Comparison of statistical pattern-recognition algorithms for hybrid processing. II. Eigenvector-based algorithm

Q. Tian, Y. Fainman, and Sing H. Lee

Department of Electrical and Computer Engineering, University of California, San Diego, La Jolla, California 92093

Received July 17, 1986; accepted May 18, 1988

The pattern-recognition algorithms based on eigenvector analysis (group 2) are theoretically and experimentally compared. Group 2 consists of Foley–Sammon (F-S) transform, Hotelling trace criterion (HTC), Fukunaga–Koontz (F-K) transform, linear discriminant function (LDF), and generalized matched filter (GMF) algorithms. It is shown that all eigenvector-based algorithms can be represented in a generalized eigenvector form. However, the calculations of the discriminant vectors are different for different algorithms. Summaries of methods of calculating the discriminant functions for the F-S, HTC, and F-K transforms are provided. Especially for the more practical, underdetermined case, where the number of training images is less than the number of pixels in each image, the calculations usually require the inversion of a large, singular pixel correlation (or covariance) matrix. We suggest solving this problem by finding its pseudoinverse, which requires inverting only the smaller, nonsingular image-correction (or covariance) matrix plus multiplying several nonsingular matrices. We also compare theoretically the classification performance with discriminant functions of the F-S, HTC, and F-K with the LDF and GMF algorithms and the linear-mapping-based algorithms with the eigenvector-based algorithms. Experimentally, we compare the eigenvector-based algorithms, using two sets of image data bases with each image consisting of 64 × 64 pixels.

1. INTRODUCTION

Many pattern-recognition algorithms and systems have been proposed and studied.[1–17] In hybrid systems, all algorithms are implemented in two steps: calculating the inner product of the discriminant vectors and the test images optically by using a computer-generated hologram and making decisions in a decision space electronically by using a microcomputer. Various algorithms differ in the ways in which they extract the discriminant vectors for incorporation into the computer-generated hologram to ensure reliable classification.

In Part I of this series[1] we compared algorithms based on linear-mapping techniques.[2–7] Here we systematically compare algorithms based on eigenvector analysis, which include Foley–Sammon (F-S) transform,[8,9] Hotelling trace criterion[10,11] (HTC), Fukunaga–Koontz (F-K) transform,[12,13] linear discriminant function[14] (LDF), and generalized matched filter[15] (GMF) algorithms. Most eigenvector-based algorithms are useful only in treating two-class problems. However, the HTC algorithm in its pooled modes can be applied to treat multiple-class problems as well.

A brief review of the algorithms to be compared is given in Section 2. A comparison is presented in Section 3. The relation between the eigenvector-based algorithms and the least-squares linear-mapping technique (LSLMT) algorithm is studied in Section 4. Section 5 provides the experimental classification results of the algorithms applied to two sets of images. Final conclusions on the comparison of the algorithms are provided in Section 6.

2. REVIEW OF EIGENVECTOR-BASED ALGORITHMS

To facilitate their comparisons we briefly review the F-S, HTC, and F-K transforms and LDF and GMF algorithms. The computational procedures for the discriminant functions of F-S, HTC, and F-K transforms will be summarized. To be consistent, we will use the same notation and definitions here as in Part I, plus the summary list of symbols given in Appendix A.

F-S Transform

The F-S transform is developed from the classical Fisher ratio method,[16] which is optimal for class separability. (The Fisher ratio is a well-known Rayleigh quotient[9].) But, besides the first F-S discriminant vector, which is exactly the same as the discriminant vector from the Fisher ratio method, we find that the second F-S discriminant vector orthonormal to the first is helpful in improving classification reliability. Based on maximizing the Fisher ratio

$$R(d) = \frac{(d^T\Delta)^2}{d^T S_2 d} = \frac{(d^T\Delta)(d^T\Delta)^T}{d^T S_2 d} = \frac{d^T S_1 d}{d^T S_2 d}, \tag{2.1}$$

the first two optimum discriminant vectors d are found as follows[9].

When S_2 is Nonsingular

Differentiating $R(d)$ from Eq. (2.1) with respect to d and setting the result equal to zero, we obtain

0740-3232/88/101670-13$02.00

$$\nabla_d R(d) = \frac{2d^T\Delta}{d^T S_2 d}\Delta - \left(\frac{d^T\Delta}{d^T S_2 d}\right)^2 2S_2 d = 0.$$

When S_2 is a nonsingular matrix and $d^T\Delta/d^T S_2 d$ is a scalar, the first discriminant vector, d_1, can be calculated from the equation above:

$$d_1 = \alpha_1 S_2^{-1}\Delta, \tag{2.2a}$$

where a normalization constant α_1 is chosen to satisfy $d_1{}^T d_1 = 1$. The second discriminant vector, d_2, can be calculated under the constraint that $d_1{}^T d_2 = 0$ by using the Lagrange method. Let C be

$$C = \frac{(d_2{}^T\Delta)^2}{d_2{}^T S_2 d_2} - \lambda d_2{}^T d_1.$$

Differentiating C with respect to d_2, we obtain

$$\frac{\partial C}{\partial d_2} = \frac{2d_2{}^T\Delta}{d_2{}^T S_2 d_2}\Delta - \left(\frac{d_2{}^T\Delta}{d_2{}^T S_2 d_2}\right)^2 2S_2 d_2 - \lambda d_1 = 0.$$

After using the constraint $d_1{}^T d_2 = 0$ to determine λ in the above equation, we finally obtain

$$d_2 = \alpha_2 S_2^{-1}\left(I - \frac{\Delta^T S_2{}^{-2}\Delta}{\Delta^T S_2{}^{-3}\Delta} S_2^{-1}\right)\Delta, \tag{2.2b}$$

where the normalization constant α_2 is again chosen such that $d_2{}^T d_2 = 1$.

When S_2 is Singular

In image classification the dimension of the images, N, is usually larger than the total number of training images, M ($=\sum_{i=1}^{K} M_i$, the underdetermined case[1]). Therefore the matrix S_2 is a singular matrix, and the inverse of S_2 does not exist. In this situation, there is no unique solution for the discriminant vector d that will maximize the Fisher ratio $R(d)$. However, in a subspace of dimension r, where r is the number of nonzero eigenvalues of the matrix S_2 ($r \le M$), unique solutions for r-discriminant vectors d can be found. A more thorough analysis on maximization of the Fisher ratio when the matrix S_2 is singular can be found in Ref. 9; here we reproduce the first two discriminant vectors:

$$d_1 = \alpha_1 S_2{}^{+}\Delta, \tag{2.3a}$$

$$d_2 = \alpha_2 S_2{}^{+}\left(I - \frac{\Delta^T S_2{}^{+2}\Delta}{\Delta^T S_2{}^{+3}\Delta} S_2{}^{+}\right)\Delta, \tag{2.3b}$$

where $S_2{}^{+}$ is the pseudoinverse of the covariance matrix S_2. $S_2{}^{+}$ can be expressed by the relation described in Appendix B and given by

$$S_2{}^{+} = W_s S_{\mathrm{IM}}{}^{-2} W_s{}^T. \tag{2.4a}$$

Using the eigenvectors E_s and eigenvalues Λ_s of the image-covariance matrix S_{IM}, we can rewrite Eq. (2.4a) as

$$S_2{}^{+} = W_s E_s \Lambda_s{}^{-2} E_s{}^T W_s{}^T, \tag{2.4b}$$

which is exactly the same equation as the one provided by the authors of Ref. 9.

Generalized Eigenvector Form of the F-S Transform

As illustrated in Ref. 17, maximizing the Fisher ratio is equivalent to finding the generalized eigenvectors d_i' and eigenvalues θ_i from

$$[S_1 - \theta_i S_2] d_i' = 0. \tag{2.5}$$

The eigenvector d_i' associated with the largest eigenvalue θ_i is the first F-S discriminant vector d_1.[9,17] To generate a series of orthonormal F-S discriminant vectors a recursive formula is given in Ref. 8. With this series of discriminant vectors a decision plane useful in interactive pattern recognition can be constructed so that a nonlinear piecewise decision boundary can be designed for classification.[18]

HTC Algorithm

The HTC[10,11] determines the optimum discriminant vectors by maximizing the trace of the matrix $S_2{}^{-1}S_1$:

$$J = \mathrm{Tr}(S_2{}^{-1}S_1), \tag{2.6}$$

where Tr denotes the trace operation.[10] This criterion implies that the distances between classes are maximized, while the distances within classes are minimized. The optimum transformation D for maximizing J is then given by the first $(K-1)(K<N)$ generalized eigenvectors of the matrix $S_2{}^{-1}S_1$, which can be determined by simultaneous diagonalization of the two symmetric matrices

$$D^T S_1 D = \Theta, \tag{2.7a}$$

$$D^T S_2 D = I. \tag{2.7b}$$

The transformation matrix $D = [d_1', d_2', \ldots, d_{K-1}']$ can be calculated in two major steps: (i) Find the eigenvectors E_s and the eigenvalues Λ_s of S_{IM}. The orthonormal vectors $(W_s E_s \Lambda_s{}^{-1/2})$ and the eigenvalues (Λ_s) of S_2 can then be calculated to satisfy

$$(\Lambda_s{}^{-1} E_s{}^T W_s{}^T) S_2 (W_s E_s \Lambda_s{}^{-1}) = I. \tag{2.8}$$

