

# Journal of Electronic Imaging

JElectronicImaging.org

## Performance evaluation of no-reference image quality metrics for face biometric images

Xinwei Liu  
Marius Pedersen  
Christophe Charrier  
Patrick Bours



Xinwei Liu, Marius Pedersen, Christophe Charrier, Patrick Bours, "Performance evaluation of no-reference image quality metrics for face biometric images," *J. Electron. Imaging* **27**(2), 023001 (2018), doi: 10.1117/1.JEI.27.2.023001.

# Performance evaluation of no-reference image quality metrics for face biometric images

Xinwei Liu,<sup>a,b,\*</sup> Marius Pedersen,<sup>b</sup> Christophe Charrier,<sup>a</sup> and Patrick Bours<sup>b</sup>

<sup>a</sup>Normandie University, UNICAEN, ENSICAEN, CNRS, GREYC, Caen, France

<sup>b</sup>NTNU—Norwegian University of Science and Technology, Gjøvik, Norway

**Abstract.** The accuracy of face recognition systems is significantly affected by the quality of face sample images. The recent established standardization proposed several important aspects for the assessment of face sample quality. There are many existing no-reference image quality metrics (IQMs) that are able to assess natural image quality by taking into account similar image-based quality attributes as introduced in the standardization. However, whether such metrics can assess face sample quality is rarely considered. We evaluate the performance of 13 selected no-reference IQMs on face biometrics. The experimental results show that several of them can assess face sample quality according to the system performance. We also analyze the strengths and weaknesses of different IQMs as well as why some of them failed to assess face sample quality. Retraining an original IQM by using face database can improve the performance of such a metric. In addition, the contribution of this paper can be used for the evaluation of IQMs on other biometric modalities; furthermore, it can be used for the development of multimodality biometric IQMs. © The Authors. Published by SPIE under a Creative Commons Attribution 3.0 Unported License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI. [DOI: [10.1117/1.JEI.27.2.023001](https://doi.org/10.1117/1.JEI.27.2.023001)]

Keywords: biometric; image quality assessment; face; performance evaluation; image-based attributes; multimodality.

Paper 170862 received Oct. 2, 2017; accepted for publication Feb. 5, 2018; published online Mar. 2, 2018.

## 1 Introduction

The face has become one of the most common and successful modalities for biometric recognition in the past decade.<sup>1</sup> As face recognition is a mature technology, it has been used in both government (e.g., Australian and New Zealand customs services called SmartGate, law enforcement agencies in the United States) and civilian (e.g., sorting photographs, security payment) applications. The study in face recognition is motivated by the need for reliable, efficient, and security recognition methods in order to perform better identification and forensic investigations.

However, face recognition is still a challenging issue when degraded face images are acquired.<sup>1</sup> In recent years, low-cost devices have enabled face recognition systems, and smartphone-based face recognition systems have received significant attention; such facts make it difficult to ensure the quality of face images. It has been proven that face sample quality has significant impact on accuracy of biometric recognition.<sup>2</sup> Low sample quality is a main reason for matching errors in biometric systems and may be the main weakness of some applications.<sup>2</sup> Biometric image quality assessment approaches are used for measuring image quality and they may help to improve system performance.

Recently, several standardizations on biometric sample quality have been finalized, especially for face modality: ISO/IEC JTC 1/SC 37 19794-5 Information technology—Biometrics—Biometric data interchange formats—Part 5: Face image data<sup>3</sup> and ISO/IEC TR 29794-5 Information technology—Biometric sample quality—Part 5: Face image data.<sup>4</sup> The standard in Ref. 3, requirements for the face image

data record format as well as the instruction of photographing high quality face images are presented. Some important aspects should be considered in order to meet the basic face image quality requirements: pose angle, facial expression, visibility of pupils and irises, focus, illumination, and so on. The standard in Ref. 4, definition and specification of methodologies for computation of objective, quantitative quality scores for facial images are proposed. Both image-based and modality-based face quality attributes are discussed in the standard.

Multimodality biometric recognition technologies become more and more popular in recent years.<sup>5</sup> However, biometric sample quality assessment methods that can be used for the evaluation of multimodality sample quality are rarely considered. It is necessary to investigate if it is possible to develop a quality metric that can assess the quality of biometric image samples from multiple modalities. Two kinds of quality attributes are usually considered when assessing biometric sample quality: image-based attributes and modality-based attributes. Image-based attributes are, for instance related to, contrast, sharpness, etc., which are presented in all image-based biometric modalities (e.g., face, iris, palm print, etc.). Modality-based attributes are dedicated for only one modality, such as pose symmetry in face biometric or eye reflection in iris biometric. Using image-based quality attributes in the quality assessment approaches make it possible to assess image-based multimodality biometric sample quality.<sup>5</sup> There are many existing image quality metrics (IQMs) that have been developed for the evaluation of natural image's quality.<sup>6</sup> Based on the availability of a reference image, IQMs can be classified into full-reference, reduced-reference, and no-reference methods.<sup>7</sup> Full-reference IQMs can assess the quality of images, which have original undistorted visual stimulus along with the distorted

\*Address all correspondence to: Xinwei Liu, E-mail: [xinwei.liu@unicaen.fr](mailto:xinwei.liu@unicaen.fr)

stimulus are available. Reduced-reference IQMs can assess the quality of images, which have the distorted stimulus and some additional information about the original stimulus available. No-reference IQMs can assess the quality of images, which have only the distorted stimulus available. According to the properties of face biometric images, only no-reference IQMs might be suitable for the assessment of face image quality. It is interesting to evaluate the performance of such kind of IQMs on face images in order to assess the possibility of developing image-based multimodality sample quality assessment metrics.

In this paper, we selected 13 no-reference IQMs to be evaluated. Three face recognition algorithms are used to evaluate biometric system performance. Face images from the GC<sup>2</sup> multimodality biometric database is used in this paper. The structure of the paper is described as follows. We first present related works and background. Then, the experimental setup followed by the experimental results and their analysis is introduced. At last the conclusion and future work are presented.

## 2 State of the Art

In this section, we present the state of the art about face image quality, redefined quality attributes for image-based biometric samples, and no-reference image quality assessment in face biometrics.

### 2.1 Face Image Quality

In Ref. 4, the face image quality is given by the relation to the use of facial images with face biometric systems. The use of low quality face images in face recognition affects the performance of the system. Currently, it is common to use image quality assessment approaches to evaluate face quality before face recognition. It can make face images differently by either applying image enhancement methods to improve image quality, choosing different recognition systems depending on face quality, or recapturing face images. Therefore, it is necessary to assess face image quality before the recognition process.

There are many factors that can affect face image quality and the performance of biometric systems. It is important to take into account image quality attributes that influence face quality. Both image-based and modality-based face quality attributes are presented in Ref. 4. Since we do not investigate modality-based attributes, we only consider the following image-based face quality attributes: (1) image resolution and size, (2) noise, (3) illumination intensity, (4) image brightness, (5) image contrast, (6) focus, blur, and sharpness, and (7) color.<sup>8</sup>

Based on the aforementioned image-based attributes, several IQMs are proposed in Ref. 4, including a noise estimation IQM,<sup>9</sup> which estimate an upper bound on the noise level from a single image based on a piecewise smooth image prior model and measured CCD camera response functions; a blur estimation IQM,<sup>10</sup> which is based on the analysis of the spread of the edges in an image for blur estimation; and a blocking artifacts estimation IQM,<sup>11</sup> which analyze blocking artifacts as components residing across two neighboring blocks and use one-dimensional pixel vectors made of pixel rows or columns across two neighboring blocks for distortion estimation.

### 2.2 Redefined Quality Attributes for Image-Based Biometric Samples

As introduced already, both image-based and modality-based quality attributes are considered in Ref. 4. Fingerprint, iris, or face images can be considered as different subspaces evolved at different places within the natural image space. Thus, using image-based quality attributes makes it possible to develop multimodality biometric sample quality assessment method. Liu et al.<sup>5</sup> suggest to employ five quality attributes when evaluating any kind of image-based biometric sample quality and they are based on the survey of state-of-the-art research works.<sup>4,12-15</sup> We apply four of them in this paper and these four image-based quality attributes and their definitions are given as

1. The contrast attribute has two aspects: local contrast and global contrast. The local contrast can be defined as the average difference between neighboring pixels' intensity. The global contrast is defined as the weighted sums of the overall local contrast for different resolutions. It is correlated to the "image contrast" attribute in Ref. 4.
2. The sharpness attribute is defined as the clarity of biometric sample structure and details. It is correlated to the "focus, blur, and sharpness" attributes in Ref. 4.
3. The luminance attribute can be defined as the intensity of the biometric sample illumination. It is correlated to the "illumination intensity" and "image brightness" attributes in Ref. 4.
4. The artifacts attribute is given as any undesired alteration in biometric sample introduced during its digital processing, such as noise, compression, and so on. It is correlated to the "noise" attribute in Ref. 4.

These are the most important image-based attributes for the evaluation of face image quality, and image-based multimodality biometric sample quality.

### 2.3 No-Reference Image Quality Assessment in Face Biometrics

In general, there are two types of no-reference IQMs: distortion-specific IQMs and generalized IQMs. Different IQMs have their pros and cons. Distortion-specific IQMs may have better performance only measuring given distortion. On the other hand, generalized IQMs can assess different types of distortions; however, they may not perform as good as distortion-specific IQMs for certain distortion. In addition, some IQMs are natural scene statistics (NSS)-based metrics and some of them have been trained on image quality databases. NSS-based IQMs can better assess the quality of natural images, and trained IQMs have better performance on images that similar to trained databases. Yet for the other types of images (nonnatural images such as synthetic ones), the performance of these IQM schemes may decrease.

There are several existing studies using no-reference IQMs to assess face sample quality. Abaza et al.<sup>16</sup> evaluated no-reference IQMs that can measure image quality factors in the context of face recognition. Then they proposed a face image quality index that combines multiple quality measures. Dutta et al.<sup>17</sup> proposed a data-driven model to predict the performance of a face recognition system based on image

quality features. They modeled the relationship between image-based quality features and recognition performance measures using a probability density function. Hua et al.<sup>18</sup> investigated the impact of out-of-focus blur on face recognition performance. Fiche et al.<sup>19</sup> introduced a blurred face recognition algorithm guided by a no-reference blur metric. From these studies we can see that no-reference IQMs can be helpful to assess the quality of face samples. The observed performance is comparable to some metrics proposed in Ref. 4 which are designed for specific face modality. However, the above-mentioned studies have two common shortages: (1) the image-based quality attributes in these studies do not cover all the five important attributes indicated in Liu et al.,<sup>5</sup> (2) the databases used in these studies contain both image-based distortions and modality-based distortions. Due to these two shortcomings, the performance of studied no-reference IQMs could be affected.

### 3 Experimental Setup

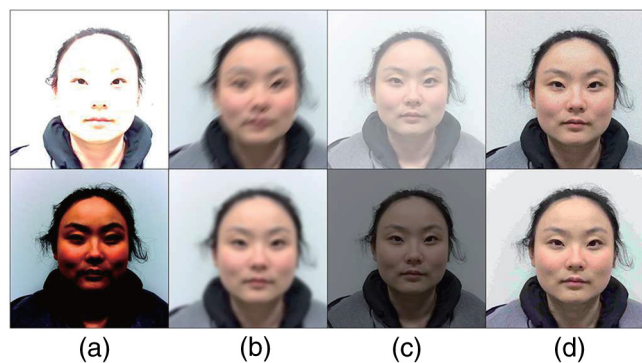
#### 3.1 Face Image Database: GC<sup>2</sup> Multimodality Biometric Database

If we want to benchmark no-reference IQMs, it is recommended to use a standard database in order to compare the results directly.<sup>2</sup> There are many existing databases in the research field, such as the color FERET database,<sup>20</sup> the Yale face database,<sup>21</sup> the AT&T face database (formerly the ORL database of faces),<sup>22</sup> and so on. For more face databases information, we refer to the face recognition homepage.<sup>23</sup> The choice of an appropriate database to be used should be made accordingly to our purpose. Since we focus on image-based quality attributes, we need to choose a specific face database that only contains image-based distortions but not including modality-based degradations. All existing face databases contain both image-based and modality-based degradations. Therefore, we create a new multiple modality biometric database named “GC<sup>2</sup> Multimodality Biometric Database.”<sup>8</sup> This database has three biometric modalities: face, contactless fingerprint, and visible wavelength iris. Three cameras are used for the acquisition: (1) a Lytro<sup>24</sup> first-generation light field camera (LFC) (11 megapixels), (2) a Google Nexus 5 embedded camera (8 megapixels), and (3) a Canon D700 with Canon EF 100 mm *f*/2.8L macro lens (18 megapixels). There are 50 subjects in the database. For the fingerprint modality, three fingers per hand and 15 sample images per finger per camera have been acquired. There are 13,500 fingerprint images in the database. For the iris modality, 15 iris samples per eye per camera have been acquired. There are 4500 iris images in the database. For the face modality, 2150 original face images are obtained in the database. In addition, we introduced different distortions to these original face images as described below. Therefore, totally 86,000 degraded face images are in the database. We only use the face modality in this paper. The acquisition is conducted in a normal office with normal luminance. The background is white and no modality-based distortions are contained in the database.

Since the face recognition application used in the paper only processes grayscale face images, we consider the four image-based quality attributes introduced in Sec. 2.2. In order to obtain image-based distortions correlated to these four attributes, we need to artificially degrade face images in the database. Inspired by the techniques used in CID:IQ

image quality database<sup>25</sup> and a similar study in biometric sample quality assessment,<sup>26</sup> we degrade face images into five degradation levels (one to five, from little degraded to highly degraded) for each distortion as the following (all image processing is conducted by using Matlab R2016 a):

- Contrast distortions. There are two kinds of contrast distortions: too low and too high contrast. We use MATLAB<sup>®</sup> function “ $J = \text{imadjust}(I, [low_{in}; high_{in}], [low_{out}; high_{out}])$ ”, which maps the values in  $I$  (original face image) to new values in  $J$  (degraded face image) such that values between  $low_{in}$  and  $high_{in}$  map to values between  $low_{out}$  and  $high_{out}$ . For low contrast, the  $low_{in}$  and  $high_{in}$  values are set to 0.2 and 1 for the degradation level 1 (little degraded),  $low_{out}$  and  $high_{out}$  values are set to 0 and 1. The  $low_{in}$  increases 0.1 for each degradation level and the other variables remain the same values. For high contrast, the  $low_{in}$  and  $high_{in}$  values are set to 0 and 0.8 for the level 1,  $low_{out}$  and  $high_{out}$  values are set to 0 and 1. The  $high_{in}$  decreases 0.1 for each degradation level and the other variables remain the same values. Examples of contrast degraded face images for degradation level 5 (highly degraded) are shown in Fig. 1(a).
- Sharpness distortions. We generate two sharpness distortions: motion blur and Gaussian blur. For motion blur, we use MATLAB<sup>®</sup> function “ $h = \text{fspecial}(\text{‘motion,’}, len, theta)$ ”, which returns a filter to the linear motion of a camera by  $len$  pixels, with an angle of  $theta$  degrees in a counterclockwise direction. The  $len$  value is set to 10 for degradation level 1 and  $theta$  is set to 45. The  $len$  increases 15 for each degradation level and the other variables remain the same values. For Gaussian blur, we use the function “ $h = \text{fspecial}(\text{‘gaussian,’}, hsize, sigma)$ ”, which returns a rotationally symmetric Gaussian lowpass filter of size  $hsize$  with standard deviation  $sigma$  (positive). The  $hsize$  value is set to [25 25] and the  $sigma$  is set to 2 for the degradation level 1. The  $sigma$  increases 2 for each degradation



**Fig. 1** Degraded face samples for degradation level 5. The first column represents too high contrast face image (upper) and too low contrast face image (lower); the second column represents motion blurred face image (upper) and Gaussian blurred face image (lower); the third column represents high luminance face image (upper) and low luminance face image (lower), and the last column represents face image contains Poisson noise (upper) and JPEG compressed face image (lower).

level and the  $hsize$  changes according to the value of  $\sigma$ . Examples of sharpness degraded face images for degradation level 5 are shown in Fig. 1(b).

