A Novel Combined Fisherfaces Framework

*Wenshu Li and Changle Zhou**

College of Computer Science and Technology, Zhejiang University, Hangzhou, China
*Department of Artificial Intelligence, Xiamen University, Xiamen, China

ABSTRACT

A novel method for Face Recognition (FR) based on a combination of global features extracted by nonlinear kernel principal component analysis and local features derived by applying Gabor wavelets is discussed. It is well known that the distribution of face images is highly nonlinear under a large variation in viewpoints. Therefore, linear methods such as principle component analysis (PCA) or linear discriminant analysis (LDA) cannot provide reliable and robust solutions for FR problems. In our framework, the proposed LDA in the unitary space makes use of the null space of the within-class scatter matrix effectively, and Global feature vectors and local feature vectors are integrated by complex vectors as input feature of the proposed LDA. The experiment results demonstrate that the proposed methodology is more effective and robust for face recognition with complex face variations.

1. INTRODUCTION

Data reduction and feature extraction in FR have been developed over the last decades. Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) are two classic tools widely used, such as PCA + LDA [1], direct LDA (D-LDA) [2, 3]. The methods do not perform well in many real-world situations, where the query test face appearance is significantly difference from the training face data due to variation in pose, lighting and expression. The limited success of these methods should be attributed to their nature [4]. As a result, it is reasonable to assume that a better solution to this inherent nonlinear problem could be achieved using nonlinear methods, such as the so-called kernel machine techniques [5].

Gabor wavelets, especially for frequency and orientation representation, have similar characteristics to those of the human visual system. Gabor wavelets can be applied locally to extract local image feature. The effect of filtering an image is to break down image content to different scales, locations, and orientation that can be extracted effectively for recognition [3, 6].

In the real world, both global scan and detailed facial feature observation happen when a human attempts to recognize a face. So, the main contributions of this paper are on two aspects: (1) 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 [7]. Nonlinear Kernel Principal Component Analysis is adapted for global feature extraction. We perform local features by Gabor wavelet that is adopted to obtain face minutia. The classical method of feature combination is to two sets of feature vectors into one union vector [8, 9]. Recently, Yang [9] proposed a feature combination strategy. Its idea is to combine two sets of feature vectors by a complex vector rather than a real union-vector. Thus, the increase of dimension is avoided corresponding of the classical method. So we make use of Yang's method for feature combinations. (2) to propose a new method in Unitary space to make use of the null space of within-class scatter matrix effectively and solve the small sample size problem of LDA. The proposed LDA method is tested and discussed on the multi-view UMIST face database [10].

2. FEATURE EXTRACTIONS

2.1. Global feature extraction

PCA is a powerful technique for extracting global structure from high-dimensional data set. Nonlinear variants of PCA have also investigated. Among these, kernel PCA can be considered as a natural generalization of PCA. Therefore, we adopt Kernel PCA method for global feature extraction. The basic concept of kernel PCA is to first map the input data space into a high dimensional feature space $\mathbb{F}$ via a non-linear mapping $\phi(\bullet)$ and perform a linear PCA in $\mathbb{F}$. The objective is that a training set, which may not be linearly separable in an input space, may be linearly separable in the mapped space.

Let $X=[x_1,x_2,...,x_M]$ be the training set in $\Re^N$, where each x_i represents a training vector, and $\Phi=[\phi(x_1),\phi(x_2),...,\phi(x_M)]$ be used to map the input data space $\Re^N$ into the feature space $\mathbb{F}$. Let $\bar{\phi}=\frac{1}{M}\sum_{i=1}^{M}\phi(x_i)$, we construct the covariance matrix in feature space $\mathbb{F}$:

$$S=\frac{1}{M}\sum_{i=1}^{M}(\phi(x_i)-\bar{\phi})(\phi(x_i)-\bar{\phi})' \quad (1)$$

