Review of "Eigenfaces vs. Fisherfaces: recognition using class specific linear projection" by P. N. Belhumeur, J. P. Hespanha and D. J. Kriegman

Jayasooryan C S (MDS202120) Krishna Gupta (MDS202122) Kumar Priyank (MDS202122) Meghna Mondal (MDS202123)

May 24, 2022

Abstract

The paper developed a face recognition algorithm which is insensitive to large variation in lighting direction and facial expression. Taking a pattern classification approach, considering each pixel in an image as a coordinate in a high-dimensional space. They take advantage of the observation that the images of a particular face, under varying illumination but fixed pose, lie in a 3D linear subspace of the high dimensional image space—if the face is a Lambertian surface without shadowing. However, since faces are not truly Lambertian surfaces and do indeed produce self-shadowing, images will deviate from this linear subspace. Rather than explicitly modeling this deviation, they linearly project the image into a subspace in a manner which discounts those regions of the face with large deviation. Their projection method is based on Fisher's Linear Discriminant and produces well separated classes in a low-dimensional subspace, even under severe variation in lighting and facial expressions. The Eigenface technique, another method based on linearly projecting the image space to a low dimensional subspace, has similar computational requirements.

We reviewed the part of the paper which talks about Correlation, Eigenfaces, and Fisherfaces methods. We were able to reproduce their results on "Variation in Facial Expression, Eye Wear, and Lighting" using Yale Face Databases.

1 Introduction

Within the last several years, numerous algorithms have been proposed for face recognition. While much progress has been made towards recognizing faces under small variations in lighting, facial expression and pose, reliable techniques for recognition under more extreme variations have proven elusive. In the paper a new approach for face recognition is outlined — one that is insensitive to large variations in lighting and facial expressions. Here, lighting variability includes not only intensity, but also direction and number of light sources. The problem statement under study is that given a set of face images labeled with the person's identity (the learning set) and an unlabeled set of face images from the same group of people (the test set), develop a face recognition algorithm which is insensitive to large variation in lighting direction and facial expression to identify each person in the test images. In the paper, the major focus is to solve the above problem using two techniques, the conventional Eigenfaces method and the relatively newer Fisherfaces method and compare results.

2 Main Results

Every image is converted into a matrix and then flattened as a vector. The problem as stated above, of classification of images, is approached within the pattern classification paradigm, considering each of the pixel values in a sample image as a coordinate in a high dimensional space (the image space). Three methods, namely Correlation Method, Eigenfaces method and Fisherfaces method, are considered to approach the problem and compare the results.

2.1 Correlation Method

Correlation Method is the simplest method. This method applies the nearest neighbor classifier in the image space. Given a matrix $X_{n\times k}$ of k images, each column vector represents an image in n-dimensional space (each pixel representing one dimension). When a new image comes, it can be classified as the image it is closest to. But since the variations between the images of the same face due to illumination and viewing direction are almost always larger than image variations due to change in face identity, therefore this method doesn't work well under varying light conditions. Also, the method is computationally expensive and requires large amounts of storage.

2.2 Eigenfaces Method

The eigenfaces method aims to achieve efficieny in computation and storage by doing dimensionality reduction. It achieves dimensionality reduction using Principal Component Analysis (PCA). PCA tries to find the lower-dimensional surface to project the high-dimensional data while explaining as much of the cumulative variance in the predictors (or variables) as possible. Consider a set of N sample images, $\{x_1, x_2, ..., x_N\}$. Assume that each image belongs to one of the c classes $\{X_1, X_2, ..., X_c\}$. Let us also consider a linear transformation mapping the original n-dimensional image space into an m-dimensional feature space, where m < n. Define new feature vectors $y_k \in R^m$ by the following linear transformation, $y_k = W^T x_k$ where $W \in R^{m\times n}$ is a matrix with orthonormal columns. If the total scatter matrix S_T is defined as : $S_T = \sum_{k=1}^{N}(x_k - \mu)(x_k - \mu)^T$ where n is the number of sample images, and $\mu \in \mathrm{R}^n$ is the mean image of all samples, then after applying the linear transformation W^T , the scatter of the transformed feature vectors $\{y_1, y_2, ...y_N\}$ is $W^T S_T W$. In PCA, the projection W_{opt} is chosen to maximize the determinant of the total scatter matrix of the projected samples, i.e.,

