Image Quality Assessment for Fake Biometric Detection: Application to Iris, Fingerprint, and Face Recognition

Javier Galbally, Sébastien Marcel, *Member, IEEE*, and Julian Fierrez

***Abstract*— To ensure the actual presence of a real legitimate trait in contrast to a fake self-manufactured synthetic or reconstructed sample is a significant problem in biometric authentication, which requires the development of new and efficient protection measures. In this paper, we present a novel software-based fake detection method that can be used in multiple biometric systems to detect different types of fraudulent access attempts. The objective of the proposed system is to enhance the security of biometric recognition frameworks, by adding liveness assessment in a fast, user-friendly, and non-intrusive manner, through the use of image quality assessment. The proposed approach presents a very low degree of complexity, which makes it suitable for real-time applications, using 25 general image quality features extracted from one image (i.e., the same acquired for authentication purposes) to distinguish between legitimate and impostor samples. The experimental results, obtained on publicly available data sets of fingerprint, iris, and 2D face, show that the proposed method is highly competitive compared with other state-of-the-art approaches and that the analysis of the general image quality of real biometric samples reveals highly valuable information that may be very efficiently used to discriminate them from fake traits.**

***Index Terms*— Image quality assessment, biometrics, security, attacks, countermeasures.**

I. Introduction

IN RECENT years, the increasing interest in the evaluation of biometric systems security has led to the creation of numerous and very diverse initiatives focused on this major field of research [1]: the publication of many research works disclosing and evaluating different biometric vulnerabilities [2], [3], the proposal of new protection methods [4], [5], related book chapters [6], the publication of several standards in the area [7], [8], the dedication of specific tracks, sessions and workshops in biometric-specific and general signal processing conferences [9], the organization of competitions focused on vulnerability assessment [10], [11], the acquisition of specific datasets [12], [13], the creation of groups and laboratories specialized in the evaluation of biometric security [14], or the existence of several European Projects with the biometric security topic as main research interest [15], [16].

All these initiatives clearly highlight the importance given by all parties involved in the development of biometrics (i.e., researchers, developers and industry) to the improvement of the systems security to bring this rapidly emerging technology into practical use.

Among the different threats analyzed, the so-called *direct* or *spoofing* attacks have motivated the biometric community to study the vulnerabilities against this type of fraudulent actions in modalities such as the iris [2], the fingerprint [17], the face [13], the signature [18], or even the gait [19] and multimodal approaches [20]. In these attacks, the intruder uses some type of synthetically produced artifact (e.g., gummy finger, printed iris image or face mask), or tries to mimic the behaviour of the genuine user (e.g., gait, signature), to fraudulently access the biometric system. As this type of attacks are performed in the analog domain and the interaction with the device is done following the regular protocol, the usual digital protection mechanisms (e.g., encryption, digital signature or watermarking) are not effective.

The aforementioned works and other analogue studies, have clearly shown the necessity to propose and develop specific protection methods against this threat. This way, researchers have focused on the design of specific countermeasures that enable biometric systems to detect fake samples and reject them, improving this way the robustness and security level of the systems.

Besides other anti-spoofing approaches such as the use of multibiometrics or challenge-response methods, special attention has been paid by researchers and industry to the *liveness detection* techniques, which use different physiological properties to distinguish between real and fake traits. Liveness assessment methods represent a challenging engineering problem as they have to satisfy certain demanding requirements [21]: (i) non-invasive, the technique should in no case be harmful for the individual or require an excessive contact with the user; (ii) user friendly, people should not be reluctant to use it; (iii) fast, results have to be produced in a

Manuscript received May 24, 2013; revised September 16, 2013; accepted November 13, 2013. Date of publication November 21, 2013; date of current version December 24, 2013. This work was supported in part by the Projects Contexts under Grant S2009/TIC-1485 from CAM, in part by Bio-Shield under Grant TEC2012-34881 from Spanish MECD, in part by TABULA RASA under Grant FP7-ICT-257289, in part by BEAT under Grant FP7-SEC-284989 from EU, and in part by Cátedra UAM-Telefónica. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Stefano Tubaro.

J. Galbally is with the Joint Research Centre, European Commission, Ispra 21027, Italy (e-mail: javier.galbally@jrc.ec.europa.es).

S. Marcel is with the IDIAP Research Institute Centre du Parc, Martigny 1920, Switzerland (e-mail: sebastien.marcel@idiap.ch).

J. Fierrez is with the Biometric Recognition Group-ATVS, EPS, Universidad Autonoma de Madrid, Madrid 28049, Spain (e-mail: julian.fierrez@uam.es).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2013.2292332

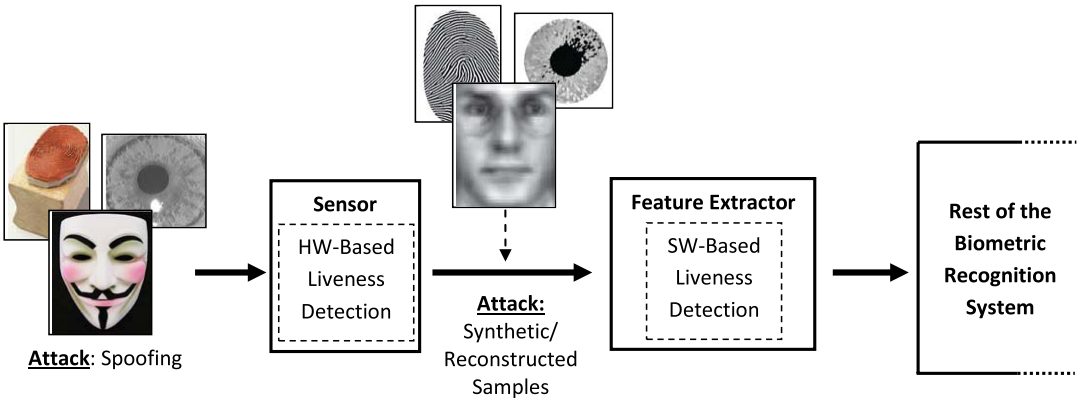


Fig. 1. Types of attacks potentially detected by hardware-based (spoofing) and software-based (spoofing + reconstructed/synthetic samples) liveness detection techniques.

very reduced interval as the user cannot be asked to interact with the sensor for a long period of time; (iv) low cost, a wide use cannot be expected if the cost is excessively high; (v) performance, in addition to having a good fake detection rate, the protection scheme should not degrade the recognition performance (i.e., false rejection) of the biometric system.

Liveness detection methods are usually classified into one of two groups (see Fig. 1): (i) *Hardware-based* techniques, which add some specific device to the sensor in order to detect particular properties of a living trait (e.g., fingerprint sweat, blood pressure, or specific reflection properties of the eye); (ii) *Software-based* techniques, in this case the fake trait is detected once the sample has been acquired with a standard sensor (i.e., features used to distinguish between real and fake traits are extracted from the biometric sample, and not from the trait itself).

The two types of methods present certain advantages and drawbacks over the other and, in general, a combination of both would be the most desirable protection approach to increase the security of biometric systems. As a coarse comparison, hardware-based schemes usually present a higher fake detection rate, while software-based techniques are in general less expensive (as no extra device is needed), and less intrusive since their implementation is transparent to the user. Furthermore, as they operate directly on the acquired sample (and not on the biometric trait itself), software-based techniques may be embedded in the feature extractor module which makes them potentially capable of detecting other types of illegal break-in attempts not necessarily classified as spoofing attacks. For instance, software-based methods can protect the system against the injection of reconstructed or synthetic samples into the communication channel between the sensor and the feature extractor [22], [23].

Although, as shown above, a great amount of work has been done in the field of spoofing detection and many advances have been reached, the attacking methodologies have also evolved and become more and more sophisticated. As a consequence, there are still big challenges to be faced in the detection of direct attacks.

One of the usual shortcomings of most anti-spoofing methods is their lack of generality. It is not rare to find that the proposed approaches present a very high performance detecting certain type of spoofs (i.e., gummy fingers made out of silicone), but their efficiency drastically drops when they are presented with a different type of synthetic trait (i.e., gummy fingers made out of gelatin). This way, their error rates vary greatly when the testing conditions are modified or if the evaluation database is exchanged. Moreover, the vast majority of current protection methods are based on the measurement of certain specific properties of a given trait (e.g., the frequency of ridges and valleys in fingerprints or the pupil dilation of the eye) which gives them a very reduced interoperability, as they may not be implemented in recognition systems based on other biometric modalities (e.g., face), or even on the same system with a different sensor.

In the present work we propose a novel software-based multi-biometric and multi-attack protection method which targets to overcome part of these limitations through the use of image quality assessment (IQA). It is not only capable of operating with a very good performance under different biometric systems (multi-biometric) and for diverse spoofing scenarios, but it also provides a very good level of protection against certain non-spoofing attacks (multi-attack). Moreover, being software-based, it presents the usual advantages of this type of approaches: fast, as it only needs one image (i.e., the same sample acquired for biometric recognition) to detect whether it is real or fake; non-intrusive; user-friendly (transparent to the user); cheap and easy to embed in already functional systems (as no new piece of hardware is required).

An added advantage of the proposed technique is its speed and very low complexity, which makes it very well suited to operate on real scenarios (one of the desired characteristics of this type of methods). As it does not deploy any trait-specific property (e.g., minutiae points, iris position or face detection), the computation load needed for image processing purposes is very reduced, using only *general* image quality measures fast to compute, combined with very simple classifiers.

It has been tested on publicly available attack databases of iris, fingerprint and 2D face, where it has reached results fully comparable to those obtained on the same databases and following the same experimental protocols by more complex trait-specific top-ranked approaches from the state-of-the-art.

The rest of the paper is structured as follows. Some key concepts about image quality assessment and the rationale

behind its use for biometric protection is given in Section II. The proposed method is described in Section III. The results for iris, fingerprint and 2D face evaluation experiments appear in Sections IV-A, IV-B, and IV-C. Conclusions are finally drawn in Section V.

II. Image Quality Assessment for Liveness Detection

The use of image quality assessment for liveness detection is motivated by the assumption that: "*It is expected that a fake image captured in an attack attempt will have different quality than a real sample acquired in the normal operation scenario for which the sensor was designed.*"

Expected quality differences between real and fake samples may include: degree of sharpness, color and luminance levels, local artifacts, amount of information found in both type of images (entropy), structural distortions or natural appearance. For example, iris images captured from a printed paper are more likely to be blurred or out of focus due to trembling; face images captured from a mobile device will probably be over- or under-exposed; and it is not rare that fingerprint images captured from a gummy finger present local acquisition artifacts such as spots and patches. Furthermore, in an eventual attack in which a synthetically produced image is directly injected to the communication channel before the feature extractor, this fake sample will most likely lack some of the properties found in natural images.

Following this "*quality-difference*" hypothesis, in the present research work we explore the potential of *general* image quality assessment as a protection method against different biometric attacks (with special attention to spoofing). As the implemented features do not evaluate any specific property of a given biometric modality or of a specific attack, they may be computed on any image. This gives the proposed method a new multi-biometric dimension which is not found in previously described protection schemes.

