Introduction à l'authentification par biométrie

Sébastien Gambs

sgambs@irisa.fr

11 janvier 2010

Authentification par biométrie

Authentification sans stocker le profil biométrique en clair

Authentification par biométrie

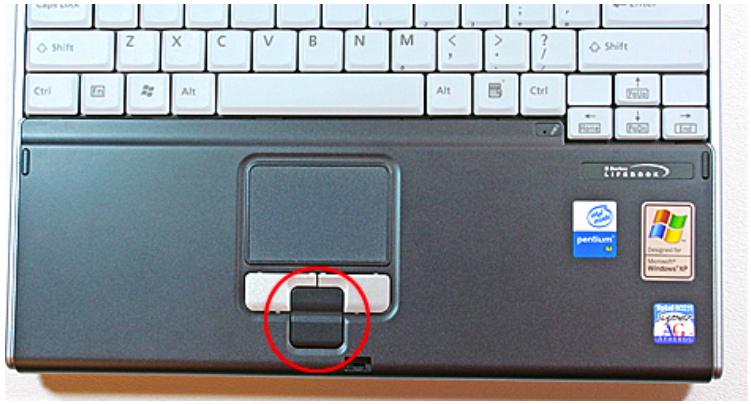
Biométrie

- Biométrie : analyse statistique des données biologiques d'un individu.
- L'*authentification par biométrie* consiste à utiliser un système de reconnaissance basé sur les caractéristiques physiques ou comportementales d'un individu pour vérifier son identité.
- Exemples : empreinte digitale, iris, visage, voix.
- Deux phases importantes :
  1. Capture des données biométriques : enregistrement de l'utilisateur en capturant un modèle de ses caractéristiques biométriques (par exemple son empreinte digitale).
  2. Vérification de son identité : le système capture des données et les compare au modèle.
     $\Rightarrow$ s'il y a correspondance l'utilisateur est authentifié
- Il est possible aussi de faire de la reconnaissance (identifier une personne parmi tous les enregistrements de la base).

Exemples d'utilisation de l'authentification par biométrie

- Contrôle d'accès à des locaux ou bâtiment.

- Authentification pour utiliser un ordinateur.



