



HAL
open science

Validation de Métriques de Qualité de Données Biométriques

Zhigang M Yao, Jean-Marie M Lebars, Christophe Charrier, Christophe
Rosenberger

► **To cite this version:**

Zhigang M Yao, Jean-Marie M Lebars, Christophe Charrier, Christophe Rosenberger. Validation de Métriques de Qualité de Données Biométriques. Colloque COMpression et REprésentation des Signaux Audiovisuels (CORESA), Nov 2017, Caen, France. hal-01591522

HAL Id: hal-01591522

<https://hal.science/hal-01591522v1>

Submitted on 5 Oct 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Validation de Métriques de Qualité de Données Biométriques

Z. Yao

J.M Lebars

C. Charrier

C. Rosenberger

Normandie Univ, UNICAEN, ENSICAEN, CNRS, GREYC, 14000 Caen, France

christophe.rosenberger@ensicaen.fr

Résumé

Le contrôle qualité de données biométriques fonctionne comme un péage pour décider si un échantillon biométrique peut être exploité pour définir le modèle d'un individu. Prendre en compte la qualité des données biométriques à l'enrôlement a un impact important sur la performance ultérieure du système biométrique. Des métriques sont proposées dans cet objectif par les chercheurs, leur pertinence doit être étudiée notamment pour évaluer leur utilité. Dans ce papier, nous proposons un cadre méthodologique permettant d'estimer la pertinence d'une métrique de qualité de données biométriques. La méthode proposée permet de comparer de façon objective deux métriques de qualité. Nous illustrons l'apport de cette méthodologie sur des empreintes digitales.

Mots clefs

Qualité d'une donnée biométrique, validation d'une métrique de qualité, empreinte digitale.

1 Introduction

La biométrie a pour objectif d'identifier ou de vérifier l'identité d'un individu à partir de caractéristiques morphologiques (visage, empreinte digitale, ...) ou comportementales (dynamique de frappe au clavier, voix, ...). A cette fin, le système biométrique exploite la référence biométrique de l'individu construite à partir d'une ou plusieurs captures de la donnée biométrique de l'individu. Afin de garantir les meilleures performances du système biométrique, on essaie en général de collecter pendant l'enrôlement la donnée biométrique avec la meilleure qualité possible. Ainsi, la photo du visage fournie lors de la création d'un passeport doit suivre plusieurs contraintes [1] pour être utilisée. De même, lors de l'acquisition d'une empreinte digitale pour le passeport, une métrique est utilisée pour garantir un niveau de qualité suffisant de la référence biométrique de l'individu [2]. Il s'agit en général de la métrique NFIQ (NIST Fingerprint Image Quality) proposée par le NIST en 2004.

La qualité d'une donnée biométrique dépend de plusieurs facteurs dont les conditions d'acquisition (environnement, bonne utilisation du capteur, ...), du capteur biométrique (problème de réglage, ...) ou de la donnée biométrique (doigt abîmé, partie du visage occultée, ...). La figure



Figure 1 – Exemples d'empreintes digitales de qualités différentes (normal, doigt sec, doigt mouillé, plis, rides)

1 présente quelques exemples d'empreinte digitale de différentes qualités. Il est clair que la performance d'un algorithme de comparaison d'empreintes digitales sera différente dans tous ces cas.

Dans la dernière décennie, plusieurs travaux de recherche ont porté sur la définition de métriques de qualité de données biométriques pour le visage [3, 4], le réseau veineux [5] et surtout l'empreinte digitale [2, 6, 7]. Plusieurs problématiques demeurent quelque ce soit la modalité biométrique considérée : comment valider la pertinence d'une métrique de qualité d'une donnée biométrique? comment démontrer objectivement la supériorité d'une métrique par rapport à une autre? La principale contribution de ce papier est de proposer un cadre méthodologique visant à quantifier la pertinence d'une métrique de qualité d'une donnée biométrique. L'approche proposée est générique et elle peut être appliquée sur n'importe quelle modalité biométrique. L'autre avantage de la méthode est de pouvoir comparer de façon objective deux métriques de qualité. La méthode permet également d'estimer dans quelle mesure une métrique est proche d'un jugement optimal, l'indicateur de pertinence est donc normalisé.

