

Blind Quality Index for Multiply Distorted Images Using Biorder Structure Degradation and Nonlocal Statistics

Yu Zhou , Leida Li , Member, IEEE, Jinjian Wu , Ke Gu , Weisheng Dong , and Guangming Shi , Senior Member, IEEE

Abstract—In the past decade, extensive image quality metrics have been proposed. The majority of them are tailored for the images that contain a specific type of distortion. However, in practice, the images are usually degraded by different types of distortions simultaneously. This poses great challenges to the existing quality metrics. Motivated by this, this paper proposes a no-reference quality index for the multiply distorted images using the biorder structure degradation and the nonlocal statistics. The design philosophy is inspired by the fact that the human visual system (HVS) is highly sensitive to the degradations of both the spatial contrast and the spatial distribution, which are prone to be changed by the joint effects of the multiple distortions. Specifically, the multiresolution representation of the image is first built by downsampling to simulate the hierarchical property of the HVS. Then, the structure degradation is calculated to measure the spatial contrast. Considering the fact that the human visual cortex has the separate mechanisms to perceive the first- and second-order structures, dubbed biorder structures, the degradations of biorder structures are calculated to account for the spatial contrast, producing the first group of the quality-aware features. Furthermore, the nonlocal self-similarity statistics is calculated to measure the spatial distribution, producing the second group of features. Finally, all the features are fed into the random forest regression model to learn the quality model for the multiply distorted images. Extensive experimental results conducted on the three public databases demonstrate the superiority of the proposed metric to the state-of-the-art metrics. Moreover, the proposed metric is also advantageous over the existing metrics in terms of the generalization ability.

Manuscript received December 10, 2017; revised March 10, 2018; accepted April 9, 2018. Date of publication April 26, 2018; date of current version October 15, 2018. This work was supported in part by the National Natural Science Foundation of China under Grant 61771473 and Grant 61379143, in part by the Six Talent Peaks High-level Talents in Jiangsu Province under Grant XYDXX-063, and in part by the Qing Lan Project. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Pablo Cesar. (*Corresponding author: Leida Li*)

Y. Zhou and L. Li are with the School of Information and Control Engineering, China University of Mining and Technology, Xuzhou 221116, China (e-mail: zhouy7476@cumt.edu.cn; lileida@cumt.edu.cn).

J. Wu, W. Dong, and G. Shi are with the School of Artificial Intelligence, Xidian University, Xi'an 710071, China (e-mail: jinjian.wu@mail.xidian.edu.cn; wsdong@mail.xidian.edu.cn; gmshi@xidian.edu.cn).

K. Gu is with the Beijing University of Technology (BJUT) Faculty of Information Technology, BJUT, Beijing 100124, China (e-mail: guke.doctor@gmail.com).

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/TMM.2018.2829607

Index Terms—Quality evaluation, multiply distorted images, spatial contrast, spatial distribution, biorder structures, nonlocal statistics.

I. INTRODUCTION

IMAGE quality assessment (IQA) has wide applications in image processing and many other applications [1]–[3], such as remote sensing and public monitoring systems [4]. Objective IQA is to design mathematic models for measuring the distortions in images [5]. The evaluation results of objective IQA are expected to be consistent with the subjective results evaluated by human eyes, as they are the ultimate receiver of images. So far, a large number of IQA metrics have sprung up, which can be divided into three types, namely full-reference (FR), reduced-reference (RR), and no-reference (NR) [2], [6], [7]. Among them, FR and RR metrics rely on the whole and partial information of the reference images for quality assessment, while NR metrics operate without any reference information. By contrast, NR metrics are more practical and are urgently required, as reference images are commonly inaccessible in practice [8].

The majority of the existing NR metrics are designed for images that are distorted by a specific type of distortions, such as blur [9], [10], and JPEG compression [11]–[16], etc. However, in practice, images are commonly subject to the processes of acquisition, compression, transmission, and storage, which easily introduce multiple distortions [17], such as, noise and blurring during acquisition, compression artifacts during compression, and transmission errors. To evaluate the multiple distortions, a natural way is to first evaluate individual distortions separately, and then integrate all distortion scores to generate the overall quality score. Unfortunately, this approach not only needs prior knowledge of the distortion types, which is usually unavailable in practice, but also ignores the joint effects of the multiple distortions. Therefore, quality metrics for multiply distorted images are in urgent need.

Motivated by this fact, we propose a blind multiply distorted image quality metric by calculating Bi-Order Structure degradation and nonlocal Statistics of images (BOSS). The proposed method is inspired by the following facts: (1) The human visual system (HVS) is adapted to perceiving image information from both spatial contrast and spatial distribution [18]. Among them, spatial contrast represents the gray level variation between the objects and the background [19], while spatial distribution

characterizes the relationship of the pixels in the neighborhood. (2) The multiple distortions have joint influence on both spatial contrast and distribution. Particularly, in this work, structure degradation and nonlocal statistics are calculated to measure the spatial contrast and spatial distribution, respectively, generating two groups of quality-aware features. More specifically, bi-order structural degradation, including first- and second-order structural degradation, is computed to measure the spatial contrast. The first-order structures represent the dominant image structures (*e.g.*, edges), while the second-order structures represent the minor structures (*e.g.*, textures) [20]. The strategy of integrating the bi-order structures is motivated by the facts that 1) the human visual cortex has been demonstrated by neurophysiology and biological vision to have the separate mechanisms to perceive the first- and second-order structures [20]; 2) textures, i.e., second-order structures, are also vital for image understanding [21], and the co-existence of multiple distortions can cause the degradations of bi-order structures. Moreover, since the HVS possesses the hierarchical property when perceiving natural scenes [22], [23], the input image is down-sampled by several times to generate the multi-resolution representation for feature extraction. Finally, the random forest (RF) regression model is employed to train the quality model [24], [25]. The contributions of the proposed method are multi-fold. First, the proposed metric simultaneously measures the structure degradation and nonlocal statistics, and the structure degradation is measured on bi-order structures. This multi-aspect evaluation strategy can better adapt to the complexity of the characteristics of multiple distortions. Second, the proposed method measures both spatial contrast and spatial distribution, which can better simulate the characteristics of the HVS.

The remainder of this paper is organized as follows: Section II reviews the related work and highlights the advantages of the proposed method. Section III gives the details of the proposed method. Then extensive experimental results and analysis are given in Section IV. Finally, Section V concludes the paper.

II. RELATED WORK

Recently, several approaches have been proposed for the quality evaluation of images with multiple distortions. In [26], a five-step quality metric was presented. For an input image, a noise estimation algorithm was first used to judge whether the image contained noise or not. If the image was noised, it was denoised first and then two evaluation metrics for sharpness and blockiness were severally utilized to measure the blur and JPEG distortions, producing the sharpness and blockiness scores. Otherwise, blur and blockiness were directly evaluated on the original image without denoising. Finally, the two scores were combined to generate the overall quality score. In [27], the authors further improved the five-step metric by further calculating the joint effects of blur and JPEG distortions, producing the six-step quality metric. In [28], a NR quality metric for multiple distortions was presented. First, the features that were sensitive to each type of distortion were first selected from three sets of natural scene statistic features, which were then encoded using a bag-of-words model. Finally, all processed features were integrated into a quality score. In [29], Li *et al.* proposed a metric for

multiply distorted images based on structure degradation. The image was first filtered using a low-pass function and down-sampled to generate multi-resolution images. The gradient map was further calculated and the Local Binary Pattern (LBP) was computed on the gradient map for structure extraction, generating the feature map. Then the gradient magnitude map was utilized as its weighting map. The histogram of the gradient-weighted map was then calculated, which was then summarized to produce the final quality score. Similarly, in [30], a method based on structure degradation was proposed. For a distorted image, the color Gaussian jet was first used for structure description. Then, the local binary pattern was used to measure the distortions.

Although the existing metrics have achieved great advances in the quality evaluation of multiply distorted images, the problem is still far from being solved. For the methods [26], [27], only three kinds of distortions were measured, namely noise, blur and JPEG distortions. However, as aforementioned, the distortions in practical images are much more diversified. Therefore, the effectiveness of these two metrics on evaluating the quality of multiply distorted images is limited. For the method [28], the distortion types in the test image have to be known in advance, which is infeasible in practice. For the method [29], only the degradation of dominant structures, which are defined as first-order structures [20], is computed as the quality feature. It ignores the impact of minor structures, i.e., second-order structures, on quality perception, which restricts the effectiveness of this method. For the method [30], the structure degradation, i.e., spatial contrast is calculated. This method ignores the fact that the HVS is also sensitive to spatial distribution [18], which also influences the image perceptual quality. By contrast, the proposed method has several advantages. First, our metric can be used to evaluate the quality of multiply distorted images without any prior knowledge of distortions, such as distortion type or distortion level. Moreover, most existing metrics only calculate the first-order structural degradation, which ignores the impact of the second-order structure. By contrast, the proposed method computes the bi-order structure degradations, which can better adapt to the complexity of the characteristics of multiple distortions. In addition, the proposed method can better characterize the HVS in terms of both spatial contrast and spatial distribution, which is highly needed in image quality evaluation.

III. PROPOSED QUALITY METRIC

In this work, the bi-order structure degradation is computed for spatial contrast measurement, while nonlocal statistics is utilized to measure spatial distribution, producing two groups of quality-aware features for multiply distorted images. Furthermore, it has been demonstrated that the HVS exhibits the hierarchical characteristic in nature, as it perceives visual information from dominant structures to minor details [22], [23]. Therefore, down-sampling conduction is used to mimic this characteristic for feature extraction by generating images with different resolutions. Finally, to map the features to an objective score, the RF is utilized to train the quality model. The flowchart of the proposed metric is shown in Fig. 1.

