

# Biometric Quality: From Assessment to Multibiometrics

Samarth Bharadwaj  
Mayank Vatsa and Richa Singh

## Abstract

Quality is an attribute or a property of an item that quantitatively measures specific aspect or content. The definition and correct method of measurement of quality of a biometric modality that is usually represented by an image is currently unclear in the research community. While a biometric image's quality is susceptible to degradation during capture and storage, it may also have low quality by its very nature. Quality of a biometric has several applications in popular research interests such as i) unconstrained biometric recognition ii) multibiometrics and iii) large-scale identity projects.

This research aims to define and demystify quality in the field of biometrics. We present a comprehensive survey of current advancements in quality assessment, starting with a concise summary of the field of *Biometrics* and recent advances and applicability of quality in *multibiometrics*. In order to understand quality assessment in biometrics, we delve into related area of image quality assessment. Further, several applications and factors that influence biometric quality are analyzed. We also investigate popular methods of evaluating quality assessment algorithms in biometrics. Finally, we explore quality in face recognition, an area that is yet to receive proportionate attention from the research community. The complexity of the problem is multiplied by the lack of consensus in literature on the definition and constitution of facial features. However, initial experiments indicate that holistic image descriptors are able to successfully encode degradations in biometric images.

## CONTENTS

<b>I</b>	<b>Introduction</b>	2
<b>II</b>	<b>Biometrics</b>	2
II-A	Recognition Pipeline . . . . .	2
II-B	Evaluation of Biometric Systems . . . . .	4
II-C	Characteristics of a Good Biometrics . . . . .	4
II-D	Fingerprint Biometrics . . . . .	5
II-E	Iris Biometrics . . . . .	6
II-F	Face Biometrics . . . . .	6
II-G	Biometric Grand Challenges . . . . .	7
<b>III</b>	<b>Multibiometrics</b>	8
III-A	Types of Multibiometric Frameworks . . . . .	8
III-B	Does Fusion Always Work? . . . . .	11
III-C	Adaptive Multibiometrics . . . . .	11
<b>IV</b>	<b>Biometric Quality</b>	12
IV-A	Image Quality Assessment . . . . .	12
IV-B	Quality of a Biometric Sample . . . . .	15
IV-C	Why is Quality Assessment Important? . . . . .	15
IV-D	Factors that Influence Biometric Quality . . . . .	16
IV-E	Evaluating Quality Assessment Techniques . . . . .	16
IV-F	Humans vs. Automatic Quality Assessment . . . . .	18
IV-G	Fingerprint Quality Assessment . . . . .	18
IV-H	Iris Quality Assessment . . . . .	19
IV-I	Face Quality Assessment . . . . .	21
IV-J	Matchscores Derived Quality . . . . .	23
<b>V</b>	<b>Novel Quality of Face Biometric</b>	23
V-A	New Insights on Face Quality . . . . .	24
V-B	Quality Framework for Biometrics . . . . .	24
V-C	Holistic Image Representations . . . . .	25
V-D	Experiments with Noisy Images . . . . .	25
V-E	Experiments with Pose . . . . .	26
<b>VI</b>	<b>Conclusion</b>	26
	<b>Appendix: Biometric Standards</b>	27
	<b>References</b>	28

## I. INTRODUCTION

UNIQUE identification of individuals is necessary in several applications of forensics, government/civil, commerce and entertainment domains; some examples are listed in Table I. Government agencies frequently perform verification of identity before providing certain information or resources, usually during entitlement programs, immigration, voter registration or welfare schemes. It is also critical to establish correct identity to ensure information security in personal device logins, network security, secure documents or databases, and medical records. Similarly, proof of identity is required in many commercial establishments such as banks, hospitals and shops. Entertainment industry also verifies user identity to provide user-based content and virtual or *augmented* reality. Biometrics, as part of *Identification Science*, is the measurement of certain key features of physical or behavioral traits to uniquely identify a person. By intelligently integrating biometrics with correct policies and procedures, it can be deployed for large scale country-wide applications to potentially eliminate the need for paper work or ID cards. It can be expected that coming years will witness several functioning, robust, and reliable large scale biometrics systems in deployment in the lines of today's U.K. IRIS project (Iris Recognition immigration System) and UAE iris based airport security system. This paper presents a review of *Biometrics* with emphasis on *Multibiometrics* and *Biometric quality*. Several modalities have emerged in biometrics in the last decade, however, three prominent modalities, namely, fingerprint, iris and face are discussed here. Further, an important aspect of biometrics and one that needs more attention from the community, is the measurement of *quality* of a biometric sample, more so in face biometrics, is discussed. In our opinion, quality of a biometric is beyond measuring quality of the image itself. While a sample's quality is susceptible to corruption during capture and storage, it may also have low quality by its very nature. With this view, we discuss a framework that categorizes existing quality measurement algorithms or techniques into perspective. It is our assertion that this will provide a coherent presentation of the current state of the art and the gaps in knowledge that need to be filled. The paper is organized as: Section II presents an overview of *biometrics*, including a brief description of three important modalities; fingerprint, iris, and face; characteristics of a biometric, and the findings of biometric grand challenges. Next, Section III presents an overview of multibiometrics which is an important applicator of biometric quality assessment. Section IV discusses quality assessment of images and biometrics with a review of recent literature in biometric quality assessment. Finally, a quality framework and our initial experimental analysis are discussed in Section V.

## II. BIOMETRICS

Biometrics has existed for over 100 years and has been used in aiding criminal investigation, identification of disaster victims and location of missing persons. Using biometrics for person identification rather than documents (passport, voter cards, ID cards) ensures against fraud by forgery or ambiguity due to lost

documents<sup>1</sup>. The advantage of biometrics is nicely summarized by Jain et al. [2] as the determination of identity based on *who she is* rather than *something she possess* (i.e. documents, smart cards) or *something she remembers* (i.e. passwords). With rapid growth in computation and development in fields such as image processing, sensor technology and pattern recognition; automatic capture, detection and measurement of biometrics has emerged into a distinct field in computer science. Advances in computing hardware capabilities also played a vital role in improvement in recognition rates, as greater number of features can now be compared in reasonable time. This continued growth has also brought several biometric technologies that are able to satisfy the characteristics put forth by [2]. While the popular biometric traits in literature continue to be fingerprint, face and iris biometrics, other human traits (both physical and psychological) have also become prevalent as biometrics. These include ear, gait, hand and finger geometry, palm print, retinal scan, voice, signature and keystrokes. A detailed discussion of these modalities can be found in [3].

### A. Recognition Pipeline

Fig. 3 illustrates the recognition pipeline of a generic biometric system. The process is inspired by human visual system which hierarchically extracts and processes data starting from coarse to dense features. The first step in a typical biometric system involves segmentation of region of interest from the input image. Next, discriminating features are extracted and matched with a stored template whose identity is known. The matching processes then produces a decision. A brief description of each of the modules is presented here, with help of existing nomenclature [4]:

- **Sensor Module:** A biometric sensor is required to capture the trait (as image) and feed to the recognition system in suitable format. Often the usability and applicability of the entire system is dependent on the quality of the sensor. A well designed sensor is one that is able to minimize failure-to-acquire rate and noise. Modern sensors are equipped with active quality feedback systems that improve data quality by user interaction.
- **Segmentation and Pre-Processing:** The sensor output is generally processed with relevant segmentation methods. This is an important step in making the system usable, for example, a camera can be placed in a corridor which captures the entire cross-sectional view at an instance; a face detection algorithm then segments the face area in the image to feed to a face recognition system. The segmented region is usually processed by an enhancement technique that aids in recognition.
- **Feature Extraction and Matching:** The next process involves extraction of certain salient features that uniquely identify the individual based on a biometric trait. For example, the position and orientation of minutia points in fingerprints. The extracted features are compared against

<sup>1</sup>Though biometric forgery is possible (using face masks, special contact lens, gummy fingers, practiced signatures or gait) there is a considerably higher cost associated than document fraud.



Fig. 1. It is expected that coming years will witness several functioning, robust and reliable large scale biometrics systems in deployment much like today's U.K.'s IRIS project (Iris Recognition immigration System), UAE's iris based airport security system, India's UID project and, US-VISIT program.(Images obtained from the internet)

TABLE I  
BIOMETRICS IS USEFUL IN SEVERAL PERSON IDENTIFICATION APPLICATIONS [1]

Forensics	Government or Civil	Commercial & Others
Corpse identification	National ID	ATM
Crime investigation	Licenses	Access controls
Parenthood determination	Welfare	E-commerce
Missing children	Border Security	Internet

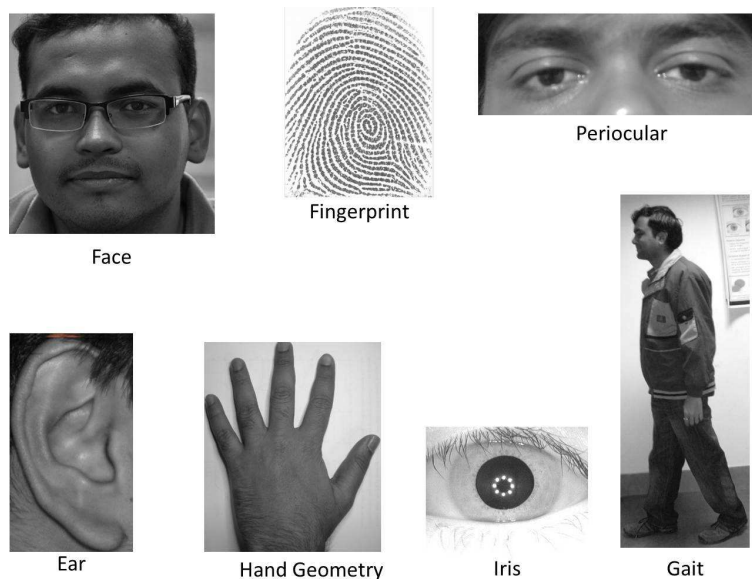


Fig. 2. Several modalities have emerged in biometrics in the last decade, however this paper chiefly discusses three prominent modalities: fingerprint, iris and face.

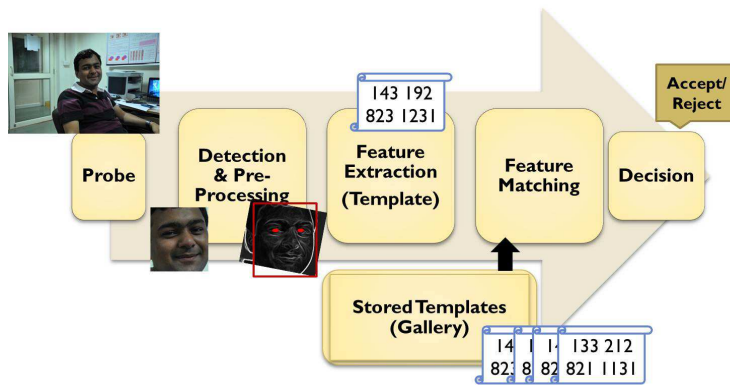


Fig. 3. Pipeline of a typical biometric system consists of a capture sequence (probe), detection & pre-processing, feature extraction, matching and decision modules.

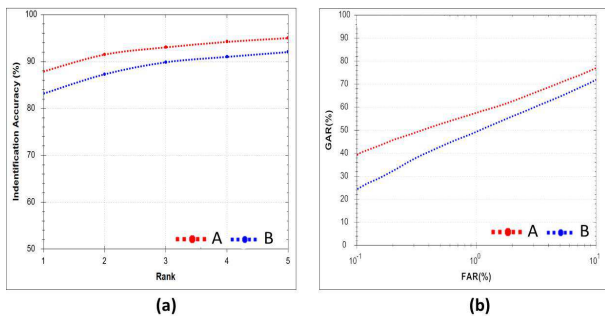


Fig. 4. (a) CMC curve, and (b) ROC curve (plotted in log-scale for better visualization of low FAR values). Recognition performance of 'A' is better than 'B'.

the stored template (whose identity is known or verified). A recognition decision is then rendered based on the application of the system.

The stored templates are referred to herewith as *gallery* and the input query is referred to as *probe*. Further, most systems present the result of matching via distance or similarity scores. Unlike matching passwords or keys, where only a *perfect* match confirms identity, biometric traits almost never match exactly. Hence, obtaining a match or non-match decision on the basis of a matchscore requires a more intricate understanding of the properties behavior of features and the distribution of matchscores.

### B. Evaluation of Biometric Systems

A biometric system is a pattern recognition system that uses biometric data as input for classification or recognition. Generally, a user must undergo an *enrollment* process in order to create the stored template. In most cases, a biometric system is used in two modes, *verification* (1:1 matching) and *identification* (1:N matching). As depicted in Fig. 5, verification is a confirmation of the identity of an individual; it is verified that the person is actually who he claims to be. For example, *Is this person Samarth?* This requires a single comparison of two templates, the stored template of the claimed identity with the query template. A biometric system with high usability is able to perform such a match instantly. In order to test the

performance of a biometric system in verification mode, it is considered a good practice to showcase a Receiver Operator characteristic curve (ROC). This is obtained by computing the following rates over a span of the discriminating threshold.

- **Genuine Accept Rate (GAR)** (or true positive rate), the percentage of correctly verified matches to the total number of correct matches
- **False Accept Rate (FAR)** (or false positive rate), the percentage of falsely accepted matches to the total number of non-matches,

A biometric system can also be used in identification mode. Identification is the process of selecting a template in the gallery that most closely resembles the probe template. This mode answers the question *who is this person?* or *Is he in the gallery set?* Use-cases include missing persons, terrorist watchlists or de-duplication of the biometric database. To present the performance of the system in identification mode, it is useful to construct a Cumulative Match Characteristic (CMC) curve; a cumulative graph of the percentage of correct matches at a particular rank order. Fig. 4 illustrates an example of CMC and ROC graphs. Herewith, the word recognition is used as a generic term to refer to either identification or verification.

### C. Characteristics of a Good Biometrics

Jain et al. [2] discuss certain fundamental properties and also some practical considerations of biometric traits and biometric systems. These properties are briefly summarized below:

- 1) **Universality:** A characteristic of all/most individuals.
- 2) **Distinctiveness:** Possess a sufficiently unique characteristic.
- 3) **Permanence:** Sufficiently invariant over a period of time.
- 4) **Collectivity:** Measurable and easily collectable characteristic of an individual.
- 5) **Performance:** Accuracy and speed of recognition, required resources operational and environmental factors that effect performance.
- 6) **Acceptability:** Social and personal acceptance of the system in society.

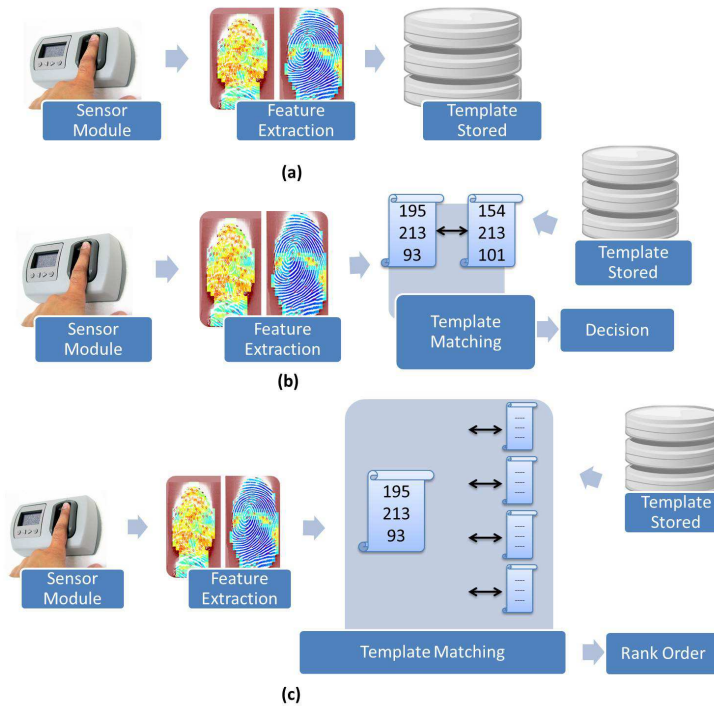


Fig. 5. Biometric system - (a) enrollment: A identity verified template is stored, recognition can be done in two modes, (b) verification (1:1) matching of templates, or (c) Identification (1:N) matching of templates, from the stored database



Fig. 6. Sample images of fingerprints with varying quality. Images obtained from various public datasets.

#### 7) **Circumvention:** No/limited spoofing or fraudulent methods of fooling the system.

Different biometric modalities exhibit advantages and disadvantages when compared based on these characteristics. Hence, there is no *ideal* biometric but several *admissible* biometric traits [3]. Therefore, relevance of a biometric modality is always derived from the intent, certain characteristics may become more important than others. Next, a brief description of biometric modalities viz. fingerprint, iris, and face, in context of these biometric characteristics is presented. A summary is presented in Table II.

