<b>0.763</b>	<b>0.904</b>	<b>1.000</b>	<b>0.991</b>	<b>1.000</b>	<b>0.965</b>	<b>0.763</b>	<b>0.640</b>
Probe View (°)	126									
Gallery View (°)	0	18	36	54	72	90	108	144	162	180
GEI [44]	0.020	0.052	0.133	0.290	0.210	0.153	0.371	0.431	0.024	0.036
MFA [26]	0.009	0.085	0.126	0.360	0.140	0.149	0.865	0.962	0.067	0.050
C3A [41]	0.535	0.541	0.576	0.664	0.827	0.933	0.988	0.994	0.740	0.573
CMCC [6]	0.240	0.281	0.330	0.450	0.550	0.752	0.968	0.971	0.781	0.199
VTM+QM [19]	0.556	0.567	0.570	0.643	0.699	0.787	0.971	0.991	0.743	0.570
SVD [20]	0.550	0.602	0.643	0.731	0.801	0.877	0.991	<b>1.000</b>	0.766	0.579
SVR [21]	0.205	0.246	0.240	0.409	0.541	0.746	0.965	0.959	0.725	0.175
MvDA [36]	0.488	0.611	0.693	0.754	0.801	0.828	0.985	0.991	0.772	0.553
Deep CNNs [24]	0.688	0.641	0.693	0.754	0.863	0.920	0.990	0.997	0.830	0.430
CPA (ours)	<b>0.711</b>	<b>0.684</b>	<b>0.798</b>	<b>0.877</b>	<b>0.921</b>	<b>0.974</b>	<b>1.000</b>	<b>1.000</b>	<b>0.912</b>	<b>0.711</b>
Probe View (°)	144									
Gallery View (°)	0	18	36	54	72	90	108	126	162	180
GEI [44]	0.048	0.137	0.234	0.214	0.056	0.052	0.060	0.266	0.052	0.125
MFA [26]	0.038	0.088	0.202	0.219	0.041	0.035	0.213	0.915	0.211	0.155
C3A [41]	0.626	0.573	0.582	0.655	0.670	0.693	0.848	0.997	0.964	0.629
CMCC [6]	0.412	0.228	0.178	0.175	0.202	0.424	0.678	0.909	0.921	0.658
VTM+QM [19]	0.617	0.544	0.553	0.594	0.602	0.637	0.769	0.956	0.886	0.602
SVD [20]	0.608	0.623	0.652	0.681	0.711	0.725	0.851	0.985	0.947	0.687
SVR [21]	0.406	0.237	0.173	0.167	0.234	0.415	0.702	0.909	0.936	0.699
MvDA [36]	0.588	0.635	0.705	0.670	0.684	0.699	0.933	0.988	0.947	0.664
Deep CNNs [24]	0.671	0.738	0.734	0.764	0.792	0.805	0.893	0.991	0.961	0.798
CPA (ours)	<b>0.746</b>	<b>0.754</b>	<b>0.798</b>	<b>0.781</b>	<b>0.798</b>	<b>0.842</b>	<b>0.965</b>	<b>1.000</b>	<b>1.000</b>	<b>0.886</b>
Probe View (°)	162									
Gallery View (°)	0	18	36	54	72	90	108	126	144	180
GEI [44]	0.129	0.335	0.133	0.048	0.036	0.036	0.020	0.044	0.056	0.512
MFA [26]	0.196	0.281	0.225	0.058	0.009	0.009	0.023	0.149	0.184	0.751
C3A [41]	0.725	0.643	0.585	0.576	0.558	0.550	0.614	0.734	0.956	0.845
CMCC [6]	0.637	0.409	0.234	0.199	0.196	0.237	0.412	0.640	0.912	0.936
VTM+QM [19]	0.757	0.608	0.591	0.585	0.556	0.585	0.655	0.804	0.985	0.857
SVD [20]	0.769	0.684	0.626	0.608	0.605	0.602	0.661	0.775	0.962	0.845
SVR [21]	0.658	0.307	0.219	0.196	0.181	0.214	0.351	0.649	0.924	0.927
MvDA [36]	0.839	0.772	0.681	0.605	0.547	0.573	0.655	0.769	0.922	0.909
Deep CNNs [24]	0.906	0.854	0.749	0.567	0.582	0.563	0.670	0.875	0.859	0.895
CPA (ours)	<b>0.921</b>	<b>0.877</b>	<b>0.763</b>	<b>0.675</b>	<b>0.614</b>	<b>0.649</b>	<b>0.754</b>	<b>0.930</b>	<b>0.991</b>	<b>0.991</b>
Probe View (°)	180									
Gallery View (°)	0	18	36	54	72	90	108	126	144	162
GEI [44]	0.379	0.109	0.020	0.012	0.016	0.012	0.020	0.020	0.028	0.286
MFA [26]	0.535	0.243	0.135	0.018	0.009	0.020	0.009	0.041	0.067	0.725
C3A [41]	0.918	0.658	0.570	0.544	0.529	0.532	0.561	0.538	0.588	0.787
CMCC [6]	0.898	0.608	0.409	0.211	0.170	0.146	0.208	0.424	0.602	0.895
VTM+QM [19]	0.936	0.760	0.655	0.585	0.518	0.512	0.550	0.570	0.667	0.857
SVD [20]	0.950	0.851	0.673	0.582	0.541	0.547	0.561	0.567	0.655	0.883
SVR [21]	0.927	0.708	0.409	0.228	0.175	0.161	0.234	0.424	0.731	0.933
MvDA [36]	0.968	0.880	0.655	0.588	0.515	0.509	0.506	0.564	0.655	0.895
Deep CNNs [24]	0.953	0.942	0.789	0.602	0.602	0.576	0.602	0.664	0.801	0.959
CPA (ours)	<b>1.000</b>	<b>0.965</b>	<b>0.807</b>	<b>0.702</b>	<b>0.605</b>	<b>0.623</b>	<b>0.623</b>	<b>0.702</b>	<b>0.816</b>	<b>0.974</b>