(ii) Transform S_1 by the vectors $(W_s E_s \Lambda_s{}^{-1})$ and calculate eigenvectors Ψ_s and eigenvalues Γ_s of the transformed matrix $(\Lambda_s{}^{-1} E_s{}^T W_s{}^T) S_1 (W_s E_s \Lambda_S{}^{-1})$

$$[(\Lambda_s{}^{-1} E_s{}^T W_s{}^T) S_1 (W_s E_s \Lambda_s{}^{-1})]\Psi_s = \Psi_s \Gamma_s. \tag{2.9}$$

The vector d for the optimum transform can be determined from

$$(d_1', d_2', \ldots, d_{K-1}') = \Psi_s{}^T \Lambda_s{}^{-1} E_s{}^T W_s{}^T. \tag{2.10}$$

For a two-class problem, two discriminant vectors (d_1', d_2') can be found, although only one of them will provide discriminant power. (The vectors $d_1', d_2', \ldots, d_{K-1}'$ are the same as the vectors $A_1, A_2, \ldots A_{K-1}$ of Ref. 10.)

F-K Algorithm

The F-K transform[12,13] is based on the Karhunen–Loéve transform and uses the training image vectors themselves instead of the mean-removed training images as the F-S and HTC transforms do. It is based on simultaneous diagonalization of the correlation matrices R and $R^{(1)}$ or $R^{(2)}$. First, find the eigenvectors E_r and eigenvalues Λ_r of R_{IM} and calculate $(WE_r\Lambda_r{}^{-1})$ to transform R into an identity matrix

$$(\Lambda_r^{-1}E_r^T W^T)R(WE_r\Lambda_r^{-1}) = I \tag{2.11a}$$

or

$$(\Lambda_r^{-1/2}\Psi_r^T)R(\Psi_r\Lambda_r^{-1/2}) = I, \tag{2.11b}$$

where Ψ_r are eigenvectors of the correlation matrix R given by $\Psi_r = WE_r\Lambda_r^{-1/2}$. By using the vectors $\Psi_r\Lambda_r^{-1/2}$, the correlation matrices $R^{(1)}$ and $R^{(2)}$ for class 1 and class 2, respectively, can be transformed into two new matrices, whose corresponding eigenvectors and eigenvalues can be calculated as

$$[\Lambda_r^{-1/2}\Psi_r^T R^{(1)}\Psi_r\Lambda_r^{-1/2}]\Psi_{r1} = \Psi_{r1}\Gamma_r, \tag{2.12a}$$

$$[\Lambda_r^{-1/2}\Psi_r^T R^{(2)}\Psi_r\Lambda_r^{-1/2}]\Psi_{r2} = \Psi_{r2}K_r. \tag{2.12b}$$

It was proved in Ref. 13 that $\Psi_{r1} = \Psi_{r2}$ and $\Gamma_r + K_r = I$. In general, if the M_1 training images for class 1 are linearly independent, the matrix $R^{(1)}$ has a rank M_1, and therefore M_1 independent discriminant vectors can be found. The two discriminant vectors with the best discrimination performance are calculated from the two eigenvectors in Ψ_{r1} associated with the largest and the smallest eigenvalues ($\psi_{r1}{}^{\max}$ and $\psi_{r1}{}^{\min}$):

$$d_1^T = \psi_{r1}{}^{\max T}\Lambda_r^{-1}E_r^T W^T, \tag{2.13a}$$

$$d_2^T = \psi_{r1}{}^{\min T}\Lambda_r^{-1}E_r^T W^T. \tag{2.13b}$$

LDF Algorithm

The LDF transform was introduced as a classification algorithm for the generalized chord transform.[14] The principle of the LDF is based on finding a vector d that maximizes the Fisher ratio of Eq. (2.1) by solving the generalized eigenvalue Eq. (2.5). If the scatter matrix S_2 is nonsingular, Eq. (2.5) can be rewritten as

$$S_2^{-1}S_1 d = \theta d. \tag{2.14}$$

The solution to Eq. (2.14) is the LDF filter given in Ref. 12, is identical to the first F-S vector described by Eq. (2.2a), and is given by

$$d = S_2^{-1}\Delta. \tag{2.15}$$

It should be mentioned that the LDF filter so determined assumes the existence of the matrix S_2^{-1} (i.e., S_2 is a nonsingular matrix). To satisfy this assumption the number of the training features (N) must be reduced in relation to the total number of training images ($M_1 + M_2$) for a two-class problem. Furthermore, the LDF algorithm provides only one discriminant vector.

GMF Algorithm

The generalized matched filter was introduced in Ref. 15 and was based on the generalized eigenvector problem given by Eq. (2.5). In order to deal with the inversion of a large matrix S_2 the authors assumed no cross correlation between pixels of the Fourier transforms of the training images, resulting in a diagonalized version of the scatter (covariance) matrix (S_2). The approximate discriminant vector can be determined from Eq. (2.15) by using the inverse of the diagonalized version of S_2. A discriminant vector so obtained has been applied to discriminate aircraft from Gaussian-noise background.

3. COMPARISON OF EIGENVECTOR-BASED ALGORITHMS

In this section we show that F-S, HTC, and F-K algorithms can be represented in a generalized eigenvector form. These algorithms are compared with one another and with the LDF and GMF algorithms in terms of their effectiveness as methods of classification.

Representing the F-S Transform in a Generalized Eigenvector Form

When the matrix S_2 is nonsingular, the generalized eigenvector form of a F-S transform [i.e., Eq. (2.5)] can be transformed to the normal eigenvalue problem

$$(S_2^{-1}S_1 - \theta_i I)d_i' = 0. \tag{3.1}$$

It can be shown that d_1 defined in Eq. (2.2a) is an eigenvector of the matrix $S_2^{-1}S_1$. From the fact that S_1 has a rank of 1, the matrix $S_2^{-1}S_1$ has only one eigenvector. The vector d_2 is calculated under the constraints that d_1 and d_2 be orthonormal.[9] Usually, in image-classification problems, the total number of training images (M) for a K-class problem ($M = \sum_{i=1}^{K} M_i$) is much smaller than the dimension of the image (N). As a result the ($N \times N$) covariance matrix S_2 is at most of rank M and, therefore, is singular. Because S_2 is singular there is no unique solution in N-dimensional space for vector d such that the Fisher ratio $R(d)$ is maximized. However, in a subspace of dimension ($K - 1$) there are unique solutions for the desired vectors d.

When the matrix S_2 is singular, the generalized eigenvector form of a F-S transform [i.e., Eq. (2.5)] can be rewritten as two simultaneous equations:

$$D^T S_1 D = \Theta,$$

$$D^T S_2 D = I, \tag{3.2}$$

where the columns of matrix D are vectors $d_1', d_2', \ldots, d_{K-1}'$, which transform S_2 into an identity matrix in a subspace of dimension $K - 1$ and simultaneously diagonalize S_1. Then choose d_1 with the largest eigenvalue θ, which is equivalent to maximizing the Fisher ratio given by Eq. (2.1). d_2 is calculated by maximizing Eq. (2.1) under the constraints that d_1 and d_2 be orthonormal.[9]

Representing the HTC Transform in a Generalized Eigenvector Form

It was discussed in Refs. 10 and 11 that, in order to maximize the trace of the matrix $S_2^{-1}S_1$, on which the HTC algorithm is based, we must diagonalize simultaneously the two matrices S_1 and S_2:

$$D^T S_1 D = \Theta, \tag{3.3}$$

$$D^T S_2 D = I. \tag{3.4}$$

According to the theorem of Ref. 11, if S_2 is a nonsingular matrix, the desired transformation is determined by the eigenvector matrix of $S_2^{-1}S_1$:

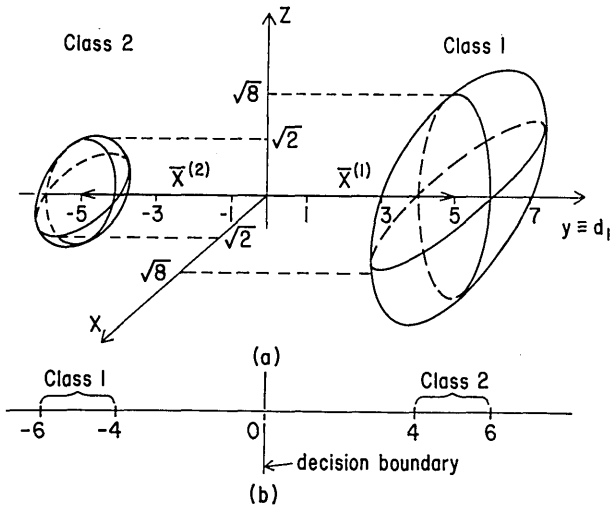

Fig. 1. (a) Graphic illustration of the numerical example for comparison between the F-S and F-K transforms: the two classes have bipolar training data, resulting in positive and negative mean vectors for the first and second classes, respectively. The F-K algorithm, based on the correlation matrices, finds the x and z directions as the best discrimination vectors. The F-S algorithm, based on covariance matrices, finds the y direction as the best discrimination vector (d_1). (b) The F-S decision plane for the example (a).

$$S_2^{-1}S_1D = D\Theta. \tag{3.5}$$

However, Eq. (3.5) can be rewritten as

$$[S_1 - \theta_i S_2]d_i' = 0, \tag{3.6}$$

where d_i' are the column vectors of D and of the corresponding eigenvalues θ_i'. This equation is exactly the same as the generalized eigenvector Eq. (2.5).

