

# Évaluation de la qualité de données biométriques

Mohamad El-Abed\*, Baptiste Hemery\*  
Christophe Charrier\*,\*\* Christophe Rosenberger\*

\*Laboratoire GREYC, ENSICAEN - Université de CAEN - CNRS  
6 Boulevard Maréchal Juin, 14000 CAEN, France  
{mohamad.elabed, baptiste.hemery, christophe.rosenberger}@ensicaen.fr  
<http://www.ecole.ensicaen.fr/~elabed/>  
<http://www.ecole.ensicaen.fr/~hemery/>  
<http://www.ecole.ensicaen.fr/~rosenber/>

\*\*Laboratoire XLIM-SIC, Université de Poitiers  
Bat. SP2MI, BP. 30179, 86962 Futuroscope Chasseneuil cedex, France  
christophe.charrier@sic.univ-poitiers.fr  
<http://greyc.stlo.unicaen.fr/charrier/>

**Résumé.** L'évaluation de la qualité des données biométriques est un facteur primordial dans le processus biométrique. Dans cet article, nous proposons une méthode générique pour évaluer la qualité des données biométriques morphologiques. Elle est basée sur l'utilisation conjointe de deux types d'informations: 1) la qualité de l'image, et 2) la qualité des paramètres extraits en utilisant le descripteur *Scale Invariant Feature Transformation* (SIFT). Cinq bases de données (quatre de visages et une d'empreintes digitales), et un système d'authentification biométrique ont été utilisés pour quantifier les performances de la méthode proposée. Les résultats expérimentaux montrent l'intérêt de la méthode proposée pour détecter plusieurs types d'altérations réelles des données, qui ont un impact majeur sur la performance globale des systèmes biométriques. Les résultats expérimentaux montrent également que la méthode proposée est plus efficace que la méthode *NIST Fingerprint Image Quality* (NFIQ) pour prédire les performances du système biométrique testé.

## 1 Introduction

Il existe trois façons génériques pour vérifier ou déterminer l'identité d'un individu : i) *ce que l'on sait* (code PIN, mot de passe, etc.), ii) *ce que l'on possède* (badge, carte à puce, etc.) et iii) *ce que l'on est* ou *ce que l'on sait faire* (empreinte digitale, dynamique de frappe au clavier, etc.). Ce dernier point fait référence à la biométrie. La biométrie consiste à vérifier ou déterminer l'identité d'un individu à partir de ses caractéristiques biologiques (comme l'ADN), comportementales (comme la voix) ou morphologiques (comme l'empreinte digitale). En comparaison aux systèmes d'authentification basés sur *ce que l'on sait* ou *ce que l'on possède*, qui offrent une réponse binaire (oui ou non), les systèmes d'authentification basés sur

## Qualité de données biométriques

la biométrie sont moins précis et donnent des réponses en terme de pourcentage de similarité (entre 0% et 100%, le 100% n'étant quasiment jamais atteint). Cette variation des résultats d'authentification d'un individu peut être due à une mauvaise interaction de l'utilisateur avec le capteur biométrique (cas d'un doigt mal positionné sur un capteur d'empreintes digitales), conditions d'acquisition différentes (cas de changements d'éclairage pour un système de reconnaissance faciale) ou utilisation de capteurs différents lors de la phase d'enrôlement et de reconnaissance. Ainsi, pour qu'un système soit opérationnel et efficace contre les divers types de bruit d'acquisition, contrôler la qualité des données acquises devient indispensable.

L'évaluation de la qualité des données biométriques est importante puisqu'elle impacte directement la performance, illustrée par le taux d'égale erreur (EER), d'un système biométrique (Grother et Tabassi, 2007). En se basant sur la notion de qualité, les données biométriques de mauvaise qualité peuvent ainsi être supprimées lors de la phase d'enrôlement ou rejetées au cours de la phase de vérification. Cette information pourra être également utilisée dans les approches multimodales (Poh et al., 2009, 2010). Par exemple, Poh et al. (2009) présentent une étude qui vise à étudier la performance des méthodes de fusion multimodale lors du changement de la qualité des données biométriques (changement qui peut être due à un changement du capteur d'acquisition). Les résultats de la campagne d'évaluation, sur 22 méthodes de fusion, ont montré que les meilleures méthodes sont celles qui exploitent la qualité lors du processus de fusion.

Dans cet article, nous proposons une nouvelle approche pour quantifier la qualité des données biométriques. L'approche proposée possède l'avantage d'être indépendante de la modalité, et du système de vérification utilisé. Nous montrons son intérêt pour détecter différents artefacts d'acquisition, qui ont un impact majeur sur la performance globale des systèmes biométriques. Dans la suite de cet article, la section 2 présente un ensemble de travaux relatifs à notre problématique. La section 3 décrit la méthode proposée. Les résultats expérimentaux sont donnés dans la section 4. Enfin, la section 5 conclut cet article par une discussion et quelques perspectives de ces travaux.

## 2 Travaux antérieurs

Dans cette section, nous présentons un aperçu des méthodes existantes destinées à évaluer la qualité des données biométriques morphologiques. Nous n'abordons pas les méthodes destinées à évaluer la qualité des données comportementales, qui sont généralement basées sur des calculs de statistiques sur les caractéristiques extraites (événements de temps, pression, *etc.*).

Alonso-Fernandez et al. (2007) ont présenté un aperçu des méthodes existantes visant à quantifier la qualité d'empreintes digitales. Les auteurs montrent l'impact des images de mauvaise qualité sur la performance globale des systèmes biométriques. Les méthodes présentées ont montré leur efficacité sur la prédiction de la qualité des images d'empreintes digitales. Cependant, ces méthodes dépendent de la modalité biométrique considérée, et ainsi ne peuvent pas être exploitables pour d'autres types de modalité (comme le visage). La métrique *NIST Fingerprint Image Quality* (NFIQ) proposée par Tabassi et Wilson (2005) est un exemple de cette famille de méthodes. Cette métrique utilise un vecteur à 11 éléments basé sur la qualité

des minuties extraites, et un processus d'apprentissage par réseaux de neurones pour prédire la classe de qualité (1 : excellente, . . . , 5 : pauvre) pour une image d'empreinte digitale.

He et al. (2008) ont proposé un modèle hiérarchique pour calculer la qualité de l'échantillon biométrique à trois niveaux : base de données ( $q_1$ ), classe ( $q_2$ ) et image ( $q_3$ ). La méthode proposée est basée sur les quantiles de la distribution de scores des utilisateurs légitimes et d'imposteurs. La méthode proposée est efficace puisqu'elle est basée sur la séparation de la distribution de scores des utilisateurs légitimes et d'imposteurs. Cependant, la méthode requière un minimum d'images pour chaque classe (*i.e.*, individu), ce qui limite son utilisation dans la pratique.

Zhang et Wang (2009) ont présenté une méthode basée sur l'hypothèse d'asymétrie du visage. La méthode utilise le descripteur SIFT pour quantifier la qualité d'une image. La méthode propose trois métriques de qualité :  $q_1$  mesure le rapport des points SIFT détectés sur les deux côtés du visage,  $q_2$  et  $q_3$  ajoutent les critères de localisation et des descripteurs SIFT sur les points d'intérêt détectés, respectivement. La méthode présentée a démontré sa robustesse face aux variations d'éclairage et de pose. D'autres méthodes basées sur l'asymétrie sont présentées dans (Gao et al., 2007; Sang et al., 2009). Cependant, ces méthodes ne peuvent pas être utilisées pour les autres types de modalité (comme l'empreinte digitale).