Assuming that the mapped data are centered, i.e. $\sum_{i=1}^{M}\phi(x_i)=0$, then $\tilde{S}=\frac{1}{M}\Phi\Phi'$. Let us form the matrix $\tilde{R}=\Phi'\Phi$. By virtue of kernel tricks, we can determine of the $M\times M$ matrix $\tilde{R}$ by $\tilde{R}_{ij}=\phi(x_i)'\phi(x_j)=k(x_i,y_j)$. In general, the assumption of

centered data in feature space made above is not reasonable. A method to center the mapped data is described here: let $\tilde{\phi}(x_j) = \phi(x_j) - 1/M * \sum_i \phi(x_i)$, $1 \le j \le M$ be the centered mapped data in the feature space. Then we centralize $\tilde{R}$ by $R = \tilde{\Phi}^{'}\tilde{\Phi} = (I - \frac{1_{M\times M}}{M})^{'}\tilde{R}(I - \frac{1_{M\times M}}{M})$ and R's eigenvectors $u_1, u_2, ..., u_m$ corresponding to m largest eigenvalues $\lambda_1, \lambda_2, ..., \lambda_m$, where I is a $M \times M$ identity matrix and $1_{M\times M}$ is a $M \times M$ matrix of all ones.

According to SVD theorem [11], [12], we get the orthonormal eigenvectors $w_1, w_2, ..., w_m$ of S. Thus, we obtain the j-th feature

$$y_j = w_j^{'}\phi(x) = \frac{1}{\sqrt{\lambda_j}} u_j^{'}[k(x_1,x), k(x_2,x), ..., k(x_M,x)], \; j = 1, ..., m \qquad (2)$$

Thus, the lower dimensional vector $Y = (y_1, y_2, ..., y_m)$ captures the most expressive features of the original data X.

2.2. Local feature extraction

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. We adopt two-dimensional Gabor Wavelets analysis to create a representation of facial features in the framework.

Gabor wavelets have been used extensively in image processing, texture analysis because of their biological relevance and computational properties. Gabor wavelets can capture the properties of spatial localization, orientation selectivity, spatial frequency selectively and quadrature phase relationship. 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 [1]:

$$\psi_{u,v}(z) = \frac{\| k_{u,v} \|^2}{\sigma^2} e^{-\frac{\|k_{u,v}\|^2 \|z\|^2}{2\sigma^2}} \left[e^{ik_{u,v}z} - e^{-\frac{\sigma^2}{2}} \right] \qquad (3)$$

$$k_{u,v} = k_v e^{i\phi_u} \qquad (4)$$

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:

$$W_{uv}(x, y) = \int I(x_1, y_1) \psi_{uv}^{*}(x - x_1, y - y_1) dx_1 dy_1 \qquad (5)$$

where $*$ indicates the complex conjugate. The Gabor wavelet transformation of a face is at the five different scale, $u \in \{0, \ldots, 4\}$ 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 Y. Before the concatenation, we firstly downsample each $W_{uv}(x, y)$ by a factor ρ to reduce the space dimension, and normalize it. We then construct a vector out of the $W_{uv}(x, y)$ by concatenating its rows (or columns). Now, let $W_{uv}^{(\rho)}(x, y)$ denote the normalized vector constructed from $W_{uv}(x, y)$ (downsampled by ρ and normalized to zero mean and unit variance). Then Gabor face feature vector $Y^{(\rho)}$ is defined as follows:

$$Y^{(\rho)} = ((W_{0,0}^{(\rho)})^{'}, (W_{0,1}^{(\rho)})^{'}, \ldots, (W_{4,7}^{(\rho)})^{'})^{'} \qquad (6)$$

where $'$ is the transpose operator.

