



HAL
open science

Quality Assessment of Image-based Biometric Information

Mohamad El-Abed, Christophe Charrier, Christophe Rosenberger

► **To cite this version:**

Mohamad El-Abed, Christophe Charrier, Christophe Rosenberger. Quality Assessment of Image-based Biometric Information. EURASIP Journal on Image and Video Processing, 2015, pp.0. hal-01101548

HAL Id: hal-01101548

<https://hal.science/hal-01101548>

Submitted on 24 Jan 2015

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.

Quality Assessment of Image-based Biometric Information

Mohamad El-Abed^a, Christophe Charrier^{b,c,d}, Christophe Rosenberger^{b,c,d}

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

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

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

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

christophe.charrier@unicaen.fr
christophe.rosenberger@ensicaen.fr

Abstract

The quality of biometric raw data is one of the main factors affecting the overall performance of biometric systems. Poor biometric samples increase the enrollment failure, and decrease the system performance. Hence, controlling the quality of the acquired biometric raw data is essential in order to have useful biometric authentication systems. Towards this goal, we present a generic methodology for the quality assessment of image-based biometric modality combining two types of information: 1) image quality and 2) pattern-based quality using the SIFT descriptor. The associated metric has the advantages of being multimodal (face, fingerprint and hand veins), and independent from the used authentication system. Six benchmark databases and one biometric verification system are used to illustrate the benefits of

the proposed metric. A comparison study with the NFIQ metric proposed by the NIST shows the benefits of the presented metric.

Keywords: Biometrics, No-Reference Image Quality Assessment, Scale-Invariant Feature Transformation (SIFT), Support Vector Machine (SVM).

1. Introduction

Biometric systems are being increasingly used in our daily life to manage the access of physical (such as border control) and logical (such as e-commerce) resources. Biometrics uses the authentication factors based on “Something that qualifies the user” and “Something the user can do”. The main benefits of this authentication method is the strong relationship between the individual and its authenticator, as well as the easiness of its use. Also, it is usually more difficult to copy the biometric characteristics of an individual than most of other authentication methods such as passwords.

Despite the advantages of biometric systems, many drawbacks decrease their proliferation. The main one is the uncertainty of the verification result. By contrast to password checking, the verification of biometric raw data is subject to errors and represented by a similarity percentage (100% is never reached). This verification inaccuracy is due to many reasons such as the variations of human characteristics (*e.g.*, occlusions [1]), environmental factors (*e.g.*, illuminations [2]) and cross-device matching [3]. This kind of acquisition artifacts may deeply affect the performance of biometric systems and hence, decrease their use in real life applications. Moreover, the impact

of quality on the system overall performance is also presented by the results of the *FVC* series of competitions (*FVC* in 2000, 2002, 2004 and 2006) [4]. More specifically, the used databases in *FVC 2004* and *FVC 2006* are more difficult than *FVC 2002* and *FVC 2000* ones, due to the perturbations deliberately introduced. The results show that the EER in average of the best matching algorithm has increased from 0.96 in *FVC 2000* and *FVC 2002* to 2.115 in *FVC 2004* and *FVC 2006*. 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 or rejected during the verification. Such information could be also used for soft biometrics and multimodal approaches [5, 6].

We present in this paper a quality assessment metric of image-based biometric raw data using both information: 1) image quality and 2) pattern-based quality using the SIFT keypoints extracted from the image. The presented metric has the advantages of being multimodal (face, fingerprint and hand veins), and independent from the used authentication system.

The outline of the paper is given as follows: Section 2 presents related previous works on quality assessment of biometric raw data. We present in Section 3 the proposed quality assessment metric. Section 4 describes the experimental results obtained for the six trial biometric databases (four for face, two for fingerprint and hand veins, respectively). A comparison study with the NFIQ¹ metric on fingerprints is given in Section 5. A conclusion

¹NIST Fingerprint Image Quality

and some perspectives of this work are given in Section 6.

2. Related works

The quality assessment of biometric raw data is receiving more and more attention in biometrics community. We present in this section an overview of existing biometric image-based quality metrics.

The quality assessment of biometric raw data is divided into three points of view as illustrated in Figure 1 [7]:

- 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.

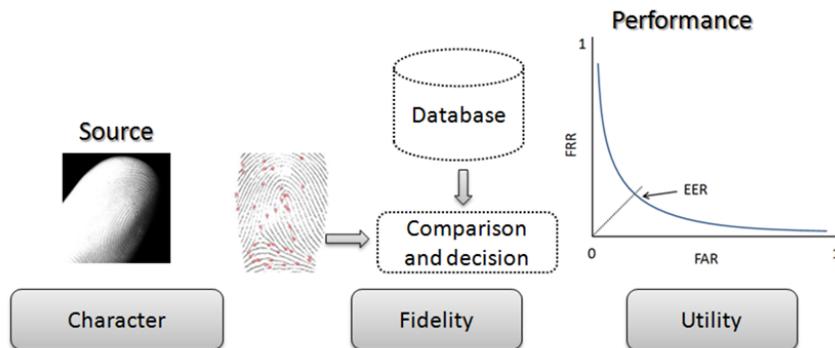


Figure 1: Quality assessment of biometric raw data: character, fidelity and utility.

In biometrics, there is an international consensus on the fact that the quality of a biometric sample should be related to its recognition performance [8]. Therefore, we present in this paper a utility-based quality assessment metric of biometric raw data. In the rest of this section, we present an overview of the existing image-based quality metrics.

Alonso-Fernandez et al. [9] present an extensive overview of existing fingerprint quality metrics which are mainly divided into three major categories:

1. Based on the use of local features of the image;
2. Based on the use of global features of the image;
3. Or addressing the problem of quality assessment as a classification problem.

The presented methods in [9] have shown their efficiency in predicting the quality of fingerprints images. However, these methods are modality-dependent, hence they cannot be used for other kinds of modalities (such as Face). An example of these metrics is the NIST Fingerprint Image Quality metric (NFIQ) [10] proposed by the NIST. NFIQ metric is dedicated to fingerprint quality evaluation.

Shen et al. [11] applied Gabor filters to identify blocks with clear ridge and valley patterns as good quality blocks. Lim et al. [12] combined local and global spatial features to detect low quality and invalid fingerprint images. Chen et al. [13] developed two new quality indices for fingerprint images. The first index measures the energy concentration in the frequency domain as a global feature. The second index measures the spatial coherence in local regions. These methods has shown their efficiency in predicting the quality of fingerprint images. However, they are didicated for fingerprint modality

and could not be used for other modalities such as veins images.

