Neurocomputing 290 (2018) 185–195

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Referenceless quality metric of multiply-distorted images based on structural degradation

Tao Dai[a], Ke Gu[b], Li Niu[c], Yong-bing Zhang[a], Weizhi Lu[a,*], Shu-Tao Xia[a]

[a] Graduate School at Shenzhen, Tsinghua University, Guangdong, Shenzhen 518055, China
[b] Faculty of Information Technology,Beijing University of Technology, Beijing 100124, China
[c] Electric and Computer Engineering department, Rice University, Houston, TX 77005, USA

ARTICLE INFO

Article history:
Received 17 June 2017
Revised 27 December 2017
Accepted 11 February 2018
Available online 21 February 2018

Communicated by X. Gao

Keywords:
Image quality assessment
No-reference
Local binary pattern
Multiple distortions
Structural degradation

ABSTRACT

Multiply-distorted images, that is, distorted by different types of distortions simultaneously, are so common in real applications. This kind of images contain multiple overlaying stages (e.g., acquisition, compression and transmission stage). Each stage will introduce a certain type of distortion, for example, sensor noise in acquisition stage and compression artifacts in compression stage. However, most current blind/no-reference image quality assessment (NR-IQA) methods are specifically designed for singly-distorted images, thus resulting in their deficiency in handling multiply-distorted images. Motivated by the hypothesis that human visual system (HVS) is adapted to the structural information in images, we attempt to assess multiply-distorted images based on structural degradation. To this end, we use both first- and high-order image structures to design a novel referenceless quality metric for multiply-distorted images. Specifically, we leverage the quality-aware features extracted from both the gradient-magnitude map and contrast-normalized map, and further improve the performance by making use of redundancy of features with random subspace method. Experimental results on popular multiply-distorted image databases verify the outstanding performance of the proposed method.

1. Introduction

Recent years have witnessed a fast development of imaging devices, such as digital cameras, and smart phones. The fast development of imaging devices has put forward the requirement for high quality of experience of end-users [1]. In practice, multiply-distorted images, namely images with multiple distortion types, are common in real applications. The different types of distortions often arise from overlaying stages, such as signal acquisition, compression and transmission stages. The appearance of multiple distortions makes the problem of blind/no-reference image quality assessment (NR-IQA) even more complex, since the modeling of the interaction between the single distortions and their interactive effects on the overall visual quality remains still challenging [2]. Besides, most existing NR-IQA methods are designed for singly-distorted images and struggle to achieve satisfying results on multiply-distorted images [3–5]. Thus, it is still demanding to develop new NR-IQA methods for multiply-distorted images.

Many early works attempt to assess the perceived quality of distorted images by considering structural information of an image. It has been widely acknowledged that structural information, like the image gradient, plays a key role in evaluating perceptual quality [6–8]. To date, there have existed many successful IQA methods based on structural degradation [4,9–11]. Taking the successful structural similarity index (SSIM) and its variants [9,10] for example, these structural-based methods achieve remarkable performance by exploring the structural information. The performance advantage comes from the fact that human visual system (HVS) is highly adapted to extract the structural information from a scene. In general, it would have two main advantages based on structural degradation to evaluate image quality: 1) it is an effective way to measure the degree of distortion based on structural degradation; 2) structural degradation contributes to distinguishing different types of image distortions. In other words, structural degradation intuitively reflects the overall quality of the perceived image.

Inspired by the above observations, some recent NR-IQA methods [4,11] have also been developed for multiply-distorted images based on the potential structural degradation, given that the image structure conveys essential visual information from a scene. In these methods, local binary pattern (LBP) operators are applied to extract structural features from different domains, such as the gradient domain [4] and the perceptual opponent-color domain [11]. Unfortunately, these methods do not fully exploit the image structural information. More precisely, most existing

* Corresponding author.
E-mail addresses: dait14@mails.tsinghua.edu.cn (T. Dai), guke.doctor@gmail.com (K. Gu), wzlusd@sz.tsinghua.edu.cn (W. Lu), xiast@sz.tsinghua.edu.cn (S.-T. Xia).

https://doi.org/10.1016/j.neucom.2018.02.050
0925-2312/

methods only utilize first-order image structures (e.g., gradient information) without taking account of high-order image structures (e.g., texture information). In practice, however, high-order image structures also play a vital role in evaluating the visual quality, due to the fact that HVS has separate mechanisms to process the first- and high-order image structures [12].

To make use of various types of image structures better, we develop a novel blind/NR-IQA metric of multiply-distorted images based on structural degradation, which considers both the first- and high-order image structures. In fact, the first-order image structures characterize the main structures of images (e.g., edges), while the high-order image structures represent fine-grained details of images (e.g., textures) [12]. Such two structural features are sensitive to different types of distortions, which thus renders them effective to reflect the perceptual quality of images. In order to utilize such image structures better, we attempt to extract a number of quality-aware features from the gradient magnitude map and contrast-normalized map, which represent the first- and high-order structural patterns of distorted images, respectively. Specially, the rotation invariant uniform LBP operator is adopted to extract the first-order structural patterns, while a generalized center-symmetric LBP (GCS-LBP) operator is employed to extract high-order structural patterns. For clarity, the main contributions of the work are summarized as below:

- Based on structural degradation, we utilize both the first- and high-order image structures to characterize the image structural features.
- Technically, when extracting the first-order structural information, we employ four directional high-pass filters (5×5) encompassing the diagonal directions to incorporate neighboring information. When extracting the high-order information, we propose a generalized center-symmetric LBP to resist the visual artifacts introduced by the contrast modifications.
- In order to avoid overfitting, we apply the random subspace method [13] in the feature space. Experimental results confirm that our proposed method acquires superior performance, by comparison with prevailing full reference (FR) and no reference IQA methods.

The rest of this paper is constructed as follows. In Section 2, we briefly review some related work on objective IQA metrics. In Section 3, we introduce the proposed IQA metric in detail, and analyze the effects of feature set. Then, experimental results are provided in Section 4. Finally, we draw the conclusions in Section 5.

2. Related work

Objective IQA algorithms aim to evaluate image visual quality automatically according to computational models, which have been successfully and widely applied in the field of image processing, such as image enhancement [14,15], image denoising [16–18], and image classification [19,20]. According to the amount of available information from the pristine image (reference image), existing objective visual quality metrics can be roughly divided into full-reference (FR), reduced-reference (RR) and no-reference (NR) IQA metrics. Most early IQA methods belong to the class of FR-IQA methods. Traditional FR-IQA metrics like peak signal-to-noise ratio (PSNR) and mean square error (MSE) are the most commonly-used IQA metric in different applications for their simplicity and efficiency. Nevertheless, it is shown that both PSNR and MSE can not coincide with human subjective ratings very well [1,21]. To overcome this drawback, more advanced FR-IQA methods that highly agree with human perceptual quality have been proposed, such as SSIM [22], OSS-SSIM [10], VIF [23], MAD [24], ADM [25], FSIM [6], GMS [7], and GMSD [8]. RR-IQA algorithms [26,27] tend to compare the partial information between the distorted image and the corresponding reference image. However, in most practical applications (e.g., denoising, enhancement), the reference image is generally unavailable, thus constraining the usage of FR- and RR-IQA methods. Hence, it is of high demand to design NR-IQA methods in practice, without access to reference image.

Most of early NR-IQA metrics have been specifically designed to reflect the visual quality of images in particular applications, like Gaussian blur (GB) [28,29], white noise (WN) [30], JPEG compression [29,31,32], and contrast distortion [33]. Thus, these methods can be viewed as distortion-specific methods by assuming that the prior knowledge of the distortion is provided beforehand or the distortion type is known in advance, which obstructs the application of these methods in the real-world scenario.

In contrast with early distortion-specific methods, recent general-purpose NR-IQA methods can handle images distorted by various distortion types. Based on whether to use subjective scores during the learning process, general-purpose NR-IQA methods can be further categorized as opinion-aware and opinion-free methods. The opinion-free methods require no access to subjective human ratings in the learning process [34,35]. One approach is proposed in [34] to fit a multivariant Gaussian (MVG) model to characterize the regularities of the natural images based on the spatial domain NSS features from natural (distortion-free) images, which is further improved in [35] by combing more perceptual features from image gradient, log-Gabor and color domains.

However, most existing NR-IQA methods fall under the category of opinion-aware methods by learning the computational models based on distorted images and the associated subjective human ratings, which generally takes two stages. The first stage is to extract quality-aware features that represent the image quality from distorted images, while the second stage is to learn a regression model to map the quality-aware features to the final quality score. Intuitively, the choice of the extracted quality-aware features play a vital role in such learning based methods. An ideal set of quality-aware features should be insensitive to image content changes (e.g., illumination change), and sensitive to diverse image distortions. One kind of popular features are based on natural scene statistics (NSS), given the assumption that natural images own regular statistics and various distortion types tend to violate such regularity to certain different degrees. According to such assumptions, many NR-IQA methods take advantage of naturalness of NSS in different domains, such as spatial NSS [36], DCT NSS [37], Wavelet NSS [38], and hybrid of several types of transform domain NSS [39]. Other efficient handcrafted features have also been designed, such as image gradients [40,41], image filter responses [42,43], and image entropies [14,44]. In [40], the authors exploit Laplacian of Gaussian responses and the joint statistics of image gradients as quality-aware features. Apart from first-order derivatives (e.g., image gradients) mentioned above, high-order derivatives of images are also extracted for NR-IQA task in [42,43]. For instance, in [44], the spatial NSS features and free energy principle based features are combined, leading to promising performance in evaluating image quality. With the rapid development of neural network, some opinion-aware methods input raw local image patches for neural networks and attempt to learn quality-aware representation [45,46] instead of designing handcrafted features. For a comprehensive review on NR-IQA methods, the readers can refer to [47].

