This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2019.2900472, IEEE Transactions on Circuits and Systems for Video Technology



# Unsupervised Blind Image Quality Evaluation via Statistical Measurements of Structure, Naturalness and Perception

Yutao Liu, Ke Gu, Yongbing Zhang, *Member, IEEE,* Xiu Li, *Member, IEEE,*
Guangtao Zhai, *Member, IEEE,* Debin Zhao, *Member, IEEE,* and Wen Gao, *Fellow, IEEE*

***Abstract*—Most of the existing blind image quality assessment (BIQA) methods belong to supervised methods, which always need a large number of image samples and expensive subjective scores for training a quality prediction model. In this paper, we focus our attention on the unsupervised BIQA methods and put forward a novel unsupervised approach. The main idea of our method is to quantify the image quality degradation through measuring the structure, naturalness and perception quality variations of the distorted image from the pristine natural images. In specific, the structure variation is captured by the deviations of the image phase congruency (PC) and gradients distributions. The naturalness variation is characterized through the distributions variations of the locally mean subtracted and contrast normalized (MSCN) coefficients and the products of pairs of the adjacent MSCN coefficients. Compared with existing unsupervised methods, we initiatively introduce the perception quality measurement into the construction of unsupervised BIQA method, which is conducted by characterizing the prediction discrepancy between the image and its brain prediction based on the free-energy principle in the newly revealed brain theory. After feature extraction, we learn a pristine multivariate Gaussian (MVG) model with the extracted features from a set of pristine natural images. The quality of a new image is finally defined as the distance between its MVG model and the learned pristine MVG model. Extensive experiments conducted on LIVE, TID2013, CSIQ, Toyama, CID2013 and the Waterloo Exploration databases demonstrate that the proposed method achieves comparative prediction performance with state-of-the-art BIQA methods.**

***Index Terms*—Blind image quality assessment (BIQA), natural scene statistics (NSS), free-energy principle.**

Manuscript received xxxx, 2018; revised xxxx, 2018; accepted xxxx, 2019. This work was supported in part by the Major State Basic Research Development Program of China (2015CB351804), the National Science Foundation of China (61703009 & 61872116 & 41876098 & 61571254), Shenzhen Science and Technology Project (JCYJ20151117173236192 & JCYJ20170817161409809), the Young Elite Scientist Sponsorship Program by China Association for Science and Technology (2017QNRC001), the Young Top-Notch Talents Team Program of Beijing Excellent Talents Funding (Grant 2017000026833ZK40). (Corresponding author: Ke Gu.)

Y. Liu, Y. Zhang and X. Li are with Graduate School at Shenzhen, Tsinghua University, Shenzhen 518055, China (e-mail: ytliu18@sz.tsinghua.edu.cn; zhang.yongbing@sz.tsinghua.edu.cn; li.xiu@sz.tsinghua.edu.cn).

K. Gu is with the Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China (e-mail: guke@bjut.edu.cn).

G. Zhai is with the Institute of Image Communication and Network Engineering, Shanghai Jiao Tong University, Shanghai 200240, China. (email: zhaiguangtao@sjtu.edu.cn).

D. Zhao is with the School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China (e-mail: dbzhao@hit.edu.cn).

W. Gao is with the School of Electrical Engineering and Computer Science, Peking University, Beijing 100871, China (e-mail: wgao@pku.edu.cn).

## I. Introduction

OBJECTIVE image quality assessment (IQA) aims to predict the image quality in consistency with subjective perception of the image quality. Depending on the accessibility of the original image, existing objective IQA methods can be classified into three categories, which are full-reference (FR) methods [1] [2] [3], reduced-reference (RR) methods [4] [5] [6] and no-reference (NR) / blind methods [7] [8] [9] [10]. In this paper, we focus our attention on the research of NR IQA methods.

NR IQA or BIQA refers to evaluating the image quality in the absence of the original image, which is highly desirable for practical use. Early NR methods are mainly distortion-specific with assuming that the image is degraded by some specific distortions, such as blur [11] [12], noise [13], JPEG compression [14], or contrast change [15], etc. Since the distortion type is known, researchers are able to define targeted features to quantify the distortions precisely so that they can evaluate the image quality desirably. However, the application scope of the distortion-specific NR methods is rather limited. Therefore, general-purpose NR methods are proposed to evaluate the image quality without restricting the distortion types in advance. Generally speaking, general-purpose NR methods can be classified into supervised methods and unsupervised methods. Supervised methods always require the subjective scores which serve as the ground truth to train the quality prediction model. The main difference of existing supervised NR methods lies in the features that they designed to quantify the image quality degradation. In [8], Moorthy et al. extracted statistical features in the wavelet domain and utilized support vector regression (SVR) to map the features onto the image quality level. In [7], Mittal et al. exploited natural scene statistics (NSS) features to characterize the image quality effectively. In [16], Wu et al. designed the joint statistics of multiple domains, such as DCT, wavelet, etc. and deduced the image quality through label transfer. In [17], Liu et al. designed the low-level and high-level statistical features to characterize the image distortions and then resorted to a neural network to map all the extracted features onto the image quality. As convolutional neural network (CNN)-based deep learning technologies achieve great success in computer vision tasks, IQA researchers have also introduced them into the design of IQA algorithms [18] [19] [20]. With elaborately-designed neural network, these methods can learn

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2019.2900472, IEEE Transactions on Circuits and Systems for Video Technology



the quality-aware features and the quality prediction model together, which saves much labor to design the features for quality evaluation.

Although the above supervised BIQA methods can achieve high prediction performance, they require a large number of image samples and the expensive subjective scores for calibrating the quality prediction module. In addition, the supervised methods may also suffer from weak generalization capability [21]. On the contrary, unsupervised NR methods don't need subjective scores during their construction process and can reveal better generalization capability. In [22], Xue et al. took the strategy of substituting the values given by the FR method FSIM for the subjective scores for learning a set of centroids and proposed a quality-aware clustering (QAC) method. Such strategy was also used in [23] and [24] for quality evaluation of the screen content images and enhanced images. However, these methods are heavily restricted by the FR methods they adopted. In [9], Mittal et al. proposed the natural image quality evaluator (NIQE), in which they fitted the locally mean subtracted and contrast normalized (MSCN) coefficients and the products of pairs of adjacent MSCN coefficients with generalized Gaussian distribution (GGD) and asymmetric generalized Gaussian distribution (AGGD) respectively to extract the quality-aware features. Then they fitted the extracted features to a multivariate Gaussian (MVG) model with a set of pristine images. The quality of a distorted image was defined as the distance between its MVG model and the pristine MVG model. In [25], Wu et al. proposed a highly efficient method, named local pattern statistics index (LPSI), in which the statistical features from binary patterns of the local image structures were extracted for quality estimation. In fact, NIQE or LPSI evaluates the image quality from only one aspect of the image, namely naturalness or structure, which is not sufficient to describe the image quality comprehensively. In [21], Zhang et al. extended NIQE to integrated local NIQE (IL-NIQE) by introducing three additional types of statistical features of gradient, Log-Gabor filter response and color into the framework and evaluating the image quality in a local manner. However, the dimension of the feature vector in IL-NIQE is very high despite applying PCA for reducing the feature vector dimension.

In this paper, we aim to introduce a novel unsupervised NR method to evaluate the image quality from a more comprehensive perspective. The proposed method works by measuring the variations of three aspects of the distorted image from the pristine images, which are structure, naturalness and the perception quality respectively. First of all, as indicated in [1], the human visual system (HVS) is highly adapted for extracting the structure information from the visual scenes. If the structure in an image is destroyed by the external distortions, the image quality will be degraded definitely. Therefore, the structure degradation degree can well reflect the quality degradation degree. Based on this concern, we perform image quality evaluation by capturing the structure variation at first. Second, there may exist some distortions that can't be well captured by the structures, e.g., the common JPEG compression. Excessive JPEG compression will largely eliminate the structures in the image, but produce unnatural blocking artifacts in the image. Therefore, we introduce the second image characteristic of naturalness for quality prediction as naturalness is an important attribute of pristine images, which is often affected by distortions. Third, real photographic images are possibly degraded by complex distortions, on which most BIQA methods employing low-level features, e.g. structure or naturalness features, still can't deliver desirable results as observed in [26]. Considering this, we attempt to measure the image quality from a higher perspective which is to gauge the perception quality variation of the distorted image from the undistorted images. Then the question turns into how to characterize the perception quality effectively. Fortunately, the newly proposed free-energy principle in brain theory and neuroscience [27] [28] offers us an answer along with a feasible way for characterizing the human perception quality. As the NSS features are quite effective in capturing distortions [7] [8] [9], we extract a set of NSS features to characterize the structure, naturalness and perception quality respectively. Among them, the structure of the image is characterized by the parameters from fitting the image phase congruency (PC) and the image gradients. The naturalness of the image is characterized by the parameters that depict the distributions of the MSCN coefficients and the adjacent MSCN coefficients products. The perception quality related features are extracted from modeling the prediction discrepancy between the image and its brain predicted version generated by sparse representation. After feature extraction, we fit the NSS features to a pristine MVG model with a set of pristine images as NIQE did. For a new given image, the distance between its MVG model and the pristine MVG model that measures the structure, naturalness and perception quality variations from the pristine images is defined to evaluate the image quality. As our method follows the design philosophy of NIQE, we name it Structure, Naturalness and Perception quality-driven NIQE or SNP-NIQE. It is worthy to mention that the novelty of SNP-NIQE over the proposed method in [17] mainly lies in four aspects. First, SNP-NIQE belongs to unsupervised BIQA methods as analyzed above, while the method in [17] still belongs to supervised methods which need subjective scores for training the quality prediction model. Second, in SNP-NIQE, we quantify the image quality from a more comprehensive perspective which characterizes the structure, naturalness and perception quality variations, while in [17], it can be deemed that only naturalness and perception quality were considered for quality evaluation. Third, in the manner of feature extraction, we take the best-fit parameters of the feature map distributions rather than directly cutting out part of the feature map distributions in [17] as the quality-aware features. At last, in the implementation of the free-energy principle for extracting the perception quality features, here we employ sparse representation strategy which has been verified resembling the perception mechanism [29] [30] rather than the autoregressive model adopted in [17]. Experimental results on six popular image databases, i.e., LIVE, TID2013, CSIQ, Toyama, CID2013 and Waterloo Exploration database, demonstrate that SNP-NIQE delivers comparative prediction performance with state-of-the-art BIQA methods.

Our contributions of this paper can be summarized as

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2019.2900472, IEEE Transactions on Circuits and Systems for Video Technology



follows. First, we establish a novel unsupervised BIQA method through the structure, naturalness and perception quality measurements, which is more comprehensive than the existing unsupervised methods. Second, to the best of our knowledge, we are the first to introduce perception quality measurement into the construction of the unsupervised BIQA model based on the free-energy principle and sparse representation. Third, our method outperforms state-of-the-art unsupervised BIQA methods and competes with classical supervised BIQA methods.

## II. The Proposed Method for BIQA

### A. Structure Statistical Modeling

In our proposed method, we first utilize the structure information to speculate the image quality. Here we employ two features, namely the image phase congruency (PC) and the image gradients, to extract the structure information of the image. These two features play complementary roles in structure extraction as PC can't reflect the contrast or luminance variation which can be captured by the gradients [2] [11].