all pairs of cross-view scenarios available in the OU-ISIR gait database.

The OU-ISIR gait database [45] is the largest database including 1912 subjects from 4 views {55°, 65°, 75°, 85°} up to now. This database was randomly split into two sets with equal size for five times. The gallery set and probe set are the same as [46]. The cross-view data from 956 subjects were used for training, and the data from one view of the remaining 956 subjects were used as the gallery samples, and the data from the another view were regarded as query samples. Ten cross validation experiments were performed, and we then

TABLE V  
EXPERIMENTAL RESULTS ON THE USF GAIT DATABASE (THE BEST AND SECOND BEST RESULTS ARE MARKED IN RED BOLD AND BLUE BOLD, RESPECTIVELY)

GEI [44]	MFA [26]	C3A [41]	CMCC [6]	VTM+QM [19]
0.83	0.84	0.88	0.85	0.89
SVD [20]	SVR [21]	MvDA [36]	Deep CNNs [24]	CPA (ours)
0.86	0.86	0.86	<b>0.97</b>	<b>0.94</b>

reported its average recognition results. The sizes of GEI in our experiments are 128 × 88 pixels. Fig. 14 shows the GEIs of 4 views.

TABLE VI  
EXPERIMENTAL RESULTS ON THE OU-ISIR GAIT DATABASE (THE BEST RECORD UNDER EACH VIEW ANGLE IS MARKED IN BOLD)

