Full-Space LDA With Evolutionary Selection for Face Recognition[1]

Xin Li, Bin Li, Hong Chen
University of Science and Technology of China
MOE-Microsoft Key Laboratory of Multimedia Computing and Communication
Hefei, 230026 Anhui, China
{simonlee,hongchen}@mail.ustc.edu.cn
binli@ustc.edu.cn

XianJi Wang, ZhengQuan Zhuang
University of Science and Technology of China
MOE-Microsoft Key Laboratory of Multimedia Computing and Communication
Hefei, 230026 Anhui, China
xjw @ mail.ustc.edu.cn
zqzhuang @ ustc.edu.cn

Abstract

Linear Discriminant Analysis (LDA) is a popular feature extraction technique for face recognition. However, it often suffers from the small sample size problem when dealing with the high dimensional face data. Some approaches have been proposed to overcome this problem, but they usually utilize all eigenvectors of null or range subspaces of within-class scatter matrix(S_W *). However, experimental results testified that not all the eigenvectors in the full space of* S_W *are positive to the classification performance, some of which might be negative. As far as we know, there have been no effective ways to determine which eigenvectors should be adopted. This paper proposes a new method EDA+Full-space LDA, which takes full advantage of the discriminative information of the null and range subspaces of* S_W *by selecting an optimal subset of eigenvectors. An Estimation of Distribution Algorithm (EDA) is used to pursuit a subset of eigenvectors with significant discriminative information in full space of* S_W *. EDA+Full-space LDA is tested on ORL face image database. Experimental results show that our method outperforms other LDA methods.*

1. Introduction

Linear Discriminant Analysis (LDA)[1] is a well-known scheme for feature extraction and dimension reduction. It has been used widely in many applications such as face recognition, image retrieval, etc. The basic idea of LDA is to find a set of projection vectors maximizing the between-class scatter matrix (S_b) while minimizing the within-class scatter matrix (S_W) in the projected feature subspace. A major drawback of LDA is that it often suffers from the small sample size (S3) problem when dealing with the high dimensional face data. When there are not enough training samples, S_W would be singular, and it would be difficult to compute the LDA vectors.

In recent years, direct linear discriminant analysis (DLDA)[2] and null-space linear discriminant analysis (NLDA)[3] have been proposed to overcome the S3 problem in the face recognition. DLDA discards the null space of S_b, since the rank of S_b is smaller than of S_W , that might lose some information of null space of S_W. NLDA extracts discriminant information from the null space of S_W, however, when the number of training sample is large, the null space of S_W becomes small, and much discriminative information outside this null space will be lost. Both DLDA and NLDA may lose some discriminative information. In order to solve the S3 problem and still preserve all discriminative information, Optimal Fisher Linear Discriminant Algorithm (OFLDA) [4] and Dual-space LDA [5] are proposed to simultaneously apply discriminant analysis in the range and null subspaces of S_W respectively, here called Full-space LDA. A common drawback of Full-space LDA is that they use all eigenvector in the range and null subspaces of S_W. They assume that keeping all eigenvector means keeping all of the discriminative information that can improve the classification accuracy efficiently.

[1] Correspondence should be addressed to Bin Li(email: binli@ustc.edu.cn).

However, from the pattern classification point of view, this assumption may not incorrect. The main reason is that not all the eigenvectors in the full space of S_W are positive to the classification performance, some of which may be passive to classification. Therefore, choosing all eigenvector of the range and null subspaces of S_W as the bases for LDA may not be optimal. As far as we known, there is no systematic way to determine which eigenvectors should be used.

Along this line, this paper focuses on the Full-space LDA as mentioned and proposes a Full-space LDA with evolutionary selection. EDA is used to pursuit a subset of eigenvectors with significant discriminative information in full space of S_W. Compared with Full-Space LDA, the proposed method can effectively eliminate less discriminatory eigenvectors and improve classification accuracy. The experiments on ORL face database clearly demonstrate its efficacy.

The rest of the paper is organized as follows: Section 2 provides the background on LDA and Full-space LDA. Section 3 describes the details of EDA+ Full-space LDA. Experimental results are reported in Section 4. Finally, we draw the conclusion in Section 5.

