Final report

# TP Biometry

*I*mSecu

*Student Author*
Irina MOSCHINI

*Teacher*
Jean-Luc DUGELAY

# Sommaire

# 1. Exercice 1 : building the eigenspace

## *1.1. Part A : computing the eigenspace A*

The maximum size of the eigenspace is given by the total number of eigenvectors, that is to say 100.
The more the eigenvalue is large, the more the associated eigenvector has an impact on the average image, that is to say it carries a lot of information about the pictures of the training set.

## *1.2. Part B : plotting the cumulative sum of eigenvalues*

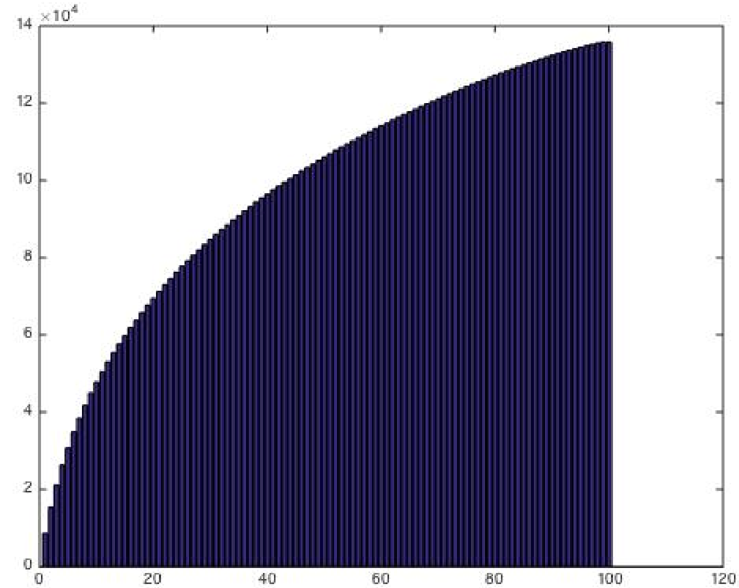


**Figure 1:** Cumulated sum of eigenvalues

We can see that the cumulative sum of eigenvalues grows very quickly at the beginning, and then increases but more slowly : some eigenvectors carry a lot of information, others almost nothing.

## *1.3. Part C : approximating s1_1.jpg*

**(a)** `10 eigenvectors`

**(b)** `20 eigenvectors`

**(c)** `30 eigenvectors`

**(d)** `40 eigenvectors`

**(e)** `50 eigenvectors`

**(f)** `60 eigenvectors`

**(g)** `70 eigenvectors`

**(h)** `80 eigenvectors`

**(i)** `90 eigenvectors`

**(j)** `100 eigenvectors`

Yes, we can rebuilt the face of the men perfectly. With 50 eigenfaces or more, the men is totally recognizable. Having more than 50 eigenfaces doesn't influence a lot the reconstructed image. With 100 eigenfaces, the resulting image is exactly the same as the input one.
It is normal as we used the images of this men to train the model, that is why we can rebuilt the picture perfectly.

