Study on Combining PCA and LDA for Face Recognition

Burra Venkata Krishna Karthik

Electrical Engineering

University of Missouri-Kansas City, MO-64110

vbcr4@mail.umkc.edu

Declaration- I Mr. Burra Venkata Krishna Karthik, hereby declare that the paper titled 'Study on Combining PCA and LDA for Face Recognition' submitted by me is based on actual and original work carried out by me. Any reference to material obtained from other sources have been duly cited and referenced. I further certify that the paper has not been published or submitted for publication anywhere else nor it will be send for publication in the future.

Abstract

Current two-dimensional face recognition approaches can obtain a good performance only under constrained environments. However, in the real applications, face appearance changes significantly due to different illumination, pose, and expression. Face recognizers based on different representations of the input face images have different sensitivity to these variations. Therefore, a combination of different face classifiers which can integrate the complementary information should lead to improved classification accuracy. We use the minimum, maximum average rule, multi fusion strategies to combine two commonly used face classifiers based on Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) representations. Experiments conducted on a face database provided by AT&T database containing a set of 40 subjects (400 face images) show that the proposed classifier combination approaches outperform individual classifiers.

Introduction

Face recognition relates image processing, pattern recognition, computer vision, statistical learning and many other important disciplines. It is also in great need and necessary of national security and public safety. Many methods have been proposed in the past decades, and some of them have been successfully applied to face recognition. In global feature approaches, face image is treated as a whole matrix, using algebraic or statistical methods such as PCA to extract global features. After feature extraction, a classifier is designed to classify these features into classes of face. However, each algorithm shows strengths and drawbacks in different situations. Various methods are proposed to further increase the recognition accuracy. An effective approach is to combine multiple classification outputs from a set of classifiers. In this paper, we study multi-classifier approaches for face recognition based on combining PCA and LDA at score level (scores of PCA and LDA should be combined using minimum, maximum, average and multi fusion strategies. By fusing the results of multi-classifiers, we expect to get a final decision with better accuracy and robustness over a single classifier. Data set is provided by AT&T database containing a set of 400 face images. There are ten different images of each of 40 distinct subjects. For some subjects, the images were taken at separate times, varying the lighting, facial expression (open/closed eyes, smiling/not smiling) and facial details (glasses/no glasses).

In this paper, in section 2, the methods used for this experiment. In section 3, the results obtained are shown. Primarily discuss about the results of receiver operation characteristics (ROC) ([2], page 388). In section 4 the conclusions are drawn on the results obtained. In section 5 the references are mentioned. Section 6 contains appendix of all the tables, figures and formulae which were used in this project.

Method

There are many classifiers which can implement face recognition, for this project we have used Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA). PCA and LDA are feature extraction techniques which try to map the data of d dimension space in k ($k<d$) dimensions with minimal loss of information. ([1], page 116).

PCA is one of the common methods used for Face recognition. PCA tries to find a linear projection which tries to maximize variance. In PCA initially the data of D dimension is projected onto the subspace having k dimensions, this projection is called Eigen space. As PCA reduces the dimensions efficiently It is generally used as a pre-processing tool for other classifiers. Initially to create the Eigen sub space we take our data and find co- variance of the centered data and from that we find Eigen values and Eigen vectors of that co-variance matrix. Then we take a look at Eigen values we select only few Eigen vectors which correspond to the highest Eigen values. We select only a few Eigen Vectors because not all Eigen Values try to maximize the variance ([1], page no 124). Then project our training data and testing data onto this reduced Eigen Subspace, by doing this we are reducing the dimensions of a d x N data to k x N (where k is number of Eigen values selected and N is number of images). Then we calculate Euclidean distances between the projected training and test data and get scores for the PCA classifier. For these scores, we create appropriate targets and from that we get training ROC.

Linear Discriminant Analysis is also a feature extraction method and like PCA, it tries to reduce the dimensions. Fisher Linear Discriminant Analysis tries to find a linear projection which maximize the between class scatter and minimize the with-in class scatter. This ratio of between class scatter and with-in class scatter is called Fisherratio. So, in general LDA tries to maximize the Fisher ratio. ([1], page no 141). So, given a data we try to find within scatter matrices and between matrices. When the data set has huge dimensions, it becomes computationally expensive to use LDA. So, we use PCA pre- processing and then project the pre-processed data onto the Fisher subspace. Then in similar fashion to PCA, we find the scores of the classifier.

Score level fusion: In score level fusion, the match outputs from multiple biometrics are combined to improve the matching performance in face recognition ([3], page no 161). The fusion of this level is the most popular approach in the biometric literature due to its straightforward process of score collection and it is also practical to be applied in multibiometric system. Moreover, the matching scores contain sufficient information to make authentic and imposter case distinguishable ([3], page no 161). However, there are some factors that can affect the combination process hence degrades the biometric performance. For example, the matching scores generated by the individual matchers may not be homogenous due to be in the different scale/range or in different probability distribution. This scheme can be applied using various techniques such as sum/average rule, min rule and max rule techniques ([4]). In the classifier-based scheme, the scores from multiple matchers are treated as a feature vector and a classifier is constructed to discriminate authentic and imposter score ([3], page no 162).

