Universal Blind Image Quality Assessment Metrics Via Natural Scene Statistics and Multiple Kernel Learning

Xinbo Gao, *Senior Member, IEEE*, Fei Gao, Dacheng Tao, *Senior Member, IEEE*, and Xuelong Li, *Fellow, IEEE*

***Abstract*— Universal blind image quality assessment (IQA) metrics that can work for various distortions are of great importance for image processing systems, because neither ground truths are available nor the distortion types are aware all the time in practice. Existing state-of-the-art universal blind IQA algorithms are developed based on natural scene statistics (NSS). Although NSS-based metrics obtained promising performance, they have some limitations: 1) they use either the Gaussian scale mixture model or generalized Gaussian density to predict the nonGaussian marginal distribution of wavelet, Gabor, or discrete cosine transform coefficients. The prediction error makes the extracted features unable to reflect the change in nonGaussianity (NG) accurately. The existing algorithms use the joint statistical model and structural similarity to model the local dependency (LD). Although this LD essentially encodes the information redundancy in natural images, these models do not use information divergence to measure the LD. Although the exponential decay characteristic (EDC) represents the property of natural images that large/small wavelet coefficient magnitudes tend to be persistent across scales, which is highly correlated with image degradations, it has not been applied to the universal blind IQA metrics; and 2) all the universal blind IQA metrics use the same similarity measure for different features for learning the universal blind IQA metrics, though these features have different properties. To address the aforementioned problems, we propose to construct new universal blind quality indicators using all the three types of NSS, i.e., the NG, LD, and EDC, and incorporating the heterogeneous property of multiple kernel learning (MKL). By analyzing how different distortions affect these statistical properties, we present two universal blind quality assessment models, NSS global scheme and NSS two-step scheme. In the proposed metrics: 1) we exploit the NG of natural images using the original marginal distribution of wavelet coefficients; 2) we measure correlations between wavelet coefficients using mutual information defined in information theory; 3) we use features of EDC in universal blind image quality prediction directly; and 4) we introduce MKL to measure the similarity of different features using different kernels. Thorough experimental results on the Laboratory for Image and Video Engineering database II and the Tampere Image Database2008 demonstrate that both metrics are in remarkably high consistency with the human perception, and overwhelm representative universal blind algorithms as well as some standard full reference quality indexes for various types of distortions.**

***Index Terms*— Exponential decay characteristic (EDC), image quality assessment (IQA), multiple kernel learning (MKL), natural scene statistics (NSS).**

Manuscript received January 25, 2013; revised May 1, 2013; accepted June 18, 2013. Date of publication July 17, 2013; date of current version November 1, 2013. This work was supported in part by the National Basic Research Program of China (973 Program) under Grant 2012CB316400, the National Natural Science Foundation of China under Grants 61125106, 61125204, 61172146, 61001203, 91120302, and 61072093, the Fundamental Research Funds for the Central Universities under Grant K5051202048, and the Shaanxi Innovative Research Team for Key Science and Technology under Grant 2012KCT-02.

X. Gao and F. Gao are with the VIPS Laboratory, School of Electronic Engineering, Xidian University, Xi'an 710071, China (e-mail: xbgao@mail.xidian.edu.cn; gaofeihifly@gmail.com).

D. Tao is with the Centre for Quantum Computation & Intelligent Systems and the Faculty of Engineering & Information Technology, University of Technology, Sydney, Ultimo, NSW 2007, Australia (e-mail: dacheng.tao@uts.edu.au).

X. Li is with the Center for OPTical IMagery Analysis and Learning (OPTIMAL), State Key Laboratory of Transient Optics and Photonics, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an 710119, Shaanxi, China (e-mail: xuelong_li@opt.ac.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/TNNLS.2013.2271356

I. Introduction

BLIND image quality assessment (IQA) metric is a computational algorithm that automatically estimates the image quality without the information about the ground truth. Or equivalently, given an input test image, we can estimate its quality using a blind IQA metric without the information about the undistorted image [1]. In recent years, many blind IQA metrics have been developed and we can group them into two classes: distortion-specific and universal metrics. Distortion-specific metrics are effective for specific distortions when the distortion information is given. In contrast, universal metrics work for different distortions. Most of the existing blind IQA metrics are distortion specific [2]–[12] and can handle particular distortions, e.g., JPEG2000 compression, JPEG compression, and blurring. Universal blind IQA models have not been well explored. Although several attempts have been proposed [13]–[17], they cannot perform as well as representative full reference (FR) and reduced reference (RR) IQA algorithms [18]–[24]. FR and RR IQA metrics, however, cannot work without a reference image. Therefore, it is important to develop universal blind IQA metrics that perform well across various types of distortions.

For universal blind quality assessment algorithms, natural scene statistics (NSS) [25] plays an important role and has been widely used in blind IQA [2], [13]–[17]. NSS is based on the hypothesis that natural images are constructed by certain criteria and have special statistics, because they can be easily distinguished from artificial images. Distortions make natural images unnatural and change NSS, so we can estimate the naturalness of a distorted natural image by measuring the

2162-237X © 2013 IEEE

deviation to particular NSS obtained from undistorted natural images.

Following the extraction of the statistical features, there are broadly two kinds of schemes that can be used to construct general-purpose blind quality metrics: the global scheme (GS) and the two-step scheme (TS). The former maps the features to a scalar quality without considering the types of distortions [13]–[15]. In contrast, the latter includes a classification stage and a quality assessment stage [16], [17]. By classification, the type of distortion of an input image is estimated. Afterward, in the quality evaluation stage, a distortion-specific blind IQA algorithm is used for each type of the distortion.

The two state-of-the-art universal blind IQA algorithms are the Distortion Identification-based Image Verity and INtegrity Evaluation (DIIVINE) [17] and BLind Image Integrity Notator using DCT Statistics-II (BLIINDS-II) [14]. In particular, for DIIVINE, Bovik *et al.* used the Gaussian scale mixture (GSM) model [26] and the generalized Gaussian distribution (GGD) [27] to simulate the non-Gaussianity (NG) [28] of the wavelet and Gabor coefficients, respectively, and employed a joint statistical model and a structural similarity index to encode the local dependency (LD) across scales, orientations, and spaces. Parameters of these models are finally exploited to estimate the image quality. DIIVINE includes a classification stage and a quality assessment stage. The type of distortion of an input image is estimated in the classification stage. Afterward, a distortion-specific blind IQA algorithm is used for each type of the distortion in the quality assessment stage. Support vector machines (SVMs) are used to classify distortions and predict the distortion-specific quality. BLIINDS-II uses GGD to model the non-Gaussian distribution of discrete cosine transform (DCT) coefficients and estimates the image quality based on their parameters. BLIINDS-II maps the features to a scalar quality without considering the types of distortions. In addition, a probabilistic graphical model is adopted to learn the mapping from these features to the scalar quality.

Although previously proposed metrics obtained promising performance, they share some of the following limitations: 1) they use either GSM or GGD to predict the non-Gaussian marginal distribution of wavelet, Gabor, or DCT coefficients. The prediction error makes the extracted features unable to reflect the change in the NG accurately. The existing algorithms use the joint statistical model and structural similarity to model the LD. Although this LD essentially encodes the information redundancy in natural images, these models do not use information divergence to measure the LD. Although the exponential decay characteristic (EDC) represents the property of natural images that large/small wavelet coefficient magnitudes tend to be persistent across scales [29], which is highly correlated with image degradations, it has not been applied to the universal blind IQA metrics; and 2) all the universal blind IQA metrics use the same similarity measure for different features to learn the universal blind IQA metrics, though these features have different properties.

To address the aforementioned problems, we develop two new universal blind IQA algorithms by integrating all the three types of NSS, i.e., the NG, LD, and EDC, involved in the following four steps: 1) we exploit the NG of natural images using the original marginal distribution of wavelet coefficients. This strategy avoids the predication error introduced by GSM and GGD. We measure correlations between wavelet coefficients using mutual information. This utilization regards the LD as information redundancy. We extract features based on EDC, and analyze the relationships between EDC and various types of distortions; and 2) we introduce multiple kernel learning (MKL) [30] to measure the similarity of different features using different kernels. In this paper, we adopt both the GS and TS to construct two universal blind IQA metrics. Extensive experiments have been conducted to verify the effectiveness of the proposed universal blind IQA algorithms.

The proposed two blind IQA algorithms are different than the representative blind IQA algorithms, i.e., BLIINDS [13], BIQI [16], DIIVINE, and BLIINDS-II, in terms of the following respects. 1) BIQI and BLIINDS-II use parameters of GGD and DIIVINE uses parameters of GSM and GGD to explore the NG, whereas the proposed IQA algorithms use the original marginal distribution to explore the NG. In contrast to DIIVINE that uses the joint statistical model and structural similarity to describe LD, our algorithms use mutual information to measure it. BLIINDS extracts features based on the structure and orientation anisotropies of natural images, but we extract features based on the NG, LD, and EDC; 2) none of existing blind IQA algorithms uses EDC; and 3) all the existing algorithms use the same similarity measure for different features, but we introduce MKL to measure the similarity of different features using different kernels.

The rest of this paper is organized as follows. Section II analyzes the statistical properties of natural images and features derived from them. In Section III, we present frameworks of the proposed quality evaluation metrics. Thorough experiments conducted on the Laboratory for Image and Video Engineering (LIVE) database II and the Tampere Image Database 2008 (TID2008) are presented and analyzed in Section IV. Section V concludes this paper.

II. NSS-Based Feature Extraction

It has been demonstrated that natural images exist strong statistical regularities that can distinguish them from the other types of images, such as infrared images, remote sensing images, and random images. To obtain NSS, an image is decomposed by a scale-space-orientation decomposition. The decomposed components preserve the structural information of the original image. Wavelet transform has been widely applied to obtain NSS [25], and this paper of NSS in the wavelet domain is systematic [31]–[36]. We thus choose to use wavelet transform in the proposed research.

According to [33], the properties of wavelet transform give the wavelet coefficients of natural images significant statistical structure, which are termed as the secondary properties: NG, clustering, and persistency. Additionally, there are two empirical tertiary properties of image wavelet coefficients: the EDC and strong persistency at finer scales. In addition, the clustering, persistency, and strong persistency at fine scales can be codified as the LD regularity, because they all present

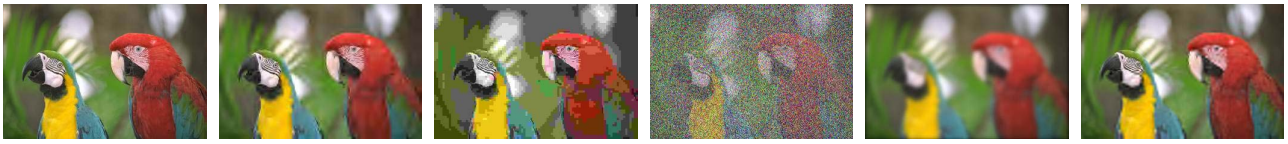


Fig. 1. Illustration of distorted images used to demonstrate the effects of distortions on NSS. From left to right: reference (REF) image, JP2k compressed image, JPEG compressed image, WN image, Gblur image, FF Rayleigh channel distorted image.

