Projet en Calcul Scientifique/Analyse de Données : Séance 1 : ACP et Puissance Itérée

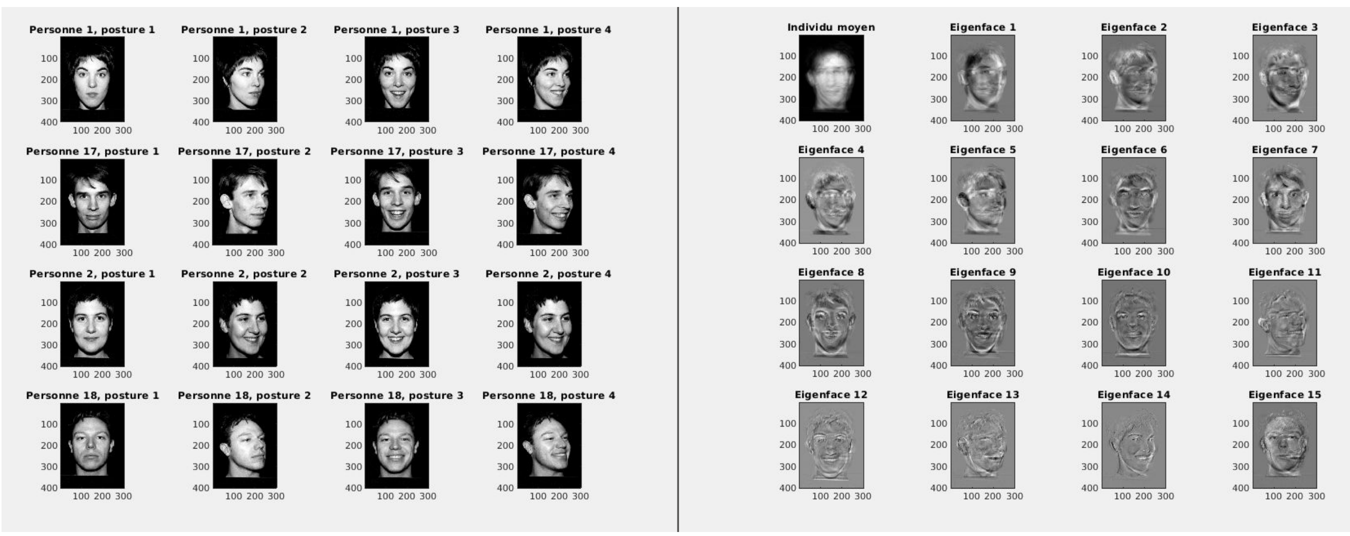


FIGURE 1 – Extraction des EigenFaces par méthode ACP

Projet réalisé par : GUILLOTIN Damien, JOURDAN Pierre-Eliot et HEURTEBISE Tom

Encadré par : LOYE Justin

Département Sciences du Numérique de l'ENSEEIHT, 26/03/2021

Table des matières

Table des figures

1 Introduction

Dans un contexte de pandémie mondiale, chacun a appris a porter quotidiennement un masque qui recouvre plus de la moitié du visage. Le projet qui nous a été confié était de reconstruire le visage de personnes occulté par un masque, le tout à partir d'une banque d'images. Enfin de réaliser ce traitement de l'image nous avons dû employer des techniques étudiées dans l'Unité d'Enseignement de Première Année en Sciences du Numérique à l'ENSEEIHT, Calcul Scientifique et Analyse de Données.
Ainsi dans ce premier rapport vous pourrez observer l'application que nous avons faite de la méthode d'Analyse en Composante Principale (ACP) étudiée en TD d'Analyse de Données. Cette dernière permettant de réduire considérablement les dimensions d'un problème et disposant d'images dont la dimension équivaut à 120 000 il était donc nécessaire d'appliquer l'ACP.
Au cours de cette première séance nous avons aussi utilisée une méthode permettant de déterminer les valeurs propres de la matrice de covariance utilisée dans l'ACP. C'est donc selon ces deux grandes parties que ce bref rapport s'articulera.

