**REVIEW**  **Open Access**

# Biometric quality: a review of fingerprint, iris, and face

Samarth Bharadwaj, Mayank Vatsa[*] and Richa Singh

## Abstract

Biometric systems encounter variability in data that influence capture, treatment, and u-sage of a biometric sample. It is imperative to first analyze the data and incorporate this understanding within the recognition system, making assessment of *biometric quality* an important aspect of biometrics. Though several interpretations and definitions of quality exist, sometimes of a conflicting nature, a holistic definition of quality is indistinct. This paper presents a survey of different concepts and interpretations of biometric quality so that a clear picture of the current state and future directions can be presented. Several factors that cause different types of degradations of biometric samples, including image features that attribute to the effects of these degradations, are discussed. Evaluation schemes are presented to test the performance of quality metrics for various applications. A survey of the features, strengths, and limitations of existing quality assessment techniques in fingerprint, iris, and face biometric are also presented. Finally, a representative set of quality metrics from these three modalities are evaluated on a multimodal database consisting of 2D images, to understand their behavior with respect to match scores obtained from the state-of-the-art recognition systems. The analysis of the characteristic function of quality and match scores shows that a careful selection of complimentary set of quality metrics can provide more benefit to various applications of biometric quality.

**Keywords:** Biometric quality assessment; Face; Fingerprint; Iris

## 1 Introduction

Biometrics, as an integral component in identification science, is being utilized in large-scale biometrics deployments such as the US Visitor and Immigration Status Indicator Technology (VISIT), UK Iris Recognition Immigration System (IRIS) project, UAE iris-based airport security system, and India's Aadhaar project. These far-reaching and inclusive delivery systems not only provide a platform to assist and enhance civilization but also offer new research directions. An important research challenge among them is the measurement of *quality* of a biometric sample. Biometric systems, like other applications of pattern recognition and machine learning, are affected by the quality of input data. Therefore, it is important to quantitatively evaluate the quality of a sample that is indicative of its ability to function as a biometric. In our opinion, quality of a biometric is beyond measuring the quality of the image itself. While a sample's quality is susceptible to

irregularities during capture or storage, it may also have low quality by its very nature. For instance, as shown in Figure 1, an input biometric sample may possess a wide range of *quality*.

Quality assessment (QA) of an *image* measures its degradation during acquisition, compression, transmission, processing, and reproduction. Several QA algorithms exist in image processing literature, which pursue different philosophies, performance, and applications. A majority of these methods are motivated towards accurate *perceptual* image quality i.e., quality as perceived by the sophisticated human visual system (HVS). These approaches require an in depth understanding of the anatomy and psychophysical functioning of the human cognitive system. Several perceptual quality metrics are surveyed by Wang and Bovik [1] and Lin and Kuo [2]. On the other hand, the quality of a biometric sample is interpreted differently throughout literature [3-10]. A summary of these interpretations is provided in Table 1. In general, biometric quality is defined as an indicator of the *usefulness* of the biometric sample for recognition,

*Correspondence: mayank@iiitd.ac.in
IIIT-D, New Delhi 110020, India

**Figure 1 Variation in quality.** A biometric system may encounter samples of a wide range of quality (Images from MBGC database). Effective quality assessment metrics that are indicative of these variations are therefore essential to an automated biometric system.

as illustrated in Figure 2. It is well established that environmental distortions such as noise, blur, and adverse illumination, affect the performance of state-of-the-art recognition algorithms. However, existing image quality metrics that measure such degradations encode only a part of the information that can measure the overall quality of a biometric sample. Hence, a clear distinction must be made between perceptual image quality assessment (PIQA) and biometric quality assessment (BQA). PIQA research attempts to understand why human subjects prefer some images to others [11,12]. The task is complex and involves multiple disciplines, including an understanding of the HVS. On the other hand, BQA provides an initial estimate of the ability of a sample to

**Table 1 Different interpretations of quality in biometrics from literature**

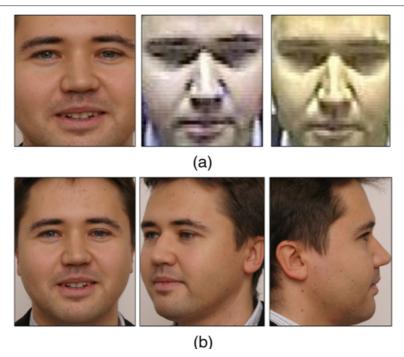| Reference | Modality | Interpretation of quality in biometrics |
|---|---|---|
| Chen et al. [3] | Fingerprint | A global measure of the strength of ridges |
| Grother and Tabassi [4] | Fingerprint | Suitability for automatic matching |
| Youmaran and Adler [5] | Face | The decrease in uncertainty of identity due to a given sample |
| Kryszczuk et al. [6] | Face | Conditionally relevant class predictors |
| Beveridge et al. [7] | Face | A measurable and actionable predictor of performance |
| ISO/IEC standards [13] | Face | Biometric data that adheres to best capture practices |
| Kalka et al. [8] | Iris | The measurement of various degradations known to affect iris recognition |
| Kumar and Zhang [9] | Knuckles | Confidence of generating reliable matching scores from the user templates |
| Poh and Kittler [10] | General framework | Degree of *extractability* of recognition features |
| BioAPI [14] | General framework | Biometric data that provides good performance for the intended purpose |

**Figure 2 Image quality vs biometric quality.** While the images (obtained from SCface database) in **(a)** are of poor *image* quality, the images in **(b)** may have lower *biometric quality*.

function as a biometric. We therefore define biometric quality as

*Quality of a biometric sample is a measure of its efficiency in aiding recognition of an individual, ideally, irrespective of the recognition system in use.*

In literature, quality assessment metrics are widely used in the formulation of biometric techniques. As illustrated in Figure 3, quality metrics can be used at various stages of the recognition pipeline to improve performance and usability of biometrics in challenging conditions. The application of quality metrics can be during both enrolment and recognition phases. Since enrolment phase is the best opportunity to re-capture a sample to maintain the overall quality of the gallery set, the quality of input sample is an important consideration. On the other hand, the quality of a probe sample during recognition phase is utilized in different methodologies to improve the recognition performance. Some important applications and evaluation metrics of quality assessment techniques in biometric systems are described here.
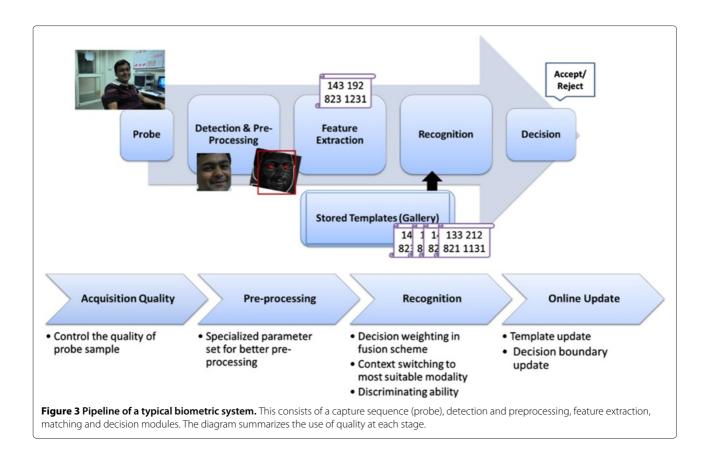
### 1.1 Quality assessment during enrolment
Quality feedback during enrolment is critical in collecting high-quality gallery data. It is common, especially in large-scale biometric systems, to have a supervised enrolment process as in the case of the India's Aadhaar project. An active quality feedback enables the collection officer

to evaluate and maintain quality standards during the enrolment process [15]. It can also be a performance measure for the collection apparatus and procedure employed for data capture [16]. Aggregated quality may also be used to create timeline along with historical or geographical meta-data for other analysis.

### 1.2 Quality assessment during recognition
Quality assessment and feedback during verification can help mitigate false alarms. A verification system can choose not to perform matching if the quality score is below a threshold, depending on the computation time of matching and the overhead of re-acquisition of data. Most modern fingerprint and iris sensors are now bundled with active quality-control mechanisms. Identification is inherently a computationally expensive process, hence, it is a good idea to use quality assessment (computationally less expensive) to improve system usability. For example, quality can be used in negative identification, where it is in the interest of the subject to provide a poor quality sample. The subject may then be persuaded to provide better quality samples without having to wait for misleading and incorrect identification result from the system. Further, in the recognition pipeline, quality is used at different stages/levels of a biometric system:
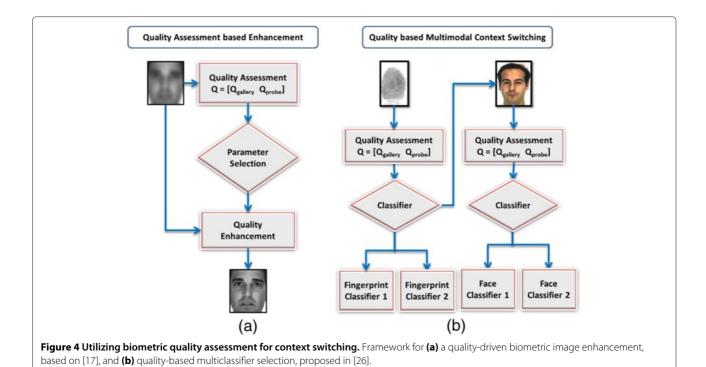
- *Preprocessing* A probe sample may contain degradations due to environmental conditions, incorrect use

**Figure 3 Pipeline of a typical biometric system.** This consists of a capture sequence (probe), detection and preprocessing, feature extraction, matching and decision modules. The diagram summarizes the use of quality at each stage.

of sensors, or transmission error. The performance of recognition systems severely depletes in such cases. Image restoration techniques can improve image quality, provided that the correct parameters are used [17]. Quality-assessment-based selection of parameters for image enhancement shows marked improvement in the recognition performance of the resultant biometric sample, when compared to using generic parameters. Also, biometric images obtained from different uncorrelated or orthogonal bands of the spectrum can provide different amounts of information, as demonstrated by Vatsa et al. [18] with the face and iris [19]. An illustration of a quality-assessment-based image enhancement framework is presented in Figure 4a.

- *Recognition* Poh et al. [20], Kryszczuk et al. [6,21], and Poh and Kittler [10] have shown that while quality assessment scores are used for perceptual understanding of the sample or performance prediction, they also possess some discriminating ability. Their experiments show that incorporating quality assessment values as additional features can improve the recognition performance. Similarly, quality-augmented product of likelihood ratio fusion scheme has shown to improve the performance [22]. Grother and Tabassi [4] have studied the relationship between quality and recognition accuracy in fingerprints and suggested

that quality scores can help in predicting the similarity scores.

- *Context switching* Context-switching frameworks dynamically select classifiers and/or distance metrics based on the quality of the sample. A serial framework for quality-based context switching is illustrated in Figure 4b. Recent literature [23-27] demonstrates the advantages of context switching of a biometric recognition pipeline based on the feedback from quality assessment algorithms. Vatsa et al. [23] propose a parallel context switching framework that uses energy in sub-bands, activity level, and pose angle for selecting the appropriate uni-modal classifier or fusion algorithm. Sellahewa and Jassim [25] present a simple thresholding-based adaptive fusion approach on illumination estimation from first-order statistics. Bhatt et al. [26] propose a serial framework of quality-based classifier selection using both image quality and biometric-specific quality metrics. Alonso-Fernandez et al. [28] present a quality-based context switching framework to improve sensor inter-operability in fingerprint biometric. Poh and Kittler [10] propose a unified framework for fusion of biometric classifiers at match score level by incorporating quality measures. This framework is based on a Bayesian perspective and can be used both as a generative and discriminative classifier.

**Figure 4 Utilizing biometric quality assessment for context switching.** Framework for **(a)** a quality-driven biometric image enhancement, based on [17], and **(b)** quality-based multiclassifier selection, proposed in [26].

- *Decision* Quality assessment scores can also aid decision-level fusion. By providing quality priors to maximize selective or cumulative combination of decision, the notion of strong or weak classifiers can become subject specific. Hence, the primary concern of using decision-level fusion schemes, discussed in [29], can also be eliminated. For rank-level fusion, Abaza and Ross [30] propose a weighted variant of boda count rank aggregation technique using quality assessment scores. An empirical evaluation [31] shows the applicability of nonlinear rank-level fusion as well, particularly in palmprint biometrics.
- *Sample update or replacement* Another interesting application of quality scores is in the replacement or addition of a confirmed probe sample to the gallery based on its quality. While this procedure has the risk of gallery contamination, it can elevate important concerns of temporal variations of biometric data, such as facial aging.
- *Decision update* Researchers are exploring the use of *online or incremental learning* approaches to improve the decision boundary of the classifiers even in deployment phase [32,33]. A major concern in such systems is to select suitable samples to learn incrementally. For instance, modifying decision boundary based on all the incoming samples may be computationally expensive. Further, online learning on outlier samples can adversely affect the system performance. One area of focus is towards using quality of the sample to determine whether the sample is suitable for classifier update.

