

FINGERPRINT QUALITY ASSESSMENT USING A NO-REFERENCE IMAGE QUALITY METRIC

Alexandre Ninassi, Christophe Charrier, Christophe Rosenberger

Mohamad El Abed

Rafik Hariri University
Meshref, Lebanon
alabedma@rhu.edu.lb

Université de Caen, UMR 6072 GREYC
ENSICAEN, UMR 6072 GREYC
CNRS, UMR 6072 GREYC
{alexandre.ninassi, christophe.rosenberger}@ensicaen.fr
christophe.charrier@unicaen.fr

ABSTRACT

The quality assessment of the acquired biometric raw data is very important as it deeply affects the performance of biometric systems and consequently their usability. Poor quality samples increase the enrolment failures, and decrease the system performance. In this paper, we present a new quality assessment metric of fingerprints. Its main originality lies in the use of a no-reference image quality metric. The proposed quality metric combines two types of parameters through a weighted sum optimized by a genetic algorithm: 1) image quality criterion and 2) pattern-based quality criteria (salient and patch-based features). BOZORTH3 matching system and the FVC2002 DB3 fingerprint database are used to clarify the benefits of the presented quality metric.

Index Terms— biometrics, fingerprint quality, genetic algorithm, NFIQ, blind image quality.

1. INTRODUCTION

Biometric authentication systems are increasingly used in real life applications such as border control, e-commerce, *etc.* Despite the obvious advantages of such systems in terms of security and usability, they are subject to errors. This verification inaccuracy is due to many reasons such as the variations of human characteristics (*e.g.*, occlusions, environmental factors (*e.g.*, illuminations [1]) and cross-device matching [2]). This kind of acquisition artifacts may deeply affect the performance of biometric systems and hence, decrease their use in real life applications. Therefore, controlling the quality of the acquired biometric raw data is considered as an essential step in both enrollment and verification phases. Using the quality information, poor quality samples can be removed during the enrollment phase or rejected during the verification. Such information could be also used for soft biometrics and multimodal approaches [3].

In this paper, we present a quality assessment metric of

fingerprints. The proposed metric uses two types of information: 1) image quality and 2) pattern-based quality. The outline of the paper is given as follows: Section 2 presents related previous works on quality assessment of biometric raw data and fingerprints. The proposed quality assessment metric is then given in Section 3. Section 4 describes the experimental results obtained for used fingerprint database. A conclusion and some perspectives of this work are given in Section 5.

2. RELATED WORKS

In this section, we present an overview of quality metrics of biometric raw data and especially fingerprints in the literature. The quality assessment of biometric raw data is divided into three points of view [4]:

- **Character:** refers to the quality of the physical features of the individual.
- **Fidelity:** refers to the degree of similarity between a biometric sample and its source.
- **Utility:** refers to the impact of the individual biometric sample on the overall performance of a biometric system.

In biometrics, there is an international consensus on the fact that the quality of a biometric sample should be related to its recognition performance [5]. Therefore, we present in this paper a utility-based quality assessment metric of fingerprint-based biometric raw data. Alonso-Fernandez *et al.* [6] present an overview of existing fingerprint quality metrics which are divided into three main approaches: 1) based on local features of the image, 2) based global features of the image, and 3) those that use a classifier in order to predict the quality of the biometric raw data. Hong *et al* [7] modeled the ridge and valley pattern as sine wave, and compute the amplitude, frequency as well as the variance of the sine wave to quantify the quality of a fingerprint. However,

these measurements cannot distinguish some invalid images from the valid ones. They classify regions of fingerprint as recoverable or unrecoverable so that if the percentages of recoverable blocks are smaller than the specified threshold, the input fingerprint is rejected as poor quality. Ratha and Bolle [8] proposed a method for image quality estimation from wavelet compressed fingerprint image. Lim and Yau [9] proposed algorithms for estimating the quality of fingerprint images in the spacial domain. They verify the repetition of ridge and valley patterns by the ratio of the eigenvalues of the grey level gradient block image. They used the quality of the feature extracted from the fingerprint image by the automatic fingerprint identification system as the benchmark to test the performance of their proposed image quality analysis procedure. Their system is used to flag poor quality fingerprint images.

