Pattern Recognition 42 (2009) 2327 – 2334

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

Feature extraction based on Laplacian bidirectional maximum margin criterion

Wankou Yang[a,*], Jianguo Wang[b], Mingwu Ren[a], Jingyu Yang[a], Lei Zhang[c], Guanghai Liu[d]

[a]School of Computer Science & Technology, Nanjing University of Science and Technology, Nanjing 210094, China
[b]Network & Education Center, Tangshan College, Tangshan 063000, China
[c]Dept. of Computing, The Hong Kong Polytechnic University, Hong Kong
[d]School Computer Science and Information Technology, Guangxi Normal University, Guangxi, China

ARTICLE INFO

Article history:
Received 28 July 2008
Received in revised form 2 March 2009
Accepted 9 March 2009

Keywords:
MMC
Bidirectional
Laplacian
Feature extraction
Face recognition

ABSTRACT

Maximum margin criterion (MMC) based feature extraction is more efficient than linear discriminant analysis (LDA) for calculating the discriminant vectors since it does not need to calculate the inverse within-class scatter matrix. However, MMC ignores the discriminative information within the local structures of samples and the structural information embedding in the images. In this paper, we develop a novel criterion, namely Laplacian bidirectional maximum margin criterion (LBMMC), to address the issue. We formulate the image total Laplacian matrix, image within-class Laplacian matrix and image between-class Laplacian matrix using the sample similar weight that is widely used in machine learning. The proposed LBMMC based feature extraction computes the discriminant vectors by maximizing the difference between image between-class Laplacian matrix and image within-class Laplacian matrix in both row and column directions. Experiments on the FERET and Yale face databases show the effectiveness of the proposed LBMMC based feature extraction method.

1. Introduction

Dimensionality reduction is an important research topic in computer vision and pattern recognition fields. The curse of high dimensionality is usually a major cause of limitations of many practical technologies, while the large quantities of features may even degrade the performances of the classifiers when the size of the training set is small compared with the number of features [1]. In the past several decades, many feature extraction methods have been proposed, and the most well-known ones are principle component analysis (PCA) and linear discriminant analysis (LDA) [2].

Un-supervised learning cannot properly model underlying structure and characteristics of different classes. Discriminant features are often obtained by supervised learning. LDA [2] is the traditional approach to learn discriminant subspace. Unfortunately, it cannot be applied directly to small size sample (SSS) problems [3] because the within-class scatter matrix is singular. As we know, face recognition is a typical small size problem. Many works have been reported to use LDA for face recognition. The most popular method, called Fisherface, was proposed by Swets et al. [4] and Belhumeur et al. [5]. In their methods, PCA is first used to reduce the dimension of the original features space to $N-c$, and the classical FLD is then applied to reduce the dimension to d ($d \leqslant c$). Since the smallest $c-1$ projection components are thrown away in the PCA step, some useful discriminatory information may be lost. On the other hand, the PCA step cannot guarantee the transformed within-class scatter matrix be nonsingular. More discussions about PCA and LDA can be found in [6].

To solve the singularity problem, a singular value perturbation can be added to the within-class scatter matrix [7]. A more systematic method is regularized discriminant analysis (RDA) [8]. In RDA, one tries to obtain more reliable estimates of the eigenvalues by correcting the eigenvalue distortion with a ridge-type regularization. Penalized discriminant analysis (PDA) is another regularized version of LDA [9,10]. The goals of PDA are not only to overcome the SSS problem but also to smooth the coefficients of discriminant vectors for better interpretation. The main problem of RDA and PDA is that they do not scale well. In applications such as face recognition, the dimensionality is often more than 10,000. It is not practical for RDA and PDA to process such large covariance matrices.

A well-known null subspace method is the LDA + PCA method [11]. When within-class scatter matrix is of full rank, LDA+PCA only calculates the maximum eigenvectors of $(S_w)^{-1}S_b$ to form the transformation matrix. Otherwise, a two-stage procedure is employed. First, the data are transformed into the null space V_0 of S_w. Second, it maximizes the between-class scatter in V_0. LDA + PCA could be sub-optimal because it maximizes the between-class scatter in the

* Corresponding author.
E-mail address: wankou_yang@yahoo.com.cn (W. Yang).

0031-3203/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2009.03.017

null space of S_w instead of the original input space. Direct LDA is another null space method that discards the null space of S_b [12]. It is achieved by diagonalizing first S_b and then S_w, which is in the reverse order of conventional simultaneous diagonalization procedure. If S_t, instead of S_w, is used in direct LDA, it is actually equivalent to the PCA+LDA. Gao et al. [13] proposed a singular value decomposition (SVD) based LDA approach to solving the single training sample per person problem for face recognition. Zhuang and Dai [14,15] developed an inverse Fisher discriminant analysis (IFDA) method. They modified the procedure of PCA and derived the regular and irregular information from the within-class scatter matrix by a new criterion called inverse Fisher discriminant criterion. Jin et al. [16] proposed the uncorrelated optimal discrimination vectors (UODV) approach which maximizes Fisher criterion simultaneously. To avoid the singularity problem of LDA, Li et al. [17] used the difference of both between-class scatter and within-class scatter as discriminant criterion, called maximum margin criterion (MMC). Since the inverse matrix does not to be constructed, the SSS problem in traditional LDA is alleviated. MMC has the advantages of effectiveness and simplicity.

The above-mentioned methods need to transform the 2D images into 1D vectors. This often leads to the so-called "curse of dimensionality" problem, which is always encountered in SSS cases such as face recognition. The matrix-to-vector transform may also cause the loss of some useful structural information embedding in the original images. To overcome the problems, Yang et al. proposed the 2-dimensional principal component analysis (2DPCA) [18]. 2DPCA is based on 2D image matrices rather than 1D vectors. That is, the image matrix does not need to be transformed into a vector. Instead, the image covariance matrix can be constructed directly from the image matrices, and its eigenvectors are derived for image feature extraction. In contrast to PCA, the size of covariance matrix using 2DPCA is much smaller. As a result, 2DPCA computes the corresponding eigenvectors more quickly than PCA. Inspired by the successful application of 2DPCA to face recognition, 2DLDA was proposed [19–22]. Recently, Zheng et al. investigated the relations between vector-based LDA and matrix-based discriminant analysis [23]. They pointed out that from the bias estimation point of view, 2DLDA might be more stable than 1DLDA.

A drawback of 2DPCA is that it needs more coefficients than PCA for image representation. Thus, 2DPCA needs more memory to store features and costs more time to classify. Zuo et al. proposed bidirectional PCA (BDPCA) [24,25] to solve this problem. BDPCA assumes that the transform kernel of PCA is separable and it is a natural extension of the classical PCA and a generalization of 2DPCA. Inspired by 2DPCA, Gao et al. [26] proposed a sequential row–column independent component analysis (RC-ICA) for face recognition.