The applications show that active involvement of quality assessment beyond the capture stage of the biometric pipeline encourages the formulation of complex and accurate biometric quality assessment. Hence, BQA is an important aspect of biometrics research that can lead towards robust and user-friendly biometric recognition systems. The aim of this survey paper is to collate different directions of quality assessment in biometrics towards a unified framework with respect to three primary modalities, *viz.*, iris, fingerprint, and face. Section 2 discusses various factors and degradations that influence quality in biometrics. Image features used in quality assessment to evaluate the effect of those degradations are also presented along with a general quality framework. Section 3 presents a review of recent literature in biometric quality assessment pertaining to fingerprint, iris, and face modalities. Evaluation protocols inspired by different applications that are indicative of the metric's performance are also presented. Section 4 presents an experimental analysis of different quality metrics and corresponding relevance to match scores providing a better understanding of the behavior of biometric quality metrics with respect to matching performance. In this experiment it is observed that in place of using an arbitrary set of quality metrics, a careful selection with respect of match scores can provide additional benefits to biometric systems. Finally, we also

discuss the salient finding from our experimental evaluations and literature as well as future scope and directions. Additionally, a brief overview of perceptual image quality assessment is presented in Appendix 1 and quality metric standards prevalent in biometrics literature are discussed in Appendix 2.

## 2 Biometric quality: factors, degradations, and features

An observer's perspective in assessing quality is an important aspect of QA [34]. For instance, the perception of an image can change with respect to the subject, the photographer, or by the interpretation of some third party. Similarly, the quality of a biometric sample can depend on acquisition system and the technology used for matching. For meaningful prediction of quality, the ideal pursuit is towards a quality metric that is consistent across any type of degradation and matching techniques. However, pragmatic solutions utilize some understanding of the degradation and matching techniques in their formulation.

This section describes the cause and effects of factors that influence quality of biometric samples. Further, the image features that are typically used in automatic image analysis of biometric samples are studied. Finally, a general framework for quality assessment in biometrics is presented.

### 2.1 Factors that influence biometric quality

It is important to appreciate the effects of various factors that affect quality to develop better assessment algorithms. While some factors are unavoidable, others may be inherent limitations of the biometric itself. These factors are either user traits or interactions between user and sensors:

- *User traits* Some important factors that influence the quality of a biometric sample during capture process can be classified as *behavioral* and *physiological* traits of the human users [35]. Behavioral traits may include motivation levels, cooperation, and fears. Physiological traits include facial hair or sensitivity to light. While some behaviors of users can be restricted, it is at the cost of usability and increased inconvenience. Further, unavoidable factors such as age, social customs, gender, and injuries can impair the quality of the captured sample. For instance, fingerprints obtained from older age groups is of lower inherent biometric quality (due to worn ridges) when using different commercial fingerprint systems [36].
- *User-sensor interaction and operational constraints* The second important factor that influences the quality of contact capture (closed/near field of view) based biometrics, such as fingerprints, palmprints, iris, and

retinal, is the interaction between users and sensors. The *usability* of the sensor is crucial to quality. Sensors with active user feedback that are portable and easy to use ensure good user-sensor interaction, resulting in better quality captures. However, environmental factors such as temperate, humidity, and background influence this interaction, adversely affecting the quality of a biometrics. Other factors that affect the quality of a biometric sample are operational constraints particularly in the use and maintenance of (touch-based) sensors and training of handlers. For instance, Aadhaar project uses different types of sensors and operational procedures in accordance with the climatic conditions of different regions of India. In such cases, controlling conditions, policies, and guidelines during operation play a significant role.

Table 2 presents some possible causes of each of the aforementioned factors. These factors have varying degrees of adversarial effect on the performance of a captured biometric sample. Uncooperative users, such as in criminal cases, pose an additional challenge to effective data collection processes. It is worthwhile to understand the different degradation processes that result from these factors.

### 2.2 Degradations in biometric images

In order to better understand quality assessment in biometrics, it might be useful to closely inspect the different artifacts that commonly manifest in biometric images. As illustrated in Figure 5, these degradations are either virtues of an image or of the biometric modality itself.

#### 2.2.1 Image-based degradations

Image degradations are manifested by the property of capture devices and conditions, irrespective of the biometric being captured:

- Blurring: Image blurring is a common phenomenon that occurs due to incorrect focus (object is outside the

**Table 2 Various behavioral, environmental, and operational factors that effect quality of biometric sample**

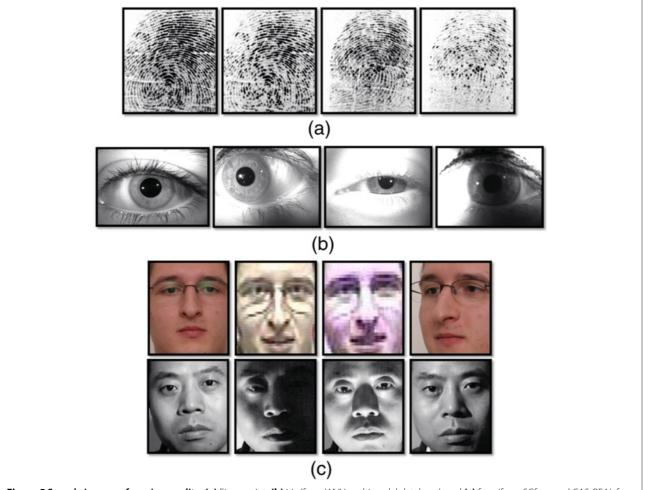| Factors | Possible causes |
| --- | --- |
| User traits | Tiredness, distractions, motivation, cooperation, fear, makeup, appearance, facial hair, clothes, or hats |
| User-sensor interactions | Indoor/outdoor, background, temperature, humidity, illumination, and ambient noise |
| Operational | Familiarity, quality feedback, sensor cleaning, supervising operator, and time between acquisition |

**Figure 5 Sample images of varying quality. (a)** Fingerprint, **(b)** iris (from WVU multimodal database), and **(c)** face (from SCface and CAS-PEAL face databases) illustrating the wide range of quality that a biometric system can encounter with different image and biometric specific degradations.

depth of field), motion, or certain environmental factors. Blurring effects edge information, which is vital to biometric recognition, particulary the minute edges of iris patterns.

- Illumination: Uniform lighting is essential for the capture of a good quality biometric. Conversely, adversely directed lighting drastically affects the performance of iris and face.
- Noise/Compression: An image may contain noise due to environmental factors, incorrect use of sensors, and transmission error. Noise contamination drastically affects the performance of recognition systems. Depending on the compression levels, various image encoding techniques produce artifacts such as blockiness and ringing effect.
- Optical distortions: Nonconformity to rectilinear projection causes distortion in the captured image. Such distortions may occur due to various environmental factors or due to the functioning of sensors. Further, difference in the sensor models also results in different

distortion profile, degrading recognition performance [37].

The aforementioned degradations usually occur due to the limitation of sensor technology or environmental conditions. As the constraints on user during capture are relaxed, the impact of these factors on the performance of systems increases drastically. Therefore, estimation and analysis of these factors are critical for building robust and nonintrusive biometric systems.

### 2.2.2 Biometric-modality-specific degradations

Biometric degradations occur as a consequence of the nature of the biometric modality being captured. For example, face and iris biometrics have multiple degrees of motion and hence pose angle at which a captured image can affect quality. Murphy-Chutorian and Trivedi [38] survey several head-pose estimation techniques. Fingerprints exhibit pose variations in terms of fingerprint orientation that may result in a partial prints. Biometric

data from unconstrained environment is plagued with occlusion or missing information. Common causes in case of face include accessories and facial hair. Erroneous data can also arise from medical conditions, scars, or skin deformations (due to temperature or dryness).

Certain degradations may be difficult to measure, for example, the aesthetic changes of the face brought about by hair style or makeup. Beveridge et al. [7] introduce the notion of *measurable covariates*, a subset of different degradations that are easy to estimate from an image. Note that measurable covariates can be properties of the image (edge density measures) or of the subject (inter-eye distance). Further, properties such as region of interest, focus of camera, and also expression, glasses, and clothing that can be controlled to some extent (at the cost of usability), are termed as *actionable*. Nonactionable covariates include age, gender, and race. Accurate assessment of measurable and actionable covariates of biometrics must be the focus of quality assessment techniques. Current research primarily focuses on using image processing techniques to assess image features that indicate quality. These different *image features* are examined next.

### 2.3 Image-based features
The aforementioned degradations manifested in biometric samples can be assessed using image features that are computationally inexpensive to compute. Automatic QA is primarily addressed by analyzing spatial and temporal features that are indicative of the image content. Features that are used extensively in current literature can be broadly divided into four categories (as shown in Figure 6):

- *Orientation* features are obtained from edges in the image. In case of the iris and face, edge information is widely used as features for recognition. Blurring, illumination, and noise degrade edge information thereby affect performance. Hence, orientation information can provide a good indication of the quality of a biometric sample.
- *Power spectrum* is a temporal measure of the power of the image signal. This measure is an indication of the amount of information present in an image region. Hence, spectral energy is often computed for different image regions to obtain local assessment of quality.
- *Intensity statistics* are direct statistical evaluation of intensities of pixels in the image. Typically, a statistical measure such as Kurtosis or point spread function (PSF) estimation is used to estimate blurring or illumination degradation in the image. The measure can then be compared to the reference values obtained from ideal images to compute the extent of degradation.

- *Wavelet transform* provides both spatial and frequency understanding of the information content in each sub-band of the image. These are particularly suited to ascertain the presence of fine *micro edges* in the iris region and to obtain local analysis of quality in different regions of an image.

In addition to the four image features, *shape* of the segmentation boundary of the biometric content of the image can also provide useful information of the quality of the sample. For instance, the circularity and pixel density of an iris segmentation are important quality measures and widely used in literature. However, we assert that the same degradations that affect recognition can also affect the segmentation performance. Hence, the performance of shape as a quality feature deteriorates rapidly with non-ideal images. In cases where color imagery is used for capture, multichannel information are also leveraged for QA. It has been reported in literature that the discriminating power of certain channels supersedes others. Therefore, quality metrics for each channel may also be considered separately. Finally, several QA techniques use multiple features to form a composite quality score via (statistical) fusion; they are referred to as *combined features*. Nonimage features such as image header information (EXIF), or cues obtained from sensor, may also be used as features for quality assessment. However, the subjective nature of these features leads to poor generalization.

### 2.4 Naturality, fidelity, and utility in biometric quality
Different QA algorithms in literature have some underlying similarities in their philosophy/approach. It might be helpful to classify existing algorithms based on these underlying principles for a thorough understanding of the current state of research and limitations of literature. Several attempts have been made at this classification; Kalka et al. [8] classified iris quality assessment algorithms into global and local algorithms. Beveridge et al. [39] classified techniques based on the properties of different covariates. Inspired by the visual quality model of Yendrikhovskij [40] (illustrated in Figure 7), this research presents three aspects of quality assessment in biometrics:

1. Biometric naturality: *the degree of apparent match of the biometric image with an internal reference of goodness.* Most of the no-reference quality assessment algorithms measure perceptual image quality, indicating the naturalness of that image. These methods [1,2,41] are based on unexpected changes in intensities or ratio of information in various spatial/temporal bands, effects that stand out in visual inspection of quality. Such metrics are adept at encoding image level degradations, such as illumination, compression artifacts, noise, and blurring.
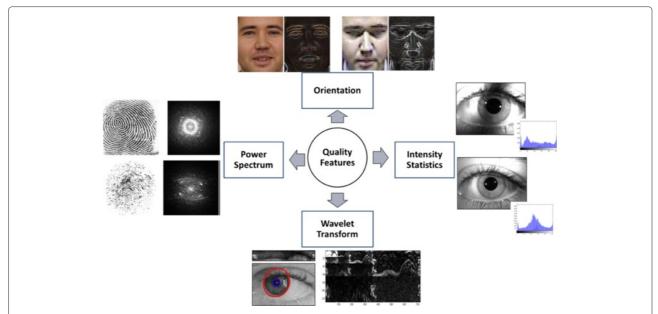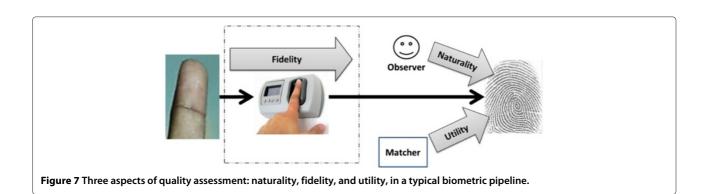
**Figure 6 Four image features are primarily used for estimating quality of biometric images.** Orientation, intensity statistics, power spectrum, and wavelet transform.

These metrics are computationally inexpensive and their performance is dependent on baseline parameters obtained from some knowledge of the intended application (internal reference of goodness).

2. Biometric fidelity: *the degree to which a biometric modality is correctly represented in the acquired image.* The quality or the extent to which the acquired image (from a sensor) successfully represents the biometric that is presented to a sensor is the measure of fidelity of a biometric sample. Measuring the fidelity is a challenging problem as there may not be additional information to verify the sample with respect to the source.

3. Biometric utility: *the degree of suitability of the sample for matching.* The utility of a biometric sample is based on its matching performance. While utility is surely dependent on the sample's naturalness and fidelity, it has been shown that (face) biometric samples of the same person captured in similar settings can exhibit marked difference in matching performance. Further, the information, while correctly captured, may be useless to the particular matcher. Hence, the utility of a biometric is often independent of the other two aspects of biometric quality.