Krichen et al. [1] present a probabilistic iris quality measure based on a Gaussian Mixture Model (GMM). The authors compared the efficiency of their metric with existing ones according two types of alterations (occlusions and blurring) which may significantly decrease the performance of iris recognition systems. Chaskar et al. [14] assessed nine quality factors of iris images such as Ideal Iris Resolution(IIR), Actual Iris Resolution (AIR), *etc.* Other iris quality metrics are presented in [15, 16]. However, these methods are used to measure the quality of iris image, and cannot be used for other types of modalities.

He et al. [17] present a hierarchical model to compute the biometric sample quality at three levels: database, class and image quality levels. The method is based on the quantiles of genuine and impostor matching score distributions. However, their model could not be used directly on a single capture (*i.e.*, requires a pre-acquired database).

Zhang and Wang [2] present an asymmetry-based quality assessment method of face images. The method uses SIFT descriptor for quality assessment. The presented method has shown its robustness against illumination and pose variations. Another asymmetry-based method is presented in [18, 19]. However, this approach supposes the asymmetry hypothesis hence, could not be used for the others types of modalities.

For the finger veins modality, very few are the existing works that predict the quality of finger veins images. We can cite the work presented by Qin et al. [20]. The authors present a quality assessment method of finger veins images based on Radon transform to detect the local vein patterns. We believe

that extensive work should be done in this area since the veins modality is considered as a promising solution to be implemented.

Discussion

Quality assessment of biometric raw data is an essential step to achieve a better accuracy in real life applications. Despite this, few researches have been conducted to this point with respect to research activities on performance side. However, most of the existing quality metrics are modality- and matcher-dependent. The others, based on the genuine and impostor matching score distributions, could not be used directly on a single capture (*i.e.*, they require a large number of captures for the same person in order to constitute its genuine score distribution). Therefore, the main contribution of this paper is the definition of a quality metric which can be considered as independent from the used matching system, and also it can be used for several biometric modalities (face, fingerprint and hand veins images). It detects with a reasonable accuracy three types of alterations that may deeply affect the global performance of the most widely used matching systems. The presented metric is not based on asymmetry hypothesis. Thus, it may be used for several types of modalities (such as fingerprint, face, hand and finger veins), and can be used directly on a single capture after training the model.

3. Developed Metric

The presented metric is designed to quantify the quality of image-based biometric data using two types of information as illustrated in Figure 2. The retained principle is as follows: using one image quality criterion (Section 3.1)

and four pattern-based quality criteria (Section 3.2), a SVM-based classification process (Section 3.3) is performed to predict the quality of the target biometric data.

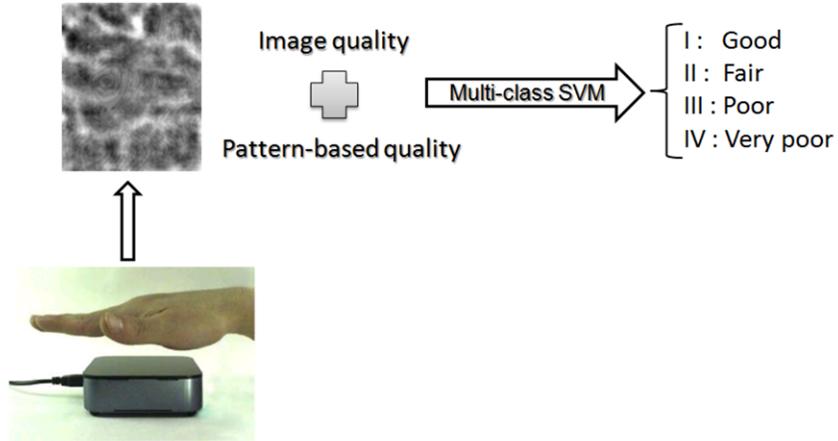


Figure 2: General scheme of the proposed quality metric.

3.1. No-reference image quality

The image quality assessment is an active research topic which is widely used to validate treatment processes applied to digital images. In the context of image compression, for example, such kind of assessment is used to quantify the quality of the reconstructed image. Existing image quality assessment metrics are divided into three categories: 1) Full-Reference (FR) quality metrics, where the target image is compared with a reference signal that is assumed to have perfect quality; 2) Reduced-Reference (RR) quality metrics, where a description of the target image is compared with a description of the reference signal; and 3) No-Reference (NR) quality metrics, where the target image is evaluated without any reference to the original one. Despite the

acceptable performance of current FR quality algorithms, the need for a reference signal limits their application, and calls for reliable no-reference algorithms.

In our study, we have used a No-Reference Image Quality Assessment (NR-IQA) index, since the reference image does not exist. The used NR-IQA method in this paper is the BLIINDS² index introduced by Saad et al. [21]. This index is based on a DCT framework. This makes it computationally convenient, uses a commonly used transform, and allows a coherent framework. The BLIINDS index is defined from four features, using 17×17 image patches centered at every pixel in the image, that are then pooled together:

1. DCT-Based Contrast feature (v_1)

Contrast is a basic perceptual attribute of an image. One may distinguish between global contrast measures and ones that are computed locally (and possibly pooled into one measure post local extraction).

The contrast of the k^{th} local DCT patch is computed as follows:

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

where N is the patch size, x_{DC} represents the DC coefficient and the set $\{x_{AC}^i | i = 1 : N\}$ represents the AC coefficients. Then, the local contrast scores from all patches of the image are then pooled together by averaging the computed values to obtain a global image contrast value v_1 :

²BLind Image Integrity Notator using DCT Statistics

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

where M is the number of local patches.

2. DCT-Based Structure Features (v_2)

Structure features are derived locally from the local DCT frequency coefficients computed on a patch k . They are based on statistical traits of the DCT histogram for which the DC coefficient is ignored. To measure these statistical traits of the DCT histograms of the patch k , its kurtosis is computed to quantify the degree of its peakedness and tail weight:

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

where μ is the mean of x_{AC} , and σ is its standard deviation. Then, the resulting values for all patches are pooled together by averaging the lowest tenth percentile of the obtained values to compute the global image kurtosis value v_2 .

3. DCT-Based Anisotropy orientation (v_3 and v_4)

It has been hypothesized that degradation processes damage a scene's directional information. Consequently, anisotropy, which is a directionally dependent quality of images, was shown by Gabarda and Cristbal [22] to decrease as more degradation is added to the image. The anisotropy measure is computed using the Renyi Entropy on DCT image patches along four different orientations $\theta = 0, 45, 90, 135$ in degrees. Each patch consists of the DCT coefficients of oriented pixel intensities. We discard the DC coefficient, since the focus is on direc-

tional information. Let the DCT coefficients of k^{th} patch of orientation θ be denoted by $P_\theta[k, j]$, where j is the frequency index of the DCT coefficient. Each DCT patch is then subjected to a normalization of the form:

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

where N is the size of the oriented k^{th} patch. Finally, the associated Renyi entropy R_θ^k is computed as

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

where $\beta > 1$. Finally, the two measures of anisotropy v_3 and v_4 are defined as

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

Due to the fact that the perception of image details depends on the image resolution, the distance from the image plane to the observer, and the acuity of the observers visual system, a multiscale approach is applied to compute the final global score as:

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

constraints by $\sum_{j=1}^4 \sum_{i=1}^L \alpha_j^i = 1$ and where L represents the number of decomposition level used. The α_j^i values are obtained using the correlation of each criterion (v_i) with the subjective notes given by human observers [21].