- **Luminance distortion.** There are two kinds of luminance distortions: too low and too high luminance. We use MATLAB<sup>®</sup> function “ $J = imadjust(I, [low_{in}; high_{in}], [low_{out}; high_{out}])$ ,” again to simulate luminance distortions. For low luminance, the  $low_{in}$  and  $high_{in}$  values are set to 0 and 1,  $low_{out}$  and  $high_{out}$  values are set to 0 and 0.9 for degradation level 1. The  $high_{out}$  decreases 0.1 for each degradation level and the other variables remain the same values. For high luminance, the  $low_{in}$  and  $high_{in}$  values are set to 0 and 1,  $low_{out}$  and  $high_{out}$  values are set to 0.1 and 1 for degradation level 1. The  $low_{out}$  increases 0.1 for each degradation level and the other variables remain the same values. Examples of luminance degraded face images for level 5 are shown in Fig. 1(c).
- **Artifacts.** We introduce two artifacts to the face images: Poisson noise and JPEG compression artifacts. We use MATLAB<sup>®</sup> function “ $J = imnoise(I, 'poisson')$ ” to add Poisson noise for degradation level 1. We add another layer of Poisson noise for each degradation level. The JPEG compression ratio is set to 0.9 for degradation level 1. The ratio decreases 0.2 for each degradation level. Examples of face images having artifacts in level 5 are shown in Fig. 1(d).

For the example of face images in all five degradation levels, we refer the reader to Fig. 12 in Appendix A.

### 3.2 No-Reference IQMs and Their Classification

Based on the survey and the availability of the source codes, we selected 13 no-reference IQMs for the performance evaluation. These IQMs have high correlation with the image-based quality attributes.<sup>5</sup> We classify these IQMs into two categories: (1) distortion specific and (2) generalized purposes holistic IQMs. In each category, we separate IQMs into two groups: NSS-based and non-NSS-based IQMs. The classification of the selected IQMs is illustrated in Table 1. As mentioned in the introduction, we consider that face images evolved in a subspace of the whole natural images space. Thus, the relevance of using NSS-based IQM methods can be investigated.

### 3.3 Face Recognition System

The open source face recognition system used in this paper is “The PhD (Pretty helpful Development functions for) face recognition toolbox,”<sup>39</sup> which is a collection of MATLAB<sup>®</sup> functions and scripts for face recognition. The toolbox was

produced as a byproduct of Štruc and Pavešić’s<sup>40</sup> research work and is freely available for download. Three face feature extraction algorithms are used:

#### 3.3.1 Kernel Fisher analysis

This feature extraction algorithm uses only kernel Fisher analysis (KFA)<sup>41</sup> on the original image without Gabor filtering (GF) technique.

#### 3.3.2 Gabor filtering + kernel Fisher analysis

In this feature extraction algorithm, a bank of complex Gabor filters defined in the spatial and frequency domains will be constructed first. Then, the algorithm computes the magnitude responses of a face image filtered with a filter bank of complex Gabor filters. The magnitude responses of the filtering operations are normalized after downscaling using zero-mean and unit variance normalization.<sup>40</sup> After that they are converted as the feature vector. Before we use the feature vector to perform face recognition, a KFA<sup>41</sup> is applied to it. The KFA method first performs nonlinear mapping from the input space to a high-dimensional feature space, and then implements the multiclass Fisher discriminant analysis in the feature space. The significance of the nonlinear mapping is that it increases the discriminating power of the KFA method, which is linear in the feature space but nonlinear in the input space. The analyzed feature vector will be finally used for face recognition.

#### 3.3.3 Phase congruency + kernel Fisher analysis

The first step in this feature extraction algorithm is the same as GF + KFA, a bank of complex Gabor filters defined in the spatial and frequency domains will be constructed first. But then the algorithm computes phase congruency (PC) features from a face image using a precomputed filter bank of complex Gabor filters.<sup>42</sup> After that they are converted as the feature vector. The feature vector is employed KFA before used for face recognition.

As described already, three face feature extraction algorithms are used in the experiment: GF + KFA, KFA, and PC + KFA. The classification method used in this paper is based on the nearest neighbor classifier.<sup>40</sup> This classification method is capable of performing comparison similarity scores between two feature vectors.

### 3.4 Approaches for the Evaluation of Face Recognition System Performance

To evaluate the performance of face recognition systems, many measures exist. Among all of them, we can consider the histograms of comparison scores. They are obtained from the genuine (comparison between samples from the same subject) and imposter (comparison between samples from different subjects) comparisons for all image samples. In general, high quality biometric samples could generate relatively “good” genuine comparison scores (in our case, a score closer to 1 the more similar the two face samples), which are well separated from imposter comparison scores.<sup>13</sup>

An IQM is useful if it can at least give an ordered indication of an eventual performance.<sup>13</sup> Rank-ordered detection error trade-off (DET) characteristics curve is one of the most commonly used and widely understood method used to evaluate the performance of quality assessment approaches.

**Table 1** Classification of the selected IQMs.

IQMs	Distortion-specific	Generalized purposes
NSS	CONTRAST <sup>27</sup>	BIQI, <sup>28</sup> BLIINDS2, <sup>29</sup> BRISQUE, <sup>30</sup> ILNIQE2 <sup>31</sup>
Non-NSS	JNBM, <sup>32</sup> DCTSP, <sup>33</sup> SH, <sup>34</sup> SSH, <sup>35</sup> PWN <sup>36</sup>	AQI, <sup>37</sup> AQIP, <sup>37</sup> SSEQ <sup>38</sup>

The DET curve used here plots false none match rate (FNMR) versus false match rate (FMR).

Grother and Tabassi<sup>13</sup> proposed to use quality-bin-based approaches to evaluate the image quality assessment methods. They believe if a certain percentage of low quality samples are excluded from the dataset, the comparison score would become “better” (closer to 1 in our case) and the equal error rate (EER) (when FMR and FNMR are equal) would decrease. We use it as one of the methods to represent the performance of no-reference IQMs. Because the scale of the quality score for each IQM is different and the linearity of the score is unknown, thus, we omit the percentile low quality samples and keep 80%, 50%, and 20% of highest quality samples from each subject for each of the trial IQM.<sup>43</sup>

## 4 Experimental Results

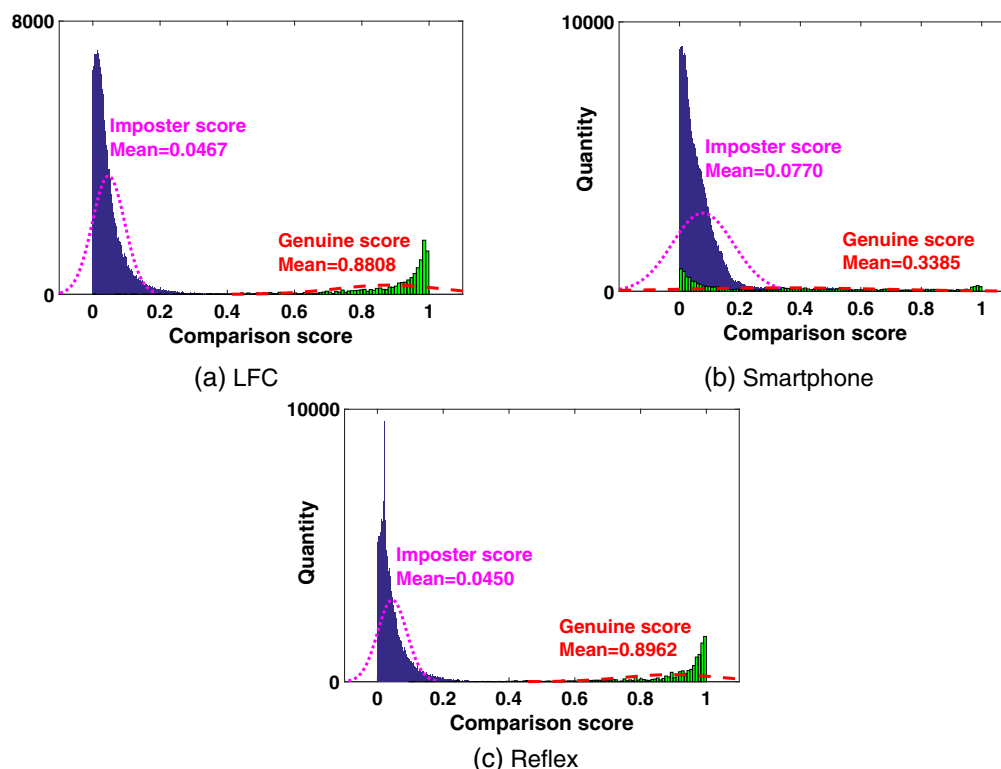
In this section, we only illustrate the results by using GF + KFA face recognition algorithm since the results from KFA and PC + KFA are very similar to GF + KFA. Since SH and SSH are similar metrics, we only show the results from SH.

### 4.1 Histogram of the Comparison Scores and Their Mean Values

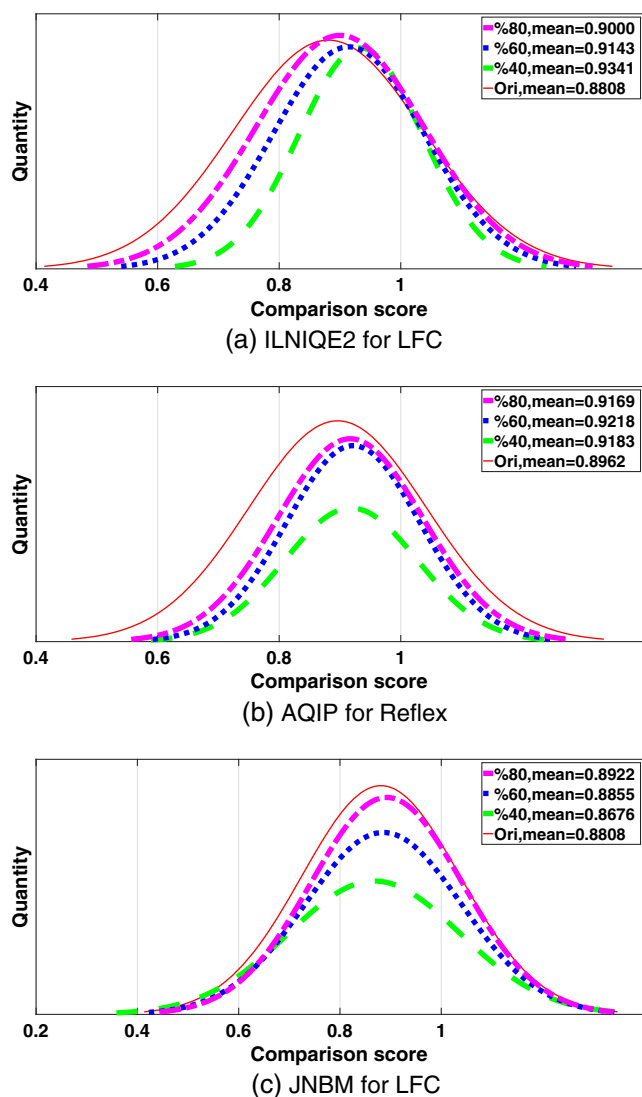
In order to evaluate the performance of the IQMs, we first plot the original comparison score by using GF + KFA recognition algorithm for three cameras in Fig. 2. The  $x$  axis represents the score and the  $y$  axis represents the quantity of the comparison. The line plots (red “- -” line for genuine comparison and magenta “-” line for imposter comparison) correspond to the fitted normal distributions. The mean value of the comparison score is given as well in Fig. 2. As mentioned before, high quality biometric samples could generate

relatively “good” genuine comparison scores (in our case, a score closer to 1 the more similar the two face samples), which are well separated from imposter comparison scores.<sup>13</sup> This can be observed from Figs. 2(a) and 2(c). However, the histograms of genuine and imposter scores as well as the mean values for smartphone are not well separated compared to LFC and reflex camera. It means that GF + KFA recognition algorithm cannot perform well on face images taken by smartphone in GC<sup>2</sup> multimodality biometric database. This can be due to the perspective distortion of the wide angle lens from smartphone. The perspective distortion can affect the performance of face recognition and needs to be compensated.<sup>44</sup> In addition, the face recognition system might be sensitive to the perspective distortion, which might be the reason why the genuine and imposter scores are not very well separated for smartphone compared to the other two cameras.

Here, we only illustrate the interesting examples in Fig. 3. The histogram of the genuine comparison score when omitting low quality samples by using 12 selected no-reference IQMs and GF + KFA recognition algorithm for three cameras are shown in Appendix B (Figs. 13–15). For each subplot in the figure, the red continuous line represents the original comparison score (the same fitted red line in Fig. 2); the magenta “-” line represents the comparison score when we omit 20% lowest quality face samples and keep the remaining 80% higher quality samples; the blue “-” line represents the comparison score when we omit 40% of the lowest quality face samples and keep the remaining 60% higher quality samples; and the green “-” line represents the comparison score when we omit 60% lowest quality face samples and keep only 40% highest quality



**Fig. 2** Comparison scores and their mean values for GF + KFA recognition algorithm for the three cameras.



**Fig. 3** Examples of comparison scores and their mean values with omitting low quality samples by using GF + KFA recognition algorithm. The  $x$  axis represents the comparison score, and the  $y$  axis represents the quantity of the genuine comparison.

samples in the database for the experiment. According to Grother's theory,<sup>13</sup> we expect to observe the fitted line moves from left to right (mean comparison value becomes closer to 1) when we keep 80%, 60%, and 40% highest quality samples.

In Fig. 3(a), by using the assessment results from ILNIQE2 (as well as BRISQUE and DCTSP in Fig. 13) for LFC (BLIINDS2, BRISQUE, ILNIQE2, and SSEQ for smartphone in Fig. 14; BIQI, BRISQUE, and DCTSP for reflex camera in Fig. 15) to omit low quality samples, we can observe the expected right shift for fitted lines (as well as the mean values). It means that these IQMs can assess face image quality and it is correlated with the performance of face recognition algorithm (GF + KFA). In Fig. 3(b), the mean values increase from keeping 80% to 60% highest quality samples by using AQIP metric for reflex camera; however, the values decrease when there is only 40% highest quality samples left. Similar observations can be found for the other two cameras. In Fig. 3(c), the mean values become

lower and lower than the original after omitting more and more low quality samples. This means that JNBM has reversed correlation with the performance of GF + KFA recognition algorithm for LFC.