2.3. Join local and global feature extraction

In unitary space C^n, the inner product is defined by

$$< X, Y >= (\bar{Y})^{'} X = Y^H X \qquad (7)$$

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

Suppose A and B are two feature vector spaces defined on the sample space Ω. We define the combined feature space $\zeta = \{\alpha + j\beta \mid \alpha \in A, \beta \in B\}$ in unitary space C^n, where $n = \max\{\dim(A), \dim(B)\}$, j is the imaginary unit [9]. 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 L class problem is considered. Then, in unitary space C^n, the between-class scatter matrix S_b, within-class scatter matrix S_w and total-class matrix are, respectively, defined as follows:

$$S_b = \sum_{i=1}^{L} p(\omega_i)(m_i - m_0)(m_i - m_0)^H \qquad (8)$$

$$S_w = \sum_{i=1}^{L} p(\omega_i) E\{(X - m_i)(X - m_i)^H / \omega_i\} \qquad (9)$$

$$S_t = S_b + S_w = E\{(X - m_i)(X - m_i)^H\} \qquad (10)$$

where $p(\omega_i)$ is the prior probability of class i; $m_i = E\{X / \omega_i\}$ is the mean vector of class i; $m_o = E\{X\} = \sum_{i=1}^{m} p(\omega_i) m_i$ is the mean of all training samples. From these equations above, it is easy to prove that S_b, S_w and S_t are nonnegative define Hermitian matrices. According to [11], we get the properties: each eigenvalue of Hermitian matrix is a real number, i.e. eigenvalues of S_b, S_w and S_t are real number.

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

Suppose P is a Hermitian matrix, then its Null space (Kernel) is defined by

$$N(P) = \{x \mid Px = 0, x \in C^n\} \tag{11}$$

Its dimension (Nullity of P) is $n - rank(P)$.

As mentioned in [2], in real space, if $q^H S_w q = 0$ and $q^H S_b q \neq 0$, the null space of S_w is very useful for discrimination. But if $q^H S_w q = 0$ and $q^H S_b q = 0$, q is not useful for discrimination. This means that not the whole null space of S_w is useful for discrimination. According to the properties of Hermitian matrix, obviously in unitary space, this idea is too true. And we can know that the Kernel of S_t is the common kernel of both $q^H S_w q = 0$ and S_w in unitary space.

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

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

(2) Keep the null space of the within-class scatter matrix. Let Q be the null space of $S_w^{'}$, then we get: $S_w^{*} = Q^H S_w^{'} Q = Q^H Y^H S_w Y Q = (YQ)^H S_w (YQ) = 0$ and $S_b^{*} = Q^H S_b^{'} Q = (YQ)^H S_b (YQ)$. YQ is the subspace of the whole null space of S_w .

(3) Diagonalize S_b : We remove the null space of S_b^{*} if it exists, and further reduce dimension if necessary. Let Ψ be the matrix whose columns are all the eigenvectors of S_b^{*} corresponding to the nonzero eigenvalues. i.e. $\Psi^T S_b^{*} \Psi = G_b > 0$. Then, the final LDA projection is: $W = YQ\Psi G_b^{-\frac{1}{2}}$.

We suppose the complex modulus of the Gabor Wavelets is denoted by α , and β denotes the feature vector Y obtained by Kernel PCA, which has been padded with trailing zeros according to the dimension of α . In unitary space C^n , according to $\zeta = \{\alpha + j\beta \mid \alpha \in A, \beta \in B\}$, we can obtain its discriminant feature vector $\Pi = W^H \zeta$. This method is called by KGLU-LDA.

When the dimensions of α and β are unequal, we know the higher-dimensional one is still more powerful than the lower-dimensional one. So, in order to eliminate the unfavorable effect resulting from an unequal dimension, our suggestion is to adopt the weighted combination form. The combination is formed by $\zeta = \{\alpha + j\theta\beta \mid \alpha \in A, \beta \in B\}$, where the weight θ is called a combination coefficient.

We discuss the estimation of the combination coefficient. The primary influential factors to select the combinational coefficient are the length and dimension of feature vectors. So, we define an experimental formula to estimate the combination coefficient:

$$\theta = \frac{n}{m} \times \frac{\|\alpha\|}{\|\beta\|} \times C_{Gabor} \tag{12}$$

where n *and* m are the dimensions of α and β , respectively. $||\bullet||_2$ denotes the 2-norm of $\bullet$. C_{Gabor} represents the multiply of the number of scales and the number of orientations.

