Blind Image Quality Assessment Using Joint Statistics of Gradient Magnitude and Laplacian Features

Wufeng Xue, Xuanqin Mou, *Member, IEEE*, Lei Zhang, *Senior Member, IEEE*, Alan C. Bovik, *Fellow, IEEE*, and Xiangchu Feng

***Abstract*— Blind image quality assessment (BIQA) aims to evaluate the perceptual quality of a distorted image without information regarding its reference image. Existing BIQA models usually predict the image quality by analyzing the image statistics in some transformed domain, e.g., in the discrete cosine transform domain or wavelet domain. Though great progress has been made in recent years, BIQA is still a very challenging task due to the lack of a reference image. Considering that image local contrast features convey important structural information that is closely related to image perceptual quality, we propose a novel BIQA model that utilizes the joint statistics of two types of commonly used local contrast features: 1) the gradient magnitude (GM) map and 2) the Laplacian of Gaussian (LOG) response. We employ an adaptive procedure to jointly normalize the GM and LOG features, and show that the joint statistics of normalized GM and LOG features have desirable properties for the BIQA task. The proposed model is extensively evaluated on three large-scale benchmark databases, and shown to deliver highly competitive performance with state-of-the-art BIQA models, as well as with some well-known full reference image quality assessment models.**

***Index Terms*— Blind image quality assessment, gradient magnitude, LOG, jointly adaptive normalization.**

I. Introduction

VISUAL signals play a profound role in our communication and interaction with the surrounding world. With the rapid development of digital imaging and network technologies, billions of digital images are available on the internet, and the number of people that share pictures on social network websites continues to increase. Indeed, it is estimated that by 2015 consumers in the U.S. will capture more than 100 billion digital pictures annually [1]. Since a variety of image distortions can be introduced during image acquisition, compression, transmission, and storage, etc., the output images may be unsatisfactory in terms of subjective quality. This is particularly true for images captured under less than ideal conditions and by low end devices such as smartphone cameras. A quantitative index on the perceptual quality of images is highly desirable for evaluating practical systems, benchmarking image processing algorithms, designing imaging systems, and monitoring image acquisition and transmission [2]. Thus, research on image quality assessment (IQA) has been extensively conducted for decades. In applications where information regarding the reference image and the distortion process is not available, the development of general purpose blind IQA (BIQA) models has become an important yet very challenging problem.

When a natural image is distorted with a known procedure, the introduced distortions can be made measureable using some specific features. For example, blur can be measured by edge width in the spatial domain [3] or kurtosis in some transformed domain [4]; blockiness introduced by JPEG/JPEG2000 compression can be quantified by statistical differences between adjacent pixels or by the zero-crossing rate around block boundaries [5]. Most of these models have been proposed to deal with a single given distortion type and they are usually referred to as distortion specific IQA models [6]. When a natural image is distorted via unknown distortion channels, the corresponding quality prediction problem becomes distortion-agnostic, and it becomes much more difficult to find specific features to measure the image quality. Fortunately, in the past two decades numerous studies have shown that high quality natural images exhibit statistical regularities [34]. When the image structures are distorted or damaged, the image statistics will be changed accordingly, making the inference of image perceptual quality possible. The quality can be measured by directly computing the distance between the statistics of high quality natural images and distorted images [7]. However, such an unsupervised BIQA model learning approach often cannot deliver

Manuscript received February 11, 2014; revised June 28, 2014; accepted August 26, 2014. Date of publication September 8, 2014; date of current version September 30, 2014. This work was supported in part by the National Natural Science Foundation of China under Grant 61172163, Grant 90920003, and Grant 61271294, in part by the Research Grants Council, Hong Kong, under Grant PolyU-5315/12E, and in part by the National Science Foundation under Grant IIS-0917175 and Grant IIS-1116656. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Erhardt Barth.

W. Xue is with the Institute of Image Processing and Pattern Recognition, Xi'an Jiaotong University, Xi'an 710049, China (e-mail: xwolfs@hotmail.com).

X. Mou is with the Institute of Image Processing and Pattern Recognition, Xi'an Jiaotong University, Xi'an 710049, China, and also with the Beijing Center for Mathematics and Information Interdisciplinary Sciences, Beijing 100190, China (e-mail: xqmou@mail.xjtu.edu.cn).

L. Zhang is with the Department of Computing, Hong Kong Polytechnic University, Hong Kong (e-mail: cslzhang@comp.polyu.edu.hk).

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

X. Feng is with the Department of Applied Mathematics, Xidian University, Xi'an, 710049, China (e-mail: xcfeng@mail.xidian.edu.cn).

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

Digital Object Identifier 10.1109/TIP.2014.2355716

high quality prediction accuracy. Alternatively, supervised BIQA model learning can help to bridge the semantic gap between image statistics and image perceptual quality. Most modern BIQA methods [8]–[12], [14]–[16] are based on the supervised learning principle, while their differences mainly lie in what statistics are employed.

Moorthy *et al.* [8] trained a support vector machine (SVM) [17] to detect image distortion types using statistical models of wavelet coefficients, and then trained a support vector regression (SVR) model [17] to predict the perceptual severity of each distortion type. A similar distortion identification and quality prediction framework is employed in [16], where the features are extracted in the wavelet domain. Saad *et al.* trained a probabilistic model based on contrast and statistical features such as kurtosis and anisotropy in the DCT domain [9], [15]. In [10], three sets of statistical features are extracted from complex wavelet transform coefficients, and then three regression models are trained on each feature set. A weighted combination of the three regression models is used to estimate image quality. In [12], a sparse representation based classifier originally developed for face recognition [13] is used to infer image quality scores. A summary of commonly used statistical features and regression algorithms for BIQA can be found in [18]. Almost all these methods follow a two stage framework: statistical feature extraction, followed by regression model learning from subjective human scores.[1] The most widely used regression algorithm is SVR with a RBF (radial basis function) kernel. The natural scene statistic (NSS) features that are employed in these BIQA models are mainly derived from the statistical distributions (histograms) of image coefficients in some bandpass transformed domain. The model parameters of best-fitted distributions (e.g., Laplacian, generalized Gaussian) to bandpass image coefficients are widely used as quality-predictive features.

Since natural images are high dimensional signals that contain a rich amount of redundancies, the extraction of statistical features may be viewed as a process of removing the redundancies to reveal the low dimensional manifold space of image perceptual quality. Complementary to distribution-based statistical description of natural images, the elementary structures of natural images can be reflected in the receptive field of retinal and cortical neurons [21]. Indeed, the principal components of natural images strongly resemble the directional derivatives of 2D Gaussian functions [19]. The work on independent components based image analysis [20] has revealed that the diverse scale- and orientation- sensitivities of retino-cortical receptive fields can also be closely modeled by Gaussian derivative functions. The work of Field and Olshausen [21] validates that natural images can be sparsely expanded over an overcomplete set of simple atoms. This finding also accounts for the 'sparsity prior' of bandpass image coefficients.

Image luminance changes convey most of the meaningful information of an image. Bandpass image responses, in particular Gaussian derivative responses, can be used to characterize various image semantic structures, such as lines, edges, corners, and blobs, etc., which closely relate to human subjective perception of image quality. Many models have been proposed to extract and analyze these kinds of 'singular' structures in natural images, such as the Gaussian smoothed gradient magnitude [23], Laplacian of Gaussian (LOG) operator [22], discrete wavelet transform (DWT) [24], discrete cosine transform (DCT) [25], first-order Markov models [36], partial differential equations (PDE) [27], and so on.

While many BIQA models use DCT, DWT and other multiscale bandpass transforms to de-correlate images, some state-of-the-art full/reduced reference IQA models [28]–[33] rely on local spatial contrast features, e.g., the gradient magnitude (GM) and the LOG to predict image quality. These two types of features share the common property that they are computed using isotropic differential operators, i.e., without angular favor. LOG filters have a center-surrounded profile that is symmetrically sensitive to intensity changes across all orientations, while GM features reflect the maximum intensity variation regardless of orientation. To the best of our knowledge, no existing general purpose BIQA models have explicitly made use of these Gaussian derivative features, despite their physiological similarities to the receptive field responses of neurons along the visual pathway. One possible exception is BRISQUE model [14], which computes center-surrounded mean-subtracted contrast normalized (MSCN) coefficients as features. Such features can be viewed as simplified LOG response signals with contrast masking. However, BRISQUE does not utilize the complementary GM-like features.

By contrast with existing NSS-based BIQA models, here we show that low-order Gaussian derivative operators, exemplified by GM and LOG, can be employed to develop high performance BIQA models. The GM and LOG features can be used to build the basic elements (i.e., local contrast) of image semantic structures, and they are hence closely related to the perceptual quality of natural images. The LOG operator responds to intensity contrast in a small spatial neighborhood, and it is a good model of the receptive field of retinal ganglion cells [22], [34]. The GM feature measures the strength of local luminance change. A contour formed by the locally maximum GM pixels may be regarded as an image edge profile. The GM and LOG features also align with the classical 'edge' and 'bar' types of features described in [35], which resemble the independent components (IC) of natural images. Rather than computing expensive IC-based NSS features, we use the easy-to-compute GM and LOG features to perform BIQA tasks. Indeed, such low-order Gaussian derivative operators have been employed in many computer vision applications, including, for example, the Harris corner detection [47], the SIFT/SUFR [48], [49] operators for object matching, and the HOG [50] features used for human detection, to name a few. The effectiveness of these Gaussian derivative based features in the above applications motivated us to introduce them into the task of BIQA.