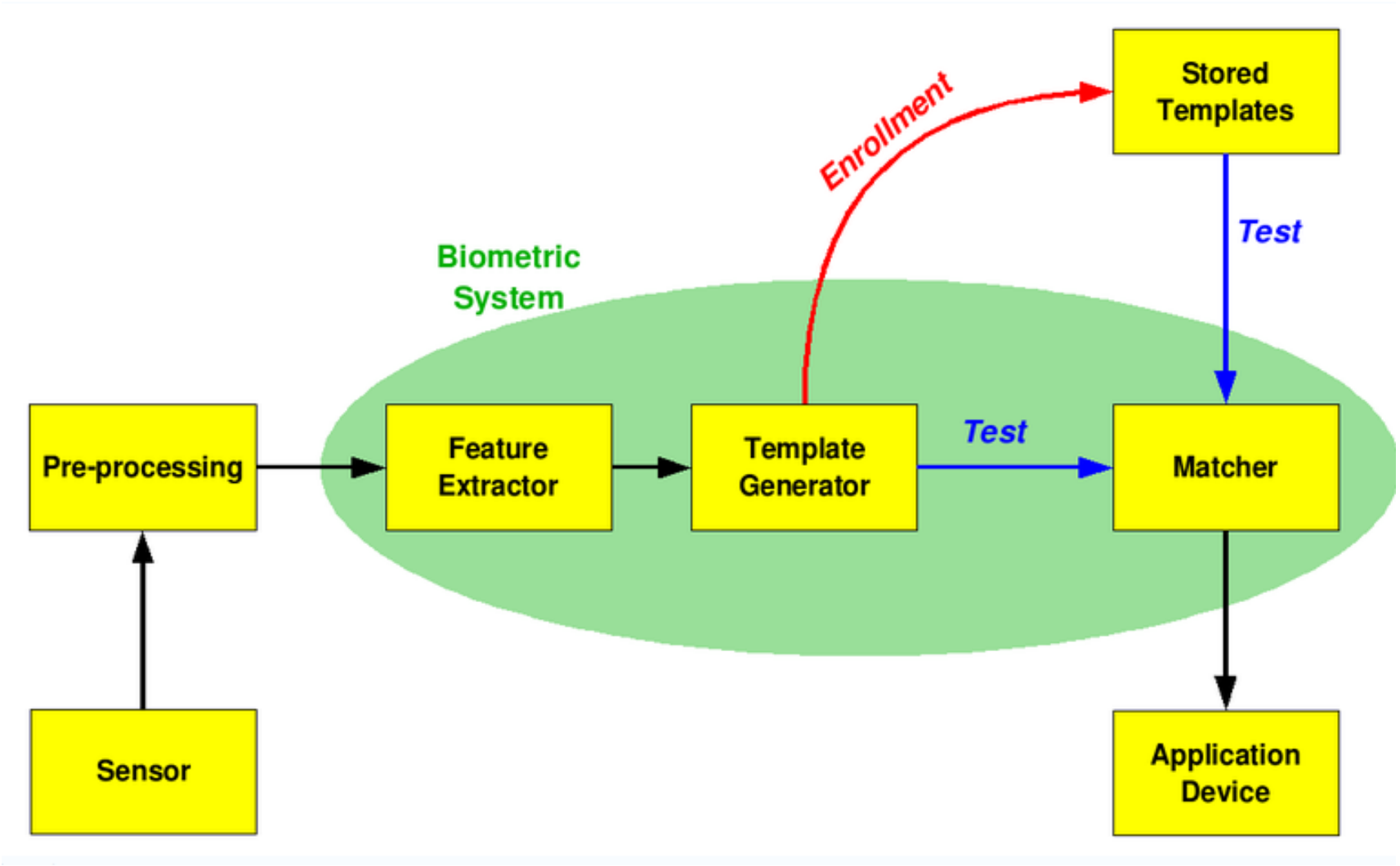
Caractéristiques des données biométriques

- Les données biométriques peuvent naturellement varier légèrement d'une fois à l'autre.
- Exemple : empreinte de voix différente à cause d'un mal de gorge ou iris dilaté différemment à cause de la prise de médicament.
- Il faut être capable de prendre en compte cette variabilité naturelle à l'intérieur du système d'authentification par biométrie.

Méthode d'identification des données biométriques

- Capture (étape 1) : on acquiert b_i le profil biométrique de l'utilisateur d'index i.
- La représentation de b_i dépend de la méthode biométrique utilisée (cela peut aller d'un vecteur de bits à une représentation beaucoup plus complexe).
- Identification (étape 2) : on mesure b', les données biométriques fraîches d'un utilisateur qui souhaite s'authentifier comme étant l'utilisateur d'index i.
- But : vérifier si b' est suffisamment similaire de b_i qu'on a stocké en mémoire.
- Si c'est le cas, on *accepte* l'utilisateur, sinon on le *rejette*.