In the current state-of-the-art, the rationale behind the use of IQA features for liveness detection is supported by three factors:

- Image quality has been successfully used in previous works for image manipulation detection [24], [25] and steganalysis [26], [27] in the forensic field. To a certain extent, many spoofing attacks, especially those which involve taking a picture of a facial image displayed in a 2D device (e.g., spoofing attacks with printed iris or face images), may be regarded as a type of image manipulation which can be effectively detected, as shown in the present research work, by the use of different quality features.
- In addition to the previous studies in the forensic area, different features measuring trait-specific quality properties have already been used for liveness detection purposes in fingerprint and iris applications [5], [28]. However, even though these two works give a solid basis to the use of image quality as a protection method in biometric systems, none of them is general. For instance, measuring the ridge and valley frequency may be a good parameter to detect certain fingerprint spoofs, but it cannot be used in iris liveness detection. On the other hand, the amount of occlusion of the eye is valid as an iris anti-spoofing mechanism, but will have little use in fake fingerprint detection.
  This same reasoning can be applied to the vast majority of the liveness detection methods found in the state-of-the-art. Although all of them represent very valuable works which bring insight into the difficult problem of spoofing detection, they fail to generalize to different problems as they are usually designed to work on one specific modality and, in many cases, also to detect one specific type of spoofing attack.
- Human observers very often refer to the "different appearance" of real and fake samples to distinguish between them. As stated above, the different metrics and methods designed for IQA intend to estimate in an objective and reliable way the perceived appearance of images by humans.

Moreover, as will be explained in Section III, different quality measures present different sensitivity to image artifacts and distortions. For instance, measures like the mean squared error respond more to additive noise, whereas others such as the spectral phase error are more sensitive to blur; while gradient-related features react to distortions concentrated around edges and textures. Therefore, using a wide range of IQMs exploiting complementary image quality properties, should permit to detect the aforementioned quality differences between real and fake samples expected to be found in many attack attempts (i.e., providing the method with multi-attack protection capabilities).

All these observations lead us to believe that there is sound proof for the "quality-difference" hypothesis and that image quality measures have the potential to achieve success in biometric protection tasks.

III. The Security Protection Method

The problem of fake biometric detection can be seen as a two-class classification problem where an input biometric sample has to be assigned to one of two classes: real or fake. The key point of the process is to find a set of discriminant features which permits to build an appropriate classifier which gives the probability of the image "realism" given the extracted set of features. In the present work we propose a novel parameterization using 25 general image quality measures.

A general diagram of the protection approach proposed in this work is shown in Fig. 2. In order to keep its generality and simplicity, the system needs only one input: the biometric sample to be classified as real or fake (i.e., the same image acquired for biometric recognition purposes). Furthermore, as the method operates on the whole image without searching for any trait-specific properties, it does not require any preprocessing steps (e.g., fingerprint segmentation, iris detection or face extraction) prior to the computation of the IQ features. This characteristic minimizes its computational load. Once the feature vector has been generated the sample is classified as real (generated by a genuine trait) or fake (synthetically produced), using some simple classifiers. In particular, for our experiments we have considered standard implementations in Matlab of the Linear Discriminant

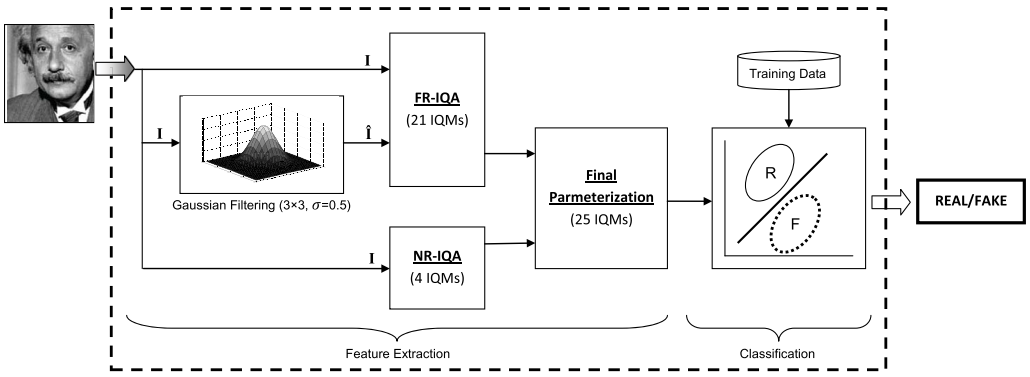


Fig. 2. General diagram of the biometric protection method based on Image Quality Assessment (IQA) proposed in the present work. IQM stands for Image Quality Measure, FR for Full-Reference, and NR for No-Reference. See Fig. 3 for a general classification of the 25 IQMs implemented. See Table I for the complete list and formal definitions of the 25 IQMs. See Section III for a more detailed description of each IQM.

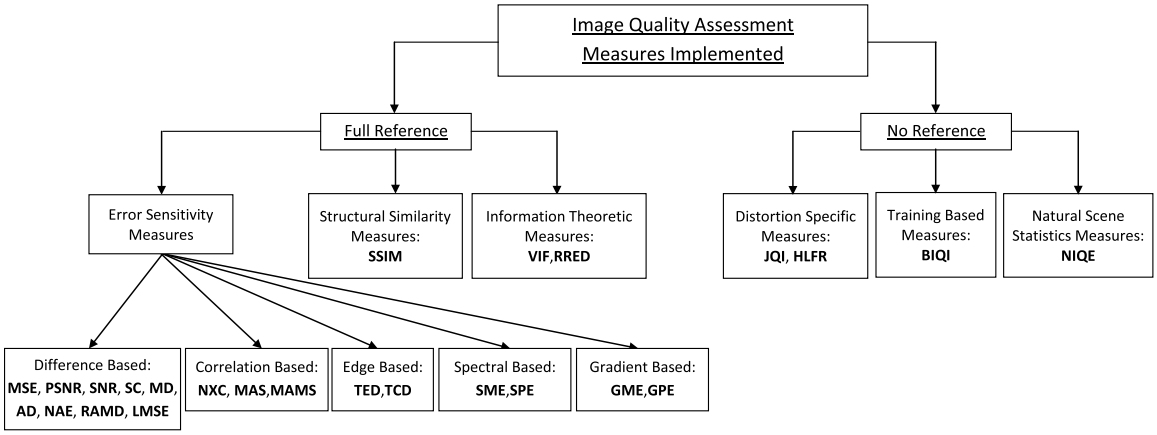


Fig. 3. Classification of the 25 image quality measures implemented in the work. Acronyms (in bold) of the different measures are explained in Table I and Section III.

Analysis (LDA) and Quadratic Discriminant Analysis (QDA) classifiers [44].

The parameterization proposed in the present work comprises 25 image quality measures both reference and blind (as will be introduced in the next sections). As it would be unfeasible to cover all the immense range of methods, approaches and perspectives proposed in the literature for IQA, the initial feature selection process to determine the set of 25 IQMs has been carried out according to four general criteria, which intend that the final method complies to the highest possible extent with the desirable requirements set for liveness detection systems (described in Section I). These four selection criteria are:

- *Performance*. Only widely used image quality approaches which have been consistently tested showing good performance for different applications have been considered.
- *Complementarity*. In order to generate a system as general as possible in terms of attacks detected and biometric modalities supported, we have given priority to IQMs based on complementary properties of the image (e.g., sharpness, entropy or structure).
- *Complexity*. In order to keep the simplicity of the method, low complexity features have been preferred over those which require a high computational load.
- *Speed*. This is, in general, closely related to the previous criterium (complexity). To assure a user-friendly non-intrusive application, users should not be kept waiting for a response from the recognition system. For this reason, big importance has been given to the feature extraction time, which has a very big impact in the overall speed of the fake detection algorithm.

The final 25 selected image quality measures are summarized in Table I. Details about each of these 25 IQMs are given in Sections III-A and III-B. For clarity, in Fig. 3 we show a diagram with the general IQM classification followed in these sections. Acronyms of the different features are highlighted in bold in the text and in Fig. 3.

A. Full-Reference IQ Measures

Full-reference (FR) IQA methods rely on the availability of a clean undistorted reference image to estimate the quality of the test sample. In the problem of fake detection addressed in this work such a reference image is unknown, as the detection system only has access to the input sample. In order to circumvent this limitation, the same strategy already successfully used for image manipulation detection in [24] and for steganalysis in [26], is implemented here.

As shown in Fig. 2, the input grey-scale image $\mathbf{I}$ (of size $N \times M$) is filtered with a low-pass Gaussian kernel ($\sigma = 0.5$ and size 3×3) in order to generate a smoothed version $\hat{\mathbf{I}}$. Then, the quality between both images ($\mathbf{I}$ and $\hat{\mathbf{I}}$) is computed according to the corresponding full-reference IQA metric.

TABLE I

LIST OF THE 25 IMAGE QUALITY MEASURES (IQMS) IMPLEMENTED IN THE PRESENT WORK AND USED FOR BIOMETRIC PROTECTION. ALL THE FEATURES WERE EITHER DIRECTLY TAKEN OR ADAPTED FROM THE REFERENCES GIVEN. IN THE TABLE: FR DENOTES FULL-REFERENCE AND NR NO-REFERENCE; $\mathbf{I}$ DENOTES THE REFERENCE CLEAN IMAGE (OF SIZE $N \times M$) AND $\hat{\mathbf{I}}$ THE SMOOTHED VERSION OF THE REFERENCE IMAGE. FOR OTHER NOTATION SPECIFICATIONS AND UNDEFINED VARIABLES OR FUNCTIONS WE REFER THE READER TO THE DESCRIPTION OF EACH PARTICULAR FEATURE IN SECTION III. ALSO, FOR THOSE FEATURES WITH NO MATHEMATICAL DEFINITION, THE EXACT DETAILS ABOUT THEIR COMPUTATION MAY BE FOUND IN THE GIVEN REFERENCES