The novelty of our work lies in that we propose to use the joint statistics of simple GM and LOG features for BIQA model learning. In particular, we propose the

[1]Strictly speaking, learning an IQA model from a training dataset with certain types of distortions cannot be said as a truly "blind" method. Considering that the majority of such methods in literature (see [8]–[10], [12], [15], [16]) are called as BIQA methods, we follow this naming in this paper.

joint adaptive normalization (JAN) operation to boost the performance of GM and LOG features on image quality prediction, and introduce the dependency index to describe the interaction between them and to refine the GM and LOG joint statistics. Our work is the first attempt to use simple LOG and GM features to conduct BIQA and it achieves leading performance in terms of quality prediction accuracy, distortion robustness and database generalization.

The rest of the paper is organized as follows. Section II presents in detail the features we used and the proposed methodology. Section III conducts extensive experiments to validate the proposed BIQA models. Section IV concludes the paper.

II. Methodology

As discussed in the Introduction section, GM and LOG features are basic elements that are commonly used to form image semantic structures. As we will see, they are also strong features to predict image local quality. In this section, we show how the joint statistics of GM and LOG can be adapted to the BIQA problem.

A. Gradient Magnitude (GM) and Laplacian of Gaussian (LOG)

Luminance discontinuities convey most of the structural information of a natural image, and they can be effectively detected from the responses of the GM and LOG operators. Denote by $\boldsymbol{I}$ an image. Its GM map can be computed as

$$\boldsymbol{G_I} = \sqrt{[\boldsymbol{I} \otimes \boldsymbol{h}_x]^2 + \left[\boldsymbol{I} \otimes \boldsymbol{h}_y\right]^2} \tag{1}$$

where "$\otimes$" is the linear convolution operator and $\boldsymbol{h}_d$, $d \in \{x, y\}$, is the Gaussian partial derivative filter applied along the horizontal (x) or vertical (y) direction:

$$\begin{aligned} \boldsymbol{h}_d(x, y|\sigma) &= \frac{\partial}{\partial d}\boldsymbol{g}(x, y|\sigma) \\ &= -\frac{1}{2\pi\sigma^2}\frac{d}{\sigma^2}\exp\left(-\frac{x^2+y^2}{2\sigma^2}\right)^{d\in\{x,y\},} \end{aligned} \tag{2}$$

where $\boldsymbol{g}(x, y|\sigma) = \frac{1}{2\pi\sigma^2}\exp\left(-\frac{x^2+y^2}{2\sigma^2}\right)$ is the isotropic Gaussian function with scale parameter σ. The LOG of image $\boldsymbol{I}$ is:

$$\boldsymbol{L_I} = \boldsymbol{I} \otimes \boldsymbol{h}_{LOG}, \tag{3}$$

where

$$\begin{aligned} \boldsymbol{h}_{LOG}(x, y|\sigma) &= \frac{\partial^2}{\partial x^2}\boldsymbol{g}(x, y|\sigma) + \frac{\partial^2}{\partial y^2}\boldsymbol{g}(x, y|\sigma) \\ &= \frac{1}{2\pi\sigma^2}\frac{x^2+y^2-2\sigma^2}{\sigma^4}\exp\left(-\frac{x^2+y^2}{2\sigma^2}\right). \end{aligned} \tag{4}$$

Filter templates of $\boldsymbol{h}_x$, $\boldsymbol{h}_y$, and $\boldsymbol{h}_{LOG}$ are displayed in Fig. 1.

The empirical marginal distributions of GM features of natural (photographic) images can be modeled as obeying a Weibull distribution [36], while those of LOG responses can be well modeled as following a generalized

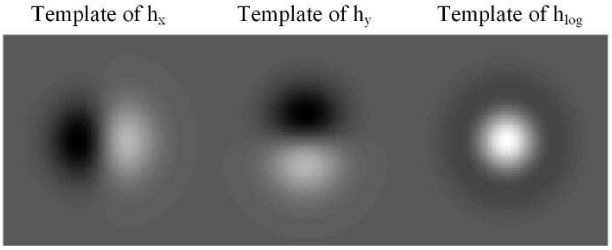


Fig. 1. The templates used for computing the GM and LOG responses.

Gaussian distribution [37]. Fig. 2 shows two natural images with different contents and a blurred chessboard image with contrast increasing linearly from top-left to bottom-right. The (cropped and zoomed) GM and LOG feature maps and their empirical distributions for the three images are depicted in the middle column of Fig. 2. For the first image *Houses*, there are many large GM coefficients and strong LOG responses, while for the second image *Hats*, there are many fine-texture induced small GM coefficients and LOG responses. Although both images are of high quality, they have very different GM and LOG distributions, implying that it is difficult to directly use the GM and LOG statistics for BIQA tasks. On the other hand, for the simulated chessboard image, the distributions of its GM and LOG features are rather similar to those of image *Hats*. Therefore, we can draw the conclusion that the marginal distributions of GM and LOG features are not stable statistical features for BIQA.

B. Joint Adaptive Normalization

GM and LOG operators could remove a significant amount of image spatial redundancies, whereas certain correlations between neighboring pixels will remain. This is also true for other bandpass feature extraction methods such as wavelet transform [38], DCT [15], MSCN [14], etc. To further remove local correlations, adaptive gain control [40] or divisive normalization models [38], [39] have been developed, both aiming to model nonlinear cortical function and to conduct objective IQA. Generally speaking, these techniques decompose an image into channels of different frequencies and orientations, then normalize each coefficient by the average energy over a local neighborhood centered at the current coefficient. Such a divisive normalization process can effectively whiten the coefficients and remove local contrast variations, resulting in a stable statistical image representation.

We decompose each image into just two channels, the GM channel and the LOG channel. As in [38] and [39], we propose to normalize the GM and LOG coefficients to obtain stable statistical image representations. Unlike [38], [39], where normalization is applied individually to each channel of different orientation and frequency, here we normalize jointly the GM and LOG channels. Let

$$\boldsymbol{F_I}(i, j) = \sqrt{\boldsymbol{G_I^2}(i, j) + \boldsymbol{L_I^2}(i, j)}. \tag{5}$$

Then a locally adaptive normalization factor is computed at each location (i, j):

$$\boldsymbol{N_I}(i, j) = \sqrt{\sum\sum_{(l,k)\in\Omega_{i,j}} \omega(l, k)\boldsymbol{F_I^2}(l, k)}, \tag{6}$$

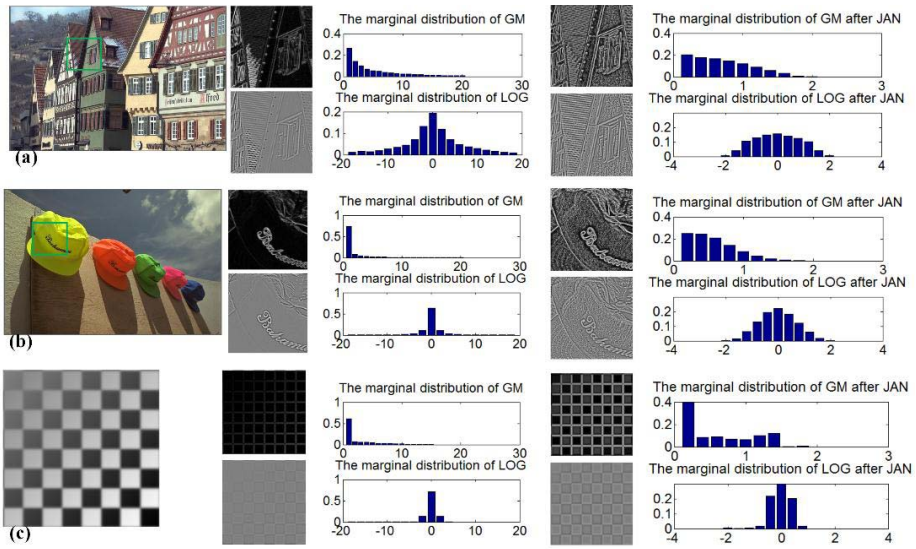


Fig. 2. The GM and LOG maps as well as their marginal distributions before (middle column) and after (right column) joint adaptive normalization. (a) *Houses*; (b) *Hats*; and (c) *Chessboard*.

