Available online at www.sciencedirect.com
ScienceDirect
Pattern Recognition 41 (2008) 778–795
PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY
www.elsevier.com/locate/pr

Feature fusion of side face and gait for video-based human identification

Xiaoli Zhou, Bir Bhanu*

Center for Research in Intelligent Systems, University of California, Riverside, CA 92521, USA

Received 5 May 2007; accepted 26 June 2007

Abstract

Video-based human recognition at a distance remains a challenging problem for the fusion of multimodal biometrics. As compared to the approach based on match score level fusion, in this paper, we present a new approach that utilizes and integrates information from side face and gait at the feature level. The features of face and gait are obtained separately using principal component analysis (PCA) from enhanced side face image (ESFI) and gait energy image (GEI), respectively. Multiple discriminant analysis (MDA) is employed on the concatenated features of face and gait to obtain discriminating synthetic features. This process allows the generation of better features and reduces the curse of dimensionality. The proposed scheme is tested using two comparative data sets to show the effect of changing clothes and face changing over time. Moreover, the proposed feature level fusion is compared with the match score level fusion and another feature level fusion scheme. The experimental results demonstrate that the synthetic features, encoding both side face and gait information, carry more discriminating power than the individual biometrics features, and the proposed feature level fusion scheme outperforms the match score level and another feature level fusion scheme. The performance of different fusion schemes is also shown as cumulative match characteristic (CMC) curves. They further demonstrate the strength of the proposed fusion scheme.

Keywords: Face recognition; Gait recognition; Multibiometrics fusion; Video based recognition at a distance

1. Introduction

Compared with the abundance of research work related to fusion at the match score level, fusion at the feature level is a relatively understudied problem because of the difficulties in practice. Multiple modalities may have incompatible feature sets and the relationship between different feature spaces may not be known [1]. Moreover, the concatenated feature vectors may lead to the problem of curse of dimensionality and it may contain noisy or redundant data, thus leading to a decrease in the performance of the classifier.

However, pattern recognition and computer vision systems that integrate information at an early stage of processing are believed to be more effective than those systems that perform integration at a later stage. Therefore, while it is relatively difficult to achieve in practice, fusion at the feature level has drawn more attention in recent years. Table 1 presents a summary of the recent work for the feature level fusion. Among the existing research work, feature concatenation is the most popular feature level fusion methodology. Some of schemes perform feature concatenation after dimensionality reduction [2–5] while others perform feature concatenation before feature selection or transformation [6,7].

In recent years, integrated face and gait recognition approaches without resorting to 3-D models have achieved some success. Most of the fusion schemes [12–16] have focused on the fusion of face and gait at the match score level and the experimental results demonstrate improved performance after fusion. Recently, Zhou and Bhanu [2] conducted feature concatenation after dimensionality reduction by the PCA and MDA combined method to fuse face and gait information at the feature level. The experimental results showed the performance improvement compared with the single biometrics, but they do not show any comparison with other schemes. Since the feature set contains richer information about the input biometrics pattern than the match score, integration at this level is expected to provide better recognition results than the match score level. Therefore, the fusion of face and gait at the feature

* Corresponding author. Tel.: +1 951 827 3954.
E-mail address: bhanu@cris.ucr.edu (B. Bhanu).

0031-3203/$30.00 © 2007 Pattern Recognition Society. Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2007.06.019

Table 1
The recent work for feature level fusion

Authors	Modalities	Methodology	Data
Yang et al. [7]	Face	PCA, K–L expansion and LDA after parallel concatenation of features	CENPARMI handwritten numeral database, NUST603 handwritten Chinese character database and ORL face image database
Kumar et al. [5]	Hand	Concatenation of geometry and texture features	1000 hand images of 100 individuals
Kinnunen et al. [4]	Voice	Concatenation of LPCC, MFCC, ARCSIN and FMT features	NIST-1999 subset
Moon et al. [8]	Fingerprint	Averaging two templates of minutiae	1149 images of 383 individuals
Feng et al. [3]	Face and palmprint	Concatenation of PCA and LDA coefficients	400 images of 40 individuals
Ross et al. [6]	Face and hand	Feature selection after concatenation of PCA and LDA coefficients	500 face and hand images of 100 individuals
Gao et al. [9]	Face and palmprints	Fusion of line features by multiview line segment Hausdorff distance	(a) 210 images of 35 individuals (b) 311 images of 35 individuals from the University of Stirling
Zhou et al. [2]	Face and gait	Feature concatenation after MDA and PCA combined method	92 video sequences of 46 individuals
Kong et al. [10]	Palmprint	Fusion of phase information from Gabor filters according to a fusion rule	9599 palmprint images of 488 different palms
Li et al. [11]	Palmprint, knuckleprint and hand shape	KPCA after fusion of kernel matrixes using decision level fusion operator	1853 right-hand images of 98 individuals

Table 2
The related work for integrating face and gait for human recognition vs. this paper

Authors	Modalities	Fusion methods	Data
Kale et al. [12]	Frontal face and gait	Hierarchical fusion and Sum/Product rule	30 subjects (number of sequences per person is not specified) and static images as the face gallery
Shakhnarovich et al. [13,14]	Frontal face and gait	Sum rule [13]	12 subjects and 2–6 sequences per person [13]
		Min, Max, Sum and Product rules [14]	26 subjects and 2–14 sequences per person [14]
Zhou et al. [2,17,19]	Side face and gait [2,17]	Feature concatenation after MDA and PCA combined method [2]	46 subjects and 2 sequences per person [2]
	Face profile and gait [19]	Sum, Product and Max rules [17]	45 subjects and 2–3 video per person [17]
		Hierarchical fusion, Sum and Product rules [19]	14 subjects and 2 sequences per person [19]
This paper	Side face and gait	MDA after concatenation of PCA-based features of side face and gait	45 individuals and 2–3 video per person

level deserves a closer study and performance comparison between different fusion schemes.

Table 2 presents a summary of related work and compares it with the work presented in this paper for the fusion of face and gait. In this paper, information related to side face and gait, from a single camera video sequence, is combined at the feature level to recognize non-cooperating individuals at a distance. We distinguish a side face from a face profile. A face profile refers to the outline of the shape of a face as seen from the side. A side face includes not only the outline of the side view of a face, but also the entire side view of eye, nose and mouth, possessing both shape and intensity information. Therefore, a side face has more discriminating power for recognition than a face profile. For side face, an enhanced side face image (ESFI), a higher resolution image compared with the image directly obtained from a single video frame, is constructed as the face template [17]. For gait, the gait energy image (GEI), which is used to characterize human walking properties, is generated as the gait template [18]. The contributions of this paper are as follows:

- A new feature level fusion scheme is proposed to fuse information from side face and gait for human recognition at a distance in a single camera scenario. Multiple discriminant analysis (MDA) is applied after the concatenation of face and gait features. This allows the generation of better discriminating features and leads to the improved performance. Face features are extracted from ESFI, which integrates face information over multiple frames in video. Gait features are extracted from GEI, a spatio-temporal compact representation of gait in video.
- The proposed feature level fusion scheme is compared with the match score level fusion schemes (Sum and Max rules) [17] and the feature level fusion scheme [2]. The basic processes of these approaches are shown in Fig. 1. The experimental results demonstrate the effectiveness of the fusion at the feature level in comparison to the match score level. The proposed feature level fusion scheme performs

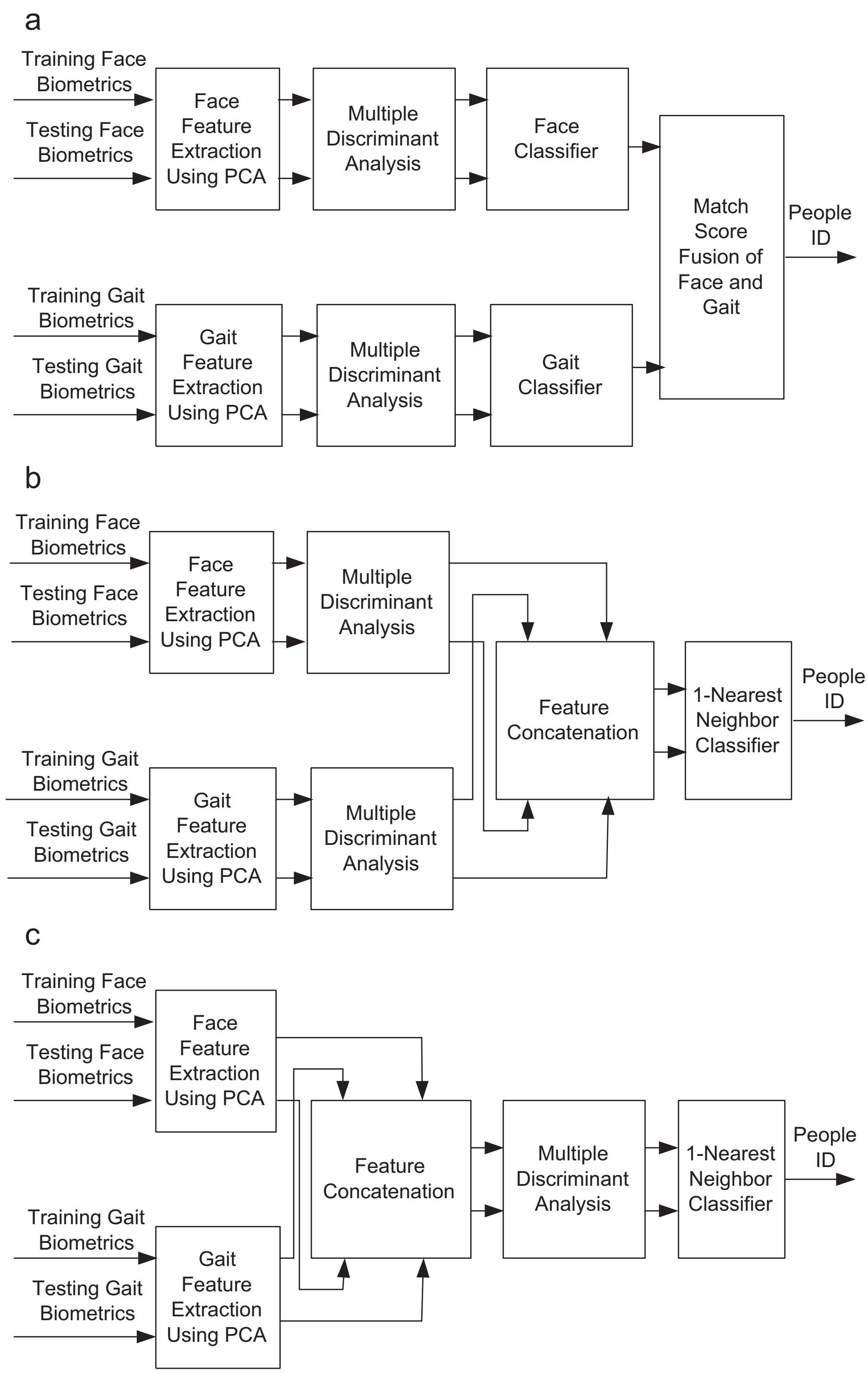

Fig. 1. The basic processes of the fusion schemes for comparison: (a) the match score level fusion scheme [17]; (b) the feature level fusion scheme [2]; and (c) the proposed feature level fusion scheme (this paper).

the best among all the compared fusion schemes. Besides the recognition rates, the performance is also compared using CMC (cumulative match characteristic) curves. They further demonstrate the strengths of the proposed fusion scheme.

