

GII Representation-Based Cross-View Gait Recognition by Discriminative Projection With List-Wise Constraints

Zhaoxiang Zhang, *Senior Member, IEEE*, Jiaxin Chen, Qiang Wu, and Ling Shao

Abstract—Remote person identification by gait is one of the most important topics in the field of computer vision and pattern recognition. However, gait recognition suffers severely from the appearance variance caused by the view change. It is very common that gait recognition has a high performance when the view is fixed but the performance will have a sharp decrease when the view variance becomes significant. Existing approaches have tried all kinds of strategies like tensor analysis or view transform models to slow down the trend of performance decrease but still have potential for further improvement. In this paper, a discriminative projection with list-wise constraints (DPLC) is proposed to deal with view variance in cross-view gait recognition, which has been further refined by introducing a rectification term to automatically capture the principal discriminative information. The DPLC with rectification (DPLCR) embeds list-wise relative similarity measurement among intraclass and inner-class individuals, which can learn a more discriminative and robust projection. Based on the original DPLCR, we have introduced the kernel trick to exploit nonlinear cross-view correlations and extended DPLCR to deal with the problem of multiview gait recognition. Moreover, a simple yet efficient gait representation, namely gait individuality image (GII), based on gait energy image is proposed, which could better capture the discriminative information for cross view gait recognition. Experiments have been conducted in the CASIA-B database and the experimental results demonstrate the outstanding performance of both the DPLCR framework and the new GII representation. It is shown that the DPLCR-based cross-view gait recognition has outperformed the-state-of-the-art approaches in almost all cases under large view variance. The combination of the GII representation and the DPLCR has further enhanced the performance to be a new benchmark for cross-view gait recognition.

Index Terms—Cross-view gait recognition, discriminative projection, gait individuality image (GII), list-wise constraints.

I. INTRODUCTION

GAIT is an attractive biometric with the advantages of being long-distance, noninvasive, and hard to hide. Despite its prevalence in video surveillance applications, gait-based recognition is still very challenging due to variations such as viewpoint, walking speed, walking surfaces, clothing, and carry conditions [1]. Among all these variations, the change of view may be the most common and critical one, considering that gaits from the same person appear to be quite different when observed from different views (see Fig. 2). On the other hand, in real scenarios, camera view is always unconstrained, making cross-view gait recognition a rather challenging problem. Taking CASIA-B [2], as shown in Fig. 1, a widely used public gait dataset for evaluating different cross-view gait recognition methods, for example, when the probe and gallery gaits are from the same view (for all 11 views, evenly distributed from 0° to 180°), even the baseline Euclidean distance [2] could achieve a rather high performance with 97.6% correct classification rates. However, the performance drops sharply from 97.6% to 4.8%, when the gallery view is 90° and the probe view varies from 90° to 36° .

A common way to deal with this problem is by transforming gait features from one view to another or exploiting their correlations before measuring gait similarities. In works [3] and [4], view transformation model (VTM) based on singular value decomposition (SVD) was introduced to handle feature transformation across views. In addition, a regression-based VTM was proposed for mappings to single-pixels [5]. These methods can cope with view variations without relying on multiple cameras or camera calibration. However, the transformation process may fail to work because either complete gait feature mapping or single-pixel mapping is sensitive to self-occlusion. An alternative is the subspace projection method, aiming to transform features from two views into a common discriminative subspace. Canonical correlations analysis (CCA), which treats features under two distinct views as two random variables and constructs projections by maximizing their statistical correlations. However, CCA is an unsupervised dimensionality reduction method without exploiting the label information, resulting in a limited recognition performance.

Manuscript received May 10, 2017; accepted September 11, 2017. Date of publication October 13, 2017; date of current version September 14, 2018. This work was supported in part by the National Natural Science Foundation of China under Grant 61773375, Grant 61375036, and Grant 61511130079, and in part by the Microsoft Collaborative Research Project. This paper was recommended by Associate Editor B. W. Schuller. (*Corresponding author: Jiaxin Chen.*)

Z. Zhang is with the Center for Excellence in Brain Science and Intelligence Technology, Chinese Academy of Sciences, Beijing 100190, China, also with the National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China, and also with the University of Chinese Academy of Sciences, Beijing 100049, China.

J. Chen is with the Department of Electrical and Computer Engineering, New York University Abu Dhabi, Abu Dhabi 129188, UAE (e-mail: jc8531@nyu.edu).

Q. Wu is with the Global Big Data Technologies Center, University of Technology Sydney, Sydney, NSW 2007, Australia.

L. Shao is with the School of Computing Sciences, University of East Anglia, Norwich NR4 7TJ, U.K.

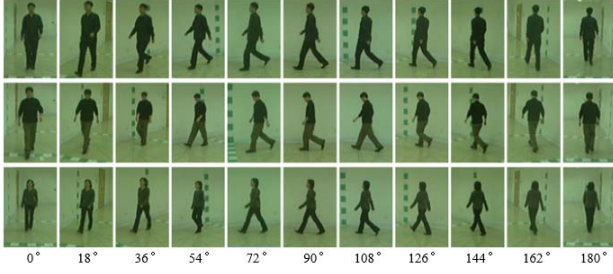


Fig. 1. Sample images from the CASIA gait database B [2] under 11 different views. The images in the same row belong to the same person.

Generally, all the aforementioned approaches attempt to explore the underlying cross-view relations, namely, transformations or correlations between features from distinct views. When the view variation is relatively small (for instance, smaller than 36°), the underlying cross-view relations would take on simple forms (for instance, approximately linear), which could be properly handled with existing methods. However, when the view variation is sufficiently large (for instance, the view change is large than 36°), the appearance of human gait would be seriously distorted. The underlying cross-view relations would become rather complicated, and take on highly nonlinear characteristics, which is fundamentally difficult to be recovered. Thus, most existing approaches focusing on feature transformation or correlation modeling always yield rather poor performance under large view variations.

To handle the cross-view gait recognition under large view variations, we propose the discriminative projection method with list-wise constraints (DPLC), which is further enhanced by adding a rectification term to be DPLC with rectification (DPLCR). Intuitively, DPLCR aims to learn a discriminative projection by maximizing the relevances of instances of the same person from distinct views while minimizing the relevance of instances of different persons in the projection subspace. To achieve this goal, we introduce the list-wise constraints, which encodes similarity information between the probe instance and any of the remaining instances. Specifically, we predefine a similarity list for each person with others, and the similarity list obtained after the discriminative projection is enforced to be as close as to the predefined one. Compared with the tractional pairwise constraint where only the similarity of an instance pair is concerned, the list-wise constraint contains more abundant similarity information, thus yields more discriminative projections. We formulate the learning of discriminative projections into an optimization problem, which can be solved by an iterative algorithm. Different from CCA, our DPLCR method utilizes the label information and seeks only one projection, instead of two projections for the probe view and the gallery view, respectively. Furthermore, in order to capture the nonlinear characteristics of the cross-view correlation under large view variations, we employ the kernel trick, and propose a kernel-DPLCR.

Moreover, we propose a simple yet efficient gait representation. The currently most widely used gait representation is gait energy image (GEI), which is the average of a silhouette

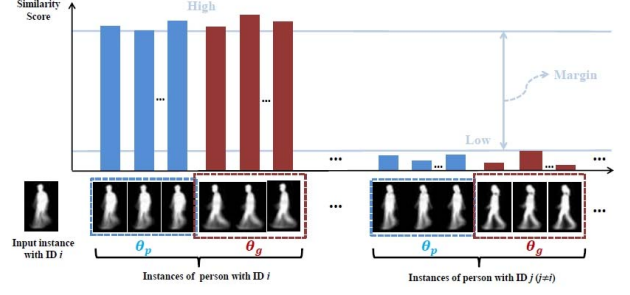


Fig. 2. Motivation of the proposed method: θ_p denotes the probe view and θ_g stands for the gallery view.

sequence. Although GEI contains discriminative information for gait recognition under the same view, it is sensitive to view variations, of which the performance drops sharply with increasing view variations. In this paper, we propose a novel gait representation based on GEI. Specifically, we first construct a synthetic common “GEI template” of a certain view by averaging GEIs from all training individual gaits captured under the corresponding camera view. We believe that the GEI template mainly contains view dependent information and has little individual identify information, which is thus less discriminative for cross-view gait recognition. It is then reasonable to subtract this common GEI template from each individual GEI, since during this operation the individual identity information is reserved while the view dependent information contained in GEI template is eliminated. Therefore, we could finally obtain a more discriminative and view-independent gait representation, which we called gait individuality image (GII) throughout this paper.

The contributions of this paper lie in twofold.

- 1) We propose a novel model for cross-view gait recognition, namely DPLCR. Different from existing methods, DPLCR employs list-wise constraints to learn a projection, which can map gait representations from different views into a common discriminative subspace. We further extend this model to multiview DPLCR, which can efficiently incorporate multiview information.
- 2) We present a new feature gait representation, called gait individual image (GII), for cross-view identity recognition. GII removes redundant view-dependent parts from GEI while reserving individual identification information, which are more discriminative and view-independent.

The remainder of this paper is organized as follows. In Section II, we briefly review the existing work for cross-view gait recognition. In Section III, we introduce the formulation of DPLCR and its optimization solution. In Section IV, the extraction of GII is described in detail. In Section VI, we evaluate the performance of the proposed method on CASIA-B gait database and USF gait database. Finally, conclusions are drawn in Section VII.

II. RELATED WORK

A. Gait Feature Extraction

Over the past decade, a large number of methods on gait feature extraction have been proposed. Generally, they can

be divided into two categories: 1) model-based methods and 2) model-free methods. Model-based methods [6]–[10] aim to model the body motion when he/she is walking so that static gait physical parameters can be extracted. However, these methods usually have limitation performance due to the highly complex structure of human body and self-occlusion.

