Journal of Zhejiang University SCIENCE
ISSN 1009-3095
http://www.zju.edu.cn/jzus
E-mail: jzus@zju.edu.cn

A novel face recognition method with feature combination*

LI Wen-shu (李文书)†[1,2], ZHOU Chang-le (周昌乐)[3], XU Jia-tuo (许家佗)[4]

([1]School of Computer Science and Technology, Zhejiang University, Hangzhou 310027, China)
([2]College of Information and Electronics, Zhejiang Sci-Tech University, Hangzhou 310018, China)
([3]Department of Artificial Intelligence, Xiamen University, Xiamen 361005, China)
([4]Shanghai University of Traditional Chinese Medicine, Shanghai 200032, China)
†E-mail: store6@163.com

Received Jan. 27, 2004; revision accepted Aug. 6, 2004

Abstract: A novel combined personalized feature framework is proposed for face recognition (FR). In the framework, the proposed linear discriminant analysis (LDA) makes use of the null space of the within-class scatter matrix effectively, and Global feature vectors (PCA-transformed) and local feature vectors (Gabor wavelet-transformed) are integrated by complex vectors as input feature of improved LDA. The proposed method is compared to other commonly used FR methods on two face databases (ORL and UMIST). Results demonstrated that the performance of the proposed method is superior to that of traditional FR approaches

Key words: Fisher discriminant criterion, Face recognition (FR), Linear discriminant analysis (LDA), Principal component analysis (PCA), Small sample size (SSS)
doi:10.1631/jzus.2005.A0454 **Document code:** A **CLC number:** TP391

INTRODUCTION

Automatic face recognition has become a very active research area in the last decade due to the new interest in, and need for, surveillance and security, telecommunication and digital libraries, human-computer intelligent interaction, and smart environments. The small sample size (SSS) problem is often encountered because the number of the samples is much smaller than the dimension of the sample space in face recognition. It results in the singularity of the within-class scatter matrix $\boldsymbol{S}_{\mathrm{w}}$ in linear discriminant analysis (LDA). Different methods proposed to solve this problem include PCA + LDA (Turk and Pentland, 1991; Belhumeur *et al.*, 1997; Swets and Weng, 1996), direct LDA (D-LDA) (Swets and Weng, 1996), DF-LDA (which combines the strengths of the D-LDA and F-LDA approaches) (Yu and Yang, 2001). It is generally believed that, when it comes to solving problems of pattern recognition, LDA-based algorithm outperforms PCA-based ones, since the former optimizes the low-dimensional representation of the objects with focus on the most discriminant feature extraction while the latter achieves simply object reconstruction (Belhumeur *et al.*, 1997; Swets and Weng, 1996; Lu *et al.*, 2003a; 2003b).

One of the most apparent disadvantages of these state-of-the-art algorithms is their poor adaptability when compared with the recognition capacity of the Human Visual System (HVS). For example, when confronting different objects, people will automatically select and combine the most salient features for the purpose of identification, which have been explored in some recent studies (Kalocsai *et al.*, 2000; Lu *et al.*, 2003b; Nastar and Mitschke, 1998). However, different subjects should have different salient features while adoption of a common feature combination for all subjects would lose useful information carried by these salient features.

In the real world, both global scan and detailed

* Project (No. 60275023) supported by the National Natural Science Foundation of China

facial feature observation are involved when a human attempts to recognize a face. So, the main contribution of this paper is to simulate such adaptability through global and local features combination, because global description and dominant feature have different contributions; holistic and local features are crucial for face recognition (Bruce *et al.*, 1998). Principal component analysis is adapted for global feature extraction. Local features are studied by using Gabor wavelet, which is adopted to obtain face minutia. The classical method of feature combination is to combine two sets of feature vectors into one union-vector (Peli *et al*., 1999; Dassigi *et al*., 2001; Li *et al*., 1995). Recently, Yang *et al*.(2003) proposed a feature combination strategy. Its idea is to combine two sets of feature vectors into a complex vector rather than a real union-vector. Thus, the increase of dimension is avoided as in the classical method. Yang's method for feature combinations method is used to propose a new method in Unitary space to make use of the null space of $\boldsymbol{S}_{\mathrm{w}}$ effectively and solve the small sample size problem of LDA.

