Blind Image Quality Assessment Using Statistical Structural and Luminance Features

Qiaohong Li, Weisi Lin, *Fellow, IEEE*, Jingtao Xu, *Member, IEEE*, and Yuming Fang

***Abstract*—Blind image quality assessment (BIQA) aims to develop quantitative measures to automatically and accurately estimate perceptual image quality without any prior information about the reference image. In this paper, we introduce a novel BIQA metric by structural and luminance information, based on the characteristics of human visual perception for distorted image. We extract the perceptual structural features of distorted image by the local binary pattern distribution. Besides, the distribution of normalized luminance magnitudes is extracted to represent the luminance changes in distorted image. After extracting the features for structures and luminance, support vector regression is adopted to model the complex nonlinear relationship from feature space to quality measure. The proposed BIQA model is called no-reference quality assessment using statistical structural and luminance features (NRSL). Extensive experiments conducted on four synthetically distorted image databases and three naturally distorted image databases have demonstrated that the proposed NRSL metric compares favorably with the relevant state-of-the-art BIQA models in terms of high correlation with human subjective ratings. The MATLAB source code and validation results of NRSL are publicly online at http://www.ntu.edu.sg/home/wslin/Publications.htm.**

***Index Terms*—Blind image quality assessment (BIQA), human visual system (HVS), no-reference (NR), structural distortion.**

I. Introduction

DIGITAL images are subject to a broad spectrum of distortions during the process of acquisition, compression, transmission and reproduction. Therefore, it is essential to guarantee the quality of image content for end-users. Subjective viewing test is a natural way to evaluate visual image quality. Despite of its high accuracy and reliability, subjective evaluation is cumbersome, expensive, time consuming, and non-reproducible, which makes it difficult to be embedded into practical applications such as real-time quality monitoring and prediction [1], [2].

Manuscript received September 23, 2015; revised April 18, 2016; accepted August 2, 2016. Date of publication August 16, 2016; date of current version November 15, 2016. This work was supported by the Ph.D. Grant from the Institute for Media Innovation, Nanyang Technological University, Singapore. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Klara Nahrstedt. *(Corresponding author: Yuming Fang.)*

Q. Li and W. Lin are with the School of Computer Engineering, Nanyang Technological University, Singapore 639798 (e-mail: qli013@e.ntu.edu.sg; wslin@ntu.edu.sg).

J. Xu is with the School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China (e-mail: xjt@bupt.edu.cn).

Y. Fang is with the School of Information Technology, Jiangxi University of Finance and Economics, Nanchang 330032, China (e-mail: fa0001ng@e.ntu.edu.sg).

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.2016.2601028

Objective image quality assessment (IQA) metrics which predict perceptual image quality using computational models serve as an effective substitute of subjective methods and are desirable in a wide range of image processing and computer vision applications. For example, there are different device parameters for digital cameras which can be tuned according to image quality [3]; image compression algorithms may use quality as the optimization guidance for quantization [4], [5]; image transmission systems can monitor quality and allocate streaming resources accordingly [6], and image recommendation systems can rank photos based on perceptual quality measure [7].

The investigations into objective IQA have led to considerable progress in the development of perceptual IQA metrics. Existing objective IQA methods can be classified as full-reference (FR), reduced-reference (RR) and no-reference (NR) methods based on the available information of original reference images [1], [2]. Most IQA models fall under the category of FR algorithms, where reference images are available for quality prediction. The mean squared error (MSE) and its corresponding peak signal-to-noise ration (PSNR) are the simplest and most widely used FR metrics due to their efficient computation and clear physical meanings. However, they are not in high agreement with human perceptual quality [1]. In the past decades, to overcome the drawbacks of these traditional metrics, various perceptual FR-IQA metrics have been proposed, such as SSIM [8], VIF [9], ADM [10], FSIM [11] and GMS [12]. For RR-IQA, partial information of reference images in the form of extracted features is required for quality evaluation [13], [14]. However, when reference image is not available as in most practical applications (e.g., transmission, denoising, enhancement), NR/Blind IQA (BIQA) would be the only possible solution to seek.

Early BIQA methods are mainly developed to evaluate perceptual quality of images distorted in a specific image processing application, such as Gaussian white noise (WN) [15], JPEG2000 (JP2K) compression [16], Gaussian blur (GB) [17] and contrast change [18]. The design of such distortion-specific BIQA model can be achieved by the priori knowledge about the corresponding distortion characteristics. A NR method for JPEG compressed images has been proposed by calculating between-block differences and in-block activities in spatial domain [19]. The quality of blurred images has been quantified by the average edge spread [20]. In [21], the pixel distortions and edge features are used to characterize the quality of JP2K compressed images. Recently, several statistical descriptors of image luminance distribution are adopted to gauge image contrast distortion [18]. Since distortion-specific BIQA methods are developed for given distortion types, their practical usage to other applications is limited. On the contrary,

general purpose BIQA methods that require no access to both reference image and distortion type can be used in various scenarios.

General purpose BIQA mainly consists of two broad ways. One is distance-based methods, which are also termed as opinion-unaware methods, as there is no need for subjective scores during the model design process. Generally, natural scene statistics (NSS) [22], [23] or perceptual features [23] are extracted from pristine natural images and used to build a clean image model. The quality of distorted image is calculated as the distance between test image features and the built clean image model [22], [23]. In [24], the authors propose to use objective scores obtained by FR-IQA metric as a substitute of subjective scores to train BIQA models. The other category is learning based models. Specifically, quality aware features are extracted from distorted images and then a regression model is learned to map these extracted features to the final quality score. This kind of BIQA methods is also termed as opinion-aware methods since subjective scores are required during the learning process. The choice of quality aware features is the determinant factor to the success of learning based BIQA methods. A good set of NR features should be some perceptually relevant features that are sensitive to a broad range of image distortion types and robust to image content variation. The most popular features adopted by BIQA methods are NSS features. These features are constructed based on the assumption that natural images are statistically regular and distortions tend to break such statistics and make the images unnatural. Thus, image quality can be quantized as image unnaturalness. BIQA methods based on DCT domain NSS [25], spatial domain NSS [22], [26], wavelet domain NSS [27], [28] and hybrid domain NSS (combination of curvelet, DCT and several types of wavelet transform) [29] have been proposed and achieve promising performance in visual quality prediction. Another kind of features used by learning based methods are perceptual features. The gradient and phase congruency features are extracted from distorted image to assess image quality [30]. In [31], a large number of low level features, texture statistics and noise/blur estimates are combined with multiple support vector regression (SVR) to build the LBIQ metric. In [32], the joint statistics of image gradients and Laplacian of Gaussian (LOG) responses are employed as quality aware features. A combination of free energy principle based features and spatial NSS features has shown some impressive results in [33]. The third kind of learning based methods directly use raw local image patches as the input and learn the quality aware representation in an unsupervised [34] or supervised [35] way. Commonly adopted regression modules in learning based methods include neural network [30], SVR [26], [28], [32] and random forest regression [36].

Despite the prominent improvement in development of BIQA methods recently, the performance of existing methods still lags behind that of FR-IQA methods. BIQA is an intrinsically challenging task due to the diverse distortion types and wide range of image content. There is still much room for the performance improvement before they are applied for practical applications. Moreover, most existing BIQA methods suffer from some limitations. As one of the pioneering BIQA methods, BIQI generally gives poor performance for visual quality prediction due to its simplistic feature extraction [27]. In DIIVINE [28], the used large number of features extracted from multi-subbands in wavelet domain slows down the method greatly. Similarly, the block-based GGD fitting in DCT domain makes BLIINDS2 difficult to predict the visual quality in real time [25]. Feature learning based methods, CORNIA [34] and QAF [36], inevitably require a large size codebook and result in high-dimension global quality aware features. Although GMLOG delivers promising performance on singly-distorted image database, its performance deteriorates considerably when tested on multiply-distorted images [37]. The combination of several RR and FR methods adopted by NFERM [33] leads to a too complicated method to be used in practice.

In this study, we develop a BIQA model by investigating the characteristics of human visual perception. Existing studies show that image structures covey the primary visual information of a scene, and the HVS is highly adapted to extract structural information for image perception and understanding [38], [39]. Here, we propose a novel effective structural feature extraction method as the spatial distribution of local binary pattern (LBP) based on the normalized luminance map to characterize the image structural information. Also, studies from neuroscience have demonstrated that the HVS is highly sensitive to luminance change of a scene, which might also cause visible distortion [40], [41]. Based on these findings, we extract a new feature to represent luminance changes in the proposed metric. After extracting the statistical structural and luminance features, we use SVR to model the complex nonlinear relationship between extracted features and human subjective ratings. Experimental results show that NRSL can achieve better performance than the relevant existing methods.

In sum, the main contribution of this study lies in three folds: first, by exploring the characteristics of the HVS, two new perceptual features are extracted to represent the structural information and luminance changes in distorted image, respectively. We demonstrate that complementary information provided by extracted statistical structural and luminance features plays an important role for image quality estimation. Second, we propose a novel perceptual BIQA metric for visual quality prediction. The proposed metric can consistently handle both synthetically-distorted and naturally-distorted images than other BIQA methods. Previous BIQA methods are mostly designed and validated on databases of images with simulated distortions. Their performance deteriorates when tested on images with realistic distortions. By contrast, the proposed method can achieve consistently better performance on both simulated and realistic distortions. Furthermore, the proposed method has the advantages of high prediction accuracy, high generalization ability, low computational complexity and low feature dimension.

