Pattern Recognition 43 (2010) 2882–2890

Contents lists available at ScienceDirect

Pattern Recognition

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

Fusion of color, local spatial and global frequency information for face recognition

Zhiming Liu, Chengjun Liu *

Department of Computer Science, New Jersey Institute of Technology, Newark, NJ 07102, USA

ARTICLE INFO

Article history:
Received 4 March 2009
Received in revised form
20 December 2009
Accepted 6 March 2010

Keywords:
Face Recognition Grand Challenge (FRGC)
Gabor image representation
Local binary patterns (LBP)
The RC_rQ hybrid color space
The YC_bC_r color space
The *YIQ* color space
The *RGB* color space

ABSTRACT

This paper presents a novel face recognition method by means of fusing color, local spatial and global frequency information. Specifically, the proposed method fuses the multiple features derived from a hybrid color space, the Gabor image representation, the local binary patterns (LBP), and the discrete cosine transform (DCT) of the input image. The novelty of this paper is threefold. First, a hybrid color space, the RC_rQ color space, is constructed by combining the *R* component image of the *RGB* color space and the chromatic component images, C_r and *Q*, of the YC_bC_r and *YIQ* color spaces, respectively. The RC_rQ hybrid color space, whose component images possess complementary characteristics, enhances the discriminating power for face recognition. Second, three effective image encoding methods are proposed for the component images in the RC_rQ hybrid color space to extract features: (i) a patch-based Gabor image representation for the *R* component image, (ii) a multi-resolution LBP feature fusion scheme for the C_r component image, and (iii) a component-based DCT multiple face encoding for the *Q* component image. Finally, at the decision level, the similarity matrices generated using the three component images in the RC_rQ hybrid color space are fused using a weighted sum rule. Experiments on the Face Recognition Grand Challenge (FRGC) version 2 Experiment 4 show that the proposed method improves face recognition performance significantly. In particular, the proposed method achieves the face verification rate (ROC III curve) of 92.43%, at the false accept rate of 0.1%, compared to the FRGC baseline performance of 11.86% face verification rate at the same false accept rate.

1. Introduction

Face recognition has become a very active research area in machine learning, pattern recognition and computer vision due to its broad applications in security and human computer interactions [28,27,4,5,20,3,17]. The recent Face Recognition Grand Challenge (FRGC) program reveals that uncontrolled illumination conditions pose grand challenges to face recognition performance [29]. As the traditional face recognition methods are inadequate in tackling these grand challenge problems, robust face recognition methods should be developed by extracting and exploiting multiple facial features efficiently. Many current face recognition methods, however, start with one single feature extraction process, such as extracting the principal components by the Eigenfaces method [33] and extracting the linear discriminating components by the Fisherfaces method [2], and then utilize one classifier, such as the nearest neighbor classifier, for face recognition. To take advantage of the rich facial features in human face images, a better face recognition methodology should consider multiple feature extraction processes, and fuse the multiple face features for improving face recognition performance. Note that fusion of multiple facial features may apply either various image features or various subspace transformed features [24,25].

The motivation of our method is to explore the complementary facial information in a novel hybrid color space in order to boost face recognition performance by fusing multiple features. The novelty of our method is three-fold. First, a novel hybrid color space, the RC_rQ color space, is constructed out of three different color spaces: the *RGB*, YC_bC_r, and *YIQ* color spaces. The RC_rQ hybrid color space, whose component images possess complementary characteristics, enhances the discriminating power for face recognition. Second, three effective image encoding methods are proposed for the component images in the RC_rQ hybrid color space: (i) a patch-based Gabor image representation for the *R* component image, (ii) a multi-resolution LBP feature fusion scheme for the C_r component image, and (iii) a component-based DCT multiple face encoding for the *Q* component image. For each method, the enhanced Fisher model (EFM) [21] is applied to extract features for classification. Finally, at the decision level, the similarity matrices generated using the three component images in the RC_rQ hybrid color space are fused using a weighted sum rule.

The Face Recognition Grand Challenge (FRGC) version 2 database [29] is applied to evaluate the effectiveness of the

* Corresponding author.
E-mail addresses: zl9@njit.edu (Z. Liu), chengjun.liu@njit.edu (C. Liu).

0031-3203/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2010.03.003

proposed method. The FRGC baseline algorithm and the biometric experimentation environment (BEE) system reveal that the FRGC version 2 Experiment 4, which is designed for indoor controlled single still image versus uncontrolled single still image, is the most challenging FRGC experiment [29]. We therefore choose the FRGC version 2 Experiment 4, which contains 12,776 training images, 16,028 controlled target images, and 8014 uncontrolled query images, to assess our proposed method. Face recognition performance is reported using the receiver operating characteristic (ROC) curves, which plot the face verification rate (FVR) versus the false accept rate (FAR). The BEE system generates three ROC curves (ROC I, ROC II, and ROC III) corresponding to the images collected within semesters, within a year, and between semesters, respectively [29]. Experimental results show that the proposed method improves face recognition performance significantly. In particular, the proposed method, which achieves the face verification rate of 92.43% at the false accept rate of 0.1%, compares favorably against the published results on the FRGC version 2 Experiment 4.

2. Background

Generally speaking, facial feature extraction methods fall into two major categories: local and holistic. Some representative local methods, such as the Gabor image representation [6,8,22] and the local binary patterns or LBP [26,1], have been shown effective for face recognition. The Gabor image representation, which captures well the salient visual features corresponding to spatial localization, orientation selectivity, and spatial frequency, displays robust characteristics in dealing with image variabilities. The LBP, which are originally introduced in [26] for texture analysis, have been successfully extended to describe faces, due to the finding that faces can be seen as a composition of micropatterns that are well described by the LBP operators [1]. Some representative holistic approaches for feature extraction are the widely used principle component analysis (PCA) [33] and linear discriminant analysis (LDA) [2,10,32]. Recently, discrete cosine transform, or DCT, has been applied as a holistic appearance-based method for face recognition [13,9]. Because DCT is asymptotically equivalent to PCA, it compresses energy of an image into a few large-magnitude coefficients. As a result, these DCT coefficients form the desirable features for pattern recognition. DCT has been demonstrated to outperform PCA in terms of computational efficiency and recognition accuracy [13].

Color information has been widely applied in face detection, but not in face recognition. Recent research reveals that different color spaces transformed from the *RGB* color space display different discriminating power for pattern recognition [11,30,20,35]. In particular, a general discriminant color model derived from the *RGB* color space has been shown more effective for face recognition than the original *RGB* color space [35]. The *R* component image in the *RGB* color space and the *V* component image in the *HSV* color space, for example, have been demonstrated to possess more discriminating power than the component images in several other color spaces for a face recognition task [30].

