

On Image Matrix Based Feature Extraction Algorithms

Liwei Wang, Xiao Wang, Jufu Feng

Center for Information Sciences

School of Electronics Engineering and Computer Sciences, Peking University

Beijing, 100871, China

{wanglw, wangxiao, fjf}@cis.pku.edu.cn

Abstract Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) are two important feature extraction methods and have been widely applied in a variety of areas. A limitation of PCA and LDA is that when dealing with image data, the image matrices must be first transformed into vectors, which are usually of very high dimensionality. This causes expensive computational cost and sometimes the singularity problem. Recently two methods called two-dimensional PCA (2DPCA) and two-dimensional LDA (2DLDA) were proposed to overcome this disadvantage by working directly on 2D image matrices without a vectorization procedure. 2DPCA and 2DLDA significantly reduce the computational effort and the possibility of singularity in feature extraction. In this paper, we show that these matrices based 2D algorithms are equivalent to special cases of image block based feature extraction, i.e. partition each image into several blocks and perform standard PCA or LDA on the aggregate of all image blocks. These results thus provide a better understanding of the 2D feature extraction approaches.

Index Terms: PCA, LDA, feature extraction, face recognition, two-dimensional PCA (2DPCA), two-dimensional LDA (2DLDA), block based feature extraction.

I. Introduction

Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) are two powerful tools for feature extraction and data representation. They have been widely used in a variety of areas such as pattern recognition, computer vision, machine learning, neural networks, bioinformatics and finite automata. Among others, a most successful application is the human face recognition. Kirby and Sirovich [1] were the first to employ Karhunen-Loeve transform to represent facial images. Their work immediately led to the PCA “Eigenface” technique [2]. Although PCA ensures the features extracted have least reconstruction error, it may not be optimal from a discrimination standpoint. Etemad et al. [3], Belhumeur et al. [4] and Zhao et al. [5] then proposed the LDA “Fisherface” method to extract features that have the best discrimination ability. Eigenface and Fisherface methods have by now become a *de facto* standard and common performance benchmarks in the field.

However, PCA and LDA have a common disadvantage when working on images. In face recognition for example, each datum is an image matrix. To perform PCA or LDA, the 2D image matrices must be previously transformed into 1D vectors. The resulting vector space is usually of a very high dimensionality: An image of 256×256 resolution then becomes a 65536 -dimensional vector. The high dimensionality makes PCA and LDA difficult to implement. For on the one hand, PCA and LDA both involve eigenvalue decomposition, which is extremely time-consuming for high dimensional vectors. On the other hand, if the dimensionality exceeds the number of data points, LDA becomes a singularity

problem and cannot be solved directly. Considerable efforts have been made to deal with these problems. For details, see [6], [7] and a large number of references therein.

Recently several authors independently proposed a technique which extracts features directly from 2D images without a vectorization preprocessing. The methods were termed two-dimensional Principal Component Analysis (2DPCA) [8], [9], and two-dimensional Linear Discriminant Analysis (2DLDA) [10], [11], [12], as generalizations of classical PCA and LDA respectively. The main idea of these 2D methods is to construct *image covariance matrix* and *image scatter matrices* directly using the original image matrices. Compared to their counterparts in PCA and LDA, i.e. vector covariance matrix and scatter matrices, the image covariance matrix and image scatter matrices have a much smaller size. Therefore, 2DPCA and 2DLDA significantly reduce the computational cost and avoid the singularity problem.

Another line of research in feature extraction and face recognition is the block-based methods. In these methods, an image is partitioned into several blocks. Often, all blocks have the same size. Features are extracted from the block images. Block-based method was first appeared in Hidden Markov Models (HMM) based face recognition [13] dating back to 1994 [14]. Fredembach et al. [15] recently presented an Eigenregion algorithm, which performs PCA on segmented image regions. Other works related to this method include [16], [17], which make use of image fragments to represent objects.

In this correspondence, we show that the 2D methods are equivalent to special cases of the block image based feature extraction. In particular, 2DPCA is equivalent

to line-based PCA. That is, if each block is an image line, and taking all these line blocks as data samples, then applying standard PCA to the aggregate of the line examples results in 2DPCA. We further show that 2DLDA is equivalent to line, column or more complex image block-based LDA depending on the dimensionality reduction. The proofs adopt a technique used in [18].