Comparing the F-S with the HTC Transform

The first discriminant vector provided by the F-S transform is identical to the one provided by the HTC transform because they are both determined from the generalized eigenvector equation, i.e., Eq. (2.5) or (3.6), and are associated with the largest eigenvalue. Although the F-S and HTC transforms provide the same first discriminant vectors, d_1, the additional discriminant vectors are calculated by quite different approaches. The HTC algorithm provides a set of discriminant vectors determined by the set of eigenvectors of $S_2^{-1}S_1$ associated with the corresponding nonzero eigenvalues. The F-S transform employs a recursive relation in Ref. 8, starting from the first discriminant vector d_1 and calculating a set of orthonormal vectors, which maximize the Fisher ratio of Eq. (2.1).

For the most realistic case of pattern recognition ($\sum_{i=1}^{K} M_i < N$) the F-S transform provides discriminant vectors with more discrimination power than the HTC transform does. For example, for a two-class problem the HTC algorithm will provide only one discriminant vector with discrimination power, because the matrix S_1 is of rank 1. The second discriminant vector provided by the HTC algorithm does not have discrimination power. In contrast, the F-S transform will generally provide more than one discriminant vector with discrimination power. The useful number of discriminant vectors for the F-S transform depends on the statistical properties of the training images. On the other hand, if S_2 is a diagonal matrix of constant elements represented by $S_2 = sI$, where s is a constant, it can be easily verified from Eq. (2.2b) that the vector d_2 is a zero vector.

Representing the F-K Transform in Generalized Eigenvector Form

For the F-K algorithm the discriminant functions are calculated by using images instead of mean-removed images as in the HTC and F-S algorithms. The F-K transform is based on simultaneous diagonalization of two matrices, R and $R^{(1)}$, which is equivalent to finding the extrema of the Rayleigh quotient

$$F(d_i) = \frac{d_i^T R^{(1)} d_i}{d_i^T R d_i}. \tag{3.7}$$

Table 1. Comparison of Eigenvector-Based F-S, HTC, and F-K Algorithms

Characteristics	Algorithms		
	F-S	HTC	F-K
Criterion	$R(d) = \dfrac{d^T S_1 d}{d^T S_2 d}$	$R(d) = \dfrac{d^T S_1 d}{d^T S_2 d}$	$F(d) = \dfrac{d^T R^{(1)} d}{d^T R d}$
Training data base	Mean-removed image vectors	Mean-removed image vectors	Image vectors
d_1	$\alpha_1 S_2^{-1}\Delta$	$\alpha_1 S_2^{-1}\Delta = A_1$	$WE_r\Lambda_r^{-1}\psi_{r1}^{\max}$
d_2	$\alpha_2 S_2^{-1}\left(I - \dfrac{\Delta^T S_2^{-2}\Delta}{\Delta^T S_2^{-3}\Delta} S_2^{-1}\right)\Delta$	A_2	$WE_r\Lambda_r^{-1}\psi_{r1}^{\min}$
Useful for number-class problem	Two-class	Multiple-class (pooled mode and pairwise mode)	Two-class
Receiver-operating-characteristic performance	0.990	0.988	0.999

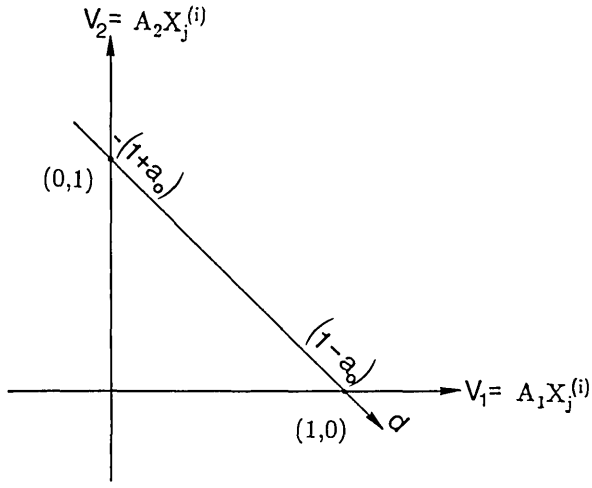

Fig. 2. Decision space (v_1, v_2) for the LSLMT algorithm. If a new decision axis with unit vector d is used, the resultant discriminant function is exactly the vector provided by the F-S algorithm.

In order to maximize Eq. (3.7) we have to find a vector that maximizes the numerator $d_i^T R^{(1)} d_i$ and minimizes the denominator $d_i^T R d_i$. This vector is the generalized eigenvector of $R^{-1}R^{(1)}$ with the maximum eigenvalue.[17] It can be represented as

$$[R^{(1)} - \theta_i R]d_i = 0, \tag{3.8}$$

where θ_i and d_i are the eigenvalue and the eigenvector, respectively. When the matrix R is nonsingular the generalized eigenvalue problem can be transformed to a normal eigenvalue problem:

$$[R^{-1}R^{(1)} - \theta_i I]d_i = 0. \tag{3.9}$$

However, in most practical cases R is a singular matrix, and Eq. (3.8) can be rewritten as two simultaneous equations:

$$D^T R^{(1)} D = \Theta,$$
$$D^T R D = I, \tag{3.10}$$

where D consists of column vectors d_i and Θ is a diagonal matrix of eigenvalues θ_i. Solving these two simultaneous equations, one obtains the discriminant vectors d_1 of Eq. (2.13a) and d_2 of Eq. (2.13b), respectively, that maximize and minimize Eq. (3.7).

The analysis of this section was based on the fact[19] that for a given Hermitian matrix $R^{(1)}$, the stationary (maximum or minimum) values of the Rayleigh quotient $d_i^T B d_i / d_i^T d_i$ can be obtained for each eigenvector d_i of matrix $R^{(1)}$. These stationary values are the corresponding eigenvalues θ_i for $d_i \neq 0$:

$$\theta_i = \frac{d_i^T R^{(1)} d_i}{d_i^T d_i}. \tag{3.11}$$

This theory can be extended[19] to a generalized eigenvector problem by replacing $d_i^T d_i$ with $d_i^T R d_i$ in the denominator of Eq. (3.11), resulting in a Rayleigh quotient $d_i^T R^{(1)} d_i / d_i^T R d_i$, as defined in Eq. (3.7).

Comparing the F-K with the F-S and HTC Transforms

The difference between the F-K and F-S and HTC algorithms arises from the fact that the F-K transform uses training images and correlation matrices to synthesize discriminant vectors but the F-S and HTC algorithms use mean-removed images and covariance matrices. One disadvantage of using correlation matrices is that, if the training data for some reason have negative values, the correlation

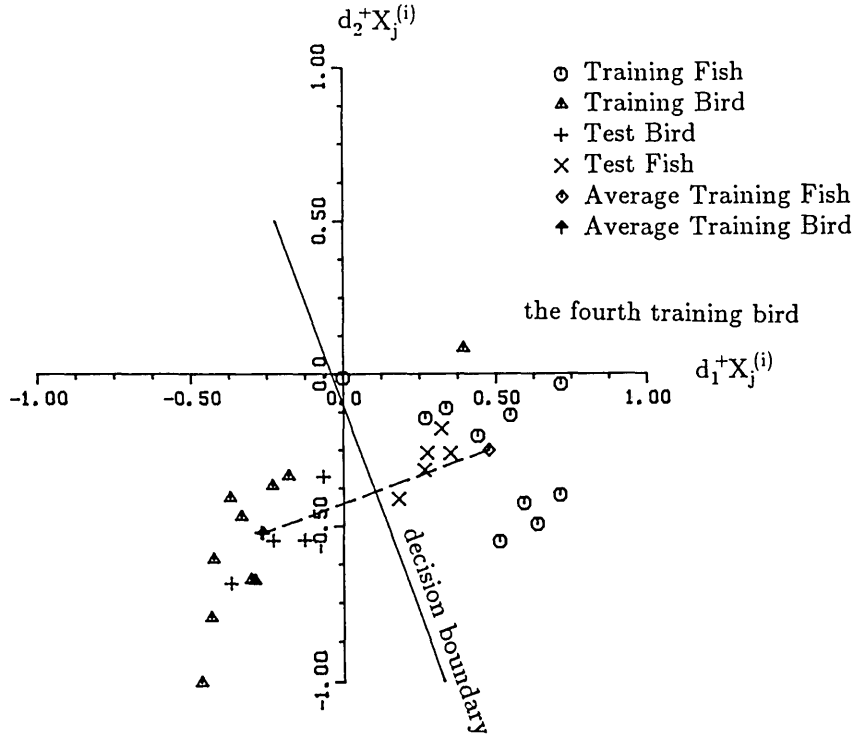

Fig. 3. Classification results from the F-S algorithm when applied to the fish–bird image data base. The decision boundary is defined by a median, normal to the line segment between the average clusters of training fish and training bird.

