Double Discriminant Analysis for Face Recognition

S.Aruna Mastani

Associate Professor, Intell Engg.College, Anantapur, INDIA

Dr.K.Soundararajan,

Rector, Jawaharlal Nehru Technological University, Anantapur, INDIA

Summary
Feature selection for face representation is one of the central issues for any face recognition system. Finding a lower dimensional feature space with enhanced discriminating power is one of the important tasks. The traditional subspace methods represent each face image as a point in the disciminant subspace that is shared by all faces of different subject (classes). Such type of representation fails to accurately represent the most discriminate features related to one class of face, so in order to extract features that capture a particular class's notion of similarity and differ much from remaining classes is modeled. In this paper we propose a new method called "Double Discriminant Analysis" Which first performs PCA (Principal Component Analysis) to reduce the sample size and extract the features that separates individual class faces maximally. Then by projecting these samples over to the null space of within class matrix the intra class variance is reduced to extract the most discriminative feature vectors for which an Individual class oriented subspace is found for each class 'i' along which the intra class variance is minimum, and separates well from all the remaining classes. This individual subspace for each class 'i' found to express most discriminative power that helps in classification and thus developing an effective face recognition system.

Key words:
Face recognition, Lined Discriminant Analysis, Double discriminant Analysis.

1. Introduction

In recent years the appearance based methods of face recognition has once again regained their significance, because of their simplicity of modeling directly on image pixels with out the need to create geometric or algebraic representation. The main disadvantage of these methods is that the accuracy of recognition degrades when any testing image is slightly varies in its appearance to those images used for training. To overcome this problem many methods has evolved which concentrates on extracting a part of the total face space where the variance between same class images is minimized and variance between different classes is maximized. These methods are known to be the subspace methods. Among the existing subspace methods Principal Component Analysis (Eigen face method) [1,2] is the first break through in the field of subspace techniques. It uses the karhunon Loeve Transforms (KLT) to produce the most expressive subspace for face representation, but is not effective in finding most Discriminant subspace, so it is mostly used a preprocessing step to reduce the high-dimensionality of raw face images. LDA (Lined Discriminant Analysis) [3] is another powerful subspace method that aids in finding the most Discriminant subspace or extracting a set of features that best separates the face classes, i.e. for instance given three persons A,B,C the features that aid in distinguishing between A and B are different from those that discriminate between B and C. LDA finds the best features that discriminate between different classes. However the efficiency of LDA degrades if the number of classes is more as it is very difficult to extract discriminants that separate each pair of input classes. The method proposed in this paper overcomes this problem by extracting a separate set of features for each class individually, Such that those features are efficient in discriminating a particular class from rest of the face classes.

2. Related work

Being an efficient technique LDA was extended/Modified by many researchers to overcome its drawbacks and strengthen its merits. Numerous methods have been proposed to solve the important problem of all face recognition system the small sample size. Swets and Weng[3,4] proposed a two stage PCA + LDA method known as Fisher face method in which PCA is used for dimension reduction so as to make 'S_W' (with in class scatter) non-singular before the application of LDA. Yang

Manuscript received February 5, 2009
Manuscript revised February 20, 2009

et.a1[5] proposed a more straight forward method that makes use of null space approach of removing null space of S_B (Between Class Scatter) and seeks a projection that minimize S_W (Called DLDA). Chen et.a1[6] proposed a method of projecting samples on to null space of S_W to make it zero and maximize S_B . S.Z. Zhou and Rama chellappa [7] proposed a different approach that uses multiple exemplars (samples) for each class while calculating the S_W and S_B instead of single exemplar (mean of class) by the LDA. This is known as Multiple exemplars Discriminant Analysis (MEDA). Bias map [8] performs a class specific LDA Inspired by the concept of above methods. Different from the above method of LDA algorithms, the DDA uses different concept in extracting a class-specific subspace for each class individual. Rest of the paper is organized as follows. In section 3 different LDA algorithms are reviewed. In section 4 DDA is proposed and discussed In section5 experimental results are presented finally section6 ends with some conclusion and the future plans.