In addition, we plot the mean values with omitting low quality samples in Fig. 4 for three cameras in order to show the overall performance of IQMs. The  $x$  axis represents the percentage of kept high quality samples and the  $y$  axis represents the mean of comparison score. The red “-” line represents the original mean of comparison score. Same findings can be obtained from Fig. 4. From the observation above, we can summarize that, based on mean comparison scores, only BRISQUE can assess face quality based on the performance of GF + KFA face recognition algorithm for all three cameras; DCTSP can assess face quality for LFC and reflex camera; ILNIQE2 can assess face quality for LFC and smartphone. The rest of the IQMs either can assess face quality based on the system performance. However, AQIP, CONTRAST, JNBM, and SH (for LFC); AQI, AQIP, CONTRAST, DCTSP, JNBM, and SH (for smartphone); AQI, CONTRAST, JNBM, and SH (for reflex camera) have reversed correlation with the performance of GF + KFA recognition algorithm according to the histogram and the mean of comparison scores.

#### 4.2 DET Curve and EER

As mentioned before, we also obtain EER as an indicator to examine the performance of IQMs. The DET curves with EER for data with and without omitting low quality face samples for three cameras by using all selected IQMs are given in Appendix C (see Figs. 16–18). Here, we only illustrate interesting examples in Fig. 5. For each subplot in Fig. 5, the red continuous line represents the original DET curve; the magenta “-” line represents the DET curve when we keep 80% highest quality face samples; the blue “:” line represents the comparison score when we keep 60% highest quality face samples; and the green “-” line represents the comparison score when we keep only 40% highest quality face samples in the database for the experiment. If a DET curve is closer to the top-right point, it means that this set of data leads to a higher face recognition performance. Meanwhile, the lower EER value the better the system performance.

From Figs. 5(a) and 5(b) we can see that, DET curves shift closer to top-right point when we keep 80%, 60%, and 40% highest quality samples using the assessment results from SSEQ (as well as BIQI and ILNIQE2 in Fig. 16) and DCTSP (as well as AQIP and BLIINDS2 for reflex camera in Fig. 18) to omit low quality samples taken by LFC and reflex camera, respectively. It means that such IQMs can assess face image quality and it is correlated with the performance of face recognition algorithm (GF + KFA). However, although the DET curves have no obvious shift but the EER values have decreased by using the assessment results from ILNIQE2 (as well as AQIP, BLIINDS2, and SSEQ) when we omit low quality samples for smartphone [see Fig. 5(c)]. In Figs. 5(d) and 5(e), EER values decrease when keeping 80% and 60% highest quality samples, but a slightly increasing can be observed when only 40% highest quality samples left in the database. We cannot conclude that such IQMs have “bad” performance because there is low percentage of

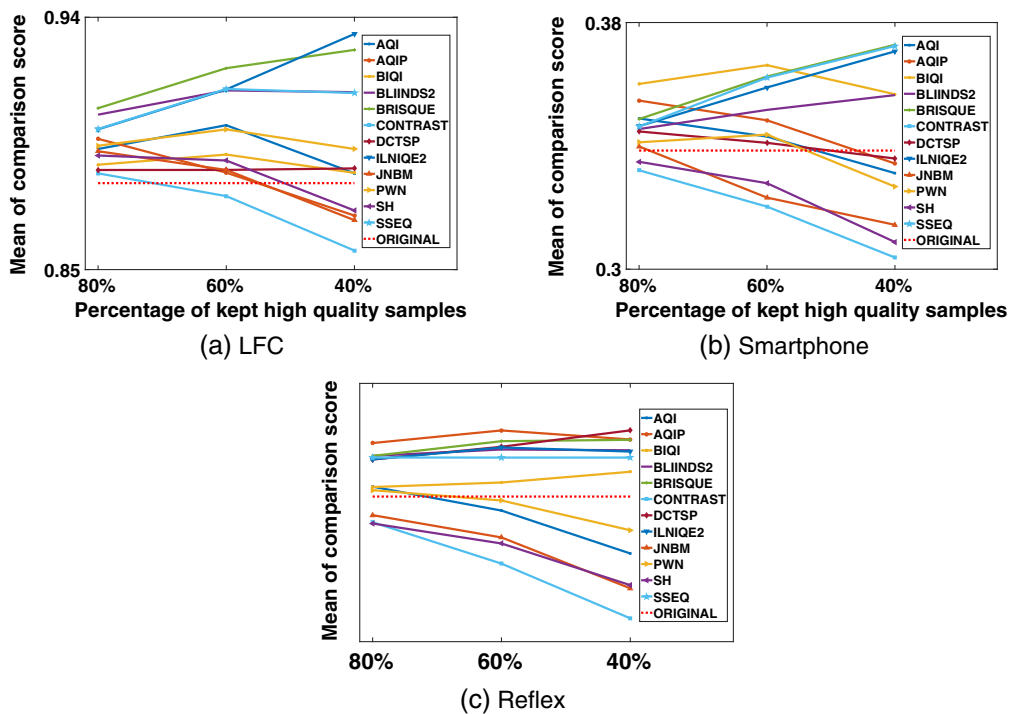


Fig. 4 Tendency of mean comparison scores with omitting low quality samples for GF + KFA recognition algorithm.

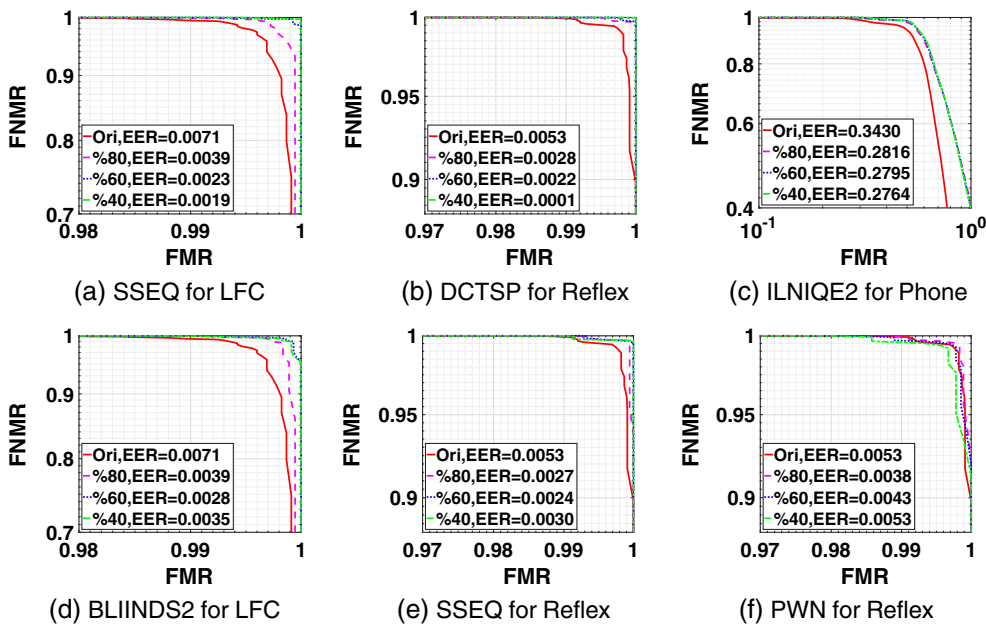


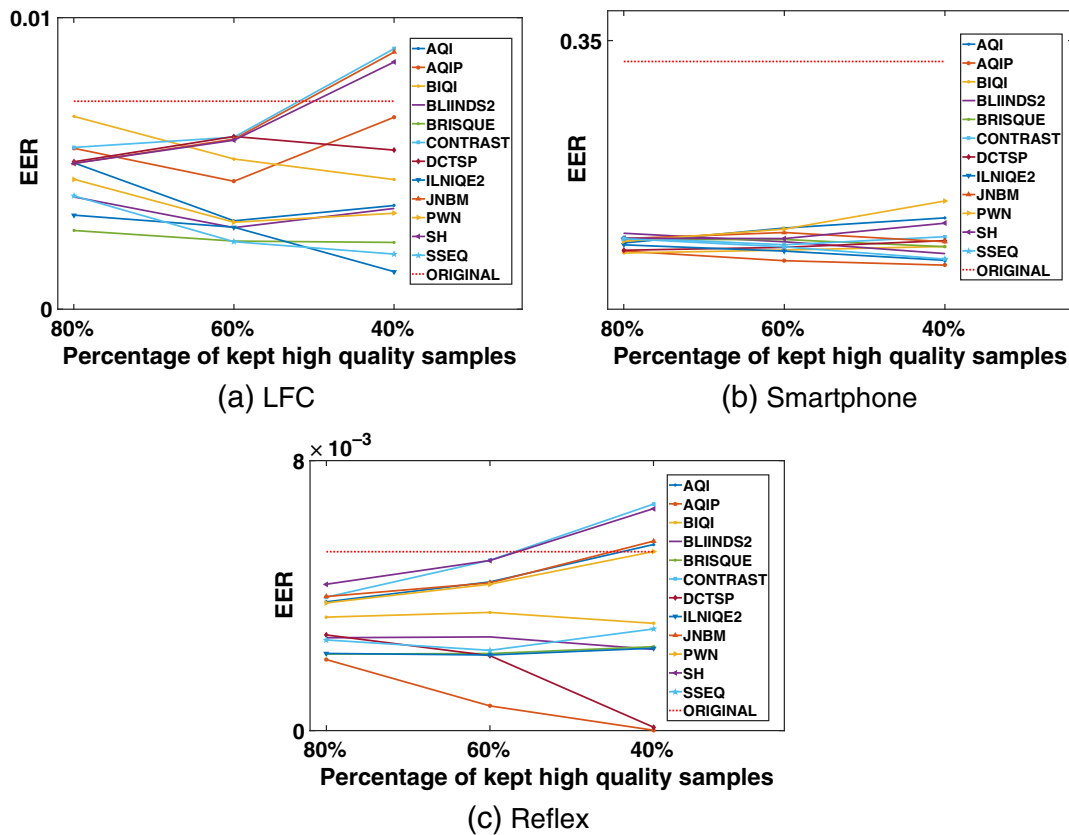
Fig. 5 Examples of DET curves with EER for comparison score with and without omitting low quality samples by using GF + KFA recognition algorithm.

samples left in the database. We can observe from Fig. 5(f) that PWN for reflex camera makes the DET curves shift to bottom-left point when omitting low quality samples. This means that such IQMs have reversed correlation with the performance of GF + KFA recognition algorithm.

In addition, we plot the tendency of EER values with omitting low quality samples in Fig. 6 for three cameras in order to show the overall performance of IQMs. The  $x$  axis

represents the percentage of kept high quality samples and the  $y$  axis represents the EER. The red “- -” line represents the original EER without omitting low quality samples. Same findings can be obtained from Fig. 6. From the observation above we can summarize that, based on DET curves and EER values, there is not a single IQM can assess face quality based on the performance of GF + KFA face recognition algorithm for all three cameras. However, ILNIQE2





**Fig. 6** Tendency of EER values with and without omitting low quality samples for GF + KFA recognition algorithm.

and SSEQ can assess face quality for LFC and smartphone; AQIP and BLIINDS2 can assess face quality for smartphone and reflex camera. The rest of the IQMs either can assess face quality for only one camera or have low ability to assess face quality based on the system performance. However, SH (for all three cameras), CONTRAST, and JNBM (for LFC and reflex camera), and AQI and PWN (for smartphone and reflex camera) have reversed correlation with the performance of GF + KFA recognition algorithm.

We also use EER values for all three cameras by omitting lowest quality face sample one by one until only one highest quality face sample left from each subject as another indicator to assess the performance of selected IQMs. The full plots are shown in Appendix D (see Figs. 19–21). Here we only give the interesting examples that illustrate the change of EER values. The  $x$  axis in Fig. 7 represents the number of omitted lowest quality samples unit. There are 40 units per captured sample image per subject (eight distortions in five levels). Each unit has 750 images (15 captured sample image per subject). The  $y$  axis represents the EER value. Here, we only illustrate the first omitted 30 units (75% of the entire number if unite) of low quality samples because when only a small part (25%) of images left in the database the change of EER values cannot be trusted. If the EER value has a smooth decreasing tendency when we omit lowest quality samples one by one, it means that the IQM used for generating the quality scores can predict the face recognition algorithm well which represents the high performance of such IQM. In Fig. 7(a) we can see that by using the assessment results from ILNIQE2 (as well as BLIINDS2, BRISQUE,

PWN, and SSEQ in Fig. 19) to omit one lowest quality sample (taken by LFC) each time, the EER curves have decreasing tendency (similar decreasing tendency for smartphone by using IQMs ILNIQE2, and AQIP, and DCTSP for reflex camera). However, in Fig. 7(b), the EER curve from BIQI for LFC camera has fluctuation all the time. But the overall trend of the curve is decreasing and the EER value became very low in the end. On the other hand, as shown in Fig. 7(c), the EER curve from DCTSP for smartphone has fluctuation all the time as well. But the overall trend of the curve is increasing and, in the end, the EER value became higher than the values in the very beginning of the curve. Finally, if we see Fig. 7(d), the EER values seem to increase when we use AQI to assess face image quality for reflex camera. From the observation above we can summarize that, based on EER values with omitting low quality face samples one by one until the best quality sample left for three cameras by using GF + KFA recognition algorithm, ILNIQE2 can assess face image quality for three cameras. The rest of the IQMs can either assess face quality for only one camera or have low ability to assess face quality based on the system performance. However, CONTRAST, JNBM, and SH have reversed correlation with the performance of GF + KFA recognition algorithm according to the results obtained from one by one omitted EER values.

### 4.3 Spearman Correlation between IQMs

Finally, we compute the Spearman correlation of quality scores between IQMs, which have a better performance as discussed above (AQIP, BLIINDS2, ILNIQE2, and SSEQ).

