

# Hybrid No-Reference Quality Metric for Singly and Multiply Distorted Images

Ke Gu, Guangtao Zhai, Xiaokang Yang, *Senior Member, IEEE*, and Wenjun Zhang, *Fellow, IEEE*

**Abstract**—In a typical image communication system, the visual signal presented to the end users may undergo the steps of acquisition, compression and transmission which cause the artifacts of blurring, quantization and noise. However, the researches of image quality assessment (IQA) with multiple distortion types are very limited. In this paper, we first introduce a new multiply distorted image database (MDID2013), which is composed of 324 images that are simultaneously corrupted by blurring, JPEG compression and noise injection. We then propose a new six-step blind metric (SISBLIM) for quality assessment of both singly and multiply distorted images. Inspired by the early human visual model and recently revealed free energy based brain theory, our method works to systematically combine the single quality prediction of each emerging distortion type and joint effects of different distortion sources. Comparative studies of the proposed SISBLIM with popular full-reference IQA approaches and start-of-the-art no-reference IQA metrics are conducted on five singly distorted image databases (LIVE, TID2008, CSIQ, IVC, Toyama) and two newly released multiply distorted image databases (LIVEMD, MDID2013). Experimental results confirm the effectiveness of our blind technique. MATLAB codes of the proposed SISBLIM algorithm and MDID2013 database will be available online at <http://gvsp.sjtu.edu.cn/>.

**Index Terms**—Image quality assessment (IQA), blind/no-reference (NR), multiply distortion types, human visual system (HVS), joint effects, free energy.

## I. INTRODUCTION

IMAGE quality assessment (IQA) is an important topic for both the scientific research and applicational development of digital image processing systems. IQA can serve as a performance measure as well as an optimization criterion for various approaches in computer vision and image processing, such as image/video compression [1]–[4], restoration [27], denoising [32] and enhancement [5]. The peak signal-to-noise ratio (PSNR) has prevailed for decades as the most popular IQA algorithm before gradually giving way to the structural similarity index (SSIM) [6] emerged a decade ago. Based on the

Manuscript received October 20, 2013; revised February 26, 2014; accepted April 8, 2014. Date of publication August 15, 2014; date of current version September 3, 2014. This work was supported in part by the National Science Foundation of China under Grant 61025005, Grant 61371146, Grant 61221001, and Grant 61390514, in part by the Foundation for the Author of National Excellent Doctoral Dissertation of China under Grant 201339, and in part by the Major State Basic Research Development Program of China (973 Program) under Grant 2010CB731401.

The authors are with the Institute of Image Communication and Information Processing, Shanghai Key Laboratory of Digital Media Processing and Transmissions, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: gukesjtuee@sjtu.edu.cn; zhaiguangtao@sjtu.edu.cn; xkyang@sjtu.edu.cn; zhangwenjun@sjtu.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/TBC.2014.2344471

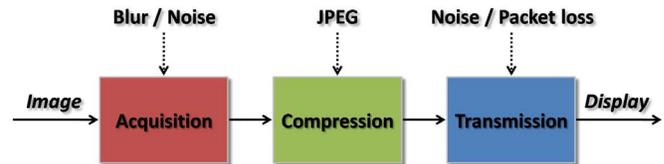


Fig. 1. Illustration of three steps (acquisition, compression, and transmission) which images usually undergo before the final display in front of consumers.

hypothesis that the perception of the human visual system (HVS) is highly capable of extracting structural information from an image, SSIM obtains better performance than PSNR on the LIVE database [7].

IQA approaches are typically divided into two categories, subjective assessment and objective assessment. According to a series of recommendations given by VQEG and ITU (e.g., the ITU-R BT.500 [8]), quite a few image quality databases with subjective ratings have been released during the past decade, and this largely promoted the research of IQA. Hundreds of IQA methods using a variety of perceptual/statistical models were developed [9]–[24]. For the scenario of full-reference (FR) IQA, very high performance in terms of the correlation between subjective scores and objective quality predictions have been achieved by those improved SSIM-type of methods, e.g., the multi-scale SSIM (MS-SSIM) [9], the natural scene statistics (NSS) inspired information content weighted SSIM (IW-SSIM) [14], and the newly proposed structural similarity weighted SSIM (SW-SSIM) [23].

Besides FR IQA algorithms, last several years have seen a surge of reduced-reference (RR) IQA metrics [25]–[29] and no-reference (NR) IQA metrics [30]–[34]. Those RR and NR IQA methods can be roughly divided into two types. The first is motivated by the recent findings in brain science, such as RR free energy based distortion metric (FEDM) and NR free energy based quality metric (NFEQM) [26]. These approaches were designed to model the internal generative mechanism of human brain. We lately proposed NR Free energy and Structural degradation model based Distortion Metric (NFSDM) [33] by integrating a pair of RR IQA algorithms (FEDM and structural degradation model (SDM) [28]).

The second type of IQA metrics targets to predict the image quality through characterizing the NSS regularity. Distortion Identification-based Image Verity and Integrity Evaluation (DIIVINE) [30] of 88 features works in the DWT domain by distortion identification before distortion-specific quality assessment. BLind Image Integrity Notator using DCT Statistics (BLIINDS-II) [31] and Blind/Referenceless Image

Spatial Quality Evaluator (BRISQUE) [32] were respectively designed in DCT and spatial domains with less features. Instead of evaluating the image quality by the regression module from subjective scores of the training samples, natural image quality evaluator (NIQE) [34] was recently developed without using human scored images.

Additionally, there exists another class of blind metrics devoted to specific distortion types [35]–[43]. In early attempts, Marziliano *et al.* developed a Blind Blur Metric (MBBM) [35] via vertical and horizontal edge detectors. Wang *et al.* proposed a No-reference JPEG-quality Evaluator (WNJE) [36] to measure blocking effects and relative blur. The fast image sharpness (FISH and FISH<sub>bb</sub>) method [37] was recently designed for blurriness estimation based on the log-energy in high frequency DWT subbands. Except blind measures for blur and JPEG compression, noise estimation of important application in denoising has nowadays been widely researched and has aroused several valid approaches, such as scale invariant based noise estimator (SINE) [38] and weak textured patches based noise evaluator (WTPNE) [39].

Despite the prosperity and successfulness of IQA studies, most image quality metrics can only deal efficiently with images of single distortion type. But the outputs of practical image processing/communication systems are usually contaminated by more than one distortion source, and to facilitate the IQA research along this line, Jayaraman *et al.* recently released a new LIVE multiply distorted image database (LIVEMD) [44] that includes two groups of doubly distorted images for two scenarios: 1) image storage, where images are first blurred and then compressed by a JPEG encoder; 2) camera image acquisition, where images are first blurred due to defocusing and then corrupted by white noise to simulate sensor noise. It is easy to imagine that multiple distortions cause bigger trouble for the HVS to interpret the useful image content.

In real image communication systems, images usually undergo three steps - acquisition, compression and transmission - before finally reaching to end consumers, as illustrated in Fig. 1. This makes images very likely to be contaminated together with artifacts of JPEG compression, blurring and noise injection. To faithfully model this real-world mixed artifacts, we in this paper introduce a new multiply distorted image database (MDID2013), which consists of 324 testing images simultaneously corrupted by three distortion types mentioned above and associated subjective human ratings obtained from twenty-five inexperienced observers.

In [45], an early model utilizes the low-pass and high-pass filters as well as non-linear transfer functions to simulate the HVS. This implies that human beings can separately perceive the degree of each type of distortion from mixed artifacts. In addition, Chandler reviewed some experiments and pointed out a few possible joint effects of various distortion types [46]. To comprehensively take into account aforementioned factors, we in this paper propose a new Six-Step BLind Metric (SISBLIM)<sup>1</sup> for quality assessment of multiply

distorted images by combining the single quality prediction of each emerging distortion type and joint effects of different distortion sources. The proposed blind IQA metric is consisted by six parts, noise estimation, image denoising, blur measure, JPEG-quality evaluator, joint effects' prediction, and HVS based fusion. Specifically, the noise level of an input image is measured first, followed by a possible denoising operation depending on the existence of additive noise. The noiseless or denoised image is then independently assessed with blur and JPEG methods. The joint effect is evaluated based on the free energy principle [26]. Finally the quality score predicted by SISBLIM is computed as a systematic integration of sub-scores for the artifacts of additive noise, blurring, JPEG compression and the joint effect. It will be shown that our blind metric is extremely effective against many mainstream FR IQA approaches and state-of-the-art NR IQA metrics on both singly and multiply distorted image quality databases.

The remainder of this paper is organized as follows. Section II first introduces the new MDID2013 database for IQA of multiple distortions. In Section III, we analyze some possible working mechanisms of the HVS when performing IQA before presenting SISBLIM in detail. In Section IV, experimental results and comparative studies using seven image databases (LIVE [7], TID2008 [48], CSIQ [49], IVC [50], Toyama [51], LIVEMD [44], and MDID2013) are reported and discussed. Finally, Section V concludes this paper.

