IEEE Trans. Pattern Anal. Machine Intell., vol. PAMI-3, no. 5, Sept. 1981.

[22] J. G. Postaire and C. Vasseur, "A fast algorithm for nonparametric probability density estimation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-4, no. 6, pp. 663-666, 1982.

[23] T. Cacoullos, "Estimation of a multivariate density," *Ann. Inst. Statist., Math.*, vol. 18, pp. 179-189, 1965.

[24] F. R. Fromm and R. A. Northouse, "Some results on nonparametric clustering of large data problems," *Proc. 1st Joint Conf. Pattern Recognition*, 1973, pp. 18-21.

[25] D. J. Eigen, F. R. Fromm, and R. A. Northouse, "Cluster analysis based on dimensional information with applications to feature selection and classification," *IEEE Trans. Syst., Man., Cybern.*, vol. SMC-4, pp. 284-294, May 1974.

[26] J. G. Postaire, "Une approche statistique unique pour l'analyse des mélanges et la détection des modes," *Rev. Statist. Appl.*, vol. 31, no. 4, pp. 17-36, 1983.

[27] A. H. Bock, *Automatishe Klassifikation*. Göttingen: Vandenhoeck & Rubrecht, 1974.

[28] J. A. Hartigan, "Clusters as modes," in *Proc. 1st Int. Symp. Data Analysis and Information*, Versailles, France, 1977.

[29] ——, "Distribution problems in clustering," in *Classification & Clustering*, J. Van Rysin, Ed. New York: Academic, 1977 pp. 45-71.

[30] B. Kopp, "Ein neuer Verfahren zur cluster analyse," *Biometrishe Meitshrift*, vol. 18, pp. 473-476, 1976.

[31] A. Rosenfeld and A. C. Kak, *Digital Picture Processing*, vol. 2. New York: Academic, 1982.

[32] J. O. Eklundh and A. Rosenfeld, "Convergence properties of relaxation," Comput. Sci. Center, Univ. Maryland, Rep. TR-701, Oct. 1978.

[33] S. W. Zucker, E. V. Krishnamurthy, and R. L. Haar, "Relaxation processes for scene labeling: Convergence, speed, and stability," *IEEE Trans. Syst., Man., Cybern.*, vol. SMC-8, no. 1, pp. 41-48, Jan. 1978.

[34] G. Fekete, J. O. Eklundh, and A. Rosenfeld, "Relaxation: Evaluation and applications," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-3, no. 4, pp. 459-469, July 1981.

[35] J. G. Postaire and C. P. A. Vasseur, "An approximate solution to normal mixture identification with application to unsupervised pattern classification," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-3, no. 2, Mar. 1981.

An Optimal Transformation for Discriminant and Principal Component Analysis

J. DUCHENE AND S. LECLERCQ

Abstract—**This correspondence proposes a general method to describe multivariate data sets by two different ways: discriminant analysis and principal component analysis. First the problem of finding K discriminant vectors in an L-class data set is solved, and compared to the solution proposed in the literature for two-class problems, and the classical solution for L-class data sets. It is shown that the method proposed in the paper is better than the classical method for L classes and that this method is a generalization of the optimal set of discriminant vectors proposed for two-class problems. Then the method is combined with a generalized principal component analysis to permit the user to define the properties of each successive computed vector. All the methods are tested using IRIS data.**

Manuscript received February 18, 1986; revised January 7, 1988. Recommended for acceptance by J. Kittler.

The authors are with the Department of Biomedical Engineering, Compiégne University, BP 649, 60206, Compiégne Cedex, France.

IEEE Log Number 8823600.

Index Terms—**Discriminant analysis, eigenvectors, feature extraction, multivariate data, pattern recognition, principal component analysis.**

I. Introduction

Pattern recognition problems often can be divided into two parts, feature extraction and classification. Many classification methods are developed in the literature in two different ways. The first set of methods works on the raw pattern matrix without any consideration to specific properties in the initial data set [1]-[3]. The objective is to find significant clusters using all the available measurements, i.e., all the parameters defining the objects or events. Then each cluster can be represented by a model, for instance by the centroid and the within-class covariance, and additional objects are classified using these characteristics. The second type of classification methods uses the pattern class information [4]. For example, measurements are made on different kinds of flowers (IRIS data).

A transformation can be applied on this pattern matrix in order to find the optimal discriminant directions. Then the classification of additional objects can be made in this optimal subspace.