3. Proposed method

The proposed method is based on the HVS's sensitivity to structural degradation. To make use of the structural information better, we extract a number of first-order and high-order features, which respectively characterize the main structures of images (e.g., edges), and fine-grained details of images (e.g., textures). Such two

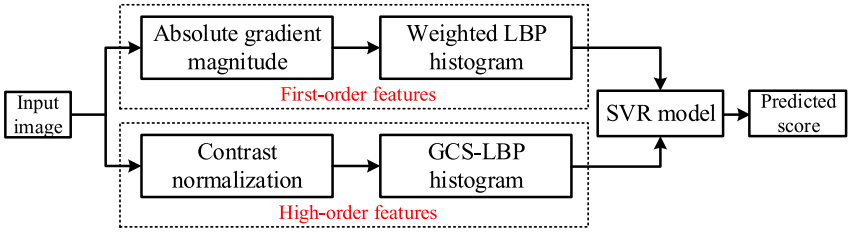


Fig. 1. Flowchart of the proposed method.

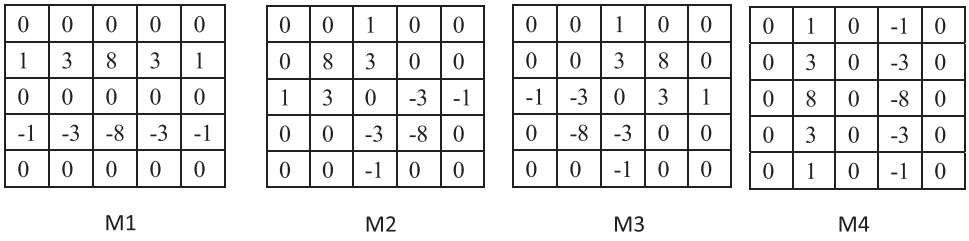


Fig. 2. Operators for calculating the gradient value.

structural features are sensitive to different types of distortions, thus rendering them effective to reflect the perceptual quality of images. We illustrate the flowchart of our method in Fig. 1, which will be detailed as follows. At first, we extract two sets of features, in which the first feature set is extracted from the image gradient-magnitude map to account for first-order image structures while the second feature set is extracted from the contrast-normalized map to account for the high-order image structures. Then, these two sets of feature are fused to represent image structural information, followed by mapping the extracted features to the final score of perceptual quality with learning methods, such as support vector regression (SVR).

A. First-order feature set

As shown in [12,41], first-order patterns characterize luminance variations, *i.e.*, image gradient information. It is well acknowledged that image gradients convey prominent visual information (e.g., edges) for human visual system [7,8], since the center-surround cell in eyes conducts a local comparison that can be treated as gradient operators [48,49]. In IQA task, gradient magnitude (GM) has been widely employed in [7,8,41,49] because different distortions can be properly reflected from the gradient magnitude. As is shown in Fig. 3, we display an intuitive example with four sample images and their corresponding GM maps (the middle row).

To calculate the image gradient, the Sobel and Prewitt filters are widely-used gradient operators, which are used to extract the first-order image structures [4,43] due to their simplicity and efficiency. Then the GM is often expressed as the root of the sum of the squares of image gradients along horizontal and vertical directions. However, using such simple filters to compute image gradients will result in some issues: (1) these filters are often too small (3×3) to encompass sufficient adjacent information; (2) these filters neglect neighboring information of some directions (e.g., diagonal directions), which also contribute to the structural information for image understanding.

To extract more neighboring information, we adopt four directional high-pass filters M_k $(k = 1, 2, 3, 4)$ (see Fig. 2) to compute the image gradients. After that, the GM map is computed as the maximum weighted average of difference of image patch centered at each pixel, i.e.,

$$G(i,j) = \max_{k=1,2,3,4} \text{mean2}(|I \otimes M_k|(i,j)), \tag{1}$$

where $|\cdot|$ denotes the absolute value operator, the symbol "$\otimes$" represents the convolution operation, and mean2$(\cdot)$ stands for the average value of a matrix. I and G denote the counterpart of GM map and its corresponding distorted image; i and j represent the pixel position within an image.

After that, image primitive microstructures (e.g., edges, lines) in the GM map can be described by the basic LBP operator [50], which can describe the relation by calculating differences between a pixel and its adjacent pixels. Using the LBP operator in the GM map, the LBP code at each pixel can be expressed as

$$\text{LBP}_{P,R} = \sum_{p=0}^{P-1} s(G_p - G_c) \cdot 2^p, \tag{2}$$

in which

$$s(G_p - G_c) = \begin{cases} 1, & G_p - G_c \geq 0 \\ 0, & \text{otherwise}, \end{cases} \tag{3}$$

where G_c and G_p are the gradient magnitudes at the center of neighborhood and its neighboring pixels; P stands for the number of neighboring pixels, and R denotes the radius of the surrounding neighborhood.

To achieve rotation invariance, a variant of LBP operator, called the rotation-invariant and uniform variant LBP, can be formulated as

$$\text{LBP}^{\text{riu2}}_{P,R} = \begin{cases} \sum_{p=0}^{P-1} s(G_p - G_c), & \text{if } \mathcal{U}(\text{LBP}_{P,R}) \leq 2 \\ P+1 & \text{otherwise}, \end{cases} \tag{4}$$

where $\mathcal{U}$ is the uniform measure that calculates the number of bitwise transitions, and $\text{LBP}^{\text{riu2}}_{P,R}$ has $\mathcal{U}$ value of at most 2. The uniform measure $\mathcal{U}$ can be computed by

$$\mathcal{U}(\text{LBP}_{P,R}) = ||s(G_0 - G_c) - s(G_{P-1} - G_c)|| + \sum_{p=1}^{P-1} ||s(G_{p-1} - G_c) - s(G_p - G_c)||. \tag{5}$$

Compared with the basic $\text{LBP}_{P,R}$ operator with 2^P patterns, the $\text{LBP}^{\text{riu2}}_{P,R}$ operator would have only $P+2$ patterns, and can detect different microstructures (e.g., edge, lines and spots). However, it only encodes the sign of the differences between the central pixel and its corresponding adjacent pixels. As shown in [51,52], the magnitude components of the differences also contribute to additional discriminant information. To encode such information, we

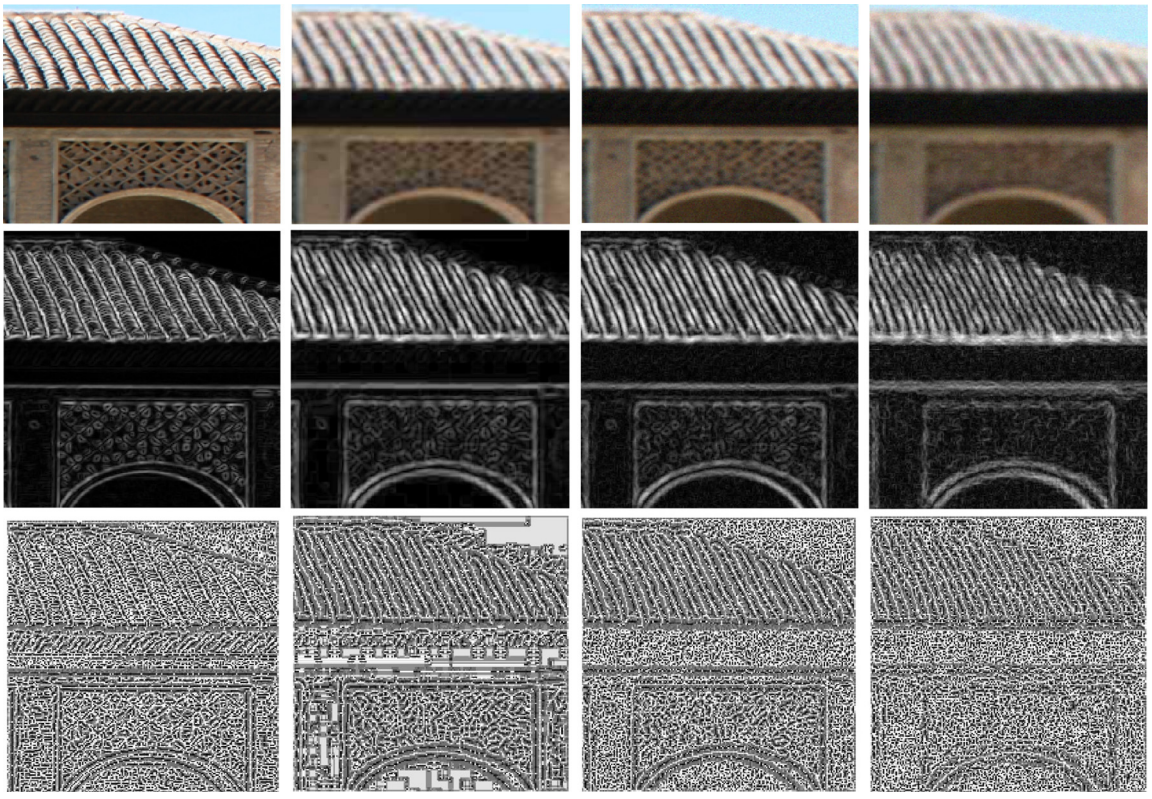


Fig. 3. Gradient magnitude and the corresponding LBP maps under different multiple distortions. (From left to right) The first to fourth columns are the results of pristine image, image with GB+JPEG, image with GB+WN, and image with GB+JPEG+WN, respectively. (From top to down) The first to third rows show sample images, the associated gradient magnitude maps, and LBP maps, respectively. .