The NIST Fingerprint Image Quality metric (NFIQ) [10] uses a vector of 11 elements based on the quality of the extracted minutiae and a learning processing by neural nets in order to optimize the combination of various image quality metrics to predict the class of the quality (1: excellent... 5: poor) for an image of a fingerprint. Recently, Merkle et al. [11] proposed a method to optimize results provided by NFIQ. Many other fingerprint quality algorithms exist [12, 13, 14]. Olsen et al [15] proposed a fingerprint quality measure based on the Gabor Filter responses, they arrived to conclude that the Gabor filter approach could be included as part of a quality feature vector in order to be able to outperform and replace the NFIQ metric. The presented methods have shown their efficiency in predicting the quality of fingerprints images. However, Alonso-Fernandez *et al.* argued that quality measures could provide complementary information, and its combination may improve the process of assessing the quality of a fingerprint image.

3. DEVELOPED METHOD

The goal of the proposed quality metric is to compute given a fingerprint image a quality value (between 0 and 100%) related to the recognition performance that could be achieved during its use as biometric reference. Our requirements is also to have a quality index linearly correlated with the matching score provided by an verification algorithm. The computation process of the quality metric is presented in Fig. 1: using one image quality criterion (Section 3.1) and the pattern-based quality criteria (Section 3.2), a genetic algorithm (Section 3.3) is performed to predict the quality of the target biometric data.

The proposed quality metric is defined as follows:

$$Q = \frac{1}{A} \sum_{i=1}^N \alpha_i \cdot C_i \quad (1)$$

Where N is the number of the retained criteria $C_i, i = 1 : N$ (in our case, $N = 11$), A normalisation constant, the α_i are the weights we will optimize with a genetic algorithm (GA).

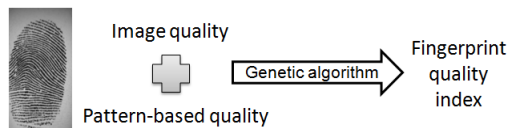


Fig. 1. General scheme of the proposed quality metric.

3.1. BLIINDS-2

Since no reference image is available for any acquired image, we used a No-Reference Image Quality Assessment (NR-IQA) algorithm to judge the quality of the resulting image. For now, NR-IQA algorithms generally follow one of three trends: 1) Distortion-specific approaches, 2) Training-based approaches and 3) Natural scene statistics (NSS) approaches. In this paper, we decided to use BLIINDS-II [16] introduced by Saad *et al.* which is a general (non-distortion specific) approach using a minimal number of features extracted entirely from the DCT-domain which is also computationally convenient. In addition, it correlates highly with human visual perception of quality. The main idea is to derive a generalized NSS-based model of local DCT coefficients, and transform the model parameters into features used for perceptual image quality score prediction. A quality score for the image is predicted by a simple Bayesian model. The Bayesian approach maximizes the probability that the image has a certain quality score given the model-based features extracted from the image.

3.2. Pattern-based criteria

We intend in this part to describe a fingerprint by considering textural features. We proposed two types of computation approach: one based on saliency points and one inside a patch.

3.2.1. Salient features

Four descriptors aiming at detecting the altered fingerprint are used with Scale Invariant Feature Transform (SIFT) [17]. SIFT descriptors provide a good description of textural information. Each image I is described by a set of invariant features: $X(I) = \{k_i = (s_i, sc_i, x_i, y_i) | i = 1 : N(I)\}$

Where:

- s_i is the 128-elements SIFT invariant descriptor computed near keypoints k_i .
- (x_i, y_i) its position in the original image I .
- sc_i its scale.
- $N(I)$ the number of detected keypoints for image.

From these features, four criteria are retained to contribute to the quality assessment of the biometric raw data:

1. The number NB of detected keypoints in the image.
2. DC coefficient of the matrix M_s (*i.e.*, Descriptors matrix).
3. The mean $MEAN_D$ and standard deviation STD_D of scales related to the keypoints detected in the image.

