A ROBUST FEATURE EXTRACTION FRAMEWORK FOR FACE RECOGNITION

Guang Dai [1], *Yuntao Qian* [2]
[1] College of Information Science & Engineering, Wenzhou University, P.R.China
[2] College of Computer Science, Zhejiang University, P.R.China

ABSTRACT

The kernel fractional-step nonlinear discriminant analysis (KF-NDA) method not only extends the fractional-step linear discriminant analysis (F-LDA) method to a nonlinear version, but also further improves the generalization ability of traditional kernel non-linear discriminant analysis (K-NDA). On the other hand, the Gabor transformed face images exhibit strong characteristics of spatial locality, scale and orientation selectivity, similar to those displayed by Gabor wavelets. Such characteristics produce salient local features that are most suitable for face recognition (FR). Hence, the augmented Gabor feature vector (AGFV) derived from a set of downsampled Gabor wavelet representations of face images is robust to the various of face images, and simultaneously exhibits the more discriminatory information. Based on the AGFV and the KF-NDA, a robust feature extraction framework, i.e., the Gabor KF-NDA (GKF-NDA), is proposed for FR. In this framework, the KF-NDA method is directly applied to extract the robust nonlinear feature from the AGFV. Experimental results tested on the popular databases show that the GKF-NDA is more effective than other existing FR approaches.

1. INTRODUCTION

Linear discriminant analysis (LDA) is a powerful tool used for dimensionality reduction and feature extraction in the face recognition (FR) system. However, the classification performance of classical LDA is often degraded by the fact that the Fisher discriminant criterion defined in the corresponding LDA is not linked to classification accuracy in the output space. To effectively overcome this problem, a fractional-step LDA (F-LDA) method has been proposed by Kothari et al [2]. In this method, a weighted between-class scatter matrix is constructed for the discriminant criterion, then the dimensionality is reduced in small fractional steps making the relevant distances be more accurately weighted. As same as classical LDA, the F-LDA cannot be directly applied to solve the small sample size problem, which is often encountered in FR tasks. To effectively solve this problem, some direct LDA (D-LDA) methods have been proposed [4,5]. Most recently, a more effective method, called direct fractional-step LDA (DF-LDA) method [6], has been proposed. This method combines the strengths of the D-LDA and F-LDA approaches, and it not only overcomes the limitation of D-LDA that the discriminant criterion is not related to classification accuracy, but also makes the F-LDA carried out in high-dimensional spaces. However, it is generally believed that the face images data distribution in practice is nonlinear and complex because of illumination, facial expression and pose variations. Hence, the DF-LDA method is also inadequate to describe the complex and nonlinear distribution, because it is still a linear technique in nature as same as other LDA [2-5].

Recently, the extensions of linear methods to nonlinear ones, using the kernel trick, have been given more attention in pattern recognition due to the high performance. As the nonlinear extension of LDA, the kernel nonlinear discriminant analysis (K-NDA) has already been shown to provide a better performance than LDA in several applications [11-13]. However, as similar as classical LDA, the generalization ability of K-NDA is degraded by the fact that the kernel Fisher discriminant criterion is not directly linked to the classification accuracy.

On the other hand, the Gabor wavelet seems to be a good approximation to the filter response profiles encountered experimentally in cortical neurons [7]. The Gabor wavelet representation of facial images exhibits strong characteristics of spatial locality, scale and orientation selectivity. As a result, these images produce salient local features that are most suitable for the FR. It is generally believed that the Gabor wavelet representation of face images should be robust to variations due to facial expression and pose changes. Most recently, Liu et al [8] proposed an augmented Gabor feature vector (AGFV) derived from a set of downsampled Gabor wavelet representations of face images for FR, and it was very robust to the variations of face images, simultaneously strengthened the discriminatory information between the different classes.

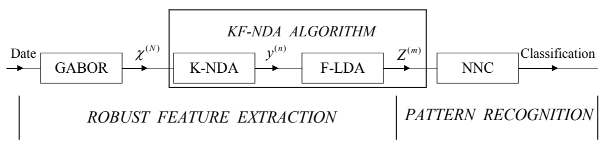


Fig.1. Face recognition system architecture based on the robust feature extraction framework. GABOR: Gabor wavelet representation, KF-NDA, K-NDA, F-LDA, NNC: nearest neighbor classifier.