Regarding the fusion of local and global features for improving face recognition performance, generally speaking, local features such as those extracted by the Gabor [22] and the LBP [1] methods are different from global features such as those derived by the Eigenfaces and the Fisherfaces methods, as different features deal with different scaled facial details. If the features are complementary to each other, the correlation between the outputs of the local and the global classifiers would be reduced. As a result, fusing the classification outputs at the decision level could lead to improved overall performance [19]. Currently, most methods extract the complementary features from the gray image only, which largely ignores the complementary characteristics provided by the different face representations. A better methodology should seek a natural yet more powerful complementary face representation by capitalizing on the multiple images in a novel color space. Toward that end, this paper presents a robust face recognition method by integrating the local and global features derived from a novel hybrid color space.

3. Multiple feature fusion method for face recognition

This section details our novel multiple feature fusion or proposed method for face recognition. The proposed method integrates color, local spatial and global frequency information for improving face recognition performance. Specifically, the color information is derived by constructing a novel hybrid color space, the RC_rQ color space, which combines the *R* component image of the *RGB* color space and the chromatic component images, C_r and *Q*, of the YC_bC_r and *YIQ* color spaces, respectively. As the hybrid color space provides the complementary and different representations for face image, the correlation among the individual component images would be reduced. As a result, the fusion of the outputs of the individual classifiers corresponding to the component images could lead to improve recognition performance [19]. The local spatial information comes from the Gabor wavelet representation of the *R* component image and the local binary patterns of the C_r component image, while the global frequency information is due to the discrete cosine transform of the *Q* component image.

3.1. A hybrid color space: RC_rQ

Color provides powerful information for object detection, indexing and retrieval, as "humans can discern thousands of color shades and intensities, compared to about only two dozen shades of gray" [12]. Color information is helpful for improving the performance of face recognition due to the complementary characteristics among the color component images. This paper addresses face recognition in a novel hybrid color space instead of the conventional color spaces. As the *R* component image in the *RGB* color space is more effective than other component images for face recognition [30], we define a new hybrid color space RC_rQ, where C_r and *Q* are from the YC_bC_r color space and the *YIQ* color space, respectively. The YC_bC_r and *YIQ* color spaces are defined as follows:

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} + \begin{bmatrix} 65.4810 & 128.5530 & 24.9660 \\ -37.7745 & -74.1592 & 111.9337 \\ 111.9581 & -93.7509 & -18.2072 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (1)$$

$$\begin{bmatrix} Y \\ I \\ Q \end{bmatrix} = \begin{bmatrix} 0.2990 & 0.5870 & 0.1140 \\ 0.5957 & -0.2745 & -0.3213 \\ 0.2115 & -0.5226 & 0.3111 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (2)$$

Fig. 1 shows the component images in the RC_rQ color space. Note that the *R* component image in Fig. 1 has the fine face region, which is suitable for the extraction of Gabor features, while the C_r and *Q* component images contain partial face contour information.

Fig. 1. RC_rQ hybrid color space. Note that the R component image has the fine face region while the C_r and Q component images contain partial face contour information.

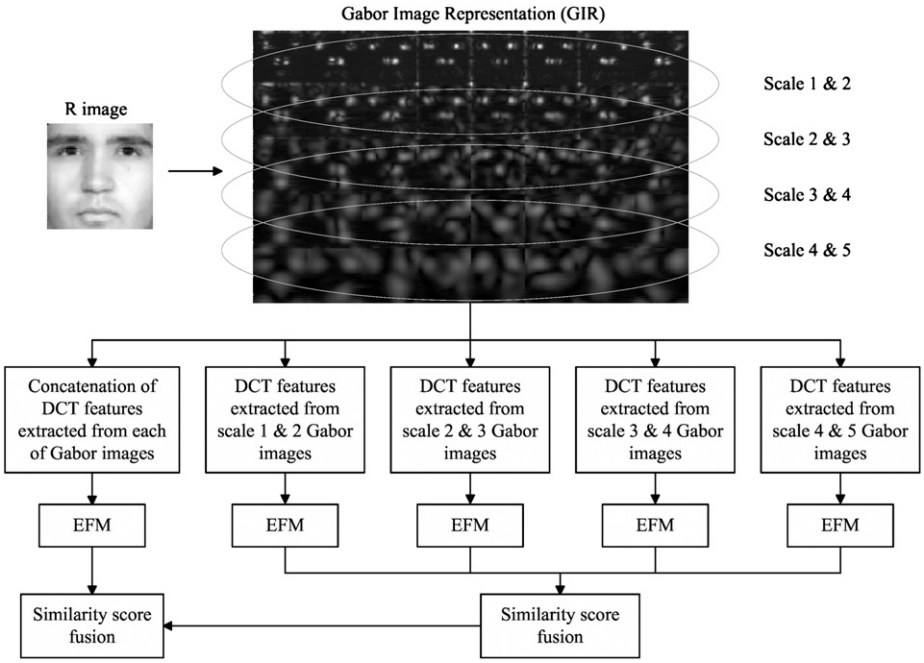

Fig. 2. The outline of face recognition using the GIR.

3.2. The patch-based Gabor image representation for the R image

The Gabor image representation (GIR) of an image captures salient visual properties such as spatial location, orientation selectivity, and spatial frequency characteristics [7]. Specifically, the GIR is the convolution of the image with a family of Gabor kernels that may be formulated as follows [7]:

$$\psi_{\mu,v}(z) = \frac{\|k_{\mu,v}\|^2}{\sigma^2} e^{-\|k_{\mu,v}\|^2\|z\|^2/2\sigma^2}[e^{ik_{\mu,v}z} - e^{-(\sigma^2/2)}] \quad (3)$$

where μ and v define the orientation and scale of the Gabor kernels, $z=(x,y)$, $\|\cdot\|$ denotes the norm operator, and the wave vector $k_{\mu,v}$ is defined as follows:

$$k_{\mu,v} = k_v e^{i\phi_\mu} \quad (4)$$

where $k_v = k_{\max}/f^v$ and $\phi_\mu = \pi\mu/8$. $k_{\max}$ is the maximum frequency, and f is the spacing factor between kernels in the frequency domain. Let $R(x,y)$ represent the R component image, the convolution of image R and a Gabor kernel $\psi_{\mu,v}$ may be formulated as follows:

$$O_{\mu,v}(z) = R(z) * \psi_{\mu,v}(z) \quad (5)$$

where $z=(x,y)$, $*$ denotes the convolution operator, and $O_{\mu,v}(z)$ is the convolution result corresponding to the Gabor kernel at orientation μ and scale v. Commonly used Gabor kernels contain five different scales, $v \in \{0,\ldots,4\}$, and eight orientations, $\mu \in \{0,\ldots,7\}$. The set $\mathcal{S} = \{O_{\mu,v}(z) : \mu \in \{0,\ldots,7\}, v \in \{0,\ldots,4\}\}$, thus, forms the GIR of the image R.

