Evolutionary Pursuit and Its Application to Face Recognition

Chengjun Liu, *Member*, *IEEE*, and Harry Wechsler, *Fellow*, *IEEE*

Abstract—This paper introduces Evolutionary Pursuit (EP) as a novel and adaptive representation method for image encoding and classification. In analogy to projection pursuit methods, EP seeks to learn an optimal basis for the dual purpose of data compression and pattern classification. The challenge for EP is to increase the generalization ability of the learning machine as a result of seeking the trade-off between minimizing the empirical risk encountered during training and narrowing the confidence interval for reducing the guaranteed risk during future testing on unseen images. Toward that end, EP implements strategies characteristic of genetic algorithms (GAs) for searching the space of possible solutions to determine the optimal basis. EP starts by projecting the original data into a lower dimensional whitened Principal Component Analysis (PCA) space. Directed but random rotations of the basis vectors in this space are then searched by GAs where evolution is driven by a fitness function defined in terms of performance accuracy ("empirical risk") and class separation ("confidence interval"). Accuracy indicates the extent to which learning has been successful so far, while separation gives an indication of the expected fitness on future trials. The feasibility of the new method has been successfully tested on face recognition where the large number of possible bases requires some type of greedy search algorithm. The particular face recognition task involves 1,107 FERET frontal face images corresponding to 369 subjects. To assess both accuracy and generalization capability, the data includes for each subject images acquired at different times or under different illumination conditions. The results reported show that EP improves on face recognition performance when compared against PCA ("Eigenfaces") and displays better generalization abilities than the Fisher linear discriminant ("Fisherfaces").

Index Terms—Evolutionary pursuit, face recognition, genetic algorithms, optimal basis, Principal Component Analysis (PCA), Fisher Linear Discriminant (FLD).

1 Introduction

Pattern recognition depends heavily on the particular choice of features used by the classifier. One usually starts with a given set of features and then attempts to derive an optimal subset of features leading to high classification performance with the expectation that similar performance will be displayed also on future trials using novel data. Standard methods approach this task by ranking the features according to some criteria such as ANOVA and/or information theory-based measures such as "infomax," and then deleting the lower ranked features. Ranking by itself is usually not enough because criteria like those listed above do not capture possible nonlinear interactions among the features, nor do they measure the ability for generalization. The process of feature selection should involve the derivation of salient features with the twin goals of reducing the amount of data used for classification and simultaneously providing enhanced discriminatory power. The search for such features is then driven by the need to increase the generalization ability of the learning machine as a result of seeking the trade-off between minimizing the empirical risk encountered during training and narrowing the confidence interval for reducing the guaranteed risk, while testing on unseen data [43]. The search for an optimal feature set amounts then to searching for an optimal basis where the feature values correspond to the projections taken along the basis axes.

The search for the best feature set—optimal basis—is analogue to finding an optimal neural code, referred to biologically as a lattice of receptive fields (RFs) ("kernels"). Several attempts have been made recently to computationally derive such an optimal neural code [31], [1], [37]. The search for such a code involves constrained optimization using design criteria, such as: 1) redundancy reduction, 2) minimization of the reconstruction error, 3) maximization of information transmission ("infomax") [24], and 4) sparseness or independence [31]. Furthermore, to the design criteria listed above one should add as an important criteria ("functionality") 5) successful pattern classification, referred to and used by Edelman [11] in the context of neural Darwinism. The search for such an optimal basis leads also to the class of Projection Pursuit (PP) methods as possible candidates for universal approximation. As an example, projection pursuit regression implements an additive model with univariate basis functions [14] [19].

The rationale behind feature extraction using an optimal basis representation is that most practical computational methods for both regression and classification use parameterization in the form of a linear combination of basis functions. This leads to a taxonomy based on the type of the basis functions used by a particular method and the corresponding optimization procedure used for parameter estimation. According to this taxonomy, most practical methods use basis function representation—those are called

- *C. Liu is with the Department of Mathematics and Computer Science, University of Missouri, St. Louis, MO 63121. E-mail: cliu@cs.umsl.edu.*
- *H. Wechsler is with the Department of Computer Science, George Mason University, Fairfax, VA 22030. E-mail: wechsler@cs.gmu.edu.*

Manuscript received 19 Nov. 1998; revised 21 July 1999; accepted 15 May 2000.
Recommended for acceptance by D. Kriegman.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number 108300.

dictionary or kernel methods, where the particular type of chosen basis functions constitutes a kernel.

Since most practical methods use nonlinear models, the determination of optimal kernels becomes a nonlinear optimization problem. When the objective function lacks an analytical form suitable for gradient descent or the computation involved is prohibitively expensive, one should use (directed) random search techniques for nonlinear optimization and variable selection similar to evolutionary computation and genetic algorithms (GAs) [17]. GAs work by maintaining a constant-sized population of candidate solutions known as individuals ("chromosomes"). The power of a genetic algorithm lies in its ability to exploit, in a highly efficient manner, information about a large number of individuals. The search underlying GAs is such that breadth and depth—exploration and exploitation—are balanced according to the observed performance of the individuals evolved so far. By allocating more reproductive occurrences to above average individual solutions, the overall effect is to increase the population's average fitness. We advance in this paper a novel and adaptive representation method and, in analogy to the pursuit methods referred to earlier, our novel method is called Evolutionary Pursuit (EP). As successful face recognition methodology depends heavily on the particular choice of the features used by the (pattern) classifier [5], [38], [3], and as the size of the original face space is very large to start with, we chose to assess the feasibility of EP using face recognition benchmark studies [26].

The outline for this paper is as follows: Section 2 provides general background on representation and discrimination coding schemes—the Principal Component Analysis (PCA) and the Fisher Linear Discriminant (FLD)—and their use for face recognition in terms of Eigenfaces [42] and Fisherfaces [2], respectively. Section 3 describes the overall solution for the face recognition problem and specifically addresses the tasks of lowering the dimensionality of the search space and finding an optimal feature set for face classification. The actual search for the optimal and reduced feature set is addressed in Section 4, where we introduce the evolutionary pursuit method and its particular implementation for face recognition. Experimental results are reported in Section 5 and they show that EP improves on face recognition performance when compared against PCA ("Eigenfaces") and displays better generalization abilities than FLD ("Fisherfaces"). The last section reviews the characteristics and merits of EP and discusses its possible impact on building further connections between functional approximation using regularization and statistical learning theory when the concern is that of reducing the guaranteed risk while testing on unseen data.

2 Representation and Discrimination Coding Schemes

Efficient coding schemes for face recognition require both low-dimensional feature representations and enhanced discrimination abilities. As this paper addresses the twin problems of lowering space dimensionality and enhancing discrimination performance, we survey first Principal Component Analysis (PCA) [21] and Fisher Linear Discriminant (FLD) [13] as encoding methods. As our benchmark studies involve face recognition, the use of PCA and FLD for face recognition in terms of Eigenfaces [42] and Fisherfaces [2], respectively, is briefly discussed, as well.

2.1 Principal Component Analysis (PCA)

One popular technique for feature selection and dimensionality reduction is PCA [15], [10]. PCA is a standard decorrelation technique and following its application, one derives an orthogonal projection basis which directly leads to dimensionality reduction, and possibly to feature selection. Let $X \in \mathbb{R}^N$ be a random vector representing an image, where N is the dimensionality of the image space. The vector is formed by concatenating the rows or the columns of the image which may be normalized to have a unit norm. The covariance matrix of X is defined as follows:

$$\Sigma_X = E\{[X - E(X)][X - E(X)]^t\}, \tag{1}$$

where $E(\cdot)$ is the expectation operator, t denotes the transpose operation, and $\Sigma_X \in \mathbb{R}^{N\times N}$. The PCA of a random vector X factorizes the covariance matrix Σ_X into the following form:

$$\Sigma_X = \Phi\Lambda\Phi^t \; with \; \Phi = [\phi_1\phi_2\ldots\phi_N], \Lambda = diag\{\lambda_1, \lambda_2, \ldots, \lambda_N\}, \tag{2}$$

where $\Phi \in \mathbb{R}^{N\times N}$ is an orthonormal eigenvector matrix and $\Lambda \in \mathbb{R}^{N\times N}$ a diagonal eigenvalue matrix with diagonal elements in decreasing order ($\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N$). $\phi_1, \phi_2, \ldots, \phi_N$ and $\lambda_1, \lambda_2, \ldots, \lambda_N$ are the eigenvectors and the eigenvalues of Σ_X, respectively.