Probe View ( $^{\circ}$ )	55			65			75			85		
Gallery View ( $^{\circ}$ )	65	75	85	55	75	85	55	65	85	55	65	75
GEI [44]	0.284	0.058	0.277	0.277	0.670	0.195	0.507	0.640	0.969	0.262	0.207	0.969
MFA [26]	0.856	0.426	0.397	0.864	0.865	0.662	0.523	0.874	0.969	0.369	0.604	0.969
C3A [41]	0.999	<b>0.996</b>	<b>0.985</b>	<b>0.999</b>	0.999	0.996	0.993	<b>1</b>	<b>1</b>	0.981	<b>0.996</b>	<b>1</b>
CMCC [6]	0.968	0.785	0.646	0.974	0.963	0.826	0.800	0.975	0.969	0.749	0.785	0.965
VTM+QM [19]	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 [20]	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 [21]	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 [36]	0.975	0.922	0.858	0.974	0.984	0.949	0.925	0.984	0.985	0.877	0.957	0.988
Deep CNNs [24]	0.983	0.960	0.805	0.963	0.973	0.833	0.942	0.978	0.851	0.900	0.960	0.984
CPA (ours)	<b>1</b>	<b>0.996</b>	<b>0.985</b>	<b>0.999</b>	<b>1</b>	<b>0.999</b>	<b>0.996</b>	<b>1</b>	<b>1</b>	<b>0.984</b>	<b>0.996</b>	<b>1</b>

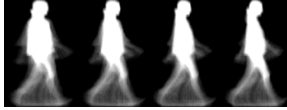


Fig. 14. GEIs of 4 views from OU-ISIR gait database.

We set  $k_1 = 1$  since there is only 1 intra-class sample for each training sample in this dataset.  $k_2$  is selected from  $\{2, 3, 4, 5\}$  through the cross validation. Basically, these four values would not hamper the performance. From Table VI, we see that the proposed CPA achieves very promising performance with approximately 99.6% accuracies for all cross-views. It outperforms other methods including GEI [44], MFA [26], C3A [41], CMCC [6], VTM+QM [19], SVD [20], SVR [21], MvDA [36], and deep CNNs [24]. Even more, the proposed CPA in cross-view cases yields slightly higher accuracy than deep CNNs in the identical-view cases. This is because the training samples for deep CNNs [24] are very limited, so the trained classifier is not sufficiently powerful to obtain satisfactory performance. As the view variation in the OU-ISIR database is relatively small, our method obtains very high recognition rates on this database.

#### H. Discussion

From above experimental results, we see that CMCC [6], VTM+QM [19], SVD [20] and SVR [21] cannot well handle the large cross-view gait recognition problem. CMCC [6] linearly transforms the gait information from one view to another. Unfortunately, the large view change may bring about non-linear diversifications, and in this case the target feature cannot be precisely constructed via linear transformation. Both VTM+QM [19] and SVD [20] need an SVD factorization on a high-dimensional matrix made up of the training samples from all views. As a result, the view-independent information and subject-independent gait information that are originally correlated are isolated. Unfortunately, both VTM+QM and SVD are sensitive to partial occlusion, and thus they are unsuitable for solving the cross-view gait recognition problems under large view changes. The accuracy of SVR [21] drops drastically when the view change gradually increases. This is because the gait information in the synthesized virtual view appears differently from the reference due to the large view variation.

TABLE VII  
COMPUTATIONAL COMPLEXITY ANALYSIS

Method	Process with high complexity	Computational complexity
MFA [26]	Intra-class and inter-class scatters, eigenvalue decomposition problem under marginal fisher criterion	$O(n^3) + O(d^3)$
C3A [41]	Eigenvalue decomposition	$O(n^3)$
CMCC [6]	Bipartite graph modeling process, bipartite graph multi-partitioning, correlation optimization by CCA and linear approximation	$O(d^2n) + O(d^3) + O((d')^3) + O(dn)$
VTM+QM [19]	Factorization process by SVD	$O(2dn^2)$
SVD [20]	Factorization process by SVD	$O(2dn^2)$
MvDA [36]	Matrix inversion and eigenvalue decomposition	$O((2d)^3)$
CPA	Eigenvalue decomposition	$O((2d)^3)$
MPA	Eigenvalue decomposition	$O((Vd)^3)$

C3A and MvDA perform better than the above CMCC, VTM+QM, SVD and SVR with significant improvements. This is because C3A [41] projects the gait features from two views into a unified subspace for achieving maximal correlation. MvDA [36] also finds a unified but discriminative subspace for bridging different views. Although C3A [41] and MvDA [36] have achieved very encouraging performance, the proposed CPA can still able to improve their recognition results. The reason lies in that our CPA models the nonlinear relationship between different views with a consideration of local nearest neighbor measurements. Consequently, the CPA is significantly better than C3A [41] and MvDA [36], and especially the advantage of CPA becomes prominent when the cross-view angle is large.