Recently, a method based on the local geometrical structure called tensor subspace analysis (TSA) [27] was proposed, which captures an optimal linear approximation to the face manifold in the sense of local isometry. However, the computational convergence of the iterative TSA algorithms is not guaranteed. To address the problem, Tao et al. proposed a tensor discriminant analysis method for feature extraction [28,29]. They proposed a convergent solution to discriminative tensor subspace selection and proved the convergence of it. In [30], Zhang et al. presented a directional multilinear ICA method by encoding the image or high-dimensional data array as a general tensor.

Recent studies have shown that the face images possibly reside on a nonlinear submanifold [31–39]. Many manifold-based learning algorithms have been proposed for discovering the intrinsic low-dimensional embedding of the original data. Among the various methods, the most well-known ones are isometric feature mapping (ISOMAP) [31], local linear embedding (LLE) [32] and Laplacian eigenmap [33]. Experiments have shown that these methods can find perceptually meaningful embedding for facial or digit images and other artificial and real-world data sets. However, how to evaluate the maps they generated on novel test data points remains unclear. He et al. [34,35] proposed the locality preserving projections (LPP), which is a linear subspace learning method derived from Laplacian eigenmap. In contrast to most manifold learning algorithms, LPP possesses a remarkable advantage that it can generate an explicit map. This map is linear and can be easily computed, like PCA and LDA. The objective function of LPP is to minimize the local scatter of the projected data.

Yang et al. [36] developed an unsupervised discriminant projection (UDP) technique for dimensionality reduction. UDP characterizes the local scatter as well as the nonlocal scatter, seeking for a projection that simultaneously maximizes the nonlocal scatter and minimizing the local scatter. Both LPP and UDP do not use the class label information and they are unsupervised methods in nature. Yan et al. proposed the marginal Fisher analysis (MFA) [37,38] and Chen et al. proposed the local discriminant embedding (LDE) [39] for feature extraction and recognition. The two methods are very similar in formulation. Both of them combine locality and class label information to represent the intraclass compactness and interclass separability. MFA and LDE take advantage of the partial structural information of classes and neighborhoods of samples; however, it is difficult to decide the number of nearest neighbors of each sample and the number of shortest pairs from different classes in MFA and LDE. In addition, the region covariance matrix (RCM) lies on the connected Riemannian manifold, instead of the subspace. RCM has many merits and is a natural feature for pattern recognition tasks. Pang et al. kernelized the RCM, formalized the similarity metric using four block matrices and obtained good results on face recognition [40].

In this paper, we propose a Laplacian bidirectional maximum margin criterion (LBMMC) for feature extraction and recognition. We formulate the Laplacian between-class scatter matrix and Laplacian within-class scatter matrix on local patches of the data by the weighted summation of distances based on image matrices. The weighted summation of distances has been successfully used in manifold learning [35,36] and can capture the underlying clustering of samples. The objective function of our proposed method is the trace difference criterion which can be directly solved by generalized eigenvalue decomposition. There is no convergence problem in our proposed method. Wang et al. [41] pointed out that the family objective functions for dimensionality reduction with trace ratio criterion can be generally transformed into the corresponding ratio trace criterion for obtaining a closed-form but approximate solution. They proved the convergence of the projection and gave the global optimality of the trace ratio value. They further extended the method into tensor space [42], but it needs much more computation and may be locally optimal in the tensor space.

Recently, Fu et al. have done some very good work in subspace learning [43–45]. In [43,44], they proposed a new criterion based on the concept of k-nearest-neighbor simplex (kNNS), which is constructed by the k-nearest-neighbors, to determine the class label of a new datum. For feature extraction, they developed a novel subspace learning algorithm, called discriminant simplex analysis (DSA), in which the within-class compactness and between-class separability are both measured by kNNS distance. In another work [45], Fu et al. proposed a new discriminant subspace learning algorithm, called correlation tensor analysis (CTA), incorporating both graph-embedded correlation mapping and discriminant analysis in a Fisher type of learning manner. The correlation metric can estimate the intrinsic angles and distance for the locally isometric embedding, which can deal with the case when Euclidean metric fails. CTA learns multiple interrelated subspaces to obtain a low-dimensional data representation reflecting both class label information and intrinsic geometric structure of the data distribution.

DSA is very similar to the feature extraction and classification method in [46], which is based on minimal local reconstruction error. In DSA, the within-class compactness and between-class separability are measured by kNNS distance. In CTA, the within-class compactness and between-class separability are measured by correlation distance. In our proposed method, the within-class compactness and between-class separability are measured by the weighted sum of distance of any two data points. All the three methods can process the case when the Euclidean distance does not work well. In DSA, the image matrix is first transformed into image vector which ignored the image structural information. In CTA, the image structural information is preserved by treating original image as tensors. In the proposed method, we directly perform the algorithm on the image matrix to preserve the image structure information. So CTA and our proposed method preserve the image structural information while DSA ignores such information.

In DSA, the objective function is transformed into the corresponding ratio trace criterion for solution. Wang et al. [41] pointed out that the solution by transforming trace ratio into ratio trace is closed-form but sub-optimal. Furthermore, Wang extended the work into tensor subspace learning [42] and proved the convergence. However, the solution proposed by Wang et al. may be locally optimal but not globally optimal and it needs much computational cost. In CTA, the objective function is in a trace ratio form. It may be transformed into a ratio trace form and it is iteratively solved as usually performed in tensor analysis. So the solution in CTA may not be globally optimal and it needs much computation. In our proposed method, the objective function is a trace difference criterion and can be directly solved by generalized eigenvalue decomposition. The optimization of the objective function is simple and does not have the convergence problem and local optimum problem.

The organization of this paper is as follows. In Section 2, we review briefly the MMC and BDPCA. In Section 3, we propose the idea and describe the new method in detail. In Section 4, experiments with face images data are presented to demonstrate the effectiveness of the new method. Conclusions are made in Section 5.

2. Related works

2.1. Outline of MMC

Suppose there are c known pattern classes $\omega_1, \omega_2, \ldots, \omega_c$, the between-class scatter matrix and within-class scatter matrix can be denoted as

$$S_b = \frac{1}{N}\sum_{i=1}^{c} l_i(m_i - m_0)(m_i - m_0)^T \tag{1}$$

$$S_w = \frac{1}{N}\sum_{i=1}^{c}\sum_{j=1}^{l_i}(x_i^j - m_i)(x_i^j - m_i)^T \tag{2}$$

where N is the total number of training samples, and l_i is the number of training samples in class i. In class i, the j-th training sample is denoted by x_i^j, the mean vector of training samples in class i is denoted by m_i and the mean vector of all training samples is m_0.

