Two-class Linear Discriminant Analysis for Face Recognition

Hazim Kemal Ekenel, Rainer Stiefelhagen
Interactive Systems Labs, Computer Science Department, Universität Karlsruhe (TH)
Am Fasanengarten 5, 76131, Karlsruhe, Germany
email: {ekenel, stiefel}@ira.uka.de

Abstract

In this paper, we present a novel face recognition system that uses two-class linear discriminant analysis for classification. In this approach a single M-class linear discriminant classifier is divided into M two-class linear discriminant classifiers. This formulation provides many advantages like more discrimination between classes, simpler calculation of projection vectors and easier update of the database with new individuals. We tested the proposed algorithm on the CMU PIE and Yale face databases. Significant performance improvements are observed, especially when the number of individuals to be classified increases.

1. Introduction

The problem of face recognition has become one of the most addressed problems in computer vision research. The intense research efforts on face recognition mainly fuels from the potential application areas, such as secure authentication, surveillance and smart environments.

Since the beginning of 1990s, appearance based approaches have been dominating the face recognition research. In Eigenfaces approach [8] the face is represented by a new orthonormal basis vector set that best reconstructs the face image in terms of mean-square error. Although this approach is very successful in representing the face images in a compact form, it has received criticism of being only a representation technique, and not aiming at classification. The approach of Fisherfaces [6], which is based on linear discriminant analysis (LDA), has been shown to diminish the effect of intra-individual variation under severe illumination and expression variations, in the classification step. LDA takes into consideration the class information and directly aims at classification. It still uses PCA for low dimensional representation and projects these low dimensional representation vectors onto a lower dimensional space where the ratio of between-class scatter to within-class scatter is maximized. Although this approach may seem more suitable for classification, it has many problems. First of all, in the absence of adequate training data, LDA may not find the best projection directions for classification [5]. Also, its performance for a large number of classes is still an open issue. Moreover, if new face classes are added to the classification task, all projection directions need to be recalculated. To address these problems, many studies have been conducted [9, 3, 4]. In [9] to eliminate the possibility of losing discriminatory information due to a separate PCA step a direct LDA algorithm is proposed. An approach that tries to refine the optimality criteria in multi-class LDA and tries to solve the small sample size problem is presented in [3]. Similarly, in [4] to improve the optimality criteria of multi-class LDA, a weighting scheme is proposed.

In this study we propose a two-class formulation of the multi-class LDA where each individual in the database has a separate projection vector which discriminates him/her from the rest of the individuals. The main rationale to resort to this classification scheme is to increase the discrimination between classes. In addition, there are many other benefits of this approach. First, the calculation of the projection vectors is much easier in two-class LDA. Second, the problem of calculating wrong projection directions due to insufficient training data is less likely to occur, since at least one of the two classes has sufficient training data. That is, it is quite reasonable to assume that the "other individuals" class has adequate training data. Third, there is no difficulty to extend the approach to work for a large number of classes. If a new class is added, only the projection vector for the new individual is required to be calculated.

The outline of the paper is as follows: In Section 2, two-class and multi-class linear discriminant analysis are explained. Experimental results are presented and discussed in Section 3. Finally, in Section 4, the conclusions are given.

2. Linear Discriminant Analysis

In the following subsections the derivation of multi-class LDA, two-class LDA, and the way they are used to identify faces are explained.

2.1. Multi-class linear discriminant analysis

The derivation of multi-class LDA is as follows (In the formulas, lowercase letters denote scalar values, bold lowercase letters denote vectors, uppercase letters denote matrices, and superscript T denotes transpose operation). Let's denote the low dimensional vector that represents face image as $\mathbf{x_k}$, the mean value of the i-th class as $\mathbf{m_i}$, and the mean value of all data as $\mathbf{m}$. The between class scatter matrix S_B can be calculated as

$$S_B = \sum_{i=1}^{c} n_i(\mathbf{m_i} - \mathbf{m})(\mathbf{m_i} - \mathbf{m})^T, \tag{1}$$

where c is the number of classes and n_i represents the number of face samples in the i-th class. The within class scatter matrix is defined as

$$S_W = \sum_{i=1}^{c} \sum_{x_k \in X} (\mathbf{x_k} - \mathbf{m_i})(\mathbf{x_k} - \mathbf{m_i})^T. \tag{2}$$