- The problem of the curse of dimensionality is reduced in two ways: (a) PCA is used to transform high dimensional face and gait templates to low dimensional feature space; (b) synthetic features are generated based on all possible combinations of face and gait features from the same video sequence.

The paper is organized as follows. Section 2 presents the overall technical approach. It introduces the construction of ESFI and GEI. It describes feature extraction from ESFI and GEI. It explains the proposed scheme to generate synthetic features for feature level fusion and classification. Section 3 provides a description of the related fusion methods [2,17] to be compared in the experimental section. In Section 4, a number of dynamic video sequences are tested using the approach presented. Experimental results are compared and discussed. Section 5 concludes the paper.

2. Technical approach

The proposed feature level fusion scheme is shown in Fig. 2. ESFI and GEI are first constructed as the face template and the gait template from a video sequence, respectively. Principal component analysis (PCA) is employed separately on face templates and gait templates to extract lower dimensional face features and gait features. MDA is then applied to the concatenated features to generate the synthetic features. Finally, the testing synthetic features are compared with the training synthetic features to evaluate the performance of the proposed approach.

2.1. ESFI construction

It is difficult to get reliable information of a side face directly from a video frame for the recognition task because of the limited resolution and small size of the face compared to the human body. To overcome this problem, we construct an ESFI, a higher resolution image compared with the image directly obtained from a single video frame, by fusing the face information from multiple video frames. The idea relies on the fact that the temporally adjacent frames in a video sequence, in which one is walking with a side view to the camera, contain slightly different, but unique information about a side face. The ESFI construction involves registration of low-resolution images, selection of aligned low-resolution images, and formation of a high-resolution image using the selected images. The details of the construction of high-resolution side face images from video sequences are described in Ref. [17]. We use the same method here and provide and example as shown in Fig. 3.

The resolution of a video frame is 480×720. The resolution of low- and high-resolution images are 68×68 and 136, respectively. Before feature extraction, all high-resolution side face images are cropped and normalized to the size of 64×32. We call these images as ESFIs. Similarly, original side face image (OSFI) is a subimage from the normalized version of the low-resolution side face image. The size of OSFI is 34×18. Fig. 3(a) shows one low-resolution face image and one reconstructed high-resolution face image. For comparison, we resize the low-resolution face image using bilinear interpolation. Fig. 3(b) shows one example of the resized OSFIs and ESFIs for comparison. Clearly, ESFIs have better quality than OSFIs. Experiments show that better face features can be extracted from constructed ESFI compared to those from the OSFIs [17].

2.2. GEI construction

GEI is a spatio-temporal compact representation of gait in video. The entire gait sequence is divided into cycles according to gait frequency and phase information. GEI reflects major shapes of silhouettes and their changes over the gait cycle. It accounts for human walking at different speeds. GEI has several advantages over the gait representation of binary silhouette sequence. GEI is not sensitive to incidental silhouette errors in the individual frames. Moreover, with such a 2D template, we do not need to consider the time moment of each frame, and the incurred errors can be, therefore, avoided.

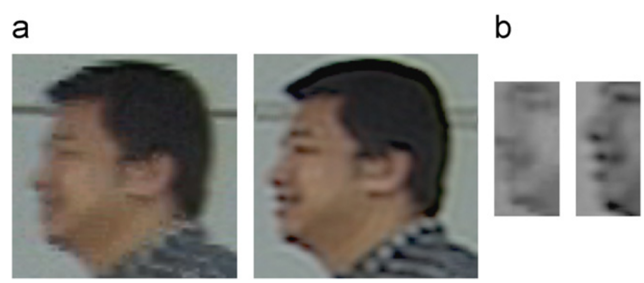

Fig. 3. (a) One resized low-resolution face image (left) and one reconstructed high-resolution face image (right) and (b) resized OSFI (left) and ESFI (right).

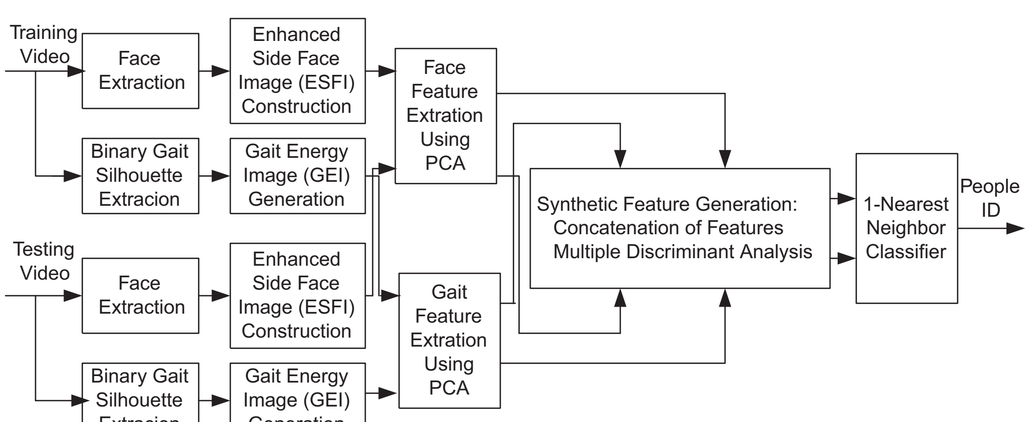

Fig. 2. The proposed feature level fusion scheme for integrating side face and gait in video.

Fig. 4. Examples of normalized and aligned silhouette images in a gait cycle. The right most image is the corresponding gait energy image (GEI).

Given the preprocessed binary gait silhouette sequence in the complete cycle(s), the gray-level GEI is obtained by averaging the normalized and aligned silhouette images in the gait cycle(s) [18]. Fig. 4 shows the sample silhouette images in a gait cycle from a person and the right most image is the corresponding GEI. The resolution of each GEI is 300×200.

2.3. Human identification using ESFI and GEI

2.3.1. Feature learning using PCA

PCA is a standard decorrelation technique [20]. The derived orthogonal projection basis leads to dimensionality reduction, and possibly to feature selection.

Let $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$, $\mathbf{x}_k \in \mathbb{R}^N$, be n random vectors representing n ESFIs or n GEIs, where N is the dimensionality of the vector obtained by concatenation of an image row-by-row. The covariance matrix is defined as $\mathbf{\Sigma_x} = E([\mathbf{x} - E(\mathbf{x})][\mathbf{x} - E(\mathbf{x})]^{\mathrm{T}})$, where $E(\cdot)$ is the expectation operator and T denotes the transpose operation. The covariance matrix $\mathbf{\Sigma_x}$ can be factorized into the following form:

$$\mathbf{\Sigma_x} = \mathbf{\Phi\Lambda\Phi}, \tag{1}$$

where $\mathbf{\Phi} = [\mathbf{\Phi}_1 \;\; \mathbf{\Phi}_2 \;\; \ldots \;\; \mathbf{\Phi}_N] \in \mathbb{R}^{N\times N}$ is the orthonormal eigenvector matrix of $\mathbf{\Sigma_x}$; $\mathbf{\Lambda} = \{\mathbf{\Lambda}_1 \;\; \mathbf{\Lambda}_2 \;\; \ldots \;\; \mathbf{\Lambda}_N\} \in \mathbb{R}^{N\times N}$ is the diagonal eigenvalue matrix of $\mathbf{\Sigma_x}$ with diagonal elements in descending order. One important property of PCA is its optimal signal reconstruction in the sense of minimum mean square error (MSE) when only a subset of principal components are used to represent the original signal. An immediate application of this property is the dimensionality reduction:

$$\mathbf{y}_k = \mathbf{P}_{pca}^{\mathrm{T}}[\mathbf{x}_k - E(\mathbf{x})], \quad k = 1, 2, \ldots, n, \tag{2}$$

where $\mathbf{P}_{pca} = [\mathbf{\Phi}_1 \;\; \mathbf{\Phi}_2 \;\; \ldots \;\; \mathbf{\Phi}_m], m \leqslant N$. The lower dimensional vector $\mathbf{y}_k \in \mathbb{R}^m$ captures the most expressive features of the original data $\mathbf{x}_k$.

2.3.2. Synthetic feature generation and classification

Let $\mathbf{f} \in \mathbb{R}^{N_1}$ and $\mathbf{g} \in \mathbb{R}^{N_2}$ be ESFI and GEI of a person represented as a vector, where N_1 and N_2 are the dimensionality of the face and the gait feature spaces, respectively. We obtain low dimensional feature vectors, $\mathbf{f}' = \mathbf{M}^f\mathbf{f}$ and $\mathbf{g}' = \mathbf{M}^g\mathbf{g}$, by using the PCA method as in Eq. (2). $\mathbf{M}^f$ and $\mathbf{M}^g$ are the PCA transformation matrices for face and gait, respectively. We choose a subset of principal components to derive the lower dimensional face and gait features, $\mathbf{f}' \in \mathbb{R}^{m_1}$ and $\mathbf{g}' \in \mathbb{R}^{m_2}$, where m_1 and m_2 are the dimensionality of the reduced face feature space and gait feature space, respectively. On the one hand, we hope to lose as little representative information of the original data as possible in the transformation from the high dimensional space to the low dimensional one. On the other hand, the eigenvectors corresponding to the small eigenvalues are excluded from the reduced space so that we can obtain more robust MDA projection as well as reduce the problem of curse of dimensionality. The eigenvalue spectrum of the covariance matrix of the training data supplies useful information regarding the choice for the dimensionality of the feature space.

Before face features and gait features are combined, the individual face features and gait features are normalized to have their values lie within similar ranges. We use a linear method [21], which provides a normalization via the respective estimates of the mean and variance. For the jth feature value in the ith feature vector w_{ij}, we have

$$\hat{w}_{ij} = \frac{w_{ij} - \bar{w}_j}{\sigma_j}, \quad i = 1, 2, \ldots, I, \quad j = 1, 2, \ldots, L, \tag{3}$$

where $\bar{w}_j = (1/I)\sum_{i=1}^{I} w_{ij}$ and $\sigma_j^2 = (1/(I-1))\sum_{i=1}^{I}(w_{ij} - \bar{w}_j)^2$. I is the number of available feature vectors and L is the number of features for each feature vector. The resulting normalized features have zero mean and unit variance.

We assume that $\hat{\mathbf{f}}$ and $\hat{\mathbf{g}}$ are face features and gait features after normalization using Eq. (3), respectively. They are concatenated to form the features as follows:

$$\mathbf{h} = [\hat{\mathbf{f}} \;\; \hat{\mathbf{g}}], \tag{4}$$

where $\mathbf{h} \in \mathbb{R}^{m_1+m_2}$. The rational behind such a simple combination is that the face and gait are viewed as carrying equally important discriminating information.

Since face and gait can be regraded as two independent biometrics in our application scenario, synchronization is totally unnecessary for them. To take advantage of information for a walking person in video, we use all possible combinations of side face features and gait features to generate the maximum number of vectors $\mathbf{h}$. Specifically, we have two feature vectors of side face and two feature vectors of gait for one person from one video. Therefore, we have four concatenated features $\mathbf{h}$ for one person from one video. It is reasonable to concatenate face and gait feature vectors in this way, since ESFI is built from multiple video frames and GEI is a compact spatio-temporal representation of gait in video. Generation of all possible vectors $\mathbf{h}$ from PCA features for side face and gait data helps to reduce the problem of curse of dimensionality for the subsequent MDA transformation.

