Blind Quality Assessment of Tone-Mapped Images Via Analysis of Information, Naturalness, and Structure

Ke Gu, Shiqi Wang, *Member, IEEE*, Guangtao Zhai, *Member, IEEE*, Siwei Ma, *Member, IEEE*, Xiaokang Yang, *Senior Member, IEEE*, Weisi Lin, *Fellow, IEEE*, Wenjun Zhang, *Fellow, IEEE*, and Wen Gao, *Fellow, IEEE*

***Abstract*—High dynamic range (HDR) imaging techniques have been working constantly, actively, and validly in the fault detection and disease diagnosis in the astronomical and medical fields, and currently they have also gained much more attention from digital image processing and computer vision communities. While HDR imaging devices are starting to have friendly prices, HDR display devices are still out of reach of typical consumers. Due to the limited availability of HDR display devices, in most cases tone mapping operators (TMOs) are used to convert HDR images to standard low dynamic range (LDR) images for visualization. But existing TMOs cannot work effectively for all kinds of HDR images, with their performance largely depending on brightness, contrast, and structure properties of a scene. To accurately measure and compare the performance of distinct TMOs, in this paper develop an effective and efficient no-reference objective quality metric which can automatically assess LDR images created by different TMOs without access to the original HDR images. Our model is shown to be statistically superior to recent full- and no-reference quality measures on the existing tone-mapped image database and a new relevant database built in this work.**

***Index Terms*—High dynamic range, image quality assessment (IQA), information entropy, no-reference (NR), statistical naturalness, structural preservation, tone mapping.**

I. Introduction

OVER the last decades, low dynamic range (LDR) images displayed on standard 8-bit monitors have dominated our daily experience with digitalized visual signals. Due to the restricted 256 levels of intensity, LDR display can possibly lead to the missing of some fine image details, which may turn out to be critical in some military and medical applications. With the rapid development of imaging and data processing technologies, in recent years much attention has been paid to high dynamic range (HDR) images that have broader dynamic ranges of intensity levels, so as to better represent luminance variations ranging from direct sunlight to faint starlight and maintain more detailed information [1]. At present, various research areas concerning HDR images abound [2], [3].

A common problem in computer graphics applications is how to visualize HDR images on standard 8-bit display devices. To address this problem, a growing number of tone mapping operators (TMOs) that convert HDR to LDR images have been devised in [4]–[6]. Due to the reduction in dynamic range, TMOs inevitably cause information loss. Therefore, on one hand the best tone-mapped image generated from the HDR version still calls for a human-assisted step, in which subjects compare a large set of tone-mapped images produced by distinct TMOs to pick the most satisfactory one, and on the other hand, an HDR tone-mapped image cannot appear as good as the HDR one displayed on monitors of more than 8 bits. But an HDR tone-mapped image via the proper tone mapping operation can provide a greater sense of luminance, contrast and details than a traditional LDR one, even though both them use the same 8-bit representation.

Nowadays, the need of image quality assessment (IQA) is grown by the rapid growth of multimedia applications, such as the optimization of video compression [7], [8], transmission [9] and enhancement [10], [11]. In the past several years, the designs of IQA were concentrating on assessing LDR images. Among them, the most popular one is perhaps the structural similarity index (SSIM) [12], which compares the similarities of the original and distorted images in light of luminance, contrast and structural information. After that, some modified SSIM-type of approaches have been designed, such as multi-scale SSIM (MS-SSIM) [13] and optimal scale selection SSIM (OSS-SSIM) [14].

The reference images, however, are not available in many cases, e.g., postprocessing systems. As a result, no-reference (NR)/blind IQA has induced many interests over the last few years. The first type of NR-IQA tries to take advantage of the underlying relationship of natural scene statistics (NSS)-based

Manuscript received January 18, 2015; revised July 28, 2015 and October 31, 2015; accepted January 11, 2016. Date of publication January 18, 2016; date of current version February 18, 2016. This work was supported in part by the Singapore MoE Tier 1 Project under Grant M4011379 and Grant RG141/14, and in part by the National Science Foundation of China under Grant 61025005, Grant 61371146, Grant 61221001, and Grant 61390514. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Christian Timmerer.

K. Gu is with the School of Computer Engineering, Nanyang Technological University, Singapore 639798, and also with Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: guke.doctor@gmail.com).

W. Lin is with the School of Computer Engineering, Nanyang Technological University, Singapore 639798 (e-mail: wslin@ntu.edu.sg).

G. Zhai, X. Yang, and W. Zhang are with the Institute of Image Communication and Information Processing, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: zhaiguangtao@sjtu.edu.cn; xkyang@sjtu.edu.cn; zhangwenjun@sjtu.edu.cn).

S. Wang, S. Ma, and W. Gao are with the Institute of Digital Media, School of Electronic Engineering and Computer Science, Peking University, Beijing 100871, China (e-mail: sqwang1986@pku.edu.cn; swma@pku.edu.cn; wgao@pku.edu.cn).

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

Digital Object Identifier 10.1109/TMM.2016.2518868

features and subjective ratings, e.g., Distortion Identification-based Image Verity and INtegrity Evaluation (DIIVINE) [15], BLind Image Integrity Notator using DCT Statistics (BLIINDS-II) [16], and Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) [17]. The second type of model systematically merges two reduced-reference (RR) quality metrics [18], [19][1] to eliminate the requirement of the lossless image [20]. The third type of blind IQA approaches works even without using human scored images, for instance, natural image quality evaluator (NIQE) [21] and quality-aware clustering (QAC) [22].

Clearly, an IQA metric suitable to LDR images and HDR images has substantial importance and wide applications. By some modifications of high contrast vision, nonlinear response to light for the full range of luminance and local adaption, Mantiuk *et al.* have improved the visual difference predictor (VDP) to develop the VDP-HDR [23], which can assess the visual quality of HDR images and LDR ones. In contrast, the VDP-HDR-2 [24] and VDP-HDR-2.2 [25] were subsequently proposed with a complete overhaul. The two metrics through mimicking the anatomy of the human visual system (HVS) have acquired fairly well correlation performance accuracy in accordance with subjective opinions in assessing both LDR and HDR images.

There have been very limited efforts devoted to the quality evaluation of a tone-mapped image compared to its associated HDR one, although the performance of any of existing TMOs is not always effective for all kinds of HDR images, which relies on brightness, contrast, structure, color and visual scene. So the systems to faithfully measure and compare the visual quality of tone-mapped images created by various TMOs are highly desirable.

Yeganeh and Wang were the pioneers paying attention to this issue [26]. They first presented a dedicated tone-mapped image database (TMID), which consists of 120 tone-mapped images created from 15 HDR versions by using eight distinct TMOs and the corresponding mean opinion scores (MOSs) recorded from 20 subjects. The TMID database offers mean human ratings of testing images, but subjective methods are considerably time-consuming, costly and labor intensive for real-time processing systems, and are thus hardly applied in automatic operator or parameter optimization to obtain the tone-mapped image of optimal quality. To that end, Yeganeh *et al.* also explored the tone-mapped quality index (TMQI) with a two-part framework. One is the modified MS-SSIM and the other comes from the NSS regulation that provides some fundamental regularities of natural images [27].

As previously mentioned, despite numerous IQA methods, it was found by tests that the TMQI and existing NR-IQA metrics do not achieve satisfactory performance on assessing the quality of tone-mapped images [28]. Accordingly, in this paper we devise an effective blind tone-mapped quality index (BTMQI) via the analysis of information, naturalness and structure. First, we assume that a high-quality tone-mapped image maintains much information [28] and we thus measure the information entropy of a set of intermediate images by darkening/brightening the original brightness of the input tone-mapped image. Second, as indicated in [29], the human visual sensation is highly adapted to the natural environment and a natural-looking LDR image is thereby more expected. Hence the departure of a tone-mapped image from natural image statistics is considered to be a significant measure of perceptual quality. Third, primary image structures contained in the HDR image should be preserved in its associated tone-mapped one, which motivates the use of the Sobel operator with a constant threshold to detect basic structures in a tone-mapped image. Lastly, a trained regression module is used to combine the aforementioned three respects of features to derive the overall quality score.