Le papier est organisé comme suit. La section 2 présente les méthodes de l'état de l'art pour analyser la performance d'une métrique de qualité d'une donnée biométrique. La

méthode proposée est décrite dans la section 3. La section 4 illustre l'application de cette méthodologie pour comparer deux métriques de qualité d'empreintes digitales. La section 5 conclue ce travail et propose différentes perspectives.

2 Travaux antérieurs

Phillips et al. [8] ont mis en évidence l'importance de l'évaluation des systèmes biométriques. Cependant, certaines études antérieures sur l'évaluation de la qualité de données biométriques ne fournissent pas de protocoles totalement ouverts [9]. Cette méthode largement utilisée (notamment pour la normalisation de métriques) repose principalement l'étude de la corrélation statistique de scores légitimes et la valeur de la métrique des échantillons. Ce type d'approche nécessite une base de données biométriques très conséquente (plusieurs millions d'échantillons) difficilement accessible aux chercheurs.

D'autres études évaluent l'intérêt d'une métrique de qualité vis à vis d'éléments d'évaluation subjective impliquant le jugement humain [10]. Ratha et Bolle ont également proposé de mesurer la pertinence d'une métrique de qualité de données biométriques par comparaison statistique (avec le facteur de corrélation de Pearson) avec les valeurs d'autres métriques de qualité [11]. De même, Shen et al. [12] ont proposé d'affecter les empreintes digitales d'une base de données suivant leur qualité évaluée par une métrique et de comparer ces classes avec d'autres générées par d'autres métriques de qualité. Ces approches ne permettent pas de quantifier la relation entre la métrique de qualité et les performances correspondantes. En outre, ces tentatives sont plus ou moins liées aux observations subjectives lors de la validation de métriques de qualité. Cependant, ce type d'opération pourrait être facilement estimée en utilisant un générateur d'empreintes synthétiques tel que SFinGe [13].

Tabassi et al. [14] définissent la qualité de l'échantillon biométrique comme un moyen de prédire la performance de reconnaissance en observant que des échantillons biométriques de bonne qualité produisent des scores légitimes relativement élevés et bien séparés des scores des imposteurs. Cependant, la prédiction dépend totalement de la performance de l'algorithme de comparaison utilisé. Grother et al. [9] ont proposé plusieurs approches d'évaluation associées aux niveaux de seuil de décision et de qualité correspondante, y compris la courbe DET et l'approche basée sur le test de Kolmogorov Smirnov (KS).

Ces dernières approches sont intéressantes mais sont difficiles à interpréter (car basées sur un test d'hypothèse), pas vraiment adaptées pour comparer quantitativement deux métriques de qualité. Nous proposons une méthode répondant à ces deux limitations et permettant également d'estimer la pertinence optimale à atteindre pour une base de données biométriques et un algorithme de comparaison.

3 Méthode proposée

Le principe de la méthode proposée consiste à évaluer la pertinence de la métrique de qualité en se focalisant sur la tâche d'enrôlement d'un utilisateur (qui correspond à l'usage le plus important des métriques de qualité). Nous supposons donc avoir une base de d'échantillons biométriques de plusieurs utilisateurs (voir Figure 2) et un algorithme de comparaison de ces échantillons.