LDA tries to find a projection direction that maximizes the ratio of between-class scatter to within-class scatter as shown below:

$$W_{opt} = argmax_W \frac{|W^T S_B W|}{|W^T S_W W|}, \tag{3}$$

where W_{opt} represents the optimal projection matrix. In its columns, it contains the generalized eigenvectors that correspond to the largest eigenvalues in

$$S_B \mathbf{w_i} = \lambda S_W \mathbf{w_i}. \tag{4}$$

If S_W is non-singular, W_{opt} can be calculated by simply computing Eigenvectors of $S_W^{-1} S_B$. Instead of computing Eigenvectors of $S_W^{-1} S_B$, it's preferable to diagonalize S_W and S_B simultaneously to obtain the same result [2].

2.2. Face recognition using multi-class LDA

To perform face recognition using multi-class LDA, at first the dimensionality of the input image should be reduced to $n - c$ to avoid the singularity of S_W (here, n is the total number of face image samples in the database, and c is the number of classes). Afterwards W_{opt} can be computed using the steps explained in Section 2.1. W_{opt} projects the $n - c$ dimensional input feature vector to the $c - 1$ dimensional decision space. When a test image arrives, it is first represented in a low dimensional form, and then it is projected to the decision space using the projection matrix, W_{opt}, computed in the training stage. In decision space, distances between the class mean vectors and the test input are calculated. Finally, the identity of the individual is assigned with the class label that has the minimum distance.

2.3. Two-class linear discriminant analysis

In the two-class case the between class scatter matrix is calculated by subtracting the mean values of the classes from each other

$$S_B = (\mathbf{m_1} - \mathbf{m_2})(\mathbf{m_1} - \mathbf{m_2})^T. \tag{5}$$

Recall that in the multi-class case only the general scatter is taken into consideration, however, this does not guarantee desired discrimination between all classes. Similar to the multi-class case, again the goal is to find a projection direction that maximizes the ratio of between-class scatter to within-class scatter

$$\mathbf{w_{opt}} = argmax_W \frac{|\mathbf{w}^T S_B \mathbf{w}|}{|\mathbf{w}^T S_W \mathbf{w}|}, \tag{6}$$

but this time, it's a vector that provides the transformation. If S_W is non-singular, $\mathbf{w_{opt}}$ can be computed by finding the eigenvectors of the Equation 4. However, since $S_B \mathbf{w_{opt}}$ is always in the direction of $\mathbf{m_1} - \mathbf{m_2}$, it is not required to compute the eigenvectors of $S_W^{-1} S_B$ [1]. Instead, $\mathbf{w_{opt}}$ can be obtained using the equation below:

$$\mathbf{w_{opt}} = S_W^{-1}(\mathbf{m_1} - \mathbf{m_2}). \tag{7}$$

2.4. Face recognition using two-class LDA

In two-class LDA, one of the classes contains face image samples of an individual, whereas the other class contains face image samples of the other individuals in the database. To perform face recognition using two-class LDA, at first the dimensionality of the input image should be reduced to $n - c$ to avoid the singularity of S_W. Afterwards $\mathbf{w_{opt}}$ can be computed for each individual, then $\mathbf{w_{opt}}$ projects the $n - c$ dimensional input feature vector to the one-dimensional decision space. When a test image arrives, it is first represented in a low dimensional form, and then it is projected to each individual's decision space, using the corresponding projection vector, $\mathbf{w_{opt,i}}$, computed in the training stage (here, the subscript i denotes the i-th class). Classification can be done by calculating the distance between the value of the test input and the mean values of the two classes for each projection. Three cases can be observed in the classification step. In the first case, at only one projected decision space, the value of the test input is closer to the individual's class than to the global class (the class that contains face image samples of the other individuals in the database). In this case we can simply choose this class as the identity of the individual. In the second case, the value of the test input is closer to an individual's class in more than one decision space. In this case, we choose the candidate that has the highest confidence score. We define the confidence score as the ratio of the distance between the value of the test input and the global class mean to the distance between the value of the test input and the individual's class mean