From the classical Fisher criterion function, we know that when the ratio of the between-class scatter to the within-class scatter is maximized, the samples can be separated easily. In this paper, the MMC based discriminant rule is defined as follows [17], which is based on the difference of between-class scatter matrix and within-class scatter matrix:

$$J_s(w) = \mathrm{tr}(W^T(S_b - S_w)W) \tag{3}$$

By the property of the extreme value of generalized Rayleigh quotient [49], the optimal projection axe is the eigenvector corresponding to the maximal eigenvalue of Eq. (3). In fact, the optimal projection axes $w_1, w_2, \ldots, w_c$ can be selected as the orthonormal eigenvectors corresponding to the first k largest eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_k$, i.e. $(S_b - S_w)w_j = \lambda_j w_j$, where $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_k$.

Comparing MMC with the classical Fisher discriminant criterion, we find that the former avoids calculating the inverse within-class scatter, i.e. $(S_w)^{-1}S_b$ is substituted by $S_b - S_w$. This can not only make the computation more efficient but also avoid the singular problem of the within-class scatter.

2.2. Bidirectional PCA

BDPCA is a straightforward image projection technique where a $k_{col} \times k_{row}$ feature matrix Y of an $m \times n$ image $X(k_{col} \ll m,\ k_{row} \ll n)$ can be obtained by

$$Y = W_{col}^T \times X \times W_{row} \tag{4}$$

where W_{col} is the column projector and W_{row} is the row projector.

Let $\{X_1, X_2, \ldots, X_N\}$ be a training set of N images. By representing the i-th image matrix X_i as an m-set of $1 \times n$ row vectors, the row total scatter matrix S_t^{row} is defined by

$$S_t^{row} = \frac{1}{Nm}\sum_{i=1}^{N}(X_i - \overline{X})^T(X_i - \overline{X}) \tag{5}$$

where $\overline{X}$ is the mean matrix of all training images. We choose the row eigenvectors corresponding to the first k_{row} largest eigenvalues of S_t^{row} to construct the row projector W_{row}

$$W_{row} = [w_1^{row}, w_2^{row}, \ldots, w_{k_{row}}^{row}] \tag{6}$$

By treating an image matrix X_i as an n-set of $m \times 1$ column vectors, the column total scatter matrix S_t^{col} is defined by

$$S_t^{col} = \frac{1}{Nn}\sum_{i=1}^{N}(X_i - \overline{X})(X_i - \overline{X})^T \tag{7}$$

We then choose the column eigenvectors corresponding to the first k_{col} largest eigenvalues of S_t^{col} to construct the column projector W_{col}

$$W_{col} = [w_1^{col}, w_2^{col}, \ldots, w_{k_{col}}^{col}] \tag{8}$$

Note that BDPCA is a generalization of 2DPCA, and 2DPCA can be regarded as a special case of BDPCA with $W_{col} = I_m$, where I_m is an $m \times m$ identity matrix [24].

3. Laplacian bidirectional maximum margin criterion

3.1. Fundamentals

The proposed LBMMC based feature extraction is a straightforward image projection method. In LBMMC, a $k_{col} \times k_{row}$ feature matrix Y of an $m \times n$ image X $(k_{col} \ll m,\ k_{row} \ll n)$ can be obtained by

$$Y = W_{col}^T \times X \times W_{row} \tag{9}$$

where W_{col} is the column projecting matrix and W_{row} is the row projecting matrix. LBMMC based feature extraction has at least three advantages over MMC based feature extraction. First, as a straightforward image projection criterion, LBMMC based feature extraction does not require mapping an image X to an image vector x. Second, LBMMC based feature extraction generally considers the structural information of images that is ignored in MMC based feature extraction. Third, LBMMC uses sample similarity weight to capture the embedding sample manifold structure. Next we discuss how to calculate W_{col} and W_{row}.

For non-Gaussian or manifold-value data, we usually deal with it from local patches because non-Gaussian data can be viewed locally Gaussian and a curved manifold can be viewed locally Euclidean [47,48]. As discussed in Section 1, the weighted summation of distances can capture the underlying clustering of samples and has been successfully used in manifold learning. In this section, furthermore, we use the weighted summation of distance to enhance the robustness of MMC. We formulate the image total Laplacian scatter matrix, image between-class Laplacian scatter matrix and image within-class Laplacian scatter matrix similarity weighting in both row and column directions; and then we propose the LBMMC for feature extraction and recognition.

Let us consider a set of N sample images $X=[X_1,\ldots,X_N]$ taken from a $(m\times n)$ dimensional image space. The similarity between two samples is defined as

$$S_{ij}=\exp(-\|x_i-x_j\|^2/t) \tag{10}$$

where x_i is the vector of image matrix X_i. Obviously, for any x_i, x_j and parameter t, $0\leqslant S_{ij}\leqslant 1$ always holds. Further, the similarity function is a strictly monotonically decreasing function with respect to the distance between two samples X_i and X_j.

The similarity matrix between any two samples in the same class can be defined as follows:

$$H_{ij}=\begin{cases} S_{ij} & \text{if } x_i \text{ and } x_j \text{ belong to the same class} \\ 0 & \text{else} \end{cases} \tag{11}$$

In the row direction, the image total scatter matrix S_t^{row}, the image within-class scatter matrix S_w^{row} and the image between-class scatter matrix S_b^{row} can be formulated as follows:

$$S_t^{row}=\frac{1}{N}\sum_{i=1}^{N}(X_i-\overline{X})^T(X_i-\overline{X})\propto\sum_{i=1}^{N}\sum_{j=1}^{N}(X_i-X_j)^T(X_i-X_j) \tag{12}$$

$$S_w^{row}=\frac{1}{N}\sum_{i=1}^{c}\sum_{j=1}^{l_i}(X_i^j-M_i)^T(X_i^j-M_i) \propto\sum_{k=1}^{c}\sum_{i=1}^{l_k}\sum_{j=1}^{l_k}(X_k^i-X_k^j)^T(X_k^i-X_k^j) \tag{13}$$

$$S_b^{row}=S_t^{row}-S_w^{row} \tag{14}$$

Using the sample similarity weight, the image total Laplacian scatter matrix can be formulated as follows:

$$LS_t^{row}=\sum_{i=1}^{N}\sum_{j=1}^{N}S_{ij}(X_i-X_j)^T(X_i-X_j) \propto\sum_{i,j}(S_{ij}X_i^TX_i-S_{ij}X_i^TX_j)=X^T(L\otimes I_m)X \tag{15}$$

where D is a diagonal matrix with D_{ii} being column (or row) sum of S, $D_{ii}=\sum_j S_{ij}$, and $L=D-S$ is the Laplacian matrix. The weight S_{ij} incurs a heavy penalty when samples X_i and X_j are far apart.

