

TP2: Implémentation d'une cascade de régresseurs pour l'alignement de points caractéristiques du visage

Fait par: Imène TARAKLI

Master ISI - 2020/2021

Objectif

L'objectif de ce TP est de réaliser une méthode d'alignement de points caractéristiques du visage à partir d'un régrésseur en une première partie puis d'une cascade de régresseurs en une seconde partie.

1 Préparation des données

La base de données fournie est constituée de 4 sous-ensembles de données: Helen, AFW, LFPW et IBug. Elle comprend des images de visages ainsi que leurs annotations (figure 1).

Cette base de données est préalablement séparée en données d'apprentissages (3180 visages) et données de test (330 visages).

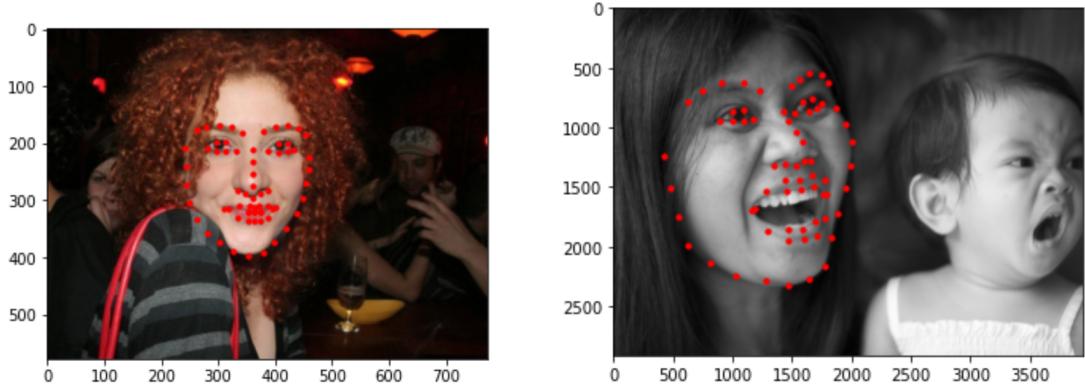


Figure 1: Exemples de visage avec leurs points caractéristiques.

1.1 Augmentation des données

Cette étape consiste à augmenter de manière artificielle le volume de données à disposition pour l'entraînement du regresseur. Cela revient à générer pour une même image divers modèles de points caractéristiques.

Pour chaque image de la base de données, on extrait la boite englobante des points caractéristiques du visage que l'on élargit de 30%. A partir de la boite élargie, on découpe l'image puis on la re-dimensionne en une taille de 128*128. La figure 2 schématisse ce processus.

Les points caractéristiques de la nouvelle image sont ensuite calculés en appliquant une translation des points fournis afin de ramener la boite englobante vers l'origine puis une mise à échelle pour que les grandeurs soient uniformes avec ceux de l'image croppée. Ces coordonnées constituent la vérité terrain de notre modèle. Ces étapes sont décrites dans la figure 3.

Pour que ces nouveaux points soient correctement positionnés sur l'image croppée réduite, il est primordiale de bien vérifier que la boîte englobante élargie se trouve bien dans l'image. Si ce