An important property of PCA is decorrelation, i.e., the components of the transformation, $X' = \Phi^t X$, are decorrelated since the covariance matrix of X' is diagonal, $\Sigma_{X'} = \Lambda$, and the diagonal elements are the variances of the corresponding components. Another property of PCA is its optimal signal reconstruction in the sense of minimum Mean Square Error (MSE) when only a subset of principal components where $P = [\phi_1\phi_2\ldots\phi_m], \quad m < N$ and $P \in \mathbb{R}^{N\times m}$ are used to represent the original signal. Following this property, an immediate application of PCA is the dimensionality reduction:

$$Y = P^t X. \tag{3}$$

The lower-dimensional vector $Y \in \mathbb{R}^m$ captures the most expressive features of the original data X. As PCA derives projection axes based on the observed variations using all the training samples, it enjoys good generalization abilities [32] for image reconstruction when tested with novel images not seen during training. The disadvantage of PCA is that it does not distinguish the different roles of the within- and the between-class variations and it treats them equally. This should lead to poor testing performance when the distributions of the face classes are not separated by the mean-difference but, instead, by the covariance-difference [15], [30], [27]. High variance by itself does not necessarily lead to good discrimination ability unless the corresponding distribution is multimodal and the modes correspond to the classes to be discriminated. One should

also be aware that as PCA encodes only for second order statistics, it lacks phase and, thus, locality information.

PCA was first applied to reconstruct human faces by Kirby and Sirovich [23]. They showed that any particular face can be economically represented along the eigenpictures coordinate space, and that any face can be approximately reconstructed by using just a small collection of eigenpictures and the corresponding projections ("coefficients"). Since eigenpictures are fairly good at representing face images, one could consider using the projections along them as classification features to recognize human faces. As a result, Turk and Pentland [42] developed a well-known face recognition method, known as Eigenfaces, where the eigenfaces correspond to the eigenvectors associated with the dominant eigenvalues of the face covariance matrix. The eigenfaces define a feature space, or "face space," which drastically reduces the dimensionality of the original space, and face detection and identification are carried out in the reduced space.

2.2 Fisher Linear Discriminant (FLD)

Another widely used technique for feature selection is the Fisher Linear Discriminant (FLD) [13]. Let $\omega_1, \omega_2, \ldots, \omega_L$ and $N_1, N_2, \ldots, N_L$ denote the classes and the number of images within each class, respectively. Let $M_1, M_2, \ldots, M_L$ and M be the means of the classes and the grand mean. The within- and between-class scatter matrices Σ_w and Σ_b are defined as follows:

$$\Sigma_w = \sum_{i=1}^{L} P(\omega_i)\Sigma_i = \sum_{i=1}^{L} P(\omega_i)E\{[X - M_i][X - M_i]^t|\omega_i\} \quad (4)$$

$$\Sigma_b = \sum_{i=1}^{L} P(\omega_i)(M_i - M)(M_i - M)^t, \quad (5)$$

where $P(\omega_i)$ and Σ_i are a priori probability and the covariance matrix of class ω_i, respectively, and $\Sigma_w, \Sigma_b \in \mathbb{R}^{N\times N}$.

FLD derives a projection matrix Ψ that maximizes the ratio $|\Psi^t\Sigma_b\Psi|/|\Psi^t\Sigma_w\Psi|$ [2]. This ratio is maximized when Ψ consists of the eigenvectors of the matrix $\Sigma_w^{-1}\Sigma_b$ [40]. The eigenvalue equation is defined as follows:

$$\Sigma_w^{-1}\Sigma_b\Psi = \Psi\Delta \quad (6)$$

where $\Psi, \Delta \in \mathbb{R}^{N\times N}$ are the eigenvector and eigenvalue matrices of $\Sigma_w^{-1}\Sigma_b$. Although both Σ_w and Σ_b are symmetric, $\Sigma_w^{-1}\Sigma_b$ is not necessarily symmetric. However, the eigenvalue and eigenvector matrices can be obtained as the result of simultaneous diagonalization of Σ_w and Σ_b [15].

FLD overcomes one of PCA's drawbacks as it can distinguish within- and between-class scatters. Furthermore, FLD induces nonorthogonal projection axes, a characteristic known to have great functional significance in biological sensory systems [9]. The drawback of FLD is that it requires large sample sizes for good generalization. One possible remedy for this drawback is to artificially generate additional data and, thus, increase the sample size [12].

FLD is behind several face recognition methods [40], [2], [12], [25]. As the original image space is high-dimensional, most methods first perform dimensionality reduction using PCA, as it is the case with the Fisherfaces method suggested by Belhumeur et al. [2]. Using similar arguments, Swets and Weng [40] point out that the Eigenfaces method only derives the Most Expressive Features (MEF). As explained earlier, such PCA inspired features do not necessarily provide for good discrimination. As a consequence, subsequent FLD projections are used to build the Most Discriminating Features (MDF) classification space. The MDF space is, however, superior to the MEF space for face recognition only when the training images are representative of the range of face (class) variations; otherwise, the performance difference between the MEF and MDF is not significant [40].

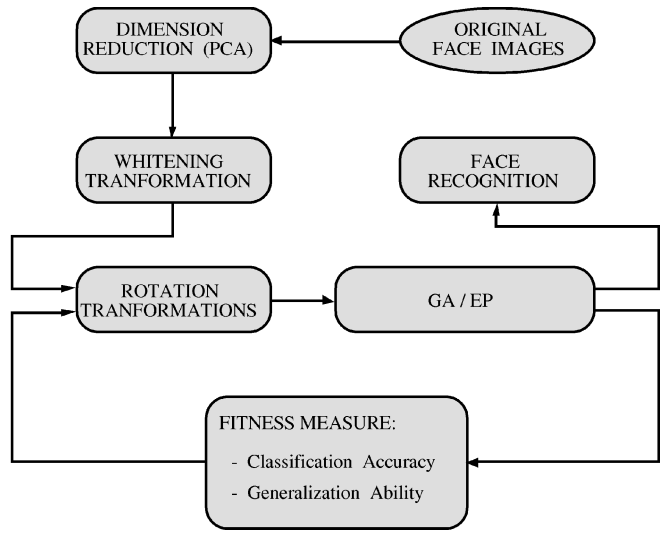

Fig. 1. Face recognition using evolutionary pursuit (EP). EP implements strategies characteristic of genetic algorithms for searching the space of possible solutions to determine the optimal basis.

3 Face Basis and Recognition

Our methodology for face recognition is shown in Fig. 1. The main thrust is to find out an optimal basis along which faces can be projected leading to a compact and efficient face encoding in terms of recognition ability. Towards that end, PCA first projects the face images into a lower-dimensional space. The next step is the whitening transformation and it counteracts the fact that the Mean-Square-Error (MSE) principle underlying PCA preferentially weights low frequencies. Directed, but random rotations of the lower-dimensional (whitened PCA) space for enhanced recognition performance are then driven by evolution and use domain specific knowledge ("fitness"). The fitness driving evolution in search of the optimal face basis considers both the recognition rates ("classification accuracy") and the scatter index ("generalization ability"). Evolution is implemented using Evolutionary Pursuit (EP) as a special form of Genetic Algorithms (GAs). Note that the reachable space of EP is increased as a result of using nonorthonormal whitening and a set of rotation transformations. One can expect better performance from nonorthogonal bases over orthogonal ones as nonorthogonality embodies a characteristic known to have great functional significance in biological sensory systems [9].

3.1 Whitening Transformation

After dimensionality reduction using PCA, the lower-dimensional feature set $Z \in \mathbb{R}^{m\times n}$ is derived as follows:

$$Z = [Y_1 Y_2 \ldots Y_n], \tag{7}$$

where n is the number of training samples and $Y_i, i = 1, 2, \ldots, n,$ comes from (3). Now, Z is subject to the whitening transformation and yields yet another feature set $V \in \mathbb{R}^{m\times n}$:

$$V = \Gamma Z, \tag{8}$$

where $\Gamma = diag\{\lambda_1^{-1/2}, \lambda_2^{-1/2}, \ldots, \lambda_m^{-1/2}\}$ and $\Gamma \in \mathbb{R}^{m\times m}$.