#	Type	Acronym	Name	Ref.	Description
1	FR	MSE	Mean Squared Error	[29]	$MSE(\mathbf{I},\hat{\mathbf{I}}) = \frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M}(\mathbf{I}_{i,j}-\hat{\mathbf{I}}_{i,j})^2$
2	FR	PSNR	Peak Signal to Noise Ratio	[30]	$PSNR(\mathbf{I},\hat{\mathbf{I}}) = 10\log(\frac{\max(\mathbf{I}^2)}{MSE(\mathbf{I},\hat{\mathbf{I}})})$
3	FR	SNR	Signal to Noise Ratio	[31]	$SNR(\mathbf{I},\hat{\mathbf{I}}) = 10\log(\frac{\sum_{i=1}^{N}\sum_{j=1}^{M}(\mathbf{I}_{i,j})^2}{N\cdot M\cdot MSE(\mathbf{I},\hat{\mathbf{I}})})$
4	FR	SC	Structural Content	[32]	$SC(\mathbf{I},\hat{\mathbf{I}}) = \frac{\sum_{i=1}^{N}\sum_{j=1}^{M}(\mathbf{I}_{i,j})^2}{\sum_{i=1}^{N}\sum_{j=1}^{M}(\hat{\mathbf{I}}_{i,j})^2}$
5	FR	MD	Maximum Difference	[32]	$MD(\mathbf{I},\hat{\mathbf{I}}) = \max\lvert\mathbf{I}_{i,j}-\hat{\mathbf{I}}_{i,j}\rvert$
6	FR	AD	Average Difference	[32]	$AD(\mathbf{I},\hat{\mathbf{I}}) = \frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M}(\mathbf{I}_{i,j}-\hat{\mathbf{I}}_{i,j})$
7	FR	NAE	Normalized Absolute Error	[32]	$NAE(\mathbf{I},\hat{\mathbf{I}}) = \frac{\sum_{i=1}^{N}\sum_{j=1}^{M}\lvert\mathbf{I}_{i,j}-\hat{\mathbf{I}}_{i,j}\rvert}{\sum_{i=1}^{N}\sum_{j=1}^{M}\lvert\mathbf{I}_{i,j}\rvert}$
8	FR	RAMD	R-Averaged MD	[29]	$RAMD(\mathbf{I},\hat{\mathbf{I}},R) = \frac{1}{R}\sum_{r=1}^{R}\max_r\lvert\mathbf{I}_{i,j}-\hat{\mathbf{I}}_{i,j}\rvert$
9	FR	LMSE	Laplacian MSE	[32]	$LMSE(\mathbf{I},\hat{\mathbf{I}}) = \frac{\sum_{i=1}^{N-1}\sum_{j=2}^{M-1}(h(\mathbf{I}_{i,j})-h(\hat{\mathbf{I}}_{i,j}))^2}{\sum_{i=1}^{N-1}\sum_{j=2}^{M-1}h(\mathbf{I}_{i,j})^2}$
10	FR	NXC	Normalized Cross-Correlation	[32]	$NXC(\mathbf{I},\hat{\mathbf{I}}) = \frac{\sum_{i=1}^{N}\sum_{j=1}^{M}(\mathbf{I}_{i,j}\cdot\hat{\mathbf{I}}_{i,j})}{\sum_{i=1}^{N}\sum_{j=1}^{M}(\mathbf{I}_{i,j})^2}$
11	FR	MAS	Mean Angle Similarity	[29]	$MAS(\mathbf{I},\hat{\mathbf{I}}) = 1-\frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M}(\alpha_{i,j})$
12	FR	MAMS	Mean Angle Magnitude Similarity	[29]	$MAMS(\mathbf{I},\hat{\mathbf{I}}) = \frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M}(1-[1-\alpha_{i,j}][1-\frac{\lvert\lvert\mathbf{I}_{i,j}-\hat{\mathbf{I}}_{i,j}\rvert\rvert}{255}])$
13	FR	TED	Total Edge Difference	[33]	$TED(\mathbf{I},\hat{\mathbf{I}}) = \frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M}\lvert\mathbf{I}_{\mathbf{E}i,j}-\hat{\mathbf{I}}_{\mathbf{E}i,j}\rvert$
14	FR	TCD	Total Corner Difference	[33]	$TCD(I,\hat{I}) = \frac{\lvert N_{cr}-\hat{N}_{cr}\rvert}{\max(N_{cr},\hat{N}_{cr})}$
15	FR	SME	Spectral Magnitude Error	[34]	$SME(\mathbf{I},\hat{\mathbf{I}}) = \frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M}(\lvert\mathbf{F}_{i,j}\rvert-\lvert\hat{\mathbf{F}}_{i,j}\rvert)^2$
16	FR	SPE	Spectral Phase Error	[34]	$SPE(\mathbf{I},\hat{\mathbf{I}}) = \frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M}\lvert\arg(\mathbf{F}_{i,j})-\arg(\hat{\mathbf{F}}_{i,j})\rvert^2$
17	FR	GME	Gradient Magnitude Error	[35]	$SME(\mathbf{I},\hat{\mathbf{I}}) = \frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M}(\lvert\mathbf{G}_{i,j}\rvert-\lvert\hat{\mathbf{G}}_{i,j}\rvert)^2$
18	FR	GPE	Gradient Phase Error	[35]	$SPE(\mathbf{I},\hat{\mathbf{I}}) = \frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M}\lvert\arg(\mathbf{G}_{i,j})-\arg(\hat{\mathbf{G}}_{i,j})\rvert^2$
19	FR	SSIM	Structural Similarity Index	[36]	See [36] and practical implementation available in [37]
20	FR	VIF	Visual Information Fidelity	[38]	See [38] and practical implementation available in [37]
21	FR	RRED	Reduced Ref. Entropic Difference	[39]	See [39] and practical implementation available in [37]
22	NR	JQI	JPEG Quality Index	[40]	See [40] and practical implementation available in [37]
23	NR	HLFI	High-Low Frequency Index	[41]	$SME(\mathbf{I}) = \frac{\sum_{i=1}^{i_l}\sum_{j=1}^{j_l}\lvert\mathbf{F}_{i,j}\rvert-\sum_{i=i_h+1}^{N}\sum_{j=j_h+1}^{M}\lvert\mathbf{F}_{i,j}\rvert}{\sum_{i=1}^{N}\sum_{j=1}^{M}\lvert\mathbf{F}_{i,j}\rvert}$
24	NR	BIQI	Blind Image Quality Index	[42]	See [42] and practical implementation available in [37]
25	NR	NIQE	Naturalness Image Quality Estimator	[43]	See [43] and practical implementation available in [37]

This approach assumes that the loss of quality produced by Gaussian filtering differs between real and fake biometric samples. Assumption which is confirmed by the experimental results given in Section IV.

1) FR-IQMs: Error Sensitivity Measures: Traditional perceptual image quality assessment approaches are based on measuring the errors (i.e., signal differences) between the distorted and the reference images, and attempt to quantify these errors in a way that simulates human visual error sensitivity features.

Although their efficiency as signal fidelity measures is somewhat controversial [45], [46], up to date, these are probably the most widely used methods for IQA as they conveniently make use of many known psychophysical features of the human visual system [47], they are easy to calculate and usually have very low computational complexity.

Several of these metrics have been included in the 25-feature parameterization proposed in the present work. For clarity, these features have been classified here into five different categories (see Fig. 3) according to the image property measured [29]:

- **Pixel Difference measures** [29], [32]. These features compute the distortion between two images on the basis of their pixelwise differences. Here we include: Mean Squared Error (**MSE**), Peak Signal to Noise Ratio (**PSNR**), Signal to Noise Ratio (**SNR**), Structural Content (**SC**), Maximum Difference (**MD**), Average Difference (**AD**), Normalized Absolute Error (**NAE**), R-Averaged Maximum Difference (**RAMD**) and Laplacian Mean Squared Error (**LMSE**). The formal definitions for each of these features are given in Table I.
  In the RAMD entry in Table I, $\max_r$ is defined as the r-highest pixel difference between two images. For the present implementation, $R = 10$.
  In the LMSE entry in Table I, $h(\mathbf{I}_{i,j}) = \mathbf{I}_{i+1,j} + \mathbf{I}_{i-1,j} + \mathbf{I}_{i,j+1} + \mathbf{I}_{i,j-1} - 4\mathbf{I}_{i,j}$.
- **Correlation-based measures** [29], [32]. The similarity between two digital images can also be quantified in

terms of the correlation function. A variant of correlation-based measures can be obtained by considering the statistics of the angles between the pixel vectors of the original and distorted images. These features include (also defined in Table I): Normalized Cross-Correlation (**NXC**), Mean Angle Similarity (**MAS**) and Mean Angle-Magnitude Similarity (**MAMS**).

In the MAS and MAMS entries in Table I, $\alpha_{i,j}$ denotes the angle between two vectors, defined as, $\alpha_{i,j} = \frac{2}{\pi} \arccos \frac{\langle \mathbf{I}_{i,j}, \hat{\mathbf{I}}_{i,j} \rangle}{||\mathbf{I}_{i,j}|| \cdot ||\hat{\mathbf{I}}_{i,j}||}$, where $\langle \mathbf{I}_{i,j}, \hat{\mathbf{I}}_{i,j} \rangle$ denotes the scalar product. As we are dealing with positive matrices $\mathbf{I}$ and $\hat{\mathbf{I}}$, we are constrained to the first quadrant of the Cartesian space so that the maximum difference attained will be $\pi/2$, therefore the coefficient $2/\pi$ is included for normalization.

- **Edge-based measures**. Edges and other two-dimensional features such as corners, are some of the most informative parts of an image, which play a key role in the human visual system and in many computer vision algorithms including quality assessment applications [33].

  Since the structural distortion of an image is tightly linked with its edge degradation, here we have considered two edge-related quality measures: Total Edge Difference (**TED**) and Total Corner Difference (**TCD**).

  In order to implement both features, which are computed according to the corresponding expressions given in Table I, we use: (i) the Sobel operator to build the binary edge maps $\mathbf{I_E}$ and $\hat{\mathbf{I}}_E$; (ii) the Harris corner detector [48] to compute the number of corners N_{cr} and $\hat{N}_{cr}$ found in $\mathbf{I}$ and $\hat{\mathbf{I}}$.
- **Spectral distance measures**. The Fourier transform is another traditional image processing tool which has been applied to the field of image quality assessment [29]. In this work we will consider as IQ spectral-related features: the Spectral Magnitude Error (**SME**) and the Spectral Phase Error (**SPE**), defined in Table I (where $\mathbf{F}$ and $\hat{\mathbf{F}}$ are the respective Fourier transforms of $\mathbf{I}$ and $\hat{\mathbf{I}}$), and arg($\mathbf{F}$) denotes phase.
- **Gradient-based measures**. Gradients convey important visual information which can be of great use for quality assessment. Many of the distortions that can affect an image are reflected by a change in its gradient. Therefore, using such information, structural and contrast changes can be effectively captured [49].

  Two simple gradient-based features are included in the biometric protection system proposed in the present article: Gradient Magnitude Error (**GME**) and Gradient Phase Error (**GPE**), defined in Table I (where $\mathbf{G}$ and $\hat{\mathbf{G}}$ are the gradient maps of $\mathbf{I}$ and $\hat{\mathbf{I}}$ defined as $\mathbf{G} = \mathbf{G}_x + i\mathbf{G}_y$, where $\mathbf{G}_x$ and $\mathbf{G}_y$ are the gradients in the x and y directions).

2) FR-IQMs: Structural Similarity Measures: Although being very convenient and widely used, the aforementioned image quality metrics based on error sensitivity present several problems which are evidenced by their mismatch (in many cases) with subjective human-based quality scoring systems [45]. In this scenario, a recent new paradigm for image quality assessment based on structural similarity was proposed following the hypothesis that the human visual system is highly adapted for extracting structural information from the viewing field [36]. Therefore, distortions in an image that come from variations in lighting, such as contrast or brightness changes (nonstructural distortions), should be treated differently from structural ones.

Among these recent objective perceptual measures, the Structural Similarity Index Measure (**SSIM**), has the simplest formulation and has gained widespread popularity in a broad range of practical applications [36], [50]. In view of its very attractive properties, the SSIM has been included in the 25-feature parameterization.

