Pattern Recognition 45 (2012) 3288–3303

Contents lists available at SciVerse ScienceDirect

Pattern Recognition

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

Extraction and fusion of partial face features for cancelable identity verification

Beom-Seok Oh [a], Kar-Ann Toh [a,*], Kwontaeg Choi [b], Andrew Beng Jin Teoh [a], Jaihie Kim [a]

[a] School of Electrical and Electronic Engineering, Yonsei University, Seoul, Republic of Korea
[b] Department of Computer Science, Yonsei University, Seoul, Republic of Korea

ARTICLE INFO

Article history:
Received 13 July 2011
Received in revised form
16 February 2012
Accepted 20 February 2012
Available online 5 March 2012

Keywords:
Face identity verification
Partial face features
Local feature extraction
Cancelable biometrics
Match scores fusion

ABSTRACT

In this paper, we propose to extract localized random features directly from partial face image matrix for cancelable identity verification. Essentially, the extracted random features consist of compressed horizontal and vertical facial information obtained from a structured projection of the raw face images. For template security reason, the face appearance information is concealed via averaging several templates over different transformations. The match score outputs of these cancelable templates are then fused through a total error rate minimization. Extensive experiments were carried out to evaluate and benchmark the performance of the proposed method based on the AR, FERET, ORL, Sheffield and BERC databases. Our empirical results show encouraging performances in terms of verification accuracy as well as satisfying four cancelable biometric properties.

1. Introduction

Face recognition is among the most attractive research topics in biometrics. Attributed to its non-intrusiveness and high universality, face recognition can be useful in various applications such as automated border control (e.g. face verification with an e-passport), identity authentication at ATM and automated access control. However, there remained many technical challenges before an automated face recognition system can be widely deployed. These challenges include high data dimensionality according to image resolution, sensitivity to imaging conditions (e.g. variations of pose, illumination, expression, etc.), small-sample-size (SSS) problem, etc. Processing a high-dimensional raw data for face recognition can be problematic since it consumes large computing resources, triggers the SSS problem [1], and often contains noise. Particularly, the large intra-identity variation caused by imaging conditions is an important problem that hurdles the wide implementation of automatic face recognition system. A solution to these high-dimensionality and sensitivity problems can be related to extraction of effective facial features which represent relevant facial information. Although many works have been done in the field during the last two decades, there remained much room for improvement.

1.1. Background

Conventional face feature extraction methods such as principal component analysis (PCA) [2] and linear discriminant analysis (LDA) [1] operate on vector based face space. They are known to work well for images with frontal view [3], and they are widely used in many research fields in pattern recognition and computer vision. However, such conventional vector based methods correspond to an extraction of global facial features which can be sensitive to imaging factors such as variations of illumination, pose, expression and so on. Besides the sensitivity issue, these methods can be computationally expensive since the vectorized face data is usually of high dimension [3].

An attempt to solve the high-dimensionality and sensitivity problems is to extract features based on partitioned image. Modular PCA (mPCA) [4] and sub-pattern PCA (SpPCA) [5] are two representative partitioning based methods. These methods not only alleviate the high-dimensionality problem [4,5], but also extract local facial features which are less sensitive to imaging conditions than that using global features [4]. However, finding an optimal number of sub-images to be partitioned is not obvious.

Another attempt is to perform feature extraction directly on an image matrix (vectorization of data is not required) as seen in two-dimensional PCA (2DPCA) [6] and two-dimensional LDA (2DLDA) [7] methods. Since this approach works on image matrix directly, the high-dimensionality problem in vector based method is well resolved and this results in a better verification performance than that of conventional PCA and LDA [6,7].

* Corresponding author. Tel.: +82 2 2123 5864; fax: +82 2 312 4584.
E-mail address: katoh@yonsei.ac.kr (K.-A. Toh).

0031-3203/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2012.02.027

Although the partition and image matrix based methods can resolve some problems observed in the conventional vector based methods, these training based methods (including the vector based methods) can be vulnerable in terms of security. Due to the permanent characteristic of biometrics and the unique template generation, once a biometric template is compromised, it is compromised forever [8]. To deal with such biometric template compromisation, cancelable biometric [9], which intentionally distorts biometrics information in a repeatable but non-reversible manner, has been proposed. A research issue in cancelable biometrics is to extract secure biometric features that satisfies several criteria such as non-invertibility, diversity, re-usability and performance [8]. In this work, we investigate into generating a face biometric template which can tolerate the sensitivity factor, reduce the dimension and satisfies these cancelable criteria.

1.2. Motivation and contributions

Random projection (RP) method [10] is among the popular choices for cancelable feature extraction [8,11,12]. It is a non-training based method and extracts random features. Although RP is widely utilized for cancelable feature extraction, it has some limitations. The extracted RP features do not contain relevant geometrical information. Moreover, RP is limited in terms of verification performance [10].

With an aim of extracting more discriminative random facial features than RP, Oh et al. [13,14] proposed two structured random projections which correspond to extraction of horizontal and vertical facial features. Since the structured projections work directly on image matrix, they include geometrical facial information where conventional RP does not have. The authors of [14] have empirically shown that the horizontal projection is more robust to large horizontal pose variation than that of the vertical projection and some unsupervised learning methods such as PCA. However, these structured projections have shown a poorer verification performance than that of LDA which is a supervised learning based method. Moreover, these projections were not explored for cancelable biometric applications in [14].

To resolve the template security issue as well as to explore into local image properties for performance enhancement, this paper investigates into extraction of cancelable sub-face image features via an image partitioning. Different from the method in [13,14], we propose to extract localized facial features by projecting the structured random matrices onto partial images. For appearance concealment, the extracted feature templates are averaged over different transformations at feature level. The averaged cancelable templates are subsequently fused at score level via a total error rate minimization which adopts a reduced multivariate (TERRM) model [15] for performance enhancement. The cancelability and robustness against external imaging factors (particularly, regarding pose variation) of the proposed projections will be evaluated under two scenarios using five databases.

The main contributions of this work are enumerated as follows: (i) proposal of a method to directly extract localized random facial features based on partial face image; (ii) proposal of new cancelable facial templates via averaging facial information over different transformations at feature level; (iii) fusion of the partial cancelable templates via TERRM at score level; (iv) provision of extensive experiments comparing the proposed method with related methods in the literature.

1.3. Organization

The rest of this paper is organized as follows: some background knowledge on related feature extraction methods which operate directly on image matrix, and some error rate measurements are briefly discussed in Section 2 for immediate reference. Section 3 describes our proposed method on extraction of partial-face image features. Extensive empirical results and discussions on experimental protocols are provided in Section 4. Section 5 provides an analysis of the proposed method in terms of feature distribution and cancelability. Some concluding remarks are given in Section 6.

2. Preliminaries

In this section, we briefly describe some feature extraction methods which operate directly on image matrix and a total error rate minimization method [15] which will be used for fusion, for immediate reference.

2.1. Principal components analysis (PCA) based on image matrix

Let $\mathbf{X}_i$ ($i=1,2,\ldots,n$ number of training samples) be p-by-q image matrices which are reshaped into vectors $\mathbf{x}_i \in \mathbb{R}^{pq\times 1}$ and stacked as $\Gamma=[\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_n]$. Consider a linear projection $\mathbf{Y}=\mathbf{U}^T\Gamma$ which maps the pq-dimensional data set Γ in image space onto a d-dimensional feature space $\mathbf{U}=[\mathbf{u}_1,\mathbf{u}_2,\ldots,\mathbf{u}_d]$, where $\mathbf{U}\in\mathbb{R}^{pq\times d}$ and $pq\gg d$. Here, the goal behind principal components analysis (PCA) [2] is to find the best projection vectors (directions) $\mathbf{U}$ that maximizes the determinant of a total scatter matrix (also called covariance matrix) across all the image samples [2]. The covariance matrix $\mathbf{G}_{PCA}$ is given by

$$\mathbf{G}_{PCA}=\sum_{i=1}^{n}(\mathbf{x}_i-\mathbf{m})(\mathbf{x}_i-\mathbf{m})^T, \tag{1}$$

where $\mathbf{G}_{PCA}\in\mathbb{R}^{pq\times pq}$, $\mathbf{m}=(1/n)\sum_{i=1}^{n}\mathbf{x}_i$ is a sample mean of training set and T denotes a transpose. Since maximizing the determinant of the projected samples is equivalent to calculating the eigenvectors of $\mathbf{G}_{PCA}$ which correspond to those largest eigenvalues, the best projection vectors $\mathbf{U}$ consist of top d number of eigenvectors.

Here, transforming image matrices $\mathbf{X}_i\in\mathbb{R}^{p\times q}$ into vectors $\mathbf{x}_i\in\mathbb{R}^{pq\times 1}$ is not only computationally expensive but also produces a high-dimensional input data which makes it difficult to estimate the covariance matrix accurately [16]. In order to alleviate the high-dimensionality problem, Yang et al. [6] proposed horizontal two-dimensional PCA (H2DPCA) which performs PCA directly on the image matrix instead of using a transformed image vector. With a similar goal of PCA, H2DPCA finds the most discriminative projection vectors $\mathbf{V}=\{\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_h\}$ $h<n$ for a linear projection $\mathbf{Y}=\mathbf{XV}$. The set of vectors $\mathbf{V}$ are the top h number of eigenvectors of a total scatter matrix of projected samples $\mathbf{G}_{H2DPCA}$ given by

$$\mathbf{G}_{H2DPCA}=\frac{1}{n}\sum_{i=1}^{n}(\mathbf{X}_i-\mathbf{M})^T(\mathbf{X}_i-\mathbf{M}), \tag{2}$$

where $\mathbf{M}=(1/n)\sum_{i=1}^{n}\mathbf{X}_i$ and $\mathbf{G}_{H2DPCA}\in\mathbb{R}^{q\times q}$ (recall that $\mathbf{X}\in\mathbb{R}^{p\times q}$). By definition of the matrix inner product in (2), this horizontal 2DPCA works only in the horizontal (row) direction of the image matrix [17].

As an alternative to H2DPCA, vertical 2DPCA (V2DPCA) [17,18] takes a transposed face image matrix $\mathbf{X}^T$ as an input, then the image covariance matrix (2) is simply re-written in an alternative form as follows:

$$\mathbf{G}_{V2DPCA}=\frac{1}{n}\sum_{i=1}^{n}(\mathbf{X}_i-\mathbf{M})(\mathbf{X}_i-\mathbf{M})^T, \tag{3}$$

where $\mathbf{G}_{V2DPCA} \in \mathbb{R}^{p\times p}$. The vertical 2DPCA works only in the vertical (column) direction of a face image matrix [17]. Similar with PCA and horizontal 2DPCA, the optimal projection vectors $\mathbf{Z}=\{\mathbf{z}_1,\mathbf{z}_2,\ldots,\mathbf{z}_f\}$ are the eigenvectors of $\mathbf{G}_{V2DPCA}$ which corresponds to the largest eigenvalues.

For a given new test image $\mathbf{X}_t$, the horizontal and vertical 2DPCA features can respectively be extracted by $\mathbf{A}=\mathbf{X}_t\mathbf{V}\in\mathbb{R}^{p\times h}$ and $\mathbf{B}=\mathbf{Z}^T\mathbf{X}_t\in\mathbb{R}^{f\times q}$, where h and f are the number of principal component vectors.

2.2. Total error rate (TER) minimization

False acceptance rate (FAR) is the rate an un-enrolled person has been falsely authorized as a genuine-user, and false rejection rate (FRR) is referred to as the rate that an enrolled person has been falsely rejected. According to [15]

$$\mathrm{FAR}=\frac{1}{m^-}\sum_{j=1}^{m^-}L(g(\mathbf{x}_j^-)\geq\tau),\quad \mathrm{FRR}=\frac{1}{m^+}\sum_{i=1}^{m^+}L(g(\mathbf{x}_i^+)<\tau), \tag{4}$$

where m^- and m^+ respectively denote the imposter and the genuine populations ($m^-+m^+=m$ is the total training population), g is a biometric matcher, τ is the decision threshold, and $L(g(\mathbf{x}_j^-)\geq\tau)=1$ whenever $g(\mathbf{x}_j^-)\geq\tau$ is true and zero otherwise (conversely, $L(g(\mathbf{x}_i^+)<\tau)=1$ whenever $g(\mathbf{x}_i^+)<\tau$ is true and zero otherwise) [15]. $\mathbf{x}^-$ denotes the feature vector of an imposter and $\mathbf{x}^+$ denotes the feature vector of a genuine-user.