### *1.4. Part D : projecting and plotting in face space*

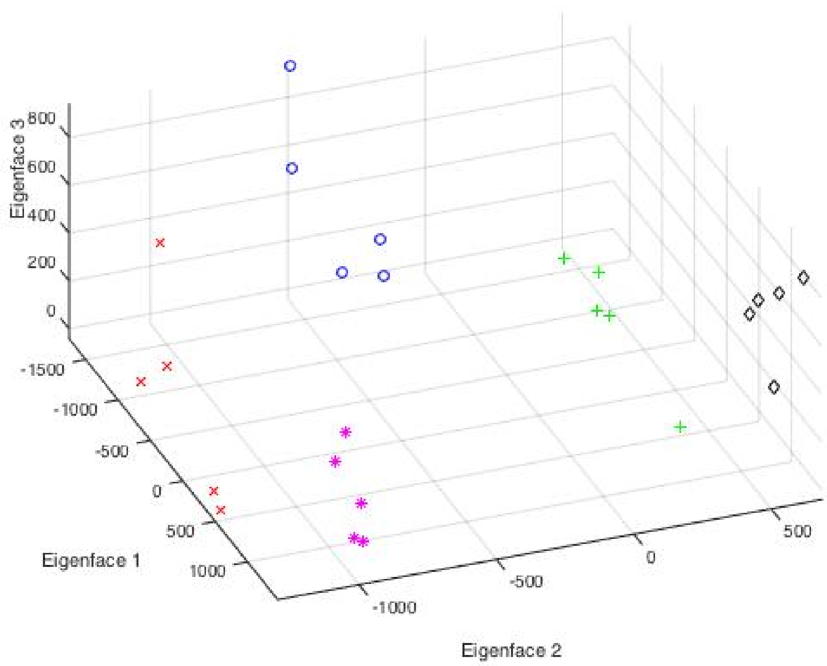


**Figure 3:** Coordinates of the five training faces of the first five training A individuals in the space spanned by the first 3 eigenfaces of space A

The figure displayed by the function `plotFirst3Coordinates` shows that for each image, the projection on the first three eigenfaces has approximatively the same coordinates. We can thus associate each individual to a region of the 3D space.
Thus, given a new image of one of these 5 individuals, projecting it on the three first eigenfaces allows to identify the individual by computing the distance from the point obtained to the other points for example.

## 2. Exercise 2 : identification

### *2.1. Part A : projecting and plotting Test A*

The classification is harder to do as the images are new, they were not used to train the algorithm.
Thus, some images are hard to classify and some individuals have similar coordinates in the space composed of the three first eigenvectors.

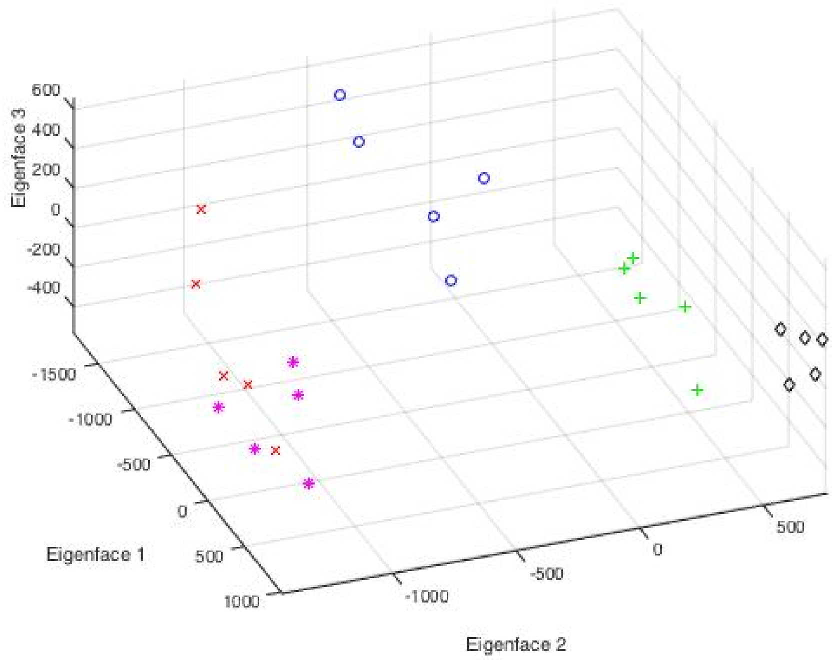


**Figure 4:** Coordinates of the five testing faces of the first five testing A individuals in the space spanned by the first 3 eigenfaces of space A

## 2.2. *Part B : approximating s1_6.jpg*

The results are a way different from the ones obtained with an image of the training set. In fact, even with 100 eigenvectors, the individual is hard to recognize : we can't rebuild it perfectly.
It is normal as this picture is not present in the training set, thus the characteristics have not been learnt by the algorithm (the eigenspace is not adapted to his face).