## II. MDID2013 DATABASE

As argued in a practical image communication system such as outlined in Fig. 1, images usually undergo the stages of acquisition, compression and transmission, are presumably

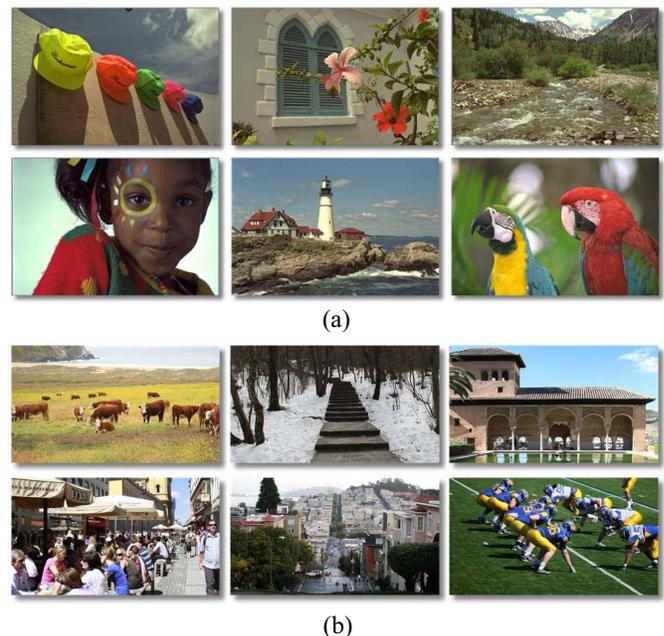


Fig. 2. Twelve lossless natural color images. (a) Six standard definition images of size  $768 \times 512$  from Kodak database [52]. (b) Six high definition images of size  $1280 \times 720$  from LIVEMD [44].

<sup>1</sup>A preliminary version of the proposed algorithm has been introduced in a conference paper [47]. The SISBLIM presented in this paper is an improved version of the FISBLIM in [47].

distorted with the artifacts of Gaussian blurring, JPEG compression and white noise injection in order. The MDID2013 database is introduced to simulate this process. Images in MDID2013 come from 12 pristine images. Fig. 2 shows the elaborately selected image sources: one half of images of size  $768 \times 512$  from Kodak database [52], and the other half images of size  $1280 \times 720$  from LIVEMD database [44]. They span a wide range of scenes, colors, illumination levels and foreground/background configurations. The overall 324 testing images are generated by successively corrupting each original image with blur, JPEG compression and noise. Details of the artifacts are given below:

- Gaussian blur: We employ Gaussian kernels of standard deviation  $\sigma_G$  with a window of size  $l_G \times l_G$  based on the Matlab *fspecial* and *imfilter* commands. Each of the R, G and B planes was blurred using the same kernel.
- JPEG: We use the Matlab *imwrite* command to create JPEG compressed images depending on  $Q$  parameters.
- White noise: We add a standard normal pdf of variance  $\sigma_N^2$  to each of the three planes R, G and B using the Matlab *imnoise* function.

The parameters stated above are the same with those used in [44], in order to make MDID2013 and LIVEMD databases complementary, since the LIVEMD includes one/double-fold artifacts whereas our MDID2013 only has three-fold distorted images. Furthermore, the utility of the same parameters is also to keep the distorted images perceptually separable from each other and from the references, and to keep the distortions within a realistic range.

We conducted the subjective viewing test with a single-stimulus (SS) method in accordance with ITU-R BT.500-12 [8]. Twenty-five inexperienced subjects participated in this test. Most of these observers were college students with different majors. To automatically display the testing images and collect the raw subjective ratings, we designed and used an interactive system similar to that in [29]. The viewing distance is fixed at four times of the image height to match the conditions in LIVEMD [44]. Note that the testing images have two kinds of image sizes, meaning that those observers have to score them at two viewing distances. To avoid frequently adjusting the distances, we divide the overall test into two consecutive parts. The first part proceeds with all the testing images of size  $768 \times 512$ , while the second part includes images of size  $1280 \times 720$ . We randomized the presentation order in each section to eliminate memory effects on the mean scores as much as possible. During the view session,

TABLE I  
SUBJECTIVE TEST CONDITIONS AND PARAMETERS

Method	Single-stimulus (SS)
Color depth	24-bits/pixel color images
Image coder	Portable Network Graphic (PNG)
Image resolution	$768 \times 512$ , $1280 \times 720$
Subjects	Twenty-five inexperienced viewers
Evaluation scales	Continuous quality scale from 0 to 1
Viewing distance	Four times the image height
Room illuminance	Dark

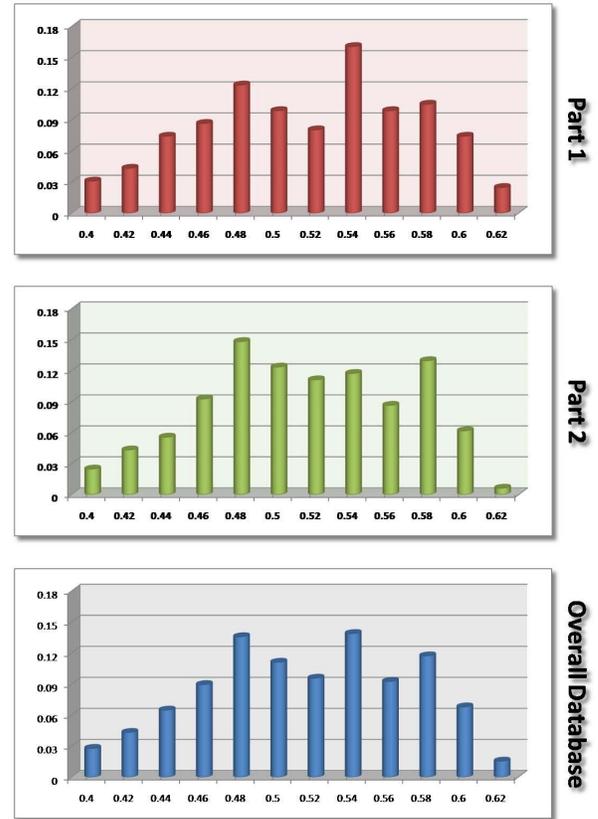


Fig. 3. Histogram of DMOS in each part and the overall database.

the subjects were asked to provide their overall perception of quality on a continuous quality scale from 0 to 1 with the precision up to 0.01%. Table I summarizes the subjective test conditions with some key parameters.

After the viewing test, we calculate the DMOS values for all the testing images. Here, we denote  $s_{ab}$  as the score provided by subject  $a$  to the testing image  $I_b$ , where  $a = \{1, \dots, 20\}$  and  $b = \{1, \dots, 324\}$ , and denote  $s'_{ab}$  as the rating of each original image. The following steps are used then:

- Outliers screening, to improve fidelity of the data.
- Differential scores computing, to acquire the difference between the pair of scores  $d_{ab} = s_{ab} - s'_{ab}$ .
- Averaging, to calculate the DMOS value for the image  $I_b$  as  $\frac{1}{N_A} \sum_x d_{ab}$ , where  $N_A$  is the number of subjects.

Distributions of DMOS scores for the two individual parts and the whole data are exhibited in Fig. 3. Notice that the shapes of distributions for two parts in Fig. 3 are quite similar. This demonstrates the visual quality of multiple distortions is almost immune to the change of image sizes.

### III. SISBLIM ALGORITHM

An early human vision model is constructed by three parts, low-pass filter, logarithmic brightness mapping, and high-pass filter [45]. Inspired by this model, we hypothesize that human beings can immediately perceive the noise level and denoise the image using the low-pass filter when watching a multiply distorted image. Thereafter, human beings can easily estimate the degree of blur and JPEG compression based on the high-pass filter. In other words, the image degradation caused

by each distortion source can be separately perceived. It is obvious that there also exist joint effects of mixed artifacts. In [46], Chandler reviewed some experiments and provided a few possible joint effects. These effects usually further lower the image quality, rendering a smaller score than the direct combination of each single quality score. In practice, the joint effect, because of masking effects, is strongly affected by the image content. Following the method in [53], ‘free energy’ is adopted to measure the descriptive complexity of the distorted image, thereby to approximate the joint effect.

Accordingly, the proposed training-free SISBLIM works to combine the single quality score of each emerging distortion type and the joint effect of mixed artifacts. Fig. 4 shows major steps of SISBLIM, which consists of noise estimation, image denoising, blur measure, JPEG-quality evaluator, joint effect’s prediction, and HVS based fusion model. The building blocks in SISBLIM are used to simulate the above-mentioned perceptual process of the HVS for multiply distorted images. For an input image signal, we first predict the noise variance. Based on this estimated noise level, we apply BM3D [54] to image denoising. Blur and JPEG artifacts are then separately assessed for the noiseless or denoised image, and the joint effect is measured by ‘free energy’. Finally, the image quality is derived by an appropriate integration of estimates of noise, blur, JPEG compression artifacts and the joint effect.

#### A. Noise Estimation

It has been found in [38] that the kurtosis values tend to be invariant across scales for a natural image, and this scale invariance will be deteriorated by the added noise. For an input image signal  $\mathbf{x}$ , the kurtosis of its noisy version  $\mathbf{y}$  can be expressed as a function of kurtosis and variance of  $\mathbf{x}$  and the variance of noise:

$$k_y = \frac{k_x(\alpha) - 3}{(1 + \frac{\sigma_n^2}{\sigma_x^2})^2} + 3 \quad (1)$$

where  $k_x$  and  $k_y$  are the kurtosis values of  $\mathbf{x}$  and  $\mathbf{y}$ , and  $\sigma_x$  and  $\sigma_n$  are the variance values of  $\mathbf{x}$  and the added noise  $\mathbf{n}$ . So the noise level  $\hat{\sigma}_n^2$  of  $\mathbf{y}$  can be estimated by minimizing:

$$\hat{\sigma}_n^2 = \arg \min_{k_x, \sigma_n^2} \sum_{i=2}^{N^2} \left| \frac{k_x - 3}{(1 + \frac{\sigma_n^2}{\sigma_{y_i}^2})^2} + 3 - k_{y_i} \right|. \quad (2)$$

where  $\sigma_{y_i}^2$  and  $k_{y_i}$  are computed from the filter responses of  $\mathbf{y}$  with each of  $N \times N$  DCT basis.

#### B. Image Denoising

If the multiply distorted image is noisy, human brains tend to denoise it at once. With the estimated noise level in Section III-A, we choose the high-performance BM3D method [54] for image denoising. The main steps are outlined as follows.