In this paper, a kernel fractional-step nonlinear discriminant analysis (KF-NDA) method, which combines the characteristics of the K-NDA and the F-LDA, has been introduced to further improve the generalization ability of K-NDA. It not only overcomes the limitation of K-NDA that the Fisher criterion defined in the K-NDA is not linked to classification accuracy, but extends the linear F-LDA method to a nonlinear one. In KF-NDA method, we first carry out a variant of K-NDA to obtain a low-dimensional subspace, where almost all classes of objects are linearly separable. Then, a weighting function is also introduced into the variant of K-NDA, so that a subsequent F-LDA step can be directly applied to carefully reorient the low-dimensional subspace leading to a set of optimal discriminant features for pattern representation. Based on the KF-NDA and the AGFV, a robust feature extraction framework, i.e., the Gabor KF-NDA (GKF-NDA), is proposed for FR, whose system architecture is shown in Fig.1. In this framework, we apply the KF-NDA method to extract the robust nonlinear feature from the AGFV that can exhibit the more discriminatory information on the basis of Gabor wavelet representation of face images. Experimental results tested on the popular databases show that the GKF-NDA is very effective for FR tasks.

2. GABOR FEATURE ANALYSIS

The Gabor wavelets have been used extensively in image processing, texture analysis because of their biological relevance and computational properties [7]. The Gabor wavelets can capture the properties of spatial localization, orientation selectivity, spatial frequency selectively, and quadrature phase relationship, and its representation has been shown to be optimal in sense of minimizing the joint two-dimensional uncertainty in space and frequency [7]. The Gabor wavelets have been found to be particularly suitable for image decomposition and representation when the goal is the derivation of local and discriminatory features. It is generally believed that the face's Gabor wavelets representation should have robust characteristics in facial expression and pose changes.

The two-dimensional Gabor wavelets can be defined:

$$\psi_{u,v}(z) = \frac{\| k_{u,v} \|^2}{\sigma^2} e^{-\frac{\|k_{u,v}\|^2 \|z\|^2}{2\sigma^2}} [e^{ik_{u,v}z} - e^{-\frac{\sigma^2}{2}}] \quad (1)$$

$$k_{u,v} = k_v e^{i\phi_u} \quad (2)$$

where u and v denote the orientation and scale of the Gabor wavelet, $z = (x, y)$, $\| \cdot \|$ denotes the norm operator, $k_v = k_{\max} / f^v$, and $\phi_u = \pi u / 8$. f is the spacing factor between kernels in the frequency domain [8]. The Gabor wavelet function can form a complete but nonorthogonal basis set.

Given any image $I(x, y)$, its Gabor wavelet transformation is then defined:

$$W_{mn}(x, y) = \int I(x_1, y_1) \psi^*_{mn}(x - x_1, y - y_1) dx_1 dy_1 \quad (3)$$

where * indicates the complex conjugate. In most cases the Gabor wavelet transformation of a face image is at the five different scales, $m \in \{0, \ldots, 4\}$, and eight different orientations, $n \in \{0, \ldots, 7\}$. Let W_{mn} denote a Gabor wavelet transformation of face image. The augmented Gabor feature vector (AGFV) of face images is defined [8]:

$$\chi^{(\rho)} = (W_{0,0}^{(\rho)t}, W_{0,1}^{(\rho)t}, \ldots, W_{4,7}^{(\rho)t})^t \quad (4)$$

where t is the transpose operator and ρ is the downsampling factor. The AGFV can encompass all the Gabor transformations of face images, and it can exhibit important discriminatory information, and simultaneously has robustness against varying expression and rotation in the face image.

3. KERNEL FRACTIONAL-STEP NONLINEAR DISCRIMINANT ANALYSIS

In order to further improve the generalization ability of K-NDA, a KF-NDA is described in this section, and simultaneously is applied to extract the robust nonlinear feature for FR. The KF-NDA has two steps: 1) firstly, a low-dimensional subspace is obtained by the variant of K-NDA step; 2) secondly, an F-LDA step is directly applied to further reduce the dimensionality of this subspace from l to l' .

