A SVM-Based Model for the evaluation of biometric sample quality

Mohamad El-Abed, Romain Giot, Baptiste Hemery, Christophe Charrier,
Christophe Rosenberger

▶ **To cite this version:**

HAL Id: hal-00992462

https://hal.science/hal-00992462

Submitted on 18 May 2014

A SVM-Based Model For The Evaluation Of Biometric Sample Quality

M. El-Abed*, R. Giot*, B. Hemery*, C. Charrier* † and C. Rosenberger*

* GREYC Laboratory, ENSICAEN - University of Caen - CNRS

6 Boulevard Maréchal Juin, 14000 CAEN, France

Email: {mohamad.elabed, romain.giot, baptiste.hemery, christophe.rosenberger}@greyc.ensicaen.fr

†XLIM-SIC laboratory, University of Poitiers

Bat. SP2MI, BP. 30179, 86962 Futuroscope Chasseneuil cedex, France

Email: christophe.charrier@sic.univ-poitiers.fr

Abstract—One of the main factors affecting the performance of biometric systems is the quality of the acquired samples. Poor-quality samples increase the enrollment failure, and decrease the system performance. Therefore, it is important for a biometric system to estimate the quality of the acquired biometric samples. Toward this goal, we present in this paper a multi-class SVM-based method to predict sample quality. The proposed method uses two types of information: the first one is based on the image quality and the second is a pattern-based quality using the SIFT keypoints extracted from the image. For the experiments, we use four large and significant face databases to show the efficiency of the proposed method in predicting the system performance illustrated by the Equal Error Rate (EER).

Index Terms—Biometrics, quality assessment, performance, Support Vector Machine (SVM), Scale-Invariant Feature Transform (SIFT).

I. INTRODUCTION

Biometrics is considered as a promising solution among traditional methods based on "what we own" (such as a key) or "what we know" (such as a password) [1]. It is based on "what we are" and "how we behave". Biometric authentication systems have many applications [2] such as border control and e-commerce. The main benefits of this technology are to provide a better security and to facilitate the authentication process for a user. By contrast to traditional authentication methods (providing a 0% verification error), biometric systems are subject to errors computed by many metrics such as Failure to Acquire Rate (FTA), False Acceptance Rate (FAR) and False Rejection Rate (FRR) [3]. This inaccuracy is due to the variations of human characteristics (such as occlusions for iris recognition systems [4]), environment factors (*i.e.*, variation of acquisition conditions such as illuminations for face recognition systems [5]) and cross-device matching [6]. These kinds of variations may deeply affect biometric raw data quality. Poor-quality samples increase the enrollment failure rate, and decrease the system performance [5]. Therefore, quality assessment is considered as an important factor to take into account in both the enrollment and verification phases. Towards this goal, we propose a multi-class SVM-based method, to assess biometric raw data, which combines the use of image and pattern-based quality. Using quality information, the bad quality samples can be removed during the enrollment step or rejected during verification. Such information could also be used in soft biometrics or multimodal approaches [7], [8].

The outline of the paper is defined as follows. We present related previous research concerning biometric raw data quality in section II. Section III presents the proposed method. Section IV illustrates the experimental results on four large face benchmark databases. We conclude and give some perspectives of this work in section V.

II. BACKGROUND

The quality assessment of biometric raw data is receiving more and more attention in biometrics community. Many quality algorithms have been developed mainly for the fingerprint modality [9], [10], face [5], [11], iris [4], voice [12] and signature signals [13]. These works have demonstrated that the performance of biometric systems is heavily affected by the quality of the acquired biometric data. Tabassi *et al.* present in [9] a method based on the measurement of the matching scores to assess fingerprint quality. The proposed method uses a black box composed of two modules, *feature extraction* and *neural network*, which associates the image quality into five classes (excellent, very good, good, fair and poor). He *et al.* [14] 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 *et al.* present in [5] 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 [15]. However, this approach supposes the asymmetry hypothesis hence, could not be used for the others types of modalities. Other efforts [7], [8] have also been focused on the incorporation of biometric quality information to multimodal fusion approaches. Poh *et al.* present in [8] a

quality-dependent evaluation campaign. It aims at assessing how well fusion algorithms can perform under changing the quality of raw biometric images. The results from this evaluation have shown that the best fusion methods are those that exploit automatically derived quality measurements.