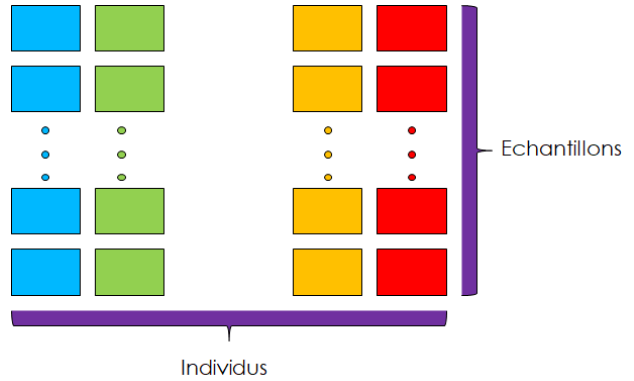
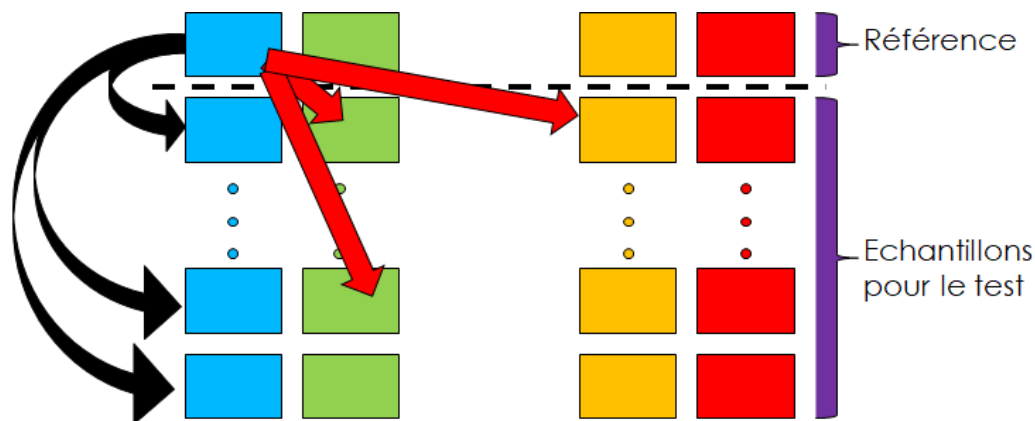


Figure 2 – Description d'une base de données biométriques.

Si l'on souhaite évaluer la performance d'un système biométrique sur cette base, il faut appliquer une heuristique pour choisir l'échantillon de référence de l'utilisateur. Une fois ce choix réalisé, nous calculons l'ensemble des scores légitimes de la base en comparant les échantillons de chaque utilisateur avec Sa référence. On fait de même pour les scores d'imposture en comparant une référence avec les échantillons de tous les autres individus de la base (voir Figure 3). Ainsi, si on a une base composée de N individus avec M échantillons par individu, par ce procédé, on génère $N \times (M - 1)$ scores légitimes et $N \times (M - 1) \times (N - 1)$ scores d'imposture. Ces scores permettent de calculer le taux de faux rejet et de fausses acceptations pour différentes valeurs du seuil de décision (à partir duquel on considère la preuve de vérification comme suffisante). Ceci permet de calculer différentes mesures classiques en biométrie : la courbe DET (évolution du faux rejet en fonction de la fausse acceptation), le taux d'égal erreur (EER) et l'aire en dessous de la courbe DET (AUC).

Revenons à présent sur le choix de l'échantillon comme référence de l'utilisateur, il existe plusieurs possibilités :

- Choix du premier échantillon comme référence de l'individu. C'est une pratique courante considérée comme un choix par défaut,
- Choix de la référence à partir d'une métrique de qualité. Dans ce cas, les échantillons sont évalués par cette métrique et le "meilleur" échantillon (au vu de cette métrique) est sélectionné comme référence de l'utilisateur,



Scores légitimes : comparaison entre un échantillon et la référence de l'utilisateur légitime

Scores d'imposture : comparaison entre un échantillon et la référence d'un autre utilisateur

Figure 3 – Explication du calcul des scores légitimes et d'imposture à partir d'une base de données biométriques.

