Capitalize on Dimensionality Increasing Techniques for Improving Face Recognition Grand Challenge Performance

Chengjun Liu

Abstract—This paper presents a novel pattern recognition framework by capitalizing on dimensionality increasing techniques. In particular, the framework integrates Gabor image representation, a novel multiclass Kernel Fisher Analysis (KFA) method, and fractional power polynomial models for improving pattern recognition performance. Gabor image representation, which increases dimensionality by incorporating Gabor filters with different scales and orientations, is characterized by spatial frequency, spatial locality, and orientational selectivity for coping with image variabilities such as illumination variations. The KFA method first performs nonlinear mapping from the input space to a high-dimensional feature space, and then implements the multiclass Fisher discriminant analysis in the feature space. The significance of the nonlinear mapping is that it increases the discriminating power of the KFA method, which is linear in the feature space but nonlinear in the input space. The novelty of the KFA method comes from the fact that 1) it extends the two-class kernel Fisher methods by addressing multiclass pattern classification problems and 2) it improves upon the traditional Generalized Discriminant Analysis (GDA) method by deriving a unique solution (compared to the GDA solution, which is not unique). The fractional power polynomial models further improve performance of the proposed pattern recognition framework. Experiments on face recognition using both the FERET database and the FRGC (Face Recognition Grand Challenge) databases show the feasibility of the proposed framework. In particular, experimental results using the FERET database show that the KFA method performs better than the GDA method and the fractional power polynomial models help both the KFA method and the GDA method improve their face recognition performance. Experimental results using the FRGC databases show that the proposed pattern recognition framework improves face recognition performance upon the BEE baseline algorithm and the LDA-based baseline algorithm by large margins.

Index Terms—Dimensionality increasing techniques, face recognition, Face Recognition Grand Challenge (FRGC), fractional power polynomial models, Gabor image representation, Kernel Fisher Analysis (KFA) method.

1 Introduction

Popular pattern recognition paradigms based on data reduction, such as redundancy reduction and dimensionality reduction, have met with difficulties in solving complex pattern recognition problems, such as the human face recognition problem. As a result, the low-dimensional pattern recognition methods based on such popular techniques as Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), Independent Component Analysis (ICA), etc., cannot achieve satisfactory pattern recognition performance, as evidenced by the recent Face Recognition Grand Challenge (FRGC) competition [33]. New paradigms in pattern recognition, such as kernel methods [6], [38] motivated by the statistical learning theory [42], emphasize the nonlinear mapping from an input space to a high-dimensional feature space. The rationale of performing such a nonlinear mapping comes from Cover's theorem on the separability of patterns, which states that "A complex pattern-classification problem cast in a high-dimensional space nonlinearly is more likely to be linearly separable than in a low-dimensional space" [15]. While these paradigms provide new means for solving pattern recognition problems, some open issues, such as kernel function selection, still require further research [6], [38].

This paper presents a new pattern recognition framework that capitalizes on some dimensionality increasing techniques, such as Gabor image representation, a novel multiclass Kernel Fisher Analysis (KFA) method, and fractional power polynomial models. As the Gabor wavelets model the receptive field profiles of cortical simple cells quite well [8], [9], Gabor image representation is applied to capture the salient visual properties such as spatial localization, orientational selectivity, spatial frequency characteristics. Note that Gabor image representation computes the convolution of an image and the Gabor filters of different scales and orientations. The novelty of the KFA method comes from the fact that it extends the two-class kernel Fisher methods [30], [5] by addressing multiclass pattern classification problems and it improves upon the traditional Generalized Discriminant Analysis (GDA) method [2] by deriving a unique solution. Note that the GDA method finds a solution that is not unique [2]. Finally, the Gabor-based KFA method applies fractional power polynomial models for improving pattern classification performance.

The feasibility of the proposed framework is assessed on face recognition problems. Face recognition has become a very active research area in recent years, driven mainly by its broad applications in human-computer interaction, homeland security, and entertainment [4], [18], [23], [22], [49], [36],

• *The author is with the Department of Computer Science, New Jersey Institute of Technology, Newark, NJ 07102. E-mail: liu@cs.njit.edu.*

Manuscript received 28 Jan. 2005; revised 22 Sept. 2005; accepted 29 Sept. 2005; published online 13 Mar. 2006.
Recommended for acceptance by J. Phillips.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number TPAMI-0062-0105.

[7]. The face recognition vendor test,[1] FRVT 2002, through extensive testing of the state-of-the-art face recognition vendor systems, identifies some challenging research areas in face recognition, such as face recognition from outdoor imagery [35]. Phillips and Newton [36] further recommend that researchers should concentrate on "face recognition problems that are harder, as defined by the image sets in the experiments and the performance by a control algorithm" rather than work on problems that have already been solved. We therefore choose two representative databases to assess the proposed framework: The FERET database [37] and the FRGC Version 1 and Version 2 databases [33]. The FERET data set used in our experiments includes 600 face images corresponding to 200 subjects. Experimental results show that 1) the KFA method achieves better face recognition performance than the GDA method and 2) the fractional power polynomial models help both the KFA method and the GDA method improve their face recognition accuracy. The data set from the FRGC Version 1 database contains 366 training images, 152 gallery images, and 608 probe images. The data set from the FRGC Version 2 database contains 12,776 training images, 16,028 controlled target images, 16,028 controlled query images for Experiment 1, and 8,014 uncontrolled query images for Experiment 4 [33]. Experimental results show that the proposed framework improves face recognition performance by large margins compared to the FRGC baseline algorithms.

2 Background

Feature extraction in pattern recognition needs to consider the effectiveness on both data representation and class separability [12], [31], [13], [43]. The Gabor image representation has become a popular representation method for pattern recognition, in general, and face recognition in particular. Lades et al. [20] applied the Gabor wavelets for face recognition by first computing the Gabor jets and then performing a flexible template matching (graph-matching) using a dynamic link architecture. Wiskott et al. [44] further expanded on the dynamic link architecture and developed a Gabor wavelet-based elastic bunch graph matching method to label and recognize human faces. Applying the 2D Gabor image representation and the labeled elastic graph matching method, Lyons et al. [29], [28] then proposed an algorithm for two-class categorization of gender, race, and facial expression. Recent research [10], [27], [24] showed that the Gabor image representation yielded better performance for face recognition than other commonly used representation methods, such as PCA [19], [41].

One commonly used classification method, the Fisher linear discriminant, or FLD (also known as, the linear discriminant analysis, or LDA), achieves high separability among different classes by deriving a projection basis that separates the different classes as far away as possible and compresses the same classes as compactly as possible [40], [3], [11]. The FLD method, when implemented in a high-dimensional PCA space, often leads to overfitting. Overfitting is more likely to occur for the small training sample size problems, such as the face recognition problem, where usually there are a large number of faces, but only a few training examples per face [34]. One solution to this drawback is to analyze the reasons for overfitting and propose new models with improved generalization abilities [26].

To account for the nonlinear interactions among patterns, the linear FLD method is extended to a nonlinear form, the kernel FLD method [30], [2], [5], [50], [32], [21]. Generally speaking, there are two types of Kernel FLD methods. The first type deals with two-class pattern classification problems and some popular methods are the Fisher discriminant analysis with kernels [30] and two variations on kernel FLD [5]. The second type addresses multiclass pattern classification problems and a representative method is the Generalized Discriminant Analysis (GDA) using a kernel approach [2]. Mika et al. [30] presented a two-class kernel FLD method whose linear classification in the feature space corresponds to a powerful nonlinear classification in the input space. The two-class kernel FLD method is competitive to some other popular classification techniques, such as AdaBoost, regularized AdaBoost, and support vector machines [30]. Baudat and Anouar [2] developed a multiclass kernel FLD method, the GDA method, which finds a nonlinear FLD solution using a kernel function. The GDA solution, however, is not unique as the GDA method chooses a particular solution that satisfies the formulation of the method. Indeed, there exists at least one such solution satisfying the equations but the solution is not unique [2].

Based on the GDA formulation, Yang [46] presented a kernel Fisherfaces method for face recognition. The kernel Fisherfaces method tries to avoid the singularity problem by adopting the same technique of the Fisherfaces method, which applies PCA first to project the pattern vectors "to a lower dimensional space so that the resulting within-class scatter matrix is nonsingular" [3]. While this technique works for the Fisherfaces method, it does not apply to the kernel Fisherfaces method because a lower dimensional space cannot make the kernel matrix nonsingular. As a matter of fact, the kernel matrix is always singular for the centered data, regardless of the dimensionality of the feature space (see Section 3.1). One possible solution to the kernel Fisherfaces method is to adopt the GDA solution. Another possibility is to apply the kernel PCA method first and then the ordinary FLD method (rather than the kernel FLD method): KPCA + FLD, as described in [45]. Zheng et al. [50] further presented a modified algorithm for the GDA method by addressing the issue of several eigenvectors associating with the same eigenvalue. Park and Park [32] recently presented a kernel nonlinear discriminant analysis method using the generalized singular value decomposition to address the singularity problem [47], [17]. Liang and Shi [21] developed a kernel uncorrelated discriminant vectors method and showed that their method performs almost the same as the GDA method.

3 Kernel Fisher Analysis

This section details a novel multiclass Kernel Fisher Analysis (KFA) method and fractional power polynomial models. The KFA method, which derives a unique solution, improves upon the GDA method [2], whose solution is not unique. This section first analyzes the GDA method and then presents the multiclass KFA method.

