IET Biometrics

Special Issue: Selected Papers from the International Workshop on Biometrics and Forensics (IWBF2017)

IET Journals
The Institution of Engineering and Technology

On view-invariant gait recognition: a feature selection solution

ISSN 2047-4938
Received on 30th July 2017
Revised 1st February 2018
Accepted on 26th February 2018
E-First on 8th May 2018
doi: 10.1049/iet-bmt.2017.0151
www.ietdl.org

Ning Jia[1] ✉, Victor Sanchez[2], Chang-Tsun Li[3]

[1]*Department of Computer Science, Durham University, Durham, UK*
[2]*Department of Computer Science, University of Warwick, Coventry, UK*
[3]*School of Computing & Mathematics, Charles Stuart University, Albury, Australia*
✉ *E-mail: ning.jia@durham.ac.uk*

Abstract: The authors present an improved feature selection solution for the view-invariant gait recognition problem, based on their previously proposed method called view-invariant feature selector (ViFS), which automatically reconstruct an optimised gallery template from a set of multi-view gallery templates. They improved ViFS by introducing a constraint to make sure that the reconstructed features have the same scale as the original features, thus reducing the number of misclassifications caused by data misalignment. They evaluate the improved ViFS on the CASIA B and OU-ISIR large population datasets by performing a wide range of comparative studies in order to explore and confirm its effectiveness. Evaluation results indicate that the proposed framework is very effective for view-invariant gait recognition tasks.

1 Introduction

The ever-growing demand of reliable human identification systems for law-enforcement, national security and commercial use has promoted the development of biometrics in the past 50 years. Advanced biometric techniques provide solutions to a wide range of challenges with the overarching intention of preventing imposters from accessing protected resources. Various biometric systems are capable of offering solid performance in real-world applications, such as fingerprint, face, and iris recognition.

Despite the fact that biometric traits are intrinsic to humans, they cannot always be easily captured by Close Circuit Television (CCTV) cameras or other types of sensors. Furthermore, it is highly preferable that the biometric analysis and authentication be performed at a distance in a non-invasive and non-obtrusive manner to avoid the *cyber arms race* (Development of deception and anti-deception techniques.). Gait, which is considered to be a behavioural biometric trait, can be measured unobtrusively at a moderate distance, thus it is predominant in remote human tracking and identification tasks. The past two decades have witnessed an important development of gait recognition systems. However, there are still important challenges that confine the practical application of gait analysis, one of which is concerned with the view angle variation between gallery data (data with known identities) and probe data (query data with unknown identities). It is therefore imperative to develop view-invariant gait recognition systems for the sake of its competitiveness in practical applications.

Gait recognition approaches can be broadly classified into two categories: *model based* and *appearance based*. Model-based gait recognition refers to identifying people by modelling their distinctive gait characteristics with underlying mathematical structures [1]. Most model-based methods rely on high-quality gait sequences captured under controlled environments (e.g. indoor environments, a close-distance between subject and camera, multi-view cameras, and depth cameras), thus they are effective at handling occlusions and changes in scale, as well as view-angle changes. However, the restrictions imposed by the underlying sensors used to acquire the data and their low tolerance to low-quality video makes these methods less applicable for outdoor gait recognition.

Appearance-based methods usually adopt gait silhouettes as the feature source to build effective gait templates. The silhouettes are obtained by subtracting the subject profile from the background using gait sequences acquired by video cameras. The classification is usually performed by measuring the pixel-to-pixel distance between gallery and probe templates. A commonly used appearance-based template is the gait energy image (GEI), which is computed as the average of the binary silhouettes from a gait cycle [2]. Experiments on several large gait datasets (over 4000 subjects) suggest that GEI is the most statistically stable and efficient template for gait recognition, while other templates such as chrono-gait image [3] and gait entropy image [4] failed to show such robustness across various datasets [5, 6]. The advantages of appearance-based approaches based on GEIs are their robustness to low-quality video and low-computational complexity. However, GEIs are usually not robust to view angle and scale changes.

View-invariant gait recognition is one of the major challenges in people identification. Many researchers have evaluated view angle transformation techniques, discriminant analysis and manifold learning approaches for cross-view gait recognition. Their proposals are usually based on a common factor; namely to establish a cross-view mapping between gallery and probe templates. However, the effectiveness of many of these proposals is restricted to small view-angle variances. A promising approach to perform view-invariant gait recognition is through multi-view feature learning.

We have previously proposed the view-invariant feature selector (ViFS) [7], which is a linear regression-based feature selector that reconstructs gallery templates from arbitrary view angles, thus minimising the cross-view variance between gallery and probe features. Within the context of multi-view gait recognition, this equals to reducing the intra-class variance. Subspace learning methods, i.e. linear discriminant analysis (LDA), have been applied to ViFS as feature enhancers to reduce the computational cost and improve the recognition accuracy. In this work, we introduce a feature scaling process to ViFS to further improve its performance by making sure that all the templates, original and reconstructed, are regularised before the similarity measurement. The scaling process reduces the noise in the reconstructed templates and thus minimises misclassifications. We test the proposed framework on the CASIA dataset B (CASIA B) and the OU-ISIR large population (OU-ISIR LP) dataset. The average recognition accuracy of our framework over 11 different views exceeds 99%.

This paper is organised as follows: Section 2 reviews the literature on view-invariant gait recognition. Section 3 briefly

explains the formulation of ViFS, as well as other fundamental concepts. Section 4 explains in detail the improvements proposed to ViFS. Section 5 presents the evaluations on the CASIA B and the OU-ISIR large population datasets. Section 6 concludes the paper.

2 Related work

In general, there are two types of view-invariant gait recognition: *cross-view recognition*, where only a single view angle is available in both, the gallery and probe sets (the view angles are different), and *multi-view recognition*, where templates from multiple view angles are available in the gallery set while the probe templates are from a single view angle, or vice versa (This case can be reversed, i.e. multi-view templates available in the probe set, while the gallery templates are from a single view angle.).

Depending on the underlying algorithms used, current view-invariant gait recognition algorithms can be classified into one of three categories: (i) those based on human models, (ii) those based on view-invariant features, and (iii) those based on unitary projections. Methods within the first category are concerned with creating models that represent the human anatomy. When multi-view gait templates can be obtained, or depth information is available, it is possible to reconstruct 3D or 2.5D models, from which arbitrary views of gait sequences can be obtained by projection, and the parameters associated with various body parts can be easily determined. Tang *et al.* [8] propose to construct parametric 3D gait models from three cameras and use partial similarity matching to improve recognition rates. Their method achieves promising results on several major gait datasets. Similarly, Luo *et al.* [9] propose to use 3D gait models and sparse representation-based classification to perform view-invariant classification. Their framework and its performance are very similar to the one prosed by Tang *et al.* [8].

Methods within the second category attempt to obtain view-invariant features from single-view gait silhouette sequences to perform recognition under lateral views, i.e. those views different from the frontal and back views. For example, Kusakunniran *et al.* [10] and Goffredo *et al.* [11] employ view-invariant gait features for cross-view recognition. In [10], the authors propose the gait texture image and apply domain transformation obtained through invariant low-rank textures to obtain common canonical side view gait features (i.e. the walking trajectory is perpendicular to the camera's viewpoint) from other view angles. Despite the good performance of this method, it is difficult to project features from the front or back views to the side view. In [11], the authors propose model-based view-invariant gait features, which use lower limb pose estimation to perform view angle rectification. However, as with other model-based methods, it is difficult to extract model's parameters (height, length of limbs, joint angles etc.) from gait sequences acquired from a distance at low resolutions and with occlusions.

