Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection

Jayasooryan C S   Krishna Gupta
Kumar Priyank   Meghna Mondal

Chennai Mathematical Institute

Paper Presentation
May 4, 2022

Introduction

- Within the last several years, numerous algorithms have been proposed for face recognition
- Problem Statement : 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), identify each person in the test images
- Objective : Develop a face recognition algorithm which is insensitive to large variation in lighting direction and facial expression

Light Variation

Figure: The same person seen under different lighting conditions can appear dramatically different: In the left image, the dominant light source is nearly head-on; in the right image, the dominant light source is from above and to the right.

Converting an Image to a Vector

- An image is an array, or a matrix, of square pixels (picture elements) arranged in columns and rows
- In a (8-bit) greyscale image each picture element has an assigned intensity that ranges from 0 to 255. Each pixel has a value from 0 (black) to 255 (white)
- In the analysis, every image is represented as a vector of length equal to number of pixels, which is obtained by flattening the matrix
- For example, if we have 150 images then it can be represented as a matrix with rows equal to the number of pixels and columns equal to 150, i.e each column will represent one image

Methods

- Problem Statement : 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), identify each person in the test images
- Methods :
  - Correlation
  - Eigenfaces
  - Fisherfaces
- The problem 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)

Correlation Method

- Simplest way of classification : Uses Nearest neighbor classifier in the image space

- If all of the images are normalized to have zero mean and unit variance, then this procedure is equivalent to choosing the image in the learning set that best correlates with the test image

Correlation Method : Disadvantages

- Simple but has many disadvantages
- Disadvantages
  - Does not work well under varying lighting conditions
  - Computationally expensive
  - Requires large amounts of storage

Eigenfaces Method

- Motivation : Achieve efficiency by doing Dimensionality Reduction
- A commonly used Dimensionality Reduction technique — particularly in face recognition - is 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

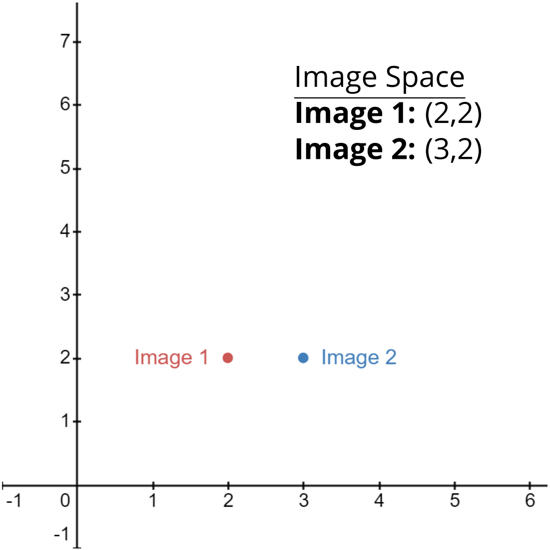
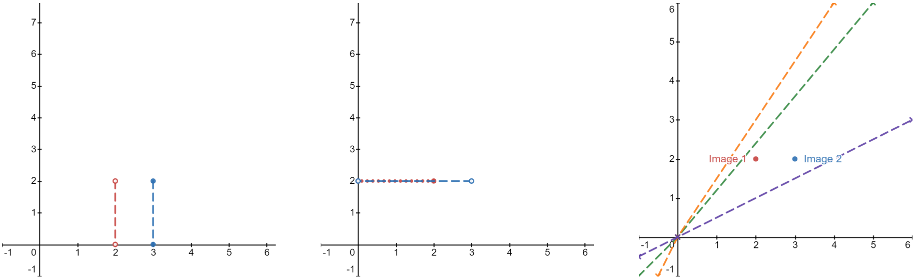


Image Space
Image 1: (2,2)
Image 2: (3,2)
Image 1
Image 2
-1
0
1
2
3
4
5
6
7

PCA : Intuition



- The two points can be projected on a 1D space in infinite ways
- The projection of the two points on X axis gives more variance than on Y axis
- PCA aims to find the lower-dimensional surface to project the high-dimensional data such that variance is maximised

Eigenfaces Method : Algorithm

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.