1. http://www.frvt.org/.

3.1 Discussion of the Generalized Discriminant Analysis Method

The GDA method is a representative multiclass kernel FLD method [2]. The GDA solution, however, is suboptimal in the sense that it is not directly derived from the optimization procedure as the kernel matrix is usually not full rank. Let K be the $M \times M$ matrix defined by the centered observations (mapped elements) in the feature space, where M is the number of the training elements (see equation (2.8) in [2]). GDA first factorizes K into the following form [2]:

$$K = U\Gamma U^t, \tag{1}$$

where Γ, as defined in [2], is a diagonal matrix of nonzero eigenvalues and U consists of the normalized eigenvectors associated with Γ. Note that the matrix K usually is not full rank. For example, if a polynomial kernel function is used to define this matrix K, as the observations (mapped elements) in the feature space are centered, the maximum value of the rank of K is $M-1$. As a result, the dimension of Γ is at most equal to $(M-1)\times(M-1)$ and the matrix U is not a square matrix. By definition, this matrix U cannot be an orthogonal matrix, as an orthogonal matrix must be a square matrix[2] [48]. Now, U^tU is an identity matrix, but UU^t usually is not an identity matrix and not full rank. The GDA optimization procedure further derives the following equation for computing the final solution α (see equation (4.1) in [2]):

$$\beta = \Gamma U^t \alpha. \tag{2}$$

From (2), the optimal solution α should be computed as follows:

$$\alpha = (UU^t)^+ U\Gamma^{-1}\beta, \tag{3}$$

where $(\cdot)^+$ is generalized inverse.

The GDA method, however, does not use (3) as the optimal solution as it involves a generalized inverse operation. Instead, the GDA method chooses the following solution [2]:

$$\alpha = U\Gamma^{-1}\beta. \tag{4}$$

It is pointed out in [2] that there exists at least one such solution that satisfies (2) while a possible solution is not unique. As UU^t usually is not an identity matrix and not full rank, the solution derived by (4) is suboptimal. Only if the matrix K were full rank (then the matrix U should be orthogonal), could (4) and (3) be identical and the GDA solution optimal. But, for the GDA method, this condition cannot be met because the data in the feature space is centered.

3.2 Multiclass Kernel Fisher Analysis

The kernel Fisher analysis, or the KFA method, derives a unique solution for multiclass pattern classification problems based on a discriminant analysis criterion in the high-dimensional feature space. Let $\omega_1, \omega_2, \ldots, \omega_L$ and $N_1, N_2, \ldots, N_L$ denote the classes and the number of samples within each class, respectively. Let $\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_M \in \mathbb{R}^N$ be the training samples in the input space and Φ be a nonlinear mapping between the input space and the feature space: $\Phi : \mathbb{R}^N \rightarrow F$. Assume the mapped data is centered (see [38] for centering data in the feature space) and let $\mathcal{D}$ represent the data matrix in the feature space:

$$\mathcal{D} = [\Phi(\mathcal{X}_1)\Phi(\mathcal{X}_2)\cdots\Phi(\mathcal{X}_M)]. \tag{5}$$

Now, define a kernel matrix, $K \in \mathbb{R}^{M\times M}$, by means of dot product in the feature space:

$$K = \mathcal{D}^t\mathcal{D}, \quad where \quad K_{ij} = \big(\Phi(\mathcal{X}_i)\cdot\Phi(\mathcal{X}_j)\big)\ i,j = 1,2,\cdots,M. \tag{6}$$

Note that, in practice, the kernel matrix K is computed by means of a kernel function rather than by explicitly implementing the nonlinear mapping Φ (see Section 3.3).

As the mapped training data is centered in the feature space, the mixture scatter matrix, S_m, and the between-class scatter matrix, S_b, are calculated as follows:

$$S_m = \mathcal{E}\big\{\Phi(\mathcal{X})\Phi(\mathcal{X})^t\big\} = \frac{1}{M}\mathcal{D}\mathcal{D}^t, \tag{7}$$

$$S_b = \sum_{i=1}^{L} P(\omega_i)\mathcal{E}\{\Phi(\mathcal{X})|\omega_i\}\mathcal{E}\{\Phi(\mathcal{X})|\omega_i\}^t = \frac{1}{M}\mathcal{D}W\mathcal{D}^t, \tag{8}$$

where $\mathcal{E}(\cdot)$ is the expectation operator, $P(\omega_i)$ is a priori probability, $\mathcal{E}\{\Phi(\mathcal{X})|\omega_i\}$ is the mean vector of class ω_i in the feature space, M is the number of training samples, and L denotes the number of classes. $W \in \mathbb{R}^{M\times M}$ is a block diagonal matrix: $W = diag\{W_1, W_2, \cdots, W_L\}$, where $W_j \in \mathbb{R}^{N_j\times N_j}$ is an $N_j \times N_j$ matrix with elements all equal to $\frac{1}{N_j}$, $j = 1,2,\cdots,L$.

A good criterion for class separability should convert these scatter matrices to a number, which becomes larger when the between-class scatter is larger or the within-class scatter is smaller [12]. One typical criterion is $J_1 = tr(S_m^{-1}S_b)$ [12]. Note that using J_1 here to derive the KFA solution in the feature space resembles using J_1 to derive the FLD solution in the input space [12]. When the mixture scatter matrix S_m is singular, an equivalent optimization criterion to J_1 is the optimization of $tr(A^tS_bA)$ with respect to A under the constraint $A^tS_mA = I$, where A is a transformation matrix and I an identity matrix whose dimension is smaller than that of S_m [12]. Optimizing the criterion J_1 or the equivalent optimization criterion leads to the following generalized eigenvalue problem [12]:

$$S_b\mathcal{V} = \lambda S_m\mathcal{V}. \tag{9}$$

From the theory of reproducing kernels, $\mathcal{V} \in F$ must lie in the span of all the training samples in the feature space [30]:

$$\mathcal{V} = \sum_{i=1}^{M} c_i\Phi(\mathcal{X}_i) = \mathcal{D}\alpha, \tag{10}$$

where $\alpha = [c_1, c_2, \cdots, c_M]^t$ and $\alpha \in \mathbb{R}^M$.

The KFA projection matrix thus consists of the eigenvectors corresponding to the largest eigenvalues of (9). However, the scatter matrices, S_m and S_b, reside in the high-dimensional feature space and are difficult to evaluate. To overcome this difficulty, we replace S_m and S_b with the kernel matrix:

$$KWK\alpha = \lambda KK\alpha. \tag{11}$$

2. Note that the term *orthonormal* is used in [2] to describe the matrix U. However, in matrix analysis, the term *orthonormal* is for describing a basis or a set of basis vectors, while the term *orthogonal* is for describing a matrix [48].

Equation (11) is derived by first plugging (7), (8), and (10) into (9); then, multiplying (9) from the left by $\mathcal{D}^t$; and finally, plugging (6) into (9). By introducing a regularization factor, we can convert the generalized eigenvalue problem of (11) into an ordinary eigenvalue problem:

$$(KK + \varepsilon I)^{-1}(KWK)\alpha = \lambda\alpha, \tag{12}$$

where ε is a small positive regularization number and I an $M \times M$ identity matrix.

The vector α should be normalized so that the KFA basis vector, $\mathcal{V}$, has unit norm:

$$\|\mathcal{V}\|^2 = \mathcal{V}^t\mathcal{V} = \alpha^t K\alpha = 1. \tag{13}$$

Let $\mathcal{V}_1, \mathcal{V}_2, \cdots, \mathcal{V}_n$ $(n \leq L-1)$ be the KFA basis vectors associated with the largest eigenvalues of (9) and $\alpha_1, \alpha_2, \cdots, \alpha_n$ be the corresponding normalized vectors of (12). The KFA projection matrix thus becomes:

$$\mathcal{P} = [\mathcal{V}_1\mathcal{V}_2 \cdots \mathcal{V}_n] = \mathcal{D}\mathcal{A}, \tag{14}$$

where $\mathcal{A} = [\alpha_1\alpha_2 \cdots \alpha_n]$ and $\mathcal{A} \in \mathbb{R}^{M\times n}$.

Let $\mathcal{X}$ be a test sample whose image in the feature space is $\Phi(\mathcal{X})$. The KFA features of $\mathcal{X}$ are derived as follows:

$$\mathcal{F} = \mathcal{P}^t\Phi(\mathcal{X}) = \mathcal{A}^t\mathcal{B}, \tag{15}$$

where $\mathcal{B} = [\Phi(\mathcal{X}_1)\cdot\Phi(\mathcal{X}) \quad \Phi(\mathcal{X}_2)\cdot\Phi(\mathcal{X}) \ \cdots \ \Phi(\mathcal{X}_M)\cdot\Phi(\mathcal{X})]^t$, and $\mathcal{F} \in \mathbb{R}^n$.

3.3 Fractional Power Polynomial Models

The KFA method is nonlinear in the input space due to the nonlinear mapping between the input space and the feature space. The advantage of applying the nonlinear mapping is that it increases the discriminating ability of a pattern classifier according to Cover's theorem on the separability of patterns [16]. The nonlinear mapping, with a possibly prohibitive computational cost, is never implemented explicitly [2], [30], [38]. But rather, the KFA method applies kernel functions in the input space to achieve the same effect of the expensive nonlinear mapping. Specifically, like the kernel FLD approach [30], the KFA method takes advantage of the Mercer equivalence condition and is feasible because the dot products in the high-dimensional feature space are replaced by a kernel function in the input space while computation is related to the number of training examples rather than the dimension of the feature space.