At first, the PC theory assumes that the structure features are extracted at those points whose Fourier components are maximal in phase [31] [32]. According to the physiological and psychophysical studies, the PC model offers a biological method that describes how mammalian visual systems extract structure features from an image [33] [34]. In this paper, we adopt the method proposed by Kovesi [34] to compute the PC map of the image.

With a 1-D signal $\mathbf{s}$, suppose $F_n^e$ and $F_n^o$ being the even- and odd-symmetric filters on scales $n$, which form a quadrature pair. Responses of each quadrature pair to $\mathbf{s}$ at position $j$ on scale $n$ can be denoted by: $[e_n(j), o_n(j)] = [s(j) * F_n^e, s(j) * F_n^o]$, the local amplitude on scale $n$ is $A_n(j) = \sqrt{e_n(j)^2 + o_n(j)^2}$. Let $E(j) = \sum_n e_n(j)$ and $O(j) = \sum_n o_n(j)$, the PC of the signal $\mathbf{s}$ can be calculated by:

$$PC(j) = \frac{U(j)}{\varepsilon + \sum_n A_n(j)} \tag{1}$$

where $U(j) = \sqrt{E^2(j) + O^2(j)}$, $\varepsilon$ is a small positive constant to keep stability. Excluding the spurious effect of noise for PC computation, the above equation can be written as:

$$PC(j) = \frac{(U(j) - T)^+}{\varepsilon + \sum_n A_n(j)} \tag{2}$$

where $T$ refers to the total noise effect, $(\cdot)^+$ represents the operation of $max(0, x)$ which guarantees the nonnegativity of the numerator. With 1-D PC definition, 2-D PC can be calculated by integrating 1-D PC of all directions:

$$PC_{2D}(\mathbf{j}) = \frac{\sum_o (U_o(\mathbf{j}) - T_o)^+}{\varepsilon + \sum_o \sum_n A_{no}(\mathbf{j})} \tag{3}$$

with $o$ representing the index of directions. Usually, a sigmoid function is introduced into 2-D PC computation to weight the PC value of each direction, namely:

$$PC_{2D}(\mathbf{j}) = \frac{\sum_o (W_o(\mathbf{j})(U_o(\mathbf{j}) - T_o)^+)}{\varepsilon + \sum_o \sum_n A_{no}(\mathbf{j})} \tag{4}$$

where $W(\mathbf{j})$ is defined as follows:

$$W(\mathbf{j}) = \frac{1}{1 + e^{g(c - l(\mathbf{j}))}} \tag{5}$$

where $c$ refers to the 'cut-off' value of the filter response spread. PC values below $c$ are to be penalized, $g$ denotes the gain factor and controls the cut-off sharpness. $l(\mathbf{j})$ denotes the spread function defined as:

$$l(\mathbf{j}) = \frac{1}{N} \frac{\sum_n A_n(\mathbf{j})}{\varepsilon + A_{max}(\mathbf{j})} \tag{6}$$

where $N$ gives the total number of scales, $A_{max}(\mathbf{j})$ refers to the maximum response amplitude at position $\mathbf{j}$.

To visually show the capability of PC in capturing the image distortions, we select three pristine images with their distorted versions from TID2013 database [35]. The investigated distortions here are JPEG compression, blur and noise. The distortion levels of the distorted versions are at the fourth degree defined in TID2013. We compute their PC maps respectively and show the results in Fig.1, in which (a)(b)(c) are the pristine images, (d)(e)(f) show the PC distributions of the pristine images and their distorted versions of (a)(b)(c), respectively. Take (d) for example which shows the PC distributions of (a) and its distorted images, it can be observed that the PC distribution is very sensitive to noise and blur as their distributions are both deviated from the original one of the pristine image. However, it is also observed that the PC distribution can't reflect the JPEG compression well as the PC distributions of the pristine image and the JPEG compressed image are very close.

To describe the PC distribution of the image, we fit it with the Weibull distribution as:

$$f(x; \lambda, k) = \begin{cases} \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} \exp\left(-\left(\frac{x}{\lambda}\right)^k\right) x \geq 0 \\ 0 \qquad\qquad\qquad\qquad\qquad\quad x < 0 \end{cases} \tag{7}$$

where $\lambda > 0$ represents the scale parameter, $k > 0$ refers to the shape parameter. $\lambda$ and $k$ that depict the PC distribution are employed as our quality-aware features.

As PC of the image is not sensitive to contrast or luminance variation, which also affects the image quality, we extract the gradients of the image for describing the structure complementarily. Here, we extracts the gradients by convolving the image with the simple but widely-used high-passing filters in image processing, which are the vertical and horizontal finite difference operators, denoted by $\mathbf{D}_v = [1, -1]^T$ and $\mathbf{D}_h = [1, -1]$, written as:

$$\mathbf{G}_v = I * \mathbf{D}_v \tag{8}$$

and

$$\mathbf{G}_h = I * \mathbf{D}_h \tag{9}$$

where $I$ refers to the input image, "$*$" refers to the convolution operation, $\mathbf{G}_v$ and $\mathbf{G}_h$ are the obtained vertical and horizontal gradients of image $I$, respectively.

The gradients distribution of the pristine image is also changed due to the external distortions. In Fig.1 (g)(h)(i), we can also find that the gradients $\mathbf{G}_v$ distributions of the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2019.2900472, IEEE Transactions on Circuits and Systems for Video Technology



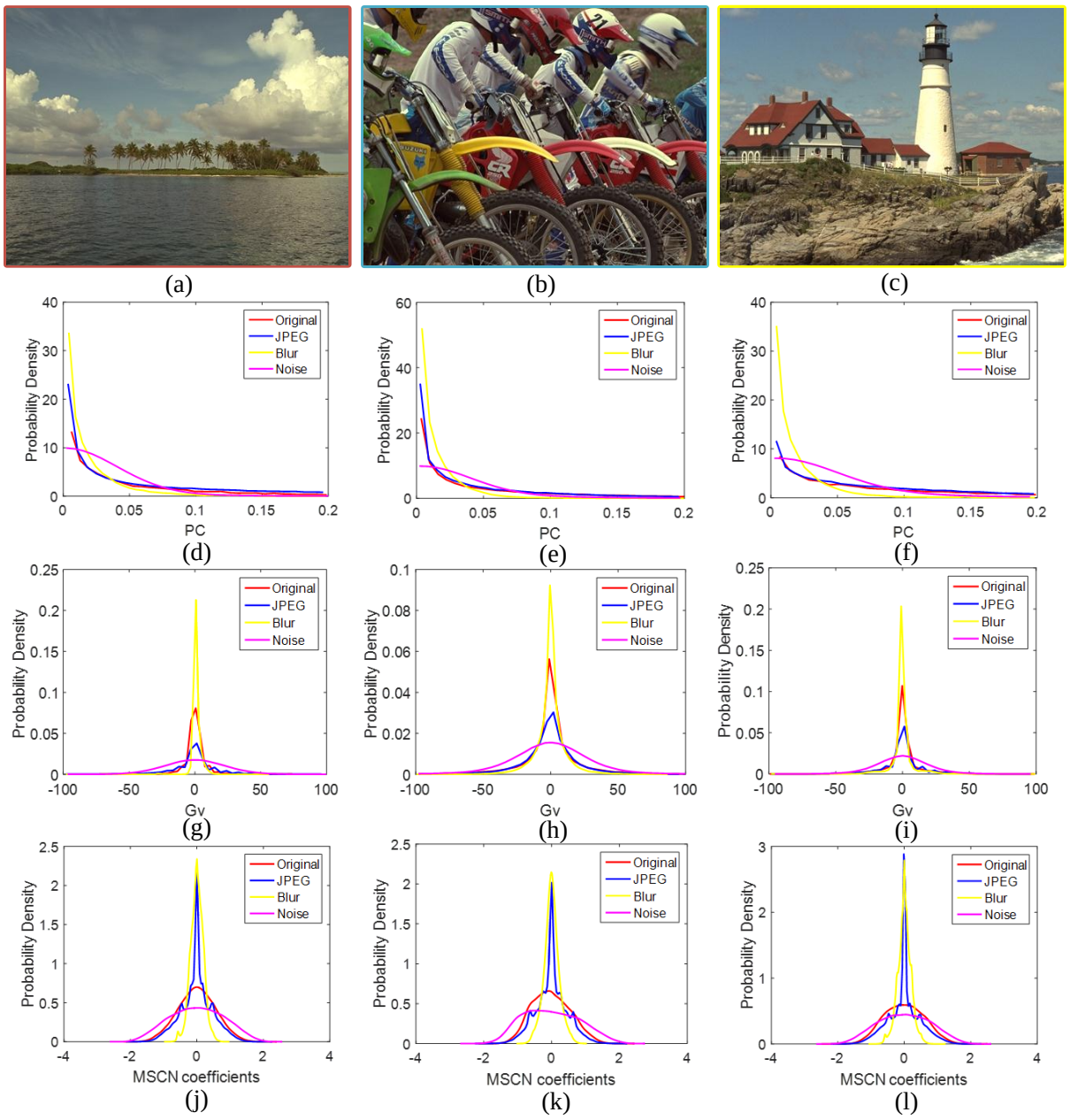



Fig. 1. An illustration of PC, $\mathbf{G}_v$ and MSCN statistics variations with regard to different kinds of distortions, i.e. JPEG, blur and noise. (a)(b)(c) are the three example pristine images; (d), (g) and (j) show the PC, $\mathbf{G}_v$ and MSCN distributions of (a) and its distorted images; (e), (h) and (k) show the PC, $\mathbf{G}_v$ and MSCN distributions of (b) and its distorted images; (f), (i) and (l) show the PC, $\mathbf{G}_v$ and MSCN distributions of (c) and its distorted images;. Note that in each plot figure, the red line represents the distribution of the pristine image, the blue line represents the distribution of the JPEG compressed image, the yellow line represents the distribution of the blurry image, the magenta line represents the distribution of the noisy image. Also note that the vertical coordinates that exhibit the PC, $\mathbf{G}_v$ and MSCN distributions of different images are also different, e.g. (d) and (e).

three pristine images are all changed due to the presence of distortions to some extent. In addition, as observed in this figure, the gradients distribution can be accurately fitted with the zero-mean GGD [36]. Taking $\mathbf{G}_v$ as example, the distribution of $\mathbf{G}_v$ can be nicely fitted with:

$$g(x;\alpha,\beta) = \frac{\alpha}{2\beta\Gamma(1/\alpha)} \exp\left(-\left(\frac{|x|}{\beta}\right)^{\alpha}\right) \tag{10}$$

where $\Gamma(\cdot)$ is the gamma function, defined as:

$$\Gamma(x) = \int_0^{\infty} \phi^{x-1} e^{-\phi} d\phi, x > 0 \tag{11}$$

where $\alpha$ is the shape parameter, $\beta$ is the standard deviation. Likewise, $\alpha$ and $\beta$ which can be estimated with moment matching-based method [37] are added to our quality-aware feature set. Correspondingly, the parameters of $\mathbf{G}_h$ are extracted and employed as our features in the same way.