3 Representation of Face images in Subspace Learning Frame Work:

The subspace learning framework has been most influential in the face recognition (FR) research. Under this framework, the problem of learning low dimensional feature representation from examples can be stated as follows : Given a training set, composed of 'C' classes with 'N_i' samples per class is where X = { 'x_m^i'}, where 'x_m^i'' denote a m^{th} sample from i^{th} class and a total of 'N' sample face images are available in the set. For computational convenience, each image is represented as a column vector of length $n = I_W \times I_h$ by lexicographic ordering of the pixel elements, i.e. $x_m^i \in R^n$, where ($I_W \times I_h$) is the image size, and R^n denotes the n-dimensional real space. Taking as input such a set X, the objective of appearance-based learning is to find, based on optimization of certain separability criteria, a transformation 'W' which produces a low dimensional feature representation 'y_m^i' = $W(x_m^i)$, $y_m^i \in R^M$ and $M << n$, intrinsic to face objects with enhanced discriminatory power for pattern classification.

Supposing that the small sample size problem exist satisfying the condition 'n > N-C' the scattering matrices S_T, S_B *and* S_W are defined as

$$S_T = \sum_{i=1}^{C} \sum_{m=1}^{Ni} (x_m^i - \mu_{full})\,(x_m^i - \mu_{full})^T$$

$$S_B = \sum_{i=1}^{C} N_i (\mu_i - \mu_{full})\,(\mu_i - \mu_{full})^T$$

$$S_W = \sum_{i=1}^{C} \sum_{m=1}^{Ni} (x_m^i - \mu_i)\,(x_m^i - \mu_i)^T$$

Where $\mu_{full} = \frac{1}{N}\sum_{i=1}^{C}\sum_{m=1}^{Ni} x_m^i$ and $\mu_i = \frac{1}{N}\sum_{m=1}^{Ni} x_m^i$

4. Review of PCA and LDA methods

4.1 PCA:

In the statistical pattern recognition literature, Principal Component Analysis (PCA) is one of the most popular tools for data reduction and feature extraction. The main idea of PCA technique is to project the samples over a subspace which maximizes the variance and minimizes the error. This is done by selecting the eigenvectors corresponding to maximum eigen values of the covariance matrix S_T defined as

$$S_T = \sum_{i=1}^{C} \sum_{m=1}^{Ni} (x_m^i - \mu_{full})\,(x_m^i - \mu_{full})^T$$

The Projection transformation W^{PCA} is formed with N-1 eigenvectors corresponding to the largest eigenvalues, and they form a low-dimensional subspace, over which the samples are projected to reduce their dimension to N-1

4.2 LDA

The conventional two-stage LDA –based algorithms, such as Fisher face [9], Direct LDA[5], Null space based LDA [6,10], etc., all lose some useful information to ensure the non singularity of with in class covariance matrix S_W by discarding a part of discriminant subspaces.

4.2.1 Basic LDA algorithm

LDA technique finds a set of basic vectors that maximizes the ratio of between class scatter to within class scatter in order to maximize the Inter class variance and minimize the Intra class variance the optimal projection matrix $W^{LDA} = (w_1\ w_2 \ldots w_{C-1})$ *is* readily computed by solving a generalized eigenvalue problem where w_i , i=1……C -1 satisfy the following criterion

$$J(W^{LDA}) = \underset{W^{LDA}}{\text{argmax}} \frac{\left|W^{LDA^T} S_B W^{LDA}\right|}{\left|W^{LDA^T} S_W W^{LDA}\right|} = optMax/1$$

If S_W is invertible, $W^{LDA^T} S_W W^{LDA}$ is always positive for every nonzero *w* since S_W is positive definite, in such case the above equation can be used directly to extract asset of optimal feature vectors. However it is not always possible in real world applications such as face recognition, that S_W is of full rank. Therefore there always exist vectors making $w^T S_W w$ be zero. These vectors turn out to be very effective for classification if they satisfy $w^T S_B w > 0$. These are the ideas of Null space based LDA methods discussed below.