The remainder of this paper is organized as follows. We briefly review 2DPCA and 2DLDA algorithms in Section II. In Section III, we establish the equivalence theorems. And finally we give a conclusion in Section IV.

II. A Brief Review of the 2D Methods

A. PCA and 2DPCA

The goal of PCA and 2DPCA is to extract features that can best approximate the data. In standard PCA, the data set consists of Euclidean vectors. Let $TS = \langle x_1, x_2, \dots, x_N \rangle$ ($x_i \in \mathbb{R}^d$) be the training sample set. PCA algorithm takes TS as input and returns M feature vectors w_1, w_2, \dots, w_M , $w_i \in \mathbb{R}^d$. The M feature vectors constitute an orthogonal basis of a linear subspace, which achieves least reconstruction error of TS among all M -dimensional subspaces. It is well-known [19] that the features w_1, w_2, \dots, w_M are the M leading eigenvectors of the sample covariance matrix

$$S = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})(x_i - \bar{x})^T. \quad (1)$$

where $\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$ is the mean of all sample vectors.

In 2DPCA however, the data are image matrices. The sample set is

$TS = \langle X(1), X(2), \dots, X(N) \rangle$, where $X(i)$ is a r -row c -column image matrix.

Instead of vectorizing each data to a $r \times c$ -dimensional vector, 2DPCA directly defines an image covariance matrix

$$S^{IM} = \frac{1}{N} \sum_{i=1}^N [X(i) - \bar{X}]^T [X(i) - \bar{X}], \quad (2)$$

where $\bar{X} = \frac{1}{N} \sum_{i=1}^N X(i)$ is the mean of all sample matrices. Note that S^{IM} has a much smaller size than the covariance matrix S of the vectorized data. The former is a $c \times c$ matrix, while the latter is $r \times c$ by $r \times c$. 2DPCA then extracts features from S^{IM} . The features are eigenvectors of S^{IM} corresponding to the M largest eigenvalues. Clearly, the features are all c -dimensional vectors.

Like PCA, features obtained by 2DPAC can be used to reconstruct the data with small error. For details, the reader is referred to [8], [9]. In Section III, we will show the equivalence of 2DPCA and line-based PCA. The reconstruction problem can therefore be understood as the reconstruction of the image lines by standard PCA.

For notational convenience, the normalizing factor $1/N$ of the image covariance matrix (see (2)) will be dropped in the rest of this paper. And when discussing 2DPCA, we will assume that, without loss of generality, data matrices have been shifted so that they have zero mean, i.e. $\bar{X} = \frac{1}{N} \sum_{i=1}^N X(i) = 0$. Accordingly, the image covariance matrix reduces to a simpler form

$$S^{IM} = \sum_{i=1}^N X(i)^T X(i). \quad (3)$$

B. LDA and 2DLDA

Both LDA and 2DLDA aim to extract features that well discriminate a set of data that belong to a number of classes. In classical LDA, the sample set is $TS = \langle x_1, x_2, \dots, x_N \rangle$, where each data is a vector in \mathbb{R}^d . The data come from K classes, denoted by C_1, C_2, \dots, C_K . Assume that class C_i contains N_i data, hence $\sum_{i=1}^K N_i = N$. The between-class scatter matrix is defined as

$$S_B = \sum_{i=1}^K N_i (\mu_i - \mu)(\mu_i - \mu)^T. \quad (4)$$

and the within-class scatter matrix is defined as

$$S_W = \sum_{i=1}^K \sum_{x_j \in C_i} (x_j - \mu_i)(x_j - \mu_i)^T. \quad (5)$$

where $\mu_i = \frac{1}{N_i} \sum_{x_j \in C_i} x_j$ is the mean of the data vector in C_i , and $\mu = \frac{1}{N} \sum_{i=1}^K \sum_{x_j \in C_i} x_j$ is the global mean.