The reason why the whitening procedure can lead to nonorthogonal basis of the overall transformation is as follows. First, let $Q \in \mathbb{R}^{m\times m}$ be a rotation matrix ($Q^tQ = QQ^t = I$) and apply Q to the feature set V. Then, using (8), (7), and (3) one derives the overall transformation matrix, $\Xi \in \mathbb{R}^{N\times m}$, which combines three transformations (dimensionality reduction, whitening, and rotation):

$$\Xi = P\Gamma Q. \tag{9}$$

Now assume the basis vectors in Ξ are orthogonal (using *proof by contradiction*):

$$\Xi^t \Xi = \Delta, \tag{10}$$

where $\Delta \in \mathbb{R}^{m\times m}$ is a diagonal matrix. Using (9) and (10), one derives the following equation:

$$\Gamma^2 = \Delta = cI, \tag{11}$$

where c is a constant and $I \in \mathbb{R}^{m\times m}$ is a unit matrix. Equation (11) holds only when all the eigenvalues are equal, and when this is not the case, the basis vectors in Ξ are not orthogonal (see Fig. 7).

3.2 Rotation Transformations

The rotation transformations are carried out in the whitened m dimensional space, in which the feature set V lies (see (8)). Let $\Omega = [\varepsilon_1 \varepsilon_2 \ldots \varepsilon_m]$ be the basis of this space where $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_m$ are unit vectors and $\Omega \in \mathbb{R}^{m\times m}$. Our evolutionary pursuit approach would, later on, search for a (reduced) subset of some basis vectors rotated from $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_m$ in terms of best discrimination and generalization performance. The rotation procedure is carried out by pairwise axes rotations. In particular, if the basis vectors ε_i and ε_j are rotated by α_k, then a new basis $\xi_1, \xi_2, \ldots, \xi_m$ is derived as follows:

$$[\xi_1 \xi_2 \ldots \xi_m] = [\varepsilon_1 \varepsilon_2 \ldots \varepsilon_m] Q_k, \tag{12}$$

where $Q_k \in \mathbb{R}^{m\times m}$ is a rotation matrix. There are a total of $M = m(m-1)/2$ rotation angles corresponding to the M pairs of basis vectors to be rotated. For the purpose of evolving an optimal basis for (face) recognition, it makes no difference if the angles are confined to $(0, \pi/2)$, since the positive directions and the order of axes are irrelevant. The overall rotation matrix $Q \in \mathbb{R}^{m\times m}$ is defined as follows:

$$Q = Q_1 Q_2 \cdots Q_{m(m-1)/2}. \tag{13}$$

3.3 Face Recognition

Let $T = [\Theta_{i_1} \Theta_{i_2} \ldots \Theta_{i_l}]$, $T \in \mathbb{R}^{m\times l}$, be the optimal basis derived by the evolutionary pursuit method (see Section 4). The new feature set $U \in \mathbb{R}^{l\times n}$ is derived as follows:

$$U = [U_1 U_2 \ldots U_n] = T^t V, \tag{14}$$

where V is the whitened feature set from (8), and $U_i \in \mathbb{R}^l, i = 1, 2, \ldots, n,$ is the feature vector corresponding to the ith face image.

Let $U_k^0, k = 1, 2, \ldots, n,$ be the prototype for class k. The classification rule is then specified as follows:

$$\|U_i - U_k^0\|_2 = \min_j \|U_i - U_j^0\|_2, \quad U_i \in \omega_k. \tag{15}$$

The new face image U_i is classified to the class ω_k from which the Euclidean distance is minimum.

4 Evolutionary Pursuit (EP)

The task for EP is to search for a face basis through the rotated axes defined in a properly whitened PCA space. Evolution is driven by a fitness function defined in terms of performance accuracy and class separation (scatter index). Accuracy indicates the extent to which learning has been successful so far, while the scatter index gives an indication of the expected fitness on future trials. Together, the accuracy and the scatter index give an indication of the overall performance ability. In analogy to the statistical learning theory [43], the scatter index is the conceptual analog for the capacity of the classifier and its use is to prevent overfitting. By combining these two terms together (with proper weights), GA can evolve balanced results and yield good recognition performance and generalization abilities.

One should also point out that just using more principal components (PCs) does not necessarily lead to better performance, since some PCs might capture the within-class scatter which is unwanted for the purpose of recognition [25], [27]. In our experiments, we searched the 20- and 30-dimensional whitened PCA spaces corresponding to the leading eigenvalues, since it is in those spaces that most of the variations characteristic of human faces occur.

4.1 Chromosome Representation and Genetic Operators

As discussed in Section 3.2, different basis vectors are derived corresponding to different sets of rotation angles. GAs are used to search among the different rotation transformations and different combinations of basis vectors in order to find out the optimal subset of vectors ("face basis"), where optimality is defined with respect to classification accuracy and generalization ability. The optimal basis is evolved from a larger vector set $\{\xi_1, \xi_2, \ldots, \xi_m\}$ rotated from a basis $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_m$ in m-dimensional space by a set of rotation angles $\alpha_1, \alpha_2, \ldots, \alpha_{m(m-1)/2}$ with each angle in the range of $(0, \pi/2)$. If the angles are discretized with small enough steps, then we can use GAs to search this discretized space. GAs require the solutions to be represented in the form of bit strings or chromosomes. If we use 10 bits (resolution) to represent each angle, then each discretized

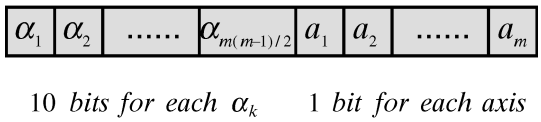

Fig. 2. Chromosome representation of rotation angles and projection axes. Ten bits are used for each angle, so that every discretized (angle) interval is less than 0.09 degrees, while one bit for each axis is indicating whether this axis is chosen as a basis vector.

(angle) interval is less than 0.09 degrees, and we need $10 * [m(m-1)/2]$ bits to represent all the angles. As we also have m basis vectors (projection axes) to choose from, another m bits should be added to the chromosome to facilitate that choice. Fig. 2 shows the chromosome representation, and $a_i, i = 1, 2, \ldots, m$, taking value of 0 or 1, indicates whether the ith vector is chosen as a basis vector.

Let N_s be the number of different choices of basis vectors in the search space. The size of the genospace, too large to search it exhaustively, is defined as follows:

$$N_s = 2^{5m(m-1)+m}. \tag{16}$$

As it searches the genospace, the GA makes its choices via genetic operators as a function of a probability distribution driven by the fitness function. The genetic operators are selection, crossover (or recombination), and mutation [17]. In our experiments, we use 1) proportionate selection: preselection of parents in proportion to their relative fitness; 2) two-point crossover: exchange the sections between the crossover points, as shown in Fig. 3; and 3) fixed probability mutation: each position of a chromosome is given a fixed probability of undergoing mutation (flipping the corresponding bit). Note that the crossover operator is not restricted to operate on an angle boundary, since any arbitrary position of a chromosome can be chosen as a crossover point.

4.2 The Fitness Function

Fitness values guide GAs on how to choose offspring for the next generation from the current parent generation. If $F \equiv \alpha_1, \alpha_2, \ldots, \alpha_{m(m-1)/2}; a_1, a_2, \ldots, a_m$ represents the parameters to be evolved by GA, then the fitness function $\zeta(F)$ is defined as follows:

$$\zeta(F) = \zeta_a(F) + \lambda\zeta_s(F), \tag{17}$$

where $\zeta_a(F)$ is the performance accuracy term, $\zeta_s(F)$ is the class separation ("generalization") term, and λ is a positive constant that determines the importance of the second term relative to the first one. In our experiments, we set $\zeta_a(F)$ to be the number of faces correctly recognized as the top choice after the rotation and selection of a subset of axes, and $\zeta_s(F)$ the scatter measurement among different classes. λ is empirically chosen such that $\zeta_a(F)$ contributes more to the fitness than $\zeta_s(F)$ does. Note that the fitness function is similar to the cost functional derived using regularization theory, a very useful tool for solving ill-posed problems in computer vision and improving the generalization ability of RBF networks [41], [36], [18]. The two terms, $\zeta_a(F)$ and $\zeta_s(F)$, put opposite pressures on the fitness function: the performance accuracy term $\zeta_a(F)$ tends to choose basis vectors which lead to small scatter, while the class separation term $\zeta_s(F)$ favors basis vectors which cause large scatter. By combining those two terms together (with proper weight λ), GA can evolve balanced results displaying good performance during both training and future test trials.