#### D. Fingerprint Biometrics

Historically, fingerprints are known to have uniquely identifiable properties for person recognition. As shown in Fig. 6, a fingerprint consists of pattern of ridges and furrows; additionally sweat pores and other *extended features* may also be visible at high resolution of capture. Several such features exist in a single fingerprint, hence the unlikeliness

of them having same position and orientation for fingerprints of different individuals, makes it a good biometric trait [5]. Jain et al. [6] present an intuition of the biology behind the creation of fingerprints, suggesting that the distribution of these features is a random biological process and unlikely to be duplicated (high *Distinctiveness*). A hierarchical ordering of features, namely Level 1, Level 2 and Level 3 features, prevalent in literature, is presented.

- 1) Level 1 features are one of three pattern - (left or right) loop, whorl and arch, which constitute 60 – 65%, 30 – 35% and 5% respectively of all fingerprints [7]. These features are a key component in fingerprint sub-categorization tasks in indexing schemes.
- 2) Level 2 features are the uniquely identifiable patterns known as *minutiae*. These points are local characteristics of a ridge, either a ridge ending or a ridge bifurcation. Minutiae are generally considered to be the most robust features in a fingerprint and are used in both automated and manual examination of fingerprints.
- 3) Level 3 features, include sweat pores, ridge contour and incipient ridges. These features are usually referred to as *extended feature* set and manifest in fingerprint images collected at a higher resolution (1000ppi).

Since ridge patterns, similar to fingerprint, can be obtained from palm and soles of feet, they may also be referred to as epidermal ridges. From the forensics point of view, those fingerprints that have been lifted from a crime scene using special equipment are known as *latent* fingerprints. These fingerprints present some difficult challenges due to the intense noise, blurring, smudging and presence of spurious minutia points. In latent prints, minutiae of are usually annotated

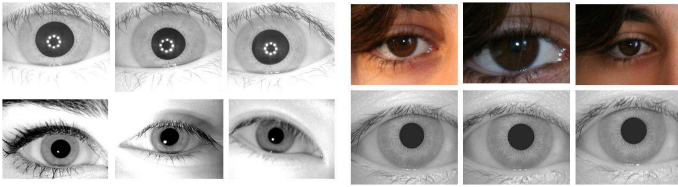


Fig. 7. Biometric samples of iris with different presentations and varying quality due to inherent or controllable traits. Images obtained from various public datasets.

manually by experts as the problem is deemed to difficult to be solved automatically. However, in recent literature several fully automatic and semi-automatic work around have been proposed that are both encouraging and insightful [8].

Some drawbacks of early fingerprint system prevented its use as a large scale biometric modality. One important factor was the difficulty in capture (*Collectability*). Wrinkled fingerprints (due to soaking), fingerprints with excessive scars (natural or otherwise) or lack of fingerprint impressions (exhibited predominantly with rural farmers, factory workers or rarely due to disease) present limitations of fingerprint technology in terms of *Universality* and *Permanence*. Recent advancements in sensors have lead to most sensors being equipped with real-time quality feedback to actively engage user in the capture process. Impression free sensor surfaces do not accumulate moisture and help avoid the cumbersome sensor cleaning processes. In recent literature, matching of fingerprint from new capture sensors are introduced, such as touchless sensors [9] and 3D-sensors [10]. However, these sensors present new challenges attributed to 3D reconstruction such as high computation requirement and cost of deployment.

Due to its history, the use of fingerprints for identification is well known in society, and entails suspension due to extensive use in criminal investigation and forensics (*low Acceptability*). On the other hand, a fingerprint sensor is relatively cheap (\$50-100) and requires reasonable computational resources when used in small and medium scales. However, cost of deployment increases substantially for large scale (nation-wide) implementation (*low Performance*). Further, fingerprint technology will always require to have a point of contact and active user involvement, as apposed to other contactless biometrics such as face biometric. Several large scale forensic and civil applications of fingerprint biometrics exist. The United States Visitor and Immigrant Status Indicator Technology (US-VISIT) captures 10 fingerprints at immigration and boarder posts for identification with terrorist watchlist (negative identification) and illegal immigration. In 2007, the system possessed over 45 million records for matching.

### E. Iris Biometrics

Iris is a muscle tissue in the eye that controls the amount of light reaching the retina. Biometric features of the iris are muscle patterns obtained from its surface that remain extremely stable over long period of time (*Universality*). Research in iris biometric began in 1985 by Flom and Safir. In 1993, Daugman [11] proposed an iris recognition algorithm, including segmentation (using integro-differential operators),

feature extraction and encoding (based on 2D- Log Gabor filters), restoration (Rubber sheet model to counteract pupil dilation) and a matching technique (using Hamming distance). Iris is captured in the infrared spectrum (750 nm wavelength) primarily to mitigate the effects of specular reflectance from moisture on the surface of the eye. The speed of matching and high dimensionality of the iris patterns makes iris biometric ideal for large scale one-to-many matching, a key advantage over other biometrics (*high Performance*). The patterns of iris present as a melancholily of texture, shown in Fig. 7. This is apposed to the well-formed ridge patterns of fingerprints. Many concerns about the *Distinctiveness* and *Permanence* of these random patterns as a biometric where put to rest with a large scale empirical study by [12]. This study consists of iris patterns captured from 632, 500 people from 152 nationalities. The total number of comparisons exceeds 200 billion. The findings indicate that the distribution of the distance scores have rapidly diminishing tails. A correct selection of matching threshold can ensure that the chance of false accept is very small even for very large datasets.

A major drawback of iris biometric continues to be the lack of usability of capture sensors. The accuracy of iris biometric is highly dependent on the amount of control over ambient illumination conditions. In order to obtain an iris image with sufficient iris texture and correct lighting, the user must interact with a controlled near-infrared sensor. Further, individuals with bushy eyelashes or puffed eyes find it hard to enroll into an iris system, effecting its *Collectivity* and *Acceptability*.

Iris biometric has shown unprecedented recognition performance on large datasets. Even identical twins or eyes of the same person do not have same iris patterns. These patterns only change with surgery and are rarely affected by glasses or contact lens. However, appearance of iris texture varies dramatically with intensity and direction of illumination. Some large-scale deployments include United Arab Emirates Iris-Guard's Homeland Security Border Control in 2001, which has apprehended 333,000 individuals to date for illegal entry to UAE. Recent literature in iris biometrics focuses on improving recognition rates with non-ideal iris imagery. While newer iris capture systems are able to capture iris texture from large distances (4m to 9m) [13], the tend to introduce noise, illumination and pose, significantly deteriorating performance of state-of-the-art recognition systems.

### F. Face Biometrics

The intuition of machines capable of vision has existed in popular culture and has been an important pursuit for researchers for decades. Face recognition is a particularly fascinating aspect of this pursuit, considering the superior performance that human beings are able to exhibits even in very non-ideal circumstances. Several cognitive and psychological studies are only beginning to analyze and replicate features used by the human brain in recognizing faces [14], [15], [16], [17]. This interest towards face recognition is reflected in the number of computer vision venues that continue to showcase research on face biometrics. Face recognition has received a huge thrust



Fig. 8. face images with different presentations and varying quality. Images obtained from FRGC dataset.

in recent years, as face biometric systems began exhibiting performance on par with other more established biometrics. Face is a non-intrusive and intuitive biometric and presents some advantages in terms of convenience of use and a general sense of *Acceptability* in comparison to fingerprint and iris biometrics. Two important stages of a face recognition system are *face detection* and *face recognition*.

- Face detection is the segmentation of the face area from the background. The performance of face detection is an important aspect of the usability of this biometric (Usability being a strong proponent in favor of this biometric). One of the leading techniques utilized for face detection is a Haar feature based adaptive boosting cascades [18]. A detailed survey of face detection, including evaluation metrics, databases and benchmarks are provided in [19]. It is widely accepted that *real-time* face detection and *tracking* in controlled indoor settings is relatively a solved problem [20]. However, the same can not be said with data captured in outdoor, non-ideal settings (such as CCTV surveillance).
- Face Recognition involves feature extraction and matching of the normalized detected face to ascertain identity. The performance of this aspect is dependent on the features extracted and the classification framework used. In normal indoor conditions with a cooperative users (neutral expression and good illumination), research has shown that automatic algorithms are beginning to outperform humans when the faces are not familiar [16]. Unlike fingerprint and iris, there is no consensus in literature on the actual matchable features of face images. Early face recognition systems used geometric distances between prominent feature such as eyes and mouth. Later research focus has shifted towards the use of *appearance* and *texture* based features for recognition. Taking advantage of high computation power available in today's computers, it is not uncommon for face biometric algorithms to have more than 4000 matchable features. A detailed discussion of several of these algorithms can be

TABLE II  
A COMPARISON OF THE CHARACTERISTICS OF FINGERPRINT, IRIS AND FACE MODALITIES. (HIGH, LOW AND MEDIUM ARE DENOTED AS  $\uparrow$ ,  $\downarrow$  AND  $\sim$ ).

Characteristic	Fingerprint	Iris	Face
Universality	$\sim$	$\uparrow$	$\uparrow$
Distinctiveness	$\uparrow$	$\uparrow$	$\downarrow$
Permanence	$\uparrow$	$\uparrow$	$\sim$
Collectivity	$\downarrow$	$\downarrow$	$\uparrow$
Performance	$\sim$	$\uparrow$	$\downarrow$
Acceptability	$\downarrow$	$\downarrow$	$\uparrow$
Circumvention (ease of)	$\sim$	$\downarrow$	$\uparrow$

found in [1].

In recent years, several researchers have shown that using multi-spectral images of face can improve accuracy in non-ideal illumination. A review of these techniques is presented in [21]. Similarly, 3D imagery for face recognition has also gained immense popularity in the research community, particularly with the advent of more accurate and cheap 3D sensors. Abate et al. [22] present a survey of 3D face recognition techniques. A further comments on current and future directions in this area of face recognition is presented in [23].

Face biometrics is still plagued by many covariates that drastically effect the recognition performance, shown in Fig. 8. The chief covariates of face recognition include illumination, pose, expression, short and long-term alterations of face (aging or disguise). A system can also be presented with plausible limitations such as twins, look-alikes, or persons who have undergone facial plastic surgery. Current literature contains several solutions to tackle these covariates, however fully reliable face recognition systems in applications like airport security are still very challenging [20].

### G. Biometric Grand Challenges

In an effort to accelerate advancements in biometrics, the government of United States (via NIST) establish several initiatives such as the *Face Recognition Technology (FERET) Evaluation 1994-96*, *Fingerprint Vendor Test and Evaluation (FpVTE) 2003*, *Face Recognition Grand Challenge (FRGC) 2004*, *Face Recognition Vendor Test (FRVT) 2000-06* and, *Iris Challenge Evaluation (ICE) 2005-06*. Through the availability of significant testing samples and challenging baseline experiments, these initiatives are designed to help spur research and engineering in the field of biometrics. Table III presents a brief summary of different versions of these challenges conducted over the years. A comprehensive analysis of *FRVT & ICE 2006* is presented by Phillips et al. [24]. Recently, these initiatives culminated to form the *Multi-biometrics Grand Challenge (MBGC) v1(2008) & v2(2009)*. This challenges is based on recognition of an individual using multiple biometric features as he walks through a portal or in uncontrolled outdoor environment (few face examples are shown in Fig. 9). A brief description of the test conditions and state-of-the-art recognition rates are discussed in Table IV.



Fig. 9. Biometric samples of face from the challenging MBGC v2.

TABLE III  
SUMMARY OF BIOMETRIC GRAND CHALLENGES

Challenge	Modality	Test Conditions	False Reject Rate % (FRR)	False Accept Rate % (FAR)
FERET 1996	Face	Illumination, expression	0.54	0.001
FRVT 2002	Face	Illumination, Outdoor/indoor	0.01	0.001
FpVTE 2003	Fingerprint	US govt, operational data	0.001	0.01
FVC 2004	Fingerprint	Skin distortions, rotation	0.2	0.2
FRVT 2006	Face	Multi-resolution, illumination,	0.01	0.001
ICE 2006	Iris	Bad quality, non-ideal imagery	0.01	0.001

TABLE IV  
PRELIMINARY RESULTS OF MBGC v2

Experiment	No. of samples	False Reject Rate % (FRR)	False Accept Rate % (FAR)
Still face (controlled) vs. HD video	1426	0.01	0.99
Still face (uncontrolled) vs. HD video	1785	0.01	0.95
Still Iris vs. NIR video	4025	0.01	0.90
<b>Fusion</b>			
Still face (controlled) and still iris	5451	0.01	0.99
Still face (uncontrolled) and still iris	5810	0.01	0.95

### III. MULTIBIOMETRICS

Uni-modal biometric systems are constrained by the limitations of the modality which maybe in terms of accuracy or usability. These constraints may hamper the use of biometric systems in security or law enforcement applications. An important direction of research and one that has obtained considerable focus in literature is the study of utilizing (*fusing*) evidence of identity from multiple biometrics, referred to as *multibiometrics*. Such systems offer additional benefits over uni-modal counterpart such as resiliency to sensor malfunction or spoofing, universality, greater resilience to noise, fault

tolerance, and improved accuracy [4].

#### A. Types of Multibiometric Frameworks

Depending on the data or process used, multibiometric approaches in the field of biometrics are divided into *categories* and *levels* respectively [4]. The five categories are described below along with a brief description of some representative literature (summarized in Table VII).

- **Multi-Sensor:** A single biometric trait is captured using multiple sensors that capture diverse information from the same presentation. For example, face is captured using a

2D camera and a 3D range sensor to perform authentication. Chen et al. [25] fuse face images obtained from thermal infrared and visible light camera by integrating evidence (at score and rank level) to improve matching performance. Marcialis et al. [26] use information from an optical and a capacitive fingerprint sensors in a fusion framework to improve recognition.

- **Multi-Sample:** Multiple images of the same modality can help improve accuracy of a biometric system. For example, multiple images (frames) of a person’s face obtained in a video sequence. This allows for more data to recognize that individual. Certain use cases of a biometric system are able to obtain such input; for example, automatic face recognition system installed over CCTV cameras are able to provide multiple frames of a single individual. In particular, Liu and Chen [27] present a video to video matching algorithm to match faces. They use an adaptive Hidden Markov Model (HMM) to combine features and match frames from the gallery and probe videos streams.
- **Multi-Instance:** Different instances of the same biometric can also be beneficial, for example, fingerprints from different fingers or iris scans from both eyes of the same person. It has been observed that the features obtained from either irides or different fingers are largely uncorrelated with one another and may provide more evidence of identity [4].
- **Multi-Modal:** Different modalities captured from the same individual can provide several layers of evidence of identity. Further, use of physically uncorrelated traits such as face and gait, is expected to improve recognition performance more than correlated traits, such as face and periocular. Further, new multi-modality biometric sensors capture different biometric modalities, such face and iris, in a single instance.
- **Multi-Algorithmic:** It is often observed that recognition algorithms that inspect different aspects of the same modality, for example texture and shape of a face, have low correlation in their matchscore/decision output. Hence, fusing them can improve overall recognition performance. This category of fusion is popular in matching periocular, fingerprints and face biometrics.

As shown in Fig. 3, there are several key stages of a biometric system. Fusion of biometric classifiers can take place at any of these key stages. In literature, these are defined to as levels of fusion. Often a fundamental issue in information fusion systems is in determining the type of information that is to be fused. The amount of available information for fusion reduces as one moves along the recognition pipeline. In biometrics, levels of fusion are described as follows:

- **Data Level:** For the categories of multi-sample or multi-instance or multi-sensor fusion, the biometric samples can be combined into one meaningful representation. For example, face images obtained from different uncorrelated or orthogonal bands of the spectrum can be fused, as demonstrated by Vatsa et al. [28]. Ratha et al. [29] propose a mosaicing scheme to integrate multiple

snapshots of fingerprints as the user rolls or slides a finger over a sensor. Yang et al. [30] propose combining multiple instances of face captured simultaneously with five cameras to create a panoramic mosaic of the face. Singh et al. [31] use multi-resolution spline technique [32] to correctly stitch and blend two separate instances of a face.

