A Feature-Enriched Completely Blind Image Quality Evaluator

Lin Zhang, *Member, IEEE,* Lei Zhang, *Senior Member, IEEE,* and Alan C. Bovik, *Fellow, IEEE*

***Abstract*—Existing blind image quality assessment (BIQA) methods are mostly *opinion-aware*. They learn regression models from training images with associated human subjective scores to predict the perceptual quality of test images. Such opinion-aware methods, however, require a large amount of training samples with associated human subjective scores and of a variety of distortion types. The BIQA models learned by opinion-aware methods often have weak generalization capability, hereby limiting their usability in practice. By comparison, *opinion-unaware* methods do not need human subjective scores for training, and thus have greater potential for good generalization capability. Unfortunately, thus far no opinion-unaware BIQA method has shown consistently better quality prediction accuracy than opinion-aware methods. Here we aim to develop an opinion-unaware BIQA method that can compete with, and perhaps outperform existing opinion-aware methods. By integrating natural image statistics features derived from multiple cues, we learn a multivariate Gaussian model of image patches from a collection of pristine natural images. Using the learned multivariate Gaussian model, a Bhattacharyya-like distance is used to measure the quality of each image patch, then an overall quality score is obtained by average pooling. The proposed BIQA method does not need any distorted sample images nor subjective quality scores for training, yet extensive experiments demonstrate its superior quality-prediction performance to state-of-the-art opinion-aware BIQA methods. The Matlab source code of our algorithm is publicly available at www.comp.polyu.edu.hk/∼cslzhang/IQA/ILNIQE/ILNIQE.htm.**

***Index Terms*—Blind image quality assessment, natural image statistics, multivariate Gaussian.**

I. Introduction

IT is a highly desirable goal to be able to faithfully evaluate the quality of output images in many applications, such as image acquisition, transmission, compression, restoration, enhancement, etc. Quantitatively evaluating an image's perceptual quality has been among the most challenging problems of modern image processing and computational vision research. Perceptual image quality assessment (IQA) methods fall into two categories: subjective assessment by humans, and objective assessment by algorithms designed to mimic the subjective judgments. Though subjective assessment is the ultimate criterion of an image's quality, it is time-consuming, cumbersome, expensive, and cannot be implemented in systems where real-time evaluation of image or video quality is needed. Hence, there has been an increasing interest in developing objective IQA methods that can automatically predict image quality in a manner that is consistent with human subjective perception.

Early no-reference IQA (NR-IQA) models commonly operated under the assumption that the image quality is affected by one or several particular kinds of distortions, such as blockiness [1], [2], ringing [3], blur [4], [5], or compression [6]–[9]. Such early NR-IQA approaches therefore extract distortion-specific features for quality prediction, based on a model of the presumed distortion type(s). Hence, the application scope of these methods is rather limited.

Recent studies on NR-IQA have focused on the so-called blind image quality assessment (BIQA) problem, where prior knowledge of the distortion types is unavailable. A majority of existing BIQA methods are "opinion aware", which means that they are trained on a dataset consisting of distorted images and associated subjective scores [10]. Representative methods belonging to this category include [11]–[18] and they share a similar architecture. In the training stage, feature vectors are extracted from the distorted images, then a regression model is learned to map the feature vectors to the associated human subjective scores. In the test stage, a feature vector is extracted from the test image and then fed into the learned regression model to predict its quality score. In [11], Moorthy and Bovik proposed a two-step framework for BIQA, called BIQI. In BIQI, given a distorted image, scene statistics are at first extracted and used to explicitly classify the distorted image into one of n distortions; then, the same set of statistics are used to evaluate the distortion-specific quality. Following the same paradigm, Moorthy and Bovik later extended BIQI to DIIVINE using a richer set of natural scene features [12]. Both BIQI and DIIVINE assume that the distortion types in the test images are represented in the training dataset, which is, however, not the case in many practical applications. By assuming that the statistics of DCT features can vary in a predictable way as the image quality changes, Saad *et al*. [13] proposed a BIQA model, called BLIINDS, by training a probabilistic model based on contrast and structure features extracted in the DCT domain. Saad *et al.* later extended BLIINDS to BLIINDS-II [14] using more sophisticated NSS-based DCT features. In [15], Mittal *et al.* used scene statistics of locally normalized luminance coefficients to quantify possible losses of naturalness in the image due to the presence of distortions, and the resulting BIQA model is referred to BRISQUE. The

This work was supported by the Natural Science Foundation of China under grant no. 61201394, the Shanghai Pujiang Program under grant no. 13PJ1408700, HK RGC GRF under grant no. PolyU 5315/12E, and the U.S. National Science Foundation under grant no. IIS-1116656.

Lin Zhang is with the School of Software Engineering, Tongji University, Shanghai 201804, China, and also with the Shenzhen Institute of Future Media Technology, Shenzhen 518055, China (email: cslinzhang@tongji.edu.cn).

Lei Zhang is with the Department of Computing, The Hong Kong Polytechnic University, Hong Kong (email: cslzhang@comp.polyu.edu.hk).

Alan C. Bovik is with the Department of Electrical and Computer Engineering, University of Texas at Austin, Austin, TX 78712 USA (email: bovik@ece.utexas.edu).

model proposed in [16] extracts three sets of features based on the statistics of natural images, distortion textures, and blur/noise; three regression models are trained for each feature set and finally a weighted combination of them is used to estimate the image quality.

In [17], Ye *et al*. proposed an unsupervised feature learning framework for BIQA, called CORNIA, which consists of the following major steps: local feature extraction, codebook construction, soft-assignment coding, max-pooling, and linear regression. In [18], Li *et al*. extracted four kinds of features from images being quality-tested: the mean value of a phase congruency [19] map computed on an image, the entropy of the phase congruency map, the entropy of the image, and the gradient of the image. A generalized regression neural network (GRNN) [20] was deployed to train the model. In [21], Zhang and Chandler extracted image quality-related statistical features in both the spatial and frequency domains. In the spatial domain, locally normalized pixels and adjacent pixel pairs were statistically modeled using log-derivative statistics; and in the frequency domain, log-Gabor filters [22] were used to extract the fine scales of the image. Based on the observation that image local contrast features convey important structural information that is related to image perceptual quality, in [23], Xue *et al*. proposed a BIQA model utilizing the joint statistics of the local image gradient magnitudes and the Laplacian of Gaussian image responses.

The opinion-aware BIQA methods discussed above require a large number of distorted images with human subjective scores to learn the regression model, which causes them to have rather weak generalization capability. In practice, image distortion types are numerous and an image may contain multiple interacting distortions. It is difficult to collect enough training samples for all such manifold types and combinations of distortions. If a BIQA model trained on a certain set of distortion types is applied to a test image containing a different distortion type, the predicted quality score will be unpredictable and likely inaccurate. Second, existing trained BIQA models have been trained on and thus are dependant to some degree on one of the available public databases. When applying a model learned on one database to another database, or to real-world distorted images, the quality prediction performance can be very poor (refer to Section IV-D for details).

Considering the shortcomings of opinion-aware BIQA methods, it is of great interest to develop "opinion-unaware" IQA models, which do not need training samples of distortions nor of human subjective scores [10]. However, while the goal of opinion-unaware BIQA is attractive, the design methodology is more challenging due to the limited available information. A few salient works have been reported along this direction. In [24], Mittal *et al*. proposed an algorithm that conducts probabilistic latent semantic analysis on the statistical features of a large collection of pristine and distorted image patches. The uncovered latent quality factors are then applied to the image patches of the test image to infer a quality score. The Natural Image Quality Evaluator (NIQE) model proposed by Mittal *et al*. [10] extracts a set of local features from an image, then fits the feature vectors to a multivariate Gaussian (MVG) model. The quality of a test image is then predicted by the distance between its MVG model and the MVG model learned from a corpus of pristine naturalistic images. However, since NIQE uses a single *global* MVG model to describe an image, useful local image information which could be used to better predict the image quality is lost. In [25], Xue *et al*. simulated a virtual dataset wherein the quality scores of distorted images are first estimated using the full reference IQA algorithm FSIM [26]. A BIQA model is then learned from the dataset by a process of patch based clustering. However, this "quality aware clustering" (QAC) method is only able to deal with four commonly encountered types of distortions; hence, unlike NIQE, QAC is not a "totally blind" BIQA method.

One distinct property of opinion-unaware BIQA methods is that they have the potential to deliver higher generalization capability than their opinion-aware counterparts due to the fact that they do not depend on training samples of distorted images and associated subjective quality scores on them. However, thus far, no opinion-unaware method has shown better quality prediction power than currently available opinion-aware methods. Thus, it is of great interest and significance to investigate whether it is possible to develop an opinion-unaware model that outperforms state-of-the-art opinion-aware BIQA models.

We make an attempt to achieve the above goal in this paper. It is commonly accepted that the statistics of a distorted image will be measurably different from those of pristine images. We use a variety of existing and new natural scene statistics (NSS) features computed from a collection of pristine natural image patches, and like NIQE, fit the extracted NSS features to an MVG model. 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 online, then compared with the learned pristine MVG model. The overall quality score of the test image is then obtained by pooling the patch scores by averaging them. We conducted an extensive series of experiments on large scale public benchmark databases, and found that the proposed opinion-unaware method exhibits superior quality prediction performance as compared to state-of-the-art opinion-aware NR IQA models, especially on the important cross-database tests that establish generalization capability.

Our work is inspired by NIQE [10]; however, it performs much better than NIQE for the following reasons. First, going beyond the two types of NSS features used in [10], we introduce three additional types of quality-aware features. Second, instead of using a single global MVG model to describe the test image, we fit the feature vector of each patch of the test image to an MVG model, and compute a local quality score on it. We believe that integrating multiple carefully selected quality-aware features that are locally expressed by a local MVG model yields a BIQA model that more comprehensively captures local distortion artifacts. We refer to this significantly improved "completely blind" image quality evaluator by the monicker Integrated Local NIQE, or IL-NIQE.

The most important message of this paper is: we demonstrate that "completely blind" opinion-unaware IQA models can achieve more robust quality prediction

performance than opinion-aware models. Such a model and algorithm can be used in innumerable practical applications. We hope that these results will encourage both IQA researchers and imaging practitioners to more deeply consider the potential of opinion-unaware "completely blind" BIQA models. To make our results fully reproducible, the Matlab source code of IL-NIQE and the associated evaluation results have been made publicly available at www.comp.polyu.edu.hk/~cslzhang/IQA/ILNIQE/ILNIQE.htm.