According to Toh et al. [15], for a classifier which is linear in its parameters ($\boldsymbol{\alpha}$), and with appropriate normalization plus inclusion of an offset term, a quadratic approximation to the step loss function $L(\cdot)$ can be utilized [15] giving a closed-form solution to TER minimization:

$$\boldsymbol{\alpha}=\left[b\mathbf{I}+\frac{1}{m^-}\sum_{j=1}^{m^-}\mathbf{P}_j^T\mathbf{P}_j+\frac{1}{m^+}\sum_{i=1}^{m^+}\mathbf{P}_i^T\mathbf{P}_i\right]^{-1}\times\left[\frac{(\tau-\eta)}{m^-}\sum_{j=1}^{m^-}\mathbf{P}_j^T+\frac{(\tau+\eta)}{m^+}\sum_{i=1}^{m^+}\mathbf{P}_i^T\right], \tag{5}$$

where $\mathbf{P}_j\triangleq\mathbf{P}(g(\mathbf{x}_j^-))$ and $\mathbf{P}_i\triangleq\mathbf{P}(g(\mathbf{x}_i^+))$ correspond to certain kernel regressors (such as a reduced polynomial model used in [19]), $g(\mathbf{x})$ is biometric classifier output and η corresponds to an arbitrary offset. $\mathbf{I}$ is an identity matrix with similar dimension to $\mathbf{P}^T\mathbf{P}$. In order to predict the label of a new sample $\mathbf{x}_t$, the estimated $\boldsymbol{\alpha}$ can be used to compute the predicted output in $f(\mathbf{x}_t)=\mathbf{P}_t\boldsymbol{\alpha}$, where $\mathbf{P}_t$ is test regressors generated from $\mathbf{x}_t$.

3. Proposed method

In this paper, we propose to extract partial-face image features based on two directional projections. Essentially, these two directional projections correspond to extraction of horizontal and vertical facial features from the two-dimensional partial-face images which are obtained via an image matrix partitioning. Next, we take an average of each directional features over several (say z number of features) transformations at feature level to generate a cancelable face template. Finally, the decision outputs of these cancelable face templates are fused at score level via a total error rate minimization that adopts a reduced multivariate (TERRM) model [15,20]. Each of these processes will be described in greater detail in the following subsections.

3.1. Directional projections on image matrix

The two directional projections correspond to a matrix that pre-multiplies the face image matrix and a matrix that post-multiplies the face image matrix. For simplicity, we will call the former $\mathbf{R}_1$ projection and the later $\mathbf{R}_2$ projection [13].

Considering the conformation of vector-matrix product, the $\mathbf{R}_1$ matrix is constructed by stacking row vectors having a similar size with the height (p) of the image $\mathbf{X}\in\mathbb{R}^{p\times q}$. Here, we can have an arbitrary k number of rows for $\mathbf{R}_1$ projection since it is pre-multiplied to the face image $\mathbf{X}$. Our design of such rows lies in the arrangement of the binary values "0" and "1" for signal blockage and passage. To extract face features which are more than a pixel wide, we construct each row with a single horizontal group of 1's with the remaining elements zeros. This is as illustrated in Fig. 1 (a). On the other hand, the $\mathbf{R}_2$ projection matrix consists a stack of (k number of) column vectors which has the same dimension as the image width (q). Similar to the $\mathbf{R}_1$ matrix, each column vector has a single vertical group of 1's with the remaining elements zeros as depicted in Fig. 1 (b). As a design parameter, the size of each group of 1's in both $\mathbf{R}_1$ and $\mathbf{R}_2$ is denoted as l. The position of each group of 1's is decided randomly. The amount of extracted features is dependent on the projection size k, and the group size l [14].

3.2. Extraction of partial-face image features

3.2.1. Extraction of vertical partial-face image features

As mentioned, the $\mathbf{R}_1$ matrix will be directly used in projecting the partial-face images. For convenience, we will call the proposed method p$\mathbf{R}_1$ (an abbreviation form of partitioned $\mathbf{R}_1$ projection) in subsequent use. Since the $\mathbf{R}_1\in\mathbb{R}^{k\times p}$ matrix is pre-multiplied to the original face image matrix $\mathbf{X}_i\in\mathbb{R}^{p\times q}$ and due to the discrete matrix size, an image can only be partitioned vertically with respect to its width size q so that multiplication of the two matrices are conformable. For simplicity as well as without considering redundant information, only non-overlapping, symmetrical and equal sizes of partitioning are considered in this work. When q is not divisible by the number of partitions, we adopt a simple procedure to take care of the partitioning. Consider, say, $q=259$. For p$\mathbf{R}_1$ projection, the last column vector (259th column) of the image matrix $\mathbf{X}_i$ is removed to facilitate an equal split of the remaining image. This removal (for odd number of images columns) and splitting procedure is continued until the required level of splitting is obtained.

Consider a $\mathbf{R}_1$ projection matrix and a vertically partitioned face image matrix $\underline{\mathbf{X}}_i\in\mathbb{R}^{p\times(q/\#\ partition)}$ where '# *partition*' indicates the number of sub-images to be partitioned, then the projected template is given by

$$\mathbf{Y}_{v_i}=\mathbf{R}_1\cdot\underline{\mathbf{X}}_i,\quad i=1,\ldots,n, \tag{6}$$

where $\mathbf{Y}_{v_i}\in\mathbb{R}^{k\times(q/\#\ partition)}$, "$\cdot$" indicates a matrix inner product and n is the total number of images in the test set. Fig. 2(a) illustrates the procedure of p$\mathbf{R}_1$ projection operating on two sub-face images with $p=56$, $q=46$, projection size $k=5$ and non-zero

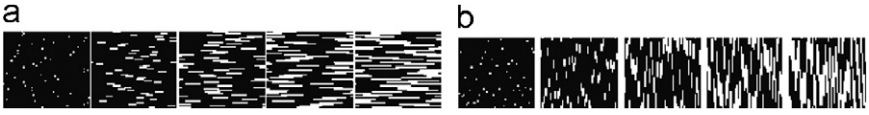

Fig. 1. Some sample examples of $\mathbf{R}_1$ (a) and $\mathbf{R}_2$ (b) matrices with different sparsities (highest sparsity on the left) where "0" and "1" pixels are shown in black and white, respectively.

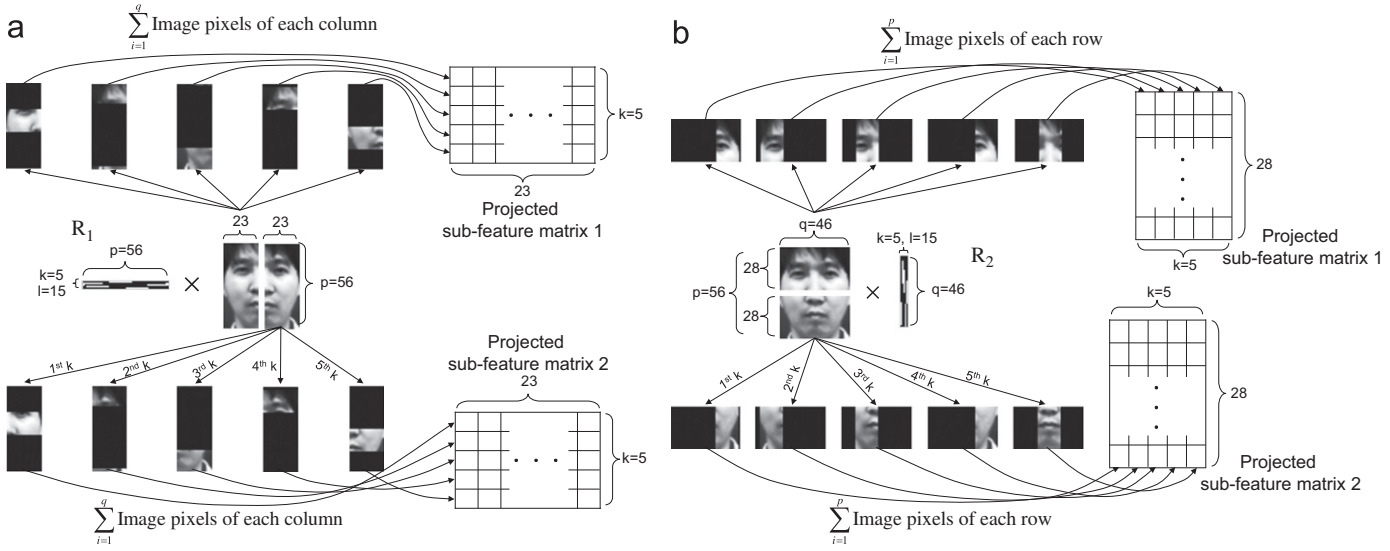

Fig. 2. An illustration of (a) p$\mathbf{R}_1$ projection on two (vertical) sub-face images and (b) p$\mathbf{R}_2$ projection on two (horizontal) sub-face images with $k=5$ and $l=15$.

group size $l=15$. As we can observe from the figure, a p$\mathbf{R}_1$ projection on each sub-image produces a feature matrix that consists of compressed vertical partial-face features.

3.2.2. Extraction of horizontal partial-face image features

Similar to p$\mathbf{R}_1$ projection, an $\mathbf{R}_2$ projection which operates on the partitioned sub-face images (will be called p$\mathbf{R}_2$ projection hereafter) is defined by

$$\mathbf{Y}_{h_i} = \underline{\mathbf{X}}_i \cdot \mathbf{R}_2, \quad i = 1, \ldots, n, \tag{7}$$

where $\underline{\mathbf{X}}_i \in \mathbb{R}^{(p/\#\ partition)\times q}$ is a horizontally partitioned sub-face image matrix, $\mathbf{R}_2 \in \mathbb{R}^{q\times k}$ and $\mathbf{Y}_{h_i} \in \mathbb{R}^{(p/\#\ partition)\times k}$ is a compressed horizontal sub-face feature matrix. Fig. 2(b) illustrates the overall procedure of p$\mathbf{R}_2$ projection. For similar reasons of simplicity as well as without considering redundant information, only non-overlapping, symmetrical and equal sizes of partitioning are considered as shown in Fig. 2(b). Similar to column splitting in p$\mathbf{R}_1$ case, the removal (for odd number of image rows) and splitting procedure is applied to p$\mathbf{R}_2$ case here.

3.3. Averaging of the directional facial templates

In order to hide the face appearance information, we propose to take an average of the extracted p$\mathbf{R}_1$ and p$\mathbf{R}_2$ templates over z number of different transformations at feature level:

$$\mathbf{Y}_{v_i} = \frac{1}{z}\sum_{j=1}^{z} \mathbf{R}_1^j \cdot \underline{\mathbf{X}}_i, \quad i = 1, \ldots, n, \tag{8}$$

where n is the total number of images in the test set. In a similar manner, the output template for p$\mathbf{R}_2$ projection can be written as

$$\mathbf{Y}_{h_i} = \frac{1}{z}\sum_{j=1}^{z} \underline{\mathbf{X}}_i \cdot \mathbf{R}_2^j, \quad i = 1, \ldots, n. \tag{9}$$

As the number of transformations z can vary, different sizes of $z \in \{1,2,\ldots,12\}$ will be evaluated to observe whether the proposed cancelable template can conceal the appearance information as well as fulfilling the four cancelable criteria such as non-invertibility, diversity, re-usability and performance [8].

3.4. Fusion of cancelable partial-templates at score level

The above extracted cancelable partial-face templates are finally fused at score level via a total error rate minimization adopting a reduced multivariate (TERRM) model [15]. 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. This Euclidean norm is known to suffer much from noise and outliers [21]. However, in view of its simplicity and intuitivity [22,23], as well as in order to compare directly with those results in [2,4–7,24], we shall adopt the Euclidean norm in our forthcoming experiments. The impacts of other distance measures such as cosine distance and correlation metric will be the subject of our future study.

Let $\mathbf{Y}_a \in \mathbb{R}^{k\times(q/\#\ partition)}$ and $\mathbf{Y}_b \in \mathbb{R}^{k\times(q/\#\ partition)}$ be the vertical partial-feature matrices based on p$\mathbf{R}_1$ projection from two images of one identity ($a=b$) (Similarly, for p$\mathbf{R}_2$ projection, $\mathbf{Y}_a \in \mathbb{R}^{(p/\#\ partition)\times k}$ and $\mathbf{Y}_b \in \mathbb{R}^{(p/\#\ partition)\times k}$). Then, the genuine-user match score is computed by

$$d(\mathbf{Y}_a, \mathbf{Y}_b) = \sqrt{\sum_{i=1}^{k}\sum_{j=1}^{q}(\mathbf{Y}_{a\,(i,j)} - \mathbf{Y}_{b\,(i,j)})^2}. \tag{10}$$

On the other hand, the match score belongs to an imposter if the images are taken from two different identities ($a \neq b$).

After generating genuine-user and imposter matching scores using (10) from each pair of cancelable partial-face templates, they are fused via TERRM [15,20]. As illustrated in Fig. 3, the number of partial-face feature scores N to be fused can vary. Various combinations of partitions and projection sizes will be investigated for fusion in our experiments to observe the most desired combination in terms of verification performance.

3.5. Summary of the proposed method

Fig. 3 shows an illustration of the overall procedure of our proposed method. The overall steps involved are summarized as follows:

(1) Generate z number of different $\mathbf{R}_1$ and $\mathbf{R}_2$ projection matrices and store them.

