Projet Calcul Scientifique et Analyse de Données
Compte-Rendu de Projet 1

Application de l'ACP : les « Eigenfaces »

François Lauriol, Priscilia Gonthier, Yael Gras
Groupe MN

25 mars 2022

Table des matières

Introduction

L'objectif du projet est de reconstituer le visage de personnes masquées à partir d'une base de donnée de visages non masqués. On dispose pour cela de 32 personnes différentes avec pour chacune 6 postures afin de constituer une base de donnée. On ne prendra pour cette première partie que 4 personnes avec 4 postures chacune pour la créer.

1 Les « Eigenfaces »

Exercice 1 : Analyse en Composantes Principales

Nous avons tout d'abord calculé les axes principaux des images d'apprentissage appelées eigenfaces, ainsi que l'individu moyen, à partir des 15 valeurs propres non nulles de la matrice de variance/covariance Σ des images de la base de donnée (figure 1).

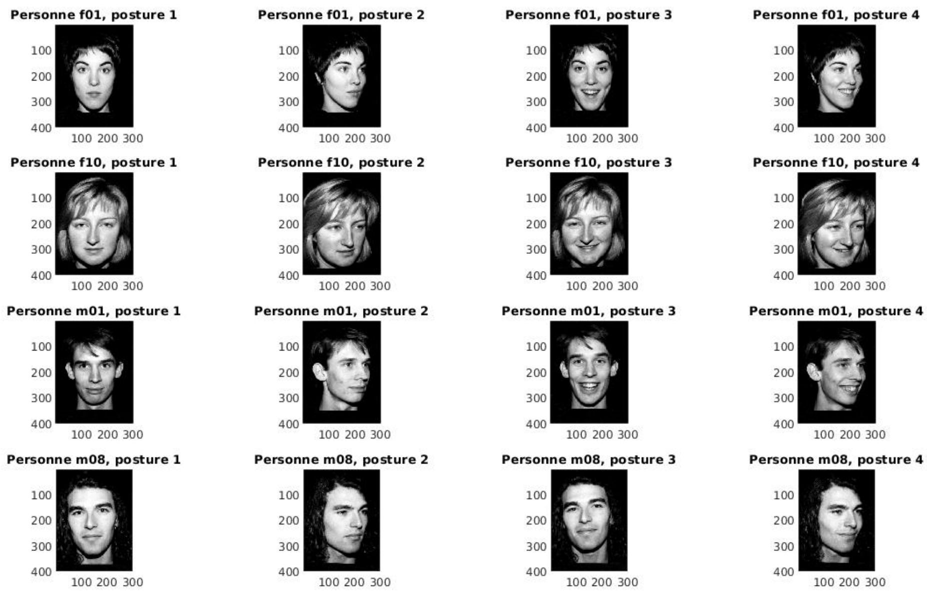

Figure 1 – Photos de data base (sans masque)

On obtient l'individu moyen et les eigenfaces de la figure 2.

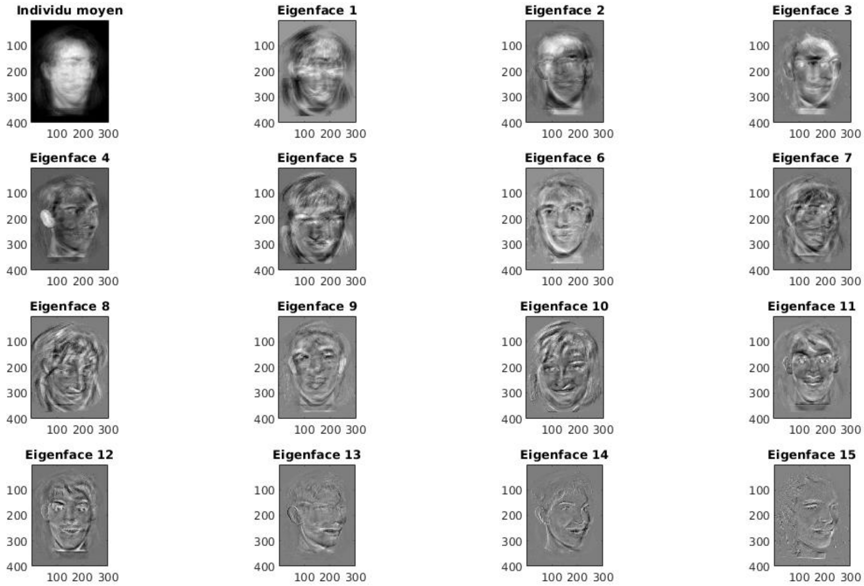

Figure 2 – Eigenfaces et individu moyen des photos de data base (sans masque)