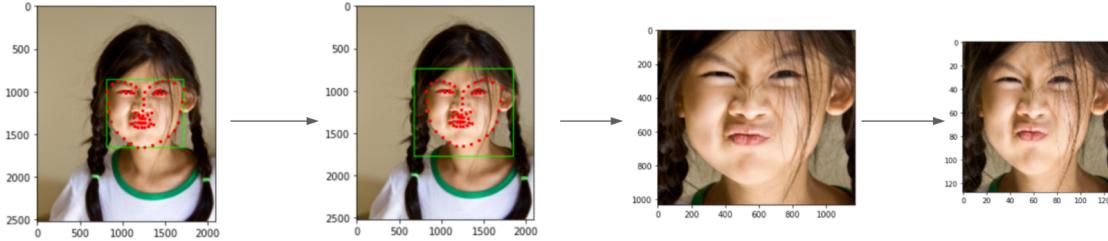


Figure 2: Pré-traitement des images

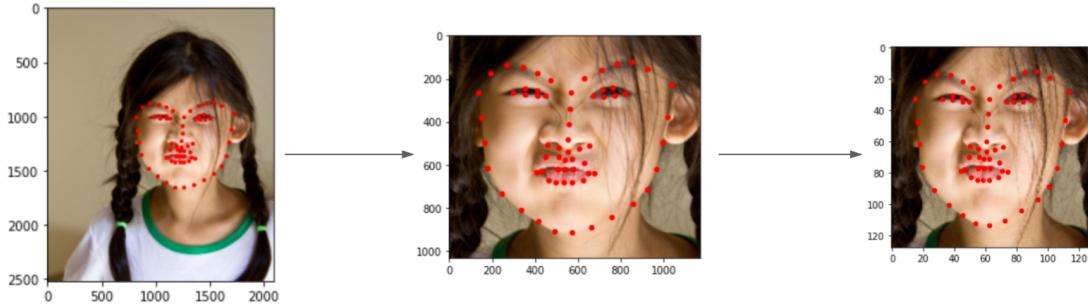


Figure 3: Coordonnées des points caractéristiques.

n'est pas le cas, il est nécessaire de réduire les dimensions de la boîte. La figure 4 illustre ceci.

Après avoir effectué ce traitement sur l'ensemble de la base de données, on calcule la position moyenne des points caractéristiques sur l'ensemble d'apprentissage. On génère par la suite 10 perturbations de ce modèle moyen. Ces perturbations sont obtenues en appliquant des translations et un facteur d'échelle de sorte à ce que le régresseur soit robuste par rapport à une imprécision de la fenêtre de détection de visage.

L'objectif de cette procédure est aussi d'augmenter la taille des données d'apprentissage. En effet, chaque image aura 10 modèles moyens perturbés. On aura donc au final $N*10$ données d'apprentissage, N étant la taille initiale.

2 Apprentissage d'un régresseur simple

2.1 Extraction de caractéristiques image

Cette étape consiste à extraire les caractéristiques qui serviront d'entrées au régresseur.

Il n'est pas possible d'utiliser de manière directe les pixels des images car les seuls points d'intérêt sont les points caractéristiques et non pas l'image entière. Ainsi, il est nécessaire d'extraire des images locales autour de ces points.

Dans ce TP, les descripteurs SIFT ont été proposés. Au voisinage de chaque point caractéristique, une fenêtre de 20 pixels est considérée depuis laquelle est calculé le descripteur SIFT.

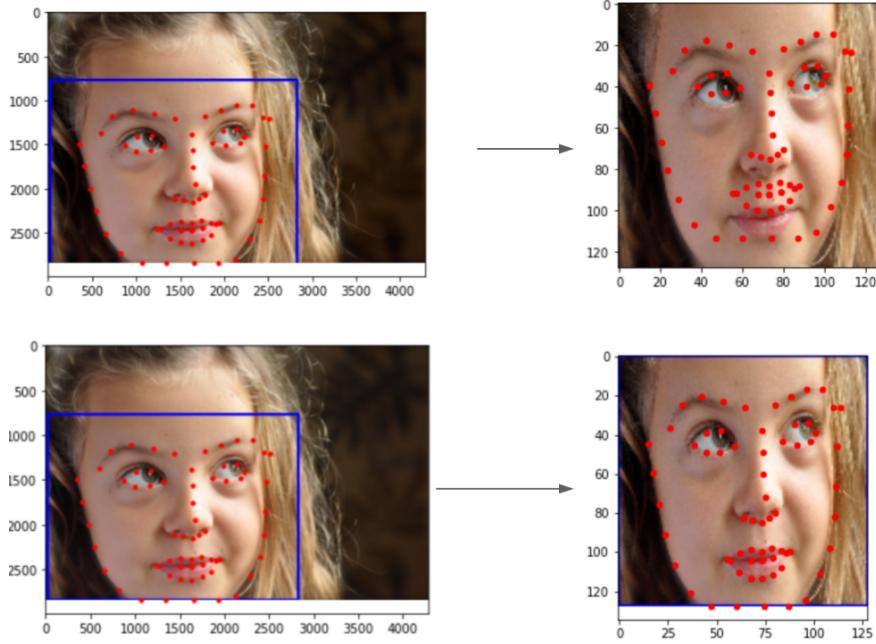


Figure 4: Coordonnées des points caractéristiques avant et après dimensionnement de la boîte englobante.