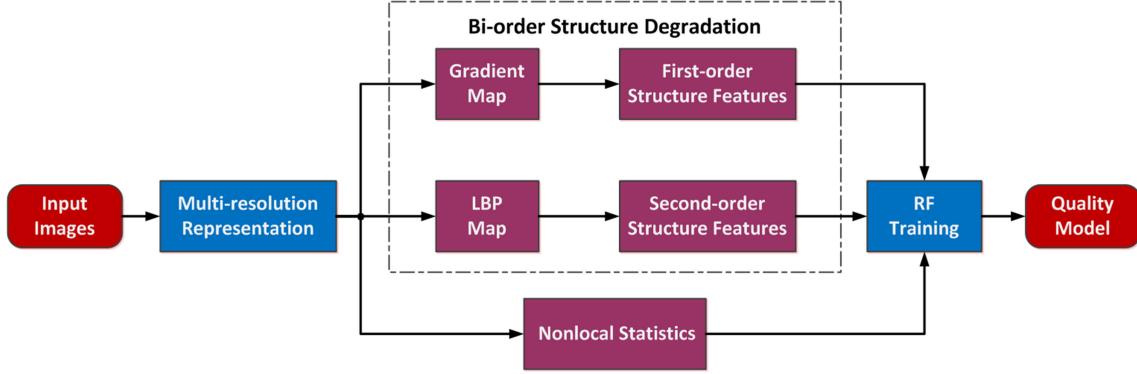


Fig. 1. Flowchart of the proposed quality prediction model.

A. Multiresolution Representation

For an input image \mathbf{I} with size $[M, N]$, to calculate the multi-resolution representation, it is down-sampled by t times. Thus, $t + 1$ images with different resolutions are obtained, which are denoted by $\mathbf{I}_1, \mathbf{I}_2, \dots, \mathbf{I}_{t+1}$, where \mathbf{I}_1 denotes the original image, while \mathbf{I}_{t+1} is the image with the lowest resolution. In this work, image \mathbf{I} is down-sampled by four times, and the scaling factor is set to 0.5.

B. Biorder Structure Degradation

The HVS has been demonstrated to be sensitive to structure degradation [29], [31]. Besides, as stated in [20], structures in an image can be classified into first- and second-order patterns. Among them, first-order structures represent the dominant structures (*e.g.*, edges), while the second-order ones denote minor structures (*e.g.*, textures). The human visual cortex has been shown to have separate mechanisms to perceiving the bi-order structures [32]. The multiple distortions in images can lead to the damage of bi-order structures, which can be seen from the example images in Fig. 2. In Fig. 2, image (a) is a reference image, and images (b) and (c) are two multiply distorted images, together with the sample regions of first- and second-order structures severally marked by purple and blue boxes. By comparing images (d)-(f), we can find that the multiple distortions jointly cause the degradations of both first- and second-order structures. Therefore, in this paper, to better employ structure information for distortion evaluation, the degradations of bi-order structures are calculated.

1) *First-order structure degradation*: First-order structures characterize dominant structures [20], [32]. Image gradient information has been demonstrated to be effective for conveying the information of main structures and spatial contrast [29], [33], so the gradient map of image \mathbf{I}_i is computed for extracting the first-order structures

$$\mathbf{G}_i = \sqrt{\mathbf{G}_x^2 + \mathbf{G}_y^2}, i \in [1, 2, \dots, t+1] \quad (1)$$

where \mathbf{G}_x and \mathbf{G}_y denote the horizontal and vertical gradients, respectively, which are calculated as

$$\mathbf{G}_x = [-1 \ 0 \ 1] * \mathbf{I}_i, \quad \mathbf{G}_y = [-1 \ 0 \ 1]^T * \mathbf{I}_i \quad (2)$$

where $*$ and T denote the convolution and transpose operators, respectively.

Fig. 3(b) and (f) illustrate the gradient maps of a reference image (a) and a multiply distorted image (e). We can observe that the gradient maps can represent the first-order structures. By comparing images (b) and (f), it can be seen that the multiple types of distortions in the image cause the change of first-order structures.

In this paper, the structure degradation is characterized by the changes of both information amount and energy on structures. Among them, the information amount is measured using the two-dimensional entropy (2D-EN) on gradient maps. Compared with the widely used Shannon entropy [5], [36], 2D-EN not only measures the information amount, but also reflects the distribution characteristics of gray values [37], [38], which can model the multiple distortions more effectively. Particularly, for a pixel $\mathbf{G}_i(x, y)$, a window $(W \times W)$ centered at $\mathbf{G}_i(x, y)$ is first determined. Then, the average gradient value of all the pixels surrounding $\mathbf{G}_i(x, y)$ in the window is calculated as

$$\mathbf{A}_i(x, y) = \frac{\left[\sum_{l=y-\frac{(W-1)}{2}}^{y+\frac{(W-1)}{2}} \sum_{k=x-\frac{(W-1)}{2}}^{x+\frac{(W-1)}{2}} \mathbf{G}_i(k, l) \right] - \mathbf{G}_i(x, y)}{W^2}. \quad (3)$$

A function $\mathbf{f}_{p,q}$ is defined to represent the number of the pixels K in the whole image that simultaneously satisfies $\mathbf{G}_{i_K} = p$ and $\mathbf{A}_{i_K} = q$. The probability is calculated as

$$\mathbf{P}_{p,q} = \frac{\mathbf{f}_{p,q}}{MN}. \quad (4)$$

The 2D-EN is then defined as

$$f_1 = - \sum_{j=0}^{255} \sum_{i=0}^{255} \mathbf{P}_{i,j} \log \mathbf{P}_{i,j}. \quad (5)$$

In this work, the window size W is set to 3.

To calculate the structure energy, the singular values are computed on the gradient map. Gradient map \mathbf{G}_i is first divided into blocks with size $b \times b$. For a block \mathbf{B}_z , it can be represented as

$$\mathbf{B}_z = \mathbf{U}_z \mathbf{S}_z \mathbf{V}_z^T, z = 1, 2, \dots, R \quad (6)$$

where R represents the total number of blocks in the image, \mathbf{U}_z and \mathbf{V}_z are two $b \times b$ square matrixes satisfying $\mathbf{U}_z^T \mathbf{U}_z = \mathbf{E}$ and $\mathbf{V}_z^T \mathbf{V}_z = \mathbf{E}$. Here, T and \mathbf{E} are the transpose operation and unitary matrix, respectively, and \mathbf{S}_z is an $b \times b$ matrix with

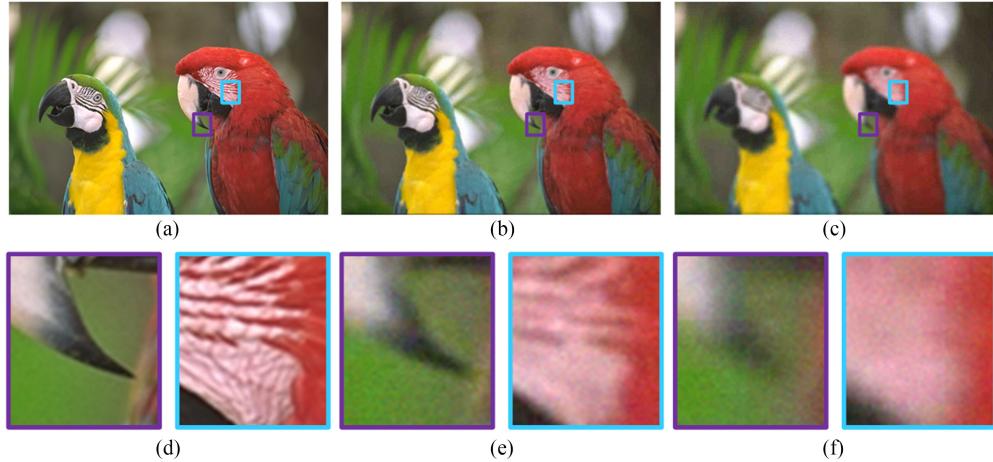


Fig. 2. Illustration of a reference image and the multiply distorted images, together with the sample regions with bi-order structure distortions marked and zoomed in. (a) Reference image; (b) Multiply distorted version of image (a) that is corrupted by Gaussian blur, JPEG compression and noise simultaneously; (c) Multiply distorted version of image (a) that is corrupted by JPEG2000 compression, Gaussian blur, JPEG compression, and noise simultaneously. The regions marked by purple boxes, namely the parrot's mouths, are the sample regions of first-order structures. The regions marked by blue boxes, namely the regions under parrot's eyes, are the sample regions of second-order structures. (d)–(f) Enlarged versions of the local regions marked in (a)–(c).

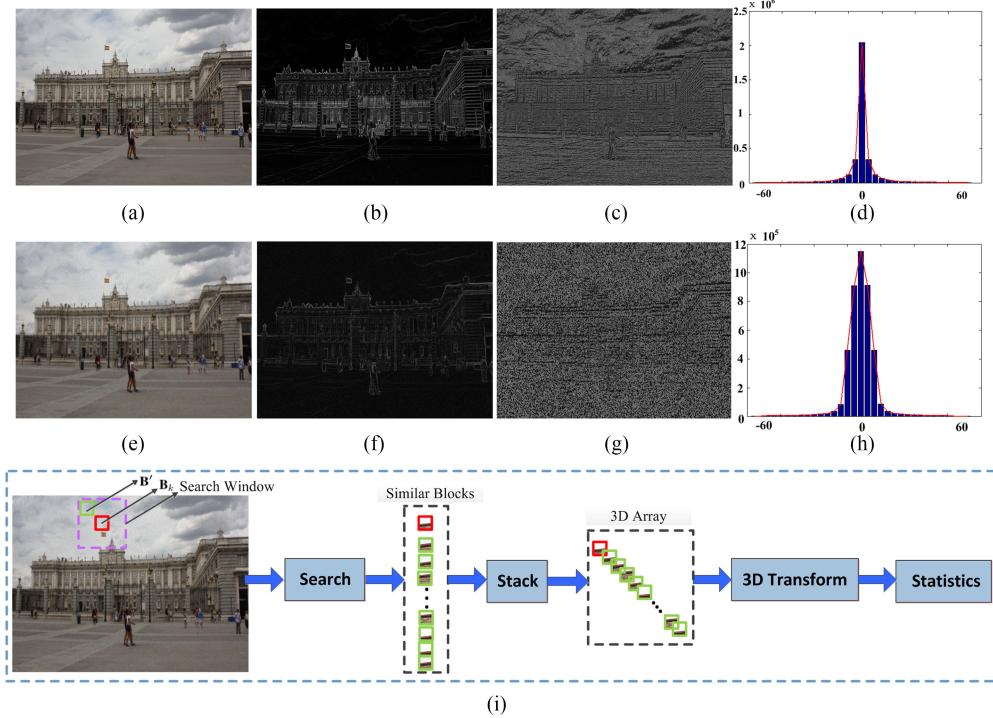


Fig. 3. Illustration of bi-order structures and calculation process of nonlocal statistics. (a) and (e) An original image and its multiply distorted version that is simultaneously distorted by noise and Gaussian blur; (b) and (f) First-order structures of (a) and (e); (c) and (g) Second-order structures of (a) and (e); (d) and (h) Nonlocal statistic histograms of images (a) and (e) using the method in [35]; (i) takes a block \mathbf{B}_k as an example to show the calculation process of nonlocal statistics.