- Choix du meilleur échantillon : En général, une base de données biométriques contient un nombre fini et faible d'échantillons M de chaque individu. Il est possible de tester toutes les possibilités, c'est à dire estimer la valeur de l'AUC obtenue pour chaque possibilité d'échantillon comme référence. Il est donc possible de déterminer le choix optimal de l'échantillon de référence au vu de la performance du système biométrique (valeur minimale de l'AUC),
- Choix du plus mauvais échantillon : on reprend la procédure précédente en prenant l'échantillon conduisant à la valeur la plus élevée de l'AUC.

En appliquant ces différentes heuristiques, on peut mesurer la performance du système biométrique pour une base de données biométriques et un algorithme de comparaison. Nous considérons dans cette étude la valeur de l'AUC qui est une mesure globale de performance à minimiser. Nous pouvons calculer $AUC_{meilleure}$ la valeur de l'AUC avec le choix optimal de la référence, $AUC_{mauvais}$ la valeur de l'AUC avec le choix de la plus mauvaise référence et $AUC_{métrique}$ la valeur de l'AUC avec le choix de la référence guidée par une métrique de qualité. Nous proposons de calculer la pertinence d'une métrique par la formule suivante :

$$P = 1 - \frac{(AUC_{métrique} - AUC_{meilleure})}{(AUC_{mauvais} - AUC_{meilleure})} \quad (1)$$

La figure 5 présente la performance obtenue sur une base de données biométriques pour différentes heuristiques de choix de la référence. En noir est représentée la courbe DET en prenant l'échantillon le plus mauvais comme ré-

férence et en vert le meilleur choix. Cette figure montre que le choix de la référence va conduire à une performance du système entre un AUC de 0.0352 à 0.2338. En utilisant deux métriques, on obtient une performance de 0.0991 (bleu) et l'autre de 0.0788 (rouge). Cette figure donne deux informations, la première est que la métrique 1 (courbe bleue) est moins efficace que la métrique 2 (courbe rouge). En effet, la métrique 2 permet une meilleure performance du système biométrique. Deuxièmement, la métrique 2 a une pertinence de $P = 78\%$ contre $P = 67\%$ pour l'autre métrique. Cela signifie qu'il y a encore des possibilités d'amélioration pour effectuer le meilleur choix de la référence.

4 Application

Nous illustrons dans cette section l'intérêt de la méthode de validation d'une métrique de qualité. Nous présentons le résultat d'analyse de la pertinence d'une métrique de qualité récente d'empreintes digitales [15]. La métrique en question consiste à mesurer le nombre de pixels de bonne qualité dans une image d'empreinte digitale. Un pixel est considéré comme de mauvaise qualité s'il fait partie de l'arrière plan de l'empreinte (après détection des crêtes de l'empreinte) ou s'il appartient à un bloc dont le gradient calculé sur l'image des crêtes n'est pas uniforme (voir Figure 6).

Nous avons utilisé 5 bases de données biométriques issues de la compétition FVC [16] avec l'algorithme de comparaison Bozorth3 proposé par le NIST [17]. Nous avons appliqué la méthodologie présentée dans la section précédente. Nous avons comparé cette métrique avec NFIQ [2].

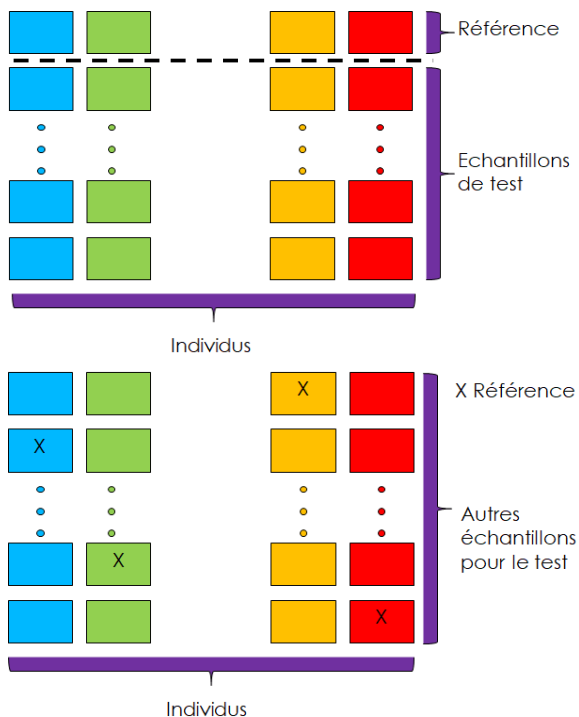


Figure 4 – Choix de l'échantillon de référence dans une base de données biométriques