$$conf(i) = \frac{|m_{i,2} - x_k|}{|m_{i,1} - x_k|}. \qquad (8)$$

In the third case, at none of the decision spaces the value of the test input is closer to the individual's class. If the conducted experiment is close-set (if only the individuals in the database are tested for identification), this case implies the detection of an individual's face image which contains a different variation in appearance apart from the ones learned from the training set. In this case, again using the proposed confidence score, the individual that has the highest confidence score is chosen as the identity of the individual. If the conducted experiment is open-set (if the individuals who are not in the database can also be tested), then this case may also imply the detection of an unknown individual.

3. Experiments

Two experiments are conducted to observe the performances of two-class LDA and multi-class LDA comparatively. Eigenfaces algorithm is also implemented as a baseline system. In the Eigenfaces approach, nearest neighbor classifier is used with the L1 norm as a distance metric, since we have observed that it produces better results than the L2 norm and cosine angle metrics.

3.1. Experiments on the Yale database

The Yale database [6] contains 165 face images of 15 individuals, where each individual has 11 face samples. Five of these are used for training. The remaining six images are used for testing. The face images are closely cropped and scaled to 64x64 pixels resolution. Sample images are shown in Fig. 1.

Fig. 2 illustrates the performances of the tested approaches with respect to varying dimensionality. In the case of PCA, dimension is the dimension of the feature vector used in classification, whereas in the case of multi-class LDA and two-class LDA, it indicates the dimension of the representation vector obtained by performing PCA on the input image before the LDA step. The actual dimension of the feature vector used in classification is 14 in multi-class LDA and one in two-class LDA. From Fig. 2, it can be observed that both types of LDAs outperform the Eigenfaces approach significantly. Two-class LDA performs slightly better than the multi-class LDA, where there is only 2.22% performance difference between the best classification scores of these two algorithms. PCA has a stable recognition rate with respect to dimension, on the other hand multi-class LDA's performance increases steadily with the increased dimensionality. Two-class LDA also performs better at high dimensions than at the lower dimensions. This observation about LDA approaches indicate that at low dimensions both types of LDA are effected by the prior PCA step where the most of the principal components are discarded which can be beneficial at the classification step.

Figure 1. Samples from the Yale database. First row: Samples from training set. Second row: Samples from test set.

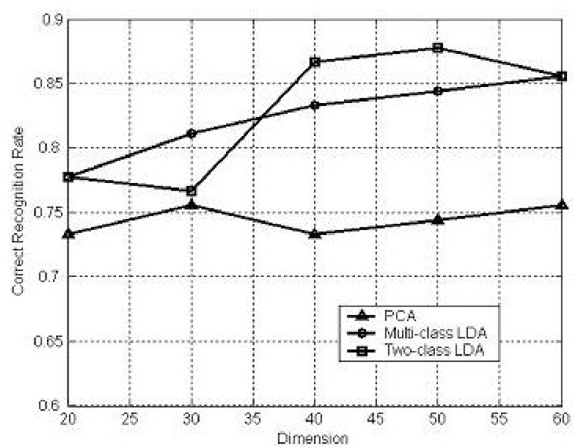

Figure 2. Correct recognition rate versus dimensionality plot -Yale database.

3.2. Experiments on the CMU PIE database

The face database is built with the samples chosen from the illumination and lights data sets of the CMU PIE database [7]. It contains 1360 face images of 68 individuals. Each individual has 20 face samples. 10 of these face samples are used for training and the remaining 10 face samples are used for testing. The face images are aligned and scaled to 64x64 pixels resolution. In Fig. 3, sample images from training and testing are shown.

Figure 3. Samples from the CMU PIE database. First row: Samples from training set. Second row: Samples from test set.

