Neurocomputing 74 (2011) 2452–2463

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Combining local face image features for identity verification

Beom-Seok Oh, Kar-Ann Toh *, Andrew Beng Jin Teoh, Jaihie Kim

School of Electrical and Electronic Engineering, Yonsei University, Seoul, Republic of Korea

ARTICLE INFO

Available online 14 May 2011

Keywords:
Face identity verification
Local feature extraction
Match scores fusion
Extreme learning machine
Total error rate minimization

ABSTRACT

With an aim of extracting robust facial features under pose variations, this paper presents two directional projections corresponding to extraction of vertical and horizontal local face image features. The matching scores computed from both horizontal and vertical features are subsequently fused at score level via an extreme learning machine that optimizes the total error rate for performance enhancement. In order to benchmark the performance, both the feature extraction and fusion results are compared with that of popular face recognition methods such as principal components analysis and linear discriminant analysis in terms of equal error rate and CPU time. Our empirical experiments using four data sets show encouraging results under considerable horizontal pose variations.

1. Introduction

Over the past few decades, face recognition remained to be among the most challenging research topic in biometrics field. The main reasons include the highly overlapping intra- and inter-identity distributions due to external imaging factors such as variations of illumination, expression, pose, age and occlusion.

According to [1], the approaches to face recognition can be divided into three categories: (1) global or holistic approach, (2) local approach and (3) hybrid approach. The global approach utilizes the entire face image information to construct features for recognition [2]. This approach shows relatively good performance for face images of frontal view [2,3]. However, the global approach often fails to yield reliable results when the face images are contaminated with illumination, pose, expression, and occlusion changes.

On the other hand, the local approach does not suffer much from the imaging factors since such variations affect the face image only partially [4]. Essentially, geometrical facial information such as width of mouth, distance between eyes, and size of nose have been adopted as features in the local methods. Some local methods also adopt blocks of appearance information such as regions of eyes, mouth, and nose as local features [5,6].

The hybrid approach utilizes both global and local facial information for feature construction and recognition. This approach is considered to be similar to that of human's recognition process [7]. One advantage of the hybrid method is the performance enhancement. However, it may incur a high computational cost due to implementation of both local and global methods.

Although the local approach works well over the global approach for images with pose, expression and illumination variations [6,8], most of the local methods require to locate the positions of facial components and extract them for feature construction [6,8,9]. Here, detecting the facial components is no trivial matter and a reliable recognition performance cannot be guaranteed if the facial components are falsely detected (or failed to detect). Therefore, it is necessary and beneficial to extract local facial information without such an additional task.

In this paper, we investigate into two directional projections which correspond to extraction of vertical and horizontal local face image features without the need to perform facial components detection. Essentially, two directional projection matrices, which consist of only horizontal and vertical groups of one's with remaining elements zero in their respective matrix entries, are proposed for these projections. Subsequently, the match score outputs obtained from both directional features are fused via an extreme learning machine that optimizes a total error rate [10] to enhance the verification performance. The robustness of the proposed projections against some external imaging factors, particularly on pose variations, will be evaluated under two scenarios using four databases.

The main contributions of this work are enumerated as follows: (i) We proposed two directional projections which correspond to extraction of vertical and horizontal local face image features; (ii) We showed that an extreme learning machine that optimizes the total error rate can be useful for scores fusion in terms of enhancing the verification performance; (iii) Empirically, we showed that extracting horizontal local facial features could be beneficial to alleviating verification error incurred from horizontal pose variations; (iv) We provided

* Corresponding author.
E-mail address: katoh@yonsei.ac.kr (K.-A. Toh).

0925-2312/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.neucom.2010.12.036

extensive experiments comparing the proposed method with related methods in the literature.

This paper is organized as follows: a brief literature survey will be provided in Section 2. In Section 3, we will briefly discuss some background knowledge on an extreme learning machine that minimizes a total error rate for scores fusion. Section 4 describes our proposed method where two directional projections are presented. This is followed by a score fusion to improve the recognition performance. Section 5 presents the data used and the experimental results with some discussions. Some concluding remarks are given in Section 6.

2. A brief literature survey

In order to gain an overview of the state-of-the-arts in the field, this section provides a brief review of existing literatures according to the above-mentioned approaches in face feature extraction. Particularly, we shall focus on the local approach where our proposed method belongs to.

2.1. Global approach

The principal component analysis (PCA) [2] and linear discriminant analysis (LDA) [11] methods are among the popular choices under the global approach. Essentially, these methods transform face images into a feature space called an eigenface for subsequent recognition [2]. However, such transformation can be computationally expensive, particularly when both the dimension of image and the sample size are large.

In order to reduce some computational cost required in the transformation, Yang et al. [12] proposed a two-dimensional PCA, called 2DPCA, by directly projecting the raw images without reshaping them into column vectors. The 2DPCA was subsequently extended to horizontal and vertical 2DPCA to cater for directional feature extraction [13]. This concept was also capitalized by Yu et al. [14] who extended the 2DLDA [15] to include directional features in a method called $(2D)^2$LDA [16]. The concept of PCA and LDA was further extended to a higher order PCA [17] and a higher order LDA [18] by the definition of tensor [17,18]. Under the tensor framework, 2DPCA and 2DLDA are special cases of second order tensor representations for face recognition [18,19].

Generally, these global approaches have been shown to achieve a reliable performance for classifying images of faces with a frontal view. However, the global approach remains to be much limited by external and imaging factors such as illumination, expression, pose, and occlusion [6,20].

2.2. Local approach

The local approach utilizes either partial geometrical facial information [5] or neighborhood topological information [21] to construct features for recognition. Since variations of local facial features caused by pose and illumination changes are relatively small as compared to that of global features, the local approach can be robust to such imaging factors [6]. In other words, the local approach may be deployed to complement the global approach to deal with the external and imaging factors. Based on the types of information adopted, the local approach can be sub-divided into three categories: (1) local geometric-feature based method, (2) local appearance based method [1,5] and (3) local manifold based method.

2.2.1. Local facial geometric-feature based method

This method uses geometrical facial features, such as the width of head or distance between eyes and so on, to construct pattern features for recognition. Ever since the successful use of an elastic bunch graph matching (EBGM [9]) for face verification using the FERET database [22], the local approach using geometrical facial features became one of the active research topics in face recognition [1]. Essentially, the geometrical features consist of a small quantity of measurements from the facial components [5]. This means that the storage cost of the system is significantly smaller than that of appearance based methods. However, implementation of these algorithms is not an easy task [1].

2.2.2. Local appearance based method

Essentially, this method uses the appearance information of face such as intensity for identity recognition. The approach, which is simple and easily implemented, appears to emerge among the dominant methods in face recognition. Based on the information extracted from the image, this method can be further classified into three categories: (i) local component-based method, (ii) local region-based method and (iii) local value-based method.

(i) *Local component-based method:* This method extracts regional parts of a face image which are occupied by facial components such as eyes, nose, and mouth to construct features for recognition. The studies conducted by Heisele et al. [6] indicate that the local component-based method can be more robust than the global approach under the condition of pose variations in face verification or identification. Moreover, the "curse of dimensionality" as seen in the global approach can be alleviated. However, since the method uses only some parts of an image for recognition, the overall configuration information of face can be lost [8].

(ii) *Local region-based method:* This method divides the face image into equal sub-regions, and uses them in a matching process for recognition [23]. Gottumukkal et al. applied the eigenface technique to these partitioned sub-regions, called modular PCA (mPCA) [4]. The mPCA outperformed a typical PCA in terms of recognition rate under the conditions of large expression and illumination variations on Yale database. According to the experiments conducted by Zou et al. [8], the local region-based method, which uses the overall shape information of face for recognition, outperforms the local component-based method.

(iii) *Local value-based method:* This method extracts local appearance information at pixel level to construct features for recognition. Achlioptas [24] proposed a sparse random projection (SRP), where a predefined number of random matrix entries were set to zero. Kim et al. [25] extracts a bunch of local features from a face image by grouping the non-zero elements generated by SRP at random locations. It has been shown that local features extracted at random positions can achieve discriminative information with a low computational cost [24,25]. However, because of the randomness, the performance is not guaranteed for individual experiments.

2.2.3. Local manifold based method

The goal of manifold learning is, essentially, to find a low-dimensional subspace that best preserves the high-dimensional topological structure of a patch (particularly a local patch) [26–28]. The local patches are subsequently rearranged (e.g. aligned) in the reduced space with global coordinates information for final classification [26–28].

Recently, two methods namely, Marginal Fisher Analysis (MFA) and Discriminative Local Alignment (DLA) have been

proposed based on graph embedding [26] and patch alignment [27] respectively. Zhou et al. [28] proposed a manifold elastic net with a sparse solution based on patch alignment. Essentially, these methods capitalized on the local manifold learning methodology for dimensionality reduction.

2.3. Hybrid approach

Similar to human's recognition mechanism [7], the hybrid approach utilizes both global and local information for face recognition [1]. Pentland et al. [29] proposed a modular Eigenspace that adopts both global and local for face recognition. In their experiments, modular eigenspaces outperforms both eigenface [2] and eigenfeatures [29] in terms of recognition rate. Recently, Al-Osaimi et al. [30] applied integrated global and local geometric 3D surface cues, rank-0 tensor fields, to 3D face recognition. Although the hybrid approach takes advantage of both global and local methods, it may incur a high computational cost due to implementation of both methods.

2.4. Summary of literature survey

The global approach shows a rather reliable performance for frontal view of face images. However, this approach can be sensitive to variations resulted from illumination, pose, expression, and occlusion. The local approach can be considered to complement the sensitivity problems caused by the global approach since any variation within local features is relatively small as compared to that of the global features. The hybrid approach combines the benefits of both local and global approaches by incurring a heavier computational cost.

Our proposed method belongs to the local region-based method since only a part of face image is extracted by a projection to construct features for face verification.

3. Preliminaries

In this section, we provide some preliminaries for immediate reference. Particularly, the extreme learning machine (ELM) which will be adopted in fusion and a total error rate minimization will be introduced.