Methods within the third category usually adopt appearance-based features, e.g. GEIs, and learn the mapping relationship of features from two different views. Makihara *et al.* [12] propose the view transformation model (VTM) to effectively project gait features between two different views. However, as with any other singular value decomposition (SVD)-based method, VTM is sensitive to noise in the training dataset and requires a huge amount of memory and high computational power to compute the matrix factorisation if the training set is large. To solve these issues, Kusakunniran *et al.* [13] use truncated SVD and LDA to enhance the performance. In [14], Muramatsu *et al.* further enhance the VTM by matching gait features locally, i.e. by separating the gait into head, torso, thigh and shank regions, in order to avoid over-fitting and reduce the influence of local-feature corruption. However, VTM-based methods still fail to provide satisfying results with large view-angle differences (over 30°) between gallery and probe.

An alternative to feature mapping is to learn a unitary subspace where features from the same subject but at different view angles are clustered together, while features from different subjects but at the same view angle are far from one another [15–17]. After learning such subspace, features from various views can be projected into it for distance matching. Within this context, Hu *et al.* [18] propose a novel unitary liner projection method named ViDP, which enables cross-view gait recognition to be conducted without knowing the query view angle. The recent work by Zhang *et al.* [19] proposes a list-wise constrained discriminative projection framework on a novel gait representation to tackle the view angle variance. Apart from reporting results for cross-view matching, they also report results for the multi-view case, which outperforms other conventional subspace learning methods.

Convolutional meural networks (CNNs) have been recently used to tackle gait recognition challenges. Alotaibi *et al.* apply a full CNN with four convolutional layers and a softmax layer for simple gait recognition tasks, i.e. matching gallery and probe data under the same view angle [20]. Yan *et al.* use a five-layer CNN with three convolutional layers and two fully connected layers for gait recognition. They also introduce a multi-task learning approach, which performs gait recognition, view angle prediction and scene prediction simultaneously. According to their findings, multi-task learning can accelerates the convergence of CNNs in the training process. However, the cross-view recognition performance of their network appears to attain small improvements compared with traditional approaches using principal component analysis (PCA) + LDA. Shiraga *et al.* successfully use a four-layer CNN, consisting of two convolutional layers and two fully connected layers, for large-scale gait recognition on the OU-ISIR large population dataset [21]. Their network has important advantages over other approaches on large-scale datasets when the view angle difference between gallery and probe data is small ($< 30^{\circ}$). They also show that CNN-based methods can significantly reduce the equal error rates and thus improve the gait verification accuracy. Feature maps learned by CNNs have strong discriminant power and thus provide robustness in gait recognition. However, performance of cross-view gait recognition with large view angle variations ($> 54^{\circ}$) is still not ideal. Wu *et al.*'s work in [22] represents the state-of-the-art of CNN-based cross-view gait recognition.

In our previous work [7], we introduce ViFS to reconstruct gallery templates from arbitrary view angles, and help to transfer the cross-view gait recognition problem to the identical-view gait recognition problem. However, despite its very good performance, some important aspects of ViFS require improvements. For example, the selected features occasionally degrade the performance compared with the performance attained when using single-view features. We realise that this is due to the fact that the ViFS does not force the features to be normalised, thus the reconstructed templates do not align with other templates. This misalignment of data inevitably introduces noise in the reconstructed templates and thus leads to misclassifications. Based on these observations, in this work we introduce a feature scaling process to ViFS to make sure that all the templates, original and reconstructed, are regularised before the similarity measurement. This is explained in detail in Section 4.

3 Background information

3.1 Gait recognition pipeline using GEIs

GEI-based gait recognition focuses on both the human body shape in the spatial domain and the movement in the temporal domain. A conventional gait recognition pipeline based on GEIs consists of three steps:

- *Acquisition of the gait signature*: This involves extracting the binary silhouette of the subject from a video sequence. Several well-known techniques have been adopted for this task, including least median of squares [23], Gaussian mixture models [24], and most recently, fully convolutional networks [25].
- *Construction of the GEIs templates*: Fig. 1 shows an example of the binary silhouettes extracted from a video sequence for a complete gait cycle (the two columns on the left hand side) and the corresponding GEI computed as the average of the binary silhouettes (right hand side). By compressing spatial and

Fig. 1 *Example GEI computed by averaging all the binary silhouettes within a gait cycle*

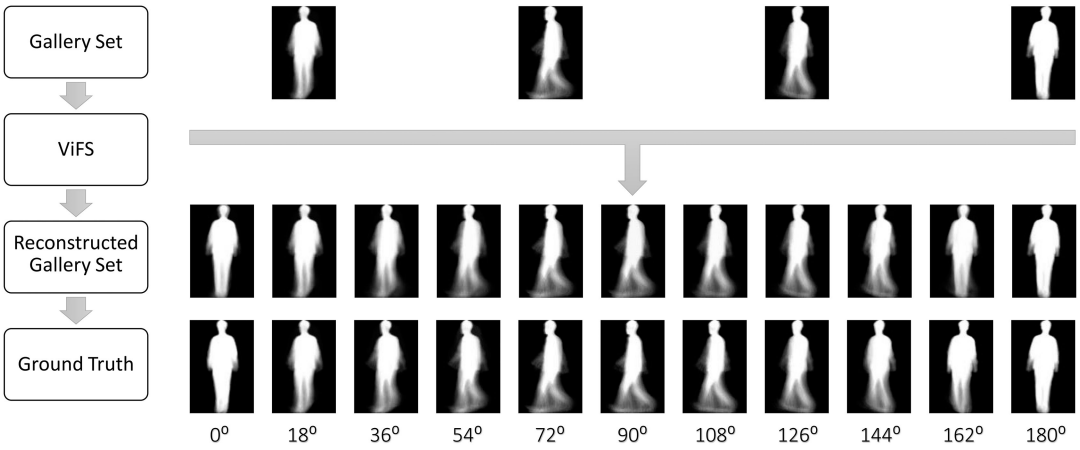

Fig. 2 *Example reconstruction by ViFS of gallery templates for missing view angles. The ground truth shows the gallery templates from all views provided by the CASIA B dataset*

temporal information into one image, a GEI is able to reduce the effect of noise and increase computational efficiency.

- *Similarity measurement:* After modelling gallery and probe data using GEIs templates, the distance between them is measured to find the matching identity.

3.2 View-invariant feature selector

Let us assume that h samples (i.e. GEIs) from h different unknown view angles are available in the gallery set $\mathcal{G} = \{\boldsymbol{x}_i\}_{i=1}^{h}$, as well as one probe sample, $\boldsymbol{y}$, from an unknown view angle in the probe set, $\mathcal{P}$. Due to the view angle difference between gallery and probe samples, the intra-class distance can be larger than the inter-class distance for the same subject, leading to misclassifications. To reduce the negative effects of view angle differences on the classification results, one can minimise the cross-view distance between gallery and probe samples. If the view angles of the gallery and probe samples are unknown, one would like to find a feature vector $\boldsymbol{w} = \{w_i\}_{i=1}^{h}$ that minimises the objective function:

$$f(\boldsymbol{w}) = \| \mathcal{G}\boldsymbol{w}^{\mathrm{T}} - \boldsymbol{y} \|^2 = \| \sum_{i=1}^{h} w_i \boldsymbol{x}_i - \boldsymbol{y} \|^2 . \tag{1}$$

The minimiser $\hat{\boldsymbol{w}}$ of $f(\boldsymbol{w})$ satisfies $\nabla f(\hat{\boldsymbol{w}}) = 0$, leading to $\nabla f(\hat{\boldsymbol{w}}) = 2\mathcal{G}^{\mathrm{T}}(\mathcal{G}\hat{\boldsymbol{w}}^{\mathrm{T}} - \boldsymbol{y})$ [26]. Then $\hat{\boldsymbol{w}}$ can be calculated as follows:

$$\hat{\boldsymbol{w}} = (\mathcal{G}^{\mathrm{T}}\mathcal{G})^{-1}\mathcal{G}^{\mathrm{T}}\boldsymbol{y} . \tag{2}$$

