International Journal of Scientific Engineering and Technology (ISSN : 2277-1581)
Volume No.1, Issue No.5, pg : 248-251 01 Nov. 2012

An Efficient Method for face Recognition Using Kernel Discriminant Analysis

Ms. Dipali M. Shimpi[1]

Computer Department, MIT, Kothrud, Pune
E-mail: dipali.shimpi@mitpune.edu.in

Abstract: The kernel approach has been proposed to solve face recognition problems under complex distribution by mapping the input space to a high-dimensional feature space. The experimental results show that the kernel-based method is a good and feasible approach to tackle the pose and illumination variations. One of the crucial factors in the kernel approach is the selection of kernel parameters, which highly affects the generalization capability and stability of the kernel-based learning methods. This paper addresses the problem of selection of Kernel parameters in Kernel Fisher Discriminant for face recognition. We propose a new criterion and derive a new formation in optimizing the parameters in RBF kernel. The proposed formulation is further integrated into a subspace LDA algorithm and a new face recognition algorithm is developed. FERET database is used for evaluation. Comparing with the existing Kernel LDA based methods with kernel parameter selected by experiment manually, the results are encouraging.

Keywords:
PCA, LDA, orthonormal matrices, feature space

1.0 INTRODUCTION

Face recognition is a highly complex and nonlinear problem because there exists so many image variations such as pose, illumination and facial expression [10]. These variations would give a nonlinear distribution of face images of an individual. In turn, linear methods such as linear discriminant analysis (LDA) [3, 4] could not provide sufficient nonlinear discriminant power to handle the problem. To overcome this limitation, Kernel approach has been proposed [5].It is also shown that kernel-based approach is a feasible approach to solve the nonlinear problems in face recognition. However, the selection of kernel functions and selection of kernel parameters is the big deal. This paper mainly focuses on the selection of kernel parameters. We select the "universal kernel" in this paper, in which there are a number of scale factors as follows

$$K(x, y) = \exp\left(-\sum \frac{(x_i - y_i)^2}{2\theta_i^2} \right) \qquad [1]$$

In fact, these factors are precisely parameters of the kernel function [1]. Existing kernel-based LDA algorithms, in general, consider $\theta = \theta_1 = \ldots = \theta_n$ and try to pick the best value for θ by experimental results. The main purpose of this paper is to find an adaptive method to optimize the kernel parameters θ_i individually for kernel-based subspace LDA algorithm.

2.0 KERNEL SUBSPACE LDA ALGORITHM

The KSLDA algorithm [11] is generalized from a novel subspace LDA algorithm [4]. KSLDA not only can solve the linear limitation on LDA-based algorithm, the subspace property but also solve the sample size problem. The basic idea of kernel-based algorithm is to mapping the nonlinear distributions higher dimensional space *F* in which classes are linear separable.

The basic idea of our proposed KSLDA method is to apply the subspace LDA method in the kernel feature space, and it consists of two steps, namely, nonlinear mapping from the input space to the feature space and applying the subspace LDA method in the feature space.

2.1 NONLINEAR MAPPING FROM THE INPUT SPACE TO THE FEATURE SPACE:

Let φ : **R**d → **F**df, x → φ(x), be a nonlinear mapping from the input space **R** to a high-dimensional feature space **F**. Considering a C-class problem in the feature space, the set of all classes is represented by C = {C1, C2, . . . , CC}, and the jth class Cj contains Nj samples, j = 1, 2 , C. Let N be the total number of training samples, i.e., N = _C j=1 Nj. Let xij be the jth sample in class i, and mi and m be the mean of the ith class and all samples, respectively. The within-class scatter matrix Sw, between-class scatter matrix Sb, and total class scatter matrix St can be formulated in the feature space

$$S_w = \frac{1}{N}\sum_{i-1}^{C}\sum_{j=1}^{N_i}\left(\varphi\left(x_j^i\right)-m_i\right)\left(\varphi\left(x_j^i\right)-m_i\right)^T$$