singular values on the diagonal

$$\mathbf{S}_z = \begin{bmatrix} \xi_1 & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & \xi_2 & 0 & \cdots & \cdots & 0 \\ 0 & 0 & \ddots & 0 & \cdots & 0 \\ 0 & \cdots & 0 & \ddots & 0 & 0 \\ 0 & \cdots & \cdots & 0 & \xi_{b-1} & 0 \\ 0 & \cdots & \cdots & \cdots & 0 & \xi_b \end{bmatrix}. \quad (7)$$

The average singular value of \mathbf{B}_z is computed as

$$\mathbf{Y}_z = \frac{\sum_{t=1}^b \xi_t}{b}. \quad (8)$$

To reduce the effect of image content on structure energy, \mathbf{Y}_z is normalized by the local variance of gradient magnitude (denoted by σ_z^2)

$$\mathbf{Y}'_z = \frac{\mathbf{Y}_z}{\sigma_z^2}. \quad (9)$$

The average value of the normalized singular values of all blocks are calculated as the second feature to measure the structure degradation

$$f_2 = \frac{\sum_{z=1}^R \mathbf{Y}'_z}{R}. \quad (10)$$

2) *Second-order structure degradation*: Second-order structures represent the minor structures, *e.g.*, textures, which also have significant impact on quality perception [32]. The LBP operator, as a local descriptor, has been demonstrated to be effective for texture representation and has been widely used in texture classification [39]. In this paper, we use LBP to extract the second-order structures. For a pixel $\mathbf{I}_i(x, y)$, the LBP value is calculated as

$$\text{LBP}_{h,r}(x, y) = \sum_{w=0}^{h-1} H(\mathbf{I}_i^w(x, y) - \mathbf{I}_i(x, y))2^w \quad (11)$$

where h and r denote the number of the equi-spaced neighbors around $\mathbf{I}_i(x, y)$, and the radium of neighbors, $\mathbf{I}_i^w(x, y)$ denotes the gray value of the neighbors. Here, h and r are set to 8 and 1, respectively, $H(\cdot)$ represents a function that is defined as

$$H(a) = \begin{cases} 1, & a \geq 0 \\ 0, & \text{otherwise.} \end{cases} \quad (12)$$

In Fig. 3, images (c) and (g) show the LBP maps of images (a) and (e). It can be observed that the multiple distortions cause the changes of second-order structures.

With the calculated LBP maps, the same two features calculated on the gradient map, namely the 2D-EN and singular value are computed. They are denoted as f_3 and f_4 , respectively.

Totally, 4 features for structure degradation are calculated on each image. Since the input image has been decomposed to five images with different resolutions, 20 features are generated for structure degradation measurement.

C. Nonlocal Statistics

Nonlocal statistics states that for a natural image, the 3D transform coefficients of all blocks, together with all coefficients of their non-local similar blocks, satisfy the General Gaussian Distribution (GGD) [35]. This statistics of natural images has been widely used for image restoration [35]. Since nonlocal statistics reveals the distribution characteristics of neighbouring pixels, and the multiple distortions in the image can disrupt this statistical regularity, we employ it to measure the spatial distribution. The calculation of nonlocal statistics is shown in Fig. 3(i). The image \mathbf{I}_i is first divided into overlapping blocks with size $b \times b$. Taking the k th block \mathbf{B}_k for example, suppose that \mathbf{L} is the search window with size $O \times O$ centered at \mathbf{B}_k , the search block \mathbf{B}' moves in the search window pixel by pixel from left to right and top to bottom. In this work, b and O are empirically set to 8 and 20, respectively. The similarity between \mathbf{B}' and \mathbf{B}_k is calculated as the average Euclidean distance

$$D = \sqrt{\sum_{y=1}^b \sum_{x=1}^b (\mathbf{B}'(x, y) - \mathbf{B}_k(x, y))^2}. \quad (13)$$

A smaller value of D indicates higher similarity between \mathbf{B}_k and \mathbf{B}' . Through the traversal calculation, v most-similar blocks with \mathbf{B}_k are determined. Then \mathbf{B}_k and v similar blocks are stacked into a 3D array, and an orthogonal 3D transform is conducted [35]. Finally, the same operation is conducted on all the blocks, and for each block, a set of transform coefficients are calculated. It was stated in [35] that the transform coefficients of all blocks, including the searched similar blocks in the whole image satisfy the GGD. In Fig. 3, images (d) and (h) are the histograms of the coefficients of natural image (a) and distorted image (e), both of which are calculated using the method in [35]. It can be found that when the image is corrupted by multiple distortions, *e.g.*, image (e), the distribution characteristic of the natural image (a) is changed.

With the inspiration of the above fact, an intuitive way is to directly use this nonlocal statistics to evaluate the distortions in an image. Unfortunately, this statistics can not effectively discriminate the distortion levels in images. To understand it better, an example is given in Fig. 4, where the nonlocal statistic model is used to calculate the histograms of a reference image and three multiply distorted images with different qualities. In Fig. 4, the first row shows a reference image (a) and three distorted images (b)-(d). The images in the second row are the statistic histograms of the coefficients calculated using the non-local statistics in [35]. From images (f)-(h), we can find that although the qualities of images (b)-(d) are different, their statistic histograms of transform coefficients are quite similar. For clarity, the curves of the histograms are shown in image (m). From the curves, it is easy to see that the statistic curves of three images with different qualities calculated using the method in [35] are quite similar. So they can not discriminate the quality differences of the images, which indicates that the statistical characteristic proposed in [35] is not quality-aware. The reason may be that the distortions in images are not evenly distributed, so some blocks are not visibly distorted, and the coefficients of these natural blocks compromise the distribution characteristics of the coefficients in distorted blocks.

To solve the problem, we improve the nonlocal statistic model by separately fitting the transform coefficients of each block, together with its similar blocks. By this means, the interference of the coefficients of different blocks is eliminated. More specifically, after the block \mathbf{B}_k and its similar blocks are stacked and processed using 3D transform, including a 2D Discrete Cosine Transform (DCT) and a 1D Haar transform, which is detailed in [35]. Then their transform coefficients are modeled by the GGD, instead of gathering coefficients of all blocks in the image together. Similarly, the coefficients of all blocks in image \mathbf{I}_i are separately fitted by the GGD, namely for each block and its similar blocks, the GGD model is severally calculated. The variance and shape-parameter calculated in the GGD fitting are used as quality-aware features of each block. Finally, the calculated features of all distorted blocks are averaged as the nonlocal statistics of the whole image. In particular, suppose the coefficient set of \mathbf{B}_k and its similar blocks is θ_k , the GGD of θ_k is defined as follows:

$$F(\theta_k, u, \sigma_k^2, \epsilon_k) = \alpha_1 e^{-[\alpha_2 |\theta_k - \mu|]^\epsilon_k} \quad (14)$$

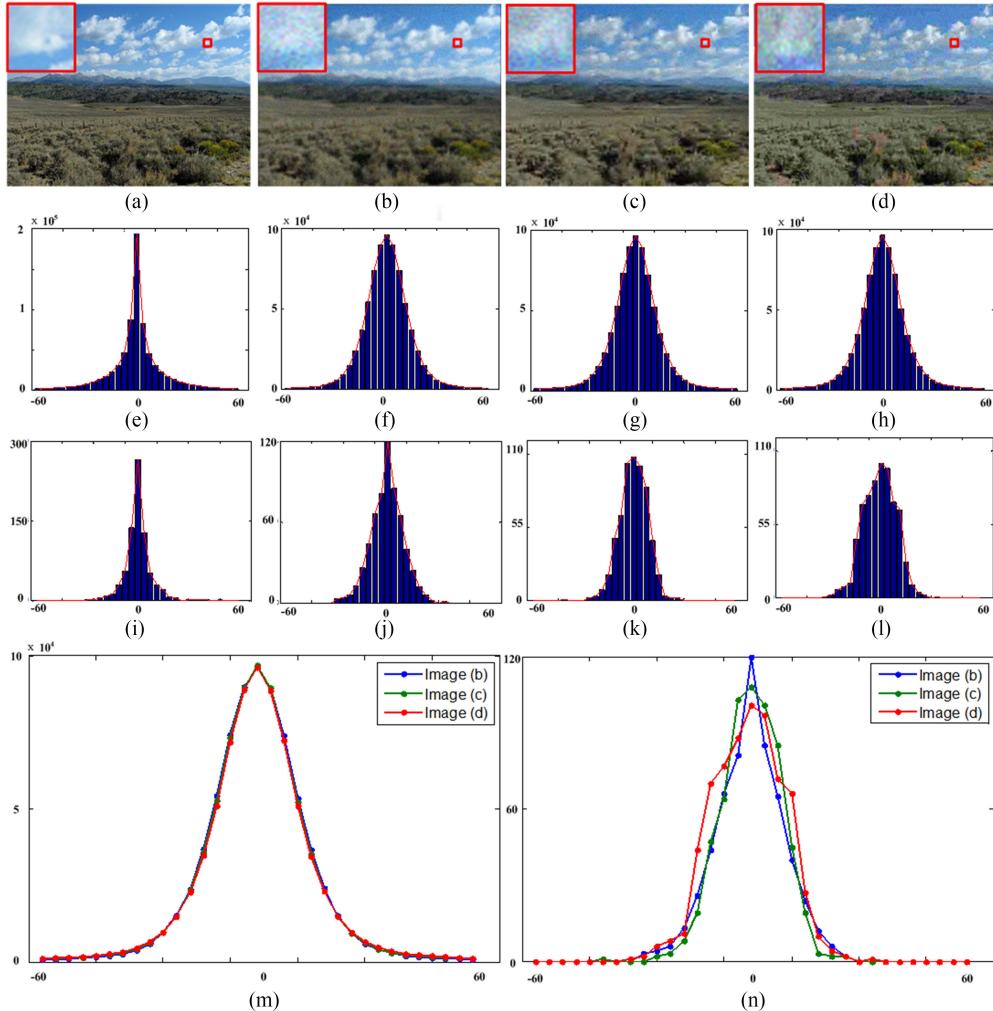


Fig. 4. Nonlocal statistic histograms using the method in [35] and the our improved method. First row: (a) Reference image; (b) Multiply distorted version of image (a) that is simultaneously distorted by Gaussian Noise (GN), Gaussian Blur (GB) and JPEG compression, MOS=4.1610; (c) Multiply distorted version of image (a) that is simultaneously distorted by GN, GB, Contrast Change (CC) and JPEG2000 compression, MOS=3.2170; (d) Multiply distorted version of image (a) that is simultaneously distorted by GN, GB, CC and JPEG compression, MOS=2.9709. Second row: Nonlocal statistic histograms using the method in [35]. Third row: Curves of the histograms using the method in [35] and our improved method. (m) Histogram curves of images (b)-(d) using the method in [35]; (n) Histogram curves of the blocks marked in images (b)-(d) using our improved method.