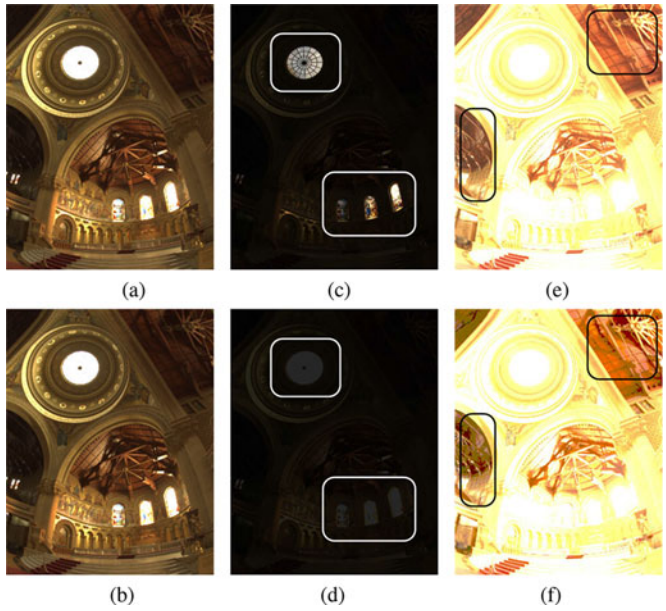


Fig. 1. Comparison of high and low dynamic range images [30]: (a)–(b) HDR image and its related normal 8-bit LDR image; (c)–(d) darkened images of (a)–(b) with 1/64th the original brightness; (e)–(f) brightened images of (a)–(b) with 32 times the original brightness. We label the obviously different regions with rectangles.

The rest of this paper is arranged as follows. Section II first describes the proposed blind IQA approach. In Section III, a comparison of our technique with TMQI and state-of-the-art NR-IQA metrics on the TMID database [26] and a new image database developed in this research is reported and analyzed. We finally conclude this paper in Section IV.

II. Blind BTMQI Approach

In order to better visualize the differences between the high and low dynamic range images, the HDR shop [30] offers an example, as provided in Fig. 1.[2] Though we cannot easily show the differences between the two images in (a) and (b), their darkened and brightened versions are able to show the differences. In Fig. 1, (c) and (d) are created by darkening (a) and (b) to 1/64th the original brightness, while (e) and (f) are generated by brightened to 32 times the original brightness. Some noticeably

[1]RR IQA works under the situation that the partial original image or some extracted features are available as auxiliary information for quality evaluation.

[2]Actually, we cannot display the HDR images in this paper, so the HDR monitor is used to show the HDR images to be captured by a screenshot tool.

distinguished regions are labeled with white rectangles in darkened images (c)–(d) and black rectangles in brightened images (e)–(f).

We find that the tone-mapped image, due to the limitation of dynamic range, cannot preserve all the information of the original HDR image. It is reasonable to suppose that a good tone-mapped image contains a great amount of information. On this basis, the first consideration of our metric assessing the quality of a tone-mapped image is to straightforwardly estimate the information volume in itself and the intermediate images created by darkening/brightening the original luminance. The intermediate images are produced by

$$I_i = \min(\max(M_i \cdot I, 0), 255) \tag{1}$$

where I is an input tone-mapped image, and M_i indicates the ith multiplier. The $\max$ and $\min$ operators are applied to clip the intermediate images into the range of $0 \sim 255$.

Next, we need to seek for a way to measure the information amount. Information entropy [31], as an important concept in statistics, is an appropriate criterion. By computing the mean unpredictability of one random signal, entropy represents its disorderly degree. More precisely, given a probability density p, we define its entropy as

$$E(p) = -\int p(x) \log p(x) dx. \tag{2}$$

To compare the differences of two probability distributions, other important metrics, such as the Kullback-Leibler (K-L) divergence and its modified symmetric formats [32], [33], are also frequently employed information-theoretic "distances". However, we find by experiments that, in contrast to entropy, the use of the K-L divergence or its symmetric versions is not able to contribute to considerable performance gain, but introduce much cost time. So we determine to adopt entropy to quantify the information volume.

A good tone-mapped image, in most cases, is generally of large entropy. We first offer a pair of tone-mapped images in Fig. 2, where (a) shows a comparatively high-quality tone-mapped image while (b) indicates a low-quality tone-mapped image. We then create a group of intermediate images with $M = \{1, n, \frac{1}{n} | n = 1.5, 2.0, \ldots, 10\}$ for each of the two images before calculate the associated entropy values. Fig. 2(c) illustrates how the entropy E varies with the changes of the multiplier M. To specify, the top red curve and bottom blue curve correspond to (a) and (b), respectively. It is easy to observe that, by a small amount of decrease/increase in luminance, entropy of (b) quickly falls down to a low level, which means that it contains less information. In comparison, (a) presents a good resistance to the fast fading of entropy. Referring to subjective human ratings in the TMID database [26], (a) indeed has a higher quality score than (b).

Two other tone-mapped images selected from the TMID database are also exhibited in Fig. 2(d)–(e) for comparison. Both the images do not have any over-exposed or under-exposed areas. The former image is of slightly better quality than the latter one. Fig. 2(f) displays the relationship between the multiplier M and the corresponding entropy E in (d)–(e), which presents

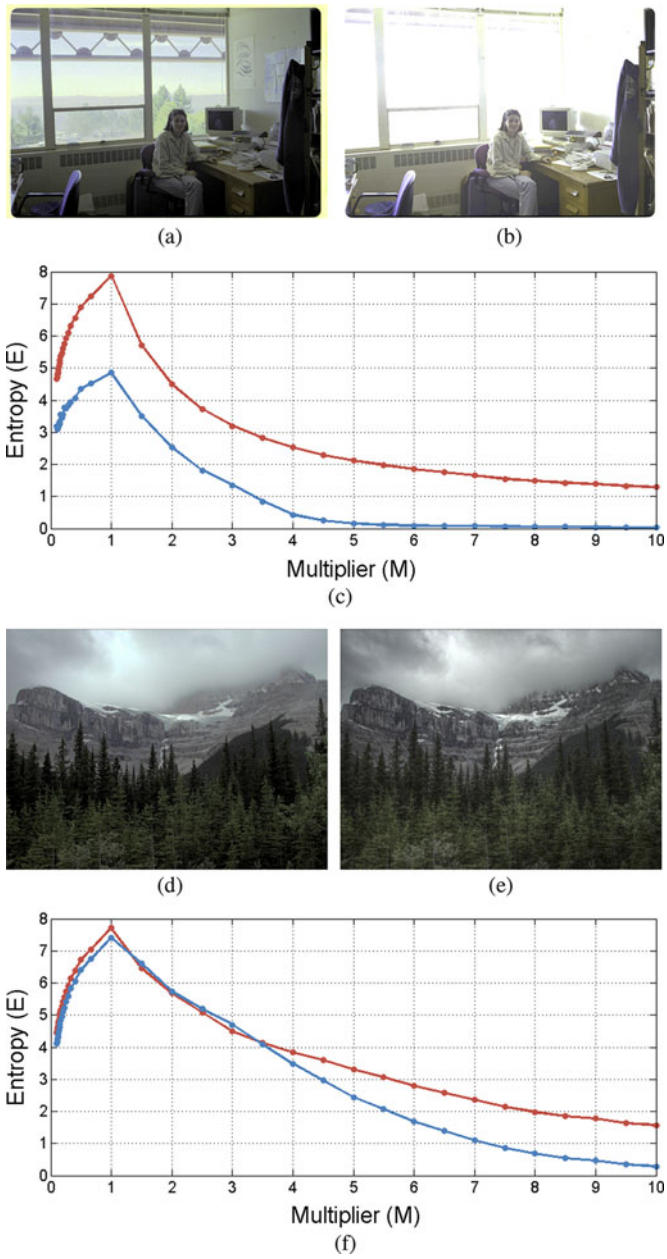


Fig. 2. Illustration of how the entropy E varies with the changes of the multiplier M. (a) A comparatively high-quality tone-mapped image. (b) A low-quality tone-mapped image. (c) The relationship between the varied multiplier M and the corresponding entropy E in (a) and (b). The top red curve is for (a), and the bottom blue one is for (b). (d)–(f) correspond to another example similar to (a)–(c).

a similar result compared to (c) although the red curve and the blur one is quite close to each other and four in 37 values reverse somewhat.