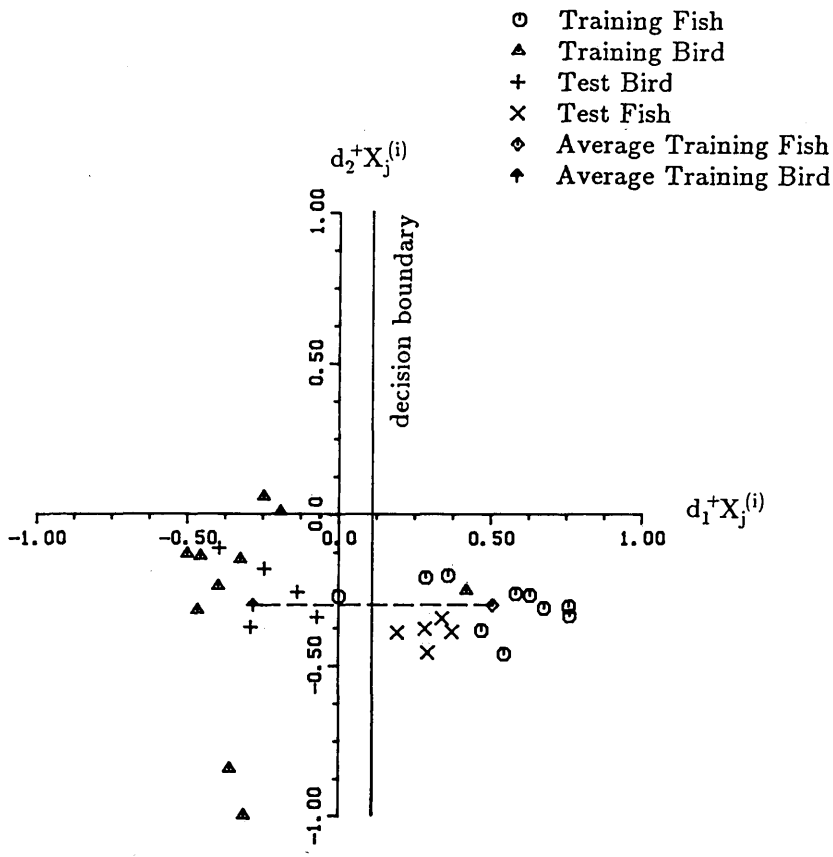

Fig. 4. Classification results from the HTC algorithm when applied to the fish–bird image data base. The decision boundary is defined by a median, normal to the line segment between the average clusters of training fish and training bird.

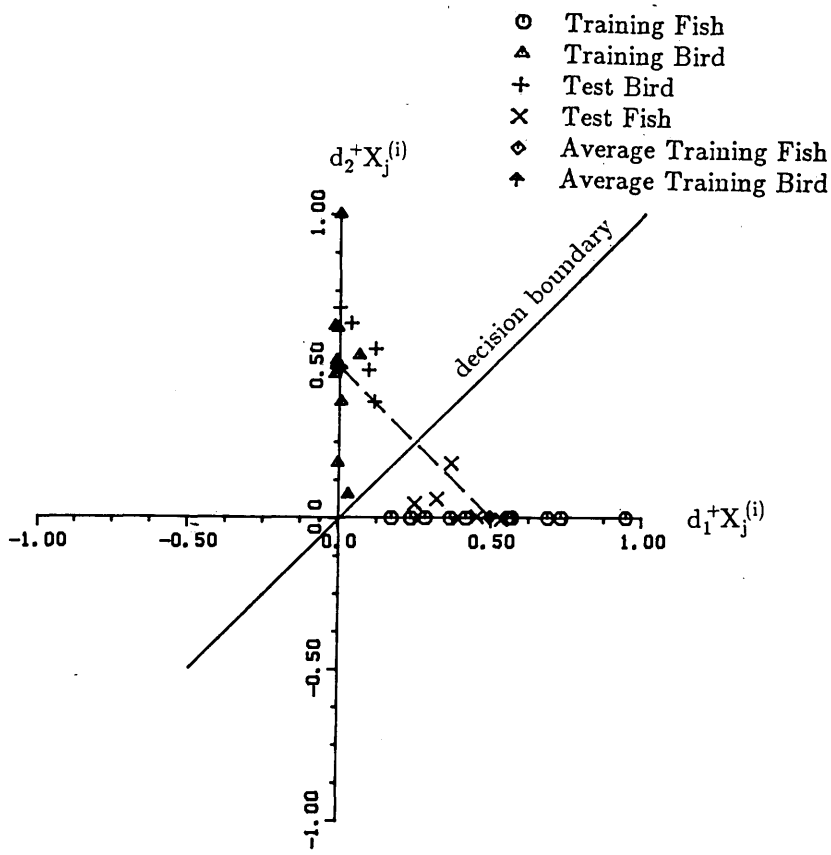

Fig. 5. Classification results from the F-K algorithm when applied to the fish–bird image data base. The decision boundary is defined by a median, normal to the line segment between the average clusters of training fish and training bird.

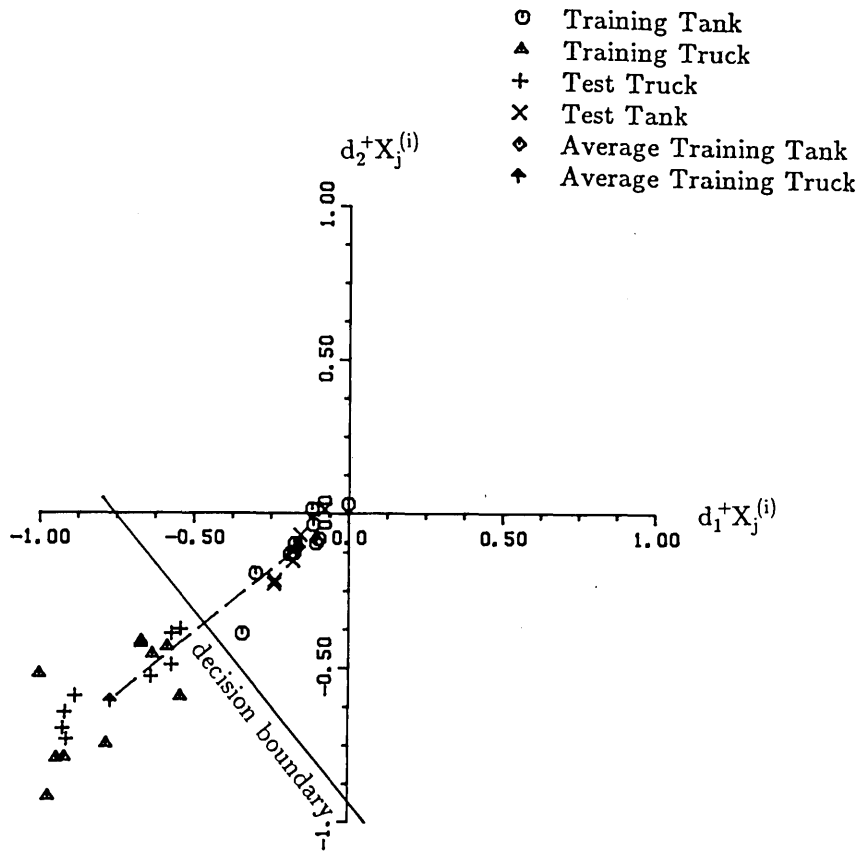

Fig. 6. Classification results from the F-S algorithm when applied to the tank–truck image data base. The decision boundary is defined by a median, normal to the line segment between the average clusters of training tank and training truck.