$$W_{opt} = \arg\max_{w} |W^T S_T W|$$

$$= [w_1, w_2, ..., w_m]$$

where, $\{w_i | i = 1, 2, ...m\}$ is the m-dimension eigen vectors of S_T corresponding to the m largest eigenvalues. If classification is done using a nearest neighbor classifier in the reduced feature space, and m is chosen to be the number of images N in the training set, then the Eigenface method is equivalent to the correlation method in the previous section.

2.3 Fisherfaces Method

The previous algorithm takes 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. Here thay argue that using class specific linear methods for dimensionality reduction and simple classifiers in the reduced feature space, one may get better recognition rates than the

Eigenface method. Fisher's Linear Discriminant (FLD) is an example of a class specific method, that tries to "shape" the scatter in order to make it more reliable for classification. Here the between-class scatter matrix be defined as

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

And the within-class scatter matrix be defined as

$$S_W = \sum_{i=1}^{c} \sum_{x_k \in X_i} (x_i - \mu_i)(x_i - \mu_i)^T \tag{2}$$

$$\begin{aligned}\mu_i &= \text{mean image of class } X_i \\ N_i &= \text{number of images in the class } X_i\end{aligned}$$

If S_w is non-singular, the optimal projection W_{opt} is given by

$$W_{opt} = arg\ \underset{W}{max} \frac{|W^T S_B W|}{|W^T S_W W|} \tag{3}$$

According to Rayleigh-Ritz quotient method, the optimization problem in Eq. (3) can be restated as:

$$\begin{aligned}&\underset{w}{max} && w^T S_B w \\ &subject\ to && w^T S_W w = 1\end{aligned}$$

The Lagrangian is:

$$L = w^T S_B w - \lambda(w^T S_W w = 1)$$

where λ is the Lagrange multiplier.

Equating the derivative of L to zero gives:

$$\begin{aligned}&\frac{\partial L}{\partial w} = 2S_B w - \lambda S_W w \underset{set}{=} 0 \\ &\Rightarrow 2S_B w = 2\lambda S_W w \\ &\Rightarrow S_B w = \lambda S_W w\end{aligned}$$

which is a generalized eigenvalue problem (S_B, S_W).

Hence,

$$\begin{aligned}W_{opt} &= arg\ \underset{W}{max} \frac{|W^T S_B W|}{|W^T S_W W|} \\ &= [w_1\ w_2\ ...\ w_m]\end{aligned}$$

here w_i is the set of generalized eigenvectors of S_B and S_W corresponding to the m largest generalized eigenvalues λ_i. There are at most $c-1$ nonzero generalized eigenvalues, so an upper bound on m is $c-1$, where c is the number of classes.

In the face recognition problems, rank of S_W is at most $N-c$. In general, the number of images in the learning set N is much smaller than the number of pixels in each image n.Hence, S_W is a $n \times n$ matrix become singular. In order to overcome the complication of a singular S_W , they proposed an alternative to the criterion in (3). This method, which we call Fisherfaces, avoids this problem by projecting the image set to a lower dimensional space,the resulting within-class scatter matrix S_W is non singular. This is achieved by using PCA to reduce the dimension of the feature space to $N-c$. Then applying the standard FLD to reduce the dimension to $c-1$. Formally,

$$W_{opt}^T = W_{FLD}^T W_{PCA}^T$$

where

$$W_{PCA} = arg \underset{W}{max}\ W^T S_T W$$
$$W_{opt} = \frac{|W^T W_{PCA}^T S_B W_{PCA} W|}{|W^T W_{PCA}^T S_W W_{PCA} W|}$$

3 Experimental Results

In this section we present and discuss a comparison of each of the previously mentioned feature extraction techniques with SVM classifiers.For this a database from the Harvard Robotics Laboratory was used in which lighting has been systematically varied. Secondly, a database at Yale was constructed that includes variation in both facial expression and lighting.