LDA chooses M optimal feature vectors so that the following objective function is maximized [20]

$$tr[(W^T S_W W)^{-1} (W^T S_B W)]. \quad (6)$$

where tr denotes the trace, and W is a column orthogonal matrix with each column a feature vector, i.e. $W = (w_1, w_2, \dots, w_M)$. The optimal features w_1, w_2, \dots, w_M are M directions. Projecting original data vectors onto these directions, the resulting within-class scatter matrix has the minimum trace while the between-class scatter matrix has maximum trace simultaneously. It is well-known that the LDA features w_1, w_2, \dots, w_M are eigenvectors corresponding to the largest M eigenvalues of the following generalized eigenvalue problem

$$S_B w_i = \lambda_i S_W w_i, \quad i = 1, 2, \dots, M$$

The 2DLDA however, deals with data in image matrices representations. Again, let the sample set be $TS = \langle X(1), X(2), \dots, X(N) \rangle$, and $X(i)$ a r -row c -column matrix. Similar to 2DPCA, 2DLDA introduces *image within-class scatter matrix* S_W^{IM} and *image between-class scatter matrix* S_B^{IM} defined as follows

$$S_W^{IM} = \sum_{i=1}^K \sum_{X(j) \in C_i} [X(j) - M(i)][X(j) - M(i)]^T, \quad (7)$$

$$S_B^{IM} = \sum_{i=1}^K N_i [M(i) - M][M(i) - M]^T. \quad (8)$$

where $M(i) = \sum_{X(j) \in C_i} X(j)$ is the mean of data matrices in class C_i , and $M = \frac{1}{N} \sum_{i=1}^K \sum_{X(j) \in C_i} X(j)$ is the global mean matrix.

2DLDA searches such optimal directions that after projecting the original data onto these directions, the resulting image within-class scatter matrix has the minimum trace, while the image between-class scatter matrix has the maximum trace. Unlike LDA which performs linear projections of vectors, 2DLDA considers bilinear projections of image matrices. More precisely, 2DLDA involves two sets of projection directions $\{u_i\}_{i=1}^{l_1}$ and $\{v_i\}_{i=1}^{l_2}$, where $u_i \in \mathbb{R}^r$ and $v_i \in \mathbb{R}^c$ ($l_1 \leq r$, $l_2 \leq c$). These vectors define two matrices $L = [u_1, u_2, \dots, u_{l_1}]$ and $R = [v_1, v_2, \dots, v_{l_2}]$. The bilinear projection of a $r \times c$ data matrix $X(i)$ is

$$Y(i) = L^T X(i) R. \quad (9)$$

where the resulting $Y(i)$ is a $l_1 \times l_2$ matrix smaller than $X(i)$. Denote \tilde{S}_W^{IM} , \tilde{S}_B^{IM} the image within-class and between-class scatter matrices of the projected data $Y(i)$'s, 2DLDA then chooses L and R so that the following objective function is

maximized

$$\begin{aligned}
& \text{tr}[(\tilde{S}_W^{IM})^{-1}(\tilde{S}_B^{IM})] \\
&= \text{tr}\left\{\left[\sum_{i=1}^K \sum_{X(j) \in C_i} L^T (X(j) - M(i)) R R^T (X(j) - M(i))^T L\right]^{-1} \cdot \right. \\
& \quad \left. \left[\sum_{i=1}^K N_i L^T (M(i) - M) R R^T (M(i) - M)^T L\right]\right\}. \tag{10}
\end{aligned}$$

However, there is no closed form solution of this optimization problem. 2DLDA instead adopts an iterative algorithm. For a fixed R , computes the optimal L , then fix L to compute best R , and this procedure is repeated for a series of rounds. It is obvious that each time the optimization is to solve a generalized eigenvalue problem (compare to (6) and see also [10]).

2DLDA has more flexibility than LDA. Note that in 2DLDA one can make different choices of the dimensionalities of the left and right projections. In Section III we will show that the choices of projection dimensionality correspond to different kinds of image blocks in block-based LDA.