FEATURE EXTRACTION METHODES

Global feature extraction

Principal component analysis (PCA) method used for global feature extraction is a powerful technique for extracting global structures from high-dimensional data set and has been widely used to reduce dimensionality and extract abstract features of faces for face recognition (Turk and Pentland, 1991; Zhao *et al*., 2000).

It is a standard decorrelation technique with its application yielding an orthogonal projection basis that directly leads to dimensionality reduction, and possibly to feature selection. Let $\boldsymbol{X}\in\Re^N$ be a random vector representing an image, where N is the dimensionality of the corresponding image space. Let $\boldsymbol{X}=[X_1, X_2, \ldots, X_T]$ be the sample set of the face images. T is the total number of the face images. Let Φ_i denote $\boldsymbol{X}_i-\bar{\boldsymbol{X}}$. Then the covariance matrix is defined as:

$$\boldsymbol{S}=\frac{1}{T}\sum_{i=1}^{T}(\boldsymbol{X}_i-\bar{\boldsymbol{X}})(\boldsymbol{X}_i-\bar{\boldsymbol{X}})^t=\frac{1}{T}\sum_{i=1}^{T}\Phi_i\Phi_i^t \quad (1)$$

where $\bar{\boldsymbol{X}}$ denotes the mean vector of all training samples. According to SVD theorem (Golub and van Loan, 1996), the orthonormal eigenvectors $w_1, w_2, \ldots, w_m$ of $\boldsymbol{S}$ corresponding to m largest eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_m$ can be obtained as

$$\boldsymbol{Y}=\boldsymbol{W}^t\boldsymbol{X} \quad (2)$$

where t is the transpose operator, $\boldsymbol{W}=[w_1, w_2, \ldots, w_m]$, $m<<N$, $\boldsymbol{W}\in\Re^{N\times m}$. The lower dimensional vector $\boldsymbol{Y}\in\Re^m$ captures the most expressive features of the original data $\boldsymbol{X}$.

Local feature analysis

As main facial features, eyes, nose and mouth often show the most distinguishable information of a given individual. However, it is very hard for computers to form a stable geometrical representation as we describe a face in our daily life. These sub-regions of face images are very small, so we adopt two-dimensional Wavelets analysis to create a representation of facial features in the framework.

We take Gabor wavelets as the basis function to create this representation, because Gabor wavelets have been used extensively in image processing, texture analysis due to their biological relevance and computational properties. Gabor wavelets can capture the properties of spatial localization, orientation selectivity, spatial frequency selectivity and quadrature phase relationship. Its representation has been shown to be optimal for minimizing the joint two-dimensional uncertainty in space and frequency (Daugman, 1988; He *et al*., 2002; Burr *et al*., 1989). The face's Gabor wavelets representation has robust characteristics in illumination and facial expression changes.

The two-dimensional Gabor Wavelets can be defined as follows (He *et al*., 2002):

$$\psi_{u,v}(z)=\frac{\|\boldsymbol{k}_{u,v}\|^2}{\sigma^2}\mathrm{e}^{-\frac{\|\boldsymbol{k}_{u,v}\|^2\|z\|^2}{2\sigma^2}}\left[\mathrm{e}^{i\boldsymbol{k}_{u,v}z}-\mathrm{e}^{-\sigma^2/2}\right] \quad (3)$$

$$\boldsymbol{k}_{u,v}=k_v\mathrm{e}^{i\phi_u} \quad (4)$$

where u and v denotes 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. Gabor wavelet function can form a complete but nonorthogonal basis set.