2 Les Eigenfaces

Tout d'abord précisons bien une phrase importante écrite dans le sujet [1] qui est : "Ce projet s'inspire d'un article [2] intitulé , écrit par Turk et Pentland et publié dans le Journal of Cognitive Neuroscience en 1881. En effet la méthode ne nous apprtient pas et est issue de travaux passés

2.1 Exercice 1 : Analyse en Composantes Principales (ACP)

2.1.1 Rappels du sujet

La première étape de cette séance de projet consistait à réaliser l'ACP d'une banque d'images constituées de n images où $n = nombreindividus * nombrepostures$.
Toutefois avant de présenter les résultats de l'ACP, rappelons le format des données qui étaient mises à notre disposition. Pour cela, notons la phrase issue de [1] qui nous dit que : "Chacune de ces n images en niveaux de gris est stockée dans une matrice bidimensionnelle de taille 300 x 400".
Le sujet [1] nous précise ainsi que chaque image est représentée par un vecteur colonne appartenant à $\mathbb{R}^{'}$ avec donc $p = 12000$.
Les données de la banque d'images à analyser se trouve donc dans une matrice X, de taille n x p qui "contient sur chaque ligne la transposée d'une image vectorisée" [1]. Pour conclure ces brefs rappels, précisons que le sujet exigeait de travailler seulement sur 4 individus lors de cette première séance dont voici les images :

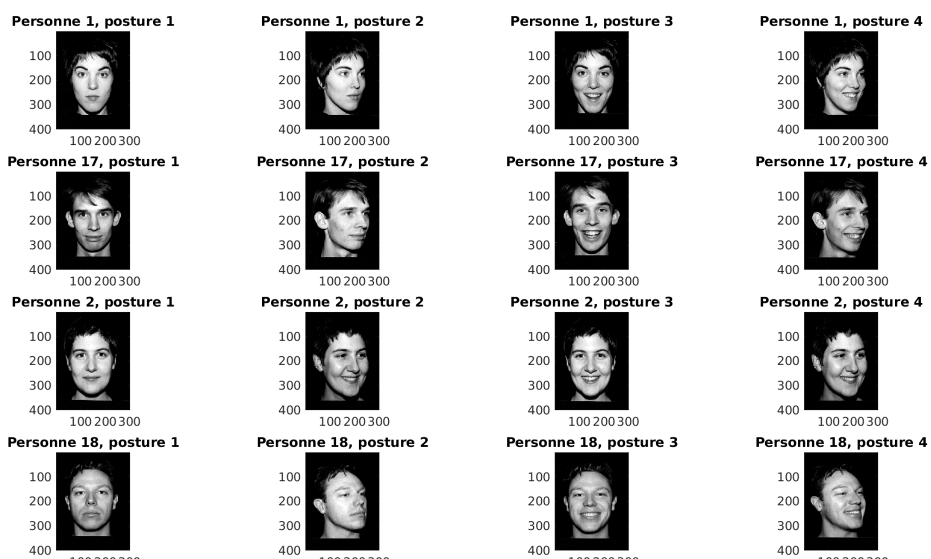


FIGURE 2 – Banque d'images

2.1.2 Question 1 : Calcul des axes principaux des images d'apprentissage

Nous vous invitons à consulter le code fourni dans l'archive et plus précisément le fichier eigenfaces.m afin d'avoir le détail de notre implémentationd de l'ACP pour cette banque d'images.
Nous vous invitons toutefois à prendre en compte le respet des remarques fournis en annexe du sujet [1] et notamment celle indiquant la nécessité de ne calculer que les $n-1$ premières valeurs propres de $\sum = X_c X_c^T/n$ et celle indiquant de calculer les couples propres de $\sum_2 = X_c^T X_c/n$ qui sont identiques à ceux de $\sum$. Dans la figure ci-dessous vous retrouverez les Eigenfaces que nous avons obtenus.