The principle subject of this correspondence is the transformation of a high-dimensional pattern matrix either for discriminating various categories of observed objects or for finding a relevant subspace of vectors. The first case corresponds to finding an optimal set of discriminant vectors, the second permits to reduce the dimension of high-dimensional data without losing significant scatter. These transformations are well known, the first as discriminant analysis [5] and the second as principal component analysis [6]. The aim of this paper is to combine the transformation methods in order to obtain information on discrimination and scatter in the same set of vectors. Furthermore, it will be proved that the classical solution for L-class discrimination problems is not necessarily the best. We provide a direct analytic solution to this problem, unlike the iterative method proposed by Okada and Tomita [7]. All the proposed methods will be illustrated using the IRIS data set [8], [9].

II. Discriminant Analysis

The objective of this method is to find the best set of discriminant vectors in order to separate predefined classes of objects or events. Each object is represented by a set of raw measurements, the number of which is frequently quite large. The most commonly used method is to obtain the vector d such that the ratio of the between-class variance to the within-class or the total projected variance is maximized [14]. Such a criterion can be expressed as

$$C_1 = \frac{d^t Bd}{d^t Td} \qquad \text{or} \qquad C_2 = \frac{d^t Bd}{d^t Wd}$$

where

B = between-class covariance matrix
W = within-class covariance matrix
T = total covariance matrix.

The two criteria C_1 and C_2 are equivalent due to the following relationship between the scatter matrices

$$T = W + B.$$

For instance, using the C_2 criterion, the best discriminant vector d_1 is provided by the expression

$$W^{-1}Bd_1 = \lambda d_1.$$

Here, d_1 is the eigenvector of $W^{-1}B$ associated with the largest eigenvalue.

If L classes were defined, it can easily be proved that at most $L - 1$ eigenvectors exist. Most of the authors choose these $L - 1$ eigenvectors as the best set of discriminant vectors [5]. However, some authors [7] showed that an orthogonal set of vectors was more powerful than the classical discriminant vectors, in terms of both discriminant ratio and mean error probability. The first vector is obtained by directly maximizing the criterion (C_1 or C_2): it is well known as the Fisher linear discriminant [8], but for classical analysis, the constraint used to compute the second discriminant vector d_2 is expressed as

$$d_2' W d_1 = 0 \quad \text{or} \quad d_2' T d_1 = 0.$$

In fact, the transformation is W- or T-orthogonal, and the example with IRIS data will show that the discrimination ratio for each discriminant vector d_n, defined as

$$\frac{d_n' B d_n}{d_n' T d_n}$$

can be greater with the proposed orthogonal transformation described below. For the two-class case, only one discriminant vector can be computed, but Foley and Sammon [10], [11] proposed an optimal set of discriminant vectors for this particular problem. The proposed transformation will include this case in a more general algorithm.

The problem is always to maximize

$$C = \frac{d_n' B d_n}{d_n' W d_n}$$

but with the constraints:

$$d_1' d_n = d_2' d_n = \cdots = d_{n-1}' d_n = 0$$

and an additional constraint on the norm of d_n, that can be chosen as

$$d_n' W d_n = 1.$$

It is always possible to restore the final norm to

$$d_n' d_n = 1.$$

after computation of the vector direction, in order to obtain an orthonormal set of vectors.

The first solution is the Fisher linear discriminant d_1:

$$W^{-1} B d_1 = \lambda_1 d_1.$$

In order to compute the nth discriminant vector, let us use the method of Lagrange multipliers to transform the C criterion including all the constraints

$$C_L = d_n' B d_n - \lambda\left[(d_n' W d_n) - 1\right] - \mu_1 d_n' d_1 - \cdots - \mu_{n-1} d_n' d_{n-1}.$$

The optimization is performed by setting the partial derivative of C_L with respect to d_n equal to zero:

$$\frac{\partial C_L}{\partial d_n} = 0 \Rightarrow 2Bd_n - 2\lambda W d_n - \mu_1 d_1 \cdots - \mu_{n-1} d_{n-1} = 0. \tag{1}$$

Multiplying the left side of (1) by d_n', we obtain

$$2d_n' B d_n - 2\lambda d_n' W d_n = 0 \Rightarrow \lambda = \frac{d_n' B d_n}{d_n' W d_n}. \tag{2}$$

Thus λ represents the expression to be maximized.