The rest of this paper is organized as follows. Section II introduces the quality-aware features used in IL-NIQE. Section III presents the detailed design of the new BIQA index IL-NIQE. Section IV presents the experimental results, and Section V concludes the paper.

II. Quality-aware NSS Features

It has been shown that natural scene statistics (NSS) are excellent indicators of the degree of quality degradation of distorted images [10]–[16]. Consequently, NSS models have been widely used in the design of BIQA algorithms. For example, parameters of the generalized Gaussian distribution (GGD) which effectively model natural image wavelet coefficients and DCT coefficients have been used as features for quality prediction [11]–[14]. In [16], a complex pyramid wavelet transform was used to extract similar NSS features. All of these NSS model based BIQA methods are opinion-aware methods, and all learn a regression model to map the extracted NSS feature vectors to subjective quality scores.

Previous studies have shown that image quality distortions are well characterized by features of local structure [27], contrast [23], [26], [27], multi-scale and multi-orientation decomposition [21], [28], and color [26]. Using these considerations, we designed a set of appropriate and effective NSS features for accomplishing opinion-unaware BIQA. To characterize structural distortion, we adopt two types of NSS features (originally proposed in [10]) derived from the distribution of locally mean subtracted and contrast normalized (MSCN) coefficients and from the distribution of products of pairs of adjacent MSCN coefficients. To more effectively characterize structural distortions and to also capture contrast distortion, we deploy quality-aware gradient features (see Sect. II-C). In order to extract quality-related multi-scale and multi-orientation image properties, we use log-Gabor filters and extract statistical features from the filter responses (see Sect. II-D). Color distortions are described using statistical features derived from the image intensity distribution in a logarithmic-scale opponent color space (see Sect. II-E). Overall, five types of features are employed. Although all of these features are well known in the NSS literature, we collectively adapt them for the task of completely-blind BIQA for the first time. Our experiments demonstrate that the new features can significantly improve image quality prediction performance (see Sect. IV-E for details).

A. Statistics of Normalized Luminance

Ruderman [29] pointed out that the locally normalized luminances of a natural gray-scale photographic image I conform to a Gaussian distribution. This normalization process can be described as:

$$\overline{I}(i,j) = \frac{I(i,j) - \mu(i,j)}{\sigma(i,j) + 1} \tag{1}$$

where i and j are spatial coordinates, and

$$\mu(i,j) = \sum_{k=-K}^{K} \sum_{l=-L}^{L} \omega_{k,l} I(i+k, j+l) \tag{2}$$

$$\sigma(i,j) = \sqrt{\sum_{k=-K}^{K} \sum_{l=-L}^{L} \omega_{k,l} [I(i+k, j+l) - \mu(i,j)]^2} \tag{3}$$

are the local image mean and contrast, where $\omega = \{\omega_{k,l} | k = -K, ..., K, l = -L, ..., L\}$ defines a unit-volume Gaussian window. The so-called MSCN coefficients $\overline{I}(i,j)$ have been observed to follow a unit normal distribution on natural images that have not suffered noticeable quality distortions [29]. This Gaussian model, however, is violated when images are subjected to quality degradations caused by common distortions. Measurements of the deviation of $\{\overline{I}(i,j)\}$ from the Gaussian model are indicative of distortion severity.

As suggested in [10], [15], we use a zero-mean generalized Gaussian distribution (GGD) to more broadly model the distribution of $\overline{I}(i,j)$ in the presence of distortion. The density function associated with the GGD is given by:

$$g(x; \alpha, \beta) = \frac{\alpha}{2\beta\Gamma(1/\alpha)} \exp\left(-\left(\frac{|x|}{\beta}\right)^{\alpha}\right) \tag{4}$$

where $\Gamma(\cdot)$ is the gamma function

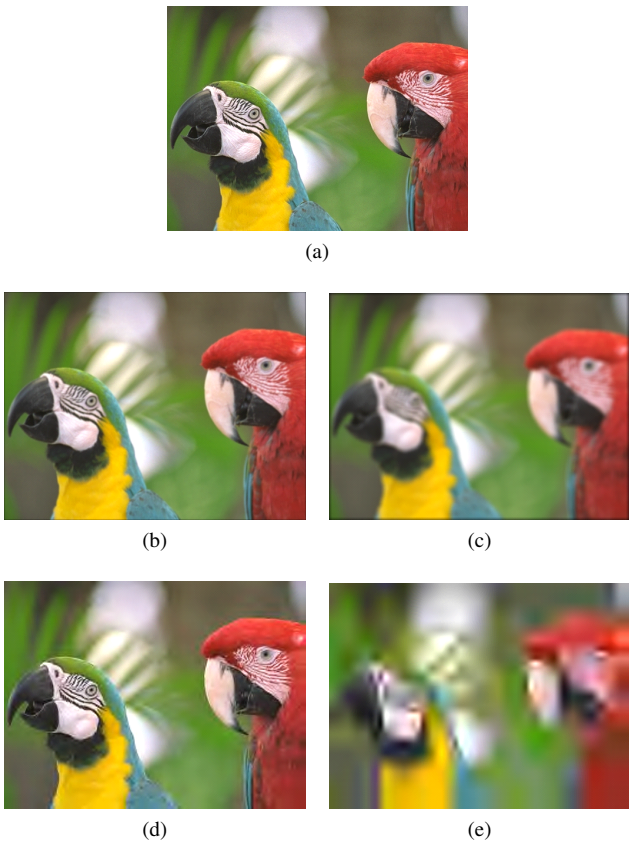
$$\Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt, x > 0. \tag{5}$$

The parameters α and β are effective "quality-aware" features that can be reliably estimated using the moment-matching based approach in [30].

B. Statistics of MSCN Products

As pointed out in [10], [15], image quality information is also captured by the distribution of the products of pairs of adjacent MSCN coefficients, in particular $\overline{I}(i,j)\overline{I}(i,j+1), \overline{I}(i,j)\overline{I}(i+1,j), \overline{I}(i,j)\overline{I}(i+1,j+1)$, and $\overline{I}(i,j)\overline{I}(i+1,j-1)$. On both pristine and distorted images, these products are well modeled as following a zero mode asymmetric GGD (AGGD) [31]:

$$g_{\alpha}(x; \gamma, \beta_l, \beta_r) = \begin{cases} \dfrac{\gamma}{(\beta_l + \beta_r)\Gamma\left(\frac{1}{\gamma}\right)} \exp\left(-\left(\dfrac{-x}{\beta_l}\right)^{\gamma}\right), \forall x \leq 0 \\ \dfrac{\gamma}{(\beta_l + \beta_r)\Gamma\left(\frac{1}{\gamma}\right)} \exp\left(-\left(\dfrac{x}{\beta_r}\right)^{\gamma}\right), \forall x > 0 \end{cases} \tag{6}$$



(a) (b) (c) (d) (e)

Fig. 1: (a) A reference image. Distorted versions of (a): (b) minor Gaussian blur, (c) severe Gaussian blur, (d) minor JPEG2K compression, and (e) severe JPEG2K compression. The subjective MOS scores of the four distorted images in (b)∼(e) are 4.6765, 2.7714, 4.5714 and 0.8235, respectively.

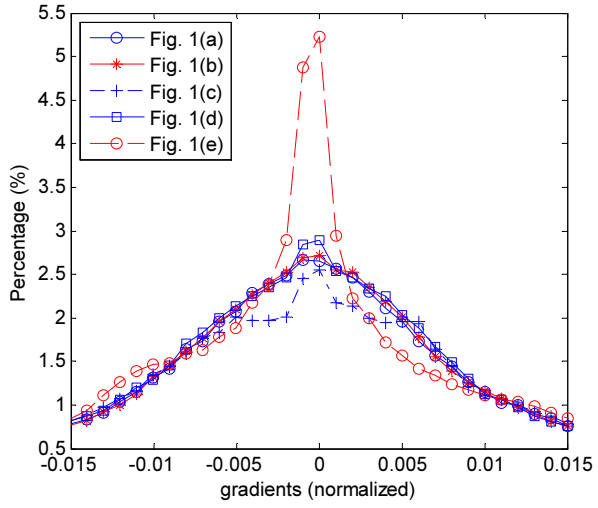
The mean of the AGGD is

$$\eta = (\beta_r - \beta_l)\frac{\Gamma\left(\frac{2}{\gamma}\right)}{\Gamma\left(\frac{1}{\gamma}\right)}. \tag{7}$$

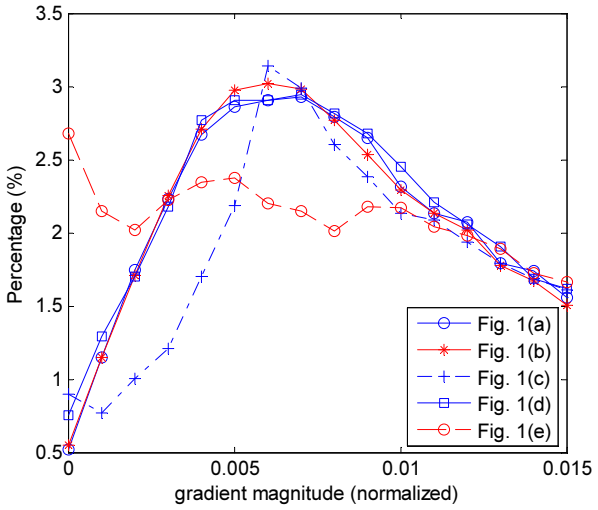
The parameters $(\gamma, \beta_l, \beta_r, \eta)$ are also powerful "quality-aware" features. By extracting these features along four orientations, 16 additional parameters are obtained.