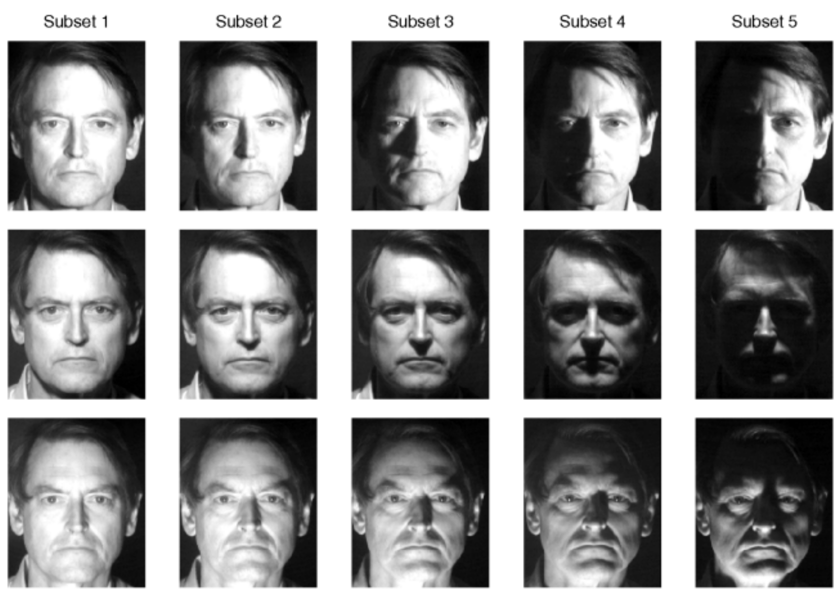

Figure 1: Example images from each subset of the Harvard Database used to test the four algorithms

- Variation in Lighting

  On each image in the Harvard DB(Fig.1) a subject's head is held steady while being illuminated by a dominant light source. The space of light source directions, which can be parameterized by spherical angles, was then sampled in 15° increments. Two experiments were performed on the Harvard Database:

  Extrapolation: When each of the methods is trained on images with near frontal illumination (Subset 1), the corresponding graph(Fig.2) shows the relative performance under extreme light source conditions.