Examples of predicted quality score using BLIINDS index are given in Figure 3. The stronger the image is degraded, the lower the quality index is.

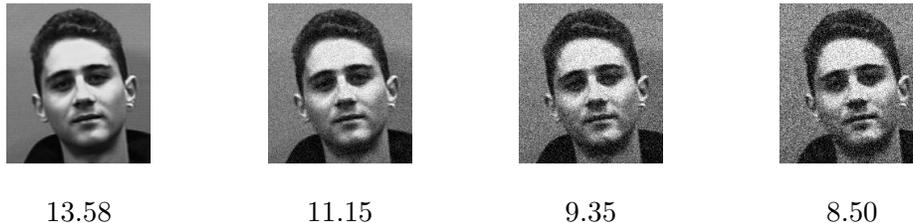


Figure 3: Examples of BLIINDS index on samples of FACES94 database. From left to right, reference image then its noisy images.

3.2. Pattern-based quality

The used pattern-based quality criteria are based on statistical measures of keypoints features. We have used this approach since keypoints features describe, in a stable manner, the regions of the image where the information is important. This approach is widely used in object [23] and biometric recognition [24] issues. For the descriptor vector computation, several methods exist in the literature such as the *Scale Invariant Feature Transform* (SIFT) [25], *Shape Contexts* [26], *Speed Up Robust Features* (SURF) [27]. In our study, we have used the SIFT algorithm since a comparison study presented by Mikolajczyk and Schmid [28] show that SIFT outperformed the other methods. SIFT algorithm has been also successfully used in biometric recognition for different modalities such as veins [24], face [29], fingerprint [30], iris [31] as well as in 3D facial recognition [32].

SIFT algorithm consists of four major stages: 1) scale-space extrema detection, 2) keypoint localization, 3) orientation assignment and 4) keypoint descriptor. In the first stage, potential interest points are identified, using a

difference-of-Gaussian function, that are invariant to scale and orientation. In the second stage, candidate keypoints are localized to sub-pixel accuracy and eliminated if found to be unstable. The third stage identifies the dominant orientations for each keypoint based on its local image patch. The keypoint descriptor in the final stage is created by sampling the magnitudes and orientations of the image gradients in a neighborhood of each keypoint and building smoothed orientation histograms that contain the important aspect of the neighborhood. Each local descriptor is composed of a 4x4 array (histogram). For each coordinate of this array, an 8 orientation vector is associated. A 128-elements ($8 \times (4 \times 4)$) vector is then built for each keypoint. In other words, each image im is described by a set of invariant features $X(im) = \{k_i = (s_i, sc_i, x_i, y_i) | i = 1 : N(im)\}$ 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 im , sc_i its scale and $N(im)$ the number of detected keypoints for image im . The features extracted are invariant to image scaling and rotation, and partially invariant to change in illumination and 3D camera viewpoint. Examples of detected SIFT keypoints are given in Figure 4. From these features, four criteria are retained (see Section 4.3.1) to contribute to the quality assessment of the biometric raw data.

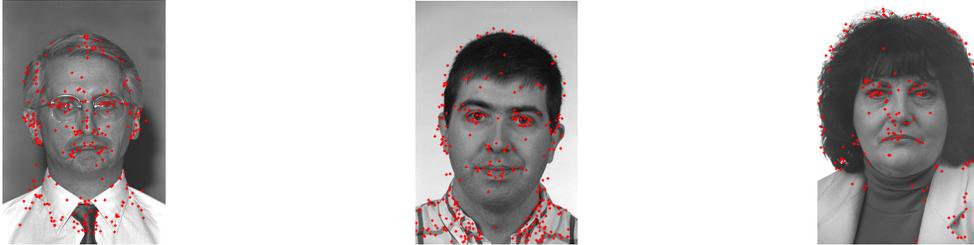


Figure 4: Examples of detected SIFT keypoints.

3.3. SVM-based classification

In order to predict biometric sample quality using both information (image quality and pattern-based quality), we use the Support Vector Machine (SVM). From all existing classification schemes, a SVM-based technique has been selected due to high classification rates obtained in previous works [33] and to their high generalization abilities. SVMs have been proposed by Vapnik [34] and are based on the structural risk minimization principle from statistical learning theory. SVMs express predictions in terms of a linear combination of kernel functions centered on a subset of the training data, known as support vectors (SV).

Suppose we have a training set $\{\mathbf{x}_i, \mathbf{y}_i\}$ where \mathbf{x}_i is the training pattern and \mathbf{y}_i the label. For problems with two classes, with the classes $y_i \in \{-1, 1\}$, a support vector machine [34] implements the following algorithm. First, the training points $\{\mathbf{x}_i\}$ are projected into a space \mathcal{H} (of possibly infinite dimension) by means of a function $\Phi(\cdot)$. The second step is to find an optimal decision hyperplane in this space. The criterion for optimality will be defined shortly. Note that for the same training set, different transformations $\Phi(\cdot)$ may lead to different decision functions. A transformation is achieved in an implicit manner using a kernel $K(\cdot, \cdot)$ and consequently the decision function can be defined as :

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

with $\alpha_i^* \in R$. The values w and b are the parameters defining the linear decision hyperplane. In SVMs, the optimality criterion to maximize is the margin, that is to say, the distance between the hyperplane and the nearest

point $\Phi(\mathbf{x}_i)$ of the training set. The α_i^* which optimize this criterion are obtained by solving the following problem :

$$\left\{ \begin{array}{l} \max_{\alpha_i} \sum_{i=1}^{\ell} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{\ell} \alpha_i \alpha_j y_i K(\mathbf{x}_i, \mathbf{x}_j y_j) \\ \text{with constraints,} \\ 0 \leq \alpha_i \leq C, \\ \sum_{i=1}^{\ell} \alpha_i y_i = 0. \end{array} \right. \quad (9)$$

where C is a penalization coefficient for data points located in or beyond the margin and provides a compromise between their numbers and the width of the margin. In this paper, we use the RBF kernel:

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

In order to train models with RBF kernels, we use a python script provided by the libsvm library [35]. This script automatically scales training and testing sets. It searches (only for the training set) the best couple (C, γ) of the kernel. The search of the best couple (C, γ) is done using a five-fold cross-validation computation.