Given an arbitrary image $I(x, y)$, its Gabor wavelet transforms is then defined to be:

$$\boldsymbol{W}_{uv}(x,y)=\int I(x_1,y_1)\psi_{uv}^{*}(x-x_1,y-y_1)\mathrm{d}x_1\mathrm{d}y_1 \quad (5)$$

where * indicates the complex conjugate. The Gabor wavelet transformation of a face is at five different scales, $u\in\{0, \ldots, 4\}$ and eight different orientations $v\in\{0, \ldots, 7\}$. To encompass different spatial frequencies, spatial localities, and orientation selectivities, we concatenate all these representation results and derive an augmented feature vector $\boldsymbol{Y}$. Before the concatenation, we firstly downsample each $\boldsymbol{W}_{uv}(x, y)$ by a factor ρ to reduce the space dimension, and normalize it. We then construct a vector out of the $\boldsymbol{W}_{uv}(x, y)$ by concatenating its rows (or columns). Now, let $\boldsymbol{W}_{uv}^{(\rho)}(x,y)$ denote the normalized vector constructed from $\boldsymbol{W}_{uv}(x, y)$ (downsampled by ρ and normalized to zero mean and unit variance). Then Gabor face feature vector $\boldsymbol{Y}^{(\rho)}$ is defined as follows:

$$\boldsymbol{Y}^{(\rho)}=((W_{0,0}^{(\rho)})^t,(W_{0,1}^{(\rho)})^t,\ldots,(W_{4,7}^{(\rho)})^t)^t \quad (6)$$

where t is the transpose operator. The Gabor face feature vector has important discriminating information, which has robustness against varying illumination and rotation in the face image (Liu and Wechsler, 2002).

Improved line discriminant analysis

In unitary space $\boldsymbol{C}^n$, the inner product is defined by

$$<\boldsymbol{X},\boldsymbol{Y}>=(\bar{\boldsymbol{Y}})^t\boldsymbol{X}=\boldsymbol{Y}^H\boldsymbol{X} \quad (7)$$

where $\boldsymbol{X},\boldsymbol{Y}\in\boldsymbol{C}^n$, and H is the denotation of conjugate transpose.

Suppose $\boldsymbol{A}$ and $\boldsymbol{B}$ are two feature vector spaces defined on the sample space Ω. We define the combined feature space $\zeta=\{\alpha+\mathrm{i}\beta\,|\,\alpha\in\boldsymbol{A},\beta\in\boldsymbol{B}\}$ in unitary space $\boldsymbol{C}^n$, where $n=\max\{\dim(\boldsymbol{A}),\dim(\boldsymbol{B})\}$, i is the imaginary unit (Yang *et al.*, 2003). If the dimensions of α and β are not equal, the lower dimensional vector α (or β) is padded with trailing zeros to length n.

Suppose the dimension of the original sample space is n, and L class problem is considered. Then, in unitary space $\boldsymbol{C}^n$, the between-class scatter matrix $\boldsymbol{S}_\mathrm{b}$, within-class scatter matrix $\boldsymbol{S}_\mathrm{w}$ and total-class matrix $\boldsymbol{S}_\mathrm{t}$ are, respectively, defined as follows:

$$\boldsymbol{S}_\mathrm{b}=\sum_{i=1}^{L}p(\omega_i)(\boldsymbol{m}_i-\boldsymbol{m}_0)(\boldsymbol{m}_i-\boldsymbol{m}_0)^H \quad (8)$$

$$\boldsymbol{S}_\mathrm{w}=\sum_{i=1}^{L}p(\omega_i)E\{(\boldsymbol{X}-\boldsymbol{m}_i)(\boldsymbol{X}-\boldsymbol{m}_i)^H/\omega_i\} \quad (9)$$

$$\boldsymbol{S}_\mathrm{t}=\boldsymbol{S}_\mathrm{b}+\boldsymbol{S}_\mathrm{w}=E\{(\boldsymbol{X}-\boldsymbol{m}_i)(\boldsymbol{X}-\boldsymbol{m}_i)^H\} \quad (10)$$