1) Obtain the basic estimation.

- Block-wise estimations. For each block in the noisy image, use block-matching to find the locations of the blocks, which are similar to the currently processed one.

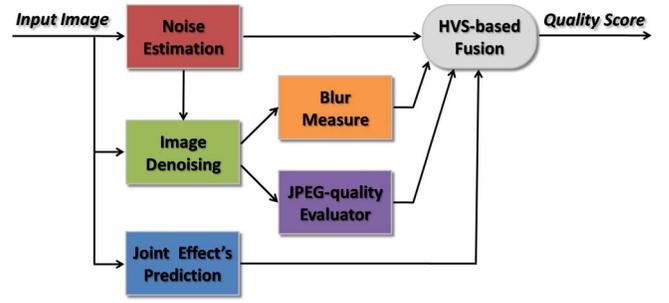


Fig. 4. Major steps of the proposed SISBLIM algorithm.

Then, use 3D transform to attenuate the noise by hard-thresholding its transform spectrum. Finally, invert 3D transform to produce estimations of all grouped blocks.

- Aggregation. Compute the basic estimation by a weighted average of obtained block-wise estimations.

2) Obtain the final estimation by using the basic estimation to further improve the grouping and to preform collaborative Wiener filtering.

- Block-wise estimations. For each block in the basic estimation, use block-matching to search for the locations of the blocks, which are similar to the currently processed one, thereby to form a pair of 3D arrays (groups). One is from the noisy image, and the other is from the basic estimation. Next, apply 3D transform on this couple of 3D arrays and perform 3D Wiener filtering using the energy spectra of the basic estimation. Eventually, invert 3D transform to produce estimations of all grouped blocks.

- Aggregation. Compute the final estimation by a weighted average of obtained block-wise estimations.

#### C. Blur Measure

The blurriness of noiseless/denoised image is measured by MBBM. Edge detection is first used to find vertical/horizontal edges in the input visual signal. Next, each row/column of the image is scanned. For pixels belonging to an edge location, the start and end positions of the edge are defined as the local extrema locations closest to the edge. The edge width is defined as the difference between the end and start positions, and is identified as the local blur measure for this edge location. At last, the global blur measure for the whole image is obtained by computing the mean value of the local blur values over all the edge locations.

#### D. JPEG-Quality Evaluator

For JPEG compression, blurring and blockiness occur together because of the coarse quantization and the independent processing of image blocks. The blurring effect is mainly introduced from the deletion of high frequency DCT coefficients while the blockiness occurs due to the discontinuity at block boundaries. On this base, WNJE performs in four steps: First, blockiness is estimated as the average differences across block boundaries:

$$B_h = \frac{1}{H(\lfloor W/8 \rfloor - 1)} \sum_{i=1}^H \sum_{j=1}^{\lfloor W/8 \rfloor - 1} |D_h(i, 8j)|; \quad (3)$$

where  $D_h(m, n) = x(m, n + 1) - x(m, n)$ ,  $n \in [1, W - 1]$  and  $H$  and  $W$  independently indicate the image height and width. Second, the average absolute difference between in-block image samples is calculated as follows:

$$A_h = \frac{1}{7} \left[ \frac{1}{H(W-1)} \sum_{i=1}^H \sum_{j=1}^{W-1} |D_h(i, j)| - B_h \right]; \quad (4)$$

Third, the horizontal zero crossing rate is computed by

$$Z_h = \frac{1}{H(W-2)} \sum_{i=1}^H \sum_{j=1}^{W-2} D_h(i, j); \quad (5)$$

Finally, the image quality prediction can be given by

$$Q_J = \phi_1 + \phi_2 (B_J)^{\theta_1} (A_J)^{\theta_2} (Z_J)^{\theta_3} \quad (6)$$

where

$$B_J = \frac{B_h + B_v}{2}, \quad A_J = \frac{A_h + A_v}{2}, \quad Z_J = \frac{Z_h + Z_v}{2}, \quad (7)$$

and  $B_v$ ,  $A_v$  and  $Z_v$  are vertical features using similar methods as  $B_h$ ,  $A_h$  and  $Z_h$ .  $\{\phi_1, \phi_2, \theta_1, \theta_2, \theta_3\}$  are model parameters to be determined later.

### E. Joint Effect's Prediction

The joint effect of mixed artifacts has an important influence on the human visual perception to the quality of the multiply distorted image. This joint effect is clearly affected by the interaction of different distortion types. Due to the existence of masking effects, this effect also substantially depends on the image content. Therefore, the two aforementioned factors should be considered to estimate the joint effect. Notice that images of more masks are usually difficult to describe on the one hand. On the other hand, we note that the perceived strength of the noise will increase when it is added to a more blurry image [46]. This phenomenon can be explained by the fact that the increasing amount of blur reduces the noise masking in the input image, thereby rendering the noise more pronounced and the image harder to characterize. Inspired by these accounts, we in this part resort to the image descriptive complexity to measure the joint effect.

The image descriptive complexity, however, is an abstract concept. To address this problem, we question into the neurological and psychophysical mechanism of human vision. Particularly, free energy is used here since it can well approximate the process of the HVS. The free energy theory was recently revealed to explain and unify brain theories in biological and physical sciences about human action, perception and learning [55]–[56]. This principle suggests that the brain always seeks the most ‘logical’ explanation of each given scene by tuning its internal generative model. The gap between the external input and its generative-model-explainable part should be related to the complexity of the given image.

Specifically, we assume for operational amenability that the internal generative model  $g$  for visual perception is parametric, which adjusts the vector  $\theta$  of parameters to explain perceived scenes. Given an image  $\mathbf{x}$ , its ‘surprise’ can be measured by

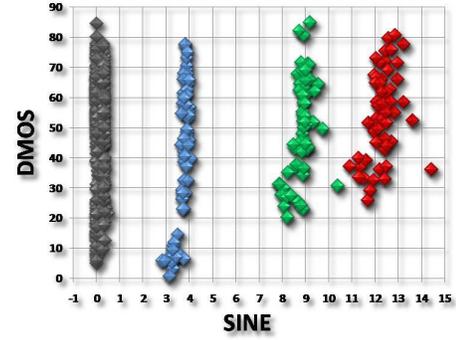


Fig. 5. Scatter plots of DMOS versus SINE [38] on LIVEMD. Red, green, blue, and black plots represent four various levels of noise.

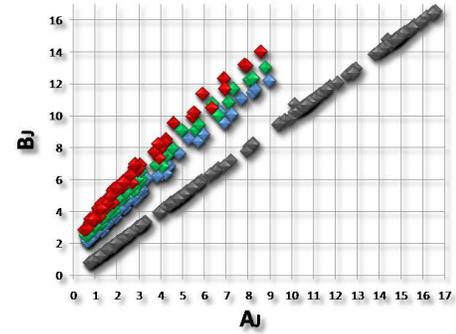


Fig. 6. Scatter plots of  $B_J$  versus  $A_J$  on LIVEMD. Red, green, blue, and black plots correspond to four different degrees of JPEG compression.

integrating the joint distribution  $p(\mathbf{x}, \theta|g)$  over the space of model parameters  $\theta$ :

$$-\log p(\mathbf{x}|g) = -\log \int p(\mathbf{x}, \theta|g) d\theta. \quad (8)$$

We introduce an auxiliary term  $q(\theta|\mathbf{x})$  into both the denominator and numerator in Eq. (8) and rewrite it as

$$-\log p(\mathbf{x}|g) = -\log \int q(\theta|\mathbf{x}) \frac{p(\mathbf{x}, \theta|g)}{q(\theta|\mathbf{x})} d\theta. \quad (9)$$

Using Jensen’s inequality, we can further get

$$-\log p(\mathbf{x}) \leq -\int q(\theta|\mathbf{x}) \log \frac{p(\mathbf{x}, \theta)}{q(\theta|\mathbf{x})} d\theta \quad (10)$$

where the right hand side is the free energy:

$$F(\theta) = -\int q(\theta|\mathbf{x}) \log \frac{p(\mathbf{x}, \theta)}{q(\theta|\mathbf{x})} d\theta. \quad (11)$$

More details of free energy can be found in [26], [55]–[56]. In this work, we define the joint effect as the free energy value on the highest  $l\%$  ‘surprise’ regions in the input image.

### F. HVS Based Fusion Model

The proposed SISBLIM is defined as a linear combination of weighted quality scores of noise, blur and JPEG