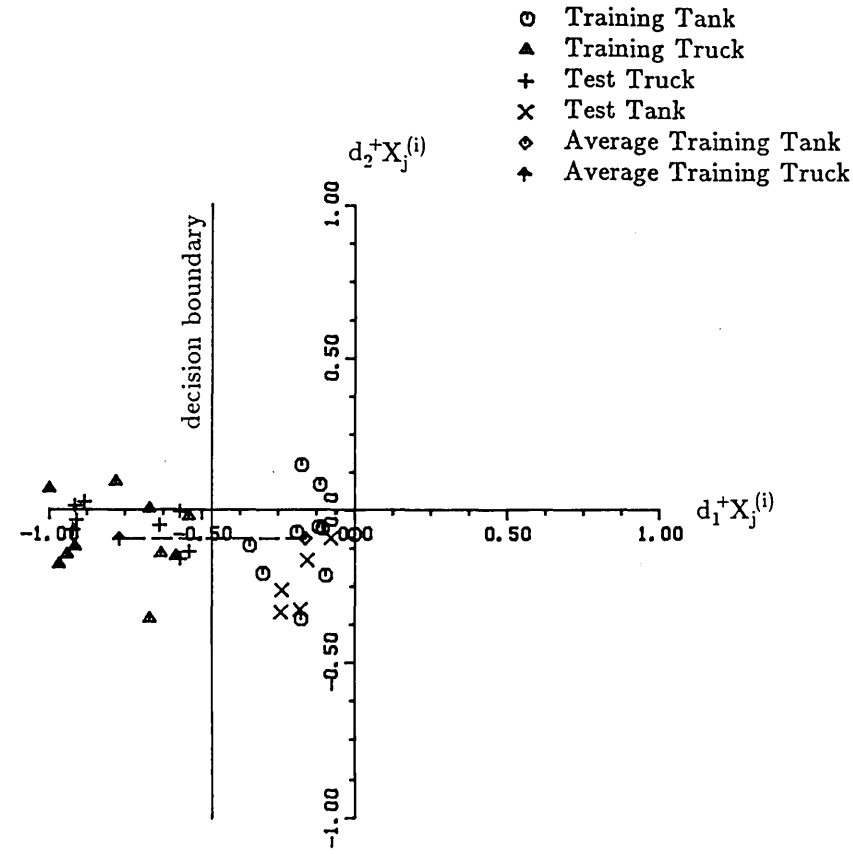

Fig. 7. Classification results from the HTC algorithm when applied to the tank–truck image data base. The decision boundary is defined by a median, normal to the line segment between the average clusters of training tank and training truck.

matrices are not so good a statistical measure as the covariance matrices. For example, in Ref. 8 a two-class classification example is presented: the mean vectors for the two classes are

$$\overline{X}^{(1)^T} = [0\ 5\ 0], \qquad \overline{X}^{(2)^T} = [0\ -5\ 0],$$

the two scatter or covariance matrices are given by

$$S_2^{(1)} = \begin{bmatrix} 8 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 8 \end{bmatrix}, \qquad S_2^{(2)} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix},$$

and the two correlation matrices are given by

$$R^{(1)} = \frac{1}{2}\begin{bmatrix} 8 & 0 & 0 \\ 0 & 26 & 0 \\ 0 & 0 & 8 \end{bmatrix}, \qquad R^{(2)} = \frac{1}{2}\begin{bmatrix} 2 & 0 & 0 \\ 0 & 26 & 0 \\ 0 & 0 & 2 \end{bmatrix}.$$

This numerical example is illustrated graphically in Fig. 1(a). The best discriminant direction seems to be direction y. However, the F-K algorithm chooses the x and z directions as the best discrimination directions because the largest difference between the correlation matrices $R^{(1)}$ and $R^{(2)}$ is along directions x and z. The correlation matrix R tends to measure the magnitude of the amplitude instead of its sign. Therefore the F-K transform cannot provide good discrimination ability when the samples have both positive and negative values, as the example above shows. Since intensities of the image pixels are always positive, this drawback will not affect classification performances in image recognition. Also, in order to overcome this sign-intensitivity problem while retaining the data-fitting capacity of the Karhunen–Loéve transform, modification of the F-K transform by removing the mean images has been proposed.[13,20]

In summary, the F-S, HTC, and F-K algorithms can all be represented in a generalized eigenvector form. There are differences, however, in the way in which they calculate the discriminant vectors; these differences are summarized in Table 1. While F-S and F-K transforms are useful for two-class problems, the HTC is useful for multiple-class ($K \geq 2$) problems.

Comparison with LDF and GMF algorithms

Both the LDF and GMF algorithms are based on computing the discriminant vectors from the generalized eigenvector problem. As has been discussed above, the most practical situation is the underdetermined problem for which the total number of training images (M) is much smaller than the image size (N), i.e., $\sum_{i=1}^{K} M_i \ll N$. In such a situation the pixel scatter matrix S_2 is singular, and the inverse of S_2^{-1} does not exist.

In order to reduce N to ensure the existence of S^{-1} in the application of the LDF algorithm,[14] a wedge-ring sampling detector has been used. The reduction of image space–bandwidth product is sometimes equivalent to some filtering and can affect the classification performances. Also, the use of a wedge-ring detector limits the LDF algorithms to magnitude information only, which may cause the algorithm to be unsuitable for treatment of three-dimensional objects, for which the phase information is important. But the most serious limitation is that the LDF algorithm provides only

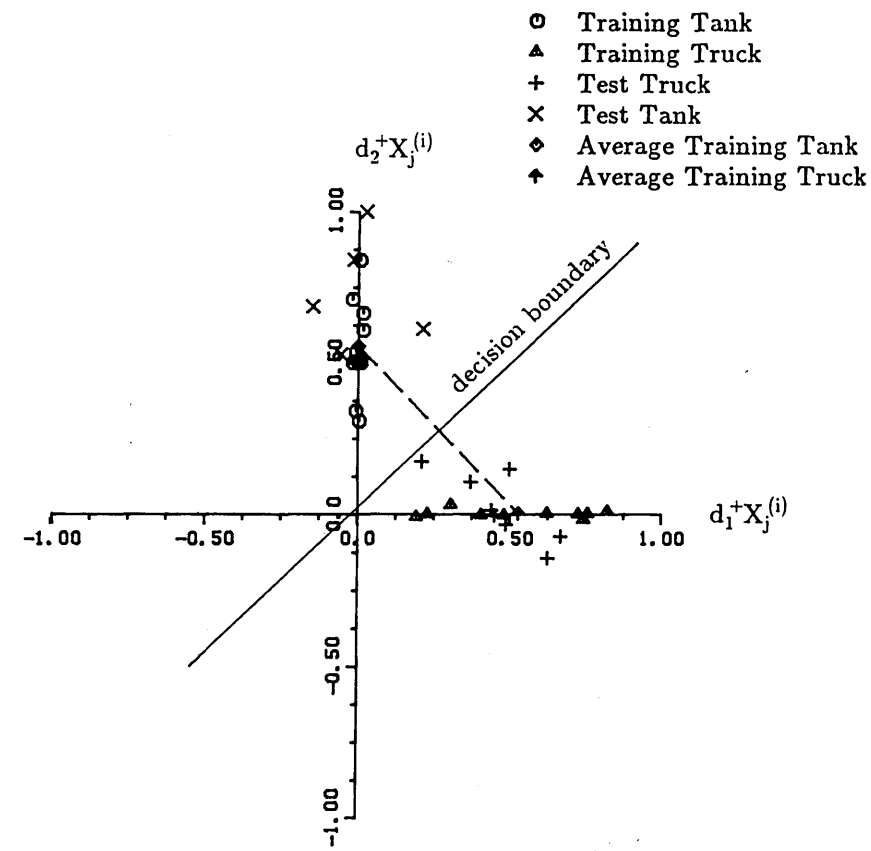

Fig. 8. Classification results from the F-K algorithm when applied to the tank–truck image data base. The decision boundary is defined by a median, normal to the line segment between the average clusters of training tank and training truck.

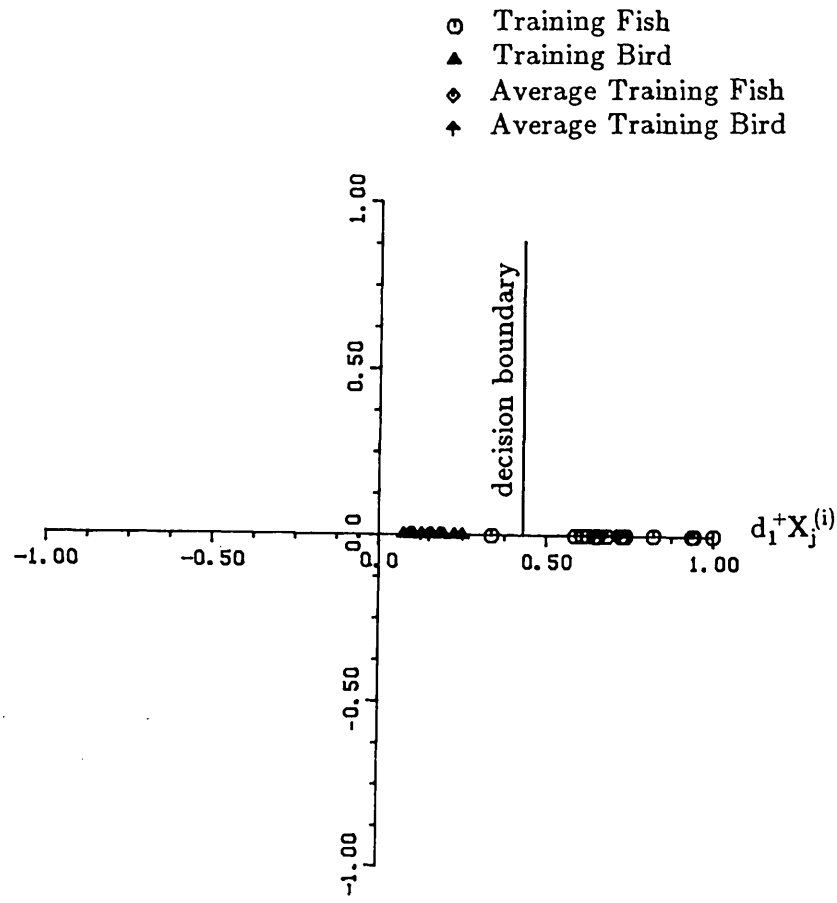

Fig. 9. Classification results from the LDF algorithm when applied to the fish–bird image data base. The decision boundary is defined by a median, normal to the line segment between the average clusters of training fish and training bird.

one discriminant vector. Its classification reliability is therefore not so good as that of the F-S transform.