where $p(\omega_i)$ is the prior probability of class i; $\boldsymbol{m}_i=E(\boldsymbol{X}/\omega_i)$ is the mean vector of class i; $\boldsymbol{m}_0=E\{\boldsymbol{X}\}=\sum_{i=1}^{m}p(\omega_i)\boldsymbol{m}_i$ is the mean of all training samples. From these equations above, it is easy to prove that $\boldsymbol{S}_\mathrm{b}$, $\boldsymbol{S}_\mathrm{w}$ and $\boldsymbol{S}_\mathrm{t}$ are nonnegative define Hermitian matrices. According to (Luo, 1992), we get the properties: each eigenvalue of the Hermitian matrix is a real number, i.e. eigenvalues of $\boldsymbol{S}_\mathrm{b}$, $\boldsymbol{S}_\mathrm{w}$ and $\boldsymbol{S}_\mathrm{t}$ are real number.

Lemma 1 In unitary space, let $\boldsymbol{Q}^H\boldsymbol{S}\boldsymbol{Q}=\Lambda$, where $\Lambda=\mathrm{diag}(a_1, a_2, \ldots, a_n)$ $(a_1>a_2>\ldots>a_n)$, $\boldsymbol{Q}=(\zeta_1, \zeta_2, \ldots, \zeta_n)$, $a_1, a_2, \ldots, a_n$ are eigenvalues of $\boldsymbol{S}$ and $\zeta_1, \zeta_2, \ldots, \zeta_n$ are associated eigenvectors. If $\boldsymbol{S}$ is Hermitian matrix and $\boldsymbol{I}$ is the identity matrix, $\boldsymbol{Q}$ is the real matrix and $\boldsymbol{Q}^H\boldsymbol{Q}=\boldsymbol{I}$.

In unitary space, the fisher discriminant criterion function can be defined too by

$$W_\mathrm{opt}=\arg\max_{W}\frac{\left|\boldsymbol{W}^H\boldsymbol{S}_\mathrm{b}\boldsymbol{W}\right|}{\left|\boldsymbol{W}^H\boldsymbol{S}_\mathrm{w}\boldsymbol{W}\right|} \quad (11)$$

where W_opt is the set of generalize eigenvectors of $\boldsymbol{S}_\mathrm{b}$ and $\boldsymbol{S}_\mathrm{w}$ corresponding to the m largest eigenvalues λ_i:

$$\boldsymbol{S}_\mathrm{b}W_i=\lambda_i\boldsymbol{S}_\mathrm{w}W_i \quad (i=1,2,\ldots,m) \quad (12)$$

It is easy to prove that the upper bounds of the rank of $\boldsymbol{S}_\mathrm{b}$, $\boldsymbol{S}_\mathrm{w}$ and $\boldsymbol{S}_\mathrm{t}$ are respectively at most $c-1$, $N-c$ and $N-1$, which are all much less than n in many practical problems.

Suppose $\boldsymbol{P}$ is a Hermitian matrix, then its null space (Kernel) is defined by

$$N(\boldsymbol{P})=\{\boldsymbol{x} \mid \boldsymbol{P}\boldsymbol{x}=0, \boldsymbol{x}\in \boldsymbol{C}^n\} \quad (13)$$

Its dimension (Nullity of $\boldsymbol{P}$) is n−rank($\boldsymbol{P}$).

As mentioned in (Belhumeur *et al.*, 1997; Chen *et al.*, 2000; Yu *et al.*, 2003), in real space, if $\boldsymbol{q}^H\boldsymbol{S}_w\boldsymbol{q}=0$ and $\boldsymbol{q}^H\boldsymbol{S}_b\boldsymbol{q}\neq 0$, the null space of $\boldsymbol{S}_w$ is very useful for discrimination. But if $\boldsymbol{q}^H\boldsymbol{S}_w\boldsymbol{q}=0$ and $\boldsymbol{q}^H\boldsymbol{S}_b\boldsymbol{q}=0$, $\boldsymbol{q}$ is not useful for discrimination. This means that not the whole null space of $\boldsymbol{S}_w$ is useful for discrimination. According to the properties of Hermitian matrix, obviously in unitary space, this idea is also true. And we can know that the Kernel of $\boldsymbol{S}_t$ is the common kernel of both $\boldsymbol{S}_b$ and $\boldsymbol{S}_w$ in unitary space. The proof is given in Appendix A.