Originally, SVMs have essentially been developed for the two classes problems. However, several approaches can be used for extending SVMs to multi-class problems. The method we use in this communication, is called *one against one*. Instead of learning N decision functions, each class is discriminated here from another one. Thus, $\frac{N(N-1)}{2}$ decision functions are learned and each of them makes a vote for the affectation of a new point \mathbf{x} . The class of this point \mathbf{x} becomes then to the majority class after the voting.

4. Experimental Results

The goal of the proposed quality metric is to detect, with a reasonable accuracy, three synthetic alterations which may deeply affect the most widely used matching systems. The proposed metric may be considered as independent from the used matching system. An example of its practical use is illustrated in Figure 5. The method predicts the alteration of the input image. Then, depending from the robustness of the used matching system against the predicted alteration, the matching system qualifies the image (good, fair, bad or very bad quality).

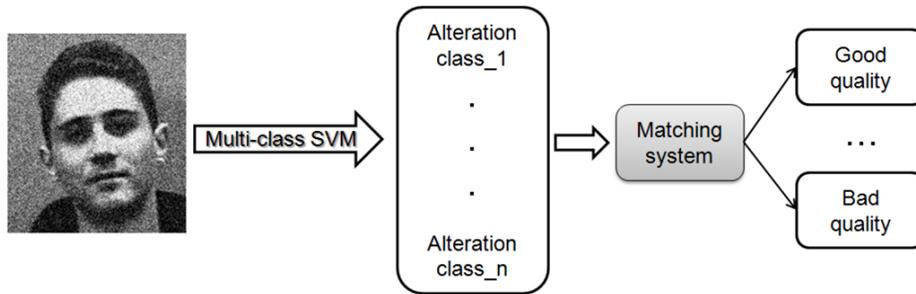


Figure 5: An example of use of the presented method.

In Section 4.1, we present the experimental protocol followed by the validation process of the proposed metric. The results are then given in Section 4.3.

4.1. Protocol

Six benchmark databases and one biometric matching algorithm are used in order to validate the proposed metric.

4.1.1. Alteration process

In this study, we introduce three types of synthetic alterations as well as three levels for each type using the MATLAB tool:

- Blurring alteration: blurring images are obtained using a two-dimensional Gaussian filter. To do so, we use the *fspecial* ('gaussian', *hsize*, σ) method which returns a rotationally symmetric Gaussian lowpass filter of size *hsize* with standard deviation σ .
- Gaussian noise alteration: noisy images are obtained using the *imnoise* (*I*, 'gaussian', *m*, *v*) method. It adds Gaussian white noise of mean *m* and variance *v* to the image *I*.
- Resize alteration: such kind of altered images are obtained using the *imresize* (*I*, *scale*, 'nearest') method. It resizes the image *I* using a nearest-neighbor interpolation.

Table 1 presents the parameters required of the used alteration MATLAB methods.

Alteration type	method	level 1	level 2	level 3
Blurring	<i>fspecial</i> ('gaussian', [7 7], σ)	$\sigma = 1$	$\sigma = 2$	$\sigma = 6$
Gaussian noise	<i>imnoise</i> (<i>I</i> , 'gaussian', 0.01, <i>v</i>)	<i>v</i> =0.003	<i>v</i> =0.01	<i>v</i> =0.017
Resize	<i>imresize</i> (<i>I</i> , <i>scale</i> , 'nearest')	<i>scale</i> = 0.8	<i>scale</i> = 0.6	<i>scale</i> = 0.4

Table 1: Parameters of the MATLAB alteration methods.

Using these alterations, the input vector to SVM is the five retained quality criteria (one for image quality and four pattern-based quality) and the output can belong to ten different classes defined as follows (see Table 2):

- class 1 illustrates a reference image.
- classes 2 to 10 illustrate 3 types of alterations and 3 levels for each type (see Section 4.1.2 for details about the introduced alterations).

Class	Description	Alteration levels
1	reference or original image	×
2, 3 and 4	blurring alteration	1, 2 and 3, respectively
5, 6 and 7	gaussian noise alteration	1, 2 and 3, respectively
8, 9 and 10	resize alteration	1, 2 and 3, respectively

Table 2: SVM classes definition.

4.1.2. Benchmark databases

In this study, we use six benchmark databases. For each database, we introduce three types of alterations (blurring, gaussian noise and resize alterations) and three levels for each type of alteration. The presented alterations are commonly realistic during the acquisition of biometric raw data, which may deeply affect the overall performance of biometric systems. Finally, we get 60 databases: 6 reference and 54 altered databases (*i.e.*, 9 for each reference database):

1. Reference databases

- **FACES94 Database [36]:** This database is composed of 152 individuals and 20 samples per individual. These images have been captured in regulated illumination and the variation of expression is moderated.



Figure 6: Samples from FACES94.

- ENSIB Database [37]: It is composed of 100 individuals and 40 samples per individual. Each sample corresponds to one pose from the left one to the right.



Figure 7: Samples from ENSIB.

- FERET Database [38, 39]: It is composed of 725 individuals with from 5 to 91 samples per individual (the average value is 11). Each sample corresponds to a pose angle, illumination and expression.



Figure 8: Samples from FERET.

- AR database [40]: It is composed of 120 individuals and 26 samples per individual. These included images captured under differ-

ent conditions of illumination, expression and occlusion.



Figure 9: Samples from AR.

- FVC2002 DB₂ database [41]: It is composed of 100 individuals and 8 samples per individual. The database was used during the Fingerprint Verification Competition (FVC2002).



Figure 10: Samples from FVC2002 DB₂.

- Hand veins database [24]: It is composed of 24 individuals and 30 samples per individual. The database has been collected by TELECOM & Management SudParis.

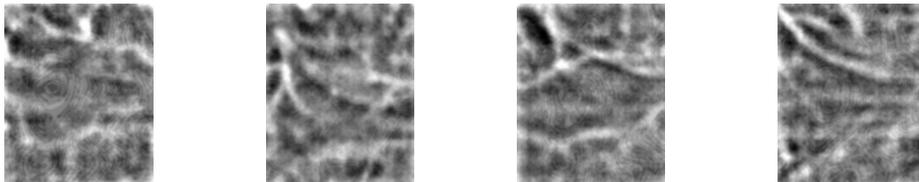


Figure 11: Samples from the hand veins database.

2. Altered databases

Using the introduced alterations presented in Section 4.1.1, we generated 54 databases from the 6 reference databases: FACES94, ENSIB, FERET, AR, FVC2002 DB₂ and the hand veins databases. Figure 12 shows these alterations on a sample from FACES94 database.



(a) Blurring alteration



(b) Gaussian noise alteration



(c) Resize alteration

Figure 12: Alterations for a reference image from FACES94. From left to right, reference image then alteration level 1, 2 and 3.