The formulation of the GMF algorithm takes another approach. It treats a singular matrix S_2 by assuming the pixel covariance matrix S_2 to be diagonal. In general the pixel scatter matrix S_2 is not diagonal, and the off-diagonal elements of the matrix are a measure of cross correlation among the different pixels. In other words, the value of the off-diagonal elements is dependent on the nature of the images to be classified. For the two classes of image defined in Ref. 15 (class 1, airplanes, and class 2, Gaussian noise), the GMF yields good classification performance. The validity of the assumptions in the GMF algorithm, however, depend on the images that must be classified.

4. COMPARISON BETWEEN EIGENVECTOR-BASED AND LINEAR-MAPPING ALGORITHMS

The relation between the LSLMT and SDF algorithms was investigated in Part I.[1] Now we show that under a certain choice of the decision space for two-class problems the two groups of algorithms, eigenvector based and linear mapping based, are directly related, although these two groups of algorithms use different routes to synthesize the discriminant vectors.

The decision space for the LSLMT algorithm is defined by the decision axes v_1 and v_2 for a two-class problem, as shown in Fig. 2. All the training images are mapped by the discriminant vectors provided by the LSLMT algorithm into two points (1, 0) and (0, 1) for classes 1 and 2, respectively. Assume that a new axis with unit vector d is used for classification; the direction and the magnitude of vector d are shown in Fig. 2. The two points mapped by such a linear-mapping algorithm will have the new coordinates $(1 - a_0)$ and $-(1 + a_0)$ in the coordinate-transformed decision space with the new axis. We shall see that this special choice of the decision space links the minimum-squared-error solution to the eigenvector-based solution (Fisher's linear discriminant[14]).

The mathematics associated with coordinate transforming the LSLMT decision space is provided in Appendix C and Ref. 17. The discriminant vector for the minimum-squared-error solution under such coordinate transformation becomes

$$A_0^T = M\left(1 - \frac{\alpha}{2}\right) S_2^{-1}\Delta. \tag{4.1}$$

Under this special choice for the decision space the minimum-squared-error discriminant function is the same as the first discriminant vector from the F-S algorithm [see Eq. (2.2a)] except for an unimportant scale factor difference. Therefore under the choice of decision space discussed, the two groups of algorithms are closely related.

It must be also noted that although both the linear-mapping-based algorithms and the eigenvector-based algorithms minimize the mean-squared deviations, the goals of

the two algorithms are quite different. The eigenvector-based algorithms in their overdetermined and underdetermined cases provide discriminant vectors, which ensure that the mean-squared deviation of the points about their cluster centers is minimized with respect to the separation of the cluster centers. Linear-mapping-based algorithms have a similar behavior in their overdetermined case under the assumption of the special choice of decision space introduced above. However, in their underdetermined case, the linear-mapping-based algorithms will provide an exact minimum-norm solution, which will map each of the training images into the centers of cluster, while the test images will be mapped into points scattered around the cluster centers. The choice of the minimum-norm solution is optimal in minimizing the deviation (scattering) of this point from the cluster center.

5. EXPERIMENTAL RESULTS AND ANALYSIS

Two image data bases are used to test the eigenvector-based algorithms. The first image data base consists of 10 fish, 10 birds for training, and 10 new fish and birds for testing, as shown in Fig. 2 of Part I.[1] The second image data base consists of 10 tanks, 10 trucks for training, and 13 new tanks and trucks for testing, as shown in Fig. 3 of Part I.[1] All the images have the 64 × 64 pixels and 8 bits of gray-level intensity, from 0 to 255. The classification results from the first image data base are shown in Figs. 3, 4, and 5 for the F-S, HTC, and F-K algorithms, respectively. The decision boundary is determined by a median, normal to a line segment between the average clusters of training fish and training birds. All the training images and test images are correctly classified except the fourth bird in the training set, which is misclassified by the F-S and the HTC algorithms (in fact the HTC also misclassified one training fish). The reason for the misclassification is that the number of the training images is much smaller than the dimension of the images. Using more training images or reducing the image resolution will improve the classification performance (see Ref. 9). Also, it has been observed that d_1 of the F-S algorithm is exactly the same as the first basis function of the HTC algorithm. However, while the second discriminant vector d_2 of the F-S algorithm possesses discriminant power, the d_2 of the HTC algorithm does not, which contributes to the vertical decision boundary in Fig. 4. The classification results from the second image data base are shown in Figs. 6, 7, and 8 for the F-S, HTC, and F-K algorithms, respectively. All the images are correctly classified.

To compare the performances of the algorithms quantitatively, we have used the experimental data of Figs. 3–5 and 6–8 to generate receiver-operating-characteristic curves[22] by varying the decision threshold for F-S, HTC, and F-K discriminants. The area under the receiver-operating-characteristics curves provides a direct quantitative performance measure of each algorithm, as summarized in Table 1. From the receiver-operating-characteristics performances of Ta-

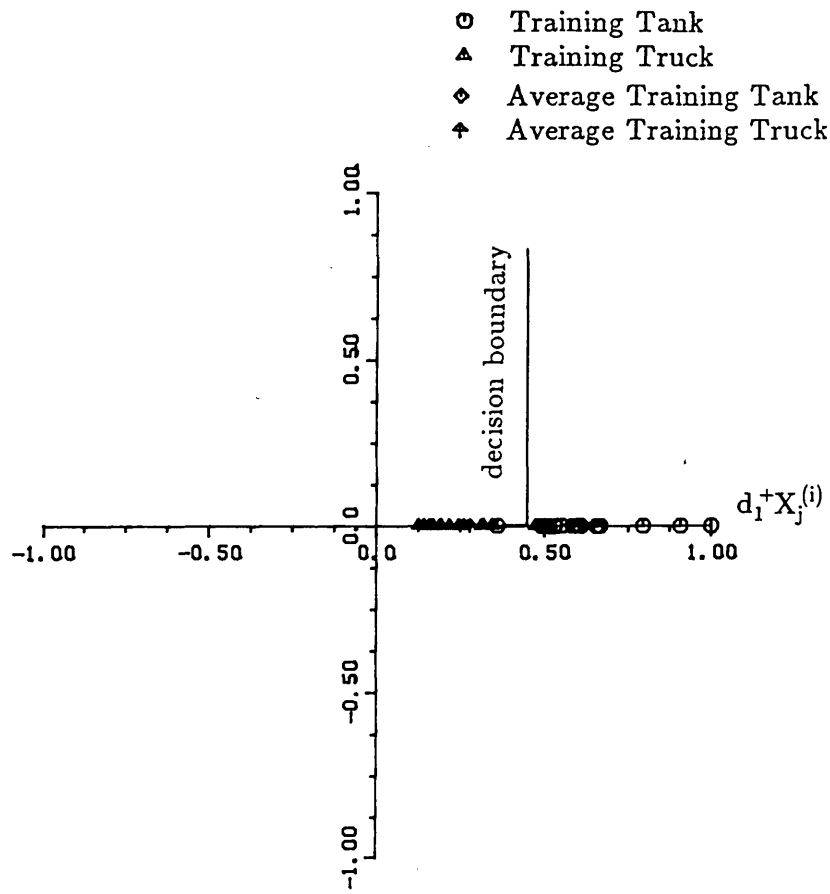

Fig. 10. Classification results from the LDF algorithm applied to the tank–truck image data base. The decision boundary is defined by a median, normal to the line segment between the average clusters of training tank and training truck.

ble 1 we conclude that the probability of the false alarm is low for all the algorithms compared (i.e., the FS, HTC, and F-K transforms).

In order to apply the LDF algorithm we have to reduce the size of the training images from $N = 4096$ (64×64) to $N \simeq 2M = 30$. Under such conditions we can ensure the existence of S_2^{-1}. The format of the training data has been arranged similarly to that described in Ref. 14: the Fourier transforms of training or test images, each of 64×64 pixels, were used to calculate the responses of a wedge-ring detector, consisting of 15 rings and 15 wedges. The LDF discriminant vectors have been calculated from Eq. (2.15) and were used later to generate the classification results shown in Figs. 9 and 10 for the fish–bird and tank–truck data bases, respectively. We observe that for the fish–bird data base one training fish is misclassified, while for the tank–truck data base one training tank and one training truck are misclassified.