## Discussion

En biométrie, les travaux effectués sur la qualité sont beaucoup moins nombreux que ceux liés à l'extraction de paramètres et la reconnaissance. La plupart de ces travaux sont dépendants de la modalité utilisée (Alonso-Fernandez et al., 2007; Tabassi et Wilson, 2005; Zhang et Wang, 2009; Gao et al., 2007; Sang et al., 2009; Krichen et al., 2007) et du système de vérification. D'autres méthodes basées sur la distribution de scores des utilisateurs légitimes et d'imposteurs existent et requièrent un minimum d'images pour chaque classe, ce qui limite leurs utilisations d'une manière directe sur les images acquises (He et al., 2008). Nous proposons une nouvelle approche pour quantifier la qualité de données biométriques. Cette approche est basée sur l'utilisation conjointe de deux types d'informations : 1) la qualité de l'image, et 2) la qualité des paramètres extraits en utilisant le descripteur SIFT (Lowe, 2004). L'approche proposée possède l'avantage d'être indépendante de la modalité et du système de vérification utilisé. Nous présentons cette méthode dans la section suivante.

## 3 Méthode développée

La méthode proposée consiste à quantifier la qualité d'une donnée biométrique en utilisant deux types d'informations complémentaires (voir figure 1). Le principe retenu est le suivant : suite au calcul d'un critère de qualité d'image (section 3.1) et de plusieurs critères de qualité du descripteur (section 3.2), un processus de classification par apprentissage statistique est opéré à partir de l'ensemble des critères calculés (section 3.3).

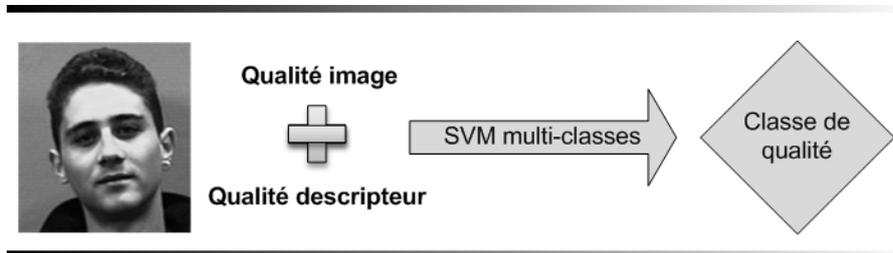


FIG. 1 – Principe de la méthode proposée.

### 3.1 Qualité image sans référence

L'évaluation de la qualité des images est utilisée pour valider un traitement appliqué sur des images numériques. Dans le cadre de la compression des images, par exemple, une telle évaluation est utilisée pour quantifier la qualité de l'image reconstruite. Les métriques de qualité sont généralement classées en trois catégories : i) les métriques de qualité avec référence complète, notées FR (*Full Reference*) (Wang et al., 2004), qui comparent l'image à évaluer avec un modèle de référence de celle-ci ; ii) les métriques de qualité avec référence réduite, notées RR (*Reduced Reference*) (Qiang et Wang, 2009), qui comparent une description de l'image à évaluer avec une description du modèle de référence ; et iii) les métriques de qualité sans référence, notées NR (*No Reference*) (Wang et al., 2002), qui quantifient la qualité de l'image à évaluer sans connaissance a priori sur celles-ci (*i.e.*, sans utilisation de modèle ou de description de référence). Dans cette étude, étant donné que le signal de référence n'est pas disponible, nous avons cherché à utiliser une métrique de qualité sans référence (NR). La plupart des métriques NR existantes dépendent de l'artefact d'acquisition (effet de bloc (Wang et al., 2000), flou (Sazzad et al., 2008), *etc.*), ce qui limite leur utilisation en pratique. D'autres méthodes (Jung et al., 2002; Charrier et al., 2006) utilisent un algorithme d'apprentissage sur des paramètres extraits. L'efficacité de ces métriques dépend ainsi de la fiabilité et de la généralisation de ces paramètres. Dans le cadre de cette thèse, nous avons utilisé l'indice *BLind Image Integrity Notator using DCT Statistics* (BLIINDS) (Saad et al., 2010) qui ne dépend pas de l'artefact d'acquisition. Cet indice exploite la notion de statistiques de scènes naturelles. L'idée principale de cette approche repose sur l'hypothèse que les fonctions du système visuel humain ont évolué en fonction du temps et sont adaptées aux statistiques du monde dans lequel l'être humain évolue. L'indice BLIINDS est basé sur le calcul de quatre facteurs de dégradation dans le domaine de la DCT<sup>1</sup> à différentes résolutions spatiales de l'image. Ces facteurs sont ensuite combinés afin de calculer la note finale de qualité. L'image est décomposée en bloc de taille  $17 \times 17$ . Les dégradations mesurées sont :

1. Distorsion de contraste ( $v_1$ ) : le contraste est une propriété intrinsèque d'une image qui désigne la différence entre les zones claires et foncées d'une image. Le contraste  $v_1$  est calculé en utilisant les valeurs de contraste local de chaque bloc. Le contraste local du  $k$ -ième bloc est donné par :

---

1. Discrete cosine transform

$$c^k(x) = \frac{1}{N} \sum_{i=1}^N \frac{x_{AC}^i}{x_{DC}} \quad (1)$$

avec  $N$  est la taille du bloc,  $x_{DC}$  représente le coefficient  $DC$  et l'ensemble  $\{x_{AC}^i | i = 1 : N\}$  représente les coefficients  $AC$ . Le contraste de l'image  $v_1$  est ainsi calculé par :

$$v_1 = \frac{1}{M} \sum_{i=1}^M c^i(x) \quad (2)$$

avec  $M$  est le nombre de blocs de l'image en question.

2. Distortion de structure ( $v_2$ ) : les caractéristiques de structure sont obtenues en utilisant le kurtosis des coefficients (non DC) de fréquences DCT, calculés sur chaque bloc. Le kurtosis (ou coefficient d'aplatissement de Pearson) correspond à une mesure de l'aplatissement de la distribution d'une variable aléatoire réelle. Le kurtosis du  $k^{\text{ème}}$  bloc est ainsi donné par :

$$\kappa^k(x_{AC}) = \frac{E(x_{AC} - \mu^4)}{\sigma^4} \quad (3)$$

avec  $\mu$  est la moyenne des coefficients  $AC$ , et  $\sigma$  son écart-type. La mesure de distortion de structure  $v_2$  est ainsi calculée par la moyenne des valeurs au dessous du 10<sup>ème</sup> percentile.

3. Anisotropie d'orientation ( $v_3$  et  $v_4$ ) : Gabarda et Cristbal (2007) montrent que la dégradation a un impact sur l'information directionnelle d'une scène. Par conséquent, l'anisotropie (qui dépend de l'information directionnelle d'une scène) est calculée en utilisant l'entropie de Rényi (qui est une généralisation de l'entropie de Shannon) sur les blocs DCT selon quatre orientations différentes  $\theta = 0, 45, 90, 135$  en degrés. Les deux mesures  $v_3$  et  $v_4$  sont calculées comme suit : les coefficients DCT du  $k^{\text{ème}}$  bloc autour de l'orientation  $\theta$  sont notés par  $P_\theta[k, j]$ , avec  $j$  est l'indice du coefficient DCT. Chaque coefficient du bloc DCT est ensuite normalisé par :

$$\tilde{P}_\theta[k, j] = \frac{P_\theta[k, j]^2}{\sum_{j=1}^N P_\theta[k, j]^2} \quad (4)$$

avec  $N$  la taille du  $k^{\text{ème}}$  bloc orienté et son entropie de Rényi  $R_\theta^k$  est défini par :

$$R_\theta^k = \frac{1}{1 - \beta} \log_2 \left( \tilde{P}_\theta[k, j]^\beta \right) \quad (5)$$

où  $\beta > 1$ . Enfin, les deux mesures basées sur l'anisotropie sont définies par :

$$v_3 = \sigma(E(R_\theta^k)) \text{ and } v_4 = \max(E(R_\theta^k)), \forall k, \forall \theta \quad (6)$$

## Qualité de données biométriques

Dans cette étude, nous fixons  $\beta$  à 3 (après essais expérimentaux). Étant donné que la perception visuelle de l'image dépend de la résolution de l'image, la distance entre le plan de l'image et l'observateur, et l'acuité des observateurs, une approche multi-échelles est appliquée afin de calculer un score global :

$$\text{BLIINDS} = \prod_{i=1}^L v_1^{\alpha_1^i} v_2^{\alpha_2^i} v_3^{\alpha_3^i} v_4^{\alpha_4^i} \quad (7)$$

avec  $\sum_{j=1}^4 \sum_{i=1}^L \alpha_j^i = 1$  et  $L$  représente le nombre de niveaux de décomposition utilisé. Les valeurs  $\alpha_j^i$  ont été obtenues en calculant la corrélation de chacun des critères ( $v_i$ ) avec les notes de qualité fournies par les observateurs humains (Saad et al., 2010). Des exemples de cette métrique sont donnés à la figure 2.