- **Feature Level:** Multi-algorithm feature level fusion, involves combining features obtained from different sources into a single template. The most prevalent method is by concatenation or averaging, combined with feature selection [33] or dimensionality reduction [34], [35], [36] to avoid performance degradation due to the curse of dimensionality. Ross et al. [?] use geometrical registration to align the coordinate systems of minutiae features obtained from multiple fingerprints and then combining them. Results indicate improved performance of mosaiced fingerprint over data level fusion. This fusion technique is only applicable for algorithms or modalities with compatible dimensionality. It may also be interesting to note that commercially, the definitions and understanding of features used in a biometric system are considered proprietary. However, new techniques for feature normalization and feature transformation may see renewed interest in this level of fusion from the community.
- **Score Level:** Fusion at score level is the most preferred form of fusion [38]. Since score level fusion offers the best tradeoff between information content and ease of fusion. This has resulted in a wide variety of fusion techniques, some popular methods of score level fusion techniques are:

- 1) **Density-based:** In statistical pattern recognition, Kittler et al. [37] developed a framework to combine evidences obtained from multiple classifiers. Consider the problem of assigning a feature set  $X$  with features  $x_j$  where  $j = 1, \dots, R$  a class  $W_r$ ,  $r = 1, \dots, M$ . Each class  $W_r$  is modeled by the PDF  $p(x_j|W_r)$  and its prior is given by  $P(W_r)$ . A class  $W_r$  is assigned to  $X$  according to Bayesian decision theory given by maximizing posteriori probability such that

$$P(W_r|x_1, \dots, x_R) \geq P(W_k|x_1, \dots, x_R) \quad (1)$$

where  $k = 1, \dots, M$ . This is known as the minimum error-rate classification rule of pattern recognition. This can be expressed as the conditional joint probability of the feature vectors using Bayes rule as:

$$P(W_r|x_1, \dots, x_R) = \frac{P(x_1, \dots, x_R|W_k)}{\sum_{l=1}^M P(x_1, \dots, x_l|W_l)} \quad (2)$$

Kittler et al. proposed several approximations together with the assumption of statistical independence of  $R$  features. Using these assumptions, five

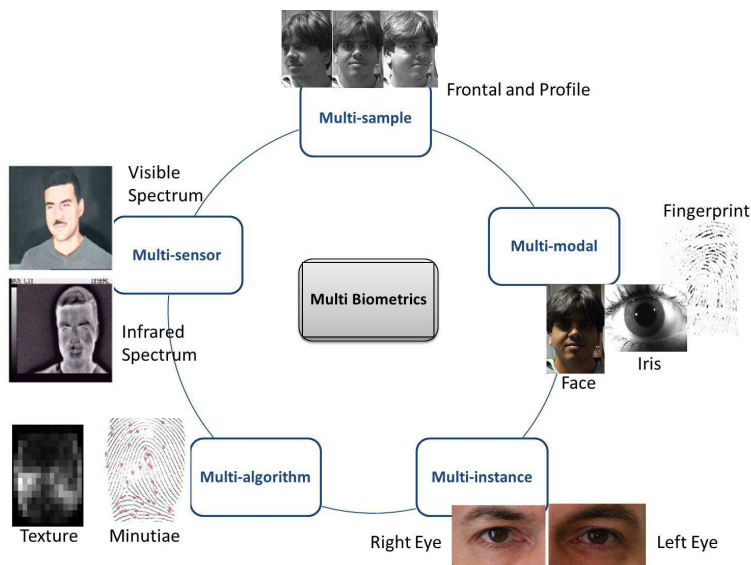


Fig. 10. Various sources available for combining evidence; multi-sample, multi-modal, multi-instance, multi-algorithm, and multi-sensor.

TABLE V  
FIVE CLASSIFIER COMBINATION RULES PROPOSED BY KITTLER ET AL. [37] A CLASS  $W_r$  IS ASSIGNED TO  $X$  ACCORDING TO THE GIVEN RULES

Name	Rule
Product Rule	$\prod_{j=1}^R P(W_r x_j) \geq \prod_{j=1}^R P(W_k x_j)$
Sum Rule	$\sum_{j=1}^R P(W_r x_j) \geq \sum_{j=1}^R P(W_k x_j)$
Max Rule	$\max_{j=1}^R P(W_r x_j) \geq \max_{j=1}^R P(W_k x_j)$
Min Rule	$\min_{j=1}^R P(W_r x_j) \geq \min_{j=1}^R P(W_k x_j)$
Median Rule	$median_{j=1}^R P(W_r x_j) \geq median_{j=1}^R P(W_k x_j)$

fusion strategies are presented. These are listed in Table V. In another approach, Nandakumar et al. [39] propose using product of likelihood ratios (PLR) estimated for each classifier using a finite Gaussian mixture model. Further, the authors extend the idea towards the product of likelihood by augmenting the ratio with quality scores.

In practical scenarios, it is often observed that the number of matchscores available to train fusion modules is small or does not adequately represent real-world data. Due to this limitation, accurate estimation of the joint conditional probability densities is not possible. In such situations matchscores are directly combined without calculation of densities. In order to achieve similar results of fusion, some transformation (normalization) must be applied to ensure all scores are in the same domain. Prevalent normalization techniques in pattern recognition are *min-max*, *decimal scaling*, *z-score*, *median & median absolute deviation*, *double sigmoid* & *tanh-estimators*. The proficiency of each of these techniques is as shown in Table VI.

- 2) **Classification-based:** All matchscores obtained from individual classifiers of a multibiometric system are combined into one feature vector that is input for a different classifier. The matchscore classifier non-linearly maps these scores to a final

TABLE VI  
IMPORTANT SCORE NORMALIZATION TECHNIQUES

Normalization	Robustness	Efficiency
min-max	No	High
decimal scaling	No	High
Z-score	No	High
median and median absolute deviation	Yes	Moderate
tanh-estimators	Yes	High

decision label. Aguilar et al. [40] proposed a support vector machine (SVM) based fusion using quality scores as feature weights. Ross et al. [41] compare matchscore fusion using sum rule, with fusion using decision trees and linear discriminant classifiers. The results indicate that the performance of classifier based fusion is limited in most cases. This is primarily attributed to lack of correct/adequate training samples for the secondary classifier to correctly learn the non-linear relationship between the individual matchscores and the fused decision.

- 3) **Belief Assignment:** Biometrics captured in non-ideal conditions may lead to conflicting decisions from different classifiers in multibiometric systems. Vatsa et al. [42] use evidence-theoretic approaches to assign *belief* to a matchscore based on estimated

densities. The authors further introduce a unification framework [43] for dynamically selecting fusion rules such as belief based sum rule and belief DSm fusion using obtained belief of individual match-scores and quality scores which are moderated with known priors.

- **Decision Level:** Fusion of different biometric systems at after obtaining their decisions has also proved to improve overall recognition performance. Particularly, if the individual pair-wise correlation of the decision outputs are low. The prominent decision level fusion techniques in literature are OR and AND rules, (weighted) majority voting, Bayesian decision fusion and Dempster-Shefer theory of evidence [4].

Another form of decision level fusion that is applicable in identification mode of a biometric system is **Rank Level** fusion. In this variant of decision level fusion, the rank ordering provided by each classifier is considered in a consensus form of combination. Borda count method is prevalent approach for such consensus based classifier combination. Several weighted variants of this technique are also proposed based on either pre-computed confidence of classifiers or by using quality assessment scores [44]. Other forms of rank level fusion use classification approaches such as logistic regression as a generalization of Borda count technique [4]. A detailed empirical evaluation of these techniques indicates towards excellent applicability of non-linear rank level fusion in palmprint biometrics [45].

Fusion have become an integral part of research in biometrics as the counter measure to inherent limitations of individual biometric traits. However, there are several issues that must be taken into account in the development of multibiometric systems such as, the cost incurred in the addition of a new entity in the framework vs. the improvement in accuracy and/or usability. The availability and the reliability of the information used for fusion is also important.

### B. Does Fusion Always Work?

In literature, there have been some criticism of fusion techniques in multibiometrics such as increase in time during enrollment (for example, multiple contact capture based modalities), computation time (processing and matching several biometric samples) and cost of deployment of system (cost of installation and maintenance of multiple sensors). Capturing multiple biometric traits also elevate concerns of privacy and discomfort to users.

Empirical analysis of various fusion schemas *disproves* the common intuition that more information is *necessarily* better than less information. Daugman [50] shows that decision level fusion can not perform better than the stronger classifier. However, the argument is made for decision level fusion where both classifier thresholds are set at equal error rates. Roli et al. [51] provide an empirical study of fusion of several imbalanced classifiers using both fixed rules (majority voting) and learned rules (weighted sum). The finding show a better performance of learnt rules. However this is governed

by the generalizability of the learning process and quality of training samples. Hence, in multibiometric systems, poor quality, limited or unrepresentative training data can cancel out any theoretical advantages of fusion. Therefore, any solution to improve performance by classifier fusion must be partially engineered.

### C. Adaptive Multibiometrics

Multibiometric systems that are intended for deployment in challenging real-world settings such as airports, national boarders and railway stations must maintain robust performance and low computation time in these non-ideal conditions. A primary concern is of degraded or missing data. The quality of probe image may degrade due to large illumination variations, improper interaction with the sensor (pose variations) or different kinds of noise or blur due to limitations of capture sensors. Further, multibiometric fusion schemes may not handle situations when the quality of probe image is not optimal and when all modalities can not be captured, thereby, performance degrades.

Marcalis et al. [52] proposed a fusion technique that eliminates the need for all biometric modalities to be captured at once. This serial fusion of face and fingerprint achieves significant reduction in the verification time while maintaining accuracy. This adaptive nature, termed *context switch* approach, is triggered by the confidence of prediction from matchscore distribution of genuine and impostor scores.

However, for effective context switching, the *trigger* must be some additional meta-information such as quality of gallery and probe sample. In recent literature of multibiometrics, quality based fusion techniques have gained enormous focus. Vatsa et al. [53] proposed a parallel context switching framework that uses image quality and case-based context switching for selecting appropriate uni-modal classifier or fusion algorithm. Bhatt et al. [54] propose a serial framework of quality based classifier selection. The ordering of the classifiers in this serial fashion is such that the strongest classifier is allowed the first attempt for identification. This approach, assuming correct selection of order of classifiers, also saves computation time. In another approach, Alonso-Fernandez et al. [55] present a quality based context switching framework that improves sensor interpretability in fingerprint biometric, when the sensor type is unknown.

The sensitivity required in a particular application of multibiometrics usually determines the usage of any particular fusion scheme and their corresponding parameters. This is usually a trade-off between efficiency and complexity. The selection can hence be modeled as an evolutionary optimization problem given an application specific reward. Kumar et al. [56] use hybrid particle swarm optimization (PSO) to adaptively combine multibiometric classifiers and discover a fusion strategy. Empirical results indicate significantly improved performance when applied at score level.

Poh and Kittler [57] have recently proposed a unified framework for the fusion of biometric classifiers at mathscore level by incorporating quality measures. The problem of quality based fusion is categorized into two approaches, *feature-based*

TABLE VII  
A REPRESENTATIVE LITERATURE REVIEW OF MULTIBIOMETRICS OF EACH CATEGORY AND LEVEL OF FUSION

Authors	Category	Level	Methodology
Marcialis and Roli [26]	Multi-Sensor	score	Optical and capacity fingerprint sensors
Chen et al. [25]	Multi-Sensor	score; rank	2D and IR camera of face
Ross and Govindarajan [46]	Multi-Sensor	feature	Redundant channels of color face image
Zhang et al. [47]	Multi-Algorithm	feature; score	multi-feature
Vatsa et al. [43]	Multi-algorithm	score	level 2 and level 3 fingerprint features
Zhou [48]	Multi-Instance	score	face sequence from video
Ratha et al. [29]	Multi-Instance	sample	integrate multiple snapshot of roll fingerprint
Singh et al. [31]	Multi-Instance	sample	panoramic face mosaic
Kale et al. [49]	Multi-Modal	score	face and gait
Nandakumar et al. [39]	Multi-Modal	score	product of likelihood ratio (PLR)

and *cluster-based*. The first, feature based approach uses quality directly as a feature, typically by concatenation with feature vector. The authors argue that due to the low discriminating capabilities of the quality metric, this approach leads to only marginal improvements, if any. The second cluster based approach partitions the input probes into quality measures based clusters. A specific fusion strategy is then designed for each partition. This closely resembles the intuition of a control parameter or *trigger*. The clustering approach is further segregated into *observable*, *hidden* and *changing* clusters, based on the type of clusters. Clusters that are *observable* can be learnt using a supervised approach from annotated labels. For example, based on subjective quality attributes or by some meta-information of gender, age or ethnicity. Clusters that are *hidden* can be discovered using unsupervised learning are based on some non-intuitive relationships of quality metric and fusion framework. The last type of clusters are *changing*, i.e., while the cluster is initially observable, it loses meaning as the system goes into deployment. Fig. 11 summarizes the categorization of score-level quality based fusion techniques. An extensive literature review of quality based fusion techniques is presented in [57].

Quality assessment is an important component of such context switching and fusion frameworks. Current research uses certain image processing algorithms that are able to assess, image degradations due to noise, compression or illumination. However, a quality metric that is biometric specific and entails a greater insight of the *usefulness* of the biometric sample in consideration, can improve the performance of these systems by providing more discernable clusters. The next sections discuss the aspect of quality with respect to biometrics.

#### IV. BIOMETRIC QUALITY

Quality is an attribute or a property of an item that quantitatively measures specific traits. The word has several connotations in business, science and philosophy. This work aims to define and demystify the meaning and understanding of quality in the field of biometrics. Further, we investigate its applicability in face recognition, an area that is yet to receive proportionate attention from the research community. First, a review of research in the image quality is first presented.

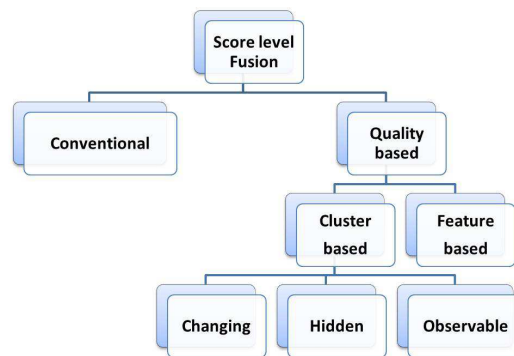


Fig. 11. A categorization of quality based matchscore level fusion techniques [57]

#### A. 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 [58]. 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 human visual system (HVS). Two distinct approaches exist in literature to model the HVS, a *bottom-up* and a *top-down* approach [58]. 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 [58]. 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** (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 [58]. The strength of error computed between the given image and the original (reference) image are 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 four phases discussed below (as shown in Fig. 12).

- 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 non-linear 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 differ 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* [58]. Most methods use the Minkowski form of pooling errors given as:

$$E = \left( \sum_m \sum_n |e(m, n)|^\beta \right)^{1/\beta}, \quad (3)$$

where  $e(m, n)$  is the normalized error of the  $n^{\text{th}}$  coefficient in the  $m^{\text{th}}$  channel of the images and  $\beta$  is a constant ranging from 1 to 4.

Watson's wavelet model [59] is based on the error visibility model. This model evaluates the subjective sensitivity of each band of the linear-phase 9/7 bi-orthogonal filters (this widely used filter is now adapted by the JPEG2000 standards [58]). These sensitivity values are not only used for quality assessment but also in image compression.

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 quality degradation. A spatial domain implementation of this idea is the *structural similarity index metrics* (SSIM) [60]. Given a distorted image ( $\mathbf{x}$ ) and reference image ( $\mathbf{y}$ ), the SSIM index of quality depends on the comparison of  $\mathbf{x}$  and  $\mathbf{y}$  by three measures: luminance, contrast, and structure.

The luminance is compared as the function  $l(\mathbf{x}, \mathbf{y})$ , given by

$$l(\mathbf{x}, \mathbf{y}) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1}, \quad (4)$$

where  $\mu_x$  and  $\mu_y$  are the mean intensities of the local luminance of  $\mathbf{x}$  and  $\mathbf{y}$  respectively.

The contrast,  $c(\mathbf{x}, \mathbf{y})$ , is given by

$$c(\mathbf{x}, \mathbf{y}) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}, \quad (5)$$

where  $\sigma_x$  and  $\sigma_y$  are the variance in intensities of the local luminance of  $\mathbf{x}$  and  $\mathbf{y}$  respectively.

The structure,  $s(\mathbf{x}, \mathbf{y})$ , is given by

$$s(\mathbf{x}, \mathbf{y}) = \frac{2\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3}, \quad (6)$$

where  $\sigma_{xy}$  is the covariance of intensities of the local luminance of  $\mathbf{x}$  and  $\mathbf{y}$ .