Discussion

The works done in quality assessment are very few in comparison to performance ones [16]. To our knowledge, most of the existing quality algorithms 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 method which can be considered as independent from the used matching system. It detects in a reasonable accuracy three types of real alterations that may deeply affect the global performance of the most widely used matching systems. The presented method 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.

III. DEVELOPED METHOD

The proposed method is designed to predict the sample quality using two types of information. The first one is based on the image quality (section III-A) and the second is a pattern-based quality using the SIFT keypoints extracted from the image (section III-B). The methodology principle is illustrated in figure 1: for an input image, the method constitutes a vector of both information and uses the multi-class Support Vector Machine (SVM) [17] classifier to assign a class for an image.
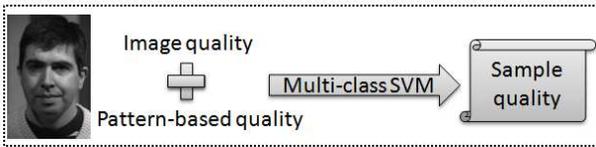

Fig. 1. General scheme of the proposed method

A. No-reference image quality

The development of general-purpose no-reference approaches to image quality assessment (NR-IQA) still lags recent advances in full-reference methods. Additionally, most no-reference or blind approaches are distortion-specific, meaning they assess only a specific type of distortion assumed present in the test image (such as blockiness, blur, or ringing). This limits their application domain. Other approaches rely on training a machine learning algorithm. These methods however, are only as effective as the features used to train their learning machines. The used NR-IQA method in this paper is the BLIINDS index introduced by Saad *et al.* [18].

This index is based on a DCT framework entirely. This makes it computationally convenient, uses a commonly used transform, and allows a coherent framework. The BLIINDS index is defined from four features that are then pooled together: i) a contrast feature υ_1, ii) a structure distortion based feature υ_2 and iii) iv) two anisotropic based measures υ_3, υ_4. 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 kth local DCT patch is computed as follows:

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

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 υ_1:

$$\upsilon_1 = \frac{1}{M} \sum_{i=1}^{M} c^i(x) \tag{2}$$

where M is the number of local patches. 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 wich 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:

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

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 υ_2. As degradation processes damage a scenes directional information, anisotropy measure, which is a directionally dependent quality of images, 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 directional information. Let the DCT coefficients of kth 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:

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

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

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

where $\beta > 1$. Finally, the two measures of anisotropy υ_3 and υ_4 are defined as

$$\upsilon_3 = \sigma(E(R_\theta^k)) \text{ and } \upsilon_4 = \max(E(R_\theta^k)), \forall k, \forall \theta \quad (6)$$

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:

$$\text{BLIINDS} = \prod_{i=1}^{M} \upsilon_1^{\alpha_1^i} \upsilon_2^{\alpha_2^i} \upsilon_3^{\alpha_3^i} \upsilon_4^{\alpha_4^i} \quad (7)$$

constraints by $\sum_{j=1}^{4} \sum_{i=1}^{M} \alpha_j^i = 1$ and where M represents the number of decomposition level used.

B. Pattern-based quality

The pattern-based quality information used to contribute to quality assessment uses the Scales Invariant Feature Transform (SIFT) descriptors [19]. SIFT algorithm consists of four major stages: i) scale-space extrema detection, ii) keypoint localization, iii) orientation assignment and iv) 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 key-point 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 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. From these features, we use the following four metrics that we considered as potentially interesting (see section IV-B) 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) and 4) 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-dimension vector containing image quality metric and four pattern-based metrics as depicted in table I.

TABLE I
THE FIVE-DIMENSIONAL VECTOR FOR PREDICTING BIOMETRIC SAMPLE QUALITY

Vector V	
Image quality metric	BLIINDS
Pattern-based metrics	keypoints
	DC coefficient
	mean of scales
	standard deviation of scales

C. Multi-class SVM 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 [20] and to their high generalization abilities. The SVMs were developed by Vapnik [17] 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). In our study, the input vector to SVM is the vector V (as depicted in table I) and the output can belong to ten different classes defined as follows (see table II):

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

TABLE II
SVM CLASSES DEFINITION

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

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 [17], [21] 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 :

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

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 [22]. This script automatically scales training and testing sets and searches 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 the majority class after the voting.

