

Mini Projet - A quoi tu ressembles sans ton masque ?

Avec la pandémie du 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.

L'opération de retrouver le visage complet 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 l'ACP et ses *eigenfaces* permettent, entre autres, de restaurer la zone dégradée.

Ce projet s'inspire d'un article intitulé *Eigenfaces for recognition*, écrit par Turk et Pentland et publié dans le *Journal of Cognitive Neuroscience* en 1991¹. 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.

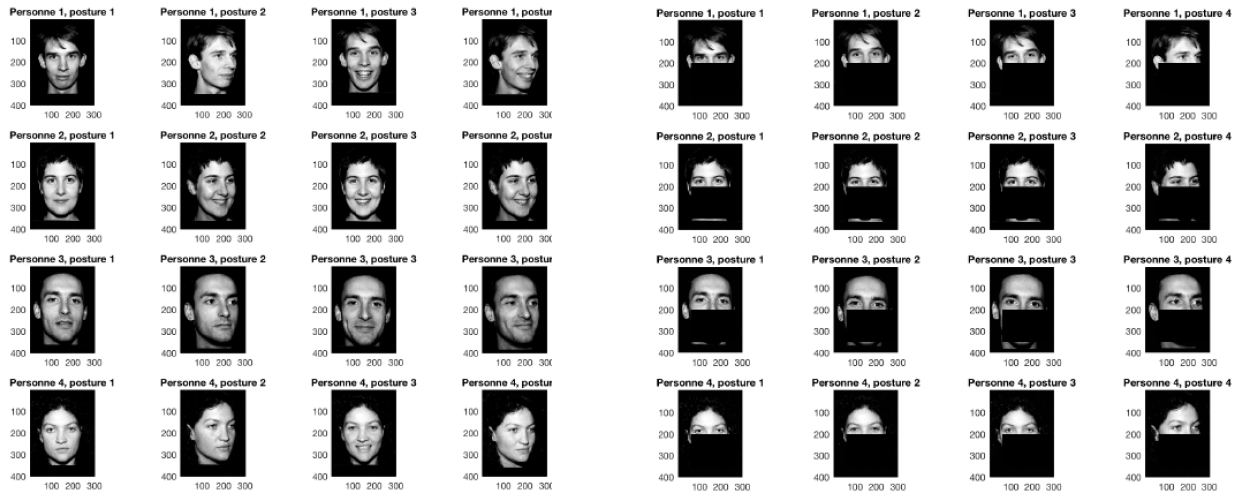


FIGURE 1 – Une base de visages et de visages masqués

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)*)². Elle comporte 32 personnes (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.

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

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

Description des données

Pour votre base d'apprentissage, vous disposez de n images de visages d'un ensemble d'individus. Chaque individu est photographié sous le même nombre de postures faciales (face, trois quart face, avec trois émotions). Chacune de ces n images en niveaux de gris est stockée dans une matrice bidimensionnelle de taille 300×400 . Ces n images constituent les *images d'apprentissage*. En les vectorisant, vous pouvez donc représenter ces images par des vecteurs colonnes de \mathbb{R}^p , où $p = 300 \times 400 = 12000$ est le nombre de pixels commun à toutes les images. Alors que dans le TP1, chaque pixel d'une image couleur constitue un point de \mathbb{R}^3 , ici c'est chaque image qui constitue un point d'un espace affine \mathbb{R}^p de dimension très élevée.

La matrice des données de la base d'apprentissage X , de taille $n \times p$, contient sur chaque ligne la transposée d'une image vectorisée.