$$= \sum_{i=1}^{C}\sum_{j=1}^{N_i}\left(\frac{1}{\sqrt{N}}\left(\varphi\left(x_j^i\right)-m_i\right)\right)\left(\frac{1}{\sqrt{N}}\left(\varphi\left(x_j^i\right)-m_i\right)\right)^T$$

$$= \varphi_w \ \varphi_w^T$$

Where Φ_w is a df x N matrix and is defined as

$$\varphi_w = \left[\bar{\varphi}_w^{\,i\,j}\right]_{i=1......C.\ j=1......N_i}$$

$$S_b = \frac{1}{N}\sum_{i=1}^{C} N_i (m_i - m)(m_i - m)^T$$

$$= \sum_{i=1}^{C}\left(\sqrt{\frac{N_i}{N}}(m_i - m)\right)\left(\sqrt{\frac{N_i}{N}}(m_i - m)\right)^T$$

$$= \varphi_b \varphi_b^T$$

where ϕ_b is a df x C matrix and is defined as

$$\varphi_b = \left[\bar{\varphi}_b^{\,i}\right]_{i=1....}C$$

$$S_t = \frac{1}{N}\sum_{i=1}^{C}\sum_{j=1}^{N_i}\left(\varphi(x_j^i) - m\right)\left(\varphi(x_j^i) - m\right)^2$$

$$= \sum_{i=1}^{C}\sum_{j=1}^{N_1}\left(\frac{1}{\sqrt{N}}\left(\varphi(x_j^i) - m\right)\right)\left(\frac{1}{\sqrt{N}}\left(\varphi(x_j^i) - m\right)\right)^T$$

$$= \varphi_t \varphi_t^T$$

where Φt is a df x N matrix and is defined as

$$\varphi_t = \left[\bar{\varphi}_t^{\,i.j}\right]_{i=1....,C,\,j=1,.......Ni}$$

For classes C_i and C_l an N_i x N_l dot-product matrix K_i is defined as

$$K_l^i = \left(K_{lk}^{ij}\right)_{\substack{j=1.......N \\ k=1.......N_i}} \text{ where}$$

$$K_{lk}^{ij} = k\left(x_j^i, x_k^l\right) = \varphi\left(x_j^i\right) - \varphi\left(x_k^l\right).$$

Then, for all C classes, we can define an N x N kernel matrix K as

$$K = \left(K_{lk}^{ij}\right)_{\substack{i=t.........C.j=1.........N_i \\ i=l,.........C.k=1........N_k}}$$

2.2 APPLYING THE SUBSPACE LDA ALGORITHM ON THE FEATURE SPACE:

Assume $S = \Phi\,\Phi^T$, $\Phi \in R^{d\times n}$ ($d \geq n$) and rank(Φ) = r. Then, the singular value decomposition $\Phi = U\begin{pmatrix}\delta \\ v\end{pmatrix}_{d\,x\,n} V^T$ Where $U = (u_1,\ldots,u_r,u_{r+1}\ldots,u_d) \in R^{d\times d}$, and $V \in R^{n\times n}$ are orthonormal matrices, Σ = diag $(\sigma_1,\ldots\sigma_r,0,\ldots.0,) \in R^{n\times n}$ $\sigma_1 \geq \sigma_2 \geq \ldots.. \geq \sigma_r > 0$, $U_1 = (u_1,\ldots,u_r,)$ and $U^T_1\ SU_1$= diag($\sigma^2_1, \sigma^2_2, \ldots\ldots, \sigma^2_r$)

***a)* Determining the sub-null-space of S_w:** The objective of this step is to determine the sub-null-space of S_w. Let $S_{wt} = \Phi^T_w\,\Phi_t \in R^{N\,xN}$. Utilizing the kernel function, we express matrix S_{wt} in terms of the kernel matrix K as follows:

$$S_{wt} = \frac{1}{N}\ (K - K.I_{NN} - \Lambda_{NN}.\ K + \Lambda_{NN}\ .K.I_{NN})$$

where I_{NN} is an N x N matrix with all terms equal to $1/N$, Λ_{NN} = diag $[\Lambda_{N1}, \Lambda_{N2}, \ldots.\Lambda_{Nc}]$ is an N x N block diagonal matrix, and Λ_{Ni} is an N_i x N_i matrix with all terms equal to 1/N. By singular value decomposition, there exist orthonormal matrices U, V $\in R^{NxN}$ and a diagonal matrix Λ = diag $(\sigma_1,\ldots\sigma_r,0,\ldots.0) \in R^{N\,x\,N}$ $(\sigma_1 \geq \sigma_2, \geq, \ldots. \geq \sigma_r > 0)$ such that $S_{wt} = U\Lambda V^T$.

