Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection

Niraj Mahajan, Samarth Singh, Shaan Ul Haque

180050069, 180050090, 180070053

Course: CS663

Instructors: Prof. Ajit Rajwade & Prof. Suyash Awate

Abstract - In this implementation, we studied and implemented the Fischer Faces algorithm [1] which uses the technique of Linear Discriminant Analysis. We used several datasets and performed various face an feature recognition experiments to compare the performance of both these methods under varying image acquisition parameters like lighting, pose, expression.

1 Introduction

Algorithms such as eigen faces, correlation and linear subspaces take advantage of the fact that under admittedly idealized conditions,the variation within class lies in a linear subspace of the image space. Hence, the classes are convex and therefore linearly separable. One can perform dimensionality reduction using linear projection and still preserve linear separability. This is a strong argument in favor of using linear methods for dimensionality reduction in the face recognition problem, at least when one seeks insensitivity to lighting conditions.

Since the learning set is labeled, it makes sense to use this information to build a more reliable method for reducing the dimensionality of the feature space. Here we argue that using class specific linear methods for dimensionality reduction and simple classifiers in the reduced feature space, one gets better recognition rates than with either the Linear Subspace method or the Eigenface method. Fisher's Linear Discriminant (FLD) is an example of a *class specific method*, in the sense that it tries to "shape" the scatter in order to make it more reliable for classification. This method selects transformation matrix in in such a way that the ratio of the between-class scatter and the within-class scatter is maximized.

Let the between-class scatter matrix be defined as

$$S_B = \sum_{i=1}^{c} N_i (\mu - \mu_i)(\mu - \mu_i)^T$$

and the within-class scatter matrix be defined as

$$S_W = \sum_{i=1}^{c} \sum_{x_k \epsilon X_i} (x_k - \mu_i)(x_k - \mu_i)^T$$

where μ_i is the mean image of class X_i, and N_i is the number of samples in class X_i. If S_W is nonsingular, the optimal projection W_{opt} is chosen as the matrix with orthonormal columns which maximizes the ratio of the determinant of the between-class scatter matrix of the projected samples to the determinant of the
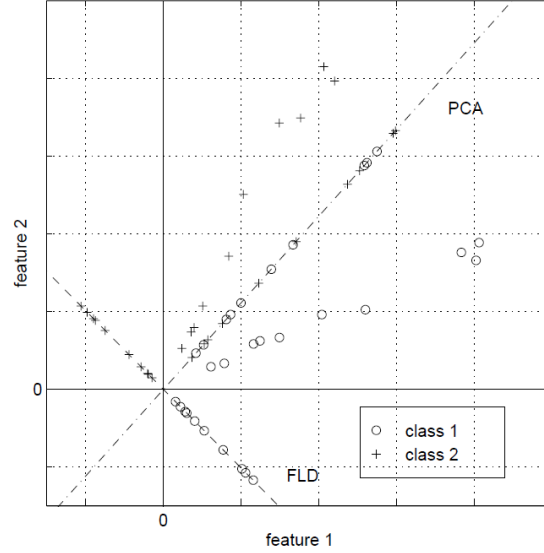
1

within-class scatter matrix of the projected samples, i.e.

$$W_{opt} = \underset{W}{\operatorname{argmax}} \frac{|W^T S_B W|}{|W^T S_W W|}$$
$$= [w_1 w_2 \ldots w_m]$$

where $\{w_i \mid i = 1, 2, \ldots m\}$ is the set of generalized eigenvectors of S_B and S_W corresponding to the m largest generalized eigenvalues $\{\lambda_i \mid i = 1, 2, \ldots m\}$, i.e.

$$S_B w_i = \lambda_i S_w w_i$$

To illustrate the advantage of Fischer Faces, we have provided with a figure taken from the paper.

Figure 1: A comparison of principal component analysis (PCA) and Fisher's linear discriminant (FLD) for a two class problem where data for each class lies near a linear subspace.

In the face recognition problem one is confronted with the difficulty that the within class scatter matrix $S_W \in R^{n \times n}$ is always singular. This stems from the fact that the rank of S_W is at most N-c, and in general, the number of images in the learning set N is much smaller than the number of pixels in each image n. This means that it is possible to choose the matrix W such that the within-class scatter of the projected samples can be made exactly zero.
In order to overcome the complication of a singular S_W, we propose an alternative to the criterion. This method, which we call Fisherfaces, avoids this problem by

2

projecting the image set to a lower dimensional space so that the resulting within-class scatter matrix S_W is nonsingular. This is achieved by using PCA to reduce the dimension of the feature space to N-c and then, applying the standard FLD reduce the dimension to c-1. More formally, W_{opt} is given by:

$$W_{opt}^T = W_{fld}^T W_{pca}^T$$

where

$$W_{pca} = \underset{W}{\text{argmax}} |W^T S_T W|$$

$$W_{fld} = \underset{W}{\text{argmax}} \frac{|W^T W_{pca}^T S_B W_{pca} W|}{|W^T W_{pca^T} S_W W_{pca} W|}$$

here, S_T is the Scatter Matrix given by:

$$S_T = \sum_{k=1}^{N} (x_k - \mu)(x_k - \mu)^T$$

Note that the optimization for W_{pca} is performed over $n \times (N - c)$ matrices with orthonormal columns, while the optimization for W_{fld} is performed over $(N-c) \times m$ matrices with orthonormal columns. In computing W_{pca} we have thrown away only the smallest c principal components.

2   Analysis of Experimental Results

In this section we performed face recognition and glasses recognition experiments on the CMU Face dataset[2], Yale dataset A, and Yale dataset B[4]. We performed a random 40-60 split for testing-training datasets and performed face recognition on all three datasets. We also performed a glasses recognition experiment on the YaleB dataset using the "leaving one out" algorithm[3].