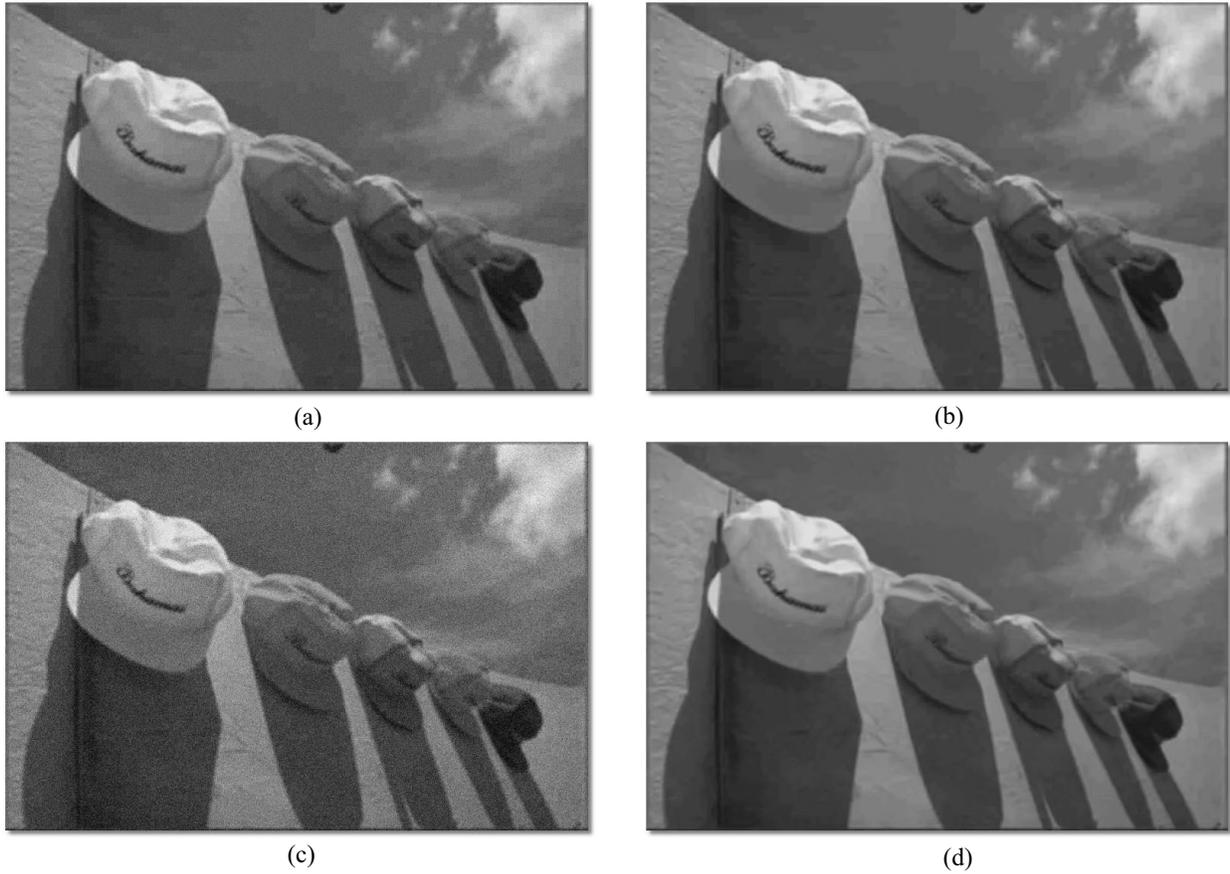


Fig. 7. Illustration of the masking effect of noise on JPEG compression artifacts in multiply-distorted images in MDID2013.

compression as well as the joint effect stated above:

$$\text{SISBLIM} = \sum_{i=\{N,B,J,F\}} (\xi_i \lambda_i) Q_i \quad (12)$$

where  $\lambda_N = \lambda_F = 1$ ,  $\lambda_J = 1 - \lambda_B$ , and  $\xi_N$ ,  $\xi_B$ ,  $\xi_J$ ,  $\xi_F$  and  $\lambda_B$  are model parameters, and  $Q_F$  indicates the joint effect measured by ‘free energy’.

The LIVEMD database [44] involves two groups of multiply distorted images, namely blur&JPEG images (225 images corrupted by blur followed by JPEG) and blur&noise images (225 images contaminated by blur followed by noise). We have mentioned that the HVS can effectively separate noise from a (multiply) distorted image. Noise estimation is almost independent of other distortion measures. As a matter of fact, the accuracy of SINE is largely immune to the influence of blur and JPEG compression. In an example shown in Fig. 5, all images in LIVEMD with four various levels of noise are represented by red, green, blue and black scatter plots, which suggests that the performance of SINE for noise estimation is rarely affected by the other two distortion types. So we first estimate the noise level ( $Q_N$ ) of a given distorted image to be an important component of SISBLIM, and then apply this estimated result to image denoising.

Although the HVS can easily separate the artifacts of blurring and blockiness, it is not an easy task for computers. In this work, we carefully compare Eq. (3) and Eq. (4), and find that the ratio of  $B_h$  to  $A_J$  can validly partition

JPEG compressed images (including images only corrupted by blockiness and images corrupted by blur and JPEG together) apart from other images. More specifically,  $B_J$  is actually computing the mean of  $|D_h|$  and  $|D_v|$  values located in the edge of all the blocks, while  $A_J$  is evaluated for the  $6 \times 6$  interior part. It is not difficult to conjecture that  $B_J$  is nearly equal to  $A_J$  for non-blockiness images, whereas  $B_J$  is larger than  $A_J$  for JPEG compressed images. Fig. 6 displays the relationship between  $B_J$  and  $A_J$  for all images in LIVEMD. Red, green, blue and black scatter plots individually indicate four different degrees of JPEG compression.

Also, it is important to notice that performance of MBBM is influenced by blockiness, since its basic idea is to measure the spread of edges in an image. We therefore only adopt  $Q_J$  with updated values of  $\{\phi_1, \phi_2, \theta_1, \theta_2, \theta_3\}$  to predict the qualities of JPEG compressed images in light of the function of  $A_J$  and  $Z_J$  in measuring blurriness, and utilize  $Q_B$  for other type of distorted images. We therefore in this paper adjust

TABLE II  
ILLUSTRATION OF DIFFERENT SISBLIM-TYPE OF METHODS WITH THE COMPONENTS USING DISTINCT DISTORTION-SPECIFIC QUALITY MEASURES

Method	MBBM [35]	FISH <sub>bb</sub> [37]
SINE [38]	SISBLIM <sub>sm</sub>	SISBLIM <sub>sb</sub>
WTPNE [39]	SISBLIM <sub>wm</sub>	SISBLIM <sub>wfb</sub>

the control parameter  $\lambda_B$  to selectively using  $Q_J$  or  $Q_B$  in the computation of SISBLIM:

$$\lambda_B = \begin{cases} 0 & \text{if } B_J/A_J \geq \Omega_{\lambda_B} \\ 1 & \text{otherwise} \end{cases} \quad (13)$$

where  $\Omega_{\lambda_B}$  is a threshold to distinguish the JPEG compressed images and other images (namely the black scatter plots and other scatter plots in Fig. 6). The value of  $\Omega_{\lambda_B}$  should be a little larger than one and in this paper we set  $\Omega_{\lambda_B} = 1.5$ .

For multiple distort types, we still have to take into account the impact of noise injection on the blockiness estimation. Fig. 7 illustrates a pair of “hats” images in MDID2013 for comparison: (a) is the image corrupted with the lowest level of blur and noise, and the highest level of JPEG compression; (b) is the denoised image from (a) using BM3D with true noise variance; (c) is the distorted image corrupted with the lowest level of blur, and the highest level of noise and JPEG compression; (d) is the denoised image from (c) with BM3D. We can easily find that (a) and (c) are contaminated by the same highest level of JPEG compression, yet the blockiness in (c) is much harder to notice than that in (a), due to the masking effect caused by the high level of noise. Equally, after image denoising, (b) exhibits obvious blockiness in the sky region, whereas (d) shows little. It is very clear that this phenomenon decides the  $\lambda_B$  value. We accordingly characterize this masking effect of noise on JPEG compression artifacts by adding a small constant (we choose 0.5) to the threshold  $\Omega_{\lambda_B}$  when obvious noise exists (e.g., the estimated noise variance is larger than 1.5).

Finally, we combine single quality prediction of each distortion type and the joint effect of mixed artifacts to derive the quality prediction of the proposed SISBLIM. To be fair, all model parameters in SISBLIM were acquired on the LIVEMD database [44], since this database is publicly available.

#### G. Fusion Model With Other Component Measures

It is very unlikely that those distortion-specific measures we have used in the SISBLIM model is the only choice. Other metrics were also tested, e.g., FISH<sub>bb</sub> [37] and WTPNE [39], as listed in Table II. SISBLIM<sub>sm</sub> is the original method. All model parameters in other three SISBLIM-type of metrics were also trained following the same procedure as SISBLIM.

### IV. EXPERIMENT RESULTS AND DISCUSSION

In this section, we testify the performance of SISBLIM and compare it with eight IQA metrics. These competitors are:

- PSNR and SSIM [6], the benchmark IQA methods with a wide application in the image processing literature.
- MS-SSIM [9], performs SSIM in each level and integrates the above outputs with psychophysical weights.
- DIIVINE [30], extracts 88 features using the NSS model to characterize the essence of natural images: 1) distortion identification; 2) distortion-specific quality assessment.
- BLINDS-II [31], relies on a simple Bayesian inference model to predict image quality based on certain features extracted using the NSS model in the DCT domain.

- BRISQUE [32] and BRISQUE-II [44], use scene statistics of locally normalized luminance coefficients to quantify possible losses of “naturalness” in images to derive a quality measure, with only 36 features. Following the method mentioned in [44], BRISQUE-II is trained on the LIVEMD database using the support vector machine (SVM) [57] with parameters supplied by the authors.

- NIQE [34], predicts the image quality nearly without any prior knowledge of contents or distortions. NIQE measures the deviations from statistical regularities in natural images.

All the six image quality databases (LIVE, TID2008, CSIQ, IVC, Toyama, LIVEMD) and our MDID2013 are used in this study. Details of MDID2013 have been illustrated in Section II. Basic information of other six databases is as follows:

- The LIVE database [7] contains five image data sets, and a total of 779 distorted images from 29 pristine images. There are five commonly encountered distortion types, and in this work, we select 465 images belonging to JPEG, white noise and Gaussian blur subsets.

- The TID2008 database [48] is the largest database including 1700 distorted images generated from 25 references with 17 distortion types at 4 distortion levels. Here, we pick 300 images corrupted by three distortion types: a) Additive noise; b) Gaussian blur; c) JPEG compression.

- The CSIQ database [49] totally consists of 866 images, which are created from 30 original images by using six types of distortions at four to five distortion levels. In our study, 450 images corrupted by additive noise, JPEG compression and Gaussian blur are used for testing.

- The IVC database [50] is consisted by 185 images generated from 10 sources. Two distortion types and their associated 70 images are applied: 1) JPEG; 2) Blurring.