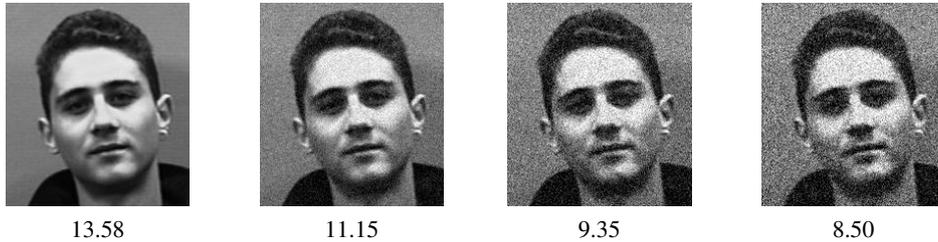


FIG. 2 – Exemples de valeurs de la métrique BLIINDS sur des images de la base de données *FACES94*. De gauche à droite, image de référence ensuite images altérées par bruit gaussien. Plus la métrique est élevée, meilleure est jugée la qualité de l'image.

### 3.2 Qualité du descripteur

La mesure de la qualité d'un descripteur est basée sur des mesures statistiques de points d'intérêt. Nous avons utilisé les points d'intérêt puisqu'ils décrivent de façon stable les régions de l'image où l'information est importante. Cette approche est généralement utilisée pour reconnaître des objets (Mansur et Kuno, 2008) et dans les algorithmes de reconnaissance biométrique (Ladoux et al., 2009). Pour le calcul du vecteur descripteur au voisinage des points détectés, il existe de nombreuses méthodes tels que SIFT (Lowe, 2004), *Shape Contexts* (Belongie et al., 2001), *Speed Up Robust Features* (SURF) (Bay et al., 2008). Parmi ces algorithmes, l'algorithme SIFT est retenu pour trois raisons principales. Premièrement, l'algorithme SIFT est invariant aux changements d'échelles et à la rotation des objets. Deuxièmement, une étude comparative (Mikolajczyk et Schmid, 2005) de différents descripteurs montre que SIFT est le plus performant. Troisièmement, l'avantage majeur de SIFT est qu'il permet une description générique de la donnée biométrique (applicable à des empreintes digitales, visages, images de veines). L'algorithme SIFT a été également utilisé par Berretti et al. (2010) dans le cas de reconnaissance faciale 3D.

L'algorithme SIFT, proposé par Lowe, est un algorithme de traitement d'images qui permet de détecter et de décrire les caractéristiques d'une image. Elle permet de transformer une

image en un ensemble de caractéristiques, chacun étant invariant aux transformations suivantes : translation de l'image, changement d'échelle (*c.-à-d.*, redimensionnement), rotation et partiellement invariant aux changement d'éclairage. Des exemples de détection de points d'intérêt sont donnés à la figure 3. En utilisant SIFT, l'image  $I$  est ainsi caractérisée par l'ensemble  $Y(I) = \{k_i = (x_i, y_i, \sigma_i, \theta_i, v_i) \mid i = 1 : N(I)\}$  avec  $N(I)$  le nombre de points d'intérêt détectés dans  $I$ ;  $(x_i, y_i)$  la position du point d'intérêt  $i$  dans  $I$ ;  $(\sigma_i, \theta_i)$  l'échelle et l'orientation du point d'intérêt  $i$ ; et  $v_i$  le vecteur (à 128 éléments) de descripteurs du point d'intérêt  $i$ . À partir de ces caractéristiques, nous avons choisi d'utiliser quatre critères qui nous ont paru pertinents (*cf.*, section 4.2.1) pour prédire la qualité du descripteur : 1) le nombre de points d'intérêt détectés dans l'image  $I$ ; 2) le coefficient DC de la matrice  $MAT$ , avec  $N(I)$  lignes et 128 colonnes, contenant les vecteurs descripteurs des points d'intérêt détectés dans  $I$ ; 3) la moyenne et 4) l'écart-type du vecteur contenant l'échelle de chaque point d'intérêt détecté dans  $I$ .

Finalement, nous disposons de cinq critères (un dédié à la qualité de l'image et quatre sur la qualité du descripteur) pour établir le niveau de qualité de données biométriques. Au lieu de faire une opération arithmétique des valeurs ainsi obtenues, nous proposons d'utiliser un algorithme de classification supervisée de la qualité à 10 classes :

- Classe 1 correspond à une image de référence (c'est à dire non altérée) ;
- Classes 2 jusqu'à 10 correspondent à 3 types d'altérations et 3 niveaux pour chaque type d'altération, respectivement. Une description détaillée des altérations introduites est donnée à la section 4.1.2.



FIG. 3 – Exemples de détection de points d'intérêt.

### 3.3 Machines à Vecteurs de Support (SVM)

Les machines à vecteurs de support ou Séparateurs à Vastes Marges est une méthode de classification par apprentissage supervisé développée par Vapnik (1995). Elle est connue sous le terme anglais par *Support Vectors Machine* (SVM). Nous avons choisi d'utiliser le classifieur SVM puisque celui-ci a montré de bonnes performances par rapport à d'autres classifieurs de la littérature (Moghaddam et Yang, 2000; Munder et Gavrilu, 2006). Le but des SVM est de classifier un objet  $x$  à l'aide d'une marge maximale associée à des vecteurs de supports et d'une fonction noyau. Cette méthode est devenue populaire du fait de ces performances à

## Qualité de données biométriques

traiter des données de grande dimension. La fonction noyau permet d'opérer un changement de repère dans un espace de plus grande dimension afin de retrouver un problème de séparation linéaire, lorsque les données ne sont pas linéairement séparables. Soit une base d'apprentissage  $S_{\text{apprentissage}} : S_{\text{apprentissage}} = \{(x_1, y_1), \dots, (x_m, y_m)\}$  composée de  $m$  couples (vecteur d'attributs, classe) avec  $x_i \in \mathbb{R}^n$  et  $y_i \in \{-1, 1\}$ . L'algorithme SVM projette les valeurs  $x_i$  dans un espace de travail  $\mathcal{H}$  ( $\phi : \mathbb{R}^n \rightarrow \mathcal{H}$ ). L'hyperplan optimal de séparation des deux classes dans l'espace  $\mathcal{H}$  est ensuite recherché. Cet hyperplan  $(w, b)$  matérialise la frontière de séparation entre les deux classes. La classe  $y$  d'un nouvel exemple  $x$  est définie par :

$$y = \langle w, \Phi(\mathbf{x}) \rangle + b = \sum_{i=1}^{\ell} \alpha_i^* y_i K(\mathbf{x}_i, \mathbf{x}) + b \quad (8)$$

avec  $\alpha_i^* \in \mathbb{R}$  et  $K(\cdot, \cdot)$  est la fonction noyau. Dans l'algorithme SVM, l'hyperplan est optimal s'il maximise la distance qui le sépare des exemples dont il est le plus proche. Cette distance est appelée marge du classifieur. Les  $\alpha_i^*$  qui maximisent le critère d'optimalité sont obtenus en résolvant :

$$\max_{\alpha_i} \sum_{i=1}^{\ell} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{\ell} \alpha_i \alpha_j y_i K(\mathbf{x}_i, \mathbf{x}_j y_j) \quad (9)$$

sous les contraintes,  $0 \leq \alpha_i \leq C$  et  $\sum_{i=1}^{\ell} \alpha_i y_i = 0$ , avec  $C$  est le coefficient de pénalisation. L'algorithme SVM de base a été développé pour les problèmes de classification à deux classes. Cependant, plusieurs approches peuvent être utilisées pour l'étendre aux problèmes multi-classes. Dans nos travaux, nous avons utilisé l'approche *un contre un* avec le critère de vote majoritaire pour la sélection de la classe finale. Nous avons utilisé un script python (*easy.py*) fourni par la librairie *libsvm* (Chang et Lin, 2001). Une recherche exhaustive (*grid-search*) est effectuée pour la recherche des deux paramètres optimums  $C$  et  $\gamma$ , et le noyau utilisé est le noyau RBF défini par :

$$k(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2) \quad (10)$$

## 4 Validation

### 4.1 Protocole expérimental

Nous présentons dans la section 4.1.1 les cinq bases de données utilisées, et dans la section 4.1.2 les trois types d'altération réelles simulant des artefacts d'acquisition introduites sur chacune de ces bases. La section 4.1.3 présente le système d'authentification biométrique utilisé, et la section 4.1.4 présente le processus de validation de la méthode proposée.