4.2.2 Direct LDA

This tries to find the optimal projection $W^{DLDA} = (w_1 w_2, \ldots w_{C-1})$ which satisfies the following criterion

$$J(W^{LDA}) = \underset{W^{DLDA}}{\arg\max} \frac{\left|W^{DLDA^T} S_B W^{DLDA}\right|}{\left|W^{DLDA^T} S_W W^{DLDA}\right|} = 1/optMin$$

4.2.3 Null Space Based LDA(NLDA)

This method [12] reduces the feature space into low dimensional subspace using PCA and then tries to find optimal projection $W^{NLDA} = (w_1 w_2, \ldots w_{C-1})$ which satisfies tne following criterion

$$J(W^{NLDA}) = \underset{W^{NLDA}}{\arg\max} \frac{\left|W^{NLDA^T} S_B W^{NLDA}\right|}{\left|W^{NLDA^T} S_W W^{NLDA}\right|} = optMax/0$$

4.2.4 Biased LDA

It finds optimal projection that separate a particular class (positive class) from all the remaining classes by considering them as single class (Negative class) .this technique is used in interactive multimedia image retrieval [8].

5. Double Discriminant Ananlysis (DDA)

5.1Algorithm

INPUT: A set o 'N' training face images { x_m^i } represented as n – dimensional vector belonging to 'C' class

OUTPUT: A lower dimensional subspaces of each class where it discriminates best from remaining classes (face images with enhanced discriminatory power with respect to individual classes.)

- Construct the sample mean of total training set μ_{full} and covariance matrix S_T and thus obtain the PCA projection matrix W^{PCA} whose columns are leading eigenvectors of S_T .Generally as the rank of S_T is *min (n, N-1)* projection matrix W^{PCA} is of dimension n x N-1.
- Project the training data samples in to PCA space thus the dimension of samples reduces to 'N-1'. each sample $y_m^i = W^{PCA^T} (x_m^i - \mu_{full})$
- Obtain the within class scattering matrix of the projected samples $S_W' = W^{PCA^T} S_W W^{PCA}$
- Evaluate the NULL space of S_W' which provide the clustering of projected samples around their respective means .
- Obtain the NULL space projection matrix W^{NULL} whose columns are (C-1) trailing eigenvectors. The dimension of projection matrix W^{NULL} is (N-1)x(C-1) project the samples in to NULL space
- The projected sample are $z_m^i = (W^{NULL})^T y_m^i$ For these projected samples obtain class specific mean μ_{Null}^i for í =1, 2,..C
- Individual class-oriented subspace for each class is generated as follows.

Consider each class 'i' as positive class and remaining classes put together as another single class called negative class obtain a positive and negative scattering matrices for each class 'í' Positive (SP_i) and Negative (SN_i) scattering matrices are defined as

$$SP_i = \sum_{m=1}^{N_i} (z_m^i - \mu_{Null}^i)(z_m^i - \mu_{Null}^i)^T$$

$$SN_i = \sum_{\substack{j=1 \\ i \neq j}}^{C} (z_m^i - \mu_{Null}^j)(z_m^i - \mu_{Null}^j)^T$$

Evaluate for each class 'í' optimum projection $W_i^{DDA} = (w_1 w_2, \ldots w_{C-1})$ which maximize

$$J(W_i^{DDA}) = \arg\max_{W_i^{DDA}} \frac{\left|W_i^{DDA^T} SN_i W_i^{DDA}\right|}{\left|W_i^{DDA} SP_i W_i^{DDA}\right|}$$

- To classify a test image '$\boldsymbol{t}$' project it over the space $K = (W^{PCA} W^{NULL})^T \boldsymbol{t}$ and then projecting the resultant image over class oriented subspaces.