The advantage of GIR stems from the integration of different spatial frequencies, spatial localities, and orientation selectivities. The GIR thus contains rich information for face recognition, which can be applied to extract features both locally and holistically. Fig. 2 shows the outline of face recognition using the GIR. For local GIR feature extraction, we separate the GIR into an ensemble of patches. The GIR is disintegrated into four patches along the horizontal direction, with the adjacent scale images forming one group. As the GIR patch images reside in a very high dimensional space (16 times the original image size), DCT is applied for dimensionality reduction for improving computational efficiency. To facilitate the DCT feature extraction, each GIR patch image is reshaped to a square, as shown in Fig. 3. After transforming the GIR patch image to the DCT domain, a frequency set selection scheme using a square mask is applied to select the low frequency feature set located in the upper-left corner. Then, the enhanced Fisher model (EFM) is used to classify these DCT feature sets. The four GIR patches generate four similarity matrices, which are fused by means of the sum rule. For holistic GIR feature extraction, the GIR is considered as a whole for classification. In particular, each of the Gabor convolved images is processed for dimensionality reduction using DCT. The DCT features derived from the 40 Gabor images are then concatenated to form an augmented vector, which is classified by the EFM. Previous

research shows that Gabor kernels with different scales help improve performance differently, which implies that different number of DCT features should be used to capture the discriminating information corresponding to the Gabor kernels with different scales. The similarity matrix generated via this approach is subsequently fused with the one generated via the local approach by means of the sum rule, as shown in Fig. 2.

3.3. The multi-resolution LBP feature fusion for the C_r image

The success of local binary patterns (LBP) in face recognition is due to its robustness in terms of gray-level monotonic transformation. In practice, the face consists of uneven skin surface, which usually leads to nonmonotonic gray-level transformation. In this case, the performance of LBP degrades significantly, while Gabor kernel filters display excellent capabilities of tolerating such variations. Compared with the R and Y images, as the C_r image lacks the detailed information of skin surface, it causes less nonmonotonic gray-level transformation. Our research reveals that LBP outperform GIR in extracting discriminating features from the C_r image for face recognition.

In a 3×3 neighborhood of an image, the basic LBP operator assigns a binary label 0 or 1 to each surrounding pixel by thresholding at the gray value of the central pixel and replacing its value with a decimal number converted from the 8-bit binary number. Formally, the LBP operator is defined as follows:

$$LBP = \sum_{p=0}^{7} 2^p s(i_p - i_c) \tag{6}$$

where $s(i_p - i_c)$ equals 1, if $i_p - i_c \geq 0$; and 0, otherwise. Two extensions of the basic LBP were further developed. The first extension allows LBP to deal with any size of neighborhoods by using circular neighborhoods and bilinearly interpolating the pixel values. The second extension defines the so called *uniform patterns*. When the binary string is considered circular, we can call LBP uniform if there are at most two bitwise transitions from 0 to 1 or vice versa. After extensions, LBP can be expressed as: $LBP^{u\,2}_{P,R}$, where P, R means P sampling points on a circle of radius R.

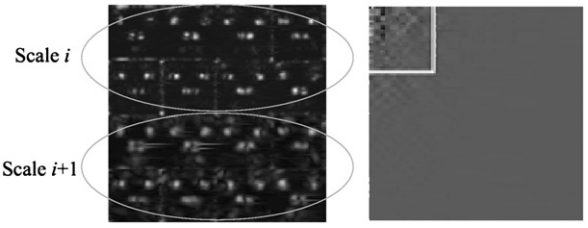

Fig. 3. A Gabor patch image and its DCT image, where $i \in \{1,\ldots,4\}$. A frequency set selection scheme, which selects different sets of the DCT coefficients to encode Gabor patch images, is used to reduce the dimensionality and decorrelate the redundancy of the GIR patch images.

Note that the Gabor image representation encompasses the features corresponding to five scales for improving face recognition performance. Inspired by this idea, we combine the multiple-resolution information from the LBP operators. First, three LBP operators, $LBP^{u\,2}_{8,1}$, $LBP^{u\,2}_{8,2}$ and $LBP^{u\,2}_{8,3}$, are used to extract the multi-resolution histogram features from the C_r image. Fig. 4 shows three local histograms derived from three scale LBP operators in a subwindow of an image. It is easy to see that these histograms are complementary to one another. Smaller scale operators extract more detailed information (microstructure) and maintain the similar profile (macrostructure) as larger operators do. Second, three global histogram features are fused to form an augmented feature. One straightforward way is to concatenate the three global histograms, corresponding to $LBP^{u\,2}_{8,1}$, $LBP^{u\,2}_{8,2}$ and $LBP^{u\,2}_{8,3}$. However, this operation will result in the problem of high dimensionality. In this paper, we propose an LBP multiple-resolution feature fusion scheme, as shown in Fig. 5. For each global LBP histogram, the EFM is used to extract features and reduce dimensionality. Let $\mathbf{X}_{h_1}$, $\mathbf{X}_{h_2}$, and $\mathbf{X}_{h_3}$ be the reduced features after the EFM process. In particular, we first normalize and then concatenate the three reduced features and derive an augmented feature vector, $\mathbf{Y} = (\mathbf{X}_{h_1} - \mu_1/\delta_1; \mathbf{X}_{h_2} - \mu_2/\delta_2; \mathbf{X}_{h_3} - \mu_3/\delta_3)$, where μ_i and δ_i are the mean and standard deviation of feature $\mathbf{X}_{h_i}$. By applying this fusion scheme, both the microstructures and the macrostructures of face image are utilized to extract the discriminating features, which contain much more face information than what a single LBP operator can provide.

3.4. The component-based DCT multiple face encoding for the Q image

In the *YIQ* color space, Y, I, and Q represent luminance, hue, and saturation, respectively. In terms of face image quality, the Q component image seems to contain too much noise to be applied directly for face recognition by local feature extraction methods, such as GIR. The rough outlines of facial components in the Q image, such as nose, eyes and mouth, however, still contain

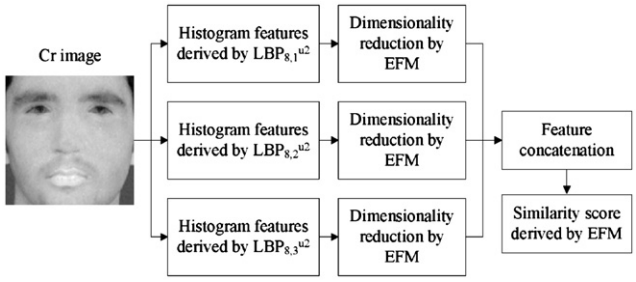

Fig. 5. Multiple resolution LBP feature fusion scheme.