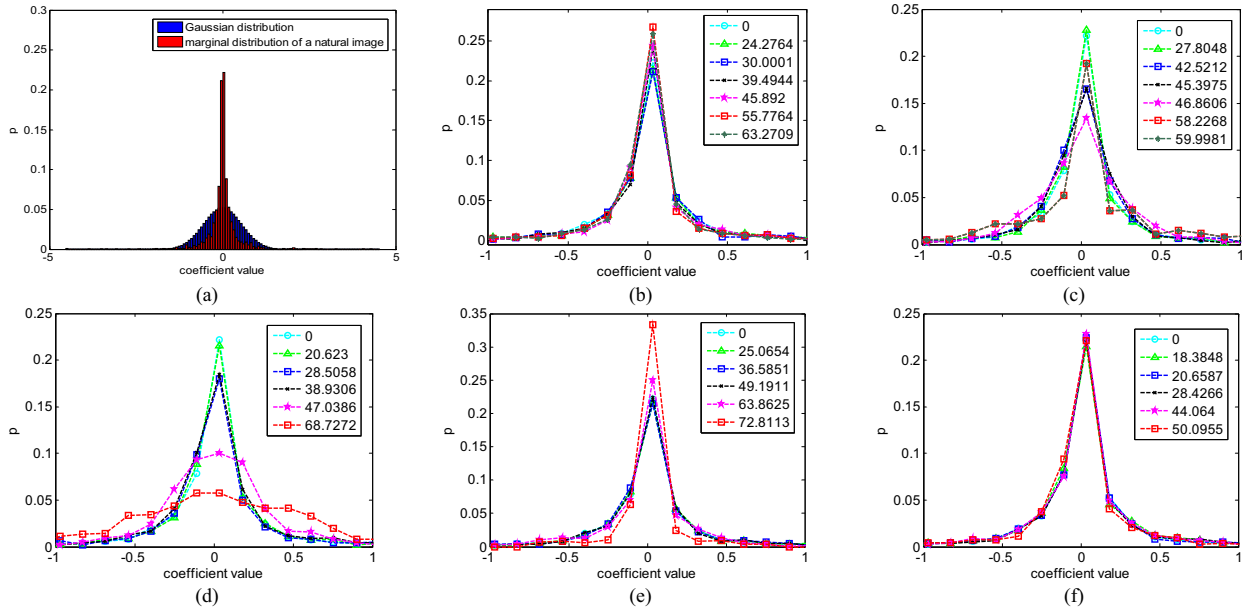


Fig. 2. Illustrations of NG of natural images and the effects of various distortions on it. (a) Comparison between the marginal distribution of wavelet coefficients of original parrots image and Gaussian distribution with the same variance. (b)–(f) Marginal distributions of wavelet coefficients of images with different extents of distortions for JP2k, JPEG, WN, Gblur, and FF. Legend: the differential mean opinion score (DMOS) of the associated image.

the high-order correlations among the wavelet coefficients of natural images.

For IQA, it is important to find some statistics that can illustrate distortions and has strong correlations with image qualities. Thus, we concentrate on the secondary and tertiary properties, which reflect the regularities of natural images. In the following sections, we introduce the features derived from NG, LD, and EDC of typical natural images and explain how distortions affect them. To show these statistical characteristics and how distortions affect them, we use a set of natural images and their distorted versions. One reference image and its distorted versions are shown in Fig. 1. As shown, we considered five common distortion types included in the LIVE database II [37], which are JPEG2000 compression (JP2k), JPEG compression (JPEG), white noise (WN), Gaussian blur (Gblur), and fast fading (FF) Rayleigh channel, in this paper.

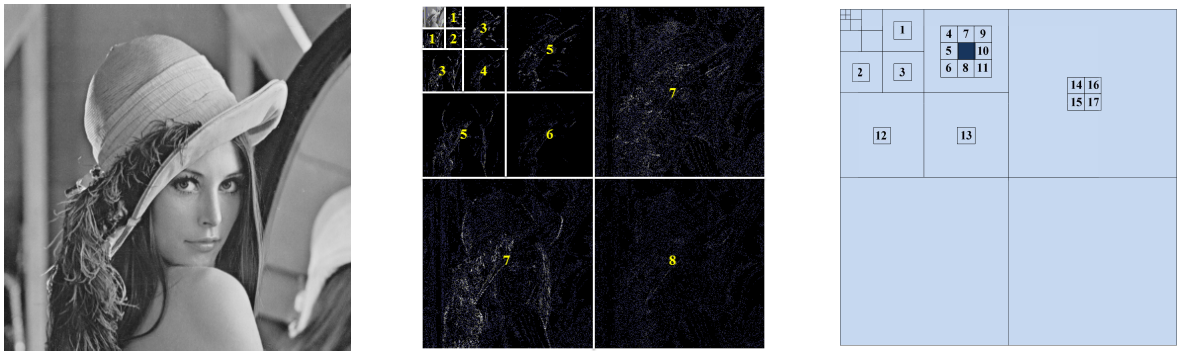
A. Non-Gaussianity

As demonstrated in [28], the wavelet coefficients of natural images have peaky and heavy-tailed marginal distributions. In other words, the distribution of wavelet coefficients of the natural image is non-Gaussian. This NG has been applied to various image processing problems, such as noise removal [34] and compression [35]. Fig. 2(a) shows the distribution of wavelet coefficients in a given subband of the reference image shown in Fig. 1(a) and a Gaussian distribution with the same variance. The distribution of wavelet coefficients has a peak in zero and a heavy tail.

Once the image is distorted, the marginal distribution will deviate from the original one, i.e., the NG changes. Fig. 2(b)–(f) shows the marginal distribution of the reference image and the distorted images with the five types of distortions, as shown in Fig. 1. The legend in Fig. 2 is the differential mean opinion score (DMOS) of the associated image. Opinion score means subjective evaluation score provided by a human observer. DMOS means the difference between the mean value of the subjective evaluation scores given by several human observers and the perfect quality. Thus, a larger DMOS shows a poorer image quality [38].

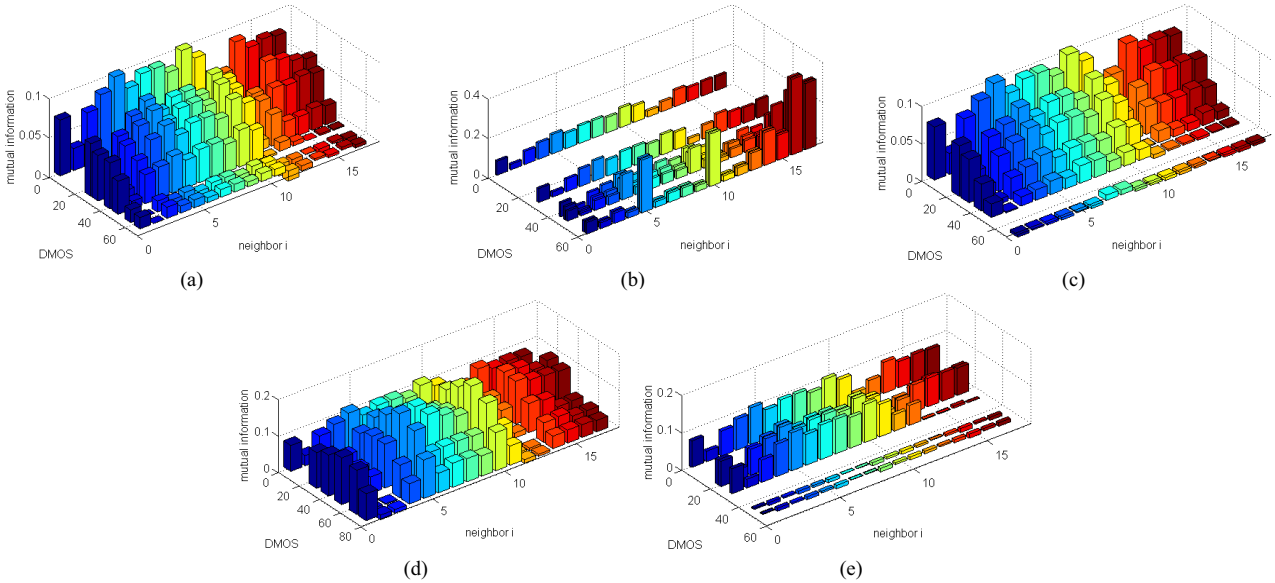
From Fig. 2, we can see that all the five types of distortions affect the marginal distribution of wavelet coefficients. For JP2k, the kurtosis of the distribution raises because most of the coefficients with low magnitudes are zero. For JPEG, blocking artifacts across block boundaries make the distribution heavier tailed. For WN, the kurtosis tends to approach to that of a Gaussian distribution. For Gblur, the magnitudes for most of the coefficients decrease, and thus the kurtosis of the distribution increases. For FF, the changes are similar to those of JP2k. For different natural images, we can obtain consistent regularities.

Considering that the raw data of the marginal distribution contain the whole information about the NG, we use the original marginal distribution of wavelet coefficients as the features to exploit the NG and to construct holistic blind IQA metrics in our framework. Given an image decomposed into L scales, for each subband in each scale, we estimate the



(a) (b) (c)

Fig. 3. LD among wavelet coefficients. (a) *Lena* image, (b) three-scale decomposition results, and (c) neighbor coefficients of a wavelet coefficient (C) in the HL subband. Dark square: C. Squares marked with numbers: its neighbor coefficients.



(a) (b) (c) (d) (e)

Fig. 4. Distributions of the mutual information versus the DMOS for the aforementioned five types of distortions. (a) JP2k. (b) JPEG. (c) WN. (d) Gblur. (e) FF.

possibility distribution, $p_k \in \mathbb{R}^{128}$, $1 \le k \le 3L$, for each subband

$$p_k = \text{histogram}(C_k) \tag{1}$$

where C_k denotes the wavelet coefficients in the kth subband and histogram$(\cdot)$ denotes the operation of calculating the histogram of random variables. Thus, we have $3L$ NG-based features. Then, we concatenate all the marginal distributions into a long feature vector

$$f_{\text{NG}} = [p_1, p_2, \ldots, p_{3L}] \tag{2}$$

for representing the NG of a natural image.