3.1. Extreme learning machine (ELM) [31]

Suppose we have m number of training samples $(\mathbf{x}_i,\mathbf{y}_i)$, $i=1,2,\ldots,m$, where $\mathbf{x}_i=[x_{i1},x_{i2},\ldots,x_{in}]^T\in\mathbb{R}^n$ is a n-dimensional input vector and $\mathbf{y}_i=[y_{i1},y_{i2},\ldots,y_{id}]^T\in\mathbb{R}^d$ is a d-dimensional output target vector. A single hidden layer feedforward network with $\tilde{N}$ hidden nodes and common activation functions (called an extreme learning machine (ELM) [31]) can be written as

$$\sum_{j=1}^{\tilde{N}}\boldsymbol{\beta}_j\phi(\mathbf{w}_j\cdot\mathbf{x}_i+b_j)=\mathbf{o}_i,\quad i=1,2,\ldots,m, \tag{1}$$

where $\mathbf{w}_j=[w_{j1},w_{j2},\ldots,w_{jn}]^T$ is a randomly fixed weight vector connecting the j-th hidden node to the input nodes, $\boldsymbol{\beta}_j=[\beta_{j1},\beta_{j2},\ldots,\beta_{jd}]^T$ is the weight vector connecting the j-th hidden node to the output nodes, b_j is a threshold for the j-th hidden node and $\mathbf{o}_i=[o_{i1},o_{i2},\ldots,o_{id}]^T$ is the d-dimensional network output.

The above m equations can be written more compactly in matrix form as $\mathbf{H}\Theta=\mathbf{Y}$, where

$$\mathbf{H}=\begin{bmatrix}\phi(\mathbf{w}_1\cdot\mathbf{x}_1+b_1) & \cdots & \phi(\mathbf{w}_{\tilde{N}}\cdot\mathbf{x}_1+b_{\tilde{N}})\\ \vdots & \cdots & \vdots\\ \phi(\mathbf{w}_1\cdot\mathbf{x}_m+b_1) & \cdots & \phi(\mathbf{w}_{\tilde{N}}\cdot\mathbf{x}_m+b_{\tilde{N}})\end{bmatrix}_{m\times\tilde{N}}, \tag{2}$$

$\Theta=[\boldsymbol{\beta}_1^T\cdots\boldsymbol{\beta}_{\tilde{N}}^T]_{\tilde{N}\times d}$ and $\mathbf{Y}=[\mathbf{y}_1^T\cdots\mathbf{y}_m^T]_{m\times d}$. Here, the i-th row of $\mathbf{H}$ indicates the i-th input data.

With the randomly fixed weights in the hidden nodes, ELM gives a closed-form solution for the output weighting parameters as

$$\Theta=\mathbf{H}^{\dagger}\mathbf{Y}=(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{Y}, \tag{3}$$

where $\mathbf{H}^{\dagger}$ is the *Moore–Penrose generalized inverse* of matrix $\mathbf{H}$. The universal approximation capability of ELM has been theoretically proven in [32,33]. Moreover, a broad range of activation function types, including non-differentiable functions, can be adopted in ELM [33,34]. We shall focus on an ELM with additive hidden node using a sigmoid activation function in this work.

3.2. Total error rates minimization

The performance of a biometric verification system is frequently measured using error rates. The false acceptance rate (FAR) is the rate that an un-enrolled person has been falsely authorized as a genuine-user. On the other hand, the false rejection rate (FRR) is referred to the rate that an enrolled person has been falsely rejected. From the FAR and FRR obtained above, total error rate (TER) can be defined as the sum of FAR and FRR [10].

For a genuine-user, the related variables have been labeled with a superscript '+' and for an imposter, the related variables have been labeled with a superscript '−'. For example, $\mathbf{x}^-\in\mathbb{R}^{m^-}$ denotes the feature vector of an imposter and $\mathbf{x}^+\in\mathbb{R}^{m^+}$ denotes the feature vector of a genuine-user, where $m=m^++m^-$ denotes the total training population.

For a binary classifier which is linear in its parameters (such as $\boldsymbol{\beta}$ in ELM) and with appropriate normalization, a quadratic approximation to the zero–one indicator step function $I(\cdot)$ can be adopted [10] giving rise to a closed-form solution for fusion:

$$\boldsymbol{\beta}=\left[\frac{1}{m^-}\sum_{j=1}^{m^-}\mathbf{H}_j^T\mathbf{H}_j+\frac{1}{m^+}\sum_{i=1}^{m^+}\mathbf{H}_i^T\mathbf{H}_i\right]^{-1}\times\left[\frac{(\tau-\eta)}{m^-}\sum_{j=1}^{m^-}\mathbf{H}_j^T\mathbf{1}_-+\frac{(\tau+\eta)}{m^+}\sum_{i=1}^{m^+}\mathbf{H}_i^T\mathbf{1}_+\right], \tag{4}$$

where $\mathbf{H}_j\triangleq\mathbf{H}(\mathbf{x}_j^-)$, $\mathbf{H}_i\triangleq\mathbf{H}(\mathbf{x}_i^+)$, $\mathbf{1}_-=[1,\ldots,1]^T\in\mathbb{N}^{m^-}$, $\mathbf{1}_+=[1,\ldots,1]^T\in\mathbb{N}^{m^+}$ (see (2) for definition of matrix $\mathbf{H}$), τ is a preset threshold and η corresponds to an arbitrary offset. The estimated parameter $\boldsymbol{\beta}$ can then be used with $\mathbf{H}_t$ constructed from new test data for class label prediction using $\mathbf{y}_t=\mathbf{H}_t\boldsymbol{\beta}$ and a threshold operation.

4. Proposed method

Our proposed method consists of two parts, namely, two directional projections at sensor level and a fusion at score level. The directional projections correspond to extraction of horizontal and vertical image patterns. Our main motivation for adopting binary projection based method can be attributed to its simplicity and computational efficiency. The resulted matching scores of these projections are finally fused via an extreme learning machine (ELM) [31] adopting a total error rate minimization [10]. Each of these parts will be described in greater detail in the following subsections.

4.1. Extraction of horizontal and vertical face features

4.1.1. Directional projection matrix

In order to extract vertical and horizontal facial components from face images, two types of directional projection matrices are

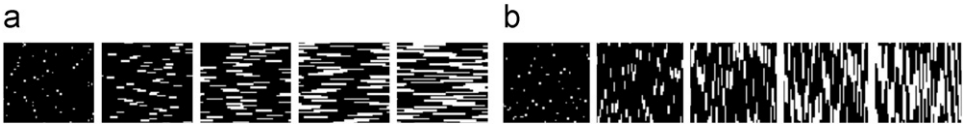


Fig. 1. Some sample examples of (a) $\mathbf{R}_1$ and (b) $\mathbf{R}_2$ projection matrices with different sparsities where the "0" and "1" pixels are shown in black and white respectively (highest sparsity on the left).

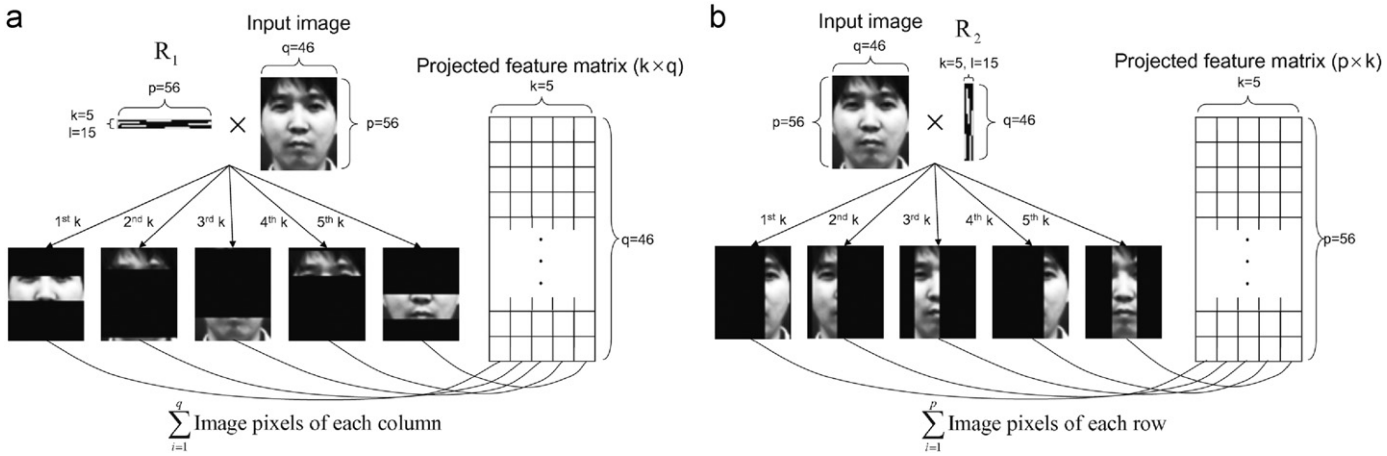


Fig. 2. Illustration of (a) $\mathbf{R}_1$ and (b) $\mathbf{R}_2$ projection procedures with $k=5$ and $l=15$.

proposed. Here, a projection matrix which consists of only horizontal groups of 1's and 0's will be called $\mathbf{R}_1$ projection matrix (see Fig. 1(a)) and that with vertical groups of 1's and 0's will be called $\mathbf{R}_2$ projection matrix (see Fig. 1(b)). The positions of these groups of 1's are determined randomly for each vector (row vectors in $\mathbf{R}_1$ and column vectors in $\mathbf{R}_2$) wherein each group consists of l number of "1" entries. Only a single group of ones appears in each vector with the remaining elements zeros. The ratio of the total size $L(L=l\times k$ for k number of projection basis) over the image size (say $p\times q$) will be called sparsity. The size of both projection matrices thus depends on the size of images used since the matrix is applied directly to the raw image. Suppose a raw image $\mathbf{X}$ has an p-by-q resolution, the size of $\mathbf{R}_1$ will be k-by-p and the size of $\mathbf{R}_2$ will be q-by-k where k is a chosen projection size. Fig. 1 shows some examples of $\mathbf{R}_1$ and $\mathbf{R}_2$ directional projection matrices with varying sparsity levels ($l\times k/p\times q$).