To find a good balance between efficacy and efficiency, we take advantage of only nine entropy values that are measured using $M = \{1, n, \frac{1}{n} | n = 3.5, 5, 6.5, 8\}$ as features. It needs to stress that we also testify and compare the performance of numerous other choices for n, and results tell that our final selection can bring about 2% performance boosts relative to the majority of choices.

Notice that the above-mentioned nine numbers are global-based entropy. Broadly speaking, the perception of the human brain to visual signals inclines to a local-and-global manner. With this concern, many recent techniques in various kinds of research directions, e.g., full-reference (FR) IQA [34] and saliency detection [35], have resulted in better performance. Hence we redefine the entropy of each of nine intermediate images stated above to be

$$E_t(I_i) = wE_g(I_i) + (1-w)E_l(I_i)^v \tag{3}$$

where $E_g(I_i)$ and $E_l(I_i)$ indicate the global and local entropy respectively; w and v are positive constants for manipulating the relative importance of the two components above and are set to be 0.59 and 1.5 in this work. Referring to the scheme in [22], the local entropy is estimated as the mean of block-based entropy, which is defined by

$$E_l(I_i) = \frac{1}{L}\sum_{j=1}^{L} E(B_{i,j}) \tag{4}$$

where $B_{i,j}$ represents the jth block of size 72×72 in the ith intermediate image; L is the number of the blocks in the image.

The second consideration of our approach is the statistical naturalness; that is we assume a high-quality tone-mapped LDR image should look natural as well. During the last few years, there has been a large body of literature dedicated to the statistics of natural images, in order to facilitate the explorations of image/video processing technologies and the understanding of biological vision [36]. Many existing NR-IQA metrics were built upon the natural scene statistics, e.g., [15]–[17]. The fundamental idea behind these models lies in that, for natural images, the normalized coefficients processed by local mean removal and divisive normalization closely follow a Gaussian distribution, whereas different distortion types or levels will reshape this distribution.

Nevertheless, we noticed that the aforesaid statistic model is not proper in assessing the visual quality of tone-mapped images, and another NSS-based model is thus in demand. A recent research of natural statistics on tone-mapped images has revealed that there exist high correlations of naturalness and image attributes, particularly the luminance and contrast [37]. Inspired by this, our BTMQI model takes into account a statistical naturalness model developed upon the average luminance and contrast [26]. In essential, this model provides an ideal tradeoff between the simplicity and the capability of capturing the crucial ingredients of image naturalness.

To be more concrete, based on natural statistics, this model is constructed by using 3,000 natural images with different categories of natural scenes, including animal, nightfall, fresco, building, grassland, tree, snow mountain, train, toy, sea, car, artifact, flower, and appliance. For each of above categories, a vast number of images were downloaded from [38],[3] and then 3,000 high-quality natural images were selected for the aforementioned fourteen content types. In each category two sample images are exhibited in Fig. 3.

[3]"Computer vision test images," [Online]. Available: http://www-2.cs.cmu.edu/afs/cs/project/cil/www/v-images.html

Fig. 3. Representative images in the chosen 3000 natural images. (a) Animal, (b) Nightfall, (c) Fresco, (d) Building, (e) Grassland, (f) Tree, (g) Snow mountain, (h) Train, (i) Toy, (j) Sea, (k) Car, (l) Artifact, (m) Flower, and (n) Appliance.

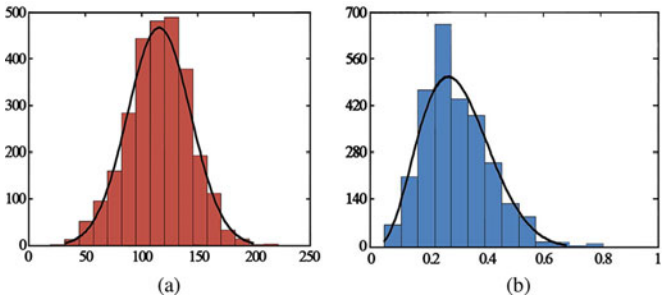


Fig. 4. Histograms of (a) means (fitted by Gaussian probability density function) and (b) standard deviations (fitted by Beta probability density function) of natural images.

For each of the natural images, we first convert it to be the gray-scale image via chromatic information removal before divided into a series of 11×11 patches. Next we compute the local mean (M_P) and local standard deviation (N_P) of each patch, so as to derive the global mean (m) and standard deviation (d) of the natural image by

$$m = \frac{1}{H}\sum_{h=1}^{H} M_P(h), \quad d = \frac{1}{H}\sum_{h=1}^{H} N_P(h) \tag{5}$$

where H stands for the number of patches in the image. In Fig. 4, we plot the two histograms of means and standard deviations of the 3,000 natural images. In general, these two measures are able to reflect the image intensity and contrast. It is viewed that the two histograms can be well fitted by a Gaussian function and a Beta probability density function defined as follows:

$$P_m(m) = \frac{1}{\sqrt{2\pi}\sigma_m}\exp\left[-\frac{m-\mu_m}{2\sigma_m^2}\right] \tag{6}$$

and

$$P_d(d) = \frac{(1-d)^{\beta_d - 1} d^{\alpha_d - 1}}{B(\alpha_d, \beta_d)} \tag{7}$$

where $B(\cdot)$ is the Beta function; the model parameters are assigned to be $\mu_m = 115.94$, $\sigma_m = 27.99$, $\alpha_d = 4.4$, and $\beta_d = 10.1$. The fitting curves are shown in Figs. 4(a)–(b). We further estimate their joint probability density function to be their product for dimensional reduction of features and finally define the

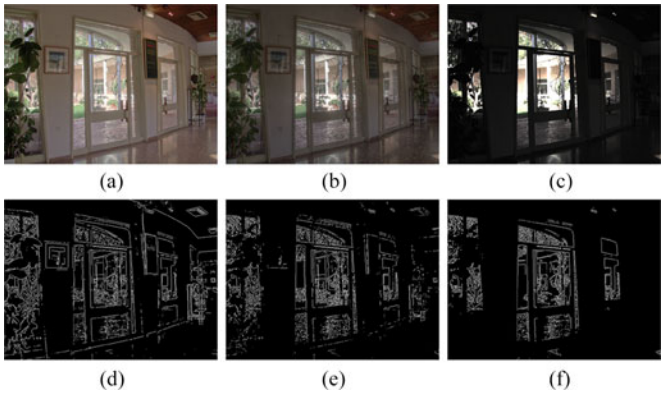


Fig. 5. Comparison of structural preservation ability: (a)–(c) indicate three tone-mapped LDR images associated to the same HDR image; (d)–(f) show the detected primary structures from (a)–(c).

statistical naturalness measure as

$$N = \frac{1}{K} P_m P_d \tag{8}$$

where K is a normalization factor changed with both P_m and P_d. In this work we define $K = \max\{P_m, P_d\}$ to make the statistical naturalness feature N normalized.

The third consideration of our approach is to calibrate the ability of a tone-mapped image to preserve main structures. In [12], Wang *et al.* indicated that the HVS highly adapts to extracting the structural information from an input visual stimulus, and they also provided a large set of examples to demonstrate that the structure-based SSIM is superior to the traditional mean-squared error (MSE) on assessing the image quality. In fact, numerous high-accuracy IQA algorithms are designed based on structural similarity [13], [14] or gradient magnitude similarity that can be regard as another measure of image structure [34], [39].