D. Discussion

The goal of the proposed method is to detect, in a reasonable accuracy, three real alterations which may deeply affect the most widely used matching systems. The proposed method may be considered as independent from the used matching system. An example of its practical use is illustrated in figure 2. The method predicts the class of the input image according to table II. Then, depending from the robustness of the used matching system against the predicted alteration, the matching system qualifies the image (good, fair or bad quality).

IV. Experimental results

In this section, we present the experimental results we obtained of the proposed method. The experimental protocol is presented in section IV-A. In order to validate our choice of the
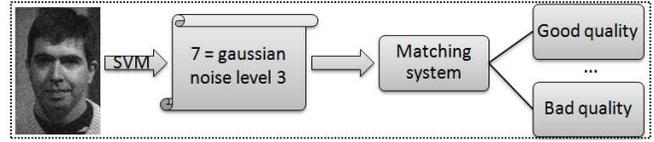

Fig. 2.   An example of use of the method

5 used metrics (1 for image quality and the rest for the pattern-based quality), we present in section IV-B the effectiveness of these metrics in detecting the three introduced alterations. In order to validate the proposed method, we present in section IV-C the SIFT matching algorithm as a use case. Finally, we show the efficiency of the proposed multi-class SVM-based quality method in predicting the performance of SIFT matching system using the Equal Error Rate (EER) metric. This error rate illustrates the overall system performance and is widely used to compare and evaluate the performance of biometric systems [3].

A. Protocol

In this study, we use four benchmark databases. For each database, we made three types of alterations (blurring, gaussian noise and resize alterations) and three levels for each type. The introduced alterations are commonly realistic during the acquisition of biometric data which may deeply affect the global performance of biometric systems. Finally, we have 40 databases: 4 reference databases and 36 altered databases (*i.e.*, 9 for each reference database):

1) Reference databases:

- FACES94 Database [23]: 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.

Fig. 3.   Samples from FACES94

- ENSIB Database [24]: It is composed of 100 individuals and 40 samples per individual. Each sample corresponds to one pose from the left one to the right.

Fig. 4.   Samples from ENSIB

- FERET Database [25], [26]: 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.

Fig. 5.   Samples from FERET

- AR database [27]: It is composed of 120 individuals and 26 samples per individual. These included images captured under different conditions of illumination, expression and occlusion.

Fig. 6.   Samples from AR

2) Altered databases: We generated 36 databases, using matlab, from the four reference databases: FACES94, ENSIB, FERET and AR databases. For each benchmark database, we generate three types of alterations and three levels for each type:

- 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 resize the image I using a nearest-neighbor interpolation.

Table III presents the parameters required of the used alteration matlab methods, and Figure 7 shows these alterations on a sample from FACES94 database.

B. Metrics behavior with alterations

In this section, we show the robustness of the used five metrics in detecting alterations presented in the previous section. To do so, we use the Pearson's correlation coefficient between two variables as defined in Eq. 11. It is defined as the covariance of the two variables divided by the product of their standard deviation:

$$Pearson(X,Y) = \frac{Cov(X,Y)}{\sigma_X \sigma_Y} \quad (11)$$

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

- $X_p = \{X_{pk}| \ k = 1 : 4\}$ where X_{p1} is the set of values of metric p for the reference databases images,

(a) blurring alteration

(b) gaussian noise alteration

(c) resize alteration

Fig. 7.   Alterations for a reference image from FACES94. From left to right, reference image then alteration level 1, 2 and 3

(X_{p2}, X_{p3}, X_{p4}) are the sets of values of metric 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, $Y = \{y_k|y_k = 1 \ for \ k = 1 : N, \ y_k = 2 \ for \ k = N + 1 : 2N, \ y_k = 3 \ for \ k = 2N + 1 : 3N \ and \ y_k = 4 \ for \ k = 3N + 1 : 4N\}$ where N is the size of the 4 reference databases.

Table IV shows that our four pattern-based metrics (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 metric 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 [18] is a multiresolution NR-IQA algorithm).