For simplicity, we assume that the feature dimensions for two different views are both  $d$ .  $n$  and  $c$  are the numbers of training samples and classes respectively. We analyze the computational complexity of the proposed CPA. It takes  $O((2d)^3)$  for eigenvalue decomposition of the matrix with the size of  $2d \times 2d$ . Further, general MPA takes  $O((Vd)^3)$  for synchronously solving  $V$  multi-view gait recognition. The number of view angles is limited in the real surveillance environment so the complexity for MPA is not going to increase a lot. We also compare the computational complexity of the proposed CPA and MPA with MFA [26], C3A [41], CMCC [6], VTM+QM [19], SVD [20] and MvDA [36]. Table VII lists their computational complexity. For CMCC [6],



$$\begin{aligned}
& \sum_{p=1}^{V-1} \sum_{q=p+1}^V \sum_{j=1}^{k_1+k_2} \left\| \mathbf{y}_{Idxi(1)}^p - \mathbf{y}_{Idxi(j+1)}^q \right\|^2 (\boldsymbol{\omega}_i)_{Idx(j)} \\
& = \text{tr} \left\{ \begin{bmatrix} \mathbf{P}^1 \\ \mathbf{P}^2 \\ \mathbf{P}^3 \\ \vdots \\ \mathbf{P}^V \end{bmatrix}^\top \begin{bmatrix} (V-1)\tilde{\mathbf{X}}_i^1 \mathbf{W}_i (\tilde{\mathbf{X}}_i^1)^\top & -\tilde{\mathbf{X}}_i^1 \mathbf{W}_i (\tilde{\mathbf{X}}_i^2)^\top & \cdots & -\tilde{\mathbf{X}}_i^1 \mathbf{W}_i (\tilde{\mathbf{X}}_i^V)^\top \\ -\tilde{\mathbf{X}}_i^2 \mathbf{W}_i (\tilde{\mathbf{X}}_i^1)^\top & (V-2)\tilde{\mathbf{X}}_i^2 \mathbf{W}_i (\tilde{\mathbf{X}}_i^2)^\top + \tilde{\mathbf{X}}_i^2 \mathbf{W}_i (\tilde{\mathbf{X}}_i^3)^\top & \cdots & -\tilde{\mathbf{X}}_i^2 \mathbf{W}_i (\tilde{\mathbf{X}}_i^V)^\top \\ -\tilde{\mathbf{X}}_i^3 \mathbf{W}_i (\tilde{\mathbf{X}}_i^1)^\top & -\tilde{\mathbf{X}}_i^3 \mathbf{W}_i (\tilde{\mathbf{X}}_i^2)^\top & (V-3)\tilde{\mathbf{X}}_i^3 \mathbf{W}_i (\tilde{\mathbf{X}}_i^3)^\top + 2\tilde{\mathbf{X}}_i^3 \mathbf{W}_i (\tilde{\mathbf{X}}_i^4)^\top & \cdots -\tilde{\mathbf{X}}_i^3 \mathbf{W}_i (\tilde{\mathbf{X}}_i^V)^\top \\ \vdots & \vdots & \vdots & \vdots \\ -\tilde{\mathbf{X}}_i^V \mathbf{W}_i (\tilde{\mathbf{X}}_i^1)^\top & -\tilde{\mathbf{X}}_i^V \mathbf{W}_i (\tilde{\mathbf{X}}_i^2)^\top & \cdots & (V-1)\tilde{\mathbf{X}}_i^V \mathbf{W}_i (\tilde{\mathbf{X}}_i^V)^\top \end{bmatrix} \begin{bmatrix} \mathbf{P}^1 \\ \mathbf{P}^2 \\ \mathbf{P}^3 \\ \vdots \\ \mathbf{P}^V \end{bmatrix} \right\}, \quad (21)
\end{aligned}$$

the computational complexity is related to the GEI segment dimension (denoted as  $d'$ ). The GEIs are first projected onto the PCA subspace to control the feature vector dimension. Thus, PCA is employed as a preprocessing step that makes  $d$  less than  $n$ . In general, there is no significant computational complexity difference among the above-mentioned methods.