Since the gallery set $\mathcal{G}$ and its covariance matrix $\mathcal{G}^{\mathrm{T}}\mathcal{G}$ are highly unlikely to be upper-triangular, we cannot solve (2) directly. Instead, we use $\boldsymbol{QR}$-factorisation, i.e. $\mathcal{G} = \boldsymbol{QR}$, to generate an orthogonal matrix $\boldsymbol{Q}$ and upper-triangular matrix $\boldsymbol{R}$ from $\mathcal{G}$. Thus (2) can be formulated as

$$\begin{aligned} \hat{\boldsymbol{w}} &= (\mathcal{G}^{\mathrm{T}}\mathcal{G})^{-1}\mathcal{G}^{\mathrm{T}}\boldsymbol{y} \\ &= ((\boldsymbol{QR})^{\mathrm{T}}(\boldsymbol{QR}))^{-1}(\boldsymbol{QR})^{\mathrm{T}}\boldsymbol{y} \\ &= \boldsymbol{R}^{-1}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{y} . \end{aligned} \tag{3}$$

We can obtain $\hat{\boldsymbol{w}}$ by solving $\boldsymbol{R}\hat{\boldsymbol{w}} = \boldsymbol{Q}^{\mathrm{T}}\boldsymbol{y}$ with back substitution. We call minimiser $\hat{\boldsymbol{w}}$ the ViFS, as it selects features from the multi-view gallery samples to reconstruct an optimal template $\hat{\mathcal{G}} = \mathcal{G}\hat{\boldsymbol{w}}^{\mathrm{T}}$ that accurately matches probe sample $\boldsymbol{y}$. In Fig. 2, we present a set of examples to demonstrate the effectiveness of ViFS for feature reconstruction. We take four samples of the same subject from the gallery set at view angles $\{18^{\circ}, 72^{\circ}, 126^{\circ}, 180^{\circ}\}$. We denote these set of samples by $\mathcal{G}$. We train ViFS so to reconstruct gallery samples representing 11 different view angles. For example, for the reconstruction of the 0° gallery sample, we use the four gallery samples and one probe sample from 0° to generate the ViFS for 0°, denoted as $\hat{\boldsymbol{w}}_0$, and obtain the reconstructed template $\hat{\mathcal{G}}_0 = \mathcal{G}\hat{\boldsymbol{w}}_0^{\mathrm{T}}$. The reconstructed gallery samples in the third row of Fig. 2 are visually similar to the ground truth samples in the fourth row of Fig. 2, suggesting that ViFS achieves an accurate view-transformation on gallery samples.

3.3 2D PCA: Yang *et al.* [27] propose the 2D extension of PCA. Consider the training set $\{\boldsymbol{I}_i | i = 1, \ldots, n\}$, where $\boldsymbol{I}_i$ is a single sample (e.g. a GEI) in 2D form with size $d_r \times d_c$, and n is the total number of samples. The image covariance matrix, $\boldsymbol{C}$, is then calculated as

$$\boldsymbol{C} = \frac{1}{n}\sum_{i=1}^{n} (\boldsymbol{I}_i - \bar{\boldsymbol{I}})^{\mathrm{T}}(\boldsymbol{I}_i - \bar{\boldsymbol{I}}), \tag{4}$$

where $\bar{\boldsymbol{I}} = (1/n)\sum_{i=1}^{n} \boldsymbol{I}_i$ is the mean value of all training samples. By performing the eigen-decomposition of $\boldsymbol{C}$, we can obtain the 2D PCA projection basis $V_{\mathrm{pca}} = \{\boldsymbol{v}_i | i = 1, \ldots, p\}$, as the p orthonormal eigenvectors corresponding to the p largest eigenvalues. Compared with the canonical PCA, 2D PCA is much more computationally efficient. For example, for GEIs of size 128×88, the covariance matrix of vectorised samples using canonical PCA has a complexity $O(2^d)$, $d = d_r \times d_c = 11264$; while the complexity of calculating the image covariance matrix, $\boldsymbol{C}$, is only $O(2^{d_r}), d_r = 128$.

In this work, we use $\boldsymbol{I}_i$ to represent sample i in 2D form. In the following, all samples are assumed to be vectorised into features vectors instead of being in 2D form. Therefore, we denote feature vector $\boldsymbol{i}$ by $\boldsymbol{x}_i$ and $\boldsymbol{y}_i$ for gallery and probe, respectively.

4 Proposed improved ViFS

In this section, we present the methodology followed to improve ViFS. Specially, we introduce a constraint in the objective function to achieve data normalisation by performing feature scaling in the minimiser $\hat{\boldsymbol{w}}$.

Table 1 Cross-view matching accuracy (%) using PCA + LDA on the CASIA B dataset

GP	0°, %	18°, %	36°, %	54°, %	72°, %	90°, %	108°, %	126°, %	144°, %	162°, %	180°, %
0°	98.00	92.00	78.00	42.00	24.00	24.00	26.00	26.00	32.00	62.00	80.00
18°	94.00	98.00	96.00	84.00	54.00	42.00	40.00	62.00	66.00	84.00	80.00
36°	82.00	98.00	98.00	98.00	96.00	78.00	72.00	88.00	86.00	74.00	56.00
54°	50.00	76.00	96.00	98.00	98.00	94.00	94.00	90.00	86.00	52.00	40.00
72°	26.00	36.00	82.00	100.00	100.00	100.00	98.00	88.00	58.00	38.00	24.00
90°	20.00	24.00	48.00	88.00	100.00	100.00	100.00	90.00	56.00	28.00	18.00
108°	20.00	28.00	64.00	92.00	100.00	100.00	100.00	98.00	92.00	40.00	28.00
126°	36.00	54.00	80.00	92.00	96.00	96.00	100.00	100.00	98.00	82.00	52.00
144°	42.00	58.00	80.00	78.00	70.00	74.00	82.00	98.00	98.00	92.00	68.00
162°	72.00	80.00	78.00	54.00	48.00	42.00	38.00	66.00	88.00	100.00	92.00
180°	86.00	70.00	56.00	34.00	14.00	22.00	18.00	30.00	64.00	94.00	100.00

4.1 Feature scaling

The reconstructed feature template of a specific view, v, is denoted as $\hat{\mathscr{G}}_v = \mathscr{G}_v \hat{\boldsymbol{w}}_v^{\mathrm{T}}$. As presented in Section 3, the ***QR***-factorisation solution does not guarantee that the $l-1$ norm of $\hat{\boldsymbol{w}}$ is equal to 1, thus it is not guaranteed that $\hat{\mathscr{G}}_v$ and $\mathscr{G}_v$ have an identical feature scale, i.e. $\max(\mathscr{G}_v) \neq \max(\hat{\mathscr{G}}_v)$ (Generally, the sparsity of the feature matrix results in $\min(\mathscr{G}_v) = \min(\hat{\mathscr{G}}_v) = 0$.). Therefore, we reformulate the object function of ViFS as

$$\arg\min_{\boldsymbol{w}} \| \mathscr{G}\boldsymbol{w}^{\mathrm{T}} - \boldsymbol{y} \|^2 \quad \text{s.t.} \| \boldsymbol{w} \|_1 = 1. \tag{5}$$

4.2 Feature enhancement using subspace learning

In order to further enhance the extracted features and increase the inter-class variance, we apply subspace learning. Since subspace learning methods are designed to project the input features into another space with lower dimensionality, the redundant information is removed and the discriminant features are preserved. Furthermore, since they are linear transformations, the computational cost and processing times are very low. In our previous work, we have shown that when used as features enhancers, locality preserving projection and LDA attain a nearly identical performance. Hence, in this paper, we only implement LDA for its low computational cost and popularity. Before applying LDA, we reduce the dimensionality of the data and make sure that the matrices are non-singular by applying PCA.