According to the ideas mentioned above, we propose an improved LDA algorithm based on eigen-analysis and simultaneous diagonalization (Swets and Weng, 1996). The whole algorithm is described as follows:

1. Diagonalize $\boldsymbol{S}_t$: find matrix $\boldsymbol{V}$ such that $\boldsymbol{V}^T\boldsymbol{S}_t\boldsymbol{V}=\Lambda$, where Λ is a diagonal matrix sorted in decreasing order. This can be done using the traditional eigen-analysis, i.e. each column of $\boldsymbol{V}$ is an eigenvector of $\boldsymbol{S}_t$, and Λ contains all the eigenvalues. Let $\boldsymbol{Y}$ be the matrix whose columns are all the eigenvectors of $\boldsymbol{S}_t$ corresponding to the nonzero eigenvalues. According to Lemma 1, we know $\boldsymbol{Y}^H\boldsymbol{Y}=\boldsymbol{I}$. Then, we get $\boldsymbol{S}'_w=\boldsymbol{Y}^H\boldsymbol{S}_w\boldsymbol{Y}$ and $\boldsymbol{S}'_b=\boldsymbol{Y}^H\boldsymbol{S}_b\boldsymbol{Y}$.

2. Keep the null space of the within-class scatter matrix. Let $\boldsymbol{Q}$ be the null space of $\boldsymbol{S}'_w$, then we get: $\boldsymbol{S}''_w=\boldsymbol{Q}^H\boldsymbol{S}'_w\boldsymbol{Q}=\boldsymbol{Q}^H\boldsymbol{Y}^H\boldsymbol{S}_w\boldsymbol{Y}\boldsymbol{Q}=(\boldsymbol{Y}\boldsymbol{Q})^H\boldsymbol{S}_w(\boldsymbol{Y}\boldsymbol{Q})=0$ and $\boldsymbol{S}''_b=\boldsymbol{Q}^H\boldsymbol{S}'_b\boldsymbol{Q}=(\boldsymbol{Y}\boldsymbol{Q})^H\boldsymbol{S}_b(\boldsymbol{Y}\boldsymbol{Q})$. $\boldsymbol{Y}\boldsymbol{Q}$ is the subspace of the whole null space of $\boldsymbol{S}_w$, and includes the most discriminant information for classification.

3. Diagonalize $\boldsymbol{S}_b$: We remove the null space of $\boldsymbol{S}''_b$ if it exists, and further reduce the dimension if necessary. Let Ψ be the matrix whose columns are all the eigenvectors of $\boldsymbol{S}''_b$ corresponding to the nonzero eigenvalues. i.e. $\Psi'\boldsymbol{S}''_b\Psi=\boldsymbol{G}_b>0$. Then, the final LDA projection is: $\boldsymbol{W}=\boldsymbol{Y}\boldsymbol{Q}\Psi\boldsymbol{G}_b^{-1/2}$.

After the normalization of global feature vectors and local feature vectors, we suppose the real part of the Gabor wavelets is denoted by α, and β denotes the imaginary part of the Gabor wavelets plus the feature vector $\boldsymbol{Y}$ obtained by PCA and padded with trailing zeros according to the dimension of α. In unitary space $\boldsymbol{C}^n$, according to the combined feature vector $\zeta=\{\alpha+\mathrm{i}\beta \mid \alpha\in\boldsymbol{A}, \beta\in\boldsymbol{B}\}$, we can obtain its discriminant feature vector $\Pi=\boldsymbol{W}^H\zeta$. This method is called GLU-LDA. As we know, the difference of feature extraction techniques of measurements might lead to the unbalance between two sets of features of a pattern. But, after the normalization process, a weighting function is used to penalize α or β, and it is true that this mean is proved by the following experiments.