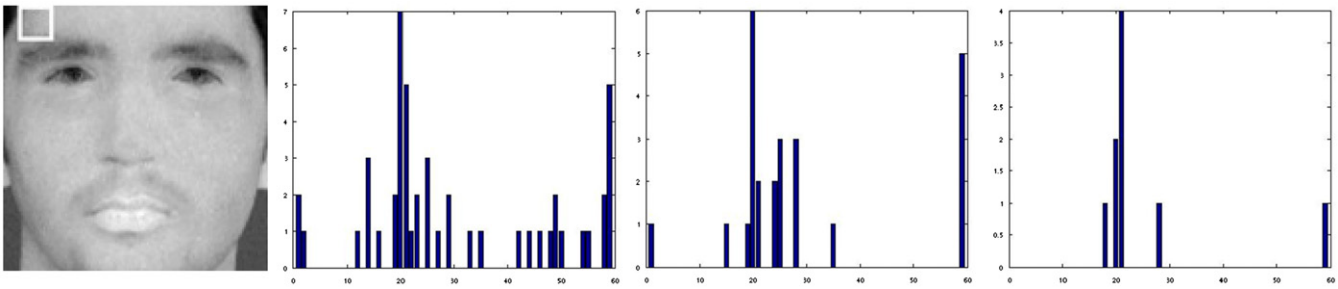

Fig. 4. A C_r image and three local histograms corresponding to the three scale operators: $LBP^{u\,2}_{8,1}$, $LBP^{u\,2}_{8,2}$, and $LBP^{u\,2}_{8,3}$, from a subwindow of 9×9 pixels.

valuable discriminating information. Based on this observation, we propose a holistic method to utilize the Q component image—the fusion of component-based DCT multiple face encoding.

Component-based methods [18,14] have been shown effective for improving face recognition performance, as the statistical variations caused by illumination and pose in each component image may be smaller than those in the whole face image. We consider a simple separation of three facial components shown in Fig. 6. As eyes and the vicinities have the most important discriminating information, they split up into the left eye component and the right eye component. While the bottom half of the face has weak discriminating capability, it is kept as an entity. Note that there are overlapping regions among the adjacent components, and each of these three components is processed by the DCT 2 (using two DCT masks) face encoding fusion scheme to generate a similarity matrix. These three similarity matrices are further fused using the sum rule to generate a new similarity matrix. The whole Q image is processed by the DCT 3 (using three DCT masks) face encoding fusion scheme to generate a similarity matrix. This similarity matrix is further fused with the similarity matrix derived using the facial components to compute the final similarity matrix.

The proposition of the DCT multiple face encoding fusion scheme is based on the observation that the reconstructed images of different DCT feature sets display different facial details. That is, these DCT feature sets are supposed to be complementary to each other. When fusing their classification outputs, the final classification result should be improved. Fig. 7 shows the outline of the three DCT face encoding fusion scheme for the whole Q image. Three masks, which are defined in the DCT domain shown in Fig. 3, are used to select three DCT feature sets. The size of the mask will be discussed in the experiment section.

3.5. Feature extraction using the enhanced Fisher model (EFM)

The EFM [21] method first applies PCA to reduce the dimensionality of the input pattern vector. Let $\mathbf{Y} \in \mathbb{R}^N$ be a random vector, whose covariance matrix is $\Sigma_{\mathbf{Y}}$. The PCA factorizes the covariance matrix $\Sigma_{\mathbf{Y}}$ into the following form:

$$\Sigma_{\mathbf{Y}} = \Phi \Lambda \Phi^t \tag{7}$$

where $\Phi = [\phi_1 \phi_2 \dots \phi_N] \in \mathbb{R}^{N \times N}$ is an orthonormal eigenvector matrix and $\Lambda = diag\{\lambda_1, \lambda_2, \dots, \lambda_N\} \in \mathbb{R}^{N \times N}$ a diagonal eigenvalue matrix. An important application of PCA is dimensionality reduction:

$$\mathbf{Z} = P^t \mathbf{Y} \tag{8}$$

where $P = [\phi_1 \phi_2 \dots \phi_m]$, $m < N$, and $P \in \mathbb{R}^{N \times m}$.

PCA is an optimal representation method in terms of mean square error, but it is not an ideal method for pattern classification. A popular classification method that achieves high separability among the different pattern classes is the Fisher linear discriminant (FLD). The FLD method applies the within-class scatter matrix, S_w, and the between-class scatter matrix, S_b, which may be formulated as follows:

$$S_w = \sum_{i=1}^{L} P(\omega_i) \mathcal{E}\{(\mathbf{Z} - M_i)(\mathbf{Z} - M_i)^t | \omega_i\} \tag{9}$$

$$S_b = \sum_{i=1}^{L} P(\omega_i)(M_i - M_0)(M_i - M_0)^t \tag{10}$$

where M_i is the mean vector of class ω_i, M_0 is the grand mean vector, $P(\omega_i)$ is the prior probability of class ω_i, and L is the number of classes. FLD derives a projection basis that maximizes the criterion $J_1 = tr(S_w^{-1} S_b)$. This criterion is maximized when Ψ consists of the eigenvectors of the matrix $S_w^{-1} S_b$:

$$S_w^{-1} S_b \Psi = \Psi \Delta \tag{11}$$

where Ψ, Δ are the eigenvector and eigenvalue matrices of $S_w^{-1} S_b$, respectively. The most discriminating features are derived by projecting the pattern vector $\mathbf{Z}$ onto the eigenvectors in Ψ:

$$\mathcal{U} = \Psi^t \mathbf{Z} \tag{12}$$

$\mathcal{U}$ thus contains the most discriminating features for pattern recognition.

The FLD method, if implemented in an inappropriate PCA space, may lead to overfitting. The EFM method [21], which applies an eigenvalue spectrum analysis criterion to choose the number of principal components to avoid overfitting, thus improves the generalization performance of the FLD. To achieve enhanced performance, the EFM method preserves a proper balance between the need that the selected eigenvalues account for most of the spectral energy of the raw data (for representational adequacy), and the requirement that the eigenvalues of the within-class scatter matrix (in the reduced PCA space) are not too small (for better generalization performance) [21].

The EFM method thus derives an appropriate low dimensional representation $\mathbf{Z}$ (see Eq. (8)) and further extracts the EFM features $\mathcal{U}$ (see Eq. (12)) from $\mathbf{Z}$ for pattern classification. Pattern classification is implemented by computing the similarity score between a target discriminating feature vector and a query discriminating feature vector. Specifically, the similarity score is calculated by the cosine similarity measure in our experiments:

$$\delta_{\cos}(\mathcal{U}, \mathcal{V}) = \frac{\mathcal{U}^t \mathcal{V}}{\|\mathcal{U}\| \|\mathcal{V}\|} \tag{13}$$

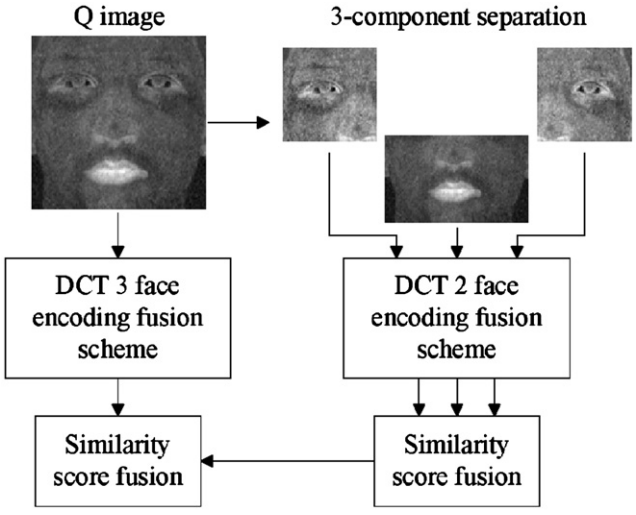

Fig. 6. Component-based DCT multiple face encoding fusion scheme.