**(a)** `10 eigenvectors`

**(b)** `20 eigenvectors`

**(c)** `30 eigenvectors`

**(d)** `40 eigenvectors`

**(e)** `50 eigenvectors`

**(f)** `60 eigenvectors`

**(g)** `70 eigenvectors`

**(h)** `80 eigenvectors`

**(i)** `90 eigenvectors`

**(j)** `100 eigenvectors`

## *2.3. Part C : computing the identification rates (first face)*

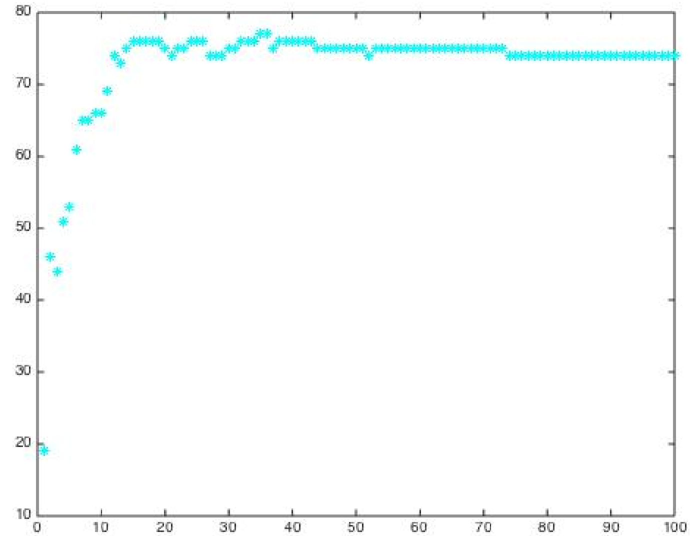


**FIGURE 6:** Identification rate as a function of the number of eigenfaces (first face)

We want the higher identification rate as possible. We can see that the highest value is reached by using between 35 or 36 eigenfaces. The identification rate is thus 77%, which is quite high.
The optimal number of eigenfaces to use is therefore 35 (the smallest number of eigenfaces that led to the best identification rate).

## *2.4. Part D : more identification rates (mean face)*

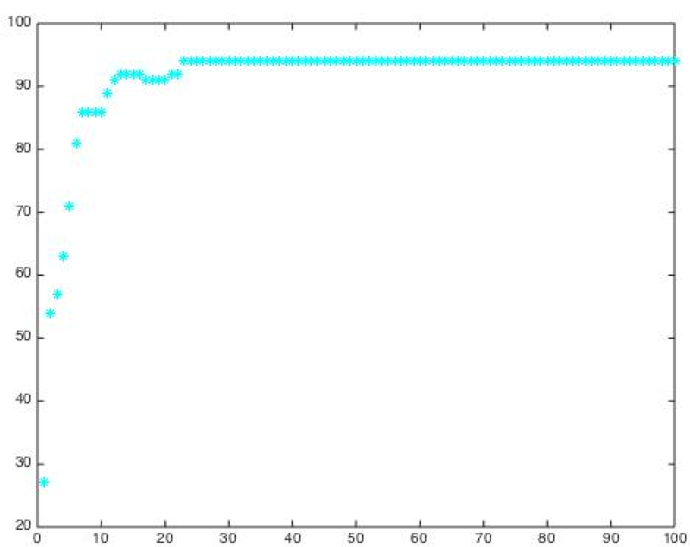


**FIGURE 7:** Identification rate as a function of the number of eigenfaces (mean face)

Taking the mean of the coordinates of the image of a same individual is very efficient as the maximum identification rate is now 94% compared to 77% previously. There are now only 6 picture that are not well classified.
The optimal number of eigenfaces to take here is now 23, which is lower than the 35 eigenvectors we needed before. This method (mean face) is therefore a way more efficient than the first one (first face). However, we can point out that all the picture can't be associated, for sure, to the right person. 6 persons can't be identified.

The cumulative identification might be powerful for some applications where we NEED to find the individual, even if we have several suspects. The police for example might need to find one person in particular and don't mind to arrest arrest a lot of people if at the end the find their target.
The probability to find the men we are looking for is therefore bigger with cumulative identification than with "classical" identification.