3.1. The Variant of K-NDA

The basic idea of K-NDA is to first map the input date x into a feature space F via a nonlinear mapping ϕ and then perform a LDA in F . Moreover, it is unnecessary to compute explicitly in F , and only compute the inner product of two vectors in F with an inner product kernel function:

$$k(x, y) = (\phi(x)^T \cdot \phi(y)) . \quad (5)$$

However, as same as classical LDA, the discriminant criterion in classical K-NDA is not directly linked to the classification error in the output space. According to the F-LDA [2], a weighed Fisher discriminant criterion in F , where a weighted between-class scatter matrix in F is used to replace the conventional between-class scatter matrix in F , can be used to solve this problem. The motivation of course being that classes, which are closer together in the input feature space F , are more likely to have more nearness or confusion in the output space and decrease the generalization ability of the K-NDA, so they should be more heavily weighted in the input feature space F . Let x be a vector of the training set X with N elements, X_i designs subsets of X with N_i elements, and $X = \bigcup_{i=1}^{c} X_i$, $X = \sum_{i=1}^{c} X_i$, $N = \sum_{i=1}^{c} N_i$, where c is the number of classes. According to [9], a weighted between-class scatter matrix in F can be defined as:

$$\hat{S}_b = \sum_{i=1}^{c-1} \sum_{j=i+1}^{c} (N_i \cdot N_j / N^2) w(d_{i,j})(m_i - m_j)(m_i - m_j)^T , \quad (6)$$

where $d_{i,j}$ is the Euclidean distance between the means of class i and class j in F , $m_i = (1/N_i) \sum_{x_j \in X_i} \phi(x_j)$ is the mean of class i in F , the weighting function $w(d_{i,j})$ is generally a monotonically decreasing function. According to [2], the weight should drop faster than the Euclidean distance between the means of class i and class j in F ; as a result, only constraint of the weighting function is $w(d_{i,j}) = d_{i,j}^{-p}$, where $p \in \mathbb{N}$ and $p \geq 3$. In addition, it is clear that $d_{i,j}$ in F can be calculated by the kernel trick as follows:

$$d_{i,j} = \| m_i - m_j \|$$

$$= \sqrt{(\sum_{i_1 \in X_i} \frac{\phi(x_{i_1})}{N_i} - \sum_{j_1 \in X_j} \frac{\phi(x_{j_1})}{N_j})^T (\sum_{i_1 \in X_i} \frac{\phi(x_{i_1})}{N_i} - \sum_{j_1 \in X_j} \frac{\phi(x_{j_1})}{N_j})} \quad (7)$$

$$= \sqrt{\sum_{i_1, i_2 \in X_i} \frac{k_{i_1 i_2}}{N_i^2} + \sum_{j_1, j_2 \in X_j} \frac{k_{j_1 j_2}}{N_j^2} - \sum_{i_1 \in X_i, j_1 \in X_j} \frac{k_{i_1 j_1}}{N_i \cdot N_j} - \sum_{i_1 \in X_i, j_1 \in X_j} \frac{k_{j_1 i_1}}{N_i \cdot N_j}}$$

where $k_{i,j} = k(x_i, x_j) = (\phi(x_i)^T \cdot \phi(x_j))$.

Hence, we propose a variant of K-NDA, where the weighted discriminant criterion in F is used to replace the conventional discriminant criterion in F , and it is expressed as:

$$\hat{J}(w) = \frac{w^T \hat{S}_b w}{w^T S_w w} \quad (8)$$

where S_w is the conventional within-class scatter matrix in F . As same as classical K-NDA, by the theory of reproducing kernel, (8) can be transformed as follows:

$$\hat{J}'(\alpha) = \frac{\alpha^T \hat{K}_b \alpha}{\alpha^T K_w \alpha} \quad (9)$$

where $K_w = (1/N) \sum_{i=1}^{c} \sum_{x_j \in X_i} (\xi_j - M_i)(\zeta_j - M_i)^T$,