Most model-free methods focus on extracting the static information of a walking person. Among them, GEI [11] have proved to be the most popular one for its simpleness and effectiveness. GEI is the average of a gait sequence in one or more cycle(s) and turns out to be robust to noise from individual image. Enhanced GEI was proposed in [12] by assigning different weights to each pixel in GEI based on its variance. Similarly, Choudhury and Tjahjadi [13] used the entropy of the limb region in GEI and applied Gaussian filter in order to achieve insensitiveness to boundary shape alterations caused by carrying conditions and clothing variations. To get more dynamic information, Zhang *et al.* [14] proposed active energy image, which is the average of the difference images between two adjacent silhouettes. Wang *et al.* [15] proposed a new gait representation based on body contour called chrono-gait image (CGI). More recently, Arora *et al.* [16] introduced a new period dependent gait representation called gait information image, which was derived by applying the concept of information set to remove the uncertainty of the pixel intensities in a silhouette image of a person over a gait cycle. Similarly, CGI is the average of body contours which are mapped into RGB space based on the time it occurs and thus reverses more temporal information. Besides, there are also gait representations based on optical flow [17], radon transform [18], and Gabor transform [19]. All these features are designed for gait recognition under the same condition and get poor performance in handling the view change situation without using view-independent methods.

B. Cross-View Gait Recognition

To meet the requirement of applications such as video surveillance, the research interests in cross-view gait recognition have been increased recently. There are three major approach categories for this issue: 1) constructing 3-D gait information by using multiple cameras or camera calibration; 2) extracting gait features robust to view change; and 3) learning projection relationship of gaits across views.

The first category is to construct 3-D gait information based on cooperative multiple cameras [20]–[23] or camera calibration [24]. Theoretically, 2-D gaits of any arbitrary view can be synthesized if 3-D model of gaits are constructed reliably. However, the needs for multiple cameras and camera calibration limit the practicability of gait recognition in real applications.

The second category is to perform cross-view gait recognition by obtaining view-invariant features. BenAbdelkader *et al.* [25] made use of self similarity plot to enhance the robustness to view changes, which achieved good performance with a limited range of view change. Kale *et al.* [26] proposed a method to obtain the side-view

gaits from any arbitrary view by using the perspective projection model. Similarly, Jean *et al.* [27] normalized the body trajectories of any arbitrary view to a standard plane so that their similarities can be compared directly. Han *et al.* [28] developed a statistical approach for view-insensitive gait recognition by analyzing the common properties of GEI along different directions. Goffredo *et al.* [29] extracted model-based gait feature, namely angular measurements and trunk spatial displacements, which are then reconstructed using view-rectification method. Although this method has outperforming results than the methods in the same category when view change is large enough, its performance may be worse than many other methods across similar views. In this category, method [29] achieves the state-of-art performance.

The third category is to transfer gait features from one view to another by learning their projection relationship before similarity measurement. Martín-Félez and Xiang [30] formulated gait recognition as a bipartite ranking problem and leveraged training samples from different persons. Liu *et al.* [31] proposed a joint subspace learning method for view-invariant gait recognition. Makihara *et al.* [3] introduced view transfer model (VTM) into cross-view gait analysis, in which SVD is applied on frequency domain features to construct the VTM. Kusakunniran *et al.* [4] adopted linear discriminant analysis (LDA) to optimize the GEI feature, and they also used the VTM based on SVD. In method [32], Zheng *et al.* further proposed a method to establish a robust VTM via robust principal component analysis. There are also methods constructing VTM based on support vector regression (SVR) [5], multilayer perceptron [33], and sparse regression [34]. Chen *et al.* [35] constructed the view transformation matrix based on projection of gravity center trajectory.

More recently, CCA is introduced to project gaits from different views onto two subspaces, in which their correlations are mutually maximized. Such methods based on CCA also belong to the third category in this paper. Compared with VTM, the CCA model can capture the projection relationship better between gait features of different views because it is more robust against feature noise and deal with feature mismatch across views better. The method in [36] model the correlations of gaits across views by using CCA on global gait feature. To further improve the performance, Kusakunniran *et al.* [37] first carried out motion co-clustering to partition the most related parts of gaits from different views into the same group. After that, CCA is applied on the different groups and the final similarity is the sum of similarities calculated from the different groups. Their method achieves the start-of-art performance of cross-view and multiview gait recognition.

III. DISCRIMINATIVE PROJECTION WITH LIST-WISE CONSTRAINTS FOR CROSS-VIEW GAIT RECOGNITION

A. Problem Formulation

For brief description, we suppose that the training data consists of instances of N_{tr} persons from the probe view θ_p and the gallery view θ_g , denoted by $X = \{X_1^T, X_2^T, \dots, X_{N_{tr}}^T\}$, where $X_i = \{x_{\theta_p,1}^i, \dots, x_{\theta_p,n_p}^i, x_{\theta_g,1}^i, \dots, x_{\theta_g,n_g}^i\}$ are $n^i = n_p^i + n_g^i$

D -dimensional features of the i th person consisting of n_p^i features from view θ_p and n_g^i features from view θ_g . For simplicity, we assume that the complete training data X is an $N \times D$ matrix, where each row x_i of X is a feature vector, and $N = \sum_{i=1}^{N_{tr}} n^i$.

As previously described, we treat the cross-view gait recognition as an image retrieval task, that is, taking X as the candidate set and every feature vector x_i as an input query. Specifically, for every input query $x_{\theta,j}^i$, or the j th instance of person i from view θ , we seek for a projection $P \in R^{D \times d}$ such that in the projection d -dimensional subspace with respect to P

$$\begin{aligned} \text{sim}(x_{\theta,m}^i P, x_{\theta_p,u}^i P) &\approx \text{sim}(x_{\theta,m}^i P, x_{\theta_g,v}^i P) \\ u &\in \{1, \dots, n_p^i\}, v \in \{1, \dots, n_g^i\}. \end{aligned} \quad (1)$$

$$\begin{aligned} \text{sim}(x_{\theta,m}^i P, x_{\theta',l}^i P) &\geq \text{sim}(x_{\theta,m}^i P, x_{\theta'',k}^i P) + \Delta \\ \theta', \theta'' &\in \{\theta^p, \theta^g\} \text{ and } i \neq j \end{aligned} \quad (2)$$

where $\text{sim}(\cdot, \cdot)$ is a similarity measure and Δ is the margin.

Intuitively, (1) indicates that the projection P would diminish the cross-view variations of the same person, and (2) implies that P could enlarge the interclass dissimilarities. Through these two inequalities, we can expect that the projection P satisfying (1) and (2) will be discriminative to recognize human identities and robust to view variations.

In this paper, we select the inner product as the similarity measure. That is to say, for two feature vectors x_i, x_j , their similarity $s_P(i, j)$ in the projected subspace with respect to P is calculated as follows:

$$s_P(i, j) = \langle x_i P, x_j P \rangle = (x_i P) \cdot (x_j P)^T = x_i P P^T x_j^T.$$

Thus, for each instance x_i , we could measure its similarities in the projected subspace with all the other instances x_1, \dots, x_N , and could obtain a similarity list $S(i) = [s_P(i, 1), \dots, s_P(i, N)]$. As shown previously, (1) and (2) could yield a discriminative projection matrix P . Concretely, in this paper we implement these two inequalities by introducing list-wise constraints, i.e., minimizing the discrepancies between $S(i)$ and the ground truth similarities $L(i) = [l(i, 1), \dots, l(i, N)]$

$$P = \underset{P \in R^{D \times d}}{\text{argmin}} \sum_{i=1}^N \|S(i) - L(i)\|_2^2. \quad (3)$$

By reformulating (3) into a more compact matrix form, we can obtain the following equivalent problem:

$$\min_{P \in R^{D \times d}} \|X P P^T X^T - L\|_F^2 \quad (4)$$

where the i th row of L is $L(i)$, and $\|\cdot\|_F$ stands for the matrix Frobenius norm.

In practice, $L(i)$ is rarely available in practice. While, in order to satisfying (1) and (2), it is usually expected that $l(i, j) = l(i, k) > l(i, l)$, where the i th, j th, and k th instances own the same human identity, and l th instance has different identity from the i th instance. Here, we predefine the following approximations:

$$L(i, j) = \begin{cases} 1, & \text{if } x_i, x_j \text{ have the same person identity} \\ 0, & \text{otherwise.} \end{cases} \quad (5)$$

From (5), it could be observed that the optimal projection P learned from (3) implicitly satisfies (1) and (2), where the margin term Δ equals 1.

However, (5) predefines L by hand, which only roughly capture the principal discriminative information. It is preferred to automatically exploit the similarity information from training data, and refine the hard assigned similarity lists in (5). Based on this motivation, we add a rectification term c_{ij} to each $L(i, j)$ defined in (5), and get the final similarities as $\tilde{L}(i, j) = L(i, j) + \beta c_{ij}$, where β is a tradeoff parameters to balance the effect of the principal similarity term $L(i, j)$ and the rectification term c_{ij} . Equation (4) is then turned into

$$\min_{P \in R^{D \times d}, C \in R^{N \times N}, C^T C = I_N} \|X P P^T X^T - (L + \beta C)\|_F^2 \quad (6)$$

where C is the rectification matrix satisfying $C(i, j) = c_{ij}$. It is worth noticing that an additional constraint $C^T C = I_N$ is added in order to prevent C from getting too large to deteriorate the principal discriminative information L and yielding meaningless similarity similarities.