use the gradient magnitude as the $\mathrm{LBP}^{\mathrm{riu2}}_{P,R}$ weight of each pixel, the normalized weighted LBP histogram can be expressed as

$$H_{\mathrm{LBP}}(h) = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N} G(i,j)\cdot f(\mathrm{LBP}^{\mathrm{riu2}}_{P,R}(G(i,j)),h)}{\sum_{i=1}^{M}\sum_{j=1}^{N} G(i,j)}, \tag{6}$$

with

$$f(m,n) = \begin{cases} 1, & m = n \\ 0, & \text{otherwise}, \end{cases} \tag{7}$$

where $h \in [0, P+1]$ is the possible LBP patterns and the gradient magnitude $G(i, j)$ is assigned to the weight of LBP histogram; M and N represent the image size. In such a way, we can extract the first-order image structures with high-contrast changes.

To illustrate the effects of the LBP operator under multiple distortions, four sample images and their associated LBP maps are exhibited in Fig. 3, in which the first row contains the source image and three distorted images corrupted by Gaussian Blur embedded by JPEG compression (GB+JPEG), Gaussian Blur embedded by white noise (GB+WN), and Gaussian Blur embedded by JPEG compression and white noise (GB+JPEG+WN), while the third row shows their corresponding LBP maps in the GM maps. As shown in Fig. 3, distinct image distortions exhibit the LBP patterns with their own characteristics, thus rendering it effective to characterize various distortions.

B. High-order feature set

As stated above, it is effective to detect the first-order patterns, which describe the luminance changes of images by a linear filter. Although linear filters can simplify calculation in the detection of image structures, they limit the types of image structures that can be obtained from image statistics. For instance, HVS can also detect fine image structures (e.g., texture), where there is no variation in mean luminance. Moreover, from the analysis in [12], we can know that such fine structures belong to high-order patterns, which can not be directly detected by linear filters.

To extract the high-order image structures, we first apply local contrast normalization operation to luminance domain due to the fact that the local contrast normalization operation has a decorrelating effect [53]. Such a nonlinear operation is often used to imitate the HVS's nonlinear masking phenomenon by mean subtraction and divisive normalization. Here, we apply similar preprocessing model as [34,36] to intensity image I to produce the normalized version $\hat{I}$ of I, *i.e.*,

$$\hat{I}(i,j) = \frac{I(i,j) - \mu(i,j)}{\sigma(i,j) + const}, \tag{8}$$

where *const* is a small positive value for avoiding division-by-zero, i and j are the pixel position indices, and the mean value and standard deviation of an image patch are computed as

$$\mu(i,j) = \sum_{x=-X}^{X}\sum_{y=-Y}^{Y} w_{x,y} I(i+x, j+y), \tag{9}$$

$$\sigma(i,j) = \sqrt{\sum_{x=-X}^{X}\sum_{y=-Y}^{Y} w_{x,y}[I(i+x,j+y) - \mu(i,j)]^2}, \tag{10}$$

where $w = \{w_{x,y} | x = -X, \ldots, X, y = -Y, \ldots, Y\}$ is a Gaussian kernel function with unit volume.

Fig. 4 displays an intuitive example of local contrast normalization. Note that the local mean subtraction process guarantees that the image patch of the response results have zero mean luminance, and thus the contrast-normalized results are invisible to linear filters. Moreover, division by local contrast strengthens image structures of low or moderate contrast. From Fig. 4, it is observed

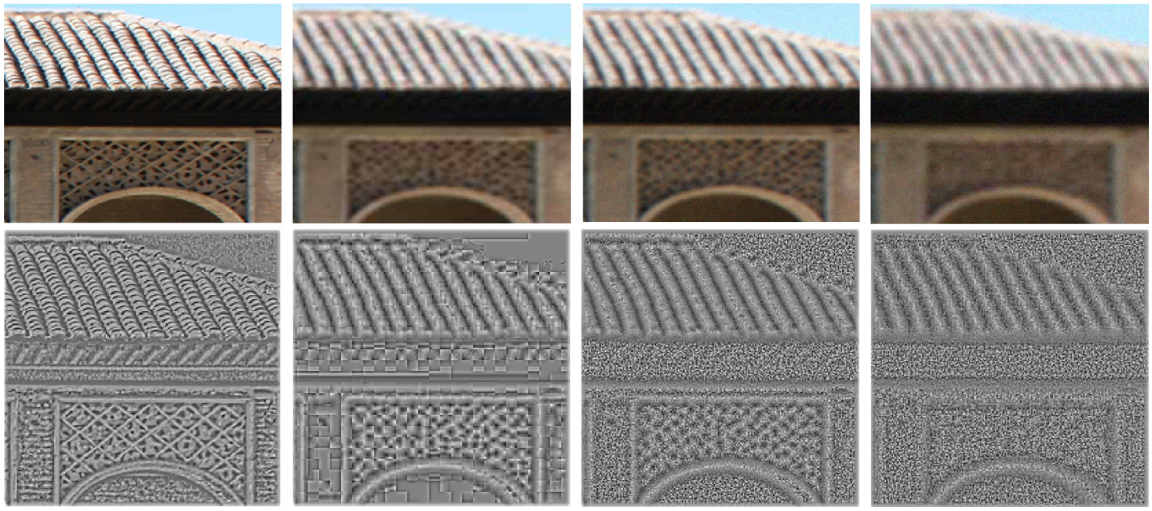


Fig. 4. Sample images and the corresponding contrast-normalized maps by local contrast normalization under different multiple distortions. (From left to right) The first to fourth columns are the results of pristine image, image with GB+JPEG, image with GB+WN, and image with GB+JPEG+WN, respectively. (From top to down) The first to second rows show sample images, and the contrast-normalized maps by local contrast normalization, respectively.

that the fine image structures (e.g., rooftop tiles) become clearer after local contrast normalization. Thus, the contrast-normalized maps of the image characterize fine-detail structures related to object surfaces, which belong to the high-order patterns. Then we attempt to extract the high-order features from the contrast-normalized map $\hat{I}$. Nevertheless, the contrast modifications tend to create visual artifacts [54]. From Fig. 4, we can observe that the noise and compression scheme patterns are also enhanced. In such cases, the standard $\mathrm{LBP}^{\mathrm{riu2}}_{P,R}$ operator is not able to characterize the quality-aware features well. To alleviate this issue, we present a novel generalized center-symmetric LBP (GCS-LBP) operator, a robust variant of the basic LBP, to extract high-order features. The GCS-LBP encoding map for each image is formulated as

$$\text{GCS-LBP}_{R,P,T} = \sum_{p=0}^{P/2-1} \hat{s}_T(|\hat{I}_p - \hat{I}_{p+(P/2)}|)2^p, \tag{11}$$

where $\hat{I}_p$ and $\hat{I}_{p+(P/2)}$ denote the center-symmetric pairs of pixels among P pixels, which are equally spaced on a neighborhood; the function $\hat{s}_T(\cdot)$ is

$$\hat{s}_T(y) = \begin{cases} 1, & y > T \\ 0, & \text{otherwise}, \end{cases} \tag{12}$$

where T is a small threshold value in the range of [0,1], and the proposed GCS-LBP has only $2^{P/2}$ histogram bins.

C. Multi-scale feature sets

In the design of our metric, a multi-scale model is also taken into account, since the recent research has revealed that HVS perceives image edges in a coarse-to-fine strategy [55]. In practice, the coarser scale is formed by low-pass filtering and then downsampled by a factor of two. In our recent work [56], we attempt to seek for the optimal scale based on the given viewing distance and the image size. As shown in [56], an appropriate scaling coefficient z is close to the square root of the ratio of the focused visual scope and the image size

$$z = \sqrt{\frac{1}{4\tan(\frac{\theta_h}{2}) \cdot \tan(\frac{\theta_w}{2})} \cdot \left(\frac{h}{d}\right)^2 \cdot \frac{w}{h}}, \tag{13}$$

where h, w, and d denote the image height, the viewing distance and image width respectively; θ_h and θ_w represent horizontal and vertical visual angles. From the analysis in [56], we found $z = 0.4955 (\approx 0.5)$ works well. Thus the used downsampling factor "2" can be viewed as approximately optimal in given viewing conditions.