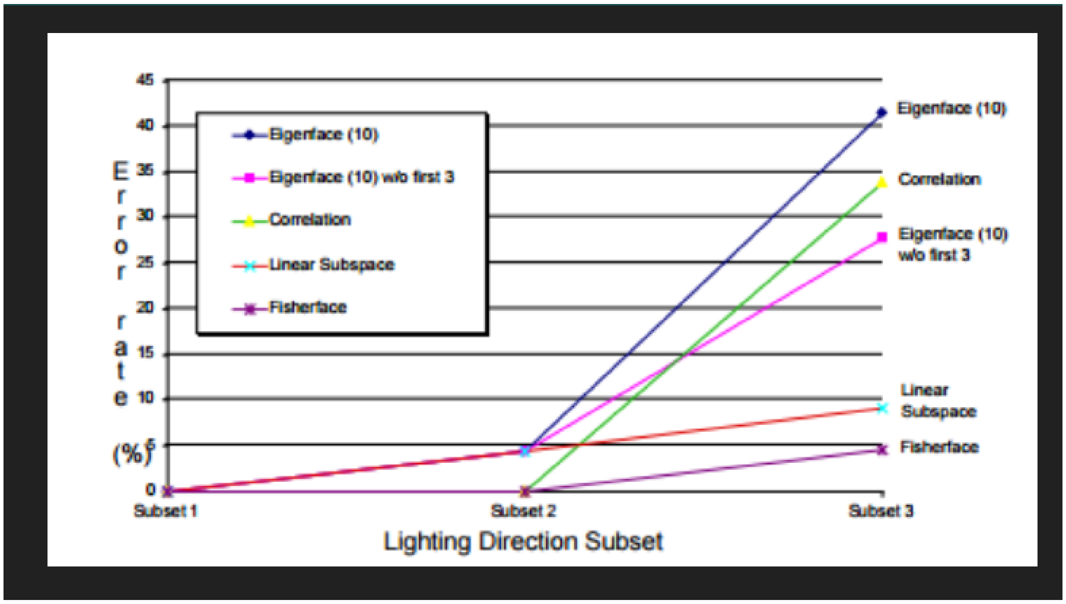

Figure 2: Each method was trained on samples from subset 1 and tested using samples from subsets 1,2,3

Interpolation: When each of the methods is trained on images from both near frontal and extreme lighting (Subsets 1 and 5), the corresponding graph(Fig.3) shows the relative performance under intermediate lighting conditions.

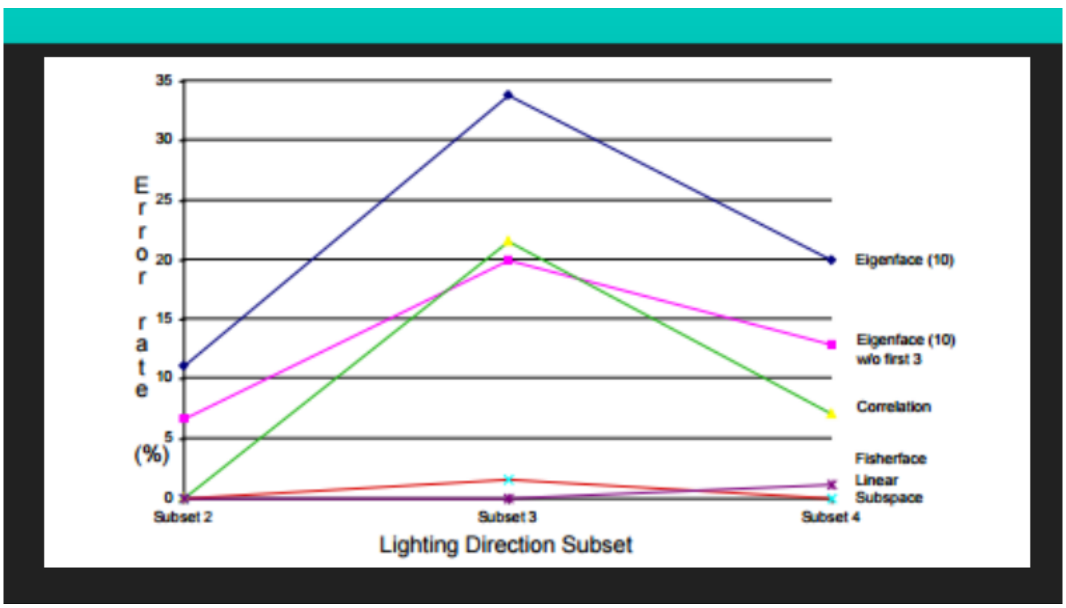

Figure 3: Each method was trained on subsets 1 & 5 and then tested on subsets 2,3,4.

These two experiments reveal a number of interesting points:

1) All of the algorithms perform perfectly when lighting is nearly frontal. However, as lighting is moved off axis, there is a significant performance difference between the two class-specific methods and the Eigenface method.

2) It has also been noted that the Eigenface method is equivalent to correlation when the number of Eigenfaces equals the size of the training set, and since performance increases with the dimension

of the eigenspace, the Eigenface method should do no better than correlation. This is empirically demonstrated as well.

3) In the Eigenface method, removing the first three principal components results in better performance under variable lighting conditions.

4) The Fisherface method had error rates lower than the Eigenface method and required less computation time.

- Variation in Facial Expression, Eye wear and lighting experiment

  Tests were designed to determine how the methods compared under a different range of conditions. For sixteen subjects, ten images were acquired during one session in front of a simple background. Subjects included females and males (some with facial hair), and some wore glasses. The images were manually centered and cropped to two different scales: The larger images included the full face and part of the background while the closely cropped ones included internal structures such as the brow, eyes, nose, mouth, and chin, but did not extend to the occluding contour.