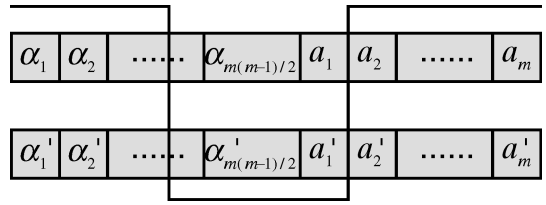

Fig. 3. Two-point crossover. The segments between the two crossover points are exchanged.

Let the rotation angle set be $\alpha_1^{(k)}, \alpha_2^{(k)}, \ldots, \alpha_{m(m-1)/2}^{(k)}$, and the basis vectors after the transformation be $\xi_1^{(k)}, \xi_2^{(k)}, \ldots, \xi_m^{(k)}$ according to (12) and (13). If GA chooses l vectors $\eta_1, \eta_2, \ldots, \eta_l$ from $\xi_1^{(k)}, \xi_2^{(k)}, \ldots, \xi_m^{(k)}$, then the new feature set, $W \in \mathbb{R}^{l \times n}$, is specified as follows:

$$W = [\eta_1 \eta_2 \ldots \eta_l]^t V, \tag{18}$$

where V is the whitened feature set (see (8)).

Let $\omega_1, \omega_2, \ldots, \omega_L$ and $N_1, N_2, \ldots, N_L$ denote the classes and number of images within each class, respectively. Let $M_1, M_2, \ldots, M_L$ and M_0 be the means of corresponding classes and the grand mean in the new feature space, $span[\eta_1, \eta_2, \ldots, \eta_l]$, we then estimate

$$M_i = \frac{1}{N_i} \sum_{j=1}^{N_i} W_j^{(i)}, \qquad i = 1, 2, \ldots, L, \tag{19}$$

where $W_j^{(i)}$, $j = 1, 2, \ldots, N_i$, represents the sample images from class ω_i, and

$$M_0 = \frac{1}{n} \sum_{i=1}^{L} N_i M_i, \tag{20}$$

where n is the total number of images for all the classes. Thus, $\zeta_s(F)$ is computed as follows:

$$\zeta_s(F) = \sqrt{\sum_{i=1}^{L} (M_i - M_0)^t (M_i - M_0)}. \tag{21}$$

Driven by the fitness function, GA evolves the optimal solution $F^o \equiv \alpha_1^o, \alpha_2^o, \ldots, \alpha_{m(m-1)/2}^o; a_1^o, a_2^o, \ldots, a_m^o$. Let Q in (13) represent this particular basis set corresponding to the rotation angles $\alpha_1^o, \alpha_2^o, \ldots, \alpha_{m(m-1)/2}^o$ (remember, $[\varepsilon_1 \varepsilon_2 \ldots \varepsilon_m]$ in (12) is a unit matrix), and let the column vectors in Q be $\Theta_1, \Theta_2, \ldots, \Theta_m$:

$$Q = [\Theta_1 \Theta_2 \ldots \Theta_m]. \tag{22}$$

If $\Theta_{i_1}, \Theta_{i_2}, \ldots, \Theta_{i_l}$ are the basis vectors corresponding to $a_1^o, a_2^o, \ldots, a_m^o$, then the optimal basis, $T \in \mathbb{R}^{m \times l}$, can be expressed as follows:

$$T = [\Theta_{i_1} \Theta_{i_2} \ldots \Theta_{i_l}], \tag{23}$$

where $i_j \in \{1, 2, \ldots, m\}$, $i_j \neq i_k$ for $j \neq k$, and $l \leq m$.

4.3 The Evolutionary Pursuit (EP) Algorithm

The evolutionary pursuit (EP) algorithm works as follows:

1. Compute the eigenvector and eigenvalue matrices, Φ and Λ, of the covariance matrix, Σ_X, using singular value decomposition (SVD) or Jacobi's method (see (2)). Choose then the first m leading eigenvectors from Φ as basis vectors and project the original image set onto those vectors to form the feature set Z (7) in this reduced PCA space.
2. Whiten the feature set Z and derive the new feature set V in the whitened PCA space (8).
3. Set $[\varepsilon_1\varepsilon_2\ldots\varepsilon_m]$ to be an $m \times m$ unit matrix: $[\varepsilon_1\varepsilon_2\ldots\varepsilon_m] = I_m$.
4. Begin the evolution loop until the stopping criterion is met—such as the fitness does not change further or the maximum number of trials is reached.
   a. Sweep the $m(m-1)/2$ pairs of axes according to a fixed order to get the rotation angle set $\alpha_1^{(k)}, \alpha_2^{(k)}, \ldots, \alpha_{m(m-1)/2}^{(k)}$ from the individual chromosome representation (see Fig. 2), and rotate the unit basis vectors, $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_m$, in this m-dimensional space to derive the new projection axes, $\xi_1^{(k)}, \xi_2^{(k)}, \ldots, \xi_m^{(k)}$, by applying (12) and (13).
   b. Compute the fitness value (17) in the feature space defined by the l projection axes, $\eta_1, \eta_2, \ldots, \eta_l$, which are chosen from the set of the rotated basis vectors $\left\{\xi_1^{(k)}, \xi_2^{(k)}, \ldots, \xi_m^{(k)}\right\}$ corresponding to the $a_i's$ from the individual chromosome representation (see Fig. 2).
   c. Find the sets of angles and the subsets of projection axes that maximize the fitness value and keep those chromosomes as the best solutions so far.
   d. Change the values of rotation angles and the subsets of the projection axes according to GA's genetic operators, and repeat the evolution loop.
5. Carry out recognition using the face basis, $T = [\Theta_{i_1}\Theta_{i_2}\ldots\Theta_{i_l}]$, evolved by the EP ((14) and (15)).

The computational complexity of the algorithm falls mainly into two parts: the PCA computation of Step 1 and the evolution loop of Step 4. In Step 1, the SVD of matrix of size $N \times N$ has the complexity of $O(N^3)$ according to [4], and the derivation of the feature set Z (7) is $O(mnN)$. In Step 4, the rotation transformations of (a) and the fitness value computations of (b) account for most of the computation. In Step 4a, each rotation transformation changes two column vectors (pairwise axes rotation), and there are $m(m-1)/2$ rotations in total, hence, the complexity is $O(m^3)$. In Step 4b, if we only count the number of multiplication, then (18) accounts for the major part of the computation with the computational complexity $O(lmn)$. The overall complexity of the evolution procedure also depends on the maximum number of trials. Note that training (Steps 1 to 4) is done offline.

5 Experiments

We assessed the feasibility and performance of our novel evolutionary pursuit method on the face recognition task, using 1,107 face images from the FERET standard facial database [34]. Robustness of the EP method is shown in terms of both absolute performance indices and comparative performance against traditional face recognition schemes such as Eigenfaces and Fisherfaces.

5.1 Face Recognition

As the new evolutionary pursuit method is tested using a large database of facial images, we review basics on face processing and the FERET standard facial database used for evaluating face recognition algorithms [34]. The face is a unique feature of human beings. Even the faces of "identical twins" differ in some respects [38]. Humans can detect and identify faces in a scene with little or no effort. This skill is quite robust, despite large changes in the visual stimulus due to viewing conditions, expression, aging, and disguises such as glasses or changes in hair style. Building automated systems that accomplish this task is, however, very difficult. There are several related face processing subproblems: 1) detection of a pattern as a face, 2) face recognition, 3) analysis of facial expressions, and 4) characterization (gender or ethnicity) based on physical features. An automated vision system that performs those operations will find countless nonintrusive applications, e.g., surveillance, criminal identification and retrieval of missing children, workstation and building security, credit card (ATM) verification, and video-document retrieval [5], [38], [33].