### *2.5. Part E : drawing identification rates as a function of N-Best*

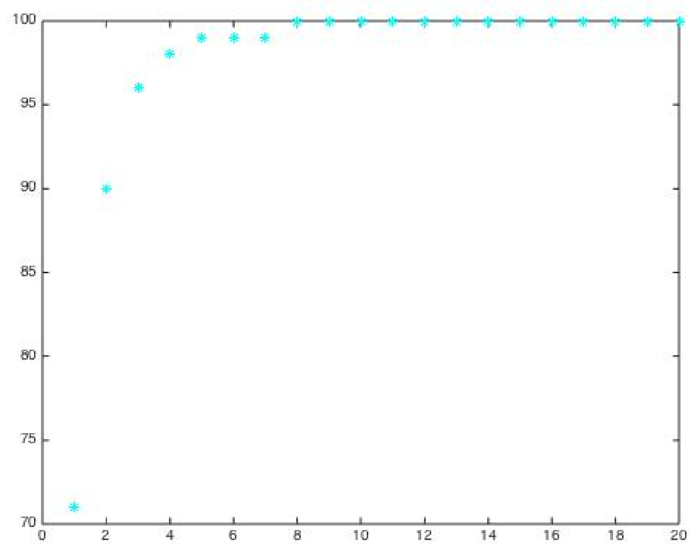


**Figure 8:** Identification rate as a function of parameter N-Best

The plot that represents the identification rate as a function of parameter NBest shows that using NBest = 8 gives a powerful result : an identification rate of 100%, that is to say that if we give the 8 individuals associated with the best ranks, the "real" men is, for sure, one of these 8 men.
Even using just the 5 individuals associated to the best ranks allows to have identification of 99%, that is to say only one mistake has been done on our testing set. This is good but some application can't take the risk to do one mistake.
A last remark is that we used 5 eigenfaces, that gives us, in the part D (mean face), 71% of good identifications which is not a lot compared to the other good results we obtained. Therefore, using more eigenfaces can led to even better result than what we get.

## 3. Exercise 3 : verification

### *3.1. Part A : computing client and impostor scores*

We can see thanks to the plot that the distance for a client can be as high as the distance for an impostor : the curves representing the distance of clients and impostors overlap (in the interval -14.5 ;-12). Thus, some people can be

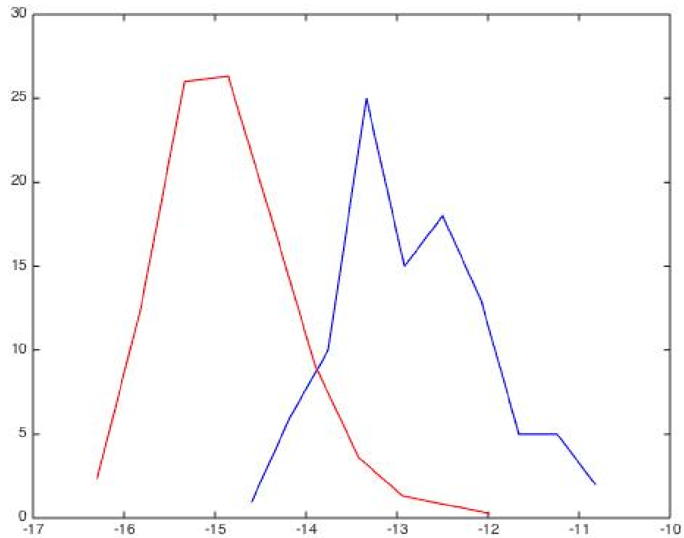


**Figure 9:** Histograms of client and impostor scores

hard to categorize and mistakes can be made in this overlapping area. The decision threshold need to be adapted to minimize the False Acceptance Rate but also the False Rejection Rate at the same time (avoid rejecting everybody).

### *3.2. Part B : plotting the Receiver Operating Characteristic (ROC) curve*

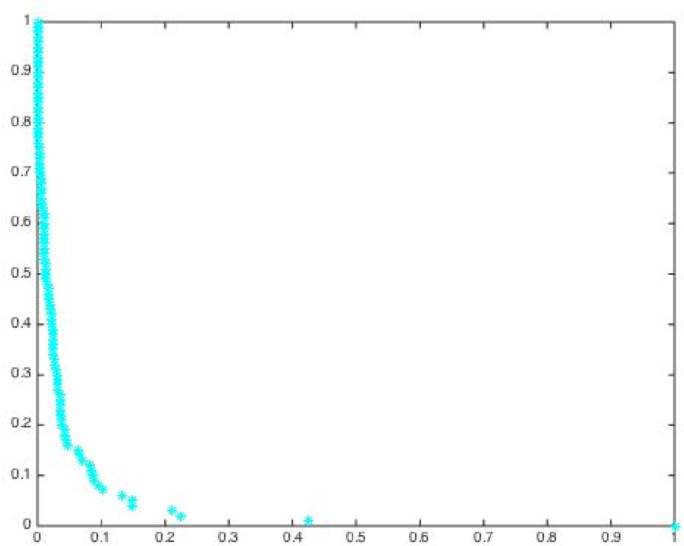


**Figure 10:** Receiver Operating Characteristic (ROC) with the 5 eigenfaces

We can see thanks to the plot that the False Acceptance Rate is very very small for a huge number of decision thresholds : it is smaller than 5% for 85 thresholds over 101. Thus, this technology can be very useful if the security has to be very high and want a impostor rate near 0% (even if the rejection rate, that mean the clients that we reject, is high).