III. The Equivalence of 2D Methods to Block-Based Approaches

A. 2DPCA

In this section we show that 2DPCA is equivalent to a special case of the image block based PCA. In particular, each block is an image line. The procedure of image line based PCA is as follows: First partition each image whose resolution is $r \times c$ into r image lines. Each image line is considered as a Euclidean vector of dimensionality c . Then the N images in the sample set are decomposed to $N \cdot r$ vectors. Finally, applying standard PCA on these $N \cdot r$ vectors, i.e. computing the

leading eigenvectors of the covariance matrix of the $N \cdot r$ data vectors, we obtain the line-based PCA features.

Theorem 1

2DPCA is equivalent to image line-based PCA. □

For a detailed proof, please see [18]. The main idea is that the image covariance matrix S^{IM} (see (3)) of 2DPCA is exactly the ordinary sample covariance matrix of all the $N \cdot r$ image lines considered as vectors in \mathbb{R}^c . More precisely, suppose that $X(1), X(2), \dots, X(N)$ are image matrices of size $r \times c$. Let $x(i)_1, x(i)_2, \dots, x(i)_r$ be the r lines of $X(i)$:

$$X(i) = \begin{bmatrix} x(i)_{11} & x(i)_{12} & \cdots & x(i)_{1c} \\ x(i)_{21} & x(i)_{22} & \cdots & x(i)_{2c} \\ \cdots & \cdots & \cdots & \cdots \\ x(i)_{r1} & x(i)_{r2} & \cdots & x(i)_{rc} \end{bmatrix} = \begin{bmatrix} x(i)_1^T \\ x(i)_2^T \\ \cdots \\ x(i)_r^T \end{bmatrix}, \quad (11)$$

and

$$x(i)_h^T = (x(i)_{h1}, x(i)_{h2}, \dots, x(i)_{hc}), \quad h = 1, 2, \dots, r, \quad i = 1, 2, \dots, N. \quad (12)$$

Then

$$S^{IM} = \sum_{i=1}^N X(i)^T X(i) = \sum_{i=1}^N \sum_{h=1}^r x(i)_h x(i)_h^T. \quad (13)$$

Hence the two methods are equivalent.

B. 2DLDA

The equivalence of 2DPCA and line-based PCA is established based upon the fact that the image covariance matrix S^{IM} is just the covariance matrix of all image lines. Note that there is a strong similarity between the image covariance matrix in 2DPCA and the image scatter matrices in 2DLDA (see (2) and (7), (8)). Hence the preceding

arguments may be applied to prove the equivalence of 2DLDA and block based LDA. However, 2DLDA is a little more complicated because of the bilinear projection involved. So to begin with we consider two simple cases of 2DLDA, for which the bilinear projections reduce to linear projections. We show that the two simple versions of 2DLDA are equivalent to line and column based LDA respectively, i.e. partition each image into r lines or c columns, and then perform LDA on these image line/column data vectors.

Theorem 2

2DLDA is equivalent to column based LDA if the right projection R is identity.

Proof

If the right projection R is identity, 2DLDA's criterion (10) then reduces to maximizing the following objective function with respect to L

$$\text{tr}\left\{L^T \left[\sum_{i=1}^K \sum_{X(j) \in C_i} (X(j) - M(i))(X(j) - M(i))^T \right] L \right\}^{-1} \cdot \left[L^T \left[\sum_{i=1}^K N_i (M(i) - M)(M(i) - M)^T \right] L \right]. \quad (14)$$

Let

$$X(j) = [x(j)_1, x(j)_2, \dots, x(j)_c],$$

$$M(i) = [m(i)_1, m(i)_2, \dots, m(i)_c],$$

$$M = [m_1, m_2, \dots, m_c].$$

where $x(j)_h$, $M(i)_h$ and m_h are the h th column vector of $X(j)$, $M(i)$ and M respectively. We thus have

$$[X(j) - M(i)][X(j) - M(i)]^T = \sum_{h=1}^c [x(j)_h - m(i)_h][x(j)_h - m(i)_h]^T. \quad (15)$$