where μ , σ_k^2 and ϵ_k denote the mean value, variance and shape-parameter, respectively. $\epsilon_k = 2$ represents Gaussian distribution, while $\epsilon_k = 1$ denotes Laplacian distribution. Since the histogram distribution is more similar to Laplacian distribution [see from Fig. 3(d)], it is employed to model the transform coefficients. Here, α_1 , α_2 are defined as

$$\alpha_1 = \frac{\alpha_2 \epsilon_k}{2\Gamma_k(1/\epsilon_k)} \quad (15)$$

$$\alpha_2 = \frac{1}{\sigma_k} \sqrt{\frac{\Gamma(3/\epsilon_k)}{\Gamma(1/\epsilon_k)}} \quad (16)$$

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

$$\Gamma_x = \int_0^\infty \psi^{x-1} e^{-\psi} d\psi, x > 0. \quad (17)$$

Since σ_k^2 and ϵ_k are related to image quality, they are used as quality-aware features of \mathbf{B}_k . The average values of σ_k^2 and ϵ_k

of all distorted blocks in image \mathbf{I}_i are computed as the features of the whole image

$$f_5 = \overline{\sigma^2} = \frac{\sum_{i=1}^C \sigma_i^2}{C} \quad (18)$$

$$f_6 = \overline{\epsilon} = \frac{\sum_{i=1}^C \epsilon_i}{C} \quad (19)$$

where C denotes the total number of distorted blocks.

Since two nonlocal statistic features are calculated on each image, and five images with different resolutions have been computed, ten nonlocal statistic features can be generated, which constitute the second group of quality-aware features.

For better illustration of the effectiveness of the improved model, the histograms of the blocks marked in Fig. 4, which are at the same position, are shown in images (j)-(l). From the enlarged versions of the marked blocks in images (b)-(d), we can see that their qualities are quite different, so the histogram curves are expected to be different. Similar to image (m), three

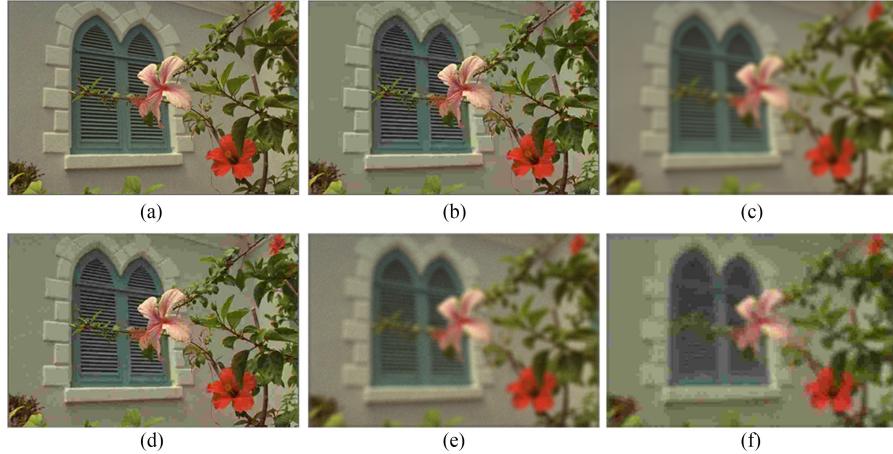


Fig. 5. Two sets of images severely distorted by single and multiple types of distortions. First row: Images with single type distortion; Second row: Multiply distorted images simultaneously corrupted by several single type distortions shown in the images in the first row.

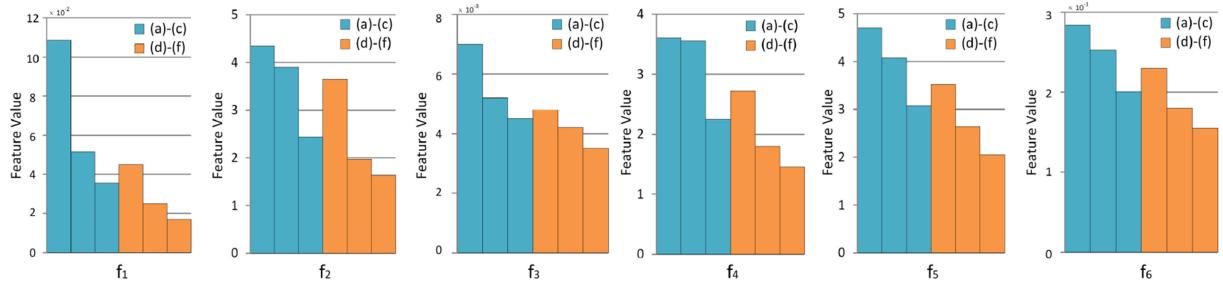


Fig. 6. Feature values of the images in Fig. 6 (f₁)–(f₆): Six proposed features extracted on the original resolution.

histogram curves are also simultaneously shown in image (n). From image (n), we can see that the improved method has better quality discrimination ability, which demonstrates the effectiveness of the improved method.

Totally, we propose six quality-aware features on the image at each resolution, i.e., f₁–f₆, including four features to measure the spatial contrast and two features to measure the spatial distribution. Among them, the spatial contrast is characterized by the bi-order structural degradation. This multi-aspect evaluation strategy can better adapt to the complexity of the characteristics of multiple distortions, such as joint effect and masking effect of one distortion on the other. To further test the performance of the proposed features, we calculate the feature values of two sets of singly and multiply distorted images. In this part, we only compare the six feature values computed on the image at the original resolution. Specifically, Fig. 5 shows two rows of images, which are corrupted by single and multiple types of distortions, respectively. It should be noted that the distortion level of each distortion type in the multiply distorted images is the same to the distortion level of the corresponding singly distorted image. By comparing the singly distorted images and the corresponding multiply distorted image, we can find that the multiple distortions in an image have joint effects and masking effects of one distortion on another. Moreover, the joint effect of multiple distortions in an image is not the simple physical addition. For example, image (d) is distorted by noise and JPEG compression, which are severally the same to the corresponding distortions in images

(a) and (b). However, we can see that the noise in image (d) is harder to be seen than the noise in image (a). This is because that the JPEG compression distortion masks the noise. Due to the joint effects, the quality of image (d) is worse than those of images (a) and (b). Fig. 6 shows the feature values of the images in Fig. 5. We can see that all the feature values of image (d) are smaller than images (a) and (b), which is consistent with the visual quality. Besides, it can be observed from Fig. 5 that the quality of images (d)–(f) is monotonously decreasing, so the feature values are expected to be monotonous. From the experimental results, we can see that the feature values of two sets of images are both monotonous, which demonstrates that the features are quality-aware and effective for measuring the joint effects and masking effects in multiply distorted images.

D. Model Training and Quality Prediction

To map the two groups of features into an overall quality score, we utilize the RF regression model [24], [25] to train the quality model. Then the trained model is employed to predict the quality score of input images. The randomness model in RF regression is the randomized node optimization model, which is calculated by node training [24]. For the *l*th node training, the problem is equivalent to the optimization of the parameters of weak learner [24]

$$f_l^* = \arg \max Z_l, f_l \in \rho_l \quad (20)$$

TABLE I
PERFORMANCES OF THE NR STATE-OF-THE-ARTS ON THREE PUBLIC MULTIPLY DISTORTED IMAGE DATABASES

Metric	Type	MDID2013 Database [27]				MLIVE Database [40]				MDID2017 Database [41]			
		PLCC	SRCC	KRCC	RMSE	PLCC	SRCC	KRCC	RMSE	PLCC	SRCC	KRCC	RMSE
BIQI [43]	G	0.3369	0.3107	0.2095	0.0478	0.7389	0.6109	0.4397	12.7439	0.6706	0.6396	0.4505	1.6344
BLIINDS [44]	G	0.7991	0.7986	0.6023	0.0308	0.8844	0.891	0.7209	7.7926	0.3084	0.2773	0.1881	2.096
BRISQUE [45]	G	0.6675	0.6826	0.4879	0.0365	0.9201	0.9049	0.7401	7.1804	0.5947	0.5931	0.4096	1.7882
DESQUE [46]	G	0.479	0.5313	0.3809	0.0388	0.9013	0.8775	0.7154	8.1919	0.8262	0.8221	0.6244	1.2578
DIIVINE [47]	G	0.4507	0.425	0.2931	0.0454	0.7234	0.6563	0.4778	13.0581	0.5691	0.5513	0.3795	1.8118
NIQE [48]	G	0.5767	0.545	0.3787	0.0415	0.8377	0.7725	0.5796	10.3292	0.6704	0.649	0.4587	1.6349
QAC [49]	G	0.2017	0.1974	0.1377	0.0498	0.3013	0.285	0.2003	18.0332	0.4964	0.4317	0.389	1.9127
IL-NIQE [50]	G	0.7001	0.7003	0.4992	0.0349	0.9087	0.9066	0.7415	7.6642	0.6899	0.6701	0.4984	1.5722
GM-LOG [51]	G	0.547	0.5354	0.3659	0.0425	0.2925	0.2475	0.1713	18.0847	0.5782	0.5796	0.3864	1.8033
NFERM [52]	G	0.8522	0.8433	0.6917	0.0295	0.8974	0.8933	0.7276	7.6918	0.4949	0.4518	0.3171	1.9146
FISMLIM [26]	MD	0.7802	0.776	0.5737	0.0318	0.8808	0.8574	0.67	8.954	0.4969	0.588	0.4306	1.9122
SISBLIM [27]	MD	0.7099	0.6934	0.4947	0.0358	0.8652	0.8572	0.6606	9.4836	0.6313	0.6545	0.4717	1.7089
GWH-GLBP [29]	MD	0.9121	0.8967	0.7219	0.0197	0.9494	0.9437	0.7978	5.866	0.8903	0.8901	0.7016	0.9902
Proposed BOSS	MD	0.9502	0.9446	0.801	0.0158	0.9549	0.9529	0.8184	5.6436	0.8997	0.8969	0.7101	0.9824

G: General-purpose metrics; MD: Multiply Distorted IQA metrics.

where ρ_l and Z_l denote the random subset and the training objective function. Z_l is defined as

$$Z_l = \sum_{d \in \eta_l} \log(|\Lambda_b(d)|) - \sum_{i \in \{L, R\}} \left(\sum_{d \in \eta_l^i} \log(|\Lambda_b(d)|) \right) \quad (21)$$

where Λ_b and η_l denote the conditional covariance matrix from probabilistic linear fitting, and the set of training data for node l , respectively. η_l^L and η_l^R are the left and right split sets. d denotes the input feature vector.