### B. Naturalness Statistical Modeling

In literature, naturalness of the image is often measured through modeling the locally mean subtracted and contrast normalized (MSCN) coefficients and the products of pairs of

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2019.2900472, IEEE Transactions on Circuits and Systems for Video Technology



adjacent MSCN coefficients, the obtained statistical features are utilized for naturalness measurement [7] [9]. Given an image $I$, its MSCN coefficients can be calculated by

$$\widehat{I}(x,y) = \frac{I(x,y) - \mu(x,y)}{\sigma(x,y) + 1} \tag{12}$$

where $\widehat{I}(x,y)$ and $I(x,y)$ represent the MSCN coefficients image and original image values at position $(x,y)$ respectively. $\mu(x,y)$ and $\sigma(x,y)$ stands for the local mean and standard deviation in a local patch centered at $(x,y)$. $\mu(x,y)$ and $\sigma(x,y)$ are respectively calculated as:

$$\mu(x,y) = \sum_{s=-S}^{S}\sum_{t=-T}^{T} \omega_{s,t} I(x+s, y+t) \tag{13}$$

$$\sigma(x,y) = \sqrt{\sum_{s=-S}^{S}\sum_{t=-T}^{T} \omega_{s,t}[I(x+s,y+t) - \mu(x,y)]^2} \tag{14}$$

where $\omega = \{\omega_{s,t} \mid s = -S, ..., S; t = -T, ..., T\}$ denotes a 2D circularly-symmetric Gaussian weighting filter.

Similarly, in Fig.1 (j)(k)(l), it's easily seen that the MSCN coefficients distributions are very indicative when the images suffer from distortions. It's worthy to note that JPEG compression can be effectively captured by the MSCN coefficients distribution. As suggested in [7] [9], the distribution of the MSCN coefficients is modeled through the zero-mean GGD and the obtained distribution parameters are employed as the features to indicate the naturalness degree.

The products of pairs of adjacent MSCN coefficients along four orientations, which are horizontal, vertical, main-diagonal and second-diagonal, namely the products of $\widehat{I}(x,y)\widehat{I}(x,y+1)$, $\widehat{I}(x,y)\widehat{I}(x+1,y)$, $\widehat{I}(x,y)\widehat{I}(x+1,y+1)$ and $\widehat{I}(x,y)\widehat{I}(x+1,y-1)$ are calculated respectively and modeled as following a zero mode asymmetric GGD (AGGD) as follows:

$$g(x;\gamma,\beta_l,\beta_r) = \begin{cases} \dfrac{\gamma}{(\beta_l+\beta_r)\Gamma(\frac{1}{\gamma})} \exp\left(-\left(\dfrac{-x}{\beta_l}\right)^{\gamma}\right) \forall x \le 0 \\ \dfrac{\gamma}{(\beta_l+\beta_r)\Gamma(\frac{1}{\gamma})} \exp\left(-\left(\dfrac{x}{\beta_r}\right)^{\gamma}\right) \forall x > 0 \end{cases} \tag{15}$$

The mean of this distribution is

$$\eta = (\beta_r - \beta_l)\frac{\Gamma(\frac{2}{\gamma})}{\Gamma(\frac{1}{\gamma})} \tag{16}$$

the model paramters $(\gamma, \beta_l, \beta_r, \eta)$ are also introduced into the quality-aware feature vector.

### C. The Perception Quality Statistical Modeling

As the image quality is yielded through the human visual perception process, we expect to quantify the image quality by characterizing the quality of human perception. In brain theory and neuroscience, the newly proposed free-energy principle offers us an executable strategy to achieve that goal. Specifically, the free-energy principle reveals the perception and understanding for the visual scenes are manipulated by an internal generative model in the brain, with which the brain generates predictions for the visual scenes in a constructive manner [38] [39] [4]. The constructive model is essentially a probabilistic model which can be decomposed into a likelihood term and a prior term. Visual perception is modeled to infer the posterior possibilities of the visual scene through inverting the likelihood term. As the internal model can't be universal, it's reasonable to suppose there exists a discrepancy between the visual scene and its brain prediction, which is believed to be very closely related to the quality of human perception [4].

Specifically, for modeling visual perception, the brain generative model that controls the visual perception process is often assumed as parametric, which explains the visual input by adjusting its parameters, denoted by $\mathcal{M}$. For convenience, we utilize $\mathbf{m}$ as the vector which is composed of the parameters in $\mathcal{M}$. Then, the 'surprise' of an input image $I$ can be calculated by integrating the joint distribution $P(I,\mathbf{m})$ over the parameter space of $\mathbf{m}$ as:

$$-\log P(I) = -\log \int P(I,\mathbf{m}) d\mathbf{m} \tag{17}$$

Here, we introduce an assistant item $\widetilde{P}(\mathbf{m}|I)$ into the right component of the above equation while still maintain its equivalence as:

$$-\log P(I) = -\log \int \widetilde{P}(\mathbf{m}|I)\frac{P(I,\mathbf{m})}{\widetilde{P}(\mathbf{m}|I)} d\mathbf{m} \tag{18}$$

here $\widetilde{P}(\mathbf{m}|I)$ can be thought of as the posterior distribution of image $I$, which is the approximate posterior distribution to the true posterior distribution $P(\mathbf{m}|I)$ when exposed to $I$. For explaining $I$, the human brain tries to minimize the divergence between the approximate $\widetilde{P}(\mathbf{m}|I)$ and the true $P(\mathbf{m}|I)$. Based on Jensen's inequality, the above equation can be written as:

$$-\log P(I) \le -\int \widetilde{P}(\mathbf{m}|I)\log\frac{P(I,\mathbf{m})}{\widetilde{P}(\mathbf{m}|I)} d\mathbf{m} \tag{19}$$

On the basis of statistical physics and thermodynamics [40], the right part of the above equation is defined as "free energy", namely:

$$F(\mathbf{m}) = -\int \widetilde{P}(\mathbf{m}|I)\log\frac{P(I,\mathbf{m})}{\widetilde{P}(\mathbf{m}|I)} d\mathbf{m} \tag{20}$$

Obviously, $F(\mathbf{m})$ denotes an upper bound of $I$'s 'surprise'. We can explain this through further derivation. As $P(I,\mathbf{m}) = P(\mathbf{m}|I)P(I)$, equation (20) can be rewritten as:

$$\begin{aligned} F(\mathbf{m}) &= \int \widetilde{P}(\mathbf{m}|I)\log\frac{\widetilde{P}(\mathbf{m}|I)}{P(\mathbf{m}|I)P(I)} d\mathbf{m} \\ &= -\log P(I) + \int \widetilde{P}(\mathbf{m}|I)\log\frac{\widetilde{P}(\mathbf{m}|I)}{P(\mathbf{m}|I)} d\mathbf{m} \\ &= -\log P(I) + \mathbf{KL}(\widetilde{P}(\mathbf{m}|I)\|P(\mathbf{m}|I)) \end{aligned} \tag{21}$$

where $\mathbf{KL}(\cdot)$ denotes the Kullback-Leibler divergence, which is nonnegative. It is easily found that the free energy of $F(\mathbf{m})$ is greater than or equal to -log$P(I)$ which accounts for the 'surprise'. When perceiving image $I$, the human brain intends to minimize $\mathbf{KL}(\widetilde{P}(\mathbf{m}|I)\|P(\mathbf{m}|I))$ of the divergence between the approximate posterior and its true posterior distributions.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2019.2900472, IEEE Transactions on Circuits and Systems for Video Technology



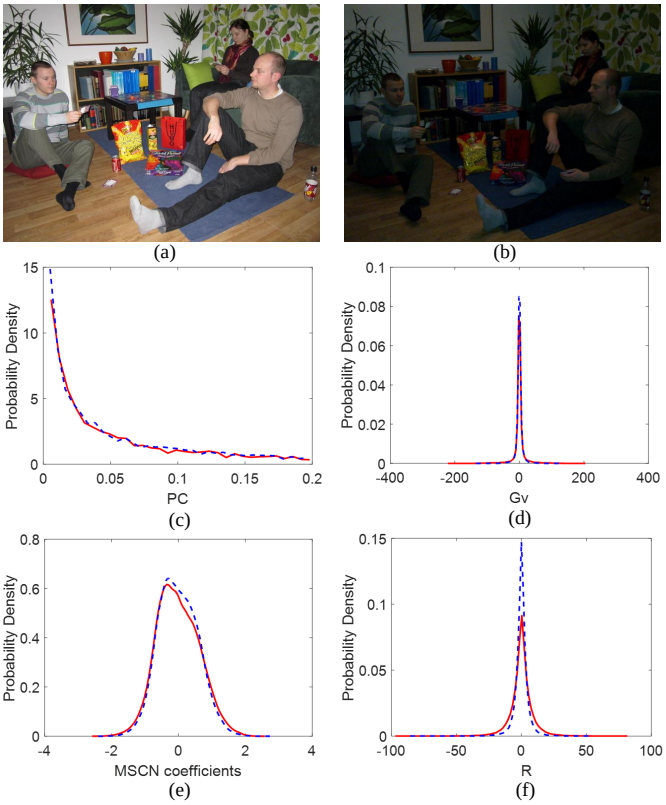



Fig. 2. An illustration of the PC, gradients, MSCN and R distributions of two images with different quality. (a) is the high-quality image (MOS:50.864), (b) is the low-quality image (MOS:24.028), (c) is the PC distribution comparison, (d) is the gradients distribution comparison, (e) is the MSCN coefficients distribution comparison, (f) is the R distribution comparison. The red solid line belongs to the high-quality image, the blue dotted line belongs to the low-quality image.

To employ the free-energy principle to characterize perception quality, we need to understand the internal generative model at first. However, the real form of the internal model is still beyond our knowledge [4]. In the literature, the autoregressive (AR) model is always used to approximate the internal model in the IQA methods [4] [41] [12]. However, in our previous work [5], we have proved sparse representation is more reasonable than AR in approximating the internal model in that sparse representation has been evidenced to resemble the strategy for representing natural images in the visual cortex of the human brain [29] [30]. Readers can refer to [5] for details about this issue. Here, we still resort to sparse representation to approximate the internal generative model. Concretely, given an image $I$, we firstly extract a patch $\mathbf{x}_k \in \mathbb{R}^{B_s}$ from $I$ with its size being $\sqrt{B_s} \times \sqrt{B_s}$ :

$$\mathbf{x}_k = \mathbf{R}_k(I) \tag{22}$$

where $\mathbf{R}_k(\cdot)$ refers to the extraction operation. The sparse representation for $\mathbf{x}_k$ via an over-complete dictionary $\mathbf{D} \in \mathbb{R}^{B_s \times M}$ can be described as:

$$\boldsymbol{\alpha}_k^* = \underset{\boldsymbol{\alpha}_k}{\operatorname{argmin}} \|\boldsymbol{\alpha}_k\|_p \quad s.t. \quad \mathbf{x}_k = \mathbf{D}\boldsymbol{\alpha}_k \tag{23}$$

where $\boldsymbol{\alpha}_k$ is the representation coefficient vector. $\|\cdot\|_p$ refers to the $l^p$ norm. We further transform the above equation to an unconstrained optimization equation:

$$\boldsymbol{\alpha}_k^* = \underset{\boldsymbol{\alpha}_k}{\operatorname{argmin}} \frac{1}{2}\|\mathbf{x}_k - \mathbf{D}\boldsymbol{\alpha}_k\|_2 + \lambda\|\boldsymbol{\alpha}_k\|_p \tag{24}$$