3) FR-IQMs: Information Theoretic Measures: The quality assessment problem may also be understood, from an information theory perspective, as an information-fidelity problem (rather than a signal-fidelity problem). The core idea behind these approaches is that an image source communicates to a receiver through a channel that limits the amount of information that could flow through it, thereby introducing distortions. The goal is to relate the visual quality of the test image to the amount of information shared between the test and the reference signals, or more precisely, the mutual information between them. Under this general framework, image quality measures based on information fidelity exploit the (in some cases imprecise) relationship between statistical image information and visual quality.

In the present work we consider two of these information-theoretic features: the Visual Information Fidelity (**VIF**) [38] and the Reduced Reference Entropic Difference index (**RRED**) [39]. Both metrics are based on the information theoretic perspective of IQA but each of them take either a global or a local approximation to the problem, as is explained below.

The VIF metric measures the quality fidelity as the ratio between the total information (measured in terms of entropy) ideally extracted by the brain from the whole distorted image and the total information conveyed within the complete reference image. This metric relies on the assumption that natural images of perfect quality, in the absence of any distortions, pass through the human visual system (HVS) of an observer before entering the brain, which extracts cognitive information from it. For distorted images, it is hypothesized that the reference signal has passed through another "distortion channel" before entering the HVS. The VIF measure is derived from the ratio of two mutual information quantities: the mutual information between the input and the output of the HVS channel when no distortion channel is present (i.e., reference image information) and the mutual information between the input of the distortion channel and the output of the HVS channel for the test image. Therefore, to compute the VIF metric, the entire reference image is required as quality is assessed on a global basis.

On the other hand, the RRED metric approaches the problem of QA from the perspective of measuring the amount of local information difference between the reference image and the projection of the distorted image onto the space of natural images, for a given subband of the wavelet domain. In essence,

the RRED algorithm computes the average difference between scaled local entropies of wavelet coefficients of reference and projected distorted images in a distributed fashion. This way, contrary to the VIF feature, for the RRED it is not necessary to have access the entire reference image but only to a reduced part of its information (i.e., quality is computed locally). This required information can even be reduced to only one single scalar in case all the scaled entropy terms in the selected wavelet subband are considered in one single block.

B. No-Reference IQ Measures

Unlike the objective reference IQA methods, in general the human visual system does not require of a reference sample to determine the quality level of an image. Following this same principle, automatic no-reference image quality assessment (NR-IQA) algorithms try to handle the very complex and challenging problem of assessing the visual quality of images, in the absence of a reference. Presently, NR-IQA methods generally estimate the quality of the test image according to some pre-trained statistical models. Depending on the images used to train this model and on the *a priori* knowledge required, the methods are coarsely divided into one of three trends [51]:

- **Distortion-specific approaches**. These techniques rely on previously acquired knowledge about the type of visual quality loss caused by a specific distortion. The final quality measure is computed according to a model trained on clean images and on images affected by this particular distortion. Two of these measures have been included in the biometric protection method proposed in the present work.
  The JPEG Quality Index (**JQI**), which evaluates the quality in images affected by the usual block artifacts found in many compression algorithms running at low bit rates such as the JPEG [40].
  The High-Low Frequency Index (**HLFI**), which is formally defined in Table I. It was inspired by previous work which considered local gradients as a blind metric to detect blur and noise [41]. Similarly, the HLFI feature is sensitive to the sharpness of the image by computing the difference between the power in the lower and upper frequencies of the Fourier Spectrum. In the HLFI entry in Table I, i_l, i_h, j_l, j_h are respectively the indices corresponding to the lower and upper frequency thresholds considered by the method. In the current implementation, $i_l = i_h = 0.15N$ and $j_l = j_h = 0.15M$.
- **Training-based approaches**. Similarly to the previous class of NR-IQA methods, in this type of techniques a model is trained using clean and distorted images. Then, the quality score is computed based on a number of features extracted from the test image and related to the general model [42]. However, unlike the former approaches, these metrics intend to provide a general quality score not related to a specific distortion. To this end, the statistical model is trained with images affected by different types of distortions.
  This is the case of the Blind Image Quality Index (**BIQI**) described in [42], which is part of the 25 feature set used in the present work. The BIQI follows a two-stage framework in which the individual measures of different distortion-specific experts are combined to generate one global quality score.
- **Natural Scene Statistic approaches**. These blind IQA techniques use *a priori* knowledge taken from natural scene distortion-free images to train the initial model (i.e., no distorted images are used). The rationale behind this trend relies on the hypothesis that undistorted images of the natural world present certain *regular* properties which fall within a certain subspace of all possible images. If quantified appropriately, deviations from the regularity of natural statistics can help to evaluate the perceptual quality of an image [43].
  This approach is followed by the Natural Image Quality Evaluator (**NIQE**) used in the present work [43]. The NIQE is a completely blind image quality analyzer based on the construction of a quality aware collection of statistical features (derived from a corpus of natural undistorted images) related to a multi variate Gaussian natural scene statistical model.

IV. Experiments and Results

The evaluation experimental protocol has been designed with a two-fold objective:

- First, evaluate the "multi-biometric" dimension of the protection method. That is, its ability to achieve a good performance, compared to other trait-specific approaches, under different biometric modalities. For this purpose three of the most extended image-based biometric modalities have been considered in the experiments: iris, fingerprints and 2D face.
- Second, evaluate the "multi-attack" dimension of the protection method. That is, its ability to detect not only spoofing attacks (such as other liveness detection-specific approaches) but also fraudulent access attempts carried out with synthetic or reconstructed samples (see Fig. 1).

With these goals in mind, and in order to achieve reproducible results, we have only used in the experimental validation publicly available databases with well described evaluation protocols. This has allowed us to compare, in an objective and fair way, the performance of the proposed system with other existing state-of-the-art liveness detection solutions.

The task in *all* the scenarios and experiments described in the next sections is to automatically distinguish between real and fake samples. As explained in Section III, for this purpose we build a 25-dimensional simple classifier based on general IQMs (see Fig. 2). Therefore, in all cases, results are reported in terms of: the False Genuine Rate (FGR), which accounts for the number of false samples that were classified as real; and the False Fake Rate (FFR), which gives the probability of an image coming from a genuine sample being considered as fake. The Half Total Error Rate (HTER) is computed as $\text{HTER} = (\text{FGR} + \text{FFR})/2$.

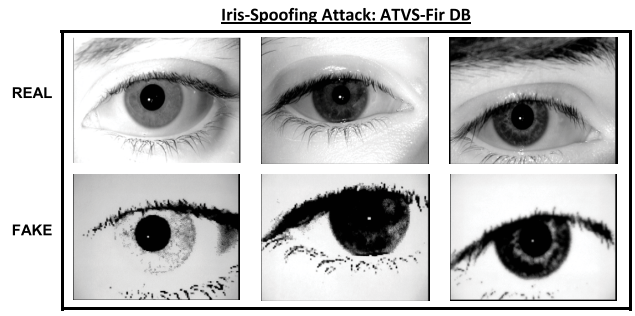


Fig. 4. Typical real iris images (top row) and their corresponding fake samples (bottom row) that may be found in the ATVS-FIr DB used in the iris-spoofing experiments. The database is available at http://atvs.ii.uam.es/.

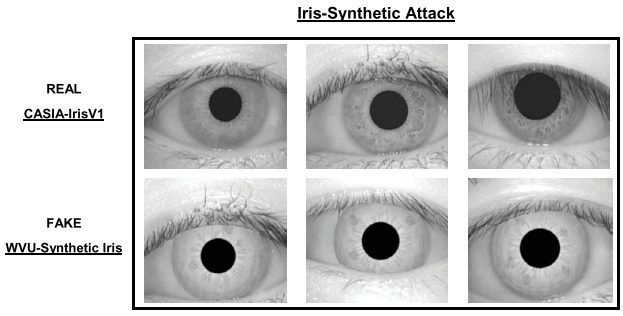


Fig. 5. Typical real iris images from CASIA-IrisV1 (top row) and fake samples from WVU-Synthetic Iris DB (bottom row), used in the iris-synthetic experiments. The databases are available at http://biometrics.idealtest.org and http://www.citer.wvu.edu/.

A. Results: Iris

For the iris modality the protection method is tested under two different attack scenarios, namely: i) spoofing attack and ii) attack with synthetic samples.

For each of the scenarios a specific pair of real-fake databases is used. Databases are divided into totally independent (in terms of users): train set, used to train the classifier; and test set, used to evaluate the performance of the proposed protection method.

In all cases the final results (shown in Table II) are obtained applying two-fold cross validation.

The classifier used for the two scenarios is based on Quadratic Discriminant Analysis (QDA) [44] as it showed a slightly better performance than Linear Discriminant Analysis (LDA), which will be used in the face-related experiments, while keeping the simplicity of the whole system.

1) Results: Iris-Spoofing: The database used in this spoofing scenario is the ATVS-FIr DB which may be obtained from the Biometric Recognition Group-ATVS.[1]

The database comprises real and fake iris images (printed on paper) of 50 users randomly selected from the BioSec baseline corpus [52]. It follows the same structure as the original BioSec dataset, therefore, it comprises 50 users $\times$ 2 eyes $\times$ 4 images $\times$ 2 sessions $=$ 800 fake iris images and its corresponding original samples. The acquisition of both real and fake samples was carried out using the LG IrisAccess EOU3000 sensor with infrared illumination which captures bmp grey-scale images of size 640×480 pixels.

In Fig. 4 we show some typical real and fake iris images that may be found in the dataset.

As mentioned above, for the experiments the database is divided into a: train set, comprising 400 real images and their corresponding fake samples of 50 eyes; and a test set with the remaining 400 real and fake samples coming from the other 50 eyes available in the dataset.

The liveness detection results achieved by the proposed approach under this scenario appear in the first row of Table II, where we can see that the method is able to correctly classify over 97% of the samples. In the last column we show the average execution time in seconds needed to process (extract the features and classify) each sample of the two considered databases. This time was measured on a standard 64-bit Windows7-PC with a 3.4 GHz processor and 16 GB RAM memory, running MATLAB R2012b.

As no other iris liveness detection method has yet been reported on the public ATVS-FIr DB, for comparison, the second row of Table II reports the results obtained on this database by a self-implementation of the anti-spoofing method proposed in [28]. It may be observed that the proposed method not only outperforms the state-of-the-art technique, but also, as it does not require any iris detection or segmentation, the processing time is around 10 times faster.

2) Results: Iris-Synthetic: In this scenario attacks are performed with synthetically generated iris samples which are injected in the communication channel between the sensor and the feature extraction module (see Fig. 1). The real and fake databases used in this case are:

- Real database: CASIA-IrisV1. This dataset is publicly available through the Biometric Ideal Test (BIT) platform of the Chinese Academy of Sciences Institute of Automation (CASIA).[2] It contains 7 grey-scale 320×280 images of 108 eyes captured in two separate sessions with a self-developed CASIA close-up camera and are stored in bmp format.
- Synthetic database: WVU-Synthetic Iris DB [23]. Being a database that contains only fully synthetic data, it is not subjected to any legal constraints and is publicly available through the CITeR research center.[3]
  The synthetic irises are generated following the method described in [23], which has two stages. In the first stage, a Markov Random Field model trained on the CASIA-IrisV1 DB is used to generate a background texture representing the global iris appearance. In the next stage, a variety of iris features such as radial and concentric furrows, collarette and crypts, are generated and embedded in the texture field. Following the CASIA-IrisV1 DB, this synthetic database includes 7 grey-scale 320×280 bmp images of 1,000 different subjects (eyes).