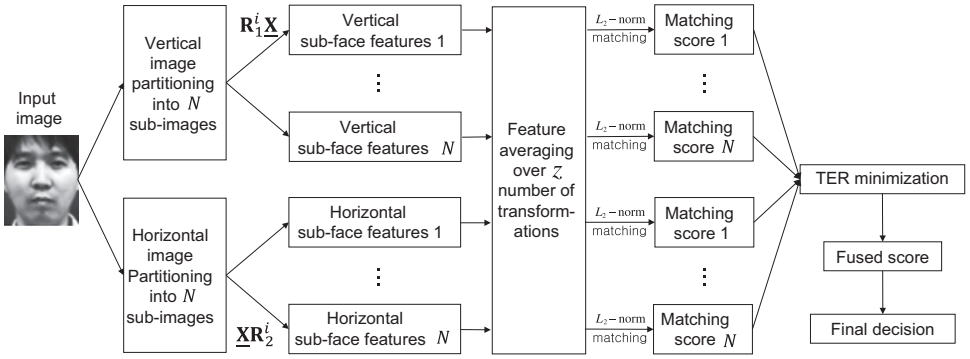

Fig. 3. An overview for the proposed p$\mathbf{R}_1$ and p$\mathbf{R}_2$ projections for $i=1,\ldots,z$, where z is the number of $\mathbf{R}_1$ and $\mathbf{R}_2$ projection matrices, and N is the number of partitioned partial-face images.

(2) Divide each face image into partial-face images $\underline{\mathbf{X}}$ according to the projection matrix size. For example, in p$\mathbf{R}_1$ projection, the face image can be partitioned vertically into arbitrary sizes. For p$\mathbf{R}_2$ projection, the partitioning of face images can only be performed horizontally.
(3) Perform z times of p$\mathbf{R}_1$ (8) and p$\mathbf{R}_2$ (9) projections on the above partial-face images using the stored z number of projection matrices to obtain the compressed directional partial-face image features.
(4) Take an average of the z number of p$\mathbf{R}_1$ and p$\mathbf{R}_2$ templates respectively at feature level.
(5) Obtain matching scores from each pair of cancelable sub-feature templates using inter- and intra-identity comparisons based on the Euclidean distance measure (10). The scores are normalized to within [0,1].
(6) Compute the fusion learning parameters $\hat{\boldsymbol{\alpha}}$ (5) based on a training set consisting of genuine-user and imposter scores. Use the estimated $\hat{\boldsymbol{\alpha}}$ to compute the predicted fusion scores for test data using $f(\mathbf{x}_t)=\mathbf{P}_t\hat{\boldsymbol{\alpha}}$ where $\mathbf{P}_t$ is a regressor matrix obtained from test data.
(7) Compute the final decision based on a preset threshold (as selected in the minimization process during training) by

$$cls(f(\mathbf{x}_t))=\begin{cases}1 & \text{if } f(\mathbf{x}_t)\geq\tau\\ 0 & \text{if } f(\mathbf{x}_t)<\tau\end{cases}$$

where $\mathbf{x}_t$ is sample test data, $f(\mathbf{x}_t)$ is the fusion classifier output and τ is the threshold.

4. Experiments

In order to evaluate the verification performance of our proposed method, extensive experiments have been conducted on five databases: AR [25], BERC [26], AT & T (ORL) [27], Sheffield (previously called UMIST database) [28] and FERET [29]. We note here that all images of these databases were taken under controlled environments. Our method thus relies upon the assumption that face detection and alignment have been performed reasonably well (i.e. within 10 pixels of spatial misalignment for an input image with 384×256 resolution based on our empirical observation) in the pre-processing step.

For benchmarking purpose, the verification results will be compared with several existing face recognition methods such as random projection (RP) [24], principal component analysis (PCA) [2], modular PCA (mPCA) [4], sub-pattern PCA (SpPCA) [5], two-dimensional PCA (2DPCA) [6] and two-dimensional linear discriminant analysis (2DLDA) [7] in terms of equal error rate (EER) and CPU time performances. The RP is included in the comparison to benchmark related random projection methods. Since PCA is among the most popular face recognition algorithms, it is chosen as a representative algorithm for one-dimensional image (vectorized) based face recognition methods. For comparability reason, the 2DPCA, 2DLDA, mPCA and SpPCA are included in our experiments to benchmark over those 2D-based and partition-based face recognition methods.

On top of the comparisons above, the proposed method will also be compared with a fusion of horizontal 2DPCA and vertical 2DPCA (will be called HV2DPCA hereafter) and a fusion of horizontal 2DLDA and vertical 2DLDA (will be called HV2DLDA hereafter) in the performance benchmarking. Similar to the proposed method, both methods adopt TERRM for fusion of the match score outputs. The purpose of conducting these comparisons is to observe performance enhancement induced by fusion of horizontal and vertical facial features at score level.

4.1. Database and pre-processing

Table 1 provides a summary of the five databases adopted in our experiments. As seen from the table, all the databases have different imaging conditions and data sizes. According to the table, AR, BERC and ORL are selected for evaluation of the proposed method under variations of illumination, expression and minor pose. The Sheffield and FERET databases are selected for evaluation under large variation in horizontal pose condition. Due to insufficient data in terms of glasses and time variations (see Table 1, G‡'s in the rightmost column), we shall exclude evaluation under these conditions.

The original RGB face images are of 768×576 resolution for the AR database, 320×240 resolution for the BERC database and 384×256 resolution for the FERET database. The original images of ORL and Sheffield databases are of 119×92 resolution, with 256 grey level per pixel. These databases are normalized by centering the eyes and the lip. Since there exists no standardized way to crop face images with pose variation, the entire face region including face contour and forehead with hair is included in the normalization process. 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 [31].

4.2. Experimental setting

4.2.1. Evaluation scenarios

Table 2 shows the configuration of our experimental scenarios for the databases utilized. As shown in the table, the proposed method will be evaluated under a shared-token scenario where

Table 1
Summary of the five adopted databases. The last three columns show configurations of the data sets selected in this work.

DB	Database descriptions			Descriptions of the data set selected in this work		
	Total number of images (subjects)	Total number of images per subject	Imaging conditions	Adopted number of images (subjects)	Adopted number of images per subject	Adopted imaging conditions (number of imgs. or note)
AR	3726 (126)	26	I/E/T/S/SG	1,680 (120)	14	I(8)/E(6)/T(2 sessions[a])
BERC	5184 (96)	54	I/E/P/W/G	1,152 (96)	12	E(6)/P(6)/I(2 conditions[b])
ORL	400 (40)	10	I/E/P/G	400 (40)	10	I(mild)/ E(minor)/G[c]/P(within $\pm 15^{\circ}$)
Sheffield	564 (20)	19–48	Large pose from profile to frontal	200 (20)	10	Large pose from profile to frontal
FERET	14,126 (1,119)	2–105	I/E/T/P/R/G	1,800 (200)	9	Large pose from ($+60^{\circ}$ to -60°)

I, illumination; E, expression; T, time; P, pose; W, wind; R, race; G, glasses; S, scarf; and SG, sunglasses. '°' indicates degree of angle and '+' and '−' respectively means a subject faces to his left and right directions [30].

[a] Half of the images (four images per identity under I and three images per identity under E) were taken at session 1. The remaining half were captured during session 2 with similar imaging conditions comparing to session 1. There was a two weeks interval between the two sessions.

[b] Half of the images (three images per identity under P and three images per identity under E) are with frontal illumination, and the remaining half are with bi-directional illumination.

[c] The 14 subjects (118 images) out of the 40 subjects (400 images) are with glasses condition.

Table 2
Experimental configuration of the scenarios along with the used databases.

Experimental scenarios		Databases
Shared-token scenario (stolen token)	Case 1: comparing $p\mathbf{R}_s$ with $\mathbf{R}_s$ [13,14]	AR, BERC
	Case 2: comparing $p\mathbf{R}_s$ with other methods	AR, ORL, BERC
	Case 3: comparing $p\mathbf{R}_s$ with other methods under large horizontal pose variation	Sheffield, FERET
User-specific token scenario		AR, BERC

Note: $s \in \{1,2\}$ is an index for the proposed method.

every identities share a similar token[1] in authentication, and a user-specific scenario where each individual user adopts a specific token.

Under the shared-token scenario, seven experiments are carried out based on the five databases to investigate the robustness of the proposed method with respect to external imaging factors under three cases: (1) comparing the proposed $p\mathbf{R}_s$ with $\mathbf{R}_s$ [13,14] where $s \in \{1,2\}$; (2) comparing the proposed $p\mathbf{R}_s$ with other methods; and (3) under large horizontal pose variation. For the first case, AR and BERC databases will be adopted. For the second case, AR, ORL and BERC databases will be used, and for the last case, FERET and Sheffield databases will be adopted as they contain large pose variation. Our purpose of evaluating case (3) is to observe the proposed method under large pose variation. Under the second scenario of a secured token, we evaluate the verification capability of the proposed method using user-specific tokens based on the AR and BERC databases.

4.2.2. Experimental protocols

(i) Protocol for shared-token scenario:

The last three columns of Table 1 show the database configurations adopted in this work. As shown in table, we have respectively selected 120 subjects (14 images per identity) for AR database, and 96 identities (12 images per subject) for BERC database. For the ORL database, the entire images are used in this work. For the Sheffield database, we utilized a subset of the database which consists of 200 images from 20 identities. For the last experiment under the shared-token scenario, 200 identities with nine images per identity were selected from the FERET database.

Half of the chosen images—from **1** to **7** in Fig. 4(a) for AR; from **a** to **f** in Fig. 4(b) for BERC; from (1) to (5) in Fig. 4(c) for ORL; (a), (c), (e), (g) and (i) in Fig. 4(d) for Sheffield; and **ba**, **bc**, **be**, **bg** and **bi** in Fig. 4 (e) for FERET — will be used for training. The remaining half — from **14** to **20** in Fig. 4(a) for AR; from **g** to **l** in Fig. 4(b) for BERC; from (6) to (10) in Fig. 4(c) for ORL; (b), (d), (f), (h) and (j) in Fig. 4(d) for Sheffield; and **bb**, **bd**, **bf** and **bh** in Fig. 4(e) for FERET—will be used for testing.

(ii) Protocol for user-specific scenario:

Under this scenario, as tabulated in Table 2, subsets of AR and BERC are adopted with the same database configurations compared to the subsets of AR and BERC under the shared-token scenario, respectively.

(iii) Protocol for test experiments:

For statistical evidence, 10 runs of two-fold cross-validation test will be conducted where the average performances are recorded. For the test performances of RP, $\mathbf{R}_1$, $\mathbf{R}_2$, $p\mathbf{R}_1$ and $p\mathbf{R}_2$, another 30 trials will be performed using 30 different RP matrices for RP, and 90 different $\mathbf{R}_1$ and $\mathbf{R}_2$ matrices for $\mathbf{R}_1$, $\mathbf{R}_2$, $p\mathbf{R}_1$ and $p\mathbf{R}_2$ (note that we average three features ($z=3$) for appearance concealment. Hence in order to generate 30 sets of averaged random features using $\mathbf{R}_1$ and $\mathbf{R}_2$, we will need 30 such random matrices for the averaging purpose. This results in 90 $\mathbf{R}_1$ and $\mathbf{R}_2$ matrices).

4.2.3. Parameter settings

(i) Parameter settings for $\mathbf{R}_1$ and $\mathbf{R}_2$ projections [13,14]

[1] The token here refers to the random seed which is used to generate the proposed $\mathbf{R}_1$ and $\mathbf{R}_2$ projection matrices.

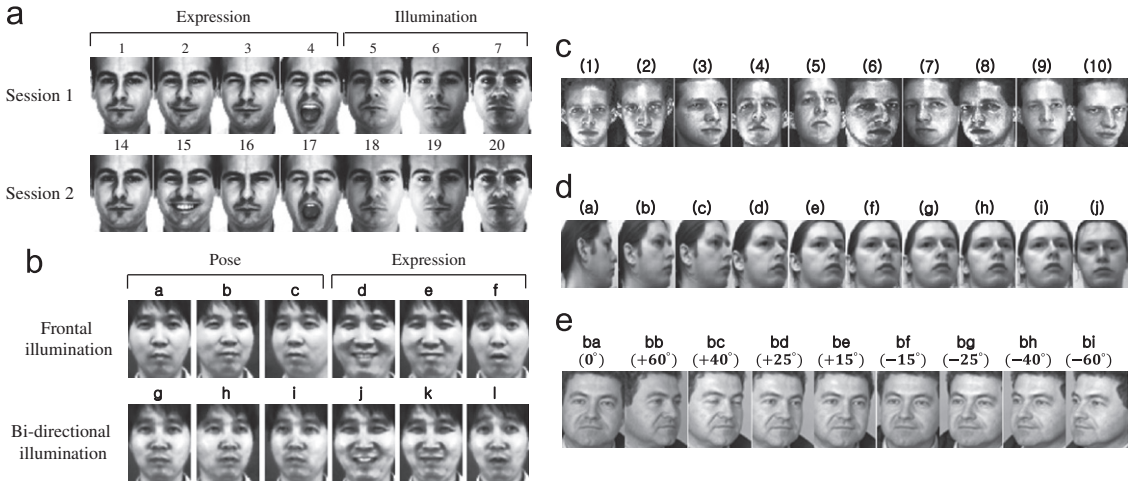

Fig. 4. Selected sample images from (a) AR, (b) BERC, (c) ORL, (d) Sheffield and (e) FERET.

Table 3
Optimal values for group size l and partition size N obtained from (30 trials using 90 different $\mathbf{R}_1$ and $\mathbf{R}_2$ matrices) 10 runs of two-fold cross-validation using only training set.