Suppose that $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_c$ and $n_1, n_2, \ldots, n_c$ denote the classes and the number of concatenated feature vectors h within each class, respectively, with $\mathbf{w} = \mathbf{w}_1 \cup \mathbf{w}_2 \cup \cdots \cup \mathbf{w}_c$ and $\hat{n} = n_1 + n_2 + \cdots + n_c$. Note that the value of $\hat{n}$ is two times of n. c is the number of classes. MDA seeks a transformation matrix $\mathbf{W}$ that maximizes the ratio of the between-class scatter matrix $\mathbf{S}_B$ to the within-class scatter matrix $\mathbf{S}_W$: $J(\mathbf{W}) = |\mathbf{W}^{\mathrm{T}}\mathbf{S}_B\mathbf{W}|/|\mathbf{W}^{\mathrm{T}}\mathbf{S}_W\mathbf{W}|$. The within-class scatter matrix is $\mathbf{S}_W = \sum_{i=1}^{c}\sum_{\mathbf{h}\in\mathbf{w}_i}(\mathbf{h}-\mathbf{M}_i)(\mathbf{h}-\mathbf{M}_i)^{\mathrm{T}}$ and the between-class scatter matrix is $\mathbf{S}_B = \sum_{i=1}^{c} n_i(\mathbf{M}_i - \mathbf{M})(\mathbf{M}_i - \mathbf{M})^{\mathrm{T}}$, where $\mathbf{M}_i = (1/n_i)\sum_{\mathbf{h}\in\mathbf{w}_i}\mathbf{h}$ and $\mathbf{M} = (1/\hat{n})\sum_{\mathbf{h}\in\mathbf{w}}\mathbf{h}$ are the means of the class i and the grand mean, respectively. $J(\mathbf{W})$ is maximized when the columns of $\mathbf{W}$ are the generalized eigenvectors of $\mathbf{S}_B$ and $\mathbf{S}_W$, which correspond to the largest generalized eigenvalues in

$$\mathbf{S}_B\mathbf{\Psi}_i = \lambda_i \mathbf{S}_W \mathbf{\Psi}_i. \tag{5}$$

There are no more than $c-1$ non-zero eigenvalues λ_i and let the corresponding eigenvectors be $\mathbf{\Psi}_i$. The transformed feature vector is obtained as follows:

$$\mathbf{z}_k = \mathbf{P}_{mda}^{\mathrm{T}}\mathbf{h}_k, \quad k = 1, 2, \ldots, \hat{n}. \tag{6}$$

where $\mathbf{P}_{mda} = [\mathbf{\Psi}_1 \ \ \mathbf{\Psi}_2 \ \ \ldots \ \ \mathbf{\Psi}_{c-1}]$ is the MDA transformation matrix. We call the lower dimensional vector $\mathbf{z}_k \in \mathbb{R}^{c-1}$ the *synthetic feature*, which captures the most discriminating power of the face and gait.

Let $\mathbf{U}_i, i = 1, 2, \ldots, c$, the mean of the training synthetic features of class i, be the prototype of class i. The unknown person is classified to class K to which the synthetic feature $\mathbf{z}$ is the nearest neighbor.

$$\|\mathbf{z} - \mathbf{U_K}\| = \min \|\mathbf{z} - \mathbf{U}_i\|. \tag{7}$$

When multiple synthetic features are obtained for one person, Eq. (7) means that the unknown person is classified to the class which has the minimum distance out of all the distances corresponding to all the classes.

3. The related fusion schemes

In this paper, we also compare the proposed feature level fusion scheme with the related fusion schemes at the match score level (Sum and Max rules) [17] and the feature level [2]. These techniques are explained in the following. PCA and MDA combined method [17] is used for feature learning from face and gait for these two fusion schemes. It is applied to face templates (ESFIs) and gait templates (GEIs) separately to get low dimensional feature representation for side face and gait. The transformed feature vector is obtained as follows:

$$\mathbf{z}_k = \mathbf{P}_{mda}^{\mathrm{T}}\mathbf{P}_{pca}^{\mathrm{T}}[\mathbf{x}_k - E(\mathbf{x})] = \mathbf{Q}[\mathbf{x}_k - E(\mathbf{x})],$$
$$k = 1, \ldots, n, \tag{8}$$

where $\mathbf{x}_k \in \mathbb{R}^N$ is the vector representing n ESFIs or n GEIs, where N is the dimensionality of the vector obtained by concatenation of an image row-by-row. $\mathbf{P}_{pca} = [\mathbf{\Phi}_1 \ \ \mathbf{\Phi}_2 \ \ \ldots \ \ \mathbf{\Phi}_m], m \leqslant \min(n, N)$ is the PCA transformation matrix, $\mathbf{P}_{mda} = [\mathbf{\Psi}_1 \ \ \mathbf{\Psi}_2 \ \ \ldots \ \ \mathbf{\Psi}_r], r \leqslant c-1$ is the MDA transformation matrix and $\mathbf{Q}$ is the overall transformation matrix. The lower dimensional vector $\mathbf{z}_k \in \mathbb{R}^r$ captures the most expressive and discriminating features of the original data $\mathbf{x}_k$.

Let $\mathbf{f} \in \mathbb{R}^{N_1}$ and $\mathbf{g} \in \mathbb{R}^{N_2}$ represent ESFI and GEI of a person, where N_1 and N_2 are the dimensionality of the face and the gait spaces, respectively. We obtain low dimensional feature vectors, $\mathbf{f}' = \mathbf{Q}^f\mathbf{f}$ and $\mathbf{g}' = \mathbf{Q}^g\mathbf{g}$, by using PCA and MDA combined method as in Eq. (8). $\mathbf{Q}^f$ and $\mathbf{Q}^g$ are the overall transformation matrices for face and gait, respectively.

3.1. Fusion at the match score level [17]

Given $\mathbf{f}'$ and $\mathbf{g}'$, the Euclidean distance for the face classifier and the gait classifier are obtained as

$$D_i^f = \|\mathbf{f}' - U_i^f\|,$$
$$D_i^g = \|\mathbf{g}' - U_i^g\|, \tag{9}$$

where U_i^f and U_i^g, $i = 1, 2, \ldots, c$, are the prototypes (mean of the features that belong to a class) of class i for face and gait, respectively. Before combination of the results of face classifier and the results of gait classifier, it is necessary to map distances obtained from the different classifiers to the same range of values. We use an exponential transformation here. Given that the distance of a probe (test data) obtained from the classifier are $D_1, D_2, \ldots, D_c$, we obtain the normalized match scores as,

$$S_i = \frac{\exp(-D_i)}{\sum_{i=1}^{c}\exp(-D_i)}, \quad i = 1, 2, \ldots, c. \tag{10}$$

After normalization using Eq. (10), the match scores of face templates and the match scores of gait templates from the same video are fused based on different match score fusion rules. We use all the possible combinations of face match scores and gait match scores to generate the maximum number of fused match scores based on the characteristics of face and gait. Specifically, we use two face match scores and two gait match scores to generate four fused match scores for one person from each video.

Since the distances representing dissimilarity become match scores by using Eq. (10), the unknown person should be classified to the class for which the fused match score is the largest. Let S_i^f and S_i^g be the normalized match scores of D_i^f and D_i^g, respectively. The unknown person is classified to class K if

$$R\{S_K^f, S_K^g\} = \max R\{S_i^f, S_i^g\}, \tag{11}$$

where $R\{,\}$ means a fusion rule. In this paper, we use Sum and Max rules. Since we obtain more than one fused match scores after fusion for one testing video sequence, Eq. (11) means the unknown person is classified to the class which gets the maximum fused match score out of all the fused match scores corresponding to all the classes.

3.2. Fusion at the feature level [2]

Before face features $\mathbf{f}'$ and gait features $\mathbf{g}'$ are combined, the individual face features and gait features are normalized to have their values lie within similar ranges using Eq. (3). We assume that $\hat{\mathbf{f}}$ and $\hat{\mathbf{g}}$ are face features and gait features after normalization using Eq. (3), respectively. They are concatenated to form the features as follows:

$$\mathbf{p} = [\hat{\mathbf{f}}\ \hat{\mathbf{g}}], \tag{12}$$

where $\mathbf{p} \in \mathbb{R}^{m_1+m_2}$. As explained in Section 2.3.2, we use all possible combinations of side face features and gait features to generate the maximum number of concatenated feature vectors based on the characteristics of face and gait. Specifically, four concatenated features are constructed based on two face features and two gait features for one person from each video.

Let $\mathbf{V}_i$, $i = 1, 2, \ldots, c$, the mean of the training synthetic features of class i, be the prototype of class i. The unknown person is classified to class K to whom the synthetic feature $\mathbf{p}$ is the nearest neighbor:

$$\|\mathbf{p} - \mathbf{V_K}\| = \min \|\mathbf{p} - \mathbf{V}_i\|. \tag{13}$$

When multiple synthetic features are obtained for one person, Eq. (13) means that the unknown person is classified to the class which has the minimum distance out of all the distances corresponding to all the classes.

4. Experimental results

4.1. Experiments and parameters

We collect 100 video sequences of 45 people using a Sony DCR-VX1000 digital video camera recorder operating at 30 frames per second. Each video sequence includes only one subject. The subject is walking in outdoor condition and expose a side view to the camera. The number of sequences per subject varies from 2 to 3. The resolution of each frame is 720 × 480. The distance between people and the video camera is about 10 feet. Fig. 5 shows some examples of the data.

We perform two experiments using two comparative data sets to test our approach and show the effect of changing clothes and changing face over time on the performance of the fusion schemes. In Experiment 1, the data consists of 90 video sequences of 45 people. Each person has two video sequences, one for the training and the other one for the testing. For the same person, the clothes are the same in the training sequence and the testing sequence. In Experiment 2, the data consists of 90 video sequences of 45 people. Each person has two video sequences, one for the training and the other one for the testing. For 10 of 45 people, the clothes are different in the training sequences and the testing sequences, and the data are collected on two separate days that are about 1 month apart. For the other 35 people, the clothes are the same in the training sequences and the testing sequences. Experiments are conducted on a 2 Dual Core AMD Opteron Processors 265 1.8 GHz Linux machine with 2 GB RAM. The simulation language is Matlab. It takes 26 s to recognize 45 people based on their ESFIs and GEIs. The computational complexity of the approach grows $O(n^3)$, where n is the number of the training samples.

Recognition performance is used to evaluate our method in the two experiments. For a video sequence, it is defined as the ratio of the number of the correctly recognized people to the number of all the people. To analyze the performance of our method more insightfully, we also provide the error index that gives the numbers of misclassified sequences. CMC curve is used to further evaluate the performance of the systems. The CMC curve returns identities associated with the K highest-scoring biometrics samples from the training data. For x axis, K rank means the K nearest neighbor method is considered for the recognition results. For y axis, the accuracy rate means the frequency when the genuine identities are included in the K nearest neighbors. The lower the rank of the genuine matching biometrics in the training data, the better the performance of identification system. Improved algorithms would result in a better CMC curve, one that would run more toward the upper left corner of the plot. For comparison, we also show the performance using face features from the OSFIs to demonstrate the performance improvement by using constructed ESFIs. The resolution of OSFI is 34 × 18. The procedures of feature extraction, synthetic feature generation and classification are the same for ESFI and OSFI. Furthermore, the proposed feature fusion scheme is compared with the single biometrics scheme where MDA is applied to the PCA features of the single biometrics, the feature level fusion scheme [2] and

Fig. 5. Two examples of video sequences.