The structural similarity (SSIM) index is given as

$$S(\mathbf{x}, \mathbf{y}) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_3)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (7)$$

Here  $C_1, C_2, C_3$  are mathematical constants. These equations are obtained from several observations of the HVS, such as *relative* sensitivity to luminance change (Weber's law) and some reasonable constraints on the similarity measure. Further, Equation 7 reduces to the Wang-Bovik index [61] at  $C_1 = C_2 = 0$ . The SSIM index parameters,  $\sigma_{xy}$ ,  $\mu_x$  and  $\mu_y$ , are computed in a local region with a sliding window, with Gaussian smoothed weights to reduce boundary effects. The performance of this algorithm far exceeds traditional metrics such as mean-squared error (MSE), as shown in Fig. 12. One major drawback of spatial domain SSIM described here is the sensitivity to distortion due to translation, rotation and scaling. One solution is to use the SSIM index formulated in the complex wavelet transform domain.

- 2) *No-Reference* (NR): Blind or no-reference quality assessment is a more difficult problem as there is no reference image for comparison. 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

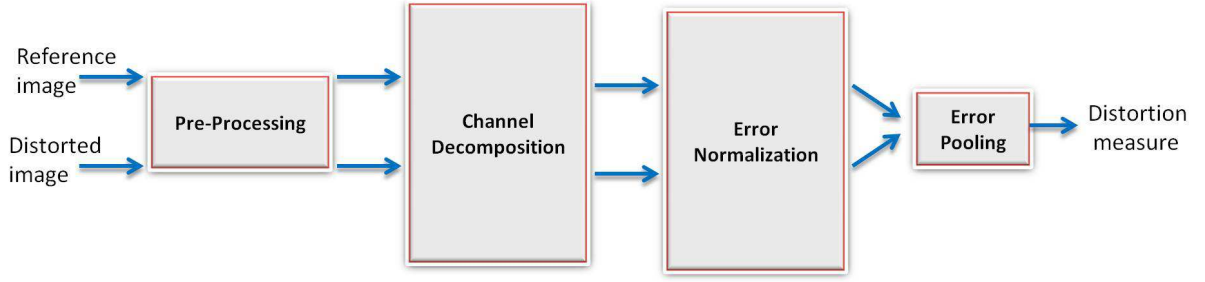


Fig. 12. A typical Full Referenced (FR) bottom-up image quality assessment system based on error visibility.

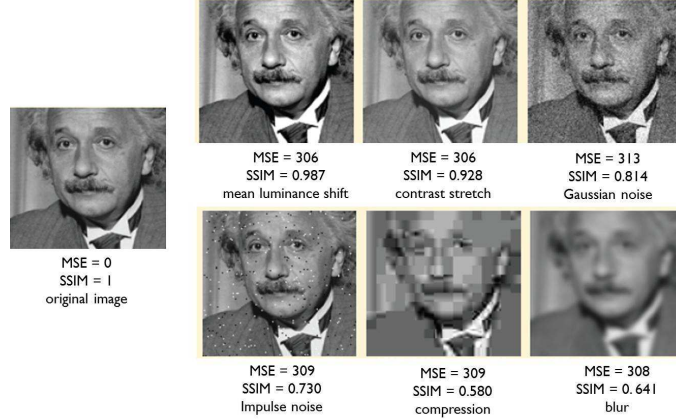


Fig. 13. Portrait image altered by various distortions. a) original image b) mean luminance shift c) contrast stretch d) impulse noise e) Gaussian noise f) blur g) compression h) spatial shift i) scale j) rotation. While most images have the same mean square error (MSE), there is a drastic difference in the visual quality. A motivating example towards SSIM quality index obtained from [60].

knowledge of the expected degradation process on the image. A NR perceptual quality assessment algorithm for JPEG compression is proposed by Wang et al. [62]. 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.

- a) Blockiness is estimated by the average intensity difference between block boundaries of the image  $x$ . For an image of size  $M \times N$ , the blockiness in horizontal direction ( $B_h$ ) is given by Equation 8:

$$B_h = \frac{1}{M([N/8] - 1)} \sum_{i=1}^M \sum_{j=1}^{[N/8]-1} |d_h(i, 8j)| \quad (8)$$

where  $d_h$  is the differentiating signal in horizontal direction  $d_h(m, n) = x(m, n+1) - x(m, n)$  for  $n \in [1, N-1]$ .

- b) Activity of the image provides insight on the effects of compression and blur in the image. Activity ( $A_h$ ) of an image of size  $M \times N$  is given by:

$$A_h = \frac{1}{7} \left\{ \frac{8}{M(N-1)} \sum_{i=1}^M \sum_{j=1}^{N-1} |d_h(i, j)| - B_h \right\} \quad (9)$$

Activity of an image may also be measured via

zero-crossing rate of the image of size  $M \times N$  and it is given by:

$$Z_h = \frac{1}{M(N-2)} \sum_{i=1}^M \sum_{j=1}^{N-2} z_h(m, n) \quad (10)$$

where,

$$z_h(m, n) = \begin{cases} 1 & \text{if horizontal ZC at } d_h(m, n) \\ 0 & \text{otherwise} \end{cases} \quad (11)$$

and  $n = [1, N-2]$ .

Similarly, blockiness, activity and zero crossing rate is measured in vertical directions as  $B_v$ ,  $A_v$  and  $Z_v$ . The overall estimation of values B, A, and Z are given by:

$$B = \frac{B_h + B_v}{2}, \quad A = \frac{A_h + A_v}{2}, \quad Z = \frac{Z_h + Z_v}{2} \quad (12)$$

Finally, the blockiness, activity and zero-crossing rate are combined to obtain quality score S,

$$S = \alpha + \beta B^{\gamma_1} A^{\gamma_2} Z^{\gamma_3} \quad (13)$$

where the parameters  $\alpha$ ,  $\beta$ ,  $\gamma_1$ ,  $\gamma_2$  and  $\gamma_3$  are the model parameters that must be estimated for a given data set. In another approach, Marziliano et al. [63] have proposed edge spread as a measure to estimate irregularities based on edges 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 [63] show that either of the two directions suffices for quality assessment.

- 3) *Reduced-Reference (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 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 pre-selected pixel values. At the receiver end, the mean-squared error (MSE) of pixel values of original and distorted image is computed to obtain quality. Gao et al. [64] 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. 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, 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.

### B. Quality of a Biometric Sample

Quality of a biometric sample is interpreted differently throughout literature [65], [66], [67], [68]. Most commonly it is assumed that image quality indicates the *usefulness* of the biometric image in recognition. It is well established that the environmental distortions such as noise, blur, adverse illumination and compression affect the performance of state-of-the-art recognition algorithms. However, existing image quality metrics encode only part of the quality information that can measure the overall quality of a biometric sample. Ideally, it is desired to design a biometric quality metric that, given an image, can measure the proficiency of an input sample in recognizing an individual (irrespective of recognition algorithms).

### C. Why is Quality Assessment Important?

Quality assessment has several applications in the field of biometrics. Some of these applications are described below with inspiration from existing literature [57], [69].

- **Enrollment Phase Quality Assessment and Assurance:** Quality feedback during biometric capture at enrollment phase is critical in collecting high quality gallery data. It is common, especially in large scale biometric systems to have a supervised enrollment process, as shown in Fig. 14. An active quality feedback enables the collection officer to maintain quality standards during the enrollment process. It can also be a performance measure for the collection apparatus and procedure employed for data capture. It helps in quality assurance and aggregated quality may also be used to create timeline along with historical or geographical meta-data for other analysis.
- **Quality Assessment in Verification:** Quality assessment and feedback during verification can help mitigate false alarms. A verification system can choose not to perform matching if quality is below a threshold, depending on the computation time of matching and the re-acquisition of data. Often, quality metrics in biometrics are evaluated by showcasing an increase in verification accuracy when low quality samples are removed.
- **Quality Assessment in Identification:** Since identification is inherently a computationally expensive process, 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 process, where it is in the interest of the subject to provide a poor quality sample. The subject may then be forced to provide better samples without having to wait for misleading and incorrect identification result from the system.
- **Differential Processing:** Extensive literature exists in demonstrating the advantages of dynamic altering or context switching of a biometric recognition pipeline based on feedback from quality assessment algorithms [53], [54], [55]. Active involvement of the quality score beyond the capture stage encourages the formulation of more complex and accurate quality assessment even at the cost of computation time. The use of quality at each stage are described as follows:
  - *Pre-processing:* Quality assessment based selection of parameters for image enhancement show marked improvement in the recognition performance of the resultant biometric sample, when compared to using generic parameters. A probe sample may contain noise due to environmental conditions, incorrect use of sensors or transmission error. The performance of recognition systems severely depletes with varying noise present in an image. Denoising techniques help in improving the recognizability of face images, provided the correct parameter are used [71]. An illustration of a quality assessment based image enhancement framework is presented in Fig. 15(a). The selection process may also be extended to using multiple enhancement algorithms as well. A more detailed discussion of the technique is presented in Section V.
  - *Recognition:* Poh et al. [72] and Kryszczuck et al. [73] have shown that quality assessment scores can



Fig. 14. A typical enrollment center of India's *Aadhaar* project to capture face, iris and fingerprint biometrics. A trained supervisor is present to ensure quality and consistency in the enrollment phase. (Image obtained from the [70])

not only be 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 recognition performance. As discussed in Section III, quality augmented product of likelihood ratio fusion scheme improves matchscore fusion [39], [74].

- *Context Switching*: As discussed in Section III-C, 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 Fig. 15.
- *Decision*: Similar to matchscore level fusion, decision level fusion can also be aided by quality assessment scores. By providing quality priors to maximize, selective or cumulative combination of decision, the notion of strong or weak classifier can become subject specific. Hence, the primary concern of using decision level fusion schemes, discussed in [50] are eliminated.
- *Sample 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 with aging.
- *Decision Update*: Currently, there is some research on *online* or *incremental learning* to improve the decision boundary of the classifiers even in deployment phase [75]. A major concern in such systems is to select suitable samples to learn incrementally. For instance, modifying decision boundary based on all incoming samples may be too computationally expensive. Further, online learning on outlier samples can adversely effect system performance. One area of focus is towards using quality of the sample to determine if the sample is suitable for classifier update.

With respect to biometric applications, accurate and prompt quality assessment of incoming samples is crucial. Several categorical attributes of a biometric, such as presence or absence of glasses, are also loosely termed as quality factors. It is interesting to note various controllable and native factors that influence the quality of a biometric sample, discussed

next.

#### D. Factors that Influence Biometric Quality

A number of factors affect the quality of a biometric sample. It is important to understand and appreciate these effects to develop better quality assessment algorithms. Of these factors, there are also some factors that are unfortunately not avoidable and are inherent limitations of the biometric itself. Alonso-Fernandez [76] presents some of these factors, listed in Table VIII.

- **User**: Some important factors that influence the quality of a biometric during capture process are *behavioral* and *physiological* traits of the human users. While many of these factors (shown in Table VIII) can be considered avoidable, it is at the cost of increased user-inconvenience. Further, some *natural* factors such as race, ethnicity, social customs and gender impair capture quality. Medical injuries or other natural abnormalities may also affect the quality of the data that is captured. These factors are compounded when user is uncooperative.
- **User-sensor interactions**: The second important factor to influence the quality of contact capture based biometrics (such as fingerprints, palm prints, iris, retinal) is the interaction between users and sensors. The *usability* of the sensor is crucial, those with active user feedback, that are portable and easy to use ensure good quality captures. However, several other *environmental* factors also influence this interaction and hence the quality of a biometrics.
- **Operational constraints**: Other factors that affect the quality of a biometric sample are operational constraints. These are mainly attributed to the use and maintenance of the sensors (particularly touch based).

#### E. Evaluating Quality Assessment Techniques

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 gold standard for comparison and testing of automatic algorithms. A set of volunteers are presented with images of different quality and their responses are aggregated using a mean operator score (MOS). A high linear correlation between the predicted quality and MOS from volunteers indicates high performance. This method can not be

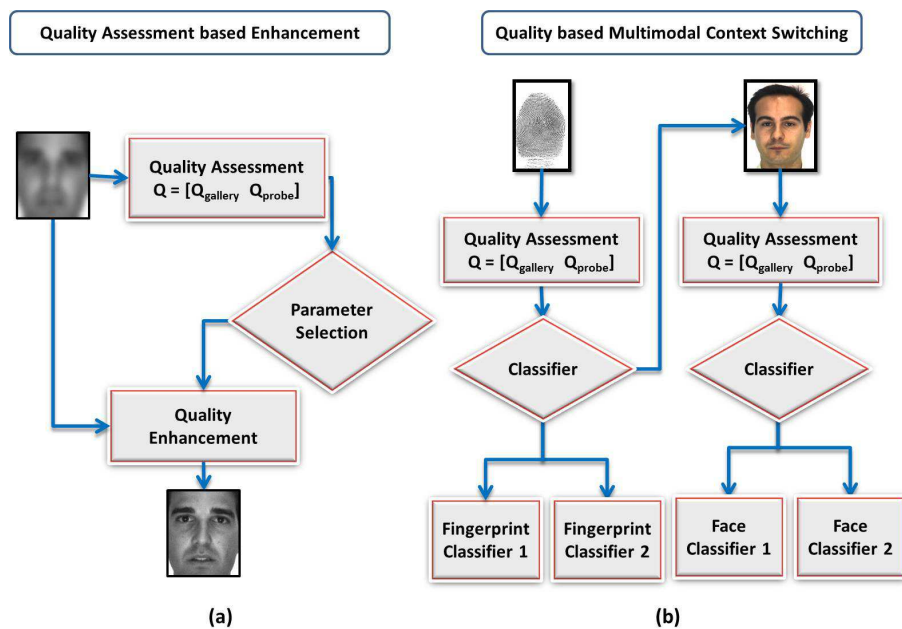


Fig. 15. Quality assessment in biometrics has several uses in context switching. Framework for a) a quality driven biometric image enhancement, based on [71] and b) quality based multi-classifier selection, proposed by [54]

TABLE VIII  
VARIOUS BEHAVIORAL, ENVIRONMENTAL AND OPERATIONAL FACTORS THAT EFFECT QUALITY OF BIOMETRIC SAMPLE

Factor	Possible Causes
Behavioral factors	Tiredness, distractions, motivation, cooperation, fear, make-up, appearance, facial hair, clothes or hats
Environmental factors	Indoor/outdoor, background, temperature, humidity, illumination, Ambient noise
Operational factors	Familiarity, quality feedback, sensor cleaning, supervising operator, time between acquisition

directly applied for biometric quality, as there is no conclusive evidence that human interpretation of quality correlates with the quality in terms of performance of a recognition algorithm. In our observation, three prominent methods of evaluation of biometric quality metrics persist in literature apart from evaluation using MOS.

- *Correlation*: As noted by [69], a biometric quality metric must be a good classifier performance predictor. With this view, a quality measure that is highly correlated (statistically) with the normalized matchscores obtained from a classifier is most desirable. Hence, several literature showcase correlation with genuine matchscores [69], [77]. Since every matchscore can be associated to quality of both gallery and probe sample, several combining methods are employed, such as  $Q_{gallery} + Q_{probe}$  or  $Q_{gallery} \times Q_{probe}$  or  $\min(Q_{gallery}, Q_{probe})$ .
- *Quality bins*: In another approach, researchers measure the impact of quality metrics by segregating the entire dataset into a number of quality bins and performing individual recognition experiments on each of them. Further, the intuition that a subset of better quality data has better recognition accuracy is substantiated. While this method is popular in the biometrics community, [65],

[66], [67], it might be difficult to attribute the difference in performance to the quality metric itself as the database can become skewed towards of easy or difficult samples.

- *Distance Metric*: Other literature use quality score to alter the distance space. For example, Chen et al. [65] 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 evaluation of quality metrics is by computing the cross-correlation between various existing metrics. However, some additional benefits of the algorithms (perhaps in terms of computation time or better correlation with MOS) must be described in order to differentiate from existing approaches. This evaluation method is used in [66].

It must be noted the best method to evaluate a quality metric is dependent on the intended application. Several quality scores in literature do not linearly correlate with matchscore, however, become good performance prediction features when used along with a non-linear mapping. Other metrics may improve performance in a modified distance space but at significant computational costs. Next, a detailed literature review of

existing quality assessment algorithms in three modalities are presented. This review includes a description of the evaluation criteria used in each of the presented research as well.

#### F. Humans vs. Automatic Quality Assessment

As mentioned previously, the performance of humans in quality assessment was considered gold standard for comparison and testing of automatic algorithms. In biometrics, it is true that experts in forensics are able to correctly predict the recognition performance of a sample based on their intuition of quality. Many datasets now exist that contain annotated human quality scores for comparison [77]. The work of Adler et al. [78] compare quality scores obtained from several human subjects and automatic algorithms for face and iris modality. They find high correlation between human subjects used in the study. Further, they also find high correlation between different quality assessment algorithms. However, they find low correlation between humans and algorithms. Some exceptions include preference by both humans and automatic algorithms for sharp images. Training and highly motivated users can provide good quality scores by observation, however it might be increasingly difficult to sustain the motivational levels to create annotations for large datasets. Hence, in general it is best to consider other evaluation techniques discussed in Section IV-E to compare and evaluate automatic quality assessment algorithms.