Database	Optimal group size, l		Number of sub-images, N	
	$\mathbf{R}_1$	$\mathbf{R}_2$	mPCA	SpPCA
AR	1	29	16	16
BERC	1	17	16	4
ORL	2	16	16	16
FERET	1	30	4	8
Sheffield	4	29	16	16

The $\mathbf{R}_1$ and $\mathbf{R}_2$ projections have two parameters namely, the projection size k and the group size l. The projection size k is the number of vectors (row vectors in $\mathbf{R}_1$ matrix and column vectors in $\mathbf{R}_2$ matrix) within the projection matrix. This k is related to the number of computations in inner product operation, and it determines the length of projected feature ($k \times 46$ in $\mathbf{R}_1$ and $56 \times k$ in $\mathbf{R}_2$). In [14], k has been empirically shown to have minor effect on the verification performance. On the other hand, it has been observed to significantly affect the computational cost. In order to balance between computational cost and verification performance, we follow [14] to adopt $k=10$ in this work.

The group size l is the number of 1's in each vector. Unlike k, l has a larger impact on recognition performance since it is directly related to the amount of information extracted from a face image block. The best values for the group size l were obtained based on 10 runs of two-fold cross-validation using only the training set by varying $l \in \{1,2,\ldots,30\}$. As seen from Table 3, the group size l is apparently dependent on the database. This is mainly due to the different imaging conditions of each database where representativeness of training data took effect. For example, the best chosen values for FERET and Sheffield databases are apparently similar since their images were both taken under large horizontal pose variation. In a similar manner, similar best l settings are observed for BERC and ORL databases where their images were acquired under variations of minor pose and expression.

For performance evaluation, we performed 30 projections using 90 different $\mathbf{R}_1$ and $\mathbf{R}_2$ matrices on top of the cross-validation in order to reduce those variations caused by the random property of $\mathbf{R}_1$ and $\mathbf{R}_2$ projection matrices. Thus, a total of 600 (10 runs $\times$ 2 fold $\times$ 30 trials) runs were performed for the cross-validation test. The optimal values are tabulated in the second and third columns of Table 3.

(ii) Parameter settings for p$\mathbf{R}_1$ and p$\mathbf{R}_2$ projections

The proposed p$\mathbf{R}_1$ and p$\mathbf{R}_2$ projections have additional parameters, the size of partitioning and the polynomial order r used in fusion. The size of partitioning is the number of image partitions in p$\mathbf{R}_1$ and p$\mathbf{R}_2$. Since $\mathbf{R}_1$ matrix is directly pre-multiplied to a face image which has a 56-by-46 resolution, the image can only be partitioned vertically such that multiplication of the two matrices are conformable. In other words, the 46-pixel width of the image is partitioned as {46(1),23(2),1(46)}, where $(\cdot)$ denotes the number of partitioned sub-images. In a similar manner, the image in p$\mathbf{R}_2$ projection can only be partitioned horizontally where the 56-pixel height of the image is partitioned as {56(1),28(2),14(4), 8(7),7(8),4(16),2(28), 1(56)} since $\mathbf{R}_2$ matrix is directly post-multiplied to the face image.

The sub-image sizes of 1(46) in p$\mathbf{R}_1$ and 4(16),2(28),1(56) in p$\mathbf{R}_2$ are too small for effective feature representation, hence we exclude them in our subsequent experiments. In addition, the number of sub-images of 8(7) and 7(8) in p$\mathbf{R}_2$ are quite similar, so only 7(8) will be experimented. p$\mathbf{R}_1$ and p$\mathbf{R}_2$ will be denoted respectively as p$\mathbf{R}_1(*)$ and p$\mathbf{R}_2(*)$ with $*$ denoting the number of partitioned sub-images in the subsequent experiments.

The polynomial order r is an adjustable parameter in a total error rate minimization that adopts a reduced multivariate polynomial (TERRM) [15]. Different polynomial orders $r \in \{1,2,\ldots,10\}$ will be experimented in our subsequent experiments to observe the effect on the verification performance. Following [20], the offset η and threshold τ are set at 0.5 where the desired target vector falls within $[0,1]$. The regularization constant is chosen at $b=0.001$ following [20].

(iii) Parameters for other methods

Similar to p$\mathbf{R}_1(*)$ and p$\mathbf{R}_2(*)$, partitioning the images into equal sub-images for mPCA [4] and SpPCA [5] methods is also limited to $N \in \{2,4,8,16,46,92,186\}$ where N denotes the number of sub-images, due to the size of adopted image. The optimal values for partition sizes N are obtained based on 10 runs of two-fold cross-validation using only training set similar to that in selecting parameters for p$\mathbf{R}_1$ and p$\mathbf{R}_2$. The optimal parameters for mPCA and SpPCA are tabulated in the last two columns of Table 3.

For fusion in HV2DPCA and HV2DLDA methods, similar to p$\mathbf{R}_s(*)$ projection, different polynomial order settings $r \in \{1,2,\ldots,10\}$ in TERRM will be experimented in our performance benchmarking.

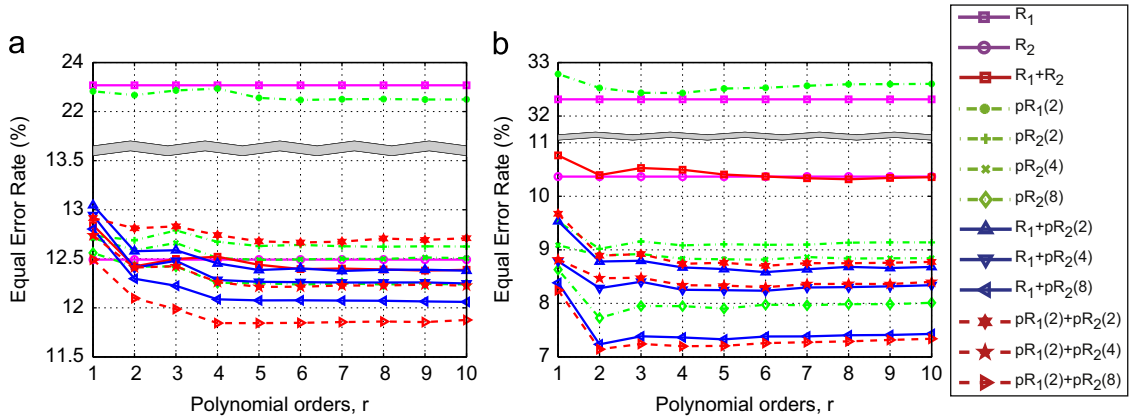

Fig. 5. Test EER performance comparison of $\mathbf{R}_1$, $\mathbf{R}_2$, $\mathbf{R}_1+\mathbf{R}_2$, $p\mathbf{R}_1(*)$ and $p\mathbf{R}_2(*)$ plotted over different polynomial orders $r \in \{1,2,\ldots,9,10\}$.

4.3. Results and discussion

4.3.1. Shared-token scenario (stolen token)

(i) Comparing the proposed $p\mathbf{R}_s$ with $\mathbf{R}_s$

Fig. 5 shows the performances of $\mathbf{R}_1$, $\mathbf{R}_2$, $\mathbf{R}_1+\mathbf{R}_2$, $p\mathbf{R}_1(*)$ and $p\mathbf{R}_2(*)$ in terms of average test EER performance, where '+' indicates a fusion at score level using TERRM. Individual $p\mathbf{R}_1(*)$ and $p\mathbf{R}_2(*)$ also indicates a fusion of '*' number of match score outputs via TERRM. For example, $p\mathbf{R}_1(2)+p\mathbf{R}_2(4)$ indicates a fusion of 6 match score outputs, and $p\mathbf{R}_2(8)$ indicates a fusion of eight match score outputs. Since $\mathbf{R}_1$ and $\mathbf{R}_2$ projections are not affected by the polynomial order r, their results are repeated 10 times in the plot to generate a line for ease of comparison.

As the figure shows, the performance of $\mathbf{R}_2$ projection is seen to significantly outperform that of $\mathbf{R}_1$ projection at the experimented projection size. Based on this, we can induce that the face images of the AR and BERC database include more discriminative information in horizontal direction ($\mathbf{R}_2$ projection) than that in vertical direction ($\mathbf{R}_1$ projection).

The results also indicate that a fusion of $\mathbf{R}_1$ and $\mathbf{R}_2$ using TERRM (denoted as $\mathbf{R}_1+\mathbf{R}_2$ in the figure) gives either similar or slightly enhanced EER performances than those of individual $\mathbf{R}_1$ and $\mathbf{R}_2$ projections over the entire range of r except at $r<5$. The $p\mathbf{R}_1(*)$ and $p\mathbf{R}_2(*)$ which fuses more image features yield an even better EER performance than that of $\mathbf{R}_1+\mathbf{R}_2$. Interestingly, the results show that the more partitions we fuse in $p\mathbf{R}_1(*)$ and $p\mathbf{R}_2(*)$ (i.e. $p\mathbf{R}_1(2)+p\mathbf{R}_2(8)$), the better the EER performance we get.

(ii) Comparing $p\mathbf{R}_s$ with other methods

Fig. 6 shows the test EER performances for RP, PCA, mPCA, SpPCA, 2DPCA, 2DLDA, HV2DPCA, HV2DLDA, $p\mathbf{R}_1(*)$ and $p\mathbf{R}_2(*)$. The performances of $p\mathbf{R}_1(*)$ and $p\mathbf{R}_2(*)$ are obtained from the variations of polynomial order $r \in \{1,2,\ldots,10\}$, where the best, mean and the worst EER performances are indicated for benchmarking purpose. The polynomial orders r for each $p\mathbf{R}_1(*)$ and $p\mathbf{R}_2(*)$ which have shown the best and the worst performances are tabulated in Table 4.

In Fig. 6(a) using AR database under Case 2, the image matrix-based methods (will be called 2D methods hereafter) such as 2DPCA, 2DLDA, HV2DPCA at $d'=\{1,2\}$, HV2DLDA and $p\mathbf{R}_1(*)+p\mathbf{R}_2(*)$ except for $p\mathbf{R}_1(2)$, significantly outperform the compared vector-based methods such as RP, PCA, mPCA and SpPCA. Among the 2D methods, HV2DLDA shows the best EER performance over entire range of d'. The best performances of 2DPCA and HV2DPCA (at $d'=2$) slightly outperform that of $p\mathbf{R}_1(*)+p\mathbf{R}_2(*)$. As mentioned in previous subsection, the proposed method yields the best performance when we fuse many features. From the best, mean and the worst EER performances among the variations of polynomial orders, we can observe that the performance deviation envelopes of $p\mathbf{R}_1(*)$ and $p\mathbf{R}_2(*)$ are small with respect to variation of r. As shown in Table 4, most of the worst EER performances come from $r<4$.

Fig. 6(b), again under Case 2 using BERC database, also shows superiority of HV2DLDA over entire range of d'. Interestingly, $p\mathbf{R}_1(*)$ and $p\mathbf{R}_2(*)$ gives comparable EER values with that of HV2DLDA except for $p\mathbf{R}_1(2)$. Moreover, the best performance of the proposed method at $p\mathbf{R}_1(2)+p\mathbf{R}_2(8)$ and $\mathbf{R}_1+p\mathbf{R}_2(8)$ outperform the best of HV2DLDA at $d'=2$ with tiny difference. The HV2DPCA outperforms 2DPCA over the entire range of d', and the best performance of HV2DPCA (observed at $d'=1$) outperforms the compared vector based methods such as RP, PCA, mPCA and SpPCA.

In Fig. 6(c) where the ORL database was experimented under Case 2, 2DLDA at $d'=2$ yields the best EER values compared to the other methods. Except at $d'=\{2,3\}$, HV2DLDA outperforms 2DLDA over the entire range of d'. The proposed method at $p\mathbf{R}_1(2)+p\mathbf{R}_2(4)$ shows slightly lower EER value than that of the best 2DPCA and HV2DPCA at $d'=2$. In contrast to the previous results (see Fig. 6(a) and (b)), interestingly, the best performance of the proposed method is obtained at $p\mathbf{R}_1(2)+p\mathbf{R}_2(4)$ rather than at $p\mathbf{R}_1(2)+p\mathbf{R}_2(8)$. Also, $p\mathbf{R}_2(4)$ slightly outperforms $p\mathbf{R}_2(8)$ in terms of EER.

To summarize the results in case 2, RP produced similar or slightly worse EER values than that of PCA (except at $d<15$ under ORL database). HV2DLDA and 2DLDA which are supervised learning methods showed superiority in terms of EER among the compared unsupervised learning methods. The proposed method showed comparable performance with that of the LDA methods except for the case of AR database.

(iii) Comparing $p\mathbf{R}_s$ with other methods under large horizontal pose variation

Under the third case of shared-token scenario (Fig. 6(d) and (e)), where we investigate the robustness of compared algorithms under large horizontal pose variation, HV2DLDA, 2DLDA and $p\mathbf{R}_1(2)+p\mathbf{R}_2(8)$ show superiority in terms of EER performance over all other compared methods. The proposed method, particularly $p\mathbf{R}_2(8)$, $\mathbf{R}_1+p\mathbf{R}_2(8)$ and $p\mathbf{R}_1(2)+p\mathbf{R}_2(8)$ on FERET database, and $p\mathbf{R}_1(2)+p\mathbf{R}_2(8)$ on Sheffield database, slightly outperform the best performance of HV2DLDA and 2DLDA (at $d'=1$ for both databases). RP yields $3\sim6\%$ better EER values than PCA, mPCA, SpPCA, 2DPCA and HV2DPCA except at $d'<6$ under FERET database. On the other hand, RP shows $2\sim6\%$ higher EER values than these compared methods under Sheffield database at $d<16$.