B. Local Dependency

Considering that LD essentially encodes the information redundancy in natural images [32], we measure the correlation among the neighbor coefficients for each subband using the mutual information [39] defined in information theory. Fig. 3(b) shows a four-scale decomposition of the *Lena* image in Fig. 3(a). Fig. 3(c) shows the neighbor coefficients of an example of wavelet coefficient in the LH (L = Low, H = High) subband. In Fig. 3(c), the dark square denotes the central coefficient, and light squares marked with numbers are its neighbor coefficients. We have 17 neighbor coefficients from different subbands in the same scale and the adjacent scales for each wavelet coefficient, including eight adjacent coefficients in the same subband, two cousin coefficients, one parent coefficient, two aunt coefficients, and four child coefficients. Let $\mathbb{C}^k$ be the set of the central coefficients in the kth subband, $\mathbb{C}_i^k$, $1 \le i \le 17$, be the set of the ith neighbors associated with $\mathbb{C}^k$. The mutual information between $\mathbb{C}^k$ and $\mathbb{C}_i^k$ is

$$I(\mathbb{C}^k \| \mathbb{C}_i^k) = \sum_{y \in \mathbb{C}^k} \sum_{x \in \mathbb{C}_i^k} p(x, y) \log \left(\frac{p(x, y)}{p(x) p(y)} \right) \tag{3}$$

where $p(x, y)$ is the joint probability distribution of $\mathbb{C}^k$ and $\mathbb{C}_i^k$. $p(x)$ and $p(y)$ are the marginal probability distribution functions of $\mathbb{C}^k$ and $\mathbb{C}_i^k$, respectively. Coefficients in the finest and coarsest scale do not have child or parent coefficients, only the median $(L - 2)$ scales are used to extract LD-based feature, thus f_{LD} is of dimensionality $17 \times 3 \times (L - 2)$.

Fig. 4 shows the distributions of the mutual information versus DMOS for the aforementioned five types of distortions.

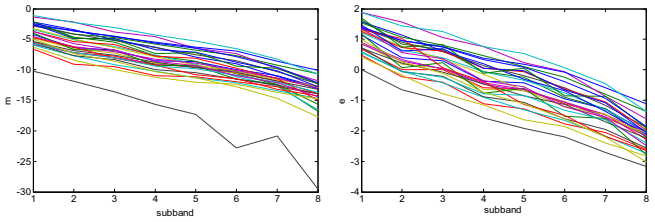


Fig. 5. Distributions of m_{EDC} and e_{EDC} of 29 reference images in the LIVE database II. Left: the plot of m_{EDC} versus subband. Right: the plot of e_{EDC} versus subband.

As shown in Fig. 4, for all the five types of distortions except JPEG, the mutual information between the central coefficient and each neighbor tends to descend with the decrease of the image quality. In contrast, the LD between the adjacent coefficients of a JPEG compressed image raises as the bit ratio becomes lower. This is because pixels in the same block in JPEG compression are equal to each other. Because f_{LD} is correlated with various distortions, it is reasonable for us to use it in the estimation of image qualities.

C. Exponential Decay Characteristic

In the 1980s, Field [31] pointed that the amplitude or energy of a natural image falls off by a factor of roughly $1/f$, wherein f is the spatial frequency. In the wavelet domain, this property shows that the energy decay exponentially as the scale becomes finer [29]. In the log domain, the relation becomes linear. We decompose an image into L scales using the wavelet transform. We use the mean of the l_2 norm of the logarithmic wavelet coefficients and the entropy of the coefficients in each subband to describe the energy. In addition, we perform a logarithmic calculation on the entropy. Because the energies in the LH and HL subbands in a given scale are almost identical, it is natural to combine their coefficients to calculate a group of features. Serial numbers of subbands when $L = 4$ are shown in Fig. 3(b). Let m_k and e_k denote the mean value and the entropy in the kth subband, respectively, $1 \leq k \leq 2L$. The mathematical expressions are as follows:

$$m_k = \frac{1}{N_k \cdot M_k} \sum_{j=1}^{N_k} \sum_{i=1}^{M_k} \log_2 \left\| C^k(i,j) \right\|_2^2 \quad \text{and} \tag{4}$$

$$e_k = \log_2 \left(-\sum_{i=1}^{N_b} p_i \ln p_i \right) \tag{5}$$

where $C^k(i,j)$ is the value of the coefficient located at (i,j) in the k th subband, m_k and N_k represent the sizes of the kth subband, p_i is the probability according to the i th bin in the marginal distribution, and N_b is the total number of bins and we set $N_b = 128$ in our experiments. Let m_{EDC} and e_{EDC} denote all the means and entropies of an image, respectively, and f_{EDC} denote the EDC-based features extracted from an image. Then, we have the following:

$$m_{\mathrm{EDC}} = [m_1, m_2, \ldots, m_{2L}] \tag{6}$$

$$e_{\mathrm{EDC}} = [e_1, e_2, \ldots e_{2L}] \quad \text{and} \tag{7}$$

$$f_{\mathrm{EDC}} = [m_{\mathrm{EDC}}, e_{\mathrm{EDC}}]. \tag{8}$$

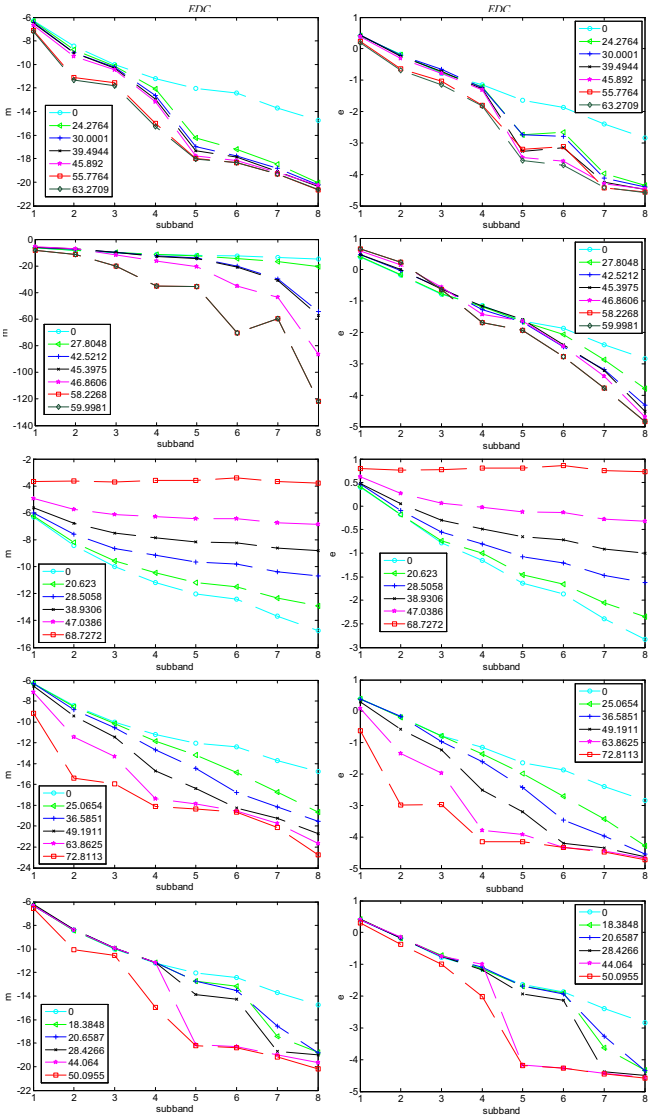


Fig. 6. Distributions of m_{EDC} and e_{EDC} of images with different types and different extents of distortions derived from one original image. From top to bottom: JP2k, JPEG, WN, Gblur, and FF. Left: m_{EDC} for the parrots image and its distorted versions. Right: e_{EDC} for the parrots image and its distorted versions. Legend: the DMOS of the associated image.

Fig. 5 shows the curves of m_{EDC} and e_{EDC} of 29 original images in the LIVE database II [37] with $L = 4$. The vertical axis of the plot represents the value of the features, and the horizontal axis represents the serial numbers of subbands. It is obvious that, for images with different contents, the distributions of m_{EDC} and e_{EDC} are similar in that they decrease linearly as the scale becomes finer.

A distorted natural image can be considered to be unnatural. Thus, the statistical characteristic may become weak or failed. Sheikh *et al.* [2] claimed that when an image was compressed using JP2k, the distribution of the amplitudes of wavelet coefficients would deviate from its original curve and the deviation would become larger as the compression rate increased. In this paper, we study the correlations between the image quality and the EDC of both m_{EDC} and e_{EDC} for the five classes of above-mentioned distortions. Fig. 6 shows the distributions of m_{EDC}

and e_{EDC} versus DMOS for the aforementioned five types of distortions. The legend in Fig. 6 is DMOS of the associated image.

For JP2k Distortion: In the JP2k compression, wavelet coefficients with low amplitudes are reduced to zero. The number of zero coefficients becomes larger as the compression rate increases. Therefore, the energy decreases with the bit rate in general. As shown in Fig. 6, the deviations between the distribution curves of m_{EDC} and e_{EDC} of the distorted image and those of the reference become larger as DMOS increases.

For JPEG Distortion: JPEG compression always introduces false edges and blur inside blocks into image contents. Because the blur considers a large portion of the regions of an image, most of the wavelet coefficients decrease in magnitude, whereas the coefficients corresponding to the false edges are large in amplitude. Thus, the mean and entropy will roughly decrease as the compression rate increases.

For WN Distortion: WN increases the number of singular points included in an image. The wavelet is adapted to capture these singularities, so the magnitude and energy of a WN image become larger. Moreover, it is notable that, for m_{EDC} and e_{EDC}, the deviation between features across different scales decreases. This can be interpreted as the fact that an image is more white as the noise becomes heavier.

For Gblur Distortion: From Fig. 6, it is easy to conclude that the mean and the entropy reduce as the distortion becomes more serious. This is mainly due to the attenuation of the energy associated with veins, while the quality degradation becomes severe, shrinking singularities contained in the image.

For FF Distortion: This is simulated by transmitting JP2k compressed images over an emulated wireless channel where bit errors are introduced. When the bit errors are at a low rate, the artifacts in the image content and the deviation of the statistical property are similar to JP2k. In contrast, when the bit error rate is high, most of the structural information will be lost and serious color distortion may occur. Therefore, m_{EDC} and e_{EDC} drop dramatically, as shown in Fig. 6.

