TP – Protection et attaque de données biométriques

Partie 1 : Exploiter des données biométriques.

Vous avez à disposition une base de données biométriques FVC.csv issue d'empreintes sous la forme CSV. Elle est composée de 100 individus avec 8 images par individu dont les caractéristiques ont été extraites par filtres de Gabor sous forme d'un vecteur de 512 valeurs réelles par images.

1 – Écrire une fonction permettant de lire le fichier CSV et de stocker les vecteurs de caractéristiques.

2 – Calculer l'EER et le seuil correspondant en utilisant la distance Euclidienne.

Partie 2 : Protection de données biométriques.

L'objectif de cette partie est de protéger ces données biométriques en garantissant 4 propriétés : l'indistinguabilité, la révocabilité, la non-inversibilité, et la performance.

3 – Écrire une fonction générant une matrice de projection N x M depuis une graine S initialisant un générateur pseudo-aléatoire. N est la taille du vecteur de caractéristiques à protéger, M la taille du gabarit produit. Les coefficients de la matrice seront :

$$\begin{cases} 1/\sqrt{M} \text{ avec probabilité } 1/2 \\ -1/\sqrt{M} \text{ avec probabilité } 1/2 \end{cases}$$

4 – Protéger les données biométriques en fixant M à 64, 128, 256, 512. Pour chaque valeur, calculer l'EER et le seuil correspondant. Que constatez-vous ?

Partie 3 : Attaque de données biométriques.

L'objectif de cette partie est d'attaquer la propriété de non-inversibilité de la protection. La protection de données biométriques doit permettre la bonne reconnaissance des individus, et ainsi les distances entre l'espace de départ et l'espace d'arrivée sont **globalement** conservées. C'est ce qu'on appelle une fonction de hachage localement sensible (LSH). Vous allez développer un algorithme d'optimisation (génétique) permettant d'attaquer un schéma biométrique.

Un algorithme génétique se déroule en plusieurs itérations, partant d'une population initiale de vecteurs aléatoires. À chaque itération, la moitié de la population précédente est sélectionnée et conservée (les vecteurs ayant le meilleur score, ici la proximité engendrée avec le gabarit cible), et se reproduit pour compléter l'autre moitié la population (par croisement en un point des vecteurs), en appliquant un faible taux de mutation aux enfants.

5 – Développez, avec un algorithme génétique, une attaque permettant de construire, depuis un gabarit, un vecteur de caractéristiques qui s'il est protégé avec la même graine permet d'être authentifié avec ce gabarit. Visualisez l'évolution de l'attaque (distance par itération, authentification ou non). Effectuez cette attaque avec les différentes valeurs de M : que constatez-vous ?

6 – Effectuez la même attaque, mais avec plusieurs gabarits cibles : l'objectif est d'en usurper un maximum ! Effectuez cette attaque avec les différentes valeurs de M : que constatez-vous ?