Le(descripteur d'un seul point caractéristique est un vecteur de taille 128. Cela est obtenu en divisant la fenêtre considérée en 4×4 échantillons et prenant en compte 8 directions par échantillon ($4 \times 4 \times 8 = 128$). La figure 5 met en avant cette procédure (source: A Short introduction to descriptors)

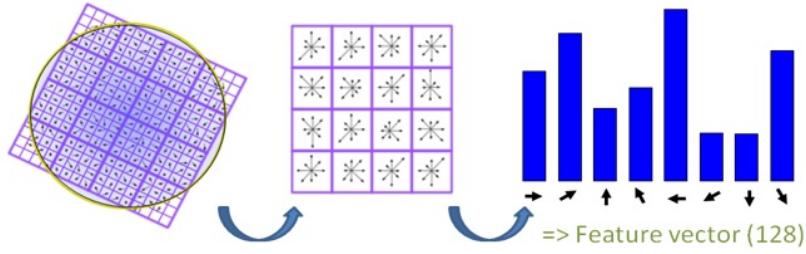


Figure 5: Extraction des descripteurs SIFT.

Après concaténation des descripteurs de chaque point caractéristique, la dimension finale du descripteur de chaque image est (8704,), soit le nombre de points (68) multiplié par la taille d'un descripteur.

Les descripteurs de chaque image sont ensuite regroupés pour former une matrice X_0 de dimension (31480, 8704).

2.2 Réduction de dimensionnalité

En machine learning, l'intérêt principal de la réduction de la dimensionnalité est d'éviter la malédiction de la dimension. En effet, quand la dimension des features est trop importante, le modèle devient de plus en plus complexe et plus il y a probabilité d'avoir de l'overfitting.

Plusieurs méthodes de réduction existent. L'Analyse en Composantes Principales (APC) permet de réduire les features en ne gardant que les variables décorrélées. L'analyse factorielle permet aussi de réduire la taille originale en exprimant la matrice complète comme le produit de deux matrices beaucoup plus petites et en y rajoutant un bruit aléatoire.

Après application d'une APC, et cela en conservant 98% de la variance totale, la matrice X_0 est réduite en une matrice \hat{X}_0 de dimension (31480, 392).

2.3 Estimation du déplacement

L'objectif de cette partie est d'estimer le déplacement du modèle moyen des points de caractéristique vers la vérité terrain. Cela revient à estimer une régression linéaire permettant de prédire le déplacement δ_s^0 :

$$\delta_s^0 = \operatorname{argmin}_{R_0} \left\| \delta_s^* - R_0 \tilde{\mathbf{Y}}_0 \right\|^2 \quad (1)$$

Où : δ_s^* est le déplacement optimal entre la VT et le modèle moyen s_0 et $\tilde{\mathbf{Y}}_0$ est la matrice \hat{X}_0 augmentée de 1 en colonne.

Le regresseur linéaire R_0 minimisant l'erreur au sens des moindres carrés est obtenu comme suit:

$$R_0 = \left(\tilde{\mathbf{Y}}_0^T \tilde{\mathbf{Y}}_0 \right)^{-1} \tilde{\mathbf{Y}}_0^T \delta_s^* \quad (2)$$

Le déplacement estimé est finalement obtenu en calculant $R_0 \tilde{\mathbf{Y}}_0$.

Finalement, le modèle d'estimation peut être résumé comme dans la figure 6.



Figure 6: Modèle d'estimation basé sur le regresseur.

En appliquant ce regresseur sur la base d'apprentissage, on obtient une erreur moyenne de : 1.22e-14. La figure 7 montre les différents résultats obtenus pour les 10 modèles moyens perturbés sur la première image de la base d'apprentissage.

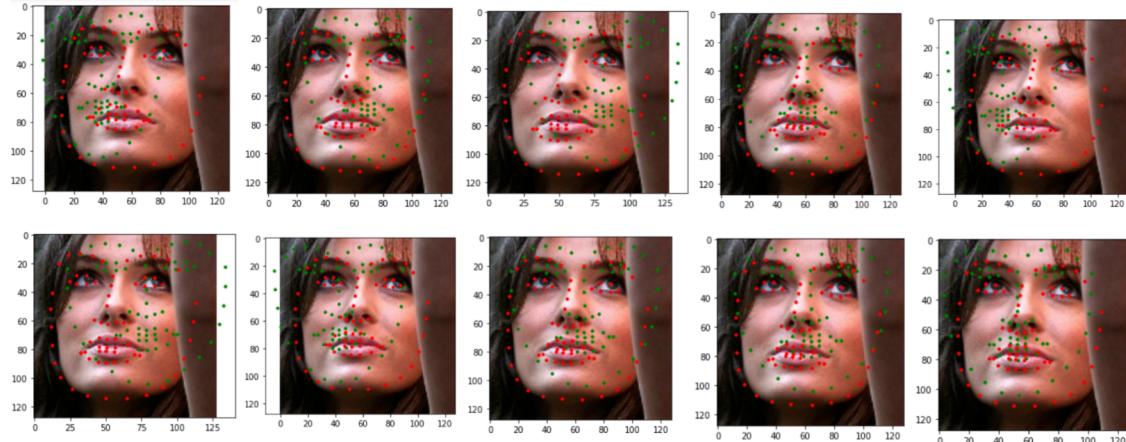


Figure 7: Résultats d'un regresseur simple sur une donnée d'apprentissage.