Multiplying the left side of (1) successively by $d_1' W^{-1}, \cdots, d_{n-1}' W^{-1}$, we now obtain a set of $n - 1$ expressions

$$\begin{aligned}
\mu_1 d_1' W^{-1} d_1 + \mu_2 d_1' W^{-1} d_2 + \cdots + \mu_{n-1} d_1' W^{-1} d_{n-1} &= 2d_1' W^{-1} B d_n \\
\mu_1 d_2' W^{-1} d_1 + \mu_2 d_2' W^{-1} d_2 + \cdots + \mu_{n-1} d_2' W^{-1} d_{n-1} &= 2d_2' W^{-1} B d_n \\
&\vdots \\
\mu_1 d_{n-1}' W^{-1} d_1 + \mu_2 d_{n-1}' W^{-1} d_2 + \cdots + \mu_{n-1} d_{n-1}' W^{-1} d_{n-1} &= 2d_{n-1}' W^{-1} B d_n.
\end{aligned}$$

Let us use matrix notation:

$$\mu^{(n-1)} = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_{n-1} \end{bmatrix} \qquad D^{(n-1)} = \begin{bmatrix} d_1' \\ d_2' \\ \vdots \\ d_{n-1} \end{bmatrix}$$

$$S^{(n-1)} = [S_{ij}^{(n-1)}], \; S_{ij}^{(n-1)} = d_i' W^{-1} d_j.$$

Using this simplified notation, the previous set of $(n - 1)$ equations can be represented in a single matrix relationship

$$S^{(n-1)} \mu^{(n-1)} = 2D^{(n-1)} W^{-1} B d_n$$

or in another form

$$\mu^{(n-1)} = 2\left[S^{(n-1)}\right]^{-1} D^{(n-1)} W^{-1} B d_n. \tag{3}$$

Let us now multiply the left side of (1) by W^{-1}

$$2W^{-1} B d_n - 2\lambda d_n - \mu_1 W^{-1} d_1 - \cdots - \mu_{n-1} W^{-1} d_{n-1} = 0.$$

This can be expressed using matrix notation as

$$2W^{-1} B d_n - 2\lambda d_n - W^{-1} [D^{(n-1)}]' \mu^{(n-1)} = 0.$$

Including (3), we obtain

$$\left(I - W^{-1}[D^{(n-1)}]' [S^{(n-1)}]^{-1} D^{(n-1)}\right) W^{-1} B d_n = \lambda d_n.$$

Considering λ as the criterion to be maximized, d_n is the eigenvector of

$$M = \left(I - W^{-1}[D^{(n-1)}]' [S^{(n-1)}]^{-1} D^{(n-1)}\right) W^{-1} B$$

associated with the largest eigenvalue of M.

Let us consider the first case $n = 1$: then $\mu^{(0)} = 0$ and M can be written in a simplified form as

$$M = W^{-1} B$$

giving the best solution as the Fisher Linear discriminant vector.

Let us now consider the two-class discrimination problem. In this case, the between-class scatter matrix can be expressed as [10]

$$B = \frac{N_1 N_2}{N} [\mu_1 - \mu_2][\mu_1 - \mu_2]'$$

where

μ_i = the mean vector of class i
N_i = the number of samples from class i
N = the total number of samples

or in another form

$$B = \frac{N_1 N_2}{N} \Delta \Delta'$$

Δ being the difference between the means of the two classes.

Then the Fisher linear discriminant can be written in the following form [10]:

$$d_1 = \alpha_1 W^{-1} \Delta \quad \text{with} \quad \alpha_1 = \frac{1}{\| W^{-1} \Delta \|}$$

and $W^{-1} B$ can be transformed as

$$W^{-1} B = \frac{N_1 N_2}{N} W^{-1} \Delta \Delta'$$

involving

$$W^{-1}Bd_n = \frac{N_1N_2}{N} W^{-1}\Delta\Delta' d_n = \frac{\alpha_n}{\alpha_1} d_1$$

with

$$\alpha_n = \frac{N_1N_2}{N} \Delta' d_n \quad \text{and} \quad W^{-1}\Delta = \frac{d_1}{\alpha_1}$$

giving directly the vector d_n

$$d_n = \frac{\alpha_n}{\lambda} \left(I - W^{-1}[D^{(n-1)}]'[S^{(n-1)}]^{-1}D^{(n-1)}\right) \frac{d_1}{\alpha_1}$$