In the row direction, the image within-class Laplacian scatter matrix LS_w^{row}

$$LS_w^{row}=\sum_{k=1}^{c}\sum_{i=1}^{l_k}\sum_{j=1}^{l_k}H_{ij}(X_k^i-X_k^j)^T(X_k^i-X_k^j) =\sum_{i=1}^{N}\sum_{j=1}^{N}H_{ij}(X_k^i-X_k^j)^T(X_k^i-X_k^j) \propto\sum_{i,j}(H_{ij}X_i^TX_i-H_{ij}X_i^TX_j)=X^T(L_w\otimes I_m)X \tag{16}$$

where D_w is a diagonal matrix with D_{wii} being the column (or row) sum of H, $D_{wii}=\sum_j H_{ij}$ and $L_w=D_w-H$ is the Laplacian matrix. The weight H_{ij} incurs a heavy penalty when samples X_i and X_j are far apart.

So in the row direction, the image between-class Laplacian scatter matrix LS_b^{row} is

$$LS_b^{row}=LS_t^{row}-LS_w^{row} \tag{17}$$

In the row direction, the image Laplacian MMC can be defined as follows:

$$J_{row}(w)=tr(W_{row}^T(LS_b^{row}-LS_w^{row})W_{row}) \tag{18}$$

We then choose the column eigenvectors corresponding to the first k_{row} largest eigenvalues of $LS_b^{row}-LS_w^{row}$ to construct the column projector W_{row}

$$W_{row}=[w_1^{row},w_2^{row},\ldots,w_{k_{row}}^{row}] \tag{19}$$

In the column direction, the image total scatter matrix S_t^{col}, the image within-class scatter matrix S_w^{col} and the image between-class scatter matrix S_b^{col} can be formulated as follows:

$$S_t^{col}=\frac{1}{N}\sum_{i=1}^{N}(X_i-\overline{X})(X_i-\overline{X})^T\propto\sum_{i=1}^{N}\sum_{j=1}^{N}(X_i-X_j)(X_i-X_j)^T \tag{20}$$

$$S_w^{col}=\frac{1}{N}\sum_{i=1}^{c}\sum_{j=1}^{l_i}(X_i^j-M_i)(X_i^j-M_i)^T\propto\sum_{k=1}^{c}\sum_{i=1}^{l_k}\sum_{j=1}^{l_k}(X_k^i-X_k^j)(X_k^i-X_k^j)^T \tag{21}$$

$$S_b^{col}=S_t^{col}-S_w^{col} \tag{22}$$

Using the sample similarity weight, the image total Laplacian scatter matrix LS_t^{col}, the image within-class Laplacian scatter matrix LS_w^{col} and the image between-class scatter matrix LS_b^{col} in the column direction can be formulated as follows:

$$LS_t^{col}=\sum_{i=1}^{N}\sum_{j=1}^{N}S_{ij}(X_i-X_j)(X_i-X_j)^T \propto\sum_{i,j}(S_{ij}X_iX_i^T-S_{ij}X_iX_j^T)=X(L\otimes I_n)X^T \tag{23}$$

$$LS_w^{col}=\sum_{k=1}^{c}\sum_{i=1}^{l_k}\sum_{j=1}^{l_k}H_{ij}(X_k^i-X_k^j)(X_k^i-X_k^j)^T =\sum_{i=1}^{N}\sum_{j=1}^{N}H_{ij}(X_k^i-X_k^j)(X_k^i-X_k^j)^T \propto\sum_{i,j}(H_{ij}X_iX_i^T-H_{ij}X_iX_j^T)=X(L_w\otimes I_n)X^T \tag{24}$$

$$LS_b^{col}=LS_t^{col}-LS_w^{col} \tag{25}$$

In the column direction, the image Laplacian MMC can be defined as follows:

$$J_{col}(w)=\mathrm{tr}(W_{col}^T(LS_b^{col}-LS_w^{col})W_{col}) \tag{26}$$

The optimal projection axes $w_1^{col},w_2^{col},\ldots,w_k^{col}$ can be selected as the orthonormal eigenvectors corresponding to the first k largest eigenvalues $\lambda_1,\lambda_2,\ldots,\lambda_k$, i.e., $(LS_b^{col}-LS_w^{col})w_j=\lambda_jw_j$, where $\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_k$. Note that any operation in column direction can be equivalently implemented by an operation in row direction by virtue of the transform operation of matrix.

So far, we have the projection matrix W_{col} and W_{row} for LBMMC. Note that if we set $W_{col}=I_m$, LBMMC will be exactly the Laplacian

2DMMC. Thus the Laplacian 2DMMC is a special case of the proposed LBMMC.

3.2. Feature extraction algorithm based on LBMMC

The proposed LBMMC based feature extraction algorithm can be summarized as follows:

Step 1: Construct the similarity matrix S and H using Eqs. (10) and (11).

Step 2: Calculate the row direction projection matrix W_{row} using Eqs. (15)–(17).

Calculate the column direction projection matrix W_{col} using Eqs. (24)–(26).

Step 3: Extract the sample feature using Eq. (9) and classify.

3.3. Connections with LLD, DLPP, ANMM and TSA

It is not hard to see that LBMMC is BDMMC if t approaches the positive infinity in the similarity functions (9). The feature extraction based on LBMMC has connections with Zhao et al.'s LLD [50], Wang et al.'s ANMM [51] and Yu et al.'s DLPP [52]. ANMM takes advantage of the partial structural information of classes and neighborhoods of samples at the same time while LBMMC based feature extraction purely explores the information of classes for discrimination. It is always difficult to decide the number of nearest neighbors of each sample and the number of shortest pairs from different classes in ANMM. LLD and DLPP ignore the structural information of images and need to transform image matrix into vector; LLD and DLPP have the singularity problem of the within-class scatter matrix; LLD and DLPP have higher computational complexity than the proposed LBMMC because they must calculate the inverse within-class scatter matrix.