#### G. Fingerprint Quality Assessment

Quality assessment of fingerprints in terms of global and local ridge quality is essential for proper functioning of the system. Poor quality fingerprint images can lead to incorrect or spurious feature (minutia) detection and hence degrade the performance of a fingerprint recognition system. These metrics are primarily used in fingerprint sensors with active quality feedback for rejection of poor quality samples. Their use is also particularly important to evaluate local unrecoverable regions of the fingerprint, as enhancement of these regions may be counter-productive. Further, these regions 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. A summary is shown in Table X.

Lim et al. [79] present a local feature based quality metric which computes *orientation certainty level* (OCL), *ridge frequency*, *ridge thickness* and *ridge-to-valley thickness ratio*. By setting various thresholds, discrete levels of local quality (good, bad, undetermined or blank) are obtained for each local block. Further, the final score is computed as a weighted combination of the block-wise quality values. OCL is a measure of spectral energy along the direction of dominant ridges (those with large energy signature in the spectrum). Ridge frequency detects any abnormal ridges (that are too close or too far from each other). The ridge thickness and ridge-to-valley ratio indicates ridges that are too thin or too wide. Interestingly, this technique is compared with the performance measures (number of detected minutiae and segmented region)

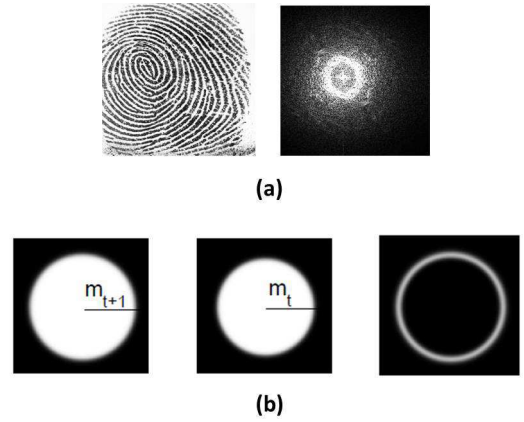


Fig. 16. a) A fingerprint image and corresponding Fourier Transform (magnitude component after shifting). The ridge information manifests as a bright band. b) Chen et al. [65] use difference of two Butterworth filters  $m_t$  and  $m_{t+1}$  to obtain a soft bandpass filter that captures the strength (and thereby quality) of the ridges.

of the FBI fingerprint system AFIS [80]. The OCL measure is also useful to compute regions of high curvature, known as singularity (delta or cores). Shen et al. [81] 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 are 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. [82] use Redundant Discrete Wavelet Transform (RDWT) to compute dominant ridge information 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. Chen et al. [83] proposed quality assessment based on the clarity of ridges and valleys. For each block of a tessellated fingerprint image, the amplitude of a sinusoidal wave that models the ridges and valleys along the normal of the local ridge direction is computed and used to cluster the pixels. Further, a distribution of the gray scale values in the local block is computed for each cluster. High clarity block regions exhibit low overlap between the gray distributions of valley and ridge are measured as high quality blocks. The method is tested on a small database of fingerprint images, with perception scores. In an another approach, Chen et al. [65] measure the quality of ridge samples by energy spectral density concentration in particular frequency bands obtained by Discrete Fourier Transform (FFT) (Equation 14). It is observed that good quality ridges manifest at a certain frequency band of the transformed fingerprint image as shown in Fig. 16.

$$F(k, l) = \frac{1}{NM} \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} f(i, j) e^{-2\pi i \left( \frac{ki}{N} + \frac{lj}{M} \right)}, \quad \iota = -1^{\frac{1}{2}}, \quad (14)$$

where  $f(i, j)$  represents the image intensity at pixel location

$(i, j)$  and  $F(k, l)$  is the corresponding Fourier transform at location  $(k, l)$ , for an image of size  $M \times N$ . The power spectrum  $P(k, l) := |F(k, l)|^2$ , encodes the geometric structure and concentration of edge information in the image. Further, a Butterworth filter ( $\mathbf{H}$ ) is applied to separate the band with ridge information.

$$H(k, l|m, n) = \frac{1}{1 + \frac{1}{m^{2n}} \left( \left( \frac{k-a}{M} \right)^2 + \left( \frac{l-b}{N} \right)^2 \right)^n}, \quad (15)$$

$$R_t(k, l) = H(k, l|m_{t+1}, n) - H(k, l|m_t, n), \quad (16)$$

$m_t = 0.06 + t \frac{0.5-0.06}{T}$  and  $t = 0, 1, 2, \dots, (T-1)$ . A total of  $T$  bands are constructed each with energy  $E_t$ , computed as  $E_t = \sum^M \sum^N R_t(k, l) P(k, l)$ . The energy in each band is further normalized as  $E_t^{norm} = \frac{E_t}{\sum_{t=1}^T E_t}$ . Finally, total energy concentration across all bands is computed as Equation 17

$$E_{total} = - \sum_{t=0}^{T-1} E_t^{norm} \log E_t^{norm} \quad (17)$$

The total energy  $E_{total}$  is a good indicator of the quality of the ridges in a fingerprint. However, this method has some limitations. Being a global quality assessment metric, the poor quality area of the fingerprint can not be localized. Further, the position of the energy bands is subject to the type of sensor used and must be empirically determined.

The most popular fingerprint quality assessment algorithm in literature is the NIST Fingerprint Image Quality (NFIQ) [84]. This approach also pioneers the use of quality metrics as performance predictor in fingerprints. A feature vector  $v$  consists of 11 quality features is 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 of size  $3 \times 3$ ). Rather than using true labels based on human perception, normalized genuine matchscores obtained from an automatic fingerprint matcher are used. In order to obtain the feature vector  $v$ , the quality of the local areas around detected minutiae points are obtained from the localized quality map, described in Table IX. A 3-layered feed forward neural network, uses the training samples  $\langle v^{(j)}, P(q^{(j)}) \rangle$ ,  $j = 1, \dots, n$ . The NFIQ quality metric has been extensively used in literature and tested across different datasets. However, the orientation estimated about the singularity points tend to fail for high curvature. Fronthaler et al. [66] 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 correlation of this quality score with NFIQ and with human annotations is high.

Alonso-Fernandez et al. [77] present a comprehensive survey

and evaluation of several fingerprint quality metrics. The primary focus is on the evaluation of quality assessment of fingerprints in terms of utility. These algorithms are segregated into global and local feature. Further, they are also segregated as direction based, Gabor filters based, pixel intensity based and power spectrum based algorithms. The classification is presented in Table X. The study uses 19200 fingerprint images collected from 200 individuals with 3 types of sensors. The primary findings the study is a high correlation of fingerprint quality metrics among themselves and with genuine matchscores. This seems to indicate that most approaches encode similar information from the fingerprint image to predict quality. Also, research into quality assessment of *latent* and 3D fingerprints are still very much open research problems.

#### H. Iris Quality Assessment

The performance of 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, 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. Most assessment techniques, therefore present several separate assessment and (often, complex) fusion techniques to result in a single score, as is the trend. A brief description of some leading iris quality assessment methods is presented in Table XI.

Chen et al. [85] 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. 2D wavelet decomposition is applied on the segmented iris region. Further, 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 (eight pixels). The energy from  $T$  concentric regions are combined into a single quality score  $Q$  as

$$Q = \frac{1}{T} \sum_{t=1}^T (m_t \times \log E_t) \quad (18)$$

where  $m_t$  is a weighting function favoring the regions closer to the pupil, and  $E_t$  is the energy from the  $t^{th}$  concentric circle given by Equation 19.

$$E_t = \frac{1}{N_t} \sum_{i=1}^{N_t} |w_{t,i}^p|^2 \quad (19)$$

here  $w_{t,i}^p$  is the product of responses of the  $i^{th}$  wavelet coefficient over  $p$  scales and  $N_t$  is the total number of coefficients. Fig. 17 illustrates the energy  $E_t$  for each concentric circle. The algorithm is evaluated by incorporating the quality values while computation of Hamming distances as follows:

TABLE IX  
THE NFIQ FINGERPRINT QUALITY METRIC USE 11 FEATURES AS INPUT TO A FEED FORWARD NEURAL NETWORK [84].

Number	Name	Feature
1	foreground	number of blocks that are quality 1 or better
2	total #of minutia	number of total minutiae found in the fingerprint
3	min05	number of minutiae that have quality 0.5 or better
4	min06	number of minutiae that have quality 0.6 or better
5	min075	number of minutiae that have quality 0.75 or better
6	min08	number of minutiae that have quality 0.8 or better
7	min09	number of minutiae that have quality 0.9 or better
8	quality zone 1	percentage of the foreground blocks of quality map with quality =1
9	quality zone 2	percentage of the foreground blocks of quality map with quality =2
10	quality zone 3	percentage of the foreground blocks of quality map with quality =3
11	quality zone 4	percentage of the foreground blocks of quality map with quality =4

TABLE X  
A REPRESENTATIVE LIST OF REPRESENTATIVE FINGERPRINT QUALITY ASSESSMENT ALGORITHMS MEASURING LOCAL/GLOBAL FEATURES

Category	Algorithm	Description	Type
Local Direction	Lim et al. (OCL)[79]	uses orientation to find valid ridge- valley patterns in local texture	Local
Gabor Filter	Shen et al. [81]	combine response of $m$ filters with different orientations	Local
Pixel Intensity	Chen et al. [83]	grey level distributions of segmented ridges	Local
Wavelet Transform	Vatsa et al. [82]	combine response from RDWT for dominant edge information	Local
Power Spectrum	Chen et al. [65]	in a ring shaped region of the spectrum	Global
Combined Features	NFIQ [84]	Amplitude, frequency and variance of sinusoid to model valid ridges	Global
Orientation Tensors	Fronthaler et al. [66]	encode orientation with parabolic symmetry features	Global

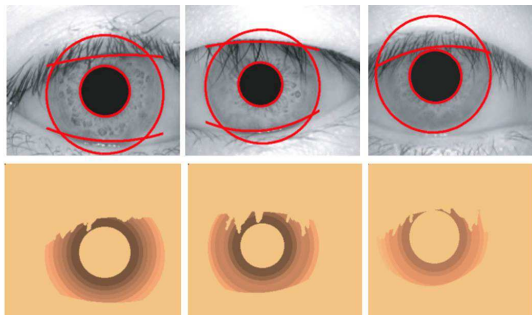


Fig. 17. Three segmented iris images with corresponding quality map depicting energy  $E_t$  of each of the concentric circles [85].

$$HD = \frac{1}{B} \frac{\sum_{i=1}^B (\sqrt{E_i^X E_i^Y} X_i \oplus Y_i)}{\sum_{i=1}^B \sqrt{E_i^X E_i^Y}} \quad (20)$$

where  $X_i$  and  $Y_i$  are the  $i^{th}$ -bits of two iris feature vectors  $X$  and  $Y$ .  $E_i^X$  and  $E_i^Y$  are the associated local quality at the  $i^{th}$ -band of  $X$  and  $Y$  respectively.

Kalka et al. [67] present quality assessment of iris images based on the evaluation of seven quality parameters and

their fusion. These individual quality scores are both image based and biometric specific in nature. The performance of the technique is found to be robust on challenging datasets and is considered to be the state-of-art in iris biometric quality assessment. The quality framework comprises of the estimation of the following covariates.

- 1) *Defocus Blur*: The most common reason for an out-of-focus image is when focus point is outside the *depth of field* during capture. Blur effect generally attenuates the high frequency components of the image. Hence, image content from these higher frequencies are analyzed from the segmented region of the iris. A  $8 \times 8$  convolution filter from the pre-processing step of Daugman [11] is used to measure the 2D spectral power of the image as given by Equation 21.

$$f(x) = 100 \frac{x^2}{(x^2 + c^2)} \quad (21)$$

where  $x$  is the total spectral power measured by the convolution filter and  $c$  is the half power, corresponding to 50%. In order to avoid spurious reading about the eyebrow region, the filter are applied locally to the bottom half of the iris region.

- 2) *Motion Blur*: Relative motion between interest point and camera leads to motion blur, generally caused due head movement during capture. Motion blur manifests as a dominant line orthogonal to the direction of the blur. Hence, blur angle is estimated by computing spectral power along quantized orientations, and the direction corresponding to the maximum power is considered as blur angle. Further the log magnitude of the highest peak in the direction of the blur corresponds to the strength of motion blur in the image.
- 3) *Off-Angle*: The rapid movement of the eyeball can lead to off-angle capture of iris. It is estimated by computing the circularity of the pupil using the magnitude of the integro-differential operators from the segmentation module of [11].
- 4) *Occlusion*: Captured iris image may be of little value if the discriminating texture is occluded due to eyelids and eyelashes. Occlusion is evaluated by applying a morphological filling operation in the circular region of the iris to combine all discontinuities in that region. An occlusion mask is then generated that indicates occluded regions in the iris image.
- 5) *Specular Reflectance*: The reflective nature of the eye can lead to specular reflectance during capture, especially in uncontrolled lighting environments. It is empirically observed that these reflections lie above 240 color value in a grey scale image, and are truncated.
- 6) *Lighting Variation*: For a consistent presentation of the iris texture, it is essential that the iris be subjected to uniform illumination. An angled or non-uniform lighting can lead to reduction in performance. To evaluate lighting variations, all region of the iris not covered by the occlusion mask computed previously, are divided into  $N$  concentric regions (or rings). Lighting variation is measured as shown in Equation 22.

$$Lighting = \sum_i^{12} (X_i - \mu)^2 N, \quad (22)$$

where  $X_{1,\dots,N}$  is the mean color value for *each* of the  $N$  regions and  $\mu$  is the mean color value of the *all* the regions.

- 7) *Pixel Count*: The amount of iris information captured is a function of the resolution of the iris image. Capturing from longer distances leads to iris images at varying resolutions. The pixel count measures ratio of the number of useful pixels (not occluded) to the total number of pixels in the segmented iris region.

The authors then use Dempster-sheffer theory based fusion to combine these individual scores to obtain a single quality value. Recent interest in non-ideal imagery has sparked research on iris recognition in the visible spectrum. Proenca [86] presents a quality assessment algorithm for operation on visible iris imagery. This method is inspired by the work of Kalka et al. [67]. Seven different quality attributes that impact recognition are identified. Using the segmentation algorithm proposed by He et al. [87], the center and biological boundaries of the iris are estimated. Further assessment of

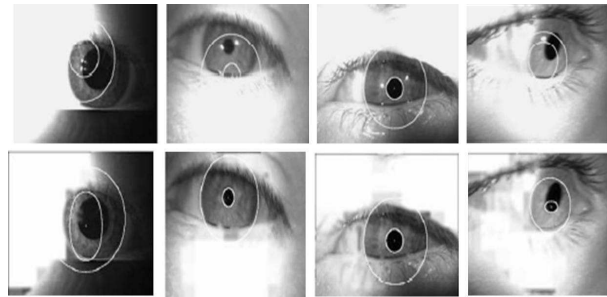


Fig. 18. Samples of poor iris segmentation on iris images obtained from a public dataset [67].

quality is performed only within this boundary to relegate the effects of eyelashes or other occlusions. Focus is determined by the energy in the high frequency components of the Fourier transform. Any motion in the iris region is treated as a linear blurring effect, and is modeled as the width of the dominating signature in the Cepstrum transform of the image. Off-angle is estimated as the circularity of the pupil region. Iris pigmentation is assessed based on an empirical relationship between pigment quality and *hue* values in the iris region. 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.

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 shown in Fig. 18, iris segmentation methods are also adversely affected by the above mentioned covariates. Further, the research community also lacks a comprehensive survey and common test-bed evaluation of existing academic and commercial 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.

### I. Face Quality Assessment

It is well established that the quality measures are an important feature of modern biometric systems. However, in face recognition literature, this area has received very little attention. Early research focuses on complete automation of essential capture guidelines in standards such as ICAO and ISO. However, these guidelines are designed for manual recognition process and provide minimal information of quality of face biometric. More research focus must be directed towards this problem, since it is observed in several empirical studies including the findings of biometric grand challenges, that the covariates of face recognition (pose, illumination, expression, noise) effect performance across different types of features or systems. A discussion of existing face quality metrics is presented here. A brief summary is also available in Table XII.