The robustness of the evolutionary pursuit method described in this paper is comparatively assessed against Eigenfaces and Fisherfaces using the U.S. Army FacE REcognition Technology (FERET) facial database [34], which has become the standard data set for benchmark studies. The FERET database consists now of 13,539 facial images corresponding to 1,565 sets (each set represents a human subject) among whom 366 sets are duplicates whose images have been acquired at a later time. The images used for our experiments are of size 256×384 with 256 gray scale levels. Since images are acquired during different photo sessions, the lighting conditions and the size of the face may vary. The diversity of the FERET database is across gender, ethnicity, and age. The image sets 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.

5.2 Experimental Results

In order to compute the within-class scatter, Fisherfaces [2] and our novel evolutionary pursuit method require at least two training images for each class. To accommodate this requirement, we chose a FERET subset of 1,107 images corresponding to 369 subjects such that there are three frontal images for each subject. The variety of the subset is such that, for the first 200 subjects, the third image is acquired at low illumination, while for the remaining 169 subjects the face images are acquired during different photo sessions and the later acquired images are referred to as duplicates. Fig. 4 shows some of the face images used in our experiments. Two images of each subject are used for training with the remaining image for testing. In other words, the training set includes 738 images while the test set 369 images. The images are cropped to the size of 64×96 and

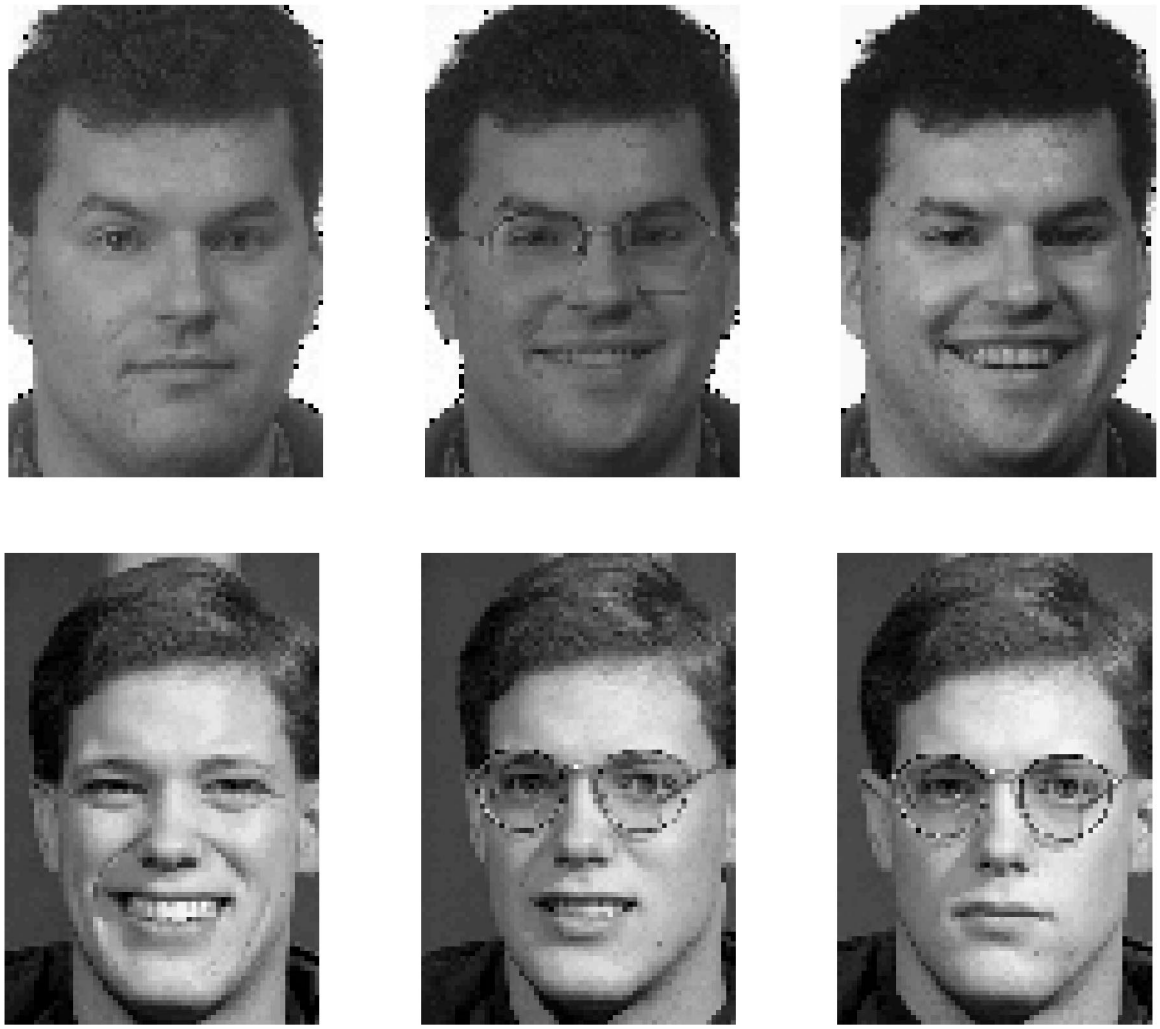

Fig. 4. Some example FERET images used in our experiments (cropped to the size of 64×96). Since the images are acquired during different photo sessions, both the lighting conditions and the facial expressions may vary.

the eye coordinates are manually detected. The background is uniform and the face images are not masked. The reasoning behind not masking the face images is that

> at least on some occasions, the processing performed by the visual system to judge identity is better characterized as "head recognition" rather than "face recognition" [39].

Masking as it usually has been implemented deletes the face outline and the effect of such deletions on recognition performance is discussed in our recent paper [29]. Shape-free face recognition methods avoid this problem by using the shape of the outline encoded by a number of control points for subsequent alignment and normalization [8].

We implemented the evolutionary pursuit (EP) algorithm and searched the spaces corresponding to $m = 20$ and $m = 30$, respectively. Note that the PCA decomposition is generated using the 738 training images and PCA reduces the dimensionality of the original image space from $N = (64 \times 96)$ to m. The first 30 eigenfaces, shown in Fig. 5, form the basis vectors used by the Eigenfaces method. Fisherfaces was implemented and experimented with, as well. Both the Eigenfaces and the Fisherfaces implementations use the Euclidean distance measure, as suggested in [42] and [2]. Table 1 shows comparative training performance, while Tables 2 and 3 give comparative testing performance. In Tables 2 and 3, top one recognition rate means the accuracy rate for the top response being correct, while top three recognition rate represents the accuracy rate for the correct response being included among the first three ranked choices.

Starting from the 20-dimensional space ($m = 20$), the evolutionary pursuit method derives 18 vectors as the optimal face basis. Fig. 6 shows the corresponding 18 basis vectors, while Fig. 7 shows the nonorthogonality of these vectors. For each row (or column), the unit bar in Fig. 7 (along the diagonal position) represents the norm of a basis vector, and the other bars correspond to the dot products of this vector and the remaining 17 basis vectors. Since some of the dot products are nonzero, these basis vectors are not orthogonal. For comparison purposes, Fig. 8 shows the 26 basis vectors found by EP when it operates in a 30-dimensional space ($m = 30$). Note that while, for PCA the basis vectors have a natural order, this is not the case with the basis derived by EP due to the rotations involved during the evolutionary procedure. The natural order characteristic of eigenfaces reflects the representational aspect of PCA and its relationship to spectral decomposition. The very first Principlal Components encode global image characteristics