Let us denote the 3D matrix containing $n_{\mathscr{T}}$ training samples by $\mathscr{T} = \{\mathscr{T}_i\}_{i=1}^{n_{\mathscr{T}}}$, with size $d_r \times d_c \times n_{\mathscr{T}}$. The eigenvectors $V_{\mathrm{pca}} = \{\boldsymbol{v}_i\}_{i=1}^{n_{\mathscr{T}}}$, as well as the corresponding eigenvalues $\lambda_{\mathrm{pca}} = \{\lambda_i\}_{i=1}^{n_{\mathscr{T}}}$, are obtained by eigen-decomposition of the covariance matrix. We select the first p eigenvectors according to:

$$\frac{\sum_{i=1}^{p} \lambda_i}{\sum_{i=1}^{n_{\mathscr{T}}} \lambda_i} > 0.99. \tag{6}$$

Thus, we obtain $\boldsymbol{V}_{\mathrm{pca}} = \{\boldsymbol{v}_i\}_{i=1}^{p}$, a $d_c \times p$ subspace projection matrix. The subspace projection is then $\mathfrak{T}_i = \mathscr{T}_i \boldsymbol{V}_{\mathrm{pca}}$, which results in matrix $\mathfrak{T} = \{\mathfrak{T}_i\}_{i=1}^{p}$. We reshape the 3D matrix $\mathfrak{T}$ to 2D form with dimensions $d_{\mathrm{pca}} = d_r \times p$ and n samples. We then use $\mathfrak{T}$ and the corresponding class labels to train the LDA projection matrix, $\boldsymbol{V}_{\mathrm{lda}}$.

Let us assume that there are h views in gallery set $\mathscr{G}$, and $n_{\mathscr{G}}$ samples in total. After obtaining the ViFS, $\hat{\boldsymbol{w}}$, the subspace projection matrices, $\boldsymbol{V}_{\mathrm{pca}}$ and $\boldsymbol{V}_{\mathrm{lda}}$, and the reconstructed gallery set, $\hat{\mathscr{G}} = \mathscr{G}\hat{\boldsymbol{w}}^{\mathrm{T}}$, we project $\hat{\mathscr{G}}$ onto the subspace matrices to obtain an enhanced gallery feature set:

$$\begin{aligned} \mathfrak{G}_{\mathrm{pca}} &= \left\{\hat{\mathscr{G}}_i \boldsymbol{V}_{\mathrm{pca}}\right\}_{i=1}^{n_{\mathscr{G}}}, \quad \text{reshape } \mathfrak{G}_{\mathrm{pca}} \text{ to } 2D, \\ \mathfrak{G}_{\mathrm{lda}} &= \boldsymbol{V}_{\mathrm{lda}}^{\mathrm{T}} \mathfrak{G}_{\mathrm{pca}}, \end{aligned} \tag{7}$$

Following the same procedure, we also obtain the enhanced probe set $\mathfrak{P}_{\mathrm{lda}}$. For simplicity, we use $\mathfrak{G}$ and $\mathfrak{P}$ to represent the enhanced gallery and probe feature sets, respectively, in the formulation of the similarity measurement.

4.3 Similarity measurement

We use the Euclidean distance to obtain matching scores between gallery and probe. The Euclidean distance between gallery feature set $\mathfrak{G}$ and probe feature set $\mathfrak{P}$ is calculated as

$$D(\mathfrak{G}_i, \mathfrak{P}_l) = \| \mathfrak{G}_i - \mathfrak{P}_l \|, \quad i = 1, \ldots, c, \tag{8}$$

where c is the number of classes, and l denotes the unknown probe data. If $D(\mathfrak{G}_k, \mathfrak{P}_l) = \min_{i=1}^{c} D(\mathfrak{G}_i, \mathfrak{P}_l)$, the probe feature vector is assigned to the same class label k of the gallery feature.

5 Performance evaluation

In this section, we validate the performance of the improved ViFS on two datasets: the CASIA Gait B and OU-ISIR Large Population Datasets. As part of these evaluations, we also analyse the trade-off between accuracy and speed of different feature enhancers when used in conjunction with the improved ViFS.

5.1 Experiments on the CASIA B dataset

The CASIA B dataset is a multi-view gait dataset that contains 124 subjects in total [28]. The size of each silhouette image is normalised to 128×88; one video sequence produces a single GEI. Since this work focuses on studying the performance of the improved ViFS across different view angles, we only employ those sequences that are not affected by changes in clothes or carrying objects. The sequences of the first 74 subjects are used for training, and the other 50 subjects are used for testing. In the testing set, each subject has six sequences; the first four sequences are regarded as gallery sequences, and the remaining two sequences as probe sequences.

We first present the performance of PCA + LDA for view-invariant gait recognition on the CASIA B dataset. This performance can be used as a baseline to assess the effectiveness of ViFS. Table 1 presents the cross-view matching results when using PCA + LDA. It is obvious that the diagonal values in the table are the highest among each row. As expected, the accuracy drops dramatically when the view angle difference is equal or above to 36°. The same pattern can be identified in Table 2, which corresponds to the case of 2D PCA + LDA.

Table 2 Cross-view matching accuracy (%) using 2D PCA + LDA on the CASIA B dataset

GP	0°, %	18°, %	36°, %	54°, %	72°, %	90°, %	108°, %	126°, %	144°, %	162°, %	180°, %
0°	98.00	86.00	72.00	36.00	18.00	22.00	22.00	24.00	32.00	60.00	74.00
18°	94.00	98.00	94.00	72.00	34.00	22.00	34.00	48.00	54.00	74.00	74.00
36°	76.00	98.00	98.00	96.00	78.00	58.00	52.00	72.00	76.00	64.00	52.00
54°	46.00	64.00	96.00	98.00	94.00	94.00	82.00	78.00	70.00	54.00	40.00
72°	30.00	38.00	72.00	86.00	100.00	100.00	98.00	80.00	58.00	32.00	24.00
90°	24.00	28.00	40.00	78.00	100.00	100.00	100.00	80.00	48.00	34.00	18.00
108°	26.00	32.00	48.00	80.00	100.00	100.00	100.00	98.00	78.00	38.00	24.00
126°	30.00	68.00	80.00	82.00	96.00	94.00	96.00	98.00	98.00	72.00	42.00
144°	38.00	64.00	72.00	70.00	72.00	70.00	80.00	98.00	96.00	96.00	60.00
162°	70.00	74.00	68.00	48.00	42.00	28.00	40.00	66.00	80.00	100.00	94.00
180°	82.00	64.00	60.00	28.00	20.00	20.00	18.00	32.00	46.00	92.00	100.00

Table 3 Recognition accuracy (%) of ViFS. Feature enhancer: PCA + LDA. Views available in the gallery set: 72° and 180°

Gallery	72°, 180°											Average, %
Probe	0°, %	18°, %	36°, %	54°, %	72°, %	90°, %	108°, %	126°, %	144°, %	162°, %	180°, %	
Baseline_1D	53.09	72.73	84.18	79.45	68.18	61.09	69.27	80.55	76.36	68.91	53.45	69.75
Baseline_2D	49.45	63.45	74.55	74.18	65.27	59.09	65.82	77.82%	74.18	64.55	51.09	65.40
ViFS1_2	88.00	90.00	96.00	100.00	100.00	100.00	100.00	98.00	92.00	100.00	100.00	96.73
OViFS1_2	**90.00**	90.00	**98.00**	100.00	100.00	100.00	100.00	98.00	92.00	100.00	100.00	**97.09**

The bold values indicates the values that outperform the previous works.