C. Gradient Statistics

The image gradient is a rich descriptor of local image structure, and hence of the local quality of an image. We have found that by introducing distortions to an image, the distributions of its gradient components (partial derivatives) and gradient magnitudes are changed. We use an example to demonstrate this fact. Fig. 1 shows five images selected from the TID2013 dataset [32]. Fig. 1(a) is a reference image while the other four are distorted versions of it: 1(b) with minor Gaussian blur, 1(c) with severe Gaussian blur, 1(d) with minor JPEG2K compression, and 1(e) with severe JPEG2K compression. The subjective scores (MOS, on a scale of 0 to 5) that were recorded on the images in Figs. 1(b), 1(c), 1(d), and 1(e) are 4.6765, 2.7714, 4.5714, and 0.8235, respectively. A higher subjective score indicates better perceptual quality. In Fig. 2(a), we plot the histograms of the gradient components of the five images in Fig. 1, while in Fig. 2(b) we plot the histograms of their gradient magnitudes. Fig. 2 reveals a number of interesting findings. First, when distortions are introduced, the empirical distributions of the image's gradient components and gradient magnitudes are affected. Secondly, more severe distortions cause greater changes in the distributions than less severe ones. The distortions applied to the images in Figs. 1(b) and 1(d) are less severe than those present in Figs. 1(c) and 1(e), and as expected, as shown in Fig. 2(a), the histograms of the images in Figs. 1(b) and 1(d) are similar to that of the reference image in Fig. 1(a), while the histograms of the images shown in Figs. 1(c) and 1(e) significantly



(a)



(b)

Fig. 2: (a) Histograms of the Gaussian-smoothed gradient components (including both I_h and I_v) computed from the five images shown in Fig. 1. (b) Histograms of the gradient magnitudes computed from the five Gaussian-smoothed images shown in Fig. 1.

deviate from that of the reference image. Similar observations can be made regarding Fig. 2(b). We have observed this statistical phenomena to broadly hold on natural photographic images, as demonstrated by our later results. Based on these observations (see also [23]), we use the empirically measured distribution parameters of image gradient components and gradient magnitudes as quality-aware NSS features for the opinion-unaware BIQA task.

We compute the (smoothed) gradient component images, denoted by I_h and I_v, by convolving I with two Gaussian derivative filters along the horizontal and vertical directions, respectively. It has been found that natural (smoothed) image gradient components are well modeled as following a GGD [33]. Thus, we use the parameters α and β computed by fitting the histograms of the gradient components I_h and I_v to the GGD model (Eq. (4)) as quality-aware NSS features.

The gradient magnitude image is computed as $\sqrt{I_h^2 + I_v^2}$. The gradient magnitudes of natural images can be well modeled as following a Weibull distribution [34]:

$$p(x; a, b) = \begin{cases} \frac{a}{b^a} x^{a-1} \exp\left(-\left(\frac{x}{b}\right)^a\right), x \geq 0 \\ 0, x < 0 \end{cases} \quad (8)$$

Larger values of the parameter a roughly correspond to more texture in the gradient magnitude map, while larger values of b imply greater local contrast [34], [35]. Recent studies in neuroscience suggest that the responses of visual neurons strongly correlate with Weibull statistics when processing images [35]. Quality degradations will alter the gradient magnitudes of an image, hence we use the parameters a and b of empirical fits of the Weibull distribution to image gradient magnitude histograms as highly relevant quality-aware NSS features in our BIQA task.

To exploit the expression of distortions in image color space, we also transform RGB images into a perceptually relevant opponent color space [36] prior to computing the NSS gradient features:

$$\begin{bmatrix} O_1 \\ O_2 \\ O_3 \end{bmatrix} = \begin{bmatrix} 0.06 & 0.63 & 0.27 \\ 0.30 & 0.04 & -0.35 \\ 0.34 & -0.6 & 0.17 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (9)$$

The weights in the above conversion are perceptually optimized on human visual data [37]. The NSS features just described are also computed on each channel O_1, O_2, and O_3, respectively, and used as quality-aware features.

D. Statistics of Log-Gabor Filter Responses

Since neurons in visual cortex respond selectively to stimulus orientation and frequency, the statistics of multi-scale, multi-orientation filter responses to an image are also useful for generating quality-aware BIQA features. Here we deploy perceptually-relevant log-Gabor filters [22] to accomplish multi-scale, multi-orientation filtering.

In the Fourier domain, a 2D log-Gabor filter can be expressed as:

$$G_2(\omega, \theta) = e^{-\frac{\left(\log\left(\frac{\omega}{\omega_0}\right)\right)^2}{2\sigma_r^2}} \cdot e^{-\frac{(\theta-\theta_j)^2}{2\sigma_\theta^2}} \quad (10)$$

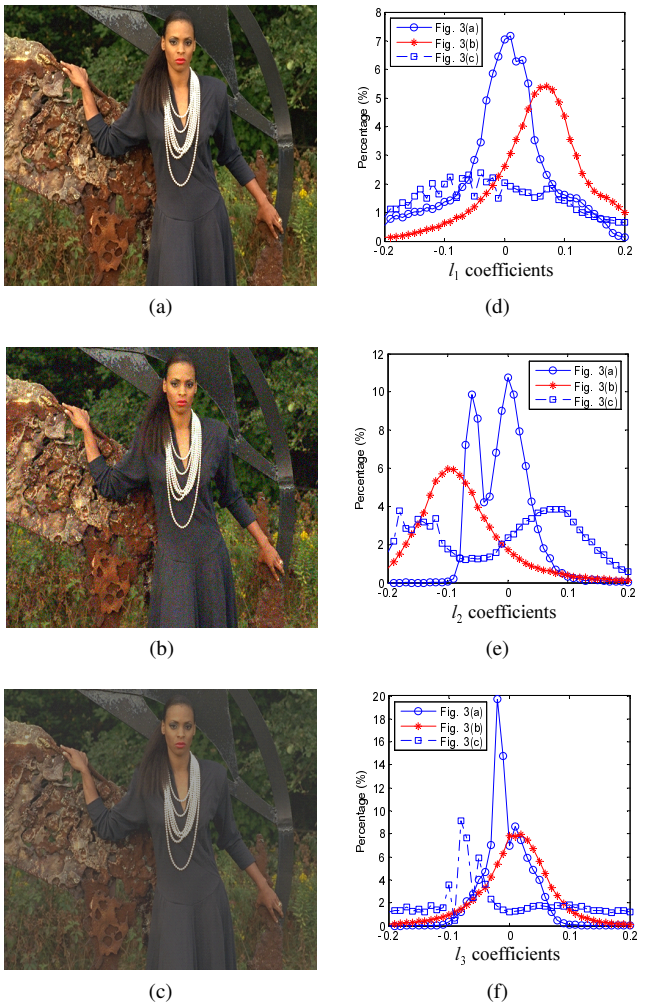


Fig. 3: (a) A reference image. Distorted versions of (a) by: (b) additive noise in the color components and (c) contrast compression. (d) Histograms of the l_1 coefficients computed from the three images shown in (a), (b), and (c). (e) and (f) are histograms of the l_2 and l_3 coefficients of the same images.

where $\theta_j = j\pi/J$, $j = \{0, 1, ..., J-1\}$ is the orientation angle, J is the number of orientations, ω_0 is the center frequency, σ_r controls the filter's radial bandwidth, and σ_θ determines the angular bandwidth of the filter. Applying log-Gabor filters having N different center frequencies and J different orientations to filter an image $f(\mathbf{x})$ yields a set of $2NJ$ responses $\{(e_{n,j}(\mathbf{x}), o_{n,j}(\mathbf{x})) : |n = 0, ..., N-1, j = 0, ..., J-1\}$, where $e_{n,j}(\mathbf{x})$ and $o_{n,j}(\mathbf{x})$ are the responses of the real and imaginary parts of a log-Gabor filter, respectively.

Given the $2NJ$ response maps $\{e_{n,j}(\mathbf{x})\}$ and $\{o_{n,j}(\mathbf{x})\}$, we extract another set of NSS features from them using the following scheme.

a) Use the GGD (Eq. (4)) to model the distributions of $\{e_{n,j}(\mathbf{x})\}$ and $\{o_{n,j}(\mathbf{x})\}$, and extract the best-fit model parameters α and β as features.

b) Use the GGD to model the smoothed directional gradient components of $\{e_{n,j}(\mathbf{x})\}$ and $\{o_{n,j}(\mathbf{x})\}$, and also use the best-fit model parameters as quality-aware NSS features.

c) Likewise, use the Weibull distribution (Eq. (8)) to model

the smoothed gradient magnitudes of $\{e_{n,j}(\mathbf{x})\}$ and $\{o_{n,j}(\mathbf{x})\}$ and take the best-fit model parameters a and b as additional NSS features.

E. Statistics of Colors

In order to further capture statistical properties that particularly pertain to color in images, we resort to a simple yet classical NSS model [38]. In [38], Ruderman *et al*. showed that in a logarithmic-scale opponent color space, the distributions of photographic image data conform well to a Gaussian probability model.

Given an RGB image having three channels $R(i,j)$, $G(i,j)$, and $B(i,j)$, first convert it into a logarithmic signal with mean subtracted:

$$\begin{aligned}\mathcal{R}(i,j) &= \log R(i,j) - \mu_R \\ \mathcal{G}(i,j) &= \log G(i,j) - \mu_G \\ \mathcal{B}(i,j) &= \log B(i,j) - \mu_B\end{aligned} \tag{11}$$

where μ_R, μ_G and μ_B are the mean values of $\log R(i,j)$, $\log G(i,j)$ and $\log B(i,j)$, respectively, over the entire image. Then, image pixels expressed in $(\mathcal{R},\mathcal{G},\mathcal{B})$ space are projected onto an opponent color space:

$$\begin{aligned}l_1(x,y) &= (\mathcal{R}+\mathcal{G}+\mathcal{B})/\sqrt{3} \\ l_2(x,y) &= (\mathcal{R}+\mathcal{G}-2\mathcal{B})/\sqrt{6} \\ l_3(x,y) &= (\mathcal{R}-\mathcal{G})/\sqrt{2}\end{aligned} \tag{12}$$

As shown in [38], the distributions of the coefficients l_1, l_2 and l_3 of natural images nicely conform to a Gaussian probability law. Thus, we use the following Gaussian model to fit the empirical density function of l_1, l_2 and l_3:

$$f(x;\zeta,\rho^2) = \frac{1}{\sqrt{2\pi}\rho}\exp\left(\frac{-(x-\zeta)^2}{2\rho^2}\right). \tag{13}$$

For each of the channels l_1, l_2 and l_3, we estimate the two model parameters ζ and ρ^2 and take them as quality-aware NSS features.

Here we use an example to show how the distributions of l_1, l_2 and l_3 vary as a function of distortion. Fig. 3(a) shows a reference image while Figs. 3(b) and 3(c) show two distorted versions of it. The image in Fig. 3(b) suffers from additive noise in the color components while Fig. 3(c) suffers from contrast compression. Figs. 3(d), 3(e) and 3(f) plot the corresponding histograms of l_1, l_2 and l_3. It may be observed that the distributions of l_1, l_2 and l_3 are significantly modified by the presence of distortions, which indicates their potential effectiveness as quality-aware NSS features for image quality prediction.