2.4 Validation sur un ensemble de test externe

Pour évaluer la capacité de généralisation de la procédure d'alignement de points caractéristiques, l'évaluation précédente n'est pas pertinente car elle repose sur le modèle moyen calculé des images d'apprentissage. Les déplacements estimés seront donc moins précis.

En appliquant le régresseur simple sur les données de test, une erreur moyenne de 0.33. Les résultats obtenus sont dans la figure 8.

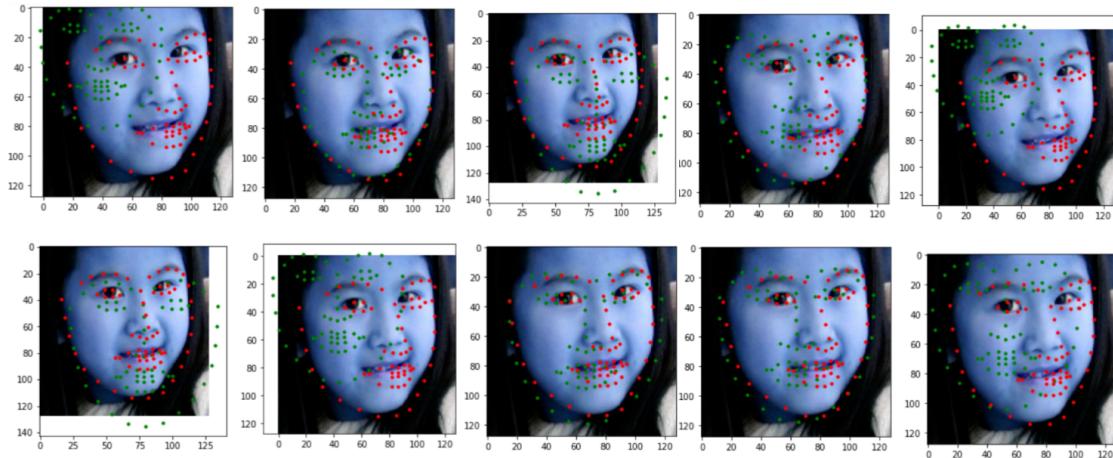


Figure 8: Résultats d'un regresseur simple sur une donnée de test.

3 Cascade de régresseurs

3.1 Itération du processus d'alignement

Dans cette partie, le processus vu précédemment va être itéré comme le montre la figure 9.

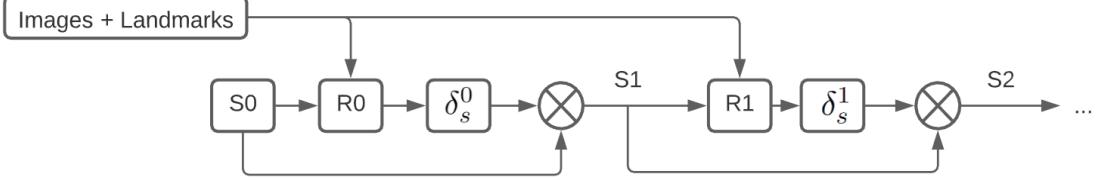


Figure 9: Cascade de régresseurs.

La figure 10 montre les résultats des 5 modèles itérés obtenus sur la première image de la base d'apprentissage. A la première itération, le déplacement estimé est assez important. Ce dernier devient de plus en plus faible au fur des itérations car il s'agit de petit ajustements. Finalement, le modèle obtenu, après 5 itérations, est très proche des landmarks d'origine comme le montre la figure 11.