One may think that LBMMC is very similar to TSA [27]. Actually they are very different. In TSA, a nearest neighbor graph is used to model the local geometrical structure. In the proposed method, we use the weighted distance sum of any two data points to preserve the local geometrical structure. In TSA, the image structural information is preserved by using the tensor space. In our proposed method, we directly work on image matrix to preserving the image structural information. The objective function in TSA is trace ratio criterion and it is iteratively solved. In LBMMC, our objective function is trace difference criterion and can be solved by generalized eigenvalue decomposition. There is no convergence problem in our proposed method. TSA needs to calculate an inverse matrix, while the proposed method does not need to calculate the inverse matrix. Thus LBMMC needs less computational cost than TSA. TSA is the extended version of LPP in tensor space, while LBMMC is the extended version of MMC in non-Euclidean space. TSA is formulated in the tensor space, while LBMMC is formulated in the image matrix. In TSA, there are no explicit rules to choose the projection vectors, while in LBMMC we can simply choose the projection vectors according to the corresponding eigenvalues.

4. Experimental results

The features of 2DPCA, BDPCA and LBMMC are matrices. The distance of two feature matrices can be calculated using either a vector-based or matrix-based method [24]. In a vector-based method, a feature matrix is first mapped to a vector and then a vector-based distance measure is used. In a matrix-based method, the distance between two feature matrices can be directly computed. In the experiments, we use the vector-based method to calculate the distance of tow feature matrices. The experiments are implemented on an AMD Athlon(tm) 64 Processor 3000+ Lenovo Computer with 512M RAM and the programming environment is MATLAB (Version 7.01).

Fig. 1. Images of one person in FERET.

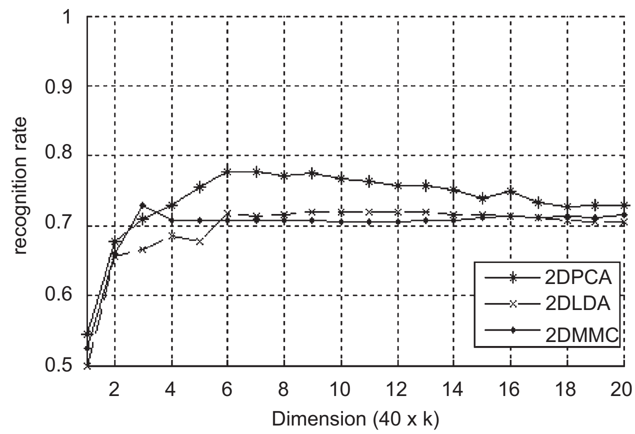

Fig. 2. The recognition rates of 2DPCA, 2DLDA, 2DMMC versus the dimensions.

4.1. Experiments on FERET database

The FERET face image database is a result of the FERET program, which was sponsored by the Department of Defense of US through the DARPA Program [53,54]. It has become a standard database for testing and evaluating state-of-the-art face recognition algorithms.

The proposed algorithm is tested on a subset of the FERET database. This subset includes 1400 images of 200 individuals (each individual has seven images). The subset involves variations in facial expression, illumination and pose. In our experiment, the facial portion of each original image was automatically cropped based on the location of eyes and the cropped image was resized to 40×40 pixels. Some example images of one person are shown in Fig. 1. In the experiment, we use the first four images per class for training and the remaining images for testing.

4.1.1. Selection of the projection axis

First, 2DPCA, 2DLDA and 2DMMC method are used for feature extraction. The number of selected eigenvectors (projection vectors) varies from 1 to 20. Here, denote by k the number of projection vectors, then the dimension of corresponding projected feature vector is $40 \times k$. Finally, a nearest neighbor classifier with cosine distance is employed to classify in the projected feature space. The recognition rates versus k are shown in Fig. 2. From Fig. 2, we can see that 2DMMC achieves the top recognition rate when k equals to 3 and 2DPCA has better performance in most cases than 2DLDA and 2DMMC.

Next, LBMMC is used for feature extraction. Let $k = 3, 4, 5, 6, 7, 8$, and denote by m the number of projection vectors in the second step of feature extraction (in the column direction). We vary m from 2 to 40 (2:2:40). A nearest neighbor classifier with cosine distance is employed for classification too. We can see that the proposed method achieves the top recognition rate when $k=6$, $m=28$. The recognition rates of the proposed method versus dimension are shown in Fig. 3. In the proposed method, the parameter t of the similarity is set as $t = 100$. Fig. 4 shows the recognition rate of the proposed method versus the parameter t. It can been seen that the it has a stable performance with various values of parameter t.

4.1.2. Comparison of the performance

In this test, PCA [5], LDA [5], 2DPCA [17], 2DLDA [18], BDPCA [24], TSA [27] and the proposed method are used for feature extraction. Note that LDA involves a PCA phase. In this phase, we keep nearly 98% image energy and select the number of principal components as $m = 250$. In the TSA algorithm, the number of iterations is taken as 10. After feature extraction, a nearest neighbor classifier with cosine distance is employed for classification. The maximal recognition rate of each method and the corresponding dimension and the running time of each phase are listed in Table 1.

From Table 1, we can see that the proposed method has the top recognition rate. LBMMC can be used for image feature extraction by reducing the dimensionality in both column and row directions, and it considers the structural information embedded in the original images. Moreover, the proposed method considers the distribution information of the original images using the sample similarity weight, which has been widely used in manifold learning.

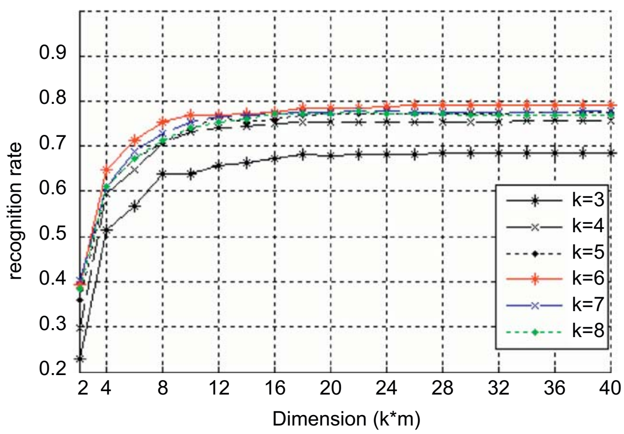

Fig. 3. The recognition rates of proposed versus the dimensions.

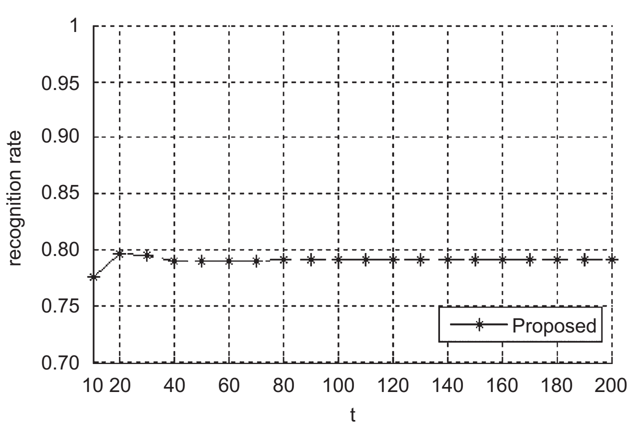

Fig. 4. The recognition rates of proposed versus the parameter *t*.