Since $S_w = \Phi_w\Phi^T_w$, by Theorem I, we get $(\Phi tV)^T S_w\ (\Phi_t V)$ = dia $(\sigma^2_1\ \ldots..,\sigma^2_r, 0,\ldots.\ 0)_{N\,x\,N}$

Let $V^1 = [\upsilon_r+1, \upsilon_r+2,\ldots.,\upsilon_N]$.

Then, the sub-null-space of S_W. Y, is given by

$Y = \Phi tV^1$. It can be seen that Y satisfies

$Y^T S_w Y = 0_{(N-r)\,x\,(N-r)}$.

b) Discarding the null-space of S_b *:* After determining the sub-null-space of S_w the projection is then determined outside the null-space of S_b,. Thus, the second step is to discard the null- space of S_b to ensure that the numerator of the Fisher Index will not be zero.

Define $S_b = Y^T S_b Y$, and then

$S_b - (Y^T\Phi_b)\ (Y^T\Phi_b)^T$.

Let $Z = Y^T\Phi_b \quad = (V^1)^T \Phi^T_t \Phi_b$,

Which is a matrix of (N - r) x C dimensions. Utilizing the kernel function, matrix Z can be expressed in terms of the kernel matrix K as follows:

$Z = (V^t)^T$ - $(K-\Lambda_{NC} - K - I_{NC}$- INN $- K\ \Delta_{NC} + I_{NN} - K\ I_{NC})$

Where ΛNC = diag $[\Lambda_{N1}, \Lambda_{N2}\ \ldots..\Lambda_N]$ is am N x C block diagonal matrix, and Λ_N is an N_j x 1 matrix with all terms equal to is an N x C matrix with each element of the column equal to and I_{NN} is a matrix with all terms equal to $1/N^2\ \sqrt{N_i}$. NC is an N x C matrix with each element of the ith column equal to $(N_i/N^3)^{1/2}$ If the norm of one row in matrix Z is very small (for example, less than le—6), then discard this row in matrix Z. Denote the number of the discarded row as r_d; $r^1 = r + r_d$ Accordingly, discard the corresponding column in matrix V^1. Rewrite the modified

International Journal of Scientific Engineering and Technology (ISSN : 2277-1581)
Volume No.1, Issue No.5, pg : 248-251 01 Nov. 2012

matrix Z and V^1 as Z' and by singular value decomposition, there exist orthonormal matrices $V^{11} \in R^{N \times (N-r1)}$. By singular value decomposition, there exist orthonormal matrices $U_b \in R^{(N-r1) \times (N-r1)}$ and $V_b \in R^{C \times C}$ such that $Z1 = U_b \Lambda_b V_b^T$.

Rewrite $Ub = (u_1,\ldots u_m, u_m+1,\ldots u_N\text{-}r^1)$ $U_b \in R^{(N-r1) \times (N-r1)}$ where u_i is the ith column of the orthonormal matrix Ub.

Denote $A = (u_1,\ldots u_m)_{(N-r1) \times m}$.

Thereby, W is the LDA projection matrix in which the Fisher criterion function

$J(W) = ti\,(W^T S_b W) / tr\{W^T S_w W)$ reaches the maximum.

3.0 EXPERIMENTAL RESULTS

FERET database is selected to evaluate the performance of our proposed optimization method on KSLDA.

3.1 DATABASES

Experiment on FERET database is performed. We have selected 70 people, 6 images for each individual. Face image variations in FERET database include pose, illumination, facial expression and aging. All images are aligned by the centers of eyes and mouth and then normalized with resolution92 112. The pixel value of each image will be normalized to 0 to 1. Images from one individual are shown below.