To consider varying image resolution and viewing distance [36], our feature sets, *i.e.*, $\mathrm{LBP}^{\mathrm{riu2}}_{P,R}$ and GCS-LBP histograms, are extracted from the GM maps and contrast-normalized maps in three scales. To be specific, the coarser scale is first processed by a low-pass filter, followed by a downsampling operation with a factor of two. Therefore, the extracted feature sets have 26 components at each scale, thus 78 components in total.

D. Analysis of LBP operators and feature sets

As stated above, we have used two kinds of LBP operators, *i.e.*, $\mathrm{LBP}^{\mathrm{riu2}}_{P,R}$ and GCS-LBP, to extract the first-order and high-order image structures, respectively. Both of these two operators belong to variants of the standard LBP. There are some similarities and differences between these two kinds of LBP. The main similarities lie in two aspects: 1) both of them compute the pixel differences between the neighboring pixels; 2) both of these two operators have much fewer patterns (histogram bins) than the standard LBP (2^P bins). The main differences between them can be summarized as follows: (1) for each pixel, GCS-LBP needs a proper threshold value T to determine whether to encode the corresponding pixel differences, while $\mathrm{LBP}^{\mathrm{riu2}}_{P,R}$ uses the sign information of pixel differences directly; (2) GCS-LBP computes the pixel differences between center-symmetric pairs of pixels along a circular window, while $\mathrm{LBP}^{\mathrm{riu2}}_{P,R}$ computes the pixel differences between the center pixel and its neighboring pixel along a circular window. Compared with $\mathrm{LBP}^{\mathrm{riu2}}_{P,R}$, GCS-LBP is tolerant to illumination changes, image noise and small perspective distortions. Therefore, it is more suitable to extract the high-order image structures in the contrast normalized map, where there are additional visual artifacts created by the contrast modifications.

For feature sets, the proposed method consists of two sets of quality-aware features, which characterize the first- and high-order information. To have an intuitive understanding of the complementary role of the first- and high-order features, we provide a few concrete examples in Fig. 5, from which we can see that the first row displays a pristine image, and its distorted versions under various multiple distortions. The second and third rows show the first-order features (weighted LBP histogram) and high-order features (GCS-LBP histogram), respectively. We can have the following observations. The $\mathrm{LBP}^{\mathrm{riu2}}_{8,1}$ have 10 distinct patterns. Among

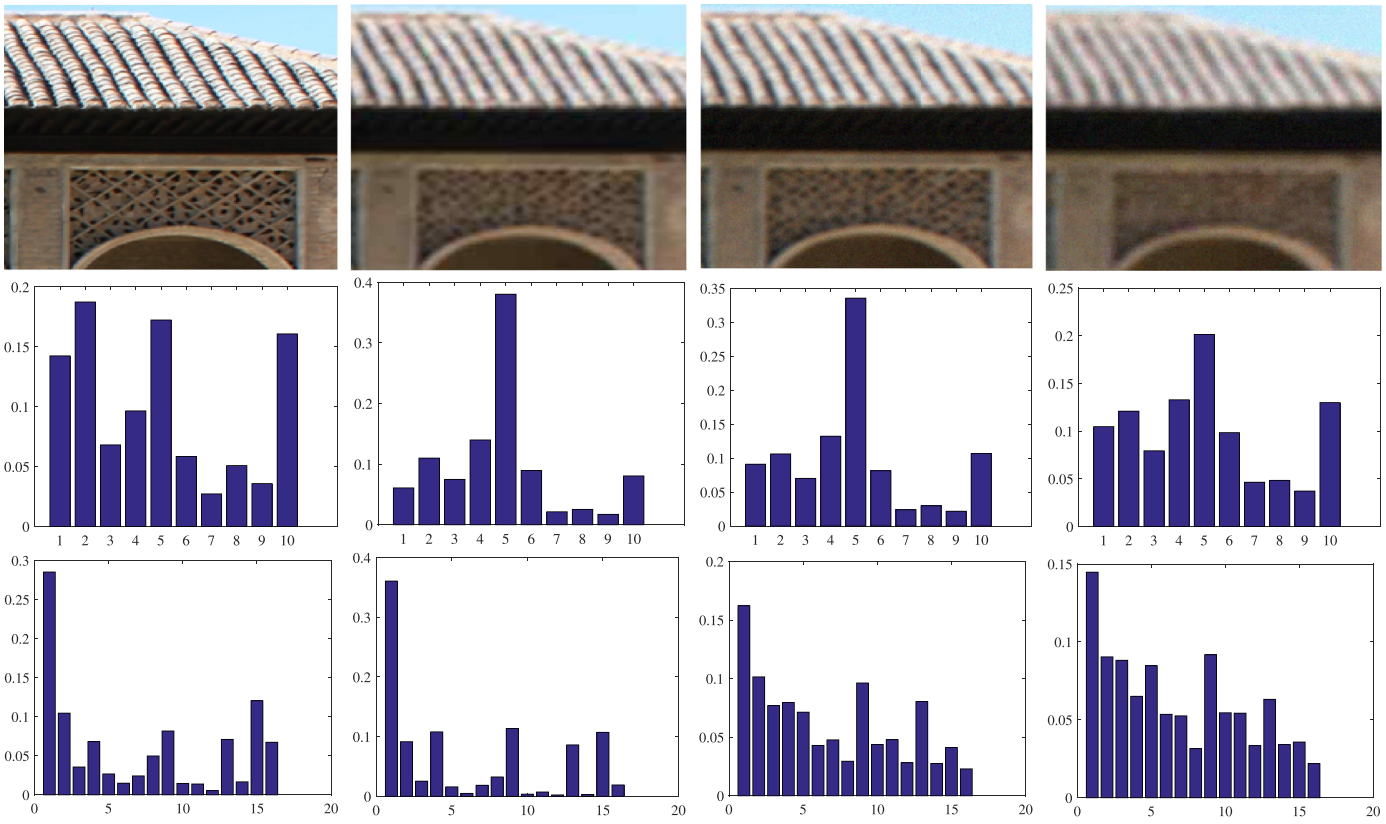


Fig. 5. Examples to illustrate the contribution of first-order and high-order information. (From left to right) The first to fourth columns are the results of pristine image, image with GB+JPEG, image with GB+WN, and image with GB+JPEG+WN, respectively. (From top to down) The first to third rows show sample images under different distortions, the associated weighted LBP maps, and GCS-LBP maps, respectively.

them, pattern 0 and 8 denote bright/dark spot or flat area; patterns 1–7 stand for edges of varying curvatures; pattern 9 represents the non-uniform patterns [50]. The GB+JPEG and GB+WN distortions smooth image edge and thus widen edge patterns, but increase pattern 5, caused by blocking effect and false edge. More multiple distortion such as GB+JPEG+WN further degrades the image information, thus leading to a more flat histogram. Similarly, the GCS-LBP has 16 distinct patterns. Among them, pattern 0 denotes the low-contrast structures (e.g., flat area), while the other patterns represents texture information of varying contrast. From Fig. 5, we can observe that the two distorted images with GB+JPEG and GB+WN share similar weighted LBP histograms, which indicates that the first-order feature can not distinguish the distortion between GB+JPEG and GB+WN well. By contrast, their GCS-LBP histograms exhibit different patterns and provide useful information for visual quality estimation. Hence, we adopt both first- and high-order feature set together in this work, since these two types of feature sets play a mutually complementary role in visual quality evaluation task.

E. Regression module for quality evaluation

After the training set is provided, the commonly used SVR [57] is used to learn the regression module in our method. The training set can be represented as $\{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_r, y_r)\}$, where $\mathbf{x}_i \in R^n$ denotes the extracted feature vector and y_i stands for its corresponding human score. The standard form of SVR can be written as follows:

$$\min_{\mathbf{w},b,\boldsymbol{\varepsilon},\boldsymbol{\varepsilon}^*} \ \frac{1}{2}\mathbf{w}^T\mathbf{w} + C\left\{\sum_{i=1}^{r}\varepsilon_i + \sum_{i=1}^{r}\varepsilon_i^*\right\} \tag{14}$$

$$\text{subject to} \quad \mathbf{w}^T\phi\mathbf{x}_i + b - y_i \le \epsilon + \varepsilon_i \tag{15}$$

$$y_i - \mathbf{w}^t\phi(\mathbf{x}_i) - b \le \epsilon + \varepsilon_i^* \tag{16}$$

$$\varepsilon_i, \varepsilon_i^* \ge 0, i = 1, \ldots, r, \tag{17}$$

where C is a trade-off parameter, and $K(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i)^T\phi(\mathbf{x}_j)$ is the kernel function.