4.1.2. Extraction of directional facial features

There is no training process required and $\mathbf{R}_1$ will be pre-multiplied directly onto test face images $\mathbf{X}_i\in\mathbb{R}^{p\times q}$, $i=1,\ldots,m$ giving:

$$\mathbf{Y}_{V_i}=\mathbf{R}_1\cdot\mathbf{X}_i,\quad i=1,\ldots,m, \tag{5}$$

where m is the total number of images in the test set. Since the projected features resulted from $\mathbf{R}_1\mathbf{X}_i$ consist of compressed (due to summations in matrix product) vertical components of the face images, we will call them compressed vertical features. The size of projected features for each image is $k\times q$. As illustrated by Fig. 2(a), this projection by pre-multiplication corresponds to a summation of vertical non-zero elements and this process progresses horizontally. This results in a set of compressed vertical face components after projection of each basis.

The $\mathbf{R}_2$ directional projection matrix, on the other hand, will be used as a post-multiplier given by:

$$\mathbf{Y}_{H_i}=\mathbf{X}_i\cdot\mathbf{R}_2,\quad i=1,\ldots,m. \tag{6}$$

Due to the matrix product with vertically random groups of ones, a compressed horizontal face image features can be extracted by (6) as Fig. 2(b) illustrates. The size of projected features resulted from $\mathbf{X}_i\mathbf{R}_2$ is $p\times k$. The amount of local features extracted is dependent on the projection size k, and the group size l. The location of the ones group in the vector decides which part of the face images will be extracted, and the number of such projections is set by k. Figs. 2(a) and (b) depict the procedures of $\mathbf{R}_1\mathbf{X}_i$ and $\mathbf{X}_i\mathbf{R}_2$ projections with $k=5$ and $l=15$, respectively. As shown in both figures, the vertical local features and horizontal local features at five locations of an image have been extracted respectively. In order to observe the effect of projection size and sparsity on verification performance, we shall vary k and l in our experiments.

4.2. Scores fusion

The compressed vertical and horizontal features extracted in the above are finally fused at score level to improve the verification performance. The matching scores, which consist of genuine-user and imposter scores, are obtained based on the Euclidean distance of intra- and inter-matchings among the template identities, respectively. For instance, consider identity indices i and j and let $\mathbf{Y}^i$ and $\mathbf{Y}^j\in\mathbb{R}^{k\times q}$ be the projected $\mathbf{R}_1$ (or $\mathbf{Y}^j\in\mathbb{R}^{p\times k}$ for $\mathbf{R}_2$ projection) feature matrices (5). If these features are extracted from two images of one identity ($i=j$), then the genuine-user match score is computed by:

$$s_n^*=\sqrt{\sum_{u=1}^{k}\sum_{v=1}^{q}(\mathbf{Y}^i_{(u,v)}-\mathbf{Y}^j_{(u,v)})^2},\quad n=1,2,\ldots,m^*,$$

$$*=\begin{cases}+ & \text{if } i=j,\\ - & \text{if } i\neq j.\end{cases} \tag{7}$$

On the other hand, the match score is an imposter if the images are taken from two different identities ($i\neq j$). These genuine-user match scores $\mathbf{x}^+=[s_1^+,s_2^+,\ldots,s_{m^+}^+]\in\mathbb{R}^{m^+}$ and imposter match scores $\mathbf{x}^-=[s_1^-,s_2^-,\ldots,s_{m^-}^-]\in\mathbb{R}^{m^-}$ resulted from $\mathbf{R}_1\mathbf{X}_i$ or $\mathbf{X}_i\mathbf{R}_2$ are then used in (2) to construct $\mathbf{H}$ for fusion. Consequently, the algorithm from (4)

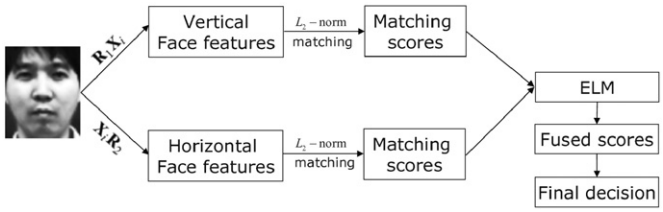


Fig. 3. Overall scheme of the proposed method.

[10] based on the total error rate minimization of ELM (TERELM) is adopted to compute the fusion parameters for future prediction.

4.3. Summary of the proposed method

The overall procedure of our proposed method is depicted in Fig. 3 and the overall steps in the proposed method are summarized as follows:

(1) Generate both $\mathbf{R}_1$ and $\mathbf{R}_2$ directional projection matrices and store them as follows:
   (a) Initialize a k-by-p matrix with all elements zero for $\mathbf{R}_1$ and q-by-k matrix for $\mathbf{R}_2$.
   (b) Generate k number of random integers $1 \le a_i \le p$, $i = 1,2,\ldots,k$ for $\mathbf{R}_1$ and $1 \le b_i \le q$ for $\mathbf{R}_2$, respectively.
   (c) For each i-th row of $\mathbf{R}_1$, replace $1 \le l \le p$ number of contiguous zero elements by 1's from the a_i-th element onwards. If $a_i + l > p$ (overflow), then replace the $(a_i + l) - p$ number of zero elements by 1's from the first element in the i-th row. The $\mathbf{R}_2$ projection matrix can be generated in a similar manner to that of $\mathbf{R}_1$, except that the maximum group size is now $l \le q$, replacing the 1's group column-wise and overflow if $(b_i + l) > q$.
   (d) Store the generated $\mathbf{R}_1$ and $\mathbf{R}_2$ projection matrices.
(2) Perform $\mathbf{R}_1\mathbf{X}_i$ and $\mathbf{X}_i\mathbf{R}_2$ using the stored projection matrices to obtain the compressed directional face image features.
(3) Obtain matching scores from both directional features by inter- and intra-identity comparisons based on the Euclidean distance measure (7). Those resulted from inter-identity comparisons are called imposter scores and those from intra-identity comparisons are called genuine-user scores. The scores are normalized to within [0,1].
(4) Compute the fusion learning parameters $\boldsymbol{\beta}$ based on (4) using a training set consisting of genuine-user and imposter scores. Use the computed $\boldsymbol{\beta}$ to compute the predicted fusion scores for test data using $\hat{g}(\mathbf{x}_t) = \mathbf{H}_t\boldsymbol{\beta}$, where $\mathbf{H}_t$ is a hidden layer output matrix obtained from test data.
(5) Compute the final decision based on a preset threshold. If $\hat{g}(\mathbf{x}_t) \ge \tau$, then the test sample $\mathbf{x}_t$ is classified as a genuine-user, otherwise an imposter, where $\hat{g}(\mathbf{x}_t)$ is the biometric classifier output and τ is the threshold.

5. Experiments

In order to evaluate the verification performance of the proposed method, four experiments were carried out under two scenarios using the ORL [35], BERC [36], Sheffield (previously called UMIST database) [37] and FERET [22] databases.

In order to benchmark the performances of our proposed methods, the verification results will be compared with several existing face recognition methods such as principal component analysis (PCA) [2], horizontal two-dimensional PCA (H2DPCA) [12], vertical 2DPCA (V2DPCA) [16], bi-directional 2DPCA ((2D)^{2}PCA) [16], modular PCA (mPCA) [4], sub-pattern PCA (SpPCA) [23], horizontal two-dimensional linear discriminant analysis (2DLDA) [15], vertical 2DLDA (V2DLDA) [15,38] and bi-directional 2DLDA ((2D)^{2}LDA) [38] in terms of the equal error rate (EER) and CPU time performances. Since PCA, 2DPCA and 2DLDA are among the popular face recognition algorithms, they are chosen as representatives for the global approach in feature extraction. For comparability reasons, the mPCA and the SpPCA are included in our experiments to benchmark over those local methods. Apart from benchmarking with 2DPCA and 2DLDA methods, the proposed method is also compared with two recently published results of Marginal Fisher Analysis (MFA) [26] and Discriminative Local Alignment (DLA) [27] in order to benchmark over state-of-the-arts and supervised learning methods.

5.1. Database and preprocessing

The ORL database [35] consists of 40 identities with 10 images per identity. The images were acquired under variations in time, minor pose (tilting and rotation up to 20°), illumination, facial expression (open/closed eyes, smiling/not smiling) and facial accessory (glasses/no glasses). The BERC face database [36] consists of 5184 images from 96 identities wherein each identity contains 54 images. The images were captured under variation of illumination, expression, minor pose (neutral, 15° of angle at left and right directions), wind, and glasses conditions. The Sheffield database [37] contains 564 images from 20 identities (mixed race and gender). The images were captured under large pose variations (multi-views from profile to frontal). The FERET database [22] contains 14 051 images from 1199 identities. The images were acquired under illumination, expression and large pose (multi-views of $\pm 65°$) changes.

The original RGB face images are of 320×240 resolution for the BERC database and 256×384 resolution for the FERET database. The original images of ORL and Sheffield databases are of 92×119 resolution, with 256 grey level per pixel. These databases are normalized by centering the eyes and the lips. Subsequently, the normalized images are converted to grey level with a 56×46 resolution. Finally, a histogram equalization was performed based on a min-max normalization [39].

5.2. Experimental settings

5.2.1. Evaluation scenarios

Table 1 shows the configuration of our two experimental scenarios along with the utilized databases. Under scenario I, the proposed method is evaluated in terms of its robustness under minor pose variations on top of expression changes using ORL and BERC database. Under scenario II, we investigate into the sensitivity to large pose variations using Sheffield and FERET databases.

(i) *Scenario I—minor pose plus expression variations*: For the ORL database under scenario I, we utilize all the 400 images (40 identities with 10 images each). Half of the 10 images per identity will be used for training, while the remaining five images will be used for testing. Fig. 4 shows some preprocessed training and test sample images from two identities.