TABLE IV
PEARSON CORRELATION COEFFICIENTS BETWEEN THE PROPOSED
METRICS AND THE THREE ALTERATIONS AMONG THE ALL DATABASES

Metrics	$\rho_{blurring}$	$\rho_{gaussian\ noise}$	ρ_{resize}
$keypoints$	-0.5728	0.3901	-0.4880
$DC\ coefficient$	-0.6155	0.5672	-0.5252
$mean\ scale$	0.7933	-0.5632	-0.3960
$std\ scale$	0.3470	-0.3467	-0.4729
$BLIINDS$	0.6316	-0.8014	-0.1018

C. Biometric matching algorithm

The matching algorithm used in this paper is a SIFT-based [19] algorithm. The matching similarity principle used is described in previous works [28]. Each image im is described by a set of invariant features $X(im)$ as described in section III-B. The verification between two images im_1 and im_2

TABLE III
PARAMETERS OF THE MATLAB ALTERATION METHODS

Alteration type	method	level 1	level 2	level 3
Blurring	fspecial ('gaussian', hsize, σ)	hsize = [7 7] and σ = 1	hsize = [7 7] and σ = 2	hsize = [7 7] and σ = 6
Gaussian noise	imnoise (I, 'gaussian', m, v)	m=0.01 and v=0.003	m=0.01 and v=0.01	m=0.01 and v=0.017
Resize	imresize (I, scale, 'nearest')	scale = 0.8	scale = 0.6	scale = 0.4

corresponds to compute a 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 [19]. Given two keypoints $x \in X(im_1)$ and $y \in X(im_2)$, we say that x is associated to y iff:

$$d(x,y) = min_{\{z \in X(im_2)\}}d(x,z) \text{ and } d(x,y) \leq C \ d(x,y\prime) \tag{12}$$

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

$$d(x,y\prime) = min_{\{z \in X(im_2), \ d(x,z)>d(x,y)\}}d(x,z) \tag{13}$$

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 $d(x,y\prime)$ is significantly greater than $d(x,y)$. The significance of the necessary gap between $d(x,y)$ and $d(x,y\prime)$ is encoded by the constant C. Then, we consider that keypoint x is matched to y iff x is associated to y and y is associated to x. Figure 8 illustrates an example of matching results resulting from an impostor and a genuine comparison. The number of associations is used here as a similarity measure.
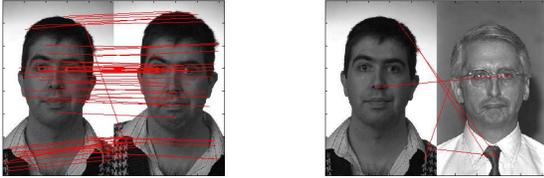

Fig. 8. Example of matching results resulting from a genuine (on the left) and an impostor comparisons (on the right)

D. Quality sets definition

The proposed multi-class SVM-based quality method predicts a class for an image. In order to quantify the efficiency of this method, we need first to define the quality sets for the used matching system. Depending on the used matching system, some alterations may not have an impact on it and others may deeply affect its performance. Therefore, we have tested the robustness of the used SIFT-based matching system against the introduced three alterations. We use the first image for the enrollment and the others for the test. We illustrate in table V the impact of adding altered images to the reference database images. We can clearly see that all the introduced

alterations have shown an impact on system performance illustrated by the Equal Error Rate (EER). Therefore, we have defined for this study the four quality sets as depicted by table VI. Certainly, we may have choose another definition of sets that will more penalize the resize alteration (*i.e.*, by putting label 9 in quality set IV) since it has the most impact on the used matching system.

TABLE V
EFFECT OF ALTERATIONS ON EACH DATABASE: VALUES OF EERS (%)

	FACES94	ENSIB	FERET	AR
original db	0.29	10.41	26	9.6
blurring alteration	1.65	15.54	29.61	13.84
gaussian noise alteration	0.76	13.3	27.81	11.72
resize alteration	9.74	18.67	34.17	18.2

TABLE VI
CATEGORY OF QUALITY

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

E. Validation of the proposed quality method

According to Gother *et al.* [29], biometric quality measurement algorithms should predict the matching performance. That is, a quality measurement algorithm 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 [3]. 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 [3]. In order to validate our method, we proceed as follows:

- Learning the multi-class SVM models: we learn four 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}). 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 used and the selection of the parameters required are presented in section III-C. Table VII illustrates the

accuracy of the produced multi-class SVM models on each benchmark database;

- Quality sets definition: we have considered four quality sets as defined in table VI;
- EER value for each quality set: in order to quantify the effectiveness of our quality method in predicting system performance, we have put each image to a quality set (according to table VI), using its predicted label by our method. Then, we have calculated the EER value for each quality set. The effectiveness of the method is quantified by how well our quality method 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).