The traditional SVR uses the whole feature set for training. For the traditional SVR, overfitting happens when the training set is relatively small compared to the high dimensionality of the feature vector. In order to avoid overfitting, we apply the random subspace method [13] which is similar to bagging. However, unlike bagging that bootstraps training samples, random subspace method performs the bootstrapping in the feature space. Specifically, we sample a small subset of features to reduce the discrepancy between the training data size and the feature vector length. Using a random sampling method, we construct a multiple number of SVRs. We then combine these SVRs to construct a more powerful model to solve the overfitting problem as well as reducing the time complexity. In this work, we adopt the radial basis function (RBF) for its simplicity and efficiency. We have trained M ($M = 5$) SVR models, where we randomly choose 60 dimensions of features for each model.

4. Experimental results

A.Databases and performance measures

The proposed method is specifically designed for multiply-distorted images. In this work, we have tested our method on two public multiply-distorted databases, *i.e.*, MLIVE [58], MDID2013

Table 1
Performance comparison with competing IQA methods. The best results are highlighted in bold.

IQA method	Type	MLIVE (450 images)			MDID2013 (324 images)			MDID2016 (1600 images)		
		SRCC	PLCC	RMSE	SRCC	PLCC	RMSE	SRCC	PLCC	RMSE
FR-IQA methods										
PSNR	FR	0.728	0.817	10.870	0.644	0.653	0.035	0.578	0.616	1.735
VSNR	FR	0.828	0.882	8.883	0.635	0.656	0.035	0.659	0.683	1.608
SSIM	FR	0.903	0.927	6.967	0.622	0.656	0.036	0.770	0.791	1.347
IW-SSIM	FR	0.911	0.939	6.636	0.888	0.887	0.023	0.889	0.896	0.975
OSS-SSIM	FR	0.919	0.931	6.682	0.764	0.731	0.034	0.692	0.706	1.558
VIF	FR	0.914	0.932	6.761	0.906	0.915	0.021	**0.930**	**0.937**	**0.771**
MAD	FR	0.895	0.915	7.606	0.857	0.862	0.023	0.724	0.744	1.471
ADM	FR	0.909	0.924	7.050	0.830	0.849	0.025	0.708	0.721	1.5255
FSIM	FR	0.895	0.917	7.039	0.750	0.770	0.031	0.887	0.897	0.974
GMS	FR	0.887	0.914	7.430	0.786	0.804	0.028	0.861	0.877	1.056
IGM	FR	0.888	0.924	7.195	0.878	0.882	0.022	0.855	0.867	1.096
VSI	FR	0.878	0.910	7.655	0.730	0.744	0.031	0.856	0.870	1.085
Pro+S-SVR	NR	**0.952**	**0.956**	**5.552**	**0.923**	**0.935**	**0.017**	**0.888**	**0.893**	**0.993**
Pro+M-SVR	NR	**0.958**	**0.960**	**5.445**	**0.929**	**0.940**	**0.017**	**0.890**	**0.897**	**0.967**
NR-IQA methods										
BIQI	NR	0.884	0.905	7.831	0.863	0.883	0.023	0.627	0.670	1.634
NIQE	NR	0.789	0.858	9.489	0.614	0.645	0.037	0.649	0.670	1.635
ILNIQE	NR	0.901	0.914	7.538	0.707	0.709	0.034	0.689	0.724	1.518
BLIINDS2	NR	0.888	0.904	7.981	0.808	0.844	0.027	0.778	0.796	1.334
DIIVINE	NR	0.866	0.899	8.256	0.836	0.846	0.027	0.551	0.565	1.817
CORNIA	NR	0.901	0.917	7.587	0.898	0.905	0.020	0.783	0.781	1.375
BRISQUE	NR	0.900	0.924	7.143	0.819	0.833	0.027	0.770	0.790	1.351
GMLOG	NR	0.834	0.873	9.165	0.825	0.831	0.026	0.850	0.857	1.132
NFERM	NR	0.898	0.917	7.459	0.855	0.871	0.024	0.451	0.495	1.914
SISBLM	NR	0.907	0.925	7.194	0.886	0.885	0.023	0.655	0.632	1.707
GWH-GLBP	NR	0.941	0.947	5.919	0.903	0.912	0.020	0.885	0.890	1.004
Pro+S-SVR	NR	**0.952**	**0.956**	**5.552**	**0.923**	**0.935**	**0.017**	**0.888**	**0.893**	**0.993**
Pro+M-SVR	NR	**0.958**	**0.960**	**5.445**	**0.929**	**0.940**	**0.017**	**0.890**	**0.897**	**0.967**

[5] and MDID2016 [59]. More information about these databases is shown as below:

- MLIVE consists of two subsets with 15 reference images in total. The first subset contains 225 images distorted by Gaussian blur embedded by JPEG (GB+JPEG). Similarly, the second subset have 225 images distorted by Gaussian blur embedded by white noise (GB+WN). These two subsets add up to 450 distorted images in total. The subjective human ratings are difference mean opinion score (DMOS) in the range of [0,100].
- MDID2013, which is created from 12 reference images, is different from the MLIVE database, since its distorted images are simultaneously distorted by three different kinds of distortions (GB+JPEG+WN). The number of distorted images is 324. Besides, the subjective human ratings in MDID2013 are DMOS in the scale of [0,1].
- MDID2016 contains five common types of distortions, i.e., WN, GB, JPEG, JPEG2000, and contrast change, created from 20 reference images. There are 1600 distorted images in total, and each distorted image is derived from degrading the reference image with random types and random levels of distortions. The subjective human ratings in MDID2016 are mean opinion score (MOS) in the range of [0, 8].

To verify the performance of our method, we compare our method with various IQA methods based on three commonly used performance measures, which are detailed as follows:

- Spearman rank-order correlation coefficient (SRCC) is one of the most widely-used performance measure in IQA task. It can be formulated as

$$\mathrm{SRCC} = 1 - \frac{6}{S(S^2-1)}\sum_{i=1}^{S} d_i^2, \tag{18}$$

where d_i denotes the difference between the objective and subjective ratings, and S denotes the image number in the testing dataset.

Table 2
Performance of the statistical significance test on the SRCC of different metrics across 1000 times. The symbols "1" or "0" indicates that our method is statistically (significance level 0.05) better, or indistinguishable than the corresponding methods.

IQA method	Type	MLIVE (450 images)			MDID2013 (324 images)		
		SRCC	PLCC	RMSE	SRCC	PLCC	RMSE
FR-IQA methods							
PSNR	FR	1	1	1	1	1	1
VSNR	FR	1	1	1	1	1	1
SSIM	FR	1	1	1	1	1	1
IW-SSIM	FR	1	1	1	1	1	1
OSS-SSIM	FR	1	1	1	1	1	1
VIF	FR	1	1	1	1	1	1
MAD	FR	1	1	1	1	1	1
ADM	FR	1	1	1	1	1	1
FSIM	FR	1	1	1	1	1	1
GMS	FR	1	1	1	1	1	1
IGM	FR	1	1	1	1	1	1
VSI	FR	1	1	1	1	1	1
GMSD	FR	1	1	1	1	1	1
NR-IQA methods							
BIQI	NR	1	1	1	1	1	1
NIQE	NR	1	1	1	1	1	1
ILNIQE	NR	1	1	1	1	1	1
BLIINDS2	NR	1	1	1	1	1	1
DIIVINE	NR	1	1	1	1	1	1
CORNIA	NR	1	1	1	1	1	1
BRISQUE	NR	1	1	1	1	1	1
GMLOG	NR	1	1	1	1	1	1
NFERM	NR	1	1	1	1	1	1
SISBLM	NR	1	1	1	1	1	1
GWH-GLBP	NR	1	1	1	1	1	1

- Pearson linear correlation coefficient (PLCC) is another important performance measure, which is calculated between subjective and objective evaluations after the nonlinear regression of