3.2.2. Patch-based features

The fingerprint image is split in 17×17 blocks. The root mean square is applied on each block to get $m \times n$ RMS values for each fingerprint. The Root Mean Square value measures the dispersion or spread of a group of data. The mathematical expression of the RMS value of an image I is:

$$RMS = \frac{1}{m \times n} \sqrt{\sum_{i=1}^m \sum_{j=1}^n I(i, j)} \quad (2)$$

Standard deviation shows how much variation or dispersion exists from the average. A low standard deviation indicates that the data points tend to be very close to the mean, whereas high standard deviation indicates that the data points are spread out over a large range of values. Kurtosis is a measure of whether the data are peaked or flat relative to a normal distribution. That is, data sets with high kurtosis tend to have a distinct peak near the mean, decline rather rapidly, and have heavy tails. Data sets with low kurtosis tend to have a flat top near the mean rather than a sharp peak. A uniform distribution would be the extreme case. A fingerprint image is finally described by different patch-based features:

1. NB_B : number of blocks in the fingerprint image
2. $MEAN_B$: average RMS value for all blocks
3. STD_B : standard deviation of RMS values for all blocks
4. MED_B : median value of RMS values for all blocks
5. STD_B : standard deviation of RMS values for all blocks
6. $SKEW_B$: skewness of RMS values for all blocks
7. $KURT_B$: kurtosis of RMS values for all blocks

In the next section, we make a brief presentation of genetic algorithms that are used to generate based on the previous features the quality index.

3.3. Genetic algorithm

In order to compute the quality assessment metric, we use genetic algorithms (GA) to compute the weights of each retained criterion. This produces a value between 0 and 100% defining the quality of a biometric image (*i.e.*, an image of poor quality produces a value near 0%). The GA proposed by Holland [18] are a general-purpose global optimization technique based on randomized search. They incorporate some aspects of iterative algorithm. A genetic algorithm is based on the idea that natural evolution is a search process that optimizes the structures it generates. An interesting characteristic of GA is their high efficiency for difficult search problems without being stuck in local extremum. In a GA, a population of individuals, described by some chromosomes, is iteratively updated by applying operators of selection, mutation and crossover to solve the problem. Each individual is evaluated by a fitness function that controls the population evolution in order to optimize it.

In our case, the genotype of individuals is a vector of 11 elements (corresponding of each weighing coefficients). The fitness function corresponds to the Pearson's correlation coefficient between the quality metric we optimize (defined by equation 1) and BOZORTH3 matching scores. In other words, the more the correlation is high, better is the prediction of the presented quality metric to the BOZORTH3 matching performance.

4. EXPERIMENTAL RESULTS

4.1. Protocol

In order to validate the proposed metric, we use the FVC 2002 DB3 database [19] composed of 100 users with 8 samples for each user. From this benchmark, we generated a synthetic altered fingerprint database simulating several acquisition artefacts.

4.1.1. Altered fingerprint database

Given one sample, we generated 3000 altered fingerprints by different artefacts: Gaussian noise (600), contrast (500), luminance (600), median blurring (20), rotation (360), scratches (200), occlusion (720). Note that this last alteration consists in to translate the fingerprint, we quantify in which measure the matcher is able to recognize the fingerprint when a part is missing and it is translated. Fig. 2 presents for one sample the different alterations. These alterations are significant considering the different artefacts we can encounter in real operations. In total, we generated 300.000 altered fingerprints.



Fig. 2. Example of alteration of a fingerprint (from left up to right bottom): original image, rotation, scratches, occlusion, contrast, blurring, noise, luminance.

4.1.2. Matching system

The matching score between the reference fingerprint image and the different distorted versions is computed with the NIST Biometric Image Software (NBIS) [20], where the MINDTCT minutiae detector and BOZORTH3 fingerprint matcher are used. The BOZORTH3 matcher is rotation and translation invariant, and uses only the location (x,y) and orientation (theta) of the minutia points to match the fingerprints.

4.1.3. Validation process

The validation process contains different steps:

1. Given the 300.000 scores (between the original and altered fingerprints), we first have to verify if the image alterations have an impact on the matching score. This will guarantee that the selected alterations provide different behaviors of the performance associated to the biometric system.
2. As a first study, we propose to compute Pearson's correlation coefficients between the parameters computed on each altered fingerprint and the associated matching score. This will allow us to identify the statistical dependency between computed features and the fingerprint quality (quantified through the matching score). We also use NFIQ as reference.
3. In order to compute the optimal weights using the GA, we split equally our database of altered fingerprint images into two sets $S_{training}$ and S_{test} (i.e., the first 50 persons for *training* and the rest for the *test*).