using the properties $d_i' d_1 = 0 \;\; \forall\, i \neq 1$ and $d_1' d_1 = 1$, we obtain

$$\frac{1}{\alpha_1} D^{(n-1)} d_1 = \begin{bmatrix} 1/\alpha_1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

giving a new expression for d_n:

$$d_n = \frac{\alpha_n}{\lambda} W^{-1} \left(\frac{1}{\alpha_1} W d_1 - [D^{(n-1)}]'[S^{(n-1)}]^{-1} \begin{bmatrix} 1/\alpha_1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \right)$$

with respect to the definition of the Fisher linear discriminant $d_1 = \alpha_1 W^{-1}\Delta$, the final form of d_n is given by

$$d_n = \frac{\alpha_n}{\lambda} W^{-1} \left(\Delta - [D^{(n-1)}]'[S^{(n-1)}]^{-1} \begin{bmatrix} 1/\alpha_1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \right)$$

This expression is exactly the form given by Foley and Sammon [11] for the determination of the optimal set of discriminant vectors. Notice that this result can be modified according to the generalized intraclass variance suggested by Anderson and Bahadur [13]. This method permits to obtain at most P orthogonal features, P being the dimension of the original space.

Finally, we have found a set of orthogonal discriminant vectors including all the particular cases previously considered. Furthermore, the determination of d_n does not involve the knowledge of any specific property for the previously computed vectors d_i, with the exception that they have to be orthogonal one to each other. That can be of greatest interest for determining a complete set of orthogonal vectors including both discrimination and best projected variance.

III. Principal Component Analysis

When a discriminant vector is found, for instance in a two-class discrimination problem, it may be of interest to find a second vector, orthogonal to the first one, maximizing the projected variance. This corresponds to a principal component analysis with a constraint of orthogonality.

It can be useful to remind us of the general solution of this analysis. The criterion is the projected variance onto the vector to be found, d, expressed as

$$\text{maximize } C = d'Td \quad \text{with} \quad d'd = 1.$$

The solution is obtained after computing the partial of C with respect to d including the constraint using the Lagrange multipliers, giving the following result

$$Td = \lambda d \quad \text{and} \quad \lambda = d'Td = C,$$

where d is the eigenvector associated with the greatest eigenvalue λ. If T is a symmetrical matrix, then the successive eigenvectors are orthogonal giving directly a set of optimal vectors.

Our problem involves computing a vector d maximizing the C criterion, but orthogonal to a given vector or a set of given vectors, for instance a subset of discriminant vectors.

A solution was given in a previous note [12] for the particular case of the computation of the best vector orthogonal to only a single previously defined vector. We propose to compute now this optimal vector in the general case of orthogonality with respect to $(n-1)$ previously defined orthogonal vectors d_i, $i = 1 \cdots n-1$.

The criterion can always be written as

$$\text{maximize } d_n' T d_n$$

subject to the constraints

$$d_n' d_i = 0 \quad \forall\, i \;\; i = 1 \cdots n-1$$
$$d_n' d_n = 1$$

knowing that

$$d_i' d_j = 0 \quad i \neq j \;\; \forall\, i, j = 1 \cdots n-1$$
$$d_i' d_i = 1 \quad \forall\, i, i = 1 \cdots n-1.$$

Using the Lagrange multipliers again, the criterion can be expressed as

$$C = d_n' T d_n - \lambda(d_n' d_n - 1) - \mu_1 d_n' d_1 - \cdots - \mu_{n-1} d_n' d_{n-1}.$$

Computing the partial of C with respect to d_n, it yields

$$\frac{\partial C}{\partial d_n} = 0 \quad 2Td_n - 2\lambda d_n - \mu_1 d_1 - \cdots - \mu_{n-1} d_{n-1} = 0. \tag{4}$$

Multiplying left side of (4) by d_n'

$$2d_n' T d_n - 2\lambda d_n' d_n = 0 \Rightarrow \lambda = d_n' T d_n$$

so λ represents the criterion to be maximized.