The remainder of this paper is organized as follows. In Section II, we describe the proposed NRSL metric based on the analysis of image structural and normalized luminance information. In Section III, we compare our method with relevant state-of-the-art IQA metrics on four synthetically-distorted image

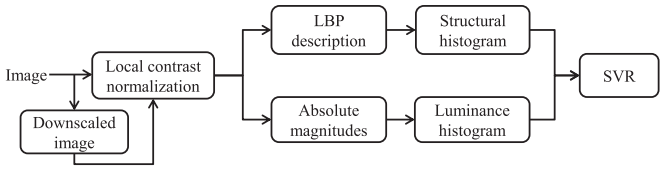


Fig. 1. Block diagram of proposed NRSL model.

databases and three authentically-distorted image databases. Section IV concludes the paper.

II. NRSL FOR BIQA

In this section, we describe the proposed NRSL metric in details. NRSL is designed by accounting for both structural and luminance degradation. Fig. 1 shows the framework of the proposed metric. First, the contrast normalization scheme is applied to the image to remove redundancy in the visual input. Second, the LBP descriptor is employed on the normalized image to extract the structural features. Furthermore, the distribution of normalized luminance magnitudes in the form of histogram is extracted from the same normalized image. Both structural and luminance features are calculated at three scales to account for the variation in viewing distance and image resolution. Finally, the extracted statistical structural and luminance features are combined together as the input to SVR to simulate the nonlinear mapping from feature space to subjective quality scores.

A. Local Contrast Normalization

Local contrast normalization has been used as a preprocessing stage to emulate the nonlinear masking of visual perception in many image processing applicaitons [42]. Generally, each coefficient is divided by the square root of Gaussian weighted combination of the squared amplitudes of its neighbors. In this study, similar preprocessing model as [22], [26] has been adopted

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

where i and j are the spatial indices of the image, and

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

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

are the local mean and standard deviation of the surrounding local patch, where $\omega = \{\omega_{k,l} | k = -K, ..., K, l = -L, ..., L\}$ defines a unit-volume Gaussian window. And the constant C is induced to guarantee numerical stability when the denominators are close to zero. Specifically, we choose

$$C = (\alpha L)^2 \tag{4}$$

where L is the dynamic range of pixel grayscale levels (255 for 8-bit grayscale image), and $\alpha \ll 1$ is a small constant.

B. The Structural Histogram

Natural images are highly structured with the manifestation of strong dependencies among their pixels. When these pixels are spatially neighboring, these high dependencies often carry important information about the structure of visual objects in natural scenes [8]. Image structures covey the primary visual information of a scene, and the HVS adaptively evolves to extract structural information for image perception and understanding [8], [43]. In this work, we extract image structural information as the spatial distribution of local inter-pixel relationship patterns.

After local contrast normalization, we obtain the normalized luminance image $\tilde{I}$. Assuming that the local patch of image $\tilde{I}$ can be characterized by joint distribution of normalized luminance values of successive pixels in this patch

$$T = t(g_c, g_0, g_1, ..., g_{P-1}) \tag{5}$$

where g_c is the normalized luminance value of the center pixel of the local patch, $\{g_0, g_1, ..., g_{P-1}\}$ are the normalized luminance values of the P circularly symmetric neighbourhood.

Without losing information, we substrate the center pixel value from the neighborhood

$$T = t(g_c, g_0 - g_c, g_1 - g_c, ..., g_{P-1} - g_c). \tag{6}$$

Analogous to [44], [45], we assume that the normalized luminance value of center pixel to be statistically independent of the differences between neighbors and center pixel, which allows for the factorization of the joint distribution

$$T \approx t(g_c) t(g_0 - g_c, g_1 - g_c, ..., g_{P-1} - g_c) \tag{7}$$

where the first term $t(g_c)$ is the marginal distribution of normalized luminance coefficients over the whole image $\tilde{I}$. The second term $t(g_0 - g_c, g_1 - g_c, ..., g_{P-1} - g_c)$ is the joint distribution of differences which describes local inter-pixel relationship. By keeping the signs of the differences, Ojala *et al.* [45] proposed the LBP operator to describe this joint distribution

$$t(s(g_0 - g_c), s(g_1 - g_c), ..., s(g_{P-1} - g_c)) \tag{8}$$

where the thresholding function $s(\cdot)$ is defined as

$$s(g_i - g_c) = \begin{cases} 1, & g_i - g_c \geq 0 \\ 0, & g_i - g_c < 0. \end{cases} \tag{9}$$

By assigning a binomial factor (2^p) for each neighbor from (8), the LBP code of one pixel is deduced as [45]

$$LBP_{P,R} = \sum_{i=0}^{P-1} s(g_i - g_c) 2^i \tag{10}$$

where P is the number of neighbors and R is the radius of the neighborhood. Given an image of size $M \times N$, after the LBP code of each pixel is identified, the structural histogram can be

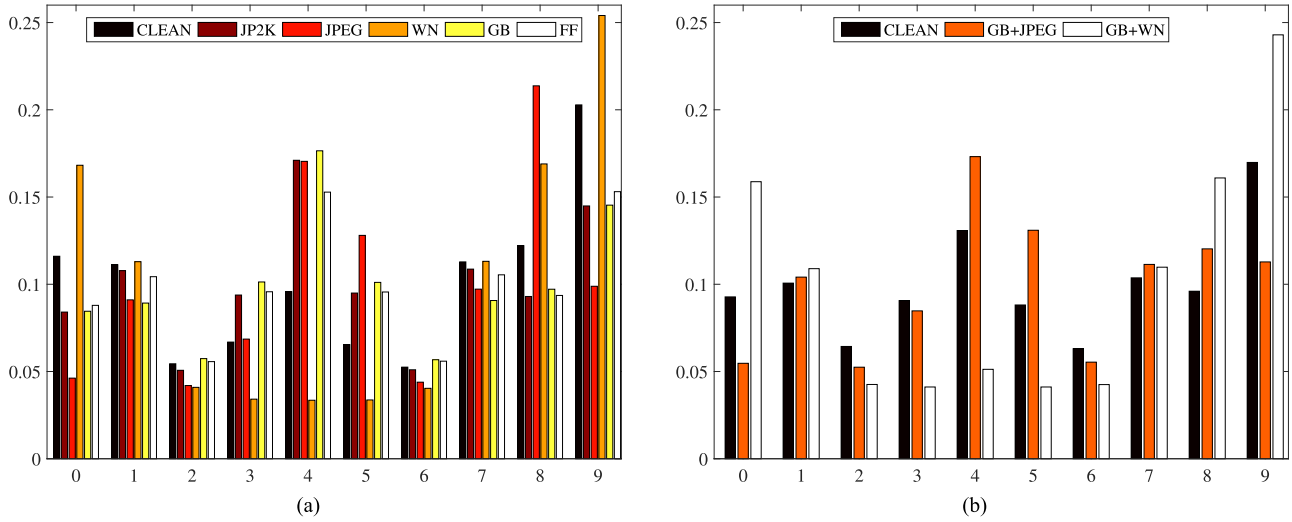


Fig. 2. Impact of distortions on structural histogram. (a) Average structural histograms on LIVE. (b) Average structural histograms on MLIVE.

built to characterize the image structural information

$$SH(k) = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N} f(LBP_{P,R}(i,j),k), k \in [0,K] \tag{11}$$

$$f(x,y) = \begin{cases} 1, & x = y \\ 0, & \text{otherwise} \end{cases} \tag{12}$$

where K is the maximum value of LBP patterns. The uniformity measure $\mathcal{U}$ is calculated as the number of bitwise transitions

$$\mathcal{U}(LBP_{P,R}) = \|s(g_{P-1} - g_c) - s(g_0 - g_c)\| + \sum_{i=0}^{P-1} \|s(g_i - g_c) - s(g_{i-1} - g_c)\|. \tag{13}$$

It was observed that certain LBP patterns can capture the fundamental structural properties, composing of the vast majority of LBP patterns in natural images (sometimes over 90%) [45]. These certain LBPs share with the same attribute: the number of spatial transitions is limited by two ($\mathcal{U} \leq 2$), which are referred as the uniform LBP patterns. To achieve rotation invariance, a locally rotation invariant uniform LBP operator can be defined as

$$LBP_{P,R}^{riu2} = \begin{cases} \sum_{i=0}^{P-1} s(g_i - g_c), & \text{if } \mathcal{U}(LBP_{P,R}) \leq 2 \\ P+1, & \text{else} \end{cases} \tag{14}$$

where superscript $riu2$ refers to the rotation invariant "uniform" patterns with $\mathcal{U}$ value less than 2. The rotation invariant uniform LBP would have $P+2$ distinct patterns ($P+1$ for the different uniform patterns and 1 for the non-uniform pattern).

The LBP patterns effectively detect image primitive microstructures, such as edges, lines, corners and spots. For example, in the case of $LBP_{8,1}$, (0) stands for bright spot, (8) denotes flat area or dark spot, and (1–7) represents edges of varying positive and negative curvature [45]. The introduction of distortion may shift the LBP pattern from one type to another. For example, blocking artifacts caused by JPEG compression may shift a flat pattern to an edge pattern, while blurring effect can change an edge pattern to flat pattern. Gaussian noise may alter the LBP pattern in a random way.

Therefore, we extract the spatial distribution of LBP patterns as an effective feature for quality perception. Since $LBP_{P,R}^{riu2}$ has a fixed set of discrete patterns ($0 \rightarrow P+1$), no quantization is required, but the different patterns are directly mapped into a histogram of $P+2$ bins. By generating the occurrence histogram, we effectively employ the statistical property of the structural information: the LBP operator detects microstructure whose underlying distribution is characterized by the histogram [45]. In this work, we set $P = 8$, thus there would be 10 bins for the structural histogram. Fig. 2(a) shows the average structural histograms of reference images and distorted images with five distortion types in LIVE database [i.e., WN, GB, JPEG, JP2K, and simulated fast fading (FF)]. The average structural histograms are generated by calculating the mean of histograms of all the images with the same distortion type. Fig. 2(b) shows the average structural histograms of reference images and distorted images with two kinds of multiple distortions in MLIVE database [i.e., GB + JPEG (Blur followed by JPEG compression), GB + WN (Blur followed by white Gaussian noise)]. As can be seen in Fig. 2, different distortions alter the structural histograms in their own characteristic ways. And structural histograms are effective to describe the impact of both single distortion and multiple distortions.