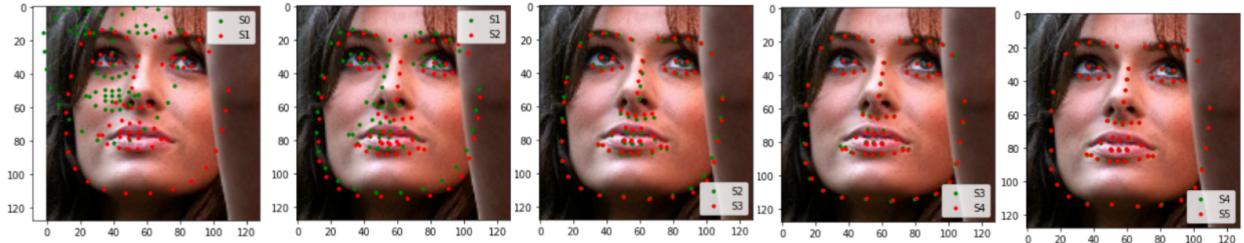


Figure 10: Résultats sur les données d'apprentissage.

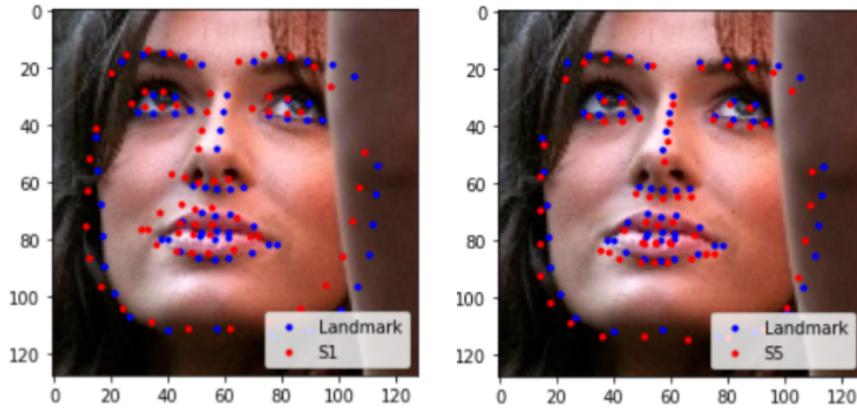


Figure 11: Comparaison entre le régresseur simple et la cascade de régresseurs.

La figure 12 montre les résultats des itérations obtenus sur la base de données de test. Le régresseur simple généralise mal l'estimation de déplacement mais cela est corrigé avec l'itération. En effet, et comme le montre la figure 13, le modèle obtenu après 5 itérations se rapproche des landmarks d'origines.

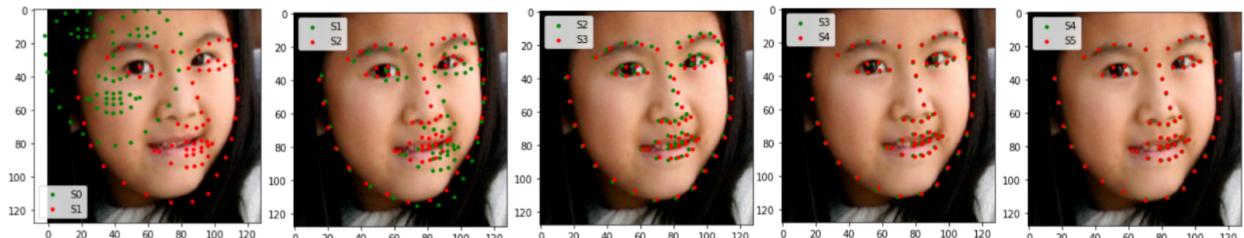


Figure 12: Résultats sur les données de test.

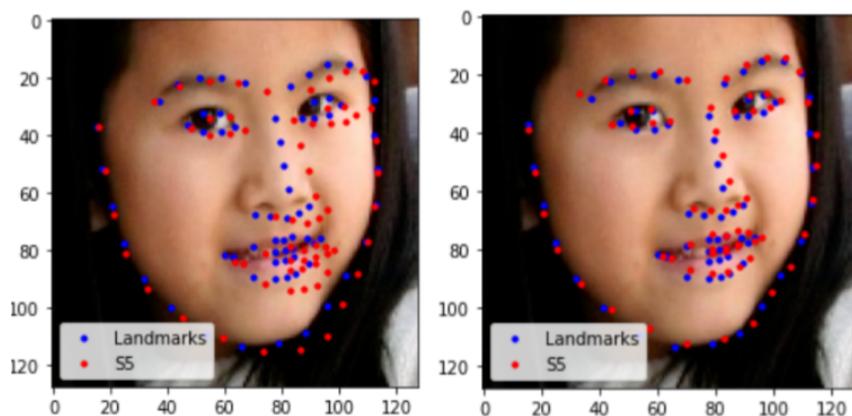


Figure 13: Comparaison entre le régresseur simple et la cascade de régresseurs.