In Fig. 5 we show some typical real and fake iris images that may be found in the CASIA-IrisV1 DB and in the WVU-Synthetic Iris DB. It may be observe that, as a consequence of the training process carried out on the

[1] http://atvs.ii.uam.es/

[2] http://biometrics.idealtest.org

[3] http://www.citer.wvu.edu/

TABLE II
RESULTS (IN PERCENTAGE) OBTAINED BY THE PROPOSED BIOMETRIC PROTECTION METHOD BASED ON IQA FOR THE TWO ATTACKING SCENARIOS CONSIDERED IN THE IRIS MODALITY: SPOOFING (TOP ROW) AND SYNTHETIC (BOTTOM ROW). FOR COMPARISON, THE MIDDLE ROW REPORTS THE RESULTS OBTAINED BY A SELF-IMPLEMENTATION OF THE ANTI-SPOOFING METHOD PRESENTED IN [28]. THE LAST COLUMN INDICATES, IN SECONDS, THE AVERAGE EXECUTION TIME TO PROCESS EACH SAMPLE

	Results: Iris			
	FFR	FGR	HTER	Av. Exec. (s)
Iris-Spoof.	4.2	0.25	2.2	0.238
Iris-Spoof. [28]	1.3	4.9	3.1	2.563
Iris-Synthetic	3.4	0.8	2.1	0.156

CASIA-IrisV1 DB, the synthetic samples are visually very similar to those of the real dataset, which makes them specially suitable for the considered attacking scenario.

The last column indicates, in seconds, the average execution time to process each sample.

In the experiments, in order to have balanced training classes (real and fake) only 54 synthetic eyes (out of the possible 1,000) were randomly selected. This way, the problem of overfitting one class over the other is avoided. The test set comprises the remaining 54 real eyes and 946 synthetic samples.

The results achieved by the proposed protection method based on IQA on this attacking scenario are shown in the bottom row of Table II. In spite of the similarity of real and fake images, the global error of the algorithm in this scenario is 2.1%.

The experiments reported in this Section IV-A show the ability of the approach to adapt to different attacking scenarios and to keep a high level of protection in all of them. Therefore, the results presented in Table II confirm the "multi-attack" dimension of the proposed method.

B. Results: Fingerprints

For the fingerprint modality, the performance of the proposed protection method is evaluated using the LivDet 2009 DB [10] comprising over 18,000 real and fake samples.

As in the iris experiments, the database is divided into a: train set, used to train the classifier; and test set, used to evaluate the performance of the protection method. In order to generate totally unbiased results, there is no overlap between both sets (i.e., samples corresponding to each user are just included in the train or the test set).

The same QDA classifier already considered in the iris-related experiments is used here.

1) Results: Fingerprints-Spoofing LivDet: The LivDet 2009 DB [10] was captured in the framework of the 2009 Fingerprint Liveness Detection Competition and it is distributed through the site of the competition.[4] It comprises

[4]http://prag.diee.unica.it/LivDet09/

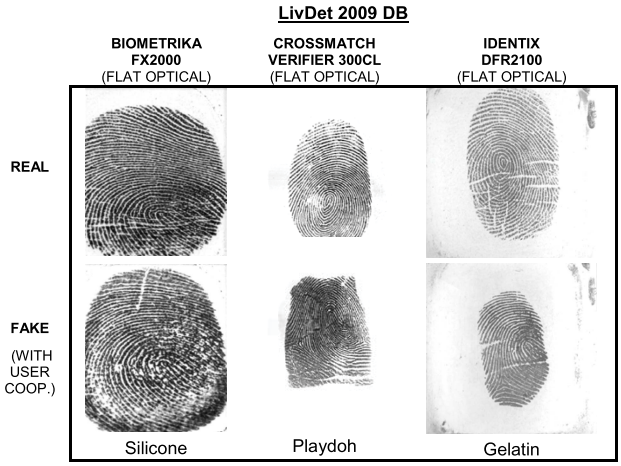


Fig. 6. Typical examples of real and fake fingerprint images that can be found in the public LivDet09 database used in the fingerprint anti-spoofing experiments. The database is available at http://prag.diee.unica.it/LivDet09/.

three datasets of real and fake fingerprints captured each of them with a different flat optical sensor: i) Biometrika FX2000 (569 dpi), ii) CrossMatch Verifier 300CL (500 dpi), and iii) Identix DFR2100 (686dpi). The gummy fingers were generated using three different materials: silicone, gelatine and playdoh, always following a consensual procedure (with the cooperation of the user). As a whole, the database contains over 18,000 samples coming from more than 100 different fingers.

Some typical examples of the images that can be found in this database are shown in Fig. 6, where the material used for the generation of the fake fingers is specified (silicone, gelatine or playdoh).

The train and test sets selected for the evaluation experiments on this database are the same as the ones used in the LivDet 2009 competition, so that the results obtained by the proposed method based on general IQA may be directly compared to the participants of the contest. The general distribution of the database in the train and test sets is specified in Table IV.

Results achieved on this database are shown in the first two rows of Table III. For clarity, only the best results achieved on LivDet09 for each of the individual datasets is given (second row). The best performance obtained by any of the reported methods on each of the three datasets is highlighted in bold in order to facilitate the comparison of the results.

In [53], a novel fingerprint liveness detection method combining perspiration and morphological features was presented and evaluated on the LivDet09 database following the same protocol (training and test sets) used in the competition. In that work, comparative results were reported with particular implementations (from the authors) of the techniques proposed in: [54], based on the wavelet analysis of the finger tip texture; [55], based on the curvelet analysis of the finger tip texture; and [56] based on the combination of local ridge frequencies and multiresolution texture analysis. In the rows 3-7 of Table III we also present these results so that they may be compared with our proposed IQA-based method (row one). In the bottom row we show the average execution time in seconds needed to process (extract the fea-

TABLE III

RESULTS (IN PERCENTAGE) OBTAINED IN THE LIVDET 2009 DB BY: THE PROPOSED BIOMETRIC PROTECTION METHOD (IQA-BASED, TOP ROW); EACH OF THE BEST APPROACHES PARTICIPATING IN LIVDET 2009 [10] (SECOND ROW); THE METHOD PROPOSED IN [53] WHICH COMBINES PERSPIRATION AND MORPHOLOGICAL FEATURES (THIRD ROW); THE METHOD PROPOSED IN [54] BASED ON THE WAVELET ANALYSIS OF THE FINGER TIP TEXTURE, ACCORDING TO AN IMPLEMENTATION FROM [53] (FOURTH ROW); THE METHOD PROPOSED IN [55] BASED ON THE CURVELET ANALYSIS OF THE FINGER TIP TEXTURE, ACCORDING TO AN IMPLEMENTATION FROM [53] (FOURTH ROW); THE METHOD PROPOSED IN [56] BASED ON THE COMBINATION OF LOCAL RIDGE FREQUENCIES AND MULTIRESOLUTION TEXTURE ANALYSIS, ACCORDING TO AN IMPLEMENTATION FROM [53] (FIFTH ROW). THE BEST PERFORMANCE REPORTED ON EACH OF THE DATASETS IS HIGHLIGHTED IN **BOLD**. THE BOTTOM ROW SHOWS, IN SECONDS, THE AVERAGE EXECUTION TIME OF THE PROPOSED METHOD TO PROCESS EACH SAMPLE OF THE THREE DATASETS

	Comparative Results: Fingerprints-LivDet09								
	Biometrika			**CrossMatch**			**Identix**		
	FFR	**FGR**	**HTER**	**FFR**	**FGR**	**HTER**	**FFR**	**FGR**	**HTER**
IQA-based	14.0	**11.6**	12.8	8.6	12.8	10.7	**1.1**	**1.4**	**1.2**
Best LivDet09 [10]	15.6	20.7	18.2	**7.4**	**11.4**	**9.4**	2.7	2.8	2.8
Marasco et al. [53]	**12.2**	13.0	**12.6**	17.4	12.9	15.2	8.3	11.0	9.7
Moon et al. [54] reported in [53]	20.8	25.0	23.0	27.4	19.6	23.5	74.7	1.6	38.2
Nikam et al. [55] reported in [53]	14.3	42.3	28.3	19.0	18.4	18.7	23.7	37.0	30.3
Abhyankar et al. [56] reported in [53]	24.2	39.2	31.7	39.7	23.3	31.5	48.4	46.0	47.2
Av. Exec. (s)	0.169			0.231			0.368		

TABLE IV

GENERAL STRUCTURE OF THE LIVDET 2009 DB. THE DISTRIBUTION OF THE FAKE SAMPLES IS GIVEN IN TERMS OF THE MATERIALS USED FOR THEIR GENERATION: *g* STANDS FOR GELATIN, *p* FOR PLAYDOH AND *s* FOR SILICONE

	LivDet 2009 DB			
	Train (Real/Fake)		Test (Real/Fake)	
	# Fingers	# Samples	# Fingers	# Samples
Biometrika	13/13	520/520s	39/13	1473/1480s
CrossMatch	35/35	1000/1000 (344g+346p+310s)	100/35	3000/3000 (1036g+1034p+930s)
Identix	63/35	750/750 (250g+250p+250s)	100/35	2250/2250 (750g+750p+750s)

tures and classify) each sample of the three datasets. This time was measured on a standard 64-bit Windows7-PC with a 3.4 GHz processor and 16 GB RAM memory, running MATLAB R2012b. Due to the high simplicity of the method, the computational cost of processing an image depends almost exclusively on the size of the sample.

The results given in Table III show that our method outperforms all the contestants in LivDet 2009 in two of the datasets (Biometrika and Identix), while its classification error is just slightly worse than the best of the participants for the Crossmatch data. The classification error rates of our approach are also clearly lower than those reported in [53] for the different liveness detection solutions tested.

The results obtained in the fingerprint-based comparative experiments strengthen the first observations made in Section IV-A about the generality of the method, which is not only capable of adapting to different biometric modalities and attacks, but it also performs better than well known methods from the state-of-the-art.

C. Results: 2D Face

The performance of the IQA-based protection method has also been assessed on a face spoofing database: the REPLAY-ATTACK DB [57] which is publicly available from the IDIAP Research Institute.[5]

The database contains short videos (around 10 seconds in mov format) of both real-access and spoofing attack attempts of 50 different subjects, acquired with a 320 × 240 resolution webcam of a 13-inch MacBook Laptop. The recordings were carried out under two different conditions: i) *controlled*, with a uniform background and artificial lighting; and ii) *adverse*, with natural illumination and non-uniform background.

Three different types of attacks were considered: i) *print*, illegal access attempts are carried out with hard copies of high-resolution digital photographs of the genuine users; ii) *mobile*, the attacks are performed using photos and videos taken with the iPhone using the iPhone screen; iii) *highdef*, similar to the mobile subset but in this case the photos and videos are displayed using an iPad screen with resolution 1024 × 768.