2. Background of LDA and Full-space LDA

2.1. LDA

Let the training set contains C classes and each class X_i has n_i samples, x_k is a sample belonging to class X_i, m_i is the center of class X_i, m is the center of the whole training set. S_W and S_b are defined as Eq. (1) and Eq. (2),

$$S_w = \sum_{i=1}^{C} \sum_{k=C_i} (x_k - m_i)(x_k - m_i)^T \quad (1)$$

$$S_b = \sum_{i=1}^{C} n_i (m_i - m)(m_i - m)^T \quad (2)$$

The total-class scatter matrix is defined as Eq. (3)

$$S_t = S_b + S_w = \sum_{i=1}^{N} (x_i - m)(x_i - m)^T \quad (3)$$

LDA method tries to find a set of projection vectors $W_{opt} = (w_1, w_2, ..., w_L)$ that maximizes the ratio of the absolute value of the between-class scatter matrix to the absolute value of the within-class scatter matrix (Fisher's criterion), as defined in Eq. (4).

$$J(W_{opt}) = \arg\max_W \frac{|W^T S_b W|}{|W^T S_W W|} \quad (4)$$

If S_W is not singular, $w_1, w_2, ..., w_L$ are the eigenvectors of $S_W^{-1} S_b$, corresponding to L (C－1) largest eigenvalues. However, when the small sample size problem occurs, S_W becomes singular and the inverse of S_W does not exist. To avoid the singularity of S_W, Fisherface[6] , DLDA and NLDA is usually adopted. A common problem with all these LDA approaches is that they all prone to lose some discriminative information in the high dimensional face space. Then Full-space LDA (OFLDA and Dual-space LDA) is proposed ,which simultaneously apply discriminant analysis in the range and null subspaces of S_W in order to preserve the all discriminative information. The experimental results in [7] show that full-space LDA outperforms the other LDAs.

2.2. Full-space LDA

Let R^d be the original sample space, V be the range subspace of S_W, and $V^{\perp}$ be the null subspace of S_W. That is

$$V = span\{\alpha_k \mid S_w \alpha_k \neq 0 \quad , k = 1, \ldots, r\} \quad (5)$$

And

$$V^{\perp} = span\{\alpha_k \mid S_w \alpha_k = 0 \quad , k = r+1, \ldots, d\} \quad (6)$$

Where $r < d$ is the rank of S_W, $\{\alpha_1, \ldots, \alpha_d\}$ is an orthonormal set, and $\{\alpha_1, \ldots, \alpha_r\}$ is the set of orhtonormal eigenvectors corresponding to the nonzero eigenvalues of S_W.

From the range and null subspaces of S_W, the LDA projection vectors can be computed according to different criterions, respectively.

$$\begin{cases} J_{range}(W_{ropt}) = \arg\max \dfrac{|W_r^T S_b W_r|}{|W_r^T S_w W_r|} \\ W_r^T S_w W_r > 0 \end{cases} \quad (7)$$

$$\begin{cases} J_{null}(W_{nopt}) = \arg\max |W_n^T S_b W_n|. \\ W_n^T S_w W_n = 0 \end{cases} \quad (8)$$

This is to say, we intend to find the discrinimant vectors of the range subspace of S_W based on the Fisher criterion and utilize $J_{null}(W_{nopt})$ to get those of the null subspace of S_W. The two sets of discriminative features are combined for recognition.

2.3. Analysis on Full-space LDA

Full-space LDA uses all eigenvectors in the range and null subspaces of S_W. It assumes that keeping all eigenvectors means keeping all of the discriminative information that can improve the classification accuracy efficiently. However, experimental results show that this is not definitely right. The main reason probably is that not all the eigenvectors in the full space of S_W are positive to the classification performance, some of which may be negative to classification. So simply using all the eigenvectors is not optimal from the point of view of pattern classification.

The ORL database with 40 persons (three training images/person and two test images/person) is used to perform an experiment as an example. Fig.1 plots accuracy rates with the different number of eigenvectors in full space of S_W, the eigenvectors of the null subspace of S_W are all selected. When the number of eigenvector in range subspace increases to 51, the accuracy reaches the 96.25%. However, when the eigenvectors are all selected, the rate is just 91.25%. This intuitive observation indicates that some eigenvectors might be negative to classification accuracy. Therefore, a strategy for selecting the eigenvectors with significant discriminative information in full space of within-class scatter matrix is required.