Alanso-Fernandez et al. [42,43] also use similar nomenclature to describe quality assessment *viewpoints*, from which the authors conclude that for fingerprint biometrics, 'utility' is of primary focus. However, it is our assertion that in order to obtain a complete understanding of the quality of a biometric sample, all three dimensions, naturality, fidelity, and utility must be evaluated. This is more pertinent for iris and face biometric, where the features are not structured as compared to fingerprints.



**Figure 7 Three aspects of quality assessment: naturality, fidelity, and utility, in a typical biometric pipeline.**

## 3 Review: quality assessment in fingerprint, iris, and face

Several techniques have been proposed in literature to assess the quality of a biometric sample that is affected by aforementioned degradations. In this section, a literature review of quality assessment algorithms pertaining to three popular modalities, *viz.*, fingerprint, iris and face, are presented, along with the review of key techniques to evaluate quality assessment algorithms.

### 3.1 Fingerprint quality assessment

Poor quality fingerprint images can lead to incorrect or spurious feature (minutia) detection (illustrated in Figure 8) and thereby degrading the performance of a fingerprint recognition system. Quality assessment of fingerprint ridge quality is essential for proper functioning of the recognition system. These metrics are primarily used in fingerprint sensors with active quality feedback for rejecting poor quality samples. Fingerprint quality is also used to evaluate local unrecoverable regions of the fingerprint, as enhancement of these regions for ridge information may be counter-productive. Further, region-wise assessment may also be useful in adaptive feature importance weighting schemes. Most fingerprint quality assessment metrics compute image properties in local regions and pool these metrics to present a single quality score. A detailed review of some seminal techniques is presented here along with a summary in Table 3.

Lim et al. [48] present a local-feature-based quality metric which computes *orientation certainty level* (OCL), *ridge frequency*, *ridge thickness*, and *ridge-to-valley thickness ratio*. Shen et al. [49] use Gabor filters for quality assessment. Fingerprint image is tessellated into blocks, and Gabor filters with different orientations is applied on each block. For high-quality blocks, response from filters of some orientations is significantly higher than others, whereas for low-quality blocks, the difference in responses from the filters is generally low. The standard deviation of the responses thus indicates local quality for each block. The aggregated local quality is compared with scores from visual inspection. Similarly, Vatsa et al. [45] use redundant discrete wavelet transform (RDWT) to compute dominant ridge activity to measure fingerprint quality. The quality metric induced huge performance improvement when incorporated into a fingerprint feature level fusion framework on a large real-world database. Olsen et al. [50] also present a quality measure based on evaluating Gabor filter responses of a fingerprint image whose performance is more robust to its parameters.

In another approach, Chen et al. [3] measure the quality of ridge samples by energy spectral density concentration in particular frequency bands obtained by discrete Fourier transform (DFT). It is observed that good quality ridges manifest at a certain frequency band of the transformed fingerprint image as shown in Figure 9.

The most popular fingerprint quality assessment algorithm in literature is the National Institute of Standards and Technology (NIST) Fingerprint Image Quality (NFIQ) [46]. This approach also pioneers the use of quality metrics as performance predictor in fingerprints. A feature vector $v$ consists of 11 quality features obtained on the basis of localized quality map per fingerprint image. The map is computed based on the local orientation, contrast, and curvature of each region of a rectangularly tessellated fingerprint image (blocks with size $3 \times 3$). Rather than using true labels based on human perception, normalized separation of genuine match score from the match score distribution obtained from an automatic fingerprint matcher is used to train a multilayered perceptron. Recently, NFIQ 2.0 [51] is introduced with a similar learning-based quality assessment
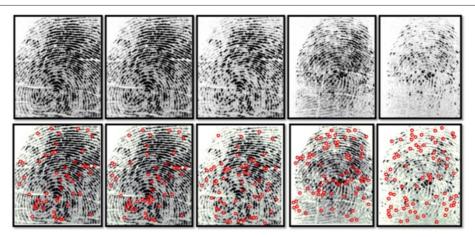


**Figure 8 Poor quality fingerprint samples often lead to spurious minutia.**

**Table 3 A representative list of fingerprint quality assessment algorithms**

| Category | Algorithm | Description | Type |
|---|---|---|---|
| Pixel intensity | Chen et al. [44] | Grey level distributions of segmented ridges | Local |
| Wavelet transform | Vatsa et al. [45] | Combined response from RDWT for dominant edge information | Local |
| Power spectrum | Chen et al. [3] | In a ring-shaped region of the spectrum | Global |
| Combined features | NFIQ [46] | Amplitude, frequency, and variance of sinusoid to model valid ridges | Global |
| Orientation tensors | Fronthaler et al. [47] | Encode orientation with parabolic symmetry features | Global |

framework in which several new image-based features are considered for inclusion, including Gabor filter responses.

The NFIQ quality metric has been extensively used in literature and tested across different datasets. However, the orientation estimated about the singularity points tends to fail for high curvature. Fronthaler et al. [47] present a solution based on characterizing orientation using parabolic symmetry features. The proposed technique first converts the image into orientation tensor representation. The orientation tensors in both horizontal and vertical direction are combined to encode the edge information obtained from the horizontal, vertical, or parabolic tensors. The information present in each local region is combined to obtain the final quality score. The paper also discusses using the same technique with higher-order orientation tensors to encode information in face images. The results indicate that correlation of this

quality score with NFIQ and with human annotations is high.

Alanso-Fernandez et al. [42] present a comparative study of several fingerprint quality metrics. These algorithms are segregated into global and local metrics depending on the nature of assessment. The study shows a high correlation of fingerprint quality metrics among themselves. This seems to indicate that the studied approaches encode similar information from the fingerprint image to predict quality. Recently, fingerprint quality computed using the ridge information in various subbands is shown to provide the best rejection criteria to improve performance [52]. The fingerprint ridge frequency and orientation were captured using short-time Fourier transform. The metric encodes the continuity of the ridge spectrum along the orientation of strong ridges in the image. In another research, self-organizing maps (SOM) are used to classify local regions of a fingerprint to
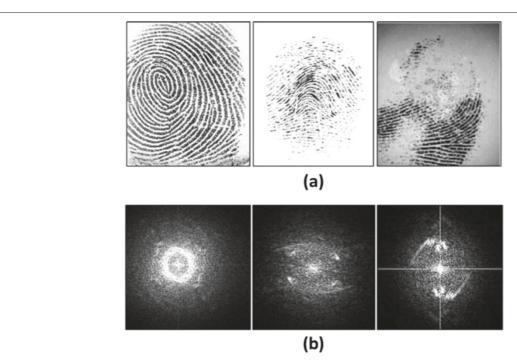


**Figure 9 A fingerprint image (a) and corresponding Fourier transform (magnitude component after shifting) (b).** The ridge information manifests as a bright band. Chen et al. [3] use the difference of two Butterworth filters to obtain a soft bandpass filter that captures the strength (and thereby quality) of the ridges.

different quality labels [53]. A SOM is trained to cluster blocks of fingerprints based on their spatial information to create a high-level representation of the fingerprint. Further, a *random forest* is used to learn the relationship between the SOM representation and actual matching performance.

The fingerprint quality assessment techniques measure consistency and strength of the ridge patterns. A direct association is made between the properties of the ridge patterns and the recognition performance of the sample. The more challenging problem of *latent* fingerprint quality assessment is also being studied [54-56]. Background noise, smudging, and partial nature of these types of fingerprints, usually obtained from crime scenes, hinder a good fit to precomputed models of ridge flows or patterns. Fingerprint quality metrics are also important for effective compression techniques [57]. Finally, quality assessment of 3D fingerprints that are obtained either from a 3D sensor or reconstructed from multiple 2D views, is an open research problem.

### 3.2 Iris quality assessment

The performance of the iris as a biometric is highly dependent on the quality of the sample. Some major covariates in iris recognition include focus and motion blur (due to hand-held sensors), off-angle (pose), occlusion (eye lashes, hair, and spectacles), dilation/constriction, and resolution. In order to compensate for these covariates, early iris capture systems were bulky and cumbersome to use. However, as newer and compact sensors with focus on usability emerge, there is greater need to measure the quality of the captured sample. Unlike fingerprints, iris patterns do not exhibit any *expected* behavior of the features, hence, quality is measured in terms of the impact of the covariate on the image. A brief description of some leading iris quality assessment methods is presented in Table 4.

Chen et al. [59] present a quality metric for iris based on the spectral energy in local regions. Firstly, iris is segmented using Canny edge detector and Hough transform. Next, occluded regions that may occur due to eyelashes are removed using intensity thresholding. The 2D Mexican hat wavelet decomposition is applied, and the product of responses from multiple scales (usually three) is used as the overall response. The iris region is partitioned into concentric bands with fixed width (8 pixels). The energy from concentric regions are separately computed and combined into a single quality score. Multiple overlapping filtering of the iris region approach is essential to encode the fine edges exhibited by the iris muscle tissue. The approach is also used for feature extraction. A similar approach is proposed by [62].

In another approach, Kalka et al. [8] present quality assessment of iris images based on the evaluation of eight quality parameters (defocus, motion blur, off-angle, occlusion, specular reflectance, illumination, and pixel count). These individual quality scores are both image-based and biometric-specific in nature. Further, Dempster-Sheffer theory-based fusion is used to combine these individual scores to obtain a single quality value. The quality measure is evaluated on the iris dataset of the West Virginia University (WVU) multimodal biometric database [63], using the quality bins approach discussed previously.

Recent interest in nonideal iris imagery has sparked research on iris recognition in the visible spectrum. Proenca [61] presents a quality assessment algorithm for operation on visible iris imagery. Similar to Kalka et al. [8], seven quality attributes that impact recognition are identified and estimated. The algorithm is tested via improvement in recognition rate when the lowest quality images from the database are ignored. The author also presents a summary of existing quality assessment algorithms for iris. In another approach, Zuo et al. [64] present an iris quality assessment technique based on match score evaluation. By utilizing precomputed distributions of genuine and imposter scores, the quality of a sample is measured by statistical fusion of two quality metrics: (a) statistical error between the distribution of genuine and imposter scores and (b) normalized difference between the sample match score and some quantile points selected from the genuine and imposter distributions. The authors later improve the approach [65] using a multivariant prediction (feed-forward neural networks) to better map quality values with matching performance. Baig et al. [66] also discuss a score level quality assessment based on Mahalanobis distance. Du et al. [67] present a feature correlation approach to assess the quality of an iris template. The measure can discriminate between natural iris patterns from the artifacts that occur during compression. It is observed that the correlation between consecutive rows of an iris template increases with compression as the less significant features are lost. The metric uses this distance measure of randomness of features as a measure of biometric quality of an iris sample.

It must be observed that the quality metrics in current literature assume accurate segmentation of the iris region as a precursor to the assessment module. However, as illustrated in Figure 10, iris segmentation methods are also adversely affected by the above-mentioned covariates. Recently, it has been shown that local quality metrics are able to predict iris segmentation performance [68]. Further, there is a lack of a benchmark approach and test-bed evaluation for academic and commercial iris quality assessment techniques. Considering the low complexity of the prevalent Hamming distance matching function, it might be interesting to consider a predictive quality assessment method similar to NFIQ.

**Table 4 A representative list of iris quality assessment algorithms**

| Category | Algorithm | Description | Type |
|---|---|---|---|
| Combined features | Daugman [58] | Focus estimate and off-angle measure by deformation function that maximizes circularity of pupil | Global |
| Power spectrum | Chen et al. [59] | Spectral energy in local regions of the iris | Local |
| Combined features | Zuo et al. [60] | Assessment of interlacing, illumination, focus, off-angle, area, blur pupil dilation | Local, global |
| Combined features | Kalka et al. [8] | Evaluation of seven quality parameters and fusing them statistically | Local, global |
| Combined features | Proenca [61] | Estimation of seven separate quality attributes that impact recognition | Local, global |

### 3.3 Face quality assessment

It is well established that quality measures are an important feature of modern face biometric systems due to the large degree of variations possible in face images (illustrated in Figure 11). However, quality assessment of faces has received comparatively less attention. Early research focuses on complete automation of essential capture guidelines in standards such as International Civil Aviation Organization (ICAO) and ISO. However, these guidelines are designed for manual recognition and provide minimal information about the quality of face biometric. More research focus must be directed towards this problem, since it has been observed in several empirical studies including the findings of biometric grand challenges that the covariates of face recognition (pose, illumination, expression, noise) affect the performance across different types of features or systems. A discussion of the existing face quality metrics is presented here and a brief summary is also available in Table 5.

#### 3.3.1 Still-face images-based techniques

Subasic et al. [69] present an evaluation scheme of a set of 17 automatic tests in conjunction with the ICAO face image presentation standards for automatic quality assessment. These tests are based on simple image processing techniques and semi-automatic annotation. The approach is tested on a set of 189 images. Further, the authors also mention some deficiencies in the ICAO standards such as lack of standard brightness, sharpness, color balance, and tolerance of background. In a similar approach, Hsu et al. [13] present a more comprehensive evaluator for the ISO/JEC 19794-5 face standards.