Note that the computation of structural similarity requires both original and distorted images, but the original references are not accessible under the blind condition. Therefore we make use of the gradient magnitude with a small threshold to search for the primary structures. Particularly, we first apply one of the most famous Sobel operator to extract the image gradient magnitude G. Because only the basic structures are wanted, a small threshold $T(= 0.05)$ is further employed to remove the fragments

$$X(n) = \begin{cases} 1, & \text{if } G(n) \geq T \\ 0, & \text{otherwise} \end{cases} \tag{9}$$

where n represents the image pixel index. As given in Fig. 5, (a)–(c) indicate three tone-mapped LDR images associated to the same HDR image, and (d)–(f) are the results. By the usage of the Sobel operator, it can be observed in (d)–(f) that the primary structures have been detected. In Fig. 5, (e) shows more structures than (f) but less than (d); for instance, the fresco hanging on the wall and the two vents in the top right corner. This means (e) has a better ability for structural preservation than (f) but worse than (d). These results are consistent with the MOS scores in [26] that (b) has a higher quality score than (c) but a lower one than (a). Accordingly, we compute the mean of X (marked as S) to measure how a tone-mapped image can preserve structures.

So far, we have extracted in total eleven features from a tone-mapped image, including $\{E_t(I_i), N, S | i = 1, 2, \ldots, 9\}$. Thereafter, we need to find a mapping learnt from the feature space to subjective human ratings with a regression module, and then use it to yield a quality score of the tone-mapped image. Of course, any regressor can be utilized. Considering the broad application of the support vector regression (SVR) [40], e.g., facet modelling [41] and visual classification [42], we also apply it following the method in [15]–[17]. Here the LIBSVM package [43] is adopted to implement SVR with a radial basis function (RBF) kernel, and more contents about training and testing will be described in Section III.

Finally, we will explain how to compute the overall quality of a tone-mapped image. For the image I, we first extract 11 features of information entropy, statistical naturalness and structural preservation. Second, the module trained on the entire images in the TMID database is used to integrate the aforesaid 11 features to predict the visual quality of the tone-mapped image I. For example, with our BTMQI technique, we estimate the objective quality scores of the tone-mapped images displayed in Fig. 5(a)–(c). The BTMQI scores are respectively 2.4, 4.0 and 7.4, having the same order with their subjective scores 2, 4.5 and 7.8. The performance of our blind IQA algorithm will be validated in the next section.

III. Experimental Results and Analysis

In this section, we will testify the correlation performance of the proposed BTMQI metric by comparison with popular and state-of-the-art blind IQA techniques on existing TMID database and a new relevant database built in our research. After that, we will analyze the advantages and disadvantages of our BTMQI model and offer several possible suggestions in the future work.

A. Experimental Protocol

We calibrate the correlation performance of the proposed BTMQI metric with some state-of-the-art IQA techniques on the TMID database [26]. The TMID database was developed at University of Waterloo, Canada. 15 HDR images were used to create a total of 120 tone-mapped LDR images with eight different TMOs. In each image set, 20 observers were invited to rank the eight testing images from the best to the worst, scored as 1 to 8. The subjective rankings for each image are then averaged to yield the final score.

The testing IQA models include six leading blind metrics and the FR TMQI method. The first model is the TMQI [26] that integrates the modified MS-SSIM [13] and a NSS-based statistical naturalness model. The second one is the DIIVINE [15], which adopts NSS-based 88 features to characterize the essence of natural images by distortion identification followed by distortion-specific visual quality metrics. The third one is the BLIINDS-II [16], which extracts 24 features in the DCT

TABLE I
PERFORMANCE COMPARISON OF BTMQI AND LEADING IQA ALGORITHMS

TMID database [26]		Four-parameter logistic function			Generalized linear models		
Model	Type	PLCC	SROCC	KROCC	PLCC	SROCC	KROCC
TMQI [26]	FR	**0.7715**	**0.7407**	**0.5585**	**0.7467**	**0.7407**	**0.5585**
DIIVINE [15]	NR	0.3791	0.3664	0.2986	0.1894	0.1711	0.1154
BLIINDS-II [6]	NR	0.5019	0.4429	0.3072	0.4914	0.4429	0.3072
BRISQUE [17]	NR	0.5481	0.4810	0.3351	0.5151	0.4810	0.3351
NFERM [20]	NR	0.3249	0.2427	0.1693	0.1630	0.2424	0.1685
NIQE [21]	NR	0.5652	0.4968	0.3495	0.4582	0.4968	0.3495
QAC [22]	NR	0.7148	0.5186	0.3597	0.6189	0.5184	0.3588
BTMQI (Pro.)	NR	**0.8541**	**0.8282**	**0.6545**	**0.8330**	**0.8282**	**0.6545**

We bold the top two models.

domain and then uses a Bayesian inference model to predict quality. The fourth one is the BRISQUE [17], which relies on scene statistics of locally normalized luminance coefficients to quantify "naturalness" losses, so as to infer a quality score. The fifth one is the NFERM [20], which was built upon an observation that there exists an approximate linear relationship of the features extracted by using two RR IQA algorithms, to eliminate the requirement of the original references. The sixth one is the NIQE [21], which does not apply any prior knowledge of contents or distortions but rather than measures the deviations from statistical regularities of natural images to predict quality. The seventh one is the QAC [22] that uses a codebook learnt from quality-aware centroids to derive the quality level of each image patch, in order to generate the overall quality score.

As suggested by video quality experts group (VQEG) [44], the nonlinear regression between subjective human ratings and objective quality predictions is computed with the four-parameter logistic function. Then three typical indices are used to compute the correlation performance. The first and second indices are the Spearman rank-order correlation coefficient (SROCC) and the Kendall's rank-order correlation coefficient (KROCC) for measuring the prediction monotonicity, and the third one is the Pearson linear correlation coefficient (PLCC) for measuring the prediction accuracy. Among the aforesaid three indices, a good objective quality metric is expected to be close to one in SROCC, KROCC and PLCC. More relevant information can be found in [39].

B. Performance Measure and Comparison

To evaluate the correlation performance of our metric, we use a training procedure to test the regression module.[4] Via the commonly used training process, we randomly divide 120 tone-mapped images in the TMID database into two subsets. One is the training set which consists of 96 images corresponding to 80% HDR images, and the other used for testing involves the rest 20%. In order to validate that the proposed BTMQI is robust across image contents and is not decided by the specific train-test split, we repeat this random 80% train—20% test procedure 1,000 times, and report the median performance value across the 1,000 iterations for the purpose of eliminating performance bias as much as possible. The results of our approach are tabulated in Table I, which indicates the substantially high performance accuracy of the proposed BTMQI.

[4]The correlation performance measure used in [26] is to compute the mean value of correlations obtained on subsets of the data, while in this paper all the correlations are computed using the whole database.

TABLE II
COMPARISON OF DISTINCT FEATURE COMBINATIONS IN BTMQI

TMID database [26]				
Model	Num	Train-Test	SROCC	KROCC
$BTMQI_9$		53%-47%	0.7455	0.5573
$BTMQI_9$		60%-40%	0.7536	0.5663
$BTMQI_9$	9	67%-33%	0.7616	0.5765
$BTMQI_9$		73%-27%	0.7698	0.5854
$BTMQI_9$		80%-20%	0.7822	0.6014
$BTMQI_9$		87%-13%	0.7882	0.6219
$BTMQI_{10}$		53%-47%	0.7778	0.5875
$BTMQI_{10}$		60%-40%	0.7843	0.5984
$BTMQI_{10}$	10	67%-33%	0.7950	0.6101
$BTMQI_{10}$		73%-27%	0.8074	0.6260
$BTMQI_{10}$		80%-20%	0.8209	0.6473
$BTMQI_{10}$		87%-13%	0.8262	0.6611
$BTMQI_{11}$		53%-47%	0.7856	0.5950
$BTMQI_{11}$		60%-40%	0.7929	0.6059
$BTMQI_{11}$	11	67%-33%	0.8030	0.6187
$BTMQI_{11}$		73%-27%	0.8150	0.6329
$BTMQI_{11}$		80%-20%	0.8282	0.6545
$BTMQI_{11}$		87%-13%	0.8382	0.6778

We further present the performance results of the seven competing IQA methods in Table I. For a fair comparison, we use the logistic function to reduce the nonlinearity of predicted scores of seven testing metrics before measure the corresponding SROCC, KROCC and PLCC values. It is clear that our technique performs better than those recently developed NR-IQA algorithms. Owing to the non-existence of the reference image, blind metrics are considered hardly matchable with FR-IQA methods. Despite this, the proposed BTMQI technique still outperforms the FR TMQI method remarkably. The success of generalized linear models (GLZ) has been already demonstrated in the qualitative assessment of multi modal distortions in digital images [45], [46]. Apart from the logistic function, we also make use of the GLZ to fit objective quality scores to subjective human ratings followed by computing the performance indices. The associated results are provided in Table I for comparison.