## VI. CONCLUSION

In this work, we have presented a novel Coupled Patch Alignment (CPA) algorithm, which is further extended to Multi-dimensional Patch Alignment (MPA), to perform cross-view gait recognition. To the best of our knowledge, this is the first devoted work on multi-dimensional patch alignment across an arbitrary number of views for gait recognition. By building local patches that consists of a sample and its intra-class and inter-class nearest-neighbors, the local relationship among training gait samples is revealed. Then, all the local patches are combined to render a unified optimization problem that can be easily solved. Thorough experiments on the CASIA(B), USF, and OU-ISIR gait databases demonstrate that our method produces favorable results to the state-of-the-arts even the view difference between the gallery and the probe is as large as  $90^\circ$ .

As discussed in this paper, CPA and MPA can be used in the monitoring scenes where GEIs under two or more known views are available for training and meanwhile the target view is also known. With the increasing demand for intelligent human identification, much efforts have been made to solve the fundamental problems in uncooperative gait recognition that the view angle of gait images is unknown but the images from one viewpoint are obtained [47], [48]. Under this situation, the target view is unknown in the training stage, while only the source view can be obtained. In the future, we will extend MPA to the situations where some subjects do not have full viewpoint images. In this case, a latent factor will be introduced to uncover the underlying structure of the missing view from the known data.

## APPENDIX

In this appendix, we show how to build the objective function for MPA. The projection matrices  $\mathbf{P}^v$  ( $v = 1, 2, \dots, V$ ) map multi-view gait samples  $\{\mathbf{x}_i^v, i = 1, 2, \dots, N$ ,

$v = 1, 2, \dots, V\}$  to  $V$  new feature vectors in the unified subspace, which is

$$\mathbf{y}_i^v = (\mathbf{P}^v)^\top \mathbf{x}_i^v, \quad v = 1, 2, \dots, V. \quad (19)$$

The objective is to minimize the distance between the projected nearest intra-class neighbors from different views, and meanwhile maximize the distance between the projected nearest inter-class neighbors from different views.

As in the CPA, the objective function becomes

$$\begin{aligned}
& \sum_{p=1}^{V-1} \sum_{q=p+1}^V \arg \min_{\mathbf{y}_i^p} \sum_{j=1}^{k_1} \left\| \mathbf{y}_i^p - (\mathbf{y}_i^q)_j \right\|^2 \\
& - \zeta \sum_{p=1}^{V-1} \sum_{q=p+1}^V \arg \max_{\mathbf{y}_i^p} \sum_{j=1}^{k_2} \left\| \mathbf{y}_i^p - (\tilde{\mathbf{y}}_i^q)_j \right\|^2. \quad (20)
\end{aligned}$$

Eq. (20) can be further expressed as (21), as shown at the top of this page, where  $\tilde{\mathbf{X}}_i^v = \begin{bmatrix} \mathbf{x}_i^v, \dots, \mathbf{x}_i^v \end{bmatrix}$ ,  $\tilde{\mathbf{X}}_i^v =$

$$\begin{bmatrix} (\mathbf{x}_i^v)_1, (\mathbf{x}_i^v)_2, \dots, (\mathbf{x}_i^v)_{k_1}, (\tilde{\mathbf{x}}_i^v)_1, (\tilde{\mathbf{x}}_i^v)_2, \dots, (\tilde{\mathbf{x}}_i^v)_{k_2} \end{bmatrix}. \quad \text{for } v = 1, 2, \dots, V, \text{ and } V \text{ represents the number of views.}$$

Here, let  $\mathbf{P}$  be a concatenation matrix  $\mathbf{P} = [(\mathbf{P}^1)^\top \dots (\mathbf{P}^V)^\top]^\top$ , and define a local multi-dimensional alignment matrix  $\mathbf{G}_i$  (see (18)). Then (21) can be further simplified as  $J_i(\mathbf{P}) = \text{tr}(\mathbf{P}^\top \mathbf{G}_i \mathbf{P})$ .