Figure 1: Images of one person from FERET database

3.2 PERFORMANCE EVALUATION

Table 1: Performance Evaluation

N	Direct LDA	Subspace LDA	Proposed Method
2	80.14%	84.71%	81.43%(15 Its.)
3	82.38%	87.24%	89.71% (14 Its.)
4	83.71%	90.86%	92.43% (17 Its.)
5	86.57%	94.00%	95.54% (15 Its.)

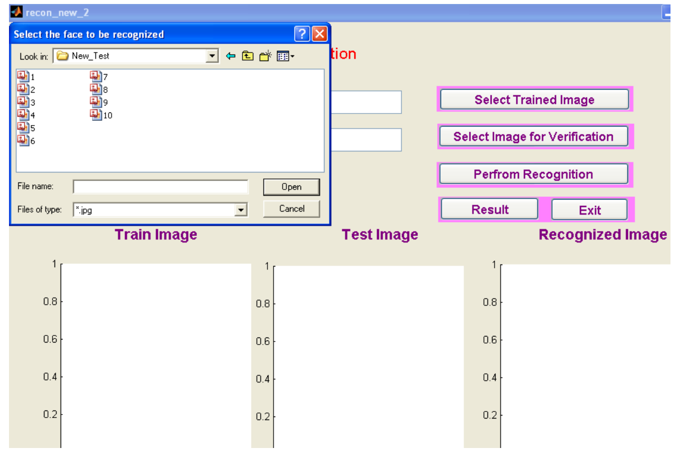
*N: number of images for training *Its. : Iterations

3.3 RECOGNITION USING KSLDA METHOD

1)Step 1: Browse for TrainDatabase.

2) Step 2: Similarly browse for TestDatabase.

3)Step 3:: Input the image number from the TestDatabase folder



4) Step 4: Press Result button and observe the result.

5) Step 5: The recognized image with the final GUI will be displayed

4.0 CONCLUSIONS

In this paper, Kernel Linear Discriminant Analysis Algorithm for face recognition works on many problems of face recognition. The KSLDA algorithm with Gaussian RBF kernel works on illumination occlusion, facial expression.The proposed *KSLDA* method has stronger generalization capability when complex variations exist in training and testing face images. The public available face database namely, the *FERET*, have been selected to evaluate the proposed method and other existing face recognition methods. When the image variations are not very large, the proposed method gives equally good results with the efficiency of 97% as compared to the current state-of-the-art methods. When the image variations are large, such as both illumination and pose variations, the proposed method clearly outperforms existing methods.

REFERENCES

[1] Milan Sonka, Vaclav Lavaca &Roger Bole, "Image Processing, analysis and Machine Vision".

[2]William Pratt, "Digital Image Processing".

[3] Rafael C. Gonzalez & Richard E.Woods, "Digital Image Processing".

[4]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.

[5] K. R. Mtiller, S. Mika, G. Riitsch,, K. Tsuda, B.Sch61kopf, "An Introduction to kernel-based learning algorithms", *IEEE Transactions on Neural Networks*, Vol. 12, pp.181-201, 2001.

[6] B. Sch61kopf, S. Mika, C. Burges, P. Knirsch, K.R. Miiller, G. R/itsch, A. Smola, "Input space vs.feature space in kernel-based methods" *IEEE Transactions on Neural Networks*, Vol. 10, No. 5, pp. 1000-1017, 1999.

[7]Xiao yang Tan, Song can Chen & Zhi-zhou,"Recognizing partially. Occluded Expression Variant Faces from Single Training Image per Person with SOM & soft K-NN Ensemble," IEEE Trans. Pattern Anal. Vol.16, No. k4, pp.875-855, July 2005.

[8] P.D.Deshpande, "Survey of Satellite Image Segmentation and Classification", NCETE2006, MAE, Alandi(D), Pune.

[9] www.sprigerlink.org

[10] www.mathtools.com

[11] www.imageprocessingplace .com