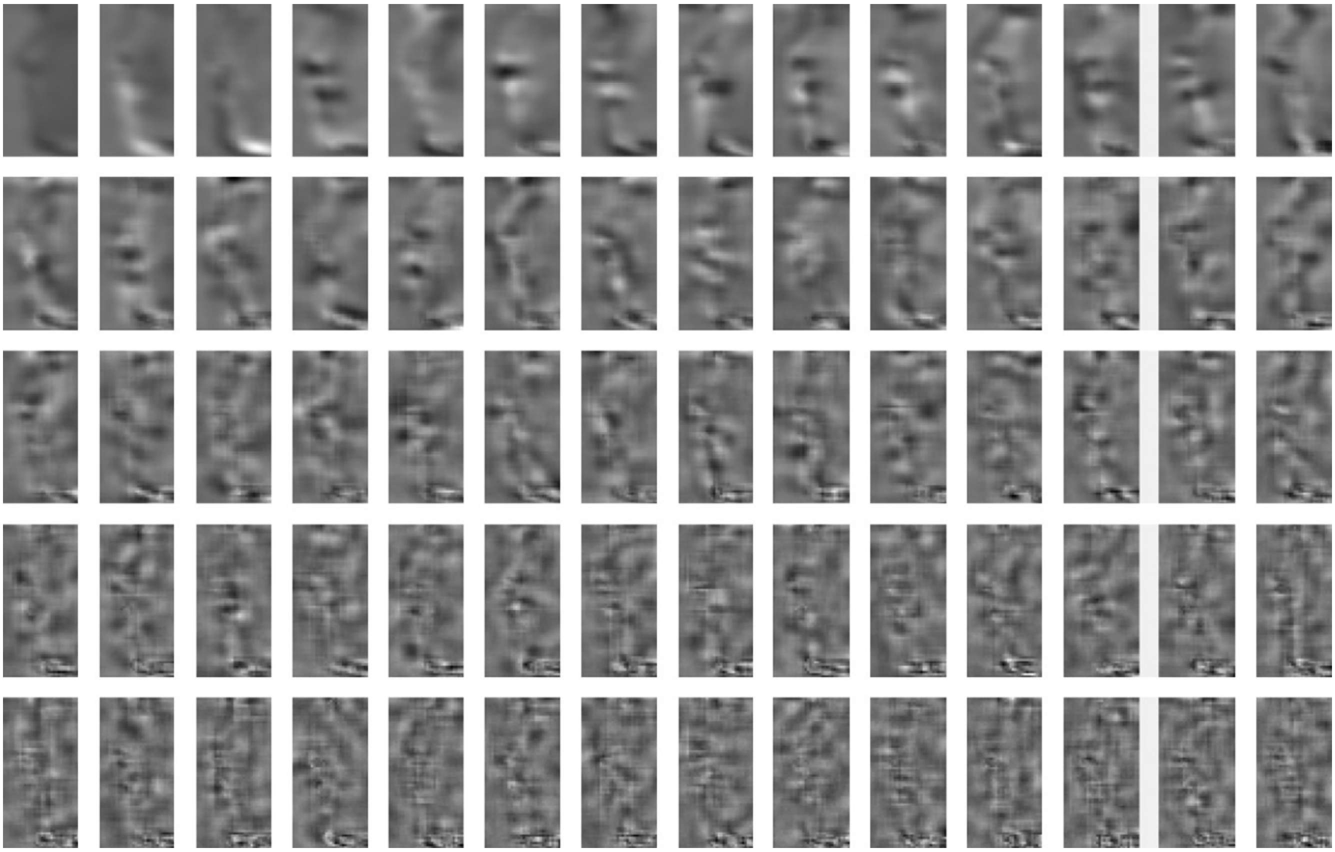

Fig. 6. The top 70 eigenvectors of face (from left to right and top to bottom).

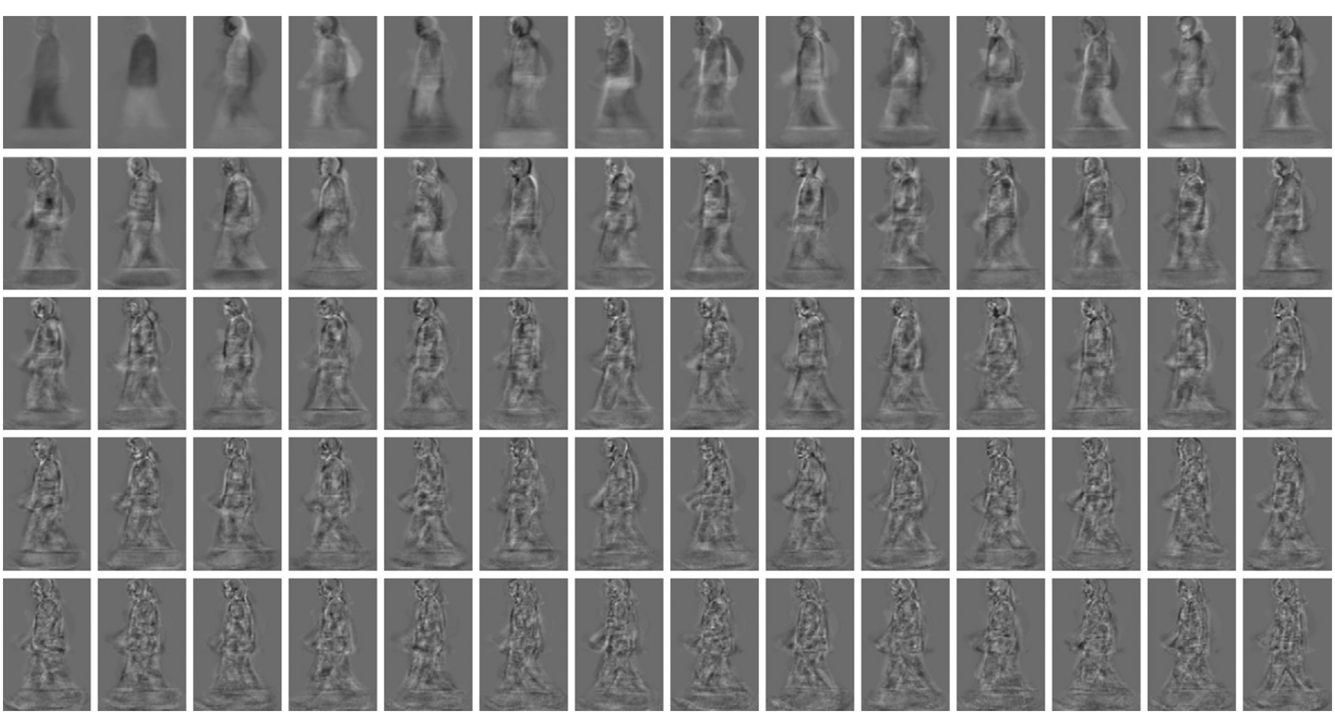

Fig. 7. First 70 eigenvectors of gait (from left to right and top to bottom).

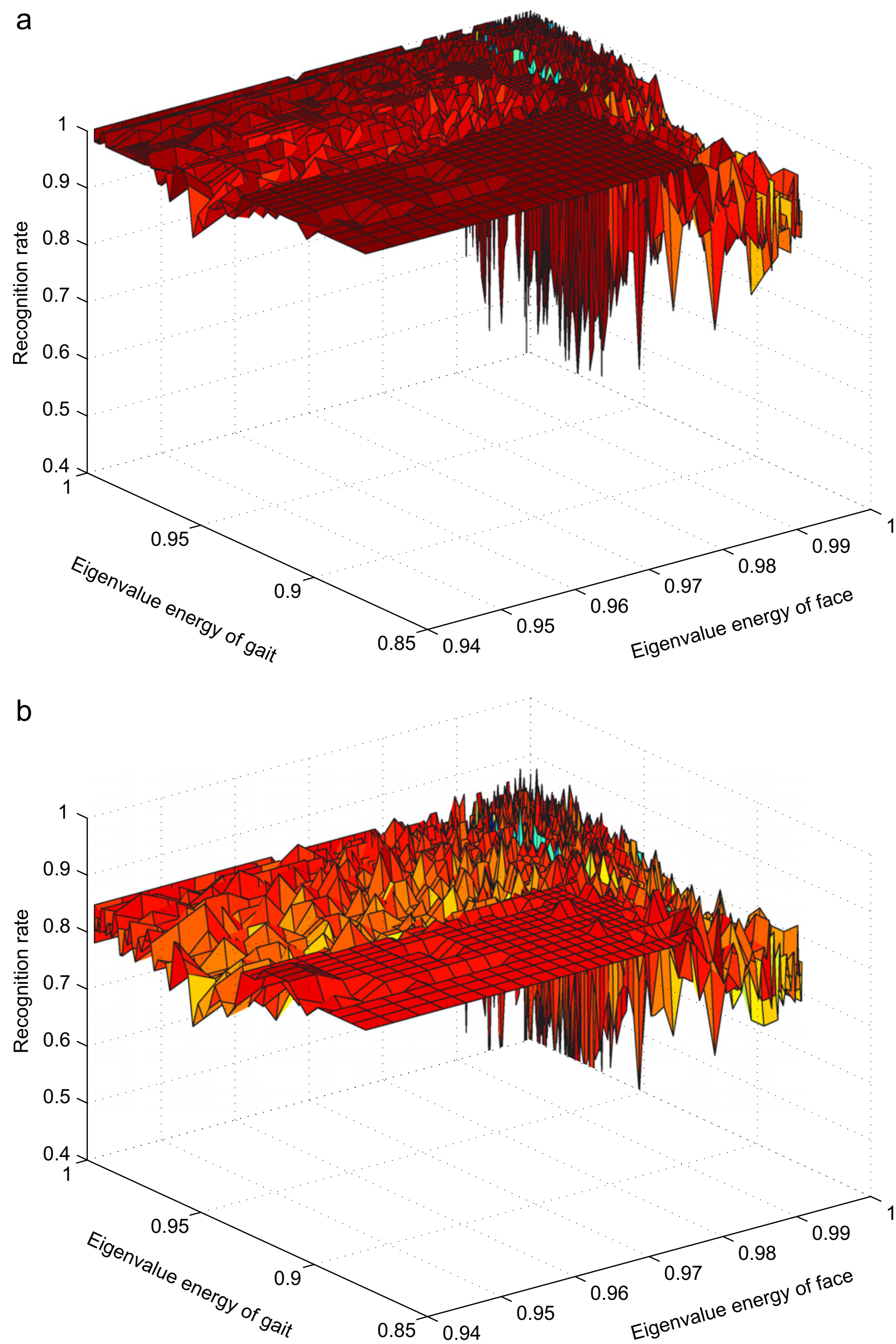

Fig. 8. (a) Performance vs. number of face and gait features based on GEI and ESFI in Experiment 1. (b) Performance vs. number of face and gait features based on GEI and ESFI in Experiment 2.

the match score level fusion schemes using Sum and Max rules [17] as explained in Section 3.

For gait, two complete walking cycles are obtained from a video sequence according to the gait frequency and gait phase. Each walking cycle consists of 20 frames. We construct two GEIs corresponding to two walking cycles from one video sequence. The resolution of each GEI is 300×200. For face, we also construct two high-resolution side face images from one video sequence. Each high-resolution side face image is built from 10 low-resolution side face images that are extracted from adjacent video frames. The resolution of low-resolution side face images is 68×68 and the resolution of reconstructed high-resolution side face images is 136×136. After normalization as indicated in Section 2.1, the resolution of ESFI is 64×32.