With the above analysis and Fig. 6, we safely reach the conclusion that EDC can reflect different kinds of distortions. Moreover, for m_{EDC} and e_{EDC}, the deviation between the spectra of distorted images and that of the ground truth monotonically increases as the image quality decreases. Specifically, as the quality of an image reduces, both m_{EDC} and e_{EDC} increase for WN distortion but decrease for the other four types of distortions. In addition, these results are supported by the experiments on various natural images.

Until now, all the features based on NSS and their relationships with various distortions have been introduced. Let f_{NSS} denote all the features extracted from an image

$$\begin{aligned} f_{\mathrm{NSS}} &= [f_{\mathrm{NG}}, f_{\mathrm{LD}}, f_{\mathrm{EDC}}] \\ &= [p_1, p_2, \ldots, p_{3L}, f_{\mathrm{LD}}, m_{\mathrm{EDC}}, e_{\mathrm{EDC}}]. \end{aligned} \tag{9}$$

III. Proposed Universal Blind IQA Metrics

Previous discussions imply that the NSS-based features are highly correlated with the image quality. Thus, we present two universal blind IQA algorithms, NSS global scheme (NSS-GS) and NSS two-step scheme (NSS-TS) based on these features. Because the features have different properties, we introduce MKL to measure the similarity of different features using different kernels. In this section, we first introduce MKL and detail the construction of kernels for the proposed IQA metrics. Afterward, we sequentially present diagrams of NSS-GS and NSS-TS. Finally, we introduce the implementation details of the proposed algorithms.

A. Multiple Kernel Learning

Kernel methods [40]–[42], such as SVMs and logistic regression, are effective for many learning problems. Besides, it has been demonstrated that using multiple kernels instead of a single one can improve the performance [43]. Thus, we introduce MKL to measure the similarity of different features using different kernels.

The prediction function in kernel methods is given by

$$f(x) = \sum_{i=1}^{l} \alpha_i K(x, x_i) + b \tag{10}$$

where α_i, $1 \leq i \leq l$, and b are coefficients learned from training examples, l is the number of training examples, and K is a given positive definite kernel associated with a reproducing kernel Hilbert space (RKHS).

In MKL, the kernel K is a convex combination of a set of basis kernels

$$K(x, x_i) = \sum_{m=1}^{M} d_m K_m(x, x_i), \quad \text{with} \quad d_m \geq 0, \sum_{m=1}^{M} d_m = 1 \tag{11}$$

where M is the total number of kernels, K_m is the mth basis kernel, and d_m is the combination weight associated with K_m, $1 \leq m \leq M$. The problem of data presentation through kernels is then transferred to the learning of d_m.

Basis kernels can be either different types of kernels or kernels of the same type but with different parameters. In the proposed research, we exploit both symmetric Kullback–Leibler divergence (SKLD) and Gaussian kernels. We exploit SKLD kernel because the feature for encoding the NG, p_k, $1 \leq k \leq 3L$, can be deemed as a probability distribution. In probability and information theories, SKLD is a natural choice for measuring the distance between two probability distributions. Therefore, we use SKLD kernel, which is defined by

$$K_{\mathrm{SKLD}}(x_i, x_j) = \frac{1}{\sqrt{2\pi\sigma_{\mathrm{S}}^2}} \exp(-D_{\mathrm{SKL}}(x_i \| x_j)/2\sigma_{\mathrm{S}}^2) \tag{12}$$

where $D_{\mathrm{SKL}}(\cdot)$ is the SKLD and σ_{S} is the bandwidth of SKLD kernel. Gaussian kernel is a common choice for kernel methods. In addition, our preliminary experiments showed that Gaussian kernel is superior to polynomial kernel, so we use Gaussian kernel, which is given by

$$K_{\mathrm{G}}(x_i, x_j) = \frac{1}{\sqrt{2\pi\sigma_{\mathrm{G}}^2}} \exp(-\|x_i - x_j\|_2^2/2\sigma_{\mathrm{G}}^2) \tag{13}$$

where σ_{G} is the bandwidth of the Gaussian kernel.

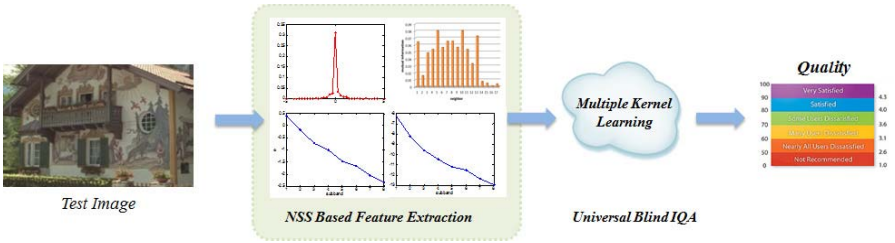


Fig. 7. Diagram of the NSS-GS method.

Each basis kernel can use all or a subset of variables that represent x. In our research, the variables in the kernels are image features. Considering the structural differences between the features based on NG, LD, and EDC, we construct a set of basis kernels for each feature, i.e., $p_1, p_2, \ldots, p_{3L}$, f_{LD}, m_{EDC} and e_{EDC}, respectively. We construct kernels in this way, because:

1) For NG-based features f_{NG}: each element, p_k, $1 \leq k \leq 3L$, represents a whole possibility distribution, so we construct a set of kernels for each possibility distribution, respectively.
2) For LD-based feature f_{LD}: all the elements in f_{LD} have the same physical interpretation and mutual information, so we treat f_{LD} as a whole and construct a set of kernels for it.
3) For EDC-based features f_{EDC}: f_{EDC} contains two types of features: m_{EDC} and e_{EDC}, we thus construct a set of kernels for m_{EDC} and e_{EDC}, respectively.

It has been shown that MKL can favorably replace cross validation on kernel parameters in [44]. Accordingly, we empirically choose a set of σ_{G} and σ_{S} from {0.001, 0.01, 0.05, 0.1, 0.5, 1, 3, 6, 10, 20, 50, 100} for Gaussian and SKLD kernels in the experiments. Therefore, we have 12 bandwidths for both Gaussian and SKLD kernels. Because we have different kinds of features and 12 different bandwidths for each of Gaussian and SKLD kernels, we present Gaussian/SKLD kernels with 12 different bandwidths in the divisive mode according to SimpleMKL. In addition, 12 Gaussian kernels using the integrated features f_{NSS} are exploited to encode the potential correlations between features derived from different statistical properties. In sum, we have the following:

1) $72L$ kernels for f_{NG}: (12 bandwidth Gaussian kernels +12 bandwidth SKLD kernels) $\times 3L$ features;
2) 24 kernels for f_{LD}: (12 bandwidth Gaussian kernels +12 bandwidth SKLD kernels) $\times$ one feature;
3) 48 kernels for f_{EDC}: (12 bandwidth Gaussian kernels +12 bandwidth SKLD kernels) $\times$ two features;
4) 12 kernels for f_{NSS}: 12 bandwidth Gaussian kernels.

Or equivalently, we have the following:

1) Gaussian kernels with 12 different bandwidths in the divisive mode and using the integrated features;
2) SKLD kernels with 12 different bandwidths in the divisive mode.

B. NSS Global Scheme

With the discussions in Section II, if we only consider the magnitudes of departures in NSS caused by distortions and ignore slight differences between different types of distortions, a unique mapping f_{NSS} to DMOS could be employed to model the numerical relationship between them. Thus, we use MKL to approximate it by a set of training data to determine the mapping from the observation. The kernels are constructed according to Part *A*. Once MKL has been trained from the training database consisting of f_{NSS} and DMOS, for an input image, we can estimate its quality index Q. We call the general blind IQA algorithm based on NSS using the GS (NSS-GS). Its diagram is shown in Fig. 7.

C. NSS Two-Step Scheme

If we consider the differences between deviations in NSS caused by different types of distortions and construct a distinct IQA algorithm for each type of distortion, it is likely we will obtain a better quality evaluation result for each distortion. For the purpose of a universal index, a classifier is added before the IQA stage to determine the distortion type of an input image. Given a test image, we first estimate its distortion type using the classification model, and then evaluate its quality using the associated distortion-specific quality metric. Here, we assume that the distortion type included in an image is unitary.

Considering the deviation between distributions of the extracted features, we use NSS-based features f_{NSS} as the input of the classifier. In the classification phase, a multiclass MKL is used. The establishment of the specific IQA metric for each distortion is similar to NSS-GS in that an MKL is used to learn the mapping from f_{NSS} to DMOS for each category of distortion except the reference. The kernels are constructed according to Part *A*. The quality of a reference image is perfect. Because this metric uses NSS and the TS, we call it as NSS-TS. It can be regarded as an extension of NSS-GS by considering more statistics. Fig. 8 shows the diagram of NSS-TS.

D. Implementation Details

The primal MKL problem is given by

$$\begin{aligned} &\min_{\{f_m\},b,\xi,d} \frac{1}{2}\sum_{m=1}^{M}\frac{1}{d_m}\|f_m\|_{H_m}^2 + C\sum_i \xi_i \\ &\text{s.t. } y_i\sum_{m=1}^{M} f_m(x_i) + y_i b \geq 1-\xi_i, \quad \xi_i > 0\ \forall i \\ &\sum_{m=1}^{M} d_m = 1 \end{aligned} \tag{14}$$

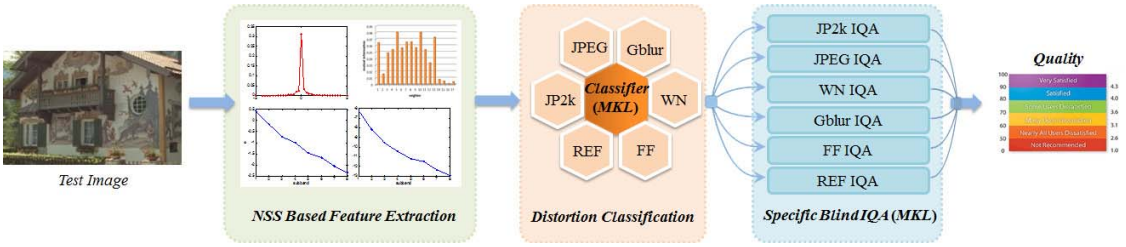


Fig. 8. Diagram of the NSS-TS method.

where each function f_m belongs to a particular RKHS H_m associated with a kernel K_m, C is a penalty parameter, and $\xi_i, \forall i$ is the training loss.