Eigenfaces Method : Algorithm

The new feature vectors

$$y_k \in R^m$$

are defined by the following linear transformation:

$$y_k = W^T x_k \qquad\qquad k = 1, 2, ...n$$

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 $m \in R^n$ 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 $W^T S_T W$.

Eigenfaces Method : Algorithm

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.

Fisherfaces: Motivation

- Face Recognition
- Naive Method: **Correlation**
  - Images are converted to vectors
  - Each image is represented by a point in *n*-dimensional image space.
  - An image in the test set is recognized (classified) by assigning to it the label of the closest point in the learning set.
- Disadvantages
  - It requires large amounts of storage to store *n*-dimensional images
  - Images space may not be tightly clustered under varying light conditions
  - Correlation is computationally expensive

Fisherfaces: Motivation

- Dimensionality reduction, $m \leq n$

$$\textit{n-dimensional} \xrightarrow[\text{Dimensionality}]{\text{reduction}} \textit{m-dimensional}$$

- Eigenface Method: Principal Component Analysis
  - Storage limitation is addressed
  - But the scatter being maximized is due not only to the between-class scatter, but also to the within-class scatter that, which make the clusters more loose

Dimenstionality Reduction

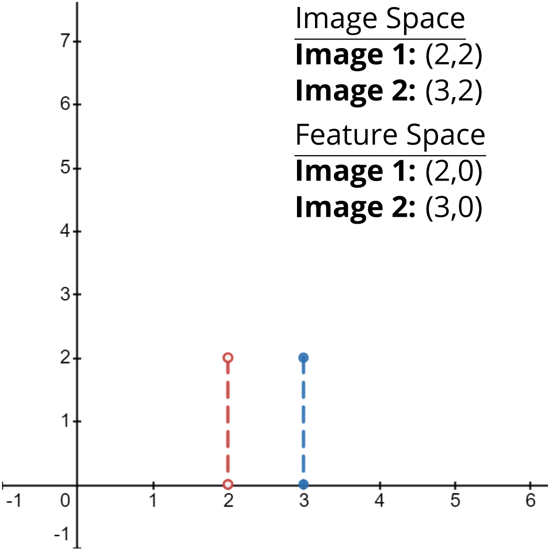


Image Space

Image 1: (2,2)

Image 2: (3,2)

Feature Space

Image 1: (2,0)

Image 2: (3,0)

- Projection Matrix

$$W_x = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$$

Dimenstionality Reduction

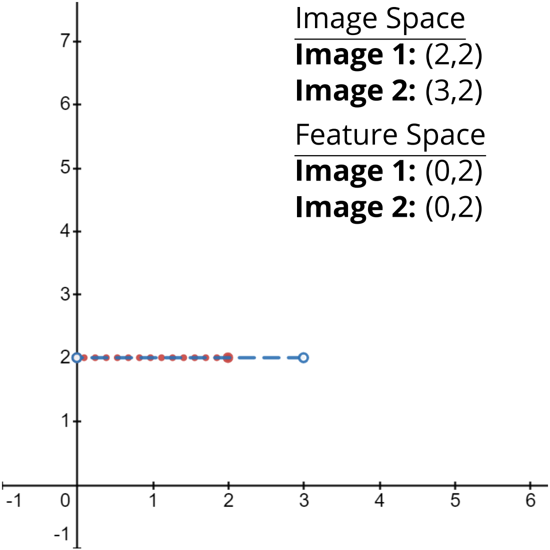


Image Space
Image 1: (2,2)
Image 2: (3,2)

Feature Space
Image 1: (0,2)
Image 2: (0,2)