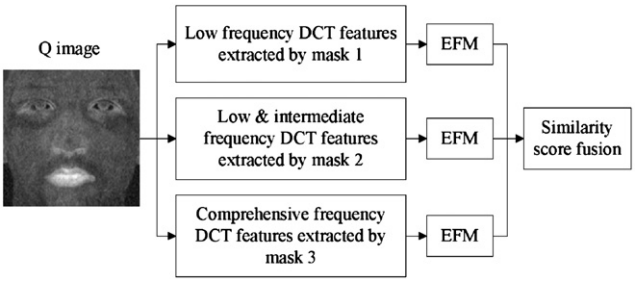

Fig. 7. DCT multiple face encoding fusion scheme.

Table 1
Number of images and image quality of the training, target, and query sets of the FRGC version 2 Experiment 4.

Data	Experiment	Set	Images	Image quality
FRGC version 2	Experiment 4	Training	12,776	Controlled or uncontrolled
		Target	16,028	Controlled
		Query	8014	Uncontrolled

where $\mathcal{U}$ and $\mathcal{V}$ are two discriminating feature vectors, and $\|\cdot\|$ denotes the norm operator.

After the three similarity matrices are generated, a decision level fusion is applied to derive the final similarity matrix based on the multiple imaging in the RC_rQ color space and the multiple feature extraction. Specifically, the three normalized similarity matrices are then fused together to form a new similarity matrix by means of a weighted summation. And finally the fused new similarity matrix is normalized using the z-score normalization technique [16] for deriving the face recognition performance.

4. Experiments

This section assesses the proposed method using the Face Recognition Grand Challenge (FRGC) version 2 database [29]. The most challenging FRGC experiment, the FRGC version 2 Experiment 4 [29], is applied to assess our proposed method. Table 1 presents the training, target, and query sets of the FRGC version 2 database for Experiment 4. In particular, the training set contains 12,776 images that are either controlled or uncontrolled. The target set has 16,028 controlled images and the query set has 8014 uncontrolled images. While the faces in the controlled images have good image resolution and good illumination, the faces in the uncontrolled images have lower image resolution and larger illumination variations. These uncontrolled factors pose grand challenges to the face recognition performance. The size of the images used in our experiments is 128×128 for the R images for Gabor feature extraction, and 64×64 for the C_r and Q images for LBP and DCT feature extraction. Face recognition performance is reported using the receiver operating characteristic (ROC) curves, which plot the face verification rate (FVR) versus the false accept rate (FAR). When a similarity matrix is presented to the BEE system, it generates three ROC curves (ROC I, ROC II, and ROC III) corresponding to the images collected within semesters, within a year, and between semesters, respectively [29]. The FRGC baseline algorithm, which in essence is a PCA algorithm optimized for large scale problems [3,29,33], provides the minimum level of performance for evaluating new face recognition methods. In particular, the FRGC baseline face verification rate (ROC III) is 11.86% at the false accept rate of 0.1%.

To evaluate the effectiveness of the new hybrid color space RC_rQ, we first conduct experiments on the R, C_r, Q and Y component images by applying the EFM method with the number of features $m=1000$ (see Eq. (8)). In particular, the face region of the R and Y images is similar to that of the C_r and Q images in Fig. 1. The spatial resolution of all the images is 64×64. The face recognition ROC curves of FRGC version 2 Experiment 4 are shown in Fig. 8. The FRGC baseline performance is also included for comparison.

The face recognition performance derived from the ROC III curves in Fig. 8 is listed in Table 2, which clearly indicates that the R component image contains much more discriminating information than the grayscale image Y does. Furthermore, by fusion at the decision level, the RC_rQ hybrid color space boosts the face recognition performance significantly. Although the C_r and Q images have lower than 60% face verification rate at 0.1% false accept rate, the overall verification performance of the RC_rQ hybrid color space is improved significantly when they are combined with the R component image. Therefore, the RC_rQ hybrid color space constitutes an excellent platform, from which one can focus on improving performance of each component image and expect to achieve good final performance by fusing their results.

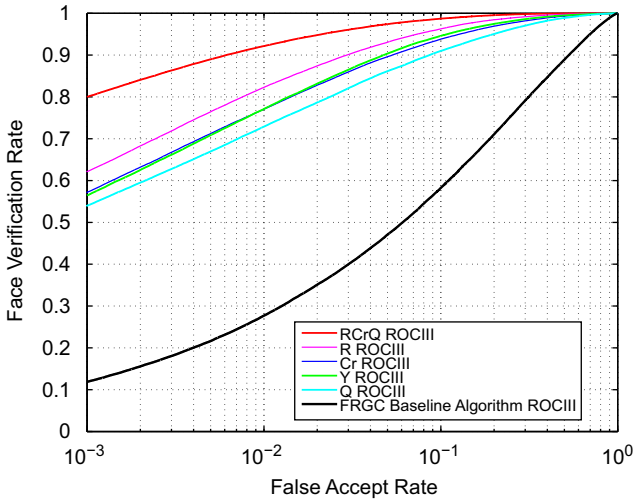

Fig. 8. FRGC version 2 Experiment 4 face recognition performance (the ROC III curve) using *Y*, *R*, *Cr*, *Q*, and *RCrQ*, respectively.

Table 2
Face verification rate (FVR) at 0.1% false accept rate (FAR) of color component images.

Color component image	FVR (ROCIII) at 0.1% FAR (%)
Y	56.41
C_r	57.17
Q	53.97
R	62.12
RC_rQ	79.98

Our patch-based Gabor image representation method extracts the DCT features from the GIR patches using a DCT mask. In particular, the size of the DCT mask is set to 64×64 for extracting the DCT features from the horizontal patches. Prior to this operation, the patches are reshaped into a square array. To evaluate the effectiveness of this reshaping idea, we conduct experiments on the patches before and after the reshaping operations. Fig. 9 shows the comparative performance, which indicates that the reshaping operation indeed helps extract more discriminating DCT features in our Gabor image representation method.