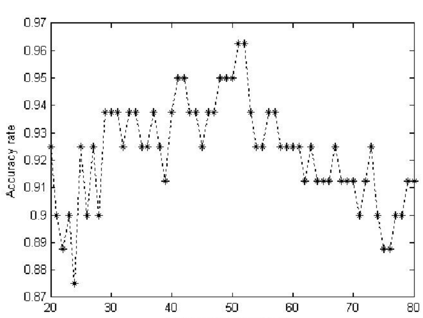


Figure 1.Accuracy rate with the different number of eigenvectors in full space of S_W, when the eigenvectors of null subspace of S_W are all selected.

3. EDA + Full-space LDA

As discussed in the previous section, we need to find the subset of eigenvectors with significant discriminative information in full space of S_W. In this section a new algorithm is proposed, which adopts Estimation of Distribution Algorithms [8] and establishes the optimal subset of eigenvectors through evolutionary selection.

EDAs emerged as a new form of evolutionary computation during the last decade. The basic idea of EDAs is to build a probabilistic model from the parental distribution in the parameter space and to generate offspring individuals by sampling from the model. The use of EDA for selecting subset of features was reported to yield speed-up in time with respect to the traditional wrapper methods for feature selection [9]. Since eigenvectors of full space of S_W are independent to each other, we adopt here the Univariate Marginal Distribution Algorithm (UMDA, [10]) to perform feature selection, which is a simple EDA based on the assumption that all variables are independent. The main scheme of the UMDA approach is shown in Figure 2.

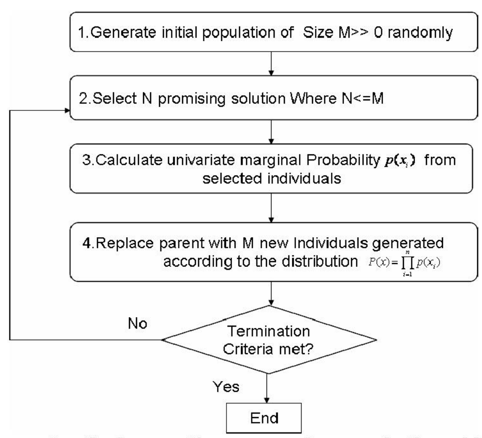


Figure 2. Schematic overview of the UMDA algorithm.

3.1. Chromosome Representation

We use binary string to represent the composition of optimal feature subset. Each bit $g_i (i = 1,2,\cdots,l)$ is corresponding to an eigenvector, that means: if $g_i = 1$, the *i*th eigenvector is selected into the optimal subset, otherwise, it is not selected. The length of chromosome is set to be l according to the number of all eigenvectors in the full space of S_W. A chromosome represents a solution of feature selection problem.

3.2. Fitness Function

Fitness function plays a crucial role in choosing offspring for the next generation from the current generation. It guides the direction of the evolution. In this paper, the fitness function is defined as Eq.(9):

$$fitness = \mu F(R) + \lambda F_{range}(G) + F_{nill}(G) \quad \textbf{(9)}$$

Where F(R) is the performance accuracy term in tuning set, $F_{range}(G)$ and $F_{null}(G)$ are the generalization terms which aim to select eigenvectors that have better generalization at the testing set. Here $F_{range}(G)$ and $F_{null}(G)$ are defined as Eq.(10) and Eq.(11):

$$F_{range}(G) = \min(D_{rb}) / \max(D_{rW}) \quad \textbf{(10)}$$

$$F_{null}(G) = \min(D_{nb}) / \max(D_{nW}) \quad \textbf{(11)}$$

Where D_{rb} and D_{rw} are the distance of between-class and within-class in the range subspace of S_W, respectively. D_{nb} and D_{nw} are the distance of between-class and within-class in the null subspace of S_W, respectively.

Weight μ and λ are empirically chosen to represent contribution of three terms to the fitness. By combining those two terms together(with proper weight μ and λ), EDA can evolve balanced results displaying good performance on both turning and testing set.