Finally, we introduce an additional regularization term $\text{tr}(P^T P)$ to avoid over-fitting, by which (4) is finally turned to

$$\min_{C^T C = I_N} \|X P P^T X^T - (L + \beta C)\|_F^2 + \lambda \cdot \text{tr}(P^T P) \quad (7)$$

where λ is the tradeoff parameter balancing the error term and the regularization term. Here, we omit the trivial constraints $P \in R^{D \times d}, C \in R^{N \times N}$ for a neat description.

As can be seen in (5) and (7), by minimizing the difference between $X P P^T X^T$ and L , the intraperson similarities (similarities between instances belonging to the same person from different views) is forced to be 1, while the interperson similarities (similarities between instances belonging to distinct persons) are forced to be 0, with a maximum margin 1. Through this way, the discriminative capability of the learned subspace projection P in handling with cross-view variations can be enhanced. On the other hand, the elements in rectification variable C perform as slack variables, allowing for violations of maximum margins. As a consequence, the generalization ability of P may be boosted. Therefore, the list-wise constraints can improve the overall performance of the projection matrix P .

B. Optimization Method

It is obviously that (7) is a nonconvex optimization problem, which means that the global optimal solution may not exist. Here, we alternatively optimize P with fixed C , and in turn optimize C with fixed P .

Concretely, we first denote the objective function by $F(P, C) = \|X P P^T X^T - (L + \beta C)\|_F^2 + \lambda \cdot \text{tr}(P^T P)$. Assuming that we have obtained values of variables P, C in the $(t-1)$ th step. For fixed $P^{(t-1)}$, the subproblem for optimizing C is as the following:

$$\begin{aligned} \min_{C^T C = I_N} & \|X P^{(t-1)} P^{(t-1)T} X^T - (L + \beta C)\|_F^2 \\ & + \lambda \cdot \text{tr}(P^{(t-1)T} P^{(t-1)}). \end{aligned} \quad (8)$$

Let $E = XP^{(t-1)}P^{(t-1)T}X^T - L$ and considering the constraint $C^TC = I_N$, it is not difficult to derive that (8) is equivalent to the following problem:

$$\max_{C^TC=I_N} \text{tr}(C^TE). \quad (9)$$

Though the problem above is not convex due to the constraint $C^TC = I_N$, (9) has a closed form solution. Suppose that the eigenvalue decomposition of E is $E = U_E \Lambda_E U_E^T$, then

$$\text{tr}(C^TE) = \text{tr}(C^T U_E \Lambda_E U_E^T) = \text{tr}(U_E^T C^T U_E \Lambda_E) \quad (10)$$

where $\Lambda_E = \text{diag}(\lambda_1, \dots, \lambda_N)$, λ_i is the i th largest eigenvalue of E . The second equality in (10) holds, due to the property $\text{tr}(AB) = \text{tr}(BA)$.

Denoting $Z = (z_{ij})_{N \times N} = U_E^T C^T U_E$, then $\text{tr}(C^TE) = \text{tr}(Z \Lambda_E) = \sum_{i=1}^N \lambda_i z_{ii}$. On the other hand, from the constraint $C^TC = I_N$, it is easy to derived that $ZZ^T = I_N$, which yields that $|z_{ii}| \leq 1$. Therefore, $\text{tr}(C^TE)$ is maximized when

$$z_{ij}^* = \begin{cases} -1, & \text{if } i = j, \text{ and } \sigma_i < 0 \\ 1, & \text{if } i = j, \text{ and } \sigma_i \geq 0 \\ 0, & \text{if } i \neq j. \end{cases} \quad (11)$$

Denoting that $Z^* = (z_{ij}^*)_{N \times N}$, then the close form solution to subproblem (8) is as follows:

$$C^* = U_E Z^* U_E^T. \quad (12)$$

By setting $C^{(t)} = C^*$, we finally obtain the update of $C^{(t)}$.

When C is fixed, for instance, $C = C^{(t)}$, the original problem (7) turns into the following subproblem:

$$\min \left\| XPP^T X^T - (L + \beta C^{(t)}) \right\|_F^2 + \lambda \cdot \text{tr}(P^T P). \quad (13)$$

The subproblem above is not convex, either. We apply the gradient descend method with backtracking line search to get a local minima where the start point is set as $P^{(t-1)}$, and obtain the update of P in the t th step

$$P^{(t)} = P^{(t-1)} - \alpha \cdot \frac{\partial F}{\partial P}(P^{(t-1)}, C^{(t)}) \quad (14)$$

where $(\partial F / \partial P)$ denotes the partial derivative of F with respect to P .

The formulation in Algorithm 1 can be further generalized to support nonlinear projection by using the kernel trick. Similar to [38], we first project the data into a reproducing kernel Hilbert space (RKHS) \mathcal{H} via a feature map ϕ with corresponding kernel function $k(x, y) = \langle \phi(x), \phi(y) \rangle_{\mathcal{H}}$. Thereafter, the data is project to \mathbb{R}^d by a linear projection $M : \mathcal{H} \rightarrow \mathbb{R}^d$. The projection matrix P then can be expressed as $P = \Phi \tilde{P}$, where $\Phi_i = \phi(x_i)$. Thus, we can reformulate (7) as an optimization problem over \tilde{P} rather than P . Since $\Phi^T \Phi \tilde{P} \tilde{P}^T \Phi^T \Phi = K \tilde{P} \tilde{P}^T K$, where $K_{i,j} = k(x_i, x_j)$, (7) could be rewritten as

$$\min_{C^TC=I_N} \left\| KPP^T K^T - (L + \beta C) \right\|_F^2 + \lambda \cdot \text{tr}(P^T P). \quad (15)$$

We can see that (15) has the same form as (7) except that X is replaced by K . Therefore, (7) can be solved by using the same approach. We summarize the overall solution in Algorithm 1. The proof of convergence of Algorithm 1 is straightforward, and we therefore omit it here.

Note that the optimization problem (13) in the initialization step has an efficient close form solution, according to [38].

Algorithm 1 (Kernel) DPLCR for Cross-View Gait Recognition

Input: training instances of N_{tr} persons from the probe view θ_p and gallery view θ_g $X = \{X_1^T, \dots, X_{N_{tr}}^T\}$ where $X_i = \{x_{\theta_p,1}^i, \dots, x_{\theta_p,n_p}^i, x_{\theta_g,1}^i, \dots, x_{\theta_g,n_g}^i\}$ (in Kernel DPLCR, $X = K$ where K is the kernel matrix), projection dimension d , tradeoff parameters β, λ .

Output: the projection matrix P .

- 1: Initialization: $t = 1$, principal similarity matrix L as defined in (5) and

$$P^{(0)} = \underset{P \in \mathbb{R}^{N \times d}}{\text{argmin}} \|XPP^T X^T - L\|_F^2 + \lambda \|P\|_F^2. \quad (16)$$

while not convergence **do**

- 2: Calculate $E = XP^{(t-1)}P^{(t-1)T}X^T - L$, do eigenvalue decomposition of E as $E = U_E \Lambda_E U_E^T$, and update the rectification term C as in (12): $C^{(t)} = U_E Z^* U_E^T$, where Z^* is defined in (11).
- 3: Calculate the step length $\alpha^{(t)}$ using backtracking line search, and update P as follows:

$$P^{(t)} = P^{(t-1)} - \alpha^{(t)} \cdot \frac{\partial F}{\partial P}(P^{(t-1)}, C^{(t)}),$$

where $F = \|KPP^T K^T - (L + \beta C)\|_F^2 + \lambda \cdot \text{tr}(P^T P)$.

- 4: $t := t + 1$.

end while

IV. PROPOSED GAIT INDIVIDUALITY IMAGE REPRESENTATION

In this section, we will introduce a novel gait representation, namely, GII, which is simple yet efficient. As GII is mostly based on GEI, a brief description for GEI is presented first. And GII is described in detail, including its motivation and intuitive explanation.

A. Gait Energy Image

Given sequences of silhouettes from multiple views, silhouette process including size normalization and alignment is carried out for GEI extraction. Two different definitions of GEI are introduced in [2] and [11]. The first one considers the gait cycle and GEI is the average image of gait sequences from one or more complete gait period(s) [11]. The second one just defined GEI as the average image of all available silhouettes [2]. In fact, there is no significant difference on performance when gait sequences are long enough according to our observation. However, the periodicity detection of gait sequences near the front view or rear view is another challenging field of gait analysis. In this paper, we use the second definition as follows:

$$G(x, y) = \frac{1}{N} \sum_{t=1}^N F_t(x, y) \quad (17)$$

where N is the number of available frames in a gait sequence, $F_t(x, y)$ is the t th frame and x and y are 2-D image coordinates.

The GEI feature contains rich information for recognition including global body shape, temporal, and spatial changes

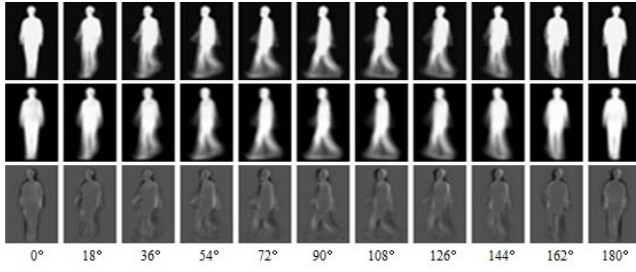


Fig. 3. Comparison of GEI, average GEI, and GII under 11 different views. All the images belong to the same person. From up to down, the images in each row correspond to GEI, average GEI, and GII, respectively. Note that the image values are normalized for the convenience of visualization.

(see Fig. 3). It is also robust against image noise from individual image. Note that view of the gait can be estimated by support vector machine or Gaussian process [36]. In this paper, we assume that all gait sequences are from known views.