## REFERENCES

- [1] D. Tao, X. Li, X. Wu, and S. J. Maybank, "General tensor discriminant analysis and Gabor features for gait recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 10, pp. 1700–1715, Oct. 2007.
- [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.*, vol. 90, pp. 87–98, Jun. 2019. doi: [10.1016/j.patcog.2019.01.017](https://doi.org/10.1016/j.patcog.2019.01.017).
- [3] W. Li, C.-C. J. Kuo, and J. Peng, "Gait recognition via GEI subspace projections and collaborative representation classification," *Neurocomputing*, vol. 275, pp. 1932–1945, Jan. 2018.
- [4] W. Li and J. Peng, "Gait recognition with a single sample per person," in *Proc. Signal Inf. Process. Assoc. Annu. Summit Conf. (APSIPA), Asia-Pacific*, Dec. 2016, pp. 1–6.
- [5] X. Ben, C. Gong, P. Zhang, R. Yan, Q. Wu, and W. Meng, "Coupled bilinear discriminant projection for cross-view gait recognition," *IEEE Trans. Circuits Syst. Video Technol.*, to be published. doi: [10.1109/TCSVT.2019.2893736](https://doi.org/10.1109/TCSVT.2019.2893736).

- [6] 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. 2013.
- [7] 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.
- [8] H. Hu, "Multiview gait recognition based on patch distribution features and uncorrelated multilinear sparse local discriminant canonical correlation analysis," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 4, pp. 617–630, Apr. 2014.
- [9] H. Hu, "Enhanced Gabor feature based classification using a regularized locally tensor discriminant model for multiview gait recognition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 7, pp. 1274–1286, Jul. 2013.
- [10] N. Liu, J. Lu, and Y. P. Tan, "Joint subspace learning for view-invariant gait recognition," *IEEE Signal Process. Lett.*, vol. 18, no. 7, pp. 431–434, Jul. 2011.
- [11] 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.
- [12] W. Kusakunniran, Q. Wu, J. Zhang, Y. Ma, and H. Li, "A new view-invariant feature for cross-view gait recognition," *IEEE Trans. Inf. Forensics Security*, vol. 8, no. 10, pp. 1642–1653, Oct. 2013.
- [13] J. Man and B. Bhanu, "Individual recognition using gait energy image," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 2, pp. 316–322, Feb. 2006.
- [14] K. Sugiura, Y. Makihara, and Y. Yagi, "Gait identification based on multi-view observations using omnidirectional camera," in *Computer Vision—ACCV*. Heidelberg, Germany: Springer Verlag, 2007, pp. 452–461.
- [15] G. Zhao, G. Liu, H. Li, and M. Pietikainen, "3D gait recognition using multiple cameras," in *Proc. IEEE Int. Conf. Autom. Face Gesture Recognit.*, Apr. 2006, pp. 529–534.
- [16] D. Muramatsu, A. Shiraiishi, Y. Makihara, M. Z. Uddin, and Y. Yagi, "Gait-based person recognition using arbitrary view transformation model," *IEEE Trans. Image Process.*, vol. 24, no. 1, pp. 140–154, Jan. 2015.
- [17] T. Connie, M. K. O. Goh, and A. B. J. Teoh, "A Grassmannian approach to address view change problem in gait recognition," *IEEE Trans. Cybern.*, vol. 47, no. 6, pp. 1395–1408, Jun. 2017.
- [18] W. Kusakunniran, Q. Wu, J. Zhang, and H. Li, "Multi-view gait recognition based on motion regression using multilayer perceptron," in *Proc. IEEE Int. Conf. Pattern Recognit.*, Aug. 2010, pp. 2186–2189.
- [19] 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.
- [20] 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 Int. Conf. Comput. Vis. Workshops*, Sep./Oct. 2009, pp. 1058–1064.
- [21] 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 Int. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 974–981.
- [22] J. Lu and Y.-P. Tan, "Uncorrelated discriminant simplex analysis for view-invariant gait signal computing," *Pattern Recognit. Lett.*, vol. 31, no. 5, pp. 382–393, 2010.