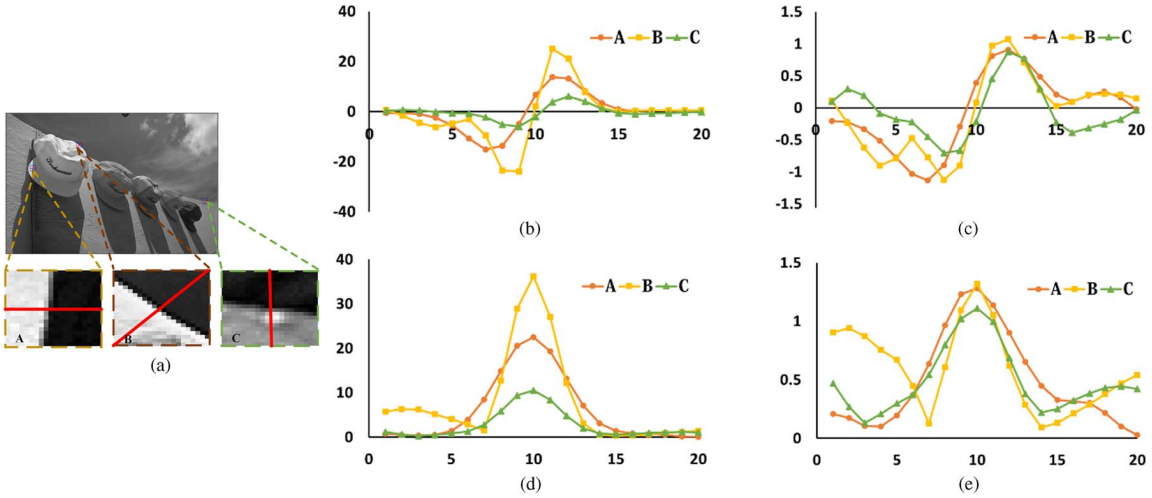


Fig. 3. Profile of GM and LOG signals along the three lines (highlighted in red) A, B and C in (a). (b) and (d) Show the GM and LOG profiles before JAN, while (c) and (e) show the corresponding profiles after JAN.

where $\Omega_{i,j}$ is a local window centered at $(i,\ j)$, $\omega(l,\ k)$ are positive weights satisfying $\Sigma_{l,k}\omega(l,k)=1$. In our implementation, we set $\omega(l,\ k)$ to be a spatially truncated Gaussian kernel rescaled to unit sum. The GM and LOG feature maps are normalized as:

$$\bar{\boldsymbol{G}}_{\boldsymbol{I}} = \boldsymbol{G}_{\boldsymbol{I}} / (\boldsymbol{N}_{\boldsymbol{I}} + \varepsilon), \tag{7}$$

$$\bar{\boldsymbol{L}}_{\boldsymbol{I}} = \boldsymbol{L}_{\boldsymbol{I}} / (\boldsymbol{N}_{\boldsymbol{I}} + \varepsilon), \tag{8}$$

where ε is a small positive constant to avoid numerical instability when $\boldsymbol{N}_{\boldsymbol{I}}$ is small. We call the above normalization procedure joint adaptive normalization (JAN).

Generally speaking, the JAN process will not change image semantic structures because it only adjusts the local image contrast scale. The benefit of JAN lies in the fact that it makes the local contrast scales of GM and LOG maps consistent across the image, and thus removes the uncertainties caused by illumination changes, varying magnitudes of edges and other structures, etc. The right column of Fig. 2 shows the GM and LOG maps of the three images after JAN. The marginal distributions of the corresponding GM and LOG maps are also shown. After JAN, the GM and LOG maps become more stationary across the whole image. The GM distributions of the two natural images *Houses* and *Hats* become very similar after JAN, though their contents are very different. So do their LOG distributions. However, for the artificial image *Chessboard*, the GM and LOG distributions after JAN become very different from those of the natural images *Houses* and *Hats*.

Let's examine more carefully how JAN adjusts the statistics of GM and LOG maps by taking image *Hats* as an example. Fig. 3(b) and 3(d) plot the GM and LOG profiles of three edges

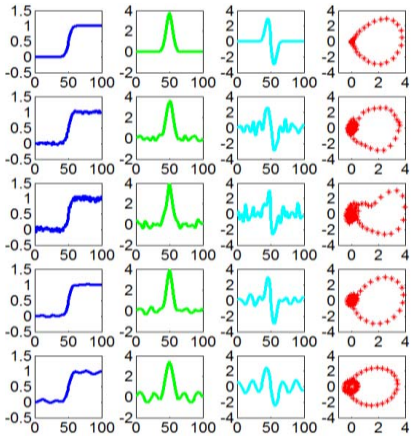


Fig. 4. A step edge signal (top row) and its two Gaussian noise corrupted versions (middle two rows) and two DCT compressed versions (bottom two rows). From left column to right column: step edge signals, their GM and LOG responses, and scatter plots of GM (x-axis) versus LOG (y-axis).

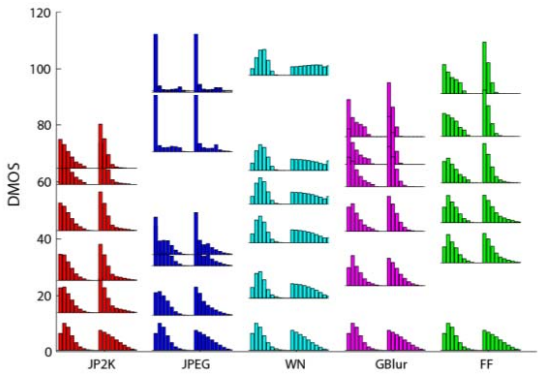


Fig. 5. Marginal probabilities P_G (shown as the first half of the histograms) and P_L (shown as the second half of the histograms) of the distorted images generated from the same reference image at different DMOS levels. The images are from the LIVE database [41], which has five types of distortions: JP2K compression, JPEG compression, white noise (WN), Gaussian blur (GB) and fast fading (FF).

(highlighted by red lines in the zoomed patches) before JAN, respectively, while Fig. 3(c) and 3(e) plot these profiles after JAN. Clearly, after JAN the differences between the profiles of the three edges are much reduced, and they tend to have more similar shapes and scales. The JAN operation reduces the dependency of image statistics on local image content.

C. Statistical Feature Description

The GM and LOG features describe local image structures from different aspects, and the interaction between them can play an important role in predicting perceptual image quality. Let us use an example to investigate the interaction between GM and LOG. The first column of Fig. 4 shows a step edge signal and several of its distorted counterparts: two Gaussian noise corrupted versions and two DCT compressed (quantized) versions. The second and third columns show the GM and LOG responses to the original and distorted step edges, while the right column of Fig. 4 shows the scatter plots of GM vs. LOG. Although distortions of the step edge can be reflected in the GM and LOG features, the GM-LOG scatter plot can better describe the signal distortions. The scatter plot of the original step signal is very smooth and regular. When the step edge is distorted, the shape of the scatter plot changes accordingly. The more severe the distortion is, the greater the change will be in the scatter plot. In addition, the change in shape of the scatter plot caused by Gaussian noise corruption is very different from that caused by DCT compression. More specifically, the scatter plots of the GM and LOG responses to the DCT compressed step edge are more regular than those of the noise corrupted ones. The above observations inspire us to explore the interaction between GM and LOG and use it to conduct BIQA.

After applying JAN to the GM and LOG features, the joint empirical distribution of $\bar{\boldsymbol{G}}_{\boldsymbol{I}}(i,j)$ and $\bar{\boldsymbol{L}}_{\boldsymbol{I}}(i,j)$ can be computed and used to learn a prediction model. We quantize $\bar{\boldsymbol{G}}_{\boldsymbol{I}}(i,j)$ into M levels $\{g_1, g_2, \ldots, g_M\}$ and $\bar{\boldsymbol{L}}_{\boldsymbol{I}}(i,j)$ into N levels $\{l_1, l_2, \ldots, l_N\}$. For conciseness of notation, we denote $\bar{\boldsymbol{G}}_{\boldsymbol{I}}(i,j)$ by G and denote $\bar{\boldsymbol{L}}_{\boldsymbol{I}}(i,j)$ by L. The joint empirical probability function of G and L can be denoted by

$$\boldsymbol{K}_{m,n} = P(G = g_m, L = l_n), m = 1, \ldots, M; n = 1, \ldots, N. \tag{9}$$

In other words, $\boldsymbol{K}_{m,n}$ is the normalized bivariate histogram of G and L. Although $\boldsymbol{K}_{m,n}$ contains a rich amount of statistical information regarding $\bar{\boldsymbol{G}}_{\boldsymbol{I}}(i,j)$ and $\bar{\boldsymbol{L}}_{\boldsymbol{I}}(i,j)$, it has a high dimensionality $(M \times N)$.

Instead of using $\boldsymbol{K}_{m,n}$ to learn the prediction model, it is desirable to extract a smaller set of quality-predictive features from $\boldsymbol{K}_{m,n}$ for this task. Intuitively, the marginal probability functions of $\bar{\boldsymbol{G}}_{\boldsymbol{I}}(i,j)$ and $\bar{\boldsymbol{L}}_{\boldsymbol{I}}(i,j)$, denoted by P_G and P_L, respectively, are straightforward choices:

$$\begin{cases} P_G(G = g_m) = \sum_{n=1}^{N} K_{m,n} \\ P_L(L = l_n) = \sum_{m=1}^{M} K_{m,n}. \end{cases} \tag{10}$$

Because of the JAN process, the marginal probability functions P_G and P_L of natural images with different contents will have similar shapes. However, when a natural image is distorted, the shapes of its P_G and P_L will deviate from those of high quality natural images. Fig. 5 shows the marginal probability functions of the distorted images of a reference image from the LIVE database [41]. To better illustrate how the marginal distributions vary with the degree of degradation, we plot the histograms of P_G and P_L across various DMOS (Difference Mean Opinion Score) levels. It can be seen that P_G and P_L gradually change with the increase of distortion level. This suggests that the shapes of P_G and P_L are predictive of image quality.