In the second experiment, 10-fold cross-validation tests are performed to reevaluate the performance of PCA, 2DPCA, LDA, 2DLDA, MMC, 2DMMC, BDPCA, TSA, and the proposed method LBMMC. In each test, four images per class are randomly chosen for training, while the remaining eight images are used for testing. All possible dimensions of the final low-dimensional representation are evaluated, and the best results are reported in Table 2. Table 2 presents the maximal average recognition rates across 10 runs for each method under the nearest neighbor classifier with cosine distance metrics. The corresponding standard deviations (std) are also presented in Table 2.

From Table 2, it can be seen that proposed method outperforms the other methods. It is worthwhile to note that there are many variants of the FERET database. It may not be enough to capture the within-class and between-class distribution information to use only four images of each class for training. Thus the LDA (Fisherface) method may work worse than PCA (Eigenfaces). This is consistent with the observation in [55] that Eigenface can outperform Fisherface when the training set is small.

4.2. Experiments on Yale database

The Yale face database contain 165 images of 15 individuals (each person providing 11 different images) under various facial expressions and lighting conditions. In our experiments, each image was manually cropped and resized to 100×80. Fig. 5 shows sample images of one person.

In the first experiment, we use the first three images per class for training and the remaining images for testing. PCA, LDA, 2DPCA, 2DLDA, BDPCA, TSA and the proposed LBMMC are used for feature extraction. In PCA and the PCA stage of LDA, we kept nearly 98% image energy and selected the number of principal components as $m = 34$. In the TSA algorithm, the number of iterations is taken to be 10. In the proposed method, the parameter t of the similarity is set as $t = 100$. Finally, a nearest neighbor classifier with cosine distance is employed. The final recognition rates are given in Table 3. As can be seen, the proposed method has the best recognition rate.

In the second experiment, 10-fold cross-validation tests are performed to evaluate the performance of PCA, 2DPCA, LDA, 2DLDA, MMC, 2DMMC, BDPCA, TSA, and the proposed method LBMMC. In each test, three images per class are randomly chosen for training, while the remaining eight images are used for testing. All possible dimensions of the final low-dimensional representation are evaluated, and the best results are reported in Table 4, which lists the maximal average recognition rates across 10 runs of each method under the nearest neighbor classifier with cosine distance metrics

Table 1
Recognition rate on the FERET database.

	PCA	LDA	MMC	2DPCA	2DLDA	2DMMC	BDPCA	TSA	LBMMC
Recognition rate	0.4967	0.6550	0.5017	0.7767	0.7200	0.7300	0.7783	0.6533	**0.7967**

Table 2
Average recognition rates and standard deviations on the FERET database.

	PCA	LDA	MMC	2DPCA	2DLDA	2DMMC	BDPCA	TSA	LBMMC
Recognition mean	0.5770	0.4563	0.6910	0.7570	0.6503	0.7470	0.6395	0.5663	**0.7647**
Recognition std	0.0938	0.904	0.1262	0.0580	0.1207	0.0744	0.1068	0.1115	**0.0550**

Fig. 5. Eleven images of one person in Yale.

Table 3
Recognition rate on the Yale database.

	PCA	LDA	MMC	2DPCA	2DLDA	2DMMC	BDPCA	TSA	LBMMC
Recognition rate	0.8417	0.8417	0.8917	0.8833	0.8750	0.8750	0.8917	0.8833	**0.9250**

Table 4
Average recognition rates and standard deviations on the Yale database.

	PCA	LDA	MMC	2DPCA	2DLDA	2DMMC	BDPCA	TSA	LBMMC
Recognition mean	0.8375	0.8667	0.8533	0.9017	0.9167	0.9150	0.8917	0.8447	**0.9317**
Recognition std	0.0343	0.0494	0.0418	0.0196	0.0192	0.0283	0.0269	0.0366	**0.0204**

and their corresponding standard deviations (std). From Table 4, it can be seen that proposed method outperforms other methods.

4.3. Evaluation of the experiments results

Is the proposed method statistically better than other methods in terms of its recognition rate? To answer this question, let us evaluate the experimental results in Table 3 using McNemar's [56–58] significance test. McNemar's test is essentially a null hypothesis statistical test based on a Bernoulli mode. If the resulting p-value is below the desired significance level (for example, 0.05), the null hypothesis is rejected and the performance difference between two algorithms is considered to be statistically significant. In the FERET and Yale databases, our proposed method outperforms TSA in the 10-fold cross-validation tests ($p = 0.000079343$ and 0.00013066, respectively). By this test, we find that the proposed method statistically significantly outperforms TSA.

5. Conclusions

We proposed a new discriminant subspace learning method, namely LBMMC, which extends MMC into non-Euclidean space. LBMMC conducts the graph embedding to keep the data relationships measured by weighted distance sum of any two data points, and it is directly performed on the image matrix to preserve the image matrix structural information. Therefore, LBMMC learns multiple interrelated subspaces to obtain a low-dimensional data representation reflecting class label information, image structural information and image intrinsic manifold structure. LBMMC avoids the calculation of inverse matrix. It was found that Laplacian 2DMMC is a special case of LBMMC. The experimental results on FERET and Yale face databases are very encouraging.

Acknowledgments

This project is supported by NSFC of China (Grant nos: 60632050, 60503026 and 60473039), Hi-Tech Research and Development Program of China (Grant no. 2006AA01Z119) and the Graduates' Research Innovation Program of Higher Education of Jiangsu Province (Grant no. CX07B_118z). We thank Prof. J.Yang for his help.