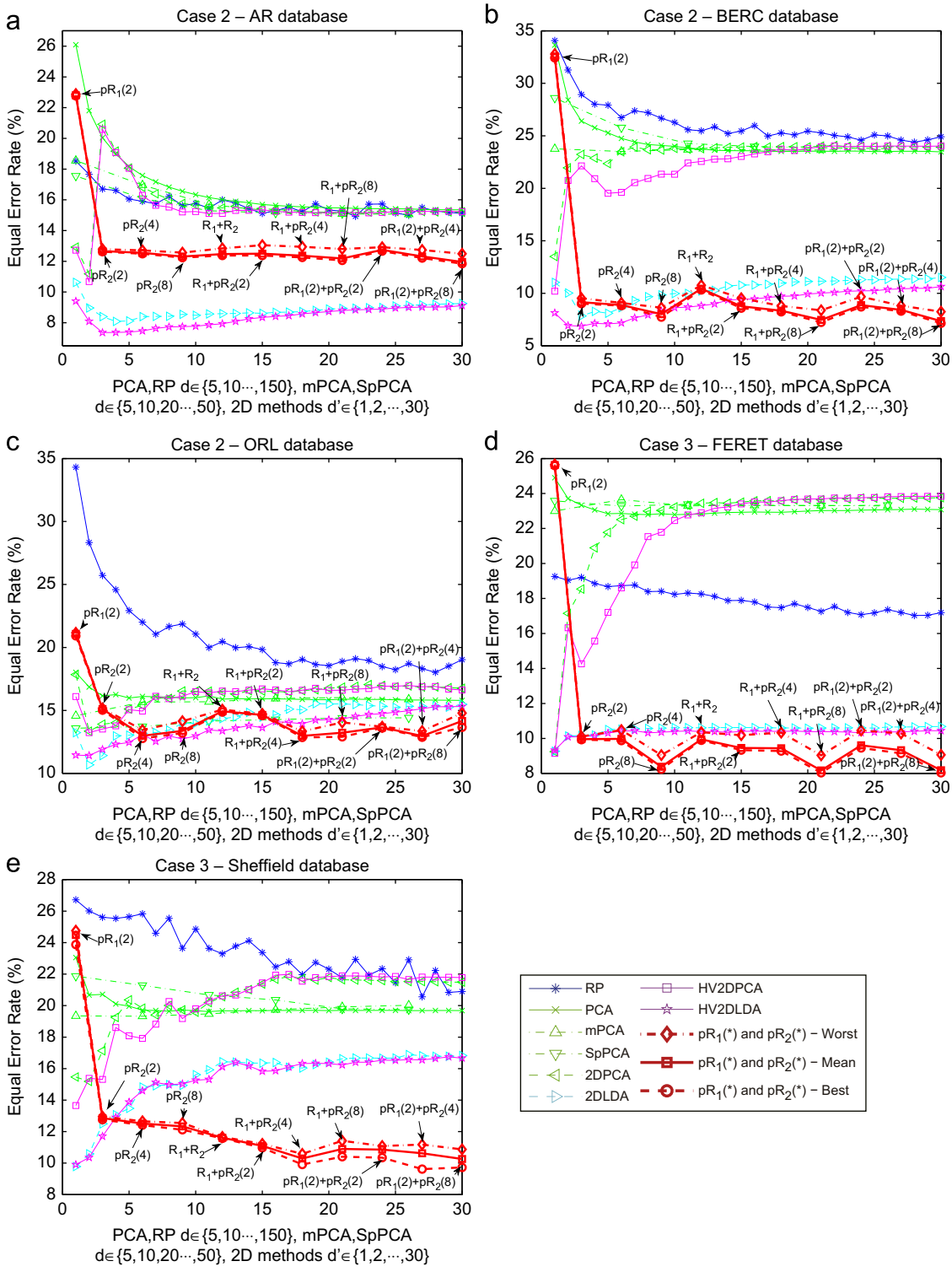

Fig. 6. Performance benchmarking plotted over different sizes of reduced dimension at $d \in \{5, 10, 15, \ldots, 150\}$ for RP and PCA; $d \in \{5, 10, 20, 30, 40, 50\}$ for mPCA and SpPCA; different number of principle component vectors $d' \in \{1, 2, \ldots, 30\}$ for 2DPCA, 2DLDA, HV2DPCA and HV2DLDA. (a) Case 2 (AR); (b) Case 2 (BERC); (c) Case 2 (ORL); (d) Case 3 (FERET); (e) Case 3 (Sheffield).

RP shows comparable performance with that of PCA and 2DPCA at $d > 15$. Similar to the previous case, $p\mathbf{R}_1(*)$ and $p\mathbf{R}_2(*)$ show a small performance deviation over different r values.

By comparing the results of Case 2 and Case 3, we observe that the proposed method seems to better tolerate horizontal pose variation than 2DPCA and 2DLDA methods (see Fig. 6(b), (d) and (e)). For example, as shown in Fig. 6(d) and (e), the proposed method is observed to be able to tolerate up to 20° of pose variation for Sheffield and FERET databases. Also, Fig. 6(b) shows that the proposed method can tolerate variations of minor pose (within $10 \sim 15°$ for BERC database) and mild illumination changes. However, from Fig. 6(a), we observe that the proposed method is sensitive to large expression and illumination changes.

(iv) $\mathbf{R}_1$ and $\mathbf{R}_2$ projections for diagonal features extraction:

In order to study the effects of diagonal and off-diagonal image features, the proposed $\mathbf{R}_1$ and $\mathbf{R}_2$ projections are applied to diagonally transformed face images following [32]. Let $\mathbf{X}_r$ and $\mathbf{X}_l$ be respectively right-skewed image (see Fig. 1(a) of [32]) and left-skewed image (see Fig. 1(b) of [32]). Then, we pre-multiplied $\mathbf{R}_1$ matrix by $\mathbf{X}_l$ to extract diagonal features (called Diagonal $\mathbf{R}_1$ in brief). In a similar manner, we post-multiplied $\mathbf{R}_2$ matrix by $\mathbf{X}_r$ which corresponds to extraction of off-diagonal features (called Diagonal $\mathbf{R}_2$ in brief). We shall use BERC and ORL databases for this study.

Fig. 7 shows the performance comparison among $\mathbf{R}_1$, $\mathbf{R}_2$, Diagonal $\mathbf{R}_1$ and Diagonal $\mathbf{R}_2$ in terms of test EER. As seen from the figure, both diagonal and off-diagonal features contain less discriminative information than that of horizontal features ($\mathbf{R}_2$). This shows that face images used in our experiments contain most discriminative information in horizontal direction among the four directions. The results also lead us to another observation that the performance of both diagonal features varies between that of $\mathbf{R}_1$ and $\mathbf{R}_2$ under various data conditions.

(v) Statistical significance test among the compared methods:

To observe whether the performance differences among the compared methods (see Fig. 6) are statistically significant, we performed a Friedman test [33]. The test compares the average rank of the compared methods. If the p-value (we calculated it using Matlab function '*friedman*()') is smaller than the critical value α, then the null hypothesis which states that all the compared methods are equivalent is rejected.

In our test, the null hypothesis is rejected at confidence level $\alpha=0.05$, since $p=7.8095\text{e}-005$. Next, we conduct a post-hoc Nemenyi test in order to analyze the differences among the compared methods. The critical distance (CD) for the Nemenyi test at $p=0.05$ is calculated as 5.3728. Following [33], the result of Nemenyi test is visualized in Fig. 8, where the compared algorithms are aligned according to their average ranks in descending order over the horizontal axis. As shown in the figure, statistically similar methods are grouped with a horizontal bar. The proposed method, 2DLDA and HV2DLDA are seen to be significantly different from RP and SpPCA in terms of verification performance. The proposed method does not show statistical significance when compared with mPCA, PCA, 2DPCA and HV2DPCA. These results show that the proposed method outperforms RP and SpPCA methods with statistical significance while showing insignificant difference in performance with that of the remaining six compared methods.

Table 4
Polynomial orders r for $\text{p}\mathbf{R}_1(*)+\text{p}\mathbf{R}_2(*)$, which have shown the best (the worst) test EER performances from $r\in\{1,2,\ldots,10\}$.

Polynomial order, r	Shared-token scenario (stolen token)				
	Case 2			Case 3	
	AR	BERC	ORL	FERET	Sheffield
$\text{p}\mathbf{R}_1(2)$	6 (4)	4 (1)	1 (4)	1 (3)	1 (10)
$\text{p}\mathbf{R}_2(2)$	8 (3)	2 (1)	1 (3)	1 (2)	1 (2)
$\text{p}\mathbf{R}_2(4)$	5 (1)	6 (1)	2 (1)	4 (1)	5 (1)
$\text{p}\mathbf{R}_2(8)$	6 (1)	2 (1)	4 (1)	5 (1)	5 (1)
$\mathbf{R}_1+\mathbf{R}_2$	9 (1)	8 (1)	6 (2)	2 (1)	6 (1)
$\mathbf{R}_1+\text{p}\mathbf{R}_2(2)$	7 (1)	6 (1)	5 (1)	2 (1)	1 (9)
$\mathbf{R}_1+\text{p}\mathbf{R}_2(4)$	10 (1)	6 (1)	2 (1)	4 (1)	2 (10)
$\mathbf{R}_1+\text{p}\mathbf{R}_2(8)$	10 (1)	2 (1)	4 (1)	5 (1)	3 (10)
$\text{p}\mathbf{R}_1(2)+\text{p}\mathbf{R}_2(2)$	6 (1)	6 (1)	5 (3)	4 (1)	1 (8)
$\text{p}\mathbf{R}_1(2)+\text{p}\mathbf{R}_2(4)$	5 (1)	6 (1)	4 (1)	4 (1)	1 (10)
$\text{p}\mathbf{R}_1(2)+\text{p}\mathbf{R}_2(8)$	4 (1)	2 (1)	2 (1)	5 (1)	1 (10)

4.3.2. *User-specific scenario*

Under this scenario, we test the proposed method in terms of user-specific tokens where each identity uses his own tokens for authentication. Because of its good verification performance among many variations of the proposed method, only $\text{p}\mathbf{R}_1(2)+\text{p}\mathbf{R}_2(8)$ will be considered for this experiment. Fig. 9 shows the average test EER values of $\text{p}\mathbf{R}_1(2)+\text{p}\mathbf{R}_2(8)$. The EER performance of the shared-token is included in the figure for comparison.

Both Fig. 9(a) on BERC database and (b) on AR database show significant performance enhancement based on user-specific tokens with respect to the shared-token case over the entire range of r. Such performance enhancement indicates an advantage of the proposed method over compared methods such as

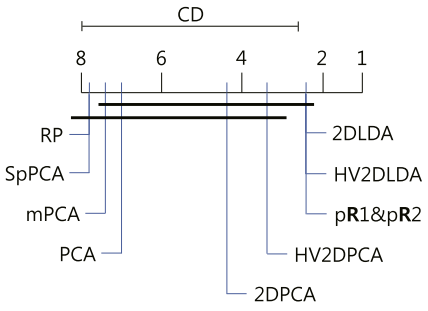

Fig. 8. A visualization of the post-hoc Nemenyi test at $p=0.05$.

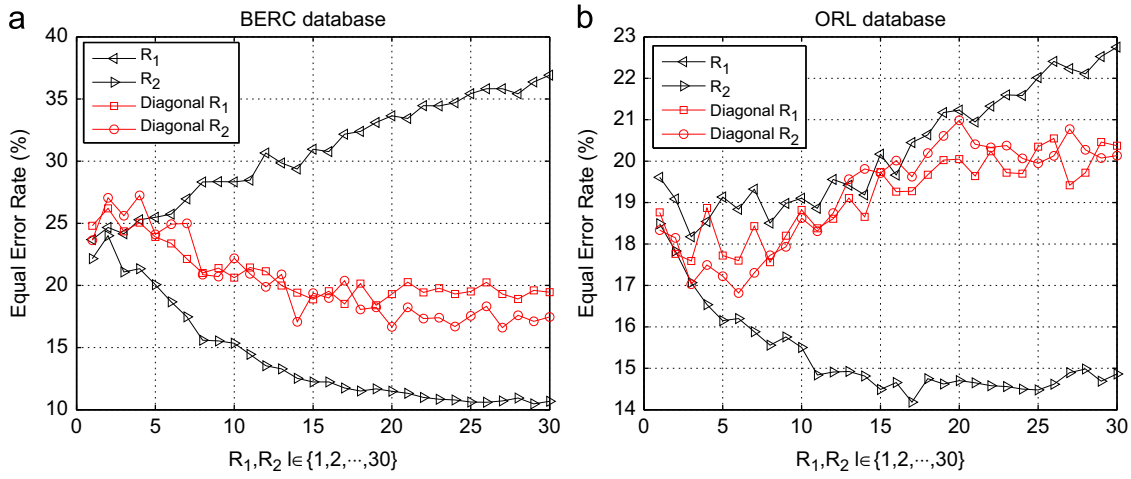

Fig. 7. Performance comparison among $\mathbf{R}_1$, $\mathbf{R}_2$, Diagonal $\mathbf{R}_1$ and Diagonal $\mathbf{R}_2$ plotted over different group sizes $l\in\{1,2,\ldots,30\}$. (a) BERC; (b) ORL.