The marginal probability functions P_G and P_L, however, do not capture the dependencies between GM and LOG. If the GM and LOG features of an image are independent, then $\boldsymbol{K}_{m,n} = P_G(G = g_m) \times P_L(L = l_n)$ for all m and n. We can define the following index to measure the dependency between GM and LOG:

$$\boldsymbol{D}_{m,n} = \frac{\boldsymbol{K}_{m,n}}{P(G = g_m) \times P(L = l_n)} \tag{11}$$

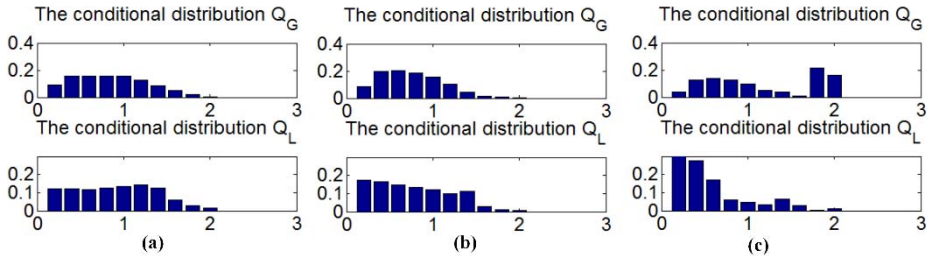


Fig. 6. Independency distribution between normalized GM and LOG features for images of (a) *Houses*, (b) *Hats*, and (c) *Chessboard*.

If the GM and LOG features of an image are independent, then $\boldsymbol{D}_{m,n}=1$ for all m and n. In practice, the GM and LOG features will have certain dependencies and $\boldsymbol{D}_{m,n}$ will take values other than 1. Directly computing and using $\boldsymbol{D}_{m,n}$ as a feature set for BIQA is not a good choice since it has the same dimension as that of $\boldsymbol{K}_{m,n}$. Instead, we can compute the dependency of each specific value $G=g_m$ against all possible values of L. Using the marginal probability $P(G=g_m)$ as a weight, define the following measure of the overall dependency of $G=g_m$ on L:

$$Q_G(G=g_m)=P(G=g_m)\cdot\frac{1}{N}\sum\nolimits_{n=1}^{N}\boldsymbol{D}_{m,n}. \tag{12}$$

Similarly, define the following measure of the overall dependency of $L=l_n$ on G:

$$Q_L(L=l_n)=P(L=l_n)\cdot\frac{1}{M}\sum\nolimits_{m=1}^{M}\boldsymbol{D}_{m,n}. \tag{13}$$

It is easy to prove that $Q_G\geq 0$ and $Q_L\geq 0$, and $\Sigma_m Q_G(G=g_m)=\Sigma_n Q_L(L=l_n)=1$. Therefore, they can be viewed as probability distributions in some sense, and we call them independency distributions.

Then Q_G and Q_L can be re-written as:

$$\begin{aligned}Q_G(G=g_m)&=\frac{1}{N}\sum\nolimits_{n=1}^{N}\frac{P(G=g_m,L=l_n)}{P(L=l_n)}\\&=\frac{1}{N}\sum\nolimits_{n=1}^{N}P(G=g_m|L=l_n),\end{aligned} \tag{14}$$

$$\begin{aligned}Q_L(L=l_n)&=\frac{1}{M}\sum\nolimits_{m=1}^{M}\frac{P(G=g_m,L=l_n)}{P(G=g_m)}.\\&=\frac{1}{M}\sum\nolimits_{m=1}^{M}P(L=l_n|G=g_m)\end{aligned} \tag{15}$$

From (14) and (15), one can see that the proposed dependency measure can be viewed as the sum of conditional probabilities of a specific value of G (or L) over variable L (or G). It describes the statistical interaction between normalized GM and LOG features. Fig. 6 plots the distributions of Q_G and Q_L for images *Houses*, *Hats* and *Chessboard*. The Q_G and Q_L of the artificial image *Chessboard* is remarkably different from those of the natural images *Houses* and *Hats*, whereas the Q_G and Q_L of *Houses* and *Hats* are quite similar. Fig. 7 plots the Q_G and Q_L of the same distorted images used in Fig. 5. One can see that Q_G and Q_L gradually change with the degree of distortion, as can be observed in Fig. 5 for the marginal distributions P_G and P_L.

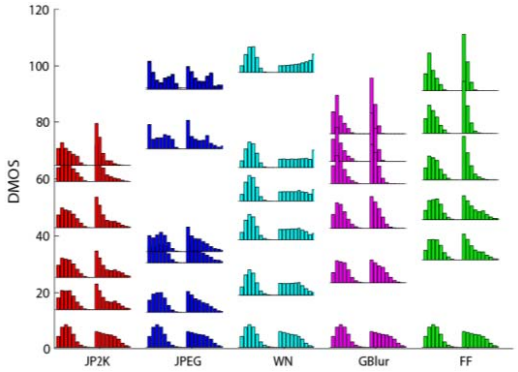


Fig. 7. The independency distributions Q_G (shown as the first half of the histograms) and Q_L (shown as the second half of the histograms) of the distorted images of a reference image at different DMOS levels. The images are from the LIVE database [41].

D. BIQA Prediction Model Learning

Based on the analyses in Section II.C, we know that the marginal distributions P_G and P_L and the independency measures Q_G and Q_L are closely related to the severity of distortions of natural images. We shall use them as statistical features to learn prediction models (i.e., regression models) for BIQA. To more comprehensively evaluate the effectiveness of the proposed statistical features, we learn three models by using different feature sets. In the 1st model, denoted by M_1, we use only the marginal distributions P_G and P_L to learn the quality prediction model; in the 2nd model, denoted by M_2, we use only the dependency measures Q_G and Q_L to learn; in the 3rd model, denoted by M_3, we use all statistical features P_G, P_L, Q_G and Q_L to learn the model.

Given the features and the DMOS scores of training images, we learn a regression function to map the features of an image to its DMOS score. The support vector regression (SVR) technique [17] is widely used to learn such regression functions [8]–[10], [14]–[16]. In this paper the ε-SVR [17] is employed for regression model learning. Given training data $\{(x_1, y_1), \ldots, (x_k, y_k)\}$, where x_i, $i=1, \ldots, k$, is the feature vector and y_i is the DMOS score, we aim to find a function to predict the score from the input feature vector: $f(x)=\langle\omega, x\rangle+b$, where $\langle\cdot,\cdot\rangle$ denotes the inner product, ω is the weight vector, and b is a bias parameter. With the constraint of flatness (small ω) and by introducing the slack variables ξ_i and ξ_i^*, ω and b can be computed by solving the following

optimization problem [17]:

$$\text{minimize } \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{k}(\xi_i + \xi_i^*)$$
$$subject\ to \begin{cases} \langle \omega, x_i\rangle - (y_i - b) \le \varepsilon + \xi_i \\ y_i - b - \langle \omega, x_i\rangle \le \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \ge 0, \end{cases} \tag{16}$$

where C is a the constant parameter to balance ω and the slack variables. The minimizer of Eq. (16) is given by [17]: $\omega = \Sigma_{i=1}^{k} t_i x_i$, where t_i is the combination coefficient.

Usually, we first map the input feature vector into a high dimensional feature space $\Phi(x)$, and then learn the regression function:

$$f(x) = \left\langle \sum_{i=1}^{k} t_i \Phi(x_i), \Phi(x) \right\rangle + b$$
$$= \sum_{i=1}^{k} t_i \langle \Phi(x_i), \Phi(x)\rangle + b. \tag{17}$$

The inner product $\langle \Phi(x_i), \Phi(x)\rangle$ can be written as a kernel function $k(x_i, x)$, and Eq. (17) becomes

$$f(x) = \sum_{i=1}^{k} t_i k(x_i, x) + b. \tag{18}$$

Introducing the kernel function $k(x_i, x)$ makes the feature mapping implicit. The radial base function (RBF) $k(x_i, x) = \exp(-\gamma(|x_i - x|)^2)$ is often used as the kernel function, where γ is the precision parameter. More details of SVR can be found in [17].

III. Experimental Results

A. Databases and Evaluation Protocols

The performance of BIQA models can be evaluated by using subjective image databases, where each image has been scored by human observers. (Difference) Mean Opinion Scores (DMOS/MOS) are usually recorded to describe how closely the predicted image quality scores by a BIQA model correlate with human judgments. Several subjective image quality evaluation databases have been established in the IQA community. Here we use the three largest and mostly widely used ones: the LIVE database [41], the CSIQ database [42] and the TID2008 database [43].

The LIVE database consists of 779 distorted images, generated from 29 original images by processing them with 5 types of distortions at various levels. The distortions involved in the LIVE database are: JPEG2000 compression (JP2K), JPEG compression (JPEG), additive white noise (WN), Gaussian blur (GB) and simulated fast fading Rayleigh channel (FF). These distortions reflect a broad range of image impairments, such as edge smoothing, block artifacts, image-dependent distortions, and additive random noise. The CSIQ database consists of 30 original images and their distorted counterparts with six types of distortions at five different distortion levels each. The TID2008 database is composed of 25 reference images and their distorted counterparts with 17 types of distortions at four levels each. For the CSIQ and TID2008 databases, we mainly consider the 4 common types of distortions that appear in the LIVE database, i.e., JP2K, JPEG, WN, and GB.

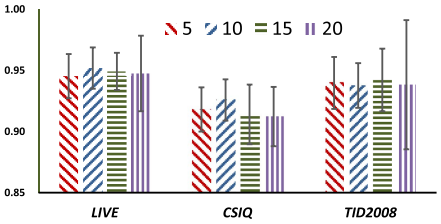


Fig. 8. SRC values of the proposed model M_3 under different numbers of bins $M = N = \{5, 10, 15, 20\}$.