Table 4 Recognition accuracy (%) of ViFS. Feature enhancer: PCA + LDA. Views available in the gallery set: 18° and 108°

Gallery	18°, 108°											Average, %
probe	0°, %	18°, %	36°, %	54°, %	72°, %	90°, %	108°, %	126°, %	144°, %	162°, %	180°, %	
ViFS1_2	92.00	98.00	98.00	100.00	98.00	98.00	100.00	100.00	96.00	90.00	70.00	94.55
OViFS1_2	92.00	98.00	98.00	100.00	**100.00**	98.00	100.00	100.00	96.00	90.00	70.00	**94.73**

The bold values indicates the values that outperform the previous works.

5.1.1 Evaluation of feature scaling: We examine the effect of feature scaling on the recognition accuracy. The feature enhancer is PCA + LDA, and the eigenvalue ratio of PCA is set to 99%. The feature dimensions are reduced from 11,264 (128×88) to 207. Since ViFS is designed to be applied to a multi-view gallery set, the combination of gallery data from different views is verified and analysed. Let us denote by ViFS1_2 the case of the original ViFS with PCA + LDA with data from two different view angles in the gallery set (Here the 1 refers to 1D PCA, and two refers to the two views available in gallery set.). Let us denote by OViFS1_2, the case of the improved ViFS with PCA + LDA with data from two different view angles in the gallery set. As shown in our previous work [7], view angles widely spread usually result in higher matching accuracy. Thus, we select two views that differ from each other as much as possible. Specifically, we evaluate view angles {72°180°} and {18°108°} in the gallery set. In other words, we make sure that we have gallery data captured from one frontal-view and one lateral-view. Tables 3 and 4 tabulate the results, along with the baseline performance. Baseline_1D refers to PCA + LDA, while Baseline_2D refers to 2D PCA + LDA. The matching accuracy values for Baseline_1D and Baseline_2D are the average of each corresponding row in Tables 1 and 2, respectively. When gallery data from two views are available, OViFS outperforms the baseline framework by ∼28%, on average. In Table 3, note that when the probe data are at 0° or 36°, OViFS outperforms the original ViFS by 2%. The underlying reason for this improvement is that the minimiser of the original ViFS for 0° is $\hat{\boldsymbol{w}}_0 = \{0.49, 0.47\}$, and $\| \hat{\boldsymbol{w}}_0 \| = 0.96$, thus the reconstructed gallery features do not align with the probe features. By normalising $\hat{\boldsymbol{w}}_0$, as in (5), $\hat{\boldsymbol{w}}_0 = \{0.51, 0.49\}$ with $\| \hat{\boldsymbol{w}}_0 \| = 1.00$. The improvement attained for the probe data at 36° can be explained in the same way. These results suggest that the proposed feature scaling method works effectively when the original ViFS minimiser fails to retain the same feature scale.

We also observe identical improvements when gallery data from three view angles are available. Let us denote this case by ViFS1_3 for the original ViFS and by OViFS1_3 for the improved ViFS. Table 5 tabulates results for gallery view angles {72°, 126°, 180°}. Note that the feature scaling method manages to improve the recognition accuracy for the 0° view angle from 84.00 to 88.00%. For this view angle, the minimiser of the original ViFS is $\hat{\boldsymbol{w}}_0 = \{0.23, 0.45, 0.29\}$, and $\| \hat{\boldsymbol{w}}_0 \| = 0.97$, which represents a situation similar to the one explained before. It is interesting to note the results for view angles 0°, 18° and 144° in Tables 3 and 5. It is well known that the frontal and back view GEIs have greater variation with other GEIs. The additional gallery data, i.e. the 126° view angle, strengthens the reconstructed features for other views (except for the front view – 0°), thus improving the overall performance.

Table 6 tabulates the highest accuracy that OViFS1_3 can achieve with gallery data from view angles {18°, 108°, 180°}. The average accuracy over all 11 views reaches 99.27%, with very good results for views angles 0°, 144° and 162°.

As expected, the performance of our improved ViFS increases as the number of different view angles are available in the gallery set, as depicted in Fig. 3. Note that OViFS1_2 achieves the poorest performance, especially for the view angles that are the most different from the two available view angles in the gallery set. OViFS1_3 achieves a high accuracy across all views, while the performance of OViFS1_4, OViFS_5 and OViFS_6 is very similar and the best. Therefore, one can conclude that if gallery data contains samples from more than three different view angles that are widely spread, our improved ViFS has a very good

Table 5 Recognition accuracy (of ViFS. Feature enhancer: PCA + LDA.

Gallery	72°, 126°, 180°											Average, %
probe	0°, %	18°, %	36°, %	54°, %	72°, %	90°, %	108°, %	126°, %	144°, %	162°, %	180°, %	
ViFS1_3	84.00	96.00	98.00	98.00	100.00	100.00	100.00	98.00	96.00	100.00	100.00	97.27
OViFS1_3	**88.00**	96.00	98.00	98.00	100.00	100.00	100.00	98.00	96.00	100.00	100.00	**97.64**

The bold values indicates the values that outperform the previous works.

Views available in the gallery set: 72°, 126° and 180°.

Table 6 Recognition accuracy (%) of ViFS. Feature enhancer: PCA + LDA.

Gallery	18°, 108° and 180°											Average, %
probe	0°, %	18°, %	36°, %	54°, %	72°, %	90°, %	108°, %	126°, %	144°, %	162°, %	180°, %	
ViFS1_3	96.00	98.00	98.00	100.00	98.00	98.00	100.00	100.00	98.00	98.00	100.00	98.55
OViFS1_3	**100.00**	98.00	98.00	100.00	98.00	98.00	100.00	100.00	**100.00**	**100.00**	100.00	**99.27**

The bold values indicates the values that outperform the previous works.

Views available in the gallery set: 18°, 108° and 180°

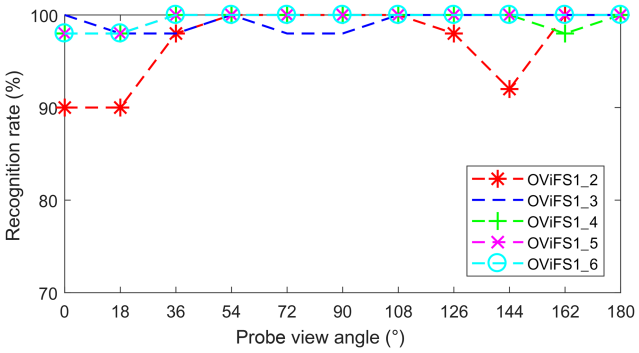

Fig. 3 *Recognition rates of the proposed improved ViFS when data from various view angles are available in the gallery set*

Table 7 Recognition accuracy (%) of OViFS for different values of e_ratio. Feature enhancer: PCA + LDA. No. views available: 3 (18°, 108° and 180°)

Gallery	18°, 108°, 180°											Average, %
probe	0°, %	18°, %	36°, %	54°, %	72°, %	90°, %	108°, %	126°, %	144°, %	162°, %	180°, %	
e_ratio = 0.8	32.00	88.00	68.00	64.00	42.00	78.00	92.00	52.00	52.00	54.00	90.00	64.73
e_ratio = 0.9	70.00	96.00	94.00	84.00	80.00	96.00	96.00	90.00	94.00	90.00	98.00	89.82
e_ratio = 0.95	88.00	98.00	96.00	98.00	94.00	96.00	98.00	94.00%	96.00	98.00	100.00	96.00
e_ratio = 0.99	100.00	98.00	98.00	100.00	98.00	98.00	100.00	100.00	100.00	100.00	100.00	**99.27**

The bold values indicates the values that outperform the previous works.

performance across all views, i.e. it achieves robust view-invariant gait recognition.