The predicted score is computed by averaging the output values of all trees

$$Q = \frac{1}{\varrho} \sum_{x=1}^{\varrho} p_x(b|d) \quad (22)$$

where ϱ and p represent the total number of regression trees and the probability density function, respectively.

IV. EXPERIMENTAL RESULTS

A. Experimental Settings

The performance of the proposed method is tested on three public multiply distorted image databases, namely MDID2013 [27], MLIVE [40], and MDID2017 [41]. The MDID2013 database consists of 324 distorted images, which are distorted by three kinds of mixed distortions, i.e., Gaussian blur, JPEG compression, and white noise. The MLIVE database consists of 450 multiply distorted images, which are distorted by two types of mixed distortions, namely Gaussian blur with JPEG compression, and Gaussian blur with white noise. The MDID2017 database consists of 1600 distorted images, which contains five types of distortions, namely Gaussian blur, JPEG compression, white noise, JPEG2000 compression and contrast distortion.

In this part, four criteria are employed for performance comparison, including Pearson Linear Correlation Coefficient (PLCC), Root Mean Square Error (RMSE), Spearman Rank order Correlation Coefficient (SRCC) and Kendalls Rank Correlation Coefficient (KRCC). PLCC and RMSE are used to measure prediction accuracy, while SRCC and KRCC are used to measure prediction monotonicity. A better metric is supposed to produce higher PLCC, SRCC, and KRCC values, and lower

RMSE value. They are computed following a five-parameter nonlinear mapping [42], [58]:

$$f(x) = \chi_1 \left(\frac{1}{2} - \frac{1}{1 + e^{\chi_2(x - \chi_3)}} \right) + \chi_4 x + \chi_5 \quad (23)$$

where x and $f(x)$ denote the predicted score and the mapped objective score, respectively, $\chi_1, \chi_2, \dots, \chi_5$ represent the fitting parameters.

B. Performance Evaluation

1) *Comparison with the state-of-the-arts:* In this part, the performance of the proposed method is first compared with three state-of-the-art Multiply Distorted (MD) IQA metrics, and ten general-purpose NR quality metrics. The three MD quality metrics are FISMLIM [26], SISBLIM [27], and GWH-GLBP [29], while the ten general-purpose NR quality metrics are BIQI [43], BLIINDS [44], BRISQUE [45], DESQUE [46], DIIVINE [47], NIQE [48], QAC [49], IL-NIQE [50], GM-LOG [51], NFERM [52]. Compared to the distortion-specific quality metrics, general-purpose quality metrics can predict image quality without the prior knowledge of the distortion type. Table I lists the experimental results, where the best performances are marked boldfaced. In the implementation, 80% images are randomly selected for model training, and the remaining 20% images are used for test. To avoid bias, the above training-test process is conducted 1000 times repeatedly. Then the median performance values are reported.

It can be seen from Table I that the proposed method achieves the best performances on all three databases, in terms of both prediction accuracy and prediction monotonicity. Specifically, (1) for the MDID2013 database [27], the proposed method is significantly better than all the state-of-the-arts. The highest PLCC and SRCC values of the existing metrics are 0.9121, and 0.8967, while the results of the proposed method are 0.9502 and 0.9446; (2) for the MLIVE database [40], among all the general-purpose quality metrics, the BRISQUE [45] method achieves the highest PLCC value and the lowest RMSE value, while the IL-NIQE [50] method achieves the highest SRCC and KRCC values. Among all existing MD quality metrics, the GWH-GLBP [29] method performs the best. However, they are still inferior to the proposed method; (3) for the MDID2017 database [41],

TABLE II
PERFORMANCES OF THE PROPOSED METHOD AND SIX FR QUALITY METRICS

Metric	MDID2013 Database [27]				MLIVE Database [40]				MDID2017 Database [41]			
	PLCC	SRCC	KRCC	RMSE	PLCC	SRCC	KRCC	RMSE	PLCC	SRCC	KRCC	RMSE
PSNR	0.5647	0.5603	0.3935	0.0419	0.7413	0.6771	0.5003	12.6934	0.6164	0.5784	0.4199	1.7350
SSIM [53]	0.5249	0.4873	0.3360	0.0433	0.8915	0.8604	0.6695	8.5692	0.8138	0.8328	0.6446	1.2804
IW-SSIM [54]	0.8514	0.8551	0.6575	0.0267	0.9110	0.8836	0.7017	7.8015	0.8983	0.8911	0.7092	0.9882
VIF [55]	0.8376	0.8444	0.6438	0.0278	0.9030	0.8823	0.6970	8.1235	0.9367	0.9306	0.7714	0.7717
FSIM [56]	0.6435	0.5818	0.3900	0.0389	0.8934	0.8637	0.6729	8.4978	0.8970	0.8873	0.7077	0.9938
GMSD [57]	0.8310	0.8310	0.6284	0.0283	0.8817	0.8519	0.6671	8.9221	0.8734	0.8611	0.6787	1.0730
Proposed BOSS	0.9502	0.9446	0.801	0.0158	0.9549	0.9529	0.8184	5.6436	0.8997	0.8969	0.7101	0.9824

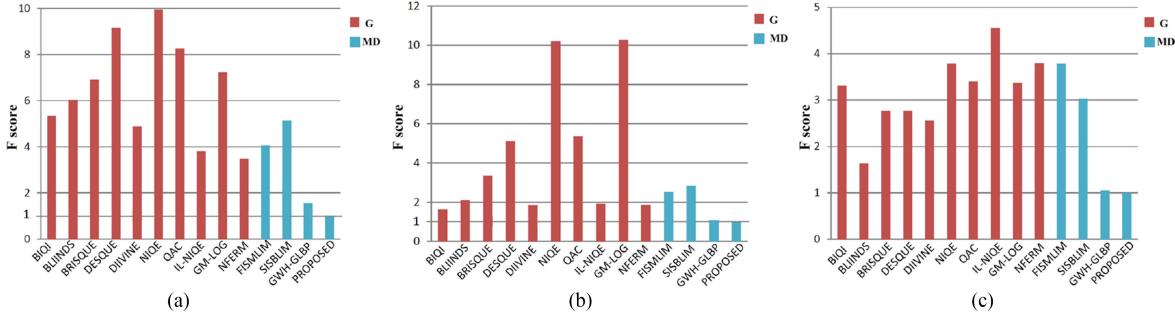


Fig. 7. F-test scores of the compared metrics against the proposed BOSS method in three public multiply distorted image databases. (a) MDID2013 database; (b) MLIVE database; (c) MDID2017 database. G: General-purpose quality metric; MD: Multiply Distorted quality metric.

the GWH-GLBP method [29] performs best among all existing MD methods and general-purpose quality metrics. By contrast, the proposed method has better performances. The above experimental results demonstrate the superiority of the proposed method.

Besides, we also compare the performance of the proposed method with those of six FR quality metrics, namely PSNR, SSIM [53], IW-SSIM [54], VIF [55], FSIM [56], and GMSD [57]. The experimental results are shown in Table II. From this table, we can see that on both MDID2013 and MLIVE databases [27], [40], the proposed method performs significantly better than all the FR quality metrics. On the MDID2017 database [41], the performance of the proposed method ranks the second, which is only worse than the VIF method [55]. However, the performances of the VIF method [55] on the other two databases are much inferior to that of the proposed method. Moreover, the proposed method does not need the original images as reference compared with these FR metrics. From these results, we know that the proposed NR metric outperforms the state-of-the-art FR image quality metrics.

2) *Statistical performance comparison*: To further testify the statistical performance of the proposed method, F-test is conducted on three databases. F-test score is defined as the squared ratio of the prediction errors (i.e., RMSE values) of a compared metric E and the proposed BOSS metric

$$F_t = \frac{(\text{RMSE}_E)^2}{(\text{RMSE}_{\text{BOSS}})^2}. \quad (24)$$

The statistics of F-test scores on three databases are illustrated in Fig. 7. From Fig. 7, we can observe that the prediction error of the proposed method is smaller than both general-purpose quality metrics and the state-of-the-art MD quality metrics on all the databases. Especially in the MDID2013 database, the

TABLE III
STATISTICAL PERFORMANCES BETWEEN BOSS AND THE STATE-OF-THE-ART QUALITY METRICS ON THREE MD IMAGE DATABASES

Metric	Type	MDID2013	MLIVE	MDID2017
BIQI [43]	G	1	1	1
BLIINDS [44]	G	1	1	1
BRISQUE [45]	G	1	1	1
DESQUE [46]	G	1	1	1
DIVINE [47]	G	1	1	1
NIQE [48]	G	1	1	1
QAC [49]	G	1	1	1
IL-NIQE [50]	G	1	1	1
GM-LOG [51]	G	1	1	1
NFERM [52]	G	1	1	1
FISMLIM [26]	MD	1	1	1
SISBLIM [27]	MD	1	1	1
GWH-GLBP [29]	MD	1	0	0

proposed method achieves much smaller prediction error than all the state-of-the-arts.

