

# Un modèle SVM pour l'évaluation de la qualité de données biométriques

Mohamad EL-ABED, Baptiste HEMERY, Christophe CHARRIER, Christophe ROSENBERGER

Université de Caen Basse-Normandie, UMR 6072 GREYC, F-14032 Caen, France

ENSICAEN, UMR 6072 GREYC, F-14050 Caen, France

CNRS, UMR 6072 GREYC, F-14032 Caen, France

{mohamad.elabed, baptiste.hemery, christophe.rosenberger}@ensicaen.fr

christophe.charrier@unicaen.fr

**Résumé** – Ce papier présente une méthode générique (*i.e.*, indépendante de la modalité) d'évaluation de la qualité de données biométriques. 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 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 *Fingerprint Image Quality* (NFIQ) proposée par le NIST pour prédire les performances du système biométrique testé.

**Abstract** – This paper presents a generic method (*i.e.*, modality-independent) to evaluate biometric raw data. Five databases (four face and one fingerprint), and a biometric authentication system are used to quantify the relevance 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 shows that the proposed method is more efficient than the *Fingerprint Image Quality* (NFIQ) method proposed by the NIST in predicting the performances of the tested system.

## 1 Introduction

L'évaluation de la qualité des données biométriques est très importante dans la mesure où celle-ci permet d'améliorer la performance d'un système biométrique [1]. En utilisant ce type d'information, les données biométriques de mauvaise qualité peuvent 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. En biométrie, les travaux effectués sur la qualité sont beaucoup moins abordés que les recherches portant sur l'extraction de paramètres et la reconnaissance. La plupart de ces travaux sont dépendants de la modalité utilisée [2, 3, 4] et du système de vérification. Un exemple des méthodes visant à quantifier la qualité d'empreintes digitales est présenté dans [5]. D'autres méthodes [6] basées sur la distribution des scores des utilisateurs légitimes et d'imposteurs requièrent un certain nombre d'images pour chaque classe, ce qui limite leur utilisation d'une manière directe sur les images acquises. L'objectif de ce papier est de présenter une nouvelle approche pour quantifier la qualité de données biométriques. L'approche proposée possède l'avantage d'être indépendante de la modalité, et du système de vérification utilisé. Dans la suite de ce papier, la section 2 décrit l'approche retenue, ainsi que dans la section 3 son protocole d'évaluation. Les résultats expérimentaux sont donnés dans la section 4. Enfin, la section 5 conclut

ce papier en fournissant quelques perspectives de ces travaux.

## 2 Méthode développée

La méthode proposée consiste à quantifier la qualité d'une donnée biométrique en utilisant deux types d'informations complémentaires. Le principe retenu est le suivant : suite au calcul d'un critère de qualité d'image (section 2.1) et de plusieurs critères de qualité du descripteur (section 2.2), un processus de classification par apprentissage statistique est opéré à partir de l'ensemble des critères calculés.

### 2.1 Qualité image sans référence

La métrique de qualité utilisée est l'indice *BLind Image Integrity Notator using DCT Statistics* (BLIINDS) [7]. Cet indice est basé sur le calcul de quatre facteurs de dégradation dans le domaine de la DCT 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  $v_1$  est calculé en utilisant les valeurs de contraste local de chaque bloc. Le contraste local du  $k$ -ième bloc est donné par :

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

avec  $N$  est la taille du bloc,  $x_{DC}$  représente le coefficient  $DC$  et l'ensemble  $\{x_{AC}^i \mid i = 1 : N\}$  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) Distorsion 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 du  $k^{\text{ème}}$  bloc est 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 distorsion 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  $v_4$ ) : les auteurs [8] 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 :

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

Dans cette étude, on fixera  $\beta$  à 3. Etant 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  $\sum_{j=1}^4 \sum_{i=1}^M \alpha_j^i = 1$  et  $L$  représente le nombre de niveaux de décomposition utilisée. 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 [7].

## 2.2 Qualité du descripteur

La mesure de la qualité d'un descripteur est basée sur des mesures statistiques de points d'intérêt car ils décrivent de façon stable les régions de l'image où l'information est importante. Pour le calcul du vecteur descripteur au voisinage des points détectés, il existe de nombreuses méthodes tels que *Scale Invariant Feature Transformation (SIFT)* [9], *Shape Contexts* [10], etc. Parmi ces algorithmes, l'algorithme SIFT [9] est retenu pour deux raisons principales. Premièrement, l'algorithme SIFT est robuste au changement d'échelle et à la rotation 2D. Deuxièmement, une étude comparative [11] de différents descripteurs montre que SIFT est le plus performant. En utilisant SIFT, l'image  $I$  est ainsi caractérisée par l'ensemble  $Y(I) = \{k_i = (x_i, y_i, e_i, o_i, v_i) \mid i = 1 : N(I)\}$  avec :  $N(I)$  le nombre de points d'intérêt détectés dans  $I$ ;  $(x_i, y_i)$  la position du point d'intérêt  $i$  dans  $I$ ;  $(e_i, o_i)$  l'échelle et l'orientation du point d'intérêt  $i$ ; et  $v_i$  le vecteur, à 128 éléments, de descripteurs du point d'intérêt  $i$ .