A query / Test image is said to belong to a class 'C' for which

C = arg min $d_i\ (K)$

Where $d_i(K) = \left\| W_i^{DDA^T} (\mu^i_{Null} - K \right\|$

As separate subspace is derived for each class, when a query /test image belongs to a trained class 'i', its distance from that class's mean is a sharp minima , i.e. the $d_i(K)$ must be at least 'r' times less than any of the distances $d_c(K)$from other classes. Where 'r' called a threshold value ranges between 0 to 1 .

$d_i\ (K) < r\ d_c(K)$ *for all* $i \neq c$ *and* $0 < r < 1$

in addition to this condition if two or more classes have almost same distance metric ,then the test image belongs to an unknown class

---

The first step of the algorithm (method) is to perform PCA on training set of image for dimension reduction and to retain the directions (features) of large between class variances among the data. Thus the projected samples are much more variant from one another with reduced dimensions. The second step is performed with an idea of clustering together the samples of each class around their respective means, by projecting the samples once again over the NULL space of within class scattering matrix (S_W'). This process brings every sample of a class closer to its mean. However the first step helps in maintaining a required distance from one cluster (class) to another. Thus the resultant samples forms most discriminating vectors which are more suitable to find class oriented subspace. In the case of As number of samples used is less than the dimensions of the sample and after performing PCA the between class scatter matrix is of dimension (N-1)x(N-1) which is of full rank, the objective of projecting S_B over null space of S_W' as in the case of DLDA[11] is not required performing explicitly the DLDA thus Discriminant analysis is performed once.

Generally LDA methods have three main drawbacks. LDA makes an error in discriminating a pair of closely related or highly similar classes. Its accuracy decreases if number of class is more as it becomes difficult of extract features that separate a pair of classes. And the last is as the mean of class is taken as prototype in representing each class, and if in case less number of samples are present in each class the calculation of discriminate direction become unstable.

The first problem is already resolved as we are using highly discriminated vectors. The next problem is solved through a method of generating class specific LDA where it is required to separate only a pair of classes at one time (i.e. Positive class and negative class) and each time one class is taken as positive class and remaining classes put together forms the second class. Thus performing Discriminant Analysis for Second time. The third problem is resolved through the concept of using all the samples of the class while estimating the discriminating directions as in[7] thus the scattering matrices defined here are different from the scattering matrices of traditional methods of LDA Here samples means of respective classes are scattered from the samples of specified class which is considered as positive. This modification is made in order to reduce the burden of heavy computation in calculating the scattering matrices, still preserving the advantage of multiple examples , As we already brought the samples closer to their means, and it is sufficient to enroll all the examples (samples) of interested class (positive)

***Fig1**:Images from Yale Database, showing the variations in lightning(row1) and expression(row2)*

6. Experimental Results

The effectiveness of the proposed method is verified by conducting extensive experiments on three well-known databases ORL [14] Yale database [13] and the UMIST[12]. We have compared the performance of DDA with that of state of art methods like LDA, DLDA,and NLDA for face recognition under slight pose and expression variations (ORL Database),and Expression and illumination variation(Yale database) and for Extreme pose variations with (UMIST database) results are tabulated.

6.1 ORL Database:

This database consists of images of 40 individuals. Each image contains a face area with little background. all the images are grey scale with a resolution of 112 x 92. The ten different images of each person include a slight variation in pose and expression, while the illumination remains constant. The total Database is divided in to two types of training sets one containing 15 classes (Persons) and the other with 40 classes. The number of samples 'k' per class varies as 3, 5, 7. In each round 'k' samples are selected randomly for training and remaining are used for testing. For each 'k' 10 test are performed and results are tabulated in table1 and table2 .this is repeated for the two types of training sets containing 15 and 40 classes. We also performed tests by varying the number of discriminate features and found maximum efficiency is achieved when optimum vectors are equal to 'number of classes-1 in all the methods, however the recognition rate is high for DDA even for less number of discriminants used for projection compared to others.