TABLE III
STATISTICAL SIGNIFICANCE COMPARISON OF BTMQI_9, BTMQI_{10}, AND BTMQI_{11} WITH Z-TEST ON THE TMID DATABASE

TMID database [26]						
$Model\text{-}1$: $Model\text{-}2$	53%-47%	60%-40%	67%-33%	73%-27%	80%-20%	87%-13%
BTMQI_{10}: BTMQI_9	+1	+1	+1	+1	+1	+1
BTMQI_{11}: BTMQI_9	+1	+1	+1	+1	+1	+1
BTMQI_{11}: BTMQI_{10}	+1	+1	+1	0	0	0

TABLE IV
MEAN COMPUTATIONAL TIME OF TESTING MODELS (IN SECOND/IMAGE) ON ALL THE 120 IMAGES IN THE TMID DATABASE

TMID database [26]								
Model	TMQI	DIIVINE	BLIINDS-II	BRISQUE	NFERM	NIQE	QAC	BTMQI
Time (sec.)	0.2588	17.772	54.805	0.2768	36.417	0.3229	0.1028	0.2425

Due to the fact that our approach is constructed by three components (i.e., information entropy, statistical naturalness and structural preservation), the contribution of each group of features deserves a quantified comparison. Apart from the 80% train—20% test procedure, other train-test procedures with 53%-47%, 60%-40%, 67%-33%, 73%-27%, 87%-13%, are used to validate and compare three distinct combinations of features in the proposed BTMQI method. We illustrate the results in Table II. BTMQI_9, BTMQI_{10} and BTMQI_{11} are respectively developed based on the first group of 9 features, the first and second groups of 10 features, and all the three groups of 11 features. Results show that a good blind quality measure has been obtained by only using the first group of 9 simple features regarding information entropy. Even with about 50% training samples, BTMQI_9 with 9 features is still statistically equivalent to the FR TMQI method. Also, we can observe that a higher performance can be acquired by adding the other two types of features regarding statistical naturalness and structural preservation. Furthermore, results also confirm that the model performance grows as the training data size increases.

Besides evaluations and comparisons based on the three performance indices mentioned above, we further apply the z-test after the Fisher transform [44] to evaluate the statistical significance on the SROCC results of testing IQA models, since the test is conducted in the same database. Results tell that the proposed BTMQI model is statistically better than all the testing state-of-the-art NR-IQA methods and the FR TMQI approach. It is worthy to emphasize that, in addition to the advanced performance, our BTMQI extracts only 11 features, much less than the NR-IQA metrics tested. In accordance with the suggestion given by the VQEG, the Bonferroni correction is also used to compare the statistical significance between each of the testing IQA models and our BTMQI metric. We can obtain the same results, confirming the effectiveness of the proposed method once more.

Furthermore, a comprehensive comparison of statistical significance among BTMQI_9, BTMQI_{10} and BTMQI_{11} with six train-test procedures is conducted using the z-test. We show the associated results in Table III. The null hypothesis shows that the mean correlation for $Model$-1 is statistically equivalent to that for $Model$-2. A value of "+1" means that $Model$-1 is statistically superior to $Model$-2, whereas a "-1" means that $Model$-1 is statistically worse than $Model$-2. A value of "0" means that $Model$-1 and $Model$-2 are statistically indistinguishable (or equivalent) to each other. From Table III, we can conclude that the BTMQI_9 with 9 features all related to information entropy has already given good results, but the 10th statistical naturalness feature completely brings performance advance in statistical significance at all the train-test procedures. In contrast, the 11th structural preservation feature also induces performance gain in statistical significance to some extent, at half of all six train-test procedures. Likewise, the Bonferroni correction is also used here to compare the statistical significance among BTMQI_9, BTMQI_{10} and BTMQI_{11} metrics with six train-test procedures. The same results with those illustrated in Table III can be derived.

A good quality metric should be simultaneously effective and efficient. Therefore we calculate the implementation time of each testing IQA model on the whole 120 tone-mapped images in the TMID database as well. The test is conducted by MATLAB on a computer with 3.40 GHz CPU processor and 4 GB memory. Table IV tabulates the average time of each metric. With a series computing, the proposed BTMQI algorithm merely takes less than one fourth second to assess an image. Since each of the features used in our method is independent of others in the computation, parallel computing is possibly introduced to reduce the implementation time to a large extent.

C. Cross-Validation Using New Database

It has been so far validated on the TMID database that our BTMQI metric has acquired superior performance and it is immune to the influence of varying image contents. Notice that, besides the two points above, a good tone-mapped IQA model should be also robust across distinct TMOs. Therefore, we in this research propose a new and complementary tone-mapped

(a) (b) (c)

Index		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Overall score	(a)	202	13	177	141	94	174	26	111	36	261	141	202	186	252	164	220
	(b)	230	0	252	115	72	238	212	102	39	204	83	221	56	182	255	139
	(c)	219	13	248	141	113	59	116	84	26	242	126	239	117	236	239	182
Mean		217	9	226	132	93	157	118	99	34	236	117	221	120	223	219	180
Standard deviation		12	6	34	12	17	74	76	11	6	24	25	15	53	30	40	33

(d)

Fig. 6. Illustration of the TMID2015 database: (a)–(c) exemplary images; (d) the final score of each testing LDR image.

image database (TMID2015), which involves three image subsets and associated 48 tone-mapped LDR images created by 16 different TMOs [47].[5] Due to the limitation of display, we show three exemplary images chosen from each data set in Fig. 6(a)–(c). Twenty people were invited to participate in our test. They are college students, consisting of 13 males and 7 females. The entire process of testing was conducted in a dark environment, according to the ITU-R BT.500-13 [48]. When scoring, each image pair is displayed in parallel on the monitor for a duration of 3 seconds.

The paired comparison method that has lately gained much attention is used to rank each pair of tone-mapped images. There are 120 testing image pairs for each subset, and 360 pairs for all the three sets. The subjects are inexperienced, so 56 image pairs, which are from 16 images in the two subsets (#5 and #10) in the TMID database, are applied to train the viewers first. The two subsets have different scenes, each one for the outdoor forest and the indoor kitchen. As thus, each observer should mark a total of 416 randomly presented image pairs during about 20 minutes. A specially designed interactive system can help users adopt only two adjacent keys to offer their opinions, without tedious mouse operations, and this makes the scoring process faster and subjective ratings more reliable. We want to stress that, with up to 16 different TMOs and the popular paired comparison method, our TMID2015 database can be considered as a good complement to the existing TMID database for cross-validation.

At the beginning, the score for each image was set as 0. During the testing, if the image A has a higher subjective score than the image B, the score of A will be increased by 1. We show the final score of each image in the TMID2015 database in Fig. 6(d). Two important conclusions can be derived from the results. First, the final scores for most images created by the same TMO is similar, which illustrates the robustness of subjective ratings in our test. Second, the effectiveness and constancy of each TMO can be measured by the mean and standard deviation values of these opinion scores. For example, as given in Fig. 6(d), the eleventh TMO has a similar mean score but much smaller standard deviation than the seventh TMO, which represents the eleventh TMO has a better constancy; the tenth TMO has a similar standard deviation but larger mean score than the eleventh TMO, which indicates the tenth TMO is more effective. In general, the performance of a TMO can be automatically and precisely judged by a good objective quality metric that has a high correlation with subjective scores.