4.1.3. Biometric matching algorithm

The used biometric matching algorithm is a SIFT-based algorithm [25]. The matching similarity principle used is described in previous works [24]. Each image im is described by a set of invariant features $X(im)$ as described in Section 3.2. The verification between two images im_1 and im_2 corresponds

to the similarity between two sets of features $X(im_1)$ and $X(im_2)$. We thus use the following matching method which is a modified version of a decision criterion first proposed by Lowe [25]. Given two keypoints $x \in X(im_1)$ and $y \in X(im_2)$, we say that x is associated to y if:

$$d(x, y) = \min_{\{z \in X(im_2)\}} d(x, z) \text{ and } d(x, y) \leq C d(x, y') \quad (11)$$

where C is an arbitrary threshold, $d(\cdot, \cdot)$ denotes the Euclidean distance between the SIFT descriptors and y' denotes any point of $X(im_2)$ whose distance to x is minimal but greater than $d(x, y)$:

$$d(x, y') = \min_{\{z \in X(im_2), d(x, z) > d(x, y)\}} d(x, z) \quad (12)$$

In other words, x is associated to y if y is the closest point from x in $X(im_2)$ according to the Euclidean distance between SIFT descriptors and if the second smallest value of this distance $d(x, y')$ is significantly greater than $d(x, y)$. The significance of the necessary gap between $d(x, y)$ and $d(x, y')$ is encoded by the constant C . Then, we consider this keypoint x is matched to y iff x is associated to y and y is associated to x . Figure 13 presents an example of matching resulting from a genuine and an impostor comparison.

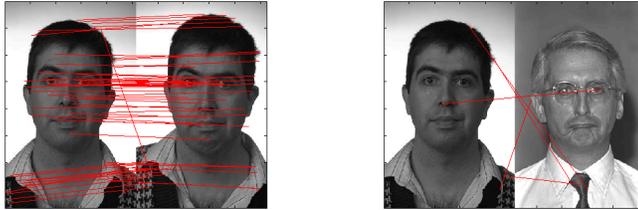


Figure 13: Example of matching results resulting from a genuine (on the left) and an impostor comparisons (on the right).

4.2. Validation process

According to Grother and Tabassi [8], 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. Therefore, we use the Equal Error Rate (EER) which illustrates the overall performance of a biometric system [42]. EER is defined as the rate when both False Acceptance Rate (FAR) and False Reject Rate (FRR) are equal: the lower EER, the more accurate the system is considered to be. In order to validate the proposed quality metric, we proceed as follows:

- Quality criteria behavior with alterations: the first step of the validation process consists of showing the robustness of the used five quality criteria in detecting the introduced alterations: blurring, Gaussian noise and resize alterations.
- Learning the multi-class SVM models: for face databases, we learn 4 multi-class SVM models using the images from the four benchmark databases (1 multi-class SVM per benchmark database illustrated by SVM_{each}), and 1 multi-class SVM model containing examples from the four benchmark databases (illustrated by SVM_{all}). For the fingerprint and hand veins databases, we learn 2 multi-class SVM models, respectively. In order to train and to test the multi-class SVM models, we split each benchmark database images into two sets $S_{training}$ and S_{test} in a balanced way (*i.e.*, both sets contain the same ratio of reference and altered images). The choice of the kernel and the selection of the parameters required are presented in Section 3.3.

- Quality sets definition: the proposed metric predicts a quality class of the target image. In order to show the utility of this metric, we need to define the quality sets for the used authentication system. Depending from the used authentication system, some alterations may have more impact on its global performance more than others. Thereafter, we use the EER to illustrate the global performance of the biometric system.
- EER value of each quality set: in order to quantify the effectiveness of our quality metric in predicting system performance, we have put each image to a quality set, using its predicted label by our metric. Then, we have calculated the EER value for each quality set. The effectiveness of the method is quantified by how well our quality metric could predict system performance among the defined quality sets. More generally speaking, the more the images are degraded, the more the performance of the overall system will be decreased (illustrated by an increase of its EER value).

4.3. Results

4.3.1. Quality criteria behavior with alterations

In this section, we show the robustness of the used criteria in detecting alterations presented in the previous section. To do so, we use the Pearson's correlation coefficient between two variables as defined in Equation 13. It is defined as the covariance of the two variables (X and Y) divided by the product of their standard deviation (σ_X and σ_Y):

$$Pearson(X, Y) = \frac{Cov(X, Y)}{\sigma_X \sigma_Y} \quad (13)$$

In order to compute the correlation of the used criteria with the three types of alterations, we define for each type of alteration and for each criterion p the variables as follows:

- $X_p = \{X_{pk} | k = 1 : 4\}$ where X_{p1} is the set of values of criterion p for the reference databases images, (X_{p2}, X_{p3}, X_{p4}) are the sets of values of criterion p for the altered databases level 1, 2 and 3, respectively.
- Alteration levels are represented by the variable Y (1: for the reference databases, 2, 3 and 4: for the altered databases level 1, 2 and 3). More precisely, $Y = \{y_k | y_k = 1 \text{ for } k = 1 : N, y_k = 2 \text{ for } k = N + 1 : 2N, y_k = 3 \text{ for } k = 2N + 1 : 3N \text{ and } y_k = 4 \text{ for } k = 3N + 1 : 4N\}$ where N is the size of the 4 reference databases.

Using the extracted SIFT keypoints and the Pearson's correlation coefficient, four pattern-based quality criteria are retained to contribute to quality assessment:

1. Keypoints: the number of keypoints detected from image im .
2. DC coefficient: DC coefficient of the matrix M_s , with $N(im)$ rows and 128 columns, related to SIFT invariant descriptor for $s_i, i = 1 : N(im)$ where $N(im)$ is the number of detected keypoints for image im .
3. Mean (μ) and
4. Standard deviation (σ) of scales: mean and standard deviation of scales related to the keypoints detected from image im .

Therefore, the vector V used to predict biometric sample quality is a five-dimensional vector containing one image quality criterion and four pattern-based criteria as depicted in Table 3.

Vector V	
Image quality criterion	BLIINDS
Pattern-based criteria	keypoints DC coefficient mean (μ) of scales standard deviation (σ) of scales

Table 3: The five-dimensional vector for predicting biometric sample quality.

Figure 14 shows that the four pattern-based criteria (keypoints, DC coefficient, mean and standard deviation of scales) are pertinent in detecting the three types of alterations: blurring, Gaussian noise and resize alterations. The image quality criterion BLIINDS has shown to be efficient (with a correlation coefficient more than 0.6) in detecting blurring and Gaussian noise alterations. For the resize alteration, BLIINDS has not shown to be efficient which is not a surprising result since resize alteration does not affect image quality (BLIINDS is a multiresolution NR-IQA algorithm). Moreover, the distortion cards given in Figure 15 show also the efficiency of BLIINDS index (*i.e.*, each feature v_i , $i = 1 : 4$) in detecting altered images.