5.1.2 Evaluation of different eigenvector ratios: We evaluate the effect of using different eigenvector ratios for the feature enhancers on the improved ViFS. This evaluation assumes that gallery data from three views angles are available: {18°, 108°, 180°}. For the case of PCA + LDA, denoted by OViFS1_3, the evaluated eigenvector ratios, e_ratio, are $\{0.8, 0.9, 0.95, 0.99\}$. Table 7 tabulates these results. As e_ratio increases, the performance improves significantly, indicating that a sufficient number of eigenvectors are required to preserve useful information from the original feature space. For the case of 2D PCA + LDA, denoted by OViFS2_3, the evaluated eigenvector ratios are e_ratio $\in \{0.7, 0.8, 0.85, 0.9, 0.95\}$. Table 8 tabulates these results. Differently from the results presented in Table 7, the average performance of OViFS2_3 has very small variance for different values of e_ratio. Note that when e_ratio is set to 0.95 or higher, the performance drops, indicating that the additional eigenvectors might introduce redundant information to the feature subspace, thus decreasing the performance. One can conclude that OViFS2_3 has a higher tolerance to this parameter change, thus providing a more stable performance than OViFS1_3. However, a well-adjusted OViFS1_3 can achieve higher accuracy, which may be suitable for tasks under controlled environments. Another benefit of using OViFS2_3 is the low computational cost. We evaluate both OViFS1_3 and OViFS2_3 on a laptop with Intel i7–6820HK and DDR4 16 GB memory. The average training time for OViFS1_3 is 51.19 s, while OViFS2_3 takes 1.57 s, ∼32 times faster than OViFS1_3. Therefore, PCA + LDA is suitable for high accuracy, while 2D PCA + LDA is suitable for real-time processing tasks.

5.1.3 Evaluation of different dataset partitions: We evaluate the influence of dataset partition on the proposed framework using OViFS1_3 with view angles {18°, 108°, 180°}. It is well known that for machine learning approaches, including the adopted subspace learning method, the amount of available training data has a huge impact on the performance. Furthermore, different dataset partition strategies are adopted by the state-of-the-art view-invariant methods, e.g. Wu *et al.* [22] evaluate the performance of CNNs with 74 subjects for training, and 50 for testing (74–50, hereafter), and with 24 subjects for training, and 100 for testing (24–100, hereafter). Therefore in this section, we evaluate OViFS using different dataset partitions.

Table 9 tabulates results for three cases: Training # 24, Training # 34 and Training # 44, where the number indicates the number of subjects used for training the framework. Note that when we reduce the number of training samples from 74 to 24, the average

Table 8 Recognition accuracy (%) of OViFS for different values of *e*_ratio. Feature enhancer: 2D PCA + LDA. No. of views available: 3 (18°, 108° and 180°)

Gallery	18°, 108°, 180°											Average, %
probe	0°, %	18°, %	36°, %	54°, %	72°, %	90°, %	108°, %	126°, %	144°, %	162°, %	180°, %	
*e*_ratio = 0.7	96.00	98.00	98.00	98.00	98.00	100.00	100.00	100.00	98.00	98.00	100.00	98.55
*e*_ratio = 0.8	98.00	98.00	96.00	98.00	98.00	100.00	100.00	100.00	98.00	98.00	100.00	98.55
*e*_ratio = 0.85	98.00	98.00	96.00	98.00	98.00	100.00	100.00	100.00	98.00	98.00	100.00	98.55
*e*_ratio = 0.9	98.00	98.00	98.00	96.00	100.00	100.00	100.00	100.00	98.00	100.00	100.00	**98.91**
*e*_ratio = 0.95	94.00	98.00	96.00	94.00	96.00	100.00	100.00	100.00	98.00	98.00	100.00	97.64

The bold values indicates the values that outperform the previous works.

Table 9 Recognition accuracy (%) of OViFS with different dataset partition strategies. Feature enhancer: PCA + LDA. No. of views available: 3 (18°, 108° and 180°)

Gallery	72°, 126°, 180°											Average, %
probe	0°, %	18°, %	36°, %	54°, %	72°, %	90°, %	108°, %	126°, %	144°, %	162°, %	180°, %	
Training # 24	88.00	97.00	98.00	96.00	77.00	97.00	99.00	96.00	87.00	93.00	99.00	93.36
Training # 34	97.78	98.89	100.00	94.44	92.22	98.89	98.89	100.00	92.22	94.44	98.89	96.97
Training # 44	96.25	100.00	100.00	97.50	95.00	98.75	98.75	98.75	91.25	98.75	98.75	97.61

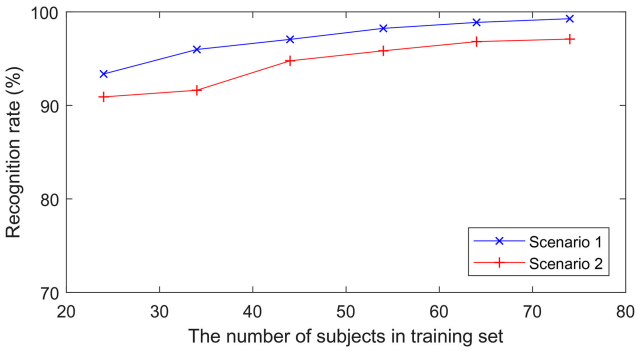

Fig. 4 *Recognition rate of OViFS1_3. Scenario 1 corresponds to gallery view angles* {72°, 126°, 180°}. *Scenario 2 corresponds to gallery view angles* {18°, 108°, 180°}

Table 10 Recognition accuracy (%) of OViFS and Tang *et al.*'s methods [8]

Method	18°, %	36°, %	54°, %	72°, %	90°, %	108°, %	126°, %	144°, %	162°, %	Average, %
Tang_9	94.00	98.00	99.00	98.00	99.00	98.00	98.00	98.00	93.00	97.30
Tang_4	91.00	98.00	92.00	98.00	94.00	98.00	93.00	98.00	90.00	94.70
OViFS1_3	98.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	**99.78**
OViFS2_3	98.00	98.00	100.00	100.00	100.00	100.00	98.00	98.00	100.00	**99.11**
OViFS1_2	92.00	98.00	98.00	100.00	100.00	100.00	98.00	100.00	100.00	**98.44**
OViFS2_2	92.00	98.00	98.00	100.00	100.00	100.00	98.00	100.00	100.00	**98.44**
OViFS1_4	99.00	100.00	94.00	99.00	93.00	89.00	98.00	97.00	94.00	**95.89**

The bold values indicates the values that outperform the previous works.

accuracy drops from 99.27% (see Table 6) to 93.36%, which is mainly caused by the poor recognition accuracy on probe data at 72°. However, as the number of training samples increases to 34, the accuracy on the 72° probe data increases to over 90%. One can then conclude that the poor accuracy of Training # 24 on the 72° view angle is due to the lack of sufficient good quality training data for this view angle, i.e. noisy data may be present causing overfitting. However, if the training set is not badly affected by noise, even with a small amount of training samples, the proposed framework can still attain a reliable performance, as the accuracy for other view angles is relatively high.

Fig. 4 shows the variation on the average performance of OViFS1_3 across all views when the number of training samples varies from 24 to 74. In general, our proposed framework is robust to training set size variation, and as expected, as the number of training samples increase, the recognition rate increases.

5.1.4 Comparison with the state of the art: Finally, we compare our improved ViFS against the state-of-the-art multi-view gait recognition approaches. Table 10 tabulates the recognition accuracy of our improved ViFS and the recently proposed methods by Tang *et al.* [8]. Tang_9 refers to the experiment that uses nine training views from 18° to 162°. Tang_4 refers to the experiment that uses four training views, {36°, 72°, 108°, 144°}. It is important to note that no results for the 0° and 180° view angles are reported in their work. We use four different settings for comparison:

OViFS1_3: PCA + LDA, *e*_ratio = 0.99, available views: 18°, 72° and 162°.