When the GIR is considered as a whole, its dimensionality is reduced using the DCT domain mask defined in Fig. 3. As different Gabor kernels contribute to the recognition performance differently, the size of masks is empirically chosen as follows to comply with such a characteristic: 8×8, 14×14, 17×17, 19×19, and 20×20, corresponding to the kernel scales from 1 to 5, respectively. The resulting feature vector with size 10,480 is then processed by EFM ($m=1250$), producing the FVR (ROCIII) of 73.30% at FAR of 0.1%. For each of the GIR patch images, the 64×64 DCT features are chosen via masking. EFM, with $m=1350$, is used to classify these features. Fig. 10 shows the ROC curves for the original R image using Gabor features. The face verification results derived from the ROC III curves are listed in Table 3. Note that the numbers in the parentheses indicate the number of

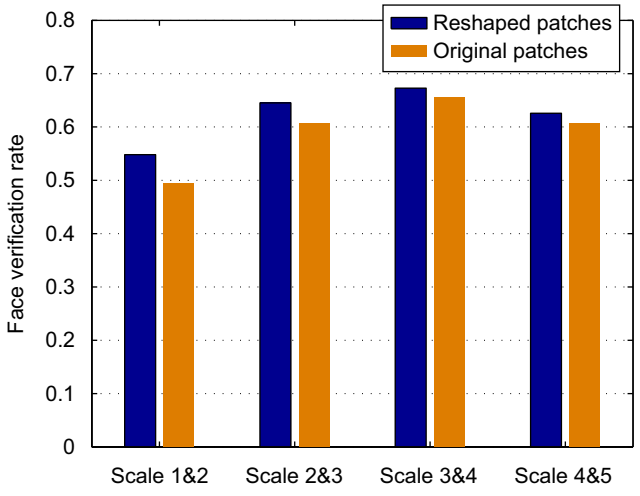

Fig. 9. The face verification rate (ROC III) of the Gabor patches at 0.1% false accept rate.

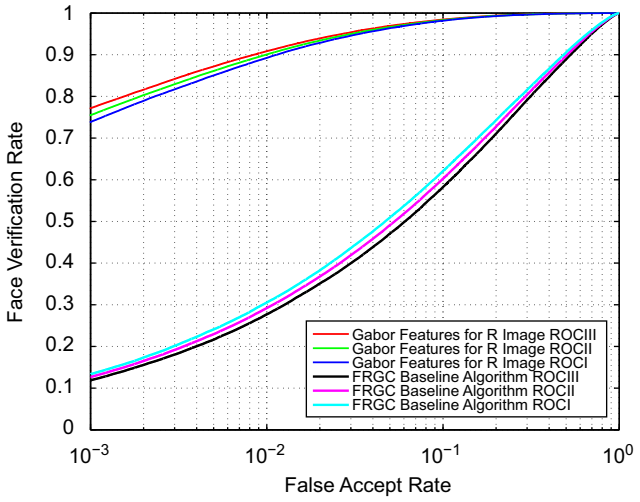

Fig. 10. FRGC version 2 Experiment 4 face recognition performance using the proposed Gabor features for the original R image. The FRGC baseline performance using gray scale images is also included for comparison.

Table 3
Face verification rate (ROC III) at 0.1% false accept rate of the R color component image.

Method	Original image (features)	Normalized image (features)
Scale 1&2	54.81% (1350)	68.27% (1350)
Scale 2&3	64.51% (1350)	72.56% (1350)
Scale 3&4	67.28% (1350)	73.99% (1350)
Scale 4&5	62.57% (1350)	69.61% (1350)
Fusion I	75.88%	82.81%
Whole GIR	73.30% (1250)	80.81% (1100)
Fusion II	77.14%	84.12%

Note that Fusion I corresponds to the fusion of the four patches, while Fusion II to the fusion of the whole GIR image and Fusion I.

features used by the EFM method. To alleviate the effect of illumination variations, an illumination normalization procedure is applied to the R images [23]. The illumination normalization is not applied to the C_r and Q images, because the irregular intensity values in C_r and Q images usually lead to unstable illumination normalization results that compromise face recognition performance. The results using the normalized R image are listed in Table 3 as well.

To extract the LBP features, we divide a face image of 64 $\times$64 into 144 (12*12) overlapping windows of 9 $\times$9 pixels (3 pixels overlapping). The EFM method (m=1350) is used to derive the discriminating features from each of the three scale LBP histograms. After concatenation, the EFM method (m=300) is applied again to process the augmented feature vector. Fig. 11 shows the FRGC version 2 Experiment 4 face recognition performance using the LBP feature fusion from the C_r image. The FRGC baseline performance using gray scale images is also included for comparison. The face verification rates (ROC III) at 0.1% false accept rate of Fig. 11 are listed in Table 4, which clearly indicate that the fusion of the LBP features help improve face recognition performance.

We now assess the Q component image using the proposed component-based DCT multiple face encoding. In particular, for the Q image with a spatial resolution of 64 $\times$64, the size of the left eye, the right eye region and the bottom half of the face is 39 $\times$39, 39 $\times$39, and 39 $\times$64, respectively. For the DCT multiple face encoding, the selection of the DCT mask sizes affects the performance after fusion. As the number of subjects of training data is 222, the rank of the between-class scatter matrix is at most 221. In order to derive the 221 EFM features, the input feature vector should reside in a space whose dimensionality is larger than 221. We therefore choose $\mathbf{M}_{15\times 15}$ as the smallest mask. The selection of the largest mask is determined by the size of the image. For the holistic image, the size of the middle mask is empirically chosen to contain the low and intermediate frequencies. This size of the middle mask is 26 $\times$26 in our experiments. Fig. 12 shows the FRGC version 2 Experiment 4 face recognition performance using the DCT features for the Q image. The FRGC baseline performance using gray scale images is also included for comparison. The face verification rates (ROC III) at 0.1% false accept rate of Fig. 12 are listed in Table 5, where the number of features for the EFM method is also included.

After generating three similarity matrices corresponding to the three component images in the RC_rQ color space, we fuse them by means of a weighted sum rule. In our experiments, we empirically set the weights to 1.0, 0.6, and 0.8, respectively, based on the different roles of the R, C_r, and Q component images for face recognition. Fig. 13 shows the FRGC version 2 Experiment 4 face recognition performance of the proposed method. The FRGC

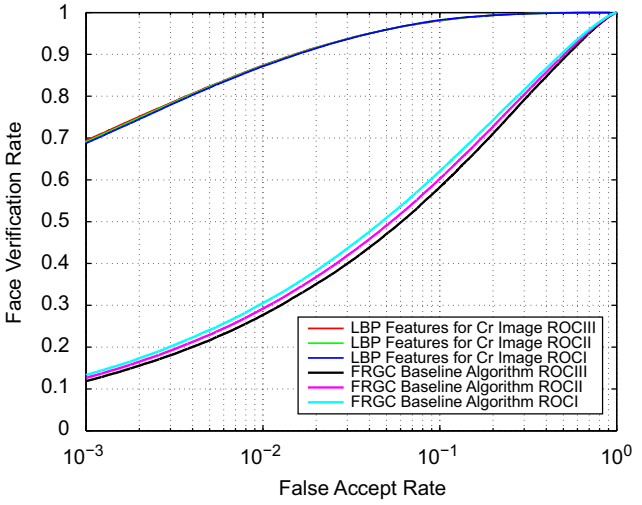

Fig. 11. FRGC version 2 Experiment 4 face recognition performance using the fused LBP features for the C_r image. The FRGC baseline performance using gray scale images is also included for comparison.