3. EXPERIMENTS

The UMIST Face Database is used to demonstrate the effectiveness of the proposed KGLU-LDA framework. It 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 is of size 112×92 . Six images per person are randomly chosen to produce a training set of 200 images. The remaining images are used to form the test set. Comparative performance is carried out against KPCA, Yang's GFF, and GFC [3]. The Nearest Neighbor Classifier rule is used for classification. In the kernel PCA, we use the Gaussian RBF $k(x,y) = \exp(-\|x-y\|^2 / \sigma^2)$.

Since the dimensions of two feature vectors are unequal, it needs to evaluate the combination coefficient before combination. According to (12), we obtain $\theta = \frac{12880}{120} \times \frac{3.8 \times 10^3}{1.1 \times 10^4} \times 40 \cong 1500$. From Table 1, we can see that estimates is just contained in the interval [600, 2600] and in this interval, classification error rates are robust with varying the value of θ . So we say that the proposed method of estimated combination coefficient is effective. We use the downsampling factor $\rho = 64$ in the augmented Gabor feature vector $Y^{(\rho)}$ because the performance differences among using three different factors ($\rho = 4, 16, 64$) are not significant by the experiment (From Fig.1, the performance is marginally less effective when the factors is 256) and it reduces to a larger extent the dimensionality of the vector space than the other two factors do.

Table 1. Classification results with varying the combination coefficient θ , when $\sigma^2 = 1e8$

θ	error rate	θ	error rate	θ	error rate
50	0.1433	1000	0.0107	2200	0.0117
100	0.0300	1200	0.0100	2400	0.0100
500	0.0150	1400	0.0100	2600	0.0100
600	0.0100	1600	0.0107	3000	0.0127
800	0.0100	1800	0.0100	30000	0.0287

To obtain the optimal σ^2 in the Gaussian RBF $k(x,y) = \exp(-\|x-y\|^2 / \sigma^2)$, a wide range of values of σ^2 is tested. From Table 2, we can easily see that the average error rate of KGLU-LDA is more stable and predictable in the interval [1e8, 5e9], when $\theta = 1700$.

Table 2. The average error rates of the KGLU-LDA with varying the parameter σ^2 in the kernel function $k(x,y)=\exp(-\|x-y\|^2/\sigma^2)$

$\sigma^2\times10^{-8}$	Error rate	$\sigma^2\times10^{-8}$	Error rate
0.1	0.1233	7.0	0.0100
0.5	0.0217	10.0	0.0100
0.8	0.0133	40.0	0.0100
0.9	0.0117	50.0	0.0100
1.0	0.0100	54.0	0.0113
2.0	0.0100	55.0	0.0120
3.0	0.0100	6.0	0.0133
4.0	0.0100	80.0	0.0167
5.0	0.0100	1000.0	0.0333

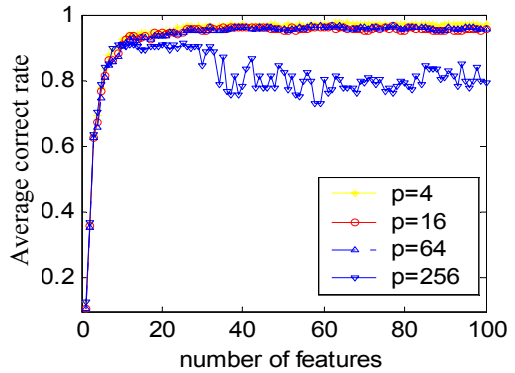

Fig. 1. Average recognition rates of KGLU-LDA as functions of the number of feature vectors with different factors ($\rho=4,16,64,256$)