Table 1
Experimental configuration of the scenarios along with the used databases.

	Databases
Scenario I (with minor pose plus expression variations)	ORL and BERC
Scenario II (with large pose variations)	Sheffield and FERET

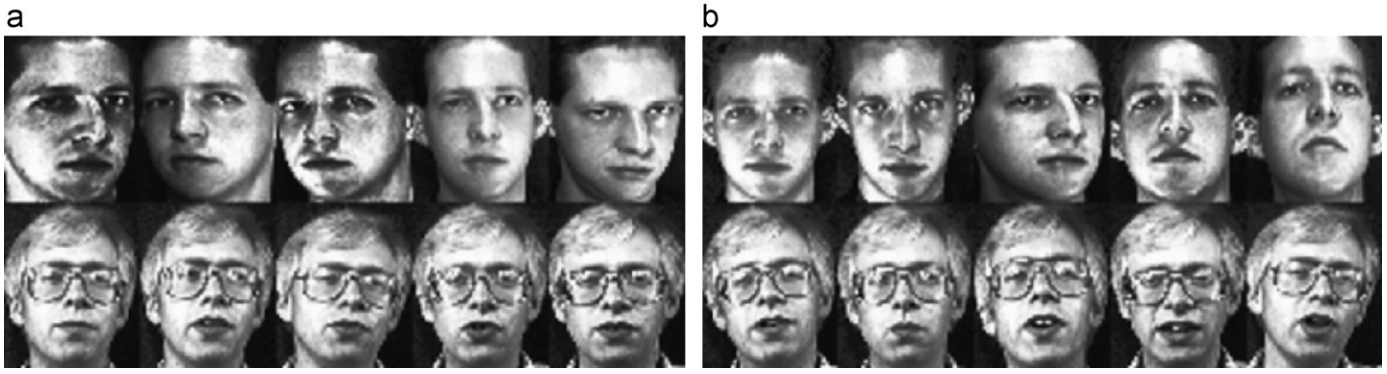


Fig. 4. Sample images of ORL database: (a) training and (b) test sets.

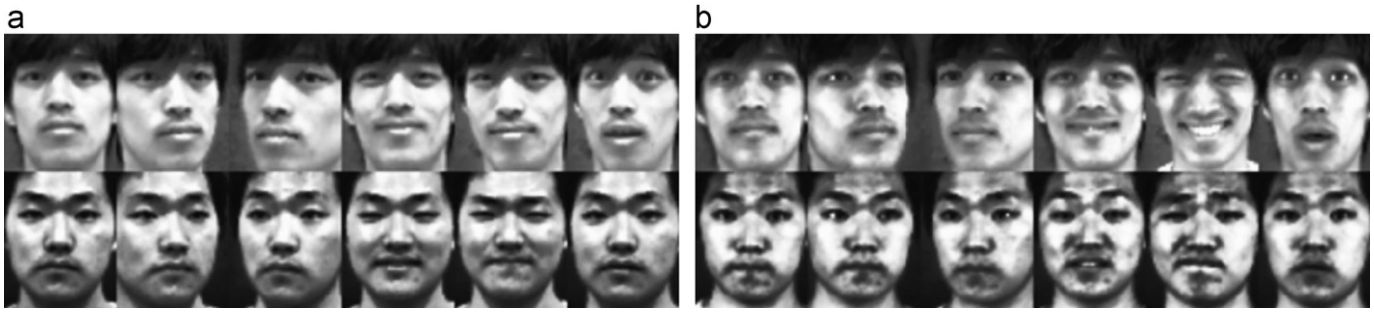


Fig. 5. Selected sample images from BERC database: (a) training and (b) test sets.

Fig. 6. (a–j) Selected sample images from the Sheffield database.

For the BERC database (again, under scenario I), we selected 96 identities with 12 images per identity which were captured under variations of illumination, expression and minor pose (neutral, 15° of angle at left and right directions). Half of the selected 12 images, which were acquired at frontal illumination with varying expression condition, will be used for training purpose. The remaining half images which were taken under bi-directional (both left and right) illuminations with similar expression variations comparing to the first half will be used for testing purpose. The dataset utilized in this scenario thus consists of 1152 (96×12) images. Fig. 5 shows some preprocessed sample images from two identities.

(ii) *Scenario II—large pose variations*: For the Sheffield database under scenario II, we used a subset of the database which consists of 200 images from 20 identities (10 images per identity). The selected images cover changes of orientation of face view from profile to frontal. Half of the selected 10 images—(a), (c), (e), (g) and (i) in Fig. 6—will be used for training purpose. The remaining half images—(b), (d), (f), (h) and (j) in Fig. 6—will be used for testing purpose. Fig. 6 shows some preprocessed sample images from one identity.

For the FERET database (again, under scenario II), we selected 200 identities with nine images per identity which were captured under large pose variations. Each of the selected images is respectively labeled as **ba** (frontal), **bb** ($+60°$), **bc**($+40°$), **bd** ($+25°$), **be**($+15°$), **bf**($-15°$), **bg**($-25°$), **bh** ($-40°$) and **bi** ($-60°$), where '°' indicates degree of angle, '+' means a subject faces to his left direction and '–' indicates a subject faces to his right direction [40]. Five images per identity—**ba**, **bc**, **be**, **bg** and **bi** in Fig. 7—will be used for training purpose. The remaining four images—**bb**, **bd**, **bf** and **bh** in Fig. 7—will be used for testing purpose. The dataset utilized in this experiment thus consists of 1800 (200×9) images. Fig. 7 shows some preprocessed sample images of one identity.

5.2.2. Evaluation protocols

Our proposed $\mathbf{R}_1$ and $\mathbf{R}_2$ projections have two parameters namely, the group size l which is the number of ones in one vector (row vector in $\mathbf{R}_1$ and column vector in $\mathbf{R}_2$), and projection size k which is the number of vectors in one projection matrix. Since the group size l is related to the amount of information extracted from a face image block, it is important in the directional matrix generation. Different group sizes $l \in \{1,2,\ldots,30\}$ will be experimented in our subsequent experiments to observe the effect on the verification performance.

Fig. 8, which is plotted over the projection size $k \in \{1,2,\ldots,30\}$, shows the average EER performances obtained from cross-validation using BERC database under scenario I at group sizes $l = 1$ and 26, respectively for $\mathbf{R}_1$ and $\mathbf{R}_2$, which have shown the best training cross-validation results. At large projection sizes $k > 20$, both $\mathbf{R}_1$ and $\mathbf{R}_2$ give better EER performances than at small

Fig. 7. Selected sample images from FERET database.

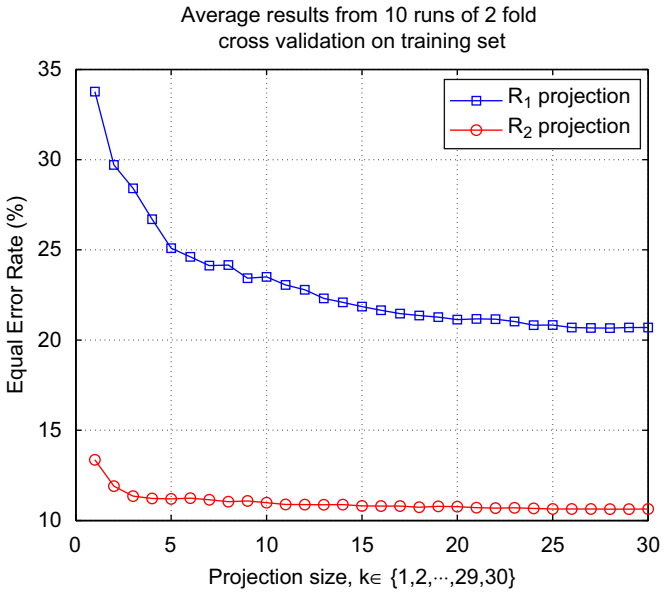


Fig. 8. Average EER performances from BERC database under scenario I plotted over different projection sizes $k \in \{1,2,\ldots,29,30\}$. The results are obtained from 10 runs of 2-folds cross-validation on training set using 30 different $\mathbf{R}_1$ and $\mathbf{R}_2$ matrices at group sizes $l=1$ for $\mathbf{R}_1$ and $l=26$ for $\mathbf{R}_2$.

projection sizes $k<5$. However, choosing a large projection size yields a large feature length ($k\times 46$ in $\mathbf{R}_1$ and $56\times k$ in $\mathbf{R}_2$). In order to balance between the performance and the feature length (computing cost), we set the projection size at $k=10$ for both $\mathbf{R}_1$ and $\mathbf{R}_2$ in our subsequent test experiments. The sizes of feature dimension produced from $\mathbf{R}_1\mathbf{X}_i$ and $\mathbf{X}_i\mathbf{R}_2, i=1,2,\ldots,m$ are thus 460 (10×46) and 560 (56×10) dimensions, respectively.

In scores fusion using extreme learning machine (ELM) [31] and total error rate minimization that adopts ELM (TERELM) [10], different numbers of hidden nodes $\tilde{N}\in\{5,10,\ldots,95,100\}$ are experimented based on 10 runs of two-fold cross-validation using only the training set to locate the optimal $\tilde{N}$. The obtained optimal values for $\tilde{N}$ are tabulated in the first column of Table 2. Following [10,31], all the inputs (matching scores) have been normalized into the range [0,1] where the desired target vector falls within $[-1,1]$, and a sigmoidal function is adopted for the activation function. For TERELM, $\tau=0$ and $\eta=1$ are set according to [10].

In order to investigate into the effect of class-specific weights in TER learning on recognition performance, 14 different weight settings (M) have been experimented as shown in Table 3. The class-specific weights are set as $1/w^-$ and $1/w^+$ according to one of the 14 choices of (w^-,w^+) pairs from the table. Only the results at optimal values of M which have been obtained based on five runs of two-fold cross-validation using only training set (see Table 2) will be reported in our test experiments.