Exercice 2 : Projection des images sur les eigenfaces

La seconde étape consiste à calculer les composantes principales et ensuite à reconstitruire les images en projetant les q premières composantes principales sur les q premières eigenfaces pour $q \in [0, 15]$ (q=0 correspond à l'individu moyen). On obtient donc au final les images figure 3.

Figure 3 – Photos des personnes sans masque après reconstitution

L'erreur quadratique moyenne entre les images originales et les images reconstruites sont tracées sur la figure 4.

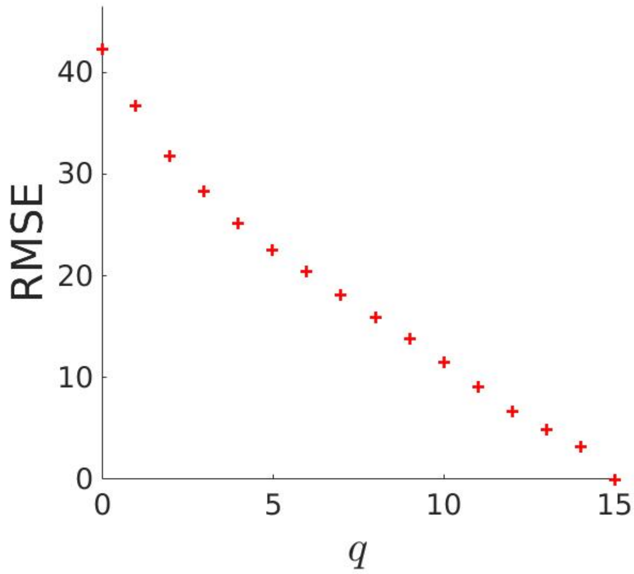

FIGURE 4 – RMSE après reconstitution à partir de l'image sans masque

L'erreur quadratique moyenne décroît jusqu'à atteindre . On remarque donc bien que les 15 eigenfaces suffisent à décrire les 16 images.

Exercice 3 : Travail sur les visages masqués

Nous disposons des images masquées à reconstruire présentes sur la figure 5. Ce sont en réalité les images précédentes auxquelles on a ajouté un rectangle noir (correspond à mettre des zéros dans les matrices d'images) pour représenter le masque.