Note that both the kernel matrix K (see (6)) and the KFA features $\mathcal{F}$ (see (15)) are defined on inner products of vectors in the high-dimensional feature space, whose computation might be prohibitively expensive. The KFA method, however, manages to compute the inner products by means of a kernel function:

$$k(\mathbf{x}, \mathbf{y}) = (\Phi(\mathbf{x})\cdot\Phi(\mathbf{y})). \tag{16}$$

For a function to be a kernel function, it has to satisfy the Mercer's condition [38], [6]:

Definition 1 (Gram Matrix). *Given a finite data set $\mathcal{X} = \{\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_M\}$ in the input space and a function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ (or $\mathbb{C}$), the $M \times M$ matrix K with elements $K_{ij} = k(\mathcal{X}_i, \mathcal{X}_j)$ is called the Gram matrix of k with respect to $\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_M$.*

Definition 2 (Kernel Function). *A sufficient and necessary condition for a symmetric function to be a kernel function is that its Gram matrix is positive semidefinite.*

Three classes of widely used kernel functions are the polynomial kernels, the Gaussian or RBF kernels, and the sigmoid kernels [38]:

$$k(\mathbf{x}, \mathbf{y}) = (\mathbf{x}\cdot\mathbf{y})^d, \tag{17}$$

$$k(\mathbf{x}, \mathbf{y}) = exp\left(-\frac{\|\mathbf{x}-\mathbf{y}\|^2}{2\sigma^2}\right), \tag{18}$$

$$k(\mathbf{x}, \mathbf{y}) = tanh(\kappa(\mathbf{x}\cdot\mathbf{y}) + \vartheta), \tag{19}$$

where $d \in \mathbf{N}$, $\sigma > 0$, $\kappa > 0$, and $\vartheta < 0$.

Note that the sigmoid kernels (see (19)) do not actually define a positive semidefinite Gram matrix, hence are not kernel functions by definition (see Definition 2) [38]. Nevertheless, the sigmoid kernels have been successfully used, in practice, such as in building support vector machines [38]. To further improve pattern classification performance, the KFA method introduces and applies the following fractional power polynomial models:

$$k(\mathbf{x}, \mathbf{y}) = sign(\mathbf{x}\cdot\mathbf{y})(abs(\mathbf{x}\cdot\mathbf{y}))^d, \tag{20}$$

where $sign(\cdot)$ is the sign function, $abs(\cdot)$ computes the absolute value, and $0 < d < 1$. Note that a fractional power polynomial does not necessarily define a kernel function as it might not define a positive semidefinite Gram matrix. A fractional power polynomial is therefore called a model rather than a kernel. The application of the fractional power polynomial models is largely motivated by the successful applications of the sigmoid kernels in practice, which are not kernel functions by definition.

3.4 The KFA Algorithm and Classification Rule

The KFA algorithm works as follows:

1. Choose a kernel function $k(\mathbf{x}, \mathbf{y})$ (see (17), (18), and (19)) or a fractional power polynomial model (see (20)).
2. Center the training data in the feature space (see [38] for centering data in the feature space).
3. Compute the kernel matrix K using the centered training data and the kernel function or the fractional power polynomial model (see (6), (16), and the kernel functions (17), (18), (19), or the fractional power polynomial model (20)).
4. Compute the bloc diagonal matrix, the W matrix (see Section 3.2).
5. Solve the eigenvalue problem of (12) and choose n eigenvectors, $\alpha_1, \alpha_2, \cdots, \alpha_n$, corresponding to the largest eigenvalues. Note that the small positive regularization number ε in (12) is empirically chosen, such as 0.001.
6. Normalize the eigenvectors, $\alpha_1, \alpha_2, \cdots, \alpha_n$, by means of (13) and derive the matrix $\mathcal{A}$ (see (14)).
7. Subtract the grand mean of the training data from every test (gallery/target or probe/query) sample in the feature space (see [38] for subtracting the grand mean in the feature space).

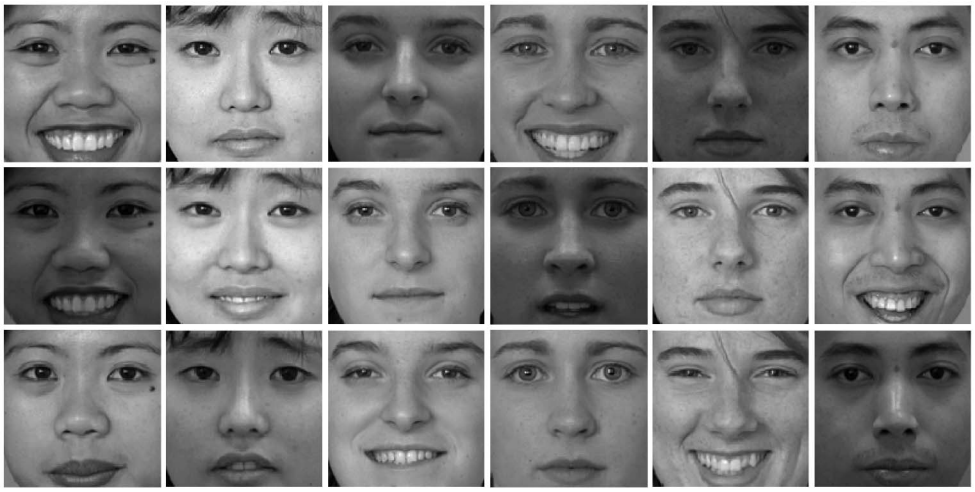

Fig. 1. Example FERET face images used in our experiments. The top two rows show the examples of the training images, while the bottom row displays the examples of the testing images.

8. Compute the vector $\mathcal{B}$ (see (15)) for every test sample using the kernel function or the fractional power polynomial model introduced in Step 1.
9. Compute the KFA features of the test sample using $\mathcal{A}$ and $\mathcal{B}$: $\mathcal{F} = \mathcal{A}^t\mathcal{B}$ (see (15)).

The computational complexity of the algorithm falls mainly into Step 3 and Step 5. Computing the kernel matrix in Step 3 has a complexity of $O(M^2)$, where M is the number of training samples. Solving the eigenvalue problem in Step 5 has a complexity of $O(M^3)$ according to LAPACK Users' Guide [1].

When a probe image is presented to the KFA classifier, it first extracts the KFA features of the probe image and then applies a similarity measure and a classification rule to recognize it. Let $\mathcal{G}_k$ be the gallery KFA feature vector for class ω_k, $k = 1, 2, \ldots, G$. The KFA method applies, then the nearest neighbor rule for classification using some similarity measure (minimum distance) δ:

$$\delta(\mathcal{P}, \mathcal{G}_k) = \min_j \ \delta(\mathcal{P}, \mathcal{G}_j) \ \longrightarrow \ \mathcal{P} \in \omega_k. \tag{21}$$

The KFA feature vector of the probe image, $\mathcal{P}$, is classified as belonging to the class of the nearest neighbor, $\mathcal{G}_k$, using the similarity measure δ. Commonly used similarity measures include the L_1 distance measure, δ_{L_1}, the L_2 distance measure, δ_{L_2}, the Mahalanobis distance measure, δ_{Md}, and the cosine similarity measure, δ_{cos} [24].

4 Experiments

The section assesses the KFA method and the proposed pattern recognition framework on face recognition using the FERET and the FRGC databases [37], [33]. The FERET database [37] consists of more than 13,000 facial images corresponding to more than 1,500 subjects. Since images are acquired during different photo sessions, the illumination conditions and the size of the face may vary. The diversity of the FERET database is across gender, ethnicity, and age. The images are acquired without any restrictions imposed on facial expression and with at least two frontal images shot at different times during the same photo session. The FERET database has become the de facto standard for evaluating face recognition technologies.

The data set used in our experiments consists of 600 FERET frontal face images corresponding to 200 subjects, such that each subject has three images of size 256×384 with 256 gray scale levels. Preprocessing normalizes the face images and extracts a facial region that contains only face so that the performance of face recognition is not affected by factors not related to face, such as hair styles. Specifically, the normalization consists of the following procedures: First, manual annotation detects the centers of the eyes; second, rotation and scaling transformations align the centers of the eyes to predefined locations and fixed interocular distance; finally, a subimage procedure crops the face image to the size of 128×128 to extract the facial region. Fig. 1 shows some example FERET images used in our experiments that are already cropped to the size of 128×128 to extract the facial region. Note that each subject has three images, which are acquired during different photo sessions under variable illumination and facial expression. As two images for each subject are randomly chosen for training and the remaining image (unseen during training) is used for testing (see Fig. 1), the KFA method has to cope with both illumination and facial expression variations.