3.3. EDA+ Full space LDA

The EDA+ Full-space LDA algorithm can be described as follows:

i. A m-dimensional PCA subspace is constructed first, and all samples are mapped into this subspace, calculate the within-class and between-class scatter matrices S'_w and S'_b ;

ii. Calculate the eigenvector matrix $P = (\alpha_1, \alpha_2, \cdots, \alpha_q, \alpha_{q+1}, \cdots \alpha_m)$ of S'_w .Suppose the first q eigenvectors of S'_w correspond to its non-zeros eigenvalues. UMDA is performed to pursuit a subset of eigenvectors with significant discriminant information in full space of within-class scatter matrix on tuning set.

Step 1 A projection matrix P_1 is formed by the eigenvectors selected by the UMDA in the range subspace of S'_w . Define $\hat{S}_b = P_1^T S'_b P_1$ and $\hat{S}_w = P_1^T S'_w P_1$, the transformation matrix U_1 is then constructed by the eigenvectors corresponding to the largest eigenvalues of $\hat{S}_W^{-1}\hat{S}_b$. The first set of discriminant vectors are given by $W_1 = U_1 P_1$.

Step 2 A second projection matrix P_2 is formed by the eigenvectors selected by the UMDA in the null subspace of S'_w . Define $\tilde{S}_b = P_2^T S'_b P_2$, the transformation matrix U_2 is then constructed by the eigenvectors corresponding to the largest eigenvalues of $\tilde{S}_b = P_2^T S'_b P_2$. The second set of discriminant vectors are given by $W_2 = U_2 P_2$.

Step 3 Fuse the two kinds of features given by W1 and W2 using normalized-distance for classification.

Step 4 Calculate fitness function on tuning set; select a number of individuals; estimate probability model; generate the new population by sampling the estimated model.

An iterative procedure repeating step 1,2,3 and 4 is carried out until a termination criterion is met.

iii Let $W_{opt1}^T = W_1^T W_{pca}^T$ and $W_{opt2}^T = W_2^T W_{pca}^T$. Then W_{opt1}^T and W_{opt2}^T are the optimal projections of EDA+Full-space LDA

4. Experimental results

In this section, we apply our method to face recognition and compare it with the existing variant LDA methods such as Fisherface, NLDA, DLDA and Full-space LDA approaches.

The proposed method is tested on ORL face database, which contains 40 people, each person has 10 different images. The images of the same person are taken at different times, under slightly varying lighting conditions and with various facial expressions. The images in the database are grayscale and the size are rescaled to be 92x112. Figure 3 show ten images of one person in ORL.

Figure 3. Ten images of one person in ORL face database

In this experiment, the ORL face database is broken into three disjoint sets: training, tuning and testing. The tuning set is used to provide tuning feedback to the UMDA, to select a subset of eigenvectors with significant discriminant information in full space of within-class scatter matrix by UMDA, as described previously in section 3.2. Once the UMDA run was finished(the optimal subset of eigenvectors are selected), the test set was used to perform unbiased testing on the subset found by UMDA. For every person, we use the first three images for training, the 4,5 images for turning and the remaining five for testing. A 1-NN(Nearest Neighbor) classifier is adopted. The parameters of UMDA used in this experiment are set as follows: population size is 500; the maximum number of generation is 30; the number of the selected promising solution is 200.The recognition results are shown in table 1.

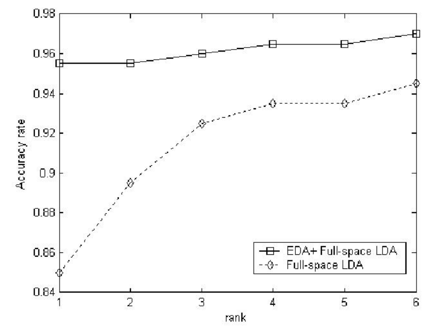
Table 1. Comparison of recognition result with Fisherface, DLDA, NLDA, Full-space LDA and proposed method

Method	Number of features	Recognition performance	
		Turning set	Testing set
Fisherface	39	0.875	0.815
NLDA	39	0.9	0.835
DLDA	39	0.8875	0.795
Full-space LDA	78	0.9125	0.85
EDA+Full-space LDA	59	0.9875	0.955

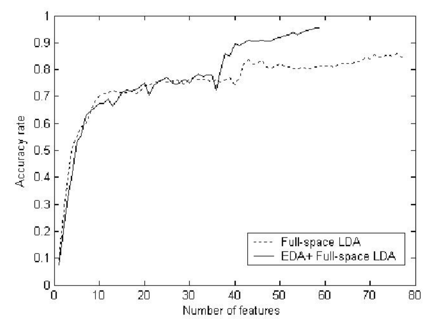
Table 1 shows the comparative tuning and testing performance of Fisherface, DLDA, NLDA, Full-space LDA and EDA+Full-space LDA. One can see that EDA+Full-space LDA outperforms the other LDA methods. The optimal number of eigenvectors be selected in the range and null subspace of S_W is presented in Table 2.