For the Experiment A, the simplest way to combine them is by voting, which corresponds to taking linear combinations of learners, so we take scores from each classifier (PCA and LDA) and use fusion techniques like average, max and min rules and calculate ROC for validation set ([1], Page no 493). For using these fusion methods, we need to make sure that the scores of all classifiers must be in the same range. We Normalize to make sure the score of both PCA and LDA remain in the range of 0 to 1. So, by this we made all the scores to be in the same range. For the Experiment B, we perform Multi-instance based system at score level (for both PCA and LDA). Here, the score belonging to each test sample should be average of the scores comparing the test sample with all the templates of a particular identity.

Results and Discussions

For implementing Face recognition in this project, we are using PCA and LDA as classifiers. For PCA the only parameter we can change to get better results is selecting the number of principal components. We observed that as the number of components increase AUC's of ROC curve decreased. This happens because the first few principal components will have much of the variance and the rest components are noisy or contain very less variance ([1], page124). A better result was obtained when 10 Principal components were used instead of 200. The best result training ROC curve AUC was 0.78659 (Refer figure 1).

For LDA, as we can have two parameters to change, one is choosing number of Eigen vectors of PCA and the other is choosing the number of components of LDA. So, we can change those parameters to get better result. As we decrease the number of PCA components, we might get better result but we shouldn't make it too low because we won't have much data to project it onto to the Fisher Subspace, so an optimum number of principal components must be chosen we used 160 Principal components. For LDA we used first 40 components of LDA. We got a better AUC using 160 components for PCA and 40 for LDA. We observed a training curve with an AUC of 0.9796 (Refer Figure 1).

Till now we discussed about each classifier results independently and from the results we can observe that LDA had better performance since it directly deals with class discrimination while PCA does not pay attention to the underlying class structure. No Free Lunch Theorem states that in any domain there is no single learning algorithm which can induce the most accurate learner, that's why we must use combinations of multiple classifiers and chose the one combination which performs the best on a separate validation set ([1], Page no 487). The simplest way to combine them is by voting, which corresponds to taking linear combinations of learners, so we take scores from each classifier and use fusion techniques like average, max and min rules and calculate ROC for validation set ([1], Page no 493). For using these fusion methods, we need to make sure that the scores of all classifiers must be in the same range. We Normalize to make sure the score of both PCA and LDA remain in the range of 0 to 1. So, by this we made all the scores to be in the same range.

Average rule of fusion is nothing but we just add the scores of classifiers combination we are trying to implement. This rule of fusion is commonly used when using linear combinations. For this rule, multiple combinations of classifiers were used and ROC's were plotted for validation set and we observe that the best AUC was for the combination of LDA and PCA where AUC is found out to be 0.94397 (Refer Table 1, Figure 1).

Max rule of fusion is taking max of each score from each classifier. Max rule gives us the most optimistic rule ([1], page 493). For these combinations of classifiers, we observe a better result of combination for PCA+LDA and AUC is 0.96992 (Refer Table 1, Figure 1). Min rule of fusion is the most pessimistic rule ([1], page 493) and we observe the worst result for the combination of LDA and PCA where AUC is found out to be 0.77648(Refer table 1, Figure 1).

AUC for PCA (without Multi-instance fusion) is found out to be 0.9224. AUC for LDA (without Multi-instance fusion) is found out to be 0.7881. AUC for PCA (with Multi-instance fusion) is found out to be 0.9876. Here, the score belonging to each test sample is average of the scores comparing the test sample with all the templates of a identity. Similarly, AUC for LDA (with Multi-instance fusion) is found to be 0.88761 where the score belonging to each test sample is average of the scores comparing the test sample with all the templates of a identity. So, we see that the performance of PCA and LDA increases after performing multi-instance fusion and comparing these two, we can see that the performance of PCA was better than LDA after performing multi instance fusion.

Conclusions

In this paper, we implemented Face recognition system using multiple learners, initially we started comparing the results of PCA and LDA classifier individually and then we used simple fusion techniques to get a better validation result. We observed that the LDA was the strongest classifier comparing to PCA. As the results, we get are data dependent we must use fusion technique to combine two classifiers. We observed a best validation result for the maximum rule with LDA and PCA combination. So, we can conclude that as LDA for this classifier is strong and combining it with PCA increased the AUC of that combination. But, in case of Experiment B, we use multi instance fusion for both PCA and LDA. We, observe that PCA with multi instance fusion gave a better result than LDA.

This data set is limited, it has only 40 subjects. Generally, in real time applications there will be thousands or millions of subjects. So, the AUC's we reported, might decrease if the subjects of data set are increased and in this data set we had 10 images of each subject but in real time applications we will have less number of images for each subject. We used simple voting techniques for combining multiple classifiers we can increase the AUC further by using techniques such as boosting or bagging. ([1], 498). We only implemented face recognition system, but in real time we should also implement Face detection. We can also make the system robust and increase the training data set images by taking the mirror of an image and then adding some noise to it. Adding a bit noise makes it more robust and that robustness is helpful in real time applications.