Let us now multiply left side of (4) successively by $d_1', \cdots, d_{n-1}'$

$$\mu_1 = 2d_1' T d_n, \cdots, \mu_{n-1} = 2d_{n-1}' T d_n$$

giving another form for (4)

$$Td_n - \lambda d_n - d_1 d_1' T d_n \cdots - d_{n-1} d_{n-1}' T d_n = 0$$

or, written in another order

$$(I - d_1 d_1' - \cdots - d_{n-1} d_{n-1}') T d_n = \lambda d_n \tag{5}$$

d_n is then the eigenvector of

$$N = (I - d_1 d_1' - \cdots - d_{n-1} d_{n-1}') T$$

associated with the greatest eigenvalue λ.

Let us consider the particular case for which $d_1 \cdots d_{n-1}$ are defined as the $(n-1)$ eigenvectors of T ("principal component" vectors)

$$\lambda_1 d_1 = Td_1, \cdots, \lambda_{n-1} d_{n-1} = Td_{n-1}$$

or

$$d_1' = \lambda_1 d_1' T^{-1}, \cdots, d_{n-1}' = \lambda_{n-1} d_{n-1}' T^{-1}.$$

Equation (5) can be transformed as

$$(T - \lambda_1 d_1 d_1' - \cdots - \lambda_{n-1} d_{n-1} d_{n-1}') d_n = \lambda_n d_n.$$

Let us define

$$X = (T - \lambda_1 d_1 d_1' - \cdots - \lambda_{n-1} d_{n-1} d_{n-1}').$$

Multiplying the right side of the above equation by d_n gives

$$Xd_n = Td_n \quad (d_1' d_n = \cdots d_{n-1}' d_n = 0)$$

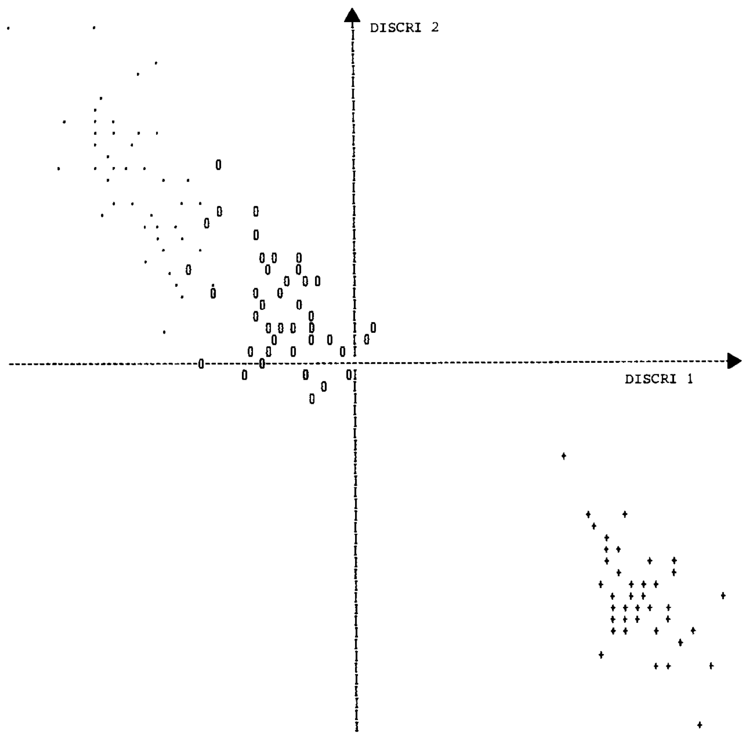

Fig. 1. Projection of the three classes of IRIS onto the best discriminant plane.

and

$$Xd_i = Td_i - \lambda_i d_i = 0.$$

d_n is both the first eigenvector of X and the nth eigenvector of T. We come back to the general solution of the principal component analysis.

Thus this method permits to define a "best variance" vector, orthogonal to a set of previously computed vectors, without using any statistical property of this set. It can be used either to carry out the classical principal component analysis or to complete a discriminant set of vectors by another orthogonal set of "best variance" vectors. All combinations are possible; in fact, the transformation using both discrimination and best variance, and computing a number of orthogonal vectors equal to the initial dimension of the space is a rotation: the initial shape of the data is not modified by the transformation, but the projection on each new vector has a specific significance.

IV. The Program "Factor" and Its Results on the IRIS-Data

The previously described methods were programmed on a microcomputer system first in compiled Basic, next in Pascal for a better execution rate. The user has to first load the data-file and to select the different classes. This routine permits to define both reference (training) and additional (test) objects: reference objects are used to define the new set of vectors, and their class has to be known before beginning the computation; on the other hand, supplementary objects are not used for determining the new vectors, but are only projected in the computed subspace.