We also use all the 17 distortion types in TID2008 to examine the generalization ability of BIQA models.

To evaluate the performance of a BIQA method, three scores that measure the consistency between the results of a BIQA model and the subjective DMOS/MOS scores are generally used: the Spearman rank order correlation coefficient (SRC), which measures the prediction monotonicity; the Pearson correlation coefficient (PCC) and the root mean squared error (RMSE), which measure the prediction accuracy. Both SRC and PCC lie in the range $[-1, 1]$. A good BIQA model should demonstrate a correlation coefficient with the subjective DMOS/MOS scores as close to 1 (or -1) as possible. The relationship between the subjective scores and the predicted scores may not be linear due to the nonlinear quality rating of human observers. As recommended by the Video Quality Expert Group [44], a nonlinear logistic regression could be built between the predicted scores and the subjective scores when calculating the indices SRC, PCC and RMSE. Denote by Q and Q_p the predicted score before and after regression, respectively. The logistic regression function is defined as follows:

$$Q_p = \beta_1\left(\frac{1}{2} - \frac{1}{\exp(\beta_2(Q - \beta_3))}\right) + \beta_4 Q + \beta_5, \tag{19}$$

where β_1, β_2, β_3, β_4 and β_5 are regression model parameters.

B. Implementation Details

When computing the GM and LOG maps, the scale parameter σ of the filters $\boldsymbol{h}_x$, $\boldsymbol{h}_y$, and $\boldsymbol{h}_{LOG}$ needs to be set. We set σ to a small value 0.5 so that fine image details can be captured in the GM and LOG feature maps. In the JAN process, the weights $\omega(k, l)$ (refer to Eq. (6)) are generated by a Gaussian kernel with scale 2σ.

When computing the joint probability $\boldsymbol{K}_{m,n}$, it is necessary to set the number of bins. In general, using a larger number of bins can lead to more accurate calculation of statistics, but this requires more samples and makes the dimension of the output features higher. For the task of image quality prediction, the goal is to use as few features as possible to achieve as high as possible prediction accuracy. Note that if the feature dimension (i.e., the number of bins) is very high, the regression model learning may become less stable. To investigate the effect of the number of bins on the quality prediction performance, we let $M = N = \{5, 10, 15, 20\}$ and compute the SRC values of the proposed scheme M_3 on the three IQA databases. The results are plotted in Fig. 8. We can see that $M = N = 10$ leads to higher and more stable

TABLE I
OVERALL PERFORMANCE OF THE COMPETING BIQA MODELS ON THE THREE DATABASES. THE RESULTS OF PSNR, SSIM AND FSIM ARE ALSO LISTED FOR REFERENCE

IQA model	LIVE (779 images)			TID2008 (384 images)			CSIQ (600 images)			Weight average	
	SRC	PCC	RMSE	SRC	PCC	RMSE	SRC	PCC	RMSE	SRC	PCC
PSNR	0.8829	0.8821	12.8983	0.8789	0.8611	0.8073	0.9292	0.8562	0.1444	0.8978	0.8687
SSIM	0.9486	0.9464	8.8035	0.9032	0.9087	0.6620	0.9362	0.9347	0.0990	0.9345	0.9342
FSIM	0.9639	0.9612	7.5461	0.9555	0.9539	0.4707	0.9629	0.9675	0.0710	0.9617	0.9617
BIQI [8]	0.8084	0.8250	15.3883	0.8438	0.8704	0.7872	0.7598	0.8353	0.1542	0.7995	0.8384
DIIVINE [16]	0.8816	0.8916	12.3294	0.8930	0.9038	0.6714	0.8697	0.9010	0.1249	0.8800	0.8974
BLIINDS2 [15]	0.9302	0.9366	9.5185	0.8982	0.9219	0.6117	0.9003	0.9282	0.1028	0.9131	0.9305
CORNIA [11]	**0.9466**	**0.9487**	**8.6969**	0.8990	0.9347	0.5669	0.8845	0.9241	0.1054	0.9151	0.9373
BRISQUE [14]	0.9430	0.9468	8.7214	**0.9357**	**0.9391**	**0.5442**	**0.9085**	**0.9356**	**0.0980**	**0.9298**	**0.9414**
M_1	0.9278	0.9329	9.8355	0.9246	0.9332	0.5711	0.9035	0.9298	0.1025	0.9189	0.9319
M_2	**0.9447**	**0.9489**	**8.6452**	**0.9278**	**0.9432**	**0.5263**	**0.9140**	**0.9408**	**0.0947**	**0.9307**	**0.9449**
M_3	**0.9511**	**0.9551**	**8.0444**	**0.9369**	**0.9406**	**0.5377**	**0.9243**	**0.9457**	**0.0909**	**0.9390**	**0.9488**

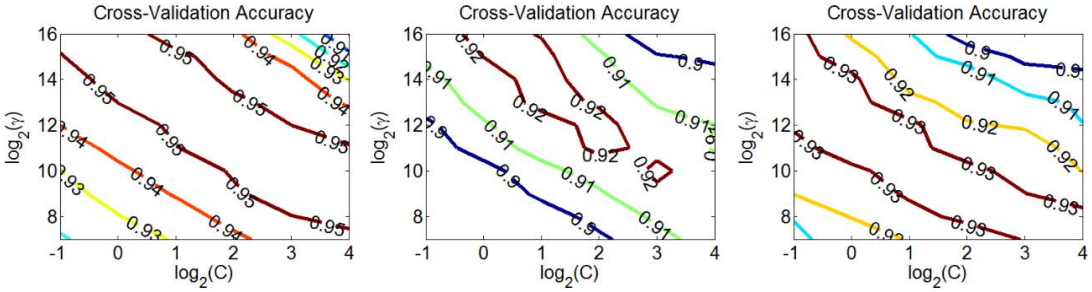


Fig. 9. The SVR parameter (C, γ) selection process for the proposed model M_3 via grid search on databases LIVE (left), CSIQ (middle) and TID2008 (right). The number on the level contour indicates the SRC value of the cross-validation.

performance across all three databases. In all the following experiments, we set $M = 10$ and $N = 10$, and thus all the statistical feature vectors P_G, P_L, Q_G and Q_L are of dimension 10.

When using the SVR to learn the regression models, the SVR parameters (C, γ) need to be set. We conducted a cross validation experiment to choose the values of (C, γ). We partitioned each database into two non-overlapped sets: the training set and the test set. More specifically, we randomly selected 80% of the reference images (and their distorted versions) as the training set and the rest as the test set. In this way, we ensured that there was no content overlap between the training set and the test set. Then the model learned from the training set was examined on the test set. This overall train-test procedure was repeated 1000 times, and the median results were reported for performance evaluation. The (C, γ) values delivering the best median result were chosen as the parameter. Fig. 9 illustrates the parameter selection process of the proposed model M_3 on the three IQA databases. The number on the level contour indicates the median SRC value of the cross-validation. Note that we scaled the subjective scores of CSIQ and TID2008 into the range of the DMOS scores of LIVE. One can see that there exists a ribbon-like region along the diagonal direction where M_3 exhibits the best SRC results on all three databases. This makes the model learning more robust to database. The optimal (C, γ) were found to be (16384, 2), (16384, 2) and (128, 16) on the LIVE, CSIQ and TID2008 databases, respectively. We used them in the following experiments. The MATLAB source code for the proposed methods can be downloaded at http://ipl.xjtu.edu.cn/ftp/xqmou/GM-LOG-BIQA.zip.

C. Performance on Individual Databases

We compared the proposed BIQA models with representative and state-of-the-art BIQA models, including BIQI [8], BLIINDS2 [15], BRISQUE [14], DIIVINE [16] and CORNIA [11]. The source codes for these competing methods were obtained from the original authors. For fair comparison, we also optimized the SVR parameters for these models by grid search. In addition, the classic PSNR, the well-known full reference IQA models Structural SIMilarity (SSIM) [45] and FSIM [31] were also included in the comparison.

We first evaluated the overall performance of the competing BIQA models on each of the three databases. The results are listed in Table I. The top three BIQA models for each index (SRC, PCC or RMSE) are highlighted in bold font. Clearly, the proposed model M_3, which employs both the marginal probabilities and the dependencies of GM and LOG features, performs the best consistently on all the databases. The proposed models M_2 and M_1 also perform very well, while M_2, which employs the dependency between GM and LOG, works slightly better than M_1, which employs only the marginal distribution of GM and LOG. This implies that the dependency statistics contain more useful information than the marginal statistics for the task of BIQA. The advantages

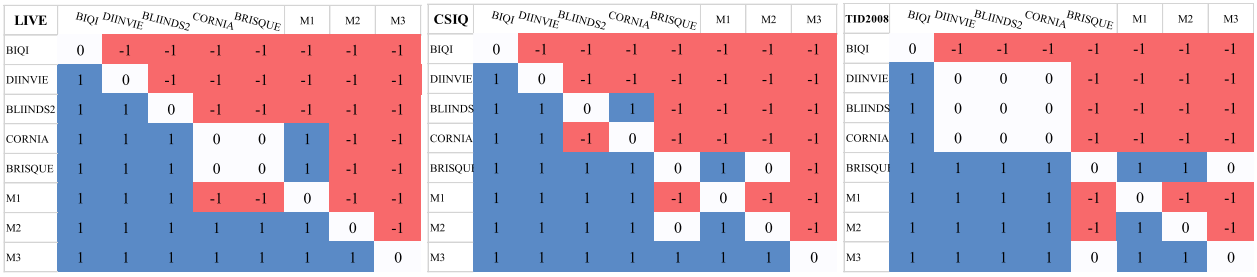


Fig. 10. Results of one sided t-test conducted by using the SRC values of competing BIQA models. A value of '1' indicates that the row model is statistically better than the column model, a value of '−1' indicates that the column model is statistically better, while a value of '0' indicates that they are statistically similar in performance.