TABLE VII
ACCURACY OF THE PRODUCED MULTI-CLASS SVM MODELS ON BOTH
TRAINING AND TEST SETS

	SVM_{each}		SVM_{all}	
	$S_{training}$	S_{test}	$S_{training}$	S_{test}
FACES94	91.01%	86.69%	85.68%	85.28%
ENSIB	97.73%	89.82%	94.92%	91.1%
FERET	82.33%	81.2%	82.29%	81.16%
AR	90.08%	89.08%	90.7%	88.92%

Tables VIII and IX illustrate the EER values for each quality set among the four used benchmark databases, using the four multi-class SVM models computed from the four benchmark databases, and the multi-class SVM model computed by examples from the four benchmark databases, respectively. The proposed quality method has shown its efficiency in predicting the used matching algorithm performance. From tables VIII and IX, we can also conclude:

- for FACES94, ENSIB and AR benchmark databases, there were no significant difference between the EERs of the reference database (*i.e.*, which is considered containing images of good quality) and the set predicted as good quality (set I) by our method;
- for the FERET benchmark database, there was a difference of 5.62% (table VIII) and 5.64% (table IX). This variation was due to the difficulties of FERET database which contains altered images by resolution. Despite this, our method has also shown to be efficient in predicting system performance in both cases.

V. CONCLUSION AND PERSPECTIVES

The evaluation of biometric raw data is receiving more and more attention in biometrics community. We present in this paper a multi-class SVM-based method for predicting the quality of biometric raw data. The method uses both kinds of information (image and pattern-based quality) to predict sample quality. In order to validate the proposed method, we use four large and significant face databases (FACES94, ENSIB, FERET and AR) and three kinds of real alterations (blurring, gaussian noise and resolution), that affecting the

TABLE VIII
EER VALUES OF THE FOUR REFERENCE DATABASES AND OF EACH
QUALITY SET AMONG THE FOUR BENCHMARK DATABASES. THESE EER
VALUES ARE COMPUTED USING THE FOUR MULTI-CLASS SVM MODELS
COMPUTED ON THE FOUR BENCHMARK DATABASES

EER values (%)					
Database	reference database	set I (good)	set II (fair)	set III (poor)	set IV (very poor)
FACES94	0.29	0.62	0.7	1.74	5.89
ENSIB	10.41	11.45	13.37	16.48	17.77
FERET	26	31.62	31.82	32.35	35.5
AR	9.6	10.26	11.82	16.15	16.81

TABLE IX
EER VALUES OF THE FOUR REFERENCE DATABASES AND OF EACH
QUALITY SET AMONG THE FOUR BENCHMARK DATABASES. THESE EER
VALUES ARE COMPUTED USING THE MULTI-CLASS SVM MODEL
COMPUTED USING EXAMPLES FROM THE FOUR BENCHMARK DATABASES

EER values (%)					
Database	reference database	set I (good)	set II (fair)	set III (poor)	set IV (very poor)
FACES94	0.29	0.62	0.72	1.77	5.79
ENSIB	10.41	11.21	13.4	16.61	17.64
FERET	26	31.64	31.78	32.27	35.69
AR	9.6	10.3	11.79	16.1	16.84

most widely used matching systems. The method has shown efficient in predicting system performance, illustrated by the EER, among the four used benchmark databases. The method can be considered as independent from the used matching system. For example, if the used matching system tolerate resize alteration, we could modify the quality sets definition by putting their labels (*i.e.*, labels 8, 9 and 10) in quality set I (*i.e.*, we consider that the altered image, by resize alteration, may be considered as of good quality for this type of matching system). On the other hand, the method is not based on asymmetry hypothesis. Thus, it may be used for several types of modalities such as hand veins. The method also could be applied directly on a single capture after training the model.

For the perspectives, we would like to quantify the effectiveness of the method on other types of modality (such as iris, fingerprint) to test if the method can be considered as modality-independent. We would also like to add another metric in order to detect luminance alteration, which also considered as an important alteration affecting biometric systems.

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).

REFERENCES

[1] J. Mahier, M. Pasquet, C. Rosenberger, and F. Cuozzo, "Biometric authentication," *Encyclopedia of Information Science and Technology*, 2008.

[2] A. K. Jain, S. Pankanti, S. Prabhakar, L. Hong, and A. Ross, "Biometrics: A grand challenge," *Pattern Recognition, International Conference*, pp. 935–942, 2004.