OViFS1_2: PCA + LDA, *e*_ratio = 0.99, available views: 72° and 162°.

OViFS2_3: 2D PCA + LDA, *e*_ratio = 0.9, available views: 18°, 72° and 162°.

OViFS2_2: 2DPCA + LDA, *e*_ratio = 0.9, available views: 18° and 108°.

Table 11 Recognition accuracy (%) of various methods for 54° gallery data. Two view angles are available in the probe set

Gallery	54°			Average, %
probe	36°, 72°, %	18°, 90°, %	0°, 108°, %	
FT-SVD	72.00	55.00	33.00	53.33
GEI-SVD	95.00	74.00	40.00	69.67
GEI-SVR	**99.00**	80.00	54.00	47.67
co-clustering	**99.00**	83.00	57.00	79.67
Zhang *et al.* (1)	95.00	82.00	58.00	78.33
Zhang *et al.* (2)	97.00	83.00	68.00	82.67
Zhang *et al.* (3)	**99.00**	85.00	72.00	85.33
OViFS1_2	97.00	**93.00**	**74.00**	**88.00**

The bold values indicates the values that outperform the previous works.

Table 12 Recognition accuracy (%) of various methods for 126° gallery data. Two view angles are available in the probe set

Gallery	126%			Average, %
probe	108%, 144%, %	90%, 162%, %	72%, 180%, %	
FT-SVD	86.00	63.00	32.00	60.33
GEI-SVD	97.00	72.00	35.00	68.00
GEI-SVR	98.00	88.00	54.00	80.00
Co-clustering	**99.00**	90.00	60.00	83.00
Zhang *et al.* (1)	97.00	85.00	60.00	80.67
Zhang *et al.* (2)	**99.00**	89.00	69.00	85.67
Zhang *et al.* (3)	**99.00**	90.00	72.00	87.00
OViFS1_2	97.00	**95.00**	**84.00**	**92.00**

The bold values indicates the values that outperform the previous works.

Table 13 Cross-view recognition accuracy (%) of 2D PCA + LDA

Gallery					Average, %
Probe	55°, %	65°, %	75°, %	85°, %	
55°	90.4	89.9	73.8	53.8	77
65°	72.2	94.8	93.1	79.6	84.9
75°	44.1	92.1	95.5	93.7	81.4
85°	35.3	76.4	94.5	96.9	75.7

OViFS1_4: PCA + LDA, $e\_ratio = 0.99$, available views: 36°, 72°, 108° and 144°.

From Table 10, one can observe that OViFS achieves higher accuracy, while requiring less view angles in the gallery set compared to Tang et al.'s work. Note that for the three-view setting, PCA + LDA achieves higher accuracy than 2D PCA + LDA, on average; while in the two-view setting, both attain identical average results.

Table 11 tabulates the recognition accuracy of OViFS_2 and different state-of-the-art methods using a 24–100 dataset partition with a gallery view angle of 54° and probe data with view angles of {36°, 72°}, {18°, 90°} and {0°, 108°}. Table 12 tabulates the recognition accuracy for the case of a gallery view angle of 54° and probe data with view angles of {108°, 144°}, {90°, 162°} and {72°, 180°}. In these tables, Zhang *et al.* (1) refers to the feature level fusion adopted by [19], Zhang *et al.* (2) refers to the score-level fusion from the same work, and Zhang *et al.* (3) refers to their multi-view DPLCR (DPLCR is the acronym of discriminative projection with list-wise constraints with rectification, which is the framework proposed by Zhang *et al.* in their paper [30]. In their paper its performance reported on the mainstream datasets is the-state-of-the-art.). From these two tables, one can observe that OViFS1_2 outperforms the state-of-the-art methods. Let us recall that ViFS is designed to match gallery data from multiple view angles with single-view angle probe data. However, it is possible for ViFS to work with the opposite situation, since the feature selection is a feature mapping process from a multi-view set to a single-view template. Note that OViFS1_2 is also more robust to large view angle variances according to the tabulated results. For example, in Table 11, when the view angle of the gallery data is 54° and the view angles of the probe data are {18°, 90°}, our framework outperforms Zhang *et al.*'s method by 8%. In Table 12, when the view angle of the gallery data is 126° and the view angles of the probe data are {72°, 180°}, OViFS_2 outperforms other methods by up to 12%.

5.2 Experiments on the OU-ISIR LP dataset

The OU-ISIR large population dataset includes more than 4000 subjects, each recorded using cameras from four view angles: 55°, 65°, 75°, and 85°. Among all the datasets commonly used for gait recognition evaluation, this dataset is one of the largest in terms of the number of subjects, thus it is more statistically reliable for performance evaluation. According to the existing evaluation protocols [21, 29, 30], a common experiment setting is to use a sub-set of 1912 subjects, which is divided into two groups, where 956 subjects are used for training and the rest for testing. We refer to this subset as the OU-ISIR LP dataset in the following discussions and results. As shown in Section 5.1, our improved ViFS when used with 2D PCA + LDA attains a strong performance on the CASIA B dataset can be trained faster and is less affected by parameters (*e_ratio* values). Hence, we focus here on evaluating OViFS2.

5.2.1 Cross-view evaluation using 2D PCA + LDA: Table 13 tabulates the recognition accuracy using 2D PCA + LDA (without OViFS) on the OU-ISIR LP dataset. The purpose of these results is to set a baseline to measure the improvements attained by OViFS2. From this table, one can observe than when the view angle difference between gallery and probe data is small (e.g. 10° or less), 2D PCA + LDA achieves a high accuracy, close to the identical-view matching results. However, when the view angle difference is larger than 20°, the accuracy decreases fast.

5.2.2 Multi-view evaluation: We evaluate OViFS2 for multi-view matching on the OU-ISIR LP dataset. We use four views, 55°, 65°, 75° and 85°, to train OViFS2_4. Here, we compare the performance of OViFS2_4 with two CNN-based approaches proposed by Wu *et al.* [22] and Shiraga *et al.* [21]. As done in [21, 22], five-fold cross-validations are employed to reduce the effect of randomness. Specifically, the training and testing sets (each contains 956 subjects) are randomly selected five times; each time we record the recognition accuracy, and the final accuracy is the average of the five experiments.

Table 14 tabulates the recognition accuracy (%) of OViFS2_4 and the CNN-based approaches. Wu *et al.* (i) refers to the case of identical-view matching, while Wu *et al.* (a) is the average matching accuracy of the gallery view angles with a certain probe view angle. The same notations apply to Shiraga *et al.*'s method. We notice that Wu *et al.*'s method achieves the highest accuracy for identical-view matching, which confirms the effectiveness of CNN to extract discriminative features, especially when sufficient number of training samples is available. Shiraga *et al.*'s method uses a shallower network than that used by Wu *et al.*, and it does not use the pair-image approach to train the network, thus its performance is lower than that attained by Wu *et al.*'s method. OViFS2_4 attains a very good performance. Moreover, its overall performance is very close to that attained by Wu *et al.*'s method for the identical-view scenario. It is worth mentioning that the training and testing time of ViFS2_4 is shorter than that required by the CNN-based approaches.