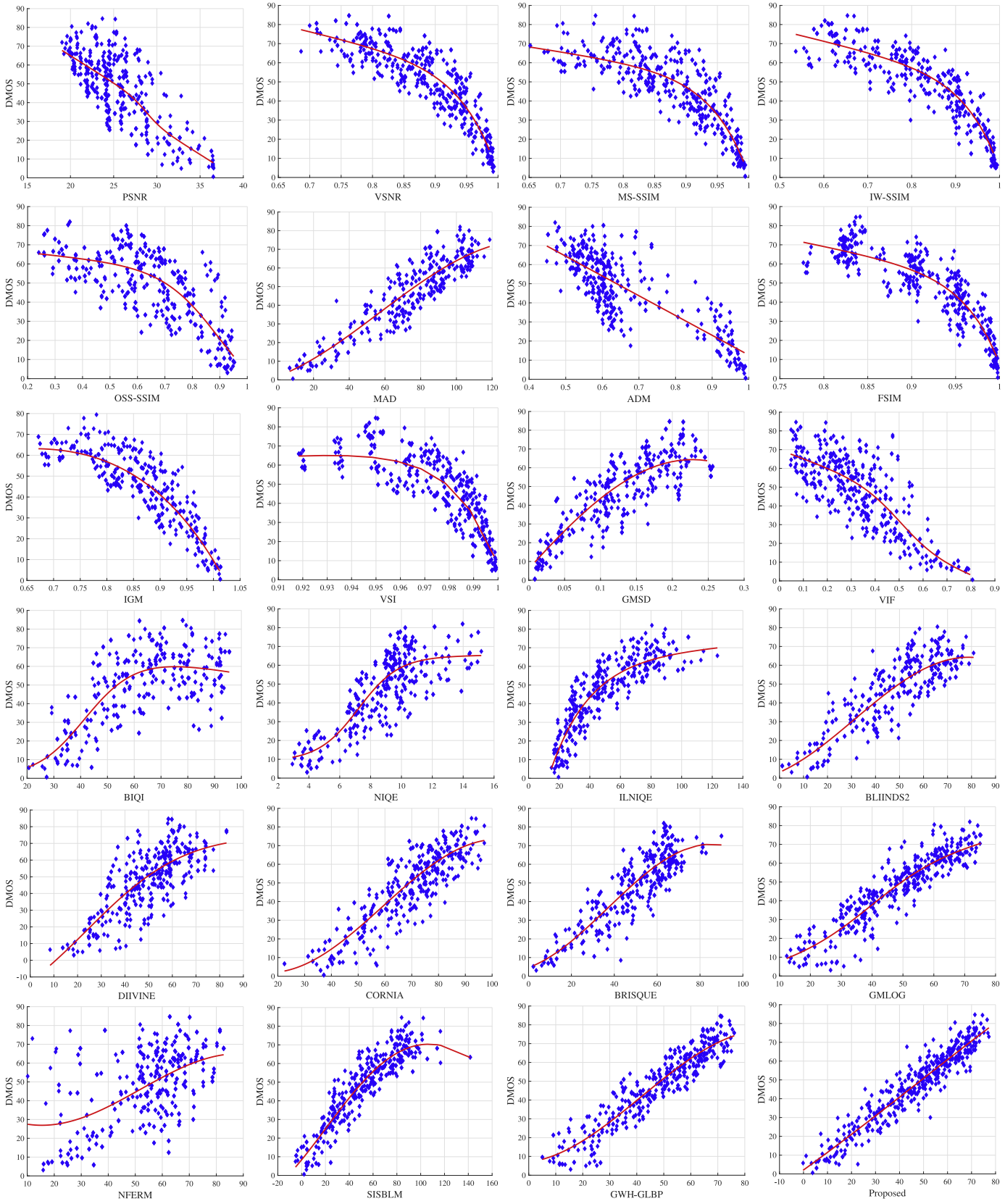


Fig. 6. Scatter plots of DMOS versus classical FR-IQA and NR-IQA metrics on MLIVE. (From left to right) The first three rows show the scatter plots of DMOS versus classical FR-IQA metrics including PSNR, VSNR, MSS-SSIM, IW-SSIM, OSS-SSIM, MAD, ADM, FSIM, IGM, VSI, and GMSD, and the proposed method. (From left to right) The last three rows show the scatter plots of DMOS versus classical NR-IQA methods including BIQI, NIQE, ILNIQE, BLIINDS2, DIIVINE, CORNIA, BRISQUE, GMLOG, NFERM, SISBLM, GWH-GLBP and the proposed method. The red lines denote the fitted curves by logistic function.

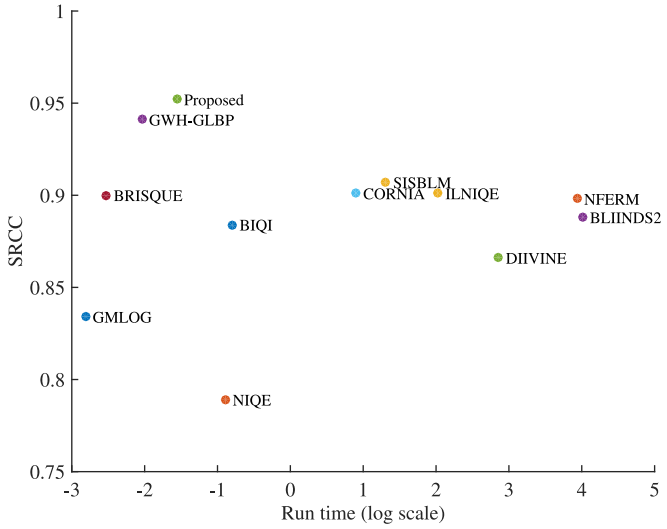


Fig. 7. SRCC of MLIVE versus running time (log scale) of various NR-IQA methods.

(21). It can be formulated as

$$\mathrm{PLCC} = \frac{\sum_i (q_i - \bar{q})(o_i - \bar{o})}{\sqrt{\sum_i (q_i - \bar{q})^2 (o_i - \bar{o})^2}}, \tag{19}$$

where o_i and $\bar{o}$ denote the i-th image's subjective score and the average of all o_i, respectively. Similarly, q_i and $\bar{q}$ denote the mapping objective scores after the nonlinear regression of (21) and the average of all q_i, respectively.

- Another important performance measure we use is root mean square error (RMSE), which measures the difference between the subjective evaluations and the mapping objective evaluations after the nonlinear mapping of (21):

$$\mathrm{RMSE} = \sqrt{\frac{1}{S}\sum (q_i - o_i)^2}. \tag{20}$$

For the above PLCC and RMSE metrics, we apply the monotonic logistic mapping as the nonlinear mapping, which can be written as

$$Q(x) = \beta_1 \left(\frac{1}{2} - \frac{1}{\exp(\beta_2 (x - \beta_3) + \beta_4 x + \beta_5)} \right), \tag{21}$$

where $\beta_i (i = 1, \ldots, 5)$ are model parameters obtained by the curve fitting process per database, $Q(x)$ represents the fitted IQA evaluation, and x denotes the original IQA evaluation.

For SVR learning on each database, distorted images of four-fifths of the reference images are selected for training, while the remaining are adopted for testing. Such trial is repeated 1000 times, and the final result is reported as the median performance.

B. Experimental setup

For the computation of the $\mathrm{LBP}^{\mathrm{riu2}}_{P,R}$ and GCS-LBP operators, P is set to 8, and R is set to 1, thus producing 10 and 16 histogram bins at each scale. The proposed quality-aware feature sets are then extracted at 3 scales. Thus, the obtained feature set consists of 78 dimensions in total. The threshold T is the main parameter for our method. In experiments, we have performed our method with different values of T in the range of [0.1, 0.3], and the performance of our method is relatively stable. Thus we set T to 0.2 in experiments. With the whole feature set, we have trained one SVR model (denoted by S-SVR). With random subspace method, we have trained 5 SVRs (denoted by M-SVR), where we randomly choose 60 dimensions of features for each SVR, and take the average results of multiple SVRs as final output.

C. Performance analysis

We have compared the proposed method with the prominent FR-IQA metrics including PSNR, VSNR [60], MS-SSIM [22], IW-SSIM [9], OSS-SSIM [10], VIF [23], MAD [24], ADM [25], FSIM [6], GMS [7], IGM [61], VSI [62], GMSD [8]. Besides, our method is also compared with the prominent NR-IQA metrics, including BIQI [63], NIQE [34], ILNIQE [35], BLIINDS2 [37], DIIVINE [38], CORNIA [45], BRISQUE [36], GMLOG [64], NFERM [44], SISBLM [5], GWH-GLBP [4]. The source codes of these IQA metrics are derived from their original authors. All the results of these methods are reported in Table 1.

Table 1 illustrates the results of SRCC, PLCC and RMSE on MLIVE, MDID2013, and MDID2016 databases. From Table 1, we have the following observations:

- Among FR-IQA methods, it is seen that the best two FR-IQA metrics on the MLIVE, MDID2013, MDID2016 databases are VIF and IW-SSIM. From Table 1, we can also observe that our method obtains better performance than VIF and IW-SSIM on MLIVE and MDID2013 databases.
- Among NR-IQA methods, it is observed that the top two on MLIVE, MDID2013, and MDID2016 are our method and GWH-GLBP. This is mainly because our method and GWH-GLBP are specifically designed for evaluating images simultaneously distorted with multiple distortions, while others belong to general-purpose NR-IQA methods.
- The results of all methods on MDID2013 are inferior to their corresponding results on MLIVE. This is because that the distorted images in MDID2013 are simultaneously distorted by three different kinds of distortions, while the distorted images in MLIVE are simultaneously distorted by two different kinds of distortions.
- With random subspace method, the performance of the proposed method can be further improved. In summary, the proposed method obtains competitive results with the prevailing FR-IQA methods, and achieves better performance than other NR-IQA methods, which demonstrates the effectiveness of the proposed method.
  metrics on multiple distortions. Also, considering SRCC, PLCC and RMSE, our method outperforms all the compared NR-IQA metrics. Despite the wonderful performance of GWH-GLBP by extracting the first-order image structures, it is still slightly inferior to the proposed method. The reason is that our method extracts more structural information including both the first- and high-order image structures.