B. Gait Individuality Image

Despite its discriminative capability in recognizing people captured from the same view, GEI suffers as the view varies. So we propose a novel representation based on GEI to bridge the gap between different views. The motivation of GII lie on that we attempt to filter the personal GEI by using a common gait template which is defined as the average GEI of a training set. In this paper, we obtain the individual gait information by simply subtracting the common gait template.

Formally, GII for person p in view v is defined as the following:

$$I_p^v(x, y) = G_p^v(x, y) - \frac{1}{M} \sum_{m=1}^M G_m^v(x, y) \quad (18)$$

where M is the total number of GEIs in training data from the probe (gallery) view, $G_m^v(x, y)$ is the m th GEI in view v . Note that $G_p^v(x, y)$ can be an instance either from training data or test data in view v .

Fig. 3 shows the comparison of GEI, average GEI, and GII under different views. Since many elements in GII are negative values, we normalize all of them to $[0, 255]$. From (18), it could be observed that elements in GII are distributed in $[-255, 255]$. Thus, values of GII can be divided into three parts: 1) large positive values, indicating the positions where the personal GEI owns more than the average GEI and corresponding to the “light” part in Fig. 3; 2) large negative values with opposite meanings, corresponding to “dark” parts in Fig. 3; and 3) values proximately equaling to 0 corresponding to the “gray” part in Fig. 3. Compare with existing features such as GEI, CGI, and GFI, the proposed features contain more abundant information for gait recognition, especially for cross-view gait recognition. Moreover, GII avoids computing the distance of the common part across views, which are less discriminative to cross-view gait recognition, and may bring interferences due to appearance distortion caused by view variations.

In detail, body parts in legs, arms, head, and contour of the trunk in GII are important parts which contain rich information for cross-view gait recognition. It can be observed from Fig. 3 that the gait appearances just change a little as the view ranges from 36° to 144° . But when it is near frontal view (0°) or rear view (180°), they are quite different. So GII has outperforming advantages across views in a certain range but still encounters problems for large view variants (e.g., view 18° versus view 90°).

V. GAIT RECOGNITION UNDER VARIOUS VIEWS

A. Cross-View Gait Recognition

As described in Algorithm 1 of Section III, we could obtain a projection matrix P for the probe view θ_p and gallery view θ_g . In the cross-view phase, given any gait instance $x_{\theta_p,j}$ from the probe view, and gait instance $x_{\theta_g,k}$ from the gallery view, we first transform them into a common subspace by P : $\tilde{x}_{\theta_p,j} = x_{\theta_p,j}P$, $\tilde{x}_{\theta_g,k} = x_{\theta_g,k}P$, and calculate the inner product in the projected space as the final similarity score: $\text{sim}(x_{\theta_p,j}, x_{\theta_g,k}) = x_{\theta_p,j}P P^T x_{\theta_g,k}^T$. Higher similarity score indicates higher probability that $x_{\theta_p,j}$, $x_{\theta_g,k}$ own the same human identity.

For kernel DPLCR, the overall transformation matrix in the RKHS space is $\tilde{P} = \Phi P$. The similarity $\text{sim}(x_{\theta_p,j}, x_{\theta_g,k})$ can be calculated by using the learned \tilde{P}

$$\begin{aligned} \text{sim}(x_{\theta_p,j}, x_{\theta_g,k}) &= \phi(x_{\theta_p,j}) \cdot \tilde{P} \tilde{P}^T \cdot \phi(x_{\theta_g,k}) \\ &= \phi(x_{\theta_p,j}) \cdot \Phi P P^T \Phi^T \cdot \phi(x_{\theta_g,k}) \\ &= \Phi_{\theta_p} P P^T \Phi_{\theta_g}^T \end{aligned}$$

where $\Phi_{\theta_p}(i) = k(x_{\theta_p,j}, x_i)$, and $\Phi_{\theta_g}(i) = k(x_{\theta_g,k}, x_i)$, and x_i is the i th instance from training data.

B. Multiview Gait Recognition

In multiview gait recognition, there are available data from at least two views for gallery and/or probe. Assuming that we have A views available for probe ($X_{\theta_{p1}}, X_{\theta_{p2}}, \dots, X_{\theta_{pA}}$), and B views available for gallery ($X_{\theta_{g1}}, X_{\theta_{g2}}, \dots, X_{\theta_{gB}}$). In this paper, we adopt three different strategies for multiview fusion: feature level fusion strategy (S1), score level fusion strategy (S2), and strategy 3 (S3) by extending the cross-view DPLCR to multiview DPLCR.

1) *S1 (Feature Level Fusion)*: This kind of fusion strategy is widely used in gait recognition [37], where probe and gallery gait feature vectors are defined as

$$\begin{aligned} X_{\theta_A} &= [X_{\theta_{p1}} : X_{\theta_{p2}} : \dots : X_{\theta_{pA}}] \\ X_{\theta_B} &= [X_{\theta_{g1}} : X_{\theta_{g2}} : \dots : X_{\theta_{gB}}]. \end{aligned} \quad (19)$$

Here, “:” means the concatenation. Then, replacing X_{θ_p} , X_{θ_g} by X_{θ_A} and X_{θ_B} , the cross-view gait recognition could be directly applied to multiview gait recognition.

2) *S2 (Score Level Fusion)*: Besides the feature level fusion, we can also solve the multiview gait recognition utilizing score level fusion. Specifically, we separately train a projection matrix $P_{\theta_{pi}, \theta_{gj}}$ for each probe and gallery view pair

$(\theta_{pi}, \theta_{gj})$ ($i \in \{1, \dots, A\}, j \in \{1, \dots, B\}$) by solving the following optimization problem using Algorithm 1 for cross-view gait recognition:

$$\min_{C_{pi,gj}^T, C_{pi,gj}=I_N} \left\| X_{pi,gj} P_{pi,gj} P_{pi,gj}^T X_{pi,gj}^T - (L + \beta C_{pi,gj}) \right\|_F^2 + \lambda \cdot \text{tr}(P_{pi,gj}^T P_{pi,gj}) \quad (20)$$

where $X_{pi,gj} = (X_1^{(pi,pj)^T}, \dots, X_{N_{tr}}^{(pi,pj)^T})$ and $X_k^{(pi,pj)} = (x_{\theta_{pi}1}^k, \dots, x_{\theta_{pi}n_{pi}^k}^k, x_{\theta_{pj}1}^k, \dots, x_{\theta_{pj}n_{gj}^k}^k)$.

After training, we could obtain $A \times B$ projections $\{P_{\theta_{pi}, \theta_{gj}} : i = 1, \dots, A; j = 1, \dots, B\}$. For two instances $\{x_{\theta_{pi}, m}\}_{i=1}^A, \{x_{\theta_{gj}, n}\}_{j=1}^B$ consisting of multiview gaits, their similarity then could be calculated as follows:

$$\text{sim}(\{x_{\theta_{pi}, m}\}_{i=1}^A, \{x_{\theta_{gj}, n}\}_{j=1}^B) = \sum_{i=1}^A \sum_{j=1}^B x_{\theta_{pi}, m} P_{\theta_{pi}, \theta_{gj}} P_{\theta_{pi}, \theta_{gj}}^T x_{\theta_{gj}, n}^T. \quad (21)$$

3) *S3 (Multiview DPLCR)*: Intuitively, since S2 learns a projection matrix for each probe and gallery view pair, it could yield more discriminative projections than S1, which takes the multiview gaits as a whole and learns a unitary projection. However, this strategy trains each $P_{\theta_{pi}, \theta_{gj}}$ separately, and has ignored the relations among these projections. For instance, intuitively the refined similarity matrix $L + \beta C$ should consistently be the same for various probe view and gallery view pairs $\{(\theta_{pi}, \theta_{gj}) : i = 1, \dots, A; j = 1, \dots, B\}$, which is not considered in S2. Here, we propose an extension of DPLCR for multiview gait recognition by co-training all projection matrices as follows:

$$\min_{C^T C = I_N} \sum_{i=1}^A \sum_{j=1}^B \left[\left\| X_{pi,gj} P_{pi,gj} P_{pi,gj}^T X_{pi,gj}^T - (L + \beta C) \right\|_F^2 + \lambda \cdot \text{tr}(P_{pi,gj}^T P_{pi,gj}) \right]. \quad (22)$$

Similar to S2, we could also obtain $A \times B$ projections $\{P_{\theta_{pi}, \theta_{gj}} : i = 1, \dots, A; j = 1, \dots, B\}$ after training. The score level fusion as shown in (21) can then be adopted for final evaluation.

It can be observed that the only difference between S2 and S3 lies on the training of the rectification matrices. In (20) the rectification matrices $\{C_{pi,gj} : i = 1, \dots, A; j = 1, \dots, B\}$ are trained independently, while in (21) $\{C_{pi,gj} : i = 1, \dots, A; j = 1, \dots, B\}$ are consistently kept the same over all probe and gallery views, and are jointly trained. Intuitively, by forcing $C_{pi,gj}$ to be constant over different pi and gj would yield the following advantage: the inconsistencies among the similarity matrices $L + C_{pi,gj}$ would be eliminated, and the performance of the score level fusion therefore could be enhanced.

Like the training process of DPLCR, we can approach a local optimal solution of (22) by alternatively optimizing $\{P_{\theta_{pi}, \theta_{gj}}\}$ and C . For fixed C , training $\{P_{\theta_{pi}, \theta_{gj}}\}$ is equivalent to training each $P_{\theta_{pi}, \theta_{gj}}$ separately, which could be implemented

Algorithm 2 (Kernel) DPLCR for Multiview Gait Recognition (Multiview DPLCR)

Input: training gaits of N_{tr} persons from A probe views $\{\theta_{pi}\}_{i=1}^A$ and B gallery view $\{\theta_{gj}\}_{j=1}^B$ $\{X_{pi,gj} : i = 1, \dots, A; j = 1, \dots, B\}$ (in Kernel Multi-view DPLCR, $X_{pi,gj} = K_{pi,gj}$ where $K_{pi,gj}$ is the corresponding kernel matrix), projection dimension d , tradeoff parameters β, λ .