The approach combines several image quality metrics and face-specific metrics using facial feature detection. While a detailed description of the evaluation metrics is lacking, the authors evaluate several linear and nonlinear fusion schemes for match score prediction. Further, the authors use a nonlinear neural network, with the proposed set of quality metrics as feature vector, to predict the match score of a commercial face matching system.

Youmaran and Adler [5] discuss information content in biometric images termed as *Biometric information* (BI). From the information theory perspective, BI is defined as the decrease in uncertainty of the identity of a person caused by the feature set. Assuming each feature to be a multivariate random variable, BI is modeled as the relative entropy $\Delta D(p||q)$ between the intra-person feature distribution $p(x)$ and the inter-person feature distribution $q(x)$.

$$\Delta D(p||q) = \int p(x) \log \frac{p(x)}{q(x)} dx \qquad (1)$$

The approach is limited by the validity of the distribution $q$ which is the model for all possible faces. While this research provides good insight into quality assessment, the algorithm is not practical to implement, since it requires a statistically valid number of samples for each subject and probe subject to estimate the distribution of subject's features. Klare and Jain [77] propose a perceived uniqueness measure of a given face sample and match scores from any face matcher. The measure computes the distance of a match score to a set of imposter scores, thus indicating face uniqueness. Gao et al. [70] proposed the use of asymmetry in LBP features [71] as a measure of the quality of face biometric. However, this
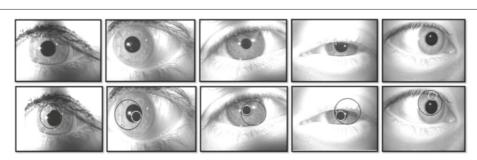


**Figure 10 Samples of poor iris segmentation on images obtained from CASIA-V4 iris database.**
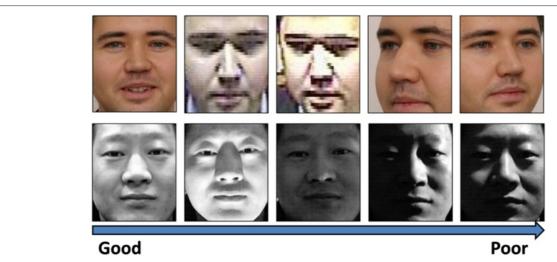
**Figure 11 Face images illustrating different levels of biometric quality.**

approach is limited in applicability as the face image must first be normalized to scale for the measurement to be accurate. The authors attempt a laborious solution of training a model for each possible scale. Zhang and Wang [72] improve on this intuition using scale invariant feature transform (SIFT) features [78]. It is suggested that illumination variation primarily affects face recognition systems. The assessment of quality is based on the assumption that given a normalized frontal face image, the location of SIFT-based feature points will be symmetric with a vertical axis. Based on this observation, quality is estimated as the ratio of the number of available points on each side of the axis. The work does not discuss any guarantee that the SIFT features are symmetric over any axis in good quality images. Further, any natural asymmetry in face, any symmetric illumination, or other noise can lead to incorrect estimation.

Recently, quality assessment in face images has renewed interest attributed to insights from the Good, Bad, and Ugly (GBU) dataset [79]. The challenging dataset used

in Face Recognition Vendor Test (FRVT) 2006 [80] consists of 9,307 frontal neutral expression face images taken in indoor or outdoor settings from 570 subjects. From this dataset, a subset of 2,170 images from 437 subjects is chosen and split into three sub-partitions (*Good, Bad and Ugly*) such that the fusion of the top three algorithms from FRVT 2006 results in GAR of 0.98, 0.80, and 0.15 at an FAR of 0.001. Further, no image appears in more than one subset and the subjects in all three partitions are the same. This unique partitioning of data enables researchers to focus on the hard matching problems of face recognition within the database. Also, this dataset can be used to better understand and model the change in *recognizability* of a subject in different environmental conditions. Phillips et al. [7,81] show that simple image quality metrics can be combined to predict face recognition performance. Using a greedy pruning approach, ranking is predicted from a quality oracle. Aggarwal et al. [82] show that good, bad, ugly pairs can be predicted by using partial least square regression between image-based features

**Table 5 A representative list of face quality assessment algorithms**

| Application | Algorithm | Description |
|---|---|---|
| Still-image | Subasic et al. [69] | Seventeen automatic tests in conjuncture with the *ICAO* face image presentation standards |
| | Hsu et al. [13] | Automatic evaluator of the ISO/JEC 19794-5 face standards |
| | Youmaran and Adler [5] | Biometric information defined from information theory |
| | Gao et al. [70] | Asymmetry in LBP features [71] as a measure of the quality |
| | Zhang et al. [72] | Asymmetry using SIFT features |
| Video-frame | Wong et al. [73] | Comparison of a facial image with *ideal* face models |
| | Nasrollahi and Moeslund [74] | Geometrical pose estimation using face bounding box |
| | Long et al. [75] | Assess sharpness, brightness, resolution, and pose in NIR videos |
| | Yao et al. [76] | Sharpness measure from frame selection |

(sharpness, hue, and intensity) and geometric attributes of a face (obtained using active appearance modeling). Hua et al. [83] use modulation transformation function to compute the sharpness in face images. Their results also indicate that sharpness is an important factor to improve face recognition results.

### 3.3.2 Video-based techniques

An important application of quality assessment in face biometrics is in video face matching [84]. Here, face recognition is performed on a video stream rather than a single still image. Some approaches of this branch of research use quality assessment for *frame selection* in order to match the best possible frame from gallery and probe face video. Wong et al. [73] present a patch-based approach using the first $d$ low-frequency components of the discrete cosine transform (DCT) obtained from each facial patch. A multivariate probabilistic model is generated using a training set of frontal faces with acceptable illumination per patch, and the probe image is compared, patch-wise, to obtain the overall quality.

The general approach for video face quality assessment is based on comparing the input face image with face models developed from *ideal* example sets. In another approach, Nasrollahi and Moeslund [74] present a simple geometrical approach based on the dimensions of the bounding box of face detection algorithm in a video face recognition system. Since pose is a primary challenge in such systems, this approach can be considered as a simple pose assessment technique. A similar approach is also used recently by Long and Li [75] for NIR video face recognition. Yao et al. [76] use a sharpness measure from frame selection for a recognition system designed for low-resolution face videos. It must be noted that while face quality assessment has received considerable attention in video face recognition research, the requirement in this particular application is for a *binary* decision (accept/reject) per video frame. Hence, such quality metrics may not sufficiently measure the quality of the face biometric sample.

The unique attribute of FRVT 2006 [80] is in providing several thought-provoking insights and directions to the problem of quality assessment in face recognition [70]. These findings are discussed by Beveridge et al. [7,39,85] with a detailed analysis of the *effect* of various *subjective* and *objective* covariates of face biometric. Current literature describes the quality of a face image as an intrinsic property of the image. Beveridge et al. [39] argue that if this intuition were true, a higher-quality sample would be consistently matched correctly. Likewise, a low-quality sample would consistently perform poorly. However, their experiments indicate that the confidence of match is dependent on the quality of both the images being matched, i.e., a considerable number of images that

are hard to recognize as part of one match pair are easy to recognize as part of other match pairs. This indicates that verification can be correctly performed if both images lie in the same quality space. The NIST Multiple-Biometric Evaluation (MBE) [86] presents six state-of-the-art commercial face recognition systems on various demographic and covariate challenges which indicate that the performance of all algorithms is affected by various factors such as gender, age, and ethnicity, apart from known covariates of pose, illumination, and expression. Hence, it follows that a quantitative measure of quality of an input face image that provides an estimate of matching performance is critical. Recently, holistic descriptors extracted from the face region are shown to be good indicators of performance of face recognition systems [87]. The low computation time of these image descriptors make them ideal features for quality assessment. Further, pseudo-labels of quality obtained from matching performance provide a direct estimate of recognizability of a given face image. Therefore, the approach is more useful than separate estimation of different covariates. The large degree of freedom of face greatly increases variability in captured information compared to other biometric modalities, making quality assessment an essential prerequisite. For face recognition systems to have robust performance outside of studio-like conditions, quality assessment of face must encapsulate the aforementioned covariates effectively.

### 3.4 Evaluating quality assessment approaches

An important aspect in the development of quality assessment algorithms is the way their performance is measured. Since the primary motivation of most image quality assessment techniques is in perceptual understanding of the image, human annotation of quality is considered as the gold standard for comparison and testing of automatic algorithms. A set of volunteers is presented with images of different quality and their responses are aggregated to a mean operator score (MOS). A high correlation between the predicted quality and MOS from volunteers indicates high performance [88]. Based on the aforementioned discussion, MOS cannot be directly applied for biometric quality, as there is no conclusive evidence that human interpretation of quality correlates with the quality in terms of the performance of a recognition algorithm. In our observation, six prominent methods of evaluation of biometric quality metrics persist in literature apart from evaluation using MOS:

- Correlation analysis: As noted by [4], a biometric quality metric must be a good classifier performance predictor. With this view, a quality measure that is highly correlated (statistically) with match scores obtained

from a classifier is the most desirable. Hence, several researchers discuss correlation with genuine match scores [42,89]. Since every match score can be associated to the quality of both gallery and probe sample, combining methods, such as $Q_{gallery} + Q_{probe}$ or $\sqrt{Q_{gallery} \times Q_{probe}}$ or $\min(Q_{gallery}, Q_{probe})$ are utilized.

- Modeling: Recently, quality metrics are utilized as predictors for dynamic processing and context switching. When correlation is established, the relationship between a series of quality scores (predictors) and associated match score (response) can be explicitly described by *modeling* using regression analysis, as shown subsequently in this research. Further, the goodness-to-fit can be evaluated by analysis of variance and inspection of residual error of fitting.
- Quality bins: In another approach, the impact of quality metrics is measured by segregating the entire dataset into a number of quality bins and performing individual recognition experiments on each of them. Further, the intuition that better quality data has better recognition accuracy is substantiated with recognition results on these quality bins [3,8,47,90].
- Distance metric: Quality score is also used to alter the feature space to improve matching. Chen et al. [3] incorporate their proposed iris quality assessment metric (computed for both gallery and probe) in the formulation of Hamming distance matcher to show improved results when compared to simple Hamming distance.
- Cross-correlation: Another possible method of evaluating quality metrics is by computing the cross-correlation between the given metric and various existing metrics [47]. In biometrics, this can be considered as a weak measure unless some additional benefits of the algorithms (in terms of computation time or better correlation with MOS) is described that differentiate from existing approaches.
- Computation time: The performance of a quality assessment algorithm in terms of computation time is an important aspect of its evaluation. In most use-cases, performing quality assessment is only meaningful when complexity is low. For instance, biometric quality assessment can only be a small overhead to the recognition pipeline. Reported computational time of a quality metric is dependent on the implementation platform and machine configuration in use. However, computational efficiency of techniques reported relative to computation time of PSNR allows for a machine-independent comparison [41].

## 4 Analysis of quality metrics

Quality metrics have been extensively used to improve the robustness and accuracy of biometric systems. Several fusion and context-switching approaches are proposed based on the intuition that quality can be indicative of the *utility* of a biometric sample. However, as discussed in Section 2, the role of a quality metric in improving the performance of a biometric system is not always implicit. Hence, an arbitrary quality metric '*q*,' defined in abstraction in various formulations of multibiometrics, must be investigated more closely. In this section, a representative set of image and biometric quality metrics is evaluated to understand their relationship with each other and with match scores. For the evaluation, match scores obtained from commercial matchers are used on WVU multimodal biometric database.

### 4.1 Database and evaluation protocol

The evaluation is performed on the WVU multimodal database [63] that contains face, fingerprint, and iris modalities. For the experiment, two images pertaining to 250 subjects (per modality) are chosen for gallery and the remaining images are used as probe. To evaluate the performance of quality metrics, three uni-modal biometric matchers are used. Fingerprint classifier used in this study is the NIST Biometric Image Software (NBIS) [91]. NBIS consists of a minutiae detector called MINDTCT and a fingerprint matching algorithm known as BOZORTH3. For face and iris biometrics, Neurotechnology [92] feature extractors and matchers are used. The performance of the matchers is illustrated in Figure 12. The varied image quality result in a considerable overlap of genuine and imposter score distributions.