$\hat{K}_b = \sum_{i=1}^{c} (N_i N_j / N^2) \cdot w(d_{i,j}) \cdot (M_i - M_j)(M_i - M_j)^T$,

with $M_i = ((1/N_i) \cdot \sum_{j=1}^{N_i} k(x_1, x_j), \ldots, (1/N_i) \cdot \sum_{j=1}^{N_i} k(x_N, x_j))^T$,

$M = ((1/N) \cdot \sum_{j=1}^{N} k(x_1, x_j), \ldots, (1/N) \cdot \sum_{j=1}^{N} k(x_N, x_j))^T$,

$\xi_j = (k(x_1, x_j), \ldots, k(x_N, x_j))^T$.

Therefore, the maximum criterion $\hat{J}'(\alpha)$ can be formed by l leading eigenvectors of $\hat{K}_w^{-1} K_b$. Suppose $\alpha_1, \ldots, \alpha_l$ constitute these l eigenvectors. Hence, the low-dimensional feature vectors of a pattern x can be given by:

$$H(x) = (\alpha_1, \ldots, \alpha_l)^T \cdot (k(x_1, x), \ldots, k(x_N, x))^T \quad (10)$$

3.2. Rotation and Reorientation of the K-NDA Subspace

Through the algorithm of the variant of K-NDA step discussed above, a low-dimensional subspace, where almost all classes of objects are linearly separable, has been spanned by $H(x)$. Hence, an F-LDA step will be directly applied to further reduce the dimensionality of this subspace from l to the required l' now.

In fact, the motivation for the F-LDA [2] comes from the following consideration: Suppose we wish to reduce the dimensionality from l to $(l-1)$ and φ_l is the corresponding eigenvector of the smallest eigenvalue of S_b, where S_b is the weighted between-class scatter matrix in the low-dimensional subspace. φ_l will be discarded when dimensionality is reduced from l to $(l-1)$. However, a problem can be encountered during the dimensionality reduction procedure. If a pair of classes is well separated in the l-dimensional input space, it is possible that the weight of the pair of classes is so small that they would heavily overlap in the $(l-1)$-dimensional space, which is orthogonal to φ_l (The weighting function defined in the weighted between-class scatter matrix is the monotonically decreasing function). The F-LDA, which makes the dimensionality reduced from l to $(l-1)$ at $r \geq 1$ fractional-steps instead of one step directly, can effectively solve this problem. F-LDA, introduces a sort of "automatic gain control", which will recompute S_b and its eigenvectors based on the changes of $w(d_{i,j})$ in the output space in each step, so that the $(l-1)$-dimensional subspace is reoriented and severe overlap between object classes in the output space is effectively avoided. Hence, the F-LDA can further accurately adjust and increase the distances of those object classes, which have smaller distances in the K-NDA-based subspace, in the output space, and simultaneously also still keep the well separability ability of the object classes, which have very big distances in the KNDA-based subspace, in the output space.

Hence, from the statements above, the KF-NDA has been constructed, and it has two novelties: 1) it can effectively improve the generalization ability of traditional K-NDA; 2) it can also overcome the limitation of the DF-LDA that it fails for a nonlinear problem, and extend the linear F-LDA method to a nonlinear version. By applying the KF-NDA to extract the robust nonlinear feature of the AGFV obtained in Section 2, a robust feature extraction framework, i.e., the Gobar KF-NDA (GKF-NDA), has also been constructed for FR, and a robust FR system architecture based on this framework is depicted in Fig.1.

4. EXPERIMENTAL RESULTS

To assess the effectiveness of the proposed framework, the experiments have been performed using the popular databases. One of the databases is the ORL database (www.uk.research.att. com). This database has 40 different persons and each one has 10 different images, including variations in pose, face expression (open or closed eyes, smiling/non-smiling) and facial details (glasses/no-glasses). All original images are scaled to 92x112 with 256-level gray scale. Fig.2 shows a part of the images in the ORL database.

Fig.2. The part of the face images in the ORL database.