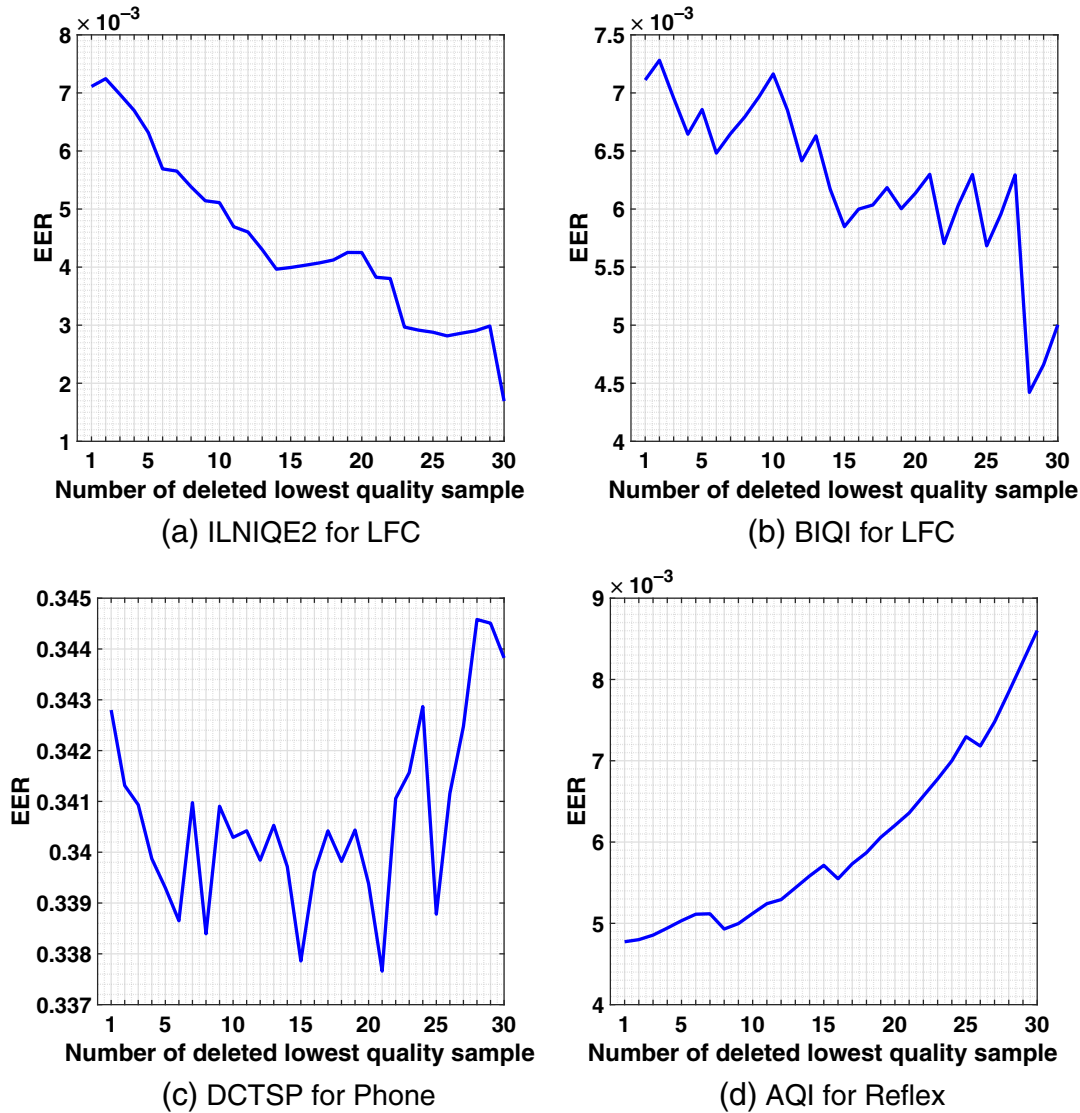


Fig. 7 Examples of EER values with omitting low quality face samples one by one until the best quality sample left by using GF + KFA recognition algorithm.

From the correlation coefficients, we can analyze whether each IQM gives the similar quality score to the same face sample image. The correlation coefficients are illustrated in Table 2. We show the coefficients for LFC in the first value in each cell, the coefficients for smartphone in the second value in each cell, and the last value in each cell for reflex camera. In Table 2 we can see that none of the IQM has high correlation with each other. It means that these

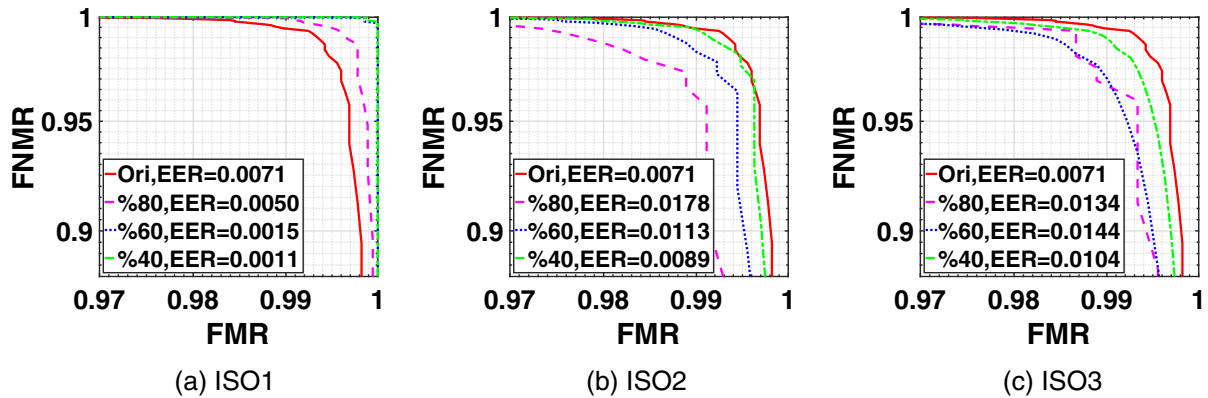
IQMs were designed for different types of distortions and they cannot be replaced by other selected IQMs.

#### 4.4 Performance Comparison between Selected IQMs and ISO-Proposed IQMs

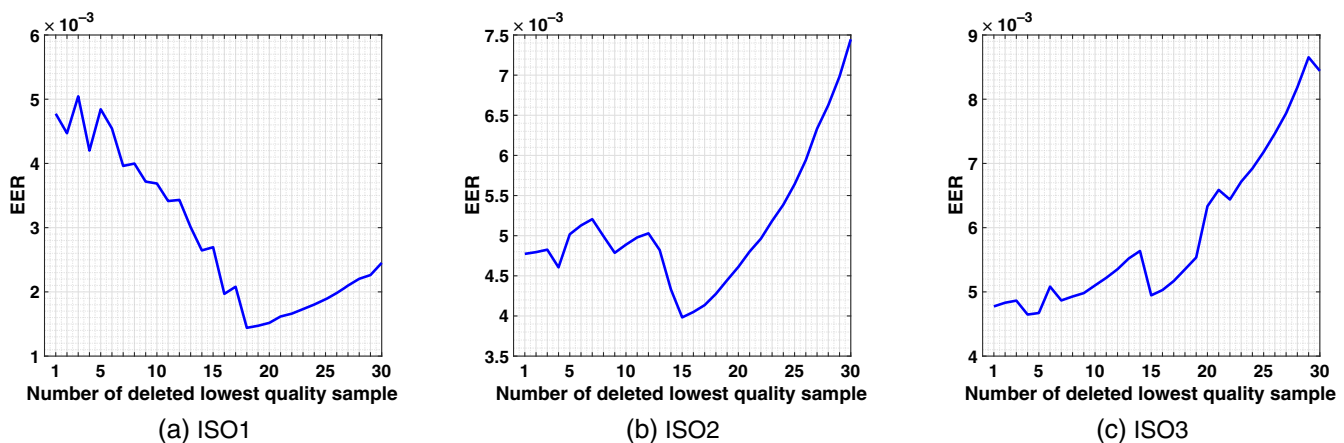
As mentioned in Sec. 2.1, several IQMs are proposed in Ref. 4. Here we compare the performance of selected IQMs and ISO-proposed IQMs: ISO1,<sup>9</sup> ISO2,<sup>10</sup> and ISO3.<sup>11</sup> In

Table 2 Spearman correlation of quality scores between IQMs. The first value in each cell represents LFC, the second value represents smartphone, and the last value represents reflex camera.

	AQIP	BLIINDS2	ILNIQE2
BLIINDS2	0.0176, 0.0406, 0.0845	1, 1, 1	0.1285, 0.0449, 0.0147
ILNIQE2	0.0703, 0.0903, 0.1138	0.1285, 0.0449, 0.0147	1, 1, 1
SSEQ	0.0185, 0.1705, 0.1589	0.1423, 0.0873, 0.1063	0.2691, 0.2623, 0.1658



**Fig. 8** DET curves with EER for comparison score with and without omitting low quality samples for LFC by using ISO-proposed IQMs (GF + KFA).



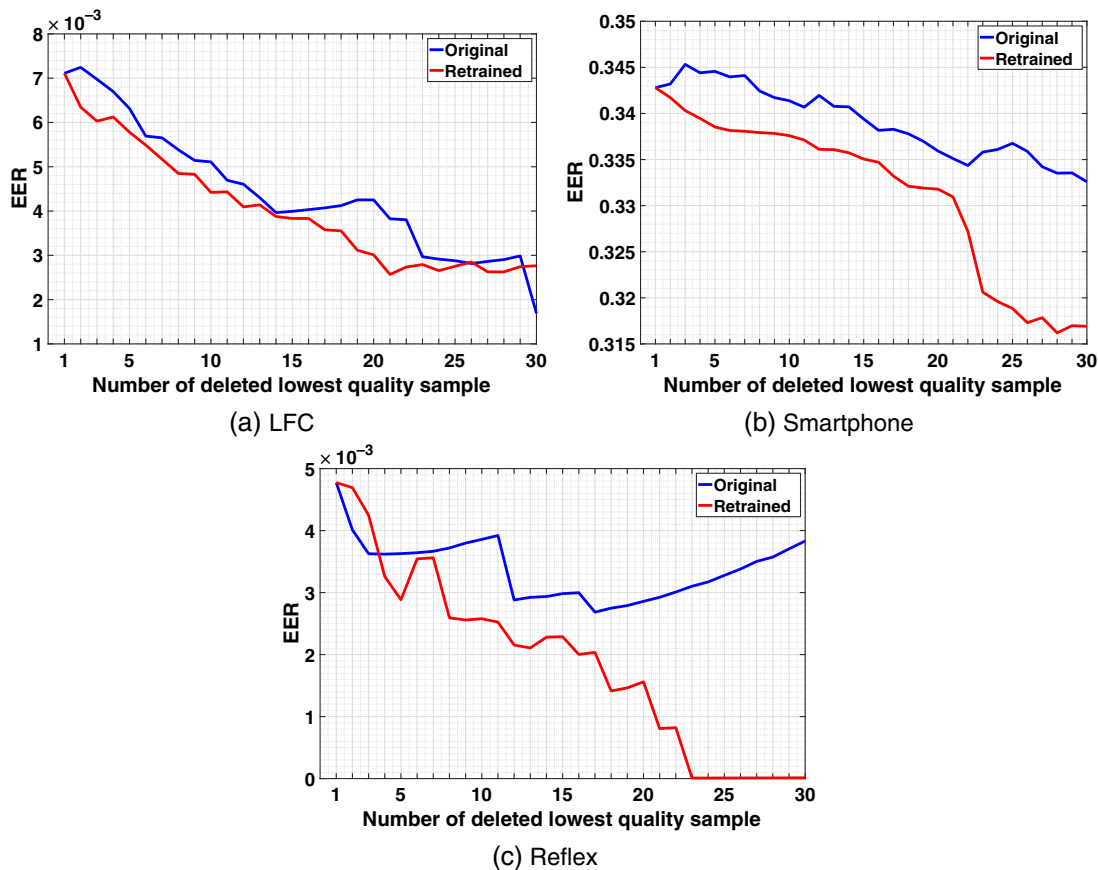
**Fig. 9** EER values with omitting low quality face samples one by one until the best quality sample left for reflex camera by using ISO-proposed IQMs (GF + KFA).

Fig. 8 we illustrate the DET curves with EER for LFC by using ISO-proposed IQMs to omit low quality samples. From Fig. 8 we can see that only ISO1 can give expected DET shift and EER decrease when we omit low quality samples in the dataset. The other two IQMs seem to have reversed correlation with the performance of GF + KFA recognition algorithm. Similar observations can be found in Fig. 9, which represent EER values with omitting low quality face samples one by one until the best quality sample left for reflex camera by using ISO-proposed IQMs. Therefore, compared to the performance evaluation results for selected no-reference IQMs, we do see there are several metrics have better performance than ISO-proposed metrics.

#### 4.5 Retraining ILNIQE2 on Face Database

From previous results, one deduced that performances of NSS-based IQM methods are competitive to predict the quality of face images guaranteeing a high level of performance of biometric systems. One specific metric, namely ILNIQE2 shows interesting results in terms of correlation between the provided quality scores and the performance results. Since this quality index has been trained on general purpose natural images, it would be interesting to investigate if results can be

improved retraining it on face images. To perform the retraining, the color FERET database<sup>20</sup> has been selected, which has 269 subjects and there are two acquisition sessions for most of subjects. For each session, 11 different sample images were acquired which contain different face angles and expressions. We use 269 images [one sample image (the front face) per subject] from the FERET database to retrain the ILNIQE2 IQM. These 269 images are all high quality face images because the ILNIQE2 metric only requires pristine images for training. The retrained metric is then used to reconstruct the experiment removing lowest quality samples one by one from each subject. The plots of EER values for three cameras are shown in Fig. 10. The blue thin lines represent the original ILNIQE2 method, and the red bold lines represent the retrained ILNIQE2 method. From Fig. 10 we can see that, after the retraining process, the overall performance of the IQM is improved because the red bold lines are under the blue thin lines. It means that the overall EER values from the retrained method are lower than the original method. In addition, the improvements for smartphone and reflex camera are greater than LFC, especially for reflex camera. By using the original ILNIQE2 to omit lowest quality samples from the database,



**Fig. 10** Comparison of EER by omitting lowest quality sample one by one using ILNIQE2 for each subject between the original method and the retrained method.

the EER values are not smoothly decreasing, moreover, the EERs increase after 18 units of lowest samples are removed for reflex camera. However, by using the retrained method, the line becomes smoother and has a decreasing tendency. Finally, the EERs reach “0” when 24 units of lowest quality samples are omitted. The difference of EERs between the original and the retrained method for reflex camera is obvious.

We would like to investigate if such improvement for reflex camera is due to the better prediction of retrained ILNIQE2 for all distortions or for some distortions. Therefore, we illustrate EER curves for single distortions in Fig. 11. Since there are eight distortions for each sample image, the total units become five instead of 40 as the case was in Fig. 10. Here we only illustrate the 80% (four units) of the EERs because when only a small part of images (e.g., 20%) left in the database, the EER is not convincing. From Fig. 11 we can see that the retraining process has little impact on high luminance and JPEG artifacts distortions. It reduces the performance of ILNIQE2 for low contrast and low luminance distortions because the average EERs are higher after retraining, but the EER curves still have a decreasing tendency. Furthermore, the retraining has a positive effect for high contrast, Gaussian blur, motion blur, and Poisson noise distortions. The EER curves for the latter three distortions have an increasing tendency for the original method; however, they have a decreasing trend after the retraining process. It is worth noting that all curves after retraining have a decreasing tendency.