En utilisant le coefficient de corrélation linéaire de Pearson, nous avons choisi d'utiliser quatre critères qui ont une forte corrélation (de 0,35 jusqu'à 0,8) avec les niveaux d'altérations. Les critères ainsi retenus pour prédire la qualité du descripteur sont : 1) le nombre de points d'intérêt détectés dans l'image  $I$ ; 2) le coefficient DC de la matrice  $M$ , 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 d'échelles des points d'intérêt détectés dans  $I$ .

Finalement, nous disposons de 5 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 l'algorithme de classification SVM [12] à 10 classes. La *classe 1* correspond à une image de référence. Tandis que les *classes allant de 2 jusqu'à 10* correspondent à 3 types d'altérations et 3 niveaux pour chaque type d'altération, respectivement. Dans cet article, 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`<sup>1</sup>. Une recherche exhaustive (*grid-search*) est effectuée pour la recherche des deux paramètres optimaux  $C$  et  $\gamma$ , et le noyau utilisé est le noyau *Radial Basis Function (RBF)*.

## 3 Protocole expérimental

Les bases de référence utilisées sont quatre bases de visages *FACES* [13], *ENSIB* [14], *FERET* [15], *AR* [16] et une d'empreintes digitales *FVC2002 DB<sub>2</sub>* [17]. Le système de vérification utilisé est un système de reconnaissance faciale (*GREYC-Face* [18]) basé sur la détection des points d'intérêt et le descripteur SIFT. 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 : alté-

1. <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>

ration par 1) mouvement (flou), 2) bruit gaussien et 3) distance d’acquisition (redimensionnement). La figure 1 illustre un exemple d’altération par flou sur une image de la base *FACES94*.



FIGURE 1 – Altération par flou : du gauche à droite, image de référence ensuite images altérées niveau 1, 2 et 3, respectivement.

Selon [1], les méthodes de qualité doivent 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 en fonction du taux d’erreur associé à cette donnée. Afin de quantifier la performance de la méthode proposée, nous procédons comme suit :

- **Apprentissage par 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 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.*, même nombre d’exemples par classe existe dans les deux ensembles).

- **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 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, le taux d’égale erreur (EER) est utilisé pour illustrer la performance globale d’un système biométrique. L’EER est égal à la valeur particulière pour laquelle le taux de fausses acceptations (FAR) et le taux de faux rejets (FRR) sont identiques. En d’autres termes, plus l’EER est faible, plus le système est performant ;

- **Corrélation des 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.

## 4 Résultats expérimentaux

### 4.1 Validation de la méthode

Les performances des six SVM multi-classes générés (cinq pour les bases de visages et un pour la base d’empreintes digitales) sont donnés dans le tableau 1. Nous avons mis le symbole « $\times$ » pour la base FVC2002  $DB_2$ , car nous avons généré un seul SVM multi-classes pour cette base. Le tableau 1 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 testé contre les altérations introduites dans la section 3.1. Pour ce faire, nous avons calculé l’EER de chaque base de références et ensuite les valeurs d’EER en ajoutant des images dégradées. Les valeurs d’EER sont calculés en utilisant la première image de référence pour l’enrôlement, et les autres pour le test. Pour chaque base utilisée, nous avons obtenues des valeurs d’EER croissantes. Cela montre que toutes les altérations introduites ont un impact sur la performance du système biométrique étudié. Par conséquent, nous définissons dans le tableau 2 les catégories de qualité retenues pour le système biométrique utilisé. La figure 2 présente les valeurs d’EER de chaque catégorie de qualité. 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 de l’EER).

	$SVM_{chaque}$		$SVM_{tout}$	
	$S_{apprentissage}$	$S_{test}$	$S_{apprentissage}$	$S_{test}$
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
$DB_2$	$\times$	$\times$	91.7	83.68

TABLE 1 – Précision (en %) des modèles SVM multi-classes sur les deux ensembles d’apprentissage ( $S_{apprentissage}$ ) et de test ( $S_{test}$ ).

Catégorie de qualité	Label prédit par le SVM multi-classes	Description
I	1	Bonne
II	2, 5 et 8	Moyenne
III	3, 6 et 9	Mauvaise
IV	4, 7 et 10	Très mauvaise

TABLE 2 – Catégories de qualité.