C. The Luminance Histogram

The HVS is also highly sensitive to luminance change of an input scene, which might cause visible distortions [12], [46]. Although local contrast normalization smoothes the luminance range in a local patch, the information of global luminance variation can still be manifested by the normalized luminance image $\tilde{I}$. Thus, we extract the luminance feature from $\tilde{I}$.

Existing studies on NSS revealed that the normalized luminance coefficients of natural images follow a Gaussian-like distribution [47]. The studies [22], [26] fit a generalized Gaussian

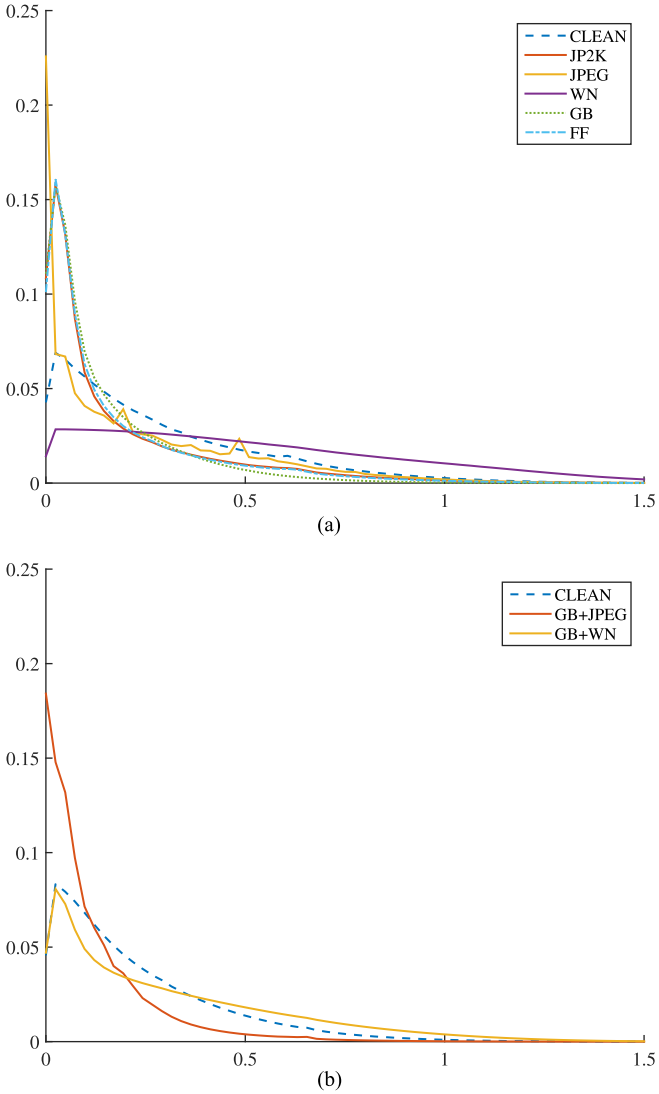


Fig. 3. Impact of distortions on luminance distribution. (a) Average luminance distributions on LIVE. (b) Average luminance distributions on MLIVE.

distribution (GGD) to the coefficients and used the GGD parameters as perceptual features. In this work, we directly employ the histogram to represent the image luminance information. It is advantageous to use the nonparametric histogram instead of the parametric GGD fitting as the quality aware features. First, it frees us from making any, possibly misguided, assumptions about the underlying feature distribution. Second, it constitutes a more accurate representation to avoid any fitting error. Third, it is more computationally efficient to calculate the histogram. Moreover, after local contrast normalization, the resultant normalized luminance coefficients form a symmetric distribution centered at 0. To reduce the histogram range, we take the magnitude operation before calculating the luminance histogram

$$\overline{I}(i,j) = \left|\tilde{I}(i,j)\right| \tag{15}$$

$$LH = \text{hist}(\overline{I}). \tag{16}$$

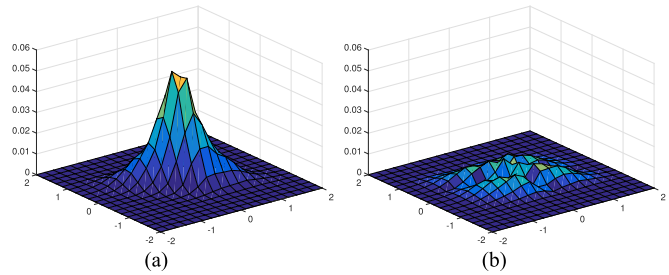


Fig. 4. Validation the independence assumption by approximating the joint distribution by the factorized marginal distribution. (a) The joint distribution $t(g_c, g_0 - g_c)$ which represents the concurrence of horizontally adjacent pixels. (b) The average error produced by factorization of $t(g_c, g_0 - g_c)$ into $t(g_c)$ and $t(g_0 - g_c)$.

Fig. 3(a) shows the average luminance histograms of the reference images and distorted images with five distortion types in LIVE database. Fig. 3(b) shows the average luminance histograms of the reference images and distorted images with two kinds of multiple distortions in MLIVE database. As shown in this figure, different distortion types show apparently different luminance changes. For example, the luminance distribution of WN is more uniform as WN introduces random disturbs, while JP2K/GB/FF make the distribution more Laplacian-like with high peak and small tail as they reduce high frequency components in images. The distribution of JPEG distorted images exhibits several peaks that are caused by the blocking artifacts. The distinct distribution changes caused by different distortion types indicate that the luminance histograms are effective to reflect the quality variation. We also observe that GB + JPEG makes the distribution more peaked as there is blurring effect for both GB and JPEG, while the luminance distribution of GB + WN resembles that of clean images since GB and WN have counter-impact on this distribution. In this case, the structural histogram may be more quality-relevant than the luminance histogram.

D. Analysis of Two Feature Sets

In the previous subsection, we have made the assumption that the marginal distribution of normalized luminance coefficients is statistically independent of the differences between neighbors and center pixel. First, we validate the feasibility of the independence assumption by approximating the joint distribution by the factorized marginal distribution. As a simplification of (7), we consider the case of one neighbor

$$t(g_c, g_0 - g_c) \approx t(g_c)t(g_0 - g_c). \tag{17}$$

Fig. 4(a) shows the average of distributions $t(g_c, g_0 - g_c)$ computed from all the images in LIVE database [48]. Fig. 4(b) shows the average error between $t(g_c, g_0 - g_c)$ and $t(g_c)t(g_0 - g_c)$. As shown in Fig. 4, the average error is relatively small in proportion to the average distribution which confirms the rationality of the independence assumption. Since the structural histogram is derived directly from the joint difference distribution, we can conclude that the structural histogram and luminance histogram are statistically independent.

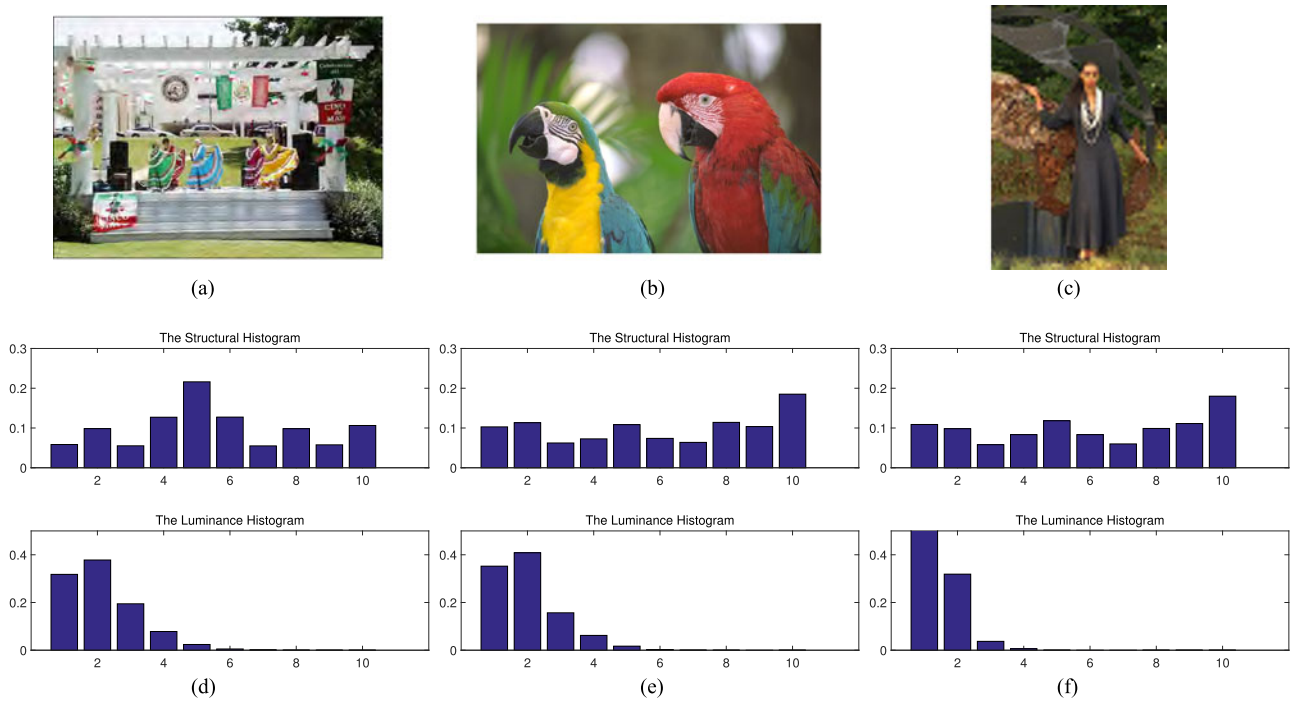


Fig. 5. Image samples to show the complementary of structural and luminance information. (a) Image with DMOS = 68.99. (b) Image with DMOS = 7.23. (c) Image with DMOS = 75.74. (d), (e), and (f) are the structural and luminance histograms of images in the first row.