In addition, access attempts in the three attack subsets (print, mobile and highdef) were recorded in two different modes depending on the strategy followed to hold the attack replay device (paper, mobile phone or tablet): i) *hand*-based and ii) *fixed*-support.

Such a variety of real and fake acquisition scenarios and conditions makes the REPLAY-ATTACK DB a unique benchmark for testing anti-spoofing techniques for face-based systems. As a consequence, the print subset was selected as the evaluation dataset in the 2011 Competition on Counter Measures to 2D Facial Spoofing Attacks [11].

Some typical images (frames extracted from the videos) from real and fake (print, mobile and highdef) access attempts that may be found in the REPLAY-ATTACK DB are shown in Fig. 7.

[5]https://www.idiap.ch/dataset/replayattack

TABLE V

RESULTS (IN PERCENTAGE) OBTAINED ON THE REPLAY-ATTACK DB BY THE PROPOSED BIOMETRIC PROTECTION METHOD FOR THE DIFFERENT SCENARIOS CONSIDERED IN THE DATASET AND FOLLOWING THE ASSOCIATED EVALUATION PROTOCOL. THE BOTTOM ROW SHOWS, IN SECONDS, THE AVERAGE EXECUTION TIME OF THE PROPOSED METHOD TO PROCESS EACH SAMPLE OF THE THREE DATASETS (THE GRANDTEST DATASET IS A COMBINATION OF THE THREE PREVIOUS ONES AS EXPLAINED IN SECT. IV-C)

	Results: Face Replay-Attack DB											
	Print			**Mobile**			**Highdef**			**Grandtest**		
	FFR	**FGR**	**HTER**	**FFR**	**FGR**	**HTER**	**FFR**	**FGR**	**HTER**	**FFR**	**FGR**	**HTER**
Hand	13.6	5.0	9.3	1.9	3.7	2.8	15.6	10.5	13.1	19.6	11.3	15.4
Fixed	11.5	5.3	8.4	2.8	4.1	3.5	8.4	9.9	9.1	13.7	11.7	12.7
All	11.6	4.1	7.9	2.4	3.9	3.2	14.0	10.2	12.1	17.9	12.5	15.2
Av. Exec. (s)	0.148			0.150			0.147					

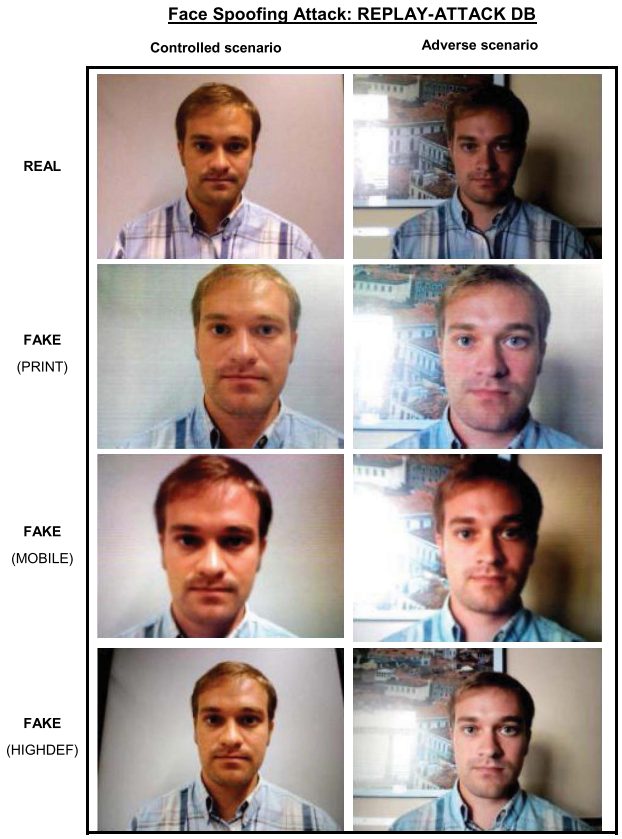


Fig. 7. Typical examples of real and fake (print, mobile and highdef) face images that can be found in the public REPLAY-ATTACK DB used in the face anti-spoofing experiments. Images were extracted from videos acquired in the two considered scenarios: controlled and adverse. The database is available at https://www.idiap.ch/dataset/replayattack.

The database has a perfectly defined associated evaluation protocol which considers three totally independent datasets (in terms of users): train, used to tune the parameters of the method; development, to fix the decision threshold; and test, where final results are computed. The protocol is released with the database and has been strictly followed in the present experiments. The general structure of the protocol is specified in Table VI.

The database is also released with face detection data. These data was used to crop and normalize all the faces to a 64×64 bounding box prior to the anti-spoofing experiments. This way the final classification results are ensured to be totally unbiased

TABLE VI

STRUCTURE OF THE EVALUATION PROTOCOLS RELEASED WITH THE REPLAY-ATTACK DB. THE DISTRIBUTION OF THE FAKE VIDEOS IS GIVEN IN TERMS OF THE PROCEDURE USED FOR THEIR ACQUISITION: HAND-HELD OR FIXED SUPPORT

	REPLAY-ATTACK DB		
	# Videos per subset (hand/fixed)		
	Train	Development	Test
Real-Accesses	60	60	80
Print-Attacks	60 (30/30)	60 (30/30)	80 (40/40)
Mobile-attacks	120 (60/60)	120 (60/60)	160 (80/80)
Highdef-attacks	120 (60/60)	120 (60/60)	160 (80/80)

and not dependent on contextual-specific artifacts such as: unwanted changes in the background; different sizes of the heads (we can see in Fig. 7 that fake faces are in general slightly bigger than the ones in real images); a black frame due to an imperfect fitting of the attack media on the capturing device screen, etc.

As the proposed IQA-based method is a single-image technique (i.e., it just needs one input image and not a sequence of them), each frame of the videos in the REPLAY-ATTACK DB has been considered as an independent sample. Therefore, classification (real or fake) is done on a frame-by-frame basis and not per video.

In Table V we show the results obtained on the test set by the proposed method using in this case a standard classifier based on Linear Discriminant Analysis (LDA), as for the face problem it showed slightly better performance than the QDA classifier used in the previous two cases (iris and fingerprints). In the bottom row we show the average execution time in seconds needed to process (extract the features and classify) each sample of the three datasets (print, mobile and highdef, as the grandtest scenario is a combination of the previous three as is explained below). As in the iris and fingerprint experiments, this time was measured on a standard 64-bit Windows7-PC with a 3.4 GHz processor and 16 GB RAM memory, running MATLAB R2012b. Recall that the print, mobile and highdef scenarios refer to the type of artifact being used as forgery and not to the acquisition device, which is de same for all cases (320×240 resolution webcam of a 13-inch MacBook Laptop). Therefore, as expected, the sample average processing time in all the datasets is almost identical.

TABLE VII
COMPARISON OF THE RESULTS (IN PERCENTAGE) OBTAINED BY THE IQA-BASED PROTECTION METHOD PROPOSED IN THE PRESENT WORK, AND THE LBP-BASED ANTI-SPOOFING TECHNIQUES DESCRIBED IN [57] (PARTIALLY BASED ON THE RESULTS REPORTED ON [58]). RESULTS ARE OBTAINED FOLLOWING THE GRANDTEST-ALL SUPPORTS PROTOCOL OF THE REPLAY-ATTACK DB

	Comparative: REP.-ATT. DB (grandtest)		
	FFR	**FGR**	**HTER**
IQA-based	17.9	12.5	15.2
LBP-based [57]	-	-	15.2
LBP-based [57], [58]	-	-	13.9

In the *grandtest* experiments (also defined in the associated protocol) the protection method is trained using data from the print, mobile and highdef scenarios, and tested also on samples from the three type of attacks. This is probably the most realistic attack case, as, in general, we cannot know *a priori* the type of artifact (paper, mobile phone or tablet) that the attacker will use to try to break into the system.

Results in Table V are also presented in terms of the type of strategy followed to hold the attack replay device: hand-based, fixed-support or all (where data of the previous two types is used).

The performance shown by the proposed algorithm in the face-based evaluation confirm the conclusions extracted from the iris and fingerprint experiments: the IQA-based protection method is able to adapt to different modalities, databases and attacks performing consistently well in all of them.

In [57] different LBP-based anti-spoofing techniques (partially based on the study presented in [58]) were tested following the exact same protocol used in the present work. Results were only reported on the grandtest scenario considering all types of supports (hand and fixed). A comparison between both protection approaches (IQA-based and LBP-based) appears in Table VII. The error rates of all methods are very similar, however, the IQA-based has the advantage of its simplicity and generality.

In the 2011 Competition on Counter Measures to 2D Facial Spoofing Attacks 2011 [11] there were several important differences with the protocol followed in the present work: i) only the print subset was used (considering both hand and fixed supports); ii) faces were not necessarily cropped and normalized (which, as mentioned before, may lead to optimistically biased results); and iii) classification was carried out on a video-basis and not frame-by-frame as in our experiments (i.e., systems in the competition exploited both spatial and temporal information). Therefore, a fully fair comparison between the competition and the present work is not possible.

However, for reference, in Table VIII we present the results obtained by the different participants in the competition compared to the performance of our method without doing the cropping and normalization of the videos. We can observe that, even though many of the contestants were using a sequence of frames to classify each video (with the complexity and speed decrease that this entails), our proposed IQA-based method performs similarly to the top ranked systems.

TABLE VIII
COMPARISON OF THE RESULTS (IN PERCENTAGE) OBTAINED BY THE IQA-BASED PROTECTION METHOD, AND THE DIFFERENT PARTICIPANTS IN THE 1ST COMPETITION ON COUNTER MEASURES TO 2D FACIAL SPOOFING ATTACKS 2011 [11]. RESULTS ARE OBTAINED ON THE PRINT SUBCORPUS OF THE REPLAY-ATTACK DB. *Motion* INDICATES THAT THE SYSTEM NEEDS TEMPORAL INFORMATION TO DETECT FAKES

	Comparative: REP.-ATT. DB (print)		
	FFR	**FGR**	**HTER**
IQA-based	0.0	1.0	0.5
AMILAB (motion) [11]	0.0	1.2	0.6
CASIA (motion) [11]	0.0	0.0	0.0
IDIAP [11]	0.0	0.0	0.0
SIANI (motion) [11]	0.0	21.2	10.6
UNICAMP (motion) [11]	1.2	0.0	0.6
UOULU [11]	0.0	0.0	0.0

Furthermore, several of the algorithms presented to the competition are based on motion-detection of the face and, therefore, their ability to detect fake access attempts carried out with replayed motion videos (mobile and highdef scenarios) would be at least under question.

It should also be noted that in many applications there is no access to a video of the user (i.e., no temporal information is available). For these scenarios, many of the anti-spoofing solutions presented at the competition (marked with *motion* in Table VIII) would not be usable as they are not designed to work on a single static face image.

D. Preliminary Feature Individuality Analysis

In this section we present a preliminary study of the discriminative power of the different quality features used in the proposed protection method. Although a deeper analysis of the features relevance for each of the considered experimental scenarios would be advisable, such a rigorous examination would represent on its own the topic for a new research work which falls out of the scope of the present contribution.