As discussed in previous sections, quality metrics can be either image-based or modality-specific. A representative set of quality metrics of both types are chosen for evaluation. Specifically, four image quality approaches and a biometric quality approach (that may each contain multiple measures) are considered for the evaluation. The abbreviations associated with each of the quality metrics are presented in Table 6 and a brief description is presented below. The techniques are all no-reference quality metrics and have low computational complexity when executed on a typical desktop machine. A detailed discussion of the computational complexity of each technique is available in the references:

- Spectral energy (SE) calculates the block-wise energy using Fourier transform components [93]. It describes abrupt changes in illumination and specular reflection. The image is tessellated into several nonoverlapping blocks, and the spectral energy is computed for each block. The value is computed as the magnitude of Fourier transform components in both horizontal and vertical directions that shows the amount of spectral energy per block.
- Marziliano et al. [94] have proposed edge spread (ES) as a measure to estimate irregularities based on edges
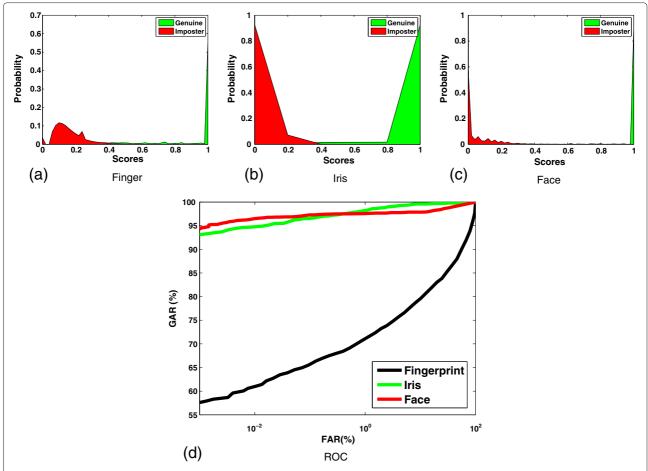
**Figure 12 Matchscores obtained for the three modalities.** Genuine and imposter score distribution for **(a)** face, **(b)** fingerprint, and **(c)** iris matchers on the WVU multimodal dataset used in this research. **(d)** Receiver operating characteristic (ROC) curve illustrates the verification performance of the respective matchers indicating the overall quality of the database.

and their adjacent regions. Specifically, it computes the effect of irregularity in an image based on the analysis of the difference in image intensity with respect to the local maxima and minima of pixel intensity at every row of the image. Edge spread can be computed in horizontal as well as vertical directions. However, the experiments in [94] show that either of the two directions suffices for quality assessment.

- A no-reference perceptual quality metric by Wang et al. [95] primarily measures compression artifacts. It is computed as the combination of blockiness and activity estimation in both horizontal and vertical directions, manifesting in three metrics: blockiness (B), activity (A), and zero-crossing rate (Z).
- A spatial domain no-reference quality assessment technique, termed *BRISQUE* (BR), proposed by Mittal et al. [41], provides a holistic assessment of *naturalness.* The quality metric is a deviation measure of a natural image from the regular statistics, indicating distortion.

Further, three modality-specific quality metrics are also used:

- Iris: Kalka et al. [8] evaluates defocus (DF), motion blur (MB), occlusion (O), illumination (I), specular reflectance (SR), and pixel count (PC). Further, a fused metric (Q) is obtained using DS-theory. The technique is discussed in Section 3.
- Fingerprint: As described in Section 3, Chen et al. [3] proposed ridge energy for fingerprint quality assessment. It is the Fourier spectrum energy computed on a frequency bandpass region where fingerprint ridges strongly manifest. In addition, a discrete quality value obtained from the NFIQ [46] tool is also utilized in this study.
- Face: For face quality assessment, geometric pose estimation (P) is computed. First, positions of eyes and mouth are estimated using corresponding Adaboost detectors [96]. Pose is estimated based on the deviation of geometric measures (inter-eye distance and

**Table 6 Various representative quality metrics considered in this study**

| Abbreviation | Quality metric |
| --- | --- |
| SE | Spectral energy |
| ES | Edge spread |
| B | Blockiness |
| A | Activity |
| ZC | Zero count |
| BR | BRISQUE |
| DF | Defocus |
| M | Motion |
| O | Occlusion |
| L | Lighting |
| S | Specular reflectance |
| PC | Pixel count |
| Q | Fused iris quality |
| RE | Ridge energy |
| NFIQ | NIST fingerprint image quality |
| P | Pose |
| F | Focus |

eye-center to mouth distance) from mean values. Additionally, focus measure (F) reported in [85] is also utilized.

### 4.2 Experimental analysis

Two key ideas are evaluated in this study: (i) the relationship between different quality metrics and (ii) the relationship of the quality of a pair of biometric samples with their match score. All match scores are converted to similarity measures for easy visualization. Some key insights can be drawn for both image-based and biometric-specific quality metrics as follows:

- Spearman correlation values for all quality metrics for face, fingerprint, and iris images are shown in Tables 7, 8, 9 respectively. The quality score from gallery and probe pair is combined as $Q = \sqrt{Q_{\text{gallery}} \times Q_{\text{probe}}}$. Low Spearman correlation is observed between the quality metrics in consideration indicating that they measure diverse aspects of quality. For instance, no-reference quality measures A in $8 \times 8$ blocks in the image. On the other hand, ES measures the gradient difference at edge boundaries, to measure blurring. Even though both are measures of blurring, the difference in approaches leads to low correlation between them.
- Scatter plot in Figures 13, 14, 15 illustrates genuine and imposter match scores against each quality metric in consideration. A three-dimensional plot of match

**Table 7 Spearman correlation between face quality scores**

| | SE | ES | B | A | ZC | P | F | BR |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| SE | 1.00 | 0.14 | −0.02 | −0.02 | −0.03 | −0.11 | −0.07 | 0.12 |
| ES | | 1.00 | −0.06 | 0.09 | −0.12 | −0.04 | 0.08 | −0.12 |
| B | | | 1.00 | 0.97 | −0.15 | −0.06 | 0.31 | −0.57 |
| A | | | | 1.00 | −0.15 | −0.07 | 0.29 | −0.56 |
| ZC | | | | | 1.00 | −0.10 | −0.11 | −0.33 |
| P | | | | | | 1.00 | 0.08 | 0.08 |
| F | | | | | | | 1.00 | −0.27 |
| BR | | | | | | | | 1.00 |

scores versus quality of gallery and probe clearly illustrates the characteristic relation between them.

- For all three modalities, no relation is observed between quality scores and imposter match scores. A similar observation is made in the case of fingerprints in [42].
- In case of certain quality scores such as Activity, Zero-Cross rate, and Focus, genuine match scores are found only in specific quality bins. Hence, any pair exhibiting quality in this range during test phase induces more confidence in matching [97]. Such simple quality measures provide an additional information to improve classification. For example, in case of A of fingerprints, the values pertaining to genuine scores are observed in the range of 15 and 25.
- For face and iris modalities, quality metrics that measure prominence of edges better map to genuine scores. For instance, ES and RE provide more confidence to genuine score than other metrics such as DF. Further, spatial no-reference measure (BR) correlates with activity measures and also characterizes the genuine scores for face and fingerprint.
- In order to evaluate the relevance of quality scores in augmenting or predicting match scores, an illustration of the cumulative density function (CDF) is presented

**Table 8 Spearman correlation between fingerprint quality scores**

| | SE | ES | B | A | ZC | RE | NFIQ | BR |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| SE | 1.00 | 0.12 | 0.12 | 0.17 | 0.00 | 0.52 | −0.02 | 0.12 |
| ES | | 1.00 | −0.21 | −0.19 | −0.18 | 0.05 | 0.03 | −0.08 |
| B | | | 1.00 | 0.94 | −0.40 | −0.03 | −0.30 | 0.61 |
| A | | | | 1.00 | −0.30 | −0.04 | −0.33 | 0.68 |
| ZC | | | | | 1.00 | 0.03 | 0.22 | −0.37 |
| RE | | | | | | 1.00 | 0.08 | 0.00 |
| NFIQ | | | | | | | 1.00 | −0.27 |
| BR | | | | | | | | 1.00 |

**Table 9 Spearman correlation between iris quality scores**

|    | SE   | ES   | B    | A     | ZC    | DF    | M     | O     | L     | S     | PC    | Q     | BR    |
|----|------|------|------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| SE | 1.00 | 0.02 | 0.25 | 0.29  | 0.18  | −0.09 | 0.03  | −0.07 | 0.01  | −0.11 | −0.10 | 0.13  | −0.05 |
| ES |      | 1.00 | 0.00 | −0.01 | −0.16 | −0.06 | −0.06 | −0.03 | 0.01  | 0.13  | 0.02  | 0.09  | −0.01 |
| B  |      |      | 1.00 | 0.97  | 0.33  | −0.53 | −0.15 | 0.05  | 0.18  | −0.08 | −0.03 | 0.48  | −0.23 |
| A  |      |      |      | 1.00  | 0.38  | −0.49 | −0.12 | −0.02 | 0.16  | −0.15 | −0.09 | 0.45  | −0.19 |
| ZC |      |      |      |       | 1.00  | −0.09 | 0.29  | −0.05 | −0.23 | −0.43 | −0.22 | 0.11  | −0.02 |
| DF |      |      |      |       |       | 1.00  | 0.12  | −0.12 | −0.20 | −0.15 | −0.13 | −0.57 | 0.14  |
| M  |      |      |      |       |       |       | 1.00  | −0.08 | −0.15 | −0.04 | −0.10 | −0.23 | 0.09  |
| O  |      |      |      |       |       |       |       | 1.00  | 0.02  | 0.56  | 0.92  | −0.27 | 0.03  |
| L  |      |      |      |       |       |       |       |       | 1.00  | 0.07  | 0.06  | −0.09 | −0.03 |
| S  |      |      |      |       |       |       |       |       |       | 1.00  | 0.74  | −0.12 | −0.06 |
| PC |      |      |      |       |       |       |       |       |       |       | 1.00  | −0.30 | 0.05  |
| Q  |      |      |      |       |       |       |       |       |       |       |       | 1.00  | −0.24 |
| BR |      |      |      |       |       |       |       |       |       |       |       |       | 1.00  |

in Figure 16. The CDF of certain quality scores are more similar to the obtained match scores, such as RE, B, O, and I as compared to ES, BR, and Z.

- To test the relationship between the quality scores and match scores obtained from each modality, a linear regression analysis is performed between the genuine scores and quality scores. As discussed previously, the quality scores from gallery and probe are combined as $Q = \sqrt{Q_{\text{gallery}} \times Q_{\text{probe}}}$. Further, the data is randomly split into nonoverlapping train and test sets. The mean squared error (MSE) of each modality, over ten times random cross-validation, is shown in Figure 17. It is observed that even with 10% of the data as training samples, genuine scores from matchers can be predicted with quality metrics using a simple linear model. To analyze the quality of fit of the regression model, analysis of variance (ANOVA) is performed to assert the effect of each quality metric in consideration as match score predictors. The analysis indicates that ES, A, DF, MB, O, PC, Q, and BR are effective with $p$ value less than 0.01 for iris modality. On the other hand, SE, ES, B, A, and Z are more effective in estimating match scores for fingerprints. We also observe that only P and ES are able to estimate match scores of the face.

In this study, it is empirically established that a direct relationship exists between certain quality metrics and match scores (which can also be viewed as classifier confidence). This encouraging result sanctions the use of quality metrics in multibiometric schemes such as quality-based fusion and context-switching. However, as observed from the scatter plots, the choice of quality metrics is an important factor.

### 4.3 Discussion

Traditional image quality metrics measures certain aspects of an image important for good visual perception. On the other hand, biometric quality assessment measures the potential of the sample for recognition. As shown in literature, such quality metrics not only help in improving data collection but also provide additional information at different stages of a biometric system. Based on the literature review and experimental analysis, here, we collate the important observations pertaining to biometric quality assessment:

- The prominent features used in quality assessment are orientation of edge features. While a strong case can be made for the performance of these features, research has shown potency of color-based and intensity-based features as well.
- There is a need for better evaluation framework for biometric quality assessment metrics. High correlation with match score performance along with statistical tests can help towards better evaluation. The good, bad, and ugly distribution of database [79] is an interesting method for evaluating the performance of quality metrics for performance prediction.
- Researchers must emphasize on the computational cost in the development of quality assessment approaches, which must be lesser or comparable to the matching time.
- Quality metrics used for quality-based multibiometric fusion approaches must be carefully selected. As discussed in Section 4, not all quality metrics are useful for match score prediction. Quality metrics that measure different kinds of degradations, including modality-specific metrics, must be considered.
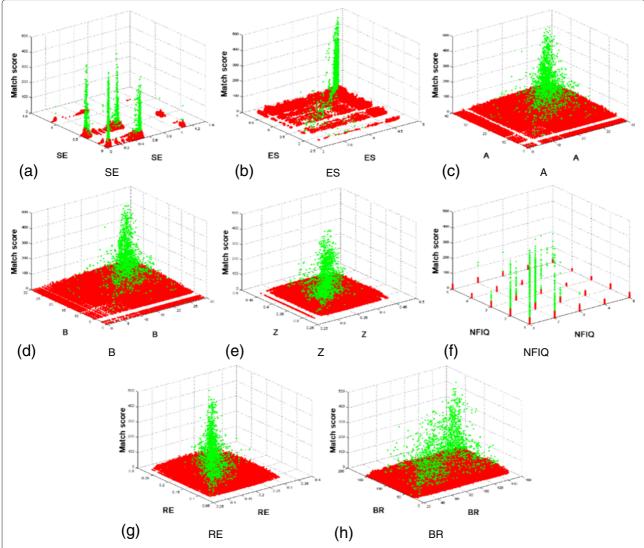
**Figure 13 Relation between match scores obtained from NBIS fingerprint matcher and various quality metrics.** Relation between match scores obtained from NBIS fingerprint matcher (*z*-axis) and various quality metrics [**(a)** SE, **(b)** ES, **(c)** A, **(d)** B, **(e)** Z, **(f)** NFIQ, **(g)** RE, **(h)** BR] for genuine (green) and imposter (red) match pairs. The *x*-axis pertains to gallery quality, while *y*-axis pertains to the probe quality. The scattering indicates that ES, A, B, Z, RE, and BR quality metrics can characterize genuine scores.