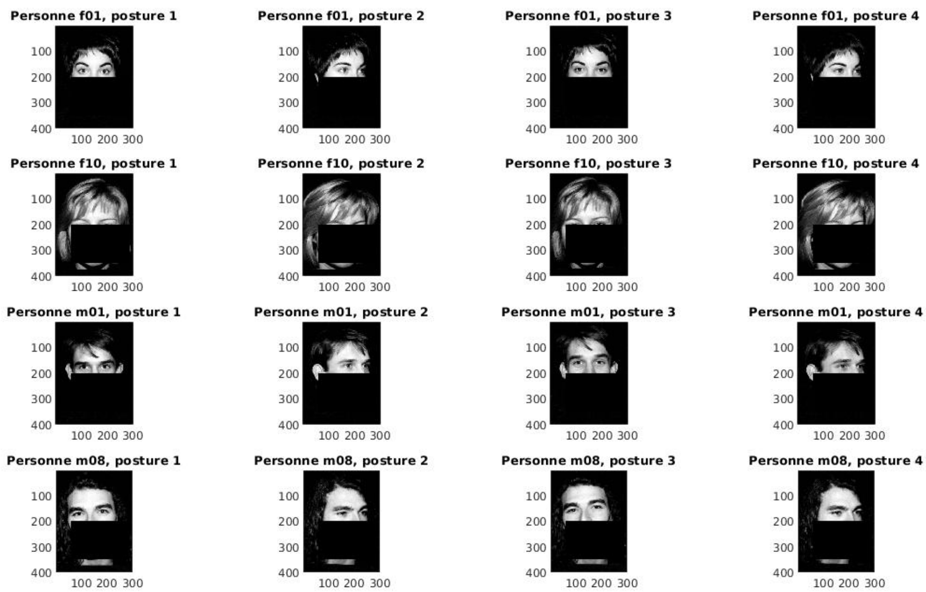

FIGURE 5 – Photos des personnes avec masque avant la reconstitution

Comme précédemment, nous avons calculé les eigenfaces et l'individu moyen sur ces visages masqués afin d'obtenir les images de la figure 6.

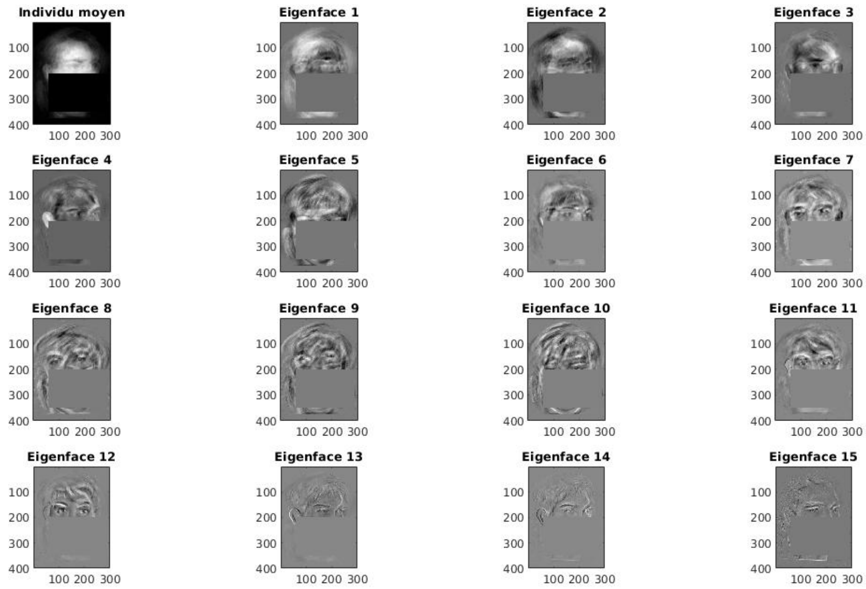

FIGURE 6 – Eigenfaces et individu moyen des photos de visages masqués

Et nous avons ensuite reconstruit les visages masqués comme pour les visages non masqués. (figure 7).

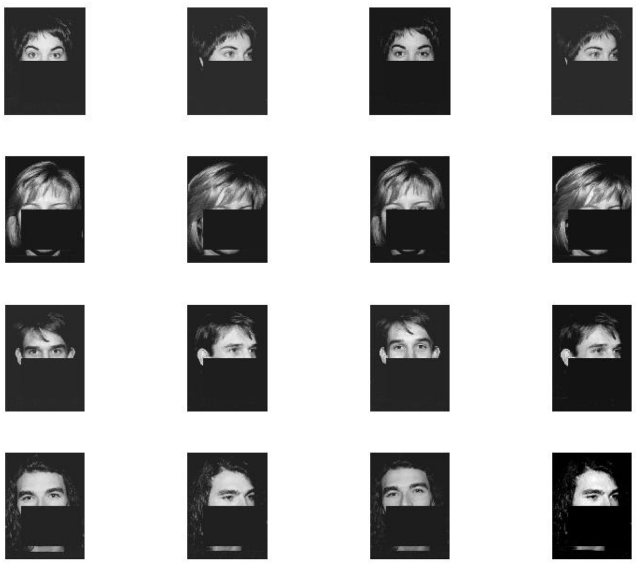

FIGURE 7 – Photos des personnes avec masque après la reconstitution

L'erreur quadratique moyenne entre les images originales masquées et les images reconstruites sont tracées sur la figure 8.