References

[1] A.K. Jain, B. Chandrasekaran, Dimensionality and sample size considerations in pattern recognition practice, in: Handbook of Statistics, North Holland, Amsterdam, 1982.
[2] R.O. Duda, P.E. Hart, Pattern Classification and Scene Analysis, Wiley, New York, 1973.
[3] S.J. Raudys, A.K. Jain, Small sample size effects in statistical pattern recognition: recommendations for practitioners, IEEE Transactions on Pattern Analysis and Machine Intelligence 13 (3) (1991) 252–264.
[4] D.L. Swets, J. Weng, Using discriminant eigenfeatures for image retrieval, IEEE Transactions on Pattern Analysis and Machine Intelligence 18 (8) (1996) 831–836.
[5] V. Belhumeur, J. Hespanha, D. Kriegman, Eigenfaces vs. Fisherfaces: recognition using class specific linear projection, IEEE Transactions on Pattern Analysis and Machine Intelligence 19 (7) (1997) 711–720.
[6] J. Yang, J.Y. Yang, Why can LDA be performed in PCA transformed space?, Pattern Recognition 36 (2) (2003) 563–566.
[7] Z.Q. Hong, J.Y. Yang, Optimal discriminant plane for a small number of samples and design method of classifier on the plane, Pattern Recognition 24 (4) (1991) 317–324.
[8] J.H. Friedman, Regularized discriminant analysis, Journal of the American Statistical Association 84 (1989) 165–175.
[9] T. Hastie, R. Tibshirani, Penalized discriminant analysis, The Annals of Statistics 23 (1995) 73–102.
[10] T. Hastie, R. Tibshirani, A. Buja, Flexible discriminant analysis by optimal scoring, Journal of the American Statistical Association 89 (1994) 1255–1270.
[11] L.F. Chen, H.Y.M. Liao, M.T. Ko, G.J. Yu, A new LDA-based face recognition system which can solve the small sample size problem, Pattern Recognition 33 (1) (2000) 1713–1726.
[12] H. Yu, J. Yang, A direct LDA algorithm for high dimensional data—with application to face recognition, Pattern Recognition 34 (10) (2001) 2067–2070.
[13] Q.X. Gao, L. Zhang, D. Zhang, Face recognition using FLDA with single training image per-person, Applied Mathematics and Computation 205 (12) (2008) 726–734.
[14] X.S. Zhuang, D.Q. Dai, Inverse Fisher discriminant criteria for small sample size problem and its application to face recognition, Pattern Recognition 38 (11) (2005) 2192–2194.
[15] X.S. Zhuang, D.Q. Dai, Improved discriminant analysis for high-dimensional data and its application to face recognition, Pattern Recognition 40 (5) (2007) 1570–1578.
[16] Z. Jin, J.Y. Yang, Z. Hu, Z. Lou, Face recognition based on the uncorrelated discrimination transformation, Pattern Recognition 34 (7) (2001) 1405–1416.
[17] H. Li, T. Jiang, K. Zhang, Efficient and robust feature extraction by maximum margin criterion, IEEE Transactions on Neural Networks 17 (1) (2006) 1157–1165.
[18] J. Yang, D. Zhang, A.F. Frangi, J.Y. Yang, Two-dimensional PCA: a new approach to appearance-based face representation and recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence 26 (1) (2004) 131–137.
[19] M. Li, B. Yuan, 2D-LDA: a statistical linear discriminant analysis for image matrix, Pattern Recognition Letters 26 (2005) 527–532.
[20] J. Yang, D. Zhang, Y. Xu, J.J. Yang, Two-dimensional discriminant transform for face recognition, Pattern Recognition 38 (2005) 1125–1129.
[21] H.L. Xiong, M.N.S. Swanmy, M.O. Ahmad, Two-dimensional FLD for face recognition, Pattern Recognition 38 (2005) 1121–1124.
[22] X.Y. Jing, H.S. Wong, D. Zhang, Face recognition based on 2D Fisherface approach, Pattern Recognition 39 (2006) 707–710.