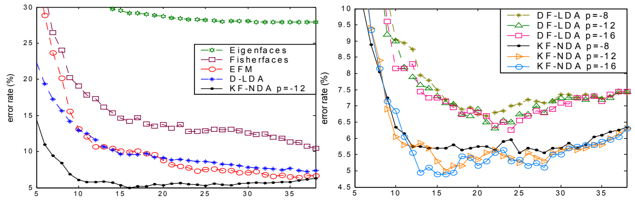
In the following experiments, we randomly select five training images and five test images per person from the ORL database, and a training set of 200 images and a test set of 200 images are created for the following experiments, and there is no overlapping between the two sets. The nearest neighbor classifier on the Euclidean distance is used for classification, and the number of fractional-step used in the DF-LDA and the KF-NDA is 30. In addition, we do each experiment on 10 times and the classification error rate results reported in this paper are an average of them, which is detailedly defined in [6,15].

4.1. Experiment 1

In this experiment, we will compare the classification error rate performance of the KF-NDA with those of the popular FR schemes, including the linear methods and the nonlinear methods. Firstly, the KF-NDA is compared with the linear methods, including: Eigenfaces [1], Fisherfaces [3], EFM [10], D-LDA [4], and DF-LDA [6]. Fig.3(a) describes the comparative results of Eigenfaces, Fisherfaces, EFM, D-LDA, and KF-NDA on the weighting function $w(d)=d^{-12}$; Fig.3(b) describes the comparative results of DF-LDA and KF-NDA on the different weighting functions $w(d)=\{d^{-8},d^{-12},d^{-16}\}$ recommended in [2]. From Fig.3, it can be seen that the error rates of the KF-NDA are lower than those of the others on the different number of the feature vectors. Secondly, the KF-NDA is compared with the nonlinear methods, including the K-PCA [14] and the K-NDA [12]. Fig.4(a) describes the comparative results of K-PCA, K-NDA and KF-NDA on the polynomial kernel $(k(y_1,y_2)=((y_1 \cdot y_2)\cdot 1e-9+1)^2)$; Fig.4(b) describes the comparative results of K-PCA, K-NDA and KF-NDA on the RBF kernel $(k(y_1,y_2)=\exp((-\|y_1-y_2\|^2)/1e9)$. From Fig.4, it can be seen that the error rates of the KF-NDA are lower than those of the other nonlinear methods on the different number of the feature vectors and it has the better generalization ability.

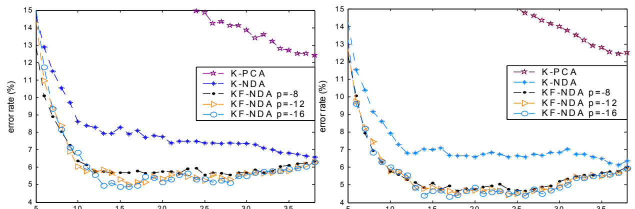
4.2. Experiment 2

To illustrate the effectiveness of the robust feature extraction framework, the following experiment will exploit those methods above on the high-dimensional AGFV obtained in Section 2, including: Gabor+Eigenfaces, Gabor+Fisherfaces, GFC [8], Gabor+D-LDA, Gabor+DF-LDA, Gabor+K-PCA, Gabor+K-NDA, GKF-NDA (In the GFC [8], the EFM [10] is applied on the AGFV; in the GKF-NDA, the KF-NDA is applied on the AGFV). For the sake of simplicity, we would only test those methods on the AGFV $\chi^{(\rho)}$ with the downsampling factor: $\rho=32$, and all Gabor wavelets used in the following experiment have the following parameters: $\sigma=2\pi$, $k_{\max}=\pi/2$, and $f=1.55$. Fig.5(a) and Fig.5(b) describe the corresponding comparative results of Fig.3(a) and Fig.3(b) on the AGFV, respectively. From Fig.5, we can see that the GKF-NDA has lower error rates than other linear methods on the AGFV. Fig.6(a) and Fig.6(b) describe the corresponding comparative results of Fig.4(a) and Fig.4(b) on the AGFV, respectively. From Fig.4, we can also see that the GKF-NDA has lower error rates than other nonlinear methods on the AGFV. In fact, the minimum error rate of the GKF-NDA in one of the ten test sets can reach 0% only using a weighting function of $w(d)=d^{-12}$ and a set of $N=16$ feature vectors, and it is a result comparable to the best results reported previously in many literatures. In addition, by comparing Fig.3 with Fig.5 and Fig.4 with Fig.6 respectively, we can also see that the Gabor representation of face images is robust to facial variations and simultaneously also exhibits more discriminatory information. In order to sufficiently illustrate the effectiveness of the GKF-NDA and the effectiveness of the Gabor wavelet representation of face images, we also calculate the average percentage of the error rates of the GKF-NDA over those of some other effective methods by $\sum_{i=5}^{38}\alpha_i/\beta_i$, where α_i and β_i are the error rates of the GKF-NDA and one of the other effective methods respectively, and i is the number of feature vectors. The results summarized in Table.1 show that the GKF-NDA not only overcomes the limitations and shortcomings of existing FR schemes, but also very effectively extracts the robust nonlinear feature for FR. According to all experimental results above, we can see that the GKF-NDA can obtain a lower error rates than those popular FR schemes, and a FR system based on the GKF-NDA is very effective. In addition, other detailed experiments have been also