4. Quality metric efficiency in prediction BOZORTH3's performance: According to Grother and Tabassi [5], biometric quality metrics should predict the matching performance. That is, a quality metric takes a biometric raw data, and produces a class or a scalar related to error rates associated to that sample. To do so, we use the obtained Pearson's correlation coefficients on both two sets $S_{training}$ and S_{test} . More generally speaking, the more the coefficient is near 1, better is the prediction of the presented quality metric to BOZORTH3's performance.

4.2. Results

First, as mentioned in the protocol, we want to verify that all the alterations have an impact on the matching score. The alteration can be realized in one direction (like blurring) or in two directions (as for example, the luminance alteration can be done by adding or suppressing some luminance). We computed the Person's correlation factor between each feature and the matching score in order to know if all the features bring some information to quantify the quality of the fingerprint. Table 1 gives the different values in comparison with NFIQ (that is surprisely low). We see that many of them have a Pearson's correlation factor upper than 0.6 (in absolute value).

NB	DC	$MEAN_D$	STD_D
-0.3079	0.7391	-0.0109	-0.2169
NB_B	$MEAN_B$	STD_B	MED_B
-0.7036	-0.2052	-0.7767	-0.1649
$KURT_B$	$SKEW_B$	BLIINDS2	NFIQ
-0.2776	0.4547	0.6548	-0.2024

Table 1. Correlation between parameters with the matching score

We used the genetic algorithm to define the optimal weighting coefficients whose values are given here: (0.014,1,-0.209,-1,-0.002,-0.596,-0.424,0.886,-0.020,-0.33,1). After training, the Pearson's correlation factor with the matching score is 0.8540 (NFIQ provides a 0.204 value). We used then the test database to assess the Pearson's correlation factor with the matching score by applying this quality index. We obtained the same value showing its generalization capability and a good behavior of the proposed quality metric. Fig. 3 shows the evolution of the matching score versus the BLIINDS-2, NFIQ and the proposed metric. The NFIQ reveals itself as a quite metric, as for example, when the NFIQ value is 3, we can obtain all the values for the matching score. Using alone the BLIINDS-2 metric, many errors occur especially when the value of the BLIINDS-2 metric is high but the matching score could be very low. The proposed metric shows a good linear relationship with the matching score. Errors (a non linear behavior) are located where the quality (considering the matching score) is very high or very low.

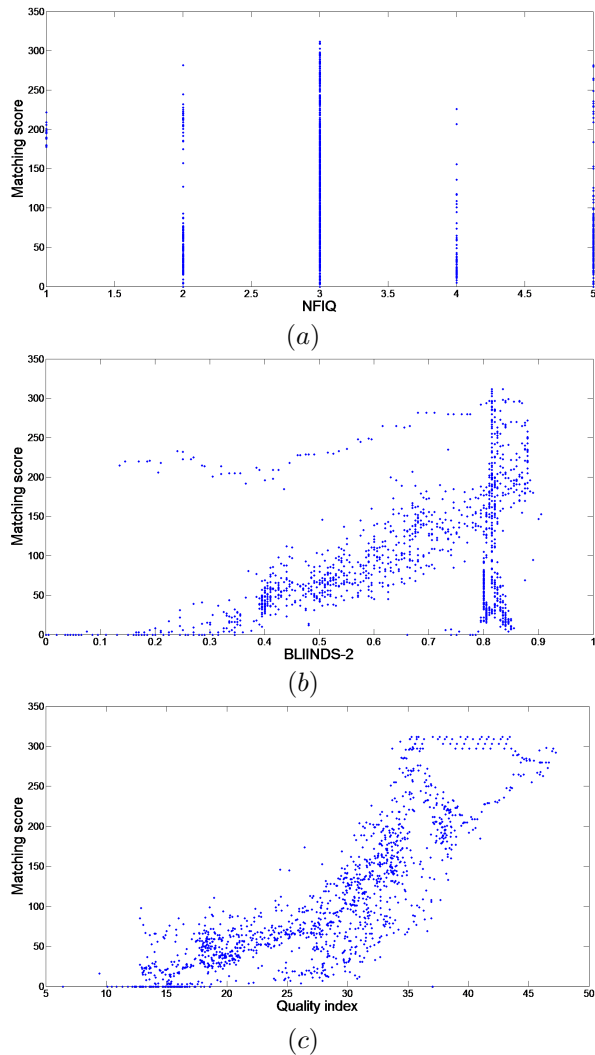


Fig. 3. Linear correlation between the matching score and (a) NFIQ, (b) BLIINDS-2, (c) proposed quality index