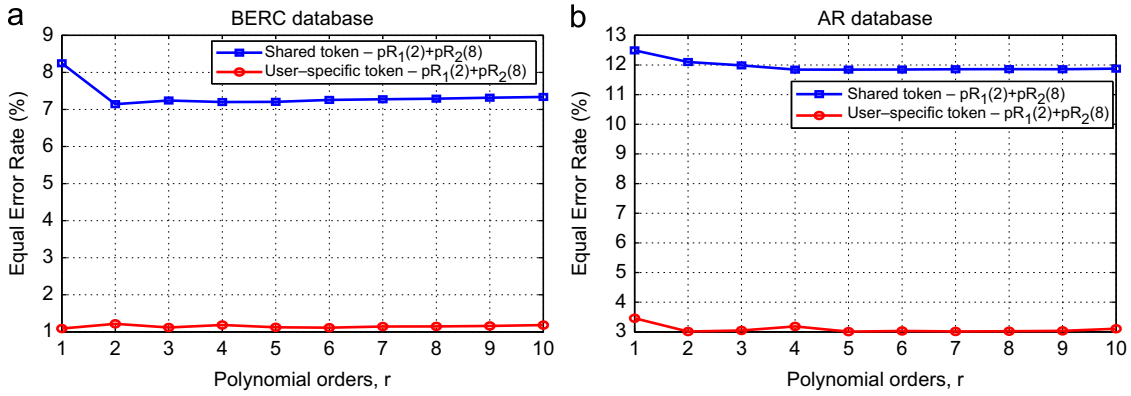

Fig. 9. Performance benchmarking of user-specific scenario versus stolen-token scenario plotted over different polynomial orders $r \in \{1,2,\ldots,9,10\}$. (a) BERC; (b) AR.

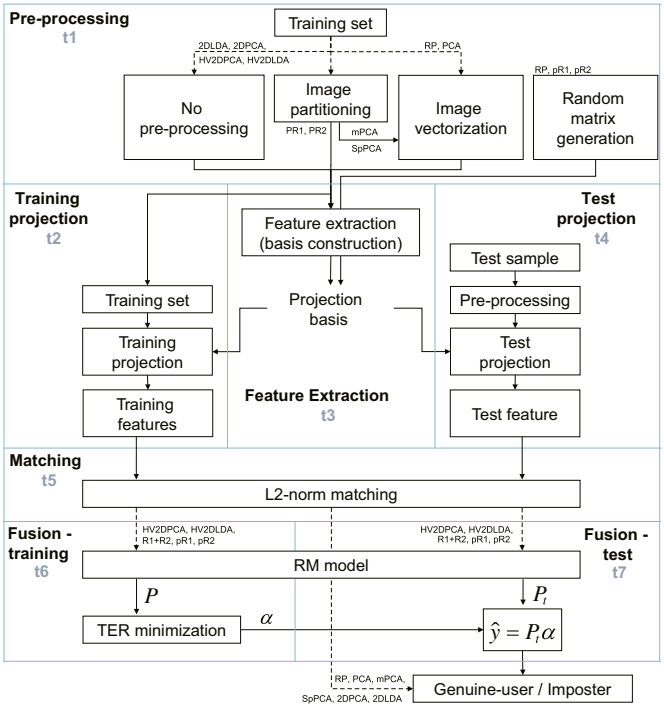

Fig. 10. Flow diagram for CPU time measurements.

Table 5
Components for measurement of training and test times for each method.

Fusion category	Method	Training time	Test time
Non-fusion based	RP	t1	t4+t5
	PCA	t1+t3	
	mPCA		
	SpPCA		
	2DPCA	t3	
	2DLDA		
Fusion based	HV2DPCA	t2+t3+t5+t6	t4+t5+t7
	HV2DLDA		
	$p\mathbf{R}_1(*)+p\mathbf{R}_2(*)$	t1+t2+t5+t6	

2DPCA, 2DLDA, HV2DPCA and HV2DLDA which cannot adopt a user-specific projection basis.

4.3.3. CPU time performance benchmarking

In this subsection, the feature extraction methods are compared in terms of their CPU times in seconds measured on a PC of 2.66 GHz with 4G RAM under Matlab platform [34]. To show the differences among the compared methods, the time was measured based on FERET database which is the largest database used in our experiments.

Fig. 10 depicts the entire components of compared methods step-by-step from pre-processing to yielding an output decision. In the figure, a solid line indicates a connection flow from a step to a next step and a dashed line indicates an optional connection for a certain algorithm. As the figure shows, different methods require different processes or steps for their training and test phases. In order to facilitate a commensurate computing cost comparison, the CPU times for each method are measured as tabulated in Table 5. We assumed here that pre-processing on database such as image normalization, conversion of color image into grey level and histogram equalization has been done beforehand.

In order to achieve a fair comparison among the methods balancing between the EER performance and computing cost, the CPU time for mPCA and SpPCA are measured based on the number of partitions N which has shown the best training EER performance as summarized in Table 3. In a similar manner, the computing cost for each HV2DPCA, HV2DLDA, $p\mathbf{R}_1(*)$ and $p\mathbf{R}_2(*)$ are measured based on the best polynomial order r (see Table 4). The test time is the CPU time consumed to test 100 samples.

Fig. 11(a) and (b) respectively shows an overview and a zoom-in view of training time, and Fig. 11 (c) shows the test times. As shown in Fig. 11(a) and (b), PCA takes the largest time to train, while RP at $d<60$ and 2DPCA at $d'>12$ take the least time. Training the PCA model is even more expensive than training the fusion methods except for $p\mathbf{R}_1(2)+p\mathbf{R}_2(8)$ and HV2DPCA and HV2DLDA at $d'>28$. Apart from PCA, fusion based methods are heavier than those non-fusion based methods in terms of training cost. mPCA and SpPCA show similar training time performance over the entire range of d.

In Fig. 11 (c), 2DPCA, 2DLDA, HV2DPCA and HV2DLDA show an increasing trend of test CPU time performance as d' increases, since adopting a high d' yields a high dimensional set of features [6,17]. The proposed method consumed a long test time when we fuse a large number of features. For example, $\mathbf{R}_1+p\mathbf{R}_2(8)$ and $p\mathbf{R}_1(2)+p\mathbf{R}_2(8)$ are about 5 times slower than $p\mathbf{R}_1(2)$ and $p\mathbf{R}_2(2)$ in testing. Because of high dimensional features of $\mathbf{R}_1$ ($d=460$) and $\mathbf{R}_2$ ($d=560$), their fusion ($\mathbf{R}_1+\mathbf{R}_2$) took about 0.5 second more than that of $p\mathbf{R}_1(8)$ ($d=70$ each) match score features. Among the compared methods, both RP and PCA are among the fastest methods during the test phase because of their low dimensional features.

4.3.4. Summary of results and observations

In the experiment which compares $\mathbf{R}_1$, $\mathbf{R}_2$, $\mathbf{R}_1+\mathbf{R}_2$, $p\mathbf{R}_1(*)$ and $p\mathbf{R}_2(*)$ in terms of EER performance (see Fig. 5), the $\mathbf{R}_2$ projection

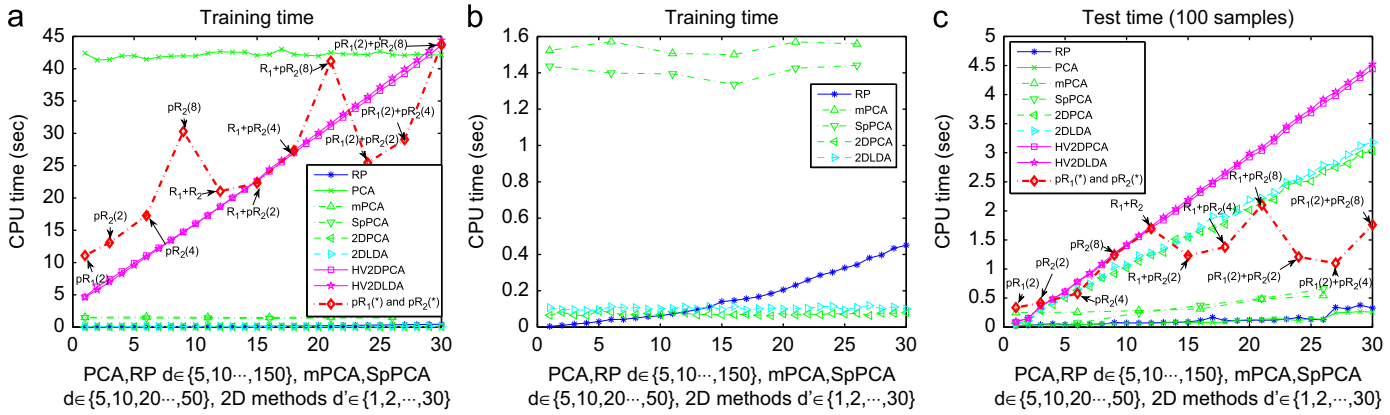

Fig. 11. Performance comparison of computing cost plotted over different sizes of reduced dimension at $d \in \{5, 10, 15, \ldots, 150\}$ for RP and PCA; at $d \in \{5, 10, 20, 30, 40, 50\}$ for mPCA and SpPCA; different number of principle component vectors at $d' \in \{1, 2, \ldots, 30\}$ for 2DPCA, 2DLDA, HV2DPCA and HV2DLDA. (a) Training time (an overview); (b) training time (a zoom-in view); (c) test time.

which corresponds to extraction of horizontal face image features outperforms the $\mathbf{R}_1$ projection which extracts the vertical face image 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.

Even though the $\mathbf{R}_1$ projection is poor in terms of verification performance itself, it is still beneficial to include $\mathbf{R}_1$ projection in fusion. For example, $p\mathbf{R}_1(2)+p\mathbf{R}_2(8)$ has shown a lower EER values than that of $\mathbf{R}_1+p\mathbf{R}_2(8)$, and $\mathbf{R}_1+p\mathbf{R}_2(8)$ showed a better performance than that of $p\mathbf{R}_2(8)$.

Under Case 2 (using BERC and ORL database) and Case 3 with shared-token scenario where the utilized images are contaminated by either minor or large pose variation, $p\mathbf{R}_1(*)$ and $p\mathbf{R}_2(*)$ yielded comparable EER values with that of 2DPCA and 2DLDA. The proposed method even overwhelms the other compared methods such as RP, PCA, mPCA and SpPCA in terms of EER performance. However, the proposed methods failed to show such good verification performance under Case 2 using AR database where the used images do not contain pose variation. These results show that the proposed method is less sensitive to pose variation (particularly horizontal pose variation) than that of the compared methods, such as 2DPCA and 2DLDA. Moreover, the proposed method is seen to outperform RP and SpPCA with statistical significance while showing no significant difference in performance with that of 2DLDA, HV2DLDA, 2DPCA, HV2DPCA, mPCA and PCA.

Under the shared-token scenario, the proposed method have shown a much better EER performance than that of RP. In addition, the proposed method that incorporated a user-specific token produced much lower EER values than that of the case where every identities utilized shared tokens. Such performance enhancement and the applicability to cancelable biometrics is a main advantage of the proposed method over the compared methods such as RP, 2DPCA and 2DLDA.

Fusion of horizontal and vertical features at score level using TERRM ($\mathbf{R}_1+\mathbf{R}_2$) overcame a performance limitation imposed by single $\mathbf{R}_1$ and $\mathbf{R}_2$. In addition, fusion of more partitioned directional features from $p\mathbf{R}_1(*)$ and $p\mathbf{R}_2(*)$ yielded a further enhancement of EER performance. However, the number of partitions appears to saturate at 8 for AR, BERC, FERET and Sheffield data sets, and at 4 for ORL data set. Moreover, the large number of features in score level fusion incurs a high computational cost (see Fig. 11). This leads us to the proposition that fusing image features from different facial parts (non-overlapping regions) at score level using TERRM with an appropriate parameter setting can significantly enhance the overall verification performance.

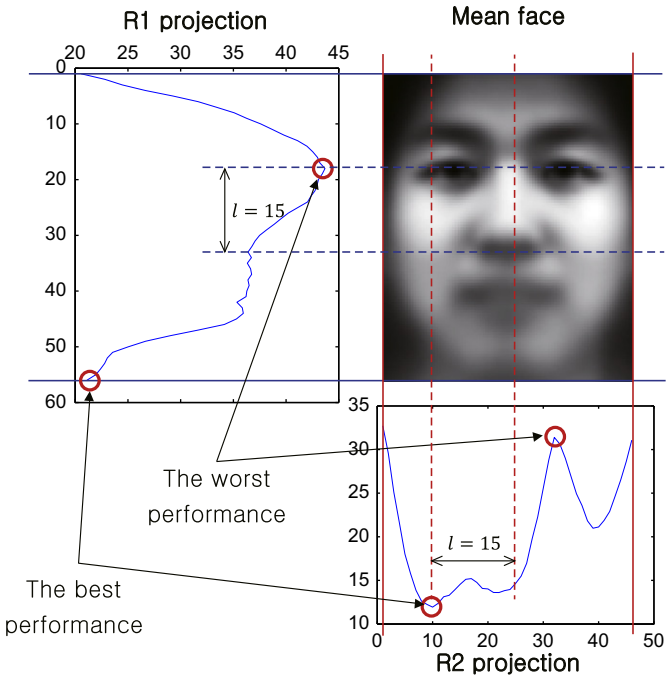

Fig. 12. A sample illustration of feature distribution in terms of EER at $k=1$ and $l=15$ on BERC database.