Fonctionnement de l'authentification par biométrie



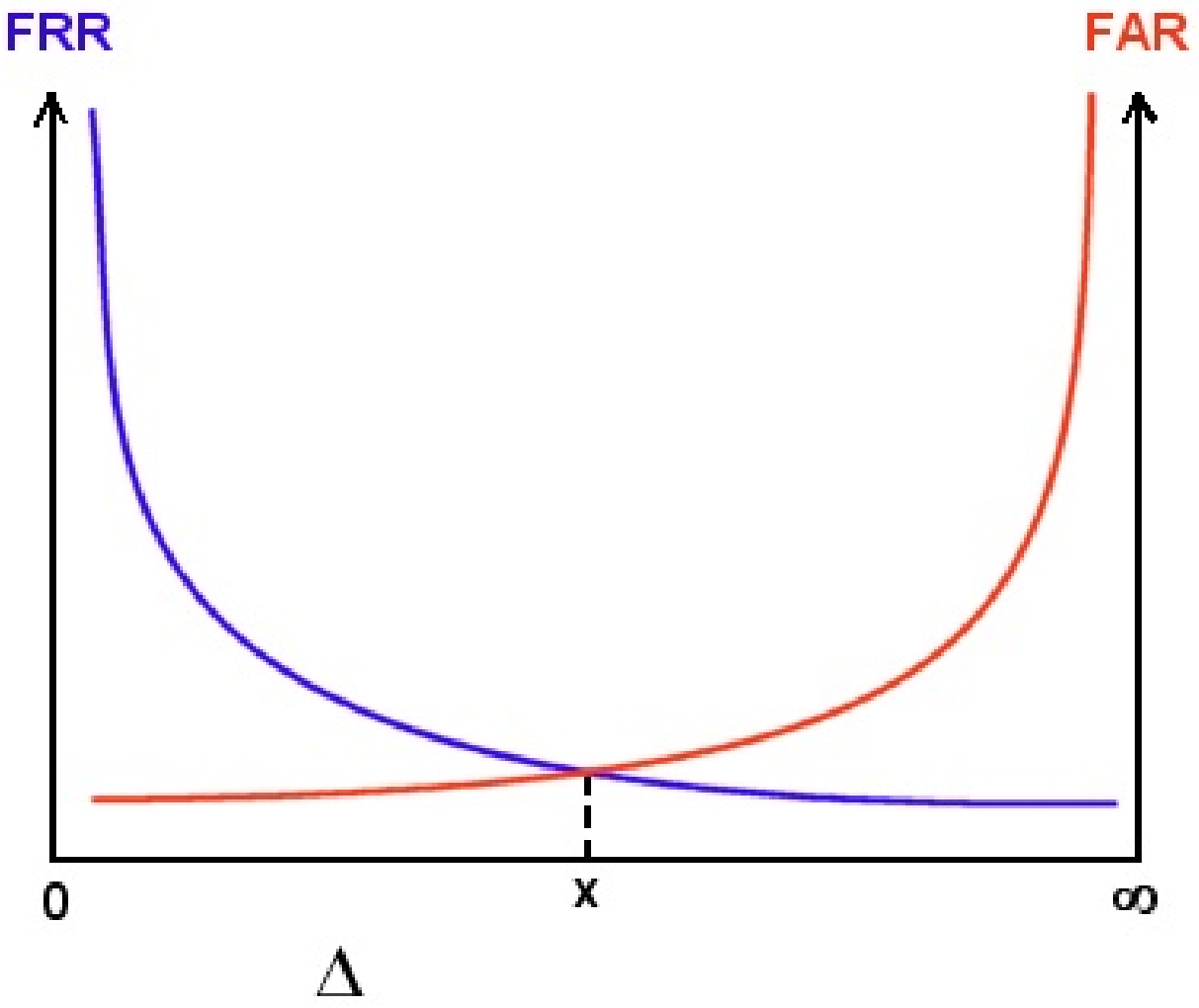
Distance entre profils biométriques

- Soit $dist(b, b')$ une mesure de distance entre deux profils biométriques b et b'.
- Exemples de mesure de distance : distance de Hamming ou distance dédition.
- Si les deux profils correspondent parfaitement si $dist(b, b') = 0$.
- En général on fixe un seuil θ tel que si $dist(b, b') < \theta$ alors on considère que les profils sont suffisamment similaires,
- sinon on considère l'authentification comme non-réussie.
- La valeur du seuil θ va dépendre de la méthode biométrique utilisée et de la variabilité naturelle à l'intérieur de la population.
- Il est possible d'utiliser une mesure de similarité au lieu d'une mesure de distance.

Erreur d'identification

- Deux types principaux d'erreurs d'identification : faux négatifs et faux positifs.
- Faux positifs : reconnaissance à tort d'un individu qui n'aurait pas du être reconnu.
- Mesuré par le taux de fausse acceptation (*FAR* pour *False Acceptance Rate* en anglais).
- Faux négatifs : non-reconnaissance d'un individu qui aurait du être reconnu.
- Mesuré par le taux de faux rejet (*FFR* pour *False Rejection Rate* en anglais).
- Baisser un des deux taux augmente l'autre.
- Exemple : on rejette toute identification où la correspondance n'est pas parfaite (baisse le FAR mais augmente le FFR).
- Une bonne méthode d'identification cherche à trouver un compromis efficace pour avoir deux taux bas.