- In differential processing techniques such as context switching, quality metrics can be important cues for selection of recognition modules. Based on the modality in consideration, additional factors such as age and gender may also be considered as cues [98].
- It is our assertion that a better understanding of the behavior of biometric quality, in terms of *naturality*, *fidelity* and *utility*, can help in the development of more meaningful quality measures. Such quality metrics may also enhance the performance of quality-based multibiometric frameworks proposed in literature.
- Face quality is affected by pose, illumination, and expression apart from image degradations such as noise and blur. Other covariates such as aging, disguise, and occlusion degrade the performance relative to a reference sample.
- The quality of a match pair is a function of the quality of both gallery and probe images [39]. Further, high-resolution frontal face images do not directly imply high-quality biometric sample or confident match.
- Important findings from the results of the FRVT 2006 [80] and MBE [86] can help towards development of better quality assessment techniques.

For instance, a slight gender bias is observed in the performance of the algorithms, with samples of female subjects performing better than male subjects in controlled environment. Also, the evaluations found that
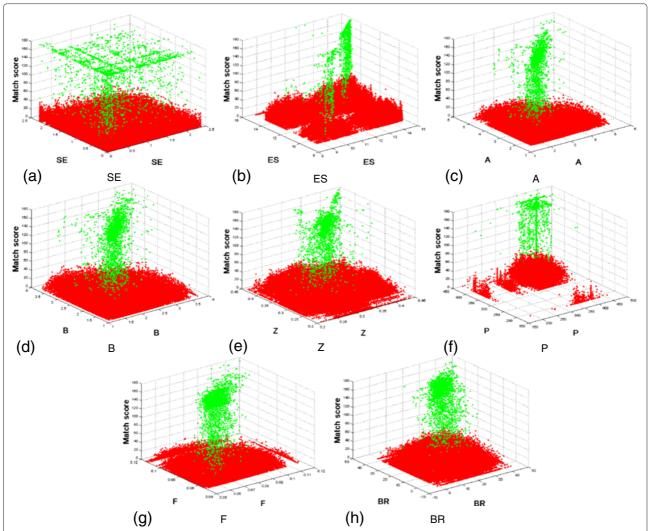
**Figure 14 Relation between match scores obtained from a commercial face matcher and various quality metrics.** Relation between match scores obtained from a commercial face matcher (*z*-axis) and various quality metrics [**(a)** SE, **(b)** ES, **(c)** A, **(d)** B, **(e)** Z, **(f)** P, **(g)** F, **(h)** BR] for genuine (green) and imposter (red) match pairs. The *x*-axis pertains to gallery quality, while *y*-axis pertains to probe quality. The scatterplot indicates that A, B, Z, F, and BR quality metrics can characterize genuine scores.

samples obtained from individuals of a certain race perform better than others, with East-Asian races performing the best.

- A strong correlation has been observed between simple image quality measures and performance of the top three algorithms of the vendor test [7]. Precisely, a high correlation has been observed between the recognition rates and a simple gradient energy-based focus measure.
- The performance of samples captured in indoor studio-like conditions is better than the performance of samples taken in uncontrolled outdoor conditions. While this result is expected, it is interesting to note that this penalty in performance decreases with relaxed false acceptance rates.

- The quality of a fingerprint sample is largely governed by the sensor in deployment. It is observed that the common factors include scars, burns, dryness, and temperature. *Auto capture* is a common feature in modern fingerprint sensors, requiring real-time quality assessment of the presented sample. Therefore, most quality metrics evaluate ridge clarity and number of detected minutia.
- The performance of iris as a biometric is hugely dependent on the quality of captured sample. The micro-features of iris texture are easily contaminated by adverse illumination, lenses, glasses, or disease. The most prevailing approach for iris quality measurement continues to be the fusion of assessment of several known quality factors.
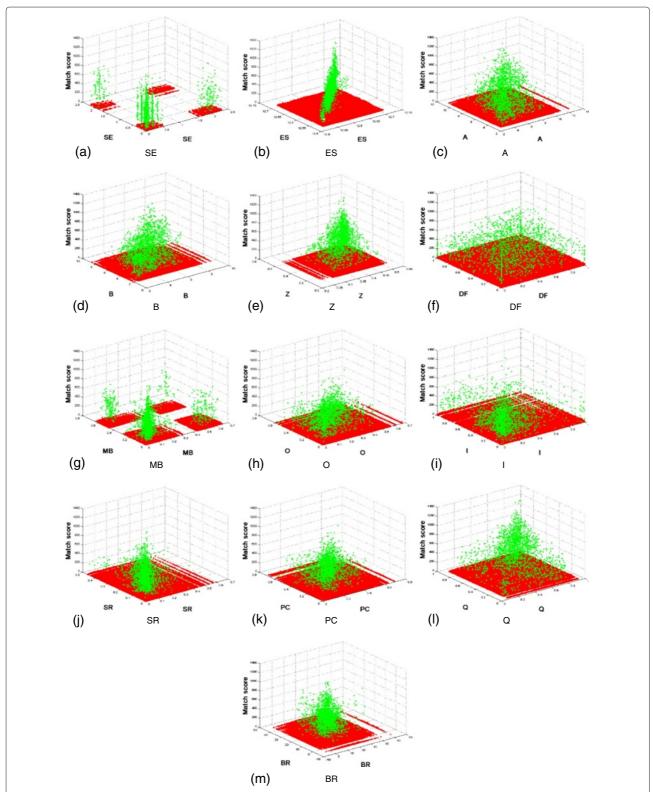
**Figure 15 Relation between match scores obtained from a commercial iris matcher and various quality metrics.** Relation between match scores obtained from a commercial iris matcher (*z*-axis) and various quality metrics [**(a)** SE, **(b)** ES, **(c)** A, **(d)** B, **(e)** Z, **(f)** DF, **(g)** MB, **(h)** O, **(i)** I, **(j)** SR, **(k)** PC, **(l)** Q, **(m)** BR] for genuine (green) and imposter (red) match pairs. The *x*-axis pertains to gallery quality while *y*-axis pertains to probe quality. The scatterplot indicates that ES, A, B, Z, SR, PC, and BR quality metrics can characterize genuine scores of match pairs. However, DF, O, I, and Q are unable to characterize genuine match scores.
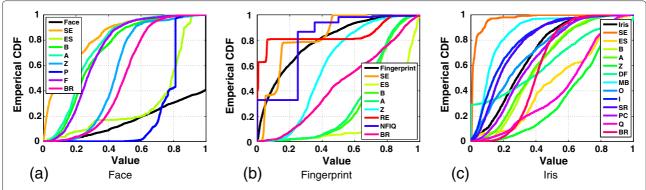
**Figure 16 The cumulative density functions (CDF) between genuine score and quality metrics for (a) face, (b) fingerprint, and (c) iris modalities.** The plots compare the distribution of each quality metric with the corresponding genuine score distribution.

- Due to the requirement of low computational time, *auto capture* in iris sensors is usually based on confidence of segmentation. A major drawback of existing approaches is in the assumption of good quality segmentation before quality assessment. However, same factors that affect biometric quality are also known to effect iris segmentation.
- Current research uses typical image processing algorithms that evaluate image degradations due to noise, compression, or illumination. However, a quality metric that entails a greater insight of the *usefulness* of the biometric sample in consideration can improve the performance of these systems by providing more discernible quality *cohorts.*

## 5   Conclusions

Quality assessment of biometric samples is an important challenge for the biometrics research community. In this survey paper, a clear distinction is made between the image quality and biometric quality of a biometric sample to capture modality-specific intuitions of quality assessment. It is our assertion that quality metrics are an important ingredient in improving the robustness of large real-world biometric systems. In an attempt to demystify the definition and work of biometric quality, several factors that affect a biometric sample are presented. Different image features utilized in literature for quality assessment, evaluation processes, and match score predictability are discussed. Further, a literature survey
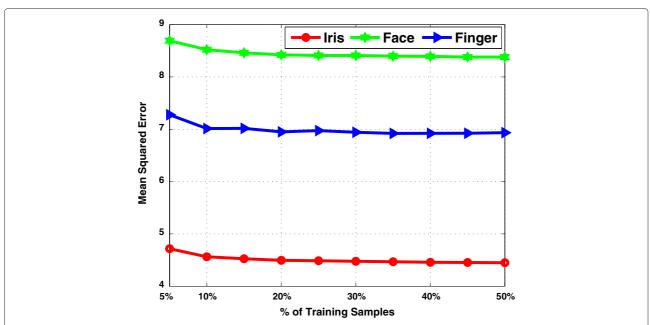


**Figure 17 Results of the regression test.** MSE of the regression test with genuine scores and quality metrics accumulated over 10 times cross-validation. Even with a small number of training samples, a linear model can predict match scores of genuine pairs showing that quality scores can be indicative of matching performance.

of the quality assessment techniques in three biometric modalities reveals that techniques often focus on naturality alone. It is imperative that quality assessment entails a notion of fidelity of capture and modality-specific utility as well. Further, the performance of a biometric quality assessment metric in terms of computational complexity must also be discussed more actively in research. The development of quality assessment algorithms of biometric samples that are computationally inexpensive to compute yet correctly encode quality will be the *sine qua non* of real-world large-scale deployments. Using quality assessment metric cannot, however, be a panacea for the recognition of poor quality images. Beveridge et al. [99] place a bound on the extent to which quality metrics can improve the performance of matching systems when they are used as performance predictors.

## Appendix 1: perceptual image quality assessment

The assessment of the quality of an *image* is important to measure and control its degradation during acquisition, compression, transmission, processing, and reproduction [1]. Several quality assessment algorithms exist in image processing literature, which pursue different philosophies, performance, and applications. A majority of these methods are motivated towards accurate *perceptual* image quality i.e., quality as perceived by the sophisticated HVS. Two distinct approaches exist in literature to model the HVS: a *bottom-up* and a *top-down* approach [1]. The first approach is based on the replication of various mechanisms of the HVS which entails a deep understanding of its anatomy and psychophysical features. Many are categorized and summarized by Wang and Bovik [1]. The second approach treats the performance of the HVS as a black box, dealing with only the input to and output from the HVS. Both approaches are important; however, optimized solutions often lie in a middle ground of both approaches to this problem.

Depending on the amount of information required, quality assessment algorithms can be segregated as *full-reference* (FR), *no-reference* (NR), and *reduced-reference* (RR) quality assessment. A detailed discussion of each of these categories is presented next.

1. *Full-reference* or FR: This category of algorithms require a distortion-free or perfect quality version of the same image, the 'original image,' in order to assess the quality of the input images. These approaches perhaps have received most interest from the community due to wide applicability in areas of quality of service (QoS) in delivery of image-based content. Most FR bottom-up quality assessment methods share a similar framework known as the *error-visibility* paradigm [1]. The strength of error computed between the given image and the original (reference) image is

weighted based on known features of the HVS. This ensures that the quality metric validates those errors which have the maximum affect on human perception. A generic error-visibility-based quality assessment framework consists of three phases discussed below:

(a) *Preprocessing*: The input reference and distorted image undergo a preprocessing stage, usually comprising of spatial registration, color space transform (to YCbCr), and filtering. It is assumed that reference and given images become properly aligned. Even small errors in registration can lead to largely incorrect prediction of quality. Sometimes, some point-wise nonlinear transformations can be applied to reduce the dynamic range of the luminance. These preprocessing techniques are also often have channel-specific parameters, as different channels have different characteristics.

(b) *Channel decomposition*: Motivated by the frequency and orientation-specific neurons in the visual cortex, the image is usually decomposed into multiple channels using decomposition techniques such as Fourier decomposition, Gabor decomposition, DCT transform, or separable wavelet transform. Each of these decomposition techniques differs in their mathematics, implementation details, and suitability to task; however, there is no clear consensus on which decomposition is better than the rest.

(c) *Error normalization and pooling*: After decomposition of both reference and given image, the error is calculated as the (weighted) difference between both sets of coefficients. These errors are often normalized in a *perceptually meaningful way* [1].

The FR top-down quality assessment algorithms have been very successful in a wide range of applications primarily due to their simplicity in design. A popular approach in literature is the *structural similarity*. This quality assessment paradigm utilizes the fact that natural images are highly structured. Hence, any unstructured information in the image is a quality degradation. A spatial domain implementation of this idea is the *structural similarity index metrics* (SSIM) [100]. Given a distorted image ($x$) and reference image ($y$), the SSIM index of quality depends on the comparison of $x$ and $y$ by three measures: luminance, contrast, and structure.

2. *No-reference* or NR: Blind or no-reference quality assessment is a more difficult problem as there is no reference image for comparison. The human visual system is able to perform blind assessment primarily due to immense prior knowledge and superior understanding of what an image is. Some distortions in an image can be assessed effectively without reference, for example, blurring and blockiness during image compression. In general, for NR quality assessment, it helps to have prior knowledge of the expected degradation process on the image. A NR perceptual quality assessment algorithm for JPEG compression is proposed by Wang et al. [95]. This method primarily measures distortions in an image due to compression (such as blockiness and blurring). It is a combination of blockiness and activity estimation in both horizontal and vertical directions.