In order to apply the GMF algorithm we calculate the pixel scatter matrix S_2 for our fish–bird data base; the result is given in Eq. (5.1), which shows that ignoring the off-diagonal elements of this scatter matrix for our data base is a rough approximation. Therefore the GMF algorithm cannot be employed for classification on our data base:

$$S_2 = \begin{bmatrix} \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & 0.22 & 0.12 & 1.0 & 0.21 & 0.42 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & 0.06 & 0.2 & 0.3 & 0.26 & 0.06 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & 0.96 & 0.16 & 0 & 0.06 & 0.01 & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{bmatrix}. \quad (5.1)$$

In summary, for the given two sets of image data bases, the F-S, HTC, and F-K algorithms provide comparably good classification. In terms of cluster analysis of the given training images, it appears that the eigenvector-based algorithms may have some advantages over the linear-mapping-based algorithms because the former, especially the F-S algorithm, can be used to observe the nature of the cluster. This is useful in determining whether the number of training images is sufficient and in designing nonlinear piecewise decision logic. On the other hand, the linear-mapping-based algorithms always map training images to fixed points in the decision plane.

6. CONCLUSIONS

The analysis has shown that the eigenvector-based algorithms can be represented in a generalized eigenvector form. However, these algorithms differ in the way in which they employ the correlation matrices or covariance matrices to calculate the discriminant vectors. Summaries of methods of calculating the discriminant vectors for the F-S, HTC, and F-K algorithms have been provided. Especially for the most practical underdetermined case ($KM \ll N$), the calculations usually require the inversion of a large singular pixel correlation (or covariance) matrix. This problem is replaced by finding the pseudoinverse of this matrix, which requires inverting only the smaller nonsingular image-correlation (or image-covariance) matrix plus multiplying several nonsingular matrices. In contrast, the LDF algorithm requires reducing the number of training features (reduced space–bandwidth product) in order to ensure that the pixel scatter matrix will be nonsingular. The performance of the GMF algorithm is dependent on the pixel correlation strength of the training images. Therefore the classification performances of both LDF and GMF algorithms are more dependent on the nature of the training data.

Also, under a certain choice of decision axis the LSLMT algorithm (linear mapping based) for the overdetermined case provides exactly the same first discriminant vector as the F-S and HTC algorithms (eigenvector based). Classification experiments on the two sets of image data bases show that the F-S, HTC, and F-K algorithms provide comparable classification performance, while the F-S algorithm provides more information on the clustering of the training images; this is valuable for an interactive pattern-recognition environment to determine statistical properties of training images and to design piecewise decision logic.

APPENDIX A: LIST OF SYMBOLS

d_1	First discriminant vector for the F-S and HTC algorithms.
d_2	Second discriminant vector for the F-S algorithm.
d	Discriminant vector.
Δ	Difference vector between the means of two classes.
$R(d)$	Fisher ratio for the F-S algorithm.
E_r, Λ_r	Eigenvectors and eigenvalues of the matrix R_{IM} (same as Ψ and Λ in Ref. 2 and E and Λ in Ref. 11).
E_s, Λ_s	Eigenvectors and eigenvalues of the matrix S_{IM} (same as E and θ in Refs. 9 and 10).
Ψ_r, Λ_r	Eigenvectors and eigenvalues of the matrix R.
Ψ_s, Λ_s	Eigenvectors and eigenvalues of the matrix $(\Lambda_s^{-1} E_s^T W_s^T) S_1 (W_s E_s \Lambda_s^{-1})$ (same as Φ and Γ in Ref. 9 and U and θ in Ref. 10).
D, Θ	Generalized eigenvectors and eigenvalues of the matrix $S_2^{-1} S_1$ (same as A and Λ in Ref. 10).
Ψ_{r1}, Γ_r	Eigenvectors and eigenvalues of the matrix $\Lambda_r^{-1/2} \Psi_r^T R^{(1)} \Psi_r \Lambda_r^{-1/2}$ (same as Ψ and Γ in Ref. 11).
Ψ_{r2}, K_r	Eigenvectors and eigenvalues of the matrix $\Lambda_r^{-1/2} \Psi_r^T R^{(2)} \Psi_r \Lambda_r^{-1/2}$ (same as Θ and M in Ref. 11).

APPENDIX B: PSEUDOINVERSE OF THE COVARIANCE MATRIX

In practice the covariance matrix

$$S_2 = W_s W_s^T \quad \text{(B1)}$$

is singular because the rank of the matrix W is at most KM ($<N$), and therefore the inverse S_2^{-1} does not exist. However, for our applications it will be useful to find a pseudoinverse of S_2 that has certain properties of the inverse matrix. By definition[12] a matrix S_2^+ is a pseudoinverse of a matrix S_2 if

$$S_2 S_2^+ S_2 = S_2, \quad \text{(B2)}$$

$$S_2^+ = U S_2^T = S_2^T V, \quad \text{(B3)}$$

where U and V are some certain matrices. To find the pseudoinverse of matrix S_2 we first can find the pseudoinverse of matrices W_s and W_s^T of Eq. (B1). By definition again

$$W_s W_s^+ W_s = W_s \tag{B4}$$

$$W_s^+ = \tilde{U} W_s^T,$$

or, combining these two equations,

$$W_s \tilde{U} W_s^T W_s = W_s. \tag{B5}$$

Multiplying Eq. (B5) from the left by W_s^T, we get

$$W_s^T W_s \tilde{U} W_s^T W_s = W_s^T W_s,$$

and solving for $\tilde{U}$ yields

$$\tilde{U} = (W_s^T W_s)^{-1}, \tag{B6}$$

where we have assumed that $(W_s^T W_s)$ is not an irregular, square matrix and has an inverse. From Eqs. (B4) we find the pseudoinverse:

$$W_s^+ = (W_s^T W_s)^{-1} W_s^T. \tag{B7}$$

Similarly, we can find the pseudoinverse of the matrix W_s^T of Eq. (B1), resulting in

$$(W_s^T)^+ = W_s (W_s^T W_s)^{-1}. \tag{B8}$$

Finally, the pseudoinverse S_2^+ given by Eq. (B1) can be rewritten as

$$\begin{aligned} S_2^+ &= (W_s W_s^T)^+ = (W_s^T)^+ W_s^+ \\ &= W_s (W_s^T W_s)^{-1} (W_s^T W_s)^{-1} W_s^T \\ &= W_s (W_s^T W_s)^{-2} W_s^T = W_s S_{\mathrm{IM}}^{-2} W_s^T, \end{aligned} \tag{B9}$$

where we have used Eqs. (B7) and (B8) and the definition of the image-covariance matrix. Equation (B9) can be rewritten in its alternative form expressed in terms of eigenvectors E_s and eigenvalues Λ_s of S_{IM}:

$$S_2^+ = W_s E_s \Lambda_s^{-2} E_s^T W_s^T. \tag{B10}$$

Equation (B10) is identical to the solution used by the authors of Ref. 9.

APPENDIX C: RELATION BETWEEN LINEAR-MAPPING AND EIGENVECTOR-BASED ALGORITHMS

The decision space for the LSLMT algorithm for a two-class problem is defined by the axes v_1 and v_2 as shown in Fig. 2. All the training images are mapped by the LSLMT algorithm into two points (1, 0) and (0, 1) for classes 1 and 2, respectively. If a new axis is defined with its direction unit vector d shown in Fig. 2, the two points mapped by the linear-mapping algorithm will have the new coordinates $(1 - a_0)$ and $-(1 + a_0)$. The $(1 \times N)$ linear transformation A_0, which maps each of the images of the first and second class into the points $(1 - a_0)$ and $-(1 + a_0)$, respectively, is

$$A_0 W = [(1 - a_0), \ldots (1 - a_0), -(1 + a_0), \ldots -(1 + a_0)], \tag{C1}$$

where the right-hand side of Eq. (C1) consists of M elements $(1 - a_0)$ and M elements $-(1 + a_0)$ and a_0 is a scalar that is calculated below.