The False Rejection Rate is therefore quite high, but some decision thresholds allow to have small FAR AND FRR. The best threshold we can keep is the one that gives a FAR = 9,63% and FRR = 8,00%.

### *3.3. Part C : drawing the Equal Error Rate (EER) curve*

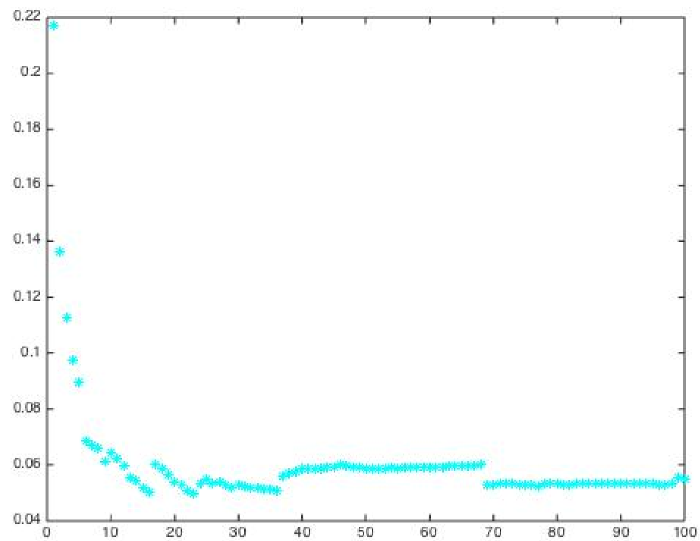


**Figure 11:** Equal Error Rate (EER) for various numbers of eigenfaces

We want the smallest value of Equal Error Rate to have good performances both on FAR and FRR.
We can see thanks to the plot that the ERR is quite high while using a very small number of eigenfaces, but very good results are obtained with more eigenfaces. The best rate we get is obtained with 23 eigenfaces, with an ERR = 4,950%. However, an interesting result is the one obtained with 16 eigenfaces, that gives an ERR of 5,00%. Doing only 5% of FAR and FRR is fair : the system is reliable.

## 4. Exercise 4 : mismatch between the eigenspace and test individuals

### *4.1. Part A : computing identification rates for set B*

The results are not that bad. The best identification rate is 85% obtained with 84 or more eigenfaces. However, using 45 eigenvectors allows to have 84% of identification.
We should use 45 eigenvectors to have good results, but will need 84 eigenvectors to have the optimal identification rate of 85%.

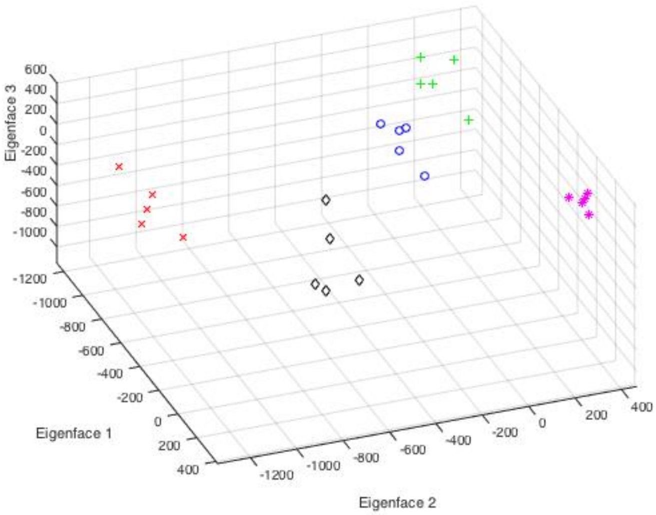


**Figure 12:** Coordinates of the five training B faces of the first five training individuals in the space spanned by the first 3 eigenfaces of space A