Furthermore, a threshold $F_{critical}$ is used to compare the statistical performances of an objective metric E and the proposed BOSS method [59]. The statistical performance of the proposed BOSS is supposed to be significantly better than that of metric E if $F_t > F_{critical}$. Two methods are regarded to be competitive if $1/F_{critical} < F_t < F_{critical}$. Otherwise, the statistical performance of BOSS is significantly worse than that of metric E . In this paper, 95 % confidence level is utilized to determine the threshold $F_{critical}$. For the MDID2013 [27], MLIVE [40], and MDID2017 [41] databases, the $F_{critical}$ values are 1.2009, 1.1679, and 1.0857, respectively. Table III summarizes the statistical performances, where “1” represents the much better statistical performance of the proposed BOSS method, while “0” represents that the statistical performances of two metrics are competitive.

It can be observed from Table III that, in three databases, the proposed method achieves significantly better statistical

TABLE IV
PERFORMANCES OF THE PROPOSED METHOD USING DIFFERENT WINDOW SIZES

Window Size (pixel)	MDID2013 Database [27]				MLIVE Database [40]				MDID2017 Database [41]			
	PLCC	SRCC	KRCC	RMSE	PLCC	SRCC	KRCC	RMSE	PLCC	SRCC	KRCC	RMSE
3×3	0.9502	0.9446	0.801	0.0158	0.9549	0.9529	0.8184	5.6436	0.8997	0.8969	0.7101	0.9824
5×5	0.9464	0.9422	0.7962	0.0160	0.9519	0.9473	0.8027	5.7772	0.8739	0.8692	0.6858	1.0638
7×7	0.9448	0.9405	0.7923	0.0164	0.9508	0.9441	0.8026	5.7957	0.8867	0.8804	0.6995	1.0223
9×9	0.9439	0.9379	0.7856	0.0165	0.9495	0.9439	0.7962	5.9505	0.8857	0.8787	0.6999	1.0131

TABLE V
PERFORMANCES OF THE PROPOSED METHOD WHEN DIFFERENT DOWNSAMPLING TIMES ARE ADOPTED

Downsampling Time	MDID2013 Database [27]				MLIVE Database [40]				MDID2017 Database [41]			
	PLCC	SRCC	KRCC	RMSE	PLCC	SRCC	KRCC	RMSE	PLCC	SRCC	KRCC	RMSE
0	0.8992	0.8832	0.7067	0.0217	0.9183	0.8901	0.7196	7.3807	0.8315	0.8225	0.6300	1.2194
1	0.9278	0.9200	0.7601	0.0191	0.9393	0.9280	0.7689	6.4476	0.8599	0.8522	0.6650	1.1230
2	0.9387	0.9307	0.7764	0.0177	0.9498	0.9430	0.7968	5.8822	0.8740	0.8687	0.6844	1.0666
3	0.9428	0.9357	0.7880	0.0167	0.9500	0.9435	0.7990	5.7867	0.8766	0.8720	0.6882	1.0585
4	0.9502	0.9446	0.801	0.0158	0.9549	0.9529	0.8184	5.6436	0.8997	0.8969	0.7101	0.9824
5	0.9424	0.9374	0.7851	0.0170	0.9513	0.9470	0.8022	5.7570	0.8745	0.8699	0.6850	1.0745

TABLE VI
PERFORMANCES OF THE PROPOSED METHOD AND THE GWH-GLBP METHOD WHEN DIFFERENT PROPORTIONS
OF IMAGES ARE SELECTED FOR MODEL TRAINING

Database	Training-Test	GWH-GLBP [29]				Proposed BOSS			
		PLCC	SRCC	KRCC	RMSE	PLCC	SRCC	KRCC	RMSE
MDID2013	80%-20%	0.9121	0.8967	0.7219	0.0197	0.9502	0.9446	0.8010	0.0158
	70%-30%	0.8585	0.8317	0.6449	0.0242	0.9371	0.9337	0.7773	0.0177
	60%-40%	0.6623	0.6642	0.4668	0.0377	0.9320	0.9298	0.7678	0.0182
	50%-50%	0.6160	0.6319	0.4690	0.0379	0.9266	0.9236	0.7576	0.0192
	40%-60%	0.4666	0.4260	0.2926	0.0446	0.9104	0.9088	0.7331	0.0209
MLIVE	80%-20%	0.9494	0.9437	0.7978	5.8660	0.9549	0.9529	0.8184	5.6436
	70%-30%	0.8149	0.8073	0.6397	11.1371	0.9457	0.9403	0.7901	6.1659
	60%-40%	0.7842	0.7924	0.5974	12.2011	0.9443	0.9413	0.7880	6.2453
	50%-50%	0.6477	0.6573	0.4697	14.4161	0.9371	0.9367	0.7795	6.5934
	40%-60%	0.4131	0.3272	0.2283	17.4939	0.9286	0.9267	0.7615	6.9858
MDID2017	80%-20%	0.8903	0.8901	0.7016	0.9902	0.8997	0.8969	0.7101	0.9824
	70%-30%	0.8304	0.8268	0.6231	1.2316	0.8951	0.8932	0.7098	0.9877
	60%-40%	0.8185	0.8178	0.6175	1.2597	0.8895	0.8836	0.7005	0.9908
	50%-50%	0.8134	0.8135	0.6092	1.2839	0.8804	0.8801	0.6945	1.0626
	40%-60%	0.7164	0.7058	0.5360	1.5307	0.8753	0.8771	0.6892	1.0816

performance than all the existing general-purpose quality metrics and two multiply distorted quality metrics, i.e., FISMLIM [26] and SISBLIM [27]. Regarding the GWH-GLBP method [29], its statistical performance is competitive with that of the proposed metric in MLIVE and MDID2017 databases, while it performs significantly worse than the proposed BOSS in the MDID2013 database. Therefore, it can be concluded that the proposed BOSS method achieves the best overall statistical performance.

C. Parameter Selection

In this part, we test the impacts of several parameters used in the proposed method on performance. These parameters include the window size W , downsampling times, and the number of training images.

1) *Impact of window size W* : When the 2D-EN is calculated, a parameter, i.e., window size W is employed. In this section, we testify the performances of the proposed method when using different window sizes. Table IV summarizes the experimental results. It can be seen from Table IV that the window size W has very little impact on the performance of the proposed method, and when the window size is set to 3×3 , the proposed method

TABLE VII
COMPONENT CONTRIBUTION FOR THE PROPOSED METHOD

Database	Feature	Criterion			
		PLCC	SRCC	KRCC	RMSE
MDID2013	First-order Structure	0.9265	0.9174	0.7548	0.0188
	Second-order Structure	0.8349	0.8261	0.6312	0.0280
	Bi-order Structure	0.9449	0.9368	0.7880	0.0165
	Nonlocal Statistics	0.9279	0.9202	0.7611	0.0186
MLIVE	Proposed	0.9502	0.9446	0.8010	0.0158
	First-order Structure	0.9328	0.9208	0.7561	6.8617
	Second-order Structure	0.8943	0.8807	0.6947	8.3860
	Bi-order Structure	0.9390	0.9316	0.7776	6.4652
MDID2017	Nonlocal Statistics	0.9134	0.9122	0.7423	7.5184
	Proposed	0.9549	0.9529	0.8184	5.6436
	First-order Structure	0.8064	0.7994	0.6069	1.2979
	Second-order Structure	0.7437	0.7291	0.5366	1.4718
	Bi-order Structure	0.8679	0.8602	0.6765	1.0943
	Nonlocal Statistics	0.7493	0.7401	0.5393	1.4565
	Proposed	0.8997	0.8969	0.7101	0.9824

achieves the best performance. Therefore, we set the window size 3×3 .

2) *Impact of downsampling times*: Table V lists the performance values of our metric with different downsampling times, where “0” in the first column represents the original resolution, namely without downsampling. From this table, we can observe that the proposed method performs the best when down-

TABLE VIII
CROSS-DATABASE EVALUATION RESULTS OF THE PROPOSED BOSS METRIC AND THE RELEVANT STATE-OF-THE-ART QUALITY METRICS

Training database	Test database	Criterion	BIQI	BLIINDS	BRISQUE	DESQUE	DIIVINE	GM-LOG	NFERM	GWH-GLBP	Proposed
MLIVE	MDID2013	PLCC	0.257	0.1474	0.6979	0.7706	0.7634	0.4461	0.5502	0.6508	0.7930
		SRCC	0.2365	0.1476	0.6964	0.7692	0.7618	0.4058	0.5529	0.6587	0.7939
		KRCC	0.1609	0.1008	0.4894	0.5657	0.5572	0.2762	0.383	0.4647	0.5872
		RMSE	0.0491	0.0503	0.0364	0.0324	0.0328	0.0455	0.0424	0.0386	0.031
	MDID2017	PLCC	0.5716	0.1796	0.5421	0.5992	0.5824	0.1557	0.3851	0.556	0.6187
		SRCC	0.5583	0.0948	0.5354	0.5843	0.5766	0.097	0.3468	0.5524	0.5986
		KRCC	0.3840	0.0633	0.3857	0.3946	0.3999	0.065	0.2356	0.3825	0.4166
		RMSE	1.8053	2.1676	1.8515	1.7877	1.7911	2.1765	2.0335	1.8315	1.731
MDID2013	MLIVE	PLCC	0.4461	0.1811	0.7067	0.3444	0.388	0.2044	0.7535	0.708	0.7676
		SRCC	0.5037	0.1842	0.7204	0.3785	0.3472	0.1354	0.7555	0.7204	0.7756
		KRCC	0.3668	0.1318	0.5207	0.2858	0.2628	0.0906	0.5631	0.5055	0.5552
		RMSE	16.9256	18.5993	13.3805	17.7551	17.4305	18.5128	12.2132	13.3562	12.1432
	MDID2017	PLCC	0.4216	0.2023	0.3379	0.3513	0.4185	0.0866	0.3157	0.3997	0.4520
		SRCC	0.3819	0.2102	0.3232	0.3101	0.4248	0.0694	0.3036	0.3099	0.4207
		KRCC	0.2701	0.1441	0.2201	0.2201	0.3143	0.0469	0.2036	0.2013	0.2814
		RMSE	1.9588	2.1579	2.0738	2.0629	2.0012	2.1951	2.0907	2.0197	1.9655
MDID2017	MLIVE	PLCC	0.6286	0.2836	0.2908	0.7375	0.5869	0.1348	0.596	0.7611	0.7813
		SRCC	0.5057	0.258	0.2853	0.6627	0.5971	0.1351	0.589	0.7505	0.768
		KRCC	0.3502	0.1761	0.1921	0.4736	0.4204	0.0907	0.433	0.5556	0.5607
		RMSE	14.7086	18.1358	18.0947	12.7725	15.3116	18.7395	15.1865	12.2680	12.0321
	MDID2013	PLCC	0.1774	0.1636	0.5884	0.1832	0.6511	0.3041	0.0123	0.7403	0.7033
		SRCC	0.1649	0.1488	0.5818	0.1778	0.6557	0.301	0.1386	0.742	0.7122
		KRCC	0.1143	0.0993	0.4008	0.122	0.4712	0.2005	0.1042	0.5491	0.4968
		RMSE	0.0500	0.0501	0.0411	0.05	0.0386	0.0484	0.0508	0.033	0.0358

sampling is conducted by four times. Therefore, this parameter value is used in our work.