Output: $A \times B$ projection matrices $\{P_{pi,gj} : i = 1, \dots, A; j = 1, \dots, B\}$.

- 1: Initialization: $t = 1$, principal similarity matrix L as defined in (5) and

$$P^{(0)} = \underset{P \in \mathbb{R}^{N \times d}}{\text{argmin}} \|X P P^T X^T - L\|_F^2 + \lambda \|P\|_F^2. \quad (23)$$

while not convergence **do**

- 2: Calculate

$E_{AB} = \sum_{i=1}^A \sum_{j=1}^B (X_{pi,gj} P_{pi,gj}^{(t-1)} P_{pi,gj}^{(t-1)T} X_{pi,gj}^T - L)$, do eigenvalue decomposition of E_{AB} as $E_{AB} = U_{E_{AB}} \Lambda_{E_{AB}} U_{E_{AB}}^T$, and update the rectification term C as in (12): $C^{(t)} = U_{E_{AB}} Z^* U_{E_{AB}}^T$, where Z^* is defined in (11) by replacing E with E_{AB} .

- 3: Calculate the step length $\alpha_{pi,gj}^{(t)}$ using backtracking line search, and update $P_{pi,gj}$ as follows:

$$P_{pi,gj}^{(t)} = P_{pi,gj}^{(t-1)} - \alpha^{(t)} \cdot \frac{\partial F_{AB}}{\partial P_{pi,gj}}(P_{pi,gj}^{(t-1)}, C^{(t)}), \quad (24)$$

where

$$F_{AB} = \sum_{i=1}^A \sum_{j=1}^B \left[\|X_{pi,gj} P_{pi,gj} P_{pi,gj}^T X_{pi,gj}^T - (L + \beta C)\|_F^2 + \lambda \cdot \text{tr}(P_{pi,gj}^T P_{pi,gj}) \right].$$

- 4: $t := t + 1$.

end while

by adopting step 3 in Algorithm 2. For fixed $\{P_{\theta_{pi}, \theta_{gj}}\}$, it is not difficult to derive that

$$C = \max_{C^T C = I_N} \text{tr}(C^T E_{A,B}) \quad (25)$$

where $E_{A,B} = \sum_{i=1}^A \sum_{j=1}^B E_{pi,gj}$, and $E_{pi,gj} = X_{pi,gj} P_{pi,gj} P_{pi,gj}^T X_{pi,gj}^T - L$. As described in Section III, (25) has a closed form solution, which could be solved by adopting step 2 in Algorithm 2.

We summarize the previously described solution to (22) in Algorithm 2, and propose the multiview DPLCR.

VI. EXPERIMENTAL RESULTS

We test the proposed method on the CASIA gait database B [2] for experimental evaluations. The CASIA gait database B is the largest public multiview gait dataset, which contains 124 subjects captured from 11 views including 0° , 18° , 36° , 54° , 72° , 90° , 108° , 126° , 144° , 162° , and 180° . Under each view, there are ten sequences for each person: two sequences when walking with a bag, two sequences when walking with a coat, and six normal walking sequences. In this

TABLE I
PERFORMANCE COMPARISON OF DIFFERENT METHODS FOR CROSS-VIEW GAIT RECOGNITION
WHEN PROBE VIEW IS 54° AND GALLERY VIEW IS FROM 0° TO 180° EXCEPT 54°

Gallery view	0°	18°	36°	72°	90°	108°	126°	144°	162°	180°
Baseline[2]	4	9	30	22	18	17	38	19	2	3
View-rectification[29]	-	-	57	65	62	63	63	-	-	-
FT-SVD[3]	8	31	72	43	28	19	24	18	14	12
GEI-SVD[4]	13	46	87	81	49	31	27	19	18	16
GEI-SVR[5]	22	64	95	93	59	51	42	27	20	21
GEI-CCA[36]	22	56	94	88	51	47	43	27	17	15
GInI-EF[16]	5	8	35	25	18	15	28	18	2	3
GInI-SF[16]	7	9	35	19	12	11	36	23	3	4
Co-clustering[37]	24	65	97	95	63	53	48	34	23	22
GEI-DPLCR	23	64	95	95	77	69	61	50	25	17
GII-DPLCR	28	67	98	97	86	76	69	58	31	20

TABLE II
PERFORMANCE COMPARISON OF DIFFERENT METHODS FOR CROSS-VIEW GAIT RECOGNITION
WHEN PROBE VIEW IS 90° AND GALLERY VIEW IS FROM 0° TO 180° EXCEPT 90°

Gallery view	0°	18°	36°	54°	72°	108°	126°	144°	162°	180°
Baseline[2]	3	4	7	17	82	88	22	2	1	1
View-rectification[29]	-	-	53	74	73	69	67	-	-	-
FT-SVD[3]	3	6	8	27	36	58	28	21	10	3
GEI-SVD[4]	7	11	22	52	75	79	45	26	12	6
GEI-SVR[5]	16	22	35	63	95	95	65	38	20	13
GEI-CCA[36]	12	18	30	55	89	91	51	32	16	8
GInI-EF[16]	2	2	7	16	80	86	14	6	2	1
GInI-SF[16]	2	2	7	16	77	88	21	5	1	2
Co-clustering[37]	18	24	41	66	96	95	68	41	21	13
GEI-DPLCR	16	24	44	74	96	97	71	43	22	14
GII-DPLCR	17	26	54	84	98	98	84	50	25	14

paper, we use all the six normal gait sequences of 124 subjects from 11 views and thus obtain $6 \times 124 \times 11 = 8184$ sequences. Sample images from the CASIA gait database B are shown in Fig. 2.

The experimental setup is similar to [37]. In each experiment, 24 subjects from the probe view and gallery view are randomly chosen as training data. And the rest 100 subjects are used for evaluating the performance of cross-view recognition. We report the mean recognition accuracy and cumulative matching score (CMS) curve after ten rounds. For all experiments in this paper, the parameters in the proposed method are selected by cross validation, and the radial basis function kernel is utilized.

A. Gait Recognition With Single Probe View

For one single probe view, we select one view degree as the probe view, and the rest 10 view degrees as gallery views. In this paper, we follow the same experimental setting as [37], and choose 54° , 90° , and 126° as probe views. Thus, there are totally 3×10 combinations of the probe view and gallery view.

To validate the performance of the proposed DPLCR method and GII feature, we report the recognition results by using DPLCR with GEI feature (denoted as GEI-DPLCR) and DPLCR with GII feature (denoted as GII-DPLCR), respectively. We compare our method with the-state-of-art methods such as co-clustering [37], GInI-EF [16], GInI-SF [16],

GEI-CCA [36], GEI-SVD [4], FT-SVD [3], GEI-SVR [5], view-rectification [29], and the baseline method [2] which match GEIs across views using Euclidean distance directly. Note that the view-rectification method [29] uses 65 subjects for testing and some of its results are not reported.

Tables I–III show the rank-1 recognition rates of different methods for probe view 54° , 90° , and 126° , respectively. It can be seen that the proposed DPLCR with GEI feature achieves competitive performance compared with the state-of-the-arts. As to the setting where the probe view is 54° , the performance of GEI-DPLCR is at least 13 percents higher than co-clustering (highest in the comparison methods), for gallery views 90° to 144° , and is slightly better for gallery views 18° and 162° . For the rest gallery views, GEI-DPLCR also achieves compatible results. As to probe view 90° , GEI-DPLCR outperforms co-clustering (highest in the comparison methods), for most gallery views except 0° and 18° , where GEI-DPLCR still achieves competitive performance. As to probe view 126° , the performance of GEI-DPLCR is at least 11 percents superior to co-clustering for gallery views 36° to 72° , and slightly higher for gallery views 18° , 90° , and 180° .

Furthermore, following [37], we also report the CMS curves for the setting where the gallery view is 90° and probe view ranges from 36° to 126° (except 90°). As shown in Fig. 4. Our method DPLCR with GEI feature (denoted by the dotted red line) always gains the best performance than the-state-of-art methods for various probe views. Therefore, we can see that

TABLE III
PERFORMANCE COMPARISON OF DIFFERENT METHODS FOR CROSS-VIEW GAIT RECOGNITION
WHEN PROBE VIEW IS 126° AND GALLERY VIEW IS FROM 0° TO 180° EXCEPT 126°

Gallery view	0°	18°	36°	54°	72°	90°	108°	144°	162°	180°
Baseline[2]	2	5	13	29	21	15	37	43	2	4
View-rectification[29]	-	-	45	57	60	70	68	-	-	-
FT-SVD[3]	4	6	7	17	16	22	55	76	26	8
GEI-SVD[4]	16	17	21	31	42	53	80	95	49	14
GEI-SVR[5]	22	26	26	42	57	78	98	98	74	19
GEI-CCA[36]	21	18	26	34	54	75	91	97	63	18
GInI-EF[16]	3	3	14	31	20	13	41	52	3	2
GInI-SF[16]	5	6	12	40	23	11	35	46	8	5
Co-clustering[37]	25	29	35	49	60	78	98	98	75	22
GEI-DPLCR	17	30	52	62	71	80	96	94	44	23
GII-DPLCR	18	33	58	68	78	83	98	95	63	27

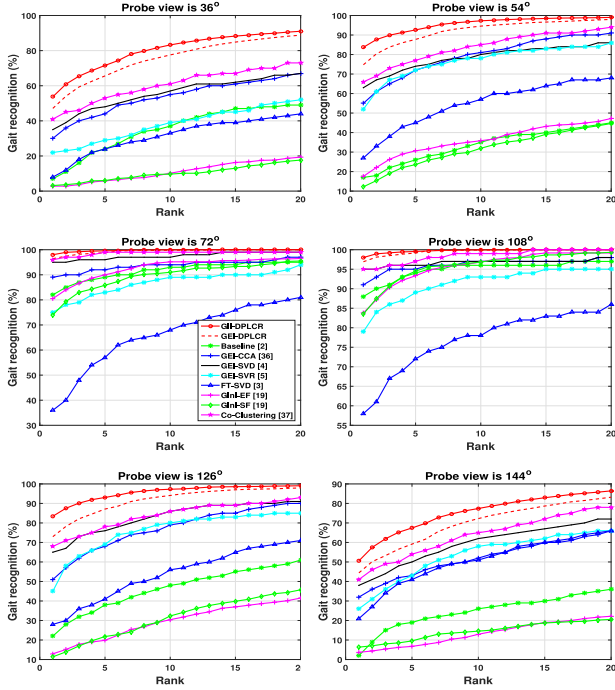


Fig. 4. CMS comparison of different methods when the gallery view is 90° and probe view ranges from 36° to 144° .

the proposed DPLCR could boost the performance of cross-view gait recognition with one single probe view based on GEI feature in most situations. The enhancement of performance by using DPLCR may due to the following.