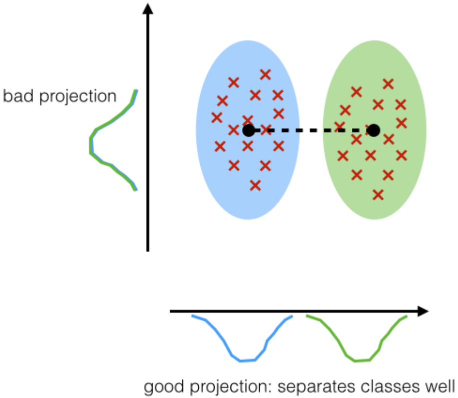
- Projection Matrix

$$W_y = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$$

Clusters

In face recognition we have two types of scattering:

1. within-class scatter; *minimize*
2. between-class scatter; *maximize*



Fisher Linear Discriminant

- The between-class scatter matrix be defined as

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

- The within-class scatter matrix be defined as

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

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

Fisher Linear Discriminant

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

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

Fisher Linear Discriminant

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

$$\max_{w} \quad w^T S_B w \tag{4}$$

$$\text{subject to} \quad w^T S_W w = 1 \tag{5}$$

The Lagrangian is:

$$L = w^T S_B w - \lambda(w^T S_W w = 1) \tag{6}$$

where λ is the Lagrange multiplier.

Fisher Linear Discriminant

Equating the derivative of L to zero gives:

$$\frac{\partial L}{\partial w} = 2S_B w - \lambda S_W w \underset{set}{=} 0 \tag{7}$$

$$\Rightarrow 2S_B w = 2\lambda S_W w \tag{8}$$

$$\Rightarrow S_B w = \lambda S_W w \tag{9}$$

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

Fisher Linear Discriminant

Hence,

$$W_{opt} = \arg\max_{W} \frac{|W^T S_B W|}{|W^T S_W W|} \tag{10}$$

$$= [w_1 \; w_2 \; \ldots \; w_m] \tag{11}$$

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.

Fisher Linear Discriminant

- In the face recognition problem, 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.
- S_W is a $n \times n$ matrix; singular.

Fisherfaces

- 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 $c - 1$.

Fisherfaces

More formally,

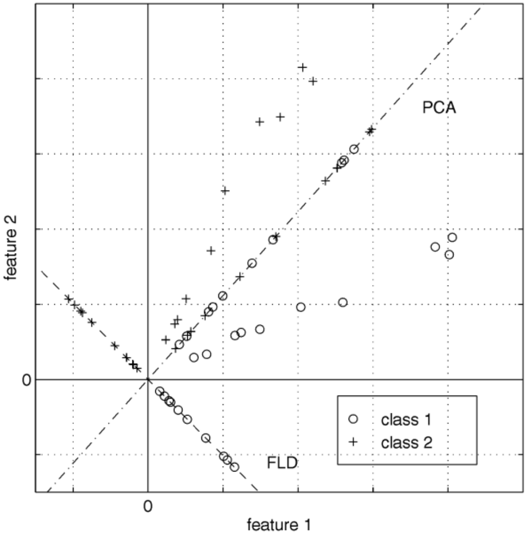
$$W_{opt}^T = W_{FLD}^T W_{PCA}^T$$

where

$$W_{PCA} = arg \max_W W^T S_T W$$

$$W_{opt} = \frac{|W^T W_{PCA}^T S_B W_{PCA} W|}{|W^T W_{PCA}^T S_W W_{PCA} W|}$$

PCA vs FLD



- 5 subsets consisting of images of 5 people with varying light were used for this experiment.
- Recognition was performed using nearest neighbor classifier.
    - Each method was trained on samples from subset 1 and tested using samples from subsets 1,2,3
    - Each method was trained on subsets 1 and 5 and then tested on subsets 2,3,4.

Experimental Results

Variation in Lighting

- **Subset 1** contains 30 images for which both the longitudinal and latitudinal angles of light source direction are within 15° of the camera axis, including the lighting direction coincident with the camera's optical axis.
- **Subset 2** contains 45 images for which the greater of the longitudinal and latitudinal angles of light source direction are 30° from the camera axis.
- **Subset 3** contains 65 images for which the greater of the longitudinal and latitudinal angles of light source direction are 45° from the camera axis.
- **Subset 4** contains 85 images for which the greater of the longitudinal and latitudinal angles of light source direction are — 60° from the camera axis.
- **Subset 5** contains 105 images for which the greater of the longitudinal and latitudinal angles of light source direction are 75° from the camera axis.