3) *Impact of Training Images*: To test the impact of the number of training images on the performance of the proposed method, we conduct an experiment by using different proportions of images to train the quality prediction model. In implementation, the proportions of images used for model training are 80%, 70%, 60%, 50%, 40%, respectively. As comparison, the performance of another training-based MD quality metric, i.e., GWH-GLBP is also tested. The training-test process is repeated 1000 times, and the median values are reported in Table VI.

From Table VI, it can be found that the proposed BOSS method performs remarkably better than GWH-GLBP. Particularly, the performance of the proposed BOSS metric only drops slightly with the decrease of training images. Although only 40% images are randomly selected for model training, the proposed method still achieves very good performance. By contrast, with the decrease of the number of training images, the GWH-GLBP method performs dramatically worse. Especially in the MDID2013 and MLIVE databases, the PLCC and SRCC values of the GWH-GLBP method are smaller than 0.5, while the performance values of proposed metric are still higher than 0.9. For the MDID2017 database, when the training images account for 40% of all images, the PLCC and SRCC values of the proposed metric are still higher than 0.85, which shows great superiority than the GWH-GLBP method [29]. These results indicate that the proposed method is much less dependent on the number of training images, and when a small number of training images are provided, the proposed method can still achieve very good performances.

D. Components Evaluation

In the proposed metric, two groups of quality-aware features are extracted, namely bi-order structure features and nonlocal

statistic features. To evaluate the contributions of each component, these two groups of features are separately used for model training and quality prediction on three databases. Moreover, the relative contributions of the first- and second-order structure features are also tested. Table VII lists the experimental results.

From Table VII, we have the following findings. (1) First-order structure contributes more than second-order structures, which is consistent with the fact that the HVS has higher sensitivity to main structures (i.e., first-order structure) than minor structures (i.e., second-order structure) [34], [60]. Even so, the second-order structures still play a significant role, because the method using bi-order structures performs better than the method that only using the first-order structures. (2) Both types of features, i.e., bi-order structure features and nonlocal statistic features, have great contributions to the proposed method. Particularly, in the MDID2013 and MLIVE databases, the performance values of the method that separately using either type of features for model training are higher than 0.9. These results demonstrate the effectiveness of the proposed features for measuring multiple distortions. (3) Bi-order structure features contribute more than nonlocal statistic features. However, the integration of two types of features, i.e., the proposed method, performs better than separately using two groups of features, which indicates the reasonability and necessity of the integration of two groups of features.

E. Generalization Ability

For a training-based quality metric, generalization ability is important in practical applications. To this end, we further evaluate the generalization ability of the proposed method using cross-database validation. Specifically, with three multiply distorted image databases, we train the model in one database, and then test the performance of the trained model in the other

two databases. For comparison, the generalization abilities of existing training-based NR IQA metrics are also tested. The experimental results are summarized in Table VIII.

It can be observed from Table VIII that the proposed BOSS metric achieves the best generalization ability in five of the six cross-database scenarios. The only exception is that when the model is trained in MDID2017 database and tested in MDID2013 database, the proposed method ranks the second, which is slightly inferior to the GWH-GLBP method. However, the GWH-GLBP method performs worse than the proposed metric in all the other cross-database validations. These results demonstrate that the proposed method has the best generalization ability.

V. CONCLUSION

In this paper, we have proposed a no-reference and robust quality index for multiply distorted images. The proposed method is inspired by the facts that 1) the human visual system is sensitive to both spatial contrast and spatial distribution, which are subjective to change by the multiple distortions, and 2) the HVS is adapted to separately perceiving bi-order structures. Based on these, we have calculated two groups of quality-aware features based on the bi-order structure degradation and the nonlocal statistics to measure the spatial contrast and spatial distribution aspects of multiple distortions. Experimental results conducted on three public multiply distorted image databases have demonstrated the superiority of the proposed method to both existing MD and general-purpose quality metrics. The proposed method has been proved to be less dependent on the number of training images. Furthermore, cross-database evaluation results have also demonstrated that the proposed method has the best generalization ability.

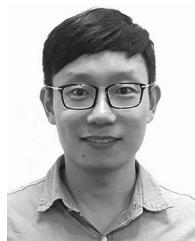
REFERENCES

- [1] W. S. Lin and C.-C. J. Kuo, "Perceptual visual quality metrics: A survey," *J. Vis. Commun. Image Represent.*, vol. 22, no. 4, pp. 297–312, May 2011.
- [2] Q. P. Jiang *et al.*, "Optimizing multistage discriminative dictionaries for blind image quality assessment," *IEEE Trans. Multimedia*, to be published, doi: 10.1109/TMM.2017.2763321, 2017.
- [3] L. Anegekuh, L. Sun, E. Jammeh, I.-H. Mkwawa, and E. Ifeatchor, "Content-based video quality prediction for HEVC encoded videos streamed over packet networks," *IEEE Trans. Multimedia*, vol. 17, no. 8, pp. 1323–1334, Aug. 2015.
- [4] X. X. Pan, F. Y. Xie, Z. G. Jiang, Z. W. Shi, and X. Y. Luo, "No-reference assessment on haze for remote-sensing images," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 12, pp. 1855–1859, Dec. 2016.
- [5] L. D. Li, W. H. Xia, W. S. Lin, Y. M. Fang, and S. Q. Wang, "No-reference and robust image sharpness evaluation based on multiscale spatial and spectral features," *IEEE Trans. Multimedia*, vol. 19, no. 5, pp. 1030–1040, May 2017.
- [6] J. J. Wu, W. S. Lin, G. M. Shi, and L. D. Li, "Orientation selectivity based visual pattern for reduced-reference image quality assessment," *Inf. Sci.*, vol. 351, pp. 18–29, Jul. 2016.
- [7] J. J. Wu *et al.*, "Visual structural degradation based reduced-reference image quality assessment," *Signal Process: Image Commun.*, vol. 47, pp. 16–27, Sep. 2016.
- [8] M. Oszust, "No-reference image quality assessment using image statistics and robust feature descriptors," *IEEE Signal Process. Lett.*, vol. 24, no. 11, pp. 1656–1660, Nov. 2017.
- [9] Q. B. Sang, H. X. Qi, X. J. Wu, C. F. Li, and A. C. Bovik, "No-reference image blur index based on singular value curve," *J. Vis. Commun. Image Represent.*, vol. 25, no. 7, pp. 1625–1630, Oct. 2014.
- [10] K. Bahrami and A. C. Kot, "A fast approach for no-reference image sharpness assessment based on maximum local variation," *IEEE Signal Process. Lett.*, vol. 21, no. 6, pp. 751–755, Jun. 2014.
- [11] L. D. Li, Y. Zhou, J. J. Wu, W. S. Lin, and H. L. Li, "GridSAR: Grid strength and regularity for robust evaluation of blocking artifacts in JPEG images," *J. Vis. Commun. Image Represent.*, vol. 30, pp. 153–163, Jul. 2015.
- [12] L. D. Li, W. S. Lin, and H. C. Zhu, "Learning structural regularity for evaluating blocking artifacts in JPEG images," *IEEE Signal Process. Lett.*, vol. 21, no. 8, pp. 918–922, Aug. 2014.
- [13] X. F. Zhang *et al.*, "High-efficiency image coding via near-optimal filtering," *IEEE Signal Process. Lett.*, vol. 24, no. 9, pp. 1403–1407, Sep. 2017.
- [14] X. F. Zhang *et al.*, "Just-noticeable difference-based perceptual optimization for JPEG compression," *IEEE Signal Process. Lett.*, vol. 24, no. 1, pp. 96–100, Jan. 2017.
- [15] X. F. Zhang *et al.*, "Low-rank decomposition-based restoration of compressed images via adaptive noise estimation," *IEEE Trans. Image Process.*, vol. 25, no. 9, pp. 4158–4171, Sep. 2016.
- [16] X. F. Zhang *et al.*, "Video compression artifact reduction via spatio-temporal multi-hypothesis prediction," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 6048–6061, Dec. 2015.
- [17] L. D. Li *et al.*, "No-reference image blur assessment based on discrete orthogonal moments," *IEEE Trans. Cybern.*, vol. 46, no. 1, pp. 39–50, Jan. 2016.
- [18] Y. P. Yan, S. L. Du, H. J. Zhang, and Y. D. Ma, "When spatial distribution unites with spatial contrast: An effective blind image quality assessment model," *IET Image Process.*, vol. 10, no. 12, pp. 1017–1028, Sep. 2016.
- [19] T. Celik and H. C. Li, "Residual spatial entropy-based image contrast enhancement and gradient-based relative contrast measurement," *J. Modern Opt.*, vol. 63, no. 16, pp. 1600–1617, Jun. 2016.
- [20] J. Larsson, M. S. Landy, and D. J. Heeger, "Orientation-selective adaptation to first- and second-order patterns in human visual cortex," *J. Neurophysiol.*, vol. 95, no. 2, pp. 862–881, Feb. 2006.
- [21] I. Bellamine and H. Tairi, "Motion detection using color structure-texture image decomposition," in *Proc. Intell. Syst. Comput. Vis.*, 2015, pp. 1–8.
- [22] D. Temel and G. Alregib, "PerSIM: Multiresolution image quality assessment in the perceptually uniform color domain," in *Proc. Int. Conf. Image*, Sep. 2015, pp. 1682–1686.
- [23] Z. L. Wan, F. Qi, Y. T. Liu, and D. B. Zhao, "Reduced reference stereoscopic image quality assessment based on entropy of classified primitives," in *Proc. Int. Conf. Multimedia Expo.*, Jul. 2017, pp. 73–78.
- [24] A. Criminisi, J. Shotton, and E. Konukoglu, "Decision forests for classification, regression, density estimation, manifold learning, and semisupervised learning," *Microsoft Res. Tech. Rep. MSR-TR-2011-114* (28).
- [25] L. J. Tang *et al.*, "An efficient and effective blind camera image quality metric via modeling quaternion wavelet coefficients," *J. Vis. Commun. Image Represent.*, vol. 49, pp. 204–212, Nov. 2017.
- [26] K. Gu *et al.*, "FISBLIM: A five-step blind metric for quality assessment of multiply distorted images," in *Proc. Workshop Signal Process. Syst.*, Oct. 2013, pp. 241–246.
- [27] K. Gu, G. T. Zhai, X. K. Yang, and W. J. Zhang, "Hybrid no-reference quality metric for singly and multiply distorted images," *IEEE Trans. Broadcast.*, vol. 60, no. 3, pp. 555–567, Sep. 2014.
- [28] Y. N. Lu, F. Y. Xie, T. L. Liu, Z. G. Jiang, and D. C. Tao, "No reference quality assessment for multiply distorted images based on an improved bag-of-words model," *IEEE Signal Process. Lett.*, vol. 22, no. 10, pp. 1811–1815, Oct. 2015.
- [29] Q. H. Li, W. S. Lin, and Y. M. Fang, "No-reference quality assessment for multiply distorted images in gradient domain," *IEEE Signal Process. Lett.*, vol. 23, no. 4, pp. 541–545, Apr. 2016.
- [30] H. Hadizadeh and I. V. Bajić, "Color gaussian jet features for no-reference quality assessment of multiply-distorted images," *IEEE Signal Process. Lett.*, vol. 23, no. 12, pp. 1717–1721, Dec. 2016.
- [31] J. J. Wu, W. S. Lin, and G. M. Shi, "Image quality assessment with degradation on spatial structure," *IEEE Signal Process. Lett.*, vol. 21, no. 4, pp. 437–440, Apr. 2014.
- [32] Q. H. Li, W. S. Lin, and Y. M. Fang, "BSD: Blind image quality assessment based on structural degradation," *Neurocomputing*, vol. 236, pp. 93–103, May 2017.
- [33] A. M. Liu, W. S. Lin, and M. Narwaria, "Image quality assessment based on gradient similarity," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1500–1512, Apr. 2012.
- [34] L. Ding, H. Huang, and Y. Zang, "Image quality assessment using directional anisotropy structure measurement," *IEEE Trans. Image Process.*, vol. 26, no. 4, pp. 1799–1809, Apr. 2017.