- The Toyama database [51] includes 168 distorted images. We choose 84 JPEG compressed images in this work.

- The LIVEMD database [44] is the first image database for multiple distortions. It has two image subsets which are created by adding different levels of noise/JPEG to blurred images, respectively. There are 225 images generated from 15 pristine images in each subset.

Using those IQA databases, we first compute the objective prediction scores of each quality metric, and use the nonlinear regression to map the scores to subjective ratings based on the four-parameter logistic function:

$$Q(\varepsilon) = \frac{\xi_1 - \xi_2}{1 + \exp(-\frac{\varepsilon - \xi_3}{\xi_4})} + \xi_2 \quad (14)$$

where  $\varepsilon$  and  $Q(\varepsilon)$  are the input score and the mapped score, and  $\xi_j$  ( $j = 1, 2, 3, 4$ ) are free parameters to be determined during the curve fitting process. We then employ five commonly used performance measures, as suggested by VQEG [58], to evaluate and compare the proposed method with competing IQA metrics. The five performance evaluations are:

- Pearson linear correlation coefficient (PLCC) is computed between human ratings and objective scores after nonlinear regression of Eq. (14). It is defined as

$$\text{PLCC} = \frac{\sum_i (q_i - \bar{q})(o_i - \bar{o})}{\sqrt{\sum_i (q_i - \bar{q})^2 (o_i - \bar{o})^2}} \quad (15)$$

TABLE III  
PERFORMANCE EVALUATIONS OF OUR SISBLIM AND COMPETING IQA METRICS. WE BOLD THE BEST TWO PERFORMED METHODS

		LIVE database (465 images) [7]					TID2008 database (300 images) [48]				
IQA metrics	Type	PLCC	SROCC	KROCC	AAE	RMS	PLCC	SROCC	KROCC	AAE	RMS
SISBLIM	NR	<b>0.9505</b>	0.9450	0.7981	<b>6.6585</b>	<b>8.5136</b>	0.8419	0.8202	0.6203	0.5670	0.7098
PSNR	FR	0.8584	0.8712	0.6859	10.958	14.056	0.8338	<b>0.8641</b>	<b>0.6792</b>	0.5380	0.7261
SSIM [6]	FR	0.8762	0.8948	0.7109	10.716	13.205	0.7164	0.7386	0.5322	0.7443	0.9177
MS-SSIM [9]	FR	0.9297	<b>0.9481</b>	<b>0.8007</b>	8.0852	10.094	<b>0.8640</b>	<b>0.8641</b>	<b>0.6672</b>	<b>0.5289</b>	<b>0.6623</b>
DIIVINE [30]	NR	0.8217	0.8304	0.6856	11.200	15.614	0.6343	0.6756	0.5067	0.7405	1.0169
BLIINDS-II [31]	NR	0.9143	0.9067	0.7369	8.6329	11.096	0.7814	0.7513	0.5601	0.6296	0.8208
BRISQUE [32]	NR	<b>0.9606</b>	<b>0.9572</b>	<b>0.8230</b>	<b>5.8660</b>	<b>7.4961</b>	<b>0.8522</b>	0.8325	0.6634	<b>0.4966</b>	<b>0.6881</b>
BRISQUE-II [44]	NR	0.7603	0.7493	0.5444	13.614	17.799	0.5435	0.4562	0.3149	0.9146	1.1040
NIQE [34]	NR	0.9162	0.9236	0.7546	8.7883	10.982	0.7301	0.7360	0.5298	0.7381	0.8987

		CSIQ database (450 images) [49]					IVC database (70 images) [50]				
IQA metrics	Type	PLCC	SROCC	KROCC	AAE	RMS	PLCC	SROCC	KROCC	AAE	RMS
SISBLIM	NR	<b>0.9347</b>	0.8957	0.7263	<b>0.0727</b>	<b>0.0954</b>	<b>0.8677</b>	<b>0.8590</b>	<b>0.6732</b>	<b>0.4548</b>	<b>0.5822</b>
PSNR	FR	0.8923	<b>0.9173</b>	0.7462	0.0844	0.1211	0.7515	0.6953	0.5153	0.6106	0.7726
SSIM [6]	FR	0.8325	0.8666	0.6673	0.1087	0.1486	0.8543	0.8163	0.6350	0.4442	0.6087
MS-SSIM [9]	FR	0.9227	<b>0.9327</b>	<b>0.7690</b>	0.0785	0.1034	<b>0.9322</b>	<b>0.9077</b>	<b>0.7521</b>	<b>0.2949</b>	<b>0.4239</b>
DIIVINE [30]	NR	0.8392	0.8320	0.6425	0.1038	0.1459	0.2949	0.2986	0.2463	1.0029	1.1189
BLIINDS-II [31]	NR	0.8735	0.8498	0.6536	0.1012	0.1306	0.4891	0.4485	0.3401	0.8796	1.0214
BRISQUE [32]	NR	<b>0.9361</b>	0.9138	<b>0.7519</b>	<b>0.0690</b>	<b>0.0943</b>	0.8308	0.8119	0.6225	0.5071	0.6517
BRISQUE-II [44]	NR	0.8129	0.7737	0.5890	0.1204	0.1562	0.6162	0.5126	0.3599	0.7811	0.9222
NIQE [34]	NR	0.8704	0.8568	0.6637	0.1029	0.1321	0.8554	0.8505	0.6591	0.4663	0.6066

		Toyama database (84 images) [51]					LIVEMD database (450 images) [44]				
IQA metrics	Type	PLCC	SROCC	KROCC	AAE	RMS	PLCC	SROCC	KROCC	AAE	RMS
SISBLIM	NR	0.8249	0.7878	0.5792	0.5582	0.6991	<b>0.8949</b>	<b>0.8781</b>	<b>0.6925</b>	<b>6.6153</b>	<b>8.4386</b>
PSNR	FR	0.3778	0.2868	0.2002	0.9836	1.1449	0.7398	0.6771	0.5003	10.281	12.724
SSIM [6]	FR	0.6542	0.6263	0.4423	0.7811	0.9352	0.7333	0.6459	0.4633	10.513	12.859
MS-SSIM [9]	FR	0.8414	0.8360	0.6434	0.5458	0.6681	0.8749	0.8392	0.6474	7.4595	9.1596
DIIVINE [30]	NR	0.7087	0.7023	0.5324	0.6606	0.8724	0.7183	0.6563	0.4778	10.235	13.157
BLIINDS-II [31]	NR	<b>0.8848</b>	<b>0.8678</b>	<b>0.6960</b>	<b>0.4473</b>	<b>0.5761</b>	0.3574	0.2464	0.1859	14.356	17.663
BRISQUE [32]	NR	<b>0.8735</b>	<b>0.8690</b>	<b>0.6856</b>	<b>0.4642</b>	<b>0.6021</b>	0.5485	0.5017	0.3644	12.809	15.813
BRISQUE-II [44]	NR	0.5381	0.4847	0.3579	0.8423	1.0422	<b>0.9349</b>	<b>0.9111</b>	<b>0.7554</b>	<b>4.6607</b>	<b>6.1865</b>
NIQE [34]	NR	0.8455	0.8378	0.6503	0.4835	0.6603	0.8389	0.7750	0.5820	8.1305	10.294

		MDID2013 database (324 images)					Average (2143 images)				
IQA metrics	Type	PLCC	SROCC	KROCC	AAE	RMS	PLCC	SROCC	KROCC	AAE	RMS
SISBLIM	NR	<b>0.8140</b>	<b>0.8079</b>	<b>0.6146</b>	<b>0.0229</b>	<b>0.0295</b>	<b>0.8920</b>	<b>0.8734</b>	<b>0.6955</b>	<b>2.9688</b>	<b>3.7896</b>
PSNR	FR	0.5607	0.5604	0.3935	0.0345	0.0421	0.7698	0.7635	0.5899	4.6934	5.9254
SSIM [6]	FR	0.4570	0.4494	0.3143	0.0377	0.0452	0.7419	0.7343	0.5518	4.7106	5.7884
MS-SSIM [9]	FR	<b>0.7435</b>	<b>0.7401</b>	<b>0.5418</b>	<b>0.0276</b>	<b>0.0340</b>	<b>0.8760</b>	<b>0.8731</b>	<b>0.6963</b>	<b>3.4465</b>	<b>4.2732</b>
DIIVINE [30]	NR	0.4471	0.4463	0.3644	0.0372	0.0455	0.6978	0.6899	0.5363	4.7695	6.4020
BLIINDS-II [31]	NR	0.2244	0.1796	0.1200	0.0416	0.0495	0.6508	0.6079	0.4711	5.0498	6.3223
BRISQUE [32]	NR	0.4133	0.2210	0.1617	0.0382	0.0463	0.7634	0.7155	0.5775	4.0870	5.1152
BRISQUE-II [44]	NR	0.5741	0.5545	0.3814	0.0345	0.0416	0.7361	0.6998	0.5279	4.1498	5.4258
NIQE [34]	NR	0.5635	0.5450	0.3787	0.0349	0.0420	0.8062	0.7891	0.6038	3.7786	4.7502

where  $o_i$  and  $\bar{o}$  are the  $i$ -th image's subjective rating and the mean of all  $o_i$ .  $q_i$  and  $\bar{q}$  are the converted objective scores after nonlinear regression and the mean of all  $q_i$ .

• Spearman rank-order correlation coefficient (SROCC) is calculated by

$$\text{SROCC} = 1 - \frac{6}{M(M^2 - 1)} \sum_{i=1}^M d_i^2 \quad (16)$$

where  $d_i$  is the difference between the  $i$ -th image's ranks in subjective and objective evaluations, and  $M$  is the image numbers in the testing database. It is a non-parametric rank-based correlation measure, independent of any monotonic nonlinear mapping between subjective and objective scores.