  The corresponding graph(Fig.4) shows the relative performance of the algorithms when applied to the Yale Database which contains variation in facial expression and lighting.

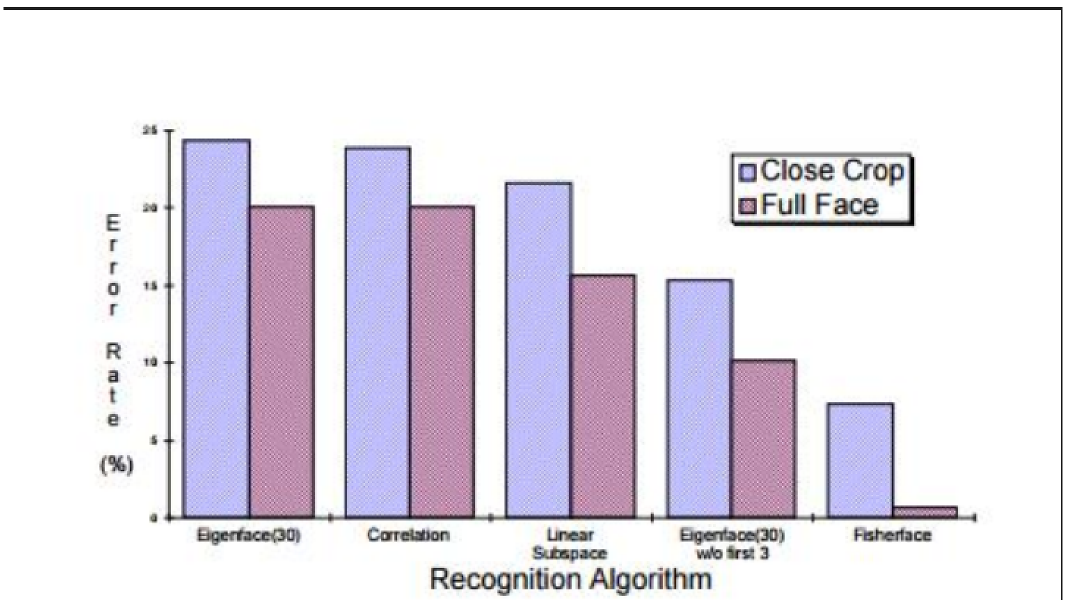

Figure 4: the relative performance of the algorithms

  All of the algorithms performed better on the images of the full face. Note that there is a dramatic improvement in the Fisherface method where the error rate was reduced from 7.3 percent to 0.6 percent. When the method is trained on the entire face, the pixels corresponding to the occluding contour of the face are chosen as good features for discriminating between individuals, i.e., the overall shape of the face is a powerful feature in face identification. As a practical note, however, it is expected that recognition rates would have been much lower for the full face images if the background or hair styles had varied and may even have been worse than the closely cropped images.

- Glasses Recognition

  In this experiment, the data set contained 36 images from a superset of the Yale Database, half with glasses. The recognition rates were obtained by cross validation, i.e., to classify the images of each person, all images of that person were removed from the database before the projection matrix W was computed.

PCA had recognition rates near chance, since, in most cases, it classified both images with and without glasses to the same class. On the other hand, the Fisherface methods can be viewed as deriving a template which is suited for finding glasses and ignoring other characteristics of the face. This conjecture is supported by observing the Fisherface corresponding to the projection matrix W

4 Results/Observations of Implementation

Following the theories and calculations explained above, we experimented with a certain set of images to infer how much the Correlation, Eigenfaces and Fischerfaces method is successful in classifying face-images .

Figure 5: Yales Dataset with variation in Facial Expression, Eyeware, Lighting

For the experimentation, we have taken up the Yales Face Database in which the variation in Facial Expression, Eye Wear and Lighting are systematically varying. The dataset contains a total of 165 images of 15 different persons (classes). Each of these persons have 11 distinct images which differ in either expression or lighting conditions. Each image is a grayscale image of dimension 243×320.