For the mPCA [4] and SpPCA [23] methods, according to the image size (56×46) used, we have seven possibilities to partition the face images into equal sub-images $N\in\{2,4,8,16,46,92,186\}$, where N denotes the number of sub-images. Similar to the above setup, the optimal value for N has been chosen based on 10 runs of two-fold cross-validation using only the training set (see Table 2).

Table 2
Optimal parameter settings found by 10 runs (except for TERELM weights tuning (M) which uses five runs) of 2-fold cross-validation on training set.

		ELM & TERELM ($\tilde{N}$)	TERELM weights tuning (M)	mPCA (N)	SpPCA (N)
Scenario I	ORL	10	1	16	16
	BERC	10	8	16	4
Scenario II	Sheffield	15	6	16	16
	FERET	10	4	4	8

Table 3
Class-specific weights setting.

Index (M)	(w^-, w^+)	*Index* (M)	(w^-, w^+)	*Index* (M)	(w^-, w^+)
0:	$(1, 1)$	5:	$(0.2m^-, m^+)$	10:	$(m^-, 0.7m^+)$
1:	(m^-, m^+)	6:	$(m^-, 0.3m^+)$	11:	$(0.7m^-, m^+)$
2:	$(m^-, 1)$	7:	$(0.3m^-, m^+)$	12:	$(m^-, 2m^+)$
3:	$(1, m^+)$	8	$(m^-, 0.5m^+)$	13:	$(2m^-, m^+)$
4:	$(m^-, 0.2m^+)$	9:	$(0.5m^-, m^+)$	-	

As for the PCA, we shall observe the EER performance with respect to the variation of the reduced dimension d. Similarly, the results of 2D methods such as 2DPCA and 2DLDA will be investigated with respect to the variation of the number of principal component vectors d'.

All the test experiments will be performed based on two-fold cross-validation for statistical evidence where the average results are recorded. Furthermore, both $\mathbf{R}_1$ and $\mathbf{R}_2$ projections will be performed 30 times using 30 different $\mathbf{R}_1$ and $\mathbf{R}_2$ projection matrices, since these matrices have a random property in its generation (the position of one's group in each vector is decided in a random manner).

5.3. Results and discussion

5.3.1. Feature extraction results

Fig. 9 shows the average test EER comparisons of un-supervised learning methods such as PCA, H2DPCA, V2DPCA, $(2D)^2$PCA, mPCA, SpPCA, $\mathbf{R}_1$ and $\mathbf{R}_2$ based on two-fold test. The PCA is plotted over different reduced dimensions $d\in\{5,10,\ldots,145,150\}$; the H2DPCA, V2DPCA and $(2D)^2$PCA are plotted over different numbers of principal component vectors at $d'\in\{1,2,\ldots,30\}$[1]; the mPCA and the SpPCA are plotted over different dimensions of

[1] The feature lengths for H2DPCA, V2DPCA and $(2D)^2$PCA are respectively $56\times d'$, $46\times d'$ and $d'\times d'$, where d' denotes the number of principal component vector [16].

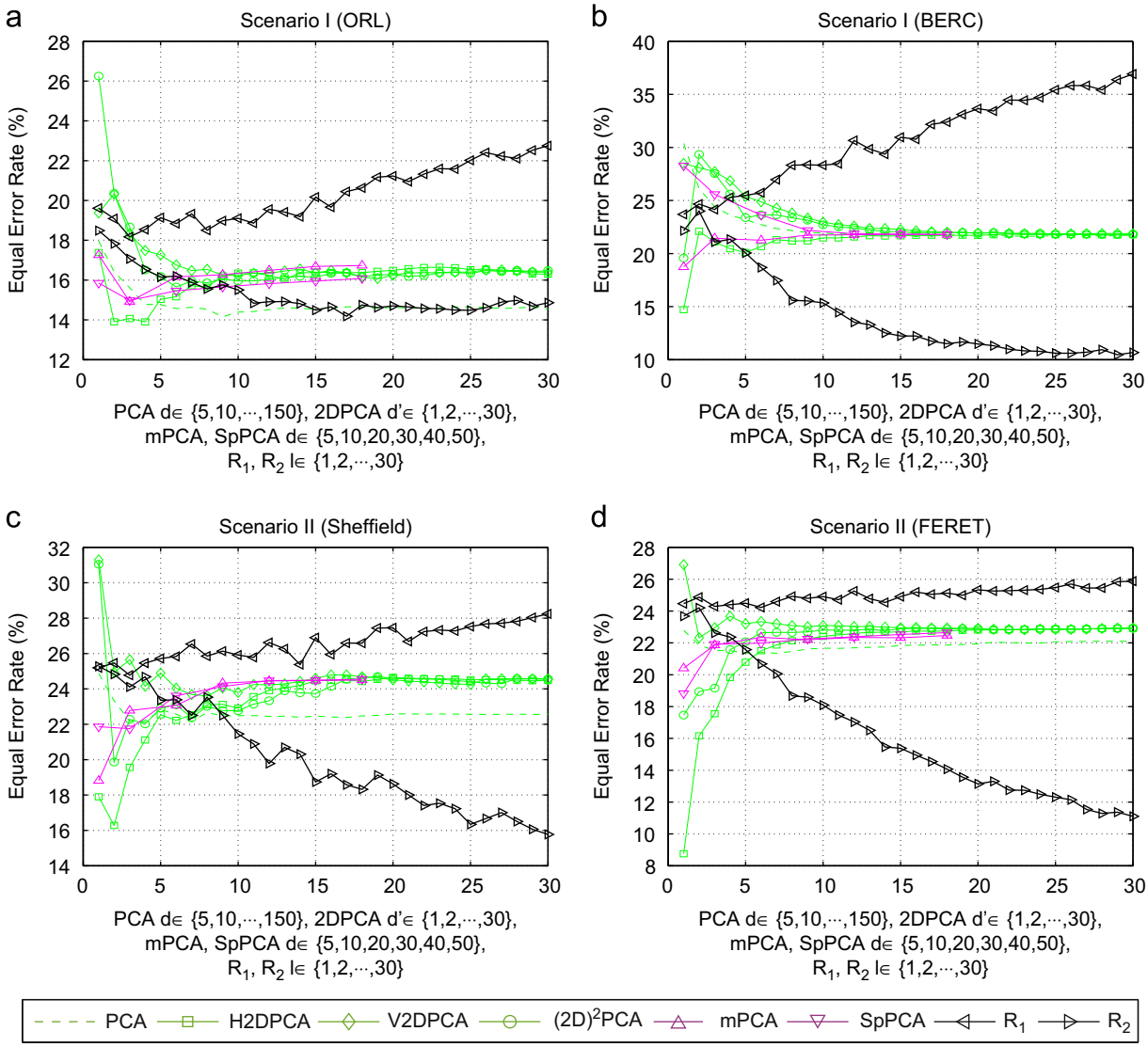


Fig. 9. Performance comparisons of feature extraction methods plotted over different sizes of reduced data at $d \in \{5,10,\ldots,150\}$ for PCA; $d \in \{5,10,20,30,40,50\}$ for mPCA and SpPCA; different number of principal component vectors $d' \in \{1,2,\ldots,30\}$ for H2DPCA, V2DPCA and $(2D)^2$PCA; different group sizes at $l \in \{1,2,\ldots,30\}$ for $\mathbf{R}_1$ and $\mathbf{R}_2$: (a) Scenario I (ORL); (b) scenario I (BERC); (c) scenario II (Sheffield); and (d) scenario II (FERET).

the reduced data at $d \in \{5,10,20,30,40,50\}$ due to the limitation of sub-image size; $\mathbf{R}_1$ and $\mathbf{R}_2$ are plotted over different group sizes at $l \in \{1,2,\ldots,30\}$.

(i) *Scenario I*: In Fig. 9(a) and (b) under scenario I, the best performances among the PCA methods were found from H2DPCA at $d'=4$ and 1, respectively. Among 2DPCA methods, V2DPCA yields the worst results. Local methods, mPCA and SpPCA, show similar EER performances with that of V2DPCA and $(2D)^2$PCA over the entire range of d and d'. With an appropriate setting of the group size l, $\mathbf{R}_2$ projection shows good performance with respect to all compared methods, while $\mathbf{R}_1$ projection shows deteriorated performances as the group size l increases. It is noted that the proposed $\mathbf{R}_2$ projection shows superiority in terms of EER over all compared methods at $l>10$ for the BERC database.

(ii) *Scenario II*: Fig. 9(c) and (d) show the superiority of H2DPCA at $d<4$ and $\mathbf{R}_2$ projection at $l>15$. Similar to the results of scenario I, V2DPCA shows the worst EER performance among the 2DPCA methods. Both mPCA and SpPCA show similar EER performances with that of V2DPCA and $(2D)^2$PCA methods under both databases. For the Sheffield database, $\mathbf{R}_2$ projection at $l>28$ slightly outperforms the best performance of H2DPCA at $d'=2$. However, H2DPCA at $d'=1$ shows about 2–3% higher EER than that of the best $\mathbf{R}_2$ projection (at $l=30$).

(iii) *Comparison with state-of-the-arts*: Apart from the above comparison with 2DPCA methods, the $\mathbf{R}_1$ and $\mathbf{R}_2$ projections are also compared with two supervised learning methods, namely MFA [26] and DLA [27]. Among the three databases (FERET [40], Yale [3] and Sheffield [41]) adopted in [27], we shall perform experiments on the Sheffield database since it is the only one that we can reproduce a similar condition regarding image cropping, normalization and selection.

Here, the experiments are performed according to the following protocol [27]: a random selection of 3, 5, 7 and 9 images per identity is used for training with half of the remaining images for parameter selection (validation set) and the other half for test. In order to avoid bothering by statistical fluctuations over random configuration of data sets and random property of the projection matrices, we performed an additional 30 trials using 30 different $\mathbf{R}_1$ and $\mathbf{R}_2$ projection matrices on top of the 10 runs that were performed in [27]. In total, we have performed 300 runs ($=10$ runs $\times$ 30 trials) for reproducible statistical evidence.

In Table 4, DLA shows the best recognition rates over the compared methods. The proposed $\mathbf{R}_2$ projection shows slightly lower recognition rates than that of DLA and outperformed that of MFA and $\mathbf{R}_1$ projection.