Table 3
Experiment 1: Single biometrics performance and error index of individuals

Performance	Biometrics		
	Original face (OSFI)	Enhanced face (ESFI)	Gait (GEI)
Recognition rate (%)	73.3	91.1	93.3
Error index	1, 6, 10, 12, 14, 18, 20, 22, 26, 28, 42, 43	13, 16, 21, 35	4, 15, 26

Table 4
Experiment 1: Fused biometrics performance and error index of individuals

Fusion method	Match score level [17]		Feature level	
	Sum rule	Max rule	Fusion scheme [2]	Fusion scheme (this paper)
OSFI & GEI				
Recognition rate (%)	93.3	93.3	97.8	97.8
Error index	4, 10, 26	4, 10, 26	26	6
ESFI & GEI				
Recognition rate (%)	95.6	97.8	100	100
Error index	4, 26	26	None	None

For the proposed feature fusion scheme, the dimensionality of the synthetic features is 44 ($c-1$), which result from applying MDA transformation to the concatenated face and gait features. The selection of eigenvectors as face features and gait features is based on both the observation and the energy criteria. Figs. 6 and 7 show the top 70 eigenvectors of face and gait, respectively. The order of eigenvectors corresponds to the descending order of the eigenvalues. The higher numbered eigenvectors seem more blotchy and it becomes more and more difficult to discern the semantics of what they are encoding. This indicates that eliminating these eigenvectors from the eigenspace should have a minimal effect on performance [22]. Meanwhile, the remaining eigenvectors should satisfy the requirement that the corresponding eigenvalues have 99% of the total energy. Furthermore, we decide to keep no more than two-thirds of the total eigenvectors to reduce the problem of curse of dimensionality. In both experiments, we retain eigenvectors corresponding to the top 59 eigenvalues as face features and the top 56 eigenvalues as gait features. In practice, the dimensionality of face and gait features may influence the performance of the proposed method.

Fig. 8 shows the performance of the proposed feature fusion scheme corresponding to the different number of the face and gait features in Experiments 1 and 2. We can see that the performance fluctuates as the number of face and gait features changes. The optimal fusion performance is achieved somewhere when the eigenvectors corresponding to the eigenvalues have between 98.5% and 99.5% of the total energy for face and gait. There is more than one optimal choice in number of the face and gait features. It is also clear that the fusion performance degrades abruptly when the number of chosen eigenvectors is below some threshold. Fig. 8 verifies the reasonableness of our choice for face and gait features.

4.1.1. Experiment 1

Forty-five people are named from 1 to 45 and each of them has two video sequences. Two GEIs and two ESFIs are constructed for each sequence. Therefore, as explained in Section 2.3.2, four synthetic features are generated based on two face features and two gait features for one person from each video. Totally, we have 180 synthetic features corresponding to 45 people in the gallery and 180 synthetic features corresponding to 45 people in the probe. Table 3 shows the performance of single biometrics. Table 4 shows the performance of fusion using different schemes. In Tables 3 and 4, the error index gives the number of misclassified sequence.

Table 3 shows that 73.3% people are correctly recognized by OSFI, 91.1% people are correctly recognized by ESFI and 93.3% people are correctly recognized by GEI. The changes of the body shape and the walking style causes gait recognition errors. We show GEIs of the people who are misclassified by the gait classifier in Fig. 9. Face is sensitive to noise as well as facial expressions, so the different condition in the training sequence and the testing sequence affects its reliability. We show ESFIs of the people who are misclassified by the face classifier in Fig. 10. Among fusion performance of ESFI and GEI in Table 4, the proposed feature fusion approach has the same performance as [2] at the best recognition rate of 100%, followed by the Max rule at 97.8% and the Sum rule at 95.6%. Fig. 11 shows people (video sequences) misclassified by integrating ESFI and GEI using the Sum and Max rules. It is clear that both of the match score level fusion schemes using Sum and Max rules misclassify the person (26), but both of the feature level fusion schemes recognize the person correctly. For fusion based on OSFI and GEI, the best performance is also achieved by the proposed feature fusion approach and [2] at 97.8%, followed by the Sum rule and the Max rule at 93.3%. Fig. 12 shows the CMC curves of Experiment 1. The CMC

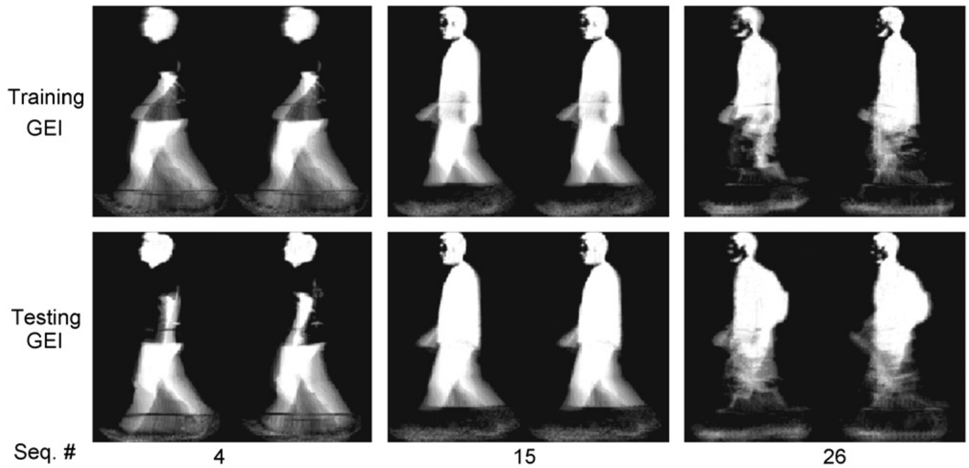

Fig. 9. Experiment 1: GEIs of people misclassified by the gait classifier (see Table 3). For each person, two GEIs of the training video sequence and two GEIs of the testing video sequence are shown for comparison.

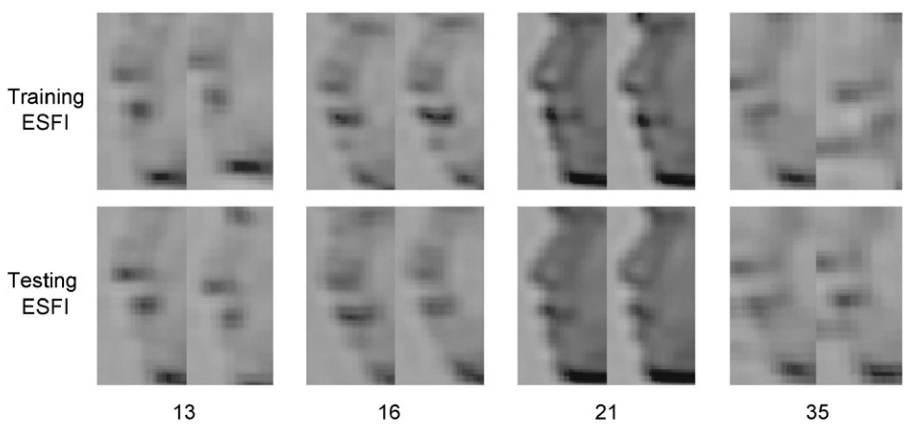

Fig. 10. Experiment 1: ESFIs of people misclassified by the face classifier (see Table 3). For each person, two ESFIs of the training video sequence and two ESFIs of the testing video sequence are shown for comparison.

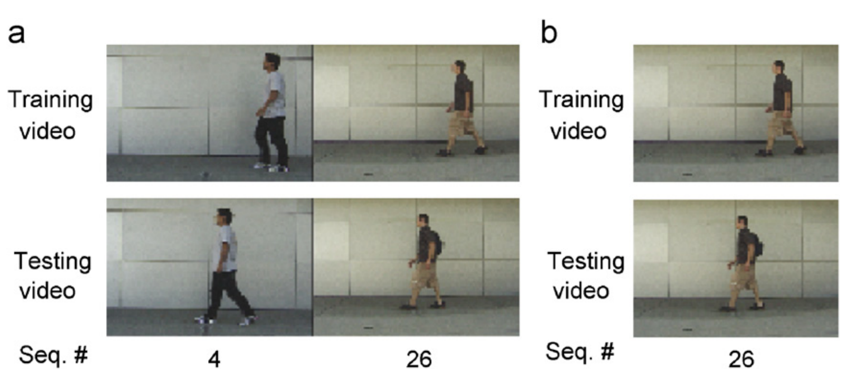

Fig. 11. Experiment 1: People misclassified by the integrated classifier based on ESFI and GEI using different fusion rules (see Table 4). For each person, one frame of the training video sequence and one frame of the testing video sequence are shown for comparison: (a) errors by the Sum rule and (b) errors by the Max rule.

curve of the proposed feature fusion scheme overlaps with the CMC curve of the approach [2]. Both of them are more toward the upper left corner of the plots compared with that of the match score fusion schemes. It is clear that the feature level fusion schemes are more effective than the match score level fusion schemes. Fusion based on ESFI and GEI always has better performance than fusion based on OSFI and GEI using the same fusion scheme. It also demonstrates that the synthetic

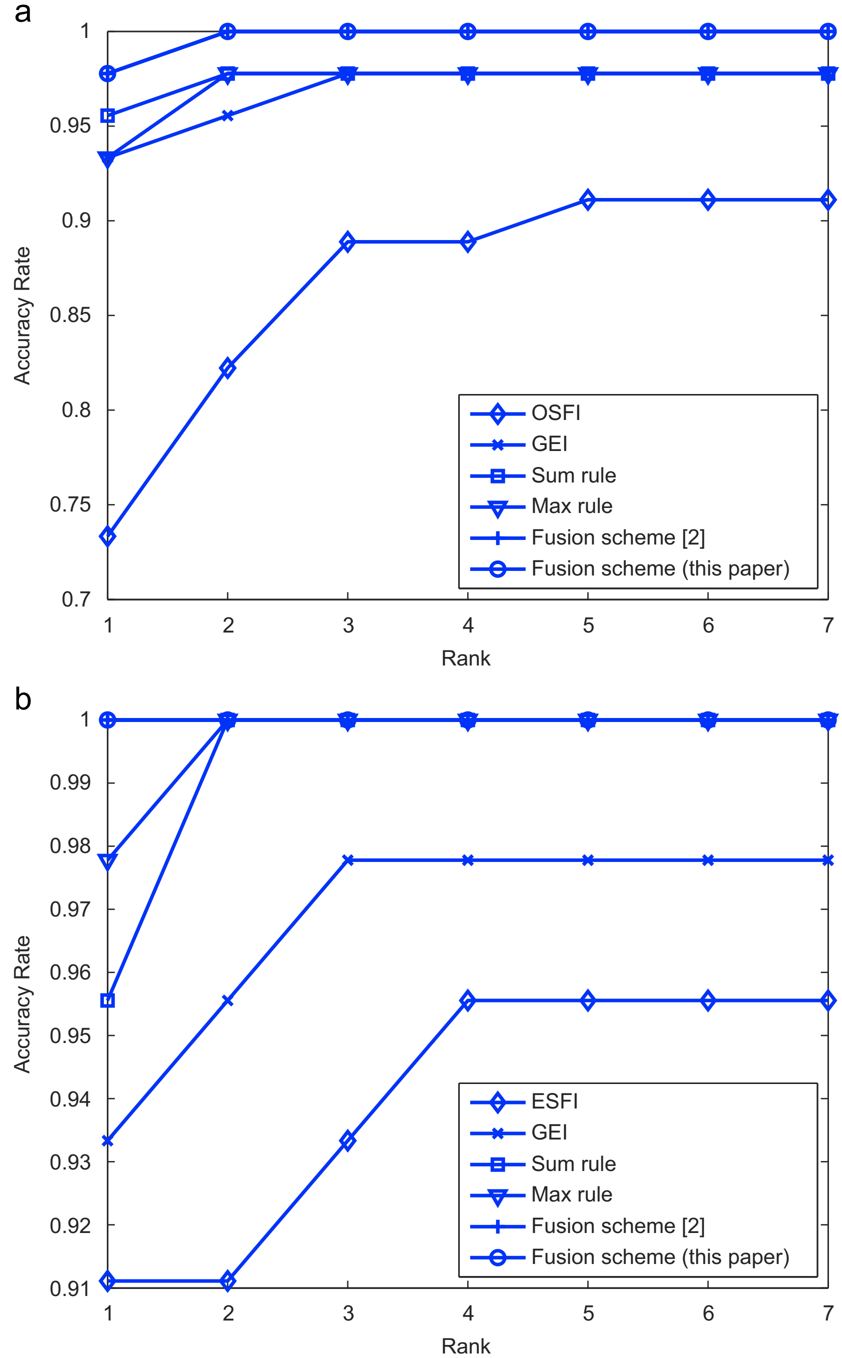

Fig. 12. Experiment 1: (a) CMC curves of the classifiers using GEI and OSFI. (b) CMC curves of the classifiers using GEI and ESFI.

features carry more discriminating power than the individual biometrics features.