• Kendall's rank-order correlation coefficient (KROCC) is another important non-parametric rank correlation metric given by

$$\text{KROCC} = \frac{M_c - M_d}{\frac{1}{2}M(M - 1)} \quad (17)$$

where  $M_c$  and  $M_d$  indicate the numbers of concordant and discordant pairs in the data set, respectively.

• Average absolute prediction error (AAE) is measured using the converted objective predictions after the nonlinear mapping of Eq. (14):

$$\text{AAE} = \frac{1}{M} \sum |q_i - o_i|. \quad (18)$$

TABLE IV  
PERFORMANCE MEASURES OF SISBLIM AND STATE-OF-THE-ART FR FSIM, GSIM, IGM, AND GMSD ON LIVEMD AND MDID2013

		LIVEMD database (450 images) [44]					MDID2013 database (324 images)				
IQA metrics	Type	PLCC	SROCC	KROCC	AAE	RMS	PLCC	SROCC	KROCC	AAE	RMS
SISBLIM	NR	0.8949	0.8781	0.6925	6.6153	8.4386	0.8140	0.8079	0.6146	0.0229	0.0295
FSIM [17]	FR	0.8932	0.8637	0.6729	6.8751	8.5048	0.6431	0.6500	0.5314	0.0319	0.0389
GSIM [18]	FR	0.8806	0.8454	0.6550	7.1803	8.9613	0.6646	0.6637	0.4600	0.0308	0.0380
IGM [20]	FR	0.8841	0.8500	0.6573	7.1725	8.8375	0.8207	0.8232	0.6237	0.0231	0.0290
GMSD [24]	FR	0.8803	0.8448	0.6548	7.1199	8.9733	0.8294	0.8283	0.6240	0.0228	0.0284

TABLE V  
PERFORMANCE EVALUATIONS OF DIFFERENT SISBLIM-TYPE OF METHODS. WE HIGHLIGHT THE BEST PERFORMED METRIC WITH BOLDFACE

		LIVE database (465 images) [7]					TID2008 database (300 images) [48]				
IQA metrics	Type	PLCC	SROCC	KROCC	AAE	RMS	PLCC	SROCC	KROCC	AAE	RMS
SISBLIM <sub>sm</sub>	NR	<b>0.9505</b>	<b>0.9450</b>	<b>0.7981</b>	<b>6.6585</b>	<b>8.5136</b>	0.8419	0.8202	0.6203	0.5670	0.7098
SISBLIM <sub>sfb</sub>	NR	0.9262	0.9149	0.7494	8.3028	10.329	0.8466	0.7811	0.5847	0.5801	0.7000
SISBLIM <sub>wm</sub>	NR	0.9454	0.9422	0.7894	7.1571	8.9259	<b>0.8509</b>	<b>0.8285</b>	<b>0.6337</b>	<b>0.5467</b>	<b>0.6909</b>
SISBLIM <sub>wfb</sub>	NR	0.9251	0.9136	0.7486	8.3555	10.407	0.8464	0.7848	0.5884	0.5803	0.7005

		CSIQ database (450 images) [49]					IVC database (70 images) [50]				
IQA metrics	Type	PLCC	SROCC	KROCC	AAE	RMS	PLCC	SROCC	KROCC	AAE	RMS
SISBLIM <sub>sm</sub>	NR	<b>0.9347</b>	<b>0.8957</b>	<b>0.7263</b>	<b>0.0727</b>	<b>0.0954</b>	<b>0.8677</b>	<b>0.8590</b>	<b>0.6732</b>	<b>0.4548</b>	<b>0.5822</b>
SISBLIM <sub>sfb</sub>	NR	0.8855	0.8475	0.6574	0.0983	0.1246	0.8755	0.8676	0.6773	0.4443	0.5658
SISBLIM <sub>wm</sub>	NR	0.9278	0.8874	0.7142	0.0780	0.1001	0.8521	0.8542	0.6632	0.4700	0.6129
SISBLIM <sub>wfb</sub>	NR	0.8814	0.8452	0.6539	0.1001	0.1267	0.8836	0.8786	0.6915	0.4239	0.5484

		Toyama database (84 images) [51]					LIVEMD database (450 images) [44]				
IQA metrics	Type	PLCC	SROCC	KROCC	AAE	RMS	PLCC	SROCC	KROCC	AAE	RMS
SISBLIM <sub>sm</sub>	NR	0.8249	0.7878	0.5792	0.5582	0.6991	<b>0.8949</b>	<b>0.8781</b>	<b>0.6925</b>	<b>6.6153</b>	<b>8.4386</b>
SISBLIM <sub>sfb</sub>	NR	<b>0.8882</b>	<b>0.8591</b>	<b>0.6659</b>	<b>0.4602</b>	<b>0.5682</b>	0.8638	0.8572	0.6606	7.8311	9.5282
SISBLIM <sub>wm</sub>	NR	0.8371	0.8005	0.5862	0.5494	0.6765	0.8936	0.8766	0.6896	6.6208	8.4904
SISBLIM <sub>wfb</sub>	NR	0.8871	0.8577	0.6642	0.4615	0.5707	0.8663	0.8613	0.6645	7.7138	9.4469

		MDID2013 database (324 images)					Average (2143 images)				
IQA metrics	Type	PLCC	SROCC	KROCC	AAE	RMS	PLCC	SROCC	KROCC	AAE	RMS
SISBLIM <sub>sm</sub>	NR	<b>0.8140</b>	<b>0.8079</b>	<b>0.6146</b>	<b>0.0229</b>	<b>0.0295</b>	<b>0.8920</b>	<b>0.8734</b>	<b>0.6955</b>	<b>2.9688</b>	<b>3.7896</b>
SISBLIM <sub>sfb</sub>	NR	0.7114	0.6899	0.4931	0.0289	0.0357	0.8888	0.8706	0.6903	3.0767	4.8885
SISBLIM <sub>wm</sub>	NR	0.7062	0.6944	0.4957	0.0294	0.0360	0.8566	0.8337	0.6452	3.5715	4.4122
SISBLIM <sub>wfb</sub>	NR	0.7709	0.7637	0.5615	0.0282	0.0348	0.8662	0.8440	0.6548	3.5739	4.4153

• Root mean-squared (RMS) error is defined as the modified version of AAE:

$$RMS = \sqrt{\frac{1}{M} \sum (q_i - o_i)^2}. \quad (19)$$

Among the five evaluations above, a value close to 1 for PLCC, SROCC, KROCC, yet close to 0 for AAE, RMS indicates superior correlation with subjective quality scores.

Table III illustrates the performance evaluations of PLCC, SROCC, KROCC, AAE and RMS (after nonlinear regression) and their average results of nine IQA approaches on seven databases. The database size-weighted average is defined as  $\bar{\delta} = \frac{\sum_i \delta_i \omega_i}{\sum_i \omega_i}$  where  $\delta_i$  ( $i = 1, \dots, 7$ ) indicates the correlation measure for each database, and  $\omega_i$  is the number of images in each database, i.e. 465 for LIVE, 300 for TID2008, 450 for CSIQ, 70 for IVC, 84 for Toyama, 450 for LIVEMD, and 324 for MDID2013. Table V reports the prediction accuracy of different SISBLIM-type of methods with distinct components (defined in Section III-G) to validate the robustness of the proposed framework. Furthermore, Table IV testifies our blind SISBLIM algorithm with state-of-the-art FR FSIM [17],

GSIM [18], IGM [20] and GMSD [24] on multiply distorted image databases (LIVEMD and MDID2013). From Tables III–V, we have the following observations:

- First, it is viewed that our algorithm has achieved very encouraging results as compared to state-of-the-art NR IQA metrics and classical FR IQA methods on single and multiple distortion types. Specifically, SISBLIM has gained better results (on average) than the powerful FR MS-SSIM and the best-performed NR BRISQUE (to date) which is trained using the entire LIVE database, and clearly outperforms the benchmark FR PSNR and SSIM and other recently proposed NR IQA metrics. Besides, note that the proposed algorithm is a little inferior to BRISQUE-II which is trained on the overall LIVEMD database and highly depends on that database, yet remarkably superior to BRISQUE-II on other singly and multiply distorted image databases. In summary, the proposed SISBLIM performs better than all testing IQA methods on average.
- Second, we can find that the proposed algorithm is extremely effective in the IQA of multiple distortions. Our SISBLIM, despite of blind metric, is superior to recently