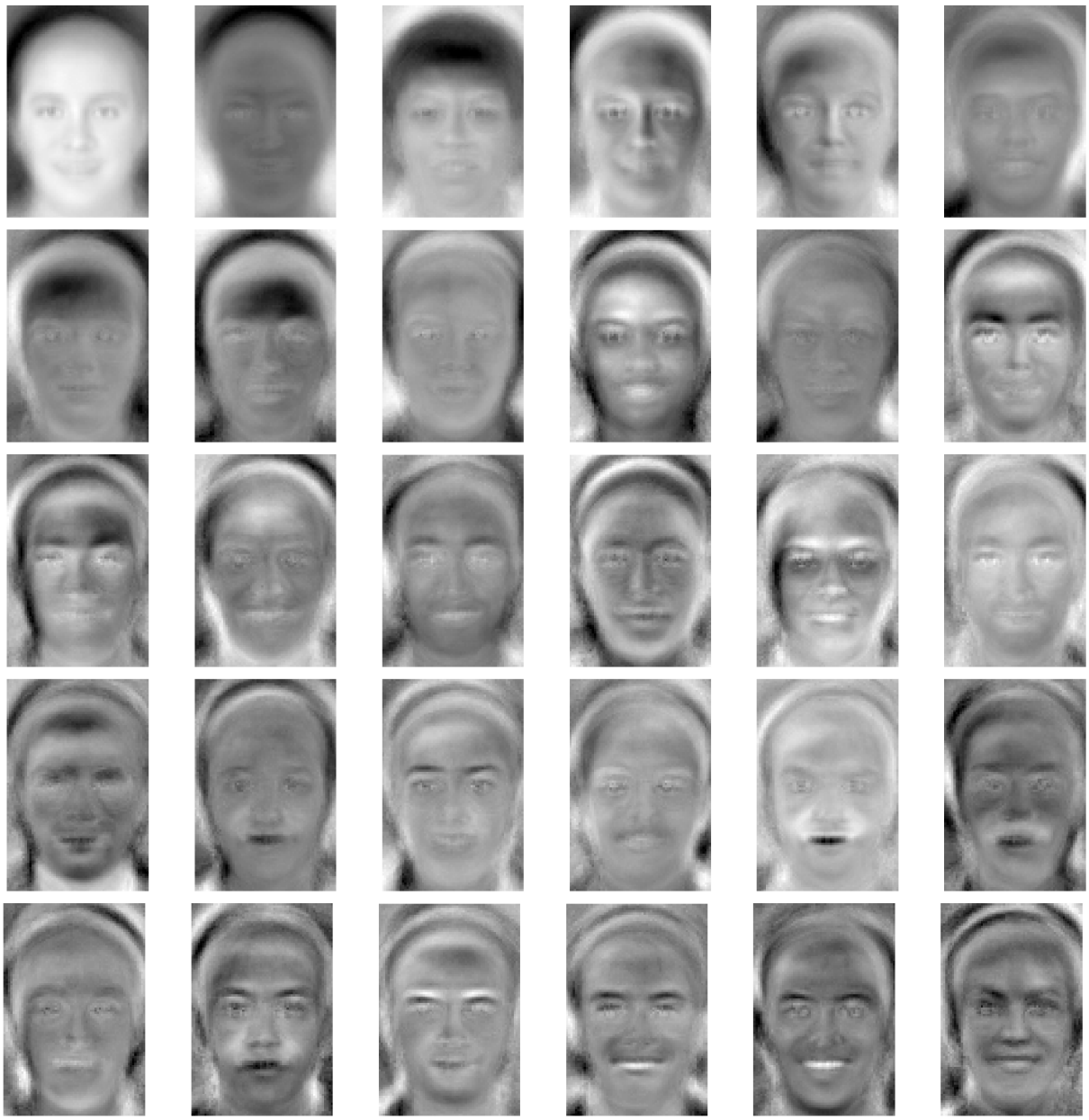

Fig. 5. The first 30 eigenfaces derived by the Eigenfaces method. The order of these vectors is listed from left to right and top to bottom corresponding to the eigenvalues in decreasing order.

in analogy to low-frequency components. EP, on the other hand, is a procedure geared primarily towards recognition and generalization, and from Fig. 6 and Fig. 8 one is inclined to conjecture that the outcome would be the derivation of features whose local contents are enhanced.

Table 1 shows the comparative training performance of Eigenfaces, Fisherfaces, and evolutionary pursuit with 18 and 26 basis vectors, respectively. The performance of Eigenfaces and Fisherfaces using 20 and 30 basis vectors is also shown in the same table, and one can see that the training performance for Fisherfaces is perfect (100 percent correct recognition rate). During testing (see Tables 2 and 3) and using 369 test images (not seen during training), the performance displayed by Fisherfaces, however, deteriorates as it lacks a good generalization ability. Both Eigenfaces and EP display better generalization when compared against Fisherfaces. In particular, Table 2 shows that when the 20-dimensional whitened PCA space is searched, EP derives 18 vectors as the optimal basis with top one recognition rate 87.80 percent compared to 81.57 percent for Eigenfaces and 79.95 percent for Fisherfaces using the same number (18) of basis vectors. When Eigenfaces and Fisherfaces use 20 basis vectors, the recognition performance is 83.47 percent and 81.84 percent, respectively, which is still lower than that of EP which uses only 18 vectors. For top three recognition rate, the EP approach again comes first and yields 95.93 percent, compared to 94.58 percent for Eigenfaces and 87.80 percent for Fisherfaces using 18 features, and 94.58 percent for the former and 90.79 percent for the latter using 30 features. EP comes out again first (see Table 3) when face recognition is carried out starting with a 30-dimensional PCA space.

TABLE 1
Comparative Training Performance for Eigenfaces, Fisherfaces, and Evolutionary Pursuit

method \axes	18	20	26	30
Eigenface	78.05%	79.40%	81.30%	80.76%
Fisherfaces	100%	100%	100%	100%
Evolutionary Pursuit	83.47%	N/A	82.66%	N/A

TABLE 2
Comparative Testing Performance for Eigenfaces, Fisherfaces, and Evolutionary Pursuit ($m = 20$)

method	# axes	top 1 recognition rate	top 3 recognition rate
Eigenface	18	81.57%	94.58%
Eigenface	20	83.47%	94.58%
Fisherfaces	18	79.95%	89.16%
Fisherfaces	20	81.84%	90.79%
Evolutionary Pursuit	18	87.80%	95.93%

TABLE 3
Comparative Testing Performance for Eigenfaces, Fisherfaces, and Evolutionary Pursuit ($m = 30$)

method	# axes	top 1 recognition rate	top 3 recognition rate
Eigenface	26	87.26%	95.66%
Eigenface	30	88.62%	95.93%
Fisherfaces	26	86.45%	93.77%
Fisherfaces	30	88.08%	95.39%
Evolutionary Pursuit	26	92.14%	97.02%

From Tables 1, 2, and 3, it becomes apparent that Fisherfaces does not display good generalization abilities, while Eigenfaces and evolutionary pursuit do. The range of training data is quite large, as it consists of both original and duplicate images acquired at a later time. As a consequence, during training, Fisherfaces performs better than both Eigenfaces and evolutionary pursuit because it overfits to a larger extent its classifier to the data. Evolutionary pursuit yields, however, improved performance over the other two methods during testing.

Table 4 shows the testing performance for EP when it operates in the 20- and 30-dimensional nonwhitened PCA spaces, respectively. Again, EP derives 18 and 26 basis vectors corresponding to the 20- and 30-dimensional PCA spaces. But the recognition rates shown in Table 4 are not as good as those shown in Tables 2 and 3, a reasonable expectation demonstrating the importance of the whitening transformation to the EP method.