5. Analysis

From the experiments, $\mathbf{R}_2$ projection is seen to significantly outperform $\mathbf{R}_1$ projection in terms of EER. In this section, we investigate into feature distributions attempting to understand the mechanism of discriminative feature extraction by $\mathbf{R}_1$ and $\mathbf{R}_2$ projections. Subsequently, the proposed method will be analyzed regarding whether it can fulfill the four requirements (properties) of cancelable biometrics [8,35].

5.1. Analysis of feature distributions

In order to observe which part of face image contains discriminative information, the feature distribution is analyzed. To do this, we generate $\mathbf{R}_1$ and $\mathbf{R}_2$ projection matrix at $k=1$ with varying group sizes $l \in \{1, 2, \ldots, 30\}$. For $\mathbf{R}_1$ projection, we generate

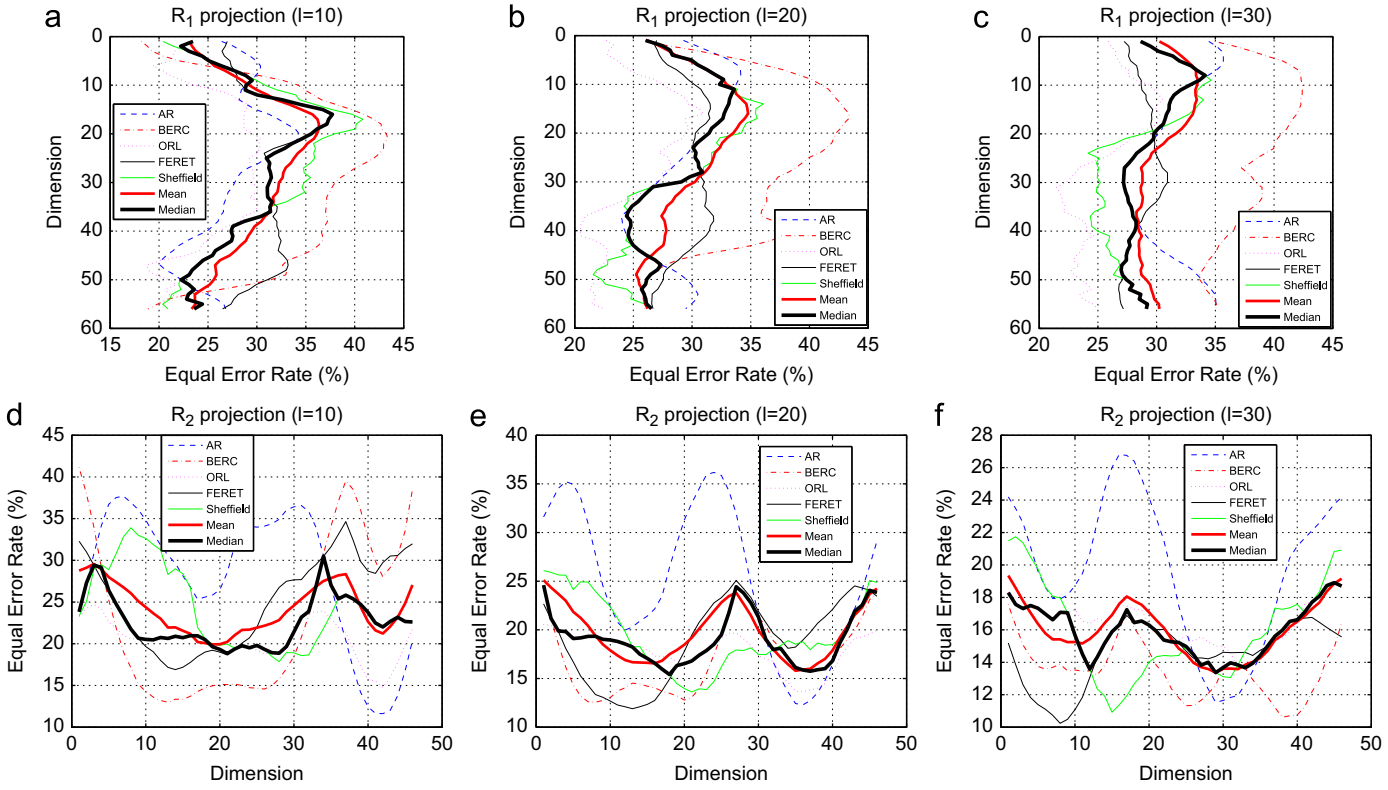

Fig. 13. Average feature distributions of $\mathbf{R}_1$ and $\mathbf{R}_2$ projections (in terms of EER) at $l \in \{10, 20, 30\}$ based on AR, BERC, ORL, FERET and Sheffield databases. (a) $\mathbf{R}_1$ projection ($l=10$); (b) $\mathbf{R}_1$ projection ($l=20$); (c) $\mathbf{R}_1$ projection ($l=30$); (d) $\mathbf{R}_2$ projection ($l=10$); (e) $\mathbf{R}_2$ projection ($l=20$); (f) $\mathbf{R}_2$ projection ($l=30$). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

56 different $\mathbf{R}_1$ matrices by shifting the 1's group pixel-by-pixel horizontally from the left most element. In a similar manner, we generate 46 different $\mathbf{R}_2$ matrices by shifting the 1's group vertically downwards from the top first element. Each of these matrices is then used for templates generation where each performance is recorded according to its starting point.

Fig. 12 shows the EER (at $l=15$) plotted over the starting point of the 1's group. The figure indicates those locations where the best and worst discriminative information are extracted by the projection matrix. The EER values resulted from $\mathbf{R}_1$ projection is drawn vertically[2] along the mean face of BERC database. In a similar manner, a plot of EER values resulted from $\mathbf{R}_2$ projection is drawn horizontally along the mean face. In the figure, a facial region which corresponds to a low EER contains more discriminative information than that with high EER values.

In Fig. 12, $\mathbf{R}_1$ projection shows the best performance when the features are extracted from the forehead including hair and jaw regions of a face image. On the other hand, the worst performance is observed when the features are extracted from facial components such as eyes, nose and lips. These results show that the forehead and jaw regions contain more discriminative information than that of other regions including regions occupied by the facial components for $\mathbf{R}_1$ projection. However, these forehead and jaw regions can vary due to variation of hair-style and jaw movement. In contrast to $\mathbf{R}_1$ projection, $\mathbf{R}_2$ projection extracts the most discriminative features from the facial components, and shows the worst performance when the features are extracted from the regions of face contour.

Fig. 13 shows the EER distribution plots for all the five databases used in our experiments (AR, BERC, ORL, FERET and Sheffield) at $l=\{10, 20, 30\}$ and $k=10$. The mean value of all the five databases is represented as a thick red line in the plot, and the median is shown as the thickest black line in the plot. As shown in the figure, all the five databases show a similar trend of results compared to that of Fig. 12. These results show that $\mathbf{R}_2$ projection extracts more relevant facial features than that of $\mathbf{R}_1$ projection.

5.2. Cancelable properties

In this subsection, the proposed $\mathbf{R}_1$ and $\mathbf{R}_2$ projections are investigated to observe whether they fulfill the requirements of cancelable biometrics [8,35]:

1. One-way transformation (non-invertibility): Non-invertibility of template computation to prevent recovery of biometric data,
2. Diversity: No same cancelable template can be used in two different applications,
3. Re-usability: Straightforward revocation and reissue in the event of compromise,
4. Performance: The cancelable biometric template should not deteriorate the verification performance.

5.2.1. Non-invertibility and performance

The proposed projections consist of a linear system. The $\mathbf{R}_1$ projection is given by $\mathbf{Y}_V = (1/z)\sum_{i=1}^{z} \mathbf{R}_1^i \mathbf{X}$, where $\mathbf{R}_1^i$ is the i-th projection of size $k \times p$ (k is the projection size, p is the column size of input data $\mathbf{X} \in \mathbb{R}^{p \times q}$, and z is the number of transformations for averaging). Consider $z=1$ and assume that only $\mathbf{X}$ is unknown.

[2] Recall that $\mathbf{R}_1$ projection corresponds to extraction of vertical face image features, while $\mathbf{R}_2$ projection extracts horizontal face image features (see Section 3 for details).

According to the projection size k, the linear system behaves as follows:

(i) If $k<p$, the number of equations ($k\times q$ number of equations) is smaller than the number of unknowns ($p\times q$ number of variables in $\mathbf{X}$), then the linear system has infinitely many solutions (called an underdetermined system).
(ii) If $k=p$, the number of equations is equal to the number of unknowns, then the linear system is determined.
(iii) If $k>p$, the number of equations is larger than the number of unknowns, then the linear system is overdetermined.

In the event that both $\mathbf{Y}_V$ and $\mathbf{R}_1$ are stolen, only the first case ($k<p$) can satisfy the mathematical non-invertibility. In other words, only when $k<p$, an unique solution does not exist for an attacker to recover the original data $\mathbf{X}$ mathematically. Next, consider the case of $z>1$. Depending on how many of the $\mathbf{R}_1^i$, $i=1,2,\ldots,z$, the attacker has stolen, two different scenarios can occur as follows:

(i) When the entire set of $\mathbf{R}_1^i$ matrices are stolen, then the number of equations and unknowns equal to that of the case $z=1$;
(ii) When $\mathbf{R}_1^i$ matrices are partially stolen, and even if the attacker knows the exact z, then the number of unknowns are increased while the number of equations is unchanged. We note here that the unrevealed $\mathbf{R}_1^i$ matrices become unknowns to the attacker. For example, if he stole just $\mathbf{R}_1^{\{1,3\}}$ matrices out of $z=4$, then the number of unknowns becomes $(z-2)\times(p\times q)$. If the exact z is not given to him, then he could not know even how many unknowns he has to recover or estimate.

The $\mathbf{R}_2$ projection is given by $\mathbf{Y}_H=(1/z)\sum_{j=1}^{z}\mathbf{X}\mathbf{R}_2^j$, where $\mathbf{R}_2^j$ is the j-th projection of size $q\times k$. Similar to the linear system of $\mathbf{R}_1$ projection, the number of equations and unknowns for $\mathbf{R}_2$ system are respectively $p\times k$ and $p\times q$ where the mathematical non-invertibility occurs when $q>k$. Moreover, the number of unknowns can be increased according to z as observed in the $\mathbf{R}_1$ system. In our performance evaluation, the size of data $\mathbf{X}$ is 56×46 and we set $k=10$ according to [14].

Next we show the non-invertibility of the proposed method in terms of image reconstruction. In order to observe the effect of averaging transformations, different number of templates used for averaging are experimented. Fig. 14(a)–(c) shows respectively a sample image from BERC database, reconstructed images from each $\mathbf{R}_1$ and $\mathbf{R}_2$ features using different number of averaging features (at $k=10$ and $l=10$), and its corresponding EER values at different group sizes $l\in\{1,10,20,30\}$ which were plotted over different number of features used in averaging. The image in Fig. 14(b) is reconstructed based on $\hat{\mathbf{X}}=\mathbf{R}_1^{-1}\cdot\mathbf{Y}_V$ for $\mathbf{R}_1$ projection and $\hat{\mathbf{X}}=\mathbf{Y}_H\cdot\mathbf{R}_2^{-1}$ for $\mathbf{R}_2$ projection, where $\hat{\mathbf{X}}$ indicates a reconstructed image. The white number inserted at the bottom of each image indicates the number of features being averaged. For example, '1:5' indicates that five feature templates have been averaged.

From Fig. 14(b), we observe that averaging more than three templates is sufficient to hide the original biometric signal from both $\mathbf{R}_1$ and $\mathbf{R}_2$ templates. In Fig. 14(c), interestingly, the EER of $\mathbf{R}_1$ projection is increased when the number of features for averaging is increased. However, the EER values of $\mathbf{R}_2$ projection is slightly decreased when the number of features for averaging is increased. These results suggest that the discriminativity of features has been accumulated during feature averaging. From these results, we observe that both $\mathbf{R}_1$ and $\mathbf{R}_2$ cancelable templates fulfill the performance property as well.

In view of computational efficiency, a large number of feature averaging increases the complexity of the proposed method. In order to balance among concealment of the original signal (satisfying the non-invertibility), complexity of the proposed method and verification performance, we shall average three distinct transformations at feature level to generate the proposed cancelable templates.

5.2.2. Diversity and re-usability

In order to investigate the diversity and re-usability of the proposed method, we setup a scenario of token compromise. Assume that some face templates from a face recognition system have been stolen by an attacker, and the system re-issues new face templates (will be called re-issued new template $\mathbf{Y}^{new}$ hereafter) to handle the attack. The attacker attempts to get an authorization from the system by using the stolen template (will be called stolen old template $\mathbf{Y}^{old}$ hereafter).