(iv) *Summary of feature extraction results*: Under scenario I on ORL database where the images were acquired under both pan and tilt pose variations, $\mathbf{R}_2$ projection and H2DPCA which correspond to extraction of horizontal image features show similar EER performances with that of PCA. On the other hand, under scenario I (BERC) and scenario II (Sheffield and FERET) where the adopted images contain either minor or large pose variations in the horizontal direction, $\mathbf{R}_2$ projection and H2DPCA yield much lower EER values than that of all other compared methods. This shows that extracting horizontal facial features could be beneficial for verification performance when the images contain horizontal pose variations.

Interestingly, most of the results show that the $\mathbf{R}_2$ projection and H2DPCA which correspond to extraction of horizontal features outperform the $\mathbf{R}_1$ projection and V2DPCA which extract the vertical facial features. This shows that the face images used in our experiments exhibit more discriminative information in the horizontal direction than that in the vertical direction.

Finally, the comparison with advanced supervised methods such as MFA and DLA shows that $\mathbf{R}_2$ projection can be a very efficient method even with its un-supervised nature.

Table 4
Best recognition rates (%) of $\mathbf{R}_1$ and $\mathbf{R}_2$ projections and other results recently reported on Sheffield database. The numbers in the parentheses are the reduced dimension for MFA; the reduced dimension, the selected parameters k_1 and k_2 for DLA; and the selected group size for $\mathbf{R}_1$ and $\mathbf{R}_2$.

Methods	The number of training samples			
	3	5	7	9
MFA [27]	79.55 (20)	90.00 (17)	94.76 (31)	96.05 (27)
DLA [26]	84.04 (40,2,1)	93.35 (44,2,2)	96.76 (33,4,5)	98.58 (24,6,5)
$\mathbf{R}_1$	69.91 (14)	82.59 (14)	89.06 (14)	92.97 (14)
$\mathbf{R}_2$	81.53 (30)	91.11 (29)	95.23 (26)	97.25 (29)

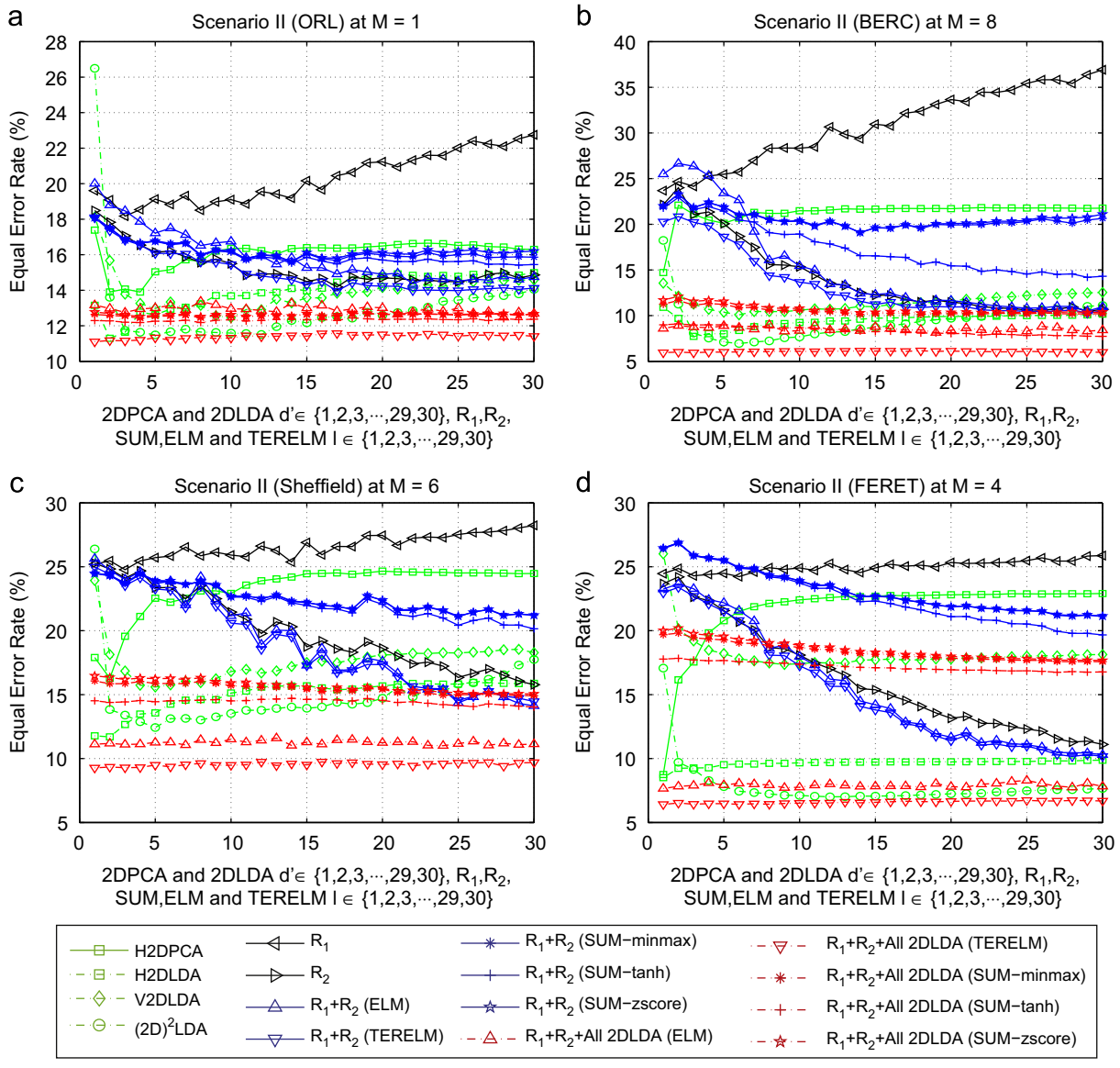


Fig. 10. Performance comparisons of feature extraction methods plotted over different group sizes at $l \in \{1,2,\ldots,30\}$. Results of H2DPCA, H2DLDA, V2DLDA and $(2D)^2$LDA are plotted over different number of principal component vectors $d' \in \{1,2,\ldots,30\}$ (a) Scenario I (ORL) at M=1; (b) Scenario I (BERC) at M=8; (c) Scenario II (Sheffield) at M=6 and (d) Scenario II (FERET) at M=4.

5.3.2. Fusion results

Fig. 10 shows the test average EER performances recorded over 30 trials of two-fold tests on fusing the match score outputs of $\mathbf{R}_1$ and $\mathbf{R}_2$ projections at different group sizes $l \in \{1,2,\ldots,30\}$. The TERELM results are obtained based on the optimal class-specific weight setting M (see Table 3) chosen from running two-fold cross-validation (five runs) using the training set. The previous H2DPCA which has shown the best EER performance among those PCA methods (see Fig. 9) and the results of the proposed $\mathbf{R}_1$ and $\mathbf{R}_2$ projections are shown in the same plot in order to compare with those performances before fusion. In order to benchmark the TERELM for scores fusion, a simple SUM rule [39] adopting three different normalization techniques [39] namely min-max, z-score and hyperbolic tangent is also experimented.

Since fusing the match score outputs using TERELM belongs to supervised learning, we shall compare the proposed TERELM with other supervised learning methods such as H2DLDA, V2DLDA and $(2D)^2$LDA. In Fig. 10, the EER performances of the compared supervised learning methods are plotted over different number of principal component vectors at $d' \in \{1,2,\ldots,30\}$. Subsequently, the match score outputs of all the three 2DLDA methods at optimal d' are fused with the match score outputs of $\mathbf{R}_1$ and $\mathbf{R}_2$ projections using TERELM. These optimal d' values were obtained based on 10 runs of two-fold cross validation using only training set as tabulated in Table 5. Similar to those results in previous section, Fig. 10(a) and (b) show the results under scenario I (ORL and BERC) and scenario II (Sheffield and FERET).

(i) *Scenario I*: In Fig. 10(a) and (b) under scenario I, fusion of $\mathbf{R}_1$ and $\mathbf{R}_2$ projections adopting TERELM (denoted as $\mathbf{R}_1+\mathbf{R}_2$ (TERELM) in the figure) shows similar or slightly better EER performances than that of $\mathbf{R}_2$ projection over the entire range of l. On the other hand, the fusion performances resulted from adopting ELM for fusion (denoted as $\mathbf{R}_1+\mathbf{R}_2$ (ELM) in the figure) appear to be over-fitted where the verification performance was not enhanced, particularly at $l<26$ for the ORL database and at $l<21$ for the BERC databases. $\mathbf{R}_1+\mathbf{R}_2$ (ELM) yields even worse EER values than that of $\mathbf{R}_1$ projection for $l<4$ for the ORL and for $l<5$ for the BERC database.

A similar trend can be observed from fusing $\mathbf{R}_1$ and $\mathbf{R}_2$ projections with all three 2DLDA methods as shown in Fig. 10(a) and (b). Here, fusion of all the five features using TERELM (denoted as $\mathbf{R}_1+\mathbf{R}_2+$All 2DLDA (TERELM) in the figure) shows an improvement of EER performances over the entire range of l while fusion using ELM (denoted as $\mathbf{R}_1+\mathbf{R}_2+$All 2DLDA (ELM) in the figure) does not show improvement over the best performed modality.

In terms of fusion techniques, the simple SUM rules (denoted respectively as SUM-minmax, SUM-zscore and SUM-tanh according to the normalization techniques) show a similar performance with that of ELM and a lower performance than that of TERELM for the ORL database case. As for the BERC database, the SUM rule shows much poorer performance than that of ELM and TERELM. Here we note that SUM-tanh outperforms the SUM-minmax and SUM-zscore particularly at $l>10$ see Fig. 10(b).