#### 4.1.1 Bases de données de référence

##### BD1 La base FACES94

La base de données FACES94 (of Essex, 1994) a été collectée sur 152 personnes (20

images par personne). Les images de cette base ont été capturées avec des conditions différentes d'expression faciale. La figure 4 présente un exemple des images de cette base.

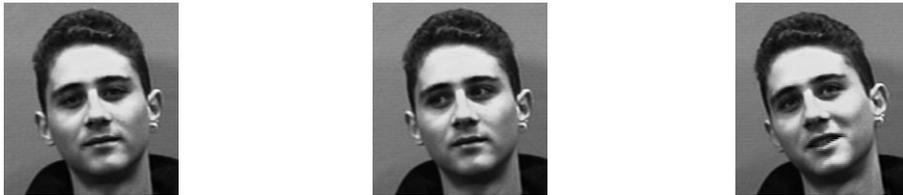


FIG. 4 – Exemple de visages de la base FACES94 (source (of Essex, 1994)).

**BD2** La base ENSIB

ENSIB (Hemery et al., 2007) est une base de données de visages collecté en 2007. Elle est composée de 100 personnes (40 images par personne), contenant des images capturées avec des conditions différentes de pose (de gauche à droite). La figure 5 illustre un exemple des images de cette base.



FIG. 5 – Exemple de visages de la base ENSIB (source (Hemery et al., 2007)).

**BD3** La base FERET

FERET (Phillips et al., 1998, 2000) est une base de données unimodale destinée à quantifier la performance des systèmes de reconnaissance faciale. Elle a été collectée sur 1199 personnes en plusieurs sessions entre août 1993 et juillet 1996. Pour certains volontaires, la différence entre la première et dernière image acquise est de plus de deux ans. Les images de cette base ont été collectées avec des conditions différentes d'expression, de pose, d'éclairage et d'âge. La figure 6 représente un exemple de visages extraits de cette base de données.



FIG. 6 – Exemple de visages de la base FERET (source (Phillips et al., 2000)).

## Qualité de données biométriques

### **BD4** *La base AR*

La base de données AR (Martinez et Benavente, 1998) a été créée par Aleix Martinez et Robert Benavente au *Computer Vision Center (CVC)*. Elle a été collectée sur 120 personnes (26 images par personne) avec des conditions différentes d'expression faciale, d'éclairage, et d'occultation (lunettes de soleil et écharpe). Un exemple de visages de cette base est présenté dans la figure 7.



FIG. 7 – Exemple de visages de la base AR (source (Martinez et Benavente, 1998)).

### **BD5** *La base FVC2002 DB<sub>2</sub>*

La base DB<sub>2</sub> est une base de données d'empreintes digitales utilisée dans la compétition *Fingerprint Verification Competition (FVC2002)* (Maio et al., 2002). Elle est composée de 100 personnes (8 images par personne). La figure 8 illustre un exemple des images de cette base.



FIG. 8 – Exemple d'empreintes digitales de la base FVC2002 DB<sub>2</sub> (source (Maio et al., 2002)).

#### **4.1.2 Bases de données dégradées**

Pour chacune des bases de référence utilisées, nous avons simulé plusieurs artefacts d'acquisition (mouvement, bruit gaussien et distance d'acquisition), et appliqué trois niveaux de dégradation pour chaque type d'altération :

- Altération par mouvement (flou) : les images altérées par flou sont obtenues par un filtre gaussien 2D en utilisant la méthode MATLAB *fspecial* ('gaussian', hsize,  $\sigma$ ) ;

- Altération par bruit gaussien : ces images sont obtenues en utilisant la méthode MATLAB *imnoise* ( $I$ , 'gaussian',  $\mu$ ,  $v$ );
- Altération par distance d'acquisition (redimensionnement) : ces images sont générées en utilisant la méthode MATLAB *imresize* ( $I$ ,  $scale$ , 'nearest').

Le tableau 1 présente la valeur des paramètres requis des méthodes MATLAB. La figure 9 illustre un exemple de ces altérations sur une image de la base de données *FACES94*.

Type d'altération	Méthode	Niveau 1	Niveau 2	Niveau 3
Flou	<i>fspecial</i>	$hsize = [7 \ 7]$ et $\sigma = 1$	$hsize = [7 \ 7]$ et $\sigma = 2$	$hsize = [7 \ 7]$ et $\sigma = 6$
Bruit gaussien	<i>imnoise</i>	$\mu = 0.01$ et $v = 0.003$	$\mu = 0.01$ et $v = 0.01$	$\mu = 0.01$ et $v = 0.017$
Redimensionnement	<i>imresize</i>	$scale = 0.8$	$scale = 0.6$	$scale = 0.4$

TAB. 1 – Paramètres des méthodes d'altérations matlab.

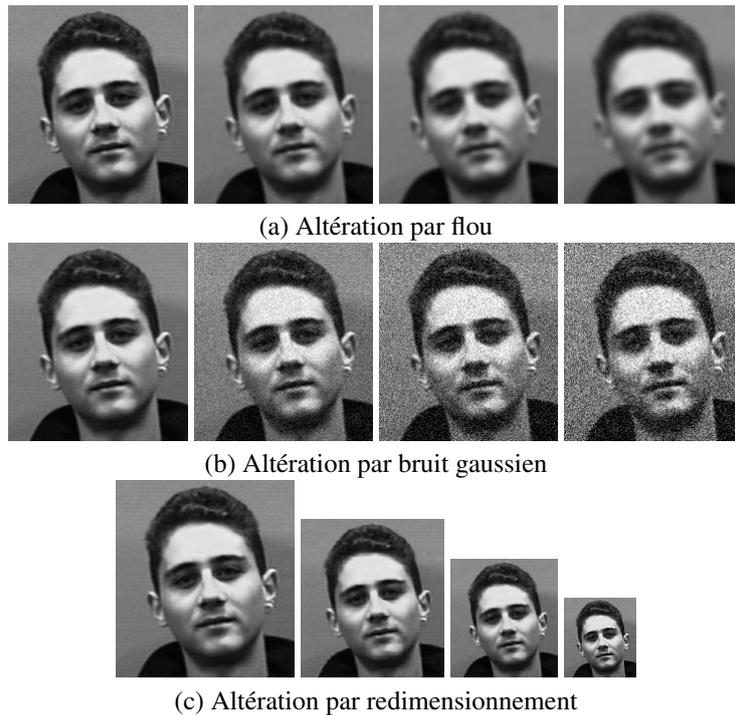


FIG. 9 – Exemple d'altérations sur une image de la base de données *FACES94*. Du gauche à droite, image de référence ensuite images altérées niveau 1, 2 et 3, respectivement.