[3] "Information technology biometric performance testing and reporting," ISO/IEC 19795-1, Tech. Rep., 2006.

[4] E. Krichen, S. Garcia Salicetti, and B. Dorizzi, "A new probabilistic iris quality measure for comprehensive noise detection," in *BTAS07*, 2007, pp. 1–6.

[5] G. Zhang and Y. Wang, "Asymmetry-based quality assessment of face images," in *ISVC '09: Proceedings of the 5th International Symposium on Advances in Visual Computing*, 2009, pp. 499–508.

[6] N. Poh, J. Kittler, and T. Bourlai, "Quality-based score normalization with device qualitative information for multimodal biometric fusion," *SMC-A*, pp. 539–554, 2010.

[7] 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 Recogn.*, pp. 1094–1105, 2010.

[8] 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," *Trans. Info. For. Sec.*, 2009.

[9] E. Tabassi and C. Wilson, "A novel approach to fingerprint image quality," in *ICIP05*, 2005, pp. 37–40.

[10] M. Vatsa, R. Singh, A. Noore, and S. Singh, "Quality induced fingerprint identification using extended feature set," in *BTAS'08: Proceedings of the 2nd IEEE international conference on Biometrics: Theory, applications and systems*, 2008, pp. 1–6.

[11] J. Sang, Z. Lei, and S. Z. Li, "Face image quality evaluation for ISO/IEC standards 19794-5 and 29794-5," in *ICB '09: Proceedings of the Third International Conference on Advances in Biometrics*, 2009, pp. 229–238.

[12] A. Harriero, D. Ramos, J. Gonzalez-Rodriguez, and J. Fierrez, "Analysis of the utility of classical and novel speech quality measures for speaker verification," in *ICB '09: Proceedings of the Third International Conference on Advances in Biometrics*, 2009, pp. 434–442.

[13] S. Muller and O. Henniger, "Evaluating the biometric sample quality of handwritten signatures," in *ICB07*, 2007, pp. 407–414.

[14] Q. He, Z. Sun, T. Tan, and Y. Zou, "A hierarchical model for the evaluation of biometric sample quality," in *ICPR08*, 2008, pp. 1–4.

[15] X. Gao, S. Li, R. Liu, and P. Zhang, "Standardization of face image sample quality," in *ICB07*, 2007.

[16] D. Maio, D. Maltoni, J. L. Wayman, and A. K. Jain, "Fvc2004: Third fingerprint verification competition," in *in Proceedings of the First International Conference on Biometric Authentication*, 2004.

[17] V. Vapnik, *The Nature of Statistical Learning Theory*. Springer, 1995.

[18] M. Saad, A. C. Bovik, and C. Charrier, "A DCT statistics-based blind image quality index," *IEEE Signal Processing Letters*, 2010.

[19] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vision*, 2004.

[20] G. Lebrun, C. Charrier, O. Lezoray, C. Meurie, and H. Cardot, "Fast pixel classification by svm using vector quantization, tabu search and hybrid color space," in *the 11th International Conference on CAIP*, 2005, pp. 685–692.

[21] B. Schölkopf and A. J. Smola, *Learning with kernels : support vector machines, regularization, optimization, and beyond*, 2002.

[22] C.-C. Chang and C.-J. Lin, *LIBSVM: a library for support vector machines*, 2001, software available at http://www.csie.ntu.edu.tw/ cjlin/libsvm.

[23] U. of Essex, "Faces94 database, face recognition data," 1994. [Online]. Available: http://cswww.essex.ac.uk/mv/allfaces/faces94.html

[24] 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.

[25] P. Phillips, H. Moon, S. Rizvi, and P. Rauss, "The FERET evaluation methodology for face recognition algorithms," *IEEE Trans. Pattern Analysis and Machine Intelligence*, 2000.

[26] P. Phillips, H. Wechsler, J. Huang, and P. Rauss, "The FERET database and evaluation procedure for face recognition algorithms," *Journal of Image and Vision Computing*, 1998.

[27] A. Martinez and R. Benavente, "The AR face database," *CVC Tech. Report*, 1998.

[28] P.-O. Ladoux, C. Rosenberger, and B. Dorizzi, "Hand vein verification system based on sift matching," in *the 3rd IAPR/IEEE International Conference on Biometrics (ICB)*, 2009.

[29] P. Grother and E. Tabassi, "Performance of biometric quality measures," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 531–543, 2007.