(ii) *Scenario II*: Fig. 10(c) and (d) show the fusion results experimented under scenario II. Except at $l \in \{1,5,6,8,11\}$ for the Sheffield database and at $2<l<10$ for the FERET database, $\mathbf{R}_1+\mathbf{R}_2$ (ELM) shows an improved EER performance over the $\mathbf{R}_2$ projection. Similar to the previous results under scenario I, $\mathbf{R}_1+\mathbf{R}_2$ (TERELM) produces better EER performances than that of $\mathbf{R}_2$ projection over the entire range of l for both databases.

For the Sheffield database, both $\mathbf{R}_1+\mathbf{R}_2+$All 2DLDA (ELM) and $\mathbf{R}_1+\mathbf{R}_2+$All 2DLDA (TERELM) outperform the best of H2DLDA at $d'=2$ over the entire range of l. However, only $\mathbf{R}_1+\mathbf{R}_2+$All 2DLDA (TERELM) outperforms the best of $(2D)^2$LDA over the entire range of l for the FERET database. Similar to the results under scenario I for BERC database, the SUM rule techniques show poorer performances than those of ELM and TERELM over the entire range of l.

(iii) *Summary of fusion results*: The above experiments show that TERELM with an appropriate setting of parameters can almost always improve the verification performance after fusion (see Fig. 10). As for ELM some over-fitting was observed. The simple SUM rule adopting three normalization techniques failed to improve the verification performance in many cases.

We have seen from the results in Fig. 9 that V2DPCA shows a lower verification performance as compared to that of H2DPCA and $(2D)^2$PCA over the entire range of d'. Similarly, V2DLDA, which corresponds to extraction of vertical facial features, shows a poor EER performance under scenario I (BERC) and scenario II (Sheffield and FERET) where the utilized images contain horizontal pose variation. Particularly, V2DLDA yields about 10% lower EER performance than that of $(2D)^2$LDA and about 8% lower (about 20% lower at $d'=1$) than that of H2DPCA over the entire range of d'. This suggests that extracting horizontal facial features could be beneficial to alleviating verification error incurred from horizontal pose variations.

5.3.3. CPU time comparison

In this sub-section, the feature extraction methods and fusion methods are compared in terms of their training and testing CPU times in seconds based on a PC of 2.66 GHz with 4G RAM under Matlab platform [42]. The CPU time was measured using the FERET database which is the largest database used in our experiments.

For $\mathbf{R}_1$ and $\mathbf{R}_2$ projections, we measured the computational time consumed in generating the directional projection matrix and recorded it as the training time. For mPCA and SpPCA, the CPU time was measured at the optimal parameter N which is the number of partitioned sub-images. The test time is the CPU time consumed in testing 100 data samples. In order to observe the effect of different parameter settings on computing cost, we measured these CPU times over variations of parameter settings. For example, the CPU time for 2D methods was measured at $d' \in \{1,2,\ldots,30\}$.

Fig. 11(a) and (c) show an overview of the training and test CPU times, whereas Fig. 11(b) and (d) show respectively their zoom-in views. In Fig. 11(a) and (b), PCA shows the highest CPU time consumption for training with respect to all other compared methods. Here, we observe that PCA is even slower than the CPU speed of fusing all the five features ($\mathbf{R}_1+\mathbf{R}_2+$All 2DLDA) over the entire variations of d and l. Due to the large number of features for fusion, $\mathbf{R}_1+\mathbf{R}_2+$All 2DLDA shows about 10 sec slower than that of $\mathbf{R}_1+\mathbf{R}_2$. SpPCA is about 0.5 sec faster than that of mPCA. Among those 2D methods, both H2DPCA and V2DPCA show the least CPU time consumption which is about twice faster than that of $(2D)^2$PCA. 2DLDA methods take up slightly more computing cost for training than that of 2DPCA methods. Among the compared methods, $\mathbf{R}_1$ and $\mathbf{R}_2$ projections take up the least CPU time in terms of training time.

As Figs. 11(c) and (d) illustrate, PCA takes the shortest test time over the entire range of d, d' and l. $(2D)^2$PCA and $(2D)^2$LDA show slightly worse CPU time performance in testing than that of PCA at $d'<12$. Since the feature length of $(2D)^2$PCA and $(2D)^2$LDA is much lower than that of other 2D methods [16,38], they are much faster in testing than other 2D methods over the entire

Table 5
The optimal number of principal component vectors for H2DLDA, V2DLDA and $(2D)^2$LDA obtained from 10 runs of two-fold cross-validation using only training set.

		H2DLDA	V2DLDA	$(2D)^2$LDA
Scenario I	ORL	2	5	5
	BERC	3	10	6
Scenario II	Sheffield	1	5	5
	FERET	1	6	10

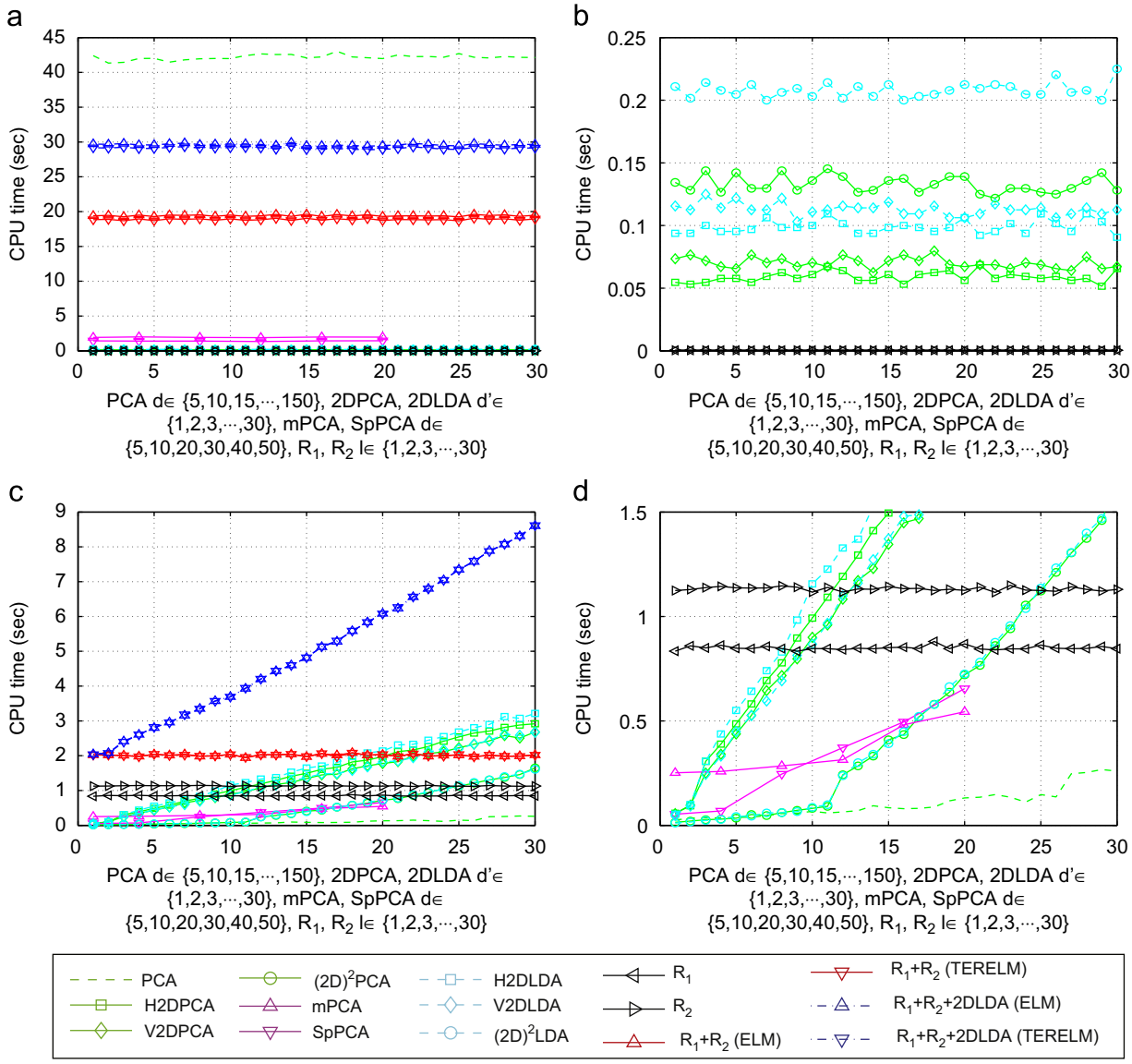


Fig. 11. Training and testing CPU times plotted over different sizes of reduced data at $d \in \{5,10\ldots,150\}$ for PCA; $d \in \{5,10,20,30,40,50\}$ for mPCA and SpPCA; different number of principal component vectors at $d' \in \{1,2\ldots,30\}$ for 2D methods; different group sizes at $l \in \{1,2,\ldots,29,30\}$ for $\mathbf{R}_1$ and $\mathbf{R}_2$; (a) training time (overview); (b) training time (zoom-in view); (c) test time (overview) and (d) test time (zoom-in view).

range of d'. Due to the difference in feature length, $\mathbf{R}_1$ projection is slightly faster than $\mathbf{R}_2$ projection. As d' increases, the testing CPU time increases due to the increase in feature length.

6. Conclusion

In this paper, we proposed two directional projections which correspond to extraction of compressed vertical and horizontal local face image features. In order to enhance the verification performance, the matching scores resulted from both individual projections were fused via an extreme learning machine that optimizes the total error rate. Empirically, we have observed that extracting horizontal local facial features could be beneficial to alleviating verification error incurred from horizontal pose variations. Moreover, a fusion of these directional features with an appropriate classifier design can overcome the performance limitation imposed by each single modality.

In our future work, an analysis of the group size l which controls the amount of information extracted would be beneficial towards a better understanding of its impact on recognition performance.

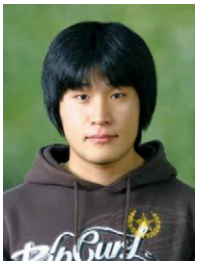
Acknowledgements

This work was supported by the National Research Foundation of Korea (NRF) through the Biometrics Research Engineering Center (BERC) at Yonsei University (Grant Number: R112002105090010 (2010) and R112002105080020 (2010)).