[23] W.S. Zheng, J.H. Lai, S.Z. Li, 1D-LDA vs. 2D-LDA: when is vector-based linear discriminant analysis better than matrix-based?, Pattern Recognition 41 (7) (2008) 2156–2172.
[24] W.M. Zuo, D. Zhang, K.Q. Wang, Bidirectional PCA with assembled matrix distance metric for image recognition, IEEE Transactions on Systems, Man, and Cybernetics—Part B 36 (4) (2006) 863–872.
[25] W.M. Zuo, D. Zhang, J. Yang, K.Q. Wang, BDPCA plus LDA: a novel fast feature extraction technique for face recognition, IEEE Transactions on Systems, Man, and Cybernetics—Part B 36 (4) (2006) 946–953.
[26] Q.X. Gao, L. Zhang, D. Zhang, Sequential row–column independent component analysis for face recognition, Neurocomputing 72 (4–6) (2009) 1152–1159.
[27] X.F. He, D. Cai, P. Niyogi, Tensor subspace analysis, in: Advances in Neural Information Proceeding Systems, vol. 18, Vancouver, Canada, 2005.
[28] D.C. Tao, X.L. Li, X.D. Wu, et al., General tensor discriminant analysis and gabor features for gait recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence 29 (10) (2007) 1700–1715.
[29] D.C. Tao, X.L. Li, W.M. Hu, et al., Supervised tensor learning, Knowledge and Information Systems (KAIS) 13 (1) (2007) 1–42.
[30] L. Zhang, Q.X. Gao, D. Zhang, Directional independent component analysis with tensor representation, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2008), 2008.
[31] J.B. Tenenbaum, V. de Silva, J.C. Langford, A global geometric framework for nonlinear dimensionality reduction, Science 290 (2000) 2319–2323.
[32] S.T. Roweis, L.K. Saul, Nonlinear dimension reduction by locally linear embedding, Science 290 (2000) 2323–2326.
[33] M. Belkin, P. Niyogi, Laplacian eigenmaps for dimensionality reduction and data representation, Neural Computation 15 (6) (2003) 1373–1396.
[34] X. He, S. Yan, Y. Hu, H. Zhang, Learning a locality preserving subspace for visual recognition, in: Proceedings of the Ninth International Conference on Computer Vision, France, October 2003, pp. 385–392.
[35] X. He, S. Yan, Y. Hu, P. Niyogi, H. Zhang, Face recognition using Laplacian faces, IEEE Transactions on Pattern Analysis and Machine Intelligence 27 (3) (2005) 328–340.
[36] J. Yang, D. Zhang, J.Y. Yang, B. Niu, Globally maximizing, locally minimizing: unsupervised discriminant projection with applications to face and palm biometrics, IEEE Transactions on Pattern Analysis and Machine Intelligence 29 (4) (2007) 650–664.
[37] S. Yan, D. Xu, B. Zhang, H.-J. Zhang, Graph embedding: a general framework for dimensionality reduction, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2005, pp. 830–837.
[38] S. Yan, D. Xu, B. Zhang, H.-J. Zhang, Graph embedding and extensions: a general framework for dimensionality reduction, IEEE Transactions on Pattern Analysis and Machine Intelligence 29 (1) (2007) 40–51.
[39] H.T. Chen, H.W. Chang, T.L. Liu, Local discriminant embedding and its variants, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR2005), vol. 2, 2005, pp. 846–853.
[40] Y.W. Pang, Y. Yuan, X.L. Li, Effective feature extraction in high-dimensional space, IEEE Transactions on Systems, Man, and Cybernetics—Part B 38 (6) (2008) 1652–1656.
[41] H. Wang, S.C. Yan, D. Xu, X.O. Tang, T.S. Huang, Trace ratio vs. ratio trace for dimensionality reduction, in: 2007 IEEE Conference on Computer Vision and Pattern Recognition (CVPR2007), vols. 1–8, 2007, pp. 108–115.
[42] H. Wang, S.C. Yan, T.S. Huang, X. Tang, A convergent solution to tensor subspace learning, in: IJCAI, 2007, pp. 629–634.
[43] Y. Fu, T.S. Huang, Image classification using correlation tensor analysis, IEEE Transactions on Image Processing 17 (2) (2008) 226–234.
[44] Y. Fu, S.C. Yan, T.S. Huang, Correlation metric for generalized feature extraction, IEEE Transactions on Pattern Analysis and Machine Intelligence 30 (12) (2008) 2229–2235.
[45] Y. Fu, S.C. Yan, T.S. Huang, Classification and feature extraction by simplexization, IEEE Transactions on Information Forensics and Security 3 (1) (2008) 91–100.
[46] J. Yang, Z. Lou, Z. Jin, J.Y. Yang, Minimal local reconstruction error measure based discriminant feature extraction and classification, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2008), vols. 1–12, 2008, pp. 173–178.
[47] M. Kirby, L. Sirovich, Application of the KL procedure for the characterization of human faces, IEEE Transactions on Pattern Analysis and Machine Intelligence 12 (1) (1990) 103–108.
[48] J.M. Lee, Riemannian Manifolds: An Introduction to Curvature, Springer, Berlin, 1997.
[49] X.D. Zhang, Matrix Analysis and Application, Tsinghua University Press, Beijing, 2004 (in Chinese).
[50] D.L. Zhao, Z.C. Liu, R. Xiao, X.O. Tang, Linear Laplacian discrimination for feature extraction, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR2007), vols. 1–8, 2007, pp. 1181–1187.
[51] F. Wang, C.S. Zhang, Feature extraction by maximizing the average neighborhood margin, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR2007), vols. 1–8, 2007, pp. 1173–1180.
[52] W.W. Yu, X.L. Teng, C.Q. Liu, Face recognition using discriminant locality preserving projection, Image and Vision computing 24 (2006) 239–248.
[53] P.J. Phillips, H. Moon, S.A. Rizvi, P.J. Rauss, The FERET evaluation methodology for face recognition algorithms, IEEE Transactions on Pattern Analysis and Machine Intelligence 22 (10) (2000) 1090–1104.
[54] P.J. Phillips, The facial recognition technology (FERET) database ⟨http://www.itl.nist.gov/iad/humanid/feret/feret_master.html⟩, 2004.
[55] A.M. Martinez, A.C. Kak, PCA versus LDA, IEEE Transactions on Pattern Analysis and Machine Intelligence 23 (2) (2001) 228–233.
[56] W. Yambor, B. Draper, R. Beveridge, Analyzing PCA-based face recognition algorithms: eigenvectors selection and distance measures, in: H. Christensen, J. Phillips (Eds.), Empirical Evaluation Methods in Computer Vision, World Scientific, Singapore, 2002.
[57] J. Devore, R. Peck, Statistics: The Exploration and Analysis of Data, third ed., Brooks Cole, 1997.
[58] B.A. Draper, K. Baek, M.S. bartlett, J.R. Beveridge, Recognizing faces with PCA and ICA, Computer Vision and Image Understanding 91 (1–2) (2003) 115–137.

About the Author—WANKOU YANG is a Ph.D. Candidate at the School of Computer Science and Technology, Nanjing University of Science and Technology, P.R. China. He received the B.S. and M.S. degrees from the Department of Computer Science, Nanjing University of Science and Technology, P.R. China, in 2002 and 2004. His research interests include pattern recognition, computer vision and digital image processing.

About the Author—JIANGUO WANG is an Assistant Professor in the Network & Education, Center Tangshan College. He received the B.S. degree in Printing Technology from Xi'an University of Technology in 1994, received the M.S. degree in Signal and Information Processing from Xi'an University of Technology in 1997. And he received the Ph.D. degree in Computer Vision & Pattern Recognition from NUST in 2008. His current interests are in the areas of pattern recognition, image processing, face detection and recognition.

About the Author—MINGWU REN is a Professor at the School of Computer Science and Technology, Nanjing University of Science and Technology, P.R. China. He received M.S. and Ph.D. degrees in Pattern Recognition and Computer Image Processing from Nanjing University of Science and Technology, P.R. China. His research interests include computer vision, digital image processing, pattern recognition and artificial control.

About the Author—JINGYU YANG received the B.S. degree in Computer Science from Nanjing University of Science and Technology (NUST), Nanjing, China. From 1982 to 1984, he was a Visiting Scientist at the Coordinated Science Laboratory, University of Illinois at Urbana-Champaign. From 1993 to 1994, he was a Visiting Professor at the School of Computer Science and Technology, Missuria University. In 1998, he acted as a Visiting Professor at Concordia University in Canada. He is currently a Professor and Chairman in the Department of Computer Science at NUST. He is the author of more than 300 scientific papers in computer vision, pattern recognition and artificial intelligence. He has won more than 20 provincial awards and national awards. His current research interests are in the areas of pattern recognition, robot vision, image processing, data fusion and artificial intelligence.

About the Author—LEI ZHANG received the B.S. degree in 1995 from Shenyang Institute of Aeronautical Engineering, Shenyang, PR China, the M.S. and Ph.D. degrees in Electrical and Engineering from Northwestern Polytechnical University, Xi'an, PR China, respectively, in 1998 and 2001. From 2001 to 2002, he was a research associate in the Dept. of Computing, The Hong Kong Polytechnic University. From January 2003 to January 2006 he worked as a Postdoctoral Fellow in the Dept. of Electrical and Computer Engineering, McMaster University, Canada. Since January 2006, he has been an Assistant Professor in the Dept. of Computing, The Hong Kong Polytechnic University. His research interests include image and video processing, biometrics, pattern recognition, multisensor data fusion and optimal estimation theory, etc.

About the Author—GUANGHAI LIU is currently an Associate Professor in the College of Computer Science and Information technology at Guangxi Normal University, Guangxi, China. He received his Ph.D. degree from the School of Computer Science and Technology at Nanjing University of Science and Technology (NUST). His current research interests are in the areas of image processing, pattern recognition and artificial intelligence.