For comparison, the first set of experiments implements the PCA method [41], using the similarity measures introduced in Section 3.4. Fig. 2 shows the face recognition performance of the PCA method using the four different similarity measures: the L_1 distance measure, the L_2 distance measure, the Mahalanobis distance measure, and the cosine similarity measure. The horizontal axis indicates the number of features used and the vertical axis represents the correct face recognition rate, which is the rate that the top response is in the correct class. As the number of training images is 400, the maximum number of features derived by PCA is 399. The Mahalanobis distance measure performs the best, followed in order by the L_1 distance measure, the L_2 distance measure, and the cosine similarity measure. The reason for such an ordering is that the Mahalanobis distance measure counteracts the fact that L_1 and L_2 distance measures in the PCA space weigh preferentially for low frequencies. As the L_2 distance measure weighs more the low frequencies than L_1 does, the L_1 distance measure should perform better than the

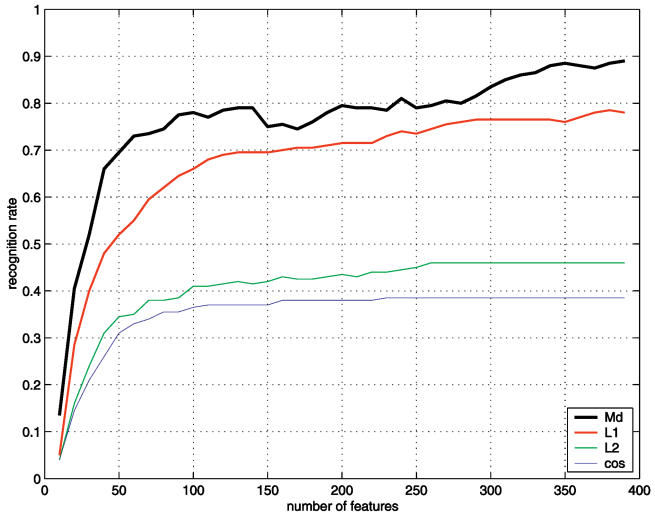

Fig. 2. Face recognition performance of the PCA method using four different similarity measures: Md (the Mahalanobis distance measure), L1 (the L_1 distance measure), L2 (the L_2 distance measure), and cos (the cosine similarity measure). Note that, when using fewer than 200 features, the best performance of PCA is below 80 percent.

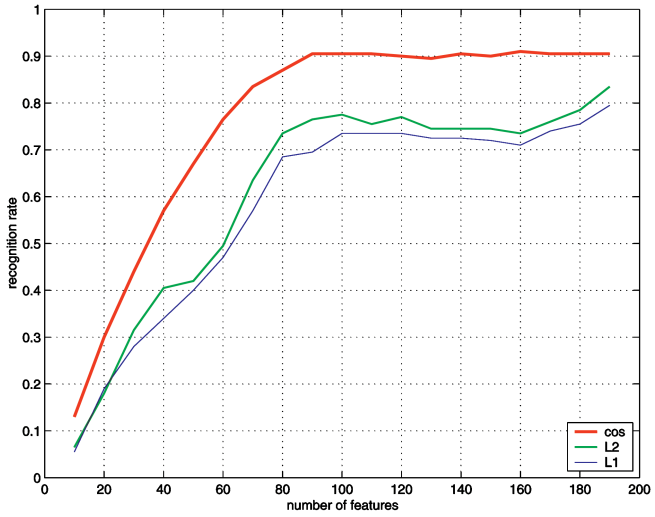

Fig. 3. Face recognition performance of the GDA method with a polynomial kernel function of degree one using three different similarity measures: cos (the cosine similarity measure), L2 (the L_2 distance measure), and L1 (the L_1 distance measure).

L_2 distance measure, a conjecture validated by our experiments. The cosine similarity measure does not compensate for the low frequency preference and it performs the worst among all the measures. Note that, when using fewer than 200 features, the best performance of PCA is still below 80 percent. The experimental results provide a baseline face recognition performance for the following comparative studies of the kernel Fisher methods, such as the GDA method [2], the GDA method with fractional power polynomial models, and the KFA method.

The second set of experiments assesses face recognition performance of the GDA method [2] with a polynomial kernel function, $k(\mathbf{x},\mathbf{y}) = (\mathbf{x}\cdot\mathbf{y})^d, d \in \mathbb{N}$. Fig. 3 shows the face recognition performance of the GDA method with a polynomial kernel of degree one, i.e., $d = 1$, using three different similarity measures: the cosine similarity measure, the L_2 distance measure, and the L_1 distance measure. The

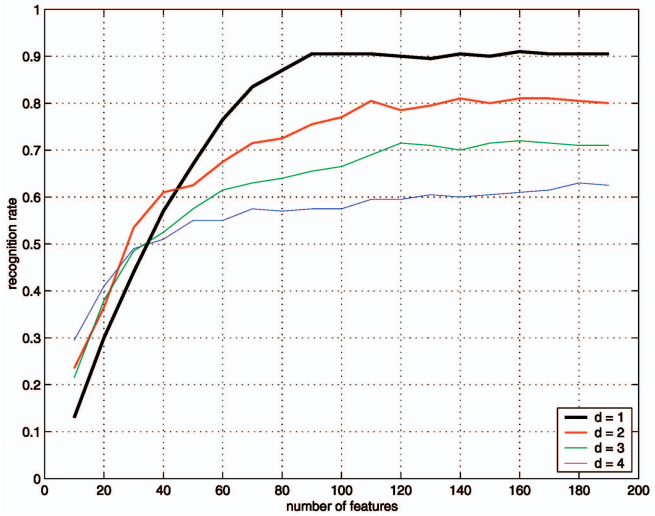

Fig. 4. Face recognition performance of the GDA method with a polynomial kernel function of four different degrees using the cosine similarity measure: $d = 1$, $d = 2$, $d = 3$, and $d = 4$.

cosine similarity measure performs the best, followed in order by the L_2 distance measure and the L_1 distance measure. We also implemented the GDA method with higher degree polynomial kernels, $d = 2, 3, 4, 5$, and the experimental results show that the cosine similarity measure consistently performs better than the L_2 and L_1 distance measures. These experiments suggest that one should use the cosine similarity measure for the following comparative assessment of the GDA method.

The third set of experiments evaluates the GDA method with a polynomial kernel function of four different degrees: $d = 1$, $d = 2$, $d = 3$, and $d = 4$. Fig. 4 shows the face recognition performance of the GDA method using the cosine similarity measure. The first order polynomial ($d = 1$) GDA method performs the best, followed in order by the second order polynomial ($d = 2$) GDA method, the third order polynomial ($d = 3$) GDA method, and the fourth order polynomial ($d = 4$) GDA method. Fig. 4 also shows that, among the four different degrees of polynomial kernels, the lower the degree is the better the GDA method performs. It thus seems natural that the GDA method should be extended to include some fractional power polynomials, whose degrees are even lower than one, i.e., $0 < d < 1$, in order to achieve better face recognition performance. Comparing Fig. 2 and Fig. 4, one can see that the GDA method with a polynomial kernel of degree one improves upon the PCA method for face recognition.

The fourth set of experiments assesses face recognition performance of the KFA method with a polynomial kernel function. In particular, Fig. 5 shows the face recognition performance of the KFA method with a polynomial kernel function of degree one, i.e., $d = 1$, using three different similarity measures: the cosine similarity measure, the L_2 distance measure, and the L_1 distance measure. Again, the cosine similarity measure performs the best, followed in order by the L_2 distance measure and the L_1 distance measure. These experiments suggest that one should use the cosine similarity measure for the following comparative assessment of the KFA method.

The fifth set of experiments evaluates the KFA method with a polynomial kernel function of four different degrees:

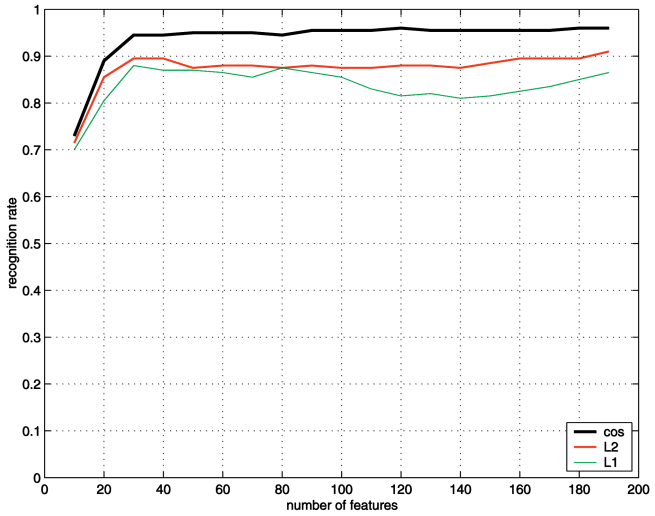

Fig. 5. Face recognition performance of the KFA method with a polynomial kernel function of degree one using three different similarity measures: cos (the cosine similarity measure), L2 (the L_2 distance measure), and L1 (the L_1 distance measure).