III. The IL-NIQE Index

Given the five types of quality-aware NSS features just described, we next derive from them a powerful opinion-unaware BIQA model called Integrated Local NIQE (IL-NIQE). First, a pristine multivariate Gaussian (MVG) model of the NSS features is learned from a collection of stored pristine images. Then, from each patch of a given test image, an MVG model is fitted to the feature vector and its local quality score is computed by comparing it with the learned pristine MVG model. Finally, the overall quality score of the test image is obtained by pooling the local quality scores.

Fig. 4: The 90 images used to learn the pristine MVG model used to create IL-NIQE.

A. Pristine MVG Model Learning

We learn a pristine MVG model to create a representation of the NSS features of natural pristine images. In IL-NIQE, the pristine MVG model serves as a "reference" against which to evaluate the quality of a given natural image patch. To learn the desired model, we collected a set of high quality natural images from the Internet. Four volunteers (postgraduate students from Tongji University, Shanghai, China) were involved and each of them was asked to search for 100 high quality

images from 4 categories: people, plants, animals, and man-made objects. This is similar to the process used to create a pristine corpus on which the NIQE index [10] was built. A large percentage of natural images fall within these categories. Then, each of the 400 collected images was visually examined by seven volunteer observers (undergraduate students from Tongji University). If no fewer than five of the seven observers found that the quality of an image was very good, then the image was retained. In the end, 90 images were thus selected as pristine images and thumbnails of all of these images are shown in Fig. 4. This process was more systematic than the image selection of the NIQE database. Note that none of the images used here can be found in any of the benchmark IQA databases that will be later used to evaluate the BIQA methods.

In IL-NIQE, there are several parameters closely related to the scale of the image, such as the parameters of log-Gabor filters. Parameters that are tuned to be optimal for one specific scale may not work well for a different scale. Like SSIM [27] and FSIM [26], a simple and practical strategy to solve this issue is to resize the images to a fixed size so that one algorithm can deal with images with different sizes. In our algorithm, each pristine image in the corpus was resized to a fixed size $P \times P$ using the bicubic interpolation method, then partitioned into patches of size $p \times p$. The NSS features described in Section II were then extracted from each patch. To make the extracted NSS features more meaningful for quality prediction, only a subset of the patches are used, based on a measure of patch contrast. The contrast at each pixel was computed as Eq. (3), then the patch contrast was computed as the sum of contrasts within each patch. Only those patches having a supra-threshold contrast greater than a threshold were selected to learn the MVG model, where the threshold was empirically determined by 78% of the peak patch contrast over each image. In order to enhance the quality prediction performance of IL-NIQE, all of the NSS features were computed over two scales (by down-sampling the images by a factor of 2) to capture multi-scale attributes of the images.

Each selected patch yields a d-dimensional feature vector by stacking all the NSS features extracted from it. As might be expected, some of the NSS features will be correlated with others (e.g., the gradient components and the gradient magnitude features). Therefore, we apply PCA to the feature vector to reduce its dimension. This can both reduce the computational cost and make the quality prediction process more efficient. Denote by $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n] \in \mathcal{R}^{d \times n}$ the matrix of feature vectors extracted from n selected image patches. By applying PCA to $\mathbf{X}$, we can learn a projection matrix $\mathbf{\Phi} \in \mathcal{R}^{d \times m}$, formed by the m ($m < d$) principle projection vectors associated with the m most significant eigenvalues of the covariance matrix of $\mathbf{X}$. Given $\mathbf{\Phi}$, each feature vector $\mathbf{x}_i$ is transformed as

$$\mathbf{x}'_i = \mathbf{\Phi}^T \mathbf{x}_i, \mathbf{x}'_i \in \mathcal{R}^{m \times 1}, i = 1, ..., n \tag{14}$$

By assuming that $\mathbf{x}'_i$, $i = 1, ..., n$, are independent samples from an m-dimensional MVG distribution, we can then learn the MVG distribution from $\{\mathbf{x}'_i\}$ using the standard maximum likelihood estimation technique. The learned MVG model is:

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{m/2}|\mathbf{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x}-\mu)^T \mathbf{\Sigma}^{-1}(\mathbf{x}-\mu)\right) \tag{15}$$

where $\mathbf{x} \in \mathcal{R}^{m \times 1}$ is the vector variable, and μ and $\mathbf{\Sigma}$ are the mean vector and the covariance matrix of $\mathbf{x}$. Note that the MVG model is fully described by the pair $(\mu, \mathbf{\Sigma})$.

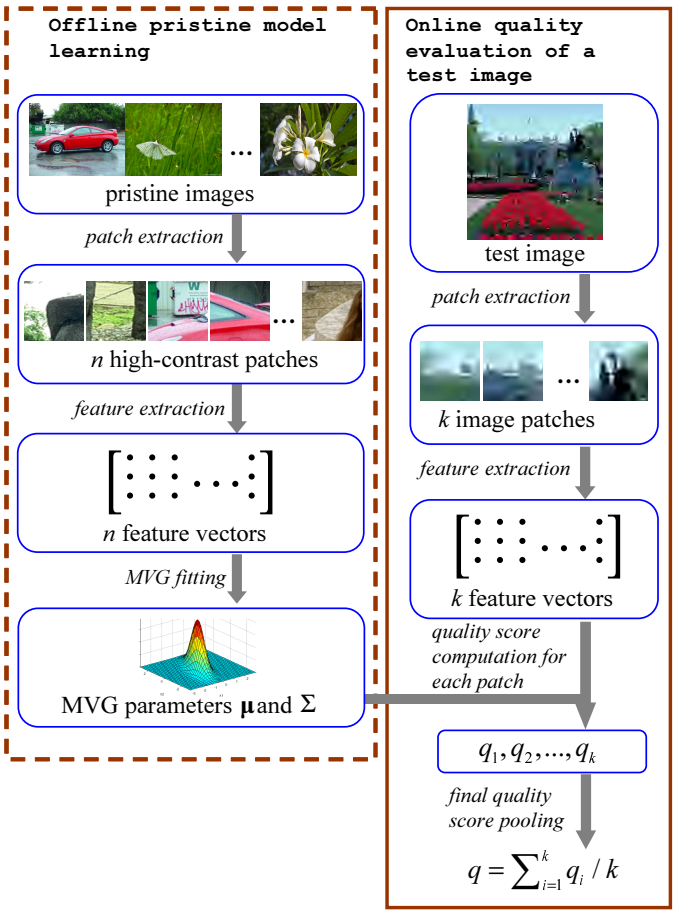


Fig. 5: Processing flow of the proposed IL-NIQE method.

B. The IL-NIQE Index

After the pristine MVG model $(\mu, \mathbf{\Sigma})$ is learned, we can use it to measure the quality of any patch in a given naturalistic test image. As in the training stage, the test image is resized to $P \times P$ and partitioned into k patches of size $p \times p$. From each patch i extract a d-dimensional NSS feature vector $\mathbf{y}_i$, then reduce the dimension of $\mathbf{y}_i$ using the pre-learned projection matrix $\mathbf{\Phi}$:

$$\mathbf{y}'_i = \mathbf{\Phi}^T \mathbf{y}_i, \mathbf{y}'_i \in \mathcal{R}^{m \times 1}, i = 1, ..., k. \tag{16}$$

Having obtained the feature set $\{\mathbf{y}'_i\}_{i=1}^k$ of a test image, we can now predict its quality score. Previous IQA studies have shown that different local regions of an image can deliver different contributions to the perception of the overall image quality [26], [39]–[43]. Therefore, each patch i is fitted by an MVG model, denoted by $(\mu_i, \mathbf{\Sigma}_i)$, which is then compared

with the pristine MVG model $(\mu, \mathbf{\Sigma})$, yielding a prediction of the local quality score of patch i. The overall quality score of the test image can then be obtained from the scores of all patches using a pooling strategy. Here we use simple average pooling.

The MVG model $(\mu_i, \mathbf{\Sigma}_i)$ could be constructed using the NSS feature vectors estimated from neighboring patches. However, this can be very costly. For simplicity, we use the NSS feature vector $\mathbf{y}_i'$ as μ_i, and the empirical covariance matrix of the feature set $\{\mathbf{y}_i'\}$ as $\mathbf{\Sigma}_i$. That is, all patches share the same covariance matrix, denoted by $\mathbf{\Sigma}'$. Then the MVG model assigned to patch i is $(\mathbf{y}_i', \mathbf{\Sigma}')$. We use the following formula to measure the distortion level of patch i:

$$q_i = \sqrt{(\mu - \mathbf{y}_i')^T \left(\frac{\mathbf{\Sigma} + \mathbf{\Sigma}'}{2}\right)^{-1} (\mu - \mathbf{y}_i')} \qquad (17)$$

which is a modified Bhattacharyya distance [10], [45] that is also used in NIQE. In summary, Eq. (17) measures the deviation of the statistics of patch i from the reference statistics pre-learned from high quality natural images. Finally, the overall quality score of the test image is pooled as the mean of $\{q_i\}$.

Fig. 5 illustrates the processing flow in IL-NIQE when computing the quality score of a given test image.

IV. Experimental Results and Discussion

A. Implementation Details

There are a few parameters in the proposed IL-NIQE algorithm. To determine the parameter values efficiently and effectively, we tuned the parameters on a subset of the TID2008 database [44]. The subset contains the first 8 reference images in TID2008 and the associated 544 distorted images. The tuning criterion is that the parameter value leading to a higher Spearman rank-order correlation coefficient (SRCC) is chosen. As a result, in our implementation, we set P (the size of the resized image) to 504, p (the patch size) to 84, and m (the dimension of PCA transformed features) to 430. The parameters related to log-Gabor filters are set as follows: $N = 3$, $J = 4$, $\sigma_r = 0.60$, $\sigma_\theta = 0.71$, $\omega_0^1 = 0.417$, $\omega_0^2 = 0.318$, and $\omega_0^3 = 0.243$, where ω_0^1, ω_0^2, and ω_0^3 represent the three center frequencies of the log-Gabor filters at three scales. The Matlab source code of the IL-NIQE algorithm can be downloaded at www.comp.polyu.edu.hk/∼cslzhang/IQA/ILNIQE/ILNIQE.htm.