To determine A_0 and a_0 we find it convenient to rewrite Eq. (C1) by moving a_0 from the right-hand to the left-hand side.

$$A' W' = [1, \ldots 1, -1, \ldots -1], \tag{C2}$$

where we have defined a $1 \times (N + 1)$ vector A' and an $(N + 1) \times 2M$ matrix W':

$$A' = [a_0, A_0]. \tag{C3}$$

$$W' = \begin{bmatrix} 1 & \ldots & 1 & 1 & \ldots & 1 \\ X_1^{(1)} & \ldots & X_M^{(1)} & X_1^{(2)} & \ldots & X_M^{(2)} \end{bmatrix}. \tag{C4}$$

We will generate the minimum-squared-error solution for the linear algebra problem of Eq. (C2) for an overdetermined case. Transposing both sides of Eq. (C2) and multiplying both sides of the result by a matrix W', we obtain

$$W' W'^T A'^T = W' \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ -1 \\ \vdots \\ -1 \end{bmatrix}. \tag{C5}$$

Using the definition of Eq. (C4), we rewrite the right-hand side of Eq. (C5):

$$W' W'^T A'^T = \begin{bmatrix} 0 \\ M\Delta \end{bmatrix}, \tag{C6}$$

where the mean difference vector Δ for a two-class problem was defined by Eq. (2.5f) of Part I.[1] Using the definition of Eq. (C4), the matrix $W' W'^T$ can be rewritten as

$$W' W'^T = \begin{bmatrix} 2M & 2M\overline{X}^T \\ 2M\overline{X} & WW^T \end{bmatrix}, \tag{C7}$$

where the mean image $\overline{X}$ has been defined by Eq. (2.2) of Part I and the correlation matrix WW^T is related to the scatter matrix S_2 by

$$\begin{aligned} WW^T &= W_s W_s^T + M(\overline{X}^{(1)} \overline{X}^{(1)^T} + \overline{X}^{(2)} \overline{X}^{(2)^T}) \\ &= S_2 + M(\overline{X}^{(1)} \overline{X}^{(1)^T} + \overline{X}^{(2)} \overline{X}^{(2)^T}). \end{aligned} \tag{C8}$$

Substituting Eq. (C8) into Eq. (C7), and Eq. (C7) into Eq. (C6), we obtain

$$\begin{bmatrix} 2M & 2M\overline{X}^T \\ 2M\overline{X} & S_2 + M(\overline{X}^{(1)} \overline{X}^{(1)^T} + \overline{X}^{(2)} \overline{X}^{(2)^T}) \end{bmatrix} \begin{bmatrix} a_0 \\ A_0^T \end{bmatrix} = \begin{bmatrix} 0 \\ M\Delta \end{bmatrix}. \tag{C9}$$

Equation (C9) is the set of linear equations from which the unknown values of the scalar a_0 and $(1 \times N)$ transformation vector A_0 can be calculated.

Calculating a_0

From the first of the set of equations in Eq. (C9), we obtain

$$2Ma_0 + 2M\overline{X}^T A_0^T = 0.$$

The scalar value a_0 is then calculated to be

$$a_0 = -\overline{X}^T A_0^T. \tag{C10}$$

Calculating the Transformation A_0

The remaining linear equations in Eq. (C9) can be rewritten as

$$2M\overline{X}a_0 + [S_2 + M(\overline{X}^{(1)}\overline{X}^{(1)^T} + \overline{X}^{(2)}\overline{X}^{(2)^T})]A_0{}^T = M\Delta. \tag{C11}$$

Substituting a_0 from Eq. (C10) into Eq. (C11) results in

$$-2M\overline{X}\overline{X}^TA_0{}^T + S_2A_0{}^T + M(\overline{X}^{(1)}\overline{X}^{(1)^T} + \overline{X}^{(2)}\overline{X}^{(2)^T})A_0{}^T = M\Delta. \tag{C12}$$

By using the definition of $\overline{X}$, $\overline{X}\overline{X}^T$ can be expressed in terms of $\overline{X}^{(1)}$ and $\overline{X}^{(2)}$:

$$\overline{X}\overline{X}^T = \frac{\overline{X}^{(1)} + \overline{X}^{(2)}}{2}\frac{\overline{X}^{(1)^T} + \overline{X}^{(2)^T}}{2}$$

$$= \tfrac{1}{4}(\overline{X}^{(1)}\overline{X}^{(1)^T} + \overline{X}^{(2)}\overline{X}^{(1)^T} + \overline{X}^{(1)}\overline{X}^{(2)^T} + \overline{X}^{(2)}\overline{X}^{(2)^T}).$$

Substituting this expression for $\overline{X}\overline{X}^T$ into Eq. (C12), we obtain

$$S_2A_0{}^T + M(\tfrac{1}{2}\overline{X}^{(1)}\overline{X}^{(1)^T} + \tfrac{1}{2}\overline{X}^{(2)}\overline{X}^{(2)^T} - \tfrac{1}{2}\overline{X}^{(1)}\overline{X}^{(2)^T} - \tfrac{1}{2}\overline{X}^{(2)}\overline{X}^{(1)^T})A_0{}^T = M\Delta.$$

Using the definition of the mean difference vector (Δ), we finally rewrite the above equation as

$$S_2A_0{}^T + \tfrac{1}{2}M\Delta\Delta^TA_0{}^T = M\Delta. \tag{C13}$$

Because $\Delta^TA_0{}^T$ is scalar,

$$\alpha \triangleq \Delta^TA_0{}^T, \tag{C14}$$

Eq. (C13) can be solved for the desired transformation vector

$$A_0{}^T = M\left(1 - \frac{\alpha}{2}\right)S_2{}^{-1}\Delta. \tag{C15}$$

A_0 is the same as the first discriminant vector from the F-S algorithm [see Eq. (2.2a)] except for an unimportant scale-factor difference. Therefore the two algorithms are closely related. Note that, in general, the minimum-squared-error solution to the overdetermined problem is strongly dependent on the choice of the decision space vectors.[17] Indeed, the decision space design described above leads the two groups of algorithms, the linear-mapping and the eigenvector-based algorithms, to provide the same discriminant vectors in their overdetermined case.

ACKNOWLEDGMENTS

The authors are grateful for the research support of the National Science Foundation under grant ECS-8303107, the U.S. Air Force Office of Scientific Research under grant AFOSR-85-0371, and the Rome Air Development Center, Hanscom Field, under grant F-19628-85-K-0039.

REFERENCES

1. Q. Tian, Y. Fainman, Z. H. Gu, and S. H. Lee, "Comparison of statistical pattern-recognition algorithms for hybrid processing. I. Linear-mapping algorithms," J. Opt. Soc. Am. **5,** 1655–1669 (1988).
2. Z. H. Gu, J. R. Leger, and S. H. Lee, "Optical implementation of the least-square linear mapping technique for image classification," J. Opt. Soc. Am. **72,** 6, 787–793 (1982).
3. Z. H. Gu and S. H. Lee, "Recognition of images of Markov-1 model by least-squares linear mapping technique," Appl. Opt. **23,** 882–827 (1984).
4. D. Casasent, W. Rozzi, and D. Fetterly, "Projection synthetic discriminant function performance," Opt. Eng. **23,** 716–720 (1984).
5. B. V. K. V. Kumar, E. Pochapsky, and D. Casasent, "Optimality considerations in modified matched spatial filters," in *Analog Optical Processing and Computing,* H. J. Caulfield, ed., Proc. Soc. Photo-Opt. Instrum. Eng. **519,** 85–93 (1984).
6. H. J. Caulfield and W. T. Maloney, "Improved discrimination in optical character recognition," Appl. Opt. **8,** 2354–2356 (1969).
7. H. J. Caulfield and R. Haimes, "Generalized matched filtering," Appl. Opt. **19,** 181–183 (1980).
8. D. H. Foley and J. W. Sammon, Jr., "An optimal set of discriminant vectors," IEEE Trans. Comput. **C-24,** 281–289 (1975).
9. Q. Tian, M. Barbero, Z. H. Gu, and S. H. Lee, "Image classification by Foley–Sammon transform," Opt. Eng. **25,** 834–840 (1986).
10. Z. H. Gu and S. H. Lee, "Optical implementation of the Hotelling trace criterion for image classification," Opt. Eng. **23,** 727–731 (1984).
11. K. Fukunaga, *Introduction to Statistical Pattern Recognition* (Academic, New York, 1972), pp. 34–35, 260–264.
12. K. Fukunaga and W. L. Koontz, "Application of the Karhunen–Loéve expansion to feature selection and ordering," IEEE Trans. Comput. **C-19,** 311–318 (1970).
13. J. R. Leger and S. H. Lee, "Image classification by an optical implementation of the Fukunaga–Koontz transform," J. Opt. Soc. Am. **72,** 556–564 (1982).
14. D. Casasent and W. T. Chang, "Generalized chord transformation for distortion-invariant optical pattern recognition," Appl. Opt. **22,** 2087–2094 (1983).
15. H. J. Caulfield and M. H. Weinberg, "Computer recognition of 2-D patterns using generalized matched filters," Appl. Opt. **21,** 1699–1704 (1982).
16. R. A. Fisher, "The use of multiple measurements in taxinomic problems," in *Contributions to Mathematical Statistics,* R. A. Fisher, ed. (Wiley, New York, 1950), pp. 32.179–32.188.
17. R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis* (Wiley-Interscience, New York, 1973), pp. 15, 117, 152–154.
18. H. C. Andrews, *Introduction to Mathematical Techniques in Pattern Recognition* (Wiley-Interscience, New York, 1972), pp. 67–74.
19. G. W. Stewart, *Introduction to Matrix Computations* (Academic, New York, 1973), p. 312.
20. P. Sanyal and D. H. Fole, "Feature selection by a modified Fugunaga-Koontz transform and its graphic interpretation," in *Proceedings of the IEEE Milwaukee Symposium on Automatic Computation and Control* (Institute of Electrical and Electronics Engineers, New York, 1976), pp. 445–452.
21. F. R. Gantmakher, *The Theory of Matrices* (Chelsea, New York, 1959), pp. 32–40.
22. H. L. Van Trees, *Detection, Estimation, and Modulation Theory* (Wiley, New York, 1968), pp. 19–52.