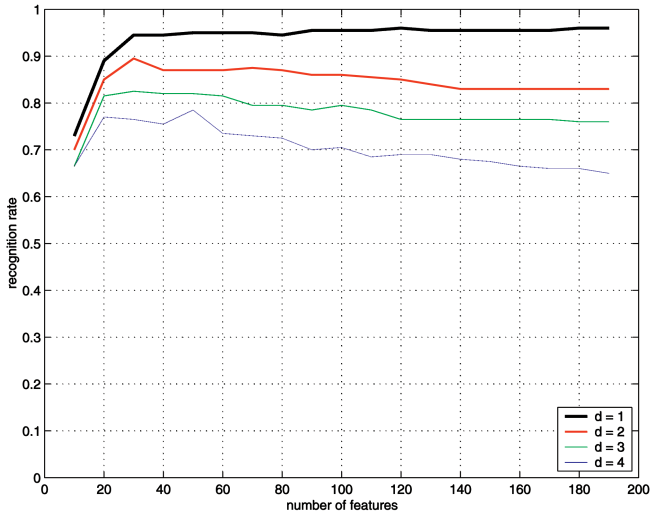

Fig. 6. Face recognition performance of the KFA method with a polynomial kernel function of four different degrees using the cosine similarity measure: $d = 1$, $d = 2$, $d = 3$, and $d = 4$.

$d = 1$, $d = 2$, $d = 3$, and $d = 4$. Fig. 6 shows the face recognition performance of the KFA method using the cosine similarity measure. The first order polynomial ($d = 1$) KFA method performs the best, followed in order by the second order polynomial ($d = 2$) KFA method, the third order polynomial ($d = 3$) KFA method, and the fourth order polynomial ($d = 4$) KFA method. Comparing Fig. 2, Fig. 4, and Fig. 6, one can see that the KFA method performs better than both the GDA method and the PCA method for face recognition. Again, Fig. 6 shows that, among the four different degrees of the polynomial kernel function, the lower the degree goes, the better performance the KFA method achieves. This finding suggests that fractional power polynomial models should be applied to further improve face recognition performance of the KFA method.

The sixth set of experiments assesses face recognition performance of the GDA method [2], the GDA method with fractional power polynomial models, and the KFA method

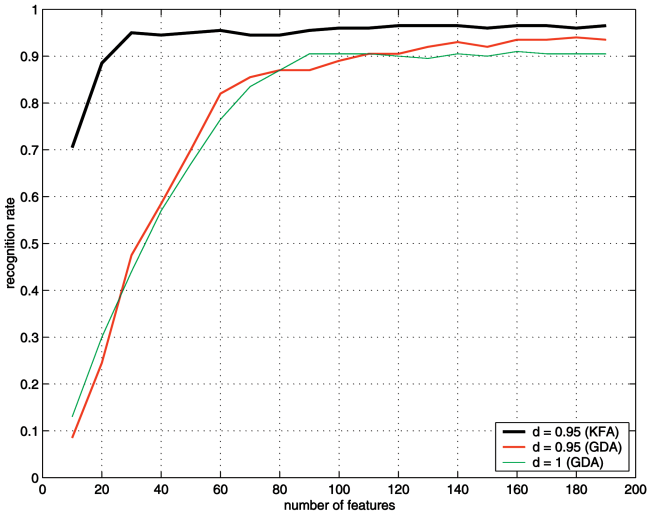

Fig. 7. Face recognition performance of the KFA method with a fractional power polynomial model of degree 0.95, the GDA method with a fractional power polynomial model of degree 0.95, and the GDA method with a polynomial kernel function of degree one.

with fractional power polynomial models, $k(\mathbf{x}, \mathbf{y}) = sign(\mathbf{x} \cdot \mathbf{y})\ (abs(\mathbf{x} \cdot \mathbf{y}))^d, 0 < d < 1$. Fig. 7 shows the face recognition performance of the KFA method with a fractional power polynomial model ($d = 0.95$), the GDA method with a fractional power polynomial model ($d = 0.95$), and the GDA method with a polynomial kernel function of degree one. Note that the degree of the fractional power polynomial model is empirically chosen. From Fig. 2, Fig. 4, Fig. 6, and Fig. 7, one can see that 1) the KFA method performs better than the GDA method, 2) both the KFA method and the GDA method improve their face recognition performance using a fractional power polynomial model, and 3) both the KFA method and the GDA method achieve better face recognition performance than the PCA method.

The experiments carried out so far using the FERET data set assess only two components of the proposed pattern recognition framework, namely, the KFA method and the fractional power polynomial models. The overall framework, which capitalizes on the dimensionality increasing techniques, such as the Gabor image representation, the KFA method, and fractional power polynomial models, will be further evaluated using the FRGC Version 1 and Version 2 databases [33]. In particular, two data sets, Data Set 1 and Data Set 2, used in our experiments come from the FRGC Version 1 and Version 2 databases, respectively. Data Set 1 contains 366 training images, 152 gallery images, and 608 probe images. These images are JPEG color images with spatial resolution of either $2,272 \times 1,704$ or $1,704 \times 2,272$. The training images are taken from 123 subjects, while the gallery and probe images are from 152 subjects that are not seen during training. The gallery images are controlled and the faces in these images have good image resolution and good illumination. The probe images, however, are uncontrolled and the faces in them have lower image resolution, larger illumination variations, and sometimes are blurring. It is these uncontrolled factors that pose grand challenges to face recognition performance. Note that the FRGC Version 1 database contains more images in the target set (943 controlled

TABLE 1
Number of Images and Image Quality (Controlled, Uncontrolled) of Training, Gallery, Probe, Target, and Query Sets of Different Experiments Corresponding to the Two Data Sets

Data	Experiment	Set	# Images	Image Quality
Data Set 1 (from FRGC Version 1)	Experiment 4	Training	366	Controlled & Uncontrolled
		Gallery	152	Controlled
		Probe	608	Uncontrolled
Data Set 2 (from FRGC Version 2)	Experiment 1	Training	12,776	Controlled & Uncontrolled
		Target	16,028	Controlled
		Query	16,028	Controlled
	Experiment 4	Training	12,776	Controlled & Uncontrolled
		Target	16,028	Controlled
		Query	8,014	Uncontrolled

images) and the query set (943 uncontrolled images). The reason of choosing the smaller gallery and probe data sets is to derive the performance curve of face recognition rate versus rank, also known as the Cumulative Match Curve (CMC). Face recognition rate is a widely used measure for evaluating face recognition performance. The more general performance curve, the Receiver Operating Characteristic (ROC) curve, which plots the Face Verification Rate (FVR) versus the False Accept Rate (FAR), is generated using the much larger FRGC Version 2 database. Data Set 2 from the FRGC Version 2 database contains 12,776 training images, 16,028 controlled target images, 16,028 controlled query images for Experiment 1, and 8,014 uncontrolled query images for Experiment 4. Note that due to physical memory limitation, only half of the 12,776 training images available in the FRGC Version 2 database are used for our experiments. For Experiment 1, the target set and the query set are identical. Table 1 shows the number of images and image quality (controlled, uncontrolled) of the training, gallery, probe, target, and query sets of different experiments.

The FRGC control algorithm, known as the BEE baseline algorithm, is a PCA algorithm that has been optimized for large scale problems [33]. In particular, the BEE baseline algorithm applies the whiten cosine distance measure for its nearest neighbor classifier [33]. The BEE baseline algorithm shows that Experiment 4, which is designed for indoor controlled single still image versus uncontrolled single still image, is the most challenging FRGC experiment. In addition to the BEE baseline algorithm, we implement an LDA-based algorithm as another baseline algorithm for comparatively assessing the performance of the proposed pattern recognition framework. The LDA-based baseline algorithm, similar to the Fisherfaces method [3], applies PCA first for dimensionality reduction and LDA second for discriminant analysis.

Image normalization first aligns the centers of the eyes to predefined locations and fixed interocular distance. In particular, the centers of the eyes are provided by the FRGC metadata and the predefined locations in the 128×128 images are (34, 29) and (34, 99). Then, a subimage procedure crops the face image to the size of 128×128 to extract the facial region. The facial region thus contains only face and the performance of face recognition is not affected by the factors not related to face, such as hair styles. The 128×128 face images undergo some traditional image processing operations, among which are image filtering for getting rid of noise and histogram processing for illumination enhancement. Common image filtering techniques include Gaussian filtering, edge-preserving smoothing, and popular histogram processing methods include histogram equalization, adaptive histogram equalization, and block histogram equalization [14]. Color image processing assesses color information for improving pattern recognition performance and different color space displays different pattern classification capabilities [39]. The optimal color space can be derived using a stochastic search method similar to the evolutionary pursuit method discussed in [25].

Gabor image representation further addresses image variabilities caused by illumination, facial expression, etc. The Gabor image representation of an image is the convolution of the image with a family of Gabor wavelets [27], [24]. This representation, characterized by spatial frequency, spatial locality, and orientational selectivity, is effective in coping with image variabilities. In particular, the Gabor filters are defined as 128×128 images matching the resolution of the face images. Due to the size of these filters, the convolution operation is implemented in the frequency domain for better computational efficiency. The 40 Gabor filters correspond to five different scales and eight different orientations with the following parameters: the spacing factor between filters in the frequency domain $\sqrt{2}$, the standard deviation 2π, and the maximum frequency $\pi/2$.