This procedure is very useful to classify objects into predefined pattern classes, which is the final objective of any pattern recognition method.

After the data file is loaded and the classes are selected, the user specifies the type of each vector of the subspace, discrimination, or best variance, and FACTOR computes the successive directions. Associated with the program GRAPH, it projects all the objects of the initial data set onto the vectors or the planes selected by the user.

This program was tested using the IRIS data: three classes of flowers were defined *a priori*, and four measurements were recorded on each object (width and length of petals and sepals).

The results are obtained by FACTOR in three different ways depending on the type of the successive vectors. Fig. 1 shows the projections onto the best discriminant plane, Fig. 2 represents the projections onto the "best variance" plane, and Fig. 3 is obtained after computing the first vector as discriminant and the second as "best variance." Table I gives information on the projected variance and the discrimination ratio for each computed vector, in each configuration. The fourth column gives the same parameters obtained by the classical discriminant analysis for three classes, i.e., the projected variance and the discrimination ratio for the two eigenvectors of the $W^{-1}B$ matrix.

Table I shows that the discriminant vectors (column 1) give a better result for discrimination. The result of the classical discriminant analysis given by Fig. 4 is to be compared to FACTOR results (Fig. 1). Notice that the two represented axes are drawn so that they appear orthogonal: in fact the angle between the two vectors was computed to be 80°.

Now compare the variance retained after the projection of the data onto the different planes: the second plane (discriminant and "best variance") provides 77 percent of the initial variance. So, this plane can be used either to estimate the shape of the data set in the initial space or it can be considered as a discriminant plane using the first axis as the Fisher linear discriminant.

V. Discussion and Conclusion

Our aim was to propose a new form of projection of data sets, combining a generalized principal componant analysis and a discriminant analysis. Even if the major goal of pattern recognition is to classify at best objects in their classes, it can be also of interest

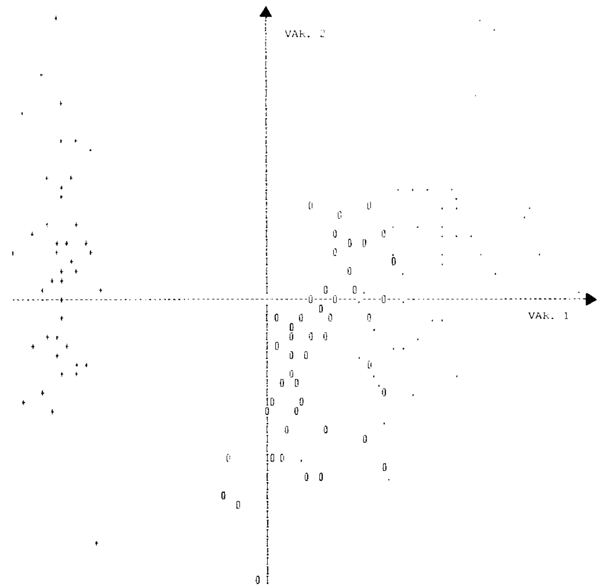

Fig. 2. Projection of the three classes of IRIS onto the first principal component plane.

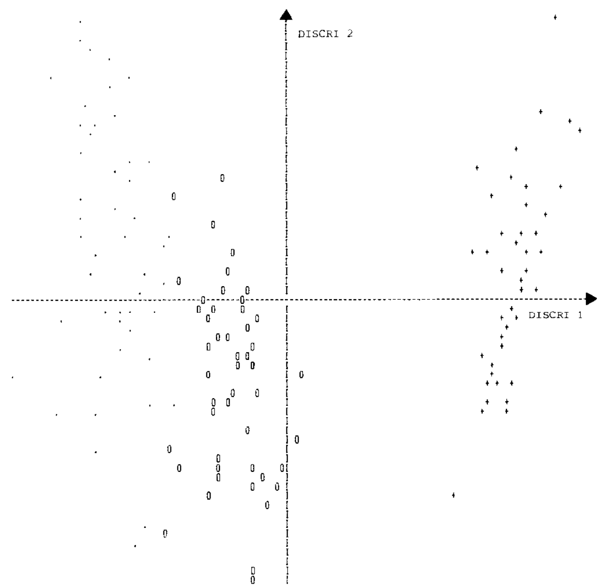

Fig. 4. Projection of the three classes of IRIS onto the classical discriminant plane.