In our research, MKL is employed to construct the universal blind IQA metric for NSS-GS, the classification model for NSS-TS, and the distortion-specific blind IQA methods for NSS-TS. The input of the MKL algorithm is the image features, and the output is the quality score or the label of the distortion type for an input test image. In classic IQA databases, such as the LIVE database II [37] and the TID2008 database [45], both the distortion type and the quality score of each image are available. In this paper, we considered five common distortion types included in the LIVE database II. We use labels {1, 2, 3, 4, 5, 6} to show JP2k, JPEG, WN, Gblur, FF, and undistorted. In the experiments, we choose part of or all the images included in the LIVE database II as the labeled data, and use them to train the proposed IQA metrics.

Several approaches have been proposed to solve the MKL problem [40], [44]. In our research, the SimpleMKL toolbox [44] has been employed. Experiments in SimpleMKL [44] show that MKL can favorably replace cross validation on kernel parameters and performs better than cross validation. This conclusion is in accordance with other works about MKL [46]–[48]. Therefore, we empirically select the parameter C of an MKL algorithm instead of using cross validation or other classical methods.

According to [44], we define the parameter C and the stopping criteria to terminate the iterative procedures before using an MKL algorithm. We adopt the settings of stopping criteria given in the SimpleMKL toolbox for their reliability. In addition, we select the optimal parameter C of MKL using the data of the LIVE database II. The MKL parameter selection method for choosing the optimal model parameter C follows the suggestion given in SimpleMKL [44] and details are listed in Table I. Table I shows the procedure for selecting the optimal C before learning the optimal parameters of the MKL algorithm.

In Step 6, if both e_{train} and e_{test} are large, the MKL algorithm is underfitted, and we can increase C to improve the prediction performance. If the trained MKL algorithm has small e_{train} but large e_{test}, it is overfitted. In this case, we decrease C. In this way, we can obtain the optimal parameter C. Afterward, we can learn the optimal $d_m, 1 \le m \le M$ and $\xi_i, \forall i$ based on the labeled data and then test the MKL algorithm on the unlabeled data. Experimental results show that the performance of the MKL algorithm is stable when C ranges during a small interval. If C is, however, far from the optimal interval, the MKL algorithm becomes overfitted or underfitted. Therefore, the value of C is important for MKL-based IQA.

TABLE I
MKL PARAMETER SELECTION METHOD

Parameter selection procedure

1. Initialize the model parameter C: the SimpleMKL toolbox suggests C=100, so we chose C=100 as the initialization;
2. Randomly choose 23 original images and their associated distorted images for training, and the rest for test;
3. Learn the combination weights d_m, $1 \le m \le M$ and $\xi_i, \forall i$ of the MKL algorithm on the training set;
4. Predict the output values for images in the test set;
5. Calculate the mean error on the training set, e_{train}, and the error on the test set, e_{test}.
6. Manually modify the parameter C according to e_{train} and e_{test}. If both e_{train} and e_{test} are large, the MKL algorithm is under-fitted, and we can increase C to improve the prediction performance. If the trained MKL algorithm has small e_{train} but large e_{test}, it is over-fitted. In this case, we decrease C;
7. Iterate steps 2-6 until both small e_{train} and e_{test} are achieved.

Outputs: Optimal parameter C of MKL.

Finally, we chose $C = 10^4$ for NSS-GS, $C = 800$ for the classification model of NSS-TS, and $C = 2000$ for all the distortion-specific IQA methods of NSS-TS in the experiments. The promising experimental results of the proposed blind IQA algorithms on various data sets demonstrate that the selected parameters are data independent.

IV. EXPERIMENTAL RESULTS

In this section, we compare the performances of NSS-GS and NSS-TS with a number of standard FR IQA metrics, e.g., peak signal-to-noise ratio (PSNR), the Structural SIMilarity index (SSIM) [18], the Visual Information Fidelity index (VIF) [19], FSIM [21] and several classical blind IQA metrics, e.g., [13] (BLIINDS), [14] (BLIINDS-II), [16] (BIQI), and [17] (DIIVINE) based on the consistency experiment. For SSIM [18], VIF [19], and FSIM [21], we used the implementations provided by the original authors, which are available at [21] and [37]. We conduct experiments according to the setting used in [16] and [17], so we can compare our algorithms with DIIVINE, BIQI, BLIINDS, and BLIINDS-II directly using the results reported in [16] and [17]. In addition, we will verify the robustness of NSS-GS and NSS-TS and test the performance of NSS-GS on images distorted by hybrid degradations.

We conduct experiments on the LIVE database II [37] and the TID2008 database [45] to verify the performance and generalization ability of the proposed algorithms. The LIVE database II and the TID2008 database are the largest two IQA databases and they are widely used in IQA publications.

The LIVE database II includes 779 distorted images in sum, which are generated from 29 original images by corrupting them with five types of distortions, i.e., JP2k (169 images), JPEG (175 images), WN (145 images), Gblur (145 images), and FF (145 images). The TID2008 database contains 1700 test images, which are derived from 25 reference images. There are 17 types of distortions for each reference image and four different scales of each type of distortion.

In the experiments, only the luminance component of an image is employed. One image is decomposed into four scales using the wavelet transform in the experiments. It is notable that the performance is not sensitive to the value of the cost. We select training samples and test samples based on the methods that are widely used in the classical blind IQA works [13]–[17]. In Parts *A* and *B*, we randomly choose 23 original images and their associated distorted images for training, and the rest for test. There are about 640 labeled images. Each labeled image has a quality score and a label of the distortion type. It is worth emphasizing that there is no overlap between the training and the test sets. We conduct the above evaluation 100 times to verify the robustness of our metrics. In Parts *C* and *D*, we test our metrics on the TID2008 database and on the images with hybrid distortions after trained on the LIVE database II. In this case, we adopt 29 original images and 779 distorted images included in the LIVE database II as the labeled data. Thus, there are 808 labeled images, and each labeled image has a quality score and the corresponding distortion type.

To verify the important dependency between the proposed NSS features and MKL methods, we test the performance of the proposed frameworks using a single-kernel learning method instead of MKL. Because SVM has the advantages of low computational cost and strong generalization ability, we adopt SVM with a radial basis function kernel as the single-kernel method in our experiments. For clarity, we refer the algorithm using SVM and the GS as NSS-GS$^{\text{SVM}}$, and the algorithm using SVM and the TS as NSS-TS$^{\text{SVM}}$. In our experiments, the LIBrary for Support Vector Machines (LIBSVM) package [49] is used to implement the SVM algorithm. The optimal parameters of SVM are learned through a fivefold cross validation on the training set.

To evaluate the performances of the proposed algorithms, four indexes are adopted as the criteria: the Pearson's linear correlation coefficient (PLCC), the Spearman's rank ordered correlation coefficient (SROCC), mean absolute error (MAE), and root-mean-square error (RMSE) between the estimated quality index, Q and the true DMOS. Before the computation of these indexes, the logistic function provided by the video quality experts group (VQEG) [50] is employed.

A. Consistency Experiment

The consistency between the estimated quality indexes and the subject evaluation results is one of the most important criteria for evaluating the performance of IQA algorithms. Fig. 9 is the scatter plot of NSS-GS and NSS-TS versus the predicted score in one trial on the entire test set. In each plot, the x-coordinate is the subjective evaluation score, DMOS, and the y-coordinate is the predicted result, Q, predicted by NSS-GS/NSS-TS. Each point corresponds to a test image. In addition, the solid line is the fitted result using the logistic function provided by VQEG. Fig. 9 thus shows the monotonicity and consistency between Q and DMOS. Fig. 9 shows that the predicted results by both NSS-GS and NSS-TS are highly consistent with the subjective evaluation scores. This demonstrates that both NSS-GS and NSS-TS have an impressive consistency with human perception.

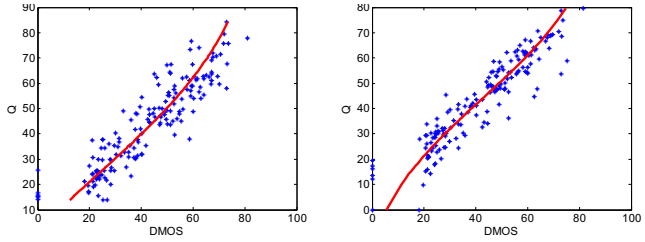


Fig. 9. Scatter plot of the predicted results Q versus DMOS over the entire test set on LIVE database II. Left: MKL-GS. Right: MKL-TS.

The performance of the metrics we compared with and the median performances of NSS-GS$^{\text{SVM}}$, NSS-TS$^{\text{SVM}}$, NSS-GS, and NSS-TS over 100 trials are presented in Table II. In addition, we have highlighted the best universal blind metric in boldface for each distortion for PLCC and SROCC, because PLCC and SROCC are widely used in IQA for performance evaluation. To highlight the superiority of the proposed blind IQA metrics over the representative blind IQA metrics, we only list the ranking of the blind IQA metrics in Table III. In particular, we rank all blind IQA metrics (except NSS-GS$^{\text{SVM}}$ and NSS-TS$^{\text{SVM}}$) according to SROCC according to [21].

From Tables II and III, we can see that both NSS-GS and NSS-TS perform consistently well in terms of correlation with human perception for all five types of distortions. For JP2k, JPEG, and WN, although the proposed are not the best among all the universal blind IQA metrics, they are still highly consistent with subjective evaluation results. For Gblur and FF, NSS-TS obtains the best performance, and NSS-GS performs very well, which is better than BIQI, BLIINDS, and DIIVINE.

Because universal blind IQA algorithms aim to work for all types of distortions, the performance across the entire test set is significant to verify whether the same score is predicted for images with the same quality but different distortion categories. Table II demonstrates that both NSS-GS and NSS-TS perform very well on the entire test set, and the samples shown in Fig. 9 corroborate the expectation that the predicted scores for images with similar qualities but different categories of distortions do not vary greatly from one another.

In addition, NSS-GS performs better than NSS-GS$^{\text{SVM}}$, and NSS-TS performs better than NSS-TS$^{\text{SVM}}$ for most of the distortions and for the whole test set. We can safely draw the conclusion that MKL can catch the numerical relationship between the new NSS features and the image quality better than the single learning kernel method. In addition, both NSS-TS$^{\text{SVM}}$ and NSS-GS$^{\text{SVM}}$ obtain comparable performances with state-of-the- art blind IQA metrics, implying the strong correlation between the proposed NSS features and the image quality.