Table 2. Comparison of number of eigenvectors in ranger and null subspace of S_W with Full-space LDA and proposed method

Method	Number of eigenvectors in ranger subspace	Number of eigenvector in null subspace
Full-space LDA	80	39
EDA+Full-space LDA	35	24



(a)



(b)

Figure 4. (a)Cumulative accuracy rates and (b)Rank-1 accuracy rates with the different number of features of Full-space LDA and EDA+Full-space LDA on ORL face database

From Fig.4(a) , by comparing the accuracy with Full-space LDA, the effectiveness of EDA+Full-space LDA can be seen. By employing the EDA to select eigenvectors in the full subspace of S_W , EDA+Full-space LDA can increase the accuracy rate from 85% to 95.5% at rank 1. Fig.4.(b) show Rank-1 accuracy rates with the different number of features. It can be seen that the recognition rate is close before the null subspace eigenvectors are utilized. It indicates that the recognition rate of the proposed method does not decrease even though the number of eigenvectors of S_W is less than the number of Full-space LDA. After the eigenvectors in null-space are considered, the accuracy rate of EDA+Full-space LDA increases obviously.

5. Conclusions

In this paper, a EDA+full-space LDA approach for the eigenvectors selection in the full space of S_W is proposed. EDA is used to pursuit a subset of eigenvectors with significant discriminative information. EDA+Full-space LDA is tested on ORL face image database. Experimental results demonstrate that our method is better than others LDA methods. In future research, we will investigate other more effective definition of fitness function for feature selection.

Acknowledgement

The work is partially supported by the Natural Science Foundation of China(NSFC) under grand No.60401015 and No. 60518002, the Natural Science Foundation of Anhui province under grand No. 050420201, and the Science Research Fund of MOE-Microsoft Key Laboratory of Multimedia Computing and Communication under grant No.05071811.

References

[1] K. Fukunaga: *Introduction to statistical pattern recognition*, Academic Press, Boston, 2nd edition, 1990.

[2] H. Yu and J. Yang: A direct lda algorithm for high-dimensional data with application to face recognition. *Pattern Recognition*, 2001.

[3] L. Chen, H. Liao, M. Ko, J. Lin, and G. Yu.:A new lda-based face recognition system which can solve the samll sample size problem. *Pattern Recognition*, 2000.

[4] J. Yang and J.Yang : Optimal FLD algorithm for facial feature extraction, *SPIE Proceedings of the Intelligent Robots and Computer Vision XX: Algorithms,Techniques, and Active Vision*, vol. 4572, pp. 438-444, 2001.

[5] X.Wang and X.Tang: Dual-space linear discrminant analysis for face recognition, *Proceeding of CVPR'04*, vol.2.pp.564-569, 2004.

[6] P.N. Belhumeur, J. Hespanha, and D.J. Kriegman.: Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Trans. on PAMI*, 19(7):711–720, 1997.

[7] Thomaz, C.E and Gillies, D.F.: A Maximum Uncertainty LDA-Based Approach for Limited Sample Size Problems — With Application to Face Recognition. *Computer Graphics and Image Processing,2005.* SIBGRAPI 2005 Page(s):89-96

[8] Larrañaga P, Lozano JA: Estimation of Distribution Algorithms. A New Tool for Evolutionary Computation. *Kluwer Academic* Publishers; 2001.

[9] Saeys Y, Degroeve S, Aeyels D, Van de Peer Y, Rouzé P: Fast feature selection using a simple Estimation of Distribution Algorithm: A case study on splice site prediction. *Bioinformatics* 2003, 19(Suppl 2):II179-II188.

[10] Muhlenbein H: The equation for response to selection and its use for prediction. *Evol Comput* 1997, 5:303-346.