in which the first part refers to the representation fidelity, the second part is the sparsity constraint for the representation vector $\boldsymbol{\alpha}_k$. $\lambda$ is a parameter that balances the importance of the two parts. $p$ takes 0 or 1. In this paper, we set p=0 and employ the OMP method [42] to solve the above optimization equation. Here, the distribution of $\boldsymbol{\alpha}_k^*$ can be regarded as the approximate posterior distribution $\widetilde{P}(\mathbf{m}|I)$. With the obtained representation vector $\boldsymbol{\alpha}_k^*$ for each $\mathbf{x}_k$, the sparse representation for image $I$ can be obtained through:

$$I' = \sum_{k=1}^{n} \mathbf{R}_k^{-1}(\mathbf{D}\boldsymbol{\alpha}_k^*) \odot \sum_{k=1}^{n} \mathbf{R}_k^{-1}(\mathbf{1}_{B_s}) \tag{25}$$

where $I'$ is the sparse representation of $I$, which serves as the brain prediction for $I$, $\mathbf{R}_k^{-1}(\cdot)$ represents the inverse operation of $\mathbf{R}_k(\cdot)$, which is to put the sparse representation $\mathbf{D}\boldsymbol{\alpha}_k^*$ of $\mathbf{x}_k$ back to image $I'$ at the location of $\mathbf{x}_k$ in image $I$, $n$ is the total number of the image patches. $\mathbf{1}_{B_s}$ refers to the vector whose values are all 1 and its size is $B_s$, "$\odot$" refers to the element-wise division operation. As the free-energy principle conjectures the divergence between the image and its brain predicted version can reflect the perception quality, we first define the divergence between $I$ and its brain prediction $I'$ by the prediction residual as:

$$R = I - I' \tag{26}$$

where $R$ represents the divergence between $I$ and $I'$, which is calculated as the subtraction of collocated pixels in $I$ and $I'$. Then $R$ can be used to indicate the perception quality variation. To illustrate this visually, we choose two real photographic images of different quality from the CID2013 database [26] and calculate their prediction residuals respectively, shown in Fig. 2, where (a) is the high-quality image whose MOS value is 50.864, (b) is the low-quality image whose MOS value is 24.028, their prediction residual R distributions are shown in (f), in which the red solid one belongs to the high-quality image and the blue dotted one belongs to the low-quality image. For comparison, we also give the PC, gradients and MSCN distributions in (c), (d), (e). From this figure, we can see that the perception quality variation can't be reflected by the PC, gradients and MSCN coefficients distributions, because the high-quality line and the low-quality line are almost coincident in (c)(d)(e). However, as observed in (f), the R distribution reveals better ability to indicate the perception quality variation of these two images. We fit the prediction residual distribution with the zero-mean GGD defined in Eq.(10), then the distribution parameters are employed as our quality-aware features to characterize the perception quality variation.

### D. Quality-aware Features Summary

Up to now, we have addressed the NSS features that characterize the image structure, naturalness and perception quality separately. For convenience, we make a brief summary

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2019.2900472, IEEE Transactions on Circuits and Systems for Video Technology



TABLE I
SUMMARY OF THE EXTRACTED QUALITY-AWARE FEATURES

| Category | Statistical Features | Dimensionality |
|---|---|---|
| Structure | PC and Gradients | 6 |
| Naturalness | MSCN coefficients | 18 |
| Perception quality | Prediction residual | 2 |

here listed in Table I, where "Dimensionality" means how many scalars are extracted in each category of the features. As observed, the feature No. of these three types of features are 6, 18, 2 respectively. As multi-scale strategy is effective in IQA [43], we extract the statistical features in two scales of the image, which are the original scale and the down-sampled scale by a factor 2. Therefore, the number of the quality-aware features in the proposed scheme reaches $(6+18+2)\times 2 = 52$. Detailed discussion about multi-scale strategy for feature extraction will be given in Section III-I.

### E. Multivariate Gaussian Model

With the designed features, we intend to learn a pristine model that serves as the "reference", compared with which to predict the quality of a new given image. To this end, we select one hundred natural pristine images from the Berkeley image segmentation database [44]. It's noted that the used pristine images are mainly common scenes in reality, which include people, animals, buildings, natural landscapes, etc.. For illustration, we show some sample images in Fig. 3. The entire image set will be uploaded at http://multimedia.sjtu.edu.cn/Data/List/Resources. With the pristine images, we extract the quality-aware features of them in the following successive steps. Firstly, we partition each image into non-overlapped patches of size $96\times 96$ and perform feature extraction on each patch, leading to a 52-dimension vector for each patch. Then we stack all the feature vectors together and fit them with an MVG density as:

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{k/2}|\mathbf{\Sigma}|^{1/2}} \exp\left( -\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right) \quad (27)$$

where $\mathbf{x} = (x_1, x_2, ..., x_{52})$ refers to the feature vector, $k = 52$, which indicates the dimension of the feature vector. The density parameters $\boldsymbol{\mu}, \mathbf{\Sigma}$ that depict the MVG density of the pristine images are employed as the reference information for quality calibration.

### F. The SNP-NIQE Index

Now we turn to define the quality of a new given image. Like the pristine MVG model extraction, we partition the new image into patches and extract the quality-aware features for each patch, then fit the stacked feature vectors with MVG model leading to its $(\boldsymbol{\mu}_d, \mathbf{\Sigma}_d)$. Then the distance between the distorted MVG model and the pristine MVG model that measures the structure, naturalness and perception quality deviations is defined to measure the image quality as:

$$Q = \sqrt{(\boldsymbol{\mu}_d - \boldsymbol{\mu})^T \left(\frac{\mathbf{\Sigma}_d + \mathbf{\Sigma}}{2}\right)^{-1} (\boldsymbol{\mu}_d - \boldsymbol{\mu})} \quad (28)$$

Fig. 3. Sample pristine images used to learn the pristine MVG model. The images include common scenes in reality, which are people, animals, buildings, natural landscapes, etc.

TABLE II
OVERALL PREDICTION PERFORMANCE COMPARISON OF UNSUPERVISED IQA METHODS ON LIVE.

| Index | LPSI [25] | QAC [22] | NIQE [9] | IL-NIQE [21] | SNP-NIQE |
|---|---|---|---|---|---|
| SRCC | 0.8181 | 0.8683 | 0.9072 | 0.8978 | **0.9082** |
| KRCC | 0.6175 | 0.6736 | 0.7290 | 0.7128 | **0.7366** |
| PLCC | 0.8280 | 0.8625 | 0.9054 | 0.9025 | **0.9069** |
| RMSE | 15.3184 | 13.8258 | 11.6021 | 11.7702 | **11.5116** |

where $(\boldsymbol{\mu}, \mathbf{\Sigma})$ are the pristine MVG model parameters. As our method follows the quality definition framework of NIQE, we name it Structure, Naturalness and Perception quality-driven NIQE or SNP-NIQE. It is noted that the closer to zero the SNP-NIQE value is, the better the image quality is.

## III. EXPERIMENTAL RESULTS

### A. Experimental Protocol

To evaluate the prediction performance of the proposed approach, we conduct extensive experiments on six well-known image quality databases, namely, LIVE [45], TID2013 [35], CSIQ [46], Toyama [47], CID2013 [26] and the Waterloo Exploration database [48]. To be specific, LIVE, TID2013, CSIQ and Toyama are classical databases in IQA research, CID2013 is a real distortion image database, dedicated to evaluating BIQA methods. The Waterloo Exploration database is a large-scale database proposed to facilitate IQA research, which contains 4744 pristine images and 94880 distorted images. In our experiments, the most commonly encountered distortion types are involved for performance evaluation as in [49] [41] [50]. They are JPEG2000 compression (JP2K), JPEG compression (JPEG), white noise (WN), gaussian blur (GB) and a Rayleigh fast fading channel distortion (FF) in LIVE database, JP2K, JPEG, WN and GB in TID2013, CSIQ and the Exploration databases, JP2K and JPEG in Toyama database.

To quantify the prediction performance of the objective IQA models, firstly, we adopt four commonly-used indexes to calculate the correlation between the subjective scores and objective scores given by the objective IQA approaches, which are Spearman Rank order Correlation coefficient (SRCC), Kendall's rank correlation coefficient (KRCC), Pearsons linear

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2019.2900472, IEEE Transactions on Circuits and Systems for Video Technology



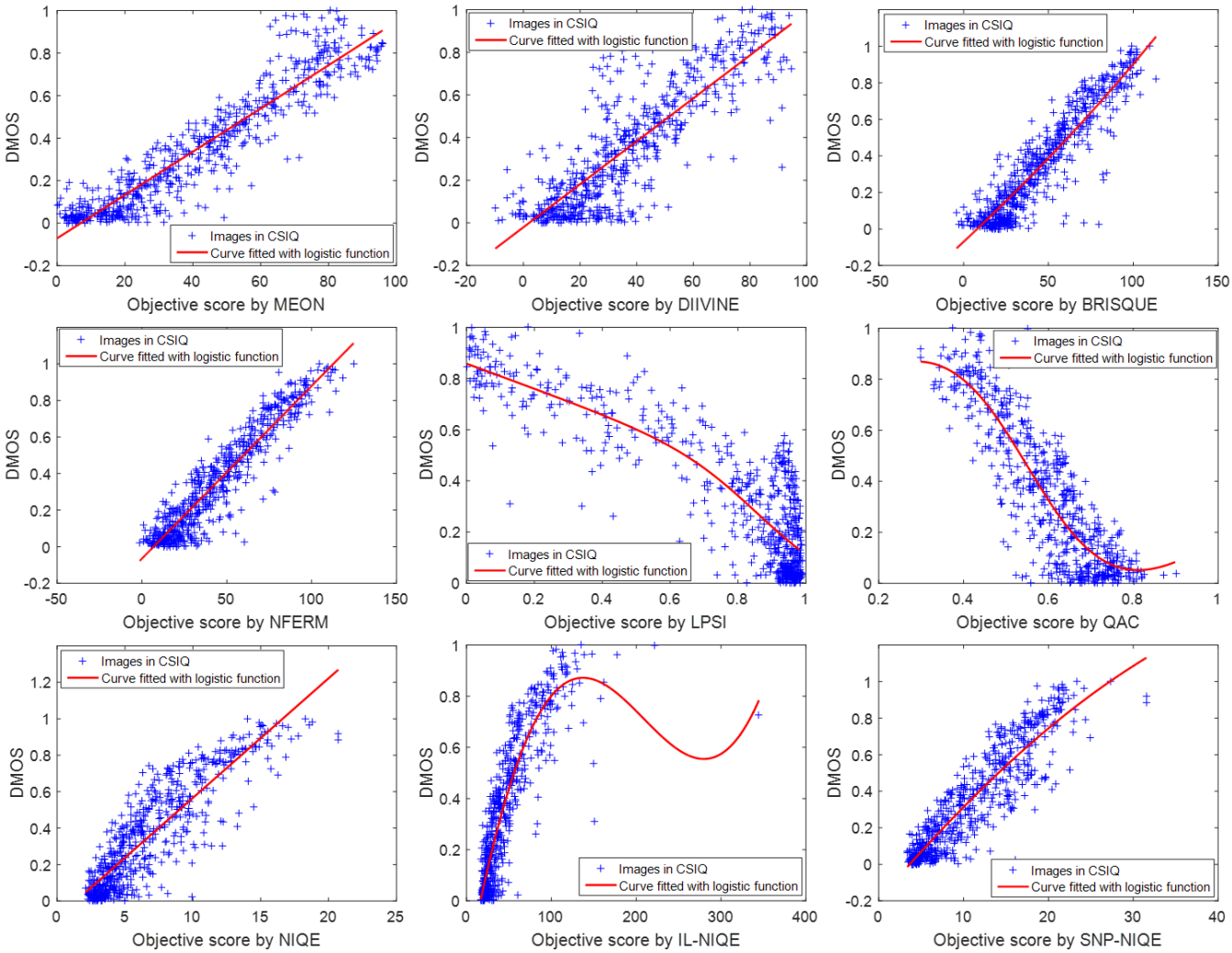



Fig. 4. Distribution diagrams of subjective DMOS values with respect to objective scores on the CSIQ database.