Fig. 8 shows some example FRGC face images used in our experiments. These images are cropped to the size of 128×128 to extract the facial region. The top row in Fig. 8 shows some example FRGC training face images: The first two images are controlled, while the remaining two images are uncontrolled. The bottom two rows in Fig. 8 display some example FRGC gallery/target and probe/query face images:

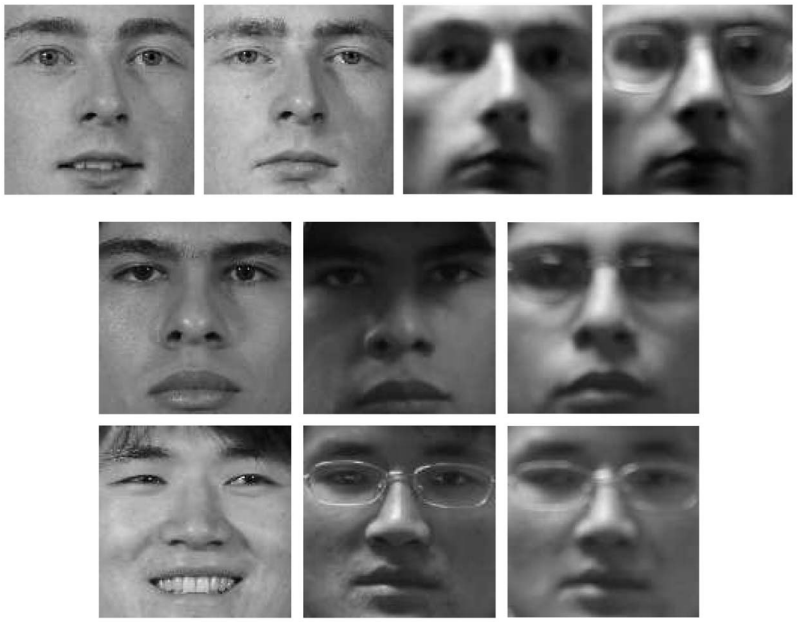

Fig. 8. Example FRGC face images used in our experiments. The images are cropped to the size of 128×128 to extract the facial region. The top row shows training images: The first two images are controlled and the remaining two images are uncontrolled. The bottom two rows display gallery/target and probe/query images: The two images in the first column are controlled gallery/target images and the remaining four images are uncontrolled probe/query images.

the two images in the first column are gallery/target images, which are controlled; the remaining four images are probe/query images, which are uncontrolled. The probe/query images display large image variabilities, such as illumination variations, low spatial resolutions, different facial expressions, glasses, and sometimes blurring.

We first assess the face recognition performance of the proposed pattern recognition framework using Data Set 1 from the FRGC Version 1 database on FRGC Experiment 4. The control algorithms are the BEE baseline algorithm and an LDA-based baseline algorithm. Note that the BEE baseline algorithm works on 150×130 masked images, while the LDA-based baseline algorithm and the proposed pattern recognition framework use the 128×128 images, as shown in Fig. 8. Fig. 9 shows the face recognition performance of the proposed method and the two baseline algorithms for FRGC Experiment 4. The horizontal axis represents the rank and the vertical axis represents the cumulative match score corresponding to the rank. As a result, the curves in Fig. 9 are also known as the Cumulative Match Curves (CMCs). The top curve represents the face recognition performance of the proposed pattern recognition framework using a fractional power polynomial model of degree 0.8 (see (20)). Note that the degree of the fractional power polynomial model is empirically chosen. The middle curve is the face recognition performance of the LDA-based algorithm, which first applies PCA to reduce the dimensionality to 180 and then LDA for feature extraction. The lower curve corresponds to the face recognition performance of the BEE baseline algorithm. Fig. 9 shows that the proposed framework with a fractional power polynomial model of degree 0.8 improves the FRGC Experiment 4 face recognition performance upon the LDA-based baseline algorithm and the BEE baseline algorithm. In particular, the proposed method achieves the rank one face recognition rate (face recognition rate of top response being correct) of 78 percent, compared to the LDA-based baseline algorithm rank one rate of 48 percent and the BEE baseline rank one rate of 37 percent.

We then assess the face recognition performance of the proposed pattern recognition framework using Data Set 2 from the FRGC Version 2 database on both FRGC Experiment 1 and Experiment 4. Experiment 1 is designed to measure face recognition performance from controlled

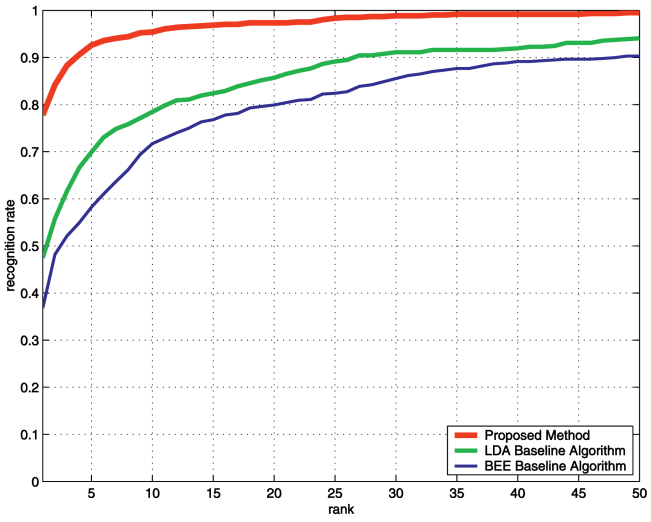

Fig. 9. Face recognition performance, the CMC curves, of the BEE baseline algorithm, the LDA-based baseline algorithm, and the proposed pattern recognition framework on FRGC Experiment 4 using Data Set 1 from the FRGC Version 1 database. The proposed framework uses a fractional power polynomial model of degree 0.8.

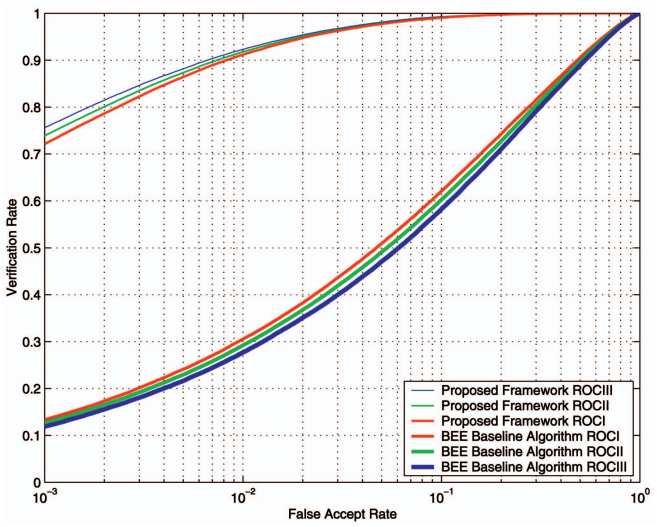

Fig. 10. Face recognition performance, the ROC curves, of the BEE baseline algorithm and the proposed pattern recognition framework on Experiment 4 using the FRGC Version 2 database. The proposed framework uses a fractional power polynomial model of degree 0.8.

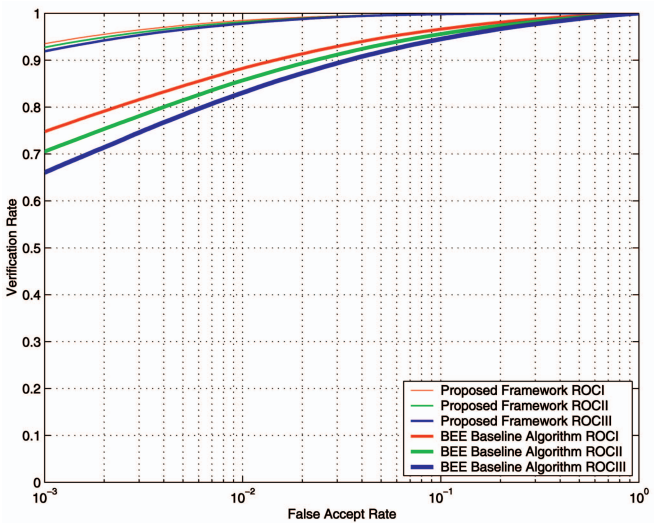

Fig. 11. Face recognition performance, the ROC curves, of the BEE baseline algorithm and the proposed pattern recognition framework on Experiment 1 using the FRGC Version 2 database. The proposed framework uses a fractional power polynomial model of degree 0.8.

frontal still images, while Experiment 4 measures performance from uncontrolled frontal still images. Thus, Experiment 4 is the most challenging FRGC experiment as it considers face recognition in uncontrolled illumination. Due to the physical memory limitation of our computers, we use only half of the total training images, i.e., 6,388 images for training. For Experiment 1, the target set and the query set are identical and each contains 16,028 controlled images. For Experiment 4, while the target set consists of 16,028 controlled images, the query set has 8,014 uncontrolled images. The ROC curve, which plots the Face Verification Rate (FVR) versus the False Accept Rate (FAR), is generated using the BEE system. The BEE system, which provides a fair and comprehensive way of conducting experimentation, 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.

Fig. 10 shows the ROC curves of the BEE baseline algorithm and the proposed pattern recognition framework on Experiment 4 using the FRGC Version 2 database. Note that the proposed framework uses a fractional power polynomial model of degree 0.8. The horizontal axis represents the false accept rate or FAR, while the vertical axis corresponds to the face verification rate or FVR. The ROC III curves in Fig. 10 show that the proposed pattern recognition framework achieves the face verification rate of 76 percent at the false accept rate of 0.1 percent compared to the face verification rate of 12 percent at the same false accept rate of the BEE baseline algorithm.[3]

To further test the generalization performance of the proposed pattern recognition framework, we straightforwardly apply the target features derived from Experiment 4 to Experiment 1, without any retraining process. Note that, for Experiment 1, the target set and the query set are identical and each contains 16,028 controlled images. Fig. 11 shows the ROC curves of the BEE baseline algorithm and the proposed pattern recognition framework on Experiment 1 using the FRGC Version 2 database. The ROC III curves in Fig. 11 show that the proposed pattern recognition framework achieves the face verification rate of 92 percent at the false accept rate of 0.1 percent, compared to the face verification rate of 66 percent at the same false accept rate of the BEE baseline algorithm.