### 4.2 Comparaison avec NFIQ

Nous avons utilisé la base d’empreintes digitales FVC2002  $DB_2$  (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 [4]. Afin de comparer ces deux algorithmes de qualité, nous avons suivi la démarche suggérée par [1]. Nous avons utilisé le test de Kolmogorov-Smirnov (KS) 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, 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 3 décrit les valeurs statistiques du test KS. La méthode proposée est efficace pour mieux séparer les deux distributions des scores intra et inter que la métrique NFIQ. Pour la catégorie IV, la méthode NFIQ est légèrement plus efficace (sta-

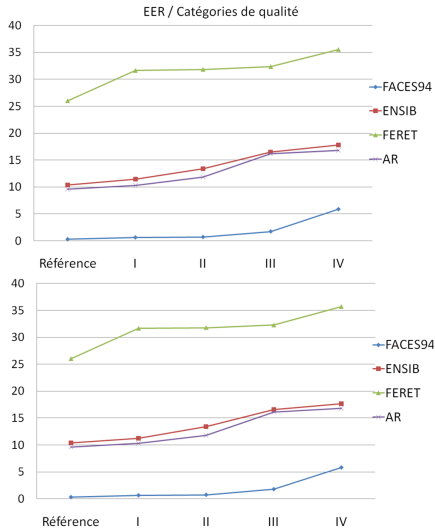


FIGURE 2 – Les valeurs d’EER des quatre bases de référence, et de chaque catégorie de qualité. Ces valeurs sont calculées en utilisant les quatre SVM multi-classes (en haut) et le SVM multi-classes généré à partir des exemples de toutes les bases (en bas), respectivement.

tistique 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
Méthode proposée	0.869	0.828	0.797	0.626
NFIQ	0.820	0.698	0.632	0.640

TABLE 3 – Comparaison entre la méthode proposée et NFIQ. Test de Kolmogorov-Smirnov (KS).

## 5 Conclusion et perspectives

Nous avons présenté dans cet article, une méthode pour prédire la qualité de données biométriques. 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>.

Pour les perspectives, nous comptons ajouter un sixième critère pour détecter l’altération par luminance, qui a un impact significatif sur la plupart des systèmes de reconnaissance faciale existants. D’autre part, la métrique NFIQ 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 est

utile.

## Références

- [1] P. Grother and E. Tabassi. Performance of biometric quality measures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29 :531–543, 2007.
- [2] E. Krichen, S. Garcia Salicetti, and B. Dorizzi. A new probabilistic iris quality measure for comprehensive noise detection. In *BTAS07*, pages 1–6, 2007.
- [3] X.F. Gao, S.Z. Li, R. Liu, and P.R. Zhang. Standardization of face image sample quality. In *International Conference on Biometrics (ICB’07)*, pages 242–251, 2007.
- [4] E. Tabassi and C.L. Wilson. A novel approach to fingerprint image quality. In *ICIP05*, pages 37–40, 2005.
- [5] F. Alonso-Fernandez, J. Fierrez, J. Ortega-Garcia, J. Gonzalez-Rodriguez, H. Fronthaler, K. Kollreider, and J. Bigun. A comparative study of fingerprint image-quality estimation methods. *IEEE Trans. on Information Forensics and Security*, 2 :734–743, 2007.
- [6] Q. He, Z.A. Sun, T.N. Tan, and Y. Zou. A hierarchical model for the evaluation of biometric sample quality. In *ICPR08*, pages 1–4, 2008.
- [7] M. Saad, A. C. Bovik, and C. Charrier. A DCT statistics-based blind image quality index. *IEEE Signal Processing Letters*, 17(6) :583–586, 2010.
- [8] S. Gabarda and G. Cristbal. Blind image quality assessment through anisotropy. *Journal of Optical Society of America*, pages B42–B51, 2007.
- [9] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision*, 60 :91 – 110, 2004.
- [10] S. Belongie, J. Malik, and J. Puzicha. Matching shapes. In *International Conference on Computer Vision*, pages 454 – 461, 2001.
- [11] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 27 :1615–1630, 2005.
- [12] V. Vapnik. *The Nature of Statistical Learning Theory*. Springer, 1995.
- [13] University of Essex. Faces94 database, face recognition data, 1994.
- [14] B. Hemery, C. Rosenberger, and H. Laurent. The ENSIB database : a benchmark for face recognition. In *ISSPA, special session "Performance Evaluation and Benchmarking of Image and Video Processing"*, 2007.
- [15] P.J. Phillips, H. Wechsler, J. Huang, and P. Rauss. The FERET database and evaluation procedure for face recognition algorithms. *Journal of Image and Vision Computing*, 16 :295–306, 1998.
- [16] A.M. Martinez and R. Benavente. The AR face database. *CVC Tech. Report*, 1998.
- [17] D. Maio, D. Maltoni, R. Cappelli, J. L. Wayman, and A. K. Jain. Fvc2002 : Second fingerprint verification competition. In *International Conference on Pattern Recognition (ICPR’02)*, volume 3, pages 811 – 814, 2002.
- [18] C. Rosenberger and L. Brun. Similarity-based matching for face authentication. *International Conference on Pattern Recognition (ICPR’2008)*, pages 1–4, 2008.