6 Conclusions

We introduced in this paper Evolutionary Pursuit (EP), a novel adaptive representation method, and showed its feasibility for the face recognition problem. In analogy to pursuit methods, EP seeks to learn an optimal basis for the dual purpose of data compression and pattern classification. The challenge for EP is to increase the generalization ability of the learning machine as a result of seeking the trade-off between minimizing the empirical risk encountered during training and narrowing the confidence interval for reducing the guaranteed risk for future testing on unseen images. Towards that end, EP implements strategies characteristic of genetic algorithms (GAs) for searching the space of possible solutions and determining an optimal basis. Within the face recognition framework, EP seeks an optimal basis for face projections suitable for compact and efficient face encoding in terms of both present and future recognition ability. Experimental results, using a large and

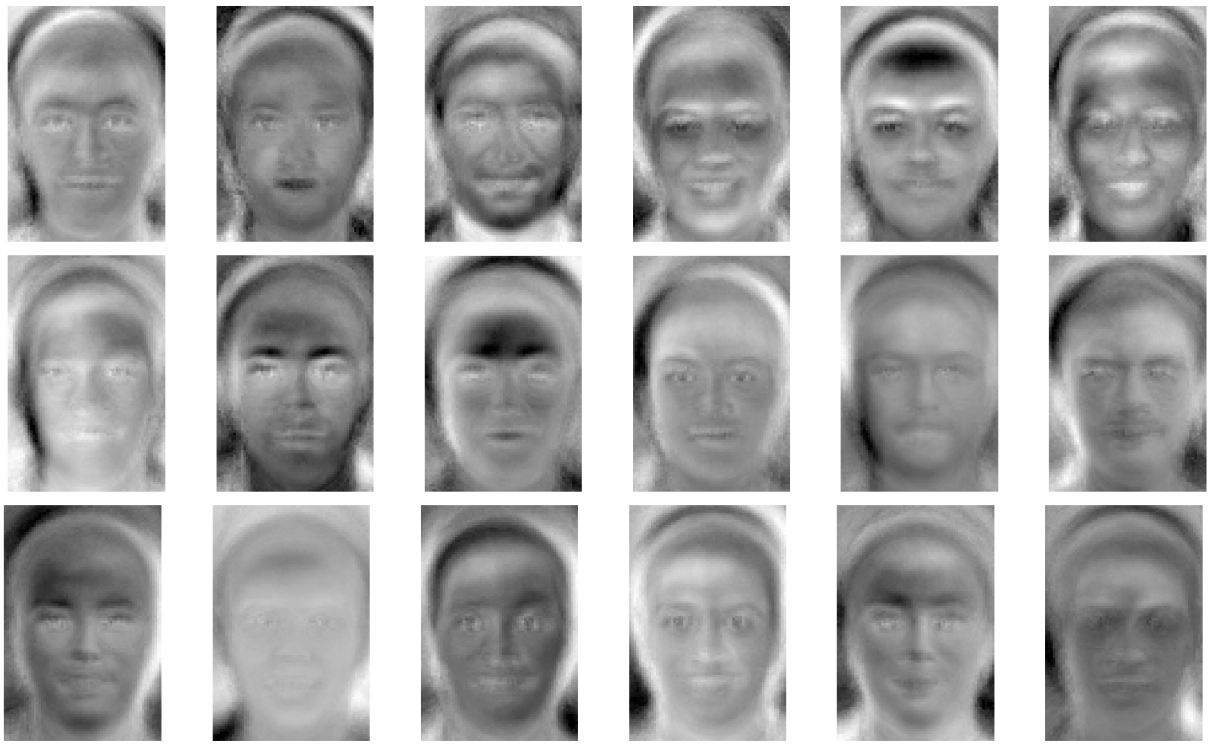

Fig. 6. Optimal basis (18 vectors) derived by the evolutionary pursuit (EP).

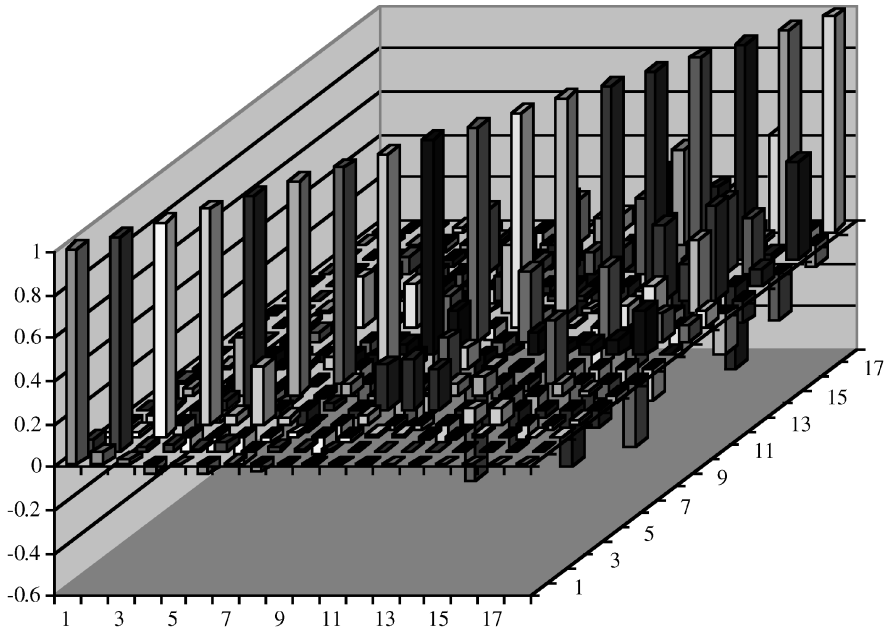

Fig. 7. Nonorthogonality of the basis vectors derived by evolutionary pursuit. For each row (or column), the unit bar (along the diagonal position) represents the norm of a basis vector, and the other bars correspond to the dot products of this vector and the remaining 17 basis vectors. Since some of the dot products are nonzero, these basis vectors are not orthogonal.

varied subset from the FERET facial database, show that the EP method compares favorably against two popular methods for face recognition—Eigenfaces and Fisherfaces.

As second order statistics provide only partial information on the statistics of both natural images and human faces, it might become necessary to incorporate higher order statistics, as well. While PCA considers the second order statistics only and it uncorrelates data, Independent Component Analysis (ICA) considers also the higher order statistics and it identifies the independent source components from their linear mixtures by minimizing the mutual information expressed as a function of high order cumulants [7], [22], [20]. ICA, thus, provides a more powerful data representation than PCA [22], [28]. EP is analogous to ICA in that both methods carry out the whitening and pairwise axes rotation transformations to derive the projection basis [7]. While ICA uses a criterion of independence or minimization of mutual information, EP is based on a

Fig. 8. Optimal basis (26 vectors) derived by the evolutionary pursuit (EP).

criterion addressing both the recognition performance and generalization ability.

The fitness function driving evolution considers both recognition rates ("performance accuracy"), i.e., empirical risk, and the scatter index, i.e., predicted risk. The fitness function is similar to the cost functional used by regularization theory for solving ill-posed problems in computer vision and improving the generalization ability of RBF networks. As the empirical and predicted risks place opposite pressures on the fitness function, it's up to GAs

TABLE 4
The Testing Performance for Evolutionary Pursuit (EP) Operating in the 20- and 30-Dimensional Nonwhitened PCA Spaces, Respectively

method	# axes	top 1 recognition rate	top 3 recognition rate
EP without whitening ($m = 20$)	18	81.03%	92.95%
EP without whitening ($m = 30$)	26	87.26%	95.66%

to evolve a well-balanced behavior displaying good performance during both training and future test trials. The prediction risk, included as a penalty, is a measure of generalization ability and is driven by the scatter index ("class separation"). In analogy to statistical learning theory, the scatter index is conceptually similar to the capacity of the classifier and its use is to prevent overfitting. One can consider the greedy search for an optimal basis leading to improved pattern separation as an attempt to redefine the search space with respect to exemplar projection axes. Overall, EP provides a new methodology for both functional approximation and pattern classification problems. A worthwhile direction for future research is to explore the role that EP can play to further support the equivalence between sparse approximation and Support Vector Machines (SVM) using the regularization theory [16]. In particular, EP could play a constructive role in terms of its ability to adaptively and efficiently search through large dictionary sets. Furthermore, one could expand on the above and also explore the role that EP can play in searching dictionary sets on a local basis. Toward that end, one should comparatively assess possible equivalence between the class of sparse representation for functional approximation and SVM (Poggio and Girosi [35]), Local Feature Analysis (LFA) (Penev and Atick [32]), Basis Pursuit (Chen and Donoho [6]), and EP as described in this paper.

ACKNOWLEDGMENTS

The authors would like to thank all the reviewers for their valuable comments. This work was partially supported by the DoD Counterdrug Technology Development Program, with the U.S. Army Research Laboratory as Technical Agent, under contract DAAL01-97-K-0118.