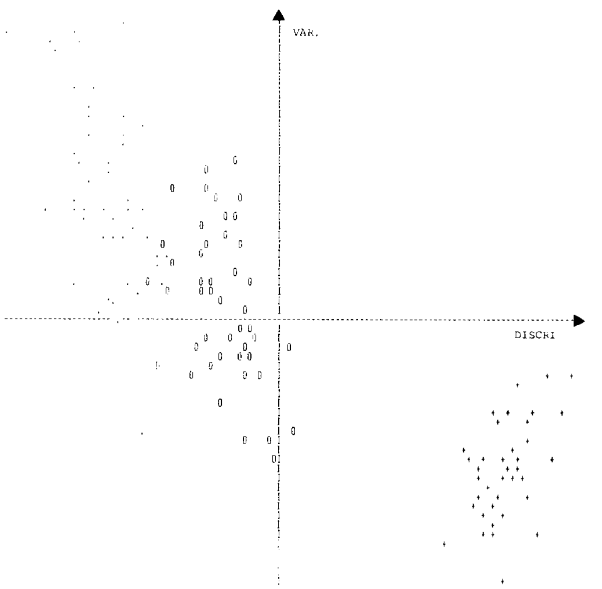

Fig. 3. Projection of the three classes of IRIS onto a mixed plane including discrimination (first axis) and best variance (second axis).

to obtain in the same subspace a good representation of the corresponding points (in the sense of best projected variance), and to show the distribution of these points into different classes (in the sense of best discrimination), especially if the percentage of correctly classified points is the same as by the classical discriminant methods. In this way, we tested our results using the leave-one-out method of error estimation, based on minimum distance and nearest-neighbor decision rules, in order to permit a comparison to the results obtained by BISWAS *et al.* [15]. The test using the minimum distance rule gives 96.7 percent of correctly classified points (CCP) for the best discriminant plane and 92 percent for the plane

TABLE I
VARIANCE AND DISCRIMINANT RATIO FOR EACH COMPUTED VECTOR (PERCENTAGE)

FEATURE OF THE PLANE		OPTIMAL DISCRI.	OPTIMAL VARIANCE	MIXED DISCRI. VARIANCE	CLASSICAL DISCRI.
FIRST VECTOR	FEATURE	DISCRI.	VARIANCE	DISCRI.	DISCRI.
	PROJECTED VARIANCE	42	73	42	42
	DISCRI. RATIO	97	–	97	97
SECOND VECTOR	FEATURE	DISCRI.	VARIANCE	VARIANCE	DISCRI.
	PROJECTED VARIANCE	13.5	23	35	3
	DISCRI. RATIO	91	–	–	23
PLANE:PROJECTED VARIANCE		55.5	96	77	45

with both discriminant and best variance features. The test using the nearest-neighbor rule gives also 96.7 percent of CCP for the first plane and 95.3 percent for the second one. These results show that the percentage of CCP is the same as for the best methods tested by BISWAS and it seems difficult to obtain better results (96.7 percent of CCP corresponds to five misclassified points). Associated with the very high discriminant ratio, these results show that the proposed projection gives the user a powerful method for discriminant analysis, then for pattern recognition.

ACKNOWLEDGMENT

The authors wish to thank Prof. J. P. Kernevez (Compiégne University) for his very useful suggestions in the mathematical developments.

REFERENCES

[1] H. Späth, *Cluster Analysis Algorithms.* New York: Wiley, 1980.
[2] J. Hartigan, *Clustering Algorithms.* New York: Wiley, 1975.
[3] G. H. Ball and D. J. Hall, "A clustering technique for summarizing multivariate data," *Behavioral Sci.*, vol. 12, pp. 153–155, 1967.
[4] K. Fukunaga and W. L. G. Koontz, "Application of the Karhunen

Loeve expansion to feature selection and ordering," *IEEE Trans. Comput.*, vol. C-19, no. 4, pp. 311-318, 1970.

[5] J. M. Romeder, *Methodes et Programmes d'Analyse Discriminante.* Paris: Dunod, 1973.

[6] C. R. Rao, "The use and interpretation of principal component analysis in applied research," *Sankhya Series A*, vol. 26, pp. 329-357, 1964.

[7] T. Okada and S. Tomita, "An optimal orthonormal system for discriminant analysis," *Pattern Recognition*, vol. 18, no. 2, pp. 139-144, 1985.