Figure

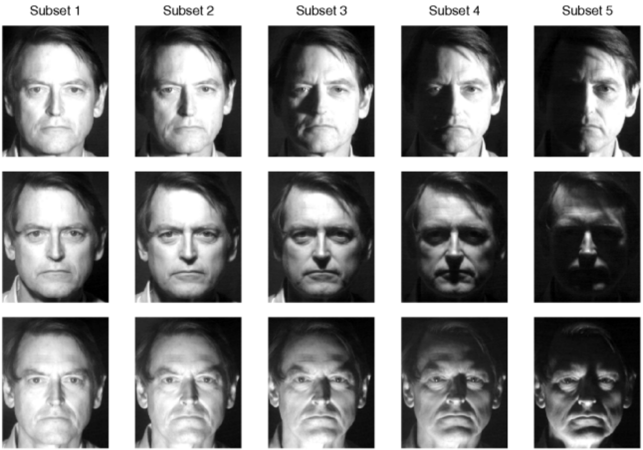


Figure: Example images from each subset of the Harvard Database used to test the four algorithms.

- Extrapolation: When each of the methods is trained on images with near frontal illumination (Subset 1), the corresponding table shows the relative performance under extreme light source conditions.
- Interpolation: When each of the methods is trained on images from both near frontal and extreme lighting (Subsets 1 and 5), the corresponding table shows the relative performance under intermediate lighting conditions.

Method	Reduced Space	Error Rate (%)		
		Subset 1	Subset 2	Subset 3
Eigenface	4	0.0	31.1	47.7
	10	0.0	4.4	41.5
Eigenface w/o 1st 3	4	0.0	13.3	41.5
	10	0.0	4.4	27.7
Correlation	29	0.0	0.0	33.9
Fisherface	4	0.0	0.0	4.6

Table

Interpolating between Subsets 1 and 5

Method	Reduced Space	Error Rate (%)		
		Subset 1	Subset 2	Subset 3
Eigenface	4	53.3	75.4	52.9
	10	11.1	33.9	20.0
Eigenface w/o 1st 3	4	31.11	60.0	29.4
	10	6.7	20	12.9
Correlation	129	0.0	21.54	7.1
Fisherface	4	0.0	0.0	1.2

- The Database contains 160 frontal face images covering 16 individuals taken under 10 different conditions: A normal image under ambient lighting, one with or without glasses, 3 images taken with different point light sources, and 5 different facial expressions.
- Recognition was performed using nearest neighbor classifier.
- The corresponding table shows the relative performance of the algorithms when applied to the Yale Database which contains variation in facial expression and lighting.

Table

Variation in Facial Expression, Eye wear and lighting experiment

Method	Reduced Space	Error Rate (%)	
		Close Crop	Full Face
Eigenface	30	24.4	19.4
Eigenface w/o 1st 30	30	15.3	10.8
Correlation	160	23.9	20.0
Fisherface	15	7.3	0.6

- Rather than selecting the classes to be individual people, the set of images can be divided into two classes: “wearing glasses” and “not wearing glasses”.
- In this experiment, the data set contained 36 images from a super set of the Yale Database, half with glasses.
- The corresponding table shows the COMPARATIVE RECOGNITION ERROR RATES FOR GLASSES/ NO GLASSES.

Table

Glasses Recognition

Method	Reduced Space	Error Rate (%)
PCA	10	52.6
Fisherface	1	5.3

Implementation

Conclusion

- All methods perform well if presented with an image in the test set which is similar to an image in the training set.
- Removing the initial three principal components improves the performance of the eigenface method in the presence of lighting variation.
- The Fisherface method is the best in handling variation in lighting and expressions.

The End

Questions? Comments?