Compromis entre FAR et FRR



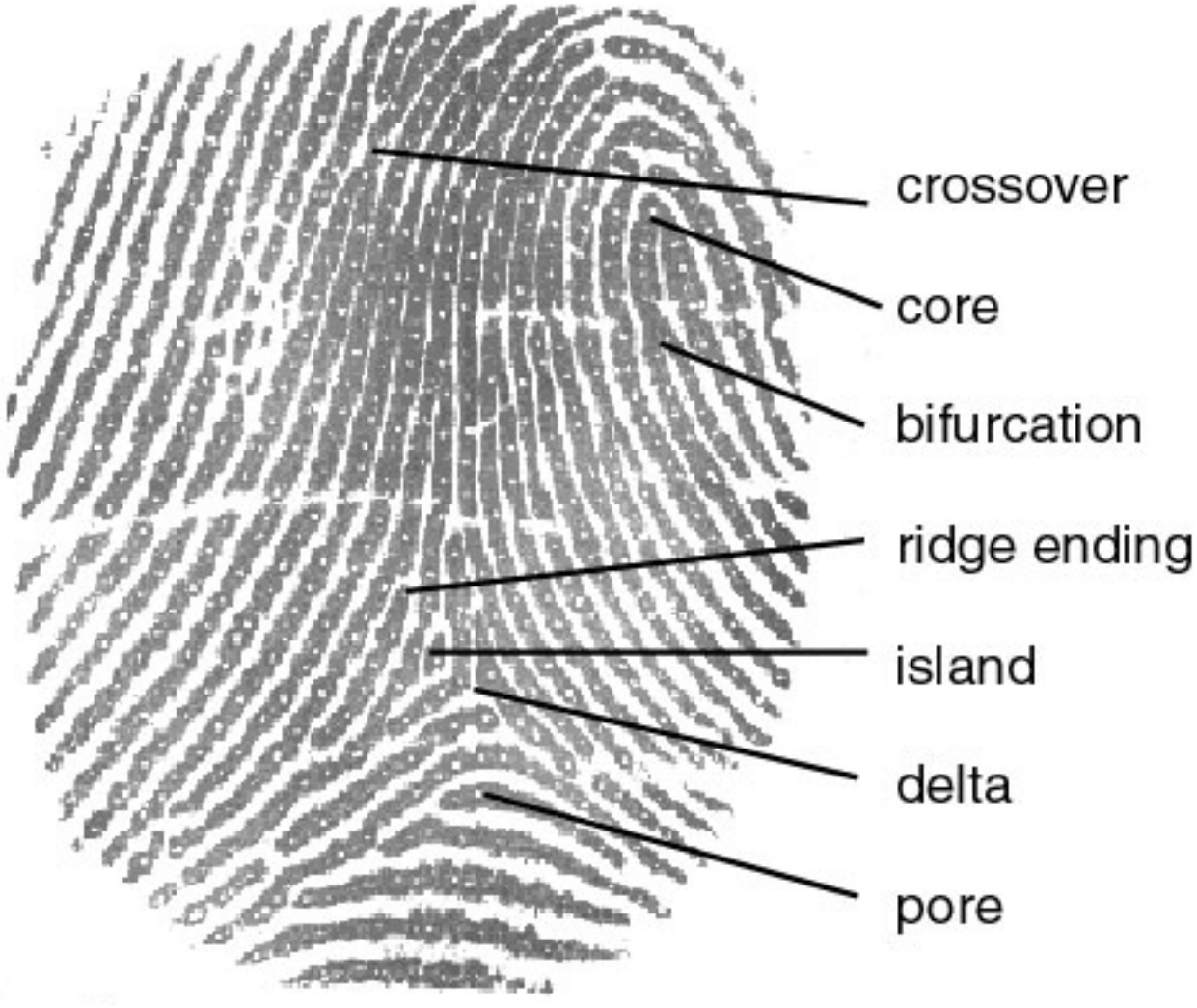
Sécurité des données biométriques

- Système sécuritaire du moment qu'il n'est pas réaliste d'acquérir les données biométriques d'un individu.
- Exemple : facile pour les empreintes digitales, je prends un café avec vous et je récupère votre gobelet ensuite.
- Conseil : pour plus de sécurité, utiliser la biométrie en combinaison avec autre méthode d'authentification (par exemple connaissance d'un mot de passe ou jeton cryptographique).