The Sequential Forward Floating Selection (SFFS) algorithm has been used to determine if certain individual features, or certain subsets of features, present a higher discrimination capability than others under the biometric security experimental framework considered in the work. The SFFS method is a deterministic, single-solution feature selection algorithm first proposed in [59], which has shown remarkable performance over other suboptimal selection schemes [60].

In the current experimental analysis, the selection criterion to be optimized by the SFFS algorithm is the HTER achieved by the system in the test set following the experimental protocols described in Sects. IV-A, IV-B and IV-C (the classifiers are the same ones used in the previous experimental sections of the work). In particular, the SFFS algorithm has been used to search for the best performing feature subsets of dimensions: 5, 10, 15 and the best overall subset regardless of its size.

For the sake of argument, the results obtained for three representative scenarios of those considered in the previous

TABLE IX
BEST PERFORMING FEATURE SUBSETS OF DIMENSIONS 5, 10, 15 AND BEST-OVERALL, FOUND USING THE SFFS ALGORITHM ACCORDING TO THE HTER ON THE TEST SET OF THE ATVS-FIR DB

	Best feature subsets - ATVS-FIr DB		
	Full Reference	**No Reference**	**HTER**
Best-5	MSE, PSNR, NAE, SME	BIQI	51.4
Best-10	MSE, PSNR, SC, NXC, SNR, SME, GPE	BIQI, NIQE, HLFI	47.5
Best-15	MSE, PSNR, SC, NXC, LMSE, NAE, SNR, SME, SPE, TCD, GPE, VIF	BIQI, NIQE, HLFI	34.1
Best-overall	All	All	2.2

TABLE X
BEST PERFORMING FEATURE SUBSETS OF DIMENSIONS 5, 10, 15 AND BEST-OVERALL, FOUND USING THE SFFS ALGORITHM ACCORDING TO THE HTER ON THE TEST SET OF THE LIVDET09 DB ACQUIRED WITH THE BIOMETRIKA SENSOR

	Best feature subsets - LivDet09 DB (Biometrika)		
	Full Reference	**No Reference**	**HTER**
Best-5	SPE, TCD, GPE	BIQI, JQI	54.7
Best-10	PSNR, AD, SPE, TED, TCD, GPE, RRED	BIQI, JQI, NIQE	45.1
Best-15	PSNR, AD, MD, RAMD, SPE, TED, TCD, GPE, SSIM, VIF, RRED	BIQI, JQI, NIQE, HLFI	31.7
Best-overall	All	All	12.8

TABLE XI
BEST PERFORMING FEATURE SUBSETS OF DIMENSIONS 5, 10, 15 AND BEST-OVERALL, FOUND USING THE SFFS ALGORITHM ACCORDING TO THE HTER ON THE TEST SET OF THE REPLAY-ATTACK DB FOR THE GRANDTEST PROTOCOL

	Best feature subsets - REPLAY-ATTACK DB (grandtest)		
	Full Reference	**No Reference**	**HTER**
Best-5	NXC, RAMD, MAS, SPE, RRED		53.5
Best-10	MSE, AD, SC, NXC, MD, RAMD, MAS, SME, SPE		48.9
Best-15	MSE, PSNR, AD, SC, NXC, MD, SNR, RAMD, MAMS, SME, SPE, TCD, GME, VIF	NIQE	38.3
Best-overall	All	All	15.2

sections are given in Tables IX–XI. Several observations may be extracted from these results:

- The most remarkable finding is that the whole group of 25 quality measures is consistently selected as the best performing feature set for all the considered scenarios and traits, showing the high complementarity of the proposed metrics for the biometric security task studied in the work.
- The first observation implies that other quality-related features could still be added to the proposed set in order to further improve its overall performance (until, eventually, adding new features starts decreasing its detection rates).
- For all cases, the best performing 5-feature and even 10-feature subsets present around a 50% HTER, which reinforces the idea that the competitive performance of the system does not rely on the high discriminative power of certain specific features but on the diversity and complementarity of the whole set.

V. CONCLUSION

The study of the vulnerabilities of biometric systems against different types of attacks has been a very active field of research in recent years [1]. This interest has lead to big advances in the field of security-enhancing technologies for biometric-based applications. However, in spite of this noticeable improvement, the development of efficient protection methods against known threats has proven to be a challenging task.

Simple visual inspection of an image of a real biometric trait and a fake sample of the same trait shows that the two images can be very similar and even the human eye may find it difficult to make a distinction between them after a short inspection. Yet, some disparities between the real and fake images may become evident once the images are translated into a proper feature space. These differences come from the fact that biometric traits, as 3D objects, have their own optical qualities (absorption, reflection, scattering, refraction), which other materials (paper, gelatin, electronic display) or synthetically produced samples do not possess. Furthermore, biometric sensors are designed to provide good quality samples when they interact, in a normal operation environment, with a real 3D trait. If this scenario is changed, or if the trait presented to the scanner is an unexpected fake artifact (2D, different material, etc.), the characteristics of the captured image may significantly vary.

In this context, it is reasonable to assume that the image quality properties of real accesses and fraudulent attacks will be different. Following this "*quality-difference*" hypothesis, in the present research work we have explored the potential of *general* image quality assessment as a protection tool against different biometric attacks (with special attention to spoofing).

For this purpose we have considered a feature space of 25 complementary image quality measures which we have combined with simple classifiers to detect real and fake access attempts. The novel protection method has been evaluated on three largely deployed biometric modalities such as the iris, the fingerprint and 2D face, using publicly available databases with well defined associated protocols. This way, the results

are reproducible and may be fairly compared with other future analogue solutions.

Several conclusions may be extracted from the evaluation results presented in the experimental sections of the article: i) The proposed method is able to consistently perform at a high level for different biometric traits ("multi-biometric"); ii) The proposed method is able to adapt to different types of attacks providing for all of them a high level of protection ("multi-attack"); iii) The proposed method is able to generalize well to different databases, acquisition conditions and attack scenarios; iv) The error rates achieved by the proposed protection scheme are in many cases lower than those reported by other trait-specific state-of-the-art anti-spoofing systems which have been tested in the framework of different independent competitions; and v) in addition to its very competitive performance, and to its "multi-biometric" and "multi-attack" characteristics, the proposed method presents some other very attractive features such as: it is simple, fast, non-intrusive, user-friendly and cheap, all of them very desirable properties in a practical protection system.

All the previous results validate the "quality-difference" hypothesis formulated in Section II: "*It is expected that a fake image captured in an attack attempt will have different quality than a real sample acquired in the normal operation scenario for which the sensor was designed.*"

In this context, the present work has made several contributions to the state-of-the-art in the field of biometric security, in particular: i) it has shown the high potential of image quality assessment for securing biometric systems against a variety of attacks; ii) proposal and validation of a new biometric protection method; iii) reproducible evaluation on multiple biometric traits based on publicly available databases; iv) comparative results with other previously proposed protection solutions.

The present research also opens new possibilities for future work, including: i) extension of the considered 25-feature set with new image quality measures; ii) further evaluation on other image-based modalities (e.g., palmprint, hand geometry, vein); iii) inclusion of temporal information for those cases in which it is available (e.g., systems working with face videos); iv) use of video quality measures for video attacks (e.g., illegal access attempts considered in the REPLAY-ATTACK DB); v) analysis of the features individual relevance.

REFERENCES

[1] S. Prabhakar, S. Pankanti, and A. K. Jain, "Biometric recognition: Security and privacy concerns," *IEEE Security Privacy*, vol. 1, no. 2, pp. 33–42, Mar./Apr. 2003.

[2] T. Matsumoto, "Artificial irises: Importance of vulnerability analysis," in *Proc. AWB*, 2004.

[3] J. Galbally, C. McCool, J. Fierrez, S. Marcel, and J. Ortega-Garcia, "On the vulnerability of face verification systems to hill-climbing attacks," *Pattern Recognit.*, vol. 43, no. 3, pp. 1027–1038, 2010.

[4] A. K. Jain, K. Nandakumar, and A. Nagar, "Biometric template security," *EURASIP J. Adv. Signal Process.*, vol. 2008, pp. 113–129, Jan. 2008.

[5] J. Galbally, F. Alonso-Fernandez, J. Fierrez, and J. Ortega-Garcia, "A high performance fingerprint liveness detection method based on quality related features," *Future Generat. Comput. Syst.*, vol. 28, no. 1, pp. 311–321, 2012.

[6] K. A. Nixon, V. Aimale, and R. K. Rowe, "Spoof detection schemes," *Handbook of Biometrics*. New York, NY, USA: Springer-Verlag, 2008, pp. 403–423.

[7] *ISO/IEC 19792:2009, Information Technology—Security Techniques—Security Evaluation of Biometrics*, ISO/IEC Standard 19792, 2009.

[8] *Biometric Evaluation Methodology. v1.0*, Common Criteria, 2002.

[9] K. Bowyer, T. Boult, A. Kumar, and P. Flynn, *Proceedings of the IEEE Int. Joint Conf. on Biometrics*. Piscataway, NJ, USA: IEEE Press, 2011.

[10] G. L. Marcialis, A. Lewicke, B. Tan, P. Coli, D. Grimberg, A. Congiu, *et al.*, "First international fingerprint liveness detection competition—LivDet 2009," in *Proc. IAPR ICIAP*, Springer LNCS-5716. 2009, pp. 12–23.

[11] M. M. Chakka, A. Anjos, S. Marcel, R. Tronci, B. Muntoni, G. Fadda, *et al.*, "Competition on countermeasures to 2D facial spoofing attacks," in *Proc. IEEE IJCB*, Oct. 2011, pp. 1–6.

[12] J. Galbally, J. Fierrez, F. Alonso-Fernandez, and M. Martinez-Diaz, "Evaluation of direct attacks to fingerprint verification systems," *J. Telecommun. Syst.*, vol. 47, nos. 3–4, pp. 243–254, 2011.

[13] A. Anjos and S. Marcel, "Counter-measures to photo attacks in face recognition: A public database and a baseline," in *Proc. IEEE IJCB*, Oct. 2011, pp. 1–7.

[14] Biometrics Institute, London, U.K. (2011). *Biometric Vulnerability Assessment Expert Group* [Online]. Available: http://www.biometricsinstitute.org/pages/biometric-vulnerability-assessment-expert-group-bvaeg.html

[15] (2012). *BEAT: Biometrices Evaluation and Testing* [Online]. Available: http://www.beat-eu.org/

[16] (2010). *Trusted Biometrics Under Spoofing Attacks (TABULA RASA)* [Online]. Available: http://www.tabularasa-euproject.org/

[17] J. Galbally, R. Cappelli, A. Lumini, G. G. de Rivera, D. Maltoni, J. Fierrez, *et al.*, "An evaluation of direct and indirect attacks using fake fingers generated from ISO templates," *Pattern Recognit. Lett.*, vol. 31, no. 8, pp. 725–732, 2010.

[18] J. Hennebert, R. Loeffel, A. Humm, and R. Ingold, "A new forgery scenario based on regaining dynamics of signature," in *Proc. IAPR ICB*, vol. Springer LNCS-4642. 2007, pp. 366–375.

[19] A. Hadid, M. Ghahramani, V. Kellokumpu, M. Pietikainen, J. Bustard, and M. Nixon, "Can gait biometrics be spoofed?" in *Proc. IAPR ICPR*, 2012, pp. 3280–3283.