4.1.2. Experiment 2

In Experiment 2, we use the same order as in Experiment 1 to name 45 people. Ten testing video sequences are substituted with the other 10 testing video sequences. The order of the 10 replaced testing video sequences are {1, 2, 5, 6, 8, 9, 10, 13, 19, 40}. Consequently, 10 out of 45 people in Experiment 2 wear different clothes in the training sequences and the testing sequences. We also construct two GEIs and two ESFIs from each sequence and therefore, as explained in Section 2.3.2, generate four synthetic features based on two face features and two gait features for one person from each video. Totally, we have 180 synthetic features corresponding to 45 people in the gallery and 180 synthetic features corresponding to 45 people in the probe. Table 5 shows the performance of single biometrics. Table 6 shows the performance of fusion using different schemes. In Tables 5 and 6, the error index gives the number of misclassified sequence.

Table 5
Experiment 2: Single biometrics performance and error index of individuals

Performance	Biometrics		
	Original face (OSFI)	Enhanced face (ESFI)	Gait (GEI)
Recognition rate (%)	64.4	80	82.2
Error index	1, 2, 5, 6, 8, 9, 13, 18, 19, 20, 26, 28, 34, 40, 42, 43	1, 2, 5, 8, 11, 13, 30, 35, 42	2, 5, 6, 8, 13, 19, 26, 40

Table 6
Experiment 2: Fused biometrics performance and error index of individuals

Fusion method	Match score level [17]		Feature level	
	Sum rule	Max rule	Fusion scheme [2]	Fusion scheme (this paper)
OSFI & GEI				
Recognition rate (%)	82.2	82.2	84.4	86.7
Error index	2, 5, 6, 8, 13, 19, 26, 40	2, 5, 6, 8, 13, 19, 26, 40	1, 2, 5, 8, 13, 19, 40	1, 2, 5, 6, 8, 19
ESFI & GEI				
Recognition rate (%)	88.9	88.9	88.9	91.1
Error index	2, 5, 6, 8, 13	2, 5, 6, 8, 13	2, 5, 8, 13, 19	2, 5, 6, 13

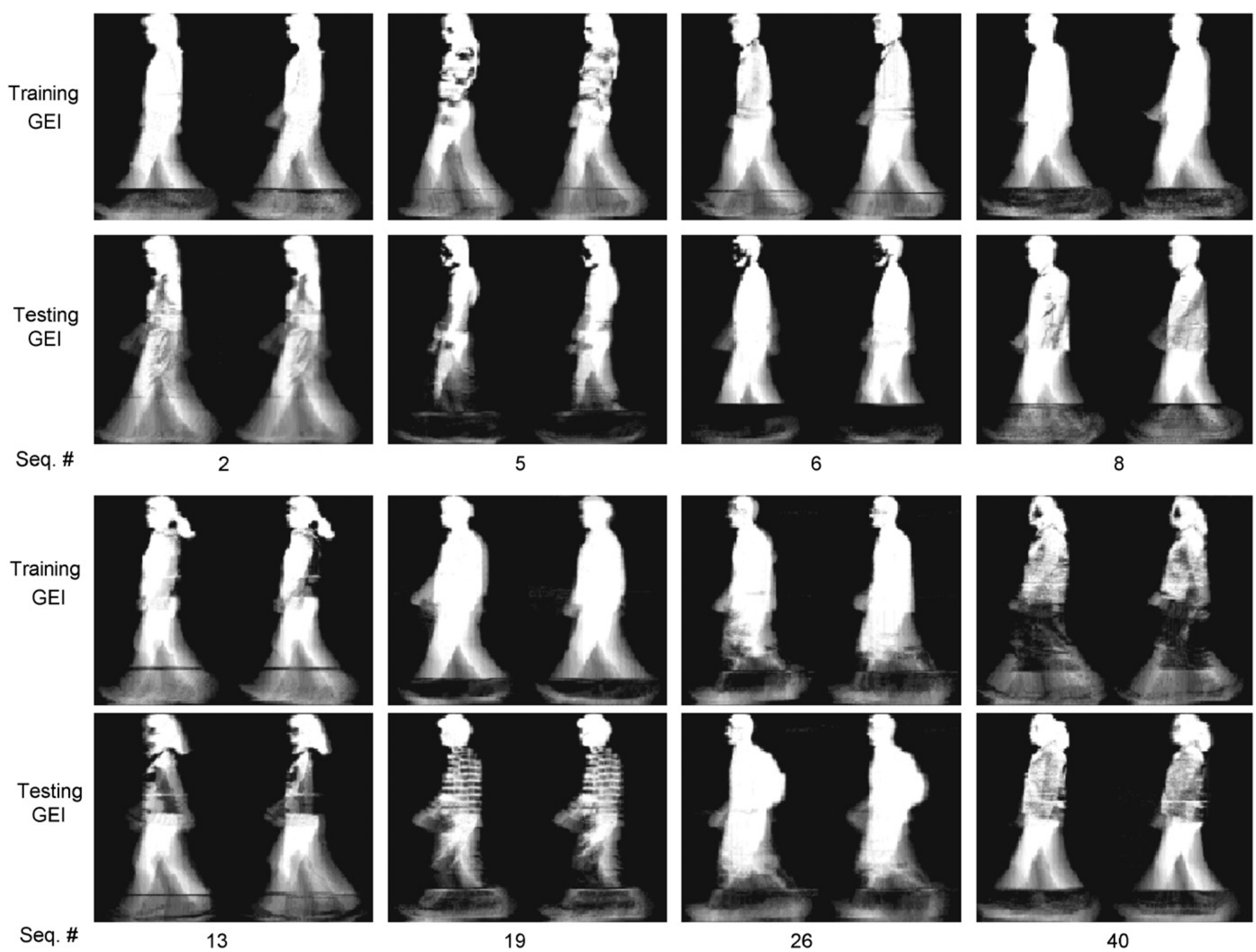

Fig. 13. Experiment 2: GEIs of people misclassified by the gait classifier (see Table 5). For each person, two GEIs of the training video sequence and two GEIs of the testing video sequence are shown for comparison.

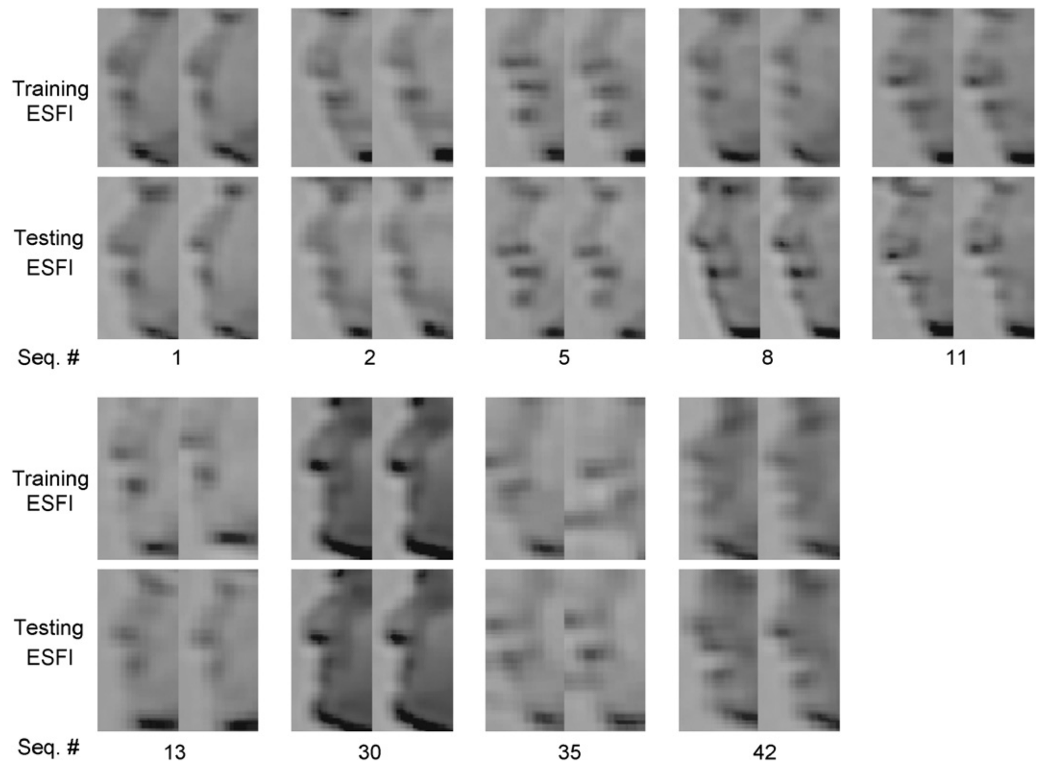

Fig. 14. Experiment 2: ESFIs of people misclassified by the face classifier (see Table 5). For each person, two ESFIs of the training video sequence and two ESFIs of the testing video sequence are shown for comparison.

Table 5 shows that 64.4% people are correctly recognized by OSFI, 80% people are correctly recognized by ESFI and 82.2% people are correctly recognized by GEI. Compared with the performance of individual biometrics in Experiment 1 in Table 3, all the performance of individual biometrics in Experiment 2 decreases to some extent. It is reasonable since the changes in the lighting conditions and the color of clothes cause changes in the segmentation of the human body. Also, changing clothes causes the difference in the shape of the training sequence and the testing sequence for the same person. Fig. 13 shows GEIs of the people who are misclassified by the gait classifier. Meanwhile, since face is sensitive to noise as well as facial expressions, the different conditions in the two video sequences that are taken at least one month apart, brings more face recognition errors. Fig. 14 shows ESFIs of the people who are misclassified by the face classifier.