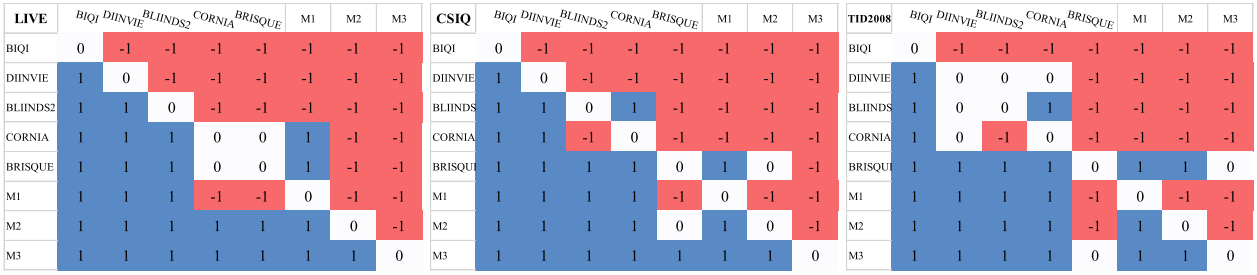


Fig. 11. Results of the Wilcoxon rank-sum test by using the SRC values of competing BIQA models. A value of '1' indicates that the row model is statistically better than the column model; a value of '−1' indicates that the column model is statistically better; and a value of '0' indicates that the two models are statistically similar in performance.

TABLE II
RANKING OF THE BIQA MODELS IN TERMS OF SRC

	LIVE	TID2008	CSIQ	weighted average	length
BIQI [8]	8	8	8	8	18
DIIVINE [16]	7	7	7	7	88
BLIINDS2 [15]	5	6	5	6	24
CORNIA [11]	**2**	5	6	5	20000
M_1	6	**3**	4	4	20
BRISQUE [14]	4	4	**3**	**3**	36
M_2	**3**	**2**	**2**	**2**	20
M_3	**1**	**1**	**1**	**1**	40

of the proposed models M_3 and M_2 over the other BIQA models (except for BRISQUE) are significant on the TID2008 and CSIQ databases. When compared with full reference IQA models, the proposed model M_3 shows clear advantages over PSNR and SSIM, and only lags the performance of FSIM a little. This is reasonable because FSIM is a state-of-the-art full reference IQA method which employs the reference image as input. In the two right most columns of Table I, we show the weighted average SRC and PCC scores of competing IQA methods over the three databases (the weights are based on the numbers of images in the three databases). (Note that the weighted average of RMSE scores cannot be computed since the subjective scores scale differently in the three databases.) We see that M_3 still performs the best among the BIQA methods in terms of weighted average SRC and PCC scores, followed by M_2 and BRISQUE.

In Table II, we rank the competing BIQA models in terms of SRC on each database. The weighted ranking on the three databases is also given. The proposed models M_3, M_2, and M_1 rank the first, the second and the fourth, respectively, demonstrating that the joint statistics of GM and LOG features are powerful predictors of natural image perceptual quality.

To determine whether the advantages of the proposed method over competing methods are statistically significant, two hypotheses tests were conducted: the one-sided t-test and the Wilcoxon rank-sum test [46]. The one-sided t-test tests the equivalence of the mean values of two samples drawn from independent populations of a normal distribution. The Wilcoxon rank-sum test tests the equivalence of the median values of two independent samples and is a nonparametric alternative to the two sample t-test. Both of the tests were performed at a significance level of 0.01 using the 1000 SRC values of all pairs of BIQA models. The null hypothesis is that the SRC values of the pair of models are drawn from populations with equal mean (t-test) or equal median (Wilcoxon rank-sum test). The alternative hypothesis is that the mean/median of one model is greater than the other. Notice that for the t-test, two assumptions must be guaranteed: independency and normality. The random split into training and testing sets ensured the independency of these values. The SRC values follow a right-skewed unimodal distribution. To ensure the normality assumption, the SRC values were firstly transformed by exponentiation [46], and then the t-test was applied to these transformed SRC values.

The results of the two tests are illustrated in Fig. 10 and Fig. 11, respectively. A value of '1' indicates that the row model is statistically better than the column model, a value of '−1' indicates that the column model is statistically better, while a value of '0' indicates that the two models have no statistical difference in performance. The two tests lead to nearly the same conclusions except for the pair of CORNIA and BLIINDS2 on TID2008 database.

TABLE III
PERFORMANCE (SRC) OF COMPETING BIQA MODELS ON INDIVIDUAL DISTORTION TYPES

		BIQI [8]	BLIINDS2 [15]	BRISQUE [14]	DIIVINE [16]	CORNIA [11]	M_1	M_2	M_3
LIVE	JP2K	0.7849	**0.9258**	0.9175	0.8418	**0.9271**	0.9127	0.9256	**0.9283**
	JPEG	0.8801	0.9500	**0.9655**	0.8926	0.9437	0.9550	**0.9589**	**0.9659**
	WN	0.9157	0.9477	**0.9789**	0.9617	0.9608	**0.9835**	0.9778	**0.9853**
	GB	0.8367	0.9132	**0.9479**	0.8792	**0.9553**	0.9150	0.9014	**0.9395**
	FF	0.7023	0.8736	0.8854	0.8202	**0.9103**	0.8556	**0.8919**	**0.9008**
CSIQ	WN	0.6000	0.8863	**0.9310**	0.8131	0.7980	**0.9444**	0.9261	**0.9406**
	JPEG	0.8384	0.9115	**0.9253**	0.8843	0.8845	**0.9186**	0.9127	**0.9328**
	JP2K	0.7573	0.8870	**0.8934**	0.8692	0.8950	0.8736	**0.9192**	**0.9172**
	GB	0.8160	**0.9152**	**0.9143**	0.8756	0.9006	0.8999	0.9052	**0.9070**
TID2008	WN	0.5368	0.7314	**0.8603**	0.7130	0.5941	0.8541	**0.9156**	**0.9068**
	GB	0.8878	**0.9176**	**0.9059**	0.8824	**0.8941**	0.8692	0.8451	0.8812
	JPEG	0.8996	0.8853	0.9103	0.9033	0.9099	**0.9379**	**0.9263**	**0.9338**
	JP2K	0.8147	0.9118	0.9044	0.9103	**0.9290**	**0.9402**	**0.9308**	0.9263
Hit-count		0	3	**9**	0	5	5	**6**	**11**
Mean		0.7900	0.8967	**0.9185**	0.8651	0.8848	0.9123	**0.9182**	**0.9281**
STD		0.116	0.055	**0.032**	0.060	0.096	0.0403	**0.0317**	**0.0274**

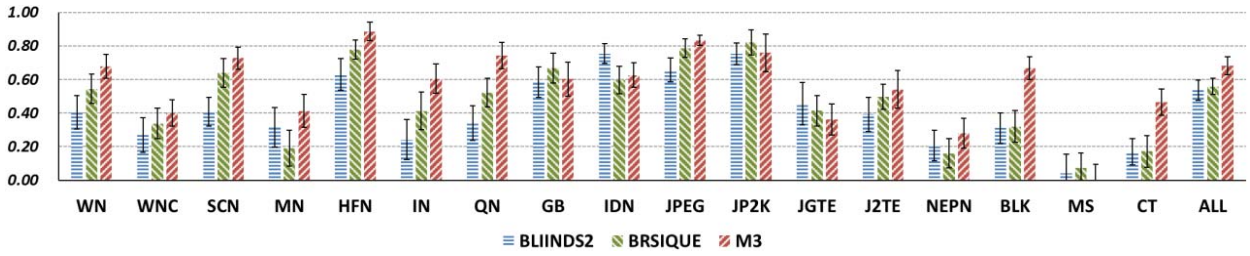


Fig. 12. Median SRC values of 1000 runs on TID2008 database for BLIINDS2, BRISQUE and the proposed model M_3. "ALL" means the overall performance on the entire database.

On LIVE and CSIQ, M_3 is statistically superior to all the other models. On TID2008, M_3 and BRISQUE have no statistically significant difference, and both of them beat the other methods. Overall, the proposed M_1, M_2, M_3, as well as the BRISQUE methods always perform better than the other methods. Interestingly, they all use isotropic IQA features. The proposed models use LOG and GM, while BRISQUE uses the MSCN feature which is a simplified version of the LOG response. However, the proposed models exploit the dependency between the LOG and GM responses, while BRISQUE utilizes the correlations between spatially neighboring coefficients of the MSCN signal.

D. Performance on Individual Distortion Type and Generalization Ability

We then tested the performance of the proposed BIQA models on each type of distortion. The results are listed in Table III. For brevity, we only present the SRC results. Similar conclusions were arrived at for PCC and RMSE. The top 3 models are highlighted in boldface. At the bottom of the table, the hit-count (i.e., the number of times ranked in the top 3 on each distortion), the average performance (Mean), and the performance standard deviation (STD) of each model are listed. For the 13 groups of distortion types in the three databases, the proposed model M_3 had the highest hit-count (i.e., 11 times), followed by BRISQUE (9 times), M_2 (6 times) and M_1 (5 times). Meanwhile, the proposed models have very high mean scores, and M_3 has the smallest STD across all the distortion groups.