[20] Z. Akhtar, G. Fumera, G. L. Marcialis, and F. Roli, "Evaluation of serial and parallel multibiometric systems under spoofing attacks," in *Proc. IEEE 5th Int. Conf. BTAS*, Sep. 2012, pp. 283–288.

[21] D. Maltoni, D. Maio, A. Jain, and S. Prabhakar, *Handbook of Fingerprint Recognition*. New York, NY, USA: Springer-Verlag, 2009.

[22] R. Cappelli, D. Maio, A. Lumini, and D. Maltoni, "Fingerprint image reconstruction from standard templates," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 9, pp. 1489–1503, Sep. 2007.

[23] S. Shah and A. Ross, "Generating synthetic irises by feature agglomeration," in *Proc. IEEE ICIP*, Oct. 2006, pp. 317–320.

[24] S. Bayram, I. Avcibas, B. Sankur, and N. Memon, "Image manipulation detection," *J. Electron. Imag.*, vol. 15, no. 4, pp. 041102-1–041102-17, 2006.

[25] M. C. Stamm and K. J. R. Liu, "Forensic detection of image manipulation using statistical intrinsic fingerprints," *IEEE Trans. Inf. Forensics Security*, vol. 5, no. 3, pp. 492–496, Sep. 2010.

[26] I. Avcibas, N. Memon, and B. Sankur, "Steganalysis using image quality metrics," *IEEE Trans. Image Process.*, vol. 12, no. 2, pp. 221–229, Feb. 2003.

[27] S. Lyu and H. Farid, "Steganalysis using higher-order image statistics," *IEEE Trans. Inf. Forensics Security*, vol. 1, no. 1, pp. 111–119, Mar. 2006.

[28] J. Galbally, J. Ortiz-Lopez, J. Fierrez, and J. Ortega-Garcia, "Iris liveness detection based on quality related features," in *Proc. 5th IAPR ICB*, Mar./Apr. 2012, pp. 271–276.

[29] I. Avcibas, B. Sankur, and K. Sayood, "Statistical evaluation of image quality measures," *J. Electron. Imag.*, vol. 11, no. 2, pp. 206–223, 2002.

[30] Q. Huynh-Thu and M. Ghanbari, "Scope of validity of PSNR in image/video quality assessment," *Electron. Lett.*, vol. 44, no. 13, pp. 800–801, 2008.

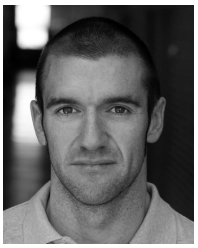
[31] S. Yao, W. Lin, E. Ong, and Z. Lu, "Contrast signal-to-noise ratio for image quality assessment," in *Proc. IEEE ICIP*, Sep. 2005, pp. 397–400.

[32] A. M. Eskicioglu and P. S. Fisher, "Image quality measures and their performance," *IEEE Trans. Commun.*, vol. 43, no. 12, pp. 2959–2965, Dec. 1995.

[33] M. G. Martini, C. T. Hewage, and B. Villarini, "Image quality assessment based on edge preservation," *Signal Process., Image Commun.*, vol. 27, no. 8, pp. 875–882, 2012.

[34] N. B. Nill and B. Bouzas, "Objective image quality measure derived from digital image power spectra," *Opt. Eng.*, vol. 31, no. 4, pp. 813–825, 1992.

[35] A. Liu, W. Lin, and M. Narwaria, "Image quality assessment based on gradient similarity," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1500–1511, Apr. 2012.
[36] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
[37] (2012). *LIVE* [Online]. Available: http://live.ece.utexas.edu/research/Quality/index.htm
[38] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Trans. Image Process.*, vol. 15, no. 2, pp. 430–444, Feb. 2006.
[39] R. Soundararajan and A. C. Bovik, "RRED indices: Reduced reference entropic differencing for image quality assessment," *IEEE Trans. Image Process.*, vol. 21, no. 2, pp. 517–526, Feb. 2012.
[40] Z. Wang, H. R. Sheikh, and A. C. Bovik, "No-reference perceptual quality assessment of JPEG compressed images," in *Proc. IEEE ICIP*, Sep. 2002, pp. 477–480.
[41] X. Zhu and P. Milanfar, "A no-reference sharpness metric sensitive to blur and noise," in *Proc. Int. Workshop Qual. Multimedia Exper.*, 2009, pp. 64–69.
[42] A. K. Moorthy and A. C. Bovik, "A two-step framework for constructing blind image quality indices," *IEEE Signal Process. Lett.*, vol. 17, no. 5, pp. 513–516, May 2010.
[43] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a 'completely blind' image quality analyzer," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 209–212, Mar. 2013.
[44] T. Hastie, R. Tibshirani, and J. Friedman., *The Elements of Statistical Learning*. New York, NY, USA: Springer-Verlag, 2001.
[45] Z. Wang and A. C. Bovik, "Mean squared error: Love it or leave it? A new look at signal fidelity measures," *IEEE Signal Process. Mag.*, vol. 26, no. 1, pp. 98–117, Jan. 2009.
[46] B. Girod, "What's wrong with mean-squared error?" in *Digital Images and Human Vision*. Cambridge, MA, USA: MIT Press, 1993, pp. 207–220.
[47] A. M. Pons, J. Malo, J. M. Artigas, and P. Capilla, "Image quality metric based on multidimensional contrast perception models," *Displays J.*, vol. 20, no. 2, pp. 93–110, 1999.
[48] C. Harris and M. Stephens, "A combined corner and edge detector," in *Proc. AVC*, 1988, pp. 147–151.
[49] J. Zhu and N. Wang, "Image quality assessment by visual gradient similarity," *IEEE Trans. Image Process.*, vol. 21, no. 3, pp. 919–933, Mar. 2012.
[50] D. Brunet, E. R. Vrscay, and Z. Wang, "On the mathematical properties of the structural similarity index," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1488–1499, Apr. 2012.
[51] M. A. Saad, A. C. Bovik, and C. Charrier, "Blind image quality assessment: A natural scene statistics approach in the DCT domain," *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3339–3352, Aug. 2012.
[52] J. Fierrez, J. Ortega-Garcia, D. Torre-Toledano, and J. Gonzalez-Rodriguez, "BioSec baseline corpus: A multimodal biometric database," *Pattern Recognit.*, vol. 40, no. 4, pp. 1389–1392, 2007.
[53] E. Marasco and C. Sansone, "Combining perspiration- and morphology-based static features for fingerprint liveness detection," *Pattern Recognit. Lett.*, vol. 33, no. 9, pp. 1148–1156, 2012.
[54] Y. S. Moon, J. S. Chen, K. C. Chan, K. So, and K. C. Woo, "Wavelet based fingerprint liveness detection," *Electron. Lett.*, vol. 41, no. 20, pp. 1112–1113, 2005.
[55] S. Nikam and S. Argawal, "Curvelet-based fingerprint anti-spoofing," *Signal, Image Video Process.*, vol. 4, no. 1, pp. 75–87, 2010.
[56] A. Abhyankar and S. Schuckers, "Fingerprint liveness detection using local ridge frequencies and multiresolution texture analysis techniques," in *Proc. IEEE ICIP*, Oct. 2006, pp. 321–324.
[57] I. Chingovska, A. Anjos, and S. Marcel, "On the effectiveness of local binary patterns in face anti-spoofing," in *Proc. IEEE Int. Conf. Biometr. Special Interest Group*, Sep. 2012, pp. 1–7.
[58] J. Maatta, A. Hadid, and M. Pietikainen, "Face spoofing detection from single images using micro-texture analysis," in *Proc. IEEE IJCB*, Oct. 2011, pp. 1–7.
[59] P. Pudil, J. Novovicova, and J. Kittler, "Flotating search methods in feature selection," *Pattern Recognit. Lett.*, vol. 15, no. 11, pp. 1119–1125, 1994.
[60] A. K. Jain and D. Zongker, "Feature selection: Evaluation, application, and small sample performance," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 2, pp. 153–158, Feb. 1997.



Javier Galbally received the M.Sc. degree in electrical engineering in 2005 from Universidad de Cantabria, Spain, and the Ph.D. degree in electrical engineering in 2009 from Universidad Autonoma de Madrid, Spain, where he worked as an Assistant Professor until 2012. In 2013, he moved to the Joint Research Centre of the European Commission where he is currently working as a Post-Doctoral Researcher. He has carried out different research internships at worldwide leading groups in biometric recognition such as BioLab from Universit di Bologna Italy, IDIAP Research Institute in Switzerland, the Scribens Laboratory at the Cole Polytechnique de Montral in Canada, or the Integrated Pattern Recognition and Biometrics Laboratory (i-PRoBe) at the West Virginia University in the USA. His research interests are mainly focused on the security evaluation of biometric systems, but also include pattern and biometric recognition, synthetic generation of biometric traits and inverse biometrics. He is actively involved in European projects focused on biometrics and is the recipient of a number of distinctions, including: IBM Best Student Paper Award at ICPR 2008, finalist of the EBF European Biometric Research Award 2009, and the Best Ph.D. Thesis Award by the Universidad Autonoma de Madrid 2010.

Sébastien Marcel received the Ph.D. degree in signal processing from Université de Rennes I in France in 2000 at CNET, the research center of France Telecom (now Orange Labs). He is a Senior Research Scientist at the Idiap Research Institute (CH) where he leads the Biometrics group and conducts research on multi-modal biometrics including face recognition, speaker recognition, vascular recognition, as well as spoofing and anti-spoofing. In 2010, he was appointed as a Visiting Professor at the University of Cagliari (IT) where he taught a series of lectures in face recognition. In 2013, he was a Lecturer at the École Polytechnique Fédérale de Lausanne where he taught "Fundamentals in Statistical Pattern Recognition." He was the main organizer of a number of special scientific events or competitive evaluations all involving biometrics, and serves as an Associate Editor for the IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY. He is also the principal investigator of international research projects including MOBIO, TABULA RASA, and BEAT. Finally, he leads the development of the Bob, the signal-processing, and machine learning toolbox.

Julian Fierrez received the M.Sc. and Ph.D. degrees in telecommunications engineering from Universidad Politecnica de Madrid, Madrid, Spain, in 2001 and 2006, respectively. Since 2002, he has been with the Biometric Recognition Group, first at Universidad Politecnica de Madrid, and since 2004 at Universidad Autonoma de Madrid, where he is currently an Associate Professor. From 2007 to 2009, he was a Visiting Researcher at Michigan State University in USA under a Marie Curie fellowship. His research interests and areas of expertise include signal and image processing, pattern recognition, and biometrics, with emphasis on signature and fingerprint verification, multi-biometrics, biometric databases, and system security. He is actively involved in European projects focused on biometrics, and is the recipient of a number of distinctions for his research, including: best Ph.D. in computer vision and pattern recognition from 2005 to 2007 by the IAPR Spanish liaison, the Motorola Best Student Paper at ICB 2006, the EBF European Biometric Industry Award 2006, the IBM best student paper at ICPR 2008, and the EURASIP Best Ph.D. Award 2012.