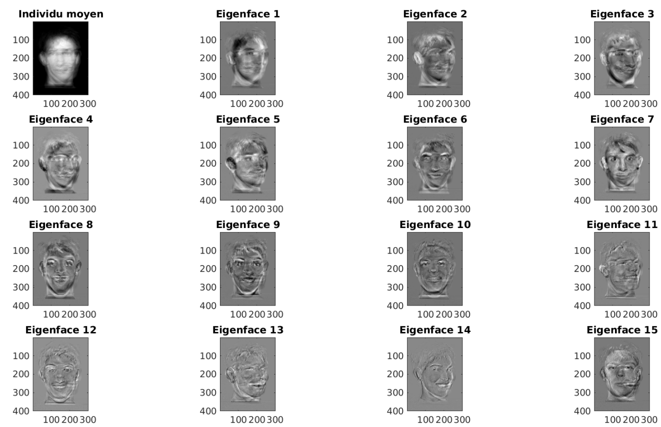


FIGURE 3 – Eigenfaces images de base

2.2 Exercice 2 : projection des images sur les eigenfaces

2.2.1 Afficher les images d'apprentissage reconstruites à l'aide des q premières eigenfaces et des q premières composantes principales

- Pour observer l'implémentation en MATLAB de cet exercice veuillez vous référer une nouvelle fois au code fourni mais en consultant cette fois le fichier projection.m. Pour ce qui est des résultats veuillez trouver l'évolution des images reconstruites en fonction de q (pour q allant de 0 à n-1) sur les figures en Annexe

- RMSE

Nous obtenons qui une erreur quadratique moyenne qui décroit comme on peut l'observer sur la figure ci dessous