To test the generalization ability of the proposed model with respect to distortion types, we conducted further experiments on the entire TID2008 database including all 17 distortions. Since the proposed M_3 makes use of GM and LOG features in the luminance channel, it is not responsive to non-structural distortions and distortions on chromatic components. Therefore, we also conducted experiments on a subset of TID2008 which includes only achromatic-structural distortions (11 distortions included). More specifically, two color aberration distortions (WNC, JGTE) and four non-structural distortions (NEPN, BLK, MS, and CT) were removed from the subset. (For more information about the distortions types, please refer to [43]). The same train-test procedure as in previously experiments was employed. Two representative methods, BRISQUE and BLIINDS2, were used for comparison.

The experimental results are illustrated in Figs. 12 and 13, respectively. When all 17 distortion types were involved, M_3 delivered better SRC values than its competitors on 12 distortions, and showed clear advantage over the competitors in terms of overall performance. However, M_3 failed to deliver good SRC values on chromatic distortions (i.e., WNC and JGTE), while its competitors also faired poorly. Low performance of M_3 and its competitors can also be observed on the non-structural distortions (NEPN, MS, CT,

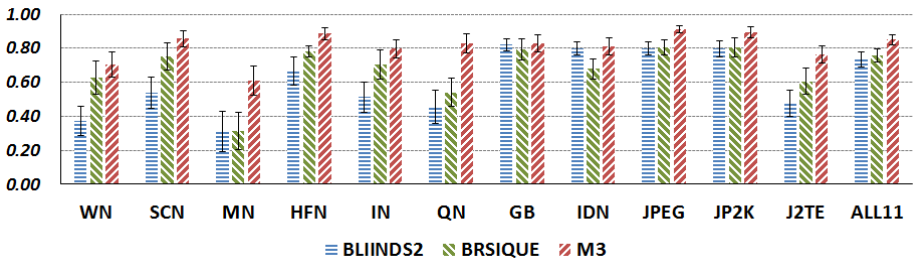


Fig. 13. Median SRC values of 1000 runs on the subset of achromatic-structural distortions (11 types in total) of TID2008 database for BLIINDS2, BRISQUE and the proposed model M_3. "ALL11" means the overall performance on this subset.

TABLE IV
PERFORMANCE (SRC) OF THE BIQA MODELS ACROSS THE THREE DATABASES

Database for Training	*Database for Testing*	*BIQI*	*BLIINDS2*	*BRISQUE*	*DIIVINE*	*CORNIA*	*M_3*
LIVE	*CSIQ*	0.7805	0.8878	0.8993	0.8571	0.8973	**0.9108**
LIVE	*TID2008*	0.8194	0.9056	0.9050	0.8599	0.8932	**0.9204**
CSIQ	*LIVE*	0.4538	0.9365	0.9311	0.8475	0.9279	**0.9459**
CSIQ	*TID2008*	0.6977	0.8005	0.8986	0.8223	0.8704	**0.9051**
TID2008	*LIVE*	0.7631	**0.9389**	0.9288	0.8658	0.9091	0.9336
TID2008	*CSIQ*	0.8009	**0.8747**	0.8665	0.8481	0.8381	0.8393

TABLE V
COMPUTATIONAL COMPLEXITY AND AVERAGE RUN-TIME (SECOND). N IS THE TOTAL NUMBER OF PIXELS IN A TEST IMAGE

	Run-time (s)	Computational complexity	Notes
BLIINDS2	123.9	$O((N/(d^2))log(N/(d^2)))$	d: block size
DIIVINE	28.2	$O(N(log(N)+m^2+N+392b))$	m: neighborhood size in DNT; b: # of bins in the 2-D histogram
CORNIA	3.246	$O(Nd^2k)$	d: block size; K: codebook size
BRISQUE	0.176	$O(d^2N)$	d: block size
M_3	0.101	$O(N(h+k))$	h: size of the derivative filter; k: size of the joint probability matrix
BIQI	0.076	$O(N)$	

etc.). On the achromatic-structural subset with 11 distortions, M_3 showed very good performance where it outperformed the two competitors on all distortion types. In terms of overall performance, M_3 achieved an SRC value of 0.8501, while BLIINDS2 and BRISQUE achieved only 0.7337 and 0.7551, respectively. M_3 demonstrated much higher generalization ability for achromatic-structural distortions than the state-of-the-art BLIINDS2 and BRISQUE models.

E. Experiments on Database Dependency

In the experiments in Sections III.C and III.D, the training samples and test samples were drawn from the same database. It is expected that the BIQA model learned from one database should be applicable to images in other databases. Therefore, to demonstrate the generality and robustness of a BIQA algorithm, it is necessary to see if satisfying results can still be obtained by applying the BIQA model learned from one database to another database. In this subsection, we conducted the following experiments. First, a BIQA model was trained on database A, then the learned model was tested on database B; next, the same BIQA model was trained on database B, then tested on database A. With the three databases, there were 6 combinations of training and test database pairs. The SRC index was used for evaluation, and the results are presented in Table IV. Again, the proposed model M_3 performed very well. It achieved the best SRC scores in 4 out of the 6 tests, and its results were very close to the best results in the other 2 tests.

F. Computational Complexity

In many practical applications it is desired to estimate the quality of an input image online. Therefore, the computational complexity is also an important factor when evaluating a BIQA model. Table V summarizes the computational complexities and the run time (the average processing time of an image from the LIVE database using the MATLAB7.10.0 programming environment) of all the competing BIQA models. One can see that BIQI is the fastest one with complexity $O(N)$, where N is the total number of image pixels. However, its performance is the worst among all the competing models. The proposed M_3 is the second fastest method. The main costs in M_3 are spent on computing the GM and LOG maps and the JAN of them, whose complexity is $O(Nh)$, and the computation of the joint probability matrix, whose complexity is $O(Nk)$, where h is the size of filters $\boldsymbol{h}_x$, $\boldsymbol{h}_y$, and $\boldsymbol{h}_{LOG}$, and k is the size of the joint probability

matrix. Hence, the overall complexity of M_3 is $O(N(h+k))$. On average, M_3 can process 10 images from the LIVE database per second using our MATLAB implementation.

IV. Conclusion

Existing BIQA models typically decompose an image into different frequency and orientation bands, and then extract statistical features from the decomposed coefficients to learn a quality prediction model. However, few BIQA models explicitly exploit simple image contrast features such as the gradient magnitude (GM) and Laplacian of Gaussian (LOG) responses, although LOG responses share similarities to human receptive field responses. Here we made the first attempt to use GM and LOG features to conduct high performance BIQA. To alleviate the effects of image content variations, we applied a joint adaptive normalization procedure to normalize the GM and LOG features and whiten the image data. Since GM and LOG features are not independent and the interaction between them can reflect local quality prediction on natural images, we proposed a simple index, called independency distribution, to measure the joint statistics of them. The proposed BIQA models employ the marginal distributions and the independency distributions of GM and LOG, and they lead to highly competitive performance with many state-of-the-art BIQA methods in terms of quality prediction accuracy, generalization ability, robustness (i.e., across-database prediction capability) and computational complexity. Encouraged by the state-of-the-art BIQA results obtained in this paper, in future work we will investigate how to use the GM and LOG features for blind local quality map estimation, which is a very useful yet very challenging research problem.

References

[1] A. C. Bovik, "Automatic prediction of perceptual image and video quality," *Proc. IEEE*, vol. 101, no. 9, pp. 2008–2024, Sep. 2013.

[2] Z. Wang, "Applications of objective image quality assessment methods," *IEEE Signal Process. Mag.*, vol. 28, no. 6, pp. 137–142, Nov. 2011.

[3] X. Marichal, W.-Y. Ma, and H. Zhang, "Blur determination in the compressed domain using DCT information," in *Proc. ICIP*, Oct. 1999, pp. 386–390.

[4] P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi, "A no-reference perceptual blur metric," in *Proc. ICIP*, 2002, pp. III-57–III-60.

[5] Y. Horita, S. Arata, and T. Murai, "No-reference image quality assessment for JPEG/JPEG2000 coding," in *Proc. EUSIPCO*, 2004, pp. 1301–1304.

[6] D. M. Chandler, "Seven challenges in image quality assessment: Past, present, and future research," *ISRN Signal Process.*, vol. 2013, Nov. 2013, Art. ID 905685.

[7] A. Mitttal, R. Soundararajan, and A. C. Bovik, "Making a 'completely blind' image quality analyzer," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 209–212, Mar. 2013.

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

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

[10] H. Tang, N. Joshi, and A. Kapoor, "Learning a blind measure of perceptual image quality," in *Proc. IEEE CVPR*, Jun. 2011, pp. 305–312.

[11] P. Ye, J. Kumar, L. Kang, and D. Doermann, "Unsupervised feature learning framework for no-reference image quality assessment," in *Proc. IEEE CVPR*, Jun. 2012, pp. 1098–1105.

[12] L. He, D. Tao, X. Li, and X. Gao, "Sparse representation for blind image quality assessment," in *Proc. IEEE CVPR*, Jun. 2012, pp. 1146–1153.

[13] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, Feb. 2009.

[14] 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.

[15] 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.

[16] 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.

[17] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statist. Comput.*, vol. 14, no. 3, pp. 199–222, 2004.