## 5 Discussion

From the overall point of view, all the selected IQMs decrease the EER when keeping 80% and 60% high quality samples in the database according to their quality assessment scores (see Fig. 6). The expected outcome is that when more low quality face samples are omitted the EER should decrease continuously. However, two kinds of unexpected outcomes are observed: (1) EER increases when more low quality samples are omitted but the EER, which is computed from the last 20% high quality face samples is still lower than the EER computed from the entire database (AQIP and DCTSP for LFC; AQI, PWN, and SH for smartphone; BRISQUE, PWN, and SSEQ for reflex camera); and (2) EER increases when omitting low quality face samples but the EER, which is computed from the last 20% high quality face samples becomes higher than the EER computed from the entire database (CONTRAST, JNBM, and SH for LFC; AQI, CONTRAST, JNBM, and SH for reflex camera). IQMs that do not belong to these two cases are then have better performance. In addition, IQMs in case (2) have lower ability to predict the performance of selected face recognition systems compared to the IQMs in case (1). When we compare the EER by omitting lowest quality sample one by one until only the highest quality face sample left from each subject we can see that, it is difficult to have a very smooth gradual declining curve. However, by using some of the IQMs to omit the lowest quality samples, we can observe that the EER curves have an obvious tendency to drop. These IQMs are BLIINDS2, BRISQUE, ILNIQE2, and SSEQ for LFC;

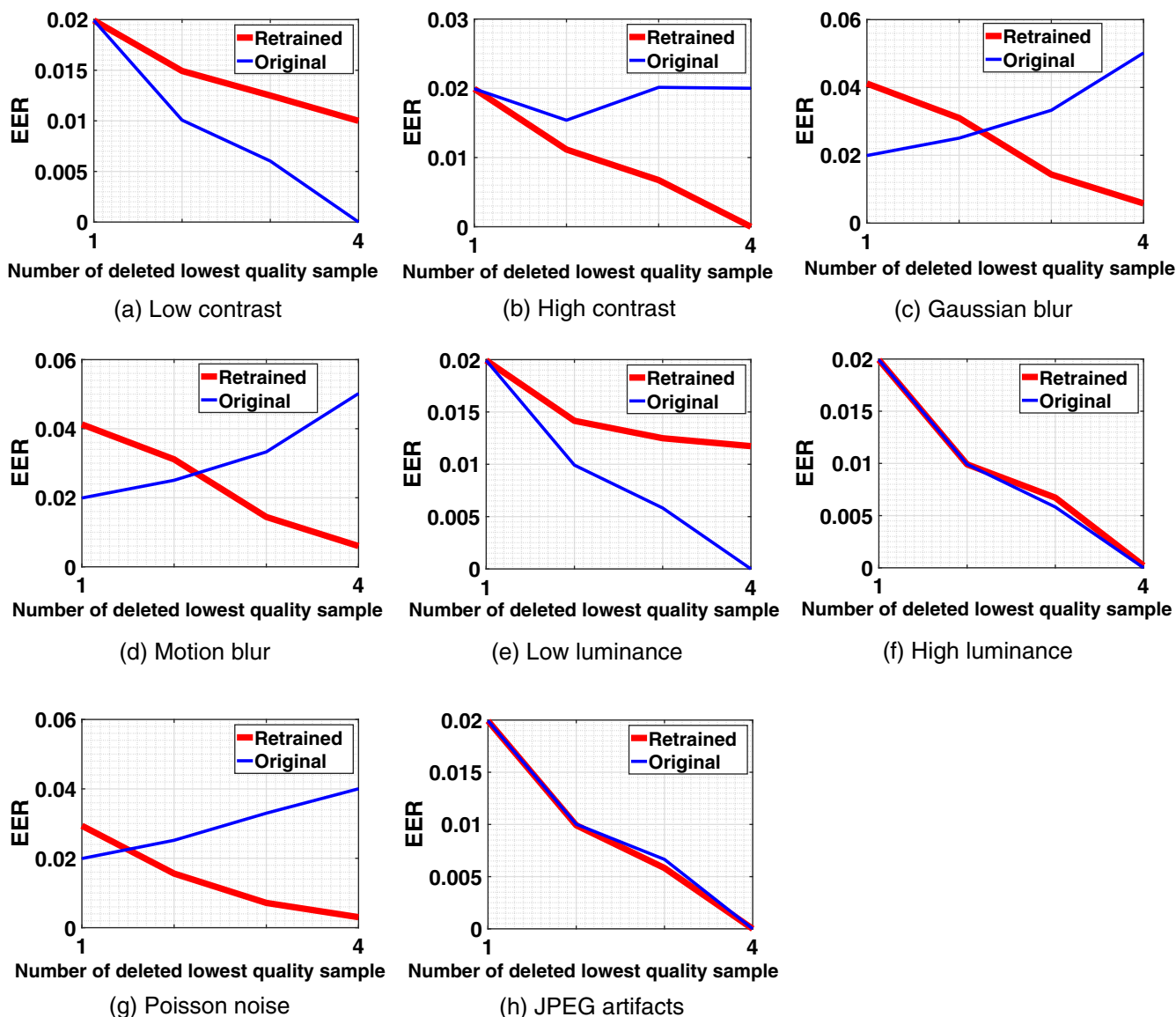


Fig. 11 Comparison of EER between the original method and the retrained method for each distortion for reflex camera.

ILNIQE2 for smartphone; AQIP and DCTSP for reflex camera. These IQMs have better overall performance than the others and they can be used for the development of new image-based multimodality biometric sample quality assessment method. In addition, although some curves may have fluctuations at some point, the general trend is still decreasing: BIQI and PWN for LFC; BIQI for smartphone; BRISQUE, ILNIQE2, and SSEQ for reflex camera. In order to improve the performance of these IQMs, an optimization process needs to be conducted. On the other hand, some IQMs lead to the gradually increasing EER curves when omitting lowest quality face samples. This outcome is the opposite of our expectation. Such IQMs are CONTRAST, JNBM, and SH for LFC; AQL, CONTRAST, JNBM, PWN, and SH for smartphone; AQL, CONTRAST, JNBM, PWN, and SH for reflex camera. They have reversed correlation with the performance of selected face recognition algorithms.

Based on the experimental results discussed we can summarize that ILNIQE2 has an overall better performance

than the other selected IQMs for all three cameras. It gives obvious decreased EER curves for LFC and smartphone when omitting lowest quality samples one by one. Its EER curve for reflex camera has fluctuations in the middle, but we can still see a decreasing trend. Several IQMs have better performance than the others for at least one camera: BLIINDS2, BRISQUE, SSEQ, AQIP, and DCTSP. Including ILNIQE2, most of these better performing IQMs are from the generalized purposes holistic category according to our classification in Table 1, except DCTSP. We introduced eight different distortions to the face sample images, and for each distortion we have five different levels of degradation. Therefore, it is not difficult to understand why some generalized purposes holistic IQMs have better performance. This is mainly due to the fact that their design is to assess the quality of an image that contains unknown and multiple distortions. However, BLIINDS2, BRISQUE, SSEQ, AQIP, DCTSP, and even for original ILNIQE2 can neither obtain a very smooth

gradually declining EER curve nor perform well for all three cameras. One of the reasons could be that most generalized purpose holistic IQMs are usually trained on natural image databases, for instance, BLIINDS2, BRISQUE, and ILNIQE2 are trained on the LIVE database.<sup>45</sup> However, not all types of distortions for image-based attributes in our dataset are introduced in the LIVE database, for example, motion blur and contrast changes are not included in the LIVE database. In addition, BLIINDS2, BRISQUE, and ILNIQE2 are also NSS-based IQMs. Face images are a subcategory of natural images so we may expect that these generalized purposes holistic and NSS-based IQMs can fail in some conditions.

As we can also see from the experimental results, by using the quality assessment scores from some IQMs to omit low quality samples, the EER increases instead of decreasing. It means that such IQMs have reversed correlation with the performance of selected face recognition system. These IQMs are CONTRAST, JNBM, and SH (for all three cameras); AQI and PWN (for smartphone and reflex camera). Except AQI, all these IQMs are from distortion-specific category, which are designed for the measurement of single type of distortion, such as JPEG or JPEG2000 compression distortions. Since these IQMs are tested under the condition of images containing only single type of distortion, they may not predict well the quality of the face samples that contain multiple distortions. If we look at the EER curves when omitting lowest quality face samples one by one using CONTRAST for LFC and smartphone, we can find that the curves have a declining tendency at a certain interval (in the beginning for LFC, in the middle for smartphone). CONTRAST is a metric used for the measurement of contrast degradation and at some point the face sample images that have contrast distortions are starting to be omitted. Before all contrast degraded face samples are omitted the EER can decrease. After there is no contrast degraded face samples left in the database, the EER stops decreasing because CONTRAST cannot predict the face image quality from other types of distortions. This explains why we can observe such phenomenon and similar observations can be found for JNBM and SH as well. When we compare the performance between selected IQMs and ISO-proposed IQMs, we can see that only ISO1 can reduce the EER when omitting low quality face samples. ISO2 and ISO3 give similar results to IQMs which have reversed correlation with the performance of selected face recognition algorithms. The reason is that they are designed for only single type of distortion and they cannot handle the multiple distortions under the condition in this paper. Thus, the findings in this paper show that this research work can be used in the future and is meaningful.

Finally, experimental results from the retrained ILNIQE2 show that the database used for training the IQMs may influence the performance. The core of the ILNIQE2 metric is a prelearned NSS fitted multivariate Gaussian (MVG) model. This model uses NSS features computed from pristine natural image (e.g., LIVE database) patches. This MVG model is therefore deployed as a pristine reference model against which to measure the quality of a given test image. On each patch of a test image, a best-fit MVG model is computed and then compared with the prelearned pristine MVG model for the calculation of the quality score. However, as

we mentioned, face images are a subcategory of natural images. It may not be appropriated to compare the best-fit MVG model from only face images with the prelearned MVG model from the entire portion of natural images. It could be explained by the variations between face images are less due to the similar structure of face images. The retrained MVG model is then more appropriate for the calculation of quality scores for face images. Therefore, the performance of retrained ILNIQE2 is better than the original.

## 6 Contribution

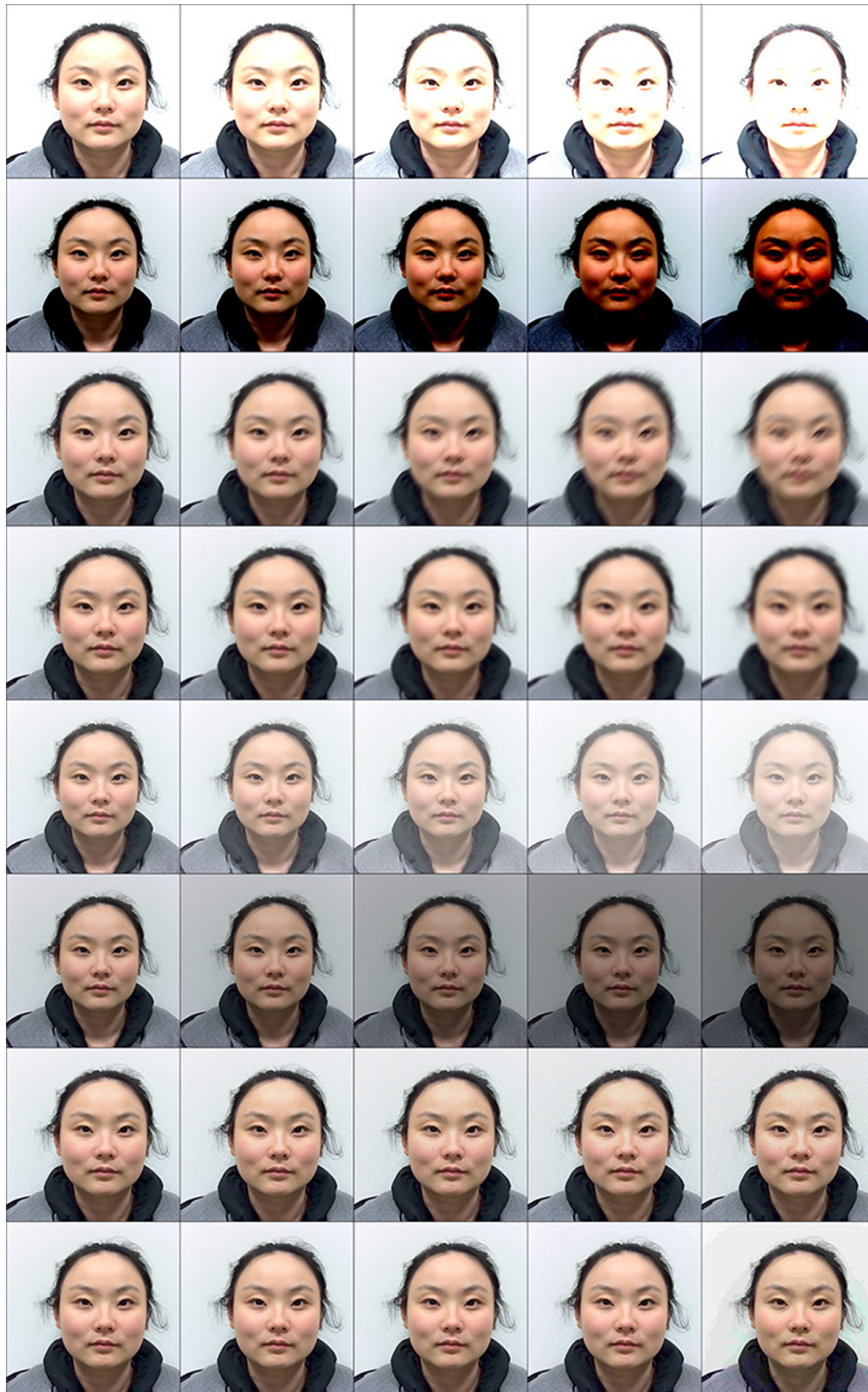
The main contribution of this paper is the evaluation of the performance and the analysis of the experimental results for no-reference IQMs designed for natural images on face samples images. Only image-based quality attributes are taken into account for both IQMs and the database used in the paper. It avoids the impact of modality-based attributes. So the evaluation results and the analysis in this paper can be used to create a common framework, which is used for the assessment of multimodality image-based biometric samples by using no-reference IQMs.

## 7 Conclusion and Future Work

In this paper, we evaluated the performance of selected no-reference IQMs for face biometric images on GC<sup>2</sup> multimodality biometric database using three face recognition algorithms. Three indicators are used to reflect the performance of IQMs according to the face recognition algorithms: histogram of mean comparison score, DET curve, and EER value. We illustrated the results by comparing between indicators with and without omitting certain percentage of low quality face samples. In addition, experiment that retraining an IQM by using only face database is conducted. From the experimental results we can conclude that, before the retraining process, ILNIQE2 has a better performance to assess the quality of face images based on the DET curves and EER values for two cameras: LFC and smartphone. The retrained ILNIQE2 metric has better performance for all three cameras and the performance for LFC and smartphone is further improved. Therefore, it is possible to use existing no-reference IQMs to assess the face sample quality, moreover, the optimization process can further improve the performance of IQMs. In general, selected distortion-specific IQMs are not as good as the selected generalized purposes holistic IQMs due to the limitation of suitable degradation. One way to improve the performance of selected IQMs is to train them on face databases, because the performance of IQMs on face images may affected by the database used for training. The aforementioned findings can be used for the development of robust quality metrics for face image quality, and furthermore, for multiple biometric modalities image quality assessment.

## Appendix A: Illustration of Distorted Face Sample Images in Five Degradation Levels

Figure 12 shows an example of distorted face image in five degradation levels.