Empreinte digitale

- Une des plus anciennes formes d'authentification biométrique.
- Représente les caractéristiques d'une empreinte digitale sous forme de points caractéristiques appelés *minuties*.
- La probabilité de trouver deux individus avec des empreintes similaires est de 1 sur 10^{24}.
- Remarque : les jumeaux ont des empreintes très proches mais non semblables.
- Avantages : facile à mettre en place et à utiliser et donne un bon taux de détection.
- Inconvénients : il est relativement facile de copier les empreintes digitales d'une personne.
- Demande aussi à faire une nouvelle capture des données si l'utilisateur se coupe ou se brûle.

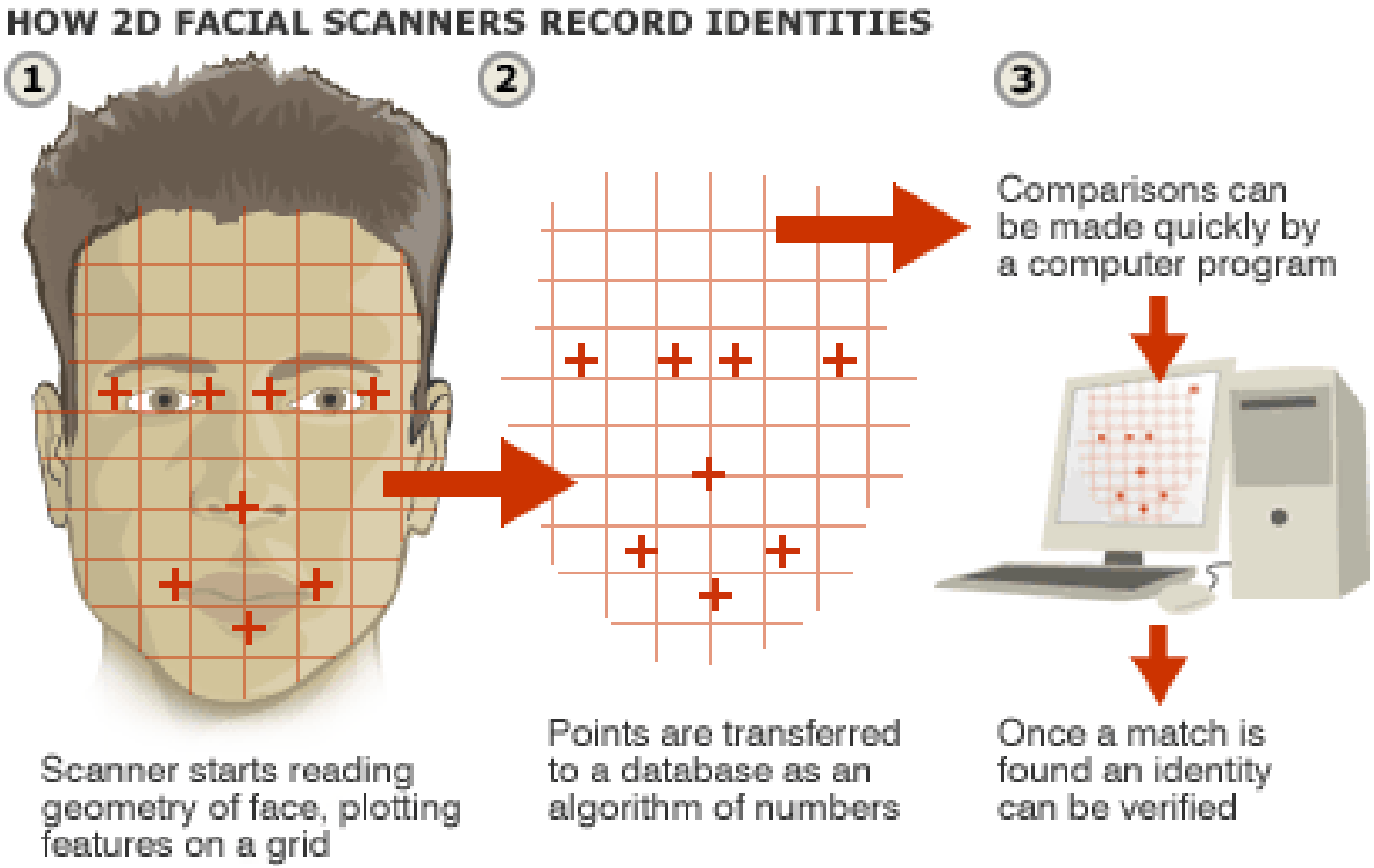
Illustration des empreintes digitales



Système de reconnaissance faciale et vocale

- Prend une empreinte (*fingerprint* en anglais) du visage ou de la voix d'un individu.
- Avantages : facile à utiliser tout comme les systèmes d'empreintes digitales.
- Inconvénients : peut être trompé par la photographie d'un visage ou l'enregistrement d'une voix.

Illustration de la reconnaissance faciale



Scan d'iris et de rétine

- Fait une capture de l'iris ou de la rétine.
- La probabilité de trouver deux individus avec des iris ayant des caractéristiques similaires est de 1 sur 10^{72}.
- L'identification peut faire jouer la lumière pour changer en temps réel l'image de l'iris mesurée.
- Avantages : méthode qui offre un des meilleurs taux d'identification et considéré comme étant un des plus sûrs du fait de la difficulté de faire une capture du comportement d'une iris.
- Même deux jumeaux possèdent un scan d'iris et de rétine différent.
- Inconvénient : procédure de capture plus contraignante.

Illustration du scan d'iris

Autres méthodes d'authentification biométriques

- Scan du réseau veineux : utiliser la structure du réseau veineux comme information biométrique.
- Pas encore tout à fait au point mais très prometteur comme technologie.
- L'authentification par biométrie s'intéresse aussi à identifier les individus par rapport à leur comportement.
- Dynamique des frappes au clavier : identification d'un individu par rapport à sa dynamique de frappe au clavier (par exemple durée entre frappes, fréquence des erreurs ou durée globale nécessaire pour taper un texte.
- Dynamique de signature : utilisée lors de l'acquisition de signatures par palette graphique en mesurant par exemple la vitesse ou encore la pression et les accélérations.