In Fig. 5, we provide a few concrete examples to show the complementary of structural and luminance features for quality assessment. Fig. 5(a) and (c) are images with low quality, and their differential mean opinion scores (DMOS) are 68.99 and 75.74, respectively. Fig. 5(b) is a high-quality image with DMOS of 7.23. Their structural and luminance histograms are shown in the second row. We compare Fig. 5(d) with Fig. 5(e) and find that these two images share similar luminance distribution, which indicates that the luminance feature cannot distinguish the distortion differences between these two images. By contrast, their structural distributions are quite different and provide valuable information for quality estimation. The ringing and blurring artifacts introduced by JP2K to Fig. 5(a) can be well captured by its structural histogram. The opposite phenomenon is observed when comparing Fig. 5(e) with Fig. 5(f). These two images share the similar structural distribution, which shows the ineffectiveness of structural feature in this case. Conversely, their luminance histograms are disparate enough to discriminate their different quality levels. Hence, we use both structural and luminance features together in this study, since they are mutually complementary for quality evaluation task. The experimental comparison in Section III-B2 also validates this conclusion.

We investigate the quality awareness of NRSL features by performing t-SNE algorithm [49]. t-SNE is an unsupervised nonlinear dimension reduction method that converts high-dimensional data to two-dimensional data that can be visualized in a scatterplot. It can well preserve the significant structure of high-dimensional data, such that similar data points in the high-dimensional feature space tend to lie close together in the low-dimensional embedding. Fig. 6 shows the 2D embedding of extracted features of NRSL, where each dot represents one image

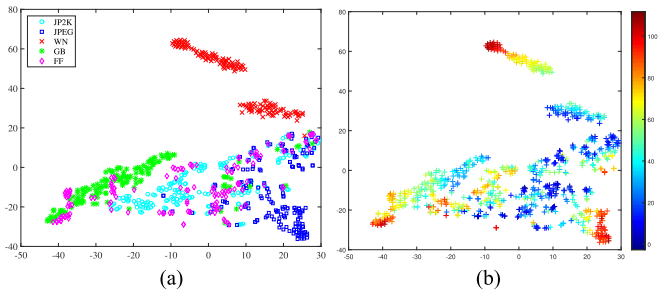


Fig. 6. t-SNE 2D embedding of the NRSL feature space on LIVE database. (a) is color-coded by distortion type. (b) is color-coded by DMOS range.

in the LIVE database, and x- and y-axis denote their coordinates in the 2D plane. Fig. 6(a) is color coded by distortion type, and Fig. 6(b) is color coded by DMOS range. As shown in Fig. 6(a), WN and JPEG are well grouped and separated from other distortions. By contrast, images corrupted by JP2K, GB, and FF tend to mix together. This seems reasonable as both JP2K and GB introduce the blurring artifacts to images, and FF is actually a multiple distortion of the combination of JP2K and transmission errors. The same or similar artifacts introduced by the three distortion types make them cluster together in the feature space. Although it is not a necessary requirement for good BIQA feature to handle the distortion classification problem, Fig. 6(a) shows that the proposed feature set can also be used to identify WN and JPEG distortion types. From Fig. 6(b), we can see that images with similar DMOS values are clustered together, no matter they are from one specific distortion type or across different distortion types. Such observation indicates that the proposed features are valuable for both distortion-specific and general purpose BIQA tasks.

E. Regression Model for Quality Prediction

SVR is widely adopted to learn the mapping function for feature pooling from the feature space to quality measure [26], [32], [50]. Considering a set of training data $\{(\boldsymbol{x}_1, y_1), ..., (\boldsymbol{x}_l, y_l)\}$, where $\boldsymbol{x}_i \in R^n$ is the extracted quality aware feature and y_i is the corresponding DMOS. Given parameters $C > 0$ and $\epsilon > 0$, the standard form of SVR is represented as [51]

$$\min_{\boldsymbol{w},b,\boldsymbol{\xi},\boldsymbol{\xi}^*} \quad \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} + C\left\{\sum_{i=1}^{l}\xi_i + \sum_{i=1}^{l}{\xi_i}^*\right\} \tag{18}$$

$$\text{subject to} \quad \boldsymbol{w}^T\phi(\boldsymbol{x}_i) + b - y_i \leq \epsilon + \xi_i \tag{19}$$

$$y_i - \boldsymbol{w}^T\phi(\boldsymbol{x}_i) - b \leq \epsilon + \xi_i^* \tag{20}$$

$$\xi_i, \xi_i^* \geq 0, i = 1, ..., l \tag{21}$$

where $K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \phi(\boldsymbol{x}_i)^T\phi(\boldsymbol{x}_j)$ is the kernel function. We use the radial basis function (RBF) kernel with the kernel function of $K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp(-\gamma\left\|\boldsymbol{x}_i - \boldsymbol{x}_j\right\|^2)$ in this work.

III. Experimental Results and Discussions

A. Experiment Protocol

1) Database Description: The comparison experiments are conducted on seven subjective image databases. Four of them are traditional synthetically-distorted image databases (i.e., LIVE [48], CSIQ [52], TID2013 [53], MLIVE [54]), which are generated by introducing graded simulated distortions onto high-quality photographs. The most commonly simulated distortions are JP2K, JPEG, WN and GB. LIVE, CSIQ and TID2013 mostly contain images corrupted by a single type of distortion. MLIVE contains images simultaneously distorted by two types of distortions. The remaining three databases are the newly emerging naturally-distorted image databases, which contain realistic distortions generated during image acquisition, processing and storage of devices (i.e., BID [55], CLIVE [56] and CID2013 [57]). For the naturally-distorted images, there are no reference images with perfect quality, and FR/RR methods are not applicable in this scenario. Also there is no definite distortion category for each image since authentic distortions are generally diverse, mixed and multipartite. These databases are summarized as follows.

1) The LIVE database includes 29 reference images and 779 distorted images corrupted by five types of distortions: JP2K, JPEG, WN, GB, and FF. Subjective quality scores are provided in the form of DMOS ranging from 0 to 100.
2) The CSIQ database includes 30 reference images and 866 distorted images corrupted by six types of distortions: JPEG, JP2K, WN, GB, pink Gaussian noise (PGN), and global contrast decrements (CTD). Subjective quality scores are provided in the form of DMOS ranging from 0 to 1.
3) The TID2013 database includes 25 reference images and 3000 distorted images corrupted by 24 types of distortions: #01 WN, #02 WN in color components, #03 spatially correlated WN, #04 masked noise, #05 high-frequency noise, #06 impulse noise, #07 quantization noise, #08 GB, #09 image denoising, #10 JPEG, #11 JP2K, #12 JPEG transmission errors, #13 JP2K transmission errors, #14 non eccentricity pattern noise, #15 local blockwise distortion of different intensity, #16 mean shift, #17 contrast change, #18 change of color saturation, #19 multiplicative Gaussian noise, #20 comfort noise, #21 lossy compression of noisy images, #22 image color quantization with dither, #23 chromatic aberrations and #24 sparse sampling and reconstruction. Subjective quality scores are provided in the form of mean opinion score (MOS) ranging from 0 to 9.
4) The MLIVE database includes 15 reference images and 450 distorted images corrupted by two types of multiple distortions: GB + JPEG and GB + WN. Subjective quality scores are provided in the form of DMOS ranging from 0 to 100.
5) The BID (realistic blur image database) database includes 586 blurred images taken under a variety of lighting conditions and exposure time. It presents typical blurring scenarios in practical applications, such as out-of-focus, simple motion, complex motion and their combination. Subjective quality scores are provided in the form of MOS ranging from 0 to 5.
6) The CLIVE (LIVE in the wild image quality challenge database) database includes 1162 distorted images captured using typical real-world mobile cameras. It contains various authentic image distortions, such as low-light blur and noise, motion blur, overexposure, underexposure, compression errors and their combination. Subjective quality scores are provided in the form of MOS ranging from 0 to 100.
7) The CID2013 database includes 474 distorted images captured using various digital cameras. The database consists of six sets (I–VI) and each set covers around 80 images captured under six typical scenes. It represents the typical distortions that depend on camera sensor type, optics and built in image processing pipeline. Subjective quality scores are provided in the form of MOS ranging from 0 to 100.

2) Performance Evaluation Criteria: A monotonic logistic function is used to provide a nonlinear mapping between objective scores and subjective scores [58]

$$f(x) = \beta_1\left(\frac{1}{2} - \frac{1}{\exp(\beta_2(x - \beta_3))}\right) + \beta_4 x + \beta_5 \tag{22}$$

where x is the original IQA score, $f(x)$ is the fitted IQA score, $\beta_j (j = 1, 2, ...5)$ are regression parameters trained per database.

The performance of IQA methods are evaluated by three different criteria: Spearman rank order correlation coefficient (SRCC) for prediction monotonicity, Pearson linear correlation coefficient (PLCC) for prediction accuracy and root mean squared error (RMSE) for prediction error. The latter two criteria were calculated after the monotonic logistic mapping.