For the fusion performance based on ESFI and GEI in Table 6, the proposed feature fusion approach achieves the best performance at 91.1%. The approach [2] has the same performance as the Sum rule and the Max rule at 88.9% [17]. We can see that a larger improvement of fusion performance is achieved by the proposed feature level fusion scheme compared with the other fusion schemes. Fig. 15 shows the people (video sequences) misclassified by integrating ESFI and GEI using different fusion rules. For fusion based on OSFI and GEI, the best performance is also achieved by the proposed feature fusion approach at 86.7%, followed by Zhou and Bhanu [2] at 84.4%, and the Sum rule and the Max rule at 82.2% [17]. Fig. 16 shows the CMC curves of Experiment 2. In Fig. 16(a), it is clear that the CMC curve of the proposed feature fusion scheme has the better performance than any other scheme. In Fig. 16(b), the accuracy rate of the proposed feature fusion scheme is lower than that of the Max rule fusion scheme at rank 4 and 5, but for the other ranks, the accuracy rate of the proposed feature fusion scheme is higher than or equal to that of the Max rule fusion scheme. Specifically, the highest accuracy rates are achieved by the proposed feature fusion scheme at rank 1 and 2, which demonstrates the better performance than other fusion schemes since the accuracy rates at low ranks are more important for a recognition system.

4.2. Discussion on experiments

The experimental results in Experiments 1 and 2 clearly demonstrate the importance of constructing ESFI. From ESFI, we can extract face features with more discriminating power. Therefore, better performance is achieved when ESFI instead of OSFI is used for all of the fusion schemes. For example, in Experiment 2, OSFI has bad performance at 64.4%, but ESFI still achieves the recognition rate of 80%. The proposed feature fusion scheme based on ESFI and GEI achieves the performance improvement of 8.9% (from 82.2% to 91.1%), while the improvement is 4.5% (from 82.2% to 86.7%) for fusion of OSFI and GEI . The results demonstrate that ESFI serves as a better face template than OSFI. The synthetic features obtained from ESFI and GEI capture more discriminating power

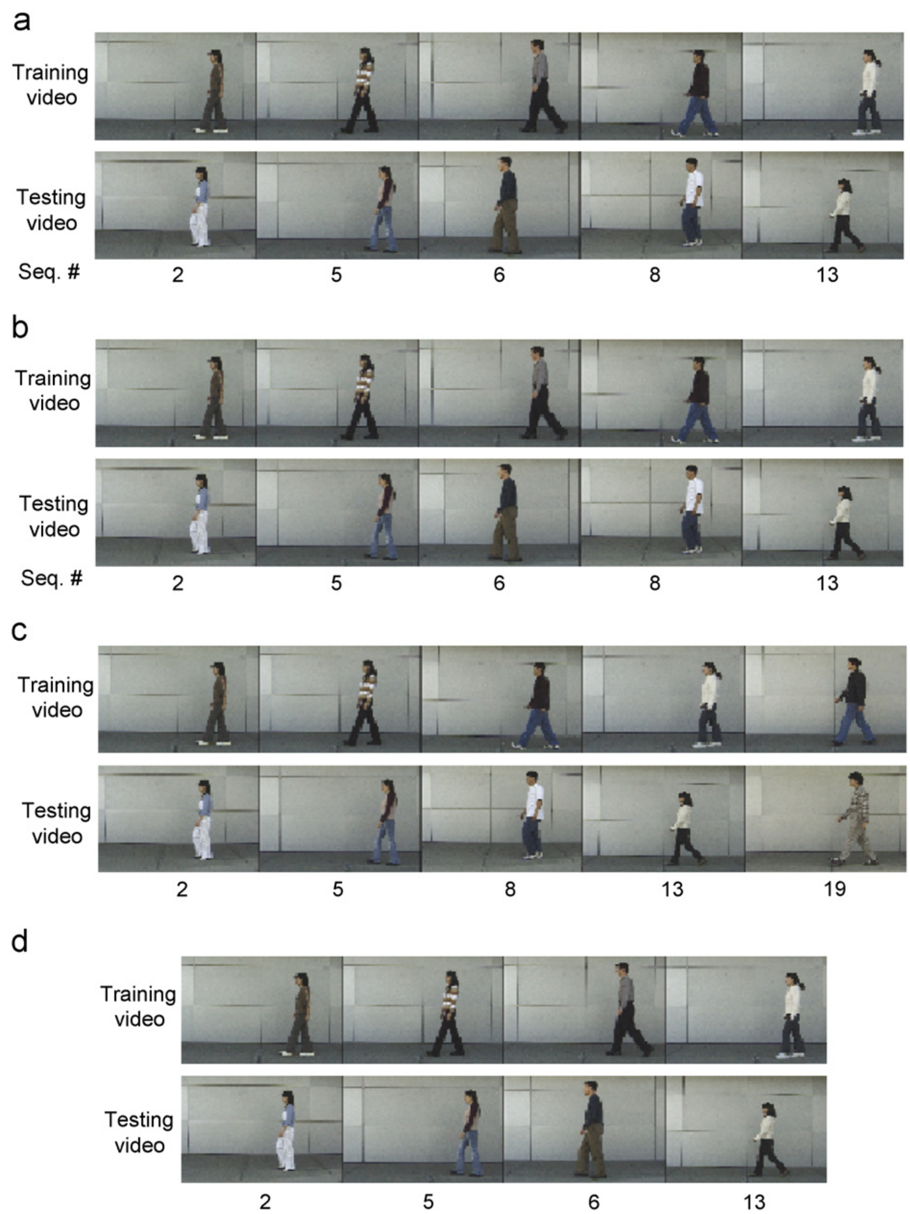

Fig. 15. Experiment 2: people misclassified by the integrated classifier based on ESFI and GEI using different fusion rules (see Table 6). For each person, one frame of the training video sequence and one frame of the testing video sequence are shown for comparison: (a) errors by the Sum rule; (b) errors by the Max rule; (c) errors by the fusion scheme [2]; and (d) errors by the proposed fusion scheme.

than that from OSFI and GEI. Consequently, the fusion based on ESFI and GEI always has better performance than fusion based on OSFI and GEI using the same fusion scheme.

In Experiment 1, the proposed feature level fusion scheme outperforms the match score level fusion schemes but has the same performance as the feature level fusion scheme [2]. For the more difficult database in Experiment 2, we can see that the proposed feature level fusion scheme outperforms all the other fusion schemes. The feature level fusion scheme [2] does not perform better than the match score level fusion schemes. Moreover, the proposed feature level fusion scheme achieves a larger performance improvement in Experiment 2 compared with the improvement in Experiment 1. Specifically, compared with the performance achieved by gait (the better performance of the two individual biometrics), the proposed scheme has an improvement of 6.7% in Experiment 1 and 8.9% in Experiment 2. All these results demonstrate the effectiveness of integrating face and gait information for human recognition using the proposed feature level fusion scheme since it outperforms the other fusion schemes and even achieves a larger improvement for the more challenging database. Furthermore, the results in both experiments indicate that the proposed feature level fusion scheme does not depend on specific features since it achieves the best fusion performance in both of cases: fusion of OSFI and GEI, and fusion of ESFI and GEI.

When Experiments 1 and 2 are compared, it is clear that the recognition rates in Experiment 2 decrease compared with Experiment 1 because of 10 out of 45 people changing their clothes

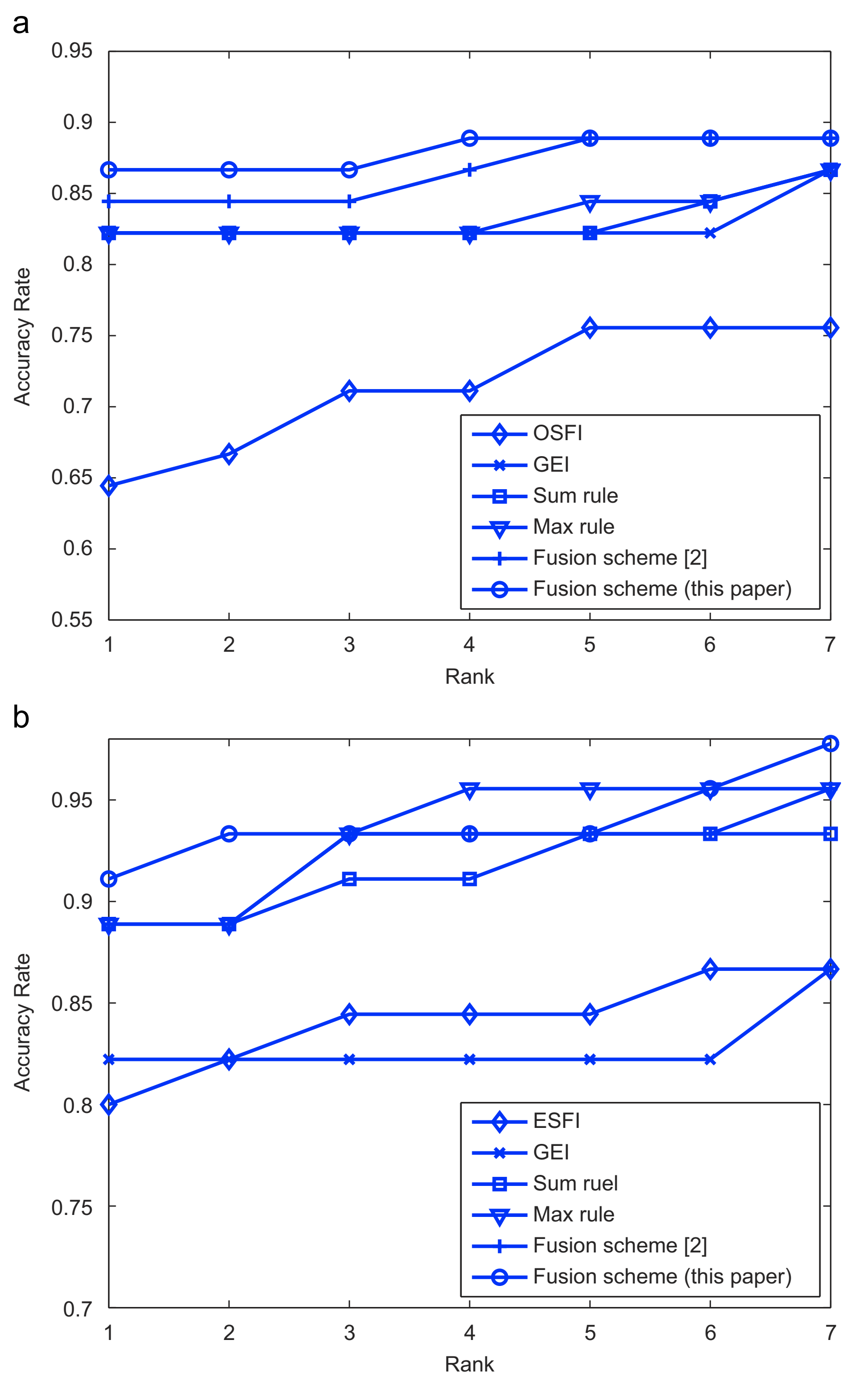

Fig. 16. Experiment 2: (a) CMC curves of the classifiers using GEI and OSFI. (b) CMC curves of the classifiers using GEI and ESFI.

in the testing sequences. As explained before, gait recognition based on GEI is not only affected by the walking style of a person, but also by the shape of human body. See Figs. 9 and 13 as examples. Face is sensitive to noise as well as facial expressions, so the different condition in the training sequence and the testing sequence affects its reliability. See Figs. 10 and 14 as examples. All these factors contribute to recognition errors of the individual classifiers. However, the fusion system based on side face and gait overcomes this problem to some extent. For example, in Experiment 2, people {2,5,6,8,13,19,26,40} are not correctly recognized by gait and the performance of gait classifier is 82.2%, but when side face information is integrated, the recognition rate is improved to 91.1%. It is because the clothes or the walking style of these people are different between the training and testing video sequences, so the gait classifier cannot recognize them correctly. However, the side face of these people does not change so much in the training and testing sequences, and it brings useful information for the fusion system and corrects some errors. On the other hand, since the face classifier is comparatively sensitive to the variation of facial expressions and noise, it cannot get a good recognition rate by itself. When gait information is combined, the better performance is achieved. For example, in Experiment 2, people {1,2,5,8,11,13,30,35,42} are not correctly recognized by face and the performance of face classifier is 80%, but when gait information is integrated, the recognition rate is improved to 91.1%. This demonstrates that the fusion system using side face and gait is effective for individual recognition in video since face and gait are two complementary biometrics. Consequently, our fusion system is relatively robust compared with the system using only one biometrics in the same scenario.