#### 4.1.3 Algorithme de vérification : GREYC-Face

C'est un système de reconnaissance faciale développé dans le laboratoire de recherche GREYC (Rosenberger et Brun, 2008). Le système utilise le descripteur SIFT pour la vérification de deux vecteurs biométriques. La vérification entre deux images  $I_1$  et  $I_2$  correspond ainsi au calcul du nombre d'associations entre les deux ensembles  $Y(I_1)$  et  $Y(I_2)$ . Une association est définie par une double mise en correspondance entre deux points d'intérêt. La méthode de mise en correspondance utilisée est celle présentée par Ladoux et al. (2009) (version modifiée de la méthode proposée par Lowe (2004)). Pour le point d'intérêt  $x$  de l'image  $I_1$ , nous recherchons le point d'intérêt  $y$  de  $I_2$  le plus proche parmi l'ensemble de points d'intérêt de  $I_2$ . Nous regardons également si le second point d'intérêt  $y'$  le plus proche est suffisamment loin de  $x$  au moyen d'une valeur seuil  $C$  :

$$d(x, y) = \min_{\{z \in Y(I_2)\}} d(x, z) \quad (11)$$

et

$$d(x, y) \leq C d(x, y') \quad (12)$$

avec

$$d(x, y') = \min_{\{z \in Y(I_2), d(x, z) > d(x, y)\}} d(x, z) \quad (13)$$

La distance  $d(., .)$  est une distance euclidienne calculée entre les deux descripteurs normalisés correspondant aux points d'intérêt. Si ces deux conditions ne sont pas remplies, alors le point  $x$  n'est pas mis en correspondance avec le point  $y$ . Une association entre les deux points d'intérêt  $x$  et  $y$  est créée si et seulement si : le point  $x$  est mis en correspondance avec le point  $y$  et le point  $y$  est mis en correspondance avec le point  $x$ . La similarité entre les deux ensembles de point est le nombre de points d'intérêt mis en correspondance. La figure 10 illustre un exemple de vérification résultant d'une tentative d'un utilisateur légitime. Le nombre d'associations est utilisé comme mesure de similarité.

#### 4.1.4 Processus de validation

Selon Grother et Tabassi (2007), les méthodes de qualité doivent être en mesure de prédire la performance des systèmes biométriques. Cela signifie qu'une méthode de qualité prend en entrée une donnée biométrique, et prédit sa catégorie de qualité lié au taux d'erreur associée à cette donnée. Afin de quantifier la performance de la méthode proposée, nous procédons comme suit :

- Apprentissage des SVM multi-classes : pour les bases de visages, nous avons généré quatre SVM multi-classes (*i.e.*, un SVM multi-classes par base), et un SVM multi-classes contenant des exemples de toutes les bases ( $SVM_{tout}$ ). Pour la base d'empreintes digitales, nous avons généré un autre SVM multi-classes. Pour apprendre et tester les différents SVM multi-classes, nous avons découpé chaque base d'images en deux ensembles  $S_{apprentissage}$  et  $S_{test}$  d'une manière équilibrée (*i.e.*, qu'il y a autant

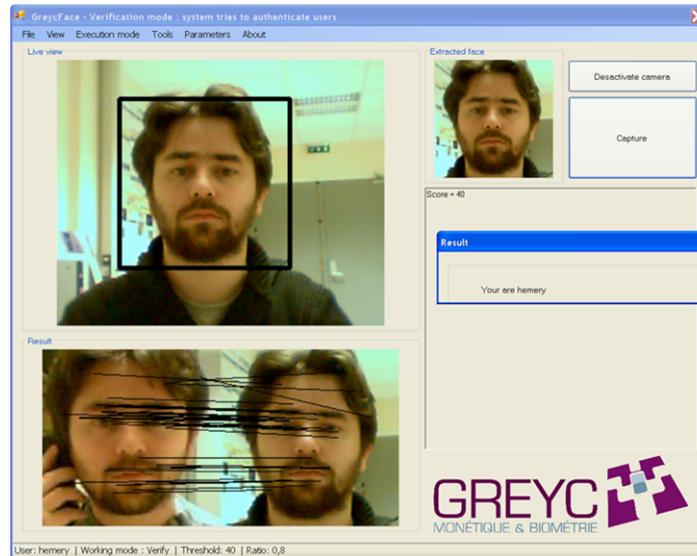


FIG. 10 – Le logiciel GREYCFace. Exemple de vérification résultant d'une tentative d'un utilisateur légitime.

d'exemples par classe dans chaque ensemble). Le choix du noyau utilisé et les paramètres requis sont présentés dans la section 3.3 ;

- Définition des catégories de qualité : la méthode SVM multi-classes proposée prédit une classe de qualité pour une image en entrée. Afin de quantifier la performance de cette méthode, nous devons tout d'abord définir les catégories de qualité pour le système de vérification utilisé. Selon le système de vérification utilisé, certaines altérations peuvent avoir un impact sur sa performance globale plus que d'autres. Dans cet article, l'EER proposé par le standard ISO/IEC 19795-1 (2006) est utilisé pour illustrer la performance globale d'un système biométrique. L'EER correspond à l'erreur constatée lorsque le système biométrique est paramétré pour laquelle le taux de fausses acceptations (FAR) et le taux de faux rejets (FRR) sont identiques. Par conséquent, plus l'EER est faible, plus le système est performant ;
- Corrélation des valeurs de l'EER avec les catégories de qualité : afin de quantifier l'efficacité de la méthode proposée pour prédire les performances du système testé, nous calculons l'EER de chaque catégorie de qualité. L'intérêt de la méthode proposée est ainsi quantifié par son efficacité pour prédire les performances du système testé. En d'autres termes, plus les données biométriques sont dégradées, plus la performance globale du système est dégradée (cela se traduit par une augmentation des valeurs de l'EER).

## 4.2 Résultats

La section 4.2.1 présente l'efficacité des cinq critères de qualité retenus en fonction des altérations. La section 4.2.2 montre l'intérêt de la méthode proposée selon le processus de validation présentée dans la section 4.1.4. Une étude comparative entre la méthode proposée et la méthode NFIQ est présentée dans la section 4.2.3.

### 4.2.1 Comportement des descripteurs en fonction des altérations

Dans cette section, nous présentons le comportement des cinq critères de qualité utilisés dans la méthode proposée avec les altérations introduites dans la section 4.1.2. Il s'agit de quantifier l'efficacité de chaque critère pour détecter les trois types d'altérations. Pour ce faire, nous utilisons le coefficient de corrélation linéaire de Pearson entre les critères de qualité utilisés et les trois types d'altérations. Le coefficient de corrélation entre deux variables aléatoires,  $X(x_1, \dots, x_p)$  et  $Y(y_1, \dots, y_p)$ , permet de quantifier la relation de dépendance qui peut exister entre ces variables. Elle est donnée par :

$$r_p = \frac{\sum_{i=1}^{i=p} (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\sum_{i=1}^{i=p} (x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^{i=p} (y_i - \bar{y})^2}} \quad (14)$$

Le coefficient de corrélation linéaire est compris entre  $-1$  et  $1$ . Plus le coefficient est proche des valeurs extrêmes  $-1$  et  $1$ , plus la corrélation linéaire entre les variables est forte. Les valeurs intermédiaires renseignent sur le degré de dépendance linéaire entre les deux variables. Une corrélation égale à  $0$  signifie que les variables sont linéairement indépendantes. Nous définissons ainsi pour chaque type d'altération et pour chaque critère  $p$  du vecteur de qualité les deux variables comme suit :

- $X_p = \{X_{pk} | k = 1 : 4\}$  où  $X_{p1}$  est l'ensemble des valeurs du critère  $p$  de toutes les images de référence,  $(X_{p2}, X_{p3}, X_{p4})$  sont les ensembles des valeurs de  $p$  de toutes les images altérées niveau 1, 2 et 3, respectivement ;
- Les niveaux d'altérations sont représentés par la variable  $Y$  (1 : pour les bases de référence, 2, 3 et 4 : pour les bases altérées niveau 1, 2 et 3, respectivement). Plus spécifiquement,  $Y = \{y_k | y_k = 1, k = 1 : N; y_k = 2, k = N + 1 : 2N; y_k = 3, k = 2N + 1 : 3N; y_k = 4, k = 3N + 1 : 4N\}$  où  $N$  correspond à la taille des quatre bases de visages de référence.