Subasic et al. [89] present an evaluation scheme of a set of seventeen automatic tests in conjunction with the ICAO face image presentation standards for automatic evaluation of

TABLE XI  
A REPRESENTATIVE LIST OF IRIS QUALITY ASSESSMENT ALGORITHMS

Algorithm	Description	Type
Daugman [11]	focus estimate and off-angle measure by deformation function that maximizes circularity of pupil	Global
Chen et al. [85]	spectral energy in local regions of the iris	Local
Zuo et al. [88]	assessment of interlacing, illumination, focus, off-angle, area, blur pupil dilation	Local, Global
Kalka et al. [67]	evaluation of seven quality parameters and fusing them statistically	Local, Global
Proenca [86]	estimation of seven separate quality attributes that impact recognition	Local, Global

quality. 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 point out 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. [90] present a more comprehensive evaluator of the ISO/JEC 19794 – 5 face standards. The approach combines several image quality metrics and face specific metrics using facial feature detection. A list of standards and the proposed approach is described in Table XIII. While a detail description of the evaluation metrics is lacking, the authors evaluate several linear and non-linear fusion schemes for matchscore prediction. Further, the authors use a nonlinear neural network, with the proposed set of quality metrics as feature vector, to predict matchscore of commercial face matching system. Recently, Youmaran and Adler [91] discuss information content in biometric images. This is termed as *Biometric information* (BI) and is defined from information theory perspective as the decrease in uncertainty of the identity of the person due to the feature set measurements. Assuming normally distributed feature sets, biometric information can be modeled as the relative entropy  $\Delta D(p||q)$  between the intra-person feature distribution  $p(x)$  and the inter-person feature distribution  $q(x)$ .

$$\Delta D(p||q) = \int p(x) \log \frac{p(x)}{q(x)} dx \quad (23)$$

Biometric information measures the biometric content of face images when using PCA and PCA + Fisher LDA features [92]. It is tested on normal and blurred face images. Initially, performance decreases but reaches steady state suggesting some of the features are unaffected by the blurring degradation. The approach is limited by the validity of the distribution  $q$  which is the model for all possible faces. While this work provides good insight into the problem of fidelity of the biometric, the algorithm is not practical to implement, since it requires a statistically valid number of samples for each subject (usually more than 30), in order to estimate the distribution of subject's features. Gao et al. [93] proposed the use of asymmetry in LBP features [94] as a measure of the quality of face biometric. However, this 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 [95] improve on this intuition using Scale Invariant Feature Transform (SIFT) features [96]. 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 about 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 classification.

An important application of quality assessment in face biometrics is in video face matching. 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. [97] 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 overall quality.

The general approach for face quality assessment in videos is based on comparison of a facial image with face models developed from *ideal* example sets. In another approach, Nasrollahi and Moeslund [98] 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 [99] for NIR video face recognition. Yao et al. [100] use 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, quality metrics may not sufficiently measure

TABLE XII  
A REPRESENTATIVE LIST OF FACE BIOMETRIC QUALITY ALGORITHMS.

Application	Algorithm	Description
Still-Image	Subasic et al. [89]	Seventeen automatic tests in conjuncture with the <i>ICAO</i> face image presentation standards
	Hsu et al. [90]	Automatic evaluator of the ISO/JEC 19794 – 5 face standards
	Youmaran and Adler [91]	Biometric information defined from information theory
	Gao et al. [93]	Asymmetry in LBP features [94] as a measure of the quality
	Zhang et al. [95]	Asymmetry using SIFT features
Video-Frame	Wong et al. [97]	Comparison of a facial image with <i>ideal</i> face models
	Nasrollahi et al. [98]	Geometrical pose estimation using face bounding box
	Long et al. [99]	Assess sharpness, brightness, resolution and pose in NIR videos
	Yao et al. [100]	Sharpness measure from frame selection

TABLE XIII  
DIFFERENT ISO METRICS AND AUTOMATIC EVALUATION APPROACH PROPOSED BY HSU ET AL. [90].

ISO	Metric	Approach
Format	Compression artifacts	Checks compression ratio and JPEG blockiness artifacts
Digital	Resolution	Locates eye positions and calculates inter-eye distance
	Grayscale density	Checks intensity dynamic range over facial area
	Video interlacing	Detects motion between interlaced fields
Photographic	Head centering	Estimates suitability of face for creating token face image
	Exposure	Checks unbalanced luminance distribution.
	Focus	Spatial sharpness and linear motion blur
	Unnatural color	Detects unnatural lighting via skin tone
Scene	Head rotation	Estimates Yaw, Pitch (from pose subspaces) and Roll angles
	Eyes closed	Detects closed eyes based on business
	Background uniformity	Spatial sharpness and linear motion blur
	Unnatural color	Analyzes gradient distribution over background
	Uniform subject lighting	Estimates lightness and its spatial distribution on face
	Facial shadows	Employs histogram-based thresholding to identify shadows
	Eye glasses and glare	Detects frame via gradient fitting; glare via blob filtering
	Hot spots	Detects bright spots on face via optimal image segmentation
Other	Face likeness	Likelihood that the given face image is a real face
	Skin texture	Measures the dispersion of the spectra of the skin

quality of the face biometric sample.

### J. Matchscores Derived Quality

Kumar and Zhang [68] present an intuition of user quality in biometrics. The quality metric is derived from the response generated from the classifier (matchscore) when presented with features obtained from different users. Hence a user specific quality metric is generated from the matchscores obtained from a training set. This quality score for each user is convolved with the actual matchscore obtained during probing. Considerable improvement in verification accuracy of hand based biometrics is observed.

In another approach Zuo et al. [101] present an iris biometric quality assessment technique based on matchscore evaluation. Assuming normal distribution of the population of genuine and imposter scores, the quality of a sample is measured as a statistical fusion of two quality metrics, namely, quality of sample index (QS) and confidence of matchscore (CS). QS is computed as the statistical error between the distribution of

genuine and imposter scores that the sample is involved in. CS of a sample is measured as the normalized difference between the sample matchscore and some quantile points selected from the genuine and imposter distribution. It is our assertion that the improved performance in iris recognition showcased in this work is attributed to the property of Hamming distance scores being distributed normally as noted by [12]. Hence, this technique may not be suitable for other distance metrics or matching algorithms.

## V. NOVEL QUALITY OF FACE BIOMETRIC

As discussed in previous sections, quality assessment of face biometric is a challenging problem that has not received the attention of the research community. The complexity of the problem is multiplied by the lack of consensus in literature on the definition and constitution of facial features. Hence, some recent research efforts towards face quality assessment are not generalizable due to assumptions made regarding the facial features. In this section, a unifying quality assessment

framework that is motivated by research in image quality assessment and the findings of large-scale experimentation of biometric grand challenges, is presented. In our opinion, the framework encompasses several aspects of quality assessment from literature and can clear the ambiguity in definition, usage and evaluation of quality metrics. Based on this model, we further our investigation on face quality metrics by presenting an evaluation of holistic image representations for quality assessment. Our finding suggest that quality and usability of a face can be encoded in super-ordinate image representations.

#### A. New Insights on Face Quality

The unique attribute of Face recognition vendor test (FRVT) 2006 is in providing several thought provoking insights and directions to the problem of quality assessment in face recognition [93]. These findings are discussed by Beveridge et al. [102], [103], [104] with a detailed analysis of the *effect* of various subjective and objective covariates of face biometric. Further, the results of FRVT 2006 also lead to the introduction of the *Good, Bad and Ugly* classification of the face databases [105]. Important findings from the results of the FRVT 2006 exercise are,

- A strong correlation between simple image quality measures and performance of the top three algorithms of the vendor test. Precisely, a correlation has been observed between the recognition rates and gradient energy of the sample image for either probe or gallery. While gradient energy is most commonly used as an indicator of whether image is in focus, this crude method can also be influenced by the presence of glasses, hair across face or harsh illumination.
- Performance on samples captured in indoor studio-like conditions is better than performance on 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 accept rates.
- 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. Finally, it has been consistently found in literature that samples obtained from individuals of a certain race perform better than others, with East-Asian races performing the best.
- The challenging dataset used during FRVT 2006 consists of 9307 frontal, neutral expression face images taken in indoor or outdoor settings, from 570 subjects. From this dataset, a subset of 2170 images from 437 subjects are 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 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.

Current literature describes the quality of a face image as an intrinsic property of the image. However, Beveridge et al. [104] 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. Beveridge et al. [103] also introduce the notion of measurable covariates of face biometric. Certain covariates may be difficult to measure, for example, the aesthetic changes to face brought about by a change in the hair style. Note that measurable covariates can be properties of the image (edge density measures) or of the subject (inter-eye distance). Image covariates such as size of face, focus of camera and subjective covariates such as expression, glasses that can be controlled to some extent are termed actionable covariates. Subjective non-actionable covariates include age, gender, and race. The focus of the research must hence be towards accurate assessment of measurable and actionable covariates. It can also be inferred that a comprehensive quality assessment framework must encode image properties such as focus, illumination and noise. Further, it must measure the applicability of the extracted features, perhaps from a general model, and must make itself relevant to the specific task for which it is employed.

#### B. Quality Framework for Biometrics

Different quality assessment algorithms in literature have some underlying similarities in their philosophy. It might be helpful to classify existing algorithms based on these underlying principles for a thorough understanding of the current state of research and gaps in literature. Several attempts have been made at this classification, Kalka et al. [67] classified iris quality assessment algorithms into global and local algorithms. Beveridge et al. [104] classify based on covariates of the biometrics. Inspired by the visual quality model of Yendrikhovskij [106], this research presents three orthogonal aspects of quality assessment in biometrics.

- 1) **Biometric Naturality:** *The degree of apparent match of the image with a viewer's internal references.* Most no-reference quality assessment algorithms, which usually correlate well with perceptual quality assessment, measure naturalness. They rely on unexpected changes in intensities or ration of information in various frequency bands, effects that tend to stand out in visual inspection of quality. Quality metrics in this dimension are adept at encoding degradation at image level, for example, illumination, compression, noise, focus, and blurring. These metrics are generally computationally inexpensive to compute, however, their performance is dependent on parameters that can be optimized with some knowledge of the intended application.
- 2) **Biometric Fidelity:** *The degree of apparent match of the acquired image and the biometric.* The quality or the

extent to which the presented biometric has been successfully captured in the representation is the measure of fidelity of the biometric. This is a difficult problem, as there is no independent method to verify the quality of the representation with the source itself. Hence, we use *biometric information* (BI) defined by Youmaran and Adler [91] in information theoretic terms as the *decrease in uncertainty of the identity of an individual* when using this representation. Here, BI is measured as a function of the properties of the entire population set.

- 3) **Biometric Utility:** *The degree of apparent suitability of the image with respect to a specific task.* The utility of a biometric sample is determined by the application. Biometric quality metric can be used for perceptual evaluation, feature-based context switching, cluster-based context switching or matching confidence. Utility is the combination of *Naturalness* and *Fidelity* of the biometric sample based on the desired application of the quality metric.

Alonso-Fernandez et al. [77] also use similar nomenclature to describe quality assessment *viewpoints*, from which they conclude that utility is of primary focus when considering quality of fingerprints. However, it is our assertion that in order to obtain a complete understanding of quality of a biometric, all three dimensions of quality must be considered for evaluation. This is more true for iris and face biometric, where the features are not structured or well-formed as compared to fingerprints.

### C. Holistic Image Representations

Holistic representation of images operate at a super-ordinate level features of the image. Research in scene recognition shows holistic representation of an image is very consistent in abstractly classifying images into broad labels. This research explores the possibility that such features can be adept at fast classification to quickly assess the biometric feature quality in terms of established training labels. First, a brief review of two prominent holistic representations is presented, namely, GIST and HOG.

Olivia and Torralba [107] have proposed a holistic representation of the spatial envelope of a scene image. Rather than viewing an image as a configuration of objects, in this model they are viewed as an unitary model. The spatial properties of the image are well preserved in such a representation of the spatial envelope (referred to as GIST). These coarse features are extracted at highly abstract level by using windowed Fourier transform. To assess the utility of a face biometric sample, we propose to use low dimensional representations of the face images. Here a set of five perceptual dimensions, namely, naturalness, openness, roughness, expansion and ruggedness are used to compute low dimensional, holistic representation of the image. The nomenclatures is derived from scene recognition research and we assert that GIST [107] can be a good descriptor for biometric quality assessment for face.

- 1) Degree of Naturalness: This spatial property describes the distribution of edges in the horizontal and vertical

orientations. It describes the presence of artificial elements such as spectacles.

- 2) Degree of Openness: The second major attribute describes the presence or lack of points of reference.
- 3) Degree of Roughness: This perceptual attribute refers to the size of the largest prominent object in the image. It evaluates the common global attributes of the image.
- 4) Degree of Expansion: This attribute describes the depth in the gradient of the space within the image.
- 5) Degree of Ruggedness: This attribute gives the deviation from horizontal by assessing the orientation of the contours of the image.

These perceptual properties are correlated with the second-order statistics and spatial arrangement of structured components in the image (for details of computing these properties, readers are referred to [107]). For a given face image  $I$  of size  $M \times M$  with  $O$  number of orientations per scale, GIST is defined as a function  $f$ ,

$$GIST_{M,O}(I) = f(N, O, R, E, Rg) \quad (24)$$

where  $N$  = Naturalness,  $O$  = Openness,  $R$  = Roughness,  $E$  = Expansion, and  $Rg$  = Ruggedness. Once the GIST descriptors are calculated, they are classified using a RBF-kernel based multi-class SVM and a quality class label  $C$  is assigned.

$$C = mSVM(GIST_{M,O}(I)) \quad (25)$$

Dalal and Triggs [108] present a global descriptor for human (pedestrian) detection in street view images, known as histogram of gradient orientations (HOG). The approach has gained immense popularity in detection of humans, vehicles and animals in still imagery and videos. This is due to the low computation time yet surprisingly high accuracy and robustness across different postures. The algorithm is based on the intuition that the shape and position of the dominant object can be understood by the distribution of orientations in local regions of the image. The research extensively describes implementation of HOG and empirically analyzes the effects of different parameters on performance. A color image is first pre-processed using gamma-correction. Unidirectional gradient kernel is applied on the image to obtain orientations. Histograms of these orientation angles are collected and normalized based on the gradient magnitude. Depending on the variant of HOG to be used, the obtained histograms are pooled over densely overlapping windows. The histograms are often normalized by  $k$ -norm operation given by  $\|v\|_k$  where  $k = 1, 2, 3$ . The normalized descriptor is used as features for classification. In this research, this descriptor is used to classify pose in conjunction with SVM.

### D. Experiments with Noisy Images

In an attempt to circumvent the cumbersome feature extraction process, low dimensional representations of the face images (GIST) is used for noise assessment. The experiments are conducted on a subset of the AR face database [109] containing 400 frontal face images pertaining to 35 subjects.

A symmetrically corrupted database is prepared that consists of eight classes of artifacts and one class representing uncor-

TABLE XIV  
DIFFERENT NOISE AND BLUR ARTIFACTS USED IN THE EXPERIMENTS  
WITH THEIR ASSOCIATED PARAMETERS.

Corruption	Parameters
White Noise	$\sigma = 0.05$
Local Variant White Noise	Dependent on local intensity
Poisson Noise	$\lambda=1$
Salt and Pepper Noise	$d = 0.05$ or 5%
Speckle	$v = 0.05$
Gaussian Blur	$\sigma = 1$
Motion Blur	$L= 1, r = 5$
Sharp	$\alpha = 0.1$

rupt images. The parameters used are presented in Table XIV. The experiment is conducted as described below.

- From the database, 50 images per artifact (corruption) class are chosen randomly for training the multi class SVM.
- GIST descriptors are computed using a bank of Gabor filters at eight orientations per scale. Image size is reduced to  $128 \times 128$  and a block size of four is used for the windowed Fourier transform
- Since the Gabor bank is computed only once for a given image size and parameters, the descriptor is computed quickly. The average time is 0.18 seconds per image using Matlab on a standard desktop PC.
- A one-versus-all SVM is used for classifying images into the nine quality bins based on the GIST features. Fig. 19 shows the confusion matrix and recognition accuracy per class of the multi class SVM classifier.
- These results indicate that the performance of the proposed method is suitable for identifying static and motion blur artifacts as well as distributed speckle noise. Further investigation on how the GIST descriptor is effected by noise can provide interesting insights towards assessment of quality.

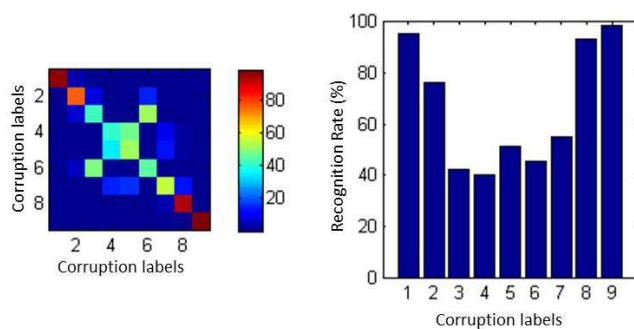


Fig. 19. a) Confusion matrix of the performance of GIST descriptor b) Accuracy of SVM in classifying each of the eight noise/blur artifacts and 1 uncorrupt class labels. Labels are defined as 1- Gaussian blur, 2- Motion blur, 3- Uncorrupt (original), 4- White noise, 5- Localvar noise, 6- Poisson noise, 7-Salt & Pepper noise, 8- Speckle noise, and 9- Sharp.