[8] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Ann. Eugen.*, vol. 7, pp. 179-188, 1936.

[9] Y. T. Chien, *Interactive Pattern Recognition.* New York: Dekker, 1978.

[10] J. W. Sammon, "An optimal discriminant plane," *IEEE Trans. Comput.*, vol. C-19, pp. 826-829, 1970.

[11] D. H. Foley and J. W. Sammon, "An optimal set of discriminant vectors," *IEEE Trans. Comput.*, vol. C-24, pp. 281-289, 1975.

[12] J. Duchene, "A significant plane for two-class discrimination problems," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-8, pp. 557-559, 1986.

[13] T. W. Anderson and X. X. Bahadur, "Classification into two multivariate normal distributions with different covariance matrices," *Ann. Math. Stat.*, vol. 33, pp. 420-431, 1962.

[14] K. V. Mardia, *Multivariate Analysis.* London: Academic, 1979.

[15] G. B. Swas, A. K. Jain, and R. C. Dubes, "Evaluation of projection algorithms," *IEEE Trans. Pattern Anal. Machine Intell*, vol. PAMI-3, pp. 701-708, 1981.

Correction to "Recognition of Noisy Subsequences Using Constrained Edit Distances"

B. J. OOMMEN

The algorithm ConstrainedDistance formally presented in the above paper[1] computed a quantity called the Constrained edit distance between a string X and another Y subject to the constraint τ.[2] However, there were some major typographical[3] and special-case errors in the final version in the above paper.[1] In particular, the initialization of the Wis plane was omitted, and the bounds on the variable e in the **For** loops were erroneously stated. Furthermore, the stated bounds on the value of the variable s for the loop which computes the Wes plane were incorrect because they could cause an inaccurate reference under a certain special case condition. Besides these, a few typographical errors were also present. Although this does **not** invalidate the results of [1], to correct the above, we have **rewritten** below the entire code in a more structured way and have modularized the initialization and the updating processes in two separate procedures called **InitializePlanes** and **ComputeDistance**, respectively. The algorithm follows.

ALGORITHM ModifiedConstrainedDistance (X, Y, d, τ, D_τ)
Input: The strings $X = x_1x_2 \cdots x_N$, $Y = y_1y_2 \cdots y_M$, the set of edit distances $d(\cdot, \cdot)$, and the constraint set τ, whose largest element is $T > 0$.
Output: The constrained edit distance $D_\tau(X, Y)$.
Method:
  InitializePlanes$(X, Y, d, Wis, Wie, Wes0)$
  ComputeDistance$(X, Y, d, \tau, Wis, Wie, Wes0, D_\tau)$
End
END ALGORITHM ModifiedConstrainedDistance
PROCEDURE InitializePlanes $(X, Y, d, Wis, Wie, Wes0)$
Input: The strings $X = x_1x_2 \cdots x_N$, $Y = y_1y_2 \cdots y_M$, and the set of edit distances $d(\cdot, \cdot)$.
Output: The initialized arrays Wis, Wie, and $Wes0$.
Method:
  $Wis(0, 0) = Wie(0, 0) = Wes0(0, 0) = 0$ /*initialize origin*/
  For $s = 1$ **to** Min$[M, N]$ **do** /*initialize s-axis*/
    $Wis(0, s) = Wes0(0, s) = Wis(0, s - 1) + d(x_s, y_s)$
  For $e = 1$ **to** N **do** /*initialize e-axis*/
    $Wie(0, e) = Wes0(e, 0) = Wie(0, e - 1) + d(x_e, \theta)$
  For $i = 1$ **to** T **do** /*initialize i-axis*/
    $Wis(i, 0) = Wie(i, 0) = Wis(i - 1, 0) + d(\theta, y_i)$

The author is with the School of Computer Science, Carleton University, Ottawa, Ont. K1S 5B6, Canada.

IEEE Log Number 8824146.

[1]B. J. Oommen, *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-9, no. 5, pp. 676-685, 1987.

[2]A preliminary version of the above paper[1] was presented at the 1987 Conference on Information Sciences and Systems, Johns Hopkins University, Baltimore, MD, in March 1987. For the sake of brevity we assume the notation and terminology of this paper.

[3]The author is grateful to Edwin T. Floyd of the Hughston Sports Medicine Foundation, Columbus, GA, for assisting him prepare this correction.

0162-8828/88/1100-0984$01.00 © 1988 IEEE