REFERENCES

[1] H.B. Barlow, "Unsupervised Learning," *Neural Computation,* vol.7 1, pp. 295-311, 1989.
[2] 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, 1997.
[3] R. Brunelli and T. Poggio, "Face Recognition: Features vs. Templates," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 15, no. 10, pp. 1,042-1,053, 1993.
[4] T.F. Chan, "An Improved Algorithm for Computing the Singular Value Decomposition," *ACM Trans. Math. Software,* vol. 8, pp. 72-83, 1982.
[5] R. Chellappa, C.L. Wilson, and S. Sirohey, "Human and Machine Recognition of Faces: A Survey," *Proc. IEEE,* vol. 83, no. 5, pp. 705-740, 1995.
[6] S. Chen and D. Donoho, "Basis Pursuit," technical report, Dept. Statistics, Stanford Univ., 1996.
[7] P. Comon, "Independent Component Analysis, A New Concept?" *Signal Processing,* vol. 36, pp. 287-314, 1994
[8] T.F. Cootes, G.J. Edwards, and C.J. Taylor, "Active Appearance Models," *Proc. Fifth European Conf. Computer Vision,* 1998.
[9] J.G. Daugman, "An Information-Theoretic View of Analog Representation in Striate Cortex," *Computational Neuroscience,* E.L. Schwartz, ed., pp. 403-424, MIT Press, 1990.
[10] K.I. Diamantaras and S.Y. Kung, *Principal Component Neural Networks: Theory and Applications.* John Wiley and Sons, 1996.
[11] G.M. Edelman, *Neural Darwinism.* Basic Books, 1987.
[12] K. Etemad and R. Chellappa, "Discriminant Analysis for Recognition of Human Face Images," *J. Optical Soc. Am. A,* vol. 14, pp. 1,724-1,733, 1997.
[13] R.A. Fisher, "The Use of Multiple Measures in Taxonomic Problems," *Ann. Eugenics,* vol. 7, pp. 179-188, 1936.
[14] J.H. Friedman and W. Stuetzle, "Projection Pursuit Regression," *J. Am. Statistical Assoc.,* vol. 76, pp. 817-823, 1981.
[15] K. Fukunaga, *Introduction to Statistical Pattern Recognition.* Academic Press, second ed., 1991.
[16] F. Girosi, "An Equivalence between Sparse Approximation and Support Vector Machines," *Neural Computation,* vol. 10, pp. 1,455-1,480, 1998.
[17] D. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning,* Addison-Wesley, 1989.
[18] S. Haykin, *Neural Networks—A Comprehensive Foundation.* Macmillan College Publishing Company, 1994.
[19] P.J. Huber, "Projection Pursuit," *Annals Statistics,* vol. 13, pp. 435-475, 1985.
[20] A. Hyvarinen and E. Oja, "A Fast Fixed-Point Algorithm for Independent Component Analysis," *Neural Computation,* vol. 9, pp. 1,483-1,492, 1997.
[21] I.T. Jolliffe, *Principal Component Analysis,* New York: Springe, 1986.
[22] J. Karhunen, E. Oja, L. Wang, R. Vigario, and J. Joutsensalo, "A Class of Neural Networks for Independent Component Analysis," *IEEE Trans. Neural Networks,* vol. 8, no. 3, pp. 486-504, 1997.
[23] 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, 1990.
[24] R. Linsker, "Self-Organization in a Perceptual Network," *Computer,* vol. 21, pp. 105-117, 1988.
[25] C. Liu and H. Wechsler, "Enhanced Fisher Linear Discriminant Models for Face Recognition," *Proc. 14th Int'l Conf. Pattern Recognition,* 1998.
[26] C. Liu and H. Wechsler, "Face Recognition Using Evolutionary Pursuit," *Proc. Fifth European Conf. Computer Vision,* 1998.
[27] C. Liu and H. Wechsler, "Probabilistic Reasoning Models for Face Recognition," *Proc. Computer Vision and Pattern Recognition,* 1998.
[28] C. Liu and H. Wechsler, "Comparative Assessment of Independent Component Analysis(IC A) for Face Recognition," *Proc. Second Int'l Conf. Audio- and Video-Based Biometric Person Authentication,* 1999.
[29] C. Liu and H. Wechsler, "Face Recognition Using Shape and Texture," *Proc. Computer Vision and Pattern Recognition,* 1999.
[30] 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.
[31] B.A. Olshausen and D.J. Field, "Emergence of Simple-Cell Receptive Field Properties by Learning a Sparse Code for Natural Images," *Nature,* vol. 381, no. 13, pp. 607-609, 1996.
[32] P.S. Penev and J.J. Atick, "Local Feature Analysis: A General Statistical Theory for Object Representation," *Network: Computation in Neural Systems,* vol. 7, pp. 477-500, 1996.
[33] P.J. Phillips, "Matching Pursuit Filters Applied to Face Identification," *IEEE Trans. Image Processing,* vol. 7, no. 8, pp. 1,150-1,164, 1998.
[34] 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.
[35] T. Poggio and F. Girosi, "A Sparse Representation for Function Approximation," *Neural Computation,* vol. 10, pp. 1,445-1,454, 1998.
[36] T. Poggio, V. Torre, and C. Koch, "Computational Vision and Regularization Theory," *Nature,* vol. 317, pp. 314-319, 1985.
[37] D. Ruderman, "The Statistics of Natural Images," *Network: Computation in Neural Systems,* vol. 5, pp. 598-605, 1994.
[38] A. Samal and P.A. Iyengar, "Automatic Recognition and Analysis of Human Faces and Facial Expression: A Survey," *Pattern Recognition,* vol. 25, no. 1, pp. 65-77, 1992.
[39] P. Sinha and T. Poggio, "I Think I Know that Face...," *Nature,* vol. 384, pp. 404, 1996.
[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, 1996.
[41] A.N. Tikhonov and V.Y. Arsenin, *Solutions of Ill-Posed Problems,* Washington: W.H. Winston, 1977.
[42] M. Turk and A. Pentland, "Eigenfaces for Recognition," *J. Cognitive Neuroscience,* vol. 13, no. 1, pp. 71-86, 1991.
[43] Y.N. Vapnik, *The Nature of Statistical Learning Theory.* Springer-Verlag, 1995.

Chengjun Liu received the PhD from George Mason University in 1999, and he is presently an assistant professor of mathematics and computer science at the University of Missouri, St. Louis. 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 models for image retrieval and face recognition based upon statistical and machine learning concepts. The class of new models includes the Probabilistic Reasoning Models (PRM), which combines the optimal representation criterion (PCA) with the optimal classification criterion (Bayes classifier), the Enhanced FLD (Fisher Linear Discriminant) Models (EFM), which improves upon the generalization ability of FLD and the Enhanced Independent Component Analysis (EICA), both using sensitivity analysis on the spectral range of eigenvalues. 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.

Harry Wechsler received the PhD in computer science from the University of California, Irvine, in 1975, and he is presently a professor of computer science at George Mason University. His research in the field of intelligent systems has been in the areas of PERCEPTION: Computer Vision (CV), Automatic Target Recognition (ATR), Signal and Image Processing (SIP), MACHINE INTELLIGENCE and LEARNING: Pattern Recognition (PR), Neural Networks (NN), Information Retrieval, Data Mining and Knowledge Discovery, EVOLUTIONARY COMPUTATION: Genetic Algorithms (GAs), Animates and Artificial Life, and HUMAN-COMPUTER INTELLIGENT INTERACTION (HCII): Face and Hand Gesture Recognition, Biometrics and Forensics, and Video Surveillance and W4. He was the director for the NATO Advanced Study Institutes (ASI) on "Active Perception and Robot Vision" (Maratea, Italy, 1989), "From Statistics to Neural Networks" (Les Arcs, France, 1993) and "Face Recognition: From Theory to Applications" (Stirling, UK, 1997), and he served as cochair for the International Conference on Pattern Recognition held in Vienna, Austria, in 1996. He has authored more than 200 scientific papers, the book *Computational Vision* (Academic Press, 1990), and was the editor for *Neural Networks for Perception* (vol. 1 and 2), published by Academic Press in 1991. He was elected as an IEEE fellow in 1992 and as an International Association of Pattern Recognition (IAPR) Fellow in 1998.