Table:1 Recognition rates of ORL Database (percentage)

Number of classes	Number of samples per class	LDA	DLDA	NLDA	DDA
15	3	84.3	88.9	90.4	93
	5	92	93.2	92.8	96
	7	97	97.8	96	97.2

Table: 2 Recognition rates of ORL Database (percentage)

Number of classes	Number of samples per class	LDA	DLDA	NLDA	DDA
40	3	76	80.1	86	94
	5	89	91.8	95.2	95.8
	7	90	94.2	96	96.8

6.2 YALE Database

This database consists of images 15 individuals each providing 11 different images. These 11 different images of each person provide extreme variations in expression and slight changes in lightning, however the pose remains constant (front view).All the images are aligned manually with respect to center points of eyes and cropped to the face area and resized to a resolution of 112 x 92. Same tests are performed as that of the ORL database and results are tabulated as below in table:3

Table 3: Recognition rates of Yale Database (Percentage)

Number of classes	Number of samples per class	LDA	DLDA	NLDA	DDA
15	3	80	82.3	90.2	93
	5	90	90.2	94.2	95.8
	7	93.8	95.1	95.2	97

6.3 UMIST Database

It is a multi view database, consisting of 575 images of 20 people, each covering a wide range of pose from profile to frontal views as well as range of appearance, each image is scaled to dimension of 112 x 92. Same tests are performed as that of the ORL database and results are tabulated as below in table: 4

Table 4: Recognition rates of UMIST database (Percentage)

Number of classes	Number of samples per class	LDA	DLDA	NLDA	DDA
15	3	56	72.9	76.4	83.23
	5	62	78.2	80	84
	7	68.7	80.8	89	87

All the methods show an increase in recognition rate with increase in number of samples per class. By comparing results obtained from two databases. One can easily, observe a decrease in recognition rate of LDA, DLDA, and NLDA methods When applied to Yale database compared to ORL. This is due to extreme variation of expression and illumination conditions present in the Yale database compared to variation of expression and mild variation of pose with constant lightning conditions of ORL. However the proposed DDA method show almost equal recognition rate irrespective of database. This shows that the proposed method is more efficient in selecting the set of most discriminate features of an every specific class that reduces the error rates of a face recognition system.

7. Conclusions

In this paper we propose a new subspace algorithm 'DDA' that in spite of incorporating all the advantages of prevailing LDA techniques adds a new concept of generating most disciminant class specific features. It utilizes the idea of NULL space analysis and proposes a straightforward method of overcoming the drawbacks of BDA when used to our application of face recognition system. Projecting all the samples over the NULL space brings them closer to their respective mean images. As a result, we defined new scattering matrices such that DDA maximizes the scatter of samples of positive class from the mean images of negative classes and minimizes the positive scatter(i.e. scatter of positive samples from their mean).this reduces the computation time. The proposed method is compared with the three existing methods LDA, DLDA, NLDA and found to be more superior and effective in extracting the most disciminant individual class specific features for developing an effective face recognition system. As a part of future work the work is progressing towards developing "*Class specific Null space DDA"* and non -linear approaches using kernel methods for both the algorithms.

References

[1] M. Turk and A. Pentland, "Eigenfaces for Recognition," *J. Cogn. Neurosci.,vol. 3, no. 1, pp. 72–86, 1991.*

[2]. M. Turk and A. Pentland. "Face recognition using eigenfaces". *In Proceedings of IEEE Computer Vision and Pattern Recognition, pages 586-590, Maui, Hawaii, Dec. 1991.*

[3]. Swets, D.L., Weng, J.: "Using Discriminant Eigen features for Image Retrieval*". IEEE Trans. Pattern Analysis and Machine Intelligence. Volume 18. Issue 8. (1996) 831-836*