3. *Reduced-reference* or RR: Quality assessment with reduced references is a relatively newer aspect of image quality assessment research. Here, the ancillary channel (usually noise-free, but not necessarily) transmits features of the original image that can be used to determine the quality of the image at the receiver end. This quality assessment paradigm is developed to monitor the quality of video streams transmitted through various noisy channels. An early technique in literature computes reference information from a random set of preselected pixel values. At the receiver end, the MSE of pixel values of the original and distorted image is be computed to obtain quality. Gao et al. [101] propose using multiscale geometrical analysis and compute a concise feature set that is normalized to improve HVS consistency. This feature vector (used as reference) encodes structural information that is perceived by HVS.

The primary method of representing biometric information of an individual is by an image. As noted above, most image quality assessment research is motivated towards *perceptual* quality of an image. Nevertheless, several important insights can be drawn from this matured research area towards a quality metric relevant to biometrics. For a detailed review of existing image quality assessment, readers are referred to [1,2]. An important difference being that biometric quality *relates* to the performance of automatic biometric systems rather than the human visual system. In fact, this constraint can have several advantages such as ease of evaluation, and algorithms can be easily tested when compared to testing with human subjects. Also, most recognition algorithms are better understood internally than the human visual system; hence, there is no need to account for various cognitive anomalies.

## Appendix 2: biometric standards

A large number of commercial and public biometric systems/solutions have lead to the standardization of several processes. This ensures inter-operability among different vendors and ensures easy integration. Here, some leading biometric standards are presented [14,102]:

1. *CBEFF*: The Common Biometrics Exchange File Format (CBEFF) [102], developed in 2001, facilitates exchange of biometric data including raw and processed biometric sample. The standardization is achieved through three major sections: Standard biometric header (SBH), Biometric Data Block (BDB), and Signature Block (SB). Further, this standard presents a nested structure with same or different modalities. This ensures a single block structure per template in multimodal or multisample systems. Within the BDB block, there is an optional field called Biometric Data Quality. The block provisions for a single scalar quantity (0 to 100) based on the ANSI/INCITS-358 standards of 2002 (discussed next). Additionally, the field also notes if the quality value is of a nonstandard variety.

2. *BioAPI*: This standard describes the specifications of an Application Programming Interface (API) in order to accommodate for a large number of biometric systems, sensors, and applications. This API is designed for system integration and application development in biometrics. The bioAPI 1.1 standard describes in Section 2.1.46 [14], a structure called bioapi_quality that indicates the quality of the biometric sample in the biometric identification record [14]. Since there is no 'universally accepted' definition of quality, bioAPI has elected to provide this structure with the goal of framing the effect of quality on usage of the vendors. The scores are based on the purpose (another structure in bioAPI called bioapi_purpose) indicted by the application (e.g., capture for enrollment/verify, capture for enrollment/identify, and capture for verify). Additionally, the demands upon the biometric vary based on the actual customer application and/or environment (i.e., a particular application usage may require higher quality samples than would normally be required by less demanding applications). Quality measurements are reported as an integral value in the range of 0 to 100. These quality scores have the following interpretation:

   - 0 to 25: Unacceptable - the biometric data cannot be used for the purpose specified by the application (bioapi_purpose). The biometric data must be replaced with a new sample.
   - 26 to 50: Marginal - the biometric data will provide poor performance for the purpose specified by the application and in most application

environments will compromise the intent of the application. the biometric data should be replaced with a new sample.

- 51 to 75: Adequate - the biometric data will provide good performance in most application environments based on the purpose specified by the application. The application should attempt to obtain higher quality data if the application developer anticipates demanding usage.

- 76 to 100: Excellent - the biometric data will provide good performance for the purpose specified by the application. The application may want to attempt to obtain better samples if the sample quality (bioapi_quality) is in the lower portion of the range (e.g., 76, 77, . . .) when convenient (e.g., during enrollment).

BioAPI states that the primary objective to include quality is to provide information on the suitability of the sample, i.e., the quality metric is used simply to decide to neglect a particular sample.

3. *e-Governance standards*: The Government of India has established biometric standards for identification and verification in various e-Governance applications [103]. These standards are largely based on the ISO/IEC 19794-5:2005 international best practices. While they are primarily designed for visual inspection, they can be improvised for future use as input to automatic systems. Further, these standards are being implemented for *Adhaar* project by the Unique Identification Authority of India (UIDAI) [104].

Biometric standardization is much needed in the community to ensure easy exchange of ideas and information, with the community still struggling with problems of interpretability. One reason could be that most standardization committees are closed grouped and are not available publicly.

### Competing interests
The authors declare that they have no competing interests.

### References
1. Z Wang, AC Bovik, Modern image quality assessment. **2**, 1–156 (2006)
2. W Lin, CCJ Kuo, Perceptual visual quality metrics: a survey. J. Vis. Comm. Image Represent. **22**(4), 297–312 (2011)
3. Y Chen, S Dass, A Jain, Fingerprint quality indices for predicting authentication performance, in *Audio and Video Based Biometric Person Authentication*. Lecture Notes in Computer Science, vol. 3546 (Springer, Berlin Heidelberg, 2005), pp. 160–170
4. P Grother, E Tabassi, Performance of biometric quality measures. EE Trans. Pattern Anal. Mach. Intell. **29**(4), 531–524 (2007)
5. R Youmaran, A Adler, Measuring biometric sample quality in terms of biometric information, in *Proceedings of Biometric Consortium* (Baltimore, Maryland, 19–21 September 2006), pp. 1–6
6. K Kryszczuk, J Richiardi, A Drygajlo, Impact of combining quality measures on biometric sample matching, in *Proceedings of IEEE International Conference on Biometrics: Theory, Applications, and Systems* (Washington DC, 28–30 September 2009), pp. 1–6
7. JR Beveridge, GH Givens, PJ Phillips, BA Draper, DS Bolme, YM Lui, FRVT 2006: quo vadis face quality. Image Vis. Comput. **28**(5), 732–743 (2010)
8. ND Kalka, J Zuo, N Schmid, B Cukic, Estimating and fusing quality factors for iris biometric images. IEEE Trans. Syst. Man Cybern. **40**(3), 509–524 (2010)
9. A Kumar, D Zhang, Improving biometric authentication performance from the user quality. IEEE Trans. Instrum. Meas. **59**(3), 730–735 (2010)
10. N Poh, J Kittler, A unified framework for multimodal biometric fusion incorporating quality measures. IEEE Trans. Pattern Anal. Mach. Intell. **34**(1), 3–18 (2011)
11. M-J Chen, AC Bovik, No-reference image blur assessment using multiscale gradient. EURASIP J. Image Video Process. **2011**(1), 1–11 (2011)
12. P Gastaldo, R Zunino, J Redi, Supporting visual quality assessment with machine learning. EURASIP J. Image Video Process. **2013**(1), 1–15 (2013)
13. RLV Hsu, J Shah, B Martin, Quality assessment of facial images, in *Proceedings of Biometric Consortium* (Baltimore, Maryland, 19–21 September 2006), pp. 1–6
14. C Tilton, The BioAPI specification, 2002. National Standards Institute (2002)
15. R Wong, N Poh, J Kittler, D Frohlich, Interactive quality-driven feedback for biometric systems, in *IEEE International Conference on Biometrics: Theory Applications and Systems* (Washington DC, USA, 27–29 Sept 2010), pp. 1–7
16. M Vatsa, R Singh, A Tiwari, S Bharadwaj, HS Bhatt, Analyzing fingerprint of Indian population using image quality: a UIDAI case study, in *Proceedings of International Workshop on Emerging Trends and Challenges in Hand-Based Biometrics* (Hong Kong, 26–29 September 2010), pp. 1–5
17. S Bharadwaj, H Bhatt, M Vatsa, R Singh, A Noore, Quality assessment based denoising to improve face recognition performance, in *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition Workshops* (Colorado Springs, 21–23 June 2011), pp. 140–145
18. M Vatsa, R Singh, A Ross, A Noore, Integrated multilevel image fusion and match score fusion of visible and infrared face images for robust face recognition. Pattern Recogn. - Special Issue on Multimodal Biometrics **41**(3), 880–893 (2008)
19. M Vatsa, R Singh, A Ross, A Noore, Quality-based fusion for multichannel iris recognition, in *Proceedings of International Conference on Pattern Recognition* (Japan, 11–15 November 2010), pp. 1314–1317
20. 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, AA Salah, T Scheidat, C Vielhauer, Benchmarking quality-dependent and cost-sensitive score-level multimodal biometric fusion algorithms. IEEE Trans. Inf. Forensics Secur. **4**(4), 849–866 (2009)
21. K Kryszczuk, A Drygajlo, Credence estimation and error prediction in biometric identity verification. Signal Process. **88**(4), 916–925 (2008)
22. K Nandakumar, Y Chen, AK Jain, SC Dass, Quality-based score level fusion in multibiometric systems, in *Proceedings of IEEE International Conference on Pattern Recognition* (Hong Kong, 20–24 August 2006), pp. 473–476
23. M Vatsa, R Singh, A Noore, A Ross, On the dynamic selection of biometric fusion algorithms. IEEE Trans. Inf. Forensics Secur. **5**(3), 470–479 (2010)
24. H Al-Assam, A Abboud, S Jassim, Hidden assumption of face recognition evaluation under different quality conditions, in *Proceedings of International Conference on Information Society* (London, UK, 27–29 June 2011), pp. 27–32
25. H Sellahewa, SA Jassim, Image-quality-based adaptive face recognition. IEEE Trans. Instrum. Meas. **59**(4), 805–813 (2010)
26. H Bhatt, S Bharadwaj, M Vatsa, R Singh, A Ross, A Noore, A framework of quality-based biometric classifier selection, in *Proceedings of IEEE/IAPR International Joint Conference on Biometrics* (Washington DC, 11–13 October 2011), pp. 1–7

27. S Paul, D Gupta, A Tiwari, Indexed search strategy for an automated biometric identification system, in *Proceedings of the International Conference of the Biometrics Special Interest Group* (Darmstadt, Germany, 6–7 Sept 2012), pp. 1–6

28. F Alonso-Fernandez, J Fierrez, D Ramos, J Gonzalez-Rodriguez, Quality-based conditional processing in multi-biometrics: application to sensor interoperability. IEEE Trans. Syst. Man Cybern. **40**(6), 1168–1179 (2010)

29. J Daugman, Combining multiple biometrics. World Wide Web electronic publication (2010). http://www.cl.cam.ac.uk/~jgd1000/combine/combine.html. 10 May 2014

30. A Abaza, A Ross, Proceedings of IEEE International Conference on Biometrics: Theory, Applications, and Systems (Italy, 2–5 June 2009), pp. 1–6

31. A Kumar, S Shekhar, Personal identification using multibiometrics rank-level fusion. IEEE Trans. Syst. Man Cybern. **41**(5), 743–752 (2011)

32. R Singh, M Vatsa, A Ross, A Noore, Biometric classifier update using online learning: a case study in near infrared face verification. Image Vis. Comput. **28**(7), 1098–1105 (2010)

33. HS Bhatt, S Bharadwaj, R Singh, M Vatsa, Noore A, A Ross, On co-training online biometric classifiers, in *Proceedings of IEEE/IAPR International Joint Conference on Biometrics* (Washington DC, 11–13 October 2011), pp. 1–7

34. B Keelan, *Handbook of Image Quality: Characterization and Prediction.* (CRC Press, USA, 2002)

35. F Alanso-Fernandez, Biometric quality assessment and its application in multimodal authentication systems. Thesis, Universidad Politécnica de Madrid, Madrid, 2007

36. SK Modi, SJ Elliott, Impact of image quality on performance: comparison of young and elderly fingerprints, in *Proceedings of International Conference on Recent Advances in Soft Computing* (United Kingdom, 10–12 July 2006), pp. 449–45

37. A Ross, R Nadgir, A thin-plate spline calibration model for fingerprint sensor interoperability. IEEE Trans. Knowl. Data Eng. **20**(8), 1097–1110 (2008)

38. E Murphy-Chutorian, MM Trivedi, Head pose estimation in computer vision: a survey. IEEE Trans. Pattern Anal. Mach. Intell. **31**(4), 607–626 (2009)

39. JR Beveridge, PJ Phillips, GH Givens, BA Draper, MN Teli, DS Bolme, When high-quality face images match poorly, in *Proceedings of IEEE Conference on Automatic Face Gesture Recognition and Workshops* (Santa Barbara, CA, March 21–25 2011), pp. 572–578

40. S Yendrikhovskij, Image quality: between science and fiction, in *Proceedings of Conference of Image Processing, Image Quality, Image Capture, Systems* (Georgia, 25–28 April 1999), pp. 173–178

41. A Mittal, AK Moorthy, AC Bovik, No-reference image quality assessment in the spatial domain. IEEE Trans. Image Process. **21**(12), 4695–4708 (2012)