TABLE II

PERFORMANCES OF PSNR, SSIM, VIF, FSIM, BLIINDS, BIQI, DIIVINE, BLIINDS-II, NSS-GS, AND NSS-TS ON LIVE DATABASE II

Distortion Types		JP2k				JPEG			
Metric	Type	PLCC	SROCC	RMSE	MAE	PLCC	SROCC	RMSE	MAE
PSNR	FR	0.896	0.890	7.187	5.528	0.860	0.841	8.170	6.380
SSIM [18]	FR	0.937	0.932	5.671	4.433	0.928	0.903	5.947	4.485
VIF [19]	FR	0.962	0.953	4.449	3.445	0.943	0.913	5.321	3.807
FSIM [21]	FR	0.962	0.982	6.705	5.592	0.952	0.963	7.457	6.119
BIQI [16]	Blind	0.750	0.736	16.540	--	0.630	0.591	24.580	--
BLIINDS [13]	Blind	0.807	0.805	14.780	--	0.597	0.552	25.320	--
DIIVINE [17]	Blind	0.922	0.913	9.660	--	0.921	0.910	12.250	--
BLIINDS-II [14]	Blind	**0.963**	**0.951**	--	--	**0.979**	**0.942**	--	--
NSS-GS^{SVM}	Blind	0.937	0.927	6.276	7.660	0.919	0.891	6.974	8.612
NSS-TS^{SVM}	Blind	0.944	0.931	5.692	7.236	0.928	0.902	6.273	7.925
NSS-GS	Blind	0.933	0.938	6.383	7.822	0.904	0.901	7.019	9.141
NSS-TS	Blind	0.947	0.931	5.792	7.169	0.933	0.915	6.333	7.912
Distortion Types		WN				Gblur			
Metric	Type	PLCC	SROCC	RMSE	MAE	PLCC	SROCC	RMSE	MAE
PSNR	FR	0.986	0.985	2.680	2.164	0.783	0.782	9.772	7.743
SSIM [18]	FR	0.970	0.963	3.916	3.257	0.874	0.894	7.639	5.760
VIF [19]	FR	0.984	0.986	2.851	2.304	0.974	0.973	3.533	2.818
FSIM [21]	FR	0.929	0.980	8.130	7.287	0.952	0.983	6.650	5.607
BIQI [16]	Blind	0.968	0.958	6.930	--	0.800	0.778	11.100	--
BLIINDS [13]	Blind	0.914	0.890	11.270	--	0.870	0.834	9.080	--
DIIVINE [17]	Blind	**0.988**	**0.984**	4.310	--	0.923	0.921	7.070	--
BLIINDS-II [14]	Blind	0.985	0.978	--	--	0.948	0.944	--	--
NSS-GS^{SVM}	Blind	0.960	0.970	4.612	6.119	0.941	0.945	5.921	7.419
NSS-TS^{SVM}	Blind	0.972	0.972	3.700	5.233	**0.953**	**0.952**	4.934	6.448
NSS-GS	Blind	0.931	0.945	5.974	8.001	0.928	0.936	6.571	8.366
NSS-TS	Blind	0.963	0.971	4.464	6.018	0.950	0.939	5.481	6.863
Distortion Types		FF				All			
Metric	Type	PLCC	SROCC	RMSE	MAE	PLCC	SROCC	RMSE	MAE
PSNR	FR	0.890	0.890	7.516	5.800	0.824	0.820	9.124	7.325
SSIM [18]	FR	0.943	0.941	5.485	4.297	0.863	0.851	8.126	6.275
VIF [19]	FR	0.962	0.965	4.502	3.547	0.950	0.953	5.024	3.887
FSIM [21]	FR	0.921	0.971	8.590	7.421	0.731	0.944	23.116	19.647
BIQI [16]	Blind	0.722	0.700	19.480	--	0.740	0.726	18.360	--
BLIINDS [13]	Blind	0.743	0.678	18.620	--	0.680	0.663	20.010	--
DIIVINE [17]	Blind	0.888	0.863	12.930	--	0.917	0.916	10.900	--
BLIINDS-II [14]	Blind	**0.944**	0.927	--	--	0.923	0.920	--	--
NSS-GS^{SVM}	Blind	0.912	0.895	6.911	9.076	0.911	0.916	5.693	7.342
NSS-TS^{SVM}	Blind	0.912	0.894	6.197	9.000	0.909	0.913	5.307	7.474
NSS-GS	Blind	0.920	0.913	6.400	8.597	0.918	0.929	5.280	7.144
NSS-TS	Blind	0.942	**0.935**	5.232	7.070	**0.926**	**0.930**	5.131	6.803

It is rather inspiring that NSS-GS and NSS-TS perform best on the entire database among all the universal blind IQA algorithms. Besides, they are even better than or comparable with the state-of-the-art reference IQA metrics, such as SSIM, VIF, and FSIM. Table IV shows the standard deviation of each criterion over 100 trials. The proposed algorithms are insensitive to data sets because the standard deviations are small. Both NSS-GS and NSS-TS obtain inspiring results because they extract image features based on all the three types of NSS, i.e., the NG, LD, and EDC, and incorporate the heterogeneous property of MKL, as analyzed in previous sections. It is notable that the performance of NSS-TS is better than that of NSS-GS, because NSS-TS considers more detailed information and can catch the numerical relationship between the statistical features and DMOS more precisely.

B. Classification Accuracy

Classification plays a significant role in NSS-TS. In addition, the classification accuracy affects the final quality prediction performance greatly. Among the 100 trials of NSS-TS, the medians classification accuracies for each type of distortion and for the entire test set are shown in Table V, and the standard deviations of classification accuracies are shown in Table VI. From Tables V and VI, we can safely draw the conclusion that the classification performance of NSS-TS is much better than DIIVINE. This is consistent with the good performance of NSS-TS.

TABLE III

RANKING OF UNIVERSAL BLIND IQA METRICS' PERFORMANCES ON THE FIVE SUBSETS AND ENTIRE DATABASE ON LIVE DATABASE II

	JP2k	JPEG	WN	Gblur	FF	All
BIQI [16]	6	5	4	6	5	5
BLIINDS [13]	5	6	6	5	6	6
DIIVINE [17]	4	3	1	4	4	4
BLIINDS-II [14]	1	1	2	1	2	3
NSS-GS	2	4	5	3	3	2
NSS-TS	3	2	3	2	1	1

TABLE IV

STANDARD DEVIATIONS OF NSS-GS' AND NSS-TS' PERFORMANCES ON THE FIVE SUBSETS AND ENTIRE DATABASE ON LIVE DATABASE II

	NSS-GS				NSS-TS			
	PLCC	SROCC	RMSE	MAE	PLCC	SROCC	RMSE	MAE
JP2k	0.025	0.025	1.034	1.287	0.027	0.033	1.173	1.447
JPEG	0.025	0.025	1.005	1.190	0.015	0.026	0.876	0.887
WN	0.023	0.021	0.883	1.224	0.015	0.013	0.963	1.158
Gblur	0.028	0.029	1.068	1.423	0.031	0.033	1.301	1.787
FF	0.036	0.036	1.203	1.652	0.025	0.030	1.016	1.427
All	0.018	0.018	0.566	0.756	0.023	0.023	0.661	0.953

TABLE V
MEDIAN CLASSIFICATION ACCURACIES (%) OF DIIVINE AND NSS-TS ON LIVE DATABASE II OVER 100 TRIALS

	JP2k	JPEG	WN	Gblur	FF	All
DIIVINE[17]	80.00	81.10	100.00	90.00	73.30	83.75
NSS-TS	94.44	94.37	96.67	93.33	93.33	93.39

TABLE VI
STANDARD DEVIANCES OF CLASSIFICATION ACCURACIES (%) OF NSS-TS ON LIVE DATABASE II OVER 100 TRAILS

	JP2k	JPEG	WN	Gblur	FF	All
NSS-TS	7.37	3.93	3.16	6.80	6.66	4.11

C. Database Independence

Because the universal blind IQA metrics are based on a learning machine, it is necessary to verify whether the parameters are overfitted and the learned model is sensitive to different databases. We train both NSS-GS and NSS-TS on the LIVE database II, and then test them on the TID2008 database. Because the features used in this paper are based on the natural images and trained on the LIVE database II. We select the 24 natural images contained in the TID2008 database and obtain the WN, Gblur, JPEG, and JP2k images, according to [17]. Table VII shows the SROCC between the predicted results and DMOS. The results of PSNR, SSIM, and DIIVINE [17] are tabulated as well for comparison.

In addition, we test NSS-GS and NSS-TS on the other 13 types of nonlearned distortions inside the TID2008 database, i.e., additive noise in color components (ANC), spatially correlated noise (SCN), masked noise (MN), high-frequency noise (HFN), impulse noise (IN), quantization noise (QN), image denoising (ID), JPEG transmission errors (JPEGTE), JP2k transmission errors (JP2kTE), noneccentricity pattern noise (NEPN), local blockwise distortions (LBD) of different intensity, intensity shift (IS), and contrast change (CC). Similarly, only images derived from the 24 natural images are considered. The SROCC between the predicted results and subjective quality scores for each nonlearned distortion is shown in Table VIII.

It is clear that NSS-GS performs better than NSS-TS on JP2k, JPEG, Gblur, and on the whole learned subsets. For WN and the nonlearned distortions are, however, similar to the learned ones [51], e.g., ANC, HFN, IN, (all of them are similar to WN), and ID (ID includes blurring artifacts), NSS-TS performs better than NSS-GS. As a whole, the performances of NSS-GS and NSS-TS therefore are comparable with the TID2008 database. Considering both the LIVE and TID2008 databases, the performance of NSS-TS is slightly better than that of NSS-GS.

Because MKL is a supervised learning technique, the proposed algorithms do not perform very well for several nonlearned degradations. They, however, still obtain promising results for most of the nonlearned distortions. This verifies the efficiency and generalization of the proposed IQA algorithms.

With the experimental results, we can safely conclude that both of the proposed metrics are database independent. Once trained on a properly prepared database, both NSS-GS and NSS-TS can be applied to images with arbitrary distortions, especially when the distortions have been covered in the training stage.

TABLE VII
SROCC BETWEEN THE PREDICTED RESULTS AND DMOS ON THE LEARNED SUBSETS OF THE TID2008 DATABASE

	JP2k	JPEG	WN	Gblur	All
PSNR	0.8250	0.8760	0.9180	0.9340	0.8700
SSIM	0.9630	0.9350	0.8170	0.9600	0.9020
DIIVINE[17]	0.9240	0.8660	0.8510	0.8620	0.8890
NSS-GS	0.9022	0.8592	0.7768	0.8742	0.8592
NSS-TS	0.8185	0.8488	0.8094	0.8465	0.8475

TABLE VIII
SROCC BETWEEN THE PREDICTED RESULTS AND DMOS ON THE NON-LEARNED SUBSETS OF THE TID2008 DATABASE

	ANC	SCN	MN	HFN	IN	QN	ID
NSS-GS	0.7128	0.6867	0.5065	0.8904	0.8685	0.6337	0.7696
NSS-TS	0.8373	0.6347	0.6224	0.8835	0.8880	0.3159	0.7977
	JPEGTE	JP2kTE	NEPN	LBD	IS	CC	
NSS-GS	0.3739	0.5996	0.0529	0.5374	0.1441	0.1680	
NSS-TS	0.0509	0.4322	0.1689	0.7528	0.3773	0.0330	

D. Hybrid Distortions

It is possible that the images are degraded by hybrid distortions, such as noise, compression, and blur. The performance on images distorted by a combination of distortions is a significant criterion for universal IQA. To evaluate the proposed algorithms, we generated a set of images with hybrid distortions. The distortions are as follows.