Table 4
Face verification rate (ROC III) at 0.1% false accept rate of C_r image.

Method	Face verification rate (%)
$LBP^{u2}_{8,1}$	61.04
$LBP^{u2}_{8,2}$	62.63
$LBP^{u2}_{8,3}$	53.84
Fusion	69.38

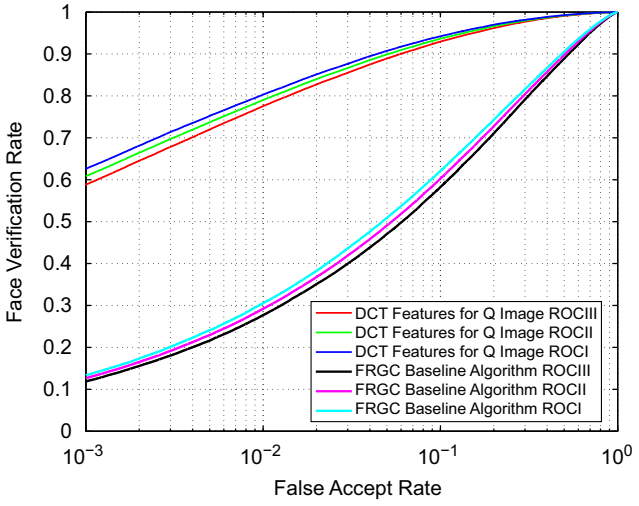

Fig. 12. FRGC version 2 Experiment 4 face recognition performance using the DCT features for the Q image. The FRGC baseline performance using gray scale images is also included for comparison.

Table 5
Face verification rate at 0.1% false accept rate of the Q color component image.

Image	Mask	Features	FVR (ROC III) at 0.1% FAR (%)			
Holistic	$\mathbf{M}_{15\times15}$	223	41.48	56.17		
	$\mathbf{M}_{26\times26}$	600	53.12			
	$\mathbf{M}_{64\times64}$	770	54.46			
LE	$\mathbf{M}_{15\times15}$	223	28.97	33.59		
	$\mathbf{M}_{39\times39}$	510	32.79			58.83
RE	$\mathbf{M}_{15\times15}$	223	34.25	38.34	54.81	
	$\mathbf{M}_{39\times39}$	510	36.41			
Bottom half	$\mathbf{M}_{15\times15}$	223	26.55	33.40		
	$\mathbf{M}_{39\times64}$	650	32.97			

Note that LE and RE represent the left eye component and the right eye component, respectively.

baseline performance using gray scale images is also included for comparison. Specifically, our proposed method, which achieves the face verification rate of 92.43% at the false accept rate of 0.1%, compares favorably against the published results on the FRGC version 2 Experiment 4. In particular, the face recognition performance of our proposed method, 92.43%, is better than the most recent best performance, 89%, reported in [31]: "on Experiment 4, the proposed method (HEC) achieves 89% verification rate at FAR=0.1%, which is 8% higher than the best known results." Our face recognition result (92.43%) significantly improves upon the 63% verification rate at 0.1% false accept rate by the kernel class-dependence feature analysis method [34], and the 74.33% verification rate at 0.1% false accept rate by the hybrid Fourier features method [15].

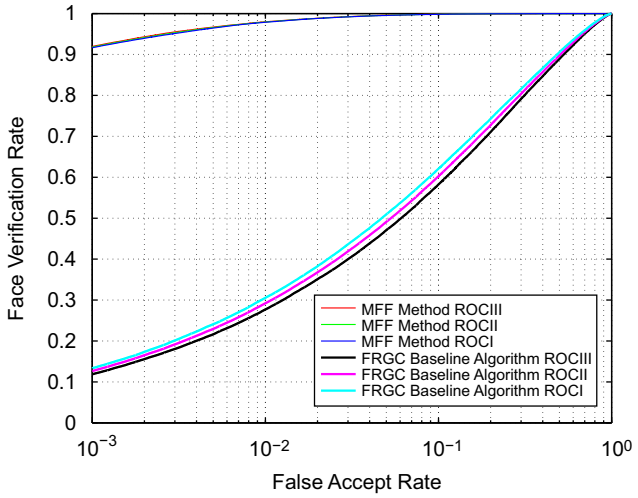

Fig. 13. FRGC version 2 Experiment 4 face recognition performance using the proposed method (MFF) with the illumination normalized R image. The FRGC baseline performance using gray scale images is also included for comparison.

5. Conclusions

A novel face recognition method by means of fusing color, local spatial and global frequency information is presented in this paper, and large scale experiments using the most challenging FRGC version 2 Experiment 4 are carried out to show the feasibility of the proposed method. The novelty of the method rests on the fusion of multiple local spatial and global frequency features in a novel hybrid color space. Specifically, the RC_rQ color space is constructed by combining the R component image of the RGB color space and the chromatic component images, C_r and Q, of the YC_bC_r and YIQ color spaces, respectively. Three effective image encoding methods are then proposed for the component images in the RC_rQ hybrid color space to extract features: (i) a patch-based Gabor image representation for the R component image, (ii) a multi-resolution LBP feature fusion scheme for the C_r component image, and (iii) a component-based DCT multiple face encoding for the Q component image. Finally, at the decision level, the similarity matrices generated using the three component images in the RC_rQ hybrid color space are fused using a weighted sum rule. Experiments show that the proposed method, which achieves the face verification rate of 92.43% at the false accept rate of 0.1%, compares favorably against the published results on the FRGC version 2 Experiment 4.

Acknowledgments

The author would like to thank the anonymous reviewers for their constructive comments and suggestions that help improve the quality of the paper.

References

[1] T. Ahonen, A. Hadid, M. Pietikainen, Face description with local binary patterns: application to face recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence 28 (12) (2006).
[2] 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.
[3] J.R. Beveridge, D. Bolme, B.A. Draper, M. Teixeira, The csu face identification evaluation system: its purpose, features, and structure, Machine Vision and Applications 16 (2) (2005) 128–138.
[4] K.W. Bowyer, K. Chang, P.J. Flynn, A survey of approaches and challenges in 3D and multi-modal 3D+2D face recognition, Computer Vision and Image Understanding 101 (1) (2006) 1–15.