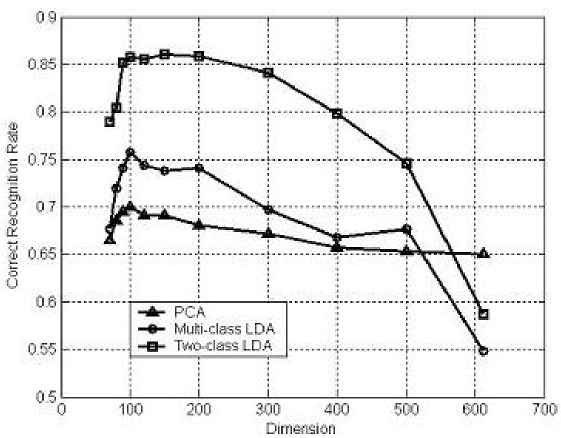

Figure 4. Correct recognition rate versus dimensionality plot -CMU PIE database.

The performance curves of the approaches with respect to varying dimensions are depicted in Fig. 4. The dimension is the dimension of the feature vector used in classification in PCA, and the reduced dimension obtained by prior PCA step in LDA approaches. The actual dimension of the feature vector used in classification is 67 in multi-class LDA and one in two-class LDA. As observed from the experiments on the Yale database, again, both types of LDA approaches outperform Eigenfaces approach significantly as long as the dimension is not very high. The performance of both of the LDA approaches increases upto a certain dimension and than drops steadily. This indicates that at very high dimensions (> 200), both of the LDA approaches suffer from "curse of dimensionality" where 10 samples for training do not suffice to estimate the best projection directions for classification. This is a different observation from the experiments on the Yale database where the dimension is limited upto a small number (i.e. 60 in the case of multi-LDA) due to limited number of training samples. The lower performance scores at lower dimensions are again due to the PCA bound imposed by the prior PCA step in the LDA approaches. As can be seen from Fig. 4, two-class LDA outperforms multi-class LDA significantly at every dimension, having a 10.29% absolute performance increase in terms of obtained best classification scores. The increase in absolute performance difference between the two approaches indicates the two-class LDA's success in handling high number of classes.

4. Conclusions

We presented a novel face recognition approach based on two-class LDA, where the traditional M-class LDA approach is converted to M 2-class classification problems. This formulation provides the benefits of more discrimination between classes, simpler computation of projection vectors and easier database update.

We tested the proposed algorithm on two separate databases, one containing a low number of classes with small amount of data that mainly consists of samples with different facial expressions and one containing higher number of classes with relatively large amount of data that mainly consists of face images under various illumination conditions. In both of the experiments two-class LDA approach performs better than the multi-class LDA. The performance difference becomes significant when the number of classes increases. This shows that two-class LDA has better discrimination capability than the multi-class LDA, especially when the number of classes is high.

References

[1] R. O. Duda and P. E. Hart. *Pattern Classification and Scene Analysis*. John Wiley Sons, 1973.

[2] K. Fukunaga. *Introduction to Statistical Pattern Recognition*. Academic Press, 1990.

[3] K. P. Juwei Lu and A. Venetsanopoulos. Face recognition using lda based algorithms. *IEEE Trans. on Neural Networks*, 14(1), 2003.

[4] D. Loog and Haeb-Umbach. Multiclass linear dimension reduction by weighted pairwise fisher criteria. *IEEE Trans. on PAMI*, 23(7), 2001.

[5] A. M. Martinez and A. C. Kak. Pca versus lda. *IEEE Trans. on PAMI*, 23(2), 2001.

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

[7] T. Sim, S. Baker, and M. Bsat. The CMU pose, illumination, and expression (PIE) database. In *Proc. of Intl. Conf. on Automatic Face and Gesture Recognition*, 2002.

[8] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Science*, pages 71–86, 1991.

[9] H. Yu and J. Yang. A direct lda algorithm for high dimensional data with application to face recognition. *Pattern Recognition*, 34, 2001.