42. F Alonso-Fernandez, J Fierrez, J Ortega-Garcia, J Gonzalez-Rodriguez, H Fronthaler, K Kollreider, J Bigun, A comparative study of fingerprint image-quality estimation methods. IEEE Trans. Inf. Forensics Secur. **2**(4), 734–743 (2007)

43. F Alonso-Fernandez, J Fierrez, J Ortega-Garcia, Quality measures in biometric systems. IEEE Secur. Privacy **10**(6), 52–62 (2012)

44. TP Chen, X Jiang, WY Yau, Fingerprint image quality analysis, in *Proceedings of IEEE International Conference on Image Processing, Singapore*, vol. 2, (24–27 October 2004), pp. 1253–1256

45. M Vatsa, R Singh, A Noore, MM Houck, Quality-augmented fusion of level-2 and level-3 fingerprint information using DSm theory. Int. J. Approximate Reasoning **50**(1), 51–61 (2009)

46. E Tabassi, CL Wilson, CI Watson. Fingerprint image quality (NISTIR 7151) (2004). ftp://sequoyah.nist.gov/pub/nist_internal_reports/ir_7151/ir_7151.pdf. 10 May 2014

47. H Fronthaler, K Kollreider, J Bigun, Automatic image quality assessment with application in biometrics, in *Proceedings of IEEE Computer Vision and Pattern Recognition Workshops* (New York, 17–22 June 2006), p. 30

48. E Lim, X Jiang, W Yau, Fingerprint quality and validity analysis, in *Proceedings of IEEE/IAPR International Conference on Image Processing* (Rochester, New York, 22–25 September 2002), pp. 469–472

49. L Shen, A Kot, W Koo, Quality measures of fingerprint images, in *Proceedings of Audio and Video Based Biometric Person Authentication* (Halmstad, Sweden, 6–8 June 2001), pp. 266–271

50. MA Olsen, H Xu, C Busch, Gabor filters as candidate quality measure for NFIQ 2.0, in *Proceedings of IAPR International Conference on Biometrics* (New Delhi, India, 29 March to 1 April 2012), pp. 158–163

51. O Barringer, E Tabassi, Fingerprint sample quality metric NFIQ 2.0, in *Proceedings of the International Conference of the Special Interest Group on Biometrics* (Darmstadt, Germany, 8–9 Sept 2011), pp. 167–171

52. K Phromsuthirak, V Areekul, Fingerprint quality assessment using frequency and orientation subbands of block-based Fourier transform, in *Proceedings of IAPR International Conference of Biometrics* (Madrid, Spain, 4–7 June 2013), pp. 1–7

53. MA Olsen, E Tabassi, A Makarov, C Busch, Self-organizing maps for fingerprint image quality assessment, in *Proceedings of IEEE International Conference of Computer Vision and Pattern Recognition Workshops* (Portland, 23–28 June 2013), pp. 1–7

54. RA Hicklin, J Buscaglia, MA Roberts, Assessing the clarity of friction ridge impressions. Forensic Sci. Int. **226**(1), 106–117 (2013)

55. S Yoon, K Cao, E Liu, AK Jain, LFIQ: Latent fingerprint image quality, in *Proceedings of IEEE International Conference on Biometrics: Theory, Applications, and Systems* (Washington DC, 29 September to 2 October 2013), p. 1

56. A Sankaran, M Vatsa, R Singh, Automated clarity and quality assessment for latent fingerprints: a preliminary study, in *Proceedings of IEEE/IAPR International Conference on Biometrics: Theory, Applications and Systems* (Washington DC, 29 September to 2 October 2013), pp. 1–7

57. Z Haddad, A Beghdadi, A Serir, A Mokraoui, Wave atoms based compression method for fingerprint images. Pattern Recognit. **46**(9), 2450–2464 (2013)

58. J Daugman, How iris recognition works. IEEE Trans. Circ. Syst. Video Tech. **14**(11), 21–30 (2004)

59. Y Chen, S Dass, A Jain, *Localized iris image quality using 2-D wavelets, Advances in Biometrics, Lecture Notes in Computer Science*, vol. 3832. (Springer, Berlin Heidelberg), pp. 373–381

60. Z Jinyu, NA Schmid, Global and local quality measures for NIR iris video, in *Proceedings of IEEE Computer Vision and Pattern Recognition Workshop* (Miami, Florida, 20–25 June 2009), pp. 120–125

61. H Proença, Quality assessment of degraded iris images acquired in the visible wavelength. IEEE Trans. Inf. Forensics Secur. **6**(1), 82–95 (2011)

62. A Abhyankar, S Schuckers, Iris quality assessment and bi-orthogonal wavelet based encoding for recognition. Pattern Recognit. **42**(9), 1878–1894 (2009)

63. S Crihalmeanu, A Ross, S Schuckers, L Hornak, *A protocol for multibiometric data acquisition, storage and dissemination*. (Technical Report, West Virginia University, 2007)

64. J Zuo, F Nicolo, NA Schmid, H Wechsler, Adaptive biometric authentication using nonlinear mappings on quality measures and verification scores, in *Proceedings of IEEE International Conference on Image Processing* (Hong Kong, 26–29 September 2010), pp. 4077–4080

65. J Zuo, NA Schmid, Adaptive quality-based performance prediction and boosting for iris authentication: methodology and its illustration. IEEE Trans. Inf. Forensics Secur. **8**(6), 1051–1060 (2013)

66. A Baig, A Bouridane, F Kurugollu, A novel modality independent score-level quality measure, in *International Symposium on Communication Systems Networks and Digital Signal Processing* (Manchester, UK, 23–25 July 2010), pp. 732-735

67. Y Du, C Belcher, Z Zhou, R Ives, Feature correlation evaluation approach for iris feature quality measure. Signal Process. **90**(4), 1176–1187 (2010)

68. F Alonso-Fernandez, J Bigun, Quality factors affecting iris segmentation and matching, in *Proceedings of IAPR International Conference of Biometrics* (Madrid, Spain, 4–7 June 2013), pp. 1–7

69. M Subasic, S Loncaric, T Petkovic, H Bogunovic, V Krivec, Face image validation system, in *Proceedings of International Symposium on Image and Signal Processing and Analysis* (Zagreb, Croatia, 15–17 September 2005), pp. 30–33

70. X Gao, S Li, R Liu, P Zhang, Standardization of face image sample quality, in *Proceedings of Advances in Biometrics*, (2007), pp. 242–251

71. T Ojala, M Pietikainen, T Maenpaa, Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans. Pattern Anal. Mach. Intell. **24**(7), 971–987 (2002)

72. G Zhang, Y Wang, Asymmetry-based quality assessment of face images, in *Advances in Visual Computing*, Las Vegas, vol. 5876 (Springer, Berlin Heidelberg, 2009), pp. 499–508

73. Y Wong, S Chen, S Mau, C Sanderson, BC Lovell, Patch-based probabilistic image quality assessment for face selection and improved video-based face recognition, in *Proceedings of IEEE International*

*Conference on Computer Vision and Pattern Recognition Workshops* (Colorado Springs, 21–23 June 2011), pp. 74–81

74. K Nasrollahi, T Moeslund, *Face quality assessment system in video sequences. Biometrics and identity management, Lecture Notes in Computer Science*, vol. 5372. (Springer, Berlin Heidelberg, 2008). pp. 10–18

75. J Long, S Li, Near infrared face image quality assessment system of video sequences, in *Proceedings of International Conference on Image and Graphics* (Hefei, China, 12–15 August 2011), pp. 275–279

76. Y Yao, B Abidi, M Abidi, *Quality assessment and restoration of face images in long range/high zoom video. Biometrics and Identity Management. Lecture Notes in Computer Science*, vol. 5372. (Springer, Berlin Heidelberg, 2007). 43–60

77. B Klare, AK Jain, Face recognition: impostor-based measures of uniqueness and quality, in *Proceedings of IEEE/IAPR International Conference on Biometrics: Theory, Applications and Systems* (Washington DC, 23–26 September 2012), pp. 23–26

78. DG Lowe, Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. **60**(2), 91–110 (2004)

79. PJ Phillips, JR Beveridge, BA Draper, G Givens, AJ O'Toole, DS Bolme, J Dunlop, YM Lui, H Sahibzada, S Weimer, An introduction to the good, the bad, and the ugly face recognition challenge problem, in *Proceedings of IEEE International Conference on Automatic Face Gesture Recognition and Workshops* (Santa Barbara, CA, 21–25 March 2011), pp. 346–353

80. PJ Phillips, WT Scruggs, AJ O'Toole, PJ Flynn, KW Bowyer, CL Schott, M Sharpe, FRVT 2006 and ICE 2006 large-scale experimental results. IEEE Trans.Pattern Anal. Mach. Intell. **32**(5), 831–846 (2010)

81. PJ Phillips, JR Beveridge, DS Bolme, BA Draper, GH Given, YM Lui, S Cheng, MN Teli, H Zhang, On the existence of face quality measures, in *Proceedings of IEEE International Conference on Biometrics: Theory, Applications and Systems* (Washington DC, 29 September to 2 October 2013), pp. 1–8

82. G Aggarwal, S Biswas, PJ Flynn, KW Bowyer, Predicting good, bad and ugly match pairs, in *Proceedings of IEEE Workshop on Applications of Computer Vision* (Breckenridge, Colorado, 9–11 January 2012), pp. 153–160

83. F Hua, P Johnson, N Sazonova, P Lopez-Meyer, S Schuckers, Impact of out-of-focus blur on face recognition performance based on modular transfer function, in *Proceedings of IEEE/IAPR International Conference on Biometrics* (New Delhi, India, 29 March to 1 April 2012), pp. 85–90

84. C Shaokang, M Sandra, TH Mehrtash, S Conrad, B Abbas, CL Brian, Face recognition from still images to video sequences: a local-feature-based framework. EURASIPJ. Image Video Process. **2011** (2011)

85. JR Beveridge, GH Givens, PJ Phillips, BA Draper, DS Bolme, YM Lui, Focus on quality, predicting FRVT 2006 performance, in *Proceedings of IEEE International Conference on Automatic Face Gesture Recognition* (Amsterdam, Netherlands, 17–19 September 2008), pp. 1–8

86. PJ Grother, GW Quinn, PJ Phillips, Report on the evaluation of 2D still-image face recognition algorithms. National Institute of Standards and Technology (2010)

87. S Bharadwaj, M Vatsa, R Singh, Can holistic representations be used for face biometric quality assessment?, in *Proceedings of IEEE International Conference on Image Processing* (Melbourne, Australia, 15–18 September 2013), pp. 1–7

88. H Choi, C Lee, No-reference image quality metric based on image classification. EURASIP J. Adv. Signal Process. **2011**(1), 1–11 (2011)

89. L Breitenbach, P Chawdhry, Image quality assessment and performance evaluation for multimodal biometric recognition using face and iris, in *Proceedings of International Symposium on Image and Signal Processing and Analysis* (Salzburg, Austria, 16–18 Sept 2009), pp. 550–555

90. J Changlong, K Hakil, X Cui, E Park, J Kim, J Hwang, S Elliott, Comparative assessment of fingerprint sample quality measures based on minutiae-based matching performance, in *International Symposium on Electronic Commerce and Security* (Sanya, China, 22–24 May 2009), pp. 309–313

91. NIST. http://www.nist.gov/itl/iad/ig/nbis.cfm. 10 May 2014

92. http://www.neurotechnology.com. 10 May 2014

93. N Nill, B Bouzas, Objective image quality measure derived from digital image power spectra. Opt. Eng. **31**(4), 813–825 (1992)

94. P Marziliano, F Dufaux, S Winkler, T Ebrahimi, A no-reference perceptual blur metric, in *Proceedings of IEEE International Conference on Image Processing, Singapore*, (Oct. 24–27 2004), pp. 57–60

95. Z Wang, AC Bovik, HR Sheikh, EP Simoncelli, Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. **13**(4), 600–612 (2004)

96. OpenCV Library. http://opencv.org/. 10 May 2014

97. X Li, Z Sun, T Tan, Predict and improve iris recognition performance based on pairwise image quality assessment, in *Proceedings of International Conference of Biometrics* (Madrid, Spain, 4–7 June 2013), pp. 1–8

98. H Iwama, M Okumura, Y Makihara, Y Yagi, The OU-ISIR gait database comprising the large population dataset and performance evaluation of gait recognition. IEEE Trans. Inf. Forensics Secur. **7**(5), 1511–1521 (2012)

99. PJ Phillips, JR Beveridge, An introduction to biometric-completeness: the equivalence of matching and quality, in *Proceedings of IEEE International Conference on Biometrics: Theory, Applications and Systems* (Washington DC, 28–30 September 2009), pp. 414–418

100. Z Wang, AC Bovik, HR Sheikh, EP Simoncelli, Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. **13**(4), 600–612 (2004)

101. X Gao, W Lu, D Tao, X Li, Image quality assessment based on multiscale geometric analysis. IEEE Trans. Image Process. **18**(7), 1409–1423 (2009)

102. What is CBEFF (Common Biometric Exchange Formats Framework)? http://www.ibia.org/cbeff/#what_is_cbeff (2004)

103. Department of Information Technology: Face image data standard for e-governance applications in India. Government of India (2005)

104. Aadhaar project. Government of India. http://uidai.gov.in/.