These results also demonstrate that the match score fusion cannot rectify the misclassification conducted by both of the face classifier and the gait classifier. For the match score fusion, people misclassified by the individual classifiers are likely to be classified correctly after fusion on the condition that there is at least one of the two classifiers that works correctly. For example, in Table 5, there are four misclassified people {2, 5, 8, 13} overlapped between classification using ESFI only and GEI only. From Table 6, we can see that the set of misclassified people {2, 5, 8, 13} are always a subset of the error indices when ESFI and GEI are combined by Sum and Max rules. However, the classifier using the synthetic features can rectify the misclassification conducted by both of the individual classifier. For example, the proposed feature level fusion scheme based on ESFI and GEI correctly recognizes the person (8) who is misclassified by both the face classifier and the gait classifier individually. It is clear that the performance of the feature level fusion mainly depends on the fused feature set while the performance of the match score level fusion mainly depends on the results of the individual biometrics classifiers. Since the fused feature set contains richer information about the input biometrics pattern than the match score, the feature level fusion is more effective than the match score level fusion when individual biometrics features are appropriately combined. Moreover, even though the proposed feature level fusion scheme has the same performance as the feature level fusion scheme [2] in Experiment 1, it outperforms the feature level fusion scheme [2] in Experiment 2, which is a more challenging case. For the feature level fusion scheme [2], the face and gait features are simply concatenated and the relationship between them are not known. For the proposed feature level fusion scheme, MDA is applied to the concatenated features of face and gait, and it improves the discriminating power of the synthetic features. The proposed feature level fusion scheme is more effective than the feature level fusion scheme [2] for the recognition task. Though the comparison of the potential to correct errors between different fusion schemes is analyzed based on the experimental results in this paper, the conclusion is applicable to other biometrics and sources of data.

5. Conclusions

The fusion of face and gait is promising in real world application because of their individual characteristics. Compared with gait, face images are readily interpretable by humans, which allows people to confirm whether a biometrics system is functioning correctly, but the appearance of a face depends on many factors: incident illumination, head pose, facial expressions, moustache/beard, eyeglasses, cosmetics, hair style, weight gain/loss, aging, and so forth. Although gait images can be easily acquired from a distance, the gait recognition is affected by clothes, shoes, carrying status and specific physical condition of an individual. The fusion system is relatively more robust compared with the system that uses only one biometrics. For example, face recognition is more sensitive to low lighting conditions, whereas gait is more reliable under these conditions. Similarly, when the walker is carrying a heavy baggage or he/she is injured, the captured face information may contribute more than gait.

In this paper, a new feature level fusion scheme is proposed to integrate information from side face and gait for recognizing individuals at a distance in video. ESFI and GEI, both of which integrate information over multiple frames in video, work as face template and gait template, respectively. Multiple discriminant analysis (MDA) is applied after the concatenation of face and gait features to generate discriminating features for improved recognition performance. The problem of curse of dimensionality is reduced since the final feature vectors used in this paper are of lower dimension than those in Ref. [2]. The experimental results show that the proposed feature level fusion scheme is effective for individual recognition in video. It outperforms the previously published fusion schemes at the match score level [17] and the feature level [2] for face- and gait-based human recognition at a distance in video.

References

[1] A.A. Ross, K. Nandakumar, A.K. Jain, Handbook of Multibiometrics, Springer, Berlin, 2006.
[2] X. Zhou, B. Bhanu, Feature fusion of face and gait for human recognition at a distance in video, in: Proceedings of the IEEE International Conference on Pattern Recognition, 2006.
[3] G. Feng, K. Dong, D. Hu, D. Zhang, When faces are combined with palmprints: a novel biometric fusion strategy, in: International Conference on Biometric Authentication, 2004, pp. 701–707.
[4] T. Kinnunen, V. Hautamaki, P. Franti, Fusion of spectral feature sets for accurate speaker identification, in: Proceedings of the International Conference on Speech and Computer, September 2004, pp. 361–365.
[5] A. Kumar, D.C.M. Wong, H.C. Shen, A.K. Jain, Personal verification using palmprint and hand geometry biometric, in: Proceedings of the Audio- and Video-based Biometric Person Authentication, 2003, pp. 668–678.
[6] A.A. Ross, R. Govindarajan, Feature level fusion of hand and face biometrics, in: Proceedings of the SPIE Conference on Biometric Technology for Human Identification II, March 2005, pp. 196–204.
[7] J. Yang, J.Y. Yang, D. Zhang, J.F. Lu, Feature fusion: parallel strategy vs. serial strategy, Pattern Recognition 38 (6) (2003) 1369–1381.
[8] Y.S. Moon, H.W. Yeung, K.C. Chan, S.O. Chan, Template synthesis and image mosaicking for fingerprint registration: an experimental study, in: IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 5, 2004, pp. 409–412.
[9] Y. Gao, M. Maggs, Feature-level fusion in personal identification, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2005.
[10] A. Kong, D. Zhang, M. Kamel, Palmprint identification using feature-level fusion, Pattern Recognition 39 (3) (2006) 478–487.
[11] Q. Li, Z. Qiu, Handmetric verification based on feature-level fusion, J. Comput. Sci. Network Security 6 (2A) (2006).

[12] A. Kale, A. Roy-chowdhury, R. Chellappa, Fusion of gait and face for human identification, in: Proceedings of Acoustics, Speech, and Signal Processing, vol. 5, 2004, pp. 901–904.
[13] G. Shakhnarovich, L. Lee, T. Darrell, Integrated face and gait recognition from multiple views, in: Proceedings of Computer Vision and Pattern Recognition, vol. 1, 2001, pp. 439–446.
[14] G. Shakhnarovich, T. Darrell, On probabilistic combination of face and gait cues for identification, in: Proceedings of Automatic Face and Gesture Recognition, vol. 5, 2002, pp. 169–174.
[15] B. Bhanu, X. Zhou, Face recognition from face profile using dynamic time warping, in: Proceedings of the International Conference on Pattern Recognition, vol. 4, 2004, pp. 499–502.
[16] X. Zhou, B. Bhanu, Integrating face and gait for human recognition, in: Proceedings of the Workshop on Biometrics Held in Conjunction with the IEEE Conference on Computer Vision and Pattern Recognition, 2006, pp. 55–62.
[17] X. Zhou, B. Bhanu, Integrating face and gait for human recognition at a distance in video, IEEE Trans. Systems Man Cybernet. Part B: Cybernet, 2007.
[18] J. Han, B. Bhanu, Individual recognition using gait energy image, IEEE Trans. Pattern Anal. Mach. Intell. 28 (2) (2006) 316–322.
[19] X. Zhou, B. Bhanu, J. Han, Human recognition at a distance in video by integrating face profile and gait, in: Proceedings of the Audio- and Video-based Biometric Person Authentication, 2005, pp. 533–543.
[20] R.C. Gonzalez, R.E. Woods, Digital Image Processing, Addison Wesley Publishing Company, Reading, MA, 1992.
[21] S. Theodorids, K. Koutroumbas, Pattern Recognition, Academic Press, New York, 1998.
[22] W.S. Yambor, B.A. Draper, J.R. Beveridge, Analyzing PCA-based face recognition algorithms: eigenvector selection and distance measures, in: H. Christensen, J. Phillips (Eds.), Empirical Evaluation Method in Computer Vision, World Scientific Press, Singapore, 2002.

About the Author—XIAOLI ZHOU received the B.S. and M.S. degrees in Electrical Engineering from Beijing University of Posts and Telecommunications, China, in 1998 and 2001, respectively. She received the Ph.D. degree in Electrical Engineering from the University of California, Riverside (UCR) in 2007. She is working on her research at the Center for Research in Intelligent Systems at UCR. Her research interests are in computer vision, pattern recognition and image processing. Her recent research has been concerned with fusion of biometrics for human recognition at a distance in video.

About the Author—BIR BHANU received the S.M. and E.E. degrees in Electrical Engineering and Computer Science from the Massachusetts Institute of Technology, Cambridge, the Ph.D. degree in Electrical Engineering from the Image Processing Institute, University of Southern California, Los Angeles, and the M.B.A. degree from the University of California, Irvine. He was the founding Professor of Electrical Engineering and served its first Chair at the University of California at Riverside (UCR). He has been the Cooperative Professor of Computer Science and Engineering and Director of Visualization and Intelligent Systems Laboratory (VISLab) since 1991. Currently, he also serves as the founding Director of an interdisciplinary Center for Research in Intelligent Systems (CRIS) at UCR. Previously, he was a Senior Honeywell Fellow at Honeywell Inc., Minneapolis, MN. He has been on the faculty of the Department of Computer Science at the University of Utah, Salt Lake City, and has worked at Ford Aerospace and Communications Corporation, CA, INRIA-France, and IBM San Jose Research Laboratory, CA. He has been the Principal Investigator of various programs for NSF, DARPA, NASA, AFOSR, ARO, and other agencies and industries in the areas of video networks, video understanding, learning and vision, image understanding, pattern recognition, target recognition, biometrics, navigation, image databases, and machine vision applications. He is the Coauthor of books on *Computational Learning for Adaptive Computer Vision* (New York: Springer-Verlag, 2007), *Evolutionary Synthesis of Pattern Recognition Systems* (New York: Springer-Verlag, 2005), *Computational Algorithms for Fingerprint Recognition* (Norwell, MA: Kluwer, 2004), *Genetic Learning for Adaptive Image Segmentation* (Norwell, MA: Kluwer, 1994), and *Qualitative Motion Understanding* (Norwell, MA; Kluwer, 1992), and the Co-editor of a book on *Computer Vision Beyond the Visible Spectrum*, (New York: Springer-Verlag, 2004). He holds 11 US and international patents and over 250 reviewed technical publications in the areas of his interest.
Dr. Bhanu has received two outstanding paper awards from the Pattern Recognition Society and has received industrial and university awards for research excellence, outstanding contributions, and team efforts. He has been on the editorial board of various journals and has edited special issues of several IEEE transactions (PAMI, IP, SMC-B, R&A, IFS) and other journals. He was General Chair for the IEEE Conference on Computer Vision and Pattern Recognition, IEEE Workshops on Applications of Computer Vision, IEEE Workshops on Learning in Computer Vision and Pattern Recognition; Chair for the DARPA Image Understanding Workshop and Program Chair for the IEEE Workshops on Computer Vision Beyond the Visible Spectrum and Multimodal Biometrics. He is a Fellow of the Institute of Electrical and Electronics Engineers (IEEE), American Association for the Advancement of Science (AAAS), International Association of Pattern Recognition (IAPR), and the International Society for Optical Engineering (SPIE).