1) *JP2k_WN*: Images were generated by adding WN on the JP2k compressed parrots image provided in the LIVE database II. The WN with a standard deviation σ_N ranging from 0.001 to 1 was added to the RGB components after scaling the values of each component between 0 and 1.
2) *JPEG_WN*: The reference image is compressed using JPEG and then distorted by the white noise. The implementation of JPEG was the *imwrite*function provided by MATLAB. In addition, the bit rates were between 0.152 and 5.50 bpp. The white noisy image was implemented using the codec introduced above.
3) *Gblur_JPEG_WN*: The R, G, and B components of a reference image were blurred using the same Gaussian kernel. Afterward, the Gaussian blurring image is compressed using JPEG and then distorted by white noise. For Gblur, the standard deviation of the Gaussian kernel σ_B was changed from 1 to 5. JPEG and WN were implemented using the same codec above.

Fig. 10 shows examples of the distorted images. Features of the three groups of images are shown in Fig. 11. In the plots for each group, the reference image is labeled as 1, and the distorted images are labeled as 2–6. From 1 to 6, the distortion becomes severer.

From Fig. 11, we can see that every type of degradation contained in the hybrid distortions can be represented by the features extracted. For degradations that have contrary effects on some of the features, they may cascade in the other features.

(a) (b) (c) (d)

Fig. 10. Illustration of images suffered from hybrid distortions. (a) JP2k_WN, 0.028 bpp, and $\sigma_{\mathrm{N}} = 0.001$ (b) JP2k_WN, 0.028 bpp, and $\sigma_{\mathrm{N}} = 0.03$. (c) JPEG_WN, 0.176 bpp, and $\sigma_{\mathrm{N}} = 1$. (d) Gblur_JPEG_WN, 0.263 bpp, $\sigma_B = 2$, and $\sigma_{\mathrm{N}} = 0.1$.

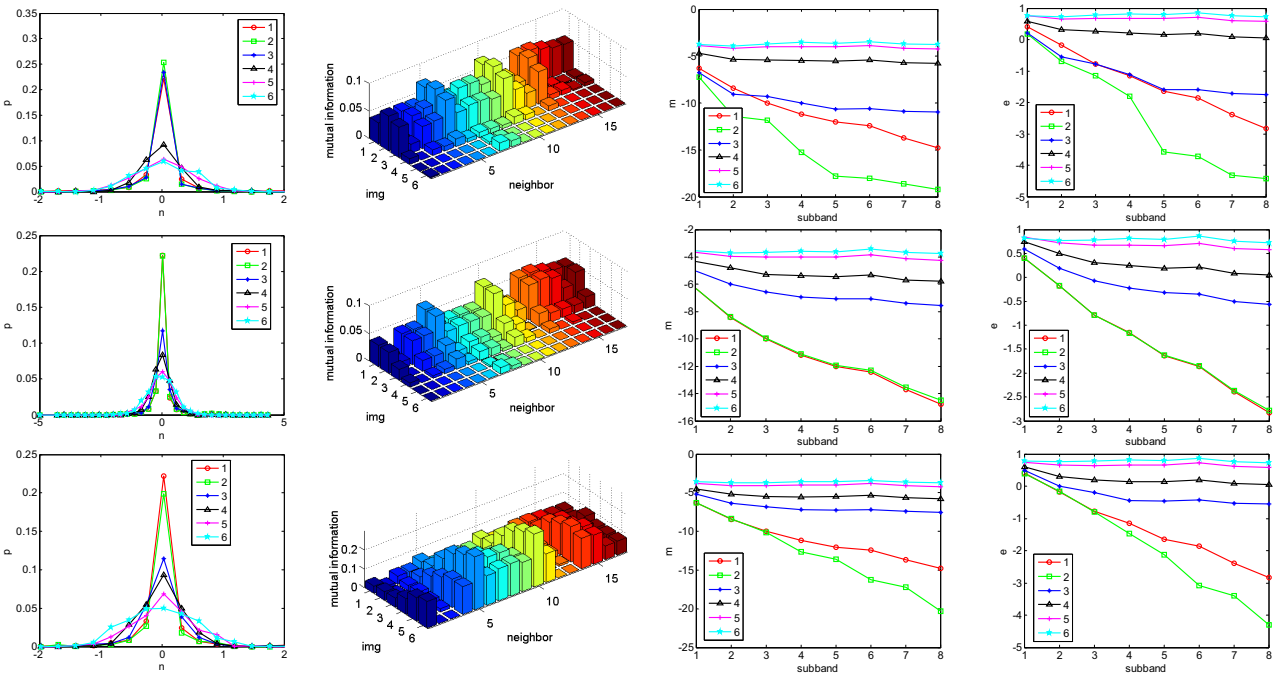


Fig. 11. Features of images distorted by a combination of degradations. From top to bottom: JP2k_WN, JPEG_WN, and Gblur_JPEG_WN. From left to right: NG, LD, and m_{EDC} and e_{EDC} for EDC.

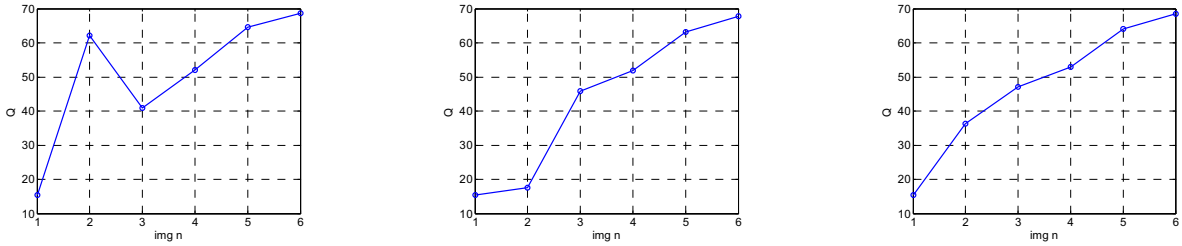


Fig. 12. Performance of NSS-GS on hybrid distortions. From left to right: JP2k&WN, JPEG_WN, and Gblur_JPEG_WN.

In addition, the deviation tends to resemble that of the dominate type of distortion. Considered JP2k_WN for example, JP2k decreases the values of m_{EDC} and e_{EDC}, whereas WN increases them, so the deviations in m_{EDC} and e_{EDC} may not change in proportion to the extent of the degradation when the deviations caused by JP2k and WN eliminate each other. Both JP2k and WN, however, decrease the correlations between adjacent coefficients. Thus, the degradation of image 3 can be captured by f_{LD}. When WN becomes much severer than JP2k, the changes in features resemble those of WN as illustrated in Section II.

We train NSS-GS on the LIVE database II, and then test it on the distorted images. To evaluate the performance, we test the monotonicity between the prediction results and the rank of the subjective evaluation results. The results are shown in Fig. 12. From the relationship between the predicted DMOS and image labels, we can see that the proposed metric performs well for most images distorted by a combination of distortions. For images whose degradations counterbalance the changes caused by each other seriously, the predicted result is inaccurate.

V. Conclusion

In this paper, the statistical properties of natural images in the wavelet domain, i.e., the NG, LD, and EDC, and their

relationships with image qualities are analyzed, resulting in two universal blind quality assessment metrics based on the GS and the TS. The comparison of both proposed algorithms with standard FR and blind IQA metrics on the LIVE database II and the TID2008 database shows they have an impressive consistency with human perception and overwhelming superiority over the standard blind IQA metrics. They even bear comparison with some FR algorithms for images corrupted by certain distortions or across the overall database. Nevertheless, they are still inferior to the best reference quality metrics. Because NSS have shown great potential in IQA, it is a valuable way to aggressively explore the relevance between the statistical properties and the image degradations for other image processing tasks, e.g., image synthesis [52], [53] and superresolution reconstruction [54], and to develop effective universal blind IQA metrics by extending the MKL framework. In addition, how to evaluate the quality of an image with mixed multidistortion precisely will be another future work.

REFERENCES

[1] Z. Wang and A. C. Bovik, *Modern Image Quality Assessment*. New York, NY, USA: Morgan & Claypool, 2006.

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

[3] P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi, "Perceptual blur and ringing metrics: Application to JPEG2000," *Signal Process., Image Commun.*, vol. 19, no. 2, pp. 163–172, Feb. 2004.

[4] S. Suresh, R. V. Babu, H. J. Kim, "No-reference image quality assessment using modified extreme learning machine classifier," *Appl. Soft Comput.*, vol. 9, no. 2, pp. 541–552, Mar. 2009.

[5] S. Suthaharan, "No-reference visually significant blocking artifact metric for natural scene images," *Signal Process.*, vol. 89, no. 8, pp. 1647–1652, Aug. 2009.

[6] R. Ferzli and L. J. Karam, "A no-reference objective image sharpness metric based on the notion of just noticeable blur (JNB)," *IEEE Trans. Image Process.*, vol. 18, no. 4, pp. 717–728, Apr. 2009.

[7] S. Wu, W. Lin, S. Xie, Z. Lu, E. P. Ong, and S. Yao, "Blind blur assessment for vision-based applications," *J. Vis. Commun. Image Represent.*, vol. 20, no. 4, pp. 231–241, May 2009.

[8] A. Ciancio, A. L. N. Targino da Costa, E. A. B. da Silva, A. Said, R. Samadani, and P. Obrador, "No-reference blur assessment of digital pictures based on multifeature classifiers," *IEEE Trans. Image Process.*, vol. 20, no. 1, pp. 64–75, Jan. 2011.

[9] S. Gabarda and G. Cristóbal, "Blind image quality assessment through anisotropy," *J. Opt. Soc. Amer. A, Opt. Image Sci. Vis.*, vol. 24, no. 12, pp. B42–B51, Dec. 2007.

[10] W. Lu, K. Zeng, D. Tao, Y. Yuan, and X. Gao, "No-reference image quality assessment in contourlet domain," *Neurocomputing*, vol. 73, nos. 4–6, pp. 784–794, Jan. 2010.