6 Conclusion

In this paper, we introduced feature scaling to improve ViFS, which is a feature selector that achieves robust view-invariant

Table 14 Multi-view recognition accuracy (%) of ViFS2_4 and other methods

Gallery	55°–85°				Average, %
probe	55°, %	65°, %	75°, %	85°, %	
Wu *et al.* (i)	**98.8**	**98.8**	**98.8**	**98.8**	**98.8**
Wu *et al.* (a)	93.5	94	94	95.8	89.9
Shiraga *et al.* (i)	94.7	95.1	95.2	94.7	94.9
Shiraga *et al.* (a)	89.2	93.3	93.3	90.5	91.6
OViFS2	**97.8**	**97.9**	**98.4**	**97.5**	**97.7**

The bold values indicates the values that outperform the previous works.

recognition by reconstructing gallery data at different view angles to be matched with single view angle probe data. The improvements introduced to ViFS in this paper normalise the associated minimiser so that the reconstructed gallery features are aligned with the probe features. To enhance the reconstructed features, our improved ViFS, denoted by OViFS, employs PCA + LDA. We evaluated OViFS with a wide range of gallery view angles, for a different number of eigenvectors. Our results showed that for high precision tasks, OViFS with PCA + LDA is most appropriate, which can attain an average recognition accuracy of 99% with gallery data from three widely spread view angles. For real-time processing, our results showed that OViFS with 2D PCA + LDA is the best choice due to its small sensitivity to parameter changes and low computational cost. Our results also showed that OViFS achieves a better performance than the state-of-the-art methods, while requiring less available gallery data from different view angles. OViFS has the potential to be used in practical scenarios when more than two cameras are available, such as in border control, smart homes and surveillance systems.

7. Acknowledgment

This work was supported by the EU Horizon 2020 – Marie Sklodowska-Curie Actions through the project entitled Computer Vision Enabled Multimedia Forensics and People Identification (Project No. 690907, Acronym: IDENTITY).

8 References

[1] Chew-Yean, Y., Nixon, M.: 'Encyclopedia of Biometrics', in Li, S.Z., Jain, A. (Eds.): 2009, pp. 633–639

[2] Han, J., Bhanu, B.: 'Individual recognition using gait energy image', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2006, **28**, (2), pp. 316–322

[3] Wang, C., Zhang, J., Wang, L., *et al.*: 'Human identification using temporal information preserving gait template', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2012, **34**, (11), pp. 2164–2176

[4] Bashir, K., Xiang, T., Gong, S.: 'Gait recognition using gait entropy image'. In Proc. 3rd Int. Conf. Crime Detection and Prevention (ICDP), December 2009, pp. 4195–4199

[5] Iwama, H., Okumura, M., Makihara, Y., *et al.*: 'The ou-isir gait database comprising the large population dataset and performance evaluation of gait recognition', *IEEE Trans. Inf. Forensics Sec.*, 2012, **7**, (5), pp. 1511–1521

[6] Makihara, Y., Matovski, D. S., Nixon, M. S., *et al.*: 'Gait recognition: databases, representations, and applications', *Wiley Encycl. Electr. Electron. Eng.*, 2015, pp. 1–15

[7] Jia, N., Li, C.T., Sanchez, V., *et al.*: 'Fast and robust framework for view-invariant gait recognition'. 2017 5th Int. Workshop on Biometrics and Forensics (IWBF), April 2017, pp. 1–6

[8] Tang, J., Luo, J., Tjahjadi, T., *et al.*: 'Robust arbitrary-view gait recognition based on 3d partial similarity matching', *IEEE Trans. Image Process.*, 2017, **26**, (1), pp. 7–22

[9] Luo, J., Tang, J., Tjahjadi, T., *et al.*: 'Robust arbitrary view gait recognition based on parametric 3d human body reconstruction and virtual posture synthesis', *Pattern Recognit.*, 2016, **60**, pp. 361–377

[10] Kusakunniran, W., Wu, Q., Zhang, J., *et al.*: 'A new view-invariant feature for cross-view gait recognition', *IEEE Trans. Inf. Forensics Sec.*, 2013, **8**, (10), pp. 1642–1653

[11] Goffredo, M., Bouchrika, I., Carter, J., *et al.*: 'Self-calibrating view-invariant gait biometrics', *IEEE Trans. Syst. Man Cybern. B: Cybern.*, 2010, **40**, (4), pp. 997–1008

[12] Makihara, Y., Sagawa, R., Mukaigawa, Y., *et al.*: 'Which reference view is effective for gait identification using a view transformation model?'. Proc. Conf. Computer Vision and Pattern Recognition Workshop (CVPRW), June 2006, pp. 45–45

[13] Kusakunniran, W., Wu, Q., Li, H., *et al.*: 'Multiple views gait recognition using view transformation model based on optimized gait energy image'. 2009 IEEE 12th Int. Conf. Computer Vision Workshops, ICCV Workshops, September 2009, pp. 1058–1064

[14] Muramatsu, D., Shiraishi, A., Makihara, Y., *et al.*: 'Gait-based person recognition using arbitrary view transformation model', *IEEE Trans. Image Process.*, 2015, **24**, (1), pp. 140–154

[15] Portillo-Portillo, J., Leyva, R., Sanchez, V., *et al.*: '*View-invariant gait recognition using a joint-DLDA framework*' (Springer International Publishing, Cham, 2016), pp. 398–408

[16] Portillo-Portillo, J., Leyva, R., Sanchez, V., *et al.*: 'Cross view gait recognition using joint-direct linear discriminant analysis', *Sensors*, 2017, **17**, (1), p. 6

[17] Portillo-Portillo, J., Leyva, R., Sanchez, V., *et al.*: 'A view-invariant gait recognition algorithm based on a joint-direct linear discriminant analysis', *Appl. Intell.*, 2017

[18] Hu, M., Wang, Y., Zhang, Z., *et al.*: 'View-invariant discriminative projection for multi-view gait-based human identification', *IEEE Trans. Inf. Forensics Sec.*, 2013, **8**, (12), pp. 2034–2045

[19] Zhang, Z., Chen, J., Wu, Q., *et al.*: 'Gii representation-based cross-view gait recognition by discriminative projection with list-wise constraints', *IEEE Trans. Cybern.*, 2017, PP, (99), pp. 1–13

[20] Alotaibi, M., Mahmood, A.: 'Improved gait recognition based on specialized deep convolutional neural networks'. Proc. IEEE Applied Imagery Pattern Recognition Workshop (AIPR), 2015, pp. 1–7

[21] Shiraga, K., Makihara, Y., Muramatsu, D., *et al.* Geinet: 'View-invariant gait recognition using a convolutional neural network'. Proc. Int. Conf. Biometrics (ICB), June 2016, pp. 1–8

[22] Wu, Z., Huang, Y., Wang, L., *et al.*: 'A comprehensive study on cross-view gait based human identification with deep CNNs', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2016, **PP, (99)**, pp. 1–1

[23] Wang, L., Tan, T., Ning, H., *et al.*: 'Silhouette analysis-based gait recognition for human identification', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2003, **25**, (12), pp. 1505–1518

[24] Sarkar, S., Phillips, P.J., Liu, Z., *et al.*: 'The humanoid gait challenge problem: data sets, performance, and analysis', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2005, **27**, (2), pp. 162–177

[25] Shelhamer, E., Long, J., Darrell, T.: 'Fully convolutional networks for semantic segmentation', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2017, **39**, (4), pp. 640–651

[26] Friedman, J., Hastie, T., Tibshirani, R.: 'The elements of statistical learning', in '*Springer series in statistics*' (Springer, Berlin, 2001), **volume 1**

[27] Yang, J., Zhang, D., Frangi, A. F., *et al.*: 'Two-dimensional PCA: a new approach to appearance-based face representation and recognition', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2004, **26**, (1), pp. 131–137

[28] Yu, S., Tan, D., Tan, T.: 'A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition'. Proc. 18th Int. Conf. Pattern Recognition (ICPR), 2006, vol. 4, pp. 441–444

[29] Mansur, A., Makihara, Y., Muramatsu, D., *et al.*: 'Cross-view gait recognition using view-dependent discriminative analysis'. Proc. IEEE Int. Joint Conf. on Biometrics (IJCB), September 2014, pp. 1–8

[30] Muramatsu, D., Makihara, Y., Yagi, Y.: 'View transformation model incorporating quality measures for cross-view gait recognition', *IEEE Trans. Cybern.*, 2016, **46**, (7), pp. 1602–1615