B. Database and Protocol

TABLE I: Benchmark IQA Datasets used to Evaluate IQA Indices

Dataset	Reference Images No.	Distorted Images No.	Distortion Types No.	Contains multiply-distortions?
TID2013	25	3000	24	YES
CSIQ	30	866	6	NO
LIVE	29	779	5	NO
LIVE MD1	15	225	1	YES
LIVE MD2	15	225	1	YES

Four large-scale benchmark IQA datasets were used to evaluate the proposed IL-NIQE index: TID2013 [32], CSIQ [46], LIVE [47], and LIVE Multiply Distorted [48]. The LIVE Multiply Distorted (MD) IQA dataset was constructed in two stages and we treat them as two separate datasets, denoted by LIVE MD1 and LIVE MD2, respectively. Information regarding the distorted image content and subjective scores of these datasets is summarized in Table X. It is worth noting that the TID2013 and LIVE MD datasets include images with multiple distortions.

We compare the proposed IL-NIQE model with five state-of-the-art opinion-aware NR-IQA methods, including BIQI [11], BRISQUE [15], BLIINDS2 [14], DIIVINE [12], and CORNIA [17] and two state-of-the-art opinion-unaware methods: NIQE [10] and QAC [25].

TABLE II: Results of Performance Evaluation on Each Individual Dataset

Datasets	Methods	80%		50%		10%	
		SRCC	PLCC	SRCC	PLCC	SRCC	PLCC
TID2013	BIQI	0.349	0.366	0.332	0.332	0.199	0.250
	BRISQUE	0.573	0.651	0.563	0.645	0.513	0.587
	BLIINDS2	0.536	0.628	0.458	0.480	0.402	0.447
	DIIVINE	0.549	0.654	0.503	0.602	0.330	0.391
	CORNIA	0.549	0.613	0.573	0.652	0.508	0.603
	NIQE	0.317	0.426	0.317	0.420	0.313	0.398
	QAC	0.390	0.495	0.390	0.489	0.372	0.435
	IL-NIQE	0.521	0.648	0.513	0.641	0.494	0.590
CSIQ	BIQI	0.092	0.237	0.092	0.396	0.020	0.311
	BRISQUE	0.775	0.817	0.736	0.781	0.545	0.596
	BLIINDS2	0.780	0.832	0.749	0.806	0.628	0.688
	DIIVINE	0.757	0.795	0.652	0.716	0.441	0.492
	CORNIA	0.714	0.781	0.678	0.754	0.638	0.732
	NIQE	0.627	0.725	0.626	0.716	0.624	0.714
	QAC	0.486	0.654	0.494	0.706	0.490	0.707
	IL-NIQE	0.822	0.865	0.814	0.854	0.813	0.852
LIVE	BIQI	0.825	0.840	0.739	0.764	0.547	0.623
	BRISQUE	0.933	0.931	0.917	0.919	0.806	0.816
	BLIINDS2	0.924	0.927	0.901	0.901	0.836	0.834
	DIIVINE	0.884	0.893	0.858	0.866	0.695	0.701
	CORNIA	0.940	0.944	0.933	0.934	0.893	0.894
	NIQE	0.908	0.908	0.905	0.904	0.905	0.903
	QAC	0.874	0.868	0.869	0.864	0.866	0.860
	IL-NIQE	0.902	0.906	0.899	0.903	0.899	0.903
MD1	BIQI	0.769	0.831	0.580	0.663	0.159	0.457
	BRISQUE	0.887	0.921	0.851	0.873	0.829	0.860
	BLIINDS2	0.885	0.925	0.841	0.879	0.823	0.859
	DIIVINE	0.846	0.891	0.805	0.836	0.631	0.675
	CORNIA	0.904	0.931	0.878	0.905	0.855	0.889
	NIQE	0.909	0.942	0.883	0.921	0.874	0.912
	QAC	0.418	0.597	0.406	0.552	0.397	0.541
	IL-NIQE	0.911	0.930	0.899	0.916	0.893	0.907
MD2	BIQI	0.897	0.919	0.835	0.860	0.769	0.773
	BRISQUE	0.888	0.915	0.864	0.881	0.849	0.867
	BLIINDS2	0.893	0.910	0.852	0.874	0.850	0.868
	DIIVINE	0.888	0.916	0.855	0.880	0.832	0.851
	CORNIA	0.908	0.920	0.876	0.890	0.843	0.866
	NIQE	0.834	0.884	0.808	0.860	0.796	0.852
	QAC	0.501	0.718	0.480	0.689	0.473	0.678
	IL-NIQE	0.928	0.915	0.890	0.895	0.882	0.896

Two commonly used metrics were employed to evaluate the performances of the competing BIQA methods. The first is the SRCC between the objective scores predicted by the BIQA models and the subjective mean opinion scores (MOS or DMOS) provided by the dataset. SRCC operates only on the rank of the data points and ignores the relative distances between data points, hence measures the prediction monotonicity of an IQA index. The second performance metric is the Pearson linear correlation coefficient (PLCC) between MOS

and the objective scores following a nonlinear regression. The nonlinear regression uses the following mapping function [47]:

$$f(x) = \beta_1 \left(\frac{1}{2} - \frac{1}{1 + \exp(\beta_2(x - \beta_3))} \right) + \beta_4 x + \beta_5 \quad (18)$$

where $\beta_i, i = 1, 2, ..., 5$ are the parameters to be fitted. A better objective IQA index is expected to have higher SRCC and PLCC values. More details regarding the underpinnings of these two performance metrics can be found in [40].

Like NIQE, the new IL-NIQE model is "completely blind". It is independent of any IQA dataset in the learning stage, and does not have any implicit or explicit restrictions on distortion type during the testing stage. In order to make a fair and comprehensive comparison with existing BIQA methods, in the following Sections IV-C and IV-D, we followed established experimental test methodologies when conducting experiments on each dataset and we also performed cross-dataset experiments.

C. Performance on Individual Datasets

We first evaluated the various BIQA models on each individual dataset. Since opinion-aware methods require distorted images to learn the model, we partitioned each dataset into a training subset and a testing subset. We report results under three partition proportions: distorted images associated with 80%, 50%, and 10% of the reference images were used for training and the remaining 20%, 50%, and 90% were used for testing. Each partition was randomly conducted 1,000 times on each dataset and the median SRCC and PLCC scores were computed as reported in Table II. Although IL-NIQE, NIQE and QAC do not need training on the dataset, we report their results on the partitioned test subset to make the comparison consistent.

The results in Table II lead us to the following conclusions. First, the prediction performance of opinion-aware methods tends to drop with decreases in proportion of the training subset. However, it is very interesting to observe that the partition ratio has little effect on the performance of the opinion-unaware methods. Indeed, when the partition ratio is low (e.g., 10%), the opinion-unaware methods, especially IL-NIQE, actually perform better than all the opinion-aware methods. Second, IL-NIQE performs much better than the other two opinion-unaware methods, NIQE and QAC. While this is to be expected given that IL-NIQE may be viewed as a feature-enriched version of NIQE, the increment in performance is quite significant. Third, on TID2013, LIVE and MD1, IL-NIQE achieves performance comparable with the opinion-aware models CORNIA, BRISQUE and BLIINDS2, and performs better than the other two opinion-aware methods over all partition ratios. Finally, IL-NIQE performs significantly better than all of the competing models on CSIQ and MD2, even though it does not need distorted images or human subjective scores on them to train the model.

D. Cross-dataset Performance Evaluation

The evaluation strategy used in Section IV-C is inadequate to evaluate the generalization capability of a BIQA model. By partitioning a single dataset into a training subset and a test subset having the same distortions, there is no demonstration of the efficacy of an IQA model on other possibly unknown distortions. Therefore, such testing methodologies actually are not "completely blind". Here, we test the generalization capability of opinion-aware models by training them on one dataset, then testing them on other datasets. Of course, for opinion-unaware methods, the training step is not needed.

The quality prediction models of the five competing opinion-aware BIQA methods, all trained on the entire LIVE dataset, were provided by the original authors of those models. We used these to test performance on the other datasets. The results are shown in Table III. For each performance measure, the two best results are highlighted in bold. Table IV tabulates the weighted-average SRCC and PLCC scores of all models over the four datasets. The weight assigned to each dataset

TABLE III: Evaluation Results when Trained on LIVE

	TID2013		CSIQ		MD1		MD2	
	SRCC	PLCC	SRCC	PLCC	SRCC	PLCC	SRCC	PLCC
BIQI	0.394	0.468	0.619	0.695	0.654	0.774	0.490	0.766
BRISQUE	0.367	0.475	0.557	0.742	0.791	0.866	0.299	0.459
BLIINDS2	0.393	0.470	0.577	0.724	0.665	0.710	0.015	0.302
DIIVINE	0.355	0.545	0.596	0.697	0.708	0.767	0.602	0.702
CORNIA	**0.429**	**0.575**	**0.663**	**0.764**	0.839	0.871	**0.841**	**0.864**
NIQE	0.311	0.398	0.627	0.716	**0.871**	**0.909**	0.795	0.848
QAC	0.372	0.437	0.490	0.708	0.396	0.538	0.471	0.672
IL-NIQE	**0.494**	**0.589**	**0.815**	**0.854**	**0.891**	**0.905**	**0.882**	**0.897**

TABLE IV: Weighted-average Performance Evaluation Based on Table III

	BIQI	BRISQUE	BLIINDS2	DIIVINE	CORNIA	NIQE	QAC	**IL-NIQE**
SRCC	0.458	0.424	0.424	0.435	0.519	0.429	0.402	**0.599**
PLCC	0.545	0.548	0.525	0.595	0.643	0.512	0.509	**0.675**

TABLE V: Evaluation Results when Trained on TID2013

	LIVE		CSIQ		MD1		MD2	
	SRCC	PLCC	SRCC	PLCC	SRCC	PLCC	SRCC	PLCC
BIQI	0.047	0.311	0.010	0.181	0.156	0.175	0.332	0.380
BRISQUE	0.088	0.108	0.639	0.728	0.625	0.807	0.184	0.591
BLIINDS2	0.076	0.089	0.456	0.527	0.507	0.690	0.032	0.222
DIIVINE	0.042	0.093	0.146	0.255	0.639	0.669	0.252	0.367
CORNIA	0.097	0.132	**0.656**	**0.750**	0.772	0.847	0.655	0.719
NIQE	**0.906**	**0.904**	0.627	0.716	**0.871**	**0.909**	**0.795**	**0.848**
QAC	0.868	0.863	0.490	0.708	0.396	0.538	0.471	0.672
IL-NIQE	**0.898**	**0.903**	**0.815**	**0.854**	**0.891**	**0.905**	**0.882**	**0.897**