**Fig. 12** Degraded face samples in five levels, the first column represents the degradation level 1 (little degraded) and the last column represents the degradation level 5 (highly degraded). The first row represents too high contrast face images; the second row represents too low contrast face images; the third row represents motion blurred face images; the fourth row represents the Gaussian blurred face images; the fifth row represents high luminance face images; the sixth row represents low luminance face images; the seventh row represents face images contain Poisson noise; the last row represents JPEG compressed face images.

### Appendix B: Figures of Histogram of the Comparison Scores and Their Mean Values

Figures 13–15 show the comparison scores and their mean values with omitting low quality samples for LFC, smartphone, and reflex camera, respectively.

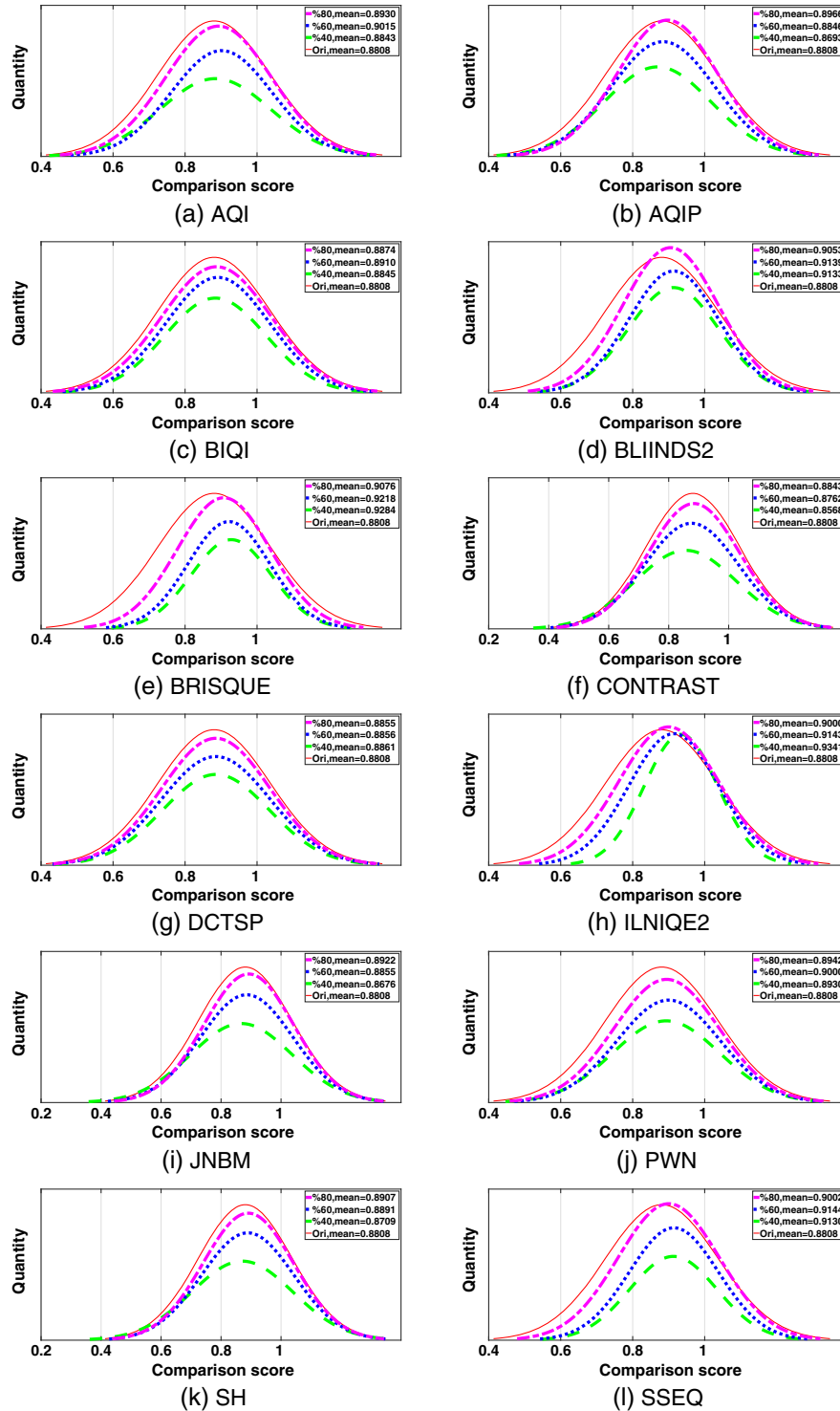


Fig. 13 Comparison scores and their mean values with omitting low quality samples for LFC by using GF + KFA recognition algorithm.



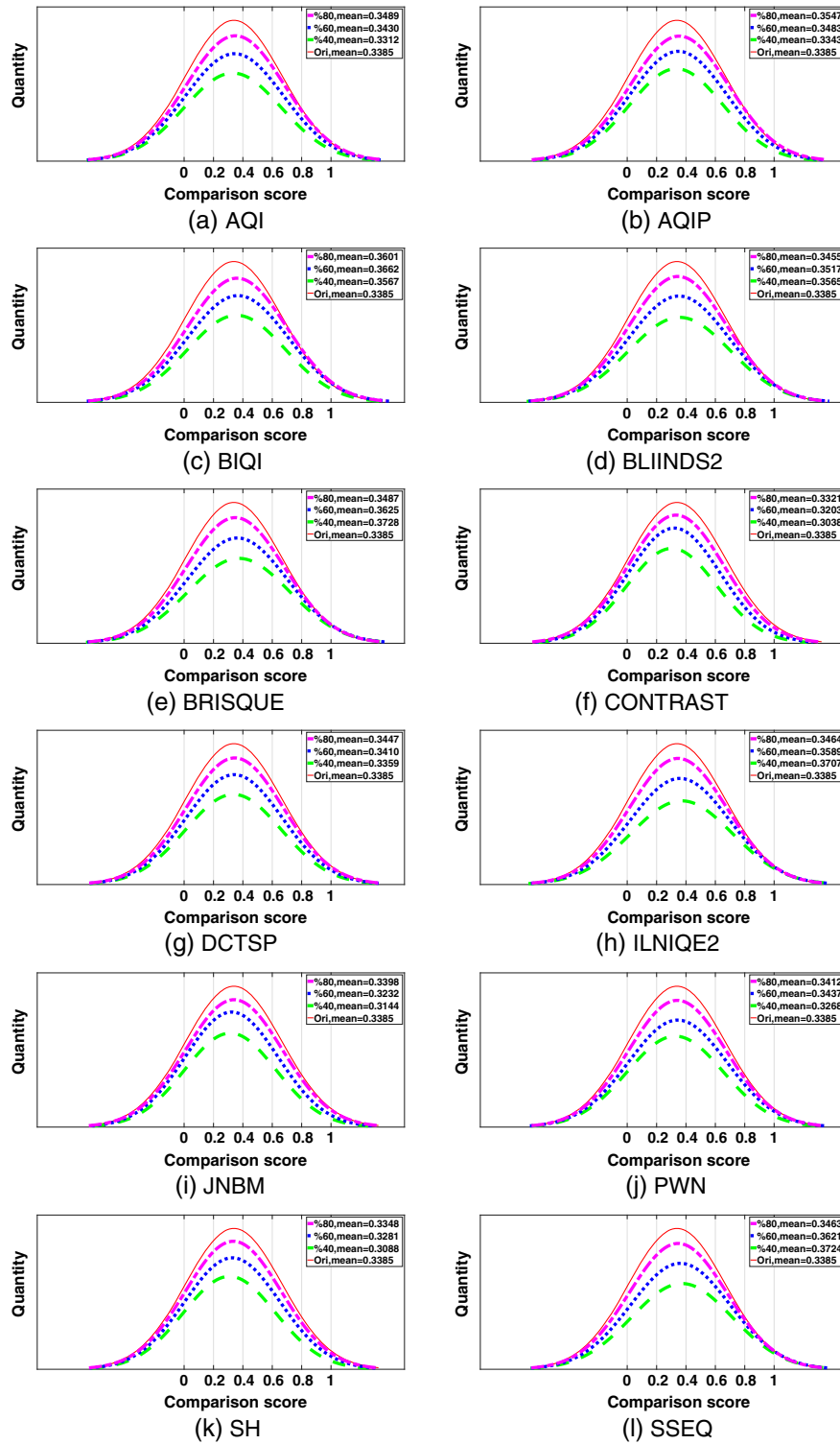


Fig. 14 Comparison scores and their mean values with omitting low quality samples for smartphone by using GF + KFA recognition algorithm.

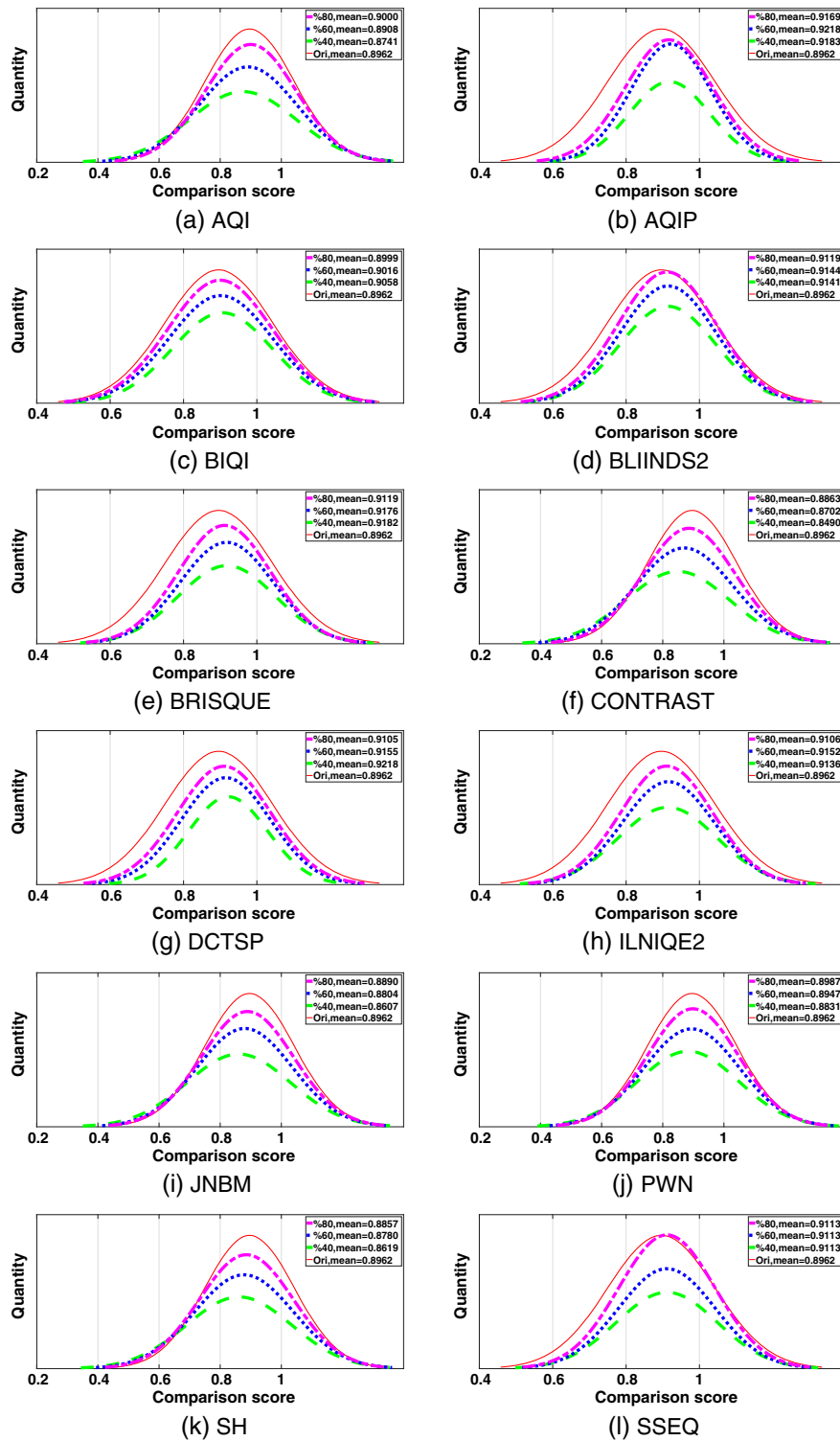


Fig. 15 Comparison scores and their mean values with omitting low quality samples for reflex camera by using GF + KFA recognition algorithm.

### Appendix C: Figures of DET Curve and EER

Figures 16–18 show the DET curves with EER for comparison score with and without omitting low quality samples for LFC, smartphone, and reflex camera, respectively.

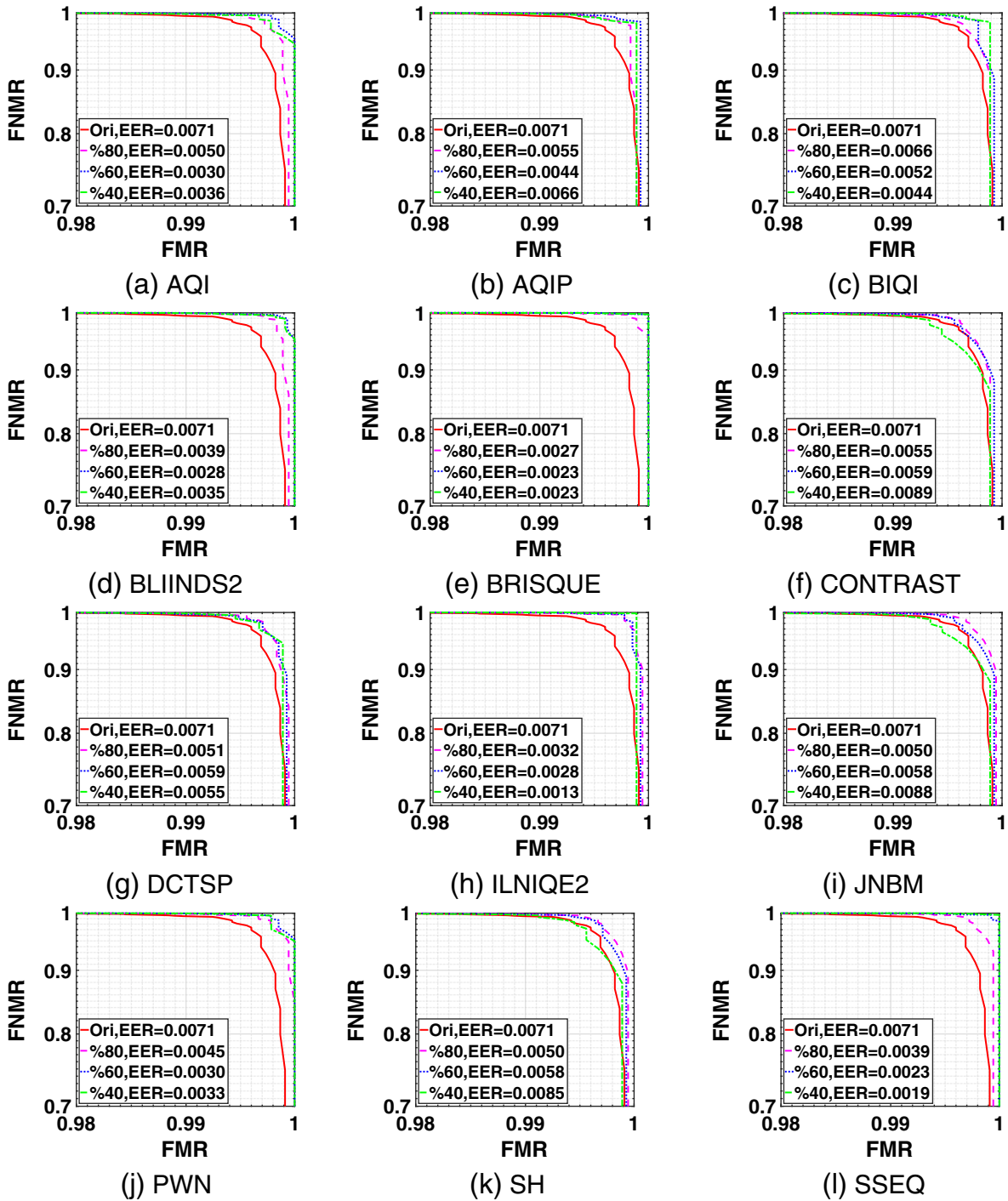


Fig. 16 DET curves with EER for comparison score with and without omitting low quality samples for LFC by using GF+ KFA recognition algorithm.