Table 2 summarizes the face recognition performance of the BEE baseline algorithm and the proposed pattern recognition framework using Data Set 1 and Data Set 2 on FRGC Experiment 1 and Experiment 4. The performance shows the rank one face recognition rate for Data Set 1 and the face verification rate at 0.1 percent false accept rate for Data Set 2 (ROC III). In particular, for Data Set 1 and the FRGC Experiment 4, the proposed method achieves the rank one face recognition rate of 78 percent, compared to the LDA-based baseline algorithm rank one rate of 48 percent and the BEE baseline rank one rate of 37percent, respectively. For Data Set 2 and FRGC Experiment 4, the proposed method achieves the face verification rate (ROC III) of 76 percent at the false accept rate of 0.1 percent, compared to the BEE baseline verification rate of 12 percent at the same false accept rate. Extending the FRGC Experiment 4 features to FRGC Experiment 1, the proposed method achieves the face verification rate (ROC III) of 92 percent at the false accept rate of 0.1 percent, compared to the BEE baseline verification rate of 66 percent at the same false accept rate.

5 Discussions

The proposed pattern recognition framework is implemented on a 3.2 GHz single processor PC, and the CPU time for running FRGC Experiment 4 using the FRGC Version 2

3. Our face recognition system based on the proposed pattern recognition framework achieved the best performance for FRGC Experiment 4—the most challenging FRGC experiment, at the Third Face Recognition Grand Challenge Workshop on 16 February 2005, McLean, Virginia (for details see http://www.bee-biometrics.org/files/presentations/FRGC-Feb05/).

TABLE 2
Face Recognition Performance of the BEE Baseline Algorithm and the Proposed Pattern Recognition Framework Using Data Set 1 and Data Set 2 on FRGC Experiment 1 and Experiment 4

Data	Experiment	Method	Performance
Data Set 1 (from FRGC Version 1)	Experiment 4	BEE Baseline	37% Rank One Rate
		LDA Baseline	48% Rank One Rate
		Proposed Method	78% Rank One Rate
Data Set 2 (from FRGC Version 2)	Experiment 4	BEE Baseline	12% FVR at 0.1% FAR
		Proposed Method	76% FVR at 0.1% FAR
	Experiment 1	BEE Baseline	66% FVR at 0.1% FAR
		Proposed Method	92% FVR at 0.1% FAR

The performance shows the rank one face recognition rate for Data Set 1 and the Face Verification Rate (FVR) at 0.1 percent False Accept Rate (FAR) for Data Set 2 (ROC III).

database is shown in Table 3. As images can be cropped in parallel, the preprocessing time can be cut in half if one uses a dual processor PC. As FRGC Experiment 1 directly applies the target features derived for Experiment 4, there is no additional CPU time cost for preprocessing, training, and feature extraction. There is only one process for Experiment 1: similarity matrix generation, which takes about 4 minutes.

Gabor image representation, one dimensionality increasing technique incorporated into the proposed framework, makes an essential contribution to the overall face recognition performance as it is able to cope with image variabilities such as illumination variations. To show the importance of Gabor image representation, we implement the KFA method with fractional power polynomial models using intensity images without any additional processing, such as the Gabor image representation, and derive the following performance: for Experiment 4, the face verification rates at the false accept rate of 0.1 percent are 31 percent, 30 percent, and 30 percent for ROC I, ROC II, and ROC III, respectively; for Experiment 1, the face verification rates at the false accept rate of 0.1 percent are 77 percent, 73 percent, and 68 percent for ROC I, ROC II, and ROC III, respectively.

TABLE 3
The CPU Time for Implementing FRGC Experiment 4 Using the FRGC Version 2 Database on a 3.2 GHz Single Processor PC

Process	CPU Time
Preprocessing (face cropping of 36,818 images)	4 hours 18 minutes
Training (6,388 training images)	20 minutes
Feature Extraction and Similarity Matrix Generation	42 minutes

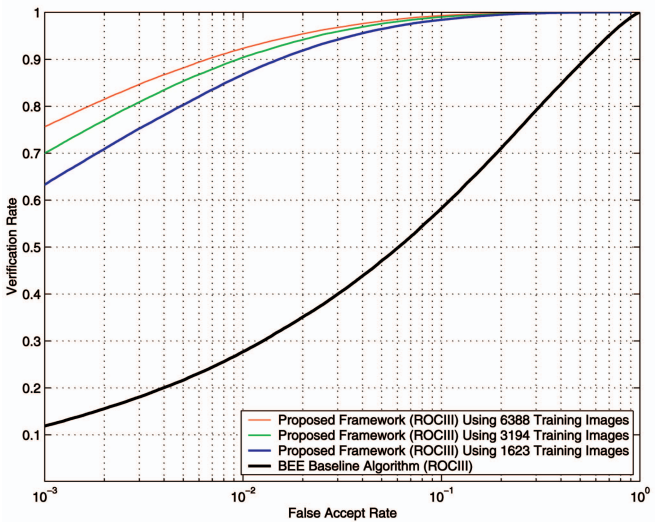

Fig. 12. Face recognition performance on Experiment 4 (the ROC III curves) of the proposed pattern recognition framework using 6,388, 3,194, and 1,623 FRGC Version 2 training images, respectively. The BEE baseline performance is included for comparison.

Another interesting issue is to study the performance of the proposed framework versus the number of training images. Toward that end, we generate three training sets by selecting half training images (round to integer numbers) for each subject from the preceding training set, starting from the original FRGC training set that contains 12,776 images. As a result, we derive three training sets consisting of 6,388, 3,194, and 1,623 images, respectively. The images in these three training sets all come from the FRGC Version 2 training set (12,776 images).[4] Fig. 12 shows the face recognition performance on Experiment 4 (the ROC III curves) of the proposed pattern recognition framework using 6,388, 3,194, and 1,623 FRGC Version 2 training images, respectively. Note that the BEE baseline performance is included for comparison. Corresponding to the three training sets with 6,388, 3,194, and 1,623 images, the face verification rates (ROC III) are 76 percent, 70 percent, 63 percent, respectively, at the false accept rate of 0.1 percent. This figure shows that the performance drops about 6 percent when the training set is reduced to half size. Fig. 13 shows the face recognition performance on Experiment 1 (the ROC III curves) of the proposed pattern recognition framework using 6,388, 3,194, and 1,623 FRGC Version 2 training images, respectively. Again, the BEE baseline performance is included for comparison. Corresponding to the three training sets with

4. The three training sets, TrainList1, TrainList2, and TrainList3, are supplemental materials which can be found at http://computer.org/tpami/archives.htm.

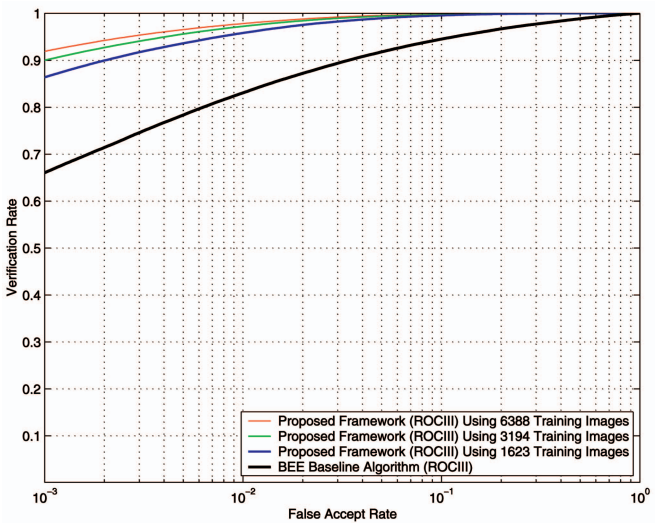

Fig. 13. Face recognition performance on Experiment 1 (the ROC III curves) of the proposed pattern recognition framework using 6,388, 3,194, and 1,623 FRGC Version 2 training images, respectively. The BEE baseline performance is included for comparison.

6,388, 3,194, and 1,623 images, the face verification rates (ROC III) are 92 percent, 90 percent, 86 percent, respectively, at the false accept rate of 0.1 percent. This figure shows that the performance drops about 3 percent when the training set is reduced to half size.

Acknowledgments

The author would like to thank the anonymous reviewers for their critical and constructive comments and suggestions. This work was partially supported by the TSWG R&D Contract N41756-03-C-4026.