NRSL was compared with nine state-of-the-art BIQA models, including NIQE [22], BIQI [27], DIIVINE [28], BLIINDS2 [25], CORNIA [34], BRISQUE [26], GMLOG [32],

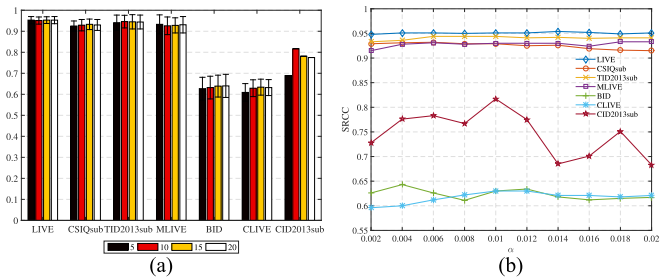


Fig. 7. Parameter choice. (a) The performance of NRSL in terms of SRCC vs. the number of bins of luminance histogram. (b) The performance of NRSL in terms of SRCC versus constant α.

NR-GLBP [59] and NFERM [33]. We also include the performance of FR-IQA methods PSNR and SSIM [8] for reference.

B. Implementation Details

1) Parameter Choice: There are two free parameters in the proposed NRSL model. The first one is the number of bins of luminance histogram. Generally, histograms with small number of bins fail to provide enough discriminative information about the distributions. By contrast, for histograms with too many bins, the average number of entries of each bin would be too small, which may cause the histograms sparse and unstable. To investigate the effect of bin number on the quality prediction performance, we try several choices of the bin number and observe SRCC performance on each database. The results are plotted in Fig. 7(a).

We can see that the bin number of 10 results in high and stable performance across all seven databases. Therefore, in our implementation, we set the number of bins of luminance histogram to 10.

Another free parameter to be determined is the α value in (4). Apart from ensuring the numerical stability, the constant α also plays an important role in mediating the contrast saturation in low contrast regions. Fig. 7(b) plots the SRCC curves against α by applying NRSL to the seven benchmark databases. We can see that for a wide range of choices for α, NRSL achieves satisfactory performance on all the databases except CID2013sub. The reason may be that there is much difference among images in CID2013 database in terms of luminance and contrast (e.g., some are captured in outdoor daylight, while others are indoor scenes with black background). Therefore, the best contrast normalization factor may vary from one scene to another. In the implementation, we set $\alpha = 0.01$ for all databases.

In addition, both the statistical structural and luminance features are calculated at three scales to account for the variations of viewing distance and image resolution [9], [60]. Besides the original image scale, the coarser scale is constructed by low-pass filtering and downsampling the image by a factor of 2.

2) Examination of Two Feature Sets: In Section II-D, we have provided some qualitative analysis on the complementary of the extracted structural and luminance features. In this part, we present the experimental comparison by using only one of the two feature sets to validate our analysis. Specifically, we tested the performance on LIVE database by splitting the database into training and testing subsets 1000 times. 80% of the distorted images are used for training and the rest for testing. There is no content overlap between these two subsets. SRCC results on LIVE database [48] are shown in Table I, where NRSL_S denotes the model which uses only structural feature and NRSL_L denotes the model using luminance feature.

TABLE I
SRCC COMPARISON ON LIVE DATABASE TO VALIDATE THE COMPLEMENTARY OF STRUCTURAL AND LUMINANCE INFORMATION

IQA model	JP2K	JPEG	WN	GB	FF	ALL
NRSL_S	0.940	0.953	0.962	0.928	0.863	0.935
NRSL_L	0.858	0.920	0.977	0.891	0.730	0.871
NRSL	0.943	0.960	0.984	0.959	0.880	0.952

From Table I, we can see that structural feature is more effective for distortion types of JP2K, JPEG, GB and FF, as such distortions greatly interfere the structural information of images; while luminance information are more effective for WN distortion. Moreover, the combination of these two feature sets always delivers higher performance than each feature set under all the cases.

C. Overall Performance on Individual Databases

First, we evaluate the overall performance of competing BIQA models on each benchmark database. As in many previous works [24], [32], [33], for CSIQ [52] and TID2013 [53] databases, we only test the four distortions that also appear in LIVE database (i.e., JP2K, JPEG, WN, GB). These four distortion types are also the most commonly encountered distortions in practical applications. We also exclude the 25th reference image with its distorted versions from TID2013 [53] as it is not a natural image. We denote the selected databases as CSIQsub and TID2013sub, which include 600 and 480 distorted images, respectively. The performance on the whole CSIQ and TID2013 databases will be presented in Section III-D. Since NRSL adopts SVR learning for quality estimation, we need to divide the database into training and testing subsets. For the four synthetically-distorted image databases (i.e., LIVE, CSIQsub, TID2013sub, MLIVE), distorted images are generated by post-processing the reference images. To achieve content independence between training and testing sets, distorted images of 80% of the reference images are used for training, and the rest are used for testing. For BID and CLIVE databases, there are no reference images. Also the images are quite different from each other in terms of content properties. Thus, we randomly split the databases into two sets: 80% for training, and 20% for testing. The training-testing split is repeated 1000 times and the median performance is reported. The CID2013 database can be grouped into two parts since Sets I–III and Sets IV–VI are rated using different subjective evaluation protocols. Sets IV–VI are used in this work for comparing the performance of BIQA methods and denoted as CID2013sub. Also, each Set includes

TABLE II
PERFORMANCE COMPARISON OF 12 IQA MODELS ON 7 BENCHMARK DATABASES

	LIVE (779)			CSIQsub (600)			TID2013sub (480)			MLIVE (450)		
IQA model	SRCC	PLCC	RMSE	SRCC	PLCC	RMSE	SRCC	PLCC	RMSE	SRCC	PLCC	RMSE
PSNR	0.884	0.882	12.808	0.928	0.853	0.146	0.928	0.918	0.555	0.725	0.815	10.934
SSIM [8]	0.939	0.934	9.738	0.923	0.933	0.101	0.918	0.939	0.481	0.901	0.925	6.969
NIQE [22]	0.909	0.909	11.376	0.883	0.892	0.126	0.819	0.831	0.779	0.793	0.860	9.437
BIQI [27]	0.821	0.838	14.874	0.819	0.873	0.137	0.818	0.856	0.722	0.883	0.905	7.833
DIIVINE [28]	0.912	0.913	11.168	0.883	0.905	0.118	0.882	0.904	0.596	0.866	0.898	8.257
BLIINDS2 [25]	0.930	0.936	9.515	0.904	0.932	0.101	0.876	0.905	0.597	0.886	0.903	8.125
CORNIA [34]	0.943	0.946	8.812	0.890	0.926	0.105	0.893	0.925	0.532	0.899	0.915	7.586
BRISQUE [26]	0.943	0.947	8.794	0.903	0.934	0.099	0.900	0.925	0.533	**0.900**	**0.922**	**7.273**
GMLOG [32]	**0.950**	**0.954**	**8.180**	0.924	0.947	0.090	**0.931**	0.943	0.464	0.831	0.871	9.197
NR-GLBP [59]	0.938	0.943	9.075	0.916	0.948	0.089	0.920	0.939	0.479	0.891	0.904	7.932
NFERM [33]	0.938	0.942	9.101	**0.929**	**0.952**	**0.084**	0.929	**0.951**	**0.436**	0.899	0.919	7.413
NRSL	**0.952**	**0.956**	**8.018**	**0.930**	**0.954**	**0.084**	**0.945**	**0.959**	**0.397**	**0.932**	**0.946**	**5.943**

	BID (586)			CLIVE (1162)			CID2013sub (233)			Weighted Average	
IQA model	SRCC	PLCC	RMSE	SRCC	PLCC	RMSE	SRCC	PLCC	RMSE	SRCC	PLCC
NIQE [22]	0.457	0.467	1.100	0.453	0.509	17.394	0.713	0.715	16.224	0.687	0.713
BIQI [27]	0.573	0.598	0.997	0.532	0.557	16.828	0.681	0.684	16.929	0.707	0.735
DIIVINE [28]	0.610	0.646	0.947	0.597	0.627	15.767	0.582	0.608	18.419	0.755	0.779
BLIINDS2 [25]	0.532	0.560	1.029	0.463	0.507	17.419	0.380	0.389	21.378	0.705	0.731
CORNIA [34]	0.625	0.646	0.949	**0.618**	**0.662**	**15.158**	0.687	0.718	16.159	**0.780**	**0.807**
BRISQUE [26]	0.581	0.605	0.990	0.607	0.645	15.450	**0.759**	**0.762**	**15.032**	0.778	0.801
GMLOG [32]	0.543	0.571	1.023	0.595	0.620	15.800	0.599	0.624	18.128	0.761	0.782
NR-GLBP [59]	**0.628**	**0.654**	**0.942**	0.612	0.634	15.622	0.625	0.656	17.510	**0.780**	0.800
NFERM [33]	0.586	0.608	0.986	0.540	0.570	16.641	0.600	0.656	17.517	0.757	0.780
NRSL	**0.638**	**0.663**	**0.931**	**0.631**	**0.654**	**15.317**	**0.817**	**0.816**	**13.399**	**0.809**	**0.825**

six scenes. Images from the same scene are quite similar to each other, and different from images from other scenes in terms of content properties. Thus, we use the leave-one-out strategy by training on images from five scenes and testing on images from the remaining scene. This is repeated until all the images have the predicted scores. Then we calculate the correlation between MOS and predicted scores. Since the training and testing sets are fixed in this scheme, the experiment is conducted only once. This experimental setup on CID2013 is also in line with several recent studies [61], [62]. There is one purely black image in CID2013 database (IS_VI_C01_D14.JPG) and several BIQA methods fail to extract effective features on it, thus we also excludes this image to form the CID2013sub with 233 images.