TABLE III

OVERALL PREDICTION PERFORMANCE COMPARISON ON TID2013, CSIQ AND TOYAMA DATABASES.

| Database | Index | MEON [20] | DIIVINE [8] | BRISQUE [7] | NFERM [41] | LPSI [25] | QAC [22] | NIQE [9] | IL-NIQE [21] | SNP-NIQE |
|---|---|---|---|---|---|---|---|---|---|---|
| **TID2013** | SRCC | **0.9012** | 0.7820 | 0.8412 | 0.8595 | 0.7046 | 0.8055 | 0.7956 | 0.8421 | **0.8565** |
| | KRCC | **0.7189** | 0.5898 | 0.6657 | 0.6958 | 0.5005 | 0.6164 | 0.5907 | 0.6536 | **0.6583** |
| | PLCC | **0.9051** | 0.7859 | 0.8677 | 0.8764 | 0.8114 | 0.8051 | 0.8066 | **0.8582** | 0.8470 |
| | RMSE | **0.5931** | 0.8626 | 0.6934 | 0.6718 | 0.8153 | 0.8273 | 0.8245 | **0.7161** | 0.7416 |
| **CSIQ** | SRCC | **0.9300** | 0.8284 | 0.9006 | 0.9142 | 0.7711 | 0.8415 | 0.8707 | 0.8801 | **0.9009** |
| | KRCC | **0.7650** | 0.6389 | 0.7360 | 0.7511 | 0.5826 | 0.6440 | 0.6848 | 0.6974 | **0.7210** |
| | PLCC | 0.9333 | 0.8454 | 0.9207 | **0.9364** | 0.8657 | 0.8736 | 0.8754 | 0.9054 | **0.9064** |
| | RMSE | 0.1015 | 0.1509 | 0.1103 | **0.0992** | 0.1415 | 0.1375 | 0.1366 | 0.1200 | **0.1194** |
| **Toyama** | SRCC | **0.8816** | 0.6416 | 0.8534 | 0.8498 | **0.8732** | 0.5189 | 0.8115 | 0.7114 | 0.8696 |
| | KRCC | **0.6955** | 0.4584 | 0.6639 | 0.6587 | **0.6885** | 0.3667 | 0.6112 | 0.5105 | 0.6811 |
| | PLCC | **0.8808** | 0.6372 | 0.8536 | 0.8517 | 0.8762 | 0.5388 | 0.8205 | 0.7247 | **0.8774** |
| | RMSE | **0.5925** | 0.9645 | 0.6519 | 0.6558 | 0.6031 | 1.0543 | 0.7155 | 0.8625 | **0.6003** |
| **AVG** | SRCC | **0.9122** | 0.7854 | 0.8709 | 0.8841 | 0.7584 | 0.7846 | 0.8332 | 0.8428 | **0.8792** |
| | KRCC | **0.7376** | 0.5956 | 0.6987 | 0.7171 | 0.5643 | 0.5964 | 0.6379 | 0.6554 | **0.6910** |
| | PLCC | **0.9152** | 0.7944 | 0.8909 | 0.9015 | 0.8457 | 0.8022 | 0.8410 | 0.8628 | **0.8791** |

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2019.2900472, IEEE Transactions on Circuits and Systems for Video Technology



correlation coefficient (PLCC) and root mean square error (RMSE) respectively on LIVE, TID2013, CSIQ and Toyama databases as subjective scores are available in these four databases. Generally, a better objective method is expected to achieve higher SRCC, KRCC and PLCC values, while lower RMSE value. Particularly, before calculating PLCC and RMSE, objective scores are suggested to be mapped to subjective scores through nonlinear regression [51]. Therefore, we apply a five-parameter logistic function to implement this, which is defined as:

$$q(z) = \beta_1 \left( \frac{1}{2} - \frac{1}{1 + \exp(\beta_2 \cdot (z - \beta_3))} \right) + \beta_4 \cdot z + \beta_5 \quad (29)$$

where $z$ and $q(z)$ are the objective score and the mapped score. $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5$ denote the parameters obtained through curve fitting.

Secondly, as the subjective scores are not available in the Exploration database, we employed Pristine/distorted image discriminability test (D-test), Listwise ranking consistency test (L-test) and Pairwise preference consistency test (P-test) introduced in [48] to compare the objective IQA methods' performance on the Exploration database. Note that higher values delivered by these three tests indicate better prediction performance.

### B. Implementation Settings

In Section II-C, we employ sparse representation to approximate the internal generative model for extracting the perception quality related features. The settings for sparse representation are set the same as in [5]. In detail, the patch size was set to 8×8, namely $B_s$ equals 64. The predefined dictionary $\mathbf{D}$ for sparse representation is instantiated with the over-complete DCT dictionary, the dimension of the dictionary is 64×144, which contains 144 atoms that can be used to represent each image patch. The over-complete DCT dictionary is created as follows: constructing a 8×12 1D-DCT $\mathbf{A}_{1D}$, in which the $k$-th atom ($k$=1,2,...,12), $a_k = \cos((i-1)(k-1)\pi/12)$, $i$=1,2,...8. All the atoms except the first one are removed by their mean value. Then $\mathbf{D}$ is generated by the Kronecker-product, namely, $\mathbf{D} = \mathbf{A}_{1D} \otimes \mathbf{A}_{1D}$ [52]. MATLAB source code of the proposed SNP-NIQE will be made publicly available at http://multimedia.sjtu.edu.cn/Data/List/Resources.

### C. Overall Prediction Performance Comparison

In this section, the overall prediction performance of the proposed SNP-NIQE with competing methods are reported. We compare SNP-NIQE with eight representative NR approaches which can be classified into two categories, namely, supervised and unsupervised. The supervised methods include MEON [20], DIIVINE [8], BRISQUE [7] and NFERM [41]. Among them, MEON is the state-of-the-art deep learning-based method and DIIVINE, BRISQUE and NFERM are mainstream traditional NR methods. The unsupervised methods include LPSI [25], QAC [22], NIQE [9] and IL-NIQE [21]. It should be clarified that the whole LIVE database is used for training for the supervised methods. Therefore, we didn't report the performance of the supervised methods on LIVE.

TABLE IV

THE D-TEST, L-TEST AND P-TEST RESULTS ON THE WATERLOO EXPLORATION DATABASE.

| Methods | D-test | L-test | P-test |
|---|---|---|---|
| MEON [20] | **0.9384** | 0.9669 | **0.9984** |
| DIIVINE [8] | 0.8538 | 0.8908 | 0.9540 |
| BRISQUE [7] | 0.9204 | **0.9772** | 0.9930 |
| NFERM [41] | 0.9068 | 0.9558 | 0.9767 |
| LPSI [25] | 0.9140 | 0.9471 | 0.9436 |
| QAC [22] | **0.9226** | 0.8699 | 0.9779 |
| NIQE [9] | 0.9109 | 0.9885 | **0.9937** |
| IL-NIQE [21] | 0.9084 | 0.9926 | 0.9927 |
| SNP-NIQE | 0.9153 | **0.9931** | 0.9936 |

TABLE V

STATISTICAL SIGNIFICANCE RESULTS (T-TEST). 1, 0, OR -1 IMPLIES SNP-NIQE IS STATISTICALLY SUPERIOR, COMPARATIVE, OR INFERIOR TO THE COMPETITOR IN EACH ROW WITH 95% CONFIDENCE.

| t-test | LIVE | TID2013 | CSIQ | Toyama |
|---|---|---|---|---|
| MEON [20] | - | -1 | -1 | 0 |
| DIIVINE [8] | - | 1 | 1 | 1 |
| BRISQUE [7] | - | -1 | -1 | 1 |
| NFERM [41] | - | -1 | -1 | 1 |
| LPSI [25] | 1 | 1 | 1 | 1 |
| QAC [22] | 1 | 1 | 1 | 1 |
| NIQE [9] | 0 | 1 | 1 | 1 |
| IL-NIQE [21] | 0 | 0 | 0 | 1 |

The overall performance results evaluated by SRCC, KRCC, PLCC and RMSE are tabulated in Table II and Table III. The experimental results on the Exploration database are listed in Table IV. Note that the best performance at each index is highlighted with boldface in Table II. In Table III and IV, we respectively bold the best results of the supervised and the unsupervised methods. "AVG" in Table III refers to the weighted average performance over the above three databases, the weights are assigned by the number of images in each image database as:

$$\bar{\xi} = \frac{\sum_i \xi_i \cdot \pi_i}{\sum_i \pi_i} \quad (30)$$

where $\bar{\xi}$ refers to the weighted average value, $\xi_i$ represents one of SRCC, KRCC and PLCC on the $i$th database, $\pi_i$ is the number of images in the $i$th database.

In Table II, it's clearly to see that SNP-NIQE achieves the best prediction performance among all the compared unsupervised methods. In Table III, for the supervised methods, it's observed that MEON achieves the best performance in most cases. For the unsupervised methods, the proposed SNP-NIQE is comparative with IL-NIQE on the TID2013 database

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2019.2900472, IEEE Transactions on Circuits and Systems for Video Technology



TABLE VII

SRCC VALUES OF THE IQA METRICS ON MOST COMMONLY ENCOUNTERED DISTORTION TYPES OF TID2013, CSIQ AND TOYAMA DATABASES.

| Database | Dis. Type | MEON [20] | DIIVINE [8] | BRISQUE [7] | NFERM [41] | LPSI [25] | QAC [22] | NIQE [9] | IL-NIQE [21] | SNP-NIQE |
|---|---|---|---|---|---|---|---|---|---|---|
| **TID2013** | AGN | **0.8797** | 0.8553 | 0.8523 | 0.8582 | 0.7690 | 0.7427 | 0.8194 | 0.8760 | **0.8856** |
| | GB | **0.8707** | 0.8344 | 0.8134 | 0.8498 | 0.8408 | 0.8464 | 0.7968 | 0.8145 | **0.8638** |
| | JPEG | **0.9104** | 0.6288 | 0.8521 | 0.8720 | **0.9123** | 0.8369 | 0.8430 | 0.8355 | 0.8791 |
| | JP2K | **0.9118** | 0.8534 | 0.8925 | 0.8097 | **0.8988** | 0.7895 | 0.8890 | 0.8581 | 0.8820 |
| | AVG | **0.8932** | 0.7930 | 0.8526 | 0.8474 | 0.8552 | 0.8039 | 0.8371 | 0.8460 | **0.8776** |
| **CSIQ** | JP2K | 0.8956 | 0.8304 | 0.8669 | **0.9048** | **0.9074** | 0.8697 | 0.9062 | 0.9059 | 0.9022 |
| | JPEG | **0.9461** | 0.7998 | 0.9092 | 0.9222 | **0.9501** | 0.9014 | 0.8832 | 0.8993 | 0.9318 |
| | GB | **0.9079** | 0.8716 | 0.9033 | 0.8964 | 0.9060 | 0.8362 | 0.8945 | 0.8576 | **0.9171** |
| | AWGN | **0.9475** | 0.8662 | 0.9253 | 0.9220 | 0.6661 | 0.8225 | 0.8097 | 0.8497 | **0.8749** |
| | AVG | **0.9243** | 0.8420 | 0.9012 | 0.9113 | 0.8574 | 0.8575 | 0.8734 | 0.8781 | **0.9065** |
| **Toyama** | JPEG | **0.8717** | 0.7023 | 0.8612 | 0.8642 | **0.9182** | 0.6714 | 0.8369 | 0.7091 | 0.8748 |
| | JP2K | **0.8969** | 0.6114 | 0.8713 | 0.8741 | 0.8438 | 0.5629 | **0.8762** | 0.7383 | 0.8701 |
| | AVG | **0.8843** | 0.6568 | 0.8662 | 0.8691 | **0.8810** | 0.6171 | 0.8566 | 0.7237 | 0.8725 |