- [35] J. Zhang, D. B. Zhao, R. Q. Xiong, S. W. Ma, and W. Gao, "Image restoration using joint statistical modeling in a space-transform domain," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 6, pp. 915–928, Jun. 2014.
- [36] Y. Z. Zhang, J. J. Wu, G. M. Shi, and X. M. Xie, "Reduce-reference image quality assessment based on entropy differences in DCT domain," in *Proc. Int. Symp. Circuits Syst.*, May 2015, pp. 2796–2799.
- [37] L. P. Zheng, H. Jiang, Q. K. Pan, and G. Y. Li, "Medical image segmentation based on an improved 2D entropy," *Proc. 4th Int. Conf. Comput. Sci. Convergence Inf. Technol.*, Nov. 2009, pp. 1596–1599.
- [38] S. H. Zhu and Y. C. Liu, "Segmenting video shots using 2D entropy model," in *Proc. 7th World Congr. Intell. Control Autom.*, Jun. 2008, pp. 5056–5061.
- [39] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002.
- [40] D. Jayaraman, A. Mittal, A. Moorthy, and A. Bovik, "Objective quality assessment of multiply distorted images," in *Proc. Conf. Rec. 46th Asilomar Conf. Signals Syst. Comput.*, Nov. 2012, pp. 1693–1697.
- [41] W. Sun, F. Zhou, and Q. Liao, "MDID: A multiply distorted image database for image quality assessment," *Pattern Recognit.*, vol. 61, pp. 153–168, Jan. 2017.
- [42] K. Gu, G. T. Zhai, W. S. Lin, and M. Liu, "The analysis of image contrast: From quality assessment to automatic enhancement" *IEEE Trans. Cybern.*, vol. 46, no. 1, pp. 284–297, Jan. 2016.
- [43] 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.
- [44] M. A. Saad and A. C. Bovik, "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.
- [45] 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.
- [46] Y. Zhang and D. M. Chandler, "No-reference image quality assessment based on log-derivative statistics of natural scenes," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 16–21.
- [47] 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, Apr. 2011.
- [48] A. Mittal, 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.
- [49] W. F. Xue, L. Zhang, and X. Q. Mou, "Learning without human scores for blind image quality assessment," in *Proc. Int. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 995–1002.
- [50] 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.
- [51] 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.
- [52] K. Gu, G. T. Zhai, X. K. Yang, and Z. Wang, "Using free energy principle for blind image quality assessment," *IEEE Trans. Multimedia*, vol. 17, no. 1, pp. 50–63, Jan. 2015.
- [53] 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.
- [54] 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.
- [55] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Trans. Image Process.*, vol. 15, no. 2, pp. 430–444, Feb. 2006.
- [56] L. Zhang, L. Zhang, X. Q. 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.
- [57] 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.
- [58] J. J. Wu *et al.*, "Visual orientation selectivity based structure description," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 4602–4613, Nov. 2015.
- [59] L. D. Li *et al.*, "Image sharpness assessment by sparse representation," *IEEE Trans. Multimedia*, vol. 18, no. 6, pp. 1085–1097, Jun. 2016.
- [60] M. David, *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco, CA, USA : W.H. Freeman, vol. 1, 1982.



Yu Zhou received the B.S. degree from the China University of Mining and Technology, Xuzhou, China, in 2014. She is currently working toward the Ph.D. degree at the School of Information and Control Engineering, China University of Mining and Technology, Xuzhou, China. Her research interests include multimedia quality assessment and perceptual image processing.



Leida Li (M'14) received the B.S. and Ph.D. degrees from Xidian University, Xi'an, China, in 2004 and 2009, respectively. From February to June 2008, he was a visiting Ph.D. student at the Department of Electronic Engineering, National Kaohsiung University of Applied Sciences, Taiwan. From January 2014 to January 2015, he was a Visiting Research Fellow with the Rapid-Rich Object Search Laboratory, School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore. From July 2016 to July 2017, he was a Senior Research Fellow with the same laboratory, Nanyang Technological University, Singapore. Currently, he is a Full Professor with the School of Information and Control Engineering, China University of Mining and Technology, Xuzhou, China. His research interests include multimedia quality assessment, information hiding, and image forensics.



Jinjian Wu received the B.Sc and Ph.D degrees from Xidian University, Xi'an, China, in 2008 and 2013, respectively. From September 2011 to March 2013, he was a Research Assistant with Nanyang Technological University, Singapore. From August 2013 to August 2014, he was a Postdoctoral Research Fellow with Nanyang Technological University. From July 2013 to June 2015, he was a Lecturer with Xidian University. Since July 2015, he has been an Associate Professor with the School of Electronic Engineering, Xidian University, Xi'an, China. His research interests include visual perceptual modeling, saliency estimation, quality evaluation, and just noticeable difference estimation. Dr. Wu has served as the Special Section Chair for IEEE Visual Communications and Image Processing 2017, Section Chair/Organizer/TPC member for ICME 2014–2015, PCM 2015–2016, ICIP 2015, and QoMEX 2016. He was awarded the best student paper of ISCAS 2013.



Ke Gu received the B.S. and PhD degrees in electronic engineering from the Shanghai Jiao Tong University, Shanghai, China, in 2009 and 2015, respectively. He works with the Beijing University of Technology, Chaoyang, China. He is currently an Associate Editor for IEEE ACCESS and is the Reviewer for the 20 top SCI journals. His research interests include image analysis, environmental perception, quality assessment, machine learning, and big data. Dr. Gu received the Best Paper Award at the IEEE International Conference on Multimedia and Expo in 2016 and the excellent Ph.D. thesis Award from the Chinese Institute of Electronics in 2016. He was the leading special session organizer in VCIP 2016 and ICIP 2017, and the TPC member of ACM-MM 2018.



Weisheng Dong received the B.S. degree in electronic engineering from the Huazhong University of Science and Technology, Wuhan, China, in 2004, and the Ph.D. degree in circuits and systems from Xidian University, Xi'an, China, in 2010. From 2009 to 2010, he was a Research Assistant with the Department of Computing, Hong Kong Polytechnic University, Hong Kong. In 2010, he joined the School of Electronic Engineering, Xidian University, Xi'an, China, as a Lecturer, where he has been a Professor since 2016. His research interests include inverse problems in image processing, sparse and low-rank representation. Dr. Dong was a recipient of the Best Paper Award at the SPIE Visual Communication and Image Processing in 2010. He is currently an Associate Editor of IEEE TRANSACTIONS ON IMAGE PROCESSING.



Guangming Shi (SM'10) received the B.S. degree in automatic control, the M.S. degree in computer control, and the Ph.D. degree in electronic information technology from Xidian University, Xi'an, China, in 1985, 1988, and 2002, respectively. He had studied at the University of Illinois and University of Hong Kong. Since 2003, he has been a Professor with the School of Electronic Engineering, Xidian University, Xi'an, China. He is currently the Academic Leader on circuits and systems, Xidian University, Xi'an, China.

His research interests include compressed sensing, brain cognition theory, multirate filter banks, image denoising, low-bitrate image and video coding, and implementation of algorithms for intelligent signal processing. He has authored or co-authored over 200 papers in journals and conferences. Prof. Shi was awarded the Cheung Kong scholar Chair Professor by the ministry of education in 2012. He served as the Chair for the 90th MPEG and 50th JPEG of the international standards organization, Technical Program Chair for FSKD06, VSPC 2009, IEEE PCM 2009, SPIE VCIP 2010, and IEEE ISCAS 2013.