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

[12] F. Gao, X. Gao, W. Lu, D. Tao, and X. Li, "An image quality assessment metric with no reference using hidden Markov tree model," *Proc. SPIE*, vol. 7744, pp. 774410-1–774410-7, Jul. 2010.

[13] M. A. Saad, A. C. Bovik, and C. Charrier, "A DCT statistics-based blind image quality index," *IEEE Signal Process. Lett.*, vol. 17, no. 6, pp. 583–586, Jun. 2010.

[14] M. A. Saad, A. C. Bovik, and C. Charrier, "DCT statistics model-based blind image quality assessment," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2011, pp. 3093–3096.

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

[16] A. K. Moorthy and A. C. Bovik, "A two-step framework for constructing blind image quality indices," *IEEE Signal Process. Lett.*, vol. 17, no. 5, pp. 513–516, May 2010.

[17] 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.

[18] 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.

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

[20] K. Okarma, "Colour image quality assessment using structural similarity index and singular value decomposition," in *Computer Vision and Graphics* (Lecture Notes in Computer Science). New York, NY, USA: Springer-Verlag, 2009, pp. 55–65.

[21] 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.

[22] S. Wang, D. Zheng, J. Zhao, W. J. Tam, and F. Speranza, "An image quality evaluation method based on digital watermarking," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 1, pp. 98–105, Jan. 2007.

[23] X. Gao, W. Lu, D. Tao, and X. Li, "Image quality assessment based on multiscale geometric analysis," *IEEE Trans. Image Process.*, vol. 18, no. 7, pp. 1409–1423, Jul. 2009.

[24] D. Tao, X. Li, W. Lu, X. Gao, "Reduced-reference IQA in contourlet domain," *IEEE Trans. Syst., Man, Cybern., Part B, Cybern.*, vol. 39, no. 6, pp. 1623–1627, Dec. 2009.

[25] A. Srivastava, A. Lee, E. Simoncelli, and S. Zhu, "On advances in statistical modeling of natural images," *J. Math. Imaging Vis.*, vol. 18, no. 1, pp. 17–33, Jan. 2003.

[26] M. J. Wainwright and E. P. Simoncelli, "Scale mixtures of Gaussians and the statistics of natural images," *Adv. Neural Inf. Process. Syst.*, vol. 12, no. 1, pp. 855–861, 2000.

[27] K. Sharifi and A. Leon-Garcia, "Estimation of shape parameter for generalized Gaussian distributions in subband decompositions of video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 5, no. 1, pp. 52–56, Feb. 1995.

[28] H. A. Chipman, E. D. Kolaczyk, and R. E. McCulloch, "Adaptive Bayesian wavelet shrinkage," *J. Amer. Stat. Assoc.*, vol. 92, no. 440, pp. 1413–1421, Dec. 1997.

[29] R. M. Dufour and E. L. Miller, "Statistical signal restoration with $1/f$ wavelet domain prior models," *Signal Process.*, vol. 78, no. 3, pp. 209–307, Nov. 1998.

[30] H. Yang, Z. Xu, J. Ye, I. King, and M. R. Lyu, "Efficient sparse generalized multiple kernel learning," *IEEE Trans. Neural Netw.*, vol. 20, no. 5, pp. 827–839, May 2009.

[31] D. J. Field, "Relationship between the statistics of natural images and the response properties of cortical cells," *J. Opt. Soc. Amer.*, vol. 4, no. 12, pp. 2379–2394, Dec. 1987.

[32] Y. Petrov and L. Zhaoping, "Local correlations, information redundancy, and sufficient pixel depth in natural images," *J. Opt. Soc. Amer. A, Opt., Image Sci. Vis.*, vol. 20, no. 1, pp. 56–66, Jan. 2003.

[33] J. K. Romberg, H. Choi, and R. G. Baraniuk, "Bayesian tree structured image modeling using wavelet domain hidden Markov models," *IEEE Trans. Image Process.*, vol. 10, no. 7, pp. 1056–1068, Jul. 2001.

[34] E. P. Simoncelli and E. H. Adelson, "Noise removal via Bayesian wavelet coring," in *Proc. 3rd Int. Conf. Image Process.*, Sep. 1996, pp. 379–383.

[35] R. L. Joshi and T. R. Fischer, "Image subband coding using arithmetic and trellis coded quantization," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 5, no. 6, pp. 515–523, Dec. 1995.

[36] M. Vasconcelos and N. Vasconcelos, "Natural image statistics and low complexity feature selection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 228–244, Feb. 2009.

[37] H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik. (2003). *LIVE Image Quality Assessment Database* [Online]. Available: http://live.ece.utexas. edu/research/quality

[38] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Trans. Image Process.*, vol. 15, no. 11, pp. 3440–3451, Nov. 2006.

[39] L. Paninski, "Estimation of entropy and mutual information," *Neural Comput.*, vol. 15, no. 6, pp. 1191–1253, Jun. 2003.

[40] K.-R. Müller, S. Mika, G. Rätsch, K. Tsuda, and B. Schölkopf, "An introduction to kernel-based learning algorithms," *IEEE Trans. Neural Netw.*, vol. 12, no. 2, pp. 181–201, Mar. 2001.

[41] P. Bouboulis, K. Slavakis, and S. Theodoridis, "Adaptive learning in complex reproducing kernel Hilbert spaces employing Wirtinger's subgradients," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 23, no. 3, pp. 425–438, Mar. 2012.

[42] S. Liwicki, S. Zafeiriou, G. Tzimiropoulos, and M. Pantic, "Efficient online subspace learning with an indefinite kernel for visual tracking and recognition," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 23, no. 10, pp. 1624–1636, Oct. 2012.
[43] G. Lanckriet, T. De Bie, N. Cristianini, M. Jordan, and W. Noble, "A statistical framework for genomic data fusion," *Bioinformatics*, vol. 20, no. 16, pp. 2626–2635, 2004.
[44] A. Rakotomamonjy, F. R. Bach, S. Canu, and Y. Grandvalet, "SimpleMKL," *J. Mach. Learn. Res.*, vol. 9, pp. 2491–2521, Sep. 2008.
[45] N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, M. Carli, and F. Battisti, "TID2008—A database for evaluation of full-reference visual quality assessment metrics," *Adv. Modern Radioelectron.*, vol. 10, no. 4, pp. 30–45, 2009.
[46] G. Lanckriet, N. Cristianini, L. ElGhaoui, P. Bartlett, and M. Jordan, "Learning the kernel matrix with semi-definite programming," *J. Mach. Learn. Res.*, vol. 5, pp. 27–72, Jan. 2004.
[47] G. Fung, M. Dundar, J. Bi, and B. Rao, "A fast iterative algorithm for Fisher discriminant using heterogeneous kernels," in *Proc. 21th Int. Conf. Mach. Learn.*, 2004, p. 40.
[48] S.-J. Kim, A. Magnani, and S. Boyd, "Optimal kernel selection in kernel Fisher discriminant analysis," in *Proc. 23th ICML*, 2006, pp. 465–472.
[49] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 27, pp. 1–27, 2011.
[50] (2000, Jun.). *Final Report From the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment* [Online]. Available: http://www.vqeg.org/
[51] T. J. Liu, W. Lin, and C.-C. J. Kuo, "Image quality assessment using multi-method fusion," *IEEE Trans. Image Process.*, vol. 22, no. 5, pp. 1793–807, May 2013.
[52] N. Wang, J. Li, D. Tao, X. Li, and X. Gao, "Heterogeneous image transformation," *Pattern Recognit. Lett.*, vol. 34, no. 1, pp. 77–84, Jan. 2013.
[53] X. Gao, N. Wang, D. Tao, and X. Li, "Face Sketch–Photo synthesis and retrieval using sparse representation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 8, pp. 1213–1226, Aug. 2012.
[54] X. Gao, K. Zhang, D. Tao, and X. Li, "Joint learning for single-image super-resolution via a coupled constraint," *IEEE Trans. Image Process.*, vol. 21, no. 2, pp. 469–480, Feb. 2012.

Xinbo Gao (M'02–SM'07) received the B.Eng., M.Sc., and Ph.D. degrees in signal and information processing from Xidian University, Xi'an, China, in 1994, 1997, and 1999, respectively.

He was a Research Fellow with the Department of Computer Science, Shizuoka University, Shizuoka, Japan, from 1997 to 1998. From 2000 to 2001, he was a Post-Doctoral Research Fellow with the Department of Information Engineering, Chinese University of Hong Kong, Shatin, Hong Kong. Since 2001, he has been with the School of Electronic Engineering, Xidian University. Currently, he is a Professor of pattern recognition and intelligent systems and the Director of the VIPS Laboratory, Xidian University. His current research interests include computational intelligence, machine learning, computer vision, pattern recognition, and wireless communications. He has published five books and over 150 technical articles in refereed journals and proceedings, including the IEEE T-IP, T-CSVT, T-NN, T-SMC, *Pattern Recognition*.

Dr. Gao is a fellow of IET. He is on the editorial boards of journals, including the *EURASIP Signal Processing* (Elsevier) and *Neurocomputing* (Elsevier). He served as a General Chair and Co-Chair or Program Committee Chair/Co-Chair or a PC Member for over 30 major international conferences.

Fei Gao received the B.Eng. degree in electronic information engineering from Xidian University, Xi'an, China, in 2009. He is currently pursuing the Ph.D. degree in intelligent information processing with the VIPS Laboratory, School of Electronic Engineering, Xidian University.

He has been a Visiting Ph.D. Student with the University of Technology, Sydney, Australia, since October 2012. His current research interests include computer vision especially for image quality assessment and its applications.

Dacheng Tao (M'07–SM'12) received the B.Eng. degree from the University of Science and Technology of China, Hefei, China, the M.Phil. degree from the Chinese University of Hong Kong, Shatin, Hong Kong, and the Ph.D. degree from the University of London, London, U.K.

He is currently a Professor with the Centre for Quantum Computation and Information Systems, Faculty of Engineering and Information Technology, University of Technology, Sydney, Australia. He mainly applies statistics and mathematics to data analysis problems in data mining, computer vision, machine learning, multimedia, and video surveillance. He has authored or co-authored more than 100 scientific articles at top venues, including the IEEE T-PAMI, T-KDE, T-NN, T-IP, NIPS, ICML, AISTATS, ICDM, IJCAI, AAAI, CVPR, ECCV, ACM T-KDD, and KDD.

Dr. Tao has received many Best Paper Awards.

Xuelong Li (M'02–SM'07–F'12) is a Full Professor with the Center for OPTical IMagery Analysis and Learning, State Key Laboratory of Transient Optics and Photonics, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an, China.