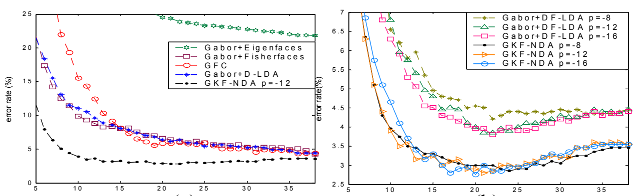
(a) (b)

Fig.3. Comparative FR performance of the KF-NDA and the various linear FR schemes using the original images, where the polynomial kernel $(k(y_1, y_2) = ((y_1 \cdot y_2)\cdot 1e-9+1)^2)$ is used for the KF-NDA.



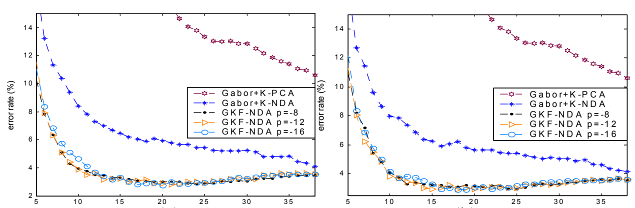
(a) (b)

Fig.4. Comparative FR performance of the KF-NDA and the various nonlinear FR schemes using the original images, where the different weighting functions $w(d) = \{d^{-8}, d^{-12}, d^{-16}\}$ is used for the KF-NDA.



(a) (b)

Fig.5. Comparative FR performance of the GKF-NDA and the various linear FR schemes using the AGFV.



(a) (b)

Fig.6. Comparative FR performance of the GKF-NDA and the various nonlinear FR schemes using the AGFV.

Tabel.1. Average percentage of error rates of GKF-NDA over those of some others effective methods, where the polynomial kernel $(k(y_1, y_2) = ((y_1 \cdot y_2)\cdot 1e-9+1)^2)$ is used for the nonlinear methods

The weighting function	$w(d)=d^{-8}$	$w(d)=d^{-12}$	$w(d)=d^{-16}$
EFM [10]	37.73%	38.39%	38.88%
D-LDA [4]	36.75%	37.37%	37.92%
DF-LDA [6]	44.67%	45.84%	46.46%
K-PCA [14]	20.66%	21.03%	21.34%
K-NDA [12]	44.80%	45.51%	46.37%
KF-NDA	57.22%	59.51%	61.16%
GFC [8]	49.12%	50.04%	50.32%
Gabor+D-LDA	50.94%	51.91%	52.51%
Gabor+DF-LDA	67.22%	71.07%	73.76%
Gabor+K-PCA	21.12%	21.53%	21.75%
Gabor+K-NDA	57.22%	58.19%	59.02%

performed on the different databases (including the FERET, Yale B and large mixed databases), all experimental results shows that the GKN-NDA is very effective (the very lower error rates can be reached only using a set of a few feature vectors). Moreover, the GKF-NDA is also far more effective than other recent nonlinear FR schemes, such as the KDDA [15], the DRDA [13]. However, we refer the reader to these results due to space limitations.