TABLE VI: Weighted-average Performance Evaluation Based on Table V

	BIQI	BRISQUE	BLIINDS2	DIIVINE	CORNIA	NIQE	QAC	**IL-NIQE**
SRCC	0.074	0.384	0.275	0.172	0.461	0.775	0.618	**0.861**
PLCC	0.250	0.491	0.349	0.251	0.527	0.821	0.744	**0.882**

TABLE VII: Analysis of Performance (SRCC) Improvement

	TID2013	CSIQ	LIVE	MD1	MD2
NIQE	0.311	0.627	0.906	0.871	0.795
L-NIQE	0.305	0.548	0.893	0.865	0.794
I-NIQE	0.463	0.770	0.883	0.879	0.876
IL-NIQE	0.494	0.815	0.898	0.891	0.882

TABLE VIII: Performance (SRCC) Evaluation of BIQA Models on Each Individual Distortion Type

Datasets	Distortion type	BIQI	BRISQUE	BLIINDS2	DIIVINE	CORNIA	NIQE	QAC	IL-NIQE
TID2013	Additive Gaussian Noise	0.7842	0.8523	0.7226	**0.8553**	0.7561	0.8194	0.7427	**0.8760**
	Additive Noise in Color Components	0.5405	0.7090	0.6497	0.7120	**0.7498**	0.6699	0.7184	**0.8159**
	Spatially Correlated Noise	0.4653	0.4908	**0.7674**	0.4626	0.7265	0.6660	0.1693	**0.9233**
	Masked Noise	0.4938	0.5748	0.5127	0.6752	**0.7262**	**0.7464**	0.5927	0.5120
	High Frequency Noise	**0.8773**	0.7528	0.8245	**0.8778**	0.7964	0.8449	0.8628	0.8685
	Impulse Noise	0.7480	0.6299	0.6501	**0.8063**	0.7667	0.7434	**0.8003**	0.7551
	Quantization Noise	0.3894	0.7984	0.7816	0.1650	0.0156	**0.8500**	0.7089	**0.8730**
	Gaussian Blur	0.7642	0.8134	**0.8557**	0.8344	**0.9209**	0.7954	0.8464	0.8142
	Image Denoising	0.4094	0.5864	0.7116	0.7231	**0.8315**	0.5903	0.3381	**0.7500**
	JPEG Compression	0.8567	0.8521	**0.8643**	0.6288	**0.8743**	0.8402	0.8369	0.8349
	JPEG2000 Compression	0.7327	0.8925	**0.8984**	0.8534	**0.9103**	0.8891	0.7895	0.8578
	JPEG Transmission Errors	0.3035	**0.3150**	0.1170	0.2387	**0.6856**	0.0028	0.0491	0.2827
	JPEG2000 Transmission Errors	0.3670	0.3594	**0.6209**	0.0606	**0.6784**	0.5102	0.4065	0.5248
	Non Eccentricity Pattern Noise	0.0073	**0.1453**	0.0968	0.0598	**0.2857**	0.0698	0.0477	0.0805
	Local Block-wise Distortions	0.0812	**0.2235**	0.2098	0.0928	0.2188	0.1269	**0.2474**	0.1357
	Mean Shift	0.0346	0.1241	0.1284	0.0104	0.0645	0.1626	**0.3060**	**0.1845**
	Contrast Change	**0.4125**	0.0403	0.1505	**0.4601**	0.1823	0.0180	0.2067	0.0141
	Change of Color Saturation	0.1418	0.1093	0.0178	0.0684	0.0807	**0.2460**	**0.3691**	0.1628
	Multiplicative Gaussian Noise	0.6424	0.7242	0.7165	**0.7873**	0.6438	0.6940	**0.7902**	0.6932
	Comfort Noise	0.2141	0.0081	0.0178	0.1156	**0.5341**	0.1548	0.1521	**0.3599**
	Lossy Compression of Noisy Images	0.5261	0.6852	0.7193	0.6327	**0.8623**	0.8011	0.6395	**0.8287**
	Color Quantization with Dither	0.6983	0.7640	0.7358	0.4362	0.2717	**0.7832**	**0.8733**	0.7487
	Chromatic Aberrations	0.5435	0.6160	0.5397	0.6608	**0.7922**	0.5612	0.6249	**0.6793**
	Sparse Sampling and Reconstruction	0.7595	0.7841	0.8164	0.8334	**0.8624**	0.8341	0.7856	**0.8650**
CSIQ	Additive Gaussian Noise	**0.8797**	**0.9252**	0.8011	0.8663	0.7458	0.8098	0.8222	0.8502
	JPEG Compression	0.8672	**0.9093**	0.9004	0.7996	**0.9075**	0.8817	0.9016	0.8991
	JPEG2000 Compression	0.7085	0.8670	0.8949	0.8308	**0.9139**	**0.9065**	0.8699	0.9063
	Additive Pink Gaussian Noise	0.3242	0.2529	0.3789	0.1766	**0.4199**	0.2993	0.0019	**0.8740**
	Gaussian Blur	0.7713	**0.9033**	0.8915	0.8716	**0.9172**	0.8953	0.8363	0.8578
	Global Contrast Decrements	**0.5855**	0.0241	0.0117	0.3958	0.3017	0.2271	0.2446	**0.5012**
LIVE	JPEG2000 Compression	—	—	—	—	—	**0.9186**	0.8621	**0.8939**
	JPEG Compression	—	—	—	—	—	**0.9412**	0.9362	**0.9418**
	White Noise	—	—	—	—	—	**0.9718**	0.9511	**0.9807**
	Gaussian Blur	—	—	—	—	—	**0.9328**	0.9134	**0.9153**
	Bit Errors in JPEG2000 Stream	—	—	—	—	—	**0.8635**	0.8231	**0.8327**
LIVE MD1	Blur + JPEG	0.6542	0.7912	0.6654	0.7080	0.8389	**0.8709**	0.3959	**0.8911**
LIVE MD2	Blur + Gaussian Noise	0.4902	0.2992	0.0151	0.6020	**0.8413**	0.7945	0.4707	**0.8824**

depends linearly on the number of distorted images contained in that dataset. We also trained the opinion-aware methods on the entire TID2013 dataset, then tested them on the other datasets. The results are shown in Tables V and VI.

From the results in Tables II∼VI, we can draw a number of interesting conclusions. First, the opinion-unaware methods NIQE and IL-NIQE exhibit clear performance advantages over their opinion-aware counterparts. (QAC is applicable to only 4 types of distortions, and hence its performance is not good.) In particular, when trained on TID2013 and then applied to other datasets, the opinion-aware methods deliver poor performance owing to their limited generalization capability. Secondly, on LIVE, IL-NIQE and NIQE achieve almost the same results, which is unsurprising given that many top models achieve high correlations on that legacy database. On the other four datasets, it can be seen that IL-NIQE always performs better than NIQE, which nicely demonstrates the much better generalization capability of IL-NIQE. Thirdly, IL-NIQE achieves the best results in nearly every scenario. In the cases where IL-NIQE is not the best, its results nearly match the best one. The superiority of IL-NIQE to other methods is clearly observable in Tables IV and VI. These results lead us to express the belief that a properly designed opinion-unaware IQA model can compete with opinion-aware methods.

E. Analysis of Performance Improvement

As compared with NIQE [10], the novelty of the IL-NIQE index lies largely in two directions. First, we enriched the feature set by introducing three new types of quality-aware NSS features. Second, instead of using a single global MVG model to describe the test image, we locally fit each patch of the test image with an MVG model. Here we explain and demonstrate the performance improvement afforded by each new aspect of IL-NIQE. Denote by L-NIQE an algorithm using the exact set of features as NIQE but using the local MVG model fit used by IL-NIQE to represent the test image. We denote by I-NIQE an algorithm that uses all five types of features in IL-NIQE but instead using a global MVG model (as in NIQE) to represent the test image. Their performances in terms of SRCC are reported in Table VII. We also present the performance of NIQE and IL-NIQE in Table VII for comparison.

From the results presented in Table VII, the following conclusions can be drawn. First, L-NIQE performs worse than NIQE, while IL-NIQE performs better than I-NIQE. This strongly suggests that the enriched set of quality-aware features not used by NIQE or L-NIQE provides the greatest boost in performance. Interestingly, global MVG model representation yields more stable performance than the local MVG-based one when incorporated into NIQE, with the converse being true for I-NIQE versus IL-NIQE. Clearly, the three new types of quality-aware features can greatly improve the performance of quality prediction. It is likely that the new features better capture detailed local expressions of distortion, given that local measurements further improve the model

performance of IL-NIQE over I-NIQE.

F. Performance on Individual Distortion Types

Though in this paper we mainly focus on studying BIQA models which are not designed for specific distortion types, it is interesting to know their performance on each individual distortion type. In this experiment, we examine the performance of competing methods on each type of distortion. For the five opinion-aware methods, we trained their quality prediction models on the entire LIVE dataset and hence we did not test them on LIVE. SRCC was used as the performance metric. The results are summarized in Table VIII. The best two results are highlighted in bold.

From the results presented in Table VIII, we make the following observations. First, for most commonly encountered distortion types, such as "additive Gaussian noise", "additive noise in color components", "spatially correlated noise", "quantization noise", "JPEG or JP2K compression", "additive pink Gaussian noise", "Gaussian blur", "blur + JPEG" and "blur + Gaussian noise", the proposed BIQA model IL-NIQE delivers very promising results. Second, on several special distortion types peculiar to TID2013, such as "non eccentricity pattern noise", "local block-wise distortion", "mean shift", "contrast change" and "change of color saturation", none of the evaluated BIQA models was able to obtain satisfying results. One possible reason is that these distortion types are hard to characterize using the features of existing BIQA models. It may also call into question the veracity of the quality scores on those distortions. In future work, we may need to investigate how to deal with these "hard" distortion types more properly.