EXPERIMENTS AND RESULTS

Two popular face databases, the ORL and the UMIST, were used to demonstrate the effectiveness of the proposed GLU-LDA framework. In the two experiments, we used the downsampling factor $\rho=64$ in the augmented Gabor feature vector $\boldsymbol{Y}^{(\rho)}$ because the performance differences using three different factors ($\rho=4$, 16, 64) were not significant by the experiment (From Fig.1, the performance is marginally less effective when the factors is 256) and reduces to a larger extent the dimensionality of the vector space compared to the other two factors.

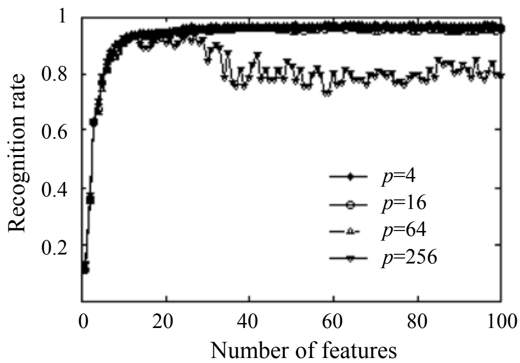

Fig.1 Recognition rates of GLU-LDA as functions of the number of feature vectors with different factors (ρ=4, 16, 64, 256)

The ORL database (ORL, 1994)

The dataset consisted of 400 frontal faces: 10 tightly cropped images of 40 subjects with variations in poses, illuminations, facial expressions and accessories. The all-original face images were sized 112×92 with 256-level gray scale. In the following experiments, we selected five training images and five testing images per person from this database, and a training set of 200 images and a testing set of remaining 200 images were created for the following experiments (the training images and testing images had no overlap between the two sets). Comparative performance was carried out against Gabor-PCA and some popular LDA schemes such as LDA (Belhumeur *et al.*, 1997), Yang's PFF (Yang *et al.*, 2003), and GFC (Liu and Wechsler, 2001; 2002). The Nearest Neighbor Classifier (NNC) rule was used for classification. In addition, since the recognition performance will be affected by the selection of training images, we did each experiment 10 times and the reported results given in this paper are their averages. Fig.1 shows that the performance of GLU-LDA is overall superior to that of the other four methods. In particular, GLU-LDA achieves 97.8% correct recognition accuracy when using only 40 features.

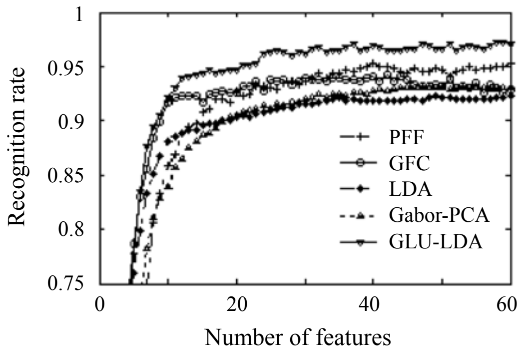

Fig2 Comparative face recognition performance of the five FR methods

The UMIST database

The UMIST (Graham *et al.*, 1998) repository is a multi-view database, consisting of 575 images of 20 people, each covering a wide range of poses from profile to frontal views. Each image in the database was of size 112×92. Six images per person were randomly chosen to produce a training set of 200 images. The remaining images were used to form the test set. Comparative performance was carried out against some popular LDA schemes such as D-LDA, Yang's GFF, GFC, and DF-LDA. The Nearest Neighbor Classifier (NNC) rule was used for classification. The average error rates of the six methods are shown in Fig.3. The performance of GLU-LDA is approximate to that of DF-LDA and overall superior to that of the other three methods.