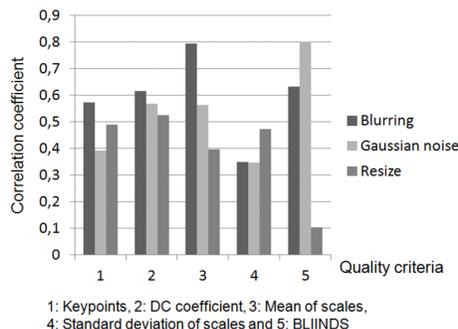


Figure 14: Pearson correlation coefficients (in absolute value) between the proposed criteria and the three alterations among the 4 face databases.

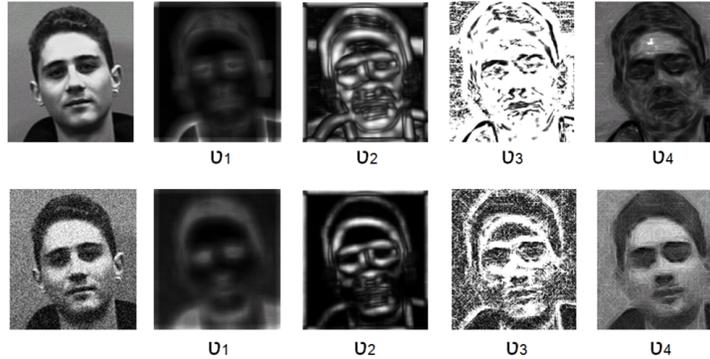


Figure 15: Distortion cards related to a slightly altered image (first row) and a highly altered one (second row), respectively. $v_i, i = 1 : 4$ correspond to the four features used in BLIINDS index computation.

4.3.2. Learning the multi-class SVM models

We learned 7 multi-class SVM models: 5 for face databases, and 2 for hand veins and fingerprint databases. Table 4 presents the accuracy of the learned multi-class SVM models on both training and test sets. We have put the symbol “ \times ” at the last two lines, since we have only 1 multi-class SVM generated per database. Table 4 shows the efficiency of the proposed metric in predicting the three synthetic alterations (blurring, Gaussian noise and resize) of data, with a valuable 10-class classification accuracy (going from 82.29% to 97.73 on the *training set*, and from 81.16% to 91.1% on the *test set*). Results for the different databases are similar but not exactly the same. The reason is related to the complexity of the databases incorporating more or less artifacts.

In order to test more the generalization of these results, we have tested the following:

- we used one of the four face databases for training, and the rest three for tests. We obtained a valuable 4-class accuracy of 87.84%. The four classes are 1: original, 2, 3 and 4: blurring, gaussian noise and resize level 3, respectively. For a 10-class classification, this accuracy decreased to 50.78% due to the different resolution of images in each database. However, the resolution of images to be used for training should be as much as similar for the test images in order to maintain a high accuracy. This is illustrated by the results of the multi-class SVM (SVM_{all}) presented in Table 4 (third and fourth columns). We have obtained good classification results since the images used for training this model contains a set of images from each database.
- we used a subset of the CASIA-FingerprintV5 [43] (first 100 persons, 5 images per person, left thumb) of 500 images. The volunteers of CASIA-FingerprintV5 were asked to rotate their fingers with various levels of pressure to generate significant intra-class variations. Using the FVC2002 DB2 database for learning the multi-class SVM, we classified each of the CASIA-FingerprintV5 500 images into the four categories presented in Table 5 (1 for good, 2 for fair, 3 for poor and 4 for very poor). We then computed the intra-class matching scores, using the matching algorithm presented in Section 4.1.3, of the CASIA-FingerprintV5 database (by taking the first image as reference and the rest for the test). Using the pearson correlation coefficient between the obtained intra-class matching scores and the four categories, we obtained a significant correlation of 0.67. This shows that images classified of good quality by the proposed method provided higher matching

scores compared to the images predicted of bad quality; which clarify the benefits of the presented method.

	SVM_{each}		SVM_{all}	
	$S_{training}$	S_{test}	$S_{training}$	S_{test}
<i>FACES94</i>	91.01	86.69	85.68	85.28
<i>ENSIB</i>	97.73	89.82	94.92	91.1
<i>FERET</i>	82.33	81.2	82.29	81.16
<i>AR</i>	90.08	89.08	90.7	88.92
<i>FVC2002 DB₂</i>	×	×	91.7	83.68
<i>Hand veins</i>	×	×	95.25	90.2

Table 4: Accuracy (in %) of the learned multi-class SVM models on both training and test sets.

4.3.3. Quality sets definition

In order to quantify the robustness of the proposed metric in predicting system performance, we need first to define the quality sets of the used biometric authentication systems. Therefore, we have tested the robustness of the used system against the three alterations presented in Section 4.1.2. The EER values are computed using the first image as a reference (single enrollment process), and the rest for the test. Figure 16 shows that all the introduced alterations have an impact on overall performance of the used authentication matching algorithm presented in Section 4.1.3. Therefore, we define in Table 5 the quality sets definition for the used matching algorithm.

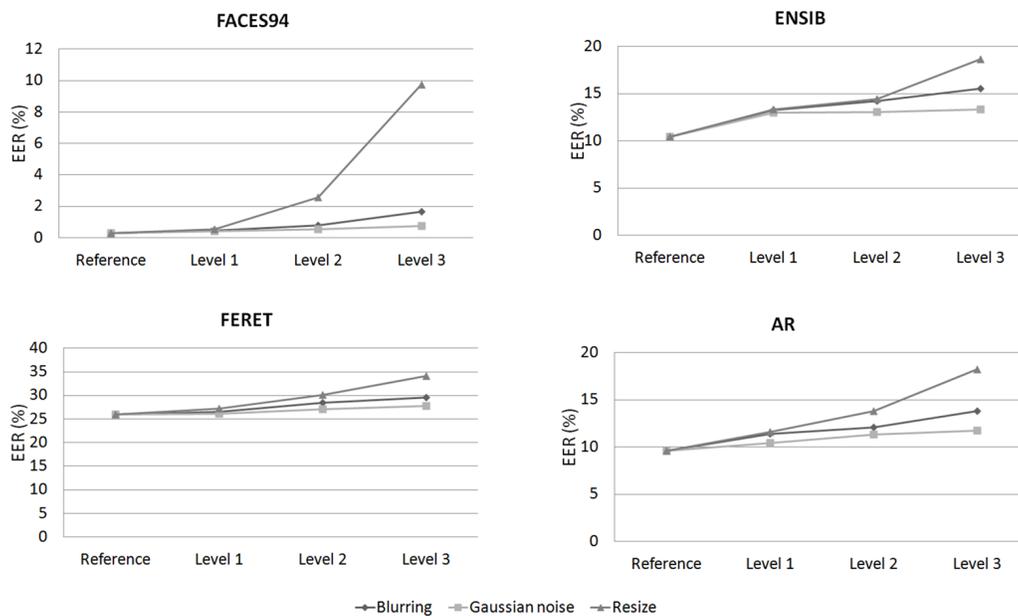


Figure 16: Impact of alterations on overall performance of the used authentication system among the 4 face databases.