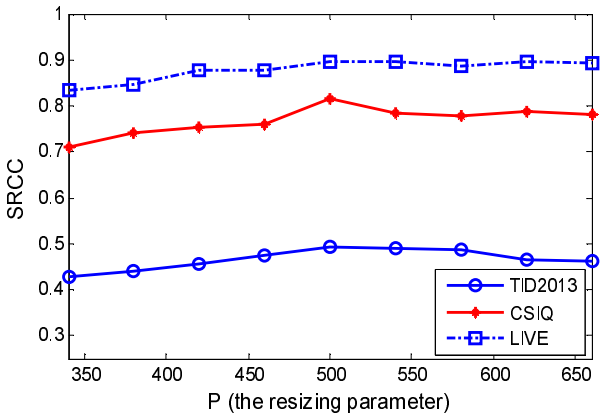
G. Performance Evaluation of Each Feature Type

In IL-NIQE, we used five types of features, including MSCN based features, MSCN products based features, gradients based features, log-Gabor responses based features, and color based features. In order to understand the relative contribution of each type of feature in IL-NIQE, we separately evaluated the performance of each feature on all the databases. SRCC is used as the performance metric. The results are reported in Table IX.

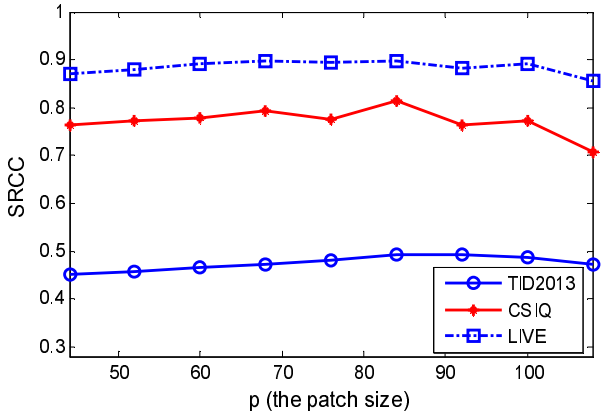
TABLE IX: Performance (SRCC) of Each of the Five Types of Features Used in IL-NIQE

Datasets	MSCN	MSCN Prod.	Gradients	log-Gabor	Color
TID2013	0.2966	0.2954	0.3701	0.4465	0.0416
CSIQ	0.5262	0.4804	0.6512	0.6418	0.0928
LIVE	0.8458	0.8401	0.6393	0.8340	0.3228
MD1	0.8321	0.8606	0.3702	0.8595	0.1110
MD2	0.6120	0.7298	0.4072	0.8714	0.1114

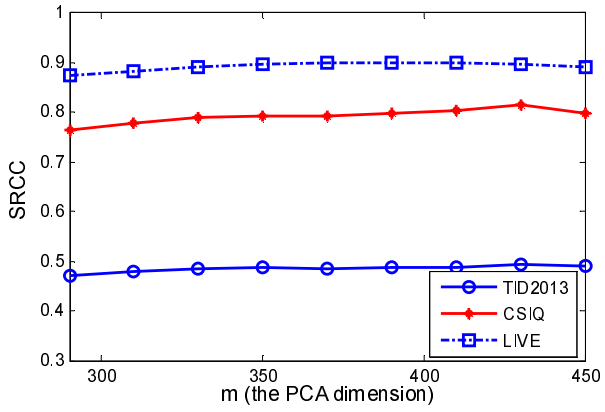
From the results shown in Table IX, we can draw the following conclusions. First, it may be seen that by using a single type of features, the BIQA performance is much worse than using the integrated features. Secondly, among the five types of features, the ones based on log-Gabor responses perform the best. It suggests that NSS features derived from



(a)



(b)



(c)

Fig. 6: The performance (SRCC) of IL-NIQE w.r.t. variations of (a) P (the resizing parameter), (b) p (the patch size), and (c) m (the PCA dimension) on TID2013, CSIQ and LIVE.

TABLE X: Time Cost of Each BIQA Model

BIQA Model	Time Cost (seconds)
BIQI	1.3861
BRISQUE	0.3332
BLIINDS2	55.6897
DIIVINE	13.1514
CORNIA	2.8978
NIQE	0.2485
QAC	0.0739
IL-NIQE	4.0924

multi-scale and multi-orientation log-Gabor responses effectively characterize image quality. Thirdly, when used alone, color-based features perform relatively poorly in regards to predicting quality distortions on these databases. However, by integrating the color features, the performance of the IL-NIQE model is indeed improved to some extent. Without color-based features, the weighted-average SRCC of the proposed model over the five datasets is 0.6369, while with color-based features its weighted-average SRCC is 0.6448.

H. Sensitivity to Parameter Variations

The parameters of the proposed IL-NIQE were tuned by a subset of TID2008 database, which contains the first 8 reference images and the associated 544 distorted images. As demonstrated in previous sub-sections, with the fixed parameters IL-NIQE performs very well on all the test datasets. In this sub-section, we further show that IL-NIQE's performance is not sensitive to the variations of parameters. Specifically, we conducted three experiments to test IL-NIQE by varying three key parameters, the resizing parameter P, the patch size p, and the PCA dimension m. The range of P is from 340 to 660 and the step size is 40. The range of p is from 44 to 108 and the step size is 8. The range of m is from 290 to 450 and the step size is 10. The results on TID2013, CSIQ and LIVE are presented in Fig. 6. SRCC is used as the performance metric.

From the results shown in Fig. 6, it can be seen that IL-NIQE's performance is robust to the parameter variations in a moderately large range. IL-NIQE has high generalization capability and does not depend on any specific test data. Even when new test data come, there is no necessary to adjust the parameter settings of IL-NIQE.

I. Computational Cost

The computational cost of each competing BIQA model was also evaluated. Experiments were performed on a standard HP Z620 workstation with a 3.2GHZ Intel Xeon E5-1650 CPU and an 8G RAM. The software platform was Matlab R2014a. The time cost consumed by each BIQA model for evaluating the quality of a 512×512 color image (taken from CSIQ) is listed in Table X. IL-NIQE has a moderate computational complexity.

V. Conclusion

We have proposed an effective new BIQA method that extends and improves upon the novel "completely blind" IQA concept introduced in [10]. The new model, IL-NIQE, extracts five types of NSS features from a collection of pristine naturalistic images, and uses them to learn a multivariate Gaussian (MVG) model of pristine images, which then serves as a reference model against which to predict the quality of the image patches. For a given test image, its patches are thus quality evaluated, then patch quality scores are averaged, yielding an overall quality score. Extensive experiments show that IL-NIQE yields much better quality prediction performance than all the compared competing methods. A significant message conveyed by this work is that "completely blind" opinion-unaware BIQA models can indeed compete with opinion-aware models. We expect that even more powerful opinion-unaware BIQA models will be developed in the near future.

References

[1] Z. Wang, A.C. Bovik, and B.L. Evans, "Blind measurement of blocking artifacts in images," in *Proc. IEEE Int'l. Conf. Image Process.*, 2000, pp. 981-984.

[2] F. Pan, X. Lin, S. Rahardja, W. Lin, E. Ong, S. Yao, Z. Lu, and X. Yang, "A locally adaptive algorithm for measuring blocking artifacts in images and videos," *Signal Process.: Image Commu.*, vol. 19, no. 6, pp. 499-506, Jul. 2004.

[3] H. Liu, N. Klomp, and I. Heynderickx, "A no-reference metric for perceived ringing artifacts in images," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 20, no. 4, pp. 529-539, Apr. 2010.

[4] R. Ferzli and L. Karam, "A no-reference objective image sharpness metric based on the notion of just noticeable blur (JNB)," *IEEE Trans. Image Process.*, vol. 18, no. 4, pp. 717-728, Apr. 2009.

[5] S. Varadarajan and L.J. Karam, "An improved perception-based no-reference objective image sharpness metric using iterative edge refinement," in *Proc. IEEE Int'l. Conf. Image Process.*, 2008, pp. 401-404.

[6] H. Sheikh, A. Bovik, and L. Cormack, "No-reference quality assessment using natural scene statistics: JPEG2000," *IEEE Trans. Image Process.*, vol. 14, no. 11, pp. 1918-1927, Nov. 2005.

[7] R. Babu, S. Suresh, and A. Perkis, "No-reference JPEG-image quality assessment using GAP-RBF," *Signal Process.*, vol. 87, no. 6, pp. 1493-1503, Jun. 2007.

[8] Z.P. Sazzad, Y. Kawayoke, and Y. Horita, "No-reference image quality assessment for JPEG2000 based on spatial features," *Signal Process.: Image Commu.*, vol. 23, no. 4, pp. 257-268, Apr. 2008.

[9] L. Liang, S. Wang, J. Chen, S. Ma, D. Zhao, and W. Gao, "No-reference perceptual image quality metric using gradient profiles for JPEG2000," *Signal Process.: Image Commu.*, vol. 25, no. 7, pp. 502-516, Aug. 2010.

[10] A. Mittal, R. Soundararajan, and A.C. Bovik, "Making a "completely blind" image quality analyzer," *IEEE Signal Process. Letters*, vol. 20, no. 3, pp. 209-212, Mar. 2013.

[11] A.K. Moorthy and A.C. Bovik, "A two-step framework for constructing blind image quality indices," *IEEE Signal Process. Letters*, vol. 17, no. 5, pp. 513-516, May 2010.

[12] A.K. Moorthy and A.C. Bovik, "Blind image quality assessment: From natural scene statistics to perceptual quality," *IEEE Trans. Image Process.*, vol. 20, no. 12, pp. 3350-3364, Dec. 2011.

[13] M.A. Saad, A.C. Bovik, and C. Charrier, "A DCT statistics-based blind image quality index," *IEEE Signal Process. Letters*, vol. 17, no. 6, pp. 583-586, Jun. 2010.

[14] M.A. Saad, A.C. Bovik, and C. Charrier, "Blind image quality assessment: A natural scene statistics approach in the DCT domain," *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3339-3352, Aug. 2012.

[15] A. Mittal, A.K. Moorthy, and A.C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4695-4708, Dec. 2012.

[16] H. Tang, N. Joshi, and A. Kapoo, "Learning a blind measure of perceptual image quality," in *Proc. IEEE Conf. Comp. Vis. Pattern Recog.*, 2011, pp. 305-312.

[17] P. Ye, J. Kumar, L. Kang, and D. Doermann, "Unsupervised feature learning framework for no-reference image quality assessment," in *Proc. IEEE Conf. Comp. Vis. Pattern Recog.*, 2012, pp. 1098-1105.

