



Application de l'ACP : A quoi tu ressembles sans ton masque ?

Avec la pandémie de la Covid-19, nous avons appris à porter des masques quotidiennement. Désormais la moitié de notre visage est cachée nous laissant le regard pour seul moyen d'expression visible. Cette opération de cacher le visage s'apparente à un domaine en traitement d'images et de vidéos appelé "inpainting". En effet, les pixels correspondant à la zone du masque (modélisé par un rectangle) sont mis à 0. Et les *eigenfaces* permettent, entre autres, de restaurer la zone dégradée.

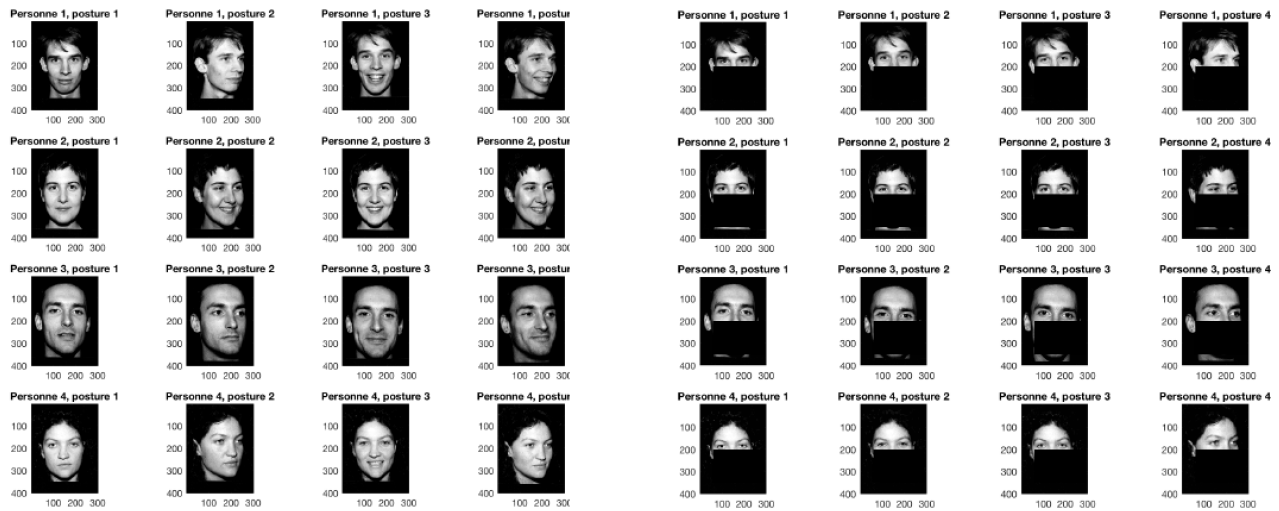


FIGURE 1 – Une base de visages et de visages masqués

Données manipulées

La base de données sur laquelle ce projet est appliqué est une collection d'images utiles pour mener des expériences en psychologie (*Psychological Image Collection at Stirling (PICS)*¹). Elle comporte 32 individus (16 hommes et 16 femmes) avec 6 postures/expressions faciales (face, trois quart face et trois émotions différentes par posture). La FIGURE 1 montre une base de $n = 16$ visages présentant 4 personnes masquées et non masquées dans 4 positions/émotions différentes relativement à la caméra.

Dans cet exemple, chaque image est de taille : 400 lignes \times 300 colonnes. Il est immédiat, par empilement des colonnes, de se ramener à la situation où chaque image est un vecteur $\mathbf{x}_i \in \mathbb{R}^p$ avec $p = 400 \times 300 = 120000$ composantes (niveaux de gris / pixels).

1. <http://pics.stir.ac.uk>

Cependant, nous disposons d'une base d'images entières (sans port du masque) pour pouvoir réaliser une reconstruction de la zone masquée.

En adoptant les notations usuelles pour l'ACP, la base d'images forme un tableau de données $\mathbf{X} \in \mathcal{M}_{\mathbb{R}}(n = 16, p = 120000)$ avec, en ligne i , le vecteur $\mathbf{x}_i^T \in \mathbb{R}^p$ qui représente la i^{eme} image de la base.

L'individu moyen $\bar{\mathbf{x}} \in \mathbb{R}^p$ est le vecteur des **moyennes arithmétiques des variables** : il représente alors, tout simplement, la version «vectorisée» de l'image moyenne $\bar{\mathbf{x}} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i$ des images de la base. Cette image moyenne est montrée en haut et à gauche de la FIGURE 2.

Il est alors possible de centrer les données (c'est-à-dire les images de la base) pour former le tableau centré $\mathbf{X}_c = \mathbf{X} - \mathbf{1}_n \bar{\mathbf{x}}^T$.

La différence avec le TP1 d'Analyse de Données où chaque pixel d'une image couleur constitue un point de \mathbb{R}^3 est qu'ici chaque image qui constitue un point d'un espace affine \mathbb{R}^p de dimension très élevée.