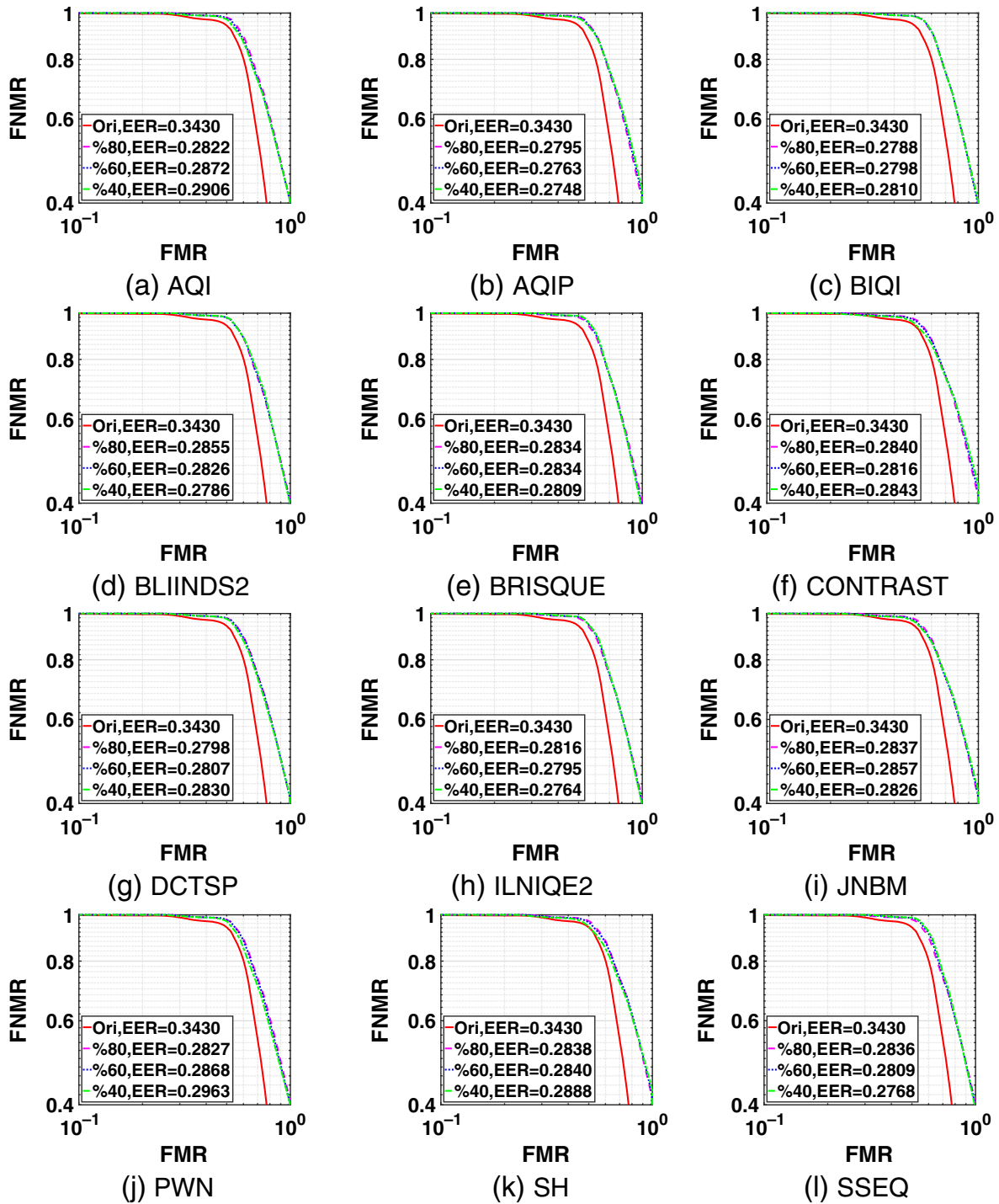


Fig. 17 DET curves with EER for comparison score with and without omitting low quality samples for phone by using GF + KFA recognition algorithm.

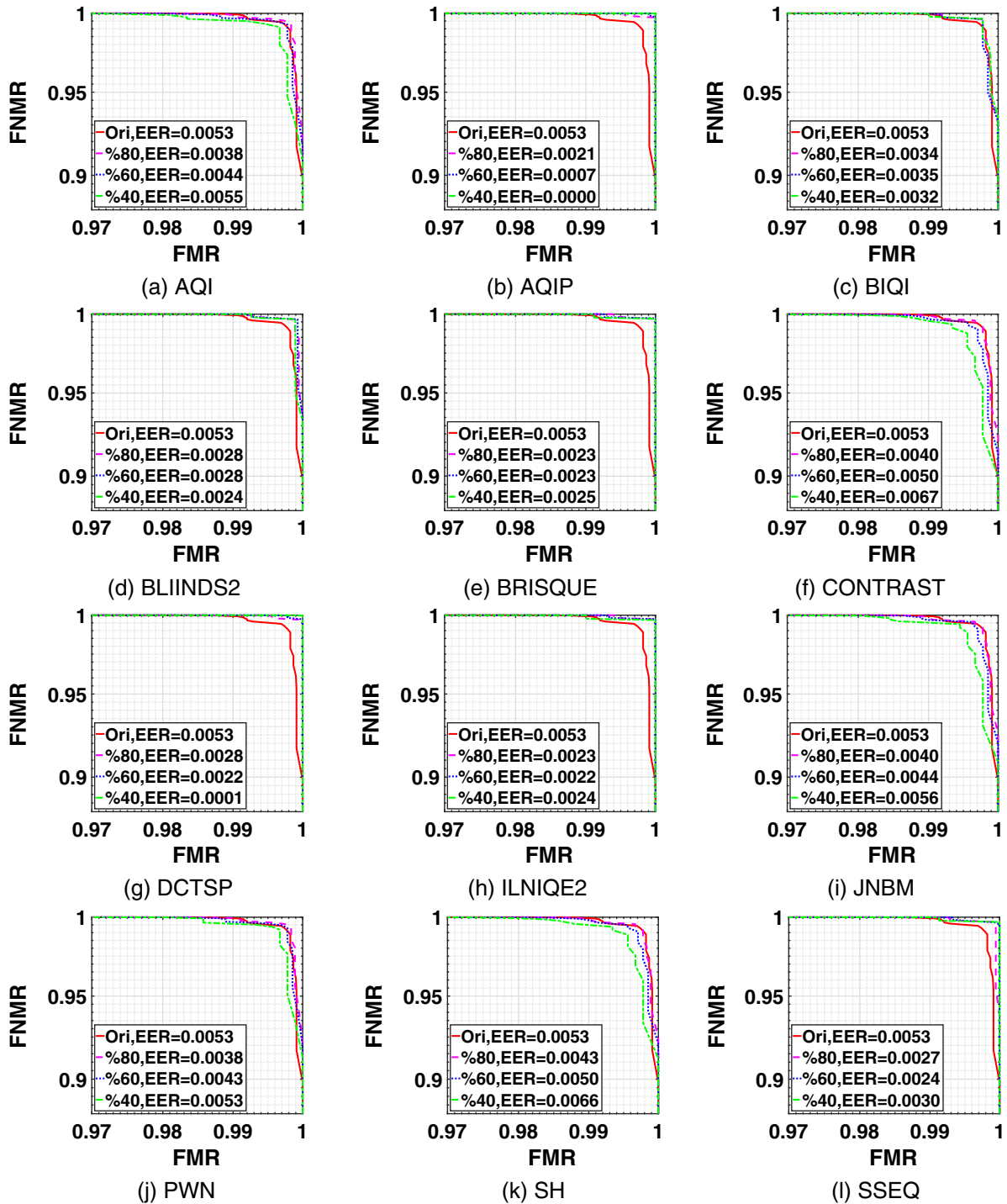
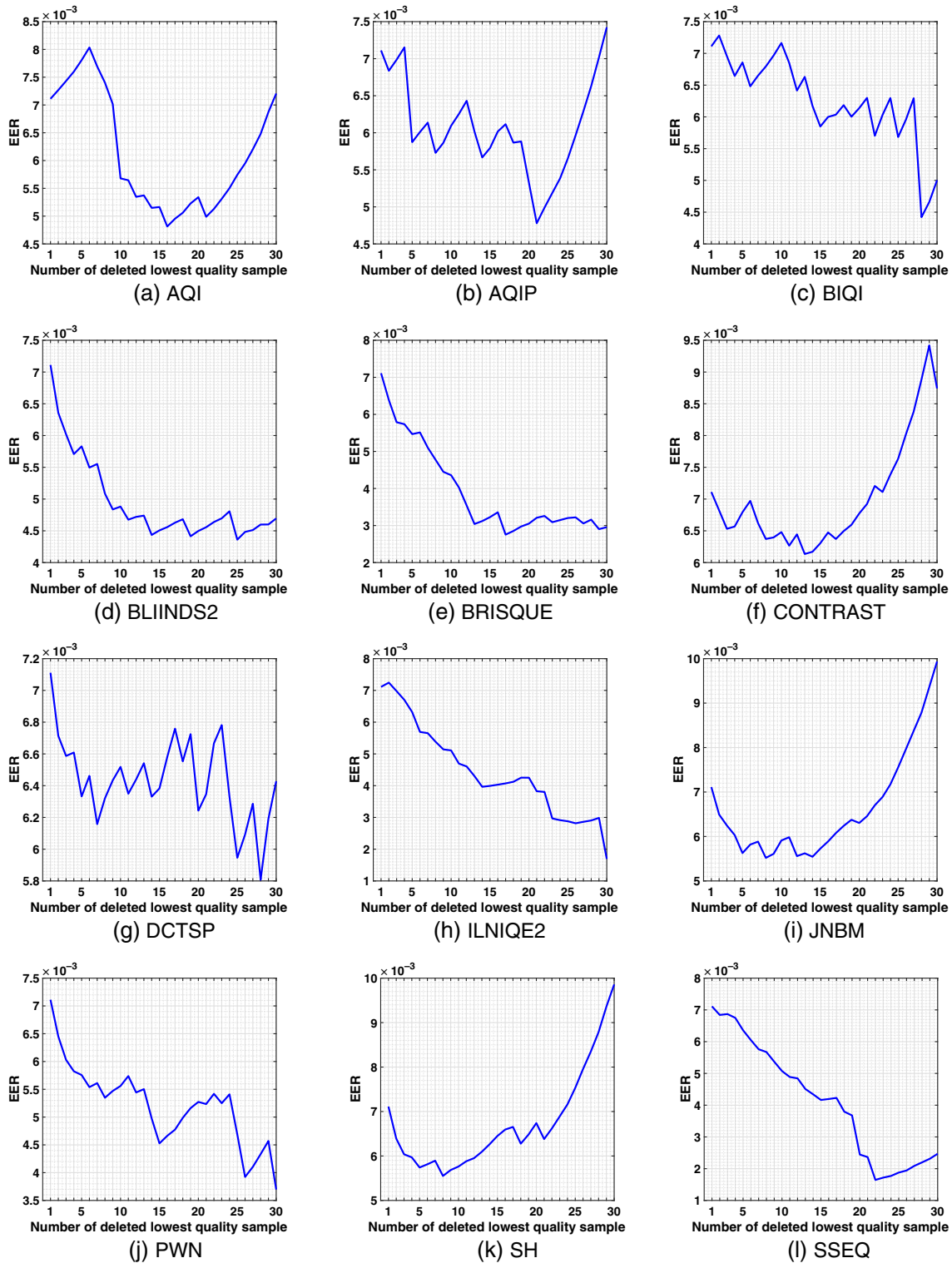


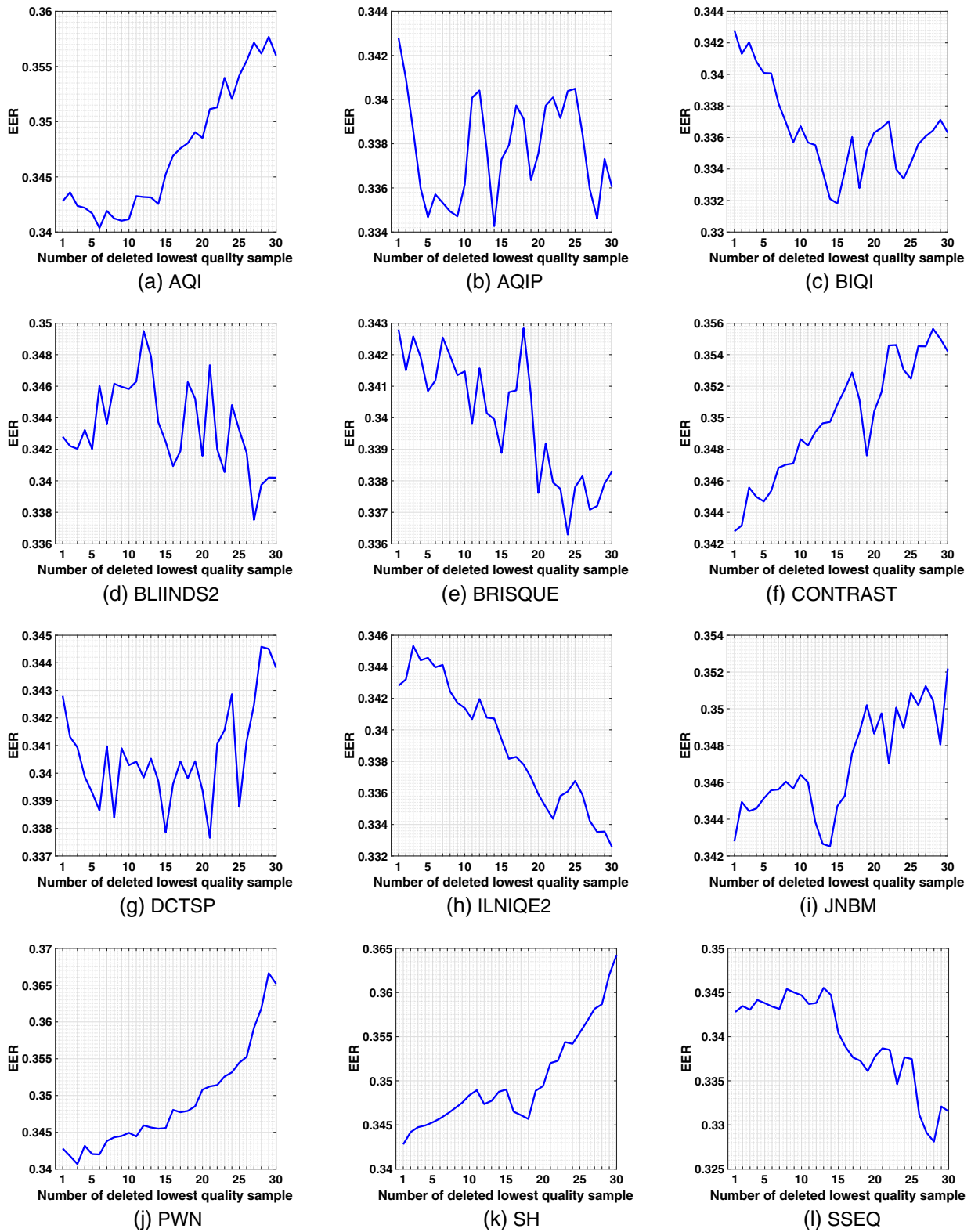
Fig. 18 DET curves with EER for comparison score with and without omitting low quality samples for reflex by using GF + KFA recognition algorithm.