The prediction performance measured by SRCC, PLCC, and RMSE criteria is listed in Table II. For each criteria, the best two BIQA models are highlighted in boldface. Moreover, in order to provide an evaluation of the overall performance of the competing IQA models, we also present their weighted-average SRCC, PLCC results over all seven databases and the weight assigned to each database is in direct proportion to the number of distorted images in that database.

Form Table II, it can be observed that NRSL performs consistently well on all the benchmark databases. Particulary, it delivers higher performance than other competitors on all seven databases. On the contrary, for other BIQA models, they may work well on some databases but fail to deliver good results on other databases. For example, CORNIA and BRISQUE can get encouraging results on LIVE and MLIVE, but they perform relatively poorly on CSIQsub and TID2013sub databases. GMLOG archives quite good performance on LIVE, CSIQsub and TID2013sub databases, but it performs rather poorly on MLIVE database. Therefore, we can conclude that objective quality scores predicted by NRSL are in higher agreement with subjective ratings than all the other BIQA models examined. However, the performance of all BIQA methods drops significantly on naturally-distorted databases compared with traditional simulated databases. This is to be expected since authentic distortions are generally more complex, diverse and multipartite, and thus the visual quality estimation requires more efforts in the community. The exsiting method NR-GLBP [59] also uses LBP for feature extraction by manually selecting several thresholds to calculate the generalized LBP on four LOG filtered images to form the quality aware features. However, since the LBP mechanism only encodes the sign of the pixel-wise difference between one pixel and its neighbors, it fails to capture the impact of luminance and contrast distortions as the magnitude information is lost. Moreover, the manually selected thresholds may not well accommodate images with different content properties. On the contrary, we effectively combine the complementary features of luminance and structural histograms in our method. The bounded performance of LBP can be improved by the luminance histogram and their combination further advances the quality prediction accuracy. The experimental results in Table II show that our method delivers better performance than NR-GLBP on all the seven benchmark databases. Also, our method has a lower feature dimensionality and faster computation time, as shown in Table VII in Section III-F.

TABLE III
STATISTICAL SIGNIFICANCE TEST

DB	NIQE	BIQI	DIIVINE	BLIINDS2	CORNIA	BRISQUE	GMLOG	NR-GLBP	NFERM
LIVE	1	1	1	1	1	1	0	1	1
CSIQsub	1	1	1	1	1	1	0	1	0
TID2013sub	1	1	1	1	1	1	1	1	1
MLIVE	1	1	1	1	1	1	1	1	1
BID	1	1	1	1	1	1	1	1	1
CLIVE	1	1	1	1	1	1	1	1	1
CID2013sub	1	1	1	1	1	1	1	1	1

1(− 1) indicates NRSL is statistically better (worse) than the method in the column. 0 indicates NRSL is statistically equivalent to the method in the column.

To further prove the superiority of NRSL over the competing BIQA methods, we calculate the statistical significance between NRSL and other BIQA methods. For the first six databases, experiments are conducted by splitting the database into training and testing sets 1000 times and the median performance is reported. Thus, we conduct the two sample t-test between the pair of SRCC values of 1000 train-test loops, which measures the equivalence of the mean values of two independent samples. For the CID2013sub database, experiments are conducted by leave-one-subset-out strategy, and predicted quality scores at all subsets are put together before calculating the final performance. Thus, we conduct the two sample F-test to the residuals of two methods, which measures the equivalence of variances of two independent samples. Both t-test and F-test are conducted at the 5% significance level, with 1 (-1) indicating that NRSL is statistically superior (inferior) to the compared method and 0 indicating that NRSL is statistically equivalent to the compared method. The results are listed in Table III. We can see that on TID2013sub, MLIVE, BID, CLIVE and CID2013sub, NRSL performs statistically better than all other BIQA methods. NRSL is on par with GMLOG on LIVE, on par with GMLOG and NFERM on CSIQsub, respectively. Generally, NRSL achieves statistically better performance than other BIQA methods.

D. Performance on Individual Distortion Type

Here, we evaluate the performance of competing IQA models on individual distortion type. Since there is no well defined distortion category for naturally-distorted images, we only conduct experiments on the four databases with simulated distortions. For BIQA models, we train on the 80% of images with various distortion types and then test on the left 20% of images with the specific distortion type. The SRCC comparison on four benchmark databases are tabulated in Table IV. There are 15 groups of distorted images in the four databases. The best two BIQA models for each distortion group are shown in boldface. It should be noted that similar results can be obtained for PLCC and RMSE criteria, thus we only list SRCC here for brevity.

From the results presented in Table IV, we can see that NRSL is among the best two metrics 12 times, followed by NFERM (7 times) and GMLOG (5 times). We also calculate the weighted mean and standard deviation of competing IQA models across all distortion groups. Among them, NRSL is with the highest mean and the lowest standard deviation across different groups. Thus, we can draw the following conclusions. In general, when distortion is of a specific type, NRSL performs the best among all BIQA models. NFERM, NR-GLBP, and BRISQUE can also achieve acceptable results. Moreover, although GMLOG delivers promising results for singly-distorted images, its performance deteriorates significantly when tested on multiply-distorted images. NRSL achieves consistently better performance on most commonly encountered distortion types.

Furthermore, in order to validate the generalization ability of NRSL to uncommon distortion types, we also compare NRSL with other five best performing BIQA models through 1000 train-test experiments on the whole CSIQ and TID2013 databases. The SRCC values of different methods are presented in Table V. From this table, we can observe that NRSL is among the best two metrics 20 times, followed by NFERM (16 times) and GMLOG (11 times).

E. Cross-Database Validation

In the previous experiments, the training and testing images are from the same database. In this subsection, we test the generalization capability of learning based BIQA models through cross-database validation. We train the BIQA models using all the images from one database and test the performance on another database. Since learning-based BIQA models require example images of same or similar distortions for training, here we only conduct cross-database validation on the three singly-distorted databases (i.e., LIVE, CSIQsub, TID2013sub) as they share similar distortion types (i.e., JP2K, JPEG, WN, GB, FF). The SRCC results of cross-database validation on three databases are provided in Table VI.

It can be seen that NRSL performs quite well in this scenario. It delivers the best performance in 5 out of 6 cases, which has demonstrated the database independency and robustness of the proposed NRSL method.

F. Computational Complexity

In many practical applications, it is much desired to have a low-complexity BIQA method which can estimate image quality in real-time. Therefore, we evaluate the computational complexity and running cost of all competing methods in Table VII. Experiments are performed on a ThinkPad T430S notebook with Intel Core i7-3520M CPU@2.9GHz and 8G RAM. The software platform is MATLAB R2014b (8.4) under Windows

TABLE IV
SRCC COMPARISON OF 12 IQA MODELS ON INDIVIDUAL DISTORTION TYPES

	D-Type	PSNR	SSIM	NIQE	BIQI	DIIVINE	BLIINDS2	CORNIA	BRISQUE	GMLOG	NR-GLBP	NFERM	NRSL
LIVE	JP2K	0.903	0.959	0.923	0.780	0.902	0.928	0.922	0.915	0.925	0.933	**0.937**	**0.943**
	JPEG	0.891	0.975	0.942	0.829	0.899	0.949	0.941	**0.963**	**0.963**	0.958	**0.964**	0.960
	WN	0.984	0.976	0.972	0.958	0.981	0.945	0.963	0.978	0.983	0.981	**0.984**	**0.984**
	GB	0.808	0.960	0.941	0.844	0.935	0.913	**0.955**	0.946	0.929	0.936	0.909	**0.959**
	FF	0.895	0.885	0.862	0.739	0.858	0.873	**0.907**	0.887	**0.899**	0.853	0.850	0.880
CSIQsub	JP2K	0.942	0.970	**0.927**	0.815	0.877	0.889	0.901	0.892	0.915	0.911	0.917	**0.921**
	JPEG	0.902	0.957	0.883	0.856	0.887	0.916	0.886	0.918	**0.936**	0.926	0.927	**0.937**
	WN	0.940	0.823	0.837	0.842	0.901	0.899	0.797	0.919	**0.939**	0.925	0.934	**0.955**
	GB	0.935	0.974	0.906	0.837	0.898	0.919	0.907	0.913	0.906	0.921	**0.924**	**0.926**
TID2013sub	JP2K	0.905	0.952	0.906	0.848	0.892	0.916	0.900	0.905	0.927	0.927	**0.938**	**0.940**
	JPEG	0.931	0.932	0.879	0.855	0.867	0.871	0.869	0.878	0.902	**0.917**	0.896	**0.920**
	WN	0.948	0.891	0.842	0.865	0.905	0.753	0.730	0.884	**0.948**	0.856	**0.939**	0.928
	GB	0.968	0.967	0.841	0.890	0.937	0.902	0.914	0.930	0.916	**0.943**	0.928	**0.958**
MLIVE	GB + JPEG	0.736	0.898	0.899	0.881	0.864	0.892	0.904	0.905	0.865	0.897	**0.919**	**0.928**
	GB + WN	0.743	0.912	0.833	0.883	0.877	0.884	0.900	0.900	0.817	**0.905**	0.887	**0.937**
	wmean	0.883	0.934	0.893	0.849	0.897	0.899	0.896	0.916	0.913	0.919	**0.923**	**0.938**
	wstd	0.080	0.043	0.042	0.048	0.031	0.041	0.053	**0.027**	0.043	0.031	0.030	**0.023**