Authentification sans stocker le profil biométrique en clair

Données biométriques et identité

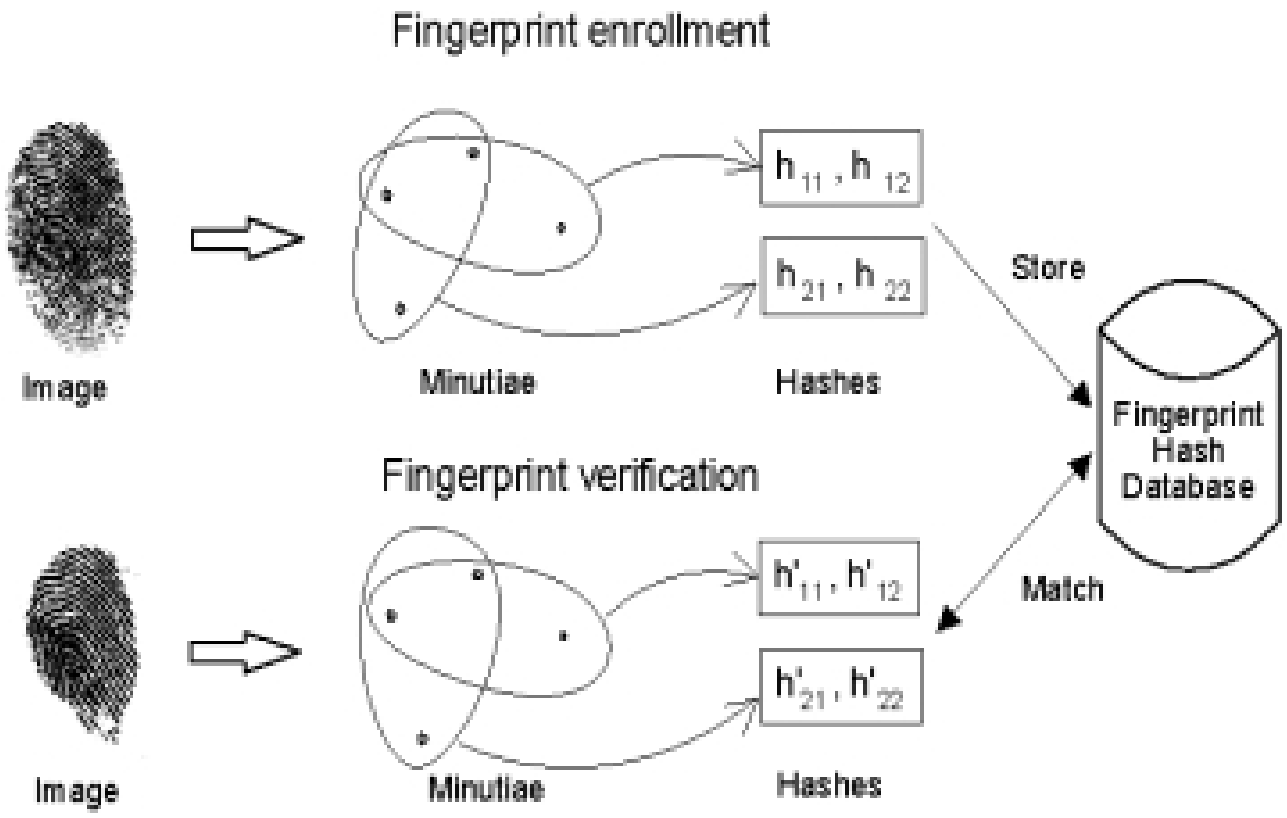
- Les données biométriques d'un individu lui étant propre, elles font partie de son identité au même titre que son nom ou sa date de naissance.
- Nécessité d'avoir un cadre législatif pour encadrer leur collecte et utilisation.
- Problème principal : comment stocker de façon sécurisé les données biométriques?
- Idéalement, on aimerait pouvoir faire de l'authentification par biométrie sans avoir à stocker les données biométriques en clair.
- Question : peut-on faire mieux que simplement en stocker une version chiffrée?

Cancellable biometrics

- Concept inventé par Ratha, Connell et Bolle en 2001.
- Idée principale : appliquer une distortion à un profil biométrique tel que :
  1. il n'est pas facile de reconstruire les données originales à partir de la version distordue,
  2. la transformation préserve la distance entre deux données biométriques proches auparavant.
- On peut ainsi stocker la version distordue sur le serveur ce qui joue le même rôle que le stockage de versions hachées des mots de passe.

Illustration des cancellable biometrics

Fonctionnement des cancellable biometrics



Fuzzy commitment

- Combine l'utilisation de codes correcteurs d'erreur et des fonctions de hachage.
- Soit b le profil biométrique d'un individu qu'on suppose pour être représenté sous la forme d'un vecteur binaire, c'est à dire $b \in \{0,1\}^n$.
- Remarque : en pratique cette hypothèse est relativement plausible pour des codes d'iris (par exemple de 512 octets) mais la méthode peut aussi être adapté à des codes fonctionnant sur de l'information géométrique (par exemple les empreintes).
- On choisit un code correcteur d'erreurs C de taille n qui est capable de corriger jusqu'à t erreurs où t est choisi empiriquement pour donner un bon compromis entre FAR et FRR.
- Une fonction de hachage universelle h est aussi utilisée.