References

[1] E. Anderson et al., *LAPACK Users' Guide,* third ed., SIAM, 1999.
[2] G. Baudat and F. Anouar, "Generalized Discriminant Analysis Using a Kernel Approach," *Neural Computation,* vol. 12, no. 10, pp. 2385-2404, 2000.
[3] P.N. Belhumeur, J.P. Hespanha, and D.J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 19, no. 7, pp. 711-720, July 1997.
[4] K.W. Bowyer, K. Chang, and P. Flynn, "A Survey of Approaches to Three-Dimensional Face Recognition," *Proc. 17th Int'l Conf. Pattern Recognition,* pp. 358-361, 2004.
[5] T. Cooke, "Two Variations on Fisher's Linear Discriminant for Pattern Recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 24, no. 2, pp. 268-273, Feb. 2002.
[6] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-Based Learning Method.* Cambridge Univ. Press, 2000.
[7] J. Daugman, "Face and Gesture Recognition: Overview," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 19, no. 7, pp. 675-676, July 1997.
[8] J.G. Daugman, "Two-Dimensional Spectral Analysis of Cortical Receptive Field Profiles," *Vision Research,* vol. 20, pp. 847-856, 1980.
[9] J.G. Daugman, "Uncertainty Relation for Resolution in Space, Spatial Frequency, and Orientation Optimized by Two-Dimensional Cortical Filters," *J. Optical Soc. Am.,* vol. 2, no. 7, pp. 1160-1169, 1985.
[10] G. Donato, M.S. Bartlett, J.C. Hager, P. Ekman, and T.J. Sejnowski, "Classifying Facial Actions," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 21, no. 10, pp. 974-989, Oct. 1999.
[11] K. Etemad and R. Chellappa, "Discriminant Analysis for Recognition of Human Face Images," *J. Optical Soc. Am. A,* vol. 14, pp. 1724-1733, 1997.
[12] K. Fukunaga, *Introduction to Statistical Pattern Recognition,* second ed. Academic Press 1990.
[13] A.S. Georghiades, P.N. Belhumeur, and D.J. Kriegman, "From Few to Many: Illumination Cone Models for Face Recognition under Variable Lighting and Pose," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 23, no. 6, pp. 643-660, June 2001.
[14] R.C. Gonzalez and R.E. Woods, *Digital Image Processing.* Prentice Hall, 2001.
[15] S. Haykin, *Neural Networks—A Comprehensive Foundation.* Macmillan College Publishing Company, 1994.
[16] S. Haykin, *Neural Networks—A Comprehensive Foundation,* second ed. Prentice Hall, 1999.
[17] P. Howland and H. Park, "Generalizing Discriminant Analysis Using the Generalized Singular Value Decomposition," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 26, no. 8, pp. 995-1006, Aug. 2004.
[18] A.K. Jain, S. Pankanti, S. Prabhakar, L. Hong, and A. Ross, "Biometrics: A Grand Challenge," *Proc. 17th Int'l Conf. Pattern Recognition,* pp. 935-942, 2004.
[19] M. Kirby and L. Sirovich, "Application of the Karhunen-Loeve Procedure for the Characterization of Human Faces," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 12, no. 1, pp. 103-108, Jan. 1990.
[20] M. Lades, J.C. Vorbruggen, J. Buhmann, J. Lange, C. von der Malsburg, R.P. Wurtz, and W. Konen, "Distortion Invariant Object Recognition in the Dynamic Link Architecture," *IEEE Trans. Computers,* vol. 42, no. 3, pp. 300-311, Mar. 1993.
[21] Z. Liang and P. Shi, "Uncorrelated Discriminant Vectors Using a Kernel Method," *Pattern Recognition,* vol. 38, no. 2, pp. 307-310, 2005.
[22] C. Liu, "A Bayesian Discriminating Features Method for Face Detection," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 25, no. 6, pp. 725-740, June 2003.
[23] C. Liu, "Enhanced Independent Component Analysis and Its Application to Content Based Face Image Retrieval," *IEEE Trans. Systems, Man, and Cybernetics, Part B: Cybernetics,* vol. 34, no. 2, pp. 1117-1127, 2004.
[24] C. Liu, "Gabor-Based Kernel PCA with Fractional Power Polynomial Models for Face Recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 26, no. 5, pp. 572-581, May 2004.
[25] C. Liu and H. Wechsler, "Evolutionary Pursuit and Its Application to Face Recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 22, no. 6, pp. 570-582, June 2000.
[26] C. Liu and H. Wechsler, "Robust Coding Schemes for Indexing and Retrieval from Large Face Databases," *IEEE Trans. Image Processing,* vol. 9, no. 1, pp. 132-137, 2000.
[27] C. Liu and H. Wechsler, "Gabor Feature Based Classification Using the Enhanced Fisher Linear Discriminant Model for Face Recognition," *IEEE Trans. Image Processing,* vol. 11, no. 4, pp. 467-476, 2002.
[28] M.J. Lyons, J. Budynek, A. Plante, and S. Akamatsu, "Classifying Facial Attributes Using a 2-D Gabor Wavelet Representation and Discriminant Analysis," *Proc. Fourth IEEE Int'l Conf. Automatic Face and Gesture Recognition,* 2000.
[29] M.J. Lyons, J. Budynek, and S. Akamatsu, "Automatic Classification of Single Facial Images," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 21, no. 12, pp. 1357-1362, Dec. 1999.
[30] S. Mika, G. Ratsch, J. Weston, B. Scholkopf, and K.R. Miller, "Fisher Discriminant Analysis with Kernels," *Neural Networks for Signal Processing IX,* Y.H. Hu, J. Larsen, E. Wilson, and S. Douglas, eds., pp. 41-48, 1999.
[31] S. Pankanti, R.M. Bolle, and A. Jain, "Guest Editors' Introduction: Biometrics—The Future of Identification," *Computer,* vol. 33, no. 2, pp. 46-49, Feb. 2000.
[32] C.H. Park and H. Park, "Nonlinear Discriminant Analysis Using Kernel Functions and the Generalized Singular Value Decomposition," *SIAM J. Matrix Analysis and Applications,* vol. 27, no. 1, pp. 87-102, 2005.
[33] P.J. Phillips, P.J. Flynn, T. Scruggs, K.W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, "Overview of the Face Recognition Grand Challenge," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* 2005.
[34] P.J. Phillips, "Matching Pursuit Filters Applied to Face Identification," *IEEE Trans. Image Processing,* vol. 7, no. 8, pp. 1150-1164, 1998.

[35] P.J. Phillips, P. Grother, R.J. Micheals, D.M. Blackburn, E. Tabassi, and J.M. Bone, "FRVT 2002: Evaluation Report," technical report, http://www.frvt.org/, Mar. 2003.
[36] P.J. Phillips and E.M. Newton, "Meta-Analysis of Face Recognition Algorithms," *Proc. Fifth Int'l Conf. Automatic Face and Gesture Recognition,* pp. 224-230, Mar. 2002.
[37] P.J. Phillips, H. Wechsler, J. Huang, and P. Rauss, "The FERET Database and Evaluation Procedure for Face-Recognition Algorithms," *Image and Vision Computing,* vol. 16, pp. 295-306, 1998.
[38] B. Scholkopf and A. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization and Beyond*. MIT Press, 2002.
[39] P. Shih and C. Liu, "Comparative Assessment of Content-Based Face Image Retrieval in Different Color Spaces," *Int'l J. Pattern Recognition and Artificial Intelligence,* vol. 19, no. 7, pp. 873-893, 2005.
[40] D.L. Swets and J. Weng, "Using Discriminant Eigenfeatures for Image Retrieval," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 18, no. 8, pp. 831-836, Aug. 1996.
[41] M. Turk and A. Pentland, "Eigenfaces for Recognition," *J. Cognitive Neuroscience,* vol. 13, no. 1, pp. 71-86, 1991.
[42] Y.N. Vapnik, *The Nature of Statistical Learning Theory,* second ed. Springer-Verlag, 1999.
[43] M.A.O. Vasilescu and D. Terzopoulos, "Multilinear Analysis of Image Ensembles: Tensorfaces," *Proc. European Conf. Computer Vision,* 2002.
[44] L. Wiskott, J.M. Fellous, N. Kruger, and C. von der Malsburg, "Face Recognition by Elastic Bunch Graph Matching," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 19, no. 7, pp. 775-779, July 1997.
[45] J. Yang, F.F. Alejandro, J. Yang, D. Zhang, and Z. Jin, "KPCA Plus LDA: A Complete Kernel Fisher Discriminant Framework for Feature Extraction and Recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 27, no. 2, pp. 230-244, Feb. 2005.
[46] M.H. Yang, "Kernel eigenfaces vs. Kernel Fisherfaces: Face Recognition Using Kernel Methods," *Proc. Fifth Int'l Conf. Automatic Face and Gesture Recognition,* May 2002.
[47] J. Ye, R. Janardan, C.H. Park, and H. Park, "An Optimization Criterion for Generalized Discriminant Analysis on Undersampled Problems," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 26, no. 8, pp. 982-994, Aug. 2004.
[48] F. Zhang, *Matrix Theory: Basic Results and Techniques.* New York: Springer, 1999.
[49] W. Zhao, R. Chellappa, J. Phillips, and A. Rosenfeld, "Face Recognition: A Literature Survey," *ACM Computing Surveys,* vol. 35, no. 4, pp. 399-458, 2003.
[50] W. Zheng, L. Zhao, and C. Zou, "A Modified Algorithm for Generalized Discriminant Analysis," *Neural Computation,* vol. 16, no. 6, pp. 1283-1297, 2004.

Chengjun Liu received the PhD degree from George Mason University in 1999, and he is presently an assistant professor of computer science at the New Jersey Institute of Technology. His research interests are in computer vision, pattern recognition, image processing, evolutionary computation, and neural computation. 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 on 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.