**Introduction**

The aim of the project is to recognize a person’s face by making a comparison between its characteristics and those of other people’s faces which are already known.

Specifically, in this second part of the project the Eigenfaces method is developed to make comparisons for face recognition, and the methodology used in Turk, M.A. Pentland, A. P. (1991) Face recognition using eigenface. in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 1991, pp. 586 – 591., is followed as reference.

Basically, what is made is to extract the most important information of a face in order to compare it with a database’s.

Mathematically, it consists in calculating the images set’s covariance matrix eigenvectors so that a space being made of them is created.

While working on image recognition the eigenvectors are called Eigenfaces and on its part, the eigenvector space, face space. Every face in the images set can be constructed as a linear combination of the vectors belonging to this, which is also the basis of the whole methodology since the recognition process refers to measuring the distance between an image and this space and depending on how long it is, to classify the face as a known or unknown one.

It has to be taken into account that the total number of possible Eigenfaces is equivalent to the number of images of the whole dataset. Likewise, the faces can be approximated using only the best M Eigenfaces, which are those with the highest eigenvalues associated with the greatest variance in the images. The number M is usually determined holistically (references)

The process of facial recognition can be summarized in these 4 steps:

1. Calculate the eigenfaces of some images, thus defining the face space.

2. When a new image is introduced, weights of incoming image and the M eigenfaces are calculated by projecting the incoming image in each Eigenface.

3. Determine if the image is a face. This is defining whether the image is relatively close to the face space or not.

4. If the new image is a face, its weight pattern is classified to a previously known or unknown person.

**Data preparation**

For the development of the second part of this project, the same group of images that had been used in the first part, was used again. As it was mentioned, the dataset is hosted in a free online database and has four different directories holding the images in different levels of difficulty. This time the directory used was Faces94, which is a collection of images consisting of a wide range of people’s pictures taken speaking in front of the camera. Because of the speech, this set is an introduction to the variation in facial expression.

Faces94 has 153 individuals images using portrait format. It contains pictures of male, female and male staff in separate directories which are all included in this study. The pictures background is plain green. It does not have any individual’s variation on head scale and image lighting, but it does have a few on head turn, tilt and slant, and considerable on facial expression.

Additionally, there is no individual hairstyle variation as the images were taken in a single session.

**Data centralization and calculation of covariance matrix**

The first step was to centralize the data in order to calculate the covariance matrix.

This was made by means of calculating the mean image for later subtracting it to the whole dataset and performing a dot product of this last matrix and its transposed to get the covariance one.

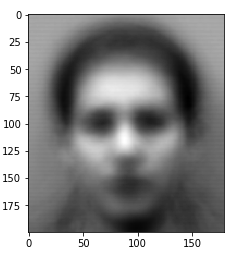


Figure 1. Mean image

The product was done to have the covariance matrix dimensions not in terms of the dataset’s images height and width but the number of images wise, as a strategy to reduce the size of the result.

**Subspaces method: Eigenfaces**

After having the covariance matrix, the next step was to compute its singular value decomposition to get the eigenvectors and eigenvalues. Then, using the results, the face space started to be created by selecting the number of components of the subspace. That selection was run using two options:

* Investigator's criteria of variability captured, where the number of principal components were obtained from a prior set representation percentage. This option raises from how disperse the values are and the percentage of variability captured, so cumulative sums of variance were performed until getting the desired representation, in this study, 98%
* Investigator's criteria of threshold contribution value, where the approach to choose the number of components was to analyze the whole variation as a percentage.

Finally, option 2 was used to set the total of components to create the face space. They were the first 305 components, with 97.59% of variability captured and from which the contribution was less than 0.01%.

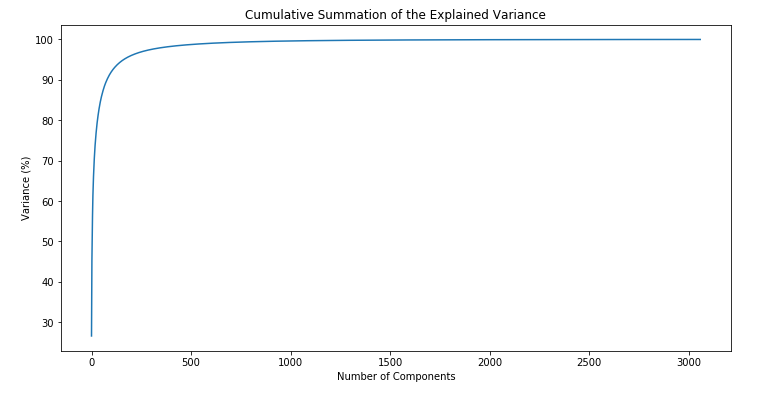


Figure 2. Cumulative summation of the explained variance

**Face space**

After setting the number of principal components that were going to be used to create the face space, the calculations were done: the dot product of the data and the eigenvectors selected was performed and the result was normalized.

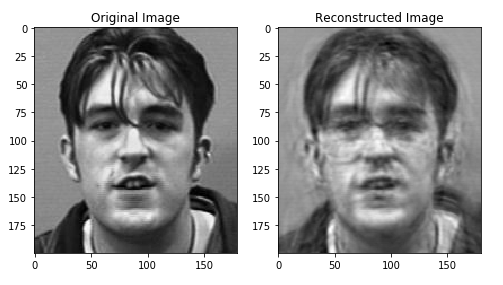


Figure 3. Face space

**Projection of an image on face space**

As every image of the dataset can be constructed as a linear combination of the face space’s vectors, some image projections were made to observe how they were reconstructed by the model.

To begin with, the weight (w) of each Eigenface in the generated subspace due to the original image, was computed. Then, that image was reconstructed by a dot product of w and the face space, and later adding the mean image.



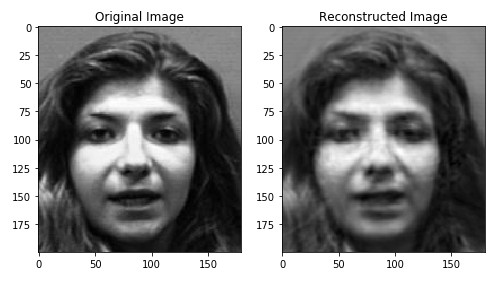


Figure 4. Reconstructed images