- 1) By using list-wise constraints, the intraclass similarities are enlarged, and the interclass similarities across distinct views are diminished in the projected subspace with maximum margins, which might strengthen the discriminative ability of the proposed method in dealing with cross-view variations. Meanwhile, the rectification term could further exploit the relative similarities and allow for violations of maximum margins, thus may enhance the generalization ability of the learned projection.
- 2) The kernelization could embed nonlinear mapping, which can better describe the complicated cross-view correlations under large view variations, compared to existing methods.

From Tables I–III and Fig. 4, we can also validate the performance of the proposed GII feature. We can see that GII-DPLCR further boosts the performance of GEI-DPLCR in all situations. For instance, for probe view 90° under gallery views 36° , 54° , and 126° , GII-DPLCR gaits more than 10 percents improvements, compared with GEI-DPLCR. And from Fig. 4, the performance of GII-DPLCR (denoted by solid red line) is obviously higher than GEI-DPLCR. We detail the comparison of GEI-DPLCR and GII-DPLCR in Fig. 5. The gain of performance by using GII feature compared with GEI may due to the fact that GII could partially eliminate the cross-view shape distortion. More specifically, GII first constructs a synthetic common GEI template of a certain view by averaging GEIs from all training individual gaits captured under the corresponding camera view. The GEI template mainly contains view dependent information and has little individual identify information, which is thus less discriminative for cross-view gait recognition. It is then reasonable to subtract this common GEI template from each individual GEI, since during this operation the individual identity information is reserved while the view dependent information contained in GEI template is eliminated. Therefore, GII could be more discriminative and view-independent, compared to GEI.

B. Gait Recognition With Multiple Probe Views

When gaits from multiple probe views are available, we evaluate the performance of DPLCR using the proposed three multiview fusion strategies following the same experimental setting as [37]. Specifically, we select 54° as the galley views under (36° , 72°), (18° , 90°), and (0° , 108°) as probe views, and 126° as the gallery view under (108° , 144°), (90° , 162°), and (72° , 162°) as probe views. Some of the experimental results are shown in Tables IV and V. We compare our method with methods FT-SVD [3], GEI-SVD [4], GEI-SVR [5], and co-clustering [37], which also report multiview to one-view performance. It can be observed that the results of multiview to one view gait recognition are all significantly better than one-to-one view situation. This is due to that more information can be exploited by our combined method when provided multiple probe views. It also can be seen that all the three fusion strategies outperforms the comparison approaches. Especially, for

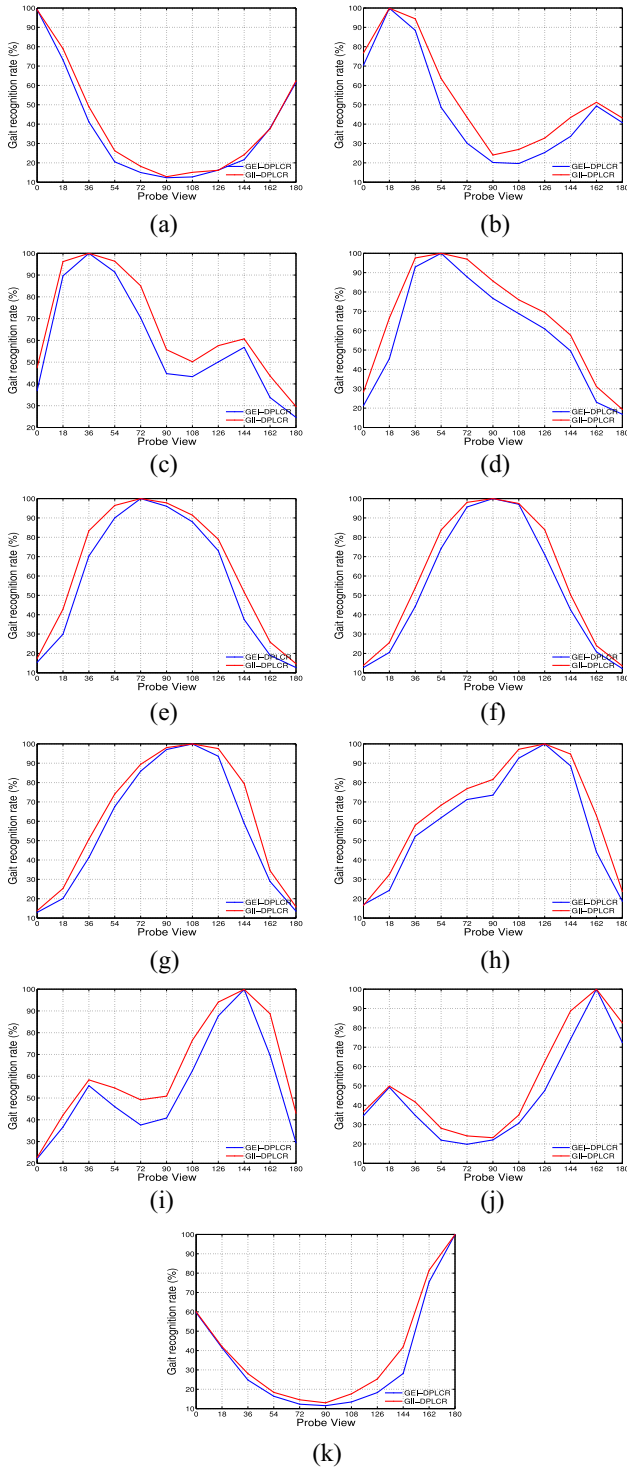


Fig. 5. Comparison of recognition accuracy at rank 1 by using GEI and GII. (a) Gallery view: 0°. (b) Gallery view: 18°. (c) Gallery view: 36°. (d) Gallery view: 54°. (e) Gallery view: 72°. (f) Gallery view: 90°. (g) Gallery view: 108°. (h) Gallery view: 126°. (i) Gallery view: 144°. (j) Gallery view: 162°. (k) Gallery view: 180°.

gallery view 54°, the performance of DPLCR using different fusion strategies is at least 8 percents higher than co-clustering (the best performance in the comparison methods) under probe views 18°, 90°, and 15 percents higher under probe views 0° and 108°. For gallery view 126°, the performance of

TABLE IV
PERFORMANCE COMPARISON OF DIFFERENT METHODS FOR MULTIVIEW TO ONE VIEW GAIT RECOGNITION WHEN GALLERY VIEW IS 54° BY USING GEI FEATURE (S1: FEATURE LEVEL FUSION, S2: SCORE LEVEL FUSION, AND S3: PROPOSED MULTIVIEW DPLCR)

Probe view	36°, 72°	36°	72°
FT-SVD[3]	72	54	56
GEI-SVD[4]	95	81	85
GEI-SVR[5]	99	92	90
Co-clustering[37]	99	92	92
Proposed+GEI (S1)	95	93	91
Proposed+GEI (S2)	97	93	91
Proposed+GEI (S3)	99	93	91
Probe view	18°, 90°	18°	90°
FT-SVD[3]	55	20	30
GEI-SVD[4]	74	34	52
GEI-SVR[5]	80	46	63
Co-clustering[37]	83	53	66
Proposed+GEI (S1)	82	46	77
Proposed+GEI (S2)	83	46	77
Proposed+GEI (S3)	85	46	77
Probe view	0°, 108°	0°	108°
FT-SVD[3]	33	5	25
GEI-SVD[4]	40	12	38
GEI-SVR[5]	54	23	42
Co-clustering[37]	57	26	44
Proposed+GEI (S1)	58	22	69
Proposed+GEI (S2)	68	22	69
Proposed+GEI (S3)	72	22	69

TABLE V
PERFORMANCE COMPARISON OF DIFFERENT METHODS FOR MULTIVIEW TO ONE VIEW GAIT RECOGNITION WHEN GALLERY VIEW IS 54° BY USING GII FEATURE (S1: FEATURE LEVEL FUSION, S2: SCORE LEVEL FUSION, AND S3: PROPOSED MULTIVIEW DPLCR)