2.1   Face Recognition

Method	Error Rate (in %)		
	CMU Face	YaleA	YaleB
Eigen Faces	0.91	16.67	45.85
Eigen Faces (w/o top 3 eigvectors)	1.82	13.33	20.24
Fischer Faces	38.64	1.67	5.87

Table 1: Error Rates of Face Recognition

2.1.1 Yale A Face Database

The yale A dataset contains 165 grayscale images in GIF format of 15 individuals.There are 11 images per subject, one per different facial expression or configuration: center-light, w/glasses, happy, left-light, w/no glasses, normal, right-light, sad, sleepy, surprised, and wink. We applied the fischer-face algorithm, reducing the dimensionality to 15 and the eigen-face algorithm, reducing the dimensionality to 30.
Since this dataset has some images with varying illumination, it is obvious that the fischer-face algorithm will outperform the eigen-face algorithm (table 1). Note that the error for the eigen-face algorithm is much higher when we do not remove the top 3 eigenvectors. This is because the face images follow a lambertian model and lighting variation can be represented in a rank 3 space. (the variation in the images due to lighting changes is much higher compared to pose and expression changes.)

2.1.2 Yale B Face Database

The Yale Face Database B contains 2415 images of 38 human subjects under varying 64 illumination conditions. We applied the fischer-face algorithm, reducing the dimensionality to 38 (number of classes) and the eigen-face algorithm, reducing the dimensionality to 50.
Since this dataset has images with just varying illuminations, the fischer-face algorithm will again outperform the eigen-face algorithm (table 1). Again, the error for the eigen-face algorithm is much higher when we do not remove the top 3 eigenvectors.

2.1.3 CMU Face Database

The CMU database consists of 20 classes each class containing 32, 128*120 .pgm format, pictures of a person in different orientation (i.e. left faced, right faced, straight and up), different expressions (i.e. angry, happy, neutral and sad) and with/without glasses. Pictures were not cropped and conatined a lot of background in them.
We divided the dataset into 70-30 ratio by randomly selecting 70% of each class for training and 30% for testing. We applied the fischer face (reducing the dimensionality to 19) and eigen face (reducing the dimensionality to 19) algorithm to this CMU Face dataset. Eigen Faces was far better than Fischer Faces (table 1). This unexpected result was due to two reasons:

- Change in Face Alignment - Changing Face Alignment breaks the assumption of high linear separability. Since Linear Discriminant Analysis ignores highly changing features of a face, changing alignment induces excessive

variation in face leading the algorithm to neglect whole face. Thus, we see a lot of error in face recognition when Fischer faces method is used.

- Non-cropped Images - Even when only frontal images are used results for Eigen Faces are better than Fischer Faces because Fischer Face neglects features of an image which vary a lot(such as eyes, mouth, face alignment, etc) and gives a lot of weightage to background due to its non variability and thus losing important information regarding the face.

2.2 Glasses Recognition

Method	Error Rate (in %) for YaleA
Eigen Faces	40.00
Eigen Faces (w/o top 3 eigvectors)	53.33
Fischer Faces	33.33

Table 2: Error Rates of Glasses Recognition

In this section we perform glasses recognition on the Yale A dataset.
The yale A dataset contains images of 15 individuals, each with and without glasses. Hence we can use these 30 images to train and test a fischerfaces algorithm on this data and compare the results with the eigenfaces algorithm

We applied the fischer-face algorithm, reducing the dimensionality to 1 and the eigen-face algorithm, reducing the dimensionality to 10.
Our task involves a binary classification over multiple people. Hence the fischerface algorithm is bound to outperform the eigenface algorithm (table 2), since the eigenface algorithm will give a high variance for different people. Since this dataset has no illumination or pose variation, removing the top3 eigenvectors will lead to loss of vital data, and this too will decrease the accuracy. Also, we notice that compared to the error rate in face recognition, the error rate for glasses recognition using the fischer method is high. This is because our dataset is really small (just 30 images). But still, the relative magnitudes of the accuracy are as per our expectation and have been explained.

3 Conclusion

We thus conclude that the Fisherface method appears to be the best at extrapolating and interpolating over variation in lighting.
Removing the largest three principal components does improve the performance of the Eigenface method in the presence of lighting variation, but harms the accuracy if lighting variations are not present. The Fisherface method appears to be the best at simultaneously handling variation in lighting and expression. Also, the fischerface method tries to ignore the features within the same class that have high variances.

4 Usage of Code

- Run the files **yaleA_tester.py**, **yaleB_tester.py**, **cmu_tester.py** (in cmu_codes direct)

- Download the data and place it in the datasets folder. Refer https://www.cse.iitb.ac.in/ nirajm/shared/663datalinks.txt for more details

References

[1] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman. "Eigenfaces vs. Fisherfaces: recognition using class specific linear projection". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19.7 (1997), pp. 711–720. DOI: 10.1109/34.598228.

[2] Dheeru Dua and Casey Graff. *CMU Face Images Dataset UCI Machine Learning Repository*. 2017. URL: http://archive.ics.uci.edu/ml.

[3] Richard O. Duda and Peter E. Hart. *Pattern classification and scene analysis / Richard O. Duda, Peter E. Hart*. English. Wiley New York, 1973, xvii, 482 pages : ISBN: 0471223611. URL: http://www.loc.gov/catdir/enhancements/fy0607/72007008-t.html.

[4] A.S. Georghiades, P.N. Belhumeur, and D.J. Kriegman. "From Few to Many: Illumination Cone Models for Face Recognition under Variable Lighting and Pose". In: *IEEE Trans. Pattern Anal. Mach. Intelligence* 23.6 (2001), pp. 643–660.