In order to evaluate statistical significance, we perform two sample T-test (significance level 0.05) between SRCC by the compared NR-IQA methods across 1000 trials. The T-test results are shown in Table 2, in which the symbol "1" or "0" indicates that the proposed method (Pro-M-SVR) is statistically better, or indistinguishable than the corresponding IQA methods. From Table 2 we can observe that our method performs statistically better than all the compared FR- and NR-IQA metrics.

Finally, we also display the scatter plots of DMOS versus objective quality evaluations of the prominent FR- and NR-IQA methods on MLIVE and MDID2013 databases, as illustrated in Fig. 6. Note that a good IQA algorithm should predict the perceptual quality consistently across subjective DMOS. From Fig. 6 we can see that the scatter plots of our method is more concentrated, and thus has better consistency than the prevailing FR- and NR-IQA metrics, which demonstrates the effectiveness of our method.

D. Computational complexity

For most of NR-IQA methods including our method, the core of them lies in feature extraction. For a fair comparison, we have compared the time cost of feature extraction of different NR-IQA

methods, as shown in Fig. 7. For all the methods, the experiments are conducted in Matlab with an Inter Core i5-3470 CPU @3.2 GHz. A scatter plot of SRCC across MLIVE dataset versus the average running time on the images with size of 768×512 is shown in Fig. 7, from which we can see that our method obtains high quality prediction accuracy with relatively low computational complexity.

5. Conclusions

In this paper, we have proposed a novel referenceless metric to evaluate the perceptual quality of multiply-distorted images based on structural degradation, inspired by the fact that image structures are crucial for visual quality perception. Specifically, our metric considers both the first- and high-oder image structures. To better exploit such image structures, we have used two kind of LBP operators to extract the first- and high-order structural information, from gradient-magnitude and contrast-normalized maps, respectively. In order to avoid overfitting, we use random subspace method in the feature space, and randomly choose portion of features for training, thus without adding the time complexity. Experiments show that the proposed method presents better prediction performance in comparison with other known FR- and NR-IQA metrics on public multiply-distorted image databases.

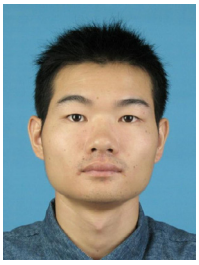
Acknowledgements

This work is supported by the National Natural Science Foundation of China under Grant nos. 61771273, the R&D Program of Shenzhen under grant Nos. JCYJ20170307153635551, JCYJ20160513103916577, and the Guangdong Special Support Program (Grant no. 2015TQ01X161).

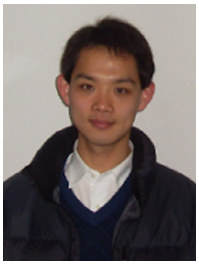
References

[1] W. Lin, C.-C. J. Kuo, Perceptual visual quality metrics: a survey, J. Vis. Commun. Image Represent. 22 (4) (2011) 297–312.
[2] D.M. Chandler, Seven challenges in image quality assessment: past, present, and future research, in: Proceedings of Signal Processing, ISRN, 2013, 2013.
[3] Y. Lu, F. Xie, T. Liu, Z. Jiang, D. Tao, No reference quality assessment for multiply-distorted images based on an improved bag-of-words model, IEEE Signal Process. Lett. 22 (10) (2015) 1811–1815.
[4] Q. Li, W. Lin, Y. Fang, No-reference quality assessment for multiply-distorted images in gradient domain, IEEE Signal Process. Lett. 23 (4) (2016) 541–545.
[5] K. Gu, G. Zhai, X. Yang, W. Zhang, Hybrid no-reference quality metric for singly and multiply distorted images, IEEE Trans. Broadcast. 60 (3) (2014) 555–567.
[6] L. Zhang, L. Zhang, X. Mou, D. Zhang, FSIM: a feature similarity index for image quality assessment, IEEE Trans. Image Process. 20 (8) (2011) 2378–2386.
[7] A. Liu, W. Lin, M. Narwaria, Image quality assessment based on gradient similarity, IEEE Trans. Image Process. 21 (4) (2012) 1500–1512.
[8] W. Xue, L. Zhang, X. Mou, A.C. Bovik, Gradient magnitude similarity deviation: a highly efficient perceptual image quality index, IEEE Trans. Image Process. 23 (2) (2014) 684–695.
[9] Z. Wang, Q. Li, Information content weighting for perceptual image quality assessment, IEEE Trans. Image Process. 20 (5) (2011) 1185–1198.
[10] K. Gu, M. Liu, G. Zhai, X. Yang, W. Zhang, Quality assessment considering viewing distance and image resolution, IEEE Trans. Broadcast. 61 (3) (2015) 520–531.
[11] H. Hadizadeh, I.V. Bajić, Color Gaussian jet features for no-reference quality assessment of multiply-distorted images, IEEE Signal Process. Lett. 23 (12) (2016) 1717–1721.
[12] J. Larsson, M.S. Landy, D.J. Heeger, Orientation-selective adaptation to first-and second-order patterns in human visual cortex, J. Neurophysiol. 95 (2) (2006) 862–881.
[13] T.K. Ho, The random subspace method for constructing decision forests, IEEE Trans. Pattern Anal. Mach. Intell. 20 (8) (1998) 832–844.
[14] K. Gu, D. Tao, J.F. Qiao, W. Lin, Learning a no-reference quality assessment model of enhanced images with big data, IEEE Trans. Neural Netw. Learn. Syst. PP (99) (2017) 1–13.
[15] T. Dai, W. Lu, W. Wang, J. Wang, S.-T. Xia, Entropy-based bilateral filtering with a new range kernel, Signal Process. 137 (2017) 223–234.
[16] H. Liang, D.S. Weller, Comparison-based image quality assessment for selecting image restoration parameters, IEEE Trans. Image Process. 25 (11) (2016) 5118–5130.
[17] Z. Zha, X. Liu, Z. Zhou, X. Huang, J. Shi, Z. Shang, L. Tang, Y. Bai, Q. Wang, X. Zhang, Image denoising via group sparsity residual constraint, in: Proceedings of 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2017, pp. 1787–1791.
[18] T. Dai, Z. Xu, H. Liang, K. Gu, Q. Tang, Y. Wang, W. Lu, S.-T. Xia, A generic denoising framework via guided principal component analysis, J. Vis. Commun. Image Represent. 48 (Supplement C) (2017) 340–352.
[19] Z. Guo, Q. Li, L. Zhang, J. You, D. Zhang, W. Liu, Is local dominant orientation necessary for the classification of rotation invariant texture? Neurocomputing 116 (2013) 182–191.
[20] B. Lei, S.A. Rahman, I. Song, Content-based classification of breath sound with enhanced features, Neurocomputing 141 (2014) 139–147.
[21] J. Qian, D. Wu, L. Li, D. Cheng, X. Wang, Image quality assessment based on multi-scale representation of structure, Digit. Signal Process. 33 (2014) 125–133.
[22] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE Trans. Image Process. 13 (4) (2004) 600–612.
[23] H.R. Sheikh, A.C. Bovik, Image information and visual quality, IEEE Trans. Image Process. 15 (2) (2006) 430–444.
[24] E.C. Larson, D.M. Chandler, Most apparent distortion: full-reference image quality assessment and the role of strategy, J. Electron. Imaging 19 (1) (2010). 011006–011006.
[25] S. Li, F. Zhang, L. Ma, K.N. Ngan, Image quality assessment by separately evaluating detail losses and additive impairments, IEEE Trans. Multimed. 13 (5) (2011) 935–949.
[26] L. Ma, S. Li, F. Zhang, K.N. Ngan, Reduced-reference image quality assessment using reorganized DCT-based image representation, IEEE Trans. Multimed. 13 (4) (2011) 824–829.
[27] J. Wu, W. Lin, G. Shi, A. Liu, Reduced-reference image quality assessment with visual information fidelity, IEEE Trans. Multimed. 15 (7) (2013) 1700–1705.
[28] N.D. Narvekar, L.J. Karam, A no-reference image blur metric based on the cumulative probability of blur detection (CPBD), IEEE Trans. Image Process. 20 (9) (2011) 2678–2683.
[29] L. Li, Y. Zhou, W. Lin, J. Wu, X. Zhang, B. Chen, No-reference quality assessment of deblocked images, Neurocomputing 177 (2016) 572–584.
[30] T. Zhu, L. Karam, A no-reference objective image quality metric based on perceptually weighted local noise, EURASIP J. Image Video Process. 2014 (1) (2014) 5.
[31] H.R. Sheikh, A.C. Bovik, L. Cormack, No-reference quality assessment using natural scene statistics: JPEG2000, IEEE Trans. Image Process. 14 (11) (2005) 1918–1927.
[32] L. Li, W. Lin, H. Zhu, Learning structural regularity for evaluating blocking artifacts in JPEG images, IEEE Signal Process. Lett. 21 (8) (2014) 918–922.
[33] K. Gu, W. Lin, G. Zhai, X. Yang, W. Zhang, C.W. Chen, No-reference quality metric of contrast-distorted images basedon information maximization, IEEE Trans. Cybern. 47 (12) (2017) 4559–4565.
[34] A. Mittal, R. Soundararajan, A.C. Bovik, Making a æcompletely blindg image quality analyzer, IEEE Signal Process. Lett. 20 (3) (2013) 209–212.
[35] L. Zhang, L. Zhang, A.C. Bovik, A feature-enriched completely blind image quality evaluator, IEEE Trans. Image Process. 24 (8) (2015) 2579–2591.
[36] A. Mittal, A.K. Moorthy, A.C. Bovik, No-reference image quality assessment in the spatial domain, IEEE Trans. Image Process. 21 (12) (2012) 4695–4708.
[37] M.A. Saad, A.C. Bovik, C. Charrier, Blind image quality assessment: a natural scene statistics approach in the dct domain, IEEE Trans. Image Process. 21 (8) (2012) 3339–3352.
[38] A.K. Moorthy, A.C. Bovik, Blind image quality assessment: from natural scene statistics to perceptual quality, IEEE Trans. Image Process. 20 (12) (2011) 3350–3364.
[39] J. Shen, Q. Li, G. Erlebacher, Hybrid no-reference natural image quality assessment of noisy, blurry, jpeg2000, and jpeg images, IEEE Trans. Image Process. 20 (8) (2011) 2089–2098.
[40] W. Xue, X. Mou, L. Zhang, A.C. Bovik, X. Feng, Blind image quality assessment using joint statistics of gradient magnitude and Laplacian features, IEEE Trans. Image Process. 23 (11) (2014) 4850–4862.
[41] Q. Li, W. Lin, Y. Fang, Bsd: Blind image quality assessment based on structural degradation, Neurocomputing 236 (2017) 93–103.
[42] J. Xu, P. Ye, Q. Li, H. Du, Y. Liu, D. Doermann, Blind image quality assessment based on high order statistics aggregation, IEEE Trans. Image Process. 25 (9) (2016) 4444–4457.
[43] Q. Li, W. Lin, Y. Fang, No-reference image quality assessment based on high order derivatives, in: Proceedings of 2016 IEEE International Conference on Multimedia and Expo (ICME), IEEE, 2016, pp. 1–6.
[44] K. Gu, G. Zhai, X. Yang, W. Zhang, Using free energy principle for blind image quality assessment, IEEE Trans. Multimed. 17 (1) (2015) 50–63.
[45] P. Ye, J. Kumar, L. Kang, D. Doermann, Unsupervised feature learning framework for no-reference image quality assessment, in: Proceedings of 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2012, pp. 1098–1105.
[46] L. Kang, P. Ye, Y. Li, D. Doermann, Convolutional neural networks for no-reference image quality assessment, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 1733–1740.
[47] R.A. Manap, L. Shao, Non-distortion-specific no-reference image quality assessment: a survey, Inf. Sci. 301 (2015) 141–160.
[48] E.H. Adelson, Lightnessperception and lightness illusions, The New Cognitive Neurosciences, 2nd Ed., 2000.
[49] K. Gu, L. Li, H. Lu, X. Min, W. Lin, A fast reliable image quality predictor by fusing micro- and macro-structures, IEEE Trans. Ind.l Electron. 64 (5) (2017) 3903–3912.