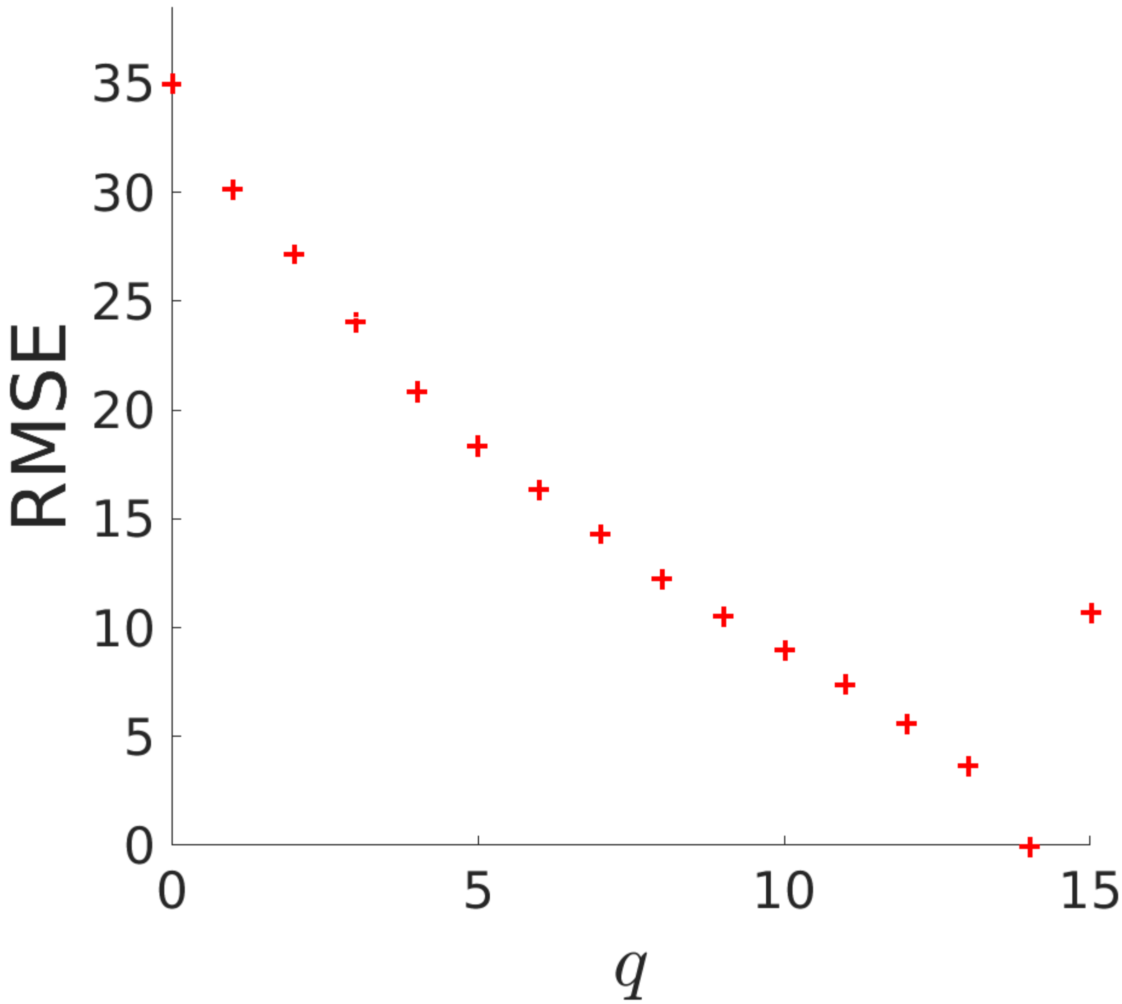


FIGURE 4 – RMSE en fonction de q

2.2.2 Question 3 : eigen_faces sur les masques

En ce qui concerne l'ACP sur les visages masqués nous vous renvoyons cette foi-ci vers le fichier eigenfaces_masque.m mais voici nos résultats :