Probe view	36°, 72°	36°	72°
FT-SVD[3]	72	54	56
GEI-SVD[4]	95	81	85
GEI-SVR[5]	99	92	90
Co-clustering[37]	99	92	92
Proposed+GII (S1)	99	96	97
Proposed+GII (S2)	99	96	97
Proposed+GII (S3)	99	96	97
Probe view	18°, 90°	18°	90°
FT-SVD[3]	55	20	30
GEI-SVD[4]	74	34	52
GEI-SVR[5]	80	46	63
Co-clustering[37]	83	53	66
Proposed+GII (S1)	91	62	85
Proposed+GII (S2)	91	62	85
Proposed+GII (S3)	92	62	85
Probe view	0°, 108°	0°	108°
FT-SVD[3]	33	5	25
GEI-SVD[4]	40	12	38
GEI-SVR[5]	54	23	42
Co-clustering[37]	57	26	44
Proposed+GII (S1)	72	28	73
Proposed+GII (S2)	78	28	73
Proposed+GII (S3)	80	28	73

DPLCR using different fusion strategies boosts co-clustering (the best performance in the comparison methods) by 2 percents under probe views 90°, 162°, and by 14 percents under probe views 0° and 108°. We also compared the performance

TABLE VI
PERFORMANCE COMPARISON OF DIFFERENT METHODS FOR MULTIVIEW
TO ONE VIEW GAIT RECOGNITION WHEN GALLERY VIEW IS 126° BY
USING GEI FEATURE (S1: FEATUER LEVEL FUSION, S2: SCORE
LEVEL FUSION, AND S3: PROPOSED MULTIVIEW DPLCR)

Probe view	108°, 144°	108°	144°
FT-SVD[3]	86	73	75
GEI-SVD[4]	97	84	92
GEI-SVR[5]	98	96	97
Co-clustering[37]	99	96	97
Proposed+GEI (S1)	97	93	92
Proposed+GEI (S2)	99	93	91
Proposed+GEI (S3)	99	93	91
Probe view	90°, 162°	90°	162°
FT-SVD[3]	63	33	31
GEI-SVD[4]	72	45	50
GEI-SVR[5]	88	65	53
Co-clustering[37]	90	68	59
Proposed+GEI (S1)	85	74	44
Proposed+GEI (S2)	89	74	44
Proposed+GEI (S3)	90	74	44
Probe view	72°, 180°	0°	108°
FT-SVD[3]	32	31	7
GEI-SVD[4]	35	31	6
GEI-SVR[5]	54	45	34
Co-clustering[37]	60	47	34
Proposed+GEI (S1)	60	71	24
Proposed+GEI (S2)	69	71	24
Proposed+GEI (S3)	72	71	24

TABLE VII
PERFORMANCE COMPARISON OF DIFFERENT METHODS FOR MULTIVIEW
TO ONE VIEW GAIT RECOGNITION WHEN GALLERY VIEW IS 126° BY
USING GII FEATURE (S1: FEATUER LEVEL FUSION, S2: SCORE
LEVEL FUSION, AND S3: PROPOSED MULTIVIEW DPLCR)

Probe view	108°, 144°	108°	144°
FT-SVD[3]	86	73	75
GEI-SVD[4]	97	84	92
GEI-SVR[5]	98	96	97
Co-clustering[37]	99	96	97
Proposed+GII (S1)	98	98	95
Proposed+GII (S2)	99	98	95
Proposed+GII (S3)	99	98	95
Probe view	90°, 162°	90°	162°
FT-SVD[3]	63	33	31
GEI-SVD[4]	72	45	50
GEI-SVR[5]	88	65	53
Co-clustering[37]	90	68	59
Proposed+GII (S1)	92	83	62
Proposed+GII (S2)	93	83	62
Proposed+GII (S3)	94	83	62
Probe view	72°, 180°	0°	108°
FT-SVD[3]	32	31	7
GEI-SVD[4]	35	31	6
GEI-SVR[5]	54	45	34
Co-clustering[37]	60	47	34
Proposed+GII (S1)	74	78	22
Proposed+GII (S2)	79	78	27
Proposed+GII (S3)	82	78	27

of different fusion strategies as shown in Tables IV–VII. As can be seen, the proposed Algorithm 2, namely Strategy 3, performs better than the feature level fusion and score level

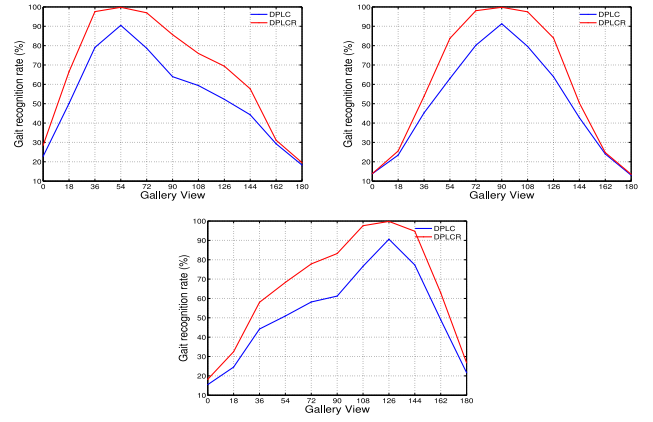


Fig. 6. CMS comparison of DPLC and DPLCR for cross-view gait recognition when probe views are 54°, 90°, and 126°.

fusion. The performance improvement indicates that by co-training the rectification term, or the unknown list-wise similarities, is superior to the straightforward feature or score level fusion.

C. Cross-View Gait Recognition Under Outdoor Environment

We further evaluate the proposed method by the USF gait database [1], which was collected in outdoor environment. Many experiments were designed to investigate the effect of five factors on performance, including surface type, shoe-wear type, weight carried, camera angle, and time. In this paper, we only adopt experiment A for evaluation since it observes view variations for gait-based identification, which is the focus of this paper. In this dataset, two gait sequences were recorded from two cameras L and R, respectively. There are 122 subjects in this dataset in which 22 ones are used for training. We also compare our method with others which also report their results on USF experiment A in Table VIII. The performance of GEI-DPLCR is better than most of the others although 2 percents worse than [37]. The reason is that DPLCR require sufficient large training data than CCA-based method for a reliable projection. In the previous experiments on the CASIA-B dataset, there are total $24 \times 6 \times 2$ (24 subjects, 6 instances, 2 views) samples for training in the one to one situation while only 22×2 (22 subjects, 2 views) on the USF gait database. While, GII-DPLCR gains a 2 percents improvement compared with co-clustering, which indicates the effectiveness of GII feature.

D. Evaluation of the Rectification Term

Experiments are conducted to evaluate the effect of the rectification term with results are shown in Table IX and Fig. 6. It can be observed that the rectification term could significantly boosts the performance of DPLC. The reason may lie on that the rectification term could exploit relative similarity information. From a statistic learning perspective, the rectification term serves as slack variables, which could enhance the generalization ability.

TABLE VIII
PERFORMANCE ON THE USF GAIT DATABASE BY DIFFERENT METHODS

Baseline[1]	MSCT+SST[39]	HMM [40]	pHMM[41]	PEI+LDA [42]	ViDP[43]	Co-clustering[37]	GEI-DPLCR	GII-DPLCR
73	80	80	85	85	83	89	87	91

TABLE IX
PERFORMANCE COMPARISON OF DPLC AND DPLCR FOR CROSS-VIEW
GAIT RECOGNITION WHEN PROBE VIEWS ARE 54°, 90°, AND 126°

Probe view		54°								
Gallery view	0°	18°	36°	72°	90°	108°	126°	144°	162°	180°
DPLC	23	50	79	79	64	60	52	44	29	18
DPLCR	28	67	98	97	86	76	69	58	31	19
Probe view		90°								
Gallery view	0°	18°	36°	54°	72°	108°	126°	144°	162°	180°
DPLC	14	23	45	63	80	80	64	43	24	13
DPLCR	14	26	54	84	98	98	84	50	25	14
Probe view		126°								
Gallery view	0°	18°	36°	54°	72°	90°	108°	144°	162°	180°
DPLC	16	24	52	62	71	74	93	89	44	19
DPLCR	17	33	58	68	77	82	97	95	63	24

VII. CONCLUSION

In this paper, we propose a novel gait feature, i.e., GII, together with a discriminative projection for cross-view gait recognition. GII is introduced to deal with small view variations by using cosine similarity measurement. For large view changes, we project the GIIs onto a common subspace such that the instances belong to the same person are closer to each other. To further improve the performance, a weighted combination model is presented. Experimental results on CASIA gait database B and USF demonstrate the superiority of the proposed method compared with the-state-of-arts in both one-to-one cross-view gait recognition and multi-to-one view gait recognition.