We plotted an sample image from the Face Database.

Now, we have flatten each image matrix to a column vector of dimension 77760×1 where each column vector represents an image. The mean image is then calculated by taking average of all the images which is then reshaped into its actual dimension. We plotted the mean image obtained here.

Now, we split the dataset into train and test set such that the train set has 150 images and the test set has 15 images respectively. Now, our goal is to check the accuracy of the different face recognition algorithms.

For that, first we have standardised the data to have mean zero and standard deviation one. Now, without any dimensionality reduction we performed the naïve Correlation method using K Neighbours Classifier and got an accuracy of 0.6 which means this algorithm has wrongly classified 6 images in the test set.

So, for the eigenfaces method we reduce the dimension to 30 (taken from the paper) and then use clustering using K Neighbours Classifier and got an accuracy of 0.8 which means this algorithm has wrongly classified 3 images in the test set.

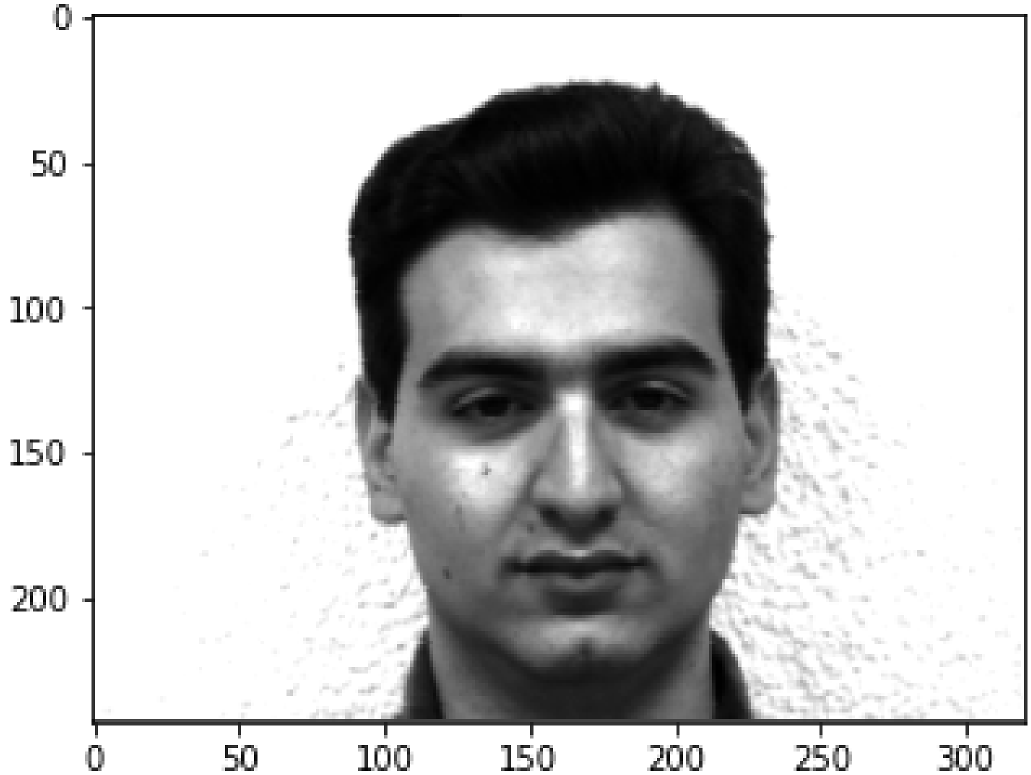

Figure 6: subject15.normal

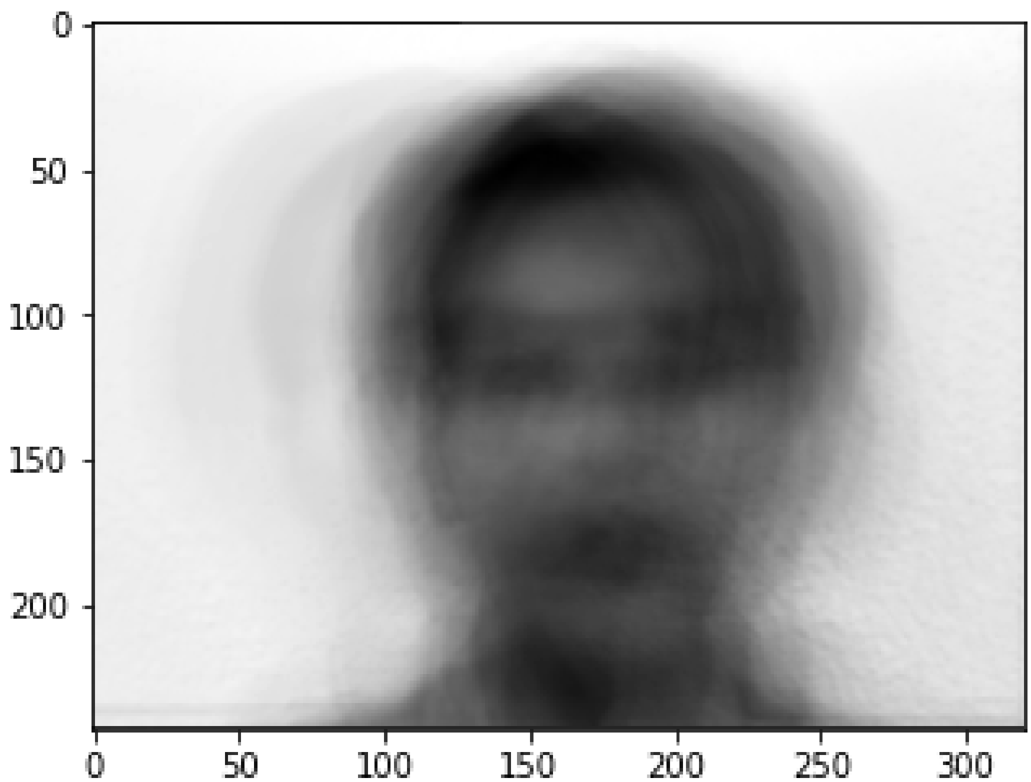

Figure 7: Mean Image

Finally, for the fischerfaces method, first we projected the images to dimension (150- 15) or 135 using PCA and then to dimension (15- 1) or 14 using LDA and then apply clustering on the reduced dimension space. So, here we get an 100% accuracy which means the Fischerface algorithm has correctly recognises all the test images.

5 Conclusion