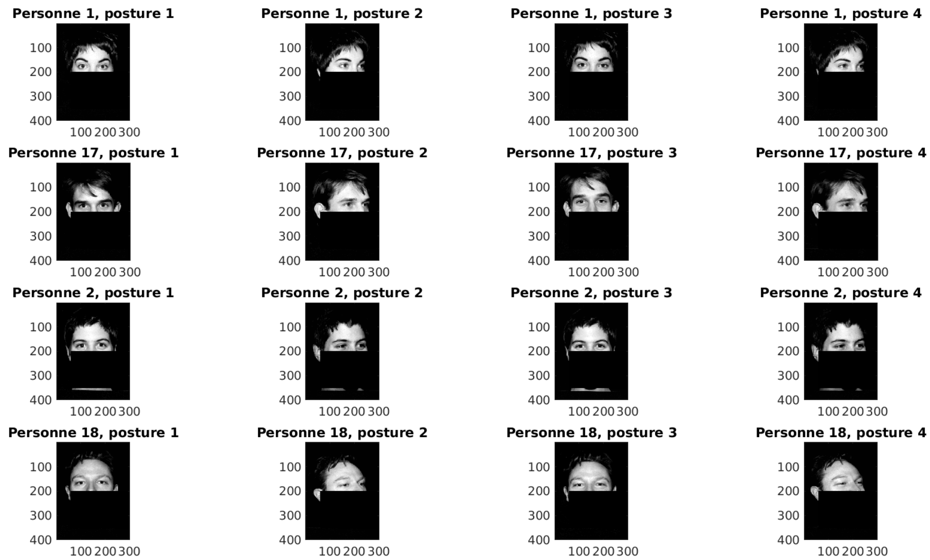


Figure 5 – Personnes masquées

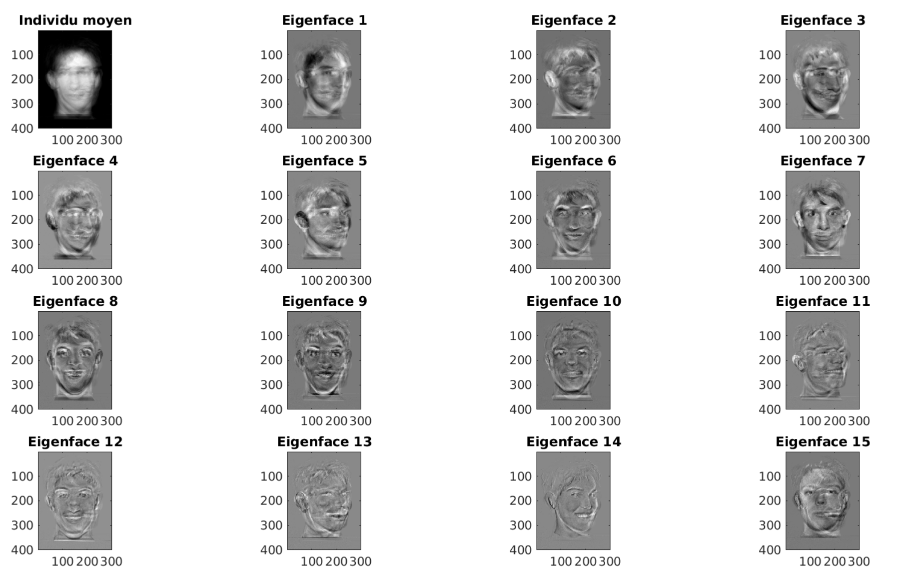


Figure 6 – Eigen faces des personnes masquées

3 L'ACP et la méthode de la puissance itérée

Dans cette partie nous nous sommes intéressés à la méthode pour déterminer les couples propres de la matrice de covariance. En effet nous avons comparé la méthode eig de MATLAB avec l'algorithme de la puissance itérée vu en TD.

3.1 Question 4 : Travailler sur $A^T A$ ou AA^T ?

Dans cette sous-partie nous allons montrer que $\forall H \in \mathbb{R}^{l \cdot k}$, les couples propres de $H^T H$ sont exactement les couples propres de HH^T.
(Pour simplifier, on travaille dans $\mathbb{R}$)
Soient $H \in \mathbb{R}^{l \cdot k}$ et $\lambda \in \mathbb{R}$ tel que $\exists u \in \mathbb{R}^*$, $H^T H u = \lambda u$
$\Leftrightarrow (H^T H u)^T = (\lambda u)^T$
$\Leftrightarrow u^T H H^T = \lambda u^T$
$\Leftrightarrow u^T (HH^T - \lambda I_n) = 0$
$\Leftrightarrow (u^T (HH^T - \lambda I_n))^T = 0^T$
$\Leftrightarrow (HH^T - \lambda I_n)^T u = 0$
$\Leftrightarrow ((HH^T)^T - (\lambda I_n)^T) u = 0$
$\Leftrightarrow (HH^T - \lambda I_n) u = 0$
$\Leftrightarrow HH^T u = \lambda u$

On en déduit donc que les couples propres sont identiques pour les deux matrices.

3.2 Question 5 : Puissance itérée

cf le code de puissance_iteree

3.3 Question 6 : est-il plus utile en theorie d'utiliser une fonction telle que eig ou la méthode de la puissance itérée dans le cadre de l'ACP effectuée ?

En théorie nous n'avons besoin que des $n-1$ premières valeurs propres de la matrice $X_c^T X_c$ et on remarque que nous controlons le nombre de couples propres produits par l'algorithme de la puissance itérée, la ou la fonction MATLAB eig nous fournit l'ensemble des valeurs propres.

A moins d'une implémentation de complexité bien inférieure pour eig (ce que nous ne sommes pas parvenus à vérifier)

il faut donc privilégier l'utilisation de la puissance itérée dans ce cas là.

3.4 Question 7 : pour la puissance itérée, sur quelle matrice doit-on appliquer la méthode pour minimiser le temps de calcul et la mémoire utilisée ?