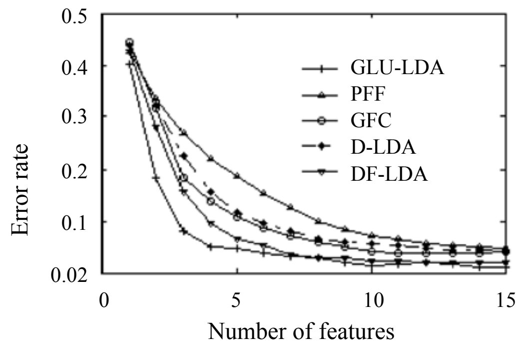

Fig.3 Comparison of error rates of the five FR methods as functions of the number of feature vectors

CONCLUSION

A novel feature extraction improved LDA method for face recognition has been proposed through combination of global and local features adaptive to each different individual. Such personalized feature integration is intended to reflect the adaptability of human vision to different subjects. The method introduced here is to combine global feature vectors (PCA-transformed) and local feature vectors (Gabor wavelet-transformed) via complex vectors as input feature of improved LDA which is to safely remove the null space of the between-class scatter matrix and to utilize the properties of Hermitian matrix. The effectiveness of the proposed method (GLU-LDA) has been demonstrated through experimentation using popular face database.

Generally speaking, the method presented here is a linear pattern recognition method. Compared with nonlinear models, a linear model is rather robust against noises and most likely will not overfit (e.g. Fig.3). But if the distribution of face patterns is highly non convex and complex such as Yale and CMU-PIE databases, nonlinear method must be researched. So, our next goal is to do this study by a nonlinear approach of GLU-LDA based on the kernel technique.

References

Belhumeur, P.N., Hespanha, J.P., Kriegman, D.J., 1997. Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. *IEEE Trans Pattern Anal Machine Intel*, **19**(7):711-720.

Bruce, V., Hancock, P., Burton, M., 1998. Human Face Perception and Identification, Face Recognition: From Theory to Application. NATO ASI Series, p.51-72.

Burr, D., Morrone, M., Spinell, D., 1989. Evidence for edge and bar detectors in human vision. *Vision Research*, **29**(4):419-431.

Chen, L.F., Liao, H.Y., Mark, K.M.T., Lin, J.C., Yu, G.J., 2000. A new LDA-based face recognition system which can solve the small sample size problem. *Pattern Recognition*, **33**(10):1713-1726.

Dassigi, V., Mann, R.C., Protopoescu, V.A., 2001. Information fusion for text classification–an experimental comparison. *Pattern Recognition*, **34** (12):2413-2425.

Daugman, J., 1988. Complete discrete 2-D Gabor transforms by Neural networks for image analysis and compression. *IEEE Trans on Acoustics, Speech, and Signal Processing*, **36**(7):1169-1179.

Golub, G.H., van Loan, C.F., 1996. Matrix Computations, Third Ed. The Johns Hopkins University Press, Baltimore, p.35-37.

Graham, D.B., Allinson, N.M., 1998. Characterizing virtual eigensignatures for general purpose face recognition. *Face Recognition: From Theory to Applications, NATO ASI Series F, Computer and Systems Sciences*, **163**:446-456.

He, C., Zheng, Y.F., Ahalt, S.C., 2002. Object tracking using the Gabor wavelet transform and the golden section algorithm. *IEEE Trans on Multimedia*, **4**(4):528-538.

Kalocsai, P., Malsburg, C., Horn, J., 2000. Face recognition by statistical analysis of feature detectors. *Image and Vision Computing*, **18**:273-281.

Li, H.Y., Deklerck, R., Cuyper, B.D., Hermanus, A., Nyssen, E., Cornelis, J., Hermanus, A., Nyssen, E., Cornelis, J., 1995. Object recognition in brain CT-scans: knowledge-based fusion of data from multiple feature extractors. *IEEE Trans on Medical Imaging*, **14**(2):212-228.

Liu, C.J., Wechsler, H., 2001. A shape- and texture-based enhanced fisher classifier for face recognition. *IEEE Trans. on Imaging Processing*, **10**(4):598-608.