### Appendix D: Figures of EER One by One

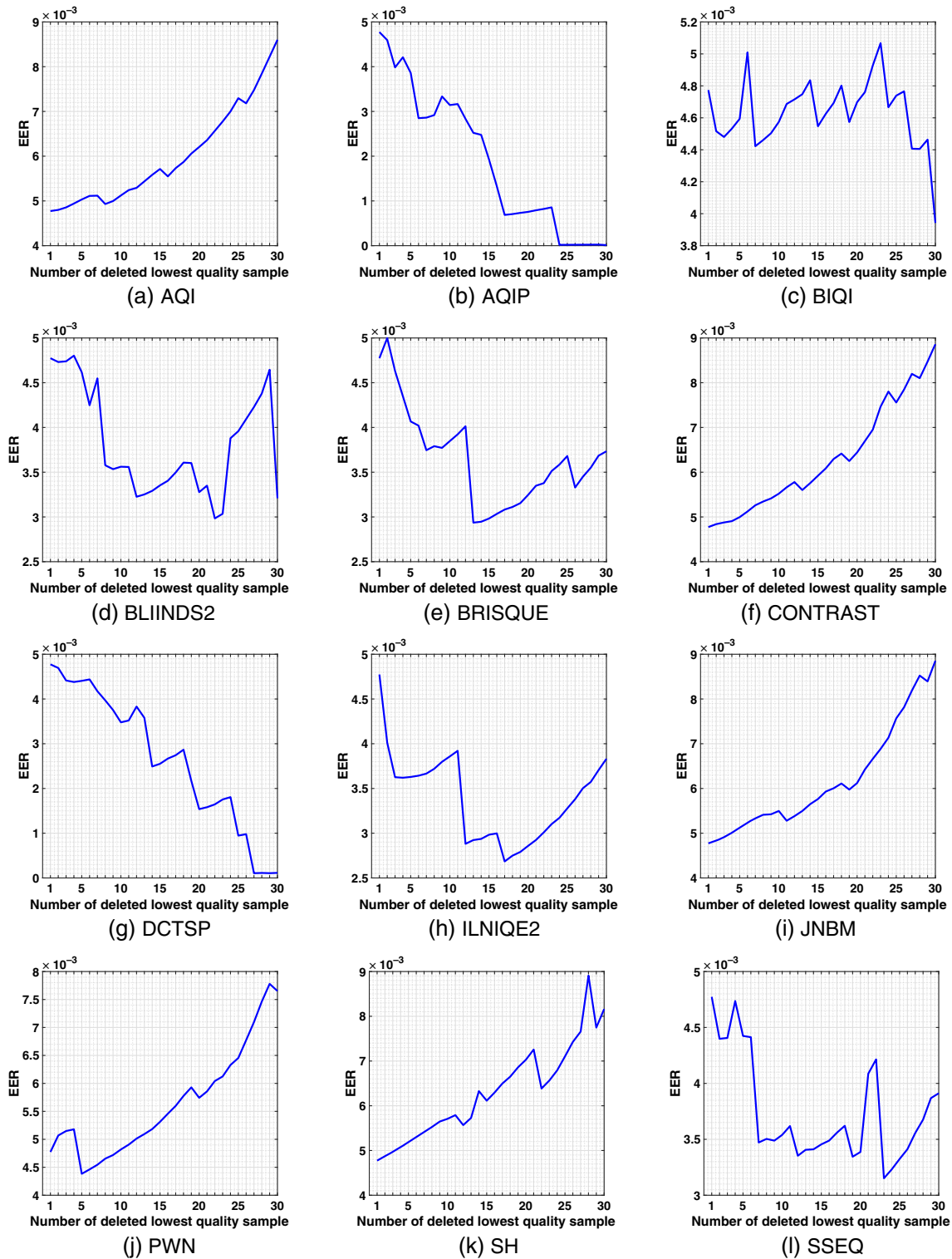
Figures 19–21 show the EER values with omitting low quality face samples one by one until the best quality sample left for LFC, smartphone, and reflex camera, respectively.



**Fig. 19** EER values with omitting low quality face samples one by one until the best quality sample left for LFC by using GF + KFA recognition algorithm.



**Fig. 20** EER values with omitting low quality face samples one by one until the best quality sample left for smartphone by using GF + KFA recognition algorithm.



**Fig. 21** EER values with omitting low quality face samples one by one until the best quality sample left for reflex camera by using GF + KFA recognition algorithm.

**Acknowledgments**

This research was supported by the Conseil Regional Basse-Normandie Grant 14P02048 and the Research Council of Norway through project number 221073: HyPerCept—Color and Quality in Higher Dimensions.

**References**

1. A. K. Jain and S. Z. Li, *Handbook of Face Recognition*, Springer, New York (2011).
2. X. Gao et al., "Standardization of face image sample quality," *Lect. Notes Comput. Sci.* **4642**, 242–251 (2007).



3. International Organization for Standardization (ISO), "Biometric data interchange formats—part 5: face image data," ISO/IEC TR 19794-5, International Electrotechnical Commission (IEC) (2011).
4. International Organization for Standardization (ISO), "Information technology—biometric sample quality—part 5: face image data," ISO/IEC TR 29794-5, International Electrotechnical Commission (IEC) (2010).
5. X. Liu, M. Pedersen, and C. Charrier, "Image-based attributes of multi-modality image quality for contactless biometric samples," in *3rd Int. Conf. on Signal Processing and Integrated Networks (SPIN)*, pp. 106–111, IEEE (2016).
6. M. Pedersen and J. Y. Hardeberg, "Survey of full-reference image quality metrics," (2009).
7. Z. Wang et al., "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process.* **13**(4), 600–612 (2004).
8. X. Liu et al., "Can no-reference image quality metrics assess contactless visible wavelength iris sample quality?" in *IEEE Int. Conf. on Image Processing (ICIP)*, pp. 1–5, IEEE (2017).
9. C. Liu et al., "Noise estimation from a single image," in *IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR)*, Vol. 1, pp. 901–908, IEEE (2006).
10. P. Marziliano et al., "A no-reference perceptual blur metric," in *Int. Conf. on Image Processing Proc.*, Vol. 3, pp. III-57–III-60, IEEE (2002).
11. J. Yang, H. Choi, and T. Kim, "Noise estimation for blocking artifacts reduction in DCT coded images," *IEEE Trans. Circuits Syst. Video Technol.* **10**(7), 1116–1120 (2000).
12. F. Alonso-Fernandez, J. Fierrez, and J. Ortega-Garcia, "Quality measures in biometric systems," *IEEE Secur. Privacy* **10**(6), 52–62 (2012).
13. P. Grother and E. Tabassi, "Performance of biometric quality measures," *IEEE Trans. Pattern Anal. Mach. Intell.* **29**(4), 531–543 (2007).
14. S. Bharadwaj, M. Vatsa, and R. Singh, "Biometric quality: a review of fingerprint, iris, and face," *EURASIP J. Image Video Process.* **2014**(1), 34 (2014).
15. M. El-Abed, C. Charrier, and C. Rosenberger, "Quality assessment of image-based biometric information," *EURASIP J. Image Video Process.* **2015**(1), 3 (2015).
16. A. Abaza et al., "Design and evaluation of photometric image quality measures for effective face recognition," *IET Biom.* **3**(4), 314–324 (2014).
17. A. Dutta, R. Veldhuis, and L. Spreeuwiers, "Predicting face recognition performance using image quality," PhD Thesis, University of Twente (2015).
18. F. Hua et al., "Impact of out-of-focus blur on face recognition performance based on modular transfer function," in *5th IAPR Int. Conf. on Biometrics (ICB)*, pp. 85–90, IEEE (2012).
19. C. Fiche, P. Ladret, and N.-S. Vu, "Blurred face recognition algorithm guided by a no-reference blur metric," *Proc. SPIE* **7538**, 75380U (2010).
20. The National Institute of Standards and Technology (NIST), "The color FERET database (version 2)," <https://www.nist.gov/itl/iad/image-group/color-feret-database> (22 November 2016).
21. "The Yale face database," <http://vision.ucsd.edu/content/yale-face-database> (22 November 2016).
22. Cambridge University Computer Laboratory, "The AT&T face database," <http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html> (22 November 2016).
23. "The face recognition homepage," <http://www.face-rec.org/databases/> (22 November 2016).
24. "Lytro," <https://www.lytro.com/> (22 November 2016).
25. X. Liu, M. Pedersen, and J. Y. Hardeberg, "CID: IQ—a new image quality database," *Lect. Notes Comput. Sci.* **8509**, 193–202 (2014).
26. X. Liu et al., "The influence of fingerprint image degradations on the performance of biometric system and quality assessment," in *Int. Conf. of the Biometrics Special Interest Group (BIOSIG)*, pp. 1–6, IEEE (2016).
27. Y. Fang et al., "No-reference quality assessment of contrast-distorted images based on natural scene statistics," *IEEE Signal Process. Lett.* **22**(7), 838–842 (2015).
28. A. K. Moorthy and A. C. Bovik, "A two-step framework for constructing blind image quality indices," *IEEE Signal Process. Lett.* **17**(5), 513–516 (2010).
29. M. A. Saad, A. C. Bovik, and C. Charrier, "DCT statistics model-based blind image quality assessment," in *18th IEEE Int. Conf. on Image Processing*, pp. 3093–3096, IEEE (2011).
30. A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Trans. Image Process.* **21**(12), 4695–4708 (2012).
31. L. Zhang, L. Zhang, and A. C. Bovik, "A feature-enriched completely blind image quality evaluator," *IEEE Trans. Image Process.* **24**(8), 2579–2591 (2015).
32. R. Ferzli and L. J. Karam, "A no-reference objective image sharpness metric based on the notion of just noticeable blur (JNB)," *IEEE Trans. Image Process.* **18**(4), 717–728 (2009).
33. Y. Han, X. Xu, and Y. Cai, "Novel no-reference image blur metric based on block-based discrete cosine transform statistics," *Opt. Eng.* **49**(5), 050501 (2010).
34. A. Leclaire and L. Moisan, "No-reference image quality assessment and blind deblurring with sharpness metrics exploiting Fourier phase information," *J. Math. Imaging Vision* **52**(1), 145–172 (2015).
35. G. Blanchet and L. Moisan, "An explicit sharpness index related to global phase coherence," in *IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1065–1068, IEEE (2012).
36. T. Zhu and L. Karam, "A no-reference objective image quality metric based on perceptually weighted local noise," *EURASIP J. Image Video Process.* **2014**(1), 1–8 (2014).
37. S. Gabarda and G. Cristóbal, "Blind image quality assessment through anisotropy," *J. Opt. Soc. Am. A* **24**(12), B42–B51 (2007).
38. L. Liu et al., "No-reference image quality assessment based on spatial and spectral entropies," *Signal Process. Image Commun.* **29**(8), 856–863 (2014).
39. Š. Vitomir, "The PHD toolbox: pretty helpful development functions for face recognition," (2012).
40. Š. Vitomir and P. Nikola, "The complete Gabor-fisher classifier for robust face recognition," *EURASIP J. Adv. Signal Process.* **2010**(1), 847680 (2010).
41. C. Liu, "Capitalize on dimensionality increasing techniques for improving face recognition grand challenge performance," *IEEE Trans. Pattern Anal. Mach. Intell.* **28**(5), 725–737 (2006).
42. P. Kovési, "Image features from phase congruency," *Videre J. Comput. Vision Res.* **1**(3), 1–26 (1999).
43. N. Poh, S. Bengio, and A. Ross, "Revisiting Doddington's zoo: a systematic method to assess user-dependent variabilities," Technical Report, IDIAP (2006).
44. J. Valente and S. Soatto, "Perspective distortion modeling, learning and compensation," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition Workshops*, pp. 9–16 (2015).
45. H. R. Sheikh et al., "Live image quality assessment database release 2," (2005).

**Xinwei Liu** received his master's degree in computer science and media technology in 2014 from Gjøvik University College, Norway. He is now approaching the end of his co-tutelle PhD in biometrics image quality assessment from both Norwegian University of Science and Technology, Gjøvik, Norway, and University of Caen, CNRS, GREYC Lab, Caen, France. His current research interests include image quality assessment, perceptual image quality, biometrics, and biometric image quality.

**Marius Pedersen** received his BSc degree in computer engineering, and MIT in Media Technology from Gjøvik University College, Norway, in 2006 and 2007, respectively. He completed his PhD program in color imaging from the University of Oslo, Norway, in 2011. Currently, he is employed as a professor at NTNU Gjøvik, Norway. He is also the director of the Norwegian Color and Visual Computing Laboratory. His work is centered on subjective and objective image quality.

**Christophe Charrier** received his PhD in computer science from the University Jean Monnet in 1998. He was a research assistant with the Laboratory of Radio Communications and Signal Processing, Laval University, Canada. He joined the GREYC Lab and the e-payment and biometrics research unit in 2008. His current research interests include image and video coding, processing, quality assessment, computational vision, and biometrics. Since 2016, he is the head of e-payment and biometrics research unit.

**Patrick Bours** received his MSc and PhD degrees in discrete mathematics from Eindhoven University of Technology in Netherlands, in 1990 and 1994, respectively. He worked 10 years at the Netherlands National Communication Security Agency in the area of crypto. He started a postdoctoral at the Norwegian Information Security Laboratory (NISlab) at Gjøvik University College in 2005. His main research interest is in behavioral biometrics. From 2012, he has a full professor position at NISlab.