On remarque en faisant tourner les deux versions de la puissance itérée qu'en cherchant les couples propres de AA^t le temps de calcul est 10 fois supérieur qu'en cherchant les couples propres de $A^t A$. Ceci n'est pas très étonnant compte tenu de la dimension de la première matrice qui est bien plus grande que la deuxième (1500 x 1500 contre 500 x 500).

Il faut donc choisir$A^t A$ pour appliquer l'algorithme de la puissance itérée !

4 Conclusion

Au cours de cette séance de projet nous avons effectué un prétraitement permettant de réduire la dimension du problème. Cependant nous sommes loin d'avoir abouti ju'squà l'objectif final du projet qui sera de reconnaitre des personnes malgré leur masque. La prochaine étape du projet consistera à améliorer l'efficacité du calcul de notre ACP, nous resterons donc encore sur le calcul de c(x) tandis que durant la dernière séance nous nous attaquerons au coeur du problème. Ci-dessous vous retrouverez un schéma bilan qui synthétise le déroulé du projet (cf [1]).

$$\mathbf{x} \in \mathbb{R}^{120000} \rightarrow \boxed{\quad \text{ACP} \quad} \rightarrow \mathbf{c}(\mathbf{x}) \in \mathbb{R}^{q \ll 120000} \rightarrow \boxed{\quad \text{h} \quad} \rightarrow h(\mathbf{x})$$

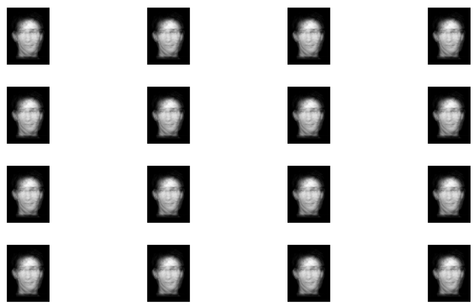
FIGURE 7 – Des données brutes x jusqu'à la reconnaissance d'invidus h(x), figure issue de [1]

Références

[1] Sujet séance-Projet 1 : Application de l'ACP, les Eigenfaces - Méthode de la puissance, 2020-2021. Sujet du projet de Calcul Scientifique/Analyse de Données.

[2] Matthew Turk and Alex Pentland. Eigenfaces for recognition. *Cognitive Neuroscience*, 3(1) :71–86, 1991.

Annexe

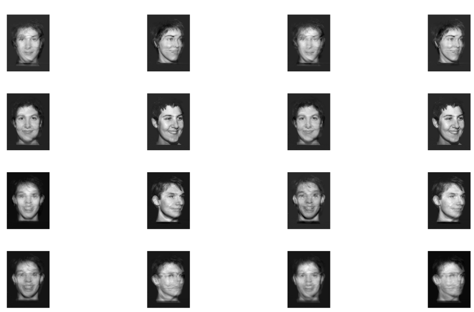


(a) q= 0

(b) q= 1

(c) q= 2

(d) q= 3



(e) q= 4

(f) q= 5

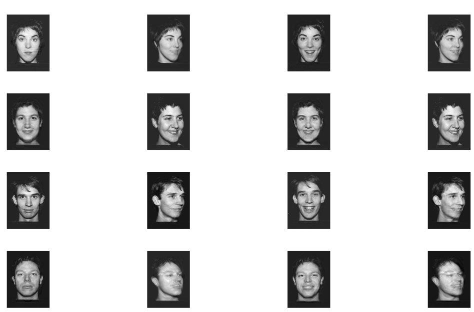
FIGURE 8 – Itération de q = 0 à q = 5

(a) q= 6

(b) q= 7

(c) q= 8

(d) q= 8



(e) q= 10

(f) q= 11

FIGURE 9 – Itération de q = 6 à q = 12

(a) q= 13

(b) q= 14

FIGURE 10 – Itération de q = 13 à q = 14