In analogy, we also have

$$[M(i) - M][M(i) - M]^T = \sum_{h=1}^c [m(i)_h - m_h][m(i)_h - m_h]^T.$$

The 2DLDA objective function (14) is then written as

$$\begin{aligned} & \text{tr} \left\{ L^T \left[\sum_{i=1}^K \sum_{X(j) \in C_i} \sum_{h=1}^c [x(j)_h - m(i)_h][x(j)_h - m(i)_h]^T \right] L \right\}^{-1} \\ & \left[L^T \left[\sum_{i=1}^K \sum_{h=1}^c N_i [m(i)_h - m_h][m(i)_h - m_h]^T \right] L \right\}. \end{aligned} \quad (16)$$

From (16) it is clear that the image within/between-class scatter matrices are exactly the within/between-class scatter matrices of all image columns $x(j)_h$, $j = 1, 2, \dots, N$, $h = 1, 2, \dots, c$. Comparing the above formula to the standard criterion of LDA (6), it is readily seen that this is the criterion of column based LDA. \square

In the same manner, we can prove the following theorem.

Theorem 3

2DLDA is equivalent to line-based LDA if the left projection L is identity. \square

As described in Section II, 2DLDA aims to choose two sets of features $L = [u_1, u_2, \dots, u_{l_1}]$ and $R = [v_1, v_2, \dots, v_{l_2}]$. The above two special cases (imposing L or R to be identity) extract only one set of features. If neither L nor R is identity, 2DLDA becomes more complicated. There is no closed form solutions of L and R simultaneously. In [10] the authors suggested a two-stage optimization algorithm: Fix R optimize L , and then fix L optimize R .

Consider the first step. If R is fixed, the optimization is very similar to the case that R is identity (see (10) and (14)). The only differences are that the terms $[X(j) - M(i)][X(j) - M(i)]^T$ and $[M(i) - M][M(i) - M]^T$ in (14) are replaced by

$[X(j) - M(i)]RR^T[X(j) - M(i)]^T$ and $[M(i) - M]RR^T[M(i) - M]^T$ if R is not identity.

If we denote

$$R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1l_2} \\ r_{21} & r_{22} & \cdots & r_{2l_2} \\ \cdots & \cdots & \cdots & \cdots \\ r_{c1} & r_{c2} & \cdots & r_{cl_2} \end{bmatrix}$$

We have

$$[X(j) - M(i)]RR^T[X(j) - M(i)]^T = \sum_{h=1}^{l_2} [\tilde{x}(j)_h - \tilde{m}(i)_h][\tilde{x}(j)_h - \tilde{m}(i)_h]^T. \quad (17)$$

and

$$[M(i) - M]RR^T[M(i) - M]^T = \sum_{h=1}^{l_2} [\tilde{m}(i)_h - \tilde{m}_h][\tilde{m}(i)_h - \tilde{m}_h]^T. \quad (18)$$

where

$$\tilde{x}(j)_h = \sum_{k=1}^c r_{kh} x(j)_h, \quad \tilde{m}(i)_h = \sum_{k=1}^c r_{kh} m(i)_h, \quad \tilde{m}_h = \sum_{k=1}^c r_{kh} m_h.$$

and $x(j)_h$, $m(i)_h$ and m_h are the h -th column vectors of $X(j)$, $M(i)$ and M respectively. The above formulae indicate that $\tilde{x}(j)_h$, $\tilde{m}(i)_h$ and \tilde{m}_h are linear combinations of all columns of image matrices $X(j)$, $M(i)$ and M respectively.

Using the arguments in previous proofs, we readily obtain the following theorem.

Theorem 4

The first optimization step of 2DLDA, i.e. fix R to choose optimal L , is equivalent to perform LDA on a set of vectors, each of which is a linear combination of the original image columns. □

Similarly, we have the following theorem.

Theorem 5

The second optimization step of 2DLDA, i.e. fix L to choose optimal R , is equivalent to perform LDA on a set of vectors, each of which is a linear combination of the original image lines. □

Theorems 2 to 5 apply to different stages of 2DLDA. We have mentioned that 2DLDA involves a series of rounds. Initially, 2DLDA assumes R to be identity and optimizes L . According to theorem 2, this is equivalent to perform ordinary LDA on all the image columns. At a further round, the current R (or L) is not identity. But due to theorem 4 (or 5), one can compute linear combinations of the image columns (rows), where the weights correspond to the entries of the current R (L). Then conducting LDA on these vectors obtains the same result as 2DLDA.