Liu, C.J., Wechsler, H., 2002. Gabor feature based classification using the enhanced fisher liner discriminant model for face recognition. *IEEE Trans on Image Processing*, **11**(4):467-475.

Lu, J.W., Plataniotis, K.N., Venetsanopoulos, A.N., 2003a. Face recognition using LDA-based algorithms. *IEEE Trans on Neural Networks*, **14**(1):195-200.

Lu, X.G., Wang, Y.H., Jain, A., 2003b. Combining classifiers for face recognition. *Inter Conf Multimedia and Expo*, **3**:6-9.

Luo, J.H., 1992. Introduction to Matrix Analysis. South China University of Technology Press, Guangzhou, p.53-69 (in Chinese).

Nastar, C., Mitschke, M., 1998. Real-time Face Recognition Using Feature Combination. IEEE Inter Conf on Automatic Face and Gesture Recognition, p.312-319.

ORL face database, 1994. AT&T Laboratories, Cambridge, U.K. Available: http://www.uk.research.att.com:pub/data/att_faces.zip.

Peli, T., Young, M., Knox, R., Ellis, K.K., Bennett, F., 1999. Feature level sensor fusion, proceedings of the SPIE sensor fusion: Architectures. *Algorithms and Applications III*, **3719**:332-339.

Swets, D., Weng, J., 1996. Using discriminant eigenfeatures for image retrieval. *IEEE Trans on Pattern Analysis and Machine Intelligence*, **18**(8):831-836.

Turk, M., Pentland, A., 1991. Eigenfaces for recognition. *J Cognitive Neurosci*, **3**(1):71-86.

Yang, J., Yang, J.Y., Zhang, D., Lu, J.F., 2003. Feature fusion: parallel strategy vs. serial strategy. *Pattern Recognition*, **36**(6):1369-1381.

Yu, H., Yang, J., 2001. A direct LDA algorithm for high-dimensional data with application to face recognition. *Pattern Recognition*, **34**(10):2067-2070.

Yu, B., Jin, L.F., Chen, P., 2003. A new LDA-based method for face recognition. *Journal of Computer-Aided Design & Computer Graphics*, **15**(3):302-306.

Zhao, W.Y., Chellappa, R., Rosenfeld, A., Phillips, P.J., 2000. Face Recognition: A Literature Survey. UMD CFAR Technical Report CAR-TR-948.

APPENDIX A

In unitary space, $\forall \boldsymbol{x} \in \boldsymbol{N}(\boldsymbol{S}_t)$, according to Eq.(10) and Eq.(13), we have

$$
\begin{aligned}
&\langle \boldsymbol{S}_t \boldsymbol{x}, \boldsymbol{x} \rangle = \boldsymbol{x}^H \boldsymbol{S}_t \boldsymbol{x} = 0 \\
&\Leftrightarrow \langle (\boldsymbol{S}_b + \boldsymbol{S}_w)\boldsymbol{x}, \boldsymbol{x} \rangle \\
&\Leftrightarrow \boldsymbol{S}_b \langle \boldsymbol{x}, \boldsymbol{x} \rangle + \boldsymbol{S}_w \langle \boldsymbol{x}, \boldsymbol{x} \rangle \\
&\Leftrightarrow \langle \boldsymbol{S}_b \boldsymbol{x}, \boldsymbol{x} \rangle + \langle \boldsymbol{S}_w \boldsymbol{x}, \boldsymbol{x} \rangle \\
&\Leftrightarrow \boldsymbol{x}^H \boldsymbol{S}_b \boldsymbol{x} + \boldsymbol{x}^H \boldsymbol{S}_w \boldsymbol{x} \\
&\Leftrightarrow \boldsymbol{x}^H \boldsymbol{S}_b \boldsymbol{x} = 0 \cap \boldsymbol{x}^H \boldsymbol{S}_w \boldsymbol{x} = 0
\end{aligned}
$$

So the proposition holds.