1 Séance 1 : Eigenfaces

Pour pouvoir remplir la zone du masque, nous nous appuyons sur le travail de M. Turk et A. Pentland intitulé «Eigenfaces for Recognition»² [2].

Il constitue une des applications les plus intéressantes et populaires de l'ACP au domaine de la reconnaissance de forme. Il s'agit tout simplement d'appliquer l'ACP à partir de données de très grandes tailles : des images de visages. Ici, l'ACP va nous permettre de réduire la dimension des images et de créer un système de reconnaissance grâce aux composantes principales pour retrouver le visage entier dans la base d'apprentissage le plus similaire au visage masqué et de permettre une reconstruction de la zone du masque.

Il est inutile et inopportun d'utiliser les $p = 120000$ niveaux de gris pour comparer l'image masquée avec chacune des images de la base d'apprentissage. L'ACP est donc un prétraitement qui consiste à réduire la dimension des données selon le schéma suivant où le prédicteur h prend en entrée la sortie de l'ACP :

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

Le vecteur caractéristique $\mathbf{c}(\mathbf{x})$ est tout simplement formé des q premières composantes principales résultant de la projection de l'image requête \mathbf{x} sur les q premiers vecteurs propres unitaires

$$\{\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_q\} \subseteq \mathbb{R}^p \text{ de } \Sigma = \frac{1}{n} (\mathbf{X}_c)^T \mathbf{X}_c.$$

Ces vecteurs propres sont des vecteurs de \mathbb{R}^{120000} est donc également des images que l'on appelle «eigenfaces». Ces images sont présentées dans la FIGURE 2.

2. <https://www.mitpressjournals.org/doi/pdfplus/10.1162/jocn.1991.3.1.71>

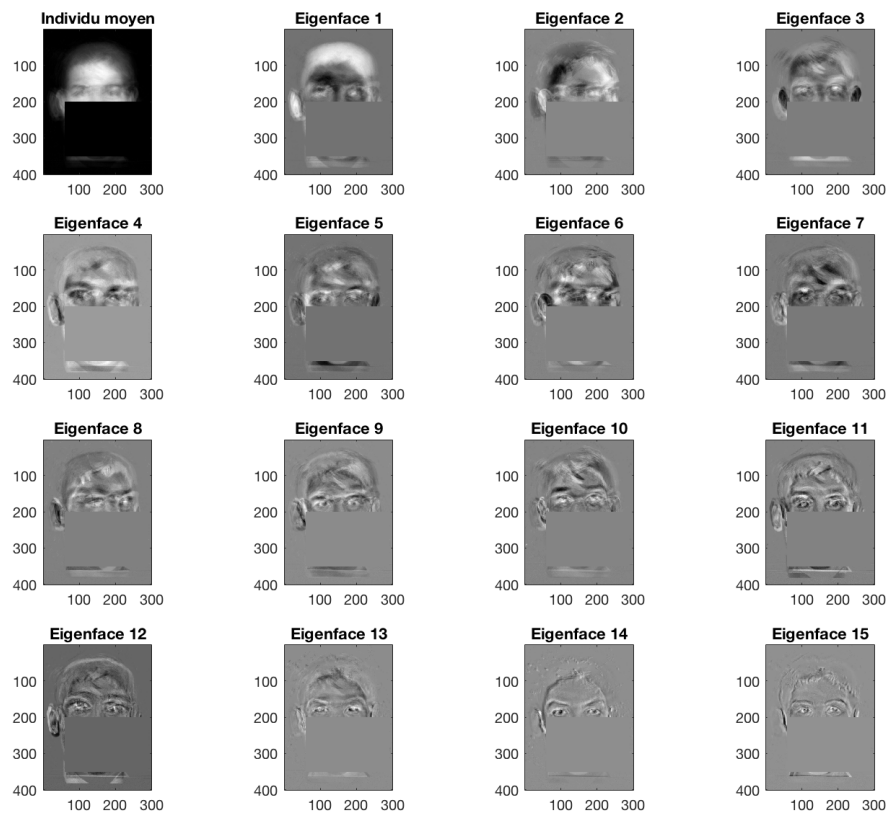


FIGURE 2 – Les «eigenfaces» masqués

2 Séance 2 : Calcul efficace des vecteurs propres principaux

La section précédente occulte toutefois un problème délicat : la grande taille de $\Sigma = \frac{1}{n}(\mathbf{X}_c)^\top \mathbf{X}_c$ ($p \times p = 120000 \times 120000$) ne permet pas de calculer ses vecteurs propres de manière aisée. Ce sera un des premiers problèmes auquel nous nous intéresserons lors de ce projet.

Comme il s'agit des vecteurs propres dominants (associés aux plus grandes valeurs propres) de la matrice Σ qui nous intéresse, des méthodes qui permettent de calculer ces vecteurs propres seront étudiées et implémentées.