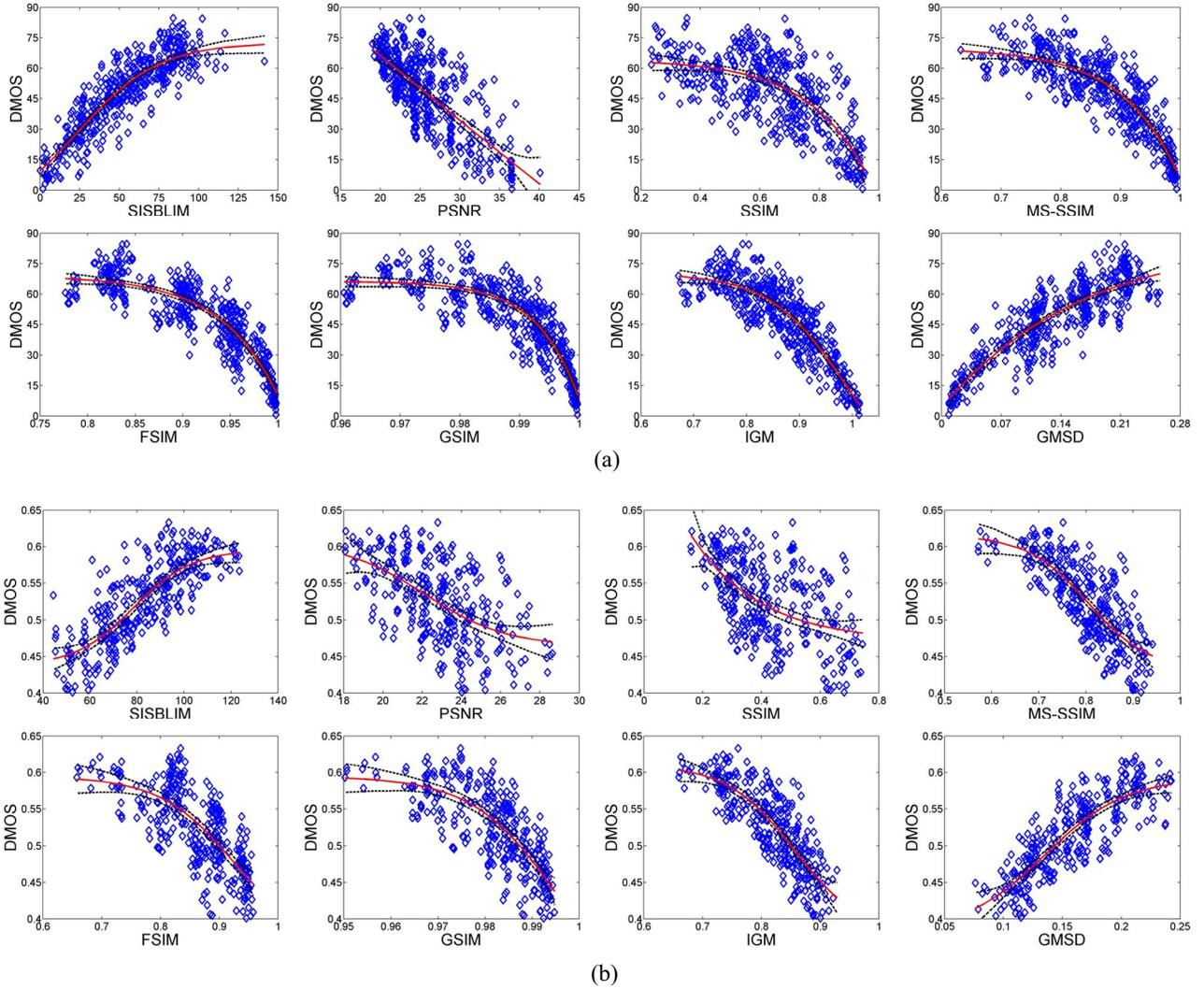


Fig. 8. Scatter plots of DMOS versus classical PSNR, SSIM, MS-SSIM, state-of-the-art FSIM, GSIM, IGM, GMSD, and our SISBLIM on LIVEMD and MDID2013. The (red) lines are curves fitted with the logistic function and the (black) dash lines are 95% confidence intervals. (a) LIVEMD database. (b) MDID2013 database.

TABLE VI

PERFORMANCE COMPARISON BETWEEN OUR SISBLIM AND OTHER METHODS WITH F-TEST (STATISTICAL SIGNIFICANCE). THE SYMBOL “1”, “0,” OR “-1” MEANS THAT SISBLIM IS STATISTICALLY (WITH 95% CONFIDENCE) BETTER, UNDISTINGUISHABLE, OR WORSE THAN THE CORRESPONDING METHODS

Database	PSNR	SSIM	MS-SSIM	DIIVINE	BLIINDS-II	BRISQUE	BRISQUE-II	NIQE
LIVE	1	1	1	1	1	0	1	1
TID2008	1	1	0	1	1	0	1	1
CSIQ	1	1	0	1	1	0	1	1
IVC	1	1	-1	1	1	1	1	1
Toyama	1	1	0	1	-1	-1	1	-1
LIVEMD	1	1	1	1	1	1	-1	1
MDID2013	1	1	1	1	1	1	1	1

developed FR IQA methods (including FSIM, GSIM, IGM, and GMSD) in assessing multiply distorted images, as listed in Table IV.

- Third, the proposed SISBLIM is a universal model since the components within our framework can be replaced with other more effective dedicated blind metrics to improve the prediction accuracy of SISBLIM. As tabulated in Table V,

SISBLIMs with the components using other distortion-specific blind measures present higher performances than the original on some databases.

- Fourth, our blind algorithm is not training based, but is motivated to model the perceptual process of the HVS. This makes SISBLIM inherently suitable for handling images with complicated distortions. Tables III and IV just report

the high correlation performance of our SISBLIM on different image quality databases.

Additionally, in this paper we also use the F-test to evaluate the statistical significance of the proposed model by computing the prediction residuals between converted objective scores and subjective ratings. We first suppose  $F$  denote the ratio of two residual variances, and  $F_c$  (determined by the number of residuals and the confidence level) be the judgement threshold. We then rule that the difference of prediction accuracy between those two metrics is significant when  $F > F_c$ . The statistical significances between our algorithm and other IQA metrics in comparison are listed in Table VI, where the symbol “1”, “0” or “-1” means that the proposed metric is statistically (with 95% confidence) better, indistinguishable, or worse than the corresponding metric, respectively. It is easy to find that our SISBLIM, despite of its training-free nature, is statistically better than the benchmark FR PSNR and SSIM and most state-of-the-art NR IQA metrics, as well as is highly on par with the powerful FR MS-SSIM.

Finally, Fig. 8 gives scatter plots of classical and recently proposed FR IQA methods and the proposed SISBLIM on multiply distorted image databases. Those approaches include PSNR, SSIM, MS-SSIM, FSIM, GSIM, IGM, and GMSD. It is apparent that the convergence and monotonicity of our blind model outperforms state-of-the-art FR IQA algorithms.

## V. CONCLUSION

This paper is devoted to the problem of quality assessment for multiply distorted images. In practice, images are usually corrupted with various kinds of distortion types simultaneously. As an important and more challenging complement to the recently released LIVEMD database, this paper first introduced a multiply distorted image database (MDID2013) with 3-fold distortions. We then modeled the process of human visual perception and designed a novel no-reference Six-Step BLind Metric (SISBLIM). The proposed algorithm is training-free and works well for single, double and triple distortion types. Experimental results on popular singly and multiply distorted image databases (LIVE, TID2008, CSIQ, IVC and Toyama, LIVEMD and MDID2013) demonstrate the superiority of SISBLIM over classical FR PSNR, SSIM, MS-SSIM, and state-of-the-art NR IQA metrics. The performance of our algorithm is also better than recently developed FR IQA methods for multiple distortions. Furthermore, it is worth mentioning that SISBLIM is a general model and other distortion measures can be integrated for improved performance.

## REFERENCES

- [1] G. Zhai, J. Cai, W. Lin, X. Yang, and W. Zhang, “Three dimensional scalable video adaptation via user-end perceptual quality assessment,” *IEEE Trans. Broadcast.*, vol. 54, no. 3, pp. 719–727, Sep. 2008.
- [2] G. Zhai *et al.*, “Cross-dimensional perceptual quality assessment for low bitrate videos,” *IEEE Trans. Multimedia*, vol. 10, no. 7, pp. 1316–1324, Nov. 2008.
- [3] 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.
- [4] S. Wang, A. Rehman, Z. Wang, S. Ma, and W. Gao, “SSIM-inspired divisive normalization for perceptual video coding,” in *Proc. IEEE Int. Conf. Image Process.*, Brussels, France, pp. 1657–1660, Sep. 2011.
- [5] D. Shaked and I. Tastl, “Sharpness measure: Towards automatic image enhancement,” in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2005.
- [6] 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.
- [7] H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik, *LIVE Image Quality Assessment Database Release 2* [Online]. Available: <http://live.ece.utexas.edu/research/quality>
- [8] ITU, *Methodology for the Subjective Assessment of the Quality of Television Pictures*, Recommendation, International Telecommunication Union/ITU Radiocommunication Sector, 2009.
- [9] 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.
- [10] H. R. Sheikh and A. C. Bovik, “Image information and visual quality,” in *Proc. IEEE Conf. Acous. Speech Sign Proc.*, May 2004, pp. 709–712.
- [11] S. Winkler and P. Mohandas, “The evolution of video quality measurement from PSNR to hybrid metrics,” *IEEE Trans. Broadcast.*, vol. 54, no. 3, pp. 660–668, Sep. 2008.
- [12] B. Ciubotaru, G.-M. Muntean, and G. Ghinea, “Objective assessment of region of interest-aware adaptive multimedia streaming quality,” *IEEE Trans. Broadcast.*, vol. 55, no. 2, pp. 202–212, Jun. 2009.
- [13] E. C. Larson and D. M. Chandler, “Most apparent distortion: Full-reference image quality assessment and the role of strategy,” *J. Electron. Imaging*, vol. 19, no. 1, Jan. 2010, Art. ID 011006.
- [14] Z. Wang and Q. Li, “Information content weighting for perceptual image quality assessment,” *IEEE Trans. Image Process.*, vol. 20, no. 5, pp. 1185–1198, May 2011.
- [15] X. Feng, T. Liu, D. Yang, and Y. Wang, “Saliency inspired full-reference quality metrics for packet-loss-impaired video,” *IEEE Trans. Broadcast.*, vol. 57, no. 1, pp. 81–88, Mar. 2011.
- [16] S. Chikkerur, V. Sundaram, M. Reisslein, and L. J. Karam, “Objective video quality assessment methods: A classification, review, and performance comparison,” *IEEE Trans. Broadcast.*, vol. 57, no. 2, pp. 165–182, Jun. 2011.
- [17] L. Zhang, L. Zhang, X. Mou, and D. Zhang, “FSIM: A feature similarity index for image quality assessment,” *IEEE Trans. Image Process.*, vol. 20, no. 8, pp. 2378–2386, Aug. 2011.
- [18] 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.
- [19] K. Gu, G. Zhai, X. Yang, L. Chen, and W. Zhang, “Nonlinear additive model based saliency map weighting strategy for image quality assessment,” in *Proc. IEEE Int. Workshop Multimedia Signal Process.*, Banff, AB, Canada, Sep. 2012, pp. 313–318.
- [20] 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.
- [21] K. Gu, G. Zhai, X. Yang, and W. Zhang, “Self-adaptive scale transform for IQA metric,” in *Proc. IEEE Int. Symp. Circuits Syst.*, Beijing, China, May 2013, pp. 2365–2368.
- [22] K. Gu, G. Zhai, X. Yang, and W. Zhang, “A new psychovisual paradigm for image quality assessment: From differentiating distortion types to discriminating quality conditions,” *Signal Image Video Process.*, vol. 7, no. 3, pp. 423–436, May 2013.
- [23] K. Gu, G. Zhai, X. Yang, W. Zhang, and M. Liu, “Structural similarity weighting for image quality assessment,” in *Proc. IEEE Int. Conf. Multimedia Expo Workshops*, San Jose, CA, USA, Jul. 2013, pp. 1–6.
- [24] W. Xue, L. Zhang, X. Mou, and A. C. Bovik, “Gradient magnitude similarity deviation: A highly efficient perceptual image quality index,” *IEEE Trans. Image Process.*, vol. 23, no. 2, pp. 684–695, Feb. 2014.
- [25] I. P. Gunawan and M. Ghanbari, “Efficient reduced-reference video quality meter,” *IEEE Trans. Broadcast.*, vol. 54, no. 3, pp. 669–679, Sep. 2008.
- [26] 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.
- [27] A. Rehman and Z. Wang, “Reduced-reference image quality assessment by structural similarity estimation,” *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3378–3389, Aug. 2012.
- [28] 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.*, Beijing, China, May 2013, pp. 1095–1098.