References

[1] W. Zhao, R. Chellappa, P. Phillips, A. Rosenfeld, Face recognition: a literature survey, ACM Computing Surveys 35 (4) (2003) 399–458 (CSUR).
[2] M. Turk, A. Pentland, Eigenfaces for recognition, Journal of Cognitive Neuroscience 3 (Winter 1991) 71–86.
[3] P.N. Belhumeur, J.P. Hespanha, D.J. Kriegman, Eigenfaces vs. Fisherfaces: recognition using class specific linear projection, IEEE Transactions on Pattern Analysis and Machine Intelligence 19 (7) (1997) 711–720.
[4] R. Gottumukkal, V.K. Asari, An improved face recognition technique based on modular PCA approach, Pattern Recognition Letters 25 (4) (2004) 429–436.
[5] R. Brunelli, T. Poggio, Face recognition: features versus templates, IEEE Transactions on Pattern Analysis and Machine Intelligence 15 (10) (1993) 1042–1052.

[6] B. Heisele, T. Serre, T. Poggio, A component-based framework for face detection and identification, International Journal of Computer Vision 74 (2) (2007) 167–181.
[7] J. Vogel, A. Schwaninger, C. Wallraven, H.H. Bülthoff, Categorization of natural scenes: local vs global information, in: Proceedings of the 3rd Symposium on Applied Perception in Graphics and Visualization, ACM, New York, NY, USA, 2006, pp. 33–40.
[8] J. Zou, Q. Ji, G. Nagy, A comparative study of local matching approach for face recognition, IEEE Transactions on Image Processing 16 (10) (2007) 2617–2628.
[9] L. Wiskott, J.-M. Fellous, N. Kuiger, C. von der Malsburg, Face recognition by elastic bunch graph matching, IEEE Transactions on Pattern Analysis and Machine Intelligence 19 (7) (1997) 775–779.
[10] K.-A. Toh, Deterministic neural classification, Neural Computation 20 (6) (2008) 1565–1595.
[11] K. Etemad, R. Chellappa, Discriminant analysis for recognition of human face images, Journal of the Optical Society of America A 14 (8) (1997) 1724–1733.
[12] J. Yang, D. Zhang, A.F. Frangi, J.-Y. Yang, Two-dimensional PCA: a new approach to appearance-based face representation and recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence 26 (1) (2004) 131–137.
[13] J. Yang, C. Liu, Horizontal and vertical 2DPCA based discriminant analysis for face verification using the FRGC version 2 database, in: International Conference on Biometrics 2007, pp. 838–847.
[14] W. Yu, Z. Wang, W. Chen, A new framework to combine vertical and horizontal information for face recognition, Neurocomputing 72 (4–6) (2009) 1084–1091.
[15] J. Ye, R. Janardan, Q. Li, Two-dimensional linear discriminant analysis, Advances in Neural Information Processing Systems 17 (2004) 1569–1576.
[16] D. Zhang, Z.-H. Zhou, (2D)^2PCA: Two-directional two-dimensional PCA for efficient face representation and recognition, Neurocomputing 69 (1–3) (2005) 224–231.
[17] M.A.O. Vasilescu, D. Terzopoulos, Multilinear analysis of image ensembles: tensorfaces, in: Proceedings of the European Conference on Computer Vision (ECCV) 2002, pp. 447–460.
[18] D. Tao, X. Li, X. Wu, S.J. Maybank, General tensor discriminant analysis and gabor features for gait recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence (2007) 1700–1715.
[19] J. Wang, A. Barreto, L. Wang, Y. Chen, N. Rishe, J. Andrian, M. Adjouadi, Multilinear principal component analysis for face recognition with fewer features, Neurocomputing 73 (2010) 1550–1555.
[20] A.M. Martínez, Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class, IEEE Transaction on Pattern Analysis and Machine Intelligence 24 (6) (2002) 748–763.
[21] L.V. der Maaten, E. Postma, H.V. den Herik, Dimensionality reduction: a comparative review, Technical Report, MICC, Maastricht University, 2008.
[22] P.J. Phillips, H. Moon, S.A. Rizvi, P.J. Rauss, The FERET evaluation methodology for face-recognition algorithms, IEEE Transactions on Pattern Analysis and Machine Intelligence 22 (10) (2000) 1090–1104.
[23] S. Chen, Y. Zhu, Subpattern-based principle component analysis, Pattern Recognition 37 (5) (2004) 1081–1083.
[24] D. Achlioptas, Database-friendly random projections, in: ACM Symposium on the Principles of Database Systems, ACE Press, 2001, pp. 274–281.
[25] Y.S. Kim, K.-A. Toh, Sparse random projection for efficient cancelable face feature extraction, in: Proceedings of the IEEE Conference on Industrial Electronics and Applications, Singapore, June 2008, pp. 2139–2144.
[26] S. Yan, D. Xu, B. Zhang, H.-J. Zhang, Q. Yang, S. Lin, Graph embedding and extensions: a general framework for dimensionality reduction, IEEE Transactions on Pattern Analysis and Machine Intelligence (2007) 40–51.
[27] T. Zhang, D. Tao, X. Li, J. Yang, Patch alignment for dimensionality reduction, IEEE Transactions on Knowledge and Data Engineering (2008) 1299–1313.
[28] T. Zhou, D. Tao, X. Wu, Manifold elastic net: a unified framework for sparse dimension reduction, Data Mining and Knowledge Discovery 22 (3) (2010) 340–371.
[29] A. Pentland, B. Moghaddam, T. Starner, View-based and modular eigenspaces for face recognition, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1994. Proceedings of CVPR'94, pp. 84–91.
[30] F. Al-Osaimi, M. Bennamoun, A. Mian, Integration of local and global geometrical cues for 3D face recognition, Pattern Recognition 41 (3) (2008) 1030–1040.
[31] G.-B. Huang, Q.-Y. Zhu, C.-K. Siew, Extreme learning machine: theory and applications, Neurocomputing 70 (1–3) (2006) 489–501.
[32] G.-B. Huang, L. Chen, C.-K. Siew, Universal approximation using incremental constructive feedforward networks with random hidden nodes, IEEE Transactions on Neural Networks 17 (4) (2006) 879–892.
[33] G.-B. Huang, L. Chen, Convex incremental extreme learning machine, Neurocomputing 70 (16–18) (2007) 3056–3062.
[34] G.-B. Huang, Q.-Y. Zhu, K. Mao, C.-K. Siew, P. Saratchandran, N. Sundararajan, Can threshold networks be trained directly? IEEE Transactions on Circuits and Systems II: Express Briefs 53 (3) (2006) 187–191.
[35] AT&T Laboratories Cambridge, Olivetti Research Laboratory (ORL) database. 〈http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html〉 (accessed June 24, 2010).
[36] S.-K. Kim, H. Lee, S. Yu, S. LEE, Robust face recognition by fusion of visual and infrared cues, in: Proceedings of IEEE Conference on Industrial Electronics and Applications, Singapore, May 2006, pp. 804–808.
[37] H. Wechsler, P. Phillips, V. Bruce, F. Fogelman-Soulie, T. Huang, Em characterizing virtual eigensignatures for general purpose face recognition, in: D.B. Graham, N.M. Allinson (Eds.), Face Recognition: From Theory to Applications, 1998, pp. 446–456.
[38] S. Noushath, G.H. Kumar, P. Shivakumara, (2D)^2LDA an efficient approach for face recognition, Pattern Recognition 39 (7) (2006) 1396–1400.
[39] A.K. Jain, K. Nandakumar, A. Ross, Score normalization in multimodal biometric systems, Pattern Recognition 38 (12) (2005) 2270–2285.
[40] National Institute of Standards and Technology (NIST), The Facial Recognition Technology (FERET) database. 〈http://www.itl.nist.gov/iad/humanid/feret/feret_master.html〉 (accessed June 24, 2010).
[41] The University of Sheffield, The Sheffield (previously UMIST) Face Database. 〈http://www.shef.ac.uk/eee/research/iel/research/face.html〉 (accessed June 24, 2010).
[42] The Math Works, MATLAB. 〈http://www.mathworks.com/〉.



Beom-Seok Oh received B.S. degree in computer science from KonKuk university, Korea, in 2008 and M.S. degree in electrical and electronic engineering from Yonsei University, Seoul, Korea, in 2010. He is currently a Ph.D. student in the School of Electrical and Electronic Engineering at Yonsei University. His research interests include biometric, pattern recognition, and machine learning.

Kar-Ann Toh is a full professor in the School of Electrical and Electronic Engineering at Yonsei University, South Korea. He received the Ph.D. degree from Nanyang Technological University (NTU), Singapore. He worked for two years in the aerospace industry prior to his post-doctoral appointments at research centres in NTU from 1998 to 2002. He was affiliated with Institute for Infocomm Research in Singapore from 2002 to 2005 prior to his current appointment in Korea. His research interests include biometrics, pattern classification, optimization and neural networks. He is a co-inventor of a US patent and has made several PCT filings related to biometric applications. Besides being an active member in publications, Dr. Toh has served as a member of technical program committee for international conferences related to biometrics and artificial intelligence. He has also served as a reviewer for international journals including several IEEE Transactions. He is currently an associate editor of Pattern Recognition Letters and a senior member of the IEEE.

Andrew Beng Jin Teoh obtained his B. Eng (Electronic) in 1999 and Ph.D. degree in 2003 from National University of Malaysia. Currently, he is an assistant professor in Department of Electrical and Electronic Engineering, Yonsei University. His research interest is in biometrics security and pattern recognition. He had published around 140 international journal and conference papers in his area.

Jaihie Kim received the PhD degree in electrical engineering in 1984 from the Case Western Reserve University, USA. Since 1984, he has been a professor in the School of Electrical and Electronic Engineering, Yonsei University in Korea, and currently he is also the Director of Biometrics Engineering Research Center in Korea. Dr. Jaihie Kim is a member of the National Academy of Engineering of Korea. His research interests include pattern recognition, computer vision, and biometrics.