Using the TMID2015 database, we testify and compare the correlation performance of our blind BTMQI model and seven competing IQA methods. As illustrated in Table V, the performance results that respectively correspond to logistic function and the GLZ on the whole database. Note that the cross validation is conducted; that is our BTMQI metric is trained first on the TMID database and then tested on the TMID2015 database. Since the correlations are obtained from completely different databases, we also use the z-test and the Bonferroni correction to measure the statistical significance between our technique and each of other testing IQA metrics. As expected, results show that the proposed model is superior to other testing ones in statistical significance, which means our BTMQI method works effectively across different image contents, TMOs and subjective assessment methods.

D. Analysis and Future Work

The main positive contribution in this paper is to propose a blind metric for tone-mapped images. To our knowledge, this model is the first of this type. By performance measures and comparisons above, the proposed blind BTMQI model is shown to be of high performance, beyond the state-of-the-art NR-IQA metrics and the FR TMQI algorithm that requires the help of HDR images, and be of a low computational cost. A good IQA algorithm is expected to be used in processing systems and applications. In contrast to the FR TMQI model which can be used to convert HDR images to standard LDR versions, our blind metric can serve as a quality monitor during the transmission and at the terminal, where the HDR images are not accessible. More analyses concerning this are given in the Appendix.

[5]All the tone-mapped images in the TMID2015 database are produced by Čadík *et al.* [47]. Our contribution is to obtain the associated subjective scores.

TABLE V
PERFORMANCE COMPARISON OF ALL TESTING IQA MODELS ON THE TMID2015 DATABASE

TMID2015 database (Four-parameter logistic function)								
Model	TMQI	DIIVINE	BLIINDS-II	BRISQUE	NFERM	NIQE	QAC	BTMQI
PLCC	**0.6417**	0.3383	0.1493	0.0143	0.1757	0.4151	0.2659	**0.6824**
SROCC	**0.5551**	0.3862	0.1235	0.1247	0.1580	0.4223	0.2204	**0.7061**
KROCC	**0.3943**	0.2697	0.0992	0.0703	0.1349	0.2857	0.1522	**0.5154**
TMID2015 database (Generalized linear models)								
Model	TMQI	DIIVINE	BLIINDS-II	BRISQUE	NFERM	NIQE	QAC	BTMQI
PLCC	**0.6118**	0.3362	0.0180	0.0147	0.2092	0.4036	0.2363	**0.6655**
SROCC	**0.5551**	0.3862	0.1189	0.1247	0.2509	0.4223	0.2204	**0.7061**
KROCC	**0.3943**	0.2697	0.0726	0.0703	0.1700	0.2857	0.1522	**0.5154**

We bold the top two methods.

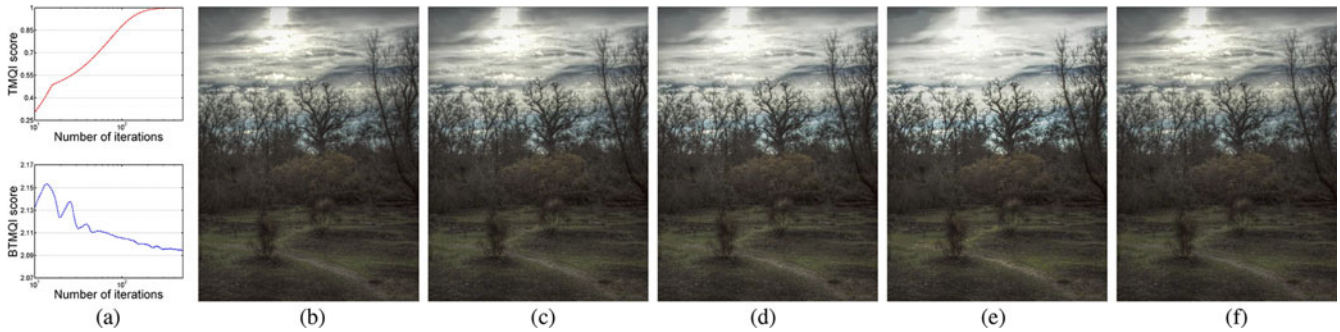


Fig. 7. (a) TMQI and BTMQI scores of each iterative image; (b) initial tone-mapped image; (c) the 10th iteration image; (d) the 50th iteration image; (e) the 200th iteration image; (f) the proposed image.

Furthermore, as mentioned previously, modern NR-IQA algorithms work ineffectively in assessing the visual quality of tone-mapped images. So the robust blind quality metric is derived by properly combining the three types of features used in our BTMQI model into existing NR-IQA metrics.

Despite the encouraging performance results achieved by our method, there are two limitations in this work. One is that the proposed BTMQI metric only takes into consideration luminance information in the tone-mapped image, and ignores the key aspect color reproduction in the perceived quality of tone-mapped images. Actually, some TMOs might produce an image with wrong colors, and thus dramatically destroy the image quality by twisting the natural appearance of the content and/or even making the representation of the content misunderstood. The other limitation is that the halo artifacts are possibly created by some local TMOs. Note that halos oftentimes produce false contours. The third type of structural preservation feature will detect these false contours, and treat them as primary structures. That is to say, the proposed BTMQI model is likely to think of a low-quality tone-mapped image with halos to be high quality. The second type of statistical naturalness feature will give the effect of halos a large punishment to the image quality, relieving the negative effect of structural preservation feature in such condition.

In the future, three improved works will be conducted. The first work is to include the chromatic component into the proposed BTMQI model. The second one is to devise a new reliable strategy or improve the existing structural preservation feature, so as to faithfully assess the quality of tone-mapped images with halos. And the third one is to develop adaptive multipliers in the computation of the first set of information entropy features in BTMQI according to the attributes of the input tone-mapped images.

IV. CONCLUSION

In this paper, we have investigated an emerging research topic—quality assessment of tone-mapped images. First, as a good complement to the existing TMID database, we have proposed a new tone-mapped image database (TMID2015) with up to 16 different TMOs in order to demonstrate the performance accuracy of IQA models across TMOs. Second, we have developed new and efficient features for measuring the information entropy of luminance-changed images. With this type of features, we have derived an effective NR IQA model that is statistically equivalent to the FR TMQI method. Third, by introducing two other kinds of classical features regarding statistical naturalness and structural preservation, we have put forward a high-accuracy blind quality measure, which is statistically superior to the state-of-the-art NR-IQA models and the TMQI metric on the TMID and TMID2015 databases, and is of a substantially low computational cost. Our source code and the TMID2015 database will be released soon at https://sites.google.com/site/guke198701/publications.

APPENDIX

A direct application of tone-mapped IQA is to select the best image from many candidates created by different TMOs,

especially when the HDR image is not accessible. Another application is to utilize the tone-mapped image quality metric as the optimization target to develop a novel TMO. Rather than proposing a computational tone-mapping framework, this application is devoted to treating tone mapping as an optimization problem in the image space and designing an iterative search approach that starts from any initial image and moves step-by-step in the image space towards the direction of improving the structural fidelity measure until a (local) maximal point is reached [49]. To specify, assuming TMQI to be the quality measure of tone mapped images, the problem of optimal tone mapping can be formulated as

$$\mathbf{Y}_{\text{opt}} = \arg\max_{\mathbf{Y}} \mathbf{Q}(\mathbf{X}, \mathbf{Y}) \tag{10}$$

where $\mathbf{Q}(\cdot)$ represents the TMQI method, and $\mathbf{X}$ and $\mathbf{Y}$ are a HDR image and its associated ton-mapped image respectively. It is noted that this is a challenging problem because of the complexity of TMQI and the high dimensionality. To do this, an iterative approach based on gradient descent is applied.

In this paper, this optimization algorithm of 500 iterations is employed to improve a tone-mapped image. We present the TMQI and BTMQI scores across each iterative image in Fig. 7(a). Despite that the TMQI score is constantly growing with the increase of iterations, our BTMQI shows different results. The proposed BTMQI gives the best score on about the 10th iteration. In reality, we find in Figs. 7(b)–(e) that the 10th iterative result shows better visual quality than others, for example, the luminance variations of clouds. This phenomenon motivates the use of BTMQI as a terminal sign to avoid the over-iteration. We illustrate the initial, 10th, 50th, 200th and BTMQI-based optimal iterative images in Figs. 7(b)–(f). The proposed BTMQI-based optimal iterative algorithm has achieved better subjective image quality.