For user authentication, the system utilizes $\mathbf{Y}^{new}$ and produces two match score distributions, namely imposter and genuine-user match scores which are respectively generated by inter- and intra-identity matching. The attacker's authorization trial can be described in terms of intra-matching between $\mathbf{Y}^{old}$ and $\mathbf{Y}^{new}$ in which a distribution of pseudo genuine-user scores is produced.

Here, we expect that the location of pseudo-genuine-user distribution is distinguishable from that of the genuine-user distribution if the diversity property is satisfied. In order to validate the point, all samples of BERC database are utilized for evaluation, and all the three distributions above are plotted in the same graph as shown in Fig. 15. The distributions of $\mathbf{R}_1$ and $\mathbf{R}_2$

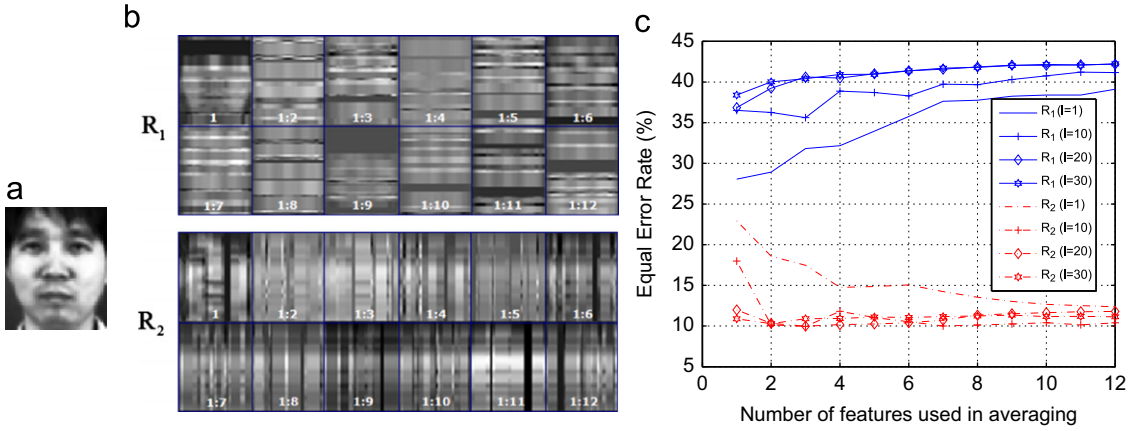

Fig. 14. Image reconstruction: (a) a sample image from BERC database; (b) reconstructed images from $\mathbf{R}_1$ and $\mathbf{R}_2$ features using different number of features for averaging; (c) the corresponding EER performances at different group sizes $l\in\{1,10,20,30\}$ plotted over different number of features in averaging.

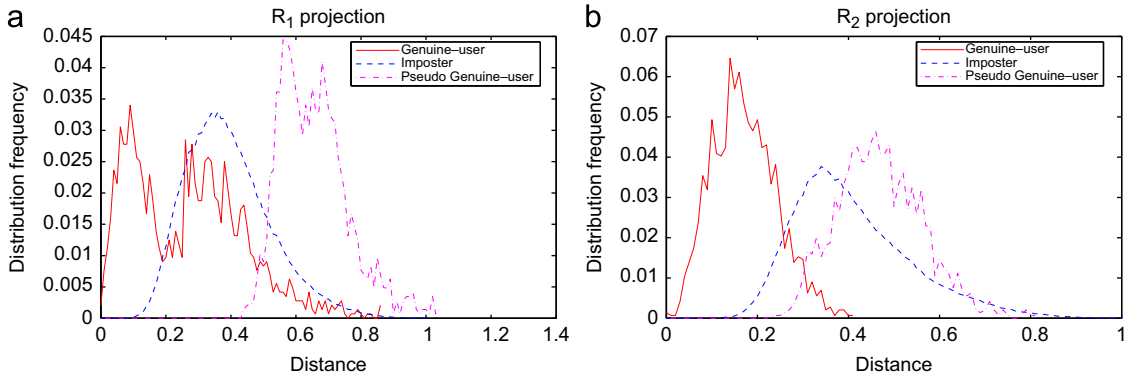

Fig. 15. Average distributions (over 30 transformations at feature level) of genuine-user scores, imposter scores and pseudo-genuine-user scores. (a) $\mathbf{R}_1$ projection; (b) $\mathbf{R}_2$ projection.

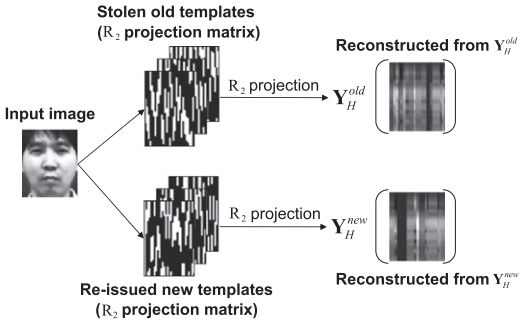

Fig. 16. An illustration for re-usability of the proposed method (e.g. $\mathbf{R}_2$ projection) in terms of image reconstruction.

projections are obtained at $l=10$ and $k=10$. For statistical evidence, 30 transformations based on 30 different projection bases (i.e. projection matrix) have been performed where the average is recorded.

In Fig. 15, the mean of pseudo genuine-user distribution is seen to be different from that of the genuine-user distribution. This indicates that both $\mathbf{R}_1$ and $\mathbf{R}_2$ projections based on a different projection basis transforms facial features into a different range of score distribution. From this result, we observe that $\mathbf{R}_1$ and $\mathbf{R}_2$ projections fulfill the diversity property.

Finally, we investigate the re-usability of the proposed method in terms of appearance difference between $\mathbf{Y}^{new}$ and $\mathbf{Y}^{old}$ templates. Fig. 16 shows the overall procedure of obtaining both templates on $\mathbf{R}_2$ projection[3] along with their reconstructed images. As illustrated in the figure, both reconstructed images look different from each other, which imply that the proposed method fulfills the re-usability property.

6. Conclusion

In this paper, we proposed to extract horizontal and vertical facial features directly from partitioned raw image matrix. In order to enhance security in identity verification, both partial features are respectively averaged at feature level over several transformations. The averaged feature templates have empirically been shown to fulfill the four criteria of cancelable biometrics. The proposed cancelable templates are finally fused at score level via a total error rate minimization which adopts a reduced multivariate model. Our empirical results on the AR, FERET, ORL, Sheffield and BERC databases showed that the proposed projections yield comparable or slightly better verification performances than that of existing methods such as 2DPCA and 2DLDA, particularly under large horizontal pose variation. The results also indicate that the proposed cancelable templates are more discriminative in terms of identity verification than that of random projection method.

In our future work, a systematic study of various directional features including different resolutions, angles and block shapes would be beneficial in terms of exploiting various possible features using a limited number of image data.

[3] Note: since results of both $\mathbf{R}_1$ and $\mathbf{R}_2$ projections are similar, only $\mathbf{R}_2$ projection is adopted here as an example.

Acknowledgments

This work was supported by the National Research Foundation of Korea Grant funded by the Korean Government (NRF-2011-013-D00099).

References

[1] 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.
[2] M. Turk, A. Pentland, Eigenfaces for recognition, Journal of Cognitive Neuroscience 3 (1) (1991) 71–86.
[3] W. Zhao, R. Chellappa, P. Phillips, A. Rosenfeld, Face recognition: a literature survey, ACM Computing Surveys (CSUR) 35 (4) (2003) 399–458.
[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] S. Chen, Y. Zhu, Subpattern-based principle component analysis, Pattern Recognition 37 (5) (2004) 1081–1083.
[6] 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.
[7] J. Ye, R. Janardan, Q. Li, Two-dimensional linear discriminant analysis, Advances in Neural Information Processing Systems 17 (2004) 1569–1576.
[8] A.B.J. Teoh, A. Goh, D.C.L. Ngo, Random multispace quantization as an analytic mechanism for biohashing of biometric and random identity inputs, IEEE Transactions on Pattern Analysis and Machine Intelligence 28 (12) (2006) 1892–1901.
[9] N. Ratha, J. Connell, R. Bolle, Enhancing security and privacy in biometrics-based authentication systems, IBM Systems Journal 40 (3) (2001) 614–634.
[10] N. Goel, G. Bebis, A. Nefian, Face recognition experiments with random projections, in: Proceedings of SPIE Conference on Biometric Technology for Human Identification, 2005, pp. 426–437.
[11] B.J. Teoh, C.T. Yuang, Cancelable biometrics realization with multispace random projections, IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics 37 (5) (2007) 1096–1106.
[12] Y. Kim, A. Teoh, K.-A. Toh, A performance driven methodology for cancelable face templates generation, Pattern Recognition 43 (7) (2010) 2544–2559.
[13] B.-S. Oh, B.-G. Choi, K.-A. Toh, Fusing horizontal and vertical components of face images for identity verification, in: Proceedings of IEEE Conference on Industrial Electronics and Applications, China, 2009, pp. 651–655.
[14] B.-S. Oh, K.-A. Toh, A.B.J. Teoh, J. Kim, Combining local face image features for identity verification, Neurocomputing 74 (16) (2011) 2452–2463.

[15] K.-A. Toh, H.-L. Eng, Between classification-error approximation and weighted least-squares learning, IEEE Transactions on Pattern Analysis and Machine Intelligence 30 (4) (2008) 658–669.
[16] H. Kong, L. Wang, E.K. Teoh, X. Li, J.-G. Wang, R. Venkateswarlu, Generalized 2D principal component analysis for face image representation and recognition, Neural Networks 18 (5–6) (2005) 585–594.
[17] D. Zhang, Z.-H. Zhou, $(2D)^2$PCA: two-directional two-dimensional PCA for efficient face representation and recognition, Neurocomputing 69 (1–3) (2005) 224–231.
[18] 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.
[19] K.-A. Toh, Q.-L. Tran, D. Srinivasan, Benchmarking a reduced multivariate polynomial pattern classifier, IEEE Transactions on Pattern Analysis and Machine Intelligence 26 (6) (2004) 740–755.
[20] K.-A. Toh, J. Kim, S. Lee, Biometric scores fusion based on total error rate minimization, Pattern Recognition 41 (3) (2008) 1066–1082.
[21] D. González-Jiménez, M. Bicego, J.W.H. Tangelder, B.A.M. Schouten, O. Ambekar, J.L. Alba-Castro, E. Grosso, M. Tistarelli, Distance measures for Gabor jets-based face authentication: a comparative evaluation, Advances in Biometrics (2007) 474–483.
[22] L. Wang, Y. Zhang, J. Feng, On the Euclidean distance of images, IEEE Transactions on Pattern Analysis and Machine Intelligence 27 (8) (2005) 1334–1339.
[23] M. D'Agostino, C. Dardanoni, Whats so special about Euclidean distance? Social Choice and Welfare 33 (2) (2009) 211–233.
[24] D. Achlioptas, Database-friendly random projections, in: ACM Symposium on the Principles of Database Systems, ACE Press, 2001, pp. 274–281.
[25] A.M. Martńez, R. Benavente, The AR Face Database, Technical Report, CVC Technical Report, 1998.
[26] 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, 2006, pp. 804–808.
[27] AT&T Laboratories Cambridge, Olivetti Research Laboratory (ORL) database, 〈http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html〉 [February 16, 2012].
[28] 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.
[29] 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.
[30] National Institute of Standards and Technology (NIST), The Facial Recognition Technology (FERET) database, URL 〈http://www.itl.nist.gov/iad/humanid/feret/feret_master.html〉 [February 16, 2012].
[31] A.K. Jain, K. Nandakumar, A. Ross, Score normalization in multimodal biometric systems, Pattern Recognition 38 (12) (2005) 2270–2285.
[32] D. Zhang, Z.-H. Zhou, S. Chen, Diagonal principal component analysis for face recognition, Pattern Recognition 39 (1) (2006) 140–142.
[33] J. Demšar, Statistical comparisons of classifiers over multiple data sets, The Journal of Machine Learning Research 7 (2006) 1–30.
[34] The MathWorks, MATLAB, URL 〈http://www.mathworks.com/〉.
[35] D. Maltoni, D. Maio, A.K. Jain, S. Prabhakar, Handbook of Fingerprint Recognition, Springer-Verlag, New York, 2003.

Beom-Seok Oh received the B.S. degree in Computer Science from KonKuk University, Korea, in 2008 and the M.S. degree in Biometrics 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 is currently an associate editor of Pattern Recognition Letters and a senior member of the IEEE.

Kwontaeg Choi received the B.S. degree in Computer Science from Hallym University, Choon-cheon, Korea in 2001, the M.S. degree in Computer Science from Yonsei University, Seoul, Korea in 2006. He received the PhD degree from the Department of Computer Science, Yonsei University, Seoul, Korea in 2011. His research interests include computer vision, pattern recognition, and face recognition.

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 the School of Electrical and Electronic Engineering, Yonsei University. His research interest is in biometrics security and pattern recognition. He had published around 180 international journal and conference papers in his area.

Jaihie Kim received the Ph.D. 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 Biometric 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.