Quality set	Predicted quality class by SVM	Description
I	1	good
II	2, 5 and 8	fair
III	3, 6 and 9	poor
IV	4, 7 and 10	very poor

Table 5: Category of quality.

4.3.4. EER value of each quality set

In order to validate the proposed quality metric in predicting the used matching algorithm performance, according to Grother and Tabassi [8], we calculate the EER value of each quality set predicted by the learned multi-class SVM models. Figure 17 shows the EER values of each quality set among

the used biometric databases. From this figure, we can deduce several points:

- The proposed metric has shown its efficiency in predicting the used matching system among the 6 biometric databases. More generally speaking, the more the images are altered, the more the EER values are increasing. This is demonstrated by the increasing curves presented in Figure 17. For the hand veins database, we have obtained a slight increase of EER values (0%: good category to 0.05%: very poor category). This result can be explained by the small size of the hand veins database (24 persons), and the robustness of the used matching system against the introduced alterations.
- The 4 curves related to the four face databases, presented in Figure 17, are computed using the 4 multi-class SVM (1 multi-class SVM per database). We have obtained similar curves using the other multi-class SVM model containing examples from the four benchmark databases. This shows the scalability of the presented metric to be used on different types of images (such as the image resolution).

5. Comparison study with NFIQ

In order to show the efficiency of the proposed metric, we present in this section a comparison study with the NIST Fingerprint Image Quality metric (NFIQ) [10]. We have used NFIQ metric proposed by the NIST, since it is the most cited at the literature for the fingerprint modality. NFIQ provides five levels of quality (NFIQ=1 indicates high quality samples, whereas

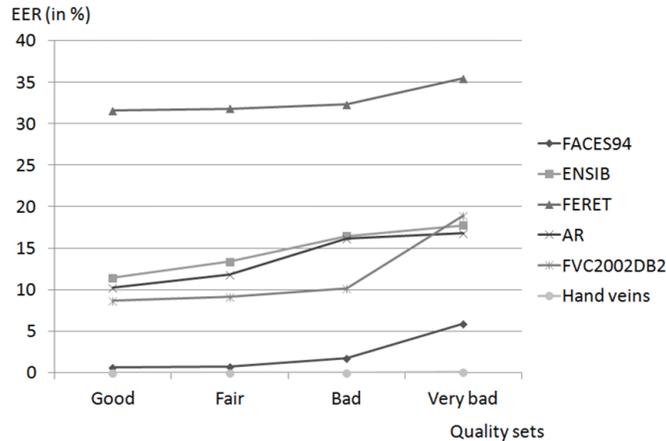


Figure 17: EER values of each quality set among the used biometric databases: face, fingerprint and hand veins databases.

NFIQ=5 indicates poor quality samples). For the comparison with the proposed method (four levels of quality), we consider that the 4th and 5th levels belong to the *very bad* quality set.

In order to compare the proposed metric with NFIQ, we use the approach suggested by Grother and Tabassi [8] when comparing quality metrics. To do so, we use the Kolmogorov-Smirnov (KS) test [44] which is a nonparametric test to measures the overlap of two distributions: in our case, distributions of scores of genuine and impostors, respectively. More generally speaking, KS test returns a value defined between 0 and 1: for better quality samples, a larger KS test statistic (*i.e.*, higher separation between genuine and impostor distributions) is expected.

Figure 18 illustrates the KS test statistic values of both quality metrics (NFIQ and the presented one). For the three quality sets (bad, fair and good), Figure 18 shows that the proposed metric (KS statistic going from

0.797 to 0.869) outperformed the NFIQ metric (KS statistic going from 0.632 to 0.82).

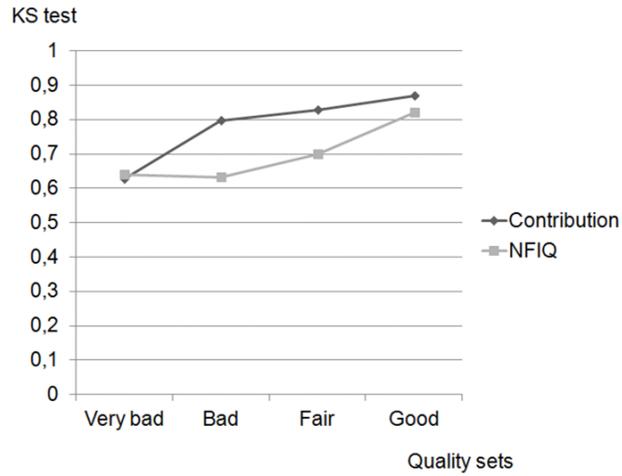


Figure 18: Comparison study between the proposed metric and NFIQ: Kolmogorov-Smirnov (KS) test statistic.

6. Conclusion and perspectives

The quality assessment of biometric raw data is a key factor to take into account during the enrollment step when using biometric systems. Such kind of information may be used to enhance the overall performance of biometric systems, as well as in fusion approaches. However, few are the works exist in comparison to the performance ones. Toward contributing in this research area, we have presented an image-based quality assessment metric of biometric raw data using two types of information (image and pattern-based quality). The proposed metric is independent from the used matching system, and could be used to several kind of modalities. Using six public biometric databases (face, fingerprint and hand veins), we have shown its efficiency in detecting three kinds of synthetic alterations (blurring, Gaussian noise and resolution).

For the perspectives of this work, we aim to add an additional quality criterion in order to detect luminance alteration, which is also considered as an important alteration affecting biometric systems (mainly, facial-based recognition systems). We aim also to compare the proposed metric with NFIQ using other kind of biometric matching algorithms (such as BOZORTH3 [45] proposed by the NIST). In addition, we are planning to test the efficiency of the presented method on altered images combining the presented alterations, which also represent another kind of real life alterations. This can be done using the presented criteria and a SVM or a genetic algorithm in order to produce an index between 0% and 100% (*i.e.*, more the index is near 100% better is the quality). Modality specific alterations could also be used to have a precise analysis of the efficiency of the proposed methodology.

Terms and definitions

Enrollment: The process of collecting biometric samples from a person and the subsequent preparation and storage of biometric reference templates representing that person's identity.

False Acceptance Rate (FAR): Rate at which an impostor is accepted by an authentication system.

False Rejection Rate (FRR): Rate at which the authorized user is rejected from the system.

Equal Error Rate (EER): This error rate corresponds to the point at which the FAR and FRR cross (compromise between FAR and FRR).