ACKNOWLEDGMENT

The authors would like to thank the associate editor and the anonymous reviewers for their constructive comments, which greatly helped in improving this paper.

REFERENCES

[1] E. Reinhard, G. Ward, S. Pattanaik, P. Debevec, W. Heidrich, and K. Myszkowski, *High Dynamic Range Imaging: Acquisition, Display, and Image-Based Lighting*, San Mateo, CA, USA: Morgan Kaufmann, 2010.

[2] R. Mantiuk, A. Efremov, K. Myszkowski, and H.-P. Seidel, "Backward compatible high dynamic range MPEG video compression," *ACM Trans. Graph.*, vol. 15, no. 9, pp. 713–723, Jul. 2006.

[3] Z. Mai *et al.*, "Optimizing a tone curve for backward-compatible high dynamic range image and video compression," *IEEE Trans. Image Process.*, vol. 20, no. 6, pp. 1558–1571, Jun. 2011.

[4] G. W. Larson, H. Rushmeier, and C. Piatko, "A visibility matching tone reproduction operator for high dynamic range scenes," *IEEE Trans. Vis. Comput. Graph.*, vol. 3, no. 4, pp. 291–306, Oct. 1997.

[5] F. Durand and J. Dorsey, "Fast bilateral filtering for the display of high dynamic range images," *ACM Trans. Graph.*, vol. 21, no. 3, pp. 257–266, Jul. 2002.

[6] F. Drago, K. Myszkowski, T. Annen, and N. Chiba, "Adaptive logarithmic mapping for displaying high contrast scenes," *Comput. Graph. Forum*, vol. 22, no. 3, pp. 419–426, Sept. 2003.

[7] S. Wang, A. Rehman, Z. Wang, S. Ma, and W. Gao, "SSIM-motivated rate distortion optimization for video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 4, pp. 516–529, Apr. 2012.

[8] S. Wang, A. Rehman, Z. Wang, S. Ma, and W. Gao, "Perceptual video coding based on SSIM-inspired divisive normalization," *IEEE Trans. Image Process.*, vol. 22, no. 4, pp. 1418–1429, Apr. 2013.

[9] H. R. Wu, A. Reibman, W. Lin, F. Pereira, and S. S. Hemami, "Perceptual visual signal compression and transmission," *Proc. IEEE*, vol. 101, no. 9, pp. 2025–2043, Sep. 2013.

[10] K. Gu, G. Zhai, X. Yang, W. Zhang, and C. W. Chen, "Automatic contrast enhancement technology with saliency preservation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 9, pp. 1480–1494, Sep. 2015.

[11] K. Gu, G. Zhai, W. Lin, and M. Liu, "The analysis of image contrast: From quality assessment to automatic enhancement," *IEEE Trans. Cybern.*, vol. 46, no. 1, pp. 284–297, Jan. 2016.

[12] 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.

[13] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multi-scale structural similarity for image quality assessment," in *Proc. IEEE Asilomar Conf. Signals, Syst., Comput.*, Nov. 2003, pp. 1398–1402.

[14] K. Gu, M. Liu, G. Zhai, X. Yang, and W. Zhang, "Quality assessment considering viewing distance and image resolution," *IEEE Trans. Broadcast.*, vol. 61, no. 3, pp. 520–531, Sep. 2015.

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

[16] 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.

[17] 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.

[18] G. Zhai, X. Wu, X. Yang, W. Lin, and W. Zhang, "A psychovisual quality metric in free-energy principle," *IEEE Trans. Image Process.*, vol. 21, no. 1, pp. 41–52, Jan. 2012.

[19] K. Gu, G. Zhai, X. Yang, and W. Zhang, "A new reduced-reference image quality assessment using structural degradation model," in *Proc. IEEE Int. Symp. Circuits Syst.*, May 2013, pp. 1095–1098.

[20] 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.

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

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

[23] R. Mantiuk, S. J. Daly, K. Myszkowski, and H.-P. Seidel, "Predicting visible differences in high dynamic range images: Model and its calibration," in *Proc. SPIE, Human Vis. Electron. Imaging X*, Mar. 2005, vol. 5666, pp. 204–214.

[24] R. Mantiuk, K. J. Kim, A. G. Rempel, and W. Heidrich, "HDR-VDP-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions," *ACM Trans. Graph.*, vol. 30, no. 4, pp. 1–13, Jul. 2011.

[25] M. Narwaria, R. Mantiuk, M. P. D. Silva, and P. L. Callet, "HDR-VDP-2.2: A calibrated method for objective quality prediction of high-dynamic range and standard images," *J. Electron. Imaging*, vol. 24, no. 1, pp. 010501–010501, Jan. 2015.

[26] H. Yeganeh and Z. Wang, "Objective quality assessment of tone-mapped images," *IEEE Trans. Image Process.*, vol. 22, no. 2, pp. 657–667, Feb. 2013.

[27] V. Mante, R. Frazor, V. Bonin, W. 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, Nov. 2005.

[28] K. Gu, G. Zhai, M. Liu, X. Yang, and W. Zhang, "Details preservation inspired blind quality metric of tone mapping methods," in *Proc. IEEE Int. Symp. Circuits Syst.*, Jun. 2014, pp. 518–521.

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

[30] "HDR shop," Inst. for Creative Technol., Univ. of Southern California, Playa Vista, CA, USA, (2013). [Online]. Available: http://www.hdrshop.com/.
[31] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 379–423, Oct. 1948.
[32] S. Kullback and R. A. Leibler, "On information and sufficiency," *Ann. Math. Statist.*, vol. 22, pp. 79–86, 1951.
[33] D. H. Johnson and S. Sinanović, "Symmetrizing the Kullback-Leibler distance," *IEEE Trans. Inf. Theory*, Mar. 2001.
[34] K. Gu, G. Zhai, X. Yang, and W. Zhang, "An efficient color image quality metric with local-tuned-global model," in *Proc. IEEE Int. Conf. Image Process.*, Oct. 2014, pp. 506–510.
[35] S. Goferman, L. Zelnik-Manor, and A. Tal, "Context-aware saliency detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 10, pp. 1915–1926, Oct. 2012.
[36] E. P. Simoncelli and B. A. Olshausen, "Natural image statistics and neural representation," *Annu. Rev. Neurosci.*, vol. 24, pp. 1193–1216, May 2001.
[37] M. Čadík and P. Slavík, "The naturalness of reproduced high dynamic range images," in *Proc. Int. Conf. Inf. Visual.*, 2005, pp. 920–925.
[38] G. Schaefer and M. Stich, "UCID—An uncompressed colour image database," in *Proc. SPIE, Storage Retrieval Methods Appl. Multimedia*, 2004, pp. 472–480, [Online]. Available: http://www-staff.lboro.ac.uk/∼cogs/datasets/UCID/ucid.html.
[39] 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.
[40] B. Schökopf, A. J. Smola, R. C. Williamson, and P. L. Bartlett, "New support vector algorithms," *Neural Comput.*, vol. 12, no. 5, pp. 1207–1245, 2000.
[41] S. Samanta and B. Chanda, "Space-time facet model for human activity classification," *IEEE Trans. Multimedia*, vol. 16, no. 6, pp. 1525–1535, Oct. 2014.
[42] Y.-C. Su, T.-H. Chiu, Y.-H. Kuo, C.-Y. Yeh, and W. H. Hsu, "Scalable mobile visual classification by kernel preserving projection over high-dimensional features," *IEEE Trans. Multimedia*, vol. 16, no. 6, pp. 1645–1653, Oct. 2014.
[43] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, 2011.
[44] VQEG, "Final report from the video quality experts group on the validation of objective models of video quality assessment," Mar. 2000. [Online]. Available: http://www.vqeg.org/.
[45] L. Janowski and Z. Papir, "Modeling subjective tests of quality of experience with a generalized linear model," in *Proc. IEEE Int. Workshops Quality Multimedia Experience*, Jul. 2009, pp. 35–40.
[46] A. Głowacz *et al.*, "Automated qualitative assessment of multi-modal distortions in digital images based on GLZ," *Ann. Telecommun.*, vol. 65, no. 1, pp. 3–17, Feb. 2010.
[47] M. Čadík, O. Hajdok, A. Lejsek, O. Fialka, M. Wimmer, A. Artusi, and L. Neumann, "Evaluation of tone mapping operators," 2010. [Online]. Available: http://cadik.posvete.cz/tmo/.
[48] *Methodology for the subjective assessment of the quality of television pictures*, ITU-R BT.500-13, 2012.
[49] K. Ma, H. Yeganeh, K. Zeng, and Z. Wang, "High dynamic range image tone mapping by maximizing tone mapped image quality index," in *Proc. IEEE Int. Conf. Multimedia Expo*, Jul. 2014 , pp. 1–6 .