TABLE VIII

SRCC VALUES OF THE IQA METRICS ON UNCOMMON DISTORTION TYPES OF TID2013 DATABASE AND CID2013 DATABASE.

| Database | Dis. Type | MEON [20] | DIIVINE [8] | BRISQUE [7] | NFERM [41] | LPSI [25] | QAC [22] | NIQE [9] | IL-NIQE [21] | SNP-NIQE |
|---|---|---|---|---|---|---|---|---|---|---|
| **TID2013** | ANC | **0.8073** | 0.7120 | 0.7090 | 0.7096 | 0.4955 | 0.7184 | 0.6699 | **0.8156** | 0.7330 |
| | SCN | **0.8853** | 0.4626 | 0.4908 | 0.2180 | 0.6968 | 0.1694 | 0.6660 | **0.9231** | 0.6495 |
| | MN | **0.6860** | 0.6752 | 0.5748 | 0.2207 | 0.0462 | 0.5927 | **0.7464** | 0.5121 | 0.7400 |
| | HFN | 0.8757 | 0.8778 | 0.7528 | **0.8814** | **0.9250** | 0.8628 | 0.8449 | 0.8683 | 0.8730 |
| | IN | **0.8697** | 0.8063 | 0.6299 | 0.1728 | 0.4324 | **0.8003** | 0.7433 | 0.7554 | 0.7997 |
| | QN | 0.0875 | 0.1650 | **0.7984** | 0.7747 | 0.8537 | 0.7089 | 0.8542 | **0.8726** | 0.8573 |
| | DEN | 0.6624 | **0.7231** | 0.5864 | 0.6389 | 0.2487 | 0.3381 | 0.5903 | **0.7491** | 0.6128 |
| | JGTE | **0.7984** | 0.2387 | 0.3150 | 0.1322 | 0.0911 | 0.0491 | 0.0028 | **0.2821** | 0.2817 |
| | J2TE | **0.4143** | 0.0606 | 0.3594 | 0.1681 | 0.6106 | 0.4065 | 0.5102 | 0.5243 | **0.5917** |
| | NEPN | 0.0054 | 0.0598 | **0.1453** | 0.0645 | 0.0520 | 0.0477 | 0.0692 | **0.0803** | 0.0149 |
| | Block | **0.2302** | 0.0928 | 0.2235 | 0.2023 | 0.1372 | **0.2474** | 0.1222 | 0.1355 | 0.0321 |
| | MS | **0.2102** | 0.0104 | 0.1241 | 0.0218 | **0.3409** | 0.3059 | 0.1614 | 0.1845 | 0.0999 |
| | CTC | 0.0984 | **0.4601** | 0.0403 | 0.2185 | 0.1992 | **0.2067** | 0.0178 | 0.0133 | 0.1562 |
| | CCS | 0.2455 | 0.0684 | 0.1093 | **0.3062** | 0.3018 | **0.3683** | 0.2425 | 0.1642 | 0.1060 |
| | MGN | **0.8132** | 0.7873 | 0.7242 | 0.7164 | 0.6959 | **0.7902** | 0.6940 | 0.6924 | 0.7401 |
| | CN | 0.0883 | 0.1156 | 0.0081 | **0.1433** | 0.0181 | 0.1521 | 0.1545 | **0.3600** | 0.2083 |
| | LCNI | **0.8266** | 0.6327 | 0.6852 | 0.6541 | 0.2356 | 0.6395 | 0.8014 | 0.8287 | **0.8300** |
| | ICQD | 0.1471 | 0.4362 | **0.7640** | 0.4790 | **0.8998** | 0.8731 | 0.7870 | 0.7486 | 0.7900 |
| | CHA | 0.6547 | **0.6608** | 0.6160 | 0.6423 | **0.6953** | 0.6249 | 0.5619 | 0.6788 | 0.6347 |
| | SSR | 0.8247 | **0.8334** | 0.7841 | 0.7850 | 0.8620 | 0.7856 | 0.8341 | **0.8650** | 0.8287 |
| | AVG | **0.5115** | 0.4439 | 0.4720 | 0.4075 | 0.4419 | 0.4844 | 0.5037 | **0.5527** | 0.5290 |
| **CID2013** | Undefined | 0.3786 | 0.4633 | 0.4419 | **0.6205** | 0.3230 | 0.0299 | 0.6539 | 0.3063 | **0.7157** |

and LPSI on the Toyama database. However, SNP-NIQE achieves the best prediction performance on CSIQ database. The average results show that SNP-NIQE still outperforms all the other unsupervised methods notably and competes with supervised BRISQUE. In Table IV, it is observed that SNP-NIQE achieves the best result in L-test on the Exploration database, which indicates the superior monotonicity of SNP-NIQE in predicting the image quality. In a word, The experimental results in Table II, III and IV verify the effectiveness and superiority of SNP-NIQE for image quality evaluation in a more comprehensive manner.

To inspect the statistical significance of the obtained results, we applied t-test on the prediction residuals of the objective methods, which are calculated as the differences between the subjective scores and the converted objective scores by Eq. (29), conforming to Gaussian distribution. The experimental results are reported in Table V, where "1", "0" and "-1" tell that the proposed SNP-NIQE is superior, comparative or inferior to the competing method in each row statistically with 95% confidence. From this table, we can see our method SNP-NIQE is statistically better than or comparative with all the unsupervised methods as the values in the lower part are

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2019.2900472, IEEE Transactions on Circuits and Systems for Video Technology



TABLE VI
PREDICTION PERFORMANCE MEASURED BY SRCC OF THE IQA METRICS ON EACH DISTORTION TYPE OF LIVE.

| Dis. Type | LPSI [25] | QAC [22] | NIQE [9] | IL-NIQE [21] | SNP-NIQE |
|---|---|---|---|---|---|
| FF | 0.7808 | 0.8231 | **0.8635** | 0.8328 | 0.8495 |
| GB | 0.9156 | 0.9134 | 0.9329 | 0.9158 | **0.9510** |
| JP2K | **0.9300** | 0.8621 | 0.9185 | 0.8942 | 0.9174 |
| JPEG | 0.9677 | 0.9362 | 0.9409 | 0.9419 | **0.9691** |
| AWGN | 0.9557 | 0.9509 | 0.9718 | **0.9807** | 0.9781 |
| AVG | 0.9099 | 0.8971 | 0.9255 | 0.9131 | **0.9330** |

TABLE IX
SRCC VALUES COMPARISON OF NIQE, IL-NIQE AND SNP-NIQE.

| Database | NIQE [9] | IL-NIQE [21] | SNP-NIQE |
|---|---|---|---|
| **LIVE** | 0.9072 | 0.8978 | **0.9082** |
| **TID2013** | 0.7956 | 0.8421 | **0.8565** |
| **CSIQ** | 0.8707 | 0.8801 | **0.9009** |
| **Toyama** | 0.8115 | 0.7114 | **0.8696** |
| **CID2013** | 0.6539 | 0.3063 | **0.7157** |

all 1 or 0, which demonstrate the superiority of SNP-NIQE statistically.

For visualization, we also provide the distribution diagrams of the subjective DMOS values with respect to objective values on the CSIQ database in Fig. 4, in which we denote the distorted images with blue "+" and the red curves are obtained in the curve fitting process. It can be observed that the blue "+" of SNP-NIQE gather much closer to the fitted curve than the competitors, like NIQE, LPSI, etc., which vividly shows the scores of SNP-NIQE correlate well with subjective DMOS values.

### *D. Performance Comparison on Individual Distortion Types*

In this section, we would like to examine the predicting ability of the IQA metrics on the individual distortion types. Here, we only adopt SRCC as the performance measure. By using KRCC, PLCC and RMSE, same conclusions can be drawn. Similarly, we only use SRCC as the performance measure in the following discussions[1]. The experimental results measured by SRCC are summarized in Table VI and Table VII respectively. Table VI lists the results on LIVE and Table VII lists the results on the other three databases. "AVG" is the direct average of the SRCC values in each database. Likewise, we bold the best performance of each row in Table VI and bold the best performance of the supervised and the unsupervised methods separately in Table VII. In Table VI, it can be observed that SNP-NIQE performs best on GB and JPEG. The "AVG" of SNP-NIQE is also the highest. In fact, we can find that the unmarked values of SNP-NIQE are all in the second or third place, which proves the advantageous ability of SNP-NIQE in evaluating the images with specific distortions. In Table VII, among the unsupervised methods, the "AVG" values indicates SNP-NIQE performs best on TID2013 and CSIQ and gains the second place on Toyama database, which manifests SNP-NIQE is also quite effective in dealing with specific distortions. Obviously, SNP-NIQE outperforms NIQE and IL-NIQE significantly as NIQE is only marked on JP2K in Toyama while IL-NIQE can't be marked. To the supervised methods, MEON is still the best method followed by NFERM. Although our method can't outperform MEON, we can find it can outperform all the other supervised BIQA methods.

[1]Readers can refer to https://pan.baidu.com/s/1UbtJKuXh18sfV2W23XzxgA for further checking the KRCC, PLCC and RMSE results.

TABLE X
FEATURE VECTOR DIMENSION AND RUNNING TIME COMPARISONS.

| Index | NIQE [9] | IL-NIQE [21] | SNP-NIQE |
|---|---|---|---|
| Feature Vector Dimension | 36 | 430 | 52 |
| Running Time (s) | 0.224 | 4.762 | 3.685 |

### *E. Generalization Capability Test*

Generalization capability is an important factor for evaluating the general-purpose NR methods. Therefore, we testify the generalization capability of all the blind models with the uncommon distortion types in TID2013 as in [41] [50]. In addition, we employ CID2013 database for this testing, which is a real distortion image database, including 474 naturally-distorted images captured by 79 different cameras or image signal processing pipelines. The experimental results are listed in Table VIII, where SRCC is utilized as the performance indicator. "Undefined" in CID2013 means the distortion type is not explicitly defined. We bold the best results of the supervised methods and unsupervised methods separately.

By analyzing Table VIII, we can have the following findings. First, comparing the best average results of the supervised methods and unsupervised methods on these two databases, we can observe the performance of unsupervised method is higher than that of supervised method, which shows better generalization capability for the unsupervised method for quality evaluation. Second, the proposed SNP-NIQE achieves the second best result at the average value on TID2013 database and the best result on CID2013 database, which reveals good generalization capability of SNP-NIQE for uncommon and real distortions.