[4] W. Zhao, R. Chellappa, and P. J. Phillips, "Subspace linear disciminant analysis for face *recognition," Center for Autom. Res., Univ. Maryland, College Park, Tech. Rep. CAR-TR-914, vol.8, 1999.*

[5]. Yang, J., Yu, H., Kunz, W.: "An Efficient LDA Algorithm for Face Recognition". *In Proceedings of the Sixth International Conference on Control, Automation, Robotics and Vision (2000)*

[6]. Chen, L.F., Liao, H.Y.M., Lin, J.C., Ko, M.T., Yu, G.J.: "A New LDA-based Face Recognition System Which Can Solve the Small Sample Size Problem*". Pattern Recognition. Volume 33. Issue 10. (2000) 1713-1726*

[7] S. K. Zhou and R. Chellappa., \Multiple-Exemplar Discriminant Analysis for Face Recognition.," in *Internationl Conference on Pattern Recognition, vol. 4, pp. 191-194, 2004.*

[8]. X. S. Zhou and T. Huang, "Small sample learning during multimedia retrieval using bias map," presented at the *IEEE Conf. Computer Vision and Pattern Recognition, Dec. 2001.*

[9] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman. "Eigenfaces vs Fisherfaces":Recognition using class specific linear projection. *IEEE Transactionson Pattern Analysis and Machine Intelligence, pp:711–720, 1997.*

[10]. L Chen, H.Liao, M. Ko,J Lin, and G.. Yu. "A new LDA-based face recognition system which can solve the small sample size problem". Pattern Recognition", pp:1713–1726, 2000.

[11]. Huang, R., Liu, Q.S., Lu, H.Q., Ma, S.D.: "Solving the Small Sample Size Problem of LDA". *In proceedings of the 16th International Conference on Pattern Recognition (ICPR'02), Quebec, Canada (2002)*

[12]Liu.W.Wang ,Y.W.Li.S.Z.Tang "Null Space Based Kernal Fisher Discriminant Analysis for Face Recognition"*In Proceedings of IEEE, International On Automatic Face And Gesture Recognition* .PP 369-374(2004)

[13]. fttp:// plucky.cs.yale.edu/cvc/pub/images/yalefacesB/..

[14]. http://www.uk.research.att.com/facedatabase.html

[15]. http://images.ee.umist.ac.uk/danny/database.html

S. Aruna Mastani received the B.E degree in Electronics and Communication Engineering from JNTU College of Engineering Anantapur in 1998. She received the M.Tech degree in Digital Systems and Computer Electronics from JNTU College of Engineering, Anantapur, India in 2002. She worked as academic assistant in JNTUCE, Anantapur. She worked as Assistant Professor in Intell Engineering College Anantapur, India from 1999 to 2004. She promoted as associate professor in 2005 and pursuing the Ph.D work in the area of Digital Image Processing.

Dr.K.Soundararajan received the B.E degree in Electronics and Communications from S.V.U, Tirupati and the M.Tech degree in Instrumentation from J.N.T.U, Kakinada. He received Ph.D from Indian Institute of Technology, Roorkee. He is having 22 years of teaching experience. He got the best teacher award for the year 2005, President of India Award in Bharat Scouts & Guides in 1968, Best Paper Award in 1990–91 from Institution of Engineers (India) for best technical paper published in Journal and the Best Teacher Award in 2006 by the State Government of Andhra Pradesh, India. He is an Expert Committee member, where he has acted as inspection committee member for AICTE (South Western Region) from 1998 till today, JNT University affiliation committee member for several colleges, Academic Audit Committee member for the three Constituent colleges of JNT University and Valuation expert for UG & PG project works of JNTU, SVU, Nargarjuna University, SK University, Osmania University and Bangalore University. He published 21 international journals/conferences and 27 national journals/conferences and he participated in 12 national seminars. He guided 2 Ph.Ds and 10 Ph.Ds of research work is under progress. Presently, he is working as Rector of Jawaharlal Nehru Technological University, Anantapur, India.