Le tableau 2 montre que les quatre critères de qualité du descripteur (Nombre de points d'intérêt, Coefficient DC, Moyenne et Ecart-type d'échelles) sont pertinents pour détecter les trois types d'altérations. Le critère de qualité image BLIINDS a montré son efficacité (avec la valeur absolue d'un coefficient de corrélation supérieur à  $0,6$ ) pour détecter les altérations par flou et bruit gaussien. Pour l'altération par redimensionnement, le tableau montre que BLIINDS ne permet pas de la détecter. Ce résultat était attendu car BLIINDS est une métrique de qualité d'image sans-référence et multi-résolutions, et que le processus de redimensionnement n'introduit aucune distortion de qualité en l'absence de référence.

Critère	Flou	Bruit gaussien	Redimensionnement
Nombre de points d'intérêt	<b>-0.5728</b>	0.3901	-0.4880
Coefficient DC	<b>-0.6155</b>	<b>0.5672</b>	<b>-0.5252</b>
Moyenne échelles	<b>0.7933</b>	<b>-0.5632</b>	-0.3960
Ecart-type échelles	0.3470	-0.3467	-0.4729
BLIINDS	<b>0.6316</b>	<b>-0.8014</b>	-0.1018

TAB. 2 – Coefficients de corrélation de Pearson entre les critères de qualité utilisés et les altérations sur les toutes les bases de visages. Les valeurs en gras correspondent à des corrélations fortes (en valeur absolue).

#### 4.2.2 Comportement de la méthode proposée

Les performances des six SVM multi-classes générés (cinq pour les bases de visages et un pour la base d'empreintes digitales) sont données dans le tableau 3. Nous avons mis le symbole «×» pour la base *FVC2002 DB<sub>2</sub>*, car nous avons généré un seul SVM multi-classes pour cette base. Le tableau 3 montre l'intérêt de la méthode proposée pour détecter les trois types d'altérations réelles (flou, bruit gaussien et redimensionnement) des données, avec des taux de bonne classification satisfaisants (de 82, 29% jusqu'à 97, 73% sur la base d'apprentissage, et de 81, 16% jusqu'à 91, 1% sur la base de test).

Afin de définir les catégories de qualité, nous avons testé la robustesse du système contre les altérations introduites dans la section 4.1.2. La figure 11 montre l'impact des images dégradées sur la performance globale (illustrée par les valeurs de l'EER) du système testé. Les valeurs de l'EER sont calculées en utilisant la première image de référence pour l'enrôlement, et les autres pour le test (procédé d'enrôlement unique). Cette figure montre que toutes les altérations introduites ont un impact sur la performance du système biométrique étudié (*i.e.*, courbes croissantes en fonction des dégradations). Par conséquent, nous définissons dans le tableau 4, les catégories de qualité retenues pour le système biométrique utilisé. La figure 12 illustre des résultats d'évaluation par la méthode proposée sur des images de la base *FACES94*. La figure 13 présente les valeurs de l'EER de chaque catégorie de qualité en utilisant les quatre SVM multi-classes (un SVM multi-classes par base), et le SVM multi-classes généré à partir des exemples de toutes les bases de visage, respectivement. La méthode proposée a montré son efficacité à prédire les performances du système testé. En d'autres termes, plus les images sont dégradées, plus la performance globale du système est dégradée (cela se traduit par une augmentation des valeurs de l'EER). À partir de la figure 13, nous pouvons également déduire deux points :

- Pour les bases *FACES94*, *ENSIB* et *AR* il n'y avait pas une différence significative entre les deux valeurs de l'EER de la base de référence et la base prédite (catégorie I) par la méthode proposée ;
- Pour la base *FERET*, il y avait une différence de 5.62% (figure 13 à gauche) et 5.64% (figure 13 à droite). Cette variation est due à la complexité de cette base (la base de référence contient beaucoup d'images altérées par résolution). Malgré cela, la méthode

## Qualité de données biométriques

a également montré son efficacité à prédire la performance du système utilisé sur cette base.

	$SVM_{chaque}$		$SVM_{tout}$	
	$S_{apprentissage}$	$S_{test}$	$S_{apprentissage}$	$S_{test}$
FACES94	91.01	86.69	85.68	85.28
ENSIB	97.73	89.82	94.92	91.1
FERET	82.33	81.2	82.29	81.16
AR	90.08	89.08	90.7	88.92
FVC2002 DB <sub>2</sub>	×	×	91.7	83.68

TAB. 3 – Précision (en %) des modèles SVM multi-classes sur les deux ensembles d'apprentissage ( $S_{apprentissage}$ ) et de test ( $S_{test}$ ).

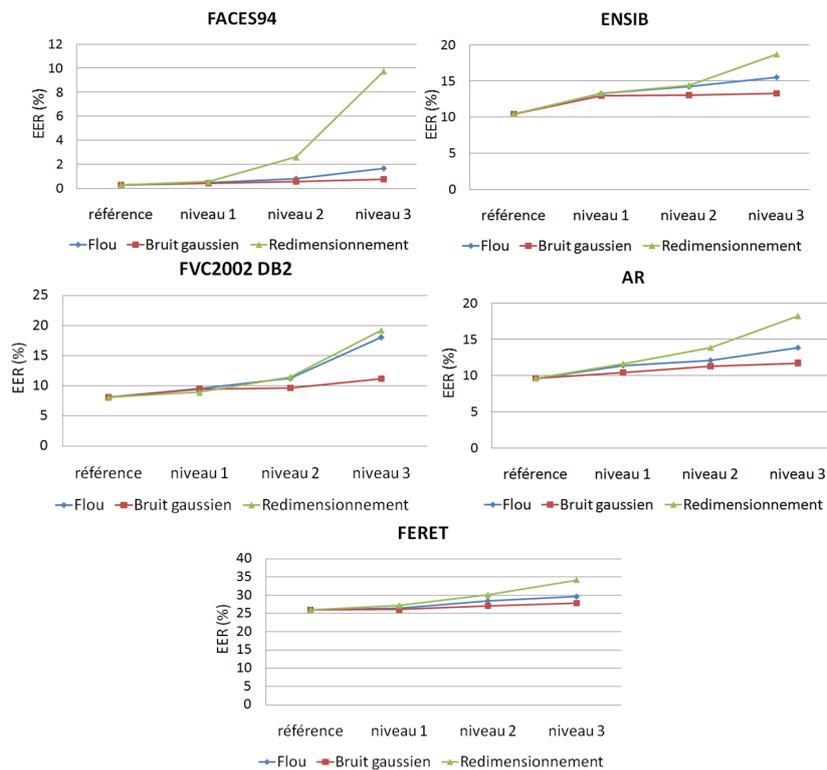


FIG. 11 – Impact des altérations sur la performance globale du système biométrique utilisé : valeurs de l'EER (en %) sur chaque base de données.

Catégorie de qualité	Label prédit par le SVM multi-classes	Description
I	I	Bonne
II	2, 5 et 8	Moyenne
III	3, 6 et 9	Mauvaise
IV	4, 7 et 10	Très mauvaise

TAB. 4 – Catégories de qualité.

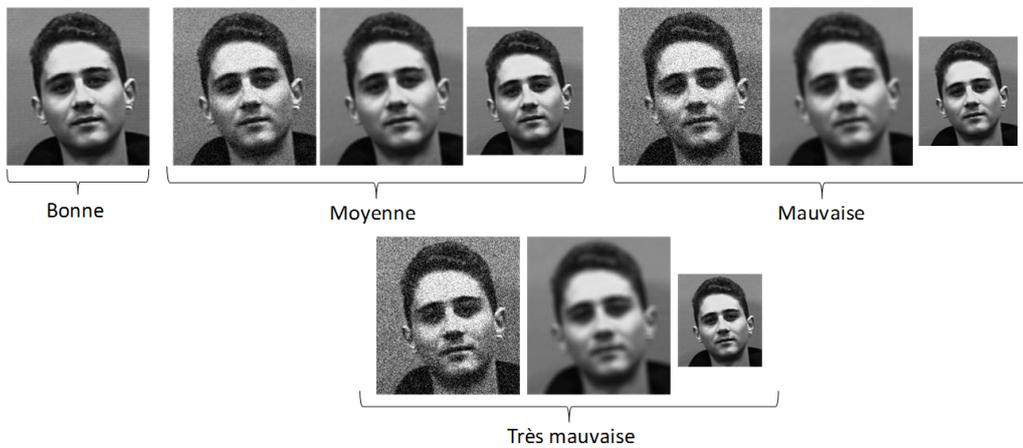


FIG. 12 – Exemple des résultats d'évaluation sur des images de la base FACES94.