REFERENCES

- [1] S. Sarkar *et al.*, “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.
- [2] 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/IAPR Int. Conf. Pattern Recognit.*, vol. 4, 2006, pp. 441–444.
- [3] Y. Makihara, R. Sagawa, Y. Mukaigawa, T. Echigo, and Y. Yagi, “Gait recognition using a view transformation model in the frequency domain,” in *Proc. Eur. Conf. Comput. Vis.*, Graz, Austria, 2006, pp. 151–163.
- [4] 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. Workshop*, Kyoto, Japan, 2009, pp. 1058–1064.
- [5] 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. Vis. Pattern Recognit.*, San Francisco, CA, USA, 2010, pp. 974–981.
- [6] J.-H. Yoo, M. S. Nixon, and C. J. Harris, “Model-driven statistical analysis of human gait motion,” in *Proc. ICIP*, 2002, pp. 1-285–1-288.
- [7] S. L. Dockstader, M. J. Berg, and A. M. Tekalp, “Stochastic kinematic modeling and feature extraction for gait analysis,” *IEEE Trans. Image Process.*, vol. 12, no. 8, pp. 962–976, Aug. 2003.
- [8] Y. Ran, I. Weiss, Q. Zheng, and L. S. Davis, “Pedestrian detection via periodic motion analysis,” *Int. J. Comput. Vis.*, vol. 71, no. 2, pp. 143–160, Feb. 2007.
- [9] A. Y. Johnson and A. F. Bobick, “A multi-view method for gait recognition using static body parameters,” in *Proc. Int. Conf. Audio Video Based Biometric Person Authentication*, 2001, pp. 301–311.
- [10] D. K. Wagg and M. S. Nixon, “On automated model-based extraction and analysis of gait,” in *Proc. IEEE Int. Conf. Automat. Face Gesture Recogn.*, Seoul, South Korea, 2004, pp. 11–16.
- [11] 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.
- [12] 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, Sep. 2008.
- [13] S. D. Choudhury and T. Tjahjedi, “Robust view-invariant multiscale gait recognition,” *Pattern Recognit.*, vol. 48, no. 3, pp. 798–811, 2015.
- [14] E. Zhang, Y. Zhao, and W. Xiong, “Active energy image plus 2DLPP for gait recognition,” *Signal Process.*, vol. 90, no. 7, pp. 2295–2302, 2010.
- [15] 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.
- [16] P. Arora, M. Hanmandlu, and S. Srivastava, “Gait based authentication using gait information image features,” *Pattern Recognit. Lett.*, vol. 68, pp. 336–342, Dec. 2015.
- [17] 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.
- [18] N. V. Boulgouris and Z. X. Chi, “Gait recognition using radon transform and linear discriminant analysis,” *IEEE Trans. Image Process.*, vol. 16, no. 3, pp. 731–740, Mar. 2007.
- [19] D. Xu, Y. Huang, Z. Zeng, and X. Xu, “Human gait recognition using patch distribution feature and locality-constrained group sparse representation,” *IEEE Trans. Image Process.*, vol. 21, no. 1, pp. 316–326, Jan. 2012.
- [20] R. Bodor, A. Drenner, D. Fehr, O. Masoud, and N. Papanikolopoulos, “View-independent human motion classification using image-based reconstruction,” *Image Vis. Comput.*, vol. 27, no. 8, pp. 1194–1206, 2009.
- [21] G. Zhao, G. Liu, H. Li, and M. Pietikainen, “3D gait recognition using multiple cameras,” in *Proc. IEEE Int. Conf. Automat. Face Gesture Recognit.*, Southampton, U.K., Apr. 2006, pp. 529–534.

- [22] G. Shakhnarovich, L. Lee, and T. Darrell, "Integrated face and gait recognition from multiple views," in *Proc. IEEE Comput. Vis. Pattern Recognit.*, 2001, pp. 1-439-1-446.
- [23] D. López-Fernández, F. J. Madrid-Cuevas, A. Carmona-Poyato, R. Muñoz-Salinas, and R. Medina-Carnicer, "A new approach for multi-view gait recognition on unconstrained paths," *J. Vis. Commun. Image Represent.*, vol. 38, pp. 396-406, Jul. 2016.
- [24] M. Goffredo, R. D. Seely, J. N. Carter, and M. S. Nixon, "Markerless view independent gait analysis with self-camera calibration," in *Proc. IEEE Int. Conf. Automat. Face Gesture Recognit.*, Amsterdam, The Netherlands, 2008, pp. 1-6.
- [25] C. BenAbdelkader, R. G. Cutler, and L. S. Davis, "Gait recognition using image self-similarity," *EURASIP J. Appl. Signal Process.*, vol. 2004, no. 4, pp. 572-585, Jan. 2004.
- [26] A. Kale, A. K. R. Chowdhury, and R. Chellappa, "Towards a view invariant gait recognition algorithm," in *Proc. IEEE Comput. Vis. Pattern Recognit.*, 2003, pp. 1-8.
- [27] F. Jean, R. Bergevin, and A. B. Albu, "Trajectories normalization for viewpoint invariant gait recognition," in *Proc. IEEE/IAPR Int. Conf. Pattern Recognit.*, Tampa, FL, USA, 2008, pp. 1-4.
- [28] J. Han, B. Bhanu, and A. K. Roy-Chowdhury, "A study on view-insensitive gait recognition," in *Proc. IEEE Int. Conf. Image Process.*, Genoa, Italy, Sep. 2005, pp. III-297-III-300.
- [29] 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.
- [30] R. Martín-Félez and T. Xiang, "Gait recognition by ranking," in *Proc. Eur. Conf. Comput. Vis.*, 2012, pp. 328-341.
- [31] 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.
- [32] S. Zheng, J. Zhang, K. Huang, R. He, and T. Tan, "Robust view transformation model for gait recognition," in *Proc. IEEE Int. Conf. Image Process.*, Brussels, Belgium, Sep. 2011, pp. 2073-2076.
- [33] 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, May 2012.
- [34] 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.
- [35] 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, Nov. 2014.
- [36] K. Bashir, T. Xiang, and S. Gong, "Gait representation using flow fields," in *Proc. Brit. Mach. Vis. Conf.*, London, U.K., 2009, pp. 1-11.
- [37] 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.
- [38] J. Chen, Z. Zhang, and Y. Wang, "Relevance metric learning for person re-identification by exploiting global similarities," in *Proc. IEEE/IAPR Int. Conf. Pattern Recognit.*, Stockholm, Sweden, 2014, pp. 1657-1662.
- [39] T. H. W. Lam, R. S. T. Lee, and D. Zhang, "Human gait recognition by the fusion of motion and static spatio-temporal templates," *Pattern Recognit.*, vol. 40, no. 9, pp. 2563-2573, Sep. 2007.
- [40] Q.-J. Zhang and S.-L. Xu, "Gait-based recognition of human using an embedded hidden Markov models," in *Proc. ICIECS*, Wuhan, China, 2009, pp. 1-4.
- [41] D. Ioannidis, D. Tzovaras, I. G. Damousis, S. Argyropoulos, and K. Moustakas, "Gait recognition using compact feature extraction transforms and depth information," *IEEE Trans. Inf. Forensics Security*, vol. 2, no. 3, pp. 623-630, Sep. 2007.
- [42] A. Roy, S. Sural, and J. Mukherjee, "Gait recognition using pose kinematics and pose energy image," *Signal Process.*, vol. 92, no. 3, pp. 780-792, Mar. 2012.
- [43] M. Hu, Y. Wang, Z. Zhang, J. J. Little, and D. Huang, "View-invariant discriminative projection for multi-view gait-based human identification," *IEEE Trans. Inf. Forensics Security*, vol. 8, no. 12, pp. 2034-2045, Dec. 2013.



Zhaoxiang Zhang (S'08-M'09-SM'15) received the B.S. degree in electronic science and technology from the University of Science and Technology of China, Hefei, China, in 2004 and the Ph.D. degree from the National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 2009.

In 2009, he joined the School of Computer Science and Engineering, Beihang University, Beijing, as an Assistant Professor, from 2009 to 2011, an Associate Professor, from 2012 to 2015, and the Vice-Director of the Department of Computer Application Technology, from 2014 to 2015. In 2015, he returned to the Institute of Automation, Chinese Academy of Sciences, as a Full Professor. His current research interests include computer vision, pattern recognition, machine learning, and brain-inspired neural network and brain-inspired learning.

Dr. Zhang is an Associate Editor or a Guest Editor of some internal journals, such as, *Neurocomputing*, *Pattern Recognition Letters*, and *IEEE ACCESS*.



Jiaxin Chen received the Ph.D. degree with the School of Computer Science and Engineering, Beihang University, Beijing, China, in 2017.

He is currently a Post-Doctoral Associate with the Multimedia and Visual Computing Laboratory, New York University Abu Dhabi, Abu Dhabi, UAE. His current research interests include pattern recognition, computer vision, and machine learning.

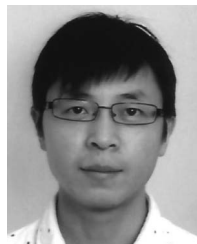


Qiang Wu received the B.Eng. and M.Eng. degrees from the Harbin Institute of Technology, Harbin, China, in 1996 and 1998, respectively, and the Ph.D. degree from the University of Technology Sydney, Ultimo, NSW, Australia, in 2004.

He is currently an Associate Professor and a Core Member of the Global Big Data Technologies Center, University of Technology Sydney. His current research interests include computer vision, image processing, pattern recognition, machine learning, and multimedia processing. His research

outcomes have been published in many premier international conferences, and the major international journals, such as the *IEEE TRANSACTIONS ON IMAGE PROCESSING*, the *IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART B: CYBERNETICS*, and the *IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY*.

Dr. Wu serves as a Reviewer for several journals and conferences, including the *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, *PR*, the *IEEE TRANSACTIONS ON IMAGE PROCESSING*, the *IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY*, the *IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART B: CYBERNETICS*, *Computer Vision and Image Understanding*, *Image and Vision Computing*, *Pattern Recognition Letters*, *Neurocomputing*, and the *EURASIP Journal on Image and Video Processing*.



Ling Shao received the B.Eng. degree in electronic and information engineering from the University of Science and Technology of China, Hefei, China, and the M.Sc. degree in medical image analysis and the Ph.D. (D.Phil.) degree in computer vision under the supervision of M. Brady from the Robotics Research Group, University of Oxford, Oxford, U.K.

He is a Professor with the School of Computing Sciences, University of East Anglia, Norwich, U.K. From 2014 to 2016, he was a Professor with

Northumbria University, Newcastle upon Tyne, U.K., from 2009 to 2014, a Senior Lecturer with the Department of Electronic and Electrical Engineering, University of Sheffield, Sheffield, U.K., and from 2005 to 2009, a Senior Scientist with Philips Research, Amsterdam, The Netherlands. His current research interests include computer vision, deep learning/machine learning, multimedia, and image/video processing.

Dr. Shao is an Associate Editor of the *IEEE TRANSACTIONS ON IMAGE PROCESSING*, the *IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS*, the *IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY*, and several other journals. He is a fellow of the British Computer Society and IET and a Life Member of the ACM.