IV. Conclusion

We established the equivalence between image matrices based and image block based feature extraction methods. Specifically, we show that 2DPCA is equivalent to image line based PCA, and 2DLDA is equivalent to line, column or their linear combination based LDA depending on the dimensionality of its left and right projection.

Acknowledgement

The authors would like to thank the reviewers and the editor for their careful and constructive comments. This work is supported by NKBRPC (2004CB318000).

References

- [1] M. Kirby and L. Sirovich, "Application of the Karhunen-Loeve Procedure for the Characterization of Human Faces," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 12, no. 1, pp. 103-108, 1990.
- [2] M.A. Turk and A.P. Pentland, "Face Recognition Using Eigenfaces," in *Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1991.
- [3] K. Etemad, R. Chellappa, "Discriminant Analysis for Recognition of Human Face Images," *J. Opt. Soc. Am. A*, vol. 14, no. 8, pp. 1724-1733, 1997.
- [4] P.N. Belhumeur, J.P. Hespanha and D.J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 19, no. 7, pp. 711-720, 1997.
- [5] W. Zhao, A. Krishnaswamy, R. Chellappa, D.L. Swets and J. Weng, "Discriminant Analysis of Principal Components for Face Recognition," In H. Wechsler, P.J. Phillips, V. Bruce and T. Huang, editors, *Face Recognition: From Theory to Applications*, pp. 73-85, Springer-Verlag, 1998.
- [6] J. H. Friedman, "Regularized Discriminant Analysis," *J. Am. Statistical Assoc.*, vol. 84, no. 405, pp. 165-175, 1989.
- [7] J. Ye, R. Janardan, C.H. Park and H. Park, "An Optimization Criterion for Generalized Discriminant Analysis on Undersampled Problems," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 26, no. 8, 2004.
- [8] J. Yang and D. Zhang, "Two-Dimensional PCA: A New Approach to Appearance-Based Face Representation and Recognition," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 26 no. 1, pp. 131-137, 2004.
- [9] J. Yang and J.Y. Yang, "From Image Vector to Matrix: a Straightforward Image Projection Technique—IMPCA vs. PCA," *Pattern Recognit.* vol. 35, no. 9, pp. 1997-1999, 2002.
- [10] J. Ye, R. Janardan and Q. Li, "Two-Dimensional Linear Discriminant Analysis," In *Proceedings Neural Information Processing Systems (NIPS)*, 2004.
- [11] J. Ye, "Generalized Low Rank Approximations of Matrices," In *Proceedings International Conference on Machine Learning (ICML)*, 2004.
- [12] M. Li and B. Yuan, "2D-LDA: A Novel Statistical Linear Discriminant Analysis for Image Matrix," *Pattern Recognit. Lett.*, vol. 26, no. 5, pp. 527-532, 2005.
- [13] M. Kim, D. Kim, and S. Lee, "Face Recognition Using the Embedded HMM with Second-Order Block-Specific Observations," *Pattern Recognit.*, vol. 36, no. 11, pp. 2723-2735, 2003.

- [14] F. Samaria and S. Young, "HMM-Based Architecture for Face Identification." *Image Vision Comput.*, vol. 12, no. 8, pp. 537-543, 1994.
- [15] C. Fredembach, M. Schroder and S. Susstrunk, "Eigenregions for Image Classification," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 26, no. 12, pp. 1645-1649, 2004.
- [16] S. Ullman, E. Sali and M. Vidal-Naquet, "A Fragment-based Approach to Object Representation and Classification," *In IWVF*, 2001.
- [17] M. Vidal-Naquet and S. Ullman, "Object Recognition with Informative Feature and Linear Classification," *In Proceedings International Conference on Computer Vision (ICCV)*, 2003.
- [18] L. Wang, X. Wang, X. Zhang and J. Feng, "The Equivalence of Two-Dimensional PCA to Line-Based PCA," *Pattern Recognit. Lett.*, vol. 26, no. 1, pp. 57-60, 2005.
- [19] I.T. Jolliffe, *Principal Component Analysis*, Springer-Verlag, New York, 1986.
- [20] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, Academic Press, San Diego, 1990.