[50] T. Ojala, M. Pietikäinen, T. Mäenpää, Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, IEEE Trans. Pattern Anal. Mach. Intell. 24 (7) (2002) 971–987.
[51] Z. Guo, L. Zhang, D. Zhang, Rotation invariant texture classification using LBP variance (LBPV) with global matching, Pattern Recogn. 43 (3) (2010a) 706–719.
[52] Z. Guo, L. Zhang, D. Zhang, A completed modeling of local binary pattern operator for texture classification, IEEE Trans. Image Process. 19 (6) (2010b) 1657–1663.
[53] S. Lyu, E.P. Simoncelli, Nonlinear image representation using divisive normalization, in: Proceedings of 2008 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2008, pp. 1–8.
[54] J. Rabin, J. Delon, Y. Gousseau, Removing artefacts from color and contrast modifications, IEEE Trans. Image Process. 20 (11) (2011) 3073–3085.
[55] H.C. Hughes, G. Nozawa, F. Kitterle, Global precedence, spatial frequency channels, and the statistics of natural images, J. Cogn. Neurosci. 8 (3) (1996) 197–230.
[56] K. Gu, G. Zhai, X. Yang, W. Zhang, Self-adaptive scale transform for IQA metric, in: Proceedings of 2013 IEEE International Symposium on Circuits and Systems (ISCAS), IEEE, 2013, pp. 2365–2368.
[57] B. Schölkopf, A.J. Smola, Learning with kernels: Support vector machines, regularization, optimization, and beyond, MIT press, 2002.
[58] D. Jayaraman, A. Mittal, A.K. Moorthy, A.C. Bovik, Objective quality assessment of multiply distorted images, in: Proceedings of 2012 Conference Record of the Forty Sixth Conference on Signals, Systems and Computers (ASILOMAR), IEEE, 2012, pp. 1693–1697.
[59] W. Sun, F. Zhou, Q. Liao, MDID: a multiply distorted image database for image quality assessment, Pattern Recogn. 61 (2017) 153–168.
[60] D.M. Chandler, S.S. Hemami, VSNR: a wavelet-based visual signal-to-noise ratio for natural images, IEEE Trans. Image Process. 16 (9) (2007) 2284–2298.
[61] J. Wu, W. Lin, G. Shi, A. Liu, Perceptual quality metric with internal generative mechanism, IEEE Trans. Image Process. 22 (1) (2013) 43–54.
[62] L. Zhang, Y. Shen, H. Li, Vsi: A visual saliency-induced index for perceptual image quality assessment, IEEE Trans. Image Process. 23 (10) (2014) 4270–4281.
[63] A.K. Moorthy, A.C. Bovik, A two-step framework for constructing blind image quality indices, IEEE Signal Process. Lett. 17 (5) (2010) 513–516.
[64] M. Zhang, C. Muramatsu, X. Zhou, T. Hara, H. Fujita, Blind image quality assessment using the joint statistics of generalized local binary pattern, IEEE Signal Process. Lett. 22 (2) (2015) 207–210.



Tao Dai received the B.S. degree in Electronic Engineering from Xidian University, China, in 2014. He is currently pursuing the Ph.D. degree in the Department of Computer Science and Technology of Tsinghua University. His research interests are in the domains of sparse representation, machine learning, image processing and computer vision (image restoration and image quality assessment).



Ke Gu received the B.S. and PhD degrees in electronic engineering from Shanghai Jiao Tong University, Shanghai, China, in 2009 and 2015. He is currently the associated editor for IEEE Access, and is the reviewer for 20 SCI journals. His research interests include quality perception and image processing. Dr. Gu received the Best Paper Award at the IEEE International Conference on Multimedia and Expo (ICME) in 2016, and received the excellent Ph.D. thesis award from the Chinese Institute of Electronics (CIE) in 2016. He is the leading special session organizer in VCIP2016 and ICIP2017.

Li Niu is currently a postdoc associate with department of electrical and computer engineering in Rice University, TX, U.S. Prior to this, he received his B.E. degree from the University of Science and Technology of China, Hefei, China, in 2011, and the Ph.D. degree with the Interdisciplinary Graduate School, Nanyang Technological University, Singapore, in 2017. He is with Beijing Key Laboratory of Computational Intelligence and Intelligent System, Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China. His current research interests include machine learning and computer vision.

Yongbing Zhang received the B.A. degree in English and the M.S. and Ph.D. degrees in computer science from the Harbin Institute of Technology, Harbin, China, in 2004, 2006, and 2010, respectively.He is currently an Associate Research Fellow with Tsinghua University, Shenzhen, China. His current research interests include video processing, image and video coding, video streaming, and transmission.Mr. Zhang was the receipient of the Best Student Paper Award at the IEEE International Conference on Visual Communication and Image Processing in 2015.

Weizhi Lu received the B.E. and M.E. degrees from Shandong University, China, in 2007 and 2010, respectively, and the Ph.D. degree from the National Institute of Applied Sciences (INSA), Rennes, France, in 2014, all in electronic engineering. He is currently with the Graduate School at Shenzhen of Tsinghua University as a postdoctoral researcher. His research interests include compressed sensing and machine learning.

Shu-Tao Xia received the B. S. degree in mathematics and the Ph.D. degree in applied mathematics from Nankai University, Tianjin, China, in 1992 and 1997, respectively. Since January 2004, he has been with the Graduate School at Shenzhen of Tsinghua University, Guangdong, China. He is now a full professor there. From March 1997 to April 1999, he was with the research group of information theory, Department of Mathematics, Nankai University, Tianjin, China. From September 1997 to March 1998 and from August to September 1998, he visited the Department of Information Engineering, The Chinese University of Hong Kong, Hong Kong. His current research interests include coding and information theory, networking, and machine learning.