- As all the images in the test set were similar to the train set images, all the methods have performed well in the test set.
- Removing the largest three principal components does improve the performance of the Eigenface method in the presence of lighting variation, as they correspond to some major features which makes difficulties for the model to classify the images.
- Out of the three methods we discussed Fischerfaces turns out to be the best method yielding optimum accuracy.

So, in this paper we get an idea to project an image to a lower dimensional space and classify them using three methods.

- Correlation - computationally heavy and not accurate for variation of lighting;
- Eigenfaces - uses PCA for dimensionality reduction;
- Fischerfaces - uses LDA for dimensionality reduction.

6 Current State of Matters

Out of the three methods discussed, Fischerfaces seems to be a perfect face recognition algorithm but this method can also fall out under extreme lighting conditions. These are actually some classical methods used previously. After that Scale Invariant Feature Transform, Speed Up Robust Features etc. were used for face recognition. But post 2012 the deep learning methods are mostly in use in this field.

7 Link to implementation

The link to the implementation is available at:Click here

8 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, vol. 19, no. 7, pp. 711-720, July 1997, doi: 10.1109/34.598228.

2. Ghojogh, B., Karray, F. Crowley, M., 2019. Eigenvalue and Generalized Eigenvalue Problems: Tutorial. ArXiv Preprint arXiv:1903.11240.

9 Work Contribution

Everyone contributed equally.