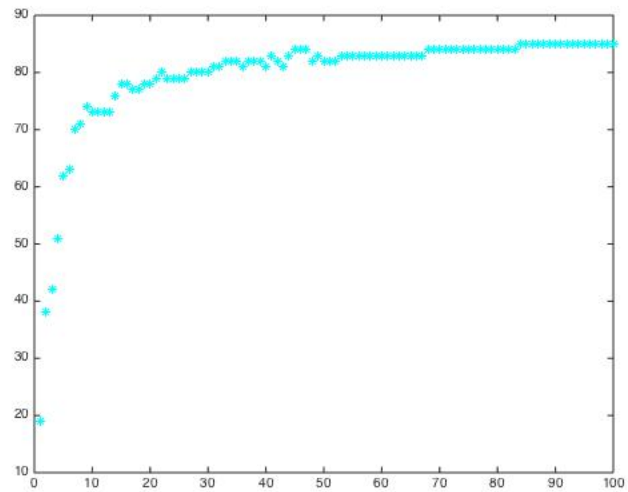


**Figure 13:** Coordinates of the five training B faces of the first five training individuals in the space spanned by the first 3 eigenfaces of space A

## *4.2. Part B : computing the eigenspace B*

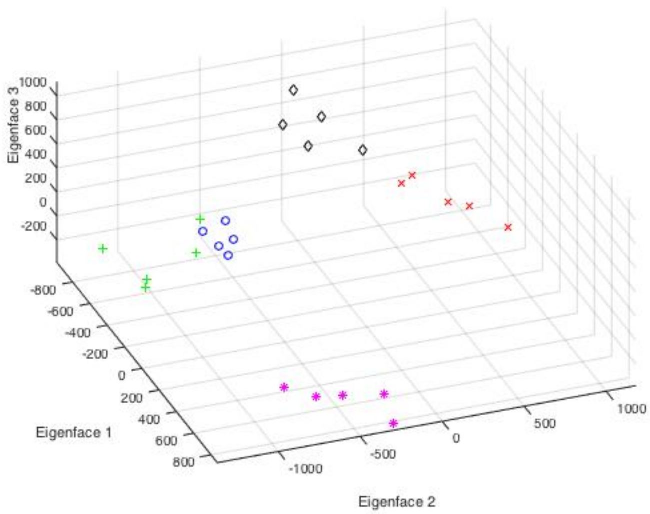


**Figure 14:** Coordinates of the five training B faces of the first five training individuals in the space spanned by the first 3 eigenfaces of space B

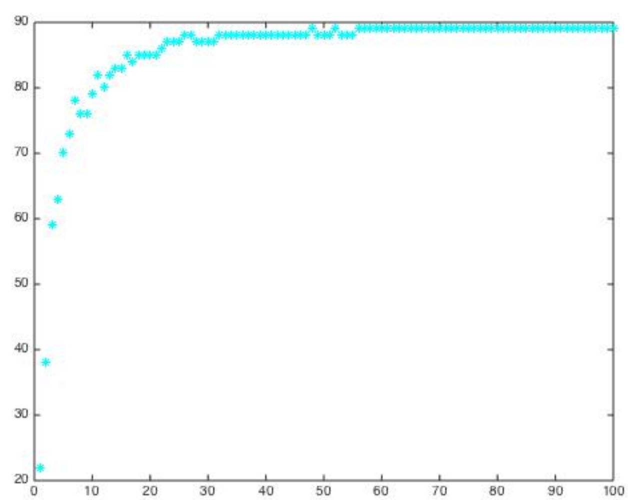


**Figure 15:** Coordinates of the five training B faces of the first five training individuals in the space spanned by the first 3 eigenfaces of space B

The best results are obtained with 48 eigenfaces, that gives 89% of identification. It is better than the results obtained with projection on space A (85%), but not extremely better, the difference is not significant. Thus, projecting on a space on which the training has not been done with the individuals that belongs to the testing can be done : results are similar.

There are some advantages to have the same individuals in the training and testing sets, but also drawbacks :
- Advantages : the results are a bit better than using a training set with other individuals than the ones in the testing set
- Drawbacks : we need to know and enroll all the suspects in the training set, which would take a lot of time and space. We would need to update all the time the training set if new people arrive, or if some leave.