[5] K.W. Bowyer, K.I. Chang, P. Yan, P.J. Flynn, E. Hansley, S. Sarkar, Multi-modal biometrics: an overview, in: Second Workshop on MultiModal User Authentication (MMUA 2006), Toulouse, France, 2006.
[6] J.G. Daugman, Two-dimensional spectral analysis of cortical receptive field profiles, Vision Research 20 (1980) 847–856.
[7] J.G. Daugman, Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional cortical filters, Journal of Optical Society of America 2 (7) (1985) 1160–1169.
[8] G. Donato, M.S. Bartlett, J.C. Hager, P. Ekman, J. Sejnowski, Classifying facial actions, IEEE Transactions on Pattern Analysis and Machine Intelligence 21 (10) (1999) 974–989.
[9] M.J. Er, W. Chen, S. Wu, High-speed face recognition based on discrete cosine transform and rbf neural networks, IEEE Transactions on Neural Networks 16 (3) (2005) 679–691.
[10] K. Etemad, R. Chellappa, Discriminant analysis for recognition of human face images, Journal of Optical Society of America A 14 (1997) 1724–1733.
[11] G.D. Finlayson, S.D. Hordley, P.M. Hubel, Color by correlation: a simple, unifying framework for color constancy, IEEE Transactions on Pattern Analysis and Machine Intelligence 23 (11) (2001) 1209–1221.
[12] R.C. Gonzalez, R.E. Woods, Digital Image Processing, second ed., Prentice Hall, Upper Saddle River, New Jersey, 2002.
[13] Z.M. Hafed, M.D. Levine, Face recognition using the discrete cosine transform, International Journal of Computer Vision 43 (3) (2001) 167–188.
[14] B. Heisele, T. Serre, T. Poggio, A component-based framework for face recognition and identification, International Journal of Computer Vision 74 (2) (2007) 167–181.
[15] W. Hwang, G. Park, J. Lee, S.C. Kee, Multiple face model of hybrid fourier feature for large face image set, in: Proceedings 2006 IEEE Conference Computer Vision and Pattern Recognition (CVPR), 2006.
[16] A.K. Jain, K. Nandakumar, A. Ross, Score normalization in multimodel biometric systems, Pattern Recognition 38 (2005) 2270–2285.
[17] A.K. Jain, S. Pankanti, S. Prabhakar, L. Hong, A. Ross, Biometrics: a grand challenge, in: Proceedings of the 17th International Conference on Pattern Recognition, 2004, pp. 935–942.
[18] T. Kim, H. Kim, W. Hwang, J. Kittler, Component-based LDA face description for image retrieval and mpeg-7 standardisation, Image and Vision Computing 23 (2005) 631–642.
[19] J. Kittler, M. Hatef, P.W. Robert, J. Matas, On combining classifiers, IEEE Transactions on Pattern Analysis and Machine Intelligence 20 (3) (1998) 226–239.
[20] C. Liu, Capitalize on dimensionality increasing techniques for improving face recognition grand challenge performance, IEEE Transactions on Pattern Analysis and Machine Intelligence 28 (5) (2006) 725–737.
[21] C. Liu, H. Wechsler, Robust coding schemes for indexing and retrieval from large face databases, IEEE Transactions on Image Processing 9 (1) (2000) 132–137.
[22] C. Liu, H. Wechsler, Gabor feature based classification using the enhanced Fisher linear discriminant model for face recognition, IEEE Transactions on Image Processing 11 (4) (2002) 467–476.
[23] Z. Liu, C. Liu, Fusion of the complementary discrete cosine features in the yiq color space for face recognition, Computer Vision and Image Understanding 111 (3) (2008) 249–262.
[24] X. Lu, Y. Wang, A.K. Jain, Combining classifiers for face recognition, in: Proceedings of the IEEE International Conference on Multimedia & Expo, July 2003.
[25] A.S. Mian, M. Bennamoun, R. Owens, An efficient multimodal 2d–3d hybrid approach to automatic face recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence 29 (11) (2007) 1927–1943.
[26] T. Ojala, M. Pietikainen, D. Harwood, A comparative study of texture measures with classification based on feature distributions, Pattern Recognition 29 (1) (1996) 51–59.
[27] A.J. OToole, H. Abdi, F. Jiang, P.J. Phillips, Fusing face recognition algorithms and humans, IEEE Transactions on Systems, Man, and Cybernetics 37 (5) (2007) 1149–1155.
[28] A.J. OToole, P.J. Phillips, F. Jiang, J. Ayyad, N. Penard, H. Abdi, Face recognition algorithms surpass humans matching faces across changes in illumination, IEEE Transactions on Pattern Analysis and Machine Intelligence 29 (9) (2007) 1642–1646.
[29] P.J. Phillips, P.J. Flynn, T. Scruggs, K.W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, W. Worek, Overview of the face recognition grand challenge, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2005.
[30] P. Shih, C. Liu, Comparative assessment of content-based face image retrieval in different color spaces, International Journal of Pattern Recognition and Artificial Intelligence 19 (7) (2005) 873–893.
[31] Y. Su, S. Shan, X. Chen, W. Gao, Hierarchical ensemble of global and local classifiers for face recognition, IEEE Transactions on Image Processing 18 (8) (2009) 1885–1896.
[32] D.L. Swets, J. Weng, Using discriminant eigen features for image retrieval, IEEE Transactions on Pattern Analysis and Machine Intelligence 18 (8) (1996) 831–836.
[33] M. Turk, A. Pentland, Eigenfaces for recognition, Journal of Cognitive Neuroscience 13 (1) (1991) 71–86.
[34] C. Xie, V. Kumar, Comparison of kernel class-dependence feature analysis (kcfa) with kernel discriminant analysis (kda) for face recognition, in: Proceedings of the IEEE on Biometrics: Theory, Application and Systems (BTAS'07), 27–29, September 2007.
[35] J. Yang, C. Liu, A general discriminant model for color face recognition, in: Proceedings of the IEEE International Conference on Computer Vision, 14–20 October 2007.

About the Author—ZHIMING LIU received his MS degree from University of Nevada, Reno, NV. He is presently a PhD student of computer science at New Jersey Institute of Technology. His research interests are in face recognition and image processing.

About the Author—CHENGJUN LIU received the Ph.D. from George Mason University in 1999, and he is presently an Associate Professor of Computer Science at New Jersey Institute of Technology. His research interests are in pattern recognition (face/iris recognition), machine learning (statistical learning, Kernel methods, similarity measures), computer vision (object/face detection, video processing), security (biometrics), and image processing (new color spaces, Gabor image representation). His recent research has been concerned with the development of novel and robust methods for image/video retrieval and object detection, tracking and recognition based upon statistical and machine learning concepts. The class of new methods he has developed includes the Bayesian discriminating features method (BDF), the probabilistic reasoning models (PRM), the enhanced Fisher models (EFM), the enhanced independent component analysis (EICA), the shape and texture-based Fisher method (STF), the Gabor–Fisher classifier (GFC), and the independent Gabor features (IGF) method. He has also pursued the development of novel evolutionary methods leading to the development of the evolutionary pursuit (EP) method for pattern recognition in general, and face recognition in particular.