Acknowledgment

Portions of the research in this paper use the CASIA-FingerprintV5 collected by the Chinese Academy of Sciences' Institute of Automation (CASIA).

References

- [1] E. Krichen, S. Garcia Salicetti, and B. Dorizzi. A new probabilistic iris quality measure for comprehensive noise detection. In *IEEE Third International Conference on Biometrics : Theory, Applications and Systems (BTAS)*, pages 1–6, 2007.
- [2] 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)*, volume 5876, pages 499–508, 2009.

- [3] 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*, 40:539–554, 2010.
- [4] R. Cappelli, D. Maio, D. Maltoni, J. L. Wayman, and A. K. Jain. Performance evaluation of fingerprint verification systems. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 28:3–18, 2006.
- [5] 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*, pages 1094–1105, 2010.
- [6] N. Poh, 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, and C. Vielhauer. Benchmarking quality-dependent and cost-sensitive score-level multimodal biometric fusion algorithms. *Transactions on Information Forensics and Security*, 4(4):849–866, 2009.
- [7] ISO/IEC 29794-1. Biometric quality framework standard, first ed. jtc1/sc37/working group 3, 2006.
- [8] P. Grother and E. Tabassi. Performance of biometric quality measures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29:531–543, 2007.
- [9] 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*, 2:734–743, 2007.

- [10] E. Tabassi and C.L. Wilson. A novel approach to fingerprint image quality. In *International Conference on Image Processing (ICIP)*, pages 37–40, 2005.
- [11] Linlin Shen, Alex Kot, and Waimun Koo. Quality measures of fingerprint images. In *Audio- and Video-based Biometric Person Authentication*, pages 266–271, 2001.
- [12] Eyung Lim, Xudong Jiang, and Wei-Yun Yau. Fingerprint quality and validity analysis. In *IEEE International Conference on Image Processing (ICIP)*, pages 469–472, 2002.
- [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)*, volume 3546, pages 160–170, 2005.
- [14] U. M. Chaskar, M. S. Sutaone, N. S. Shah, and Jaison.T. Iris image quality assessment for biometric application. *IJCSI International Journal of Computer Science Issues*, 9:474–478, 2012.
- [15] Y. Chen, S.C. Dass, and A.K. Jain. Localized iris image quality using 2-d wavelets. In *International Conference on Biometrics (ICB)*, 2006.
- [16] Nathan D. Kalka, Jinyu Zuo, Natalia A. Schmid, and Bojan Cukic. Estimating and fusing quality factors for iris biometric images. *Trans. Sys. Man Cyber. Part A*, 40:509–524, 2010.
- [17] Q. He, Z.A. Sun, T.N. Tan, and Y. Zou. A hierarchical model for the evaluation of biometric sample quality. In *International Conference on Pattern Recognition (ICPR)*, pages 1–4, 2008.

- [18] 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.
- [19] J. Sang, Z. Lei, and S. Z. Li. 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)*, pages 229–238, 2009.
- [20] Huafeng Qin, Sheng Li, Alex C. Kot, and Lan Qin. Quality assessment of finger-vein image. In *Signal & Information Processing Association Annual Summit and Conference (APSIPA ASC)*, pages 1–4, 2012.
- [21] 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.
- [22] S. Gabarda and G. Cristbal. Blind image quality assessment through anisotropy. *Journal of Optical Society of America*, pages B42–B51, 2007.
- [23] P. Parisot. *Suivi d'objets dans des sequences d'images de scenes dformables*. PhD thesis, Institut de recherche en Informatique de Toulouse, 2009.
- [24] P.-O. Ladoux, C. Rosenberger, and B. Dorizzi. Palm vein verification system based on sift matching. In *the 3rd IAPR/IEEE International Conference on Biometrics (ICB'09)*, pages 1290–1298, 2009.
- [25] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision (IJCV)*, 60:91 – 110, 2004.
- [26] S. Belongie, J. Malik, and J. Puzicha. Matching shapes. In *International Conference on Computer Vision*, pages 454 – 461, 2001.

- [27] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool. Speeded-Up Robust Features (SURF). *Computer Vision and Image Understanding*, 110:346 – 359, 2008.
- [28] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 27:1615–1630, 2005.
- [29] Dakshina Ranjan Kisku, Ajita Rattani, Enrico Grosso, and Massimo Tistarelli. Face identification by sift-based complete graph topology. In *5th IEEE International Workshop on Automatic Identification Advanced Technologies (AUTOID'07)*, pages 63–68, 2007.
- [30] Unsang Park, Sharath Pankanti, and A. K. Jain. Fingerprint verification using sift features. In *Proc. SPIE 6944, Biometric Technology for Human Identification*, 2008.
- [31] F. Alonso-Fernandez, P. Tome-Gonzalez, V. Ruiz-Albacete, and J. Ortega-Garcia. Iris recognition based on sift features. In *IEEE Proc. Intl. Conf. on Biometrics, Identity and Security (BIDS)*, pages 1–8, 2009.
- [32] S. Berretti, A. D. Bimbo, P. Pala, B. B. Amor, and M. Daoudi. A set of selected sift features for 3D facial expression recognition. In *Proceedings of the 20th International Conference on Pattern Recognition (ICPR)*, pages 4125–4128, 2010.
- [33] Gilles Lebrun, Christophe Charrier, Olivier Lezoray, C. Meurie, and Hubert Cardot. Fast pixel classification by svm using vector quantization, tabu search and hybrid color space. In *the 11th International Conference on Computer Analysis of Images and Pattern (CAIP)*, pages 685–692, 2005.
- [34] V. Vapnik. *The Nature of Statistical Learning Theory*. Springer-Verlag, 1995.

- [35] Chih-Chung Chang and Chih-Jen Lin. *LIBSVM: a library for support vector machines*, 2001. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- [36] University of Essex. Faces94 database, face recognition data, 1994.
- [37] B. Hemery, C. Rosenberger, and H. Laurent. 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"*, 2007.
- [38] P.J. Phillips, H. Moon, S.A. Rizvi, and P.J. Rauss. The FERET evaluation methodology for face-recognition algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 22(10):1094–1104, 2000.
- [39] 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.
- [40] A.M. Martinez and R. Benavente. The AR face database. *CVC Tech. Report*, 1998.
- [41] 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.
- [42] ISO/IEC 19795-1. Information technology – biometric performance testing and reporting – part 1: Principles and framework, 2006.
- [43] Casia-fingerprintv5. <http://biometrics.idealtest.org/>.

- [44] G. Saporta. *Probabilits, Analyse des donnees et Statistiques*. Editions Technip, 1990.
- [45] C. I. Watson, M. D. Garris, E. Tabassi, C. L. Wilson, R. M. McCabe, S. Janet, and K. Ko. Users's Guide to NIST Biometric Image Software (NBIS). Technical report, National Institute of Standards and Technology (NIST), 2007.