[18] C. Li, A.C. Bovik, and X. Wu, "Blind image quality assessment using a general regression neural network," *IEEE Trans. Neural Netw.*, vol. 22, no. 5, pp. 793-799, May 2011.

[19] P. Kovesi, "Image features from phase congruency," *Videre: J. Comput. Vis. Res.*, vol. 1, no. 3, pp. 1-26, Jan. 1999.

[20] D.F. Specht, "A general regression neural network," *IEEE Trans. Neural Netw.*, vol. 2, no. 6, pp. 568-576, Nov. 1991.

[21] Y. Zhang and D.M. Chandler, "No-reference image quality assessment based on log-derivative statistics of natural scenes," *J. Electron. Imaging*, vol. 22, no. 4, pp. 043025:1-23, Dec. 2013.

[22] D.J. Field, "Relations between the statistics of natural images and the response properties of cortical cells," *J. Opt. Soc. Am. A*, vol. 4, no. 12, pp. 2379-2394, Dec. 1987.

[23] W. Xue, X. Mou, L. Zhang, A.C. Bovik, and X. Feng, "Blind image quality assessment using joint statistics of gradient magnitude and Laplacian features," *IEEE Trans. Image Process.*, vol. 23, no. 11, pp. 4850-4862, Nov. 2014.

[24] A. Mittal, G.S. Muralidhar, J. Ghosh, and A.C. Bovik, "Blind image quality assessment without human training using latent quality factors," *IEEE Signal Process. Letters*, vol. 19, no. 2, pp. 75-78, Feb. 2012.
[25] W. Xue, L. Zhang, and X. Mou, "Learning without human scores for blind image quality assessment," in *Proc. IEEE Conf. Comp. Vis. Pattern Recog.*, 2013, pp. 995-1002.
[26] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: A feature similarity index for image quality assessment," *IEEE Trans. Image Process.*, vol. 20, no. 8, pp. 2378-2386, Aug. 2011.
[27] Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600-612, Apr. 2004.
[28] J. Shen, Q. Li, and G. Erlebacher, "Hybrid no-reference natural image quality assessment of noisy, blurry, JPEG2000, and JPEG images," *IEEE Trans. Image Process.*, vol. 20, no. 8, pp. 2089-2098, Aug. 2011.
[29] D.L. Ruderman, "The statistics of natural images," *Netw. Comput. Neural Syst.*, vol. 5, no. 4, pp. 517-548, Apr. 1994.
[30] K. Sharifi and A. Leon-Garcia, "Estimation of shape parameter for generalized Gaussian distributions in subband decompositions of video," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 5, no. 1, pp. 52-56, Feb. 1995.
[31] N.E. Lasmar, Y. Stitou, and Y. Berthoumieu, "Multiscale skewed heavy tailed model for texture analysis," in *Proc. IEEE Int'l. Conf. Image Process.*, 2009, pp. 2281-2284.
[32] N. Ponomarenko, O. Ieremeiev, V. Lukin, K. Egiazarian, L. Jin, J. Astola, B. Vozel, K. Chehdi, M. Carli, F. Battisti, and C.-C.J. Kuo, "Color image database TID2013: Peculiarities and preliminary results," in *Proc. Euro. Workshop on Visual Info. Process.*, 2013, pp. 106-111.
[33] T. Pouli, D.W. Cunningham, and E. Reinhard, "A survey of image statistics relevant to computer graphics," *Computer Graphics Forum*, vol. 30, no. 6, pp. 1761-1788, Sep. 2011.
[34] J.M. Geusebroek and A.W.M. Smeulders, "A six-stimulus theory for stochastic texture," *Int. J. Comp. Vis.*, vol. 62, no. 1, pp. 7-16, Apr. 2005.
[35] H.S. Scholte, S. Ghebreab, L. Waldorp, A.W.M. Smeulders, and V.A.F. Lamme, "Brain responses strongly correlate with Weibull image statistics when processing natural images," *J. Vision*, vol. 9, no. 4, pp. 29.1-15, Apr. 2009.
[36] J.M. Geusebroek, R.V.D. Boomgaard, and A.W.M. Smeulders, "Color invariance," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 23, no. 12, pp. 1338-1350, Dec. 2001.
[37] J.M. Geusebroek, R.V.D. Boomgaard, A.W.M. Smeulders, and A. Dev, "Color and scale: The spatial structure of color images," in *Proc. Euro. Conf. Comp. Vis.*, 2000, pp. 331-341.
[38] D.L. Ruderman, T.W. Cronin, and C. Chiao, "Statistics of cone response to natural images: Implications for visual coding," *J. Opt. Soc. Am. A*, vol. 15, no. 8, pp. 2036-2045, Aug. 1998.
[39] L. Zhang and H. Li, "SR-SIM: A fast and high performance IQA index based on spectral residual," in *Proc. IEEE Int'l. Conf. Image Process.*, 2012, pp. 1473-1476.
[40] Z. Wang and Q. Li, "Information content weighting for perceptual image quality assessment," *IEEE Trans. Image Process.*, vol. 20, no. 5, pp. 1185-1198, May 2011.
[41] A.K. Moorthy and A.C. Bovik, "Visual importance pooling for image quality assessment," *IEEE J. Selected Topics in Signal Process.*, vol. 3, no. 2, pp. 193-201, Apr. 2009.
[42] H. Liu and I. Heynderickx, "Visual attention in objective image quality assessment: Based on eye-tracking data," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 21, no. 7, pp. 971-982, Jul. 2011.
[43] E.C. Larson and D.M. Chandler, "Unveiling relationships between regions of interest and image fidelity metrics," in *Proc. Visual Communications and Image Process.*, 2008, pp. 68222A:1-16.
[44] N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, M. Carli, and F. Battisti, "TID2008łA database for evaluation of full-reference visual quality assessment metrics," *Adv. Modern Radioelectron.*, vol. 10, pp. 30-45, Jul. 2009.
[45] K. Fukunaga, *Introduction to Statistical Pattern Recognition*. Academic Press, 1972.
[46] E.C. Larson and D.M. Chandler, "Most apparent distortion: Full-reference image quality assessment and the role of strategy," *J. Electronic Imaging*, vol. 19, no. 1, pp. 001006:1-21, Mar. 2010.
[47] H.R. Sheikh, M.F. Sabir, and A.C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Trans. Image Process.*, vol. 15, no. 11, pp. 3440-3451, Nov. 2006.
[48] D. Jayaraman, A. Mittal, A.K. Moorthy, and A.C. Bovik, "Objective quality assessment of multiply distorted images," in *Proc. Asilomar Conf. Signals, Systems, and Computers*, 2012, pp. 1693-1697.

Lin Zhang (M'11) received the B.Sc. and M.Sc. degrees from the Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, China, in 2003 and 2006, respectively. He received the Ph.D. degree from the Department of Computing, the Hong Kong Polytechnic University, Hong Kong, in 2011. From March 2011 to August 2011, he was a Research Assistant with the Department of Computing, the Hong Kong Polytechnic University. In Aug. 2011, he joined the School of Software Engineering, Tongji University, Shanghai, China, where he is currently an Associate Professor. His current research interests include biometrics, pattern recognition, computer vision, and perceptual image/video quality assessment.

Lei Zhang (M'04-SM'14) received the B.Sc. degree in 1995 from Shenyang Institute of Aeronautical Engineering, Shenyang, P.R. China, the M.Sc. and Ph.D degrees in Control Theory and Engineering from Northwestern Polytechnical University, Xian, P.R. China, respectively in 1998 and 2001. From 2001 to 2002, he was a research associate in the Dept. of Computing, The Hong Kong Polytechnic University. From Jan. 2003 to Jan. 2006 he worked as a Postdoctoral Fellow in the Dept. of Electrical and Computer Engineering, McMaster University, Canada. In 2006, he joined the Dept. of Computing, The Hong Kong Polytechnic University, as an Assistant Professor. Since Sept. 2010, he has been an Associate Professor in the same department. His research interests include Computer Vision, Pattern Recognition, Image and Video Processing, and Biometrics, etc. Dr. Zhang has published more than 200 papers in those areas. By 2015, his publications have been cited more than 12,000 times in literature. Dr. Zhang is currently an Associate Editor of *IEEE Trans. on Image Processing*, *IEEE Trans. on CSVT* and *Image and Vision Computing*. He was awarded the 2012-13 Faculty Award in Research and Scholarly Activities. More information can be found in his homepage http://www4.comp.polyu.edu.hk/∼cslzhang/.

Alan C. Bovik (S'80-M'84-SM'89-F'96) holds the Cockrell Family Endowed Regents Chair in Engineering at The University of Texas at Austin, where he is Director of the Laboratory for Image and Video Engineering (LIVE) in the Department of Electrical and Computer Engineering and the Institute for Neuroscience. His research interests include image and video processing, computational vision, and visual perception. He has published over 700 technical articles in these areas and holds several U.S. patents. His publications have been cited more than 40,000 times in the literature, his current H-index is about 75, and he is listed as a Highly-Cited Researcher by Thompson Reuters. His several books include the companion volumes *The Essential Guides to Image and Video Processing* (Academic Press, 2009).

Dr. Bovik has received a number of major awards from the IEEE Signal Processing Society, including: the Society Award (2013); the Technical Achievement Award (2005); the Best Paper Award (2009); the Signal Processing Magazine Best Paper Award (2013); the Education Award (2007); the Meritorious Service Award (1998) and (co-author) the Young Author Best Paper Award (2013). He also was named recipient of the Honorary Member Award of the Society for Imaging Science and Technology for 2013, received the SPIE Technology Achievement Award for 2012, and was the IS&T/SPIE Imaging Scientist of the Year for 2011. He is also a recipient of the Hocott Award for Distinguished Engineering Research from the Cockrell School of Engineering at The University of Texas at Austin (2008) and the Distinguished Alumni Award from the University of Illinois at Champaign-Urbana (2008). He is a Fellow of the IEEE, the Optical Society of America (OSA), and the Society of Photo-Optical and Instrumentation Engineers (SPIE). He co-founded and was the longest-serving Editor-in-Chief of the *IEEE Transactions on Image Processing* (1996-2002), created and served as the first General Chair of the *IEEE International Conference on Image Processing*, held in Austin, Texas, in November, 1994. Dr. Bovik is a registered Professional Engineer in the State of Texas.