[18] P. Ye and D. Doermann, "No-reference image quality assessment using visual codebooks," *IEEE Trans. Image Process.*, vol. 21, no. 7, pp. 3129–3138, Jul. 2012.

[19] P. J. B. Hancock, R. J. Baddeley, and L. S. Smith, "The principal components of natural images," *Netw., Comput. Neural Syst.*, vol. 3, no. 1, pp. 61–70, 1992.

[20] A. J. Bell and T. J. Sejnowski, "The 'independent components' of natural scenes are edge filters," *Vis. Res.*, vol. 37, no. 23, pp. 3327–3338, 1997.

[21] D. J. Field and B. A. Olshausen, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, vol. 381, no. 6583 pp. 607–609, 1996.

[22] D. Marr and E. Hildreth, "Theory of edge detection," *Proc. Roy. Soc. London B, Biol. Sci.*, vol. 207, no. 1167, pp. 187–217, 1980.

[23] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-8, no. 6, pp. 679–698, Nov. 1986.

[24] S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 7, pp. 674–693, Jul. 1989.

[25] S. A. Khayam, "The discrete cosine transform (DCT): Theory and application," Dept. Elect. Comput. Eng., Michigan State Univ., Lansing, MI, USA, Tech. Rep. ECE 802-602, 2003.

[26] C. A. Bouman, K. Sauer, and S. Saquib, "Markov random fields and stochastic image models," in *Proc. ICIP*, Oct. 1995, pp. 1–103.

[27] T. F. Chan, J. Shen, and L. Vese, "Variational PDE models in image processing," *Notices Amer. Math. Soc.*, vol. 50, no. 1, pp. 14–26, 2003.

[28] M. Zhang, X. Mou, and D. Zhang, "Non-shift edge based ratio (NSER): An image quality assessment metric based on early vision features," *IEEE Signal Process. Lett.*, vol. 18, no. 5, pp. 315–318, May 2011.

[29] W. Xue and X. Mou, "An image quality assessment metric based on non-shift edge," in *Proc. ICIP*, Sep. 2011, pp. 3309–3312.

[30] G.-H. Chen, C.-L. Yang, and S.-L. Xie, "Gradient-based structural similarity for image quality assessment," in *Proc. ICIP*, Oct. 2006, pp. 2929–2932.

[31] 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.

[32] W. Xue and X. Mou, "Reduced reference image quality assessment based on Weibull statistics," in *Proc. IEEE QoMEX*, Jun. 2010, pp. 1–6.

[33] W. Xue, L. Zhang, X. Mou, and A. C. Bovik, "Gradient magnitude similarity deviation: A highly efficient perceptual image quality index," *IEEE Trans. Image Process.*, vol. 23, no. 2, pp. 684–695, Feb. 2014.

[34] D. L. Ruderman, "The statistics of natural images," *Netw., Comput. Neural Syst.*, vol. 5, no. 4, pp. 517–548, 1994.

[35] D. H. Hubel and T. N. Wiesel, "Brain mechanisms of vision," *Sci. Amer.*, vol. 241, no. 3, pp. 1–150, 1979.

[36] J.-M. Geusebroek and A. W. M. Smeulders, "A six-stimulus theory for stochastic texture," *Int. J. Comput. Vis.*, vol. 62, nos. 1–2, pp. 7–16, 2005.

[37] J. Huang and D. Mumford, "Statistics of natural images and models," in *Proc. IEEE CVPR*, Jun. 1999, pp. 541–547.

[38] M. J. Wainwright, O. Schwartz, and E. P. Simoncelli, "Natural image statistics and divisive normalization," in *Probabilistic Models of the Brain: Perception and Neural Function*. Cambridge, MA, USA: MIT Press, 2002, pp. 203–222.

[39] S. Lyu and E. P. Simoncelli, "Nonlinear image representation using divisive normalization," in *Proc. IEEE CVPR*, Jun. 2008, pp. 1–8.

[40] D. J. Heeger, "Normalization of cell responses in cat striate cortex," *Vis. Neurosci.*, vol. 9, no. 2, pp. 181–197, 1992.

[41] H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik. *Live Image Quality Assessment Database Release 2*. [Online]. Available: http://live.ece.utexas.edu/research/quality, accessed Sep. 1, 2011.

[42] E. C. Larson and D. M. Chandler, "Most apparent distortion: Full-reference image quality assessment and the role of strategy," *J. Electron. Imag.*, vol. 19, no. 1, p. 011006, Jan. 2010.

[43] N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, M. Carli, and F. Battisti, "TID2008—A database for evaluation of full-reference visual quality assessment metrics," *Adv. Modern Radio Electron.*, vol. 10, no. 4, pp. 30–45, 2009.

[44] (Aug. 2003). *VQEG, Final Report From the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment—Phase II*. [Online]. Available: http://www.vqeg.org/

[45] 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.

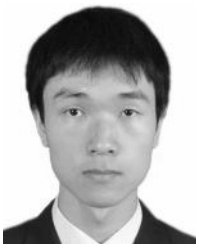
[46] D. J. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures*. London, U.K.: Chapman & Hall, 2003.

[47] C. Harris and M. Stephens, "A combined corner and edge detector," in *Proc. Alvey Vis. Conf.*, 1988, pp. 147–151.

[48] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.

[49] H. Bay, T. Tuytelaars, and L. Van Gool, "SURF: Speeded up robust features," in *Proc. ECCV*, 2006, pp. 404–417.

[50] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. CVPR*, Jun. 2005, pp. 886–893.



Wufeng Xue received the B.Sc. degree in automatic engineering from the School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an, China, in 2009, where he is currently pursuing the Ph.D. degree with the Institute of Image Processing and Pattern Recognition. His research interest focuses on perceptual quality of visual signals.

Xuanqin Mou (M'08) has been with the Institute of Image Processing and Pattern Recognition (IPPR), Electronic and Information Engineering School, Xi'an Jiaotong University, Xi'an, China, since 1987, where he has been an Associate Professor since 1997, and a Professor since 2002. He is currently the Director of IPPR. He served as a member of the 12th Expert Evaluation Committee for the National Natural Science Foundation of China, and the 5th and 6th Executive Committee of the China Society of Image and Graphics, and the Vice President of the Shaanxi Image and Graphics Association. He has authored or co-authored over 200 peer-reviewed journal or conference papers. He was a recipient of the Yung Wing Award for Excellence in Education, the KC Wong Education Award, the Technology Academy Award for Invention by the Ministry of Education of China, and the Technology Academy Awards from the Government of Shaanxi Province, China.

Lei Zhang (M'04–SM'14) received the B.Sc. degree from the Shenyang Institute of Aeronautical Engineering, Shenyang, China, in 1995, the M.Sc. and Ph.D. degrees in control theory and engineering from Northwestern Polytechnical University, Xi'an, China, in 1998 and 2001, respectively. From 2001 to 2002, he was a Research Associate with the Department of Computing, Hong Kong Polytechnic University, Hong Kong. From 2003 to 2006, he was a Post-Doctoral Fellow with the Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON, Canada. In 2006, he joined the Department of Computing, Hong Kong Polytechnic University, as an Assistant Professor, where he has been an Associate Professor with the Department of Computing since 2010. His research interests include image and video processing, computer vision, pattern recognition, and biometrics. He has authored about 200 papers in those areas. He is currently an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY and *Image and Vision Computing*. He was a recipient of the 2012–13 Faculty Award in Research and Scholarly Activities.

Alan C. Bovik (S'80–M'81–SM'89–F'96) currently holds a position of the Ernest J. Cockrell Endowed Chair in Engineering with the University of Texas at Austin, Austin, TX, USA, where he is also the Director of the Laboratory for Image and Video Engineering. He is a faculty member with the Department of Electrical and Computer Engineering and the Center for Perceptual Systems, Institute for Neuroscience. His research interests include image and video processing, computational vision, and visual perception. He has authored over 700 technical articles in these areas, and holds several U.S. patents. His publications have been cited over 35 000 times in the literature, his current H-index is over 70, and he is listed as a Highly-Cited Researcher by Thompson Reuters. His several books include the companion volumes entitled *The Essential Guides to Image and Video Processing* (Academic Press, 2009).

He was a recipient of 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 Distinguished Lecturer Award (2000), the Meritorious Service Award (1998), and (co-author) the Young Author Best Paper Award (2013). He was also a recipient of the Honorary Member Award of the Society for Imaging Science and Technology in 2013, the SPIE Technology Achievement Award in 2012, and the IS&T/SPIE Imaging Scientist of the Year Award in 2011. He is a fellow of the Optical Society of America and the Society of Photo-Optical and Instrumentation Engineers. He was the Editor-in-Chief of the IEEE TRANSACTIONS ON IMAGE PROCESSING from 1996 to 2002, and the Founding General Chair of the First IEEE International Conference on Image Processing, held in Austin, TX, USA, in 1994.

Dr. Bovik is a registered Professional Engineer in the State of Texas and a frequent Consultant to legal, industrial, and academic institutions.

Xiangchu Feng received the B.S. degree in computational mathematics from Xi'an Jiaotong University, Xi'an, China, in 1984, and the M.S. and Ph.D. degrees in applied mathematics from Xidian University, Xi'an, in 1989 and 1999, respectively. He is currently a Professor of Applied Mathematics with Xidian University. His research interests include numerical analysis, wavelets, and partial differential equations for image processing.