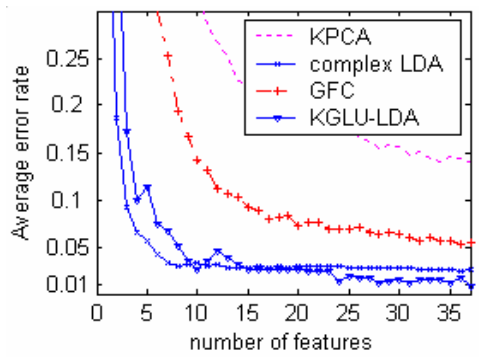

Fig. 2. Comparison of error rates obtained by the four FR methods as functions of the number of feature vectors

The average error rates of the four methods are shown in Fig.2, when $\sigma^2=1e9$, and $\theta=1500$. We can see that when the number of feature is less than 10, the performance of yang's complex LDA is good. However, when the number of feature is more than 10, the performance of KGLU-LDA is overall superior to that of the other three methods. In particular, KGLU-LDA achieves 0.01% average error recognition rate when using only 37 features.

By the above results, we can easily see that the proposed method is more effective and predictable.

4. CONCLUSIONS

A personalized combination of global features extracted by nonlinear kernel principal component analysis and local features derived by applying Gabor wavelets is discussed. 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 and local feature vectors 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 has been demonstrated through experimentation using UMIST face database. KGLU-LDA is more effective and predictable.

5. REFERENCES

1. P.N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Machine Intel.*, vol.19, no.7, pp. 711–720, 1997.
2. H. Yu and J. Yang, "A direct LDA algorithm for high-dimensional data with application to face recognition," *Pattern Recognition*, vol. 34,pp. 2067–2070, 2001.
3. Chenjun Liu and Harry Wechsler, "Gabor feature based classification using the enhanced fisher liner discriminant model for face recognition," *IEEE Trans on Image processing*, vol. 11, no. 4, pp. 467-475,2002.
4. Juwei Lu, Kostantinos N. Plataniotis, and Anastasios N. Venetsanopoulos, "Face recognition using LDA-based algorithms," *IEEE Trans. On Neural Networks*, vol.14, no. 1, pp. 195-200, 2003
5. K.-R. Müller, S. Mika, G. Rätsch, K.suda, and B. Schölkopf, "An introduction to kernel-based learning algorithms," *IEEE Trans. Neural Networks*, vol. 12, pp. 181–201,2001.
6. P. Kalocsai, C. Malsburg, J. Horn, "Face recognition by statistical analysis of feature detectors," *Image and Vision Computing*, vol.18, pp. 273-281, 2000.
7. Bruce, V., Hancock, P. and Burton, M. "Human face perception and identification, face recognition: from theory to application," *NATO ASI Series*, Springer, pp. 51-72, 1998.
8. V. Dassigi, R.C. Mann, V.A. Protopoescu, "Information fusion for text classi0cation—an experimental comparison," *Pattern Recognition*, vol. 34, no. 12, pp. 2413–2425,2001.
9. Jian Yang, Jingyu Yang, David Zhang and Jianfeng Lu, "Feature fusion: parallel strategy vs. serial strategy," *Pattern Recognition*, vol.36, no. 6, pp. 1369-1381, 2003.
10. D. B. Graham and N. M. Allinson, "Characterizing virtual eigen signatures for general purpose face recognition," in Face Recognition: From Theory to Applications, H. Wechsler, P. J. Phillips, V. Bruce, F. Fogelman-Soulie and T. S. Huang, Eds., NATO ASI Series F, Computer and Systems Sciences, Vol. 163, pp 446-456,1998.
11. G.H. Golub, C.F. Van Loan, Matrix Computations, Third ed., the Johns Hopkins University Press, Baltimore, 1996.
12. B. Schölkopf, A. J. Smola, and K. R. Müller, "Nonlinear component analysis as a kernel eigenvalue Problem," *Neural Computation*, vol. 10, pp. 1299–1319,1998.