Ces méthodes s'appuient sur un espace de vecteurs que nous allons faire converger vers l'espace des vecteurs propres liés aux valeurs propres dominantes. Elles utilisent le quotient de Rayleigh ([1]) et peuvent être vues comme une extension à plusieurs vecteurs de la méthode de la puissance itérée.

Vous allez implémenter en MATLAB différentes versions qui seront de plus en plus efficaces et qui permettront d'arrêter les calculs quand le «bon» nombre de vecteurs propres sera atteint. Ce «bon» nombre pourra être déterminé par le pourcentage d'information (contraste) que vous souhaitez atteindre avec votre ACP.

3 Séance 3 : Reconnaissance et reconstruction

A partir des composantes principales de la base des visages masqués, on a défini ainsi :

$$\mathbf{c}(\mathbf{x}) \stackrel{\text{déf}}{=} \begin{bmatrix} \mathbf{w}_1^\top \mathbf{x} \\ \mathbf{w}_2^\top \mathbf{x} \\ \vdots \\ \mathbf{w}_q^\top \mathbf{x} \end{bmatrix}$$

On dit que $\mathbf{c}(\mathbf{x})$ est une représentation compacte préservant au mieux l'information contenue dans \mathbf{x} avec seulement q composantes principales. La prédiction h peut alors être définie de diverses manières. Parmi les possibilités les plus simples, on peut chercher les images de la base dont les représentations compactes sont les plus proches (au sens d'une distance d bien choisie) de celle associée à l'image requête :

$$h(\mathbf{x}) = \arg \min_{i \in 1 \dots n} d(\mathbf{c}(\mathbf{x}_i), \mathbf{c}(\mathbf{x}))$$

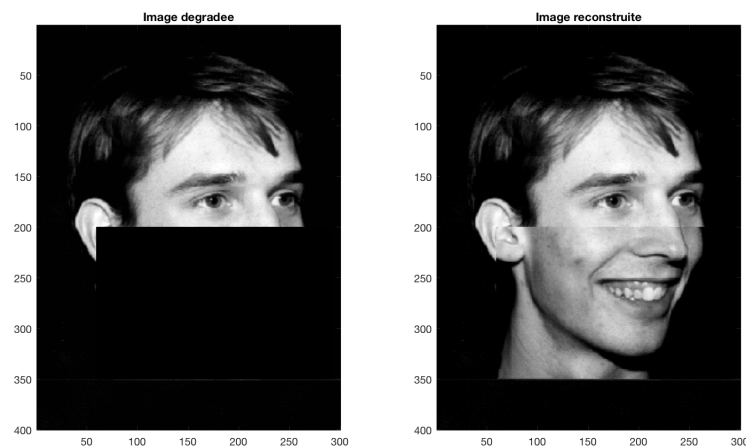
avec $d(\mathbf{c}(\mathbf{x}_i), \mathbf{c}(\mathbf{x})) = \|\mathbf{c}(\mathbf{x}_i) - \mathbf{c}(\mathbf{x})\|_2$.

Cette méthode de classification est appelée aussi *k-plus proches voisins* et sera l'objet du TP4 d'Analyse de Données.

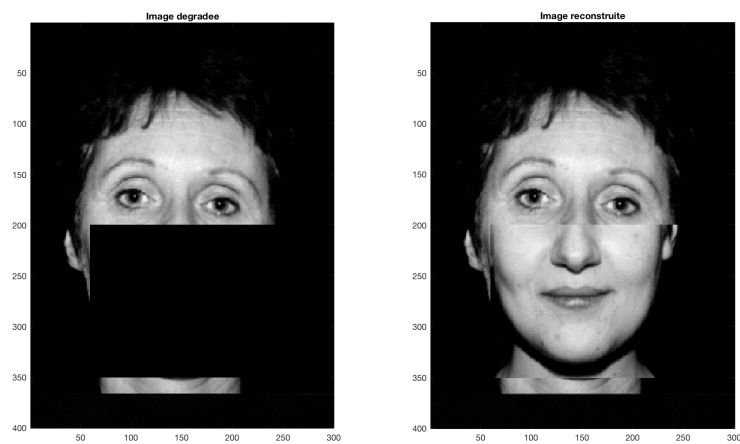
La reconstruction peut être réalisée en remplaçant la zone correspondant au masque par la zone de l'image de la base d'apprentissage de visages entiers la plus proche dans l'espace défini par les eigenfaces masqués.

Voici deux exemples de reconstruction d'images après reconnaissance :

- personne présente dans la base d'apprentissage mais avec une posture/émotion différente :



— personne ne se trouvant pas dans la base d'apprentissage :



4 Bibliography

- [1] G. W. Stewart. *Matrix Algorithms : Volume 2, Eigensystems*. Society for Industrial and Applied Mathematics (SIAM), 2001.
- [2] Matthew Turk and Alex Pentland. Eigenfaces for recognition. *J. Cognitive Neuroscience*, 3(1) :71–86, 1991.