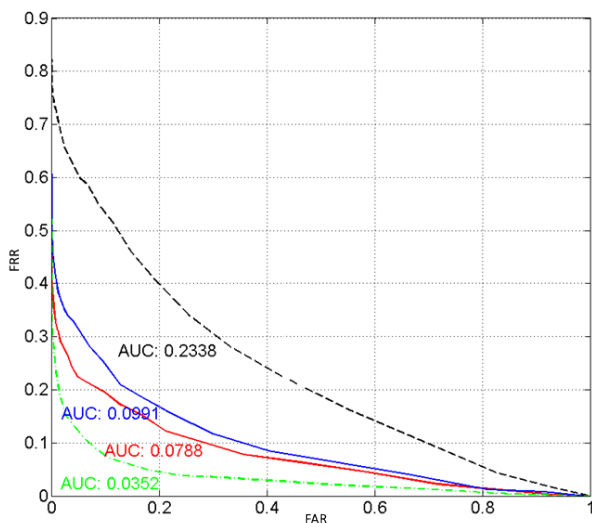


Figure 5 – Représentation de la performance en fonction du choix de la référence : plus mauvais choix (noir), meilleur choix (vert), choix par la métrique 1 (bleu) et choix par la métrique 2 (rouge).

Le tableau 5 présente la valeur obtenue de l'AUC sur les 5 bases de données biométriques en réalisant le choix de la référence de chaque individu à partir de la métrique NFIQ. Nous utilisons la métrique MSEG [15]. On peut constater

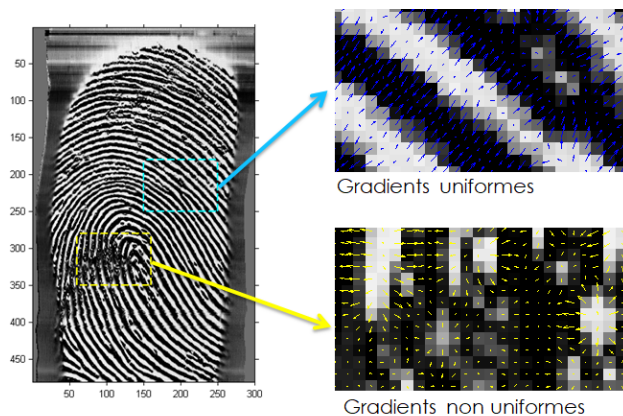


Figure 6 – Principe de la métrique de qualité d'une empreinte digitale.

que cette seconde métrique permet d'obtenir de meilleures performances sur les 3 dernières bases de données biométriques. Sur les deux premières, NFIQ donne un résultat légèrement meilleur sur des bases plus simples (avec un AUC très proche de 0%). Dans la mesure où plusieurs bases de données biométriques ont été utilisées, on peut conclure que cette métrique propose une amélioration de NFIQ.

5 Conclusion et perspectives

Nous avons proposé dans ce papier une méthodologie d'analyse de la pertinence d'une métrique de qualité de données biométriques. Le principe est de quantifier la performance du choix de l'échantillon de référence d'un individu. Nous avons pu mettre en évidence le caractère générique de l'approche et l'intérêt opérationnel de pouvoir quantifier la pertinence d'une métrique. L'intérêt de cette méthodologie a été illustrée pour comparer deux métriques de l'état de l'art sur cinq bases d'empreintes digitales.