References

1) Ethem Alpaydin, Introduction to machine Learning, 3rd edition, MIT press, 2014
2) J. C. Príncipe, et al., Neural and adaptive systems: fundamentals through simulations. New York: Wiley, 1999.
3) Jaafar, Haryati, and Dzati Athiar Ramli. "A review of multibiometric system with fusion strategies and weighting factor." *International Journal of Computer Science Engineering (IJCSE)* 2.4 (2013): 158-165.
4) M. Parviz, and M.S. Moin, "Boosting Approach for Score Level Fusion In Multimodal Biometrics Based On AUC Maximization", Journal of Information Hiding and Multimedia Signal Processing, Vol. 2, No.1, 2011, pp. 51-60.
5) Matthew Turk and Alex Pentland, "Eigen Faces for Recognition", Journal of Cognitive Neuroscience Volume 3, Number 1, 1991.

Appendices

Table 1- Performance Metrics for PCA, LDA and Multi Classifier system

Rule	Classifiers	EER	AUC	Decidability
Average	LDA+PCA	0.0355064	0.94397	3.244
Max	LDA+PCA	0.096538	0.96992	2.722
Min	LDA+PCA	0.032128	0.77648	0.9843
PCA only	PCA	0.015462	0.78659	2.433
LDA only	LDA	0.029122	0.97969	1.224

Table 2- Performance Metrics for PCA and LDA with and without Multi instance fusion

Rule	Classifiers	EER	AUC	Decidability
LDA with Multi instance fusion	LDA	0.024162	0.88761	1.2246
PCA with Multi instance fusion	PCA	0.013722	0.98767	3.6831
LDA without Multi instance fusion	LDA	0.015765	0.7881	0.9986
PCA without multi instance fusion	PCA	0.035776	0.9224	3.221

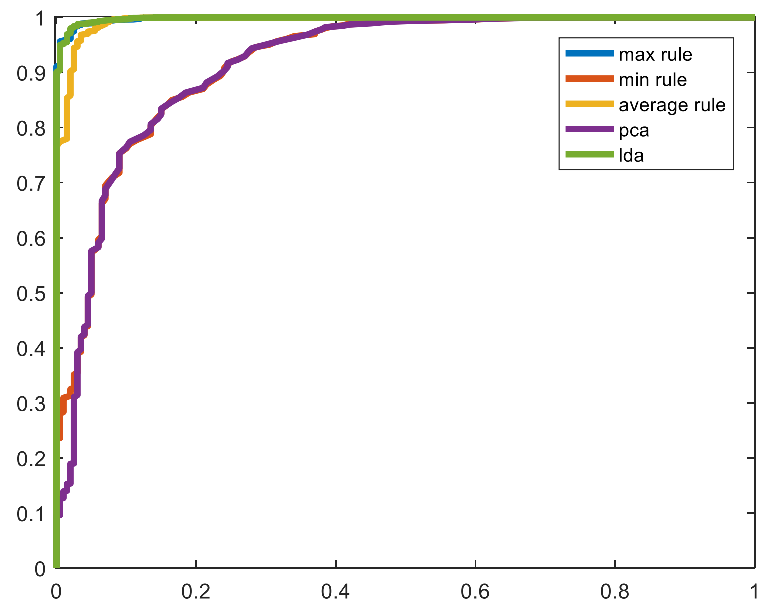

Fig1: 5 ROC curves for performance of PCA, LDA and MSC rules

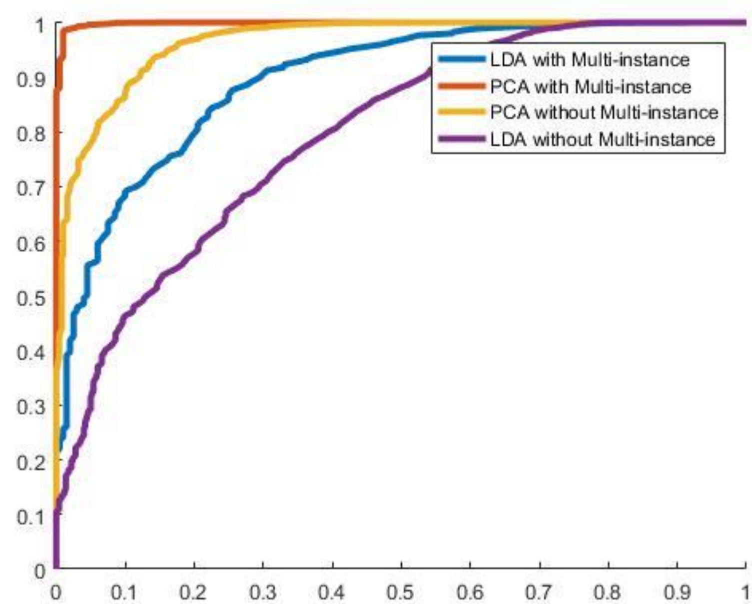

Fig2: 4 ROC curves for performance of PCA, LDA with and without multifusion