- [23] 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.
- [24] 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.
- [25] T. Zhang, D. Tao, X. Li, and J. Yang, "Patch alignment for dimensionality reduction," *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 9, pp. 1299–1313, Sep. 2009.
- [26] S. Yan, D. Xu, B. Zhang, H.-J. Zhang, Q. Yang, and S. Lin, "Graph embedding and extensions: A general framework for dimensionality reduction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 1, pp. 40–51, Jan. 2007.
- [27] 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.
- [28] D. S. Matovski, M. S. Nixon, S. Mahmoodi, and J. N. Carter, "The effect of time on gait recognition performance," *IEEE Trans. Inf. Forensics Security*, vol. 7, no. 2, pp. 543–552, Apr. 2012.
- [29] A. F. Bobick and A. Y. Johnson, "Gait recognition using static, activity-specific parameters," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, vol. 1, Dec. 2001, pp. I-423–I-430.
- [30] W. Kusakunniran, Q. Wu, J. Zhang, and H. Li, "Gait recognition across various walking speeds using higher order shape configuration based on a differential composition model," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 42, no. 6, pp. 1654–1668, Dec. 2012.
- [31] S. Huang, A. Elgammal, J. Lu, and D. Yang, "Cross-speed gait recognition using speed-invariant gait templates and globality–locality preserving projections," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 10, pp. 2071–2083, Oct. 2015.
- [32] 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.
- [33] J. Zhang, J. Pu, C. Chen, and R. Fleischer, "Low-resolution gait recognition," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 40, no. 4, pp. 986–996, Aug. 2010.
- [34] N. Akae, A. Mansur, Y. Makihara, and Y. Yagi, "Video from nearly still: An application to low frame-rate gait recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 1537–1543.
- [35] 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, Nov. 2009.
- [36] 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./Oct. 2014, pp. 1–8.
- [37] 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.
- [38] 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](https://doi.org/10.1109/TCSVT.2017.2760835).
- [39] S. Yu, H. Chen, E. B. G. Reyes, and N. Poh, "GaitGAN: Invariant gait feature extraction using generative adversarial networks," in *Proc. CVPR Workshops*, Jul. 2017, pp. 532–539.
- [40] D. R. Hardoon, S. Szedmak, and J. Shawe-Taylor, "Canonical correlation analysis: An overview with application to learning methods," *Neural Comput.*, vol. 16, no. 12, pp. 2639–2664, 2004.
- [41] 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.
- [42] X. Ben, W. Meng, and R. Yan, "Dual-ellipse fitting approach for robust gait periodicity detection," *Neurocomputing*, vol. 79, pp. 173–178, Mar. 2012.
- [43] X. Ben, W. Meng, K. Wang, and R. Yan, "An adaptive neural networks formulation for the two-dimensional principal component analysis," *Neural Comput. Appl.*, vol. 27, no. 5, pp. 1245–1261, Jul. 2016.
- [44] 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. IEEE Int. Conf. Pattern Recognit.*, vol. 4, Aug. 2006, pp. 441–444.
- [45] 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.
- [46] [Online]. Available: <http://www.am.sanken.osaka-u.ac.jp/BiometricDB/dataset/GaitLP/Benc-hmarks.html>
- [47] R. Martín-Félez and T. Xiang, "Uncooperative gait recognition by learning to rank," *Pattern Recognit.*, vol. 47, no. 12, pp. 3793–3806, Dec. 2014.
- [48] N. Takemura, Y. Makihara, D. Muramatsu, T. Echigo, and Y. Yagi, "Multi-view large population gait dataset and its performance evaluation for cross-view gait recognition," *IPSI Trans. Comput. Vis. Appl.*, vol. 10, no. 1, pp. 1–14, 2018.