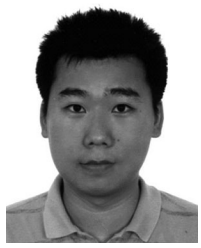
Ke Gu received the B.S. and Ph.D. degrees in electronic engineering from Shanghai Jiao Tong University, Shanghai, China, in 2009 and 2015.

From July to November 2014, he was a Visiting Student with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada. From December 2014 to January 2015, he was a Visiting Student with the School of Computer Engineering, Nanyang Technological University, Singapore. From February to March 2015, he was a Visiting Student with the Department of Computer Science and Technology, Peking University, Beijing, China. His research interests include quality assessment, contrast enhancement, and visual saliency.

Mr. Gu is a Reviewer for the IEEE TRANSACTIONS ON IMAGE PROCESSING, the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, the IEEE TRANSACTIONS ON MULTIMEDIA, the IEEE TRANSACTIONS ON CYBERNETICS, the IEEE TRANSACTIONS ON BROADCASTING, *Signal Processing Letters*, *Neurocomputing*, *Signal Processing: Image Communication*, *Journal of Visual Communication and Image Representation*, *Signal*, *Image and Video Processing*, and *IET Image Processing*.

Shiqi Wang (M'15) received the B.S. degree in computer science from the Harbin Institute of Technology, Harbin, China, in 2008, and the Ph.D. degree in computer application technology from Peking University, Beijing, China, in 2014.

He is currently a Postdoc Fellow with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada. From April 2011 to August 2011, he was with Microsoft Research Asia, Beijing, China, as an Intern. He has proposed more than 20 technical proposals to ISO/MPEG, ITU-T, and AVS video coding standards. His current research interests include video compression and image/video quality assessment.



Guangtao Zhai (M'10) received the B.E. and M.E. degrees from Shandong University, Shandong, China, in 2001 and 2004, respectively, and the Ph.D. degree from Shanghai Jiao Tong University, Shanghai, China, in 2009. From 2008 to 2009, he was a Visiting Student with the Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON, Canada, where he was a Post-Doctoral Fellow from 2010 to 2012.

From 2012 to 2013, he was a Humboldt Research Fellow with the Institute of Multimedia Communication and Signal Processing, Friedrich Alexander University of Erlangen-Nuremberg, Erlangen, Germany. He is currently a Research Professor with the Institute of Image Communication and Information Processing, Shanghai Jiao Tong University. His research interests include multimedia signal processing and perceptual signal processing.

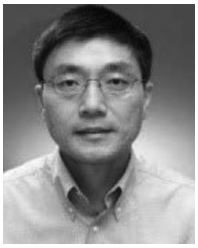
Prof. Zhai was the recipient of the National Excellent Ph.D. Thesis Award from the Ministry of Education of China in 2012.

Siwei Ma (M'12) received the B.S. degree from Shandong Normal University, Jinan, China, in 1999, and the Ph.D. degree in computer science from the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China, in 2005.

From 2005 to 2007, he was a Post-Doctoral Researcher with the University of Southern California, Los Angeles, CA, USA. He then joined the Institute of Digital Media, Department of Electrical Engineering and Computer Science, Peking University, Beijing, China, where he is currently an Associate Professor. He has authored or coauthored over 70 technical articles in refereed journals and proceedings in the areas of image and video coding, video processing, video streaming, and transmission.

Xiaokang Yang (A'10–SM'04) received the B.S. degree from Xiamen University, Xiamen, China, in 1994, the M.S. degree from the Chinese Academy of Sciences, Shanghai, China, in 1997, and the Ph.D. degree from Shanghai Jiao Tong University, Shanghai, China, in 2000.

He is currently a Full Professor and Deputy Director of the Institute of Image Communication and Information Processing, Department of Electronic Engineering, Shanghai Jiao Tong University. From September 2000 to March 2002, he was a Research Fellow with the Centre for Signal Processing, Nanyang Technological University, Singapore. From April 2002 to October 2004, he was a Research Scientist with the Institute for Infocomm Research, Singapore. He has authored or coauthored over 80 refereed papers, and has filed six patents. His current research interests include video processing and communication, media analysis and retrieval, perceptual visual processing, and pattern recognition.



Weisi Lin (S'91–M'92–SM'00–F'16) received the Ph.D. degree from King's College London, London, U.K, in 1993.

He is currently an Associate Professor with the School of Computer Engineering, Nanyang Technological University, Singapore. He has authored or coauthored more than 340 refereed papers published in international journals and conferences. His research interests include image processing, visual quality evaluation, and perception-inspired signal modeling.

Prof. Lin is a Fellow of Institution of Engineering Technology and an Honorary Fellow of the Singapore Institute of Engineering Technologists. He has been on the Editorial Board of the IEEE TRANSACTIONS ON IMAGE PROCESSING, the IEEE TRANSACTIONS ON MULTIMEDIA (2011–2013), the IEEE SIGNAL PROCESSING LETTERS, and the *Journal of Visual Communication and Image Representation*. He has been elected as an APSIPA Distinguished Lecturer (2012, 2013). He served as a Technical-Program Chair for the Pacific-Rim Conference on Multimedia 2012, the IEEE INTERNATIONAL CONFERENCE ON MULTIMEDIA AND EXPO 2013, and the International Workshop on Quality of Multimedia Experience 2014.

Wenjun Zhang (M'02–SM'10–F'12) received the B.S., M.S., and Ph.D. degrees in electronic engineering from Shanghai Jiao Tong University, Shanghai, China, in 1984, 1987, and 1989, respectively.

From 1990 to 1993, he was a Post-Doctoral Fellow with the Philips Kommunikation Industrie AG, Nuremberg, Germany. He joined the Faculty of Shanghai Jiao Tong University in 1993 and became a Full Professor with the Department of Electronic Engineering in 1995. He has authored or coauthored more than 90 papers published in international journals and conferences, and holds more than 40 patents. His main research interests include digital video coding and transmission, multimedia semantic processing, and intelligent video surveillance.

Prof. Zhang is a Chief Scientist of the Chinese National Engineering Research Centre of Digital Television (NERC-DTV) and the Chair of the Future of Broadcast Television Initiative (FOBTV) Technical Committee.

Wen Gao (S'87–M'88–SM'05–F'09) received the Ph.D. degree in electronics engineering from the University of Tokyo, Tokyo, Japan, in 1991.

He is currently a Professor of Computer Science with Peking University, Beijing, China. Before joining Peking University, he was a Professor of Computer Science with the Harbin Institute of Technology, Harbin, China, from 1991 to 1995, and a Professor with the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China. He has authored or coauthored five books and over 600 technical articles in refereed journals and conference proceedings in the areas of image processing, video coding and communication, pattern recognition, multimedia information retrieval, multimodal interface, and bioinformatics.

Dr. Gao has served or serves on the Editorial Board of several journals, such as the IEEE TRANSACTIONS ON IMAGE PROCESSING, the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, the IEEE TRANSACTIONS ON MULTIMEDIA, the IEEE TRANSACTIONS ON AUTONOMOUS MENTAL DEVELOPMENT, the *EURASIP Journal of Image Communications*, and the *Journal of Visual Communication and Image Representation*. He has chaired a number of prestigious international conferences on multimedia and video signal processing, such as the IEEE ICME and ACM Multimedia, and also has served on the advisory and technical committees of numerous professional organizations.