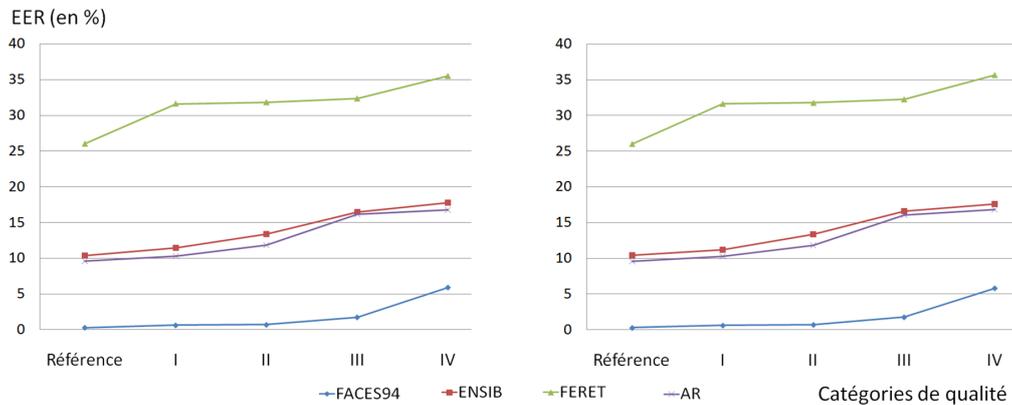


FIG. 13 – Les valeurs de l'EER des quatre bases de référence, et de chaque catégorie de qualité. Ces valeurs sont calculées en utilisant les quatre SVM multi-classes (à gauche) et le SVM multi-classes généré à partir des exemples de toutes les bases (à droite), respectivement.

### 4.2.3 Étude comparative entre la méthode proposée et NFIQ

Nous avons utilisé la base d’empreintes digitales *FVC2002 DB<sub>2</sub>* (avec ses images altérées, le nombre total des images ainsi utilisées est égal à 8000) pour comparer la méthode proposée avec la métrique de qualité NFIQ (Tabassi et Wilson, 2005) proposée par le NIST. Nous avons choisi NFIQ puisque cette dernière est la plus citée dans la littérature dans le cas d’empreintes digitales. Afin de comparer ces deux algorithmes de qualité, nous avons suivi la démarche suggérée par Grother et Tabassi (2007). Nous avons utilisé le test de Kolmogorov-Smirnov (KS) (Saporta, 1990) pour mesurer le chevauchement des deux distributions de scores des utilisateurs légitimes (scores intra) et d’imposteurs (scores inter). Ce test statistique retourne une valeur définie entre 0 et 1 : une valeur proche de 0 signifie que les deux distributions sont homogènes (*i.e.*, dépendantes), tandis qu’une valeur proche de 1 signifie que les deux distributions sont indépendantes. Ainsi, plus les images sont de bonne qualité, plus une valeur statistique KS importante (proche de 1) est attendue.

Le tableau 5 décrit les valeurs statistiques du test KS. La méthode proposée a montré son efficacité pour mieux séparer les deux distributions des scores intra et inter que la métrique NFIQ. Pour la catégorie IV (*i.e.*, images de très mauvaise qualité), la méthode NFIQ est légèrement plus efficace (statistique KS égale à 0,64) que la méthode proposée (statistique KS égale à 0,626) pour séparer la distribution des scores intra et inter. Tandis que, pour les autres trois catégories de qualité (I, II et III), la méthode proposée (statistiques KS allant de 0,797 jusqu’à 0,869) est nettement meilleure que la méthode NFIQ (statistiques KS allant de 0,632 jusqu’à 0,82).

Méthode	Catégorie I	Catégorie II	Catégorie III	Catégorie IV
SVM multi-classes	0.869	0.828	0.797	0.626
NFIQ	0.82	0.698	0.632	0.64

TAB. 5 – Comparaison entre la méthode proposée et NFIQ. Test de Kolmogorov-Smirnov (KS) pour un intervalle de confiance égal à 95%.

## 5 Conclusion et perspectives

Nous avons présenté dans cet article, une méthode pour prédire la qualité de données biométriques morphologiques. La méthode proposée utilise deux types d’informations complémentaires : 1) la qualité de l’image, et 2) la qualité des paramètres extraits en utilisant le descripteur SIFT. L’approche proposée est indépendante de la modalité utilisée, et du système de vérification. Nous avons montré son intérêt pour détecter trois types d’altérations réelles (flou, bruit gaussien et redimensionnement) des données, qui ont un impact majeur sur la performance globale des systèmes biométriques. Nous avons également montré que la méthode proposée est plus efficace que la méthode de qualité NFIQ sur la base d’empreintes digitales *FVC2002 DB<sub>2</sub>*. À noter que la métrique NFIQ comparée avec notre méthode est basée sur la qualité des minuties extraites. Ainsi, une étude comparative entre la méthode proposée et NFIQ en utilisant d’autres systèmes d’authentification basés sur les minuties extraites (tels que

le système de vérification BOZORTH3 (Watson et al., 2007) développé par le NIST) s'avère utile. Ceci constitue une perspective de ce travail. Nous comptons également ajouter un sixième critère pour détecter l'altération par luminance, qui a un impact significatif sur la plupart des systèmes de reconnaissance faciale existants.

## Notions utiles

### – Enrôlement

L'enrôlement est la première phase de tout système biométrique. Il s'agit de l'étape pendant laquelle un utilisateur est enregistré dans le système pour la première fois. Elle est commune à la vérification et l'identification. Pendant l'enrôlement, la caractéristique biométrique est mesurée en utilisant un capteur biométrique afin d'extraire une représentation numérique. Cette représentation est ensuite réduite, en utilisant un algorithme d'extraction bien défini, afin de réduire la quantité de données à stocker pour ainsi faciliter la vérification et l'identification. Dépendant de l'application et du niveau de sécurité souhaité, le modèle biométrique retenu, est stocké soit dans une base de données centrale soit sur un élément personnel propre à chaque personne ;

### – Vérification

La vérification d'identité consiste à contrôler si l'individu utilisant le système est bien la personne qu'il prétend être. Le système compare l'information biométrique acquise avec le modèle biométrique correspondant stocké dans la base de données, on parle de test  $I : I$ . Dans ce cas, le système renvoie uniquement une décision binaire (oui ou non) pouvant être pondérée. Le processus de vérification peut être formalisé comme suit :  
Soit le vecteur d'entrée  $C_U$  définissant les caractéristiques biométriques de l'utilisateur  $U$  extraites par le système, et  $M_U$  son modèle biométrique stocké dans la base de données, le système retourne une valeur booléenne suite au calcul de la fonction  $f$  définie par :

$$f(C_U, M_U) = \begin{cases} 1 & \text{si } S(C_U, M_U) \geq \tau \\ 0 & \text{sinon} \end{cases} \quad (15)$$

où  $S$  est la fonction de similarité définissant la correspondance entre les deux vecteurs biométriques, et  $\tau$  le seuil de décision à partir duquel les deux vecteurs sont considérés comme identiques.

## Remerciements

Les auteurs tiennent à remercier le Ministère d'Enseignement Supérieur et de la Recherche Français pour son soutien financier de ce travail.