Les perspectives de cette étude visent à proposer une métrique de qualité d'empreinte digitale par optimisation de la valeur P .

Références

- [1] Matteo Ferrara, Annalisa Franco, et Davide Maltoni. Evaluating systems assessing face-image compliance with icao/iso standards. Dans *European Workshop on Biometrics and Identity Management*, pages 191–199. Springer, 2008.
- [2] E Tabassi, CL Wilson, et C Watson. Fingerprint image quality (nfiq). nistir 7151. august 2004, 2011.
- [3] Kamal Nasrollahi et Thomas B Moeslund. Face quality assessment system in video sequences. Dans *European Workshop on Biometrics and Identity Management*, pages 10–18. Springer, 2008.

Métrique	00DB2	02DB2	04DB1	04DB2	04DB3
NFIQ	0.22%	0.11%	2.66%	3.86%	1.89%
MSEG	0.10%	0.20%	1.93%	3.24%	1.51%

Tableau 1 – Valeur de l’AUC obtenue sur 5 bases de données biométriques en réalisant le choix de la référence de chaque individu à partir des métriques NFIQ et MSEG.



Figure 7 – Exemples d’empreinte digitale issus des 5 bases de données biométriques.

[4] Pankaj Wasnik, Kiran B Raja, Raghavendra Ramachandra, et Christoph Busch. Assessing face image quality for smartphone based face recognition system. Dans *Biometrics and Forensics (IWBF), 2017 5th International Workshop on*, pages 1–6. IEEE, 2017.

[5] Huafeng Qin et Mounim A El Yacoubi. Deep representation for finger-vein image quality assessment. *IEEE Transactions on Circuits and Systems for Video Technology*, 2017.

[6] Zhigang Yao, Jean-Marie Le Bars, Christophe Charrier, et Christophe Rosenberger. Fingerprint quality assessment combining blind image quality, texture and minutiae features. Dans *ICISSP 2015*, 2015.

[7] Xinwei Liu, Marius Pedersen, Christophe Charrier, Patrick Bours, et Christoph Busch. The influence of fingerprint image degradations on the performance of biometric system and quality assessment. Dans *Biometrics Special Interest Group (BIOSIG), 2016 International Conference of the*, pages 1–6. IEEE, 2016.

[8] P Jonathon Phillips, Alvin Martin, Charles L Wilson, et Mark Przybocki. An introduction evaluating biometric systems. *Computer*, 33(2) :56–63, 2000.

[9] Patrick Grother et Elham Tabassi. Performance of biometric quality measures. *IEEE transactions on pattern analysis and machine intelligence*, 29(4), 2007.

[10] Tai Pang Chen, Xudong Jiang, et Wei-Yun Yau. Fingerprint image quality analysis. Dans *Image Processing, 2004. ICIP’04. 2004 International Conference on*, volume 2, pages 1253–1256. IEEE, 2004.

[11] Nalini K Ratha et Ruud Bolle. *Fingerprint image quality estimation*. IBM TJ Watson Research Center, 1999.

[12] LinLin Shen, Alex Kot, et Wai Mun Koo. Quality measures of fingerprint images. Dans *AVBPA*, pages 266–271. Springer, 2001.

[13] Raffaele Cappelli, D Maio, et D Maltoni. Sfinge : an approach to synthetic fingerprint generation. Dans *International Workshop on Biometric Technologies (BT2004)*, pages 147–154, 2004.

[14] Elham Tabassi et Patrick Grother. Fingerprint image quality. Dans *Encyclopedia of Biometrics*, pages 482–490. Springer, 2009.

[15] Z Yao, Christophe Charrier, et Christophe Rosenberger. Pixel pruning for fingerprint quality assessment. Dans *International Biometric Performance Testing Conference (IBPC)*, 2016.

[16] Fingerprint verification competition databases.

[17] Patricia A Flanagan. Nist biometric image software (nbis). 2010.