TABLE V
SRCC COMPARISON ON THE WHOLE CSIQ AND TID2013 DATABASES

	CSIQ (866)							TID2013 (3000)								
BIQA model	WN	JPEG	JP2K	PGN	GB	CTD	ALL	#1	#2	#3	#4	#5	#6	#7	#8	#9
CORNIA [34]	0.763	0.842	0.869	0.567	0.854	0.533	0.733	0.550	0.209	0.717	0.360	0.797	0.585	0.727	0.840	0.721
BRISQUE [26]	0.682	0.846	0.817	0.743	0.808	0.396	0.740	0.706	**0.523**	0.776	0.295	0.836	**0.802**	0.682	0.861	0.500
GMLOG [32]	**0.804**	0.864	0.890	0.774	0.857	0.562	0.804	0.748	**0.591**	0.769	**0.491**	0.875	0.693	**0.833**	0.878	0.721
NR-GLBP [59]	0.657	0.879	0.858	0.787	0.862	**0.697**	0.804	0.664	0.466	0.759	0.081	0.728	0.620	0.728	0.827	0.721
NFERM [33]	0.731	**0.890**	**0.908**	**0.797**	**0.879**	0.628	**0.810**	**0.851**	0.520	**0.846**	**0.521**	**0.894**	**0.857**	**0.785**	**0.888**	**0.741**
NRSL	**0.810**	**0.903**	**0.912**	**0.836**	**0.896**	**0.659**	**0.851**	**0.813**	0.457	**0.867**	0.393	**0.902**	0.787	0.700	**0.886**	**0.795**

	TID2013 (3000)															
BIQA model	#10	#11	#12	#13	#14	#15	#16	#17	#18	#19	#20	#21	#22	#23	#24	ALL
CORNIA [34]	0.806	0.800	**0.595**	0.654	0.157	0.016	0.177	**0.262**	**0.170**	0.407	**0.541**	**0.696**	0.649	0.689	0.874	0.629
BRISQUE [26]	0.790	0.779	0.254	0.723	**0.213**	0.197	**0.217**	0.079	**0.113**	0.674	0.198	0.627	0.849	**0.724**	0.811	0.567
GMLOG [32]	**0.823**	0.872	0.400	**0.731**	0.190	**0.318**	0.119	**0.224**	-0.121	0.701	0.202	0.664	**0.886**	0.648	**0.915**	**0.679**
NR-GLBP [59]	**0.844**	0.867	**0.440**	0.594	**0.226**	0.204	0.105	0.123	-0.023	0.580	**0.447**	0.507	0.762	**0.748**	0.830	0.560
NFERM [33]	0.797	**0.920**	0.381	0.718	0.176	0.081	**0.238**	0.056	-0.029	**0.762**	0.206	0.401	0.848	0.684	0.878	0.652
NRSL	0.818	**0.891**	0.345	**0.805**	0.117	**0.323**	0.136	0.194	-0.110	**0.753**	0.434	**0.751**	**0.866**	0.694	**0.887**	**0.661**

TABLE VI
SRCC COMPARISON ON CROSS-DATABASE VALIDATION

Train DB	Test DB	NIQE	BIQI	DIIVINE	BLIINDS2	CORNIA	BRISQUE	GMLOG	NR-GLBP	NFERM	NRSL
LIVE	CSIQsub	0.869	0.762	0.871	0.901	0.898	0.890	0.897	0.911	0.907	**0.917**
	TID2013sub	0.811	0.821	0.865	0.855	0.879	0.878	0.907	**0.916**	0.913	**0.916**
CSIQsub	LIVE	0.905	0.755	0.801	0.894	0.920	0.919	0.903	0.917	0.918	**0.921**
	TID2013sub	0.811	0.738	0.862	0.765	0.852	0.874	0.879	0.893	0.904	**0.921**
TID2013sub	LIVE	**0.905**	0.742	0.828	0.894	0.907	0.877	0.889	0.899	0.838	0.896
	CSIQsub	0.869	0.669	0.823	0.864	0.859	0.861	0.794	0.821	0.863	**0.875**

7 Home Premium. The MATLAB source codes of all the competing IQA models are obtained from original authors except for NR-GLBP, which is implemented by ourselves according to [59]. The feature extraction time consumed by each IQA model for estimating the quality of one 512×512 image is listed in the second column. Assuming N is the number of pixels and d is the window size for local normalization, we can observe that the proposed method NRSL is quite efficient with the time complexity of $\mathcal{O}(N(d^2+10))$, where $\mathcal{O}(d^2N)$ is for the local contrast normalization, $\mathcal{O}(8N)$ is for the LBP calculation (8 is the number of neighbors for LBP), and $\mathcal{O}(2N)$ is for calculation of the two histograms. More specifically, it can process 10 images per second, which meets the time requirement in most image processing applications.

TABLE VII
COMPUTATIONAL COMPLEXITY AND RUN-TIME COMPARISON OF BIQA MODELS

BIQA model	Time(s)	Time complexity [32]
NIQE	0.23	$\mathcal{O}(d^2 N)$ d: window size
BIQI	0.05	$\mathcal{O}(N)$
DIIVINE	15.52	$\mathcal{O}(N(\log(N) + m^2 + N + 392b))$ m: DNT neighborhood size; b: 2D histogram bin number
BLIINDS2	61.39	$\mathcal{O}((N/d^2)\log(N/d^2))$ d: window size
CORNIA	2.45	$\mathcal{O}(d^2 K N)$ d: window size; K: codebook size
BRISQUE	0.08	$\mathcal{O}(d^2 N)$ d: window size
GMLOG	0.06	$\mathcal{O}(N(h+k))$ h: filter size; k: probability matrix size
NR-GLBP	1.65	$\mathcal{O}(4N(h+5t))$ h: filter size; t: number of threshold
NFERM	54.03	$\mathcal{O}(d^2 N \log(N))$ d: window size of AR model
NRSL	0.11	$\mathcal{O}(N(d^2+10))$ d: window size

IV. CONCLUSION

In this paper, we have proposed a novel blind image quality assessment algorithm based on structural and luminance information. Unlike previous BIQA models normally generate features by fitting transformed image coefficients to specific probability distribution, in the proposed method, we utilize two statistical distributions on normalized luminance map to characterize the HVS-sensitive features. After local contrast normalization, the structural and luminance histograms are extracted to construct the image quality aware features. The structural histogram captures the influence of various distortions on the inter-pixel relationship patterns, while the luminance histogram describes the global variation pattern of pixel-wise luminance value. Our research findings suggest that complementary information of structural and luminance features plays an important role for quality assessment tasks. Extensive experimental results on seven large-scale public IQA benchmark databases have demonstrated that the proposed method NRSL is highly competitive to the state-of-the-art BIQA methods in terms of prediction accuracy, distortion consistency, and database independency. In the future, we would like to extend NRSL to account for distortions in chromatic component and try to build a blind quality assessment metric for video sequences.

REFERENCES

[1] W. Lin and C.-C. J. Kuo, "Perceptual visual quality metrics: A survey," *J. Visual Commun. Image Represent.*, vol. 22, no. 4, pp. 297–312, 2011.

[2] Z. Wang and A. Bovik, "Reduced- and no-reference image quality assessment," *IEEE Signal Process. Mag.*, vol. 28, no. 6, pp. 29–40, Nov. 2011.

[3] J. Alakarhu, "Image sensors and image quality in mobile phones," in *Proc. Int. Image Sensor Workshop*, 2007, pp. 1–4.

[4] G. Zhai *et al.* "Cross-dimensional perceptual quality assessment for low bit-rate videos," *IEEE Trans. Multimedia*, vol. 10, no. 7, pp. 1316–1324, Nov. 2008.

[5] F. Zhang, L. Ma, S. Li, and K. N. Ngan, "Practical image quality metric applied to image coding," *IEEE Trans. Multimedia*, vol. 13, no. 4, pp. 615–624, Aug. 2011.

[6] L. Anegekuh, L. Sun, E. Jammeh, I.-H. Mkwawa, and E. Ifeachor, "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.

[7] A. Gaur and K. Mikolajczyk, "Ranking images based on aesthetic qualities," in *Proc. 22nd Int. Conf. Pattern Recog.*, Aug. 2014, pp. 3410–3415.

[8] 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.

[9] H. Sheikh and A. Bovik, "Image information and visual quality," *IEEE Trans. Image Process.*, vol. 15, no. 2, pp. 430–444, Feb. 2006.

[10] S. Li, F. Zhang, L. Ma, and K. N. Ngan, "Image quality assessment by separately evaluating detail losses and additive impairments," *IEEE Trans. Multimedia*, vol. 13, no. 5, pp. 935–949, Oct. 2011.

[11] L. Zhang, D. Zhang, X. Mou, and D. Zhang, "FSIM: A feature similarity index for image quality assessment," *IEEE Trans. Image Process.*, vol. 20, no. 8, pp. 2378–2386, Aug. 2011.

[12] A. Liu, W. Lin, and M. Narwaria, "Image quality assessment based on gradient similarity," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1500–1512, Apr. 2012.

[13] L. Ma, S. Li, F. Zhang, and K. N. Ngan, "Reduced-reference image quality assessment using reorganized DCT-based image representation," *IEEE Trans. Multimedia*, vol. 13, no. 4, pp. 824–829, Aug. 2011.

[14] J. Wu, W. Lin, G. Shi, and A. Liu, "Reduced-reference image quality assessment with visual information fidelity," *IEEE Trans. Multimedia*, vol. 15, no. 7, pp. 1700–1705, Nov. 2013.

[15] T. Zhu and L. Karam, "A no-reference objective image quality metric based on perceptually weighted local noise," *EURASIP J. Image Video Process.*, vol. 2014, no. 1, pp. 1–8, 2014.