## Références

- Alonso-Fernandez, F., J. Fierrez, J. Ortega-Garcia, J. Gonzalez-Rodriguez, H. Fronthaler, K. Kollreider, et J. Bigun (2007). A comparative study of fingerprint image-quality estimation methods. *IEEE Transactions on Information Forensics and Security* 2, 734–743.
- Bay, H., A. Ess, T. Tuytelaars, et L. Van Gool (2008). Speeded-Up Robust Features (SURF). *Computer Vision and Image Understanding* 110, 346 – 359.
- Belongie, S., J. Malik, et J. Puzicha (2001). Matching shapes. In *International Conference on Computer Vision*, pp. 454 – 461.
- Berretti, S., A. D. Bimbo, P. Pala, B. B. Amor, et M. Daoudi (2010). A set of selected sift features for 3D facial expression recognition. In *Proceedings of the 20th International Conference on Pattern Recognition (ICPR)*, pp. 4125–4128.
- Chang, C.-C. et C.-J. Lin (2001). *LIBSVM : a library for support vector machines*. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- Charrier, C., G. Lebrun, et O. Lezoray (2006). A machine learning-based color image quality metric. In *Third European Conference on Color Graphics, Imaging, and Vision*, pp. 251–256.
- Gabarda, S. et G. Cristbal (2007). Blind image quality assessment through anisotropy. *Journal of Optical Society of America*, B42–B51.
- Gao, X., S. Li, R. Liu, et P. Zhang (2007). Standardization of face image sample quality. In *International Conference on Biometrics (ICB'07)*, pp. 242–251.
- Grother, P. et E. Tabassi (2007). Performance of biometric quality measures. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 29, 531–543.
- He, Q., Z. Sun, T. Tan, et Y. Zou (2008). A hierarchical model for the evaluation of biometric sample quality. In *International Conference on Pattern Recognition (ICPR)*, pp. 1–4.
- Hemery, B., C. Rosenberger, et H. Laurent (2007). The ENSIB database : a benchmark for face recognition. In *International Symposium on Signal Processing and its Applications (ISSPA), special session "Performance Evaluation and Benchmarking of Image and Video Processing"*.
- ISO/IEC 19795-1 (2006). Information technology – biometric performance testing and reporting – part 1 : Principles and framework.
- Jung, M., D. Lger, et M. Gzalet (2002). Univariant assessment of the quality of images. *Journal of Electronic Imaging* 11(3), 354–364.
- Krichen, E., S. Garcia Salicetti, et B. Dorizzi (2007). A new probabilistic iris quality measure for comprehensive noise detection. In *IEEE Third International Conference on Biometrics : Theory, Applications and Systems (BTAS)*, pp. 1–6.
- Ladoux, P.-O., C. Rosenberger, et B. Dorizzi (2009). Palm vein verification system based on sift matching. In *the 3rd IAPR/IEEE International Conference on Biometrics (ICB'09)*, pp. 1290–1298.
- Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision (IJCV)* 60, 91 – 110.
- Maio, D., D. Maltoni, R. Cappelli, J. L. Wayman, et A. K. Jain (2002). Fvc2002 : Second

- fingerprint verification competition. In *International Conference on Pattern Recognition (ICPR'02)*, Volume 3, pp. 811 – 814.
- Mansur, A. et Y. Kuno (2008). Specific and class object recognition for service robots through autonomous and interactive methods. *IEICE Transactions on Information and Systems E91-D*, 1793–1803.
- Martinez, A. et R. Benavente (1998). The AR face database. *CVC Tech. Report*.
- Mikolajczyk, K. et C. Schmid (2005). A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis & Machine Intelligence* 27, 1615–1630.
- Moghaddam, B. et M. Yang (2000). Sex with support vector machines. In *Neural Information Processing Systems*, pp. 960–966.
- Munder, S. et D. Gavrilu (2006). An experimental study on pedestrian classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 28(11), 1863–1868.
- of Essex, U. (1994). Faces94 database, face recognition data.
- Phillips, P., H. Moon, S. Rizvi, et P. Rauss (2000). The FERET evaluation methodology for face-recognition algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)* 22(10), 1094–1104.
- Phillips, P., H. Wechsler, J. Huang, et P. Rauss (1998). The FERET database and evaluation procedure for face recognition algorithms. *Journal of Image and Vision Computing* 16, 295–306.
- Poh, N., T. Bourlai, et J. Kittler (2010). A multimodal biometric test bed for quality-dependent, cost-sensitive and client-specific score-level fusion algorithms. *Pattern Recognition*, 1094–1105.
- Poh, N., T. Bourlai, J. Kittler, L. Allano, F. Alonso-Fernandez, O. Ambekar, J. Baker, B. Dorizzi, O. Fatukasi, J. Fierrez, H. Ganster, J. Ortega-Garcia, D. Maurer, A. A. Salah, T. Scheidat, et C. Vielhauer (2009). Benchmarking quality-dependent and cost-sensitive score-level multimodal biometric fusion algorithms. *Transactions on Information Forensics and Security* 4(4), 849–866.
- Qiang, L. et Z. Wang (2009). Reduced-reference image quality assessment using divisive normalization-based image representation. *IEEE Journal of Selected Topics in Signal Processing* 3, 202–211.
- Rosenberger, C. et L. Brun (2008). Similarity-based matching for face authentication. *International Conference on Pattern Recognition (ICPR'08)*, 1–4.
- Saad, M., A. C. Bovik, et C. Charrier (2010). A DCT statistics-based blind image quality index. *IEEE Signal Processing Letters* 17(6), 583–586.
- Sang, J., Z. Lei, et S. Z. Li (2009). Face image quality evaluation for ISO/IEC standards 19794-5 and 29794-5. In *Proceedings of the Third International Conference on Advances in Biometrics (ICB)*, pp. 229–238.
- Saporta, G. (1990). *Probabilités, Analyse des données et Statistiques*. Editions Technip.
- Sazzad, Z. M. P., Y. Kawayoke, et Y. Horita (2008). No-reference image quality assessment for JPEG2000 based on spatial features. *Signal Processing : Image Communication* 23, 257–268.

- Tabassi, E. et C. Wilson (2005). A novel approach to fingerprint image quality. In *International Conference on Image Processing (ICIP)*, pp. 37–40.
- Vapnik, V. (1995). *The Nature of Statistical Learning Theory*. Springer-Verlag.
- Wang, Z., A. C. Bovik, et B. L. Evans (2000). Blind measurement of blocking artifacts in images. In *IEEE International Conference on Image Processing (ICIP)*, Volume 3, pp. 981–984.
- Wang, Z., A. C. Bovik, H. R. Sheikh, et E. P. Simoncelli (2004). Image quality assessment : From error visibility to structural similarity. *IEEE Transactions on Image Processing* 13(4), 600–612.
- Wang, Z., H. R. Sheikh, et A. C. Bovik (2002). No-reference perceptual quality assessment of jpeg compressed images. In *IEEE International Conference on Image Processing*, Volume 1, pp. 477–480.
- Watson, C. I., M. D. Garris, E. Tabassi, C. L. Wilson, R. M. McCabe, S. Janet, et K. Ko (2007). Users's Guide to NIST Biometric Image Software (NBIS). Technical report, National Institute of Standards and Technology (NIST).
- Zhang, G. et Y. Wang (2009). Asymmetry-based quality assessment of face images. In *Proceedings of the 5th International Symposium on Advances in Visual Computing (ISVC)*, Volume 5876, pp. 499–508.

## Summary

The quality assessment of biometric raw data is a primary factor within the biometric process. In this paper, we propose a generic method to evaluate biometric raw data. It is based on the joint use of two types of informations: 1) the image quality, and 2) the pattern-based quality using the Scale Invariant Feature Transformation (SIFT) descriptor. Five databases (four face and one fingerprint), and a biometric authentication system are used to quantify the efficiency of the proposed method. Experimental results show the efficiency of the proposed method in detecting several types of real alterations, that may deeply influence the overall performance of biometric systems. It also show that the proposed method is more efficient than the NIST Fingerprint Image Quality (NFIQ) method in predicting the performances of the tested system.