5. CONCLUSIONS

This paper introduces a robust feature extraction framework for FR. In this framework, the KF-NDA method, which has better generalization ability than classical K-NDA, is applied directly to extract the robust nonlinear feature of the AGFV that is robust to the variations of face images and exhibits the more discriminatory information. Experimental results tested on the popular databases reveal that this framework is very effective for FR. In addition, two next goals are to further improve the effectiveness of this framework: 1) one next goal is to further search for an optimal and sparse code resulting from the Gabor wavelet representation of face images, before forming the AGFV; 2) another next goal is to explore methods for incorporating prior knowledge about FR in the KF-NDA method by constructing appropriate kernel functions, and simultaneously optimize some nonlinear parameters that significantly influence the performance of the FR system. (This work is supported by the National Natural Science Foundation of P.R.China no.60103018)

6. REFERENCES

[1] M.A.Turk, A.P.Pentland, "Eigenfaces for recognition," *J.Cogn. Neurosci.*, vol.3, no.1, pp.71–86, 1991.

[2] R.Lotlikar, R.Kothari, "Fractional-Step dimensionality reduction," *IEEE Trans. Pattern Anal. Machine Intell.,* vol.22, no.6, pp.623-627, June, 2000.

[3] P.N.Belhumeur, J.P.Hespanha, D.J.Kriegman, "Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Machine Intell.*, vol.19, pp.711–720, May 1997.

[4] H.Yu, J.Yang, "A direct LDA algorithm for high-dimensional data with application to face recognition," *Pattern Recognit.*, vol.34, pp.2067–2070, 2001.

[5] L.-F.Chen, H.-Y.M.Liao, M.-T.Ko, J.-C.Lin, G.-J.Yu, "A new LDA-based face recognition system which can solve the small sample size problem," *Pattern recognit.*, vol.33, pp.1713-1726, 2000.

[6] J.W.Lu, K.N.Plataniotis, A.N.Venetsanopoulos, "Face recognition using LDA-based algorithms," *IEEE trans, neural networks*, vol.14, pp.195-200, 2003.

[7] B.S.Manjunath, W.Y.Ma, "Texture Feature for Browsing and Retrieval of Image Data," *IEEE trans. Pattern Anal. Machine Intell.*, vol.18, pp.837-842, August. 1996.

[8] C.J.Liu, H.Wechsler, "Gabor feature based classification using the enhance fisher linear discriminant model for face recognition," *IEEE Trans. Image Processing,* vol.11, pp.467-476. 2002.

[9] M.Loog, P.R.W.Duin, R.Haeb-Umbach, "Multiclass linear dimension reduction by weighted pairwise Fisher criteria," *IEEE Trans. Pattern Anal. Machine Intell.*, vol.23, no.7, pp.762-766, July, 2001.

[10] C.J.Liu, H.Wechsler, "Robust coding schemes for indexing and retrieval from large face recognition," *IEEE Trans. Image Processing*, vol.9, pp.132-137, Jan, 2000.

[11] S.Mika, G.Ratsch, J.Weston, B.Scholkopf, K.-R.Muller, "Fisher discriminant analysis with kernels," *in Proc. Neural Networks for Signal Processing IX*, Y.-H.Hu, J.Larsen, E.Wilson, and S.Douglas, Eds:IEEE, 1999, pp.41–48.

[12] Q.S.Liu, R.Huang, H.Q.Lu, S.D.Ma, "Face recognition using kernel based Fisher discriminant analysis," *in Proc. 5th IEEE Int. Conf. Automatic Face and Gesture Recognition*, 2002.

[13] J.W.Lu, K.N.Plataniotis, A.N.Venetsanopoulos,"Regularized discriminant analysis for the small sample size problem in face recognition," *Pattern Recognition Letter*, vol.24, pp.3079-3087, 2003.

[14] B.Scholkopf, A.Smola, K.R.Muller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Comput.,* vol.10, pp.1299-1319, 1999.

[15] J.W.Lu, K.N.Plataniotis, A.N.Venetsanopoulos, "Face recognition using kernel direct discriminant analysis algorithms," *IEEE Trans. Neural Networks,* vol.14, pp.117-126, 2003.