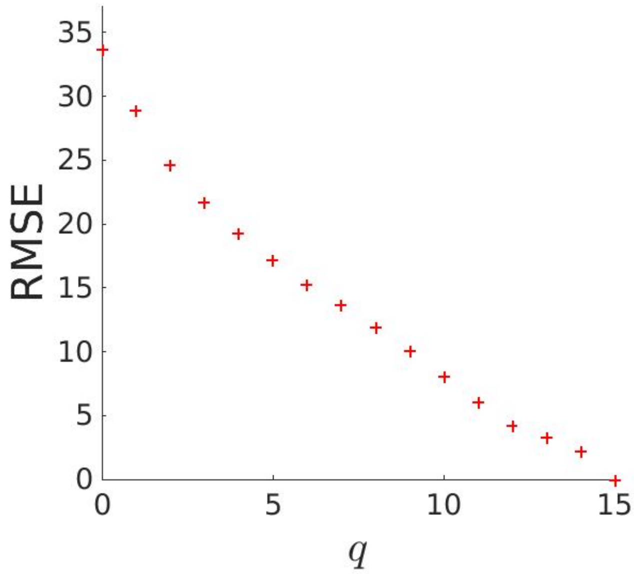

FIGURE 8 – RMSE après reconstitution à partir de l'image avec masque

On remarque comme précédemment que l'erreur quadratique décroît et que 15 eigenfaces suffisent à retrouver les images d'origine.

2 L'ACP et la méthode de la puissance itérée

Question 4 :

Soit λ valeur propre de $H^T H$

Soit $x \in \mathbb{R}^n$ vecteur propre de $H^T H$ on a donc

$$\begin{aligned} H^T H x &= \lambda x \\ \Leftrightarrow H H^T H x &= H \lambda x \\ \Leftrightarrow H H^T (H x) &= \lambda (H x) \end{aligned}$$

Donc λ est aussi une valeur propre de HH^T associée au vecteur propre Hx. Nous avons ainsi que toute valeur propre de $H^T H$ est valeur propre de HH^T. Et les vecteurs propres de HH^T se déduisent de ceux de $H^T H$ en les multipliant par H.

Question 5 :

En lançant le script puissance_iteree.m nous obtenons les résultats suivants :

Erreur pour la méthode avec la grande matrice = 9.915e-09
Erreur pour la méthode avec la petite matrice = 9.910e-09
Écart relatif entre les deux valeurs propres trouvées = 9.17e-09

Temps pour une itération avec la grande matrice = 1.472e-03
Temps pour une itération avec la petite matrice = 1.247e-04

Valeur propre dominante (méthode avec la grande matrice) = 9.326e+04
Valeur propre dominante (méthode avec la petite matrice) = 9.326e+04
Valeur propre dominante (fonction eig) = 9.326e+04

Question 6 :

Après recherche nous avons trouvé que la complexité de la fonction *eig* de Matlab est en $O(q^3)$ or la complexité de notre algorithme de la puissance itérée est d'après calcul en $O(q^2 \times it_{max})$. Ainsi si $it_{max} \leq q$ alors l'algorithme de la puissance itérée est plus rapide que la fonction *eig* de Matlab. Cependant, réduire la valeur d'it_{max} peut engendrer des imprécisions sur les valeurs trouvées comme valeurs propres et vecteurs propres.

Question 7 :

Il vaut mieux appliquer la méthode sur la plus petite matrice afin de minimiser le temps de calcul et la mémoire utilisée. En effet le temps de calcul avec la petite matrice est de $1,746 \times 10^{-4}$ contre $6,954 \times 10^{-3}$ pour la grande matrice. La petite matrice ayant une taille inférieure à la grande matrice (dans notre exemple 500×500 contre 1500×1500) la place en mémoire sera réduite en utilisant la petite matrice, c'est à dire $A^T A$. De plus l'erreur pour la méthode avec la petite matrice est équivalente à celle de la grande matrice.