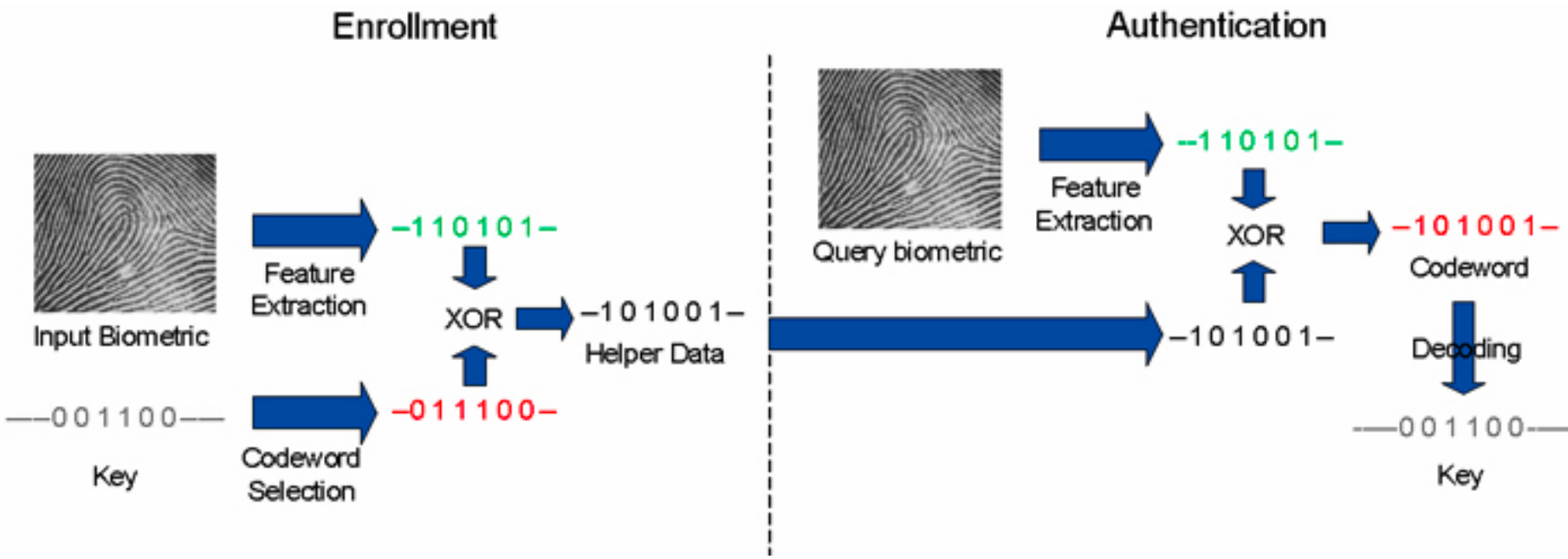
Fuzzy commitment (phase de capture)

- Le modèle biométrique b de l'utilisateur est mesuré par un scanner.
- Un mot de code c est choisi uniformément au hasard parmi tous les mots de code possibles de C et $z = c \oplus b$ est calculé.
- On calcule aussi $h(c)$ la valeur hachée de c.
- $h(c)$ et z seront gardées en mémoire par le serveur.
- Question : pourquoi ces valeurs sont elles sécuritaires du point de vue du profil biométrique?

Fuzzy commitment (phase de vérification)

- Soit b' le profil biométrique d'un utilisateur qui vient dêtre fraîchement mesuré.
- Le serveur envoie la valeur z au client qui peut calculé alors $z \oplus b'$ et le décoder vers le mot de code le plus proche c'.
- Le client calcule $h(c')$ et l'envoie au serveur.
- Le serveur considère que l'authentification est réussie si $h(c') = h(c)$ et rejette l'utilisateur sinon.

Fonctionnement des fuzzy commitments



Fuzzy extractors

- Cryptographique primitive fonctionnant sur le même principe que les fuzzy commitment.
- Permet d'extraire une chaîne aléatoire *rand* uniformément distribuée de taille l à partir d'un profil biométrique de taille n (pour $l < n$).
- Deux étapes :
  1. Génération : prend comme entrée un profil biométrique b et produit en sortie une chaîne aléatoire *rand* et une chaîne d'aide p.
  2. Récupération : prend comme entrée un profil biométrique b' proche du profil original ($dist(b, b') < t$) et la chaîne d'aide p et produit en sortie la chaîne aléatoire *rand*.
- Applications possibles : génération d'une clé à partir des données biométriques d'un individu qui peut servir pour le chiffrement ou pour l'authentification.