1 Les «Eigenfaces» par analyse en composantes principales

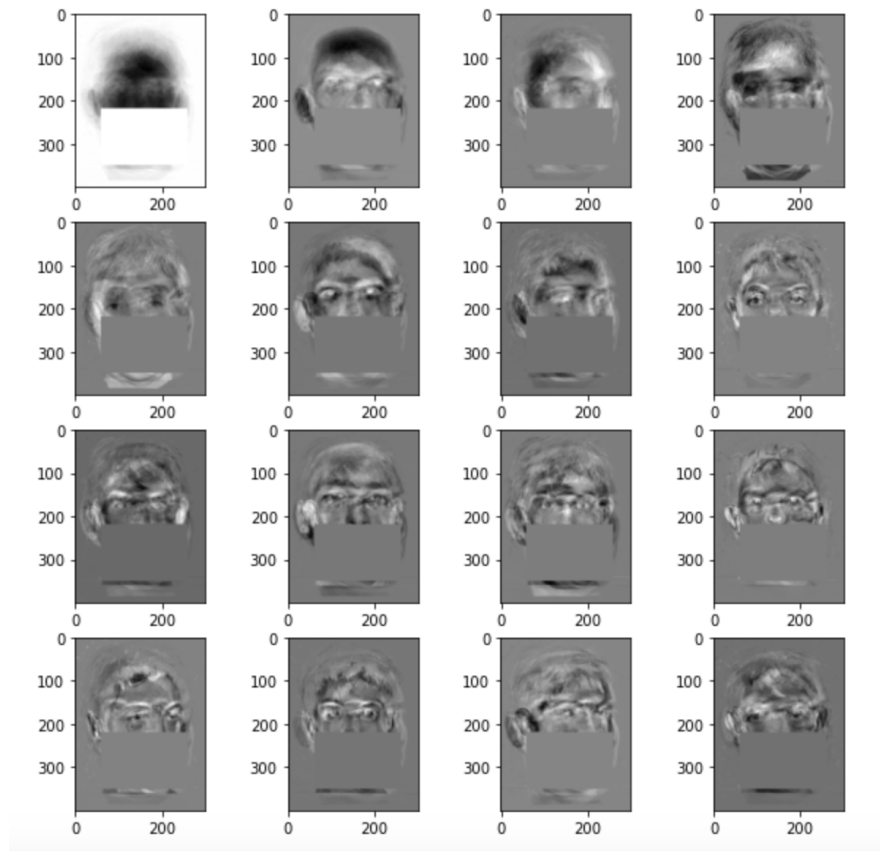


FIGURE 2 – Les «eigenfaces»

Question : Complétez la fonction `eigenfaces`, qui vise à calculer les axes principaux des images d'apprentissage à partir des vecteurs propres associés aux $n - 1$ valeurs propres non nulles de la matrice de variance/covariance Σ des données. Ces axes principaux sont appelés *eigenfaces* (FIGURE 2) par Turk et Pentland, par contraction des mots anglais *eigenvectors* et *faces*.

Remarque : la fonction `np.linalg.eig` ne peut pas être directement appliquée à Σ . En effet, sa taille $p \times p$ est gigantesque ($p = 120000$). Or, pour une matrice M quelconque, $M^\top M$ et MM^\top ont les mêmes valeurs propres *non nulles*. On peut donc appliquer la fonction `eig` à $\Sigma_2 = X_c X_c^\top / n$, de taille $n \times n$ beaucoup plus petite, pour calculer les valeurs propres non nulles de Σ .

2 Reconnaissance et reconstruction

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 la plus similaire au visage masqué et de permettre une reconstruction de la zone du masque (en faisant le lien avec le visage sans masque correspondant).

Pour réaliser cette reconnaissance, 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 de la matrice Σ correspondant aux images masquées

2.1 Reconnaissance

A partir des composantes principales de la base des visages avec masque, on a défini ainsi $\mathbf{c}(\mathbf{x})$, une représentation de \mathbf{x} , image requête (ou de test), dans cette base :

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

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 \mathbf{x}_i de la base dont les représentations compactes sont les plus proches de celle associée à l'image requête \mathbf{x} , ce qui revient à minimiser la fonction $h(\mathbf{x})$ définie par :

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

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

La fonction $h(\mathbf{x})$ est aussi appelée **classifieur**. Elle représente ici la méthode des 1-plus proche voisin vue lors du TP2. Ce classifieur peut être étendu à d'autres méthodes comme la version générale des k -plus proches voisins que vous pouvez réutiliser pour cette partie du projet.

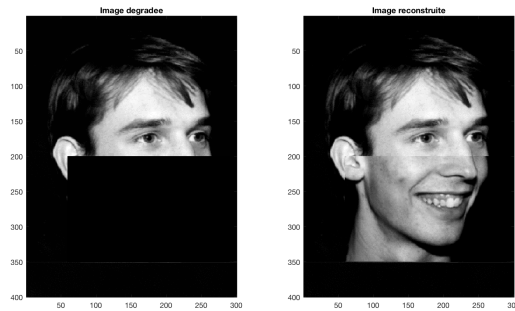
Question : En tirant aléatoirement une image de test, parmi les 32 personnes et les 6 postures faciales disponibles dans la base de données complète, complétez une fonction pour trouver l'individu (personne+posture) dans la base d'apprentissage qui est le plus proche de l'image de test (vous pouvez utiliser et adapter la fonction `kppv` que vous avez écrite lors du TP2).

2.2 Reconstruction

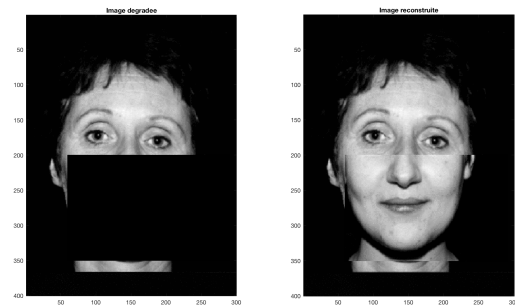
Question : A partir de la question précédente, implémentez la reconstruction de la zone du masque en remplaçant la zone correspondant au masque par la zone de l'image de la base d'apprentissage de visages entiers la plus proche dans l'espace défini par les eigenfaces masqués.

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 :



2.3 Evaluation

Pour évaluer la reconnaissance du classifieur, vous utiliserez les images reconstruites correspondant à plusieurs requêtes.

En réutilisant les outils qui vous ont été fournis ou que vous avez développés dans les sections précédentes, imaginez une évaluation de la reconstruction.

3 Travail demandé

- Réalisez une étude sur la reconnaissance et la reconstruction des zones masquées
- Une archive comportant le notebook et un rapport au format pdf de 5 pages maximum sera à rendre sous moodle.
- Ce travail est à réaliser en binomes.
- Deadline : 10 novembre