### E. Experiments with Pose

Pose estimation is a challenging problem in face recognition and several solutions have been proposed based on facial symmetry, orientation of nose region, shape of face, and 3D reconstruction. In this research, we present a simple learning based approach to pose estimation using the HOG descriptor. The experiment is conducted on the MultiPIE dataset [110], We use a sub-sample of nine viewpoints with all 10 illumination conditions and 4 sessions. The frontal illumination images of a subject in all nine viewpoints are shown in Fig. 20.

- From the MultiPIE dataset, 30% of the total subjects from session 1 are chosen for training. Further, only two randomly selected images per user, to avoid over-fitting. Hence, 2772 images are used in the training phase. The remaining 313560 images are used for testing.
- HOG descriptor is computed for all images, with 9 histogram bins and  $3 \times 3$  block size.
- A one-versus-all SVM is used for classifying images into one of the 9 classes based on the HOG features. The results are presented in Fig. 20.
- The results show excellent classification performance when pose estimation is viewed as a supervised learning problem using simple descriptors. Further, the nature of the result prompt towards broader class labels for higher accuracy. A modified approach known as PHOG (HOG descriptor over three levels of Gaussian pyramid (down sampling) and concatenates the features) is also used. This approach used with SVM and same training and testing samples, yields improved results as shown in Fig. 20.

Further, the pose estimation approach is evaluated with the smaller SC-Face dataset [111] in order to assess generality. This dataset also consists of high resolution face images in different pose angles.

- This database, consists of 130 subjects and 9 poses. All images of 13 subjects are used for training and the remaining 117 subjects for testing. All 9 poses from the dataset for a single subject are shown in Fig. 21(a).
- Similar to the previous experiment, a one-versus-all SVM is used for classifying each image into one of the nine classes based on the HOG features. The results are shown in Fig. 21(b). Further improvement is achieved with PHOG descriptors, as shown in Fig. 21(c).

These initial results on the use of holistic descriptors such as GIST and HOG for various image and biometric quality factors show promise towards a robust and computationally inexpensive solution to the important problem of quality assessment in face biometrics. By further evaluating the effect of each class label (of either noise or pose) on recognition accuracy, the techniques described above can be used for classifier performance prediction.

## VI. CONCLUSION

Quality assessment of biometric samples has received little attention by the biometrics research community. Often, no clear distinction is made between image and biometric quality. It is our assertion that quality metrics are an important ingredient to

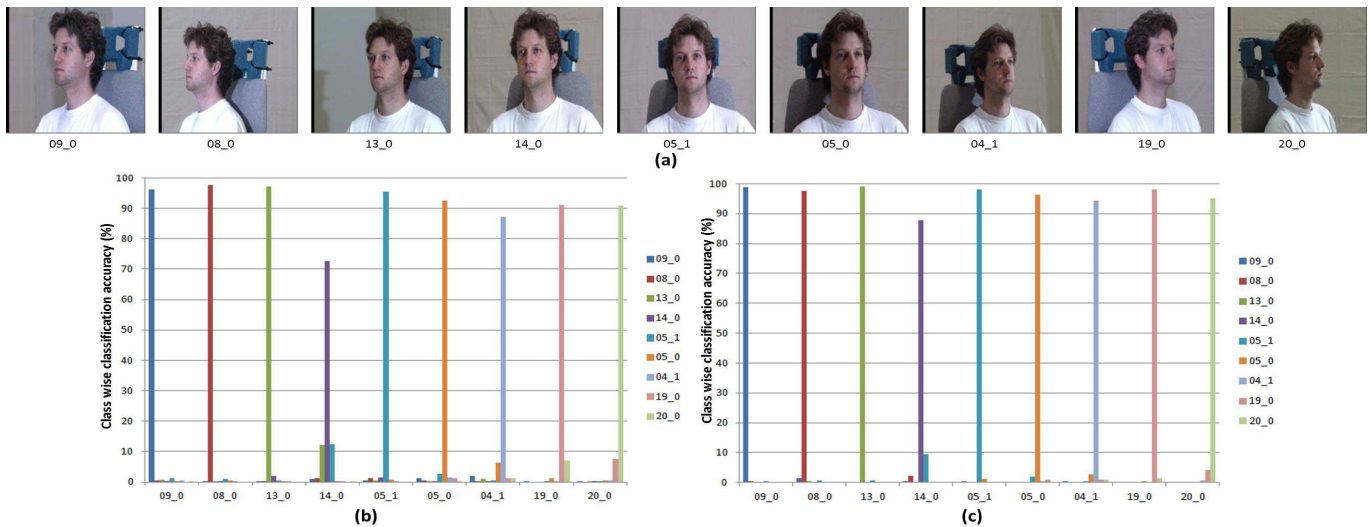


Fig. 20. a) Frontal illumination images of a subject in all nine viewpoints from the MultiPIE dataset, b) a bar graph of the HOG and SVM pose estimator, and c) similar graph with PHOG descriptor.

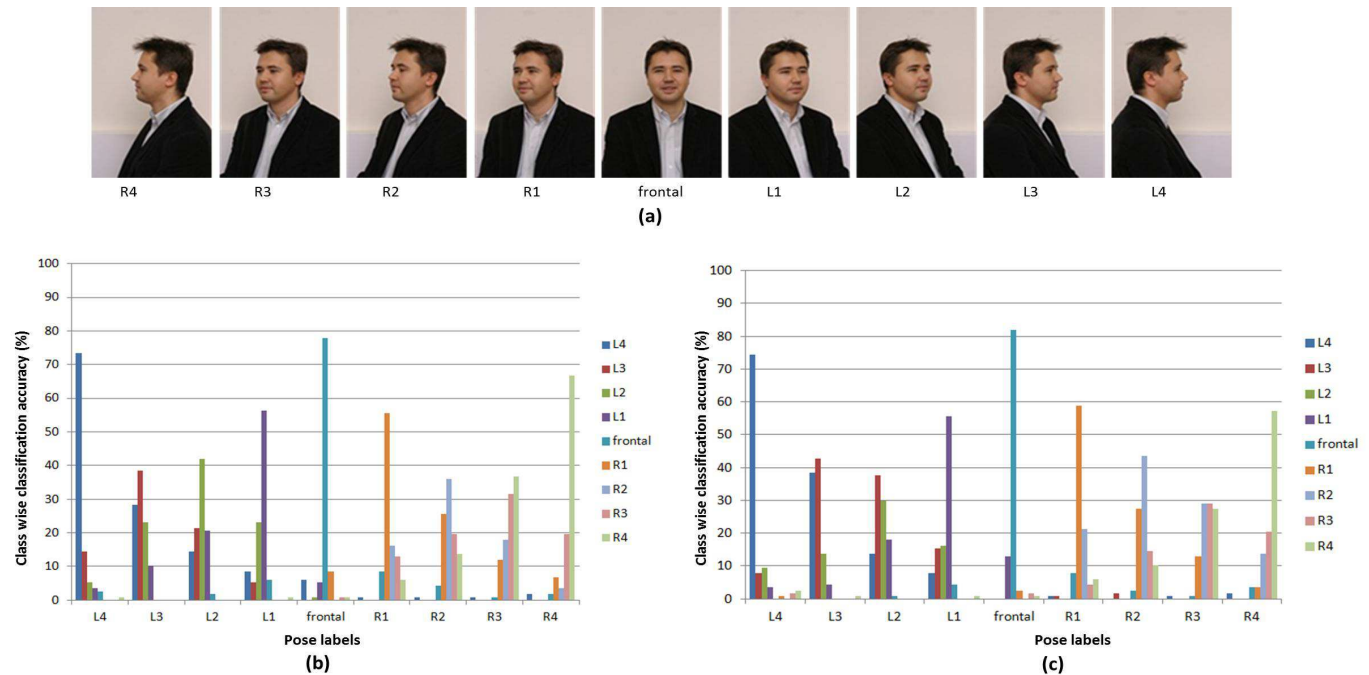


Fig. 21. a) All nine poses from the SC-Face dataset for a single subject, b) a bar graph of the HOG and SVM pose estimator on the SC-Face dataset, and c) A similar bar graph with PHOG and SVM

improve the robustness of large real-world biometric systems. In an attempt to demystify the definition and the working of biometric quality, three distinct quality dimensions, namely, Naturality, Fidelity and Utility are redefined specifically for biometrics.

Several gaps in quality assessment research exist for biometrics. Firstly, the definition of quality specific for biometrics must be established, attributing them more towards performance prediction of the classifiers than human perception evaluation. Secondly, the performance and methodology of the quality metric must entail the utility of the biometric as well. Finally, the performance in terms of computational complexity

must also be discussed more actively in research. A renewed interest in the development of quality assessment algorithms of biometric samples that are computationally inexpensive to compute yet correctly encode quality will be the linchpin of real-world and large scale biometric deployments.

## APPENDIX BIOMETRIC STANDARDS

A large number of commercial and public biometric systems/solutions has lead to standardization of several processes. This ensures inter-operability among different vendors and ensures easy integration. Here, some leading biometric standards

are presented [112], [113]:

- 1) **CBEFF**: The Common Biometrics Exchange File Format (CBEFF) [112], developed in 2001, facilitates exchange of biometric data including raw, 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 multi-modal or multi-sample systems. Within the BDB block, there is an optional field called Biometric Data Quality. The block provisions for a single scalar quantity (0-100) based on the ANSI/INCITS-358 standards of 2002 (discussed next). Additionally, the field also notes if the quality value is of a non-standard 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 [113], a structure called bioapi\_quality that indicates the quality of the biometric sample in the biometric identification record [113]. 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; 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 0 – 100 except as follows:

- value of –1: bioapi\_quality was not set by the vendor.
- value of –2: bioapi\_quality is not supported by the vendor.

quality scores in the range 0 – 100 have the following interpretation:

- 0 – 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 – 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 – 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 – 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**: Government of India has established biometric standards for identification and verification in various e-Governance applications [114]. 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) [70].

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

## REFERENCES

- [1] W. Zhao, R. Chellappa, P.J. Phillips, and A. Rosenfeld, “Face recognition: A literature survey”, *ACM Computing Surveys*, vol. 35, no. 4, pp. 399–458, 2003.
- [2] A.K. Jain, A. Ross, and S. Prabhakar, “An introduction to biometric recognition”, *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 1, pp. 4–20, 2004.
- [3] A.K. Jain, P. Flynn, and A. Ross, *Handbook of Biometrics*, Springer, 2008.
- [4] A. Ross, K. Nandakumar, and A.K. Jain, *Handbook of Multibiometrics*, Springer, 2006.
- [5] S.C. Dass, Y. Zhu, and A.K. Jain, “Statistical models for assessing the individuality of fingerprints”, in *Proceedings of IEEE Workshop on Automatic Identification Advanced Technologies*, 2005, pp. 3–9.
- [6] A.K. Jain, Y. Chen, and M. Demirkus, “Pores and ridges: High-resolution fingerprint matching using level 3 features”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 1, pp. 15–27, 2007.
- [7] J. Frenette, “Fingerprint patterns”, [http://odec.ca/projects/2004/fren4j0/public\\_html/fingerprint\\_patterns.htm](http://odec.ca/projects/2004/fren4j0/public_html/fingerprint_patterns.htm), 2004.
- [8] A.K. Jain and J. Feng, “Latent fingerprint matching”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 1, pp. 88–100, 2011.
- [9] B.Y. Hiew, A.B.J. Teoh, and Y.H. Pang, “Touch-less fingerprint recognition system”, in *Proceedings of IEEE Workshop on Automatic Identification Advanced Technologies*, 2007, pp. 24–29.
- [10] D. Maltoni, D. Maio, and A.K. Jain, *Handbook of Fingerprint*, Springer, 2006.

- [11] J. Daugman, "How iris recognition works", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 11, pp. 21–30, 2004.
- [12] J. Daugman, "Probing the uniqueness and randomness of iriscodes: Results from 200 billion iris pair comparisons", *Proceedings of the IEEE*, vol. 94, no. 11, pp. 1927–1935, 2006.
- [13] H. Proenca, S. Filipe, R. Santos, J. Oliveira, and L.A. Alexandre, "The ubiris.v2: A database of visible wavelength iris images captured on-the-move and at-a-distance", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 8, pp. 1529–1535, 2010.
- [14] T. Serre, L. Wolf, and T. Poggio, "Object recognition with features inspired by visual cortex", in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2005, vol. 2, pp. 994–1000.
- [15] N. Pinto and D.D. Cox, "Beyond simple features: A large-scale feature search approach to unconstrained face recognition", in *Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition*, 2011, pp. 8–15.
- [16] A.J. O'Toole, P.J. Phillips, F.Jiang, J.Ayyad, N.Penard, and H. Abdi, "Face recognition algorithms surpass humans matching faces over changes in illumination", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 9, pp. 1642–1646, 2007.
- [17] P. Sinha, B. Balas, Y. Ostrovsky, and R. Russell, "Face recognition by humans: Nineteen results all computer vision researchers should know about", *Proceedings of the IEEE*, vol. 94, no. 11, pp. 1948–1962, 2006.
- [18] P. Viola and M. Jones, "Robust real-time object detection", *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137–154, 2002.
- [19] Ming-Hsuan Y., D.J. Kriegman, and N. Ahuja, "Detecting faces in images: a survey", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 1, pp. 34–58, 2002.
- [20] S. Li and A.K. Jain, *Handbook of Face Recognition*, Springer, 2006.
- [21] S.G. Kong, J. Heo, B.R. Abidi, J. Paik, and M.A. Abidi, "Recent advances in visual and infrared face recognition: a review", *Computer Vision and Image Understanding*, vol. 97, no. 1, pp. 103–135, 2005.
- [22] A.F. Abate, M. Nappi, D. Riccio, and G. Sabatino, "2D and 3D face recognition: A survey", *Pattern Recognition Letters*, vol. 28, no. 14, pp. 1885–1906, 2007.
- [23] K.W. Bowyer, K. Chang, and P. Flynn, "A survey of approaches and challenges in 3D and multi-modal 3D and 2D face recognition", *Computer Vision and Image Understanding*, vol. 101, no. 1, pp. 1–15, 2006.
- [24] P.J. Phillips, W.T. Scruggs, A.J. O'Toole, P.J. Flynn, K.W. Bowyer, C.L. Schott, and M. Sharpe, "FRVT 2006 and ICE 2006 large-scale experimental results.", *IEEE Transactions On Pattern Analysis And Machine Intelligence*, vol. 32, no. 5, pp. 831–846, 2010.
- [25] X. Chen, P.J. Flynn, and K.W. Bowyer, "IR and visible light face recognition", *Computer Vision and Image Understanding*, vol. 99, no. 3, pp. 332–358, 2005.
- [26] G.L. Marcialis and F. Roli, "Fingerprint verification by fusion of optical and capacitive sensors", *Pattern Recognition Letters*, vol. 25, no. 8, pp. 1315–1322, 2004.
- [27] X. Liu and T. Chen, "Video-based face recognition using adaptive hidden markov models", in *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, 2003, vol. 1, p. 340.
- [28] M. Vatsa, R. Singh, A. Ross, and A. Noore, "Integrated multilevel image fusion and match score fusion of visible and infrared face images for robust face recognition", *Pattern Recognition - Special Issue on Multimodal Biometrics*, vol. 41, no. 3, pp. 880–893, 2008.
- [29] N.K. Ratha, J.H. Connell, and R.M. Bolle, "Image mosaicing for rolled fingerprint construction", in *Proceedings of IEEE/IAPR International Conference on Pattern Recognition*, 1998, vol. 2, pp. 1651–1653.
- [30] F. Yang, M. Paindavoin, H. Abdi, and A. Monopoli, "Development of a fast panoramic face mosaicking and recognition system", *Optical Engineering*, vol. 44, no. 8, pp. 087, 2005.
- [31] R. Singh, M. Vatsa, A. Ross, and A. Noore, "A mosaicing scheme for pose-invariant face recognition", *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 37, no. 5, pp. 1212–1225, 2007.
- [32] P.J. Burt and E.H. Adelson, "A multiresolution splines with application to image mosaics", *ACM transactions on Graphics*, vol. 2, no. 4, pp. 217–236, 1983.
- [33] A. Ross and R. Govindarajan, "Feature level fusion using hand and face biometrics", in *Proceedings of SPIE Conference on Biometric Technology for Human Identification*, 2005, vol. 70, pp. 196–204.
- [34] Y. Yao, X.Y. Jing, and H.S. Wong, "Face and palmprint feature level fusion for single sample biometrics recognition", *European Symposium on Artificial Neural Networks*, vol. 70, no. 79, pp. 1582–1586, 2007.
- [35] J. Yang, J. Yang, D. Zhang, and J. Lu, "Feature fusion: parallel strategy vs. serial strategy", *Pattern Recognition*, vol. 36, no. 6, pp. 1369–1381, 2003.
- [36] M.I. Ahmad, W.L. Woo, and S.S. Dlay, "Multimodal biometric fusion at feature level: Face and palmprint", in *Proceedings of International Symposium on Communication Systems Networks and Digital Signal Processing*, 2010, pp. 801–805.
- [37] J. Kittler, M. Hatef, R.P.W. Duin, and J. Matas, "On combining classifiers", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 3, pp. 226–239, 1998.
- [38] A. Ross and A. Jain, "Multibiometric systems", *Communications of ACM*, vol. 47, no. 1, pp. 34–40, 2004.
- [39] K. Nandakumar, Yi Chen, S.C. Dass, and A.K. Jain, "Likelihood ratio-based biometric score fusion", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 2, pp. 342–347, 2008.
- [40] J. Fierrez-Aguilar, J. Ortega-Garcia, J. Gonzalez-Rodriguez, and J. Bigun, "Discriminative multimodal biometric authentication based on quality measures", *Pattern Recognition*, vol. 38, no. 5, pp. 777–779, 2005.
- [41] A. Ross and A. Jain, "Information fusion in biometrics", *Pattern Recognition Letters*, vol. 24, pp. 2115–2125, 2003.
- [42] M. Vatsa, R. Singh, A. Noore, and S.K. Singh, "Belief function theory based biometric match score fusion: Case studies in multi-instance and multi-unit iris verification", in *Proceedings of IEEE International Conference on Advances in Pattern Recognition*, 2009, pp. 433–436.
- [43] M. Vatsa, R. Singh, and A. Noore, "Unification of evidence-theoretic fusion algorithms: A case study in level-2 and level-3 fingerprint features", *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, vol. 39, no. 1, pp. 47–56, 2009.
- [44] A. Abaza and A. Ross, "Quality based rank-level fusion in multibiometric systems", in *Proceedings of IEEE International Conference on Biometrics: Theory, Applications, and Systems*, 2009, pp. 1–6.
- [45] A. Kumar and S. Shekhar, "Personal identification using multibiometrics rank-level fusion", *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 41, no. 5, pp. 743–752, 2011.
- [46] A. Ross and R. Govindarajan, "Feature level fusion using hand and face biometrics", in *Proceedings of SPIE*, 2005, vol. 196, p. 5779.
- [47] C. Zhang, K. Uchimura, C. Zhang, and G. Koutaki, "3d face recognition using multi-level multi-feature fusion", in *Proceedings of Symposium on Image and Video Technology*, 2010, pp. 21–26.
- [48] S. Zhou, "Probabilistic recognition of human faces from video", *Computer Vision and Image Understanding*, vol. 91, no. 1-2, pp. 214–245, 2003.
- [49] A. Kale, A.K. Roychowdhury, and R. Chellappa, "Fusion of gait and face for human identification", in *Proceedings of IEEE International*