- [29] K. Gu, G. Zhai, X. Yang, W. Zhang, and M. Liu, "Subjective and objective quality assessment for images with contrast change," in *Proc. IEEE Int. Conf. Image Process.*, Melbourne, VIC, Australia, Sep. 2013, pp. 383–387.
- [30] 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.
- [31] 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.
- [32] 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.
- [33] K. Gu, G. Zhai, X. Yang, W. Zhang, and L. Liang, "No-reference image quality assessment metric by combining free energy theory and structural degradation model," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, San Jose, CA, USA, Jul. 2013, pp. 1–6.
- [34] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a 'completely blind' image quality analyzer," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 209–212, Mar. 2013.
- [35] P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi, "A no-reference perceptual blur metric," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2002, pp. 57–60.
- [36] Z. Wang, H. R. Sheikh, and A. C. Bovik, "No-reference perceptual quality assessment of JPEG compressed images," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2002, pp. 477–480.
- [37] P. Vu and D. M. Chandler, "A fast wavelet-based algorithm for global and local image sharpness estimation," *IEEE Signal Process. Lett.*, vol. 19, no. 7, pp. 423–426, Jul. 2012.
- [38] D. Zoran and Y. Weiss, "Scale invariance and noise in natural images," in *Proc. IEEE Int. Conf. Comput. Vis.*, Kyoto, Japan, Sep. 2009, pp. 2209–2216.
- [39] X. Liu, M. Tanaka, and M. Okutomi, "Noise level estimation using weak textured patches of a single noisy image," in *Proc. IEEE Int. Conf. Image Process.*, Orlando, FL, USA, Sep. 2012, pp. 665–668.
- [40] H. R. Sheikh, A. C. Bovik, and L. K. Cormack, "No-reference quality assessment using natural scene statistics: JPEG2000," *IEEE Trans. Image Process.*, vol. 14, no. 11, pp. 1918–1927, Nov. 2005.
- [41] H. Liu, N. Klomp, and I. Heynderickx, "A no-reference metric for perceived ringing artifacts in images," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 4, pp. 529–539, Apr. 2010.
- [42] G. Zhai and X. Wu, "Noise estimation using statistics of natural images," in *Proc. IEEE Int. Conf. Image Process.*, Brussels, Belgium, Sep. 2011, pp. 1857–1860.
- [43] 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.*, Melbourne, VIC, Australia, Jun. 2014, pp. 518–521.
- [44] D. Jayaraman, A. Mittal, A. K. Moorthy, and A. C. Bovik, "Objective quality assessment of multiply distorted images," in *Proc. IEEE Asilomar Conf. Signals Syst. Comput.*, Pacific Grove, CA, USA, Nov. 2012, pp. 1693–1697.
- [45] D. J. Granrath, "The role of human visual models in image processing," *Proc. IEEE*, vol. 69, no. 5, pp. 552–561, May 1981.
- [46] D. M. Chandler, "Seven challenges in image quality assessment: Past, present, and future research," *ISRN Signal Process.*, vol. 2013, 2013, Art. ID 905685.
- [47] K. Gu *et al.*, "FISBLIM: A five-step blind metric for quality assessment of multiply distorted images," in *Proc. IEEE Workshop Signal Process. Syst.*, Taipei City, China, Oct. 2013, pp. 241–246.
- [48] N. Ponomarenko *et al.*, "TID2008—A database for evaluation of full-reference visual quality assessment metrics," *Adv. Mod. Radioelectron.*, vol. 10, pp. 30–45, 2009.
- [49] E. C. Larson and D. M. Chandler, *Categorical Image Quality (CSIQ) Database* [Online]. Available: <http://vision.okstate.edu/csiq>
- [50] A. Ninassi, P. Le Callet, and F. Atrousseau, *Subjective Quality Assessment-IVC Database* [Online]. Available: <http://www2.irccyn.ec-nantes.fr/ivcdb>
- [51] Y. Horita, K. Shibata, Y. Kawayoke, and Z. M. P. Sazzad, *MICT Image Quality Evaluation Database* [Online]. Available: <http://mict.eng.u-toyama.ac.jp/mict/index2.html>
- [52] *Kodak Lossless True Color Image Suite* [Online]. Available: <http://r0k.us/graphics/kodak/>
- [53] G. Zhai, X. Wu, and J. Wang, "On monotonicity of image quality metrics," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, Kyoto, Japan, Mar. 2012, pp. 1157–1160.
- [54] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2095, Aug. 2007.
- [55] K. Friston, J. Kilner, and L. Harrison, "A free energy principle for the brain," *J. Physiol. Paris*, vol. 100, nos. 1–3, pp. 70–87, Sep. 2006.
- [56] K. Friston, "The free-energy principle: A unified brain theory?" *Nature Rev. Neurosci.*, vol. 11, pp. 127–138, Feb. 2010.
- [57] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, Apr. 2011, Art. ID 27.
- [58] VQEG. (2000, Mar.). *Final Report from the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment* [Online]. Available: <http://www.vqeg.org/>



**Ke Gu** received the B.S. degree in electronic engineering from Shanghai Jiao Tong University, Shanghai, China, in 2009, and is currently pursuing the Ph.D. degree.

His research interests include image quality assessment, contrast enhancement, and visual saliency detection.

Mr. Gu is the reviewer for the IEEE SIGNAL PROCESSING LETTERS and the *Journal of Visual Communication and Image Representation*. In 2014, he was a Visiting Student at the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada.



**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, where he is currently a Research Professor with the Institute of Image Communication and Information Processing. From 2006 to 2007, he was a Student Intern with the Institute for Infocomm Research, Singapore. From 2007 to 2008, he was a Visiting Student with the School of Computer Engineering, Nanyang Technological University, Singapore. 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 was the recipient of the Award of National Excellent Ph.D. Thesis from the Ministry of Education of China in 2012. His research interests include multimedia signal processing and perceptual signal processing.



**Xiaokang Yang** (M'00–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, 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 2000 to 2002, he was a Research Fellow with Centre for Signal Processing, Nanyang Technological University, Singapore. From 2002 to 2004, he was a Research Scientist with the Institute for Infocomm Research, Singapore. He has published 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.

Dr. Yang actively participates with the International Standards such as MPEG-4, JVT, and MPEG-21. He received the Microsoft Young Professorship Award 2006, the Best Young Investigator Paper Award at IS&T/SPIE International Conference on Video Communication and Image Processing (VCIP2003), and awards from A-STAR and Tan Kah Kee foundations. He is a member of Visual Signal Processing and Communications Technical Committee of the IEEE Circuits and Systems Society. He was the Special Session Chair of Perceptual Visual Processing of IEEE ICME2006. He is the Local Co-Chair of ChinaCom2007 and the Technical Program Co-Chair of IEEE SiPS2007.



**Wenjun Zhang** (M'02–SM'10–F'11) 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.

He worked as a Post-Doctoral Fellow at Philips Kommunikation Industrie AG, Nuremberg, Germany, from 1990 to 1993, where he was actively involved in developing HD-MAC system. He joined the Faculty of Shanghai Jiao Tong University in 1993 and became a Full Professor with the Department of Electronic Engineering in 1995. As

the National HDTV TEEG Project Leader, he successfully developed

the first Chinese HDTV prototype system in 1998. He was one of the main contributors to the Chinese Digital Television Terrestrial Broadcasting Standard issued in 2006 and is leading team in designing the next generation of broadcast television system in China from 2011. He holds over 40 patents and published over 90 papers in international journals and conferences. His main research interests include digital video coding and transmission, multimedia semantic processing, and intelligent video surveillance.

Dr. Zhang is a Chief Scientist of the Chinese National Engineering Research Centre of Digital Television (NERC-DTV), an Industry/Government Consortium with DTV Technology Research and Standardization and the Chair of Future of Broadcast Television Initiative Technical Committee.