5. CONCLUSION AND PERSPECTIVES

In this paper, we address the problem of quantifying the quality of a fingerprint given one image. We proposed a no-reference quality index combining two types of information (image quality criterion and pattern-based criteria). Experimental results that have been realized on a large amount of altered fingerprints show the benefit of the proposed metric in comparison with NFIQ. The presented metric also provides a value between 0 and 100% (that is more precise than a 5-scale evaluation). Perspectives of this work concern the use of other texture features (such as Gabor) for comparison issues, and to quantify in which measure this metric could be generalized to other biometric modalities.

6. REFERENCES

- [1] G. Zhang and Y. Wang, "Asymmetry-based quality assessment of face images," in *Proceedings of the 5th International Symposium on Advances in Visual Computing (ISVC)*, 2009, vol. 5876, pp. 499–508.
- [2] N. Poh, J.V. Kittler, and T. Bourlai, "Quality-based score normalization with device qualitative information for multimodal biometric fusion," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 40, pp. 539–554, 2010.
- [3] N. Poh, T. Bourlai, and J. Kittler, "A multimodal biometric test bed for quality-dependent, cost-sensitive and client-specific score-level fusion algorithms," *Pattern Recognition*, pp. 1094–1105, 2010.
- [4] ISO/IEC 29794-1, "Biometric quality framework standard, first ed. jtc1/sc37/working group 3," 2006.
- [5] P. Grother and E. Tabassi, "Performance of biometric quality measures," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, pp. 531–543, 2007.
- [6] 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 Transactions on Information Forensics and Security*, vol. 2, pp. 734–743, 2007.
- [7] L. Hong, Y. Wan, and A. Jain, "Fingerprint image enhancement: Algorithm and performance evaluation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 20, no. 8, pp. 777–789, 1998.
- [8] N.K. Ratha and R. Bolle, *Fingerprint image quality estimation*, IBM TJ Watson Research Center, 1999.
- [9] E. Lim, X. Jiang, and W. Yau, "Fingerprint quality and validity analysis," in *Image Processing. 2002. Proceedings. 2002 International Conference on*. IEEE, 2002, vol. 1, pp. I–469.
- [10] E. Tabassi and C.L. Wilson, "A novel approach to fingerprint image quality," in *International Conference on Image Processing (ICIP)*, 2005, pp. 37–40.
- [11] J. Merkle, M. Schwaiger, and M. Breitenstein, "Towards improving the nist fingerprint image quality nfiq algorithm.," in *International Conference BIOSIG*, 2010.
- [12] L. Shen, A. C. Kot, and W. M. Koo, "Quality measures of fingerprint images," in *Proceedings of the 3rd International Conference on Audio- and Video-Based Biometric Person Authentication (AVBPA)*, 2001, pp. 266–271.
- [13] Y. Chen, S. C. Dass, and A. K. Jain, "Fingerprint quality indices for predicting authentication performance," in *5th International Conference Audio- and Video-Based Biometric Person Authentication (AVBPA)*, 2005, vol. 3546, pp. 160–170.
- [14] S. Lee, C. Lee, and J. Kim, "Model-based quality estimation of fingerprint images," in *IAPR/IEEE International Conference on Biometrics (ICB'06)*, 2006, pp. 229–235.
- [15] M. A. Olsen, H. Xu, and C. Busch, "Gabor filters as candidate quality measure for nfiq 2.0," in *NIST International Biometric Performance Testing Conference*, 2012, p. 6.
- [16] M. Saad, A. C. Bovik, and C. Charrier, "DCT-statistics model-based blind image quality assessment," in *IEEE International Conference on Image Proc. (ICIP)*, Brussels, Belgium, 2011, pp. 3154–3157.
- [17] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision (IJCV)*, vol. 60, pp. 91–110, 2004.
- [18] J. H. Holland, "Genetic algorithm," *Scientific American Journal*, 1992.
- [19] 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)*, 2002, vol. 3, pp. 811–814.
- [20] "Nbis (nist biometric image software) release 4.0.1," <http://www.nist.gov/itl/iad/ig/nbis.cfm>, 2011.