- Conference on Acoustics, Speech, and Signal Processing*, 2004, vol. 5, pp. V-901-4.
- [50] J. Daugman, "Combining multiple biometrics", <http://www.cl.cam.ac.uk/~jgd1000/combine/combine.html>, 2010.
- [51] F. Roli, J. Kittler, G. Fumera, and D. Muntoni, "An experimental comparison of classifier fusion rules for multimodal personal identity verification systems", 2002.
- [52] G. Marcialis, F. Roli, and L. Didaci, "Personal identity verification by serial fusion of fingerprint and face matchers", *Pattern Recognition*, vol. 42, no. 11, pp. 2807-2817, 2009.
- [53] M. Vatsa, R. Singh, A. Noore, and A. Ross, "On the dynamic selection of biometric fusion algorithms", *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 3, pp. 470-479, 2010.
- [54] H. Bhatt, S. Bharadwaj, M. Vatsa, R. Singh, A. Ross, and A. Noore, "A framework of quality-based biometric classifier selection", in *Proceedings of IEEE/IAPR International Joint Conference on Biometrics*, 2011.
- [55] F. Alonso-Fernandez, J. Fierrez, D. Ramos, and J. Gonzalez-Rodriguez, "Quality-based conditional processing in multi-biometrics: Application to sensor interoperability", *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, vol. 40, no. 6, pp. 1168-1179, 2010.
- [56] A. Kumar, V. Kanhangad, and D. Zhang, "A new framework for adaptive multimodal biometrics management", *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 1, pp. 92-102, 2010.
- [57] N. Poh and J. Kittler, "A unified framework for multimodal biometric fusion incorporating quality measures", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 99, pp. 1, 2011.
- [58] Z. Wang and A.C. Bovik, *Modern Image Quality Assessment*, vol. 2, Synthesis Lectures on Image, Video, and Multimedia Processing, 2006.
- [59] A.B. Watson, G.Y. Yang, J.A. Solomon, and J. Villasenor, "Visibility of wavelet quantization noise", *IEEE Transactions on Image Processing*, vol. 6, no. 8, pp. 1164-1175, 1997.
- [60] Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli, "Image quality assessment: from error visibility to structural similarity", *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600-612, 2004.
- [61] Z. Wang and A.C. Bovik, "A universal image quality index", *Signal Processing Letters*, vol. 9, no. 3, pp. 81-84, 2002.
- [62] Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli, "Image quality assessment: from error visibility to structural similarity", *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600-612, 2004.
- [63] P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi, "A no-reference perceptual blur metric", in *Proceedings of IEEE International Conference on Image Processing*, 2004, pp. 57-60.
- [64] X. Gao, W. Lu, D. Tao, and X. Li, "Image quality assessment based on multiscale geometric analysis", *IEEE Transactions on Image Processing*, vol. 18, no. 7, pp. 1409-1423, 2009.
- [65] Y. Chen, S. Dass, and A. Jain, "Fingerprint quality indices for predicting authentication performance", in *Audio and Video Based Biometric Person Authentication*, vol. 3546 of *Lecture Notes in Computer Science*, pp. 160-170, 2005.
- [66] H. Fronthaler, K. Kollreider, and J. Bigun, "Automatic image quality assessment with application in biometrics", in *Proceedings of IEEE Computer Vision and Pattern Recognition Workshops*, p. 30, 2006.
- [67] N. D. Kalka, J. Zuo, N. Schmid, and B. Kukic, "Estimating and Fusing Quality Factors for Iris Biometric Images", *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 40, no. 3, pp. 509-524, 2010.
- [68] A. Kumar and D. Zhang, "Improving biometric authentication performance from the user quality", *IEEE Transactions on Instrumentation and Measurement*, vol. 59, no. 3, pp. 730-735, 2010.
- [69] G. Patrick and T. Elham, "Performance of biometric quality measures", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 4, pp. 531-524, 2007.
- [70] Government of India, "Aadhaar project", <http://uidai.gov.in/>.
- [71] S. Bharadwaj, H. Bhatt, M. Vatsa, R. Singh, and A. Noore, "Quality assessment based denoising to improve face recognition performance", in *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition Workshops*, 2011, pp. 140-145.
- [72] N. Poh, T. Bourlai, J. Kittler, L. Allano, F. Alonso-fernandez, O. Ambekar, J. Baker, B. Dorizzi, O. Fatukasi, J. Fierrez, H. Ganster, J. Ortega-garcia, D. Maurer, A.A. Salah, T. Scheidat, and C. Vielhauer, "Benchmarking quality-dependent and cost-sensitive score-level multimodal biometric fusion algorithms", *IEEE Transactions on Information Forensics and Security*, vol. 4, no. 4, pp. 849-866, 2009.
- [73] K. Kryszczuk, J. Richiardi, and A. Drygajlo, "Impact of combining quality measures on biometric sample matching", in *Proceedings of IEEE International Conference on Biometrics: Theory, Applications, and Systems*, 2009, pp. 1-6.
- [74] K. Nandakumar, Yi Chen, A.K. Jain, and S.C. Dass, "Quality-based score level fusion in multibiometric systems", in *Proceedings of IEEE/IAPR International Conference on Pattern Recognition*, 2006, vol. 4, pp. 473-476.
- [75] "Biometric classifier update using online learning: A case study in near infrared face verification", *Image and Vision Computing*, vol. 28, no. 7, pp. 1098-1105, 2010.
- [76] F. Alonso-Fernandez, "Biometric quality assessment and its application in multimodal authentication systems", *Thesis*, 2007.
- [77] F. Alonso-Fernandez, J. Fierrez, J. Ortega-Garcia, J. Gonzalez-Rodriguez, H. Fronthaler, K. Kollreider, and J. Bigun, "A comparative study of fingerprint image-quality estimation methods", *IEEE Transactions on Information Forensics and Security*, vol. 2, no. 4, pp. 734-743, 2007.
- [78] A. Adler and T. Dembinsky, "Human vs. automatic measurement of biometric sample quality", in *Proceedings of Canadian Conference on Electrical and Computer Engineering*, 2006, pp. 2090-2093.
- [79] E. Lim, X. Jiang, and W. Yau, "Fingerprint quality and validity analysis", in *Proceedings of IEEE/IAPR International Conference on Image Processing*, 2002, vol. 1, pp. 469-472.
- [80] Federal Bureau of Investigation (FBI), "The Integrated Automated Fingerprint Identification System (AFIS)", [http://www.fbi.gov/about-us/cjis/fingerprints\\_biometrics/iafis/iafis](http://www.fbi.gov/about-us/cjis/fingerprints_biometrics/iafis/iafis).
- [81] L. Shen, A. Kot, and W. Koo, "Quality measures of fingerprint images", in *Proceedings of Audio and Video Based Biometric Person Authentication*, vol. 2091, pp. 266-271, 2001.
- [82] M. Vatsa, R. Singh, A. Noore, and M.M. Houck, "Quality-augmented fusion of level-2 and level-3 fingerprint information using DSM theory", *International Journal of Approximate Reasoning*, vol. 50, no. 1, pp. 51 - 61, 2009.
- [83] T.P. Chen, X. Jiang, and W.Y. Yau, "Fingerprint image quality analysis", in *Proceedings of IEEE International Conference on Image Processing*, 2004, vol. 2, pp. 1253-1256.
- [84] E. Tabassi, C.L. Wilson, and C.I. Watson, "Fingerprint image quality (NISTIR 7151)", [ftp://sequoyah.nist.gov/pub/nist\\_internal\\_reports/ir\\_7151/ir\\_7151.pdf](ftp://sequoyah.nist.gov/pub/nist_internal_reports/ir_7151/ir_7151.pdf), 2004.
- [85] Y. Chen, S. Dass, and A. Jain, "Localized iris image quality using 2-D wavelets", in *Proceedings of Advances in Biometrics*, vol. 3832, pp. 373-381, 2005.
- [86] H. Proena, "Quality assessment of degraded iris images acquired in the visible wavelength", *IEEE Transactions on Information Forensics and Security*, vol. 6, no. 1, pp. 82-95, 2011.
- [87] Z. He, T. Tan, Z. Sun, and X. Qiu, "Toward accurate and fast

- iris segmentation for iris biometrics”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 9, pp. 1670–1684, 2009.
- [88] Z. Jinyu and N.A. Schmid, “Global and local quality measures for nir iris video”, in *Proceedings of IEEE Computer Vision and Pattern Recognition Workshop*, 2009, vol. 0, pp. 120–125.
- [89] M. Subasic, S. Loncaric, T. Petkovic, H. Bogunovic, and V. Krivec, “Face image validation system”, in *Proceedings of International Symposium on Image and Signal Processing and Analysis*, 2005, pp. 30–33.
- [90] R.L.V. Hsu, J. Shah, and B. Martin, “Quality assessment of facial images”, in *Proceedings of Biometric Consortium Conference*, 2006, pp. 1–6.
- [91] R. Youmaran and A. Adler, “Measuring biometric sample quality in terms of biometric information”, in *Proceedings of Biometric Consortium*, 2006, pp. 1–6.
- [92] A.M. Martinez and A.C. Kak, “PCA versus LDA”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 2, pp. 228–233, 2001.
- [93] X. Gao, S. Li, R. Liu, and P. Zhang, “Standardization of face image sample quality”, in *Proceedings of Advances in Biometrics*, vol. 4642, pp. 242–251, 2007.
- [94] T. Ojala, M. Pietikainen, and T. Maenpaa, “Multiresolution gray-scale and rotation invariant texture classification with local binary patterns”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971–987, 2002.
- [95] G. Zhang and Y. Wang, “Asymmetry-based quality assessment of face images”, in *Advances in Visual Computing*, 2009, vol. 5876, pp. 499–508.
- [96] David G. Lowe, “Distinctive image features from scale-invariant keypoints”, *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.
- [97] Y. Wong, S. Chen, S. Mau, C. Sanderson, and B.C. 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*, 2011, pp. 74–81.
- [98] K. Nasrollahi and T. Moeslund, “Face quality assessment system in video sequences”, in *Biometrics and Identity Management*, vol. 5372 of *Lecture Notes in Computer Science*, pp. 10–18, 2008.
- [99] J. Long and S. Li, “Near infrared face image quality assessment system of video sequences”, in *Proceedings of International Conference on Image and Graphics*, 2011, pp. 275–279.
- [100] Y. Yao, B. Abidi, and M. Abidi, “Quality assessment and restoration of face images in long range/high zoom video”, in *Face Biometrics for Personal Identification*, R. Hammoud, B. Abidi, and M. Abidi, Eds., Signals and Communication Technology, pp. 43–60. Springer Berlin Heidelberg, 2007.
- [101] J. Zuo, F. Nicolo, N.A. Schmid, and H. Wechsler, “Adaptive biometric authentication using nonlinear mappings on quality measures and verification scores”, in *IEEE International Conference on Proceedings of Image Processing*, 2010, pp. 4077–4080.
- [102] J.R. Beveridge, G.H. Givens, P.J. Phillips, B.A. Draper, D.S. Bolme, and Y.M.Lui, “Focus on quality, predicting FRVT 2006 performance”, in *Proceedings of IEEE International Conference on Automatic Face Gesture Recognition*, 2008, pp. 1–8.
- [103] J.R. Beveridge, G.H. Givens, P.J. Phillips, B.A. Draper, D.S. Bolme, and Y.M.Lui, “FRVT 2006: Quo vadis face quality”, *Image and Vision Computing*, vol. 28, no. 5, pp. 732–743, 2010.
- [104] J.R. Beveridge, P.J. Phillips, G.H. Givens, B.A. Draper, M.N. Teli, and D.S. Bolme, “When high-quality face images match poorly”, in *Proceedings of IEEE Conference on Automatic Face Gesture Recognition and Workshops*, 2011, pp. 572–578.
- [105] P.J. Phillips, J.R. Beveridge, B.A. Draper, G. Givens, A.J. O’Toole, D.S. Bolme, J. Dunlop, Y.M. Lui, H. Sahibzada, and S. Weimer, “An introduction to the good, the bad,& the ugly face recognition challenge problem”, in *Proceedings of IEEE International Conference on Automatic Face Gesture Recognition and Workshops*, 2011, pp. 346–353.
- [106] S. Yendrikhovskij, “Image quality: Between science and fiction”, in *Proceedings of Image Processing, Image Quality, Image Capture, Systems Conference*, 1999, pp. 173–178.
- [107] A. Oliva and A. Torralba, “Modeling the shape of the scene: A holistic representation of the spatial envelope”, *International Journal of Computer Vision*, vol. 42, no. 3, pp. 145–175, 2001.
- [108] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection”, in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2005, vol. 1, pp. 886–893.
- [109] A.M. Martinez and R. Benevento, “The AR face database”, *CVC Technical Report #24*, 1998.
- [110] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker, “Multi-PIE”, *Image and Vision Computing*, vol. 28, no. 5, pp. 807–813, 2010.
- [111] M. Grgic, K. Delac, and S. Grgic, “SCface surveillance cameras face database”, *Multimedia Tools and Applications*, vol. 51, pp. 863–879, 2011.
- [112] “What is cbeff (common biometric exchange formats framework)?”, [http://www.ibia.org/cbeff/#what\\_is\\_cbeff](http://www.ibia.org/cbeff/#what_is_cbeff), 2004.
- [113] C. Tilton, “The BioAPI specification”, *National Standards Institute*, 2002.
- [114] Department of Information Technology, “Face image data standard for e-governance applications in india”, in *Government of India*, 2005.