### *F. Detailed Comparison with NIQE and IL-NIQE*

As NIQE, IL-NIQE and SNP-NIQE work under the same framework, we want to compare them in more detail. Here, we compare them from three aspects, which are the prediction accuracy, the dimension of the quality-aware feature vector and the running time. The prediction accuracy measured by SRCC are summarized in Table IX, where we bold the best performed method on each database. The feature vector dimension and the running time comparisons are listed in Table X. The running time is measured in seconds and recorded as follows.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2019.2900472, IEEE Transactions on Circuits and Systems for Video Technology



TABLE XI
PERFORMANCE CONTRIBUTION OF EACH TYPE OF FEATURES AND THEIR COMBINATIONS IN SNP-NIQE.

| Database | PC&Gradients | MSCN | R | PC&Gradients+MSCN | PC&Gradients+R | MSCN+R | All |
|---|---|---|---|---|---|---|---|
| **LIVE** | 0.7528 | 0.9046 | 0.7191 | 0.9051 | 0.8019 | 0.9048 | **0.9082** |
| **TID2013** | 0.6627 | 0.8069 | 0.5121 | 0.8414 | 0.6773 | 0.8220 | **0.8565** |
| **CSIQ** | 0.6689 | 0.8823 | 0.6905 | 0.8877 | 0.7571 | 0.8936 | **0.9009** |
| **Toyama** | 0.5841 | 0.8217 | 0.5896 | 0.8518 | 0.7071 | 0.8520 | **0.8696** |
| **CID2013** | 0.6250 | 0.6643 | 0.6330 | 0.6853 | 0.6526 | 0.6942 | **0.7157** |

We ran NIQE, IL-NIQE and SNP-NIQE respectively on the entire TID2013 database (3000 images of size $512 \times 384$) and the average time was calculated for comparison. Note that the hardware platform is Thinkpad X220 computer with 2.5GHz CPU and 4G RAM. The software platform is Matlab R2012a. In Table IX, compared with NIQE, IL-NIQE performs better on TID2013 and CSIQ databases, while on LIVE, Toyama and CID2013, IL-NIQE can't give better results, especially on CID2013 database. By contrast, SNP-NIQE exceeds NIQE on all databases and it can outperform NIQE by a large margin, which fully demonstrates that the features of SNP-NIQE are much more effective than that of NIQE and IL-NIQE to capture the image quality degradation. From Table X, we can see although the feature vector dimensions of IL-NIQE and SNP-NIQE are both higher than NIQE, the dimension of SNP-NIQE is 52, which is much less than 430 of IL-NIQE. In addition, the running time of SNP-NIQE is also less than IL-NIQE. Therefore, we can conclude that SNP-NIQE is much more promising than IL-NIQE.

### *G. Contribution of The Quality-aware Features to The Prediction Performance*

In the proposed SNP-NIQE, we evaluate the image quality from three aspects, which are structure, naturalness and the perception quality, thus we extracted the gradients and PC, MSCN and prediction residual statistical features to characterize the structure, naturalness and the perception quality, respectively. Hence, it's interesting to understand the individual contribution of each type of features and their combinations to the final prediction performance. Toward this end, we conducted experiments to check the prediction performance of each type of features and their combinations. The experimental results in terms of SRCC are listed in Table XI, where "R" refers to the prediction residual on behalf of the perception quality. "+" refers to combining the two types of features. "All" refers to that all types of features are included for quality evaluation. The best performance in each row is stressed with boldface.

From this table, we can draw some important conclusions. First, the structure features alone lead to moderate prediction performance. It is observed that the performance of the structure features is relatively low on the Toyama database. This is because half of this image database are JPEG compressed images, while this distortion type that will introduce extra structures can't be precisely described only by the structure features. Second, the MSCN features that characterize the naturalness earn the highest performance across all the databases among the three types of features. In addition, the performance of "PC&Gradients+R" without MSCN features is the lowest among the combinations of two types of features on each database. Therefore, we can conclude that naturalness measurement plays the leading role in image quality evaluation. Third, the performance of the R features on behalf of the perception quality is very encouraging which can be comparative with the structure features, given that only two features are extracted. Such experimental results also confirm the potential of exploring the mechanism of the human visual perception system for image quality evaluation. We envision that with deeper exploration of the human visual system, the prediction accuracy of perception quality measurement can be further improved so that our proposed method can be further perfected. Fourth, the combination of two types of features can achieve better prediction performance than any single type of features therein and the combination of all the three types of features achieves the best performance, which indicates the features we extracted that measure the structure, naturalness and the perception quality can work cooperatively for image quality evaluation.

### *H. Robustness to Different Pristine Image Datasets and Image Numbers*

To construct SNP-NIQE, we learn a reference MVG model from a set of pristine images, compared with which to define the quality of a new image. In this subsection, we will verify that it's robust to learn the pristine model from different pristine image sets and with different image numbers. Except for training the pristine model from the Berkeley segmentation dataset, we also selected two other datasets for obtaining the pristine model. The first one is from the Waterloo Exploration Database. We selected the first 100 pristine images from the Exploration database and got a pristine image dataset. The other one is General100 dataset [53], which contains 100 images of good quality proposed for super resolution training. In each dataset, we selected the first 10, 20,...,100 images respectively to train a series of MVG models for quality determination. The experiments were conducted on LIVE database and the prediction performance of these pristine models are shown in Fig. 5, where we employ SRCC as the performance measure. It's observed that the prediction performance of these three datasets changes slightly as the image number varies. In addition, there is no much difference among the three datasets.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2019.2900472, IEEE Transactions on Circuits and Systems for Video Technology



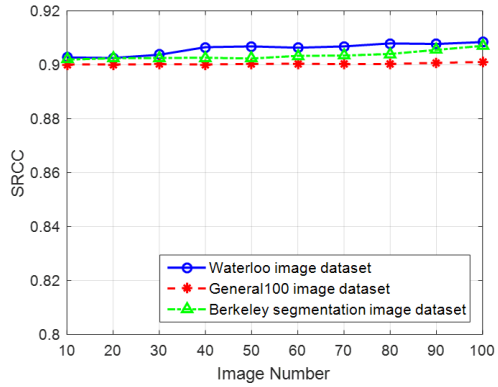



Fig. 5. Prediction performance variation on LIVE of pristine MVG models learned from different image datasets with different image numbers.

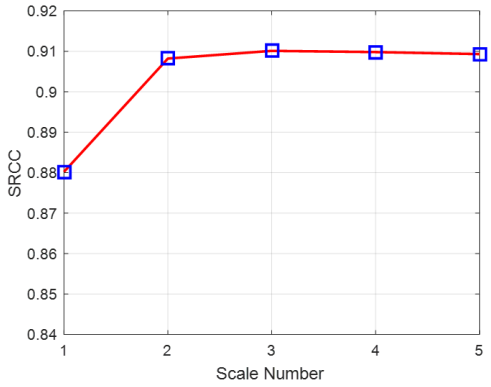



Fig. 6. Prediction performance on LIVE with regard to different scale numbers.

Therefore, we verify that different image datasets and image numbers for training the pristine model have little influence on the final prediction performance.

### *I. Evaluation of Multi-scale Strategy in SNP-NIQE*

In SNP-NIQE, multi-scale strategy is adopted for extracting the quality-aware features. Specifically, we carried out feature extraction in two scales, which are the original scale and the down-sampled scale by the factor of 2. In this subsection, we care about whether more scales can boost the performance of the proposed method. Therefore, we down-sampled the original image by 2 iteratively until we got 5 scales, when the down-sampling factor attains 16. The experimental results of different scale numbers tested on LIVE is shown in Fig. 6. It's clearly seen that when the scale number reaches 2, the performance increases notably. However, when the scale number is greater than 2, the prediction performance changes slightly. Therefore, by comprehensive consideration of the prediction performance and the computational complexity, we set the default scale number to 2 in SNP-NIQE.

### *J. Application Methodology of SNP-NIQE for Video Quality Evaluation*

As SNP-NIQE can effectively evaluate the image quality, it can be applied to video quality evaluation. Different from image quality evaluation, video quality evaluation should not only characterize the spatial distortions of each frame, but also measure the temporal distortions, such as jitter, motion inconsistency, etc. caused by motion prediction error. For measuring the spatial distortions, existing IQA techniques can be applied directly as each frame can be considered as an independent image. For measuring the temporal distortions, different strategies have been proposed. On one hand, dedicated features derived from the successive frame difference or motion cues were exploited to effectively characterize the temporal distortions [54] [55]. On the other hand, rather than measuring the temporal distortions explicitly, some works made efforts to investigate different perceptual pooling strategies to integrate the spatial features of each frame together to evaluate the video quality [56] [57]. As the proposed SNP-NIQE is proved to be an effective indicator to the image quality, it can be used to measure the spatial distortions of each frame and then integrated with temporal distortion measures to evaluate the entire video quality. These works will be investigated in our future work.

## IV. Conclusion

In this paper, we have concentrated our efforts on the BIQA research and constructed a novel unsupervised model named SNP-NIQE, in which we treated the image quality evaluation problem through measuring the structure, naturalness and the perception quality variations due to the distortions. Thus, three types of effective NSS features were designed and extracted to characterize the structure, naturalness and the perception quality respectively. After feature extraction, we learned a pristine MVG model with the quality-aware features from a set of pristine images which serves as the "reference" for quality prediction. The distance between the MVG model of the question image and the learned pristine MVG model is defined to measure the question image quality. Extensive experiments conducted on six popular image databases demonstrate that the proposed SNP-NIQE achieves comparative prediction performance with state-of-the-art NR IQA methods.

## References

[1] 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, 2004.

[2] 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, 2011.

[3] K. Gu, L. Li, H. Lu, X. Min, and W. Lin, "A fast reliable image quality predictor by fusing micro-and macro-structures," *IEEE Trans. Ind. Electron.*, vol. 64, no. 5, pp. 3903–3912, 2017.

[4] 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, 2012.

[5] Y. Liu, G. Zhai, K. Gu, X. Liu, D. Zhao, and W. Gao, "Reduced-reference image quality assessment in free-energy principle and sparse representation," *IEEE Trans. Multimedia*, vol. 20, no. 2, pp. 379–391, 2018.

[6] S. Wang, K. Gu, X. Zhang, W. Lin, S. Ma, and W. Gao, "Reduced-reference quality assessment of screen content images," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 1, pp. 1–14, 2018.

[7] 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, 2012.

[8] 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, 2011.

[9] 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, 2013.

[10] Q. Wu, H. Li, K. N. Ngan, and K. Ma, "Blind image quality assessment using local consistency aware retriever and uncertainty aware evaluator," *IEEE Trans. Circuits Syst. Video Technol.*, pp. 1–1, 2017.

[11] Y. Liu, K. Gu, G. Zhai, X. Liu, D. Zhao, and W. Gao, "Quality assessment for real out-of-focus blurred images," *J. Vis. Commun. Image Represent.*, vol. 46, pp. 70–80, 2017.

[12] K. Gu, G. Zhai, W. Lin, X. Yang, and W. Zhang, "No-reference image sharpness assessment in autoregressive parameter space," *IEEE Trans. Image Process.*, vol. 24, no. 10, pp. 3218–3231, 2015.