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

[17] N. D. Narvekar and L. J. Karam, "A no-reference image blur metric based on the cumulative probability of blur detection (CPBD)," *IEEE Trans. Image Process.*, vol. 20, no. 9, pp. 2678–2683, Sep. 2011.

[18] Y. Fang *et al.*, "No-reference quality assessment of contrast-distorted images based on natural scene statistics," *IEEE Signal Process. Lett.*, vol. 22, no. 7, pp. 838–842, Jul. 2015.

[19] Z. Wang, H. R. Sheikh, and A. C. Bovik, "No-reference perceptual quality assessment of JPEG compressed images," in *Proc. Int. Conf. Image Process.*, 2002, vol. 1, pp. I–477.

[20] E. Ong *et al.*, "A no-reference quality metric for measuring image blur," in *Proc. 7th Int. Symp. Signal Process. Appl.*, Jul. 2003, vol. 1, pp. 469–472.

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

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

[23] L. Zhang, L. Zhang, and A. Bovik, "A feature-enriched completely blind image quality evaluator," *IEEE Trans. Image Process.*, vol. 24, no. 8, pp. 2579–2591, Aug. 2015.

[24] W. Xue, L. Zhang, and X. Mou, "Learning without human scores for blind image quality assessment," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2013, pp. 995–1002.

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

[26] 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.

[27] 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.

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

[29] J. Shen, Q. Li, and G. Erlebacher, "Hybrid no-reference natural image quality assessment of noisy, blurry, JPEG2000, and JPEG images," *IEEE Trans. Image Process.*, vol. 20, no. 8, pp. 2089–2098, Aug. 2011.

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

[31] H. Tang, N. Joshi, and A. Kapoor, "Learning a blind measure of perceptual image quality," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2011, pp. 305–312.

[32] 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.

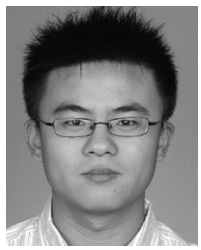
[33] K. Gu, G. Zhai, X. Yang, and W. Zhang, "Using free energy principle for blind image quality assessment," *IEEE Trans. Multimedia*, vol. 17, no. 1, pp. 50–63, Jan. 2015.
[34] P. Ye, J. Kumar, L. Kang, and D. Doermann, "Unsupervised feature learning framework for no-reference image quality assessment," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2012, pp. 1098–1105.
[35] L. Kang, P. Ye, Y. Li, and D. Doermann, "Convolutional neural networks for no-reference image quality assessment," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2014, pp. 1733–1740.
[36] L. Zhang, Z. Gu, X. Liu, H. Li, and J. Lu, "Training quality-aware filters for no-reference image quality assessment," *IEEE Multimedia*, vol. 21, no. 4, pp. 67–75, Oct. 2014.
[37] Q. Li, W. Lin, and Y. 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.
[38] H. Nothdurft, "Sensitivity for structure gradient in texture discrimination tasks," *Vision Res.*, vol. 25, no. 12, pp. 1957–1968, 1985.
[39] A. B. Watson and J. A. Solomon, "Model of visual contrast gain control and pattern masking," *JOSA A*, vol. 14, no. 9, pp. 2379–2391, 1997.
[40] R. A. Frazor and W. S. Geisler, "Local luminance and contrast in natural images," *Vision Res.*, vol. 46, no. 10, pp. 1585–1598, 2006.
[41] V. Mante, R. A. Frazor, V. Bonin, W. S. Geisler, and M. Carandini, "Independence of luminance and contrast in natural scenes and in the early visual system," *Nature Neurosci.*, vol. 8, no. 12, pp. 1690–1697, 2005.
[42] S. Lyu and E. P. Simoncelli, "Nonlinear image representation using divisive normalization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2008, pp. 1–8.
[43] J. Wu, W. Lin, G. Shi, and A. Liu, "Perceptual quality metric with internal generative mechanism," *IEEE Trans. Image Process.*, vol. 22, no. 1, pp. 43–54, Jan. 2013.
[44] T. Ojala, K. Valkealahti, E. Oja, and M. Pietikinen, "Texture discrimination with multidimensional distributions of signed gray-level differences," *Pattern Recog.*, vol. 34, no. 3, pp. 727–739, 2001.
[45] T. Ojala, M. Pietikäinen, and T. Mäenpää, "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.
[46] J. Wu, W. Lin, G. Shi, X. Wang, and F. Li, "Pattern masking estimation in image with structural uncertainty," *IEEE Trans. Image Process.*, vol. 22, no. 12, pp. 4892–4904, Dec 2013.
[47] D. L. Ruderman, "The statistics of natural images," *Netw.: Comput. Neural Syst.*, vol. 5, no. 4, pp. 517–548, 1994.
[48] H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik, "LIVE image quality assessment database release 2," 2005. [Online]. Available: http://live.ece.utexas.edu/research/quality/subjective.htm
[49] L. Van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, no. 2579–2605, p. 85, 2008.
[50] M. Narwaria and W. Lin, "Objective image quality assessment based on support vector regression," *IEEE Trans. Neural Netw.*, vol. 21, no. 3, pp. 515–519, Mar. 2010.
[51] B. Schlkopf and A. J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Cambridge, MA, USA: MIT Press, 2002.
[52] E. C. Larson and D. Chandler, "Categorical image quality (CSIQ) database," 2010. [Online]. Available: http://vision.okstate.edu/csiq
[53] N. Ponomarenko *et al.*, "Image database TID2013: Peculiarities, results and perspectives," *Signal Process.: Image Commun.*, vol. 30, pp. 57–77, 2015.
[54] 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.
[55] A. Ciancio *et al.*, "No-reference blur assessment of digital pictures based on multifeature classifiers," *IEEE Trans. Image Process.*, vol. 20, no. 1, pp. 64–75, Jan. 2011.
[56] D. Ghadiyaram and A. C. Bovik, "Massive online crowdsourced study of subjective and objective picture quality," *IEEE Trans. Image Process.*, vol. 25, no. 1, pp. 372–387, Jan. 2016.
[57] T. Virtanen, M. Nuutinen, M. Vaahteranoksa, P. Oittinen, and J. Hkkinen, "CID2013: A database for evaluating no-reference image quality assessment algorithms," *IEEE Trans. Image Process.*, vol. 24, no. 1, pp. 390–402, Jan. 2015.
[58] "Final report from the video quality experts group on the validation of objective models of video quality assessment," VQEG, Boulder, CO, USA, Mar. 2000.
[59] M. Zhang, C. Muramatsu, X. Zhou, T. Hara, and H. Fujita, "Blind image quality assessment using the joint statistics of generalized local binary pattern," *IEEE Signal Process. Lett.*, vol. 22, no. 2, pp. 207–210, Feb. 2015.
[60] Z. Wang, E. Simoncelli, and A. Bovik, "Multiscale structural similarity for image quality assessment," in *Proc. 37th Asilomar Conf. Signals, Syst. Comput.*, Nov. 2003, vol. 2, pp. 1398–1402.
[61] M. A. Saad, P. Corriveau, and R. Jaladi, "Objective consumer device photo quality evaluation," *IEEE Signal Process. Lett.*, vol. 22, no. 10, pp. 1516–1520, Oct. 2015.
[62] M. Nuutinen, T. Virtanen, and P. Oittinen, "Image feature subsets for predicting the quality of consumer camera images and identifying quality dimensions," *J. Electron. Imag.*, vol. 23, no. 6, p. 061111, 2014.

Qiaohong Li received the B.E. and M.E. degrees in information and communication engineering from the Beijing University of Posts and Telecommunications, Beijing, China, and is currently working toward the Ph.D. degree in computer engineering from Nanyang Technological University, Singapore.

Her research interests include image quality assessment, speech quality assessment, computer vision, and visual perceptual modelling.

Weisi Lin (S'91–M'92–SM'00–F'16) received the B.Sc. degree in electronics and the M.Sc. degree in digital signal processing from Zhongshan University, Guangzhou, China, in 1982 and 1985, respectively, and the Ph.D. degree in computer vision from King's College London, London, U.K., in 1992.

He was involved in teaching and research with Zhongshan University; Shantou University, Shantou, China; Bath University, Bath, U.K.; the National University of Singapore, Singapore; the Institute of Microelectronics, Singapore; and the Institute for Infocomm Research, Singapore. He was the Laboratory Head of Visual Processing and the Acting Department Manager of Media Processing, Institute for Infocomm Research, Singapore. He is currently an Associate Professor with the School of Computer Engineering, Nanyang Technological University, Singapore. His current research interests include image processing, perceptual modeling, video compression, multimedia communication, and computer vision.



Jingtao Xu (S'14–M'16) received the B.S. degree in electronic information science and technology from Beijing Normal University, Beijing, China, in 2008, and is currently working toward the Ph.D. degree at the Beijing University of Posts and Telecommunications, Beijing.

From 2013 to 2015, he was a Visiting Student with the Institute for Advanced Computer Studies, University of Maryland at College Park, College Park, MD, USA. His research interests include image quality assessment, video quality assessment, and machine learning.

Yuming Fang received the B.E. degree from Sichuan University, Chengdu, China, the M.S. degree from the Beijing University of Technology, Beijing, China, and the Ph.D. degree from Nanyang Technological University, Singapore.

He is currently an Associate Professor with the School of Information Technology, Jiangxi University of Finance and Economics, Nanchang, China. He was previously was a Visiting Researcher with the IRCCyN laboratory, PolyTech Nantes, Nantes, France; the University of Nantes, Nantes, France; National Tsinghua University, Hsinchu, Taiwan; and the University of Waterloo, Waterloo, ON, Canada. His research interests include visual attention modeling, visual quality assessment, image retargeting, computer vision, and 3D image/video processing.