[13] D. Zoran and Y. Weiss, "Scale invariance and noise in natural images," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2009, pp. 2209–2216.

[14] 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.*, 2002, pp. 477–480.

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

[16] Q. Wu, H. Li, F. Meng, K. N. Ngan, B. Luo, C. Huang, and B. Zeng, "Blind image quality assessment based on multichannel feature fusion and label transfer," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 3, pp. 425–440, 2016.

[17] Y. Liu, K. Gu, S. Wang, D. Zhao, and W. Gao, "Blind quality assessment of camera images based on low-level and high-level statistical features," *IEEE Trans. Multimedia*, 2018.

[18] J. Kim and S. Lee, "Fully deep blind image quality predictor," *IEEE J. Sel. Topics Signal Process*, vol. 11, no. 1, pp. 206–220, 2017.

[19] J. Guan, S. Yi, X. Zeng, W. K. Cham, and X. Wang, "Visual importance and distortion guided deep image quality assessment framework," *IEEE Trans. Multimedia*, vol. 19, no. 11, pp. 2505–2520, 2017.

[20] K. Ma, W. Liu, K. Zhang, Z. Duanmu, Z. Wang, and W. Zuo, "End-to-end blind image quality assessment using deep neural networks," *IEEE Trans. Image Process.*, vol. 27, no. 3, pp. 1202–1213, 2018.

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

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

[23] K. Gu, J. Zhou, J. Qiao, G. Zhai, W. Lin, and A. Bovik, "No-reference quality assessment of screen content pictures," *IEEE Trans. Image Process.*, 2017.

[24] K. Gu, D. Tao, J.-F. Qiao, and W. Lin, "Learning a no-reference quality assessment model of enhanced images with big data," *IEEE Trans. Neural Netw. Learn. Syst.*, 2017.

[25] Q. Wu, Z. Wang, and H. Li, "A highly efficient method for blind image quality assessment," in *IEEE Int. Conf. Image Process.*, 2015, pp. 339–343.

[26] T. Virtanen, M. Nuutinen, M. Vaahteranoksa, P. Oittinen, and J. Häkkinen, "Cid2013: a database for evaluating no-reference image quality assessment algorithms," *IEEE Trans. Image Process.*, vol. 24, no. 1, pp. 390–402, 2015.

[27] K. Friston, J. Kilner, and L. Harrison, "A free energy principle for the brain," *J. Physiol. Paris*, vol. 100, no. 1, pp. 70–87, 2006.

[28] K. Friston, "The free-energy principle: a unified brain theory?" *Nature Rev. Neurosci.*, vol. 11, no. 2, pp. 127–138, 2010.

[29] B. A. Olshausen and D. J. Field, "Sparse coding with an overcomplete basis set: A strategy employed by v1?" *Vision research*, vol. 37, no. 23, pp. 3311–3325, 1997.

[30] B. A. Olshausen *et al.*, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, vol. 381, no. 6583, pp. 607–609, 1996.

[31] M. C. Morrone and D. Burr, "Feature detection in human vision: A phase-dependent energy model," *Proceedings of the Royal Society of London B: Biological Sciences*, vol. 235, no. 1280, pp. 221–245, 1988.

[32] M. C. Morrone, J. Ross, D. C. Burr, and R. Owens, "Mach bands are phase dependent," *Nature*, vol. 324, no. 6094, pp. 250–253, 1986.

[33] M. C. Morrone and R. A. Owens, "Feature detection from local energy," *Pattern recognition letters*, vol. 6, no. 5, pp. 303–313, 1987.

[34] P. Kovesi, "Image features from phase congruency," *Videre: Journal of computer vision research*, vol. 1, no. 3, pp. 1–26, 1999.

[35] N. Ponomarenko, O. Ieremeiev, V. Lukin, K. Egiazarian, L. Jin, J. Astola, B. Vozel, K. Chehdi, M. Carli, F. Battisti *et al.*, "Color image database tid2013: Peculiarities and preliminary results," in *Proc. 4th Eur. Workshop Vis. Inf. Process.*, 2013, pp. 106–111.

[36] J. Zhang, D. Zhao, R. Xiong, S. Ma, and W. Gao, "Image restoration using joint statistical modeling in a space-transform domain," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 6, pp. 915–928, June 2014.

[37] 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, 1995.

[38] K. Friston, J. Kilner, and L. Harrison, "A free energy principle for the brain," *Journal of Physiology-Paris*, vol. 100, no. 1, pp. 70–87, 2006.

[39] K. Friston, "The free-energy principle: a unified brain theory?" *Nature Reviews Neuroscience*, vol. 11, no. 2, pp. 127–138, 2010.

[40] M. Moore, "Statistical mechanics: A set of lectures," 1974.

[41] K. Gu, G. Zhai, X. Yang, and W. Zhang, "Using free energy principle for blind image quality assessment," *IEEE Trans. Multimedia*, vol. 17, no. 1, pp. 50–63, 2015.

[42] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, 2007.

[43] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Proc. IEEE Asilomar Conf. Signals, Syst., Comput.*, vol. 2, 2003, pp. 1398–1402.

[44] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *IEEE Int. Conf. Comput. Vision*, vol. 2, 2001, pp. 416–423.

[45] L. C. H. R. Sheikh, Z. Wang and A. C. Bovik, "Live image quality assessment database release 2," 2006.

[46] E. C. Larson and D. Chandler, "Categorical image quality (csiq) database," *Online, http://vision. okstate. edu/csiq*, 2010.

[47] Y. Horita, K. Shibata, Y. Kawayoke, and Z. P. Sazzad, "Mict image quality evaluation database," *[Online], http://mict. eng. u-toyama. ac. jp/mictdb. html*, 2011.

[48] K. Ma, Z. Duanmu, Q. Wu, Z. Wang, H. Yong, H. Li, and L. Zhang, "Waterloo exploration database: New challenges for image quality assessment models," *IEEE Trans. Image Process.*, vol. 26, no. 2, pp. 1004–1016, Feb 2017.

[49] W. Xue, X. Mou, L. Zhang, A. C. Bovik, and X. Feng, "Blind image quality assessment using joint statistics of gradient magnitude and laplacian features," *IEEE Trans. Image Process.*, vol. 23, no. 11, pp. 4850–4862, 2014.

[50] Q. Li, W. Lin, J. Xu, and Y. Fang, "Blind image quality assessment using statistical structural and luminance features," *IEEE Trans. Multimedia*, vol. 18, no. 12, pp. 2457–2469, 2016.

[51] A. M. Rohaly, J. Libert, P. Corriveau, A. Webster *et al.*, "Final report from the video quality experts group on the validation of objective models of video quality assessment," *ITU-T Standards Contribution COM*, pp. 9–80, 2000.

[52] M. Elad, *Sparse and Redundant Representations*. Springer, 2010.

[53] C. Dong, C. C. Loy, and X. Tang, "Accelerating the super-resolution convolutional neural network," *in Proceedings of European Conference on Computer Vision (ECCV)*, 2016.

[54] A. Mittal, M. A. Saad, and A. C. Bovik, "A completely blind video integrity oracle," *IEEE Trans. Image Process.*, vol. 25, no. 1, pp. 289–300, Jan 2016.

[55] L. Xu, W. Lin, L. Ma, Y. Zhang, Y. Fang, K. N. Ngan, S. Li, and Y. Yan, "Free-energy principle inspired video quality metric and its use in video coding," *IEEE Trans. Multimedia*, vol. 18, no. 4, pp. 590–602, April 2016.

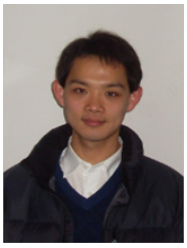
[56] Z. Chen, N. Liao, X. Gu, F. Wu, and G. Shi, "Hybrid distortion ranking tuned bitstream-layer video quality assessment," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 6, pp. 1029–1043, June 2016.

[57] J. You, J. Korhonen, A. Perkis, and T. Ebrahimi, "Balancing attended and global stimuli in perceived video quality assessment," *IEEE Trans. Multimedia*, vol. 13, no. 6, pp. 1269–1285, Dec 2011.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2019.2900472, IEEE Transactions on Circuits and Systems for Video Technology



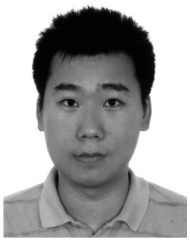
**Yutao Liu** received the B.S., M.S. and Ph.D. degrees in computer science from Harbin Institute of Technology(HIT), Harbin, China, in 2011, 2013 and 2018, respectively. From 2014 to 2016, he was with the Institute of Image Communication and Information Processing, Shanghai Jiao Tong University, Shanghai, as a research assistant. He is currently a Post-Doctoral Fellow with the Graduate School at Shenzhen, Tsinghua University. His research interests include image quality assessment and perceptual image processing.



**Ke Gu** received the B.S. and Ph.D. degrees in electronic engineering from Shanghai Jiao Tong University, Shanghai, China, in 2009 and 2015, respectively. He is currently a Professor with the Beijing University of Technology, Beijing, China. His research interests include environmental perception, image processing, quality assessment, and machine learning. He received the Best Paper Award from the IEEE Transactions on Multimedia (T-MM), the Best Student Paper Award at the IEEE International Conference on Multimedia and Expo (ICME) in 2016, and the Excellent Ph.D. Thesis Award from the Chinese Institute of Electronics in 2016. He was the Leading Special Session Organizer in the VCIP 2016 and the ICIP 2017, and serves as a Guest Editor for the Digital Signal Processing Journal. He is currently an Associate Editor for the IEEE ACCESS and IET Image Processing (IET-IP), and an Area Editor for the Signal Processing Image Communication (SPIC). He is a Reviewer for 20 top SCI journals.

**Yongbing Zhang** received the B.A. degree in English and the M.S. and Ph.D degrees in computer science from the Harbin Institute of Technology, Harbin, China, in 2004, 2006, and 2010, respectively. He joined Graduate School at Shenzhen, Tsinghua University, Shenzhen, China in 2010, where he is currently an associate professor. He was the receipt of the Best Student Paper Award at IEEE International Conference on Visual Communication and Image Processing in 2015. His current research interests include signal processing, image/video coding, and machine learning.

**Xiu Li** (M'15) received her Ph.D.degree in computer integrated manufacturing in 2000. Since then, she has been worked in Tsinghua University. Her research interests are in the areas of data mining, deep learning, computer vision and image processing.



**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, Germany. He received 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.

**Debin Zhao** (M'11) received the B.S., M.S., and Ph.D. degrees in computer science from the Harbin Institute of Technology, China in 1985, 1988, and 1998, respectively. He is currently a Professor with the Department of Computer Science, Harbin Institute of Technology. He has published over 200 technical articles in refereed journals and conference proceedings in the areas of image and video coding, video processing, video streaming and transmission, and pattern recognition.

**Wen Gao** (M'92-SM'05-F'09) received the Ph.D. degree in electronics engineering from the University of Tokyo, Tokyo, Japan, in 1991. He is currently a Professor of Computer Science with Peking University, Beijing, China. Before joining Peking University, he was a Professor of Computer Science with the Harbin Institute of Technology, Harbin, China, from 1991 to 1995, and a Professor with the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China. He has authored five books and more than 600 technical articles in refereed journals and conference